Computational Biochemical Study

of the Prebiotic Selection of Nucleic Acids

By
Lázaro Andrés Monteserín Castanedo

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Applied Sciences.

January, 2024, Halifax, Nova Scotia

Approved: Dr. Chérif F. Matta
           Supervisor

Approved: Dr. Stacey D. Wetmore
           Examiner

Approved: Dr. Genlou Sun
           Committee member

Approved: Dr. Cory Pye
           Committee member

Date: January 31, 2024

Computational Biochemical Study
of the Prebiotic Selection of Nucleic Acids

by

Lázaro Andrés Monteserín Castanedo

Abstract

This thesis addresses fundamental questions related to the prebiotic evolutionary selection of the building blocks of nucleic acids. The structural tendencies and propensities in today's nucleic acids are rationalized based on thermodynamics as a principal driver of evolutionary selection. The free energies of the possible reaction paths available to prebiotic Nature are calculated from quantum chemistry. As one example (of many), the β-anomers of the nucleosides(tides) - predominant in modern nucleic acids - are found to be slightly more stable than their α-counterparts. This small thermodynamic advantage operating over millennia may have contributed to the observed dominance of today's canonical forms. Calculations also suggest the possibility that non-canonical N-(2-aminoethyl)glycine (AEG) and glycerol nucleosides(tides) may have assisted in the synthesis of today's nucleosides(tides) if the prebiotic environment has been aqueous. Energetic comparisons of ancestral nucleic acids containing arsenate instead of phosphate indicate no thermodynamic advantage for the phosphate, raising an important open question as to the reason for Nature's selection of the latter. It is also found, computationally, that barbituric acid may have well been a prebiotic precursor of today's nucleobases reinforcing earlier proposals. A more fundamental question may be about the choice of nucleic acids as the carriers of genetic information, in the first place, instead of other contenders such as proteins. A partial answer is formulated by proposing a quantitative account of the "*value*" of information as a new dimension to be added to the traditional "*amount*" (bits) in Shannon's information theory. Thus, the thesis addresses certain aspects of evolutionary biochemistry from the standpoint of thermodynamics under differing conditions of solvation. Meanwhile, the rates (kinetics) were not considered in this work since the synthetic and mechanistic steps from reactants to products of most of the proposed reactions remain largely unknown. Several other potential factors have not been considered but, with these variables being constant, our results remain valid and so are the questions they open for future investigations.

January 31, 2024

Computational Biochemical Study
of the Prebiotic Selection of Nucleic Acids

by

Lázaro Andrés Monteserín Castanedo

Long Summary of the Thesis

The present thesis addresses a set of fundamental questions related to the prebiotic origins of the building blocks of nucleic acids. The author of this thesis seeks to shed light on the following type of questions: can a plausible molecular evolution pathway be proposed to explain the emergence of nucleosides and nucleotides as they exist in today's nucleic acids and a proto-RNA based solely on thermodynamic (free energies) considerations? Is the evolutionary selection of, primarily, the β-anomers of the nucleosides(tides) in preference to the α-counterparts consistent with the differences in their energetic stabilities? Why has thymine been chosen to be predominantly incorporated in today's DNA while uracil is typically found in RNA? And, at a more fundamental level, why, in fact, have nucleic acids been selected as the carriers of genetic information rather than other aperiodic biopolymer such as proteins?

To address these (and other related) questions, the present author relies primarily on theoretical (quantum) computational chemistry. This choice is mainly driven by the difficulty in recreating the variety of potential prebiotic Earth conditions in a laboratory. From the chest of tools of theoretical chemistry, this work is based on statistical thermodynamics, semiempirical electronic structure methods such as PM7, and on density functional theory (DFT). From quantum and computational chemistry, total and free energies associated with the molecular structures of the building blocks of nucleic acids have been accurately calculated. The free energies were then fed to statistical mechanics to estimate the thermodynamic probabilities of the synthesis of nucleosides(tides) given postulated pathways. To run these calculations, an in-house Linux cluster was used along with the resources of Canada's national high-performance computing (HPC) system (*Compute Canada* which starting in June 2022 was replaced by the *Digital Research Alliance of Canada* (*The Alliance*)). *Compute Canada/The Alliance* officials estimate that, in 2022 (only), the author's calculations used more than ¼ of a million CPU hours (*ca.* 235,000 CPU hours ≈ 27 CPU years).

Two pathways were proposed for the formation of the nucleosides(tides): one may be termed "*classic pathway*" and the other "*alternative pathway*". The difference between two pathways is in the *order of addition* of the reactants. In the classic pathway, the condensation reaction of the *trifunctional connector* (TC), e.g., D-ribofuranose and the *nitrogenous base* (*recognition unit* (RU)) is carried out followed by the condensation of the *ionized linker* (IL), e.g., the phosphate ion. Since free energy is a state function, the change in free energies of two such pathways do not depend on the order of steps, however, the first condensation reaction can *trap* the intermediate's geometry in a different local minimum on the potential energy hypersurface of leading to differences in the geometries of the products and hence differences in the overall energies of reaction. This does not contradict in any way Hess' law since the *products have different geometries* depending on the order of addition and, hence, also different energies.

It is found that, overall, there is a slight energetic advantage of the *β*-anomers of nucleotides (the dominant form in today's nucleic acids) over their *α*-anomers counterparts. This observation holds for the modeling of these building blocks in both vacuum and implicit (continuum) aqueous solvation. Furthermore, for non-canonical nucleosides(tides) (those with alternative components (TCs, RUs and ILs)), some differences are observed favoring the dominant anomeric form observed today in the genetic material. *This small thermodynamic advantage operating over millennia may have contributed in the observed dominance of canonical β-anomers as we know them today*.

The classic pathway for the synthesis of the canonical building blocks of nucleic acids is favored in vacuum (but not in implicit solvation). The latter reversal of behavior on going from gas-phase to aqueous-phase is aligned with the well-known "*water problem*" which stipulates that the nucleotides are hydrolyzed (hence are unstable) in aqueous environment and that all laboratory attempts to synthesize nucleotides in aqueous environment have, as a result, very low reaction yields. This also suggests that the prebiotic environment may have had a low dielectric constant, i.e., may have been non-polar.

In the case of non-canonical nucleosides and nucleotides, either pathway in both aqueous and gas-phases favors the synthesis of the barbituric acid C^5-glycosylated (BA-C^5). Overall, the synthesis of barbituric acid (BA)-C^5-nucleosides(tides) is the most thermodynamically favored over all other canonical and several non-canonical nucleosides(tides) by typically > 10 kJ/mol (2.4 kcal/mol). *These results point toward barbituric acid as one of the potential prebiotic precursors of today's nucleobases, reinforcing a thesis that has already been proposed in the literature*.

The synthesis of nucleosides(tides) containing glycerol and N-(2-aminoethyl)glycine (AEG) is estimated to be considerably more favored compared to other trifunctional connectors, including the canonical D-ribofuranose and D-2'-deoxyribofuranose. In other words, the synthesis of AEG-nucleosides in implicit aqueous solvation, and glycerol-nucleosides(tides) in both vacuum and aqueous media, are thermodynamically favored by ≈ -20 kJ/mol. *These results suggest that non-canonical AEG and glycerol nucleosides(tides) may have assisted the synthesis of today's nucleosides(tides) in case the prebiotic environment has been aqueous*.

In present day, there are bacteria that live near deep ocean hydrothermal vents that may contain arsenic-based nucleic acids (or arsenate groups as ionized linker instead of phosphate), a few authors have wondered whether early forms of nucleic acid may have had As in place of P. The author reformulated this question in energetic terms. The results of the calculations in this thesis show that, surprisingly, there are almost *no differences in energies of reactions* upon replacing HPO_4^{2-} (hydrogen phosphate ion) by $HAsO_4^{2-}$ (hydrogen arsenate ion) in the TC of the nucleotides. This finding is true for both sets of calculations whether in aqueous environment or in vacuum. This unexpected finding means that *the possibility of an ancestral nucleic acid containing As instead of P cannot be ruled-out on thermodynamic grounds*. But why has the phosphate been selected in the grand majority of contemporary organisms? This is an open question that we leave for future investigations. Nevertheless, perhaps *a tentative answer could be related to the higher resistance to hydrolysis by phosphate*.

In-line with the general interest of the author of this thesis, *Shannon's information theory* is reviewed in relation to its capability of quantifying the "*amount*" of information (in bits) per symbol in a given coding language. The relevant chapter of the thesis contains a brief review of the important work of Lila Gatlin and of Mikhail Volkenstein on the capacity to convey information in a "4-letter nucleic acid language" where the bases/nucleotides are the symbols compared to that of a "20-letter protein language" where the individual amino acid residues are the symbols. The 4-letter language has less information carrying capacity per symbol than its 20-letter counterpart and,

hence, the effect of a mutation is, in general, less dramatic in a stretch of nucleic acid than the more information-dense language of protein. To this day, information theory has focused solely on the *amount* of information (the number of bits) in a given message of letter. Following the lead of Volkenstein, we present here a few original thoughts on the possible manners in which we can also take the *value* of information into quantitative account. The value of information is as important as its amount if not more especially for the living system. This value is related to the redundancy of the information and the effect of receiving it and its translation into physical "effects". This work is still in its early stages, however, several open questions in this domain are outlined in the relevant part of this thesis.

In closing, this thesis addresses certain aspect of evolutionary biochemistry from the standpoint of thermodynamics, that is, the spontaneity of the set of reactions being considered under differing conditions of solvation. The advantage of such approach is, as mentioned above, that the free energies of reactions are independent of the particular synthetic steps taking one from a set of reactants to a set of products. This can be an advantage since in many cases such detailed synthetic and mechanistic steps are largely unknown. A slight thermodynamic advantage will tend to favor the accumulation of a given species in nature at the expense of another produced via a competing reaction channel. Meanwhile, one must not forget the other side of the coin in chemical reactivity, which is its *dynamic* aspect. That latter aspect is captured by transition state theory and the kinetics of chemical reactions. In order to study the rate of the reactions that led to today's observed nucleic acids, we need a detailed synthetic route(s) and study the slow (rate-determining) step in any give pathway. This knowledge is presently largely unavailable and is certainly an alley for future explorations.

Thus, what has been left out in this work is *kinetics aspects*, but this is not all. Other factors that were not considered include – but are not limited to - the Earth's temperature, the presence of ions, uneven surfaces, pH variations, different atmospheric compositions, the flooding of the surface of planet Earth with UV radiation and/or cosmic showers, *etc*. These are all valid and important factors that were not considered, in part, due to the extreme complication that this would have entailed and the combinatorial explosion of the various sets of such conditions. However, with all these potential variables kept constant, our results remain valid and so are the questions they open for future investigations.

January 31, 2024

Declaration

Lázaro Andrés Monteserín Castanedo declares no conflict of interest and agrees to share any materials and/or information related with this thesis upon request.

Lázaro Andrés Monteserín Castanedo

PhD Candidate, M.Sc., B.Sc.

Department of Chemistry, Saint Mary's University

923 Robie Street, Halifax, Nova Scotia, Canada B3H 3C3

Dedication

To my wife Simone and our baby; "the loves of my life" and in memory of my mother María de Lourdes who taught me that in life great achievements often require great sacrifices.

Acknowledgement

Gratitude is one of the qualities that make us a better human being and today, I am grateful for many things.

¡*Muchas gracias*! to my Mother, María de Lourdes Castanedo González. I am the man that I am today because of you. I wish you could see me today. I remember how you always said that my future was outside of Cuba, you were completely right, *Mamá*. I hope that wherever you are, you are looking upon me and that I have been able to make you proud. I believe that a way of honoring our loved ones' memory is to follow their teachings and I hope to always follow the path that you created for me through your own sacrifices.

I am beyond grateful to my wife, Simone Monteserín Castanedo, for her infinite support and patience during this long journey. I can now say that you are no longer a "PhD widow." Do you remember when I told you I was thinking of pursuing my PhD and asked for your thoughts? You said: "you should go for it but, wait, how long is that going to be?" Thank you for your encouragement, for your guidance when I was lost and simply for being always truly you. I am the luckiest man on Earth to be chosen by you and I love you more than anything. You are going to be an amazing mother! I can't wait to embark on this new journey of becoming a parent with you. To you, I dedicate this verse paraphrased from the poem "*Con tantos palos que te dio la vida*" by Fayad Jamis: "…*con tantos palos que me dio la vida y aún no me canso de decir: te quiero*". I cannot thank enough my Canadian parents, Rozanne and Mac Persad for their unconditional love and support.

I am forever indebted to my Supervisor, Dr. Chérif Farid Matta (*Mount Saint Vincent* and *Saint Mary's* Universities). I still remember when I first came to Canada with my backpack full of books and $20 USD in my pocket. You were waiting for me at the airport at midnight and asked if I had brought a luggage, to which I responded "yes" without understanding the question. Twenty-minutes later, we were waiting at the baggage carousel for an imaginary luggage that never arrived. Chérif, you are more than a Supervisor, you are a close friend. Thanks for your support and guidance in personal and academic matters. Thanks for correcting my English, it has improved because of you. I hope I have helped you to improve your Spanish! ¡*En fin, muchas gracias de todo corazón*!

List of Figures

List of Tables

$(\Delta G^{\circ}_{\beta\alpha})$. All results are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The Integral Equation Formalism of the Polarizable Continuum Model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

Glossary of Acronyms

Gyr	**G**iga**yea**r
NAs	**N**ucleic **A**cids
aa	**a**mino **a**cids
RNA	**R**ibo**N**ucleic **A**cid
DNA	**D**eoxyribo**N**ucleic **A**cid
mRNA	**m**essenger RNA
tRNA	**t**ransfer RNA
rRNA	**r**ibosomal RNA
RU	**R**ecognition **U**nit
TC	**T**rifunctional **C**onnector
IL	**I**onized **L**inker
A	**A**denine
G	**G**uanine
C	**C**ytosine
T	**T**hymine
U	**U**racil
WC	**W**atson & **C**rick
CHs	**C**arbo**H**ydrate**s**
UV	**U**ltra**V**iolet
MM	**M**ela**M**ine
CA	**C**yanuric **A**cid
TAP	2, 4, 6-**T**ri**A**mino**P**yrimidine
BA	**B**arbituric **A**cid
C-BAMP	5-ribofuranosyl-**C**-**BA**rbiturate-5'-**M**ono**P**hosphate
MMP	N-ribofuranosyl-**M**elamine-5'-**M**ono**P**hosphate
AFM	**A**tomic **F**orce **M**icroscopy

^{1}H-NMR	proton **N**uclear **M**agnetic **R**esonance
TAPAS	TAP with succinic acid
CyCo6	CA linked with hexanoic acid or butyric acid
AMP	**A**denosine 5'-**M**ono**P**hosphate
DAP	2, 6-**D**i**A**mino**P**urine
CACo4	**C**yanuric **A**cid derivative with a short butanoic acid tail
XNA	*Xeno*-**N**ucleic **A**cids
5-MR	5-**M**ember **R**ing
6-MR	6-**M**ember **R**ing
(*p*)-RNA	**p**yranosyl-RNA
(*p*)-DNA	**p**yranosyl-DNA
GNA	**G**lycerol **N**ucleic **A**cids
PNA	**P**eptide **N**ucleic **A**cids
AEG	N-(2-**A**mino**E**thyl)**G**lycine
aegPNA	PNA containing AEG
LC-MS	**L**iquid **C**hromatography-**M**ass **S**pectrometry
Å	**Å**ngstrom
T_m	melting point
T_8	**8**-**T**-oligonucleotide
U_8	**8**-**U**-oligonucleotide
K	**K**elvin
CPG	**C**ontrolled **P**ore **G**lass
bPNAs	dipeptide **PNAs**
tPNAs	**t**hioester **PNAs**
ANAs	**A**lanyl **N**ucleic **A**cid
ED	**E**thylen**D**iamine
Ac	**Ac**etyl
ECI	**E**nantiomeric **C**ross-**I**nhibition

P	Phosphorous
As	Arsenic
S	**S**outh
N	**N**orth
CP	**C**remer & **P**ople
dr	**D**-2'-deoxy**r**ibofuranoses
r	D-**r**ibofuranoses
UNG	**U**-**N**-**G**lycosylase
MUG	DNA-**G**lycosylase
MM	**M**olecular **M**echanics
MD	**M**olecular **D**ynamics
QM	**Q**uantum **M**echanics
DFT	**D**ensity **F**unctional **T**heory
PES	**P**otential **E**nergy hyper**S**urface
K.E	**K**inetic **E**nergy
a.u	**a**tomic **u**nits
E	**E**nergies
B-O	**B**orn-**O**ppenheimer
HF	**H**artree-**F**ock
SCF	**S**elf- **C**onsistent **F**ield
post-HF	**p**ost-**H**artree-**F**ock
XRD	**X**-**R**ay **D**iffraction
HK	**H**ohenberg-**K**ohn
KS	**K**ohn and **S**ham
H-GGA	**H**ybrid-**G**eneralized **G**radient **A**pproximation
GGA	**G**eneralized **G**radient **A**pproximation
LSDA	**L**ocal-**S**pin-**D**ensity **A**pproximation
vdW	**v**an **d**er **W**aals

CIEFPCM	dielectric **C**onductor **P**olarized **C**ontinuum **M**odel
ZDO	**Z**ero **D**ifferential **O**verlap approximation
AO	**A**tomic **O**rbitals
BSSE	**B**asis **S**et **S**uperposition **E**rror
NDDO	**N**eglect of **D**iatomic **D**ifferential **O**verlap
AM1	**A**ustin **M**odel 1
PM6-D	**P**arametric **M**ethod 6 **D**ispersive
PM7	**P**arametric **M**ethod 7
AUE	**A**verage **U**nsigned **E**rrors
MMH	**M**ultiple **M**inimum **H**ypersurface
GB	**G**eneralized **B**orn
IEFPCM	**P**olarized **C**ontinuum **M**odel
QM-SCF	**Q**uantum **M**echanics **S**elf-**C**onsistent **F**ield
PME	**P**article **M**esh **E**wald
SAS	**S**olvent **A**ccessible **S**urface
NAD+	**N**icotinamide **A**denine **D**inucleotide
COSMO	**CO**nductor-like **S**creening **MO**del
MOPAC	**M**olecular **O**rbital **PAC**kage
IEFPCM	**I**ntegral **E**quation **F**ormalism variant of the "*Polarizable Continuum*
ΔE	difference in the total energies
$\Delta E_{(ZPE)}$	difference in the total **E**nergies corrected for **Z**ero-**P**oint vibrational
$\Delta H°$	difference in enthalpy at normal conditions (293.215K, 1 atm)
$\Delta G°$	difference in Gibbs free energies at normal conditions
drMP	2'-**d**eoxy**r**ibose 5'-**M**ono**P**hosphate
rMP	**r**ibose 5'-**M**ono**P**hosphate
dA	2'-**d**eoxy**A**denosine
dG	2'-**d**eoxy**G**uanidine
dC	2'-**d**eoxy**C**ytidine

dT	2'-**de**oxy**T**hymidine
dU	2'-**de**oxy**U**ridine
rA	**A**denosine
rG	**G**uanidine
rC	**C**ytidine
rT	**T**hymidine
rU	**U**ridine
dAMP	2'-**de**oxy**A**denosine 5'-**M**ono**P**hosphate
dGMP	2'-**de**oxy**G**uanidine 5'-**M**ono**P**hosphate
dCMP	2'-**de**oxy**C**ytidine 5'-**M**ono**P**hosphate
dTMP	2'-**de**oxy**T**hymidine 5'-**M**ono**P**hosphate
dUMP	2'-**de**oxy**U**ridine 5'-**M**ono**P**hosphate
AMP	**A**denosine 5'-**M**ono**P**hosphate
GMP	**G**uanidine 5'-**M**ono**P**hosphate
CMP	**C**ytidine 5'-**M**ono**P**hosphate
TMP	**T**hymidine 5'-**M**ono**P**hosphate
UMP	**U**ridine 5'-**M**ono**P**hosphate
Ns	**N**ucleo**s**ide
Nt	**N**ucleo**t**ide
5'-SMP	5'-**S**ugar **M**ono**P**hosphate
NMPs	ribo**N**ucleosides 5'-**M**ono**P**hosphate
AEG	N-(2-**A**mino**E**thyl)-**G**lycine
H_2CO	formaldehyde
P-form	**P**yranose form
F-form	**F**uranose form
2dRibf	D-2'-**de**oxy**Ribof**uranose
D-Ribf	D-**Ribof**uranose
D-Tho	D-**Thre**o**s**e

D-2dRib	D-**2**'-**d**eoxy**Rib**opyranose
D-Rib	D-**R**ibopyranose
CH2-COOH	acetyl group
NPT	**N**ormal **P**ressure and **T**emperature
FT-MW	**F**ourier **T**ransform - **M**icro**W**ave
NMR	**N**uclear **M**agnetic **R**esonance
HPLC	**H**igh **P**erformance **L**iquid **C**hromatography
MS	**M**ass **S**pectrometry
LC	**L**iquid **C**hromatography
RNP	**R**ibo**N**uleo**P**rotein
dUTPase	dUTP nucleotidohydrolase
UDG	**U**racil-**N**-**G**lycosylase
PBS	**P**hosphate **B**uffered **S**aline
ddNs	2',3'-**di**-(**d**eoxy)**N**ucleoside**s**
$^2T_3$	2'-endo-**T**wist-3'-exo
$^3T_2$	3'-endo-**T**wist-2'-exo
$^1C_4$	1'-endo-**C**hair-4'-exo
$^4C_1$	4'-endo-**C**hair-1'-exo
$H_2PO_4^-$	(di)hydrogen-Phosphate ion
$H_2AsO_4^-$	(di)hydrogen-Arsenate ion
FSA	**F**unctional **S**imilarity of **A**mino acid
Trp	**Tryp**tophan
Ala	**Ala**nine
ncDNA	**n**on-**c**oding **DNA**
ATPase	**ATP** synth**ase**
Ns(t)s	**N**ucleo**s**ides(**t**ide**s**)
Ns(t)ANA	**N**ucleo**s**ides(**t**ide**s**) **ANA**logs
XNA	**X**eno-**N**ucleic **A**cids

FDA	United States **F**ood and **D**rug **A**dministration
SELEX	**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment
PCR	**P**olymerase **C**hain **R**eaction
oligoNts	**oligoN**ucleo**t**ide**s**
ML	**M**achine **L**earning
DNN	**D**eep **N**eural **N**etworks
AEGIS	**A**rtificially **E**xpanded **G**enetic **I**nformation **S**ystem
OTA	**O**chra**T**oxin **A**
GAI	**G**enerative **A**rtificial **I**ntelligence
DL	**D**eep **L**earning
S_i	**S**tructural information entropy
SIC	**S**tructural **I**nformation **C**ontent
QTAIM	**Q**uantum **T**heory of **A**toms in **M**olecules
QED	**Q**uantitative **E**stimate of **D**rug-likeness index
QM/MM	**Q**uantum **M**echanics/**M**olecular **M**echanics
S_T	**T**animoto **S**imilarity index
ONIOM	**O**ur own **N**-layered **I**ntegrated molecular **O**rbital and **M**olecular
CFI	**C**anada **F**oundation for **I**nnovation
CPD	**C**yclobutene **P**yrimidine **D**imers
$[B(OH)_4]^-$	borate ion
H_2NCOH	Formamide
cGMP	**c**yclic 3', 5' **G**uanosine **M**ono**P**hosphate
POPC	**P**almitoyl-**O**leoyl**P**hosphatidyl**C**holine
POPA	**P**almitoyl-**O**leoylphosph**A**tidic **A**cid
LPC	**L**yso**P**hosphatidyl**C**holine
MALDI-TOF	**M**atrix-**A**ssisted **L**aser **D**esorption/**I**onization-**T**ime **O**f **F**light

Table of Contents

Chapter 1

Introduction

"It was a biochemical Jackson Pollock: a field of strings, tangles, loops"

Carl Zimmer (2021),

Life's edge: searching for what it means to be alive. New York, NY, Dutton

1.1   Theories on the prebiotic origin of the first biomolecules of life

It is considered that the Earth originated about 4.53 billion years (Gyr) ago [1], but when and how did the first biomolecules of life emerged in the prebiotic earth? Dohm and Maruyama proposed in 2015 the concept of "Habitable Trinity" as one of the minimum requirements for the emergence of life. It consists in the coexistence of atmosphere, water and landmass with continual material circulation between the three of them that is driven by the sun. The main elements for life (C, N, H, O and nutrients) are precisely provided from the three components of the Trinity [2].

Life is generally characterized by 1) compartmentation: ability to keep its components together and distinguish itself from the environment, 2) replication: the ability to process and transmit heritable information and 3) metabolism: the ability to capture energy and matter [2]. The biopolymers of life (proteins, nucleic acids and lipids) are the molecules that make these processes happen.

The first biomolecules and life may have emerged in some moment between 3.5 - 4 Gyr [3] [3], but how did this happen?

There are three main theories for the origin of life:

1) Life may have originated in the hydrothermal systems in the primitive oceans. This is based on the discovery of thermophilic organisms in the deep-sea vents in the late 1970s [4, 5, 6].

2) The materials essential for the origin of life came from outer space. Extraterrestrial objects (meteorites, comets and interplanetary dust particles) delivered essential chemical elements, e.g., carbon and basic organic molecules such as purines and pyrimidines for the ensemble into more complex biopolymers [7, 1, 8]. For example, carbonaceous chondrites that come from asteroidal

bodies contain 5% of carbon and are rich in a mixture of organic matter that includes amino acids, purine and pyrimidine nitrogenous bases, monocarboxylic acids and sugar-like molecules [9].

3) The most accepted theory for the origin of life is the theory of chemical evolution proposed by Aleksandr I. Oparin and John Burton S. Haldane in 1924-1925 called the "prebiotic soup". This theory proposed that in a primitive reductive atmosphere more simple organic compounds went through a series of chemical reactions to give origin to more complex biomolecules [10, 11].

1.2 "RNA world hypothesis" and the chemical nature of nucleic acids

Which biomolecule of life emerged first? Some literature make reference for example to proteins as the first biomolecules, based essentially on the works by Stanley Miller and Harold Urey in 1953 [12, 13] on the synthesis of **a**mino **a**cids (aa) in chemical conditions that simulated the "prebiotic soup" [14].

Regardless of Miller-Urey's experiment most literature agrees that **N**ucleic **A**cids (NAs) were the first biomolecules that emerged in the prebiotic earth. Specifically, **R**ibo**N**ucleic **A**cid (RNA) has been accepted as the most likely first biopolymer. This has been known as the "RNA world" hypothesis [15, 16, 17].

Unlike DNA or proteins that can store information or have catalytic activity respectively, RNA can have both biological functions simultaneously. **M**essenger RNA (mRNA) can be obtained from a single strand of DNA through transcription and it contains the genetic information that can be translated into amino acids by **t**ransfer RNA (tRNA). On the other hand **r**ibosomal RNA (rRNA) catalyzes the formation of peptide bonds in the ribosomes [18, 19, 20, 21].

These biochemical properties make RNA the ideal candidate for the first primitive biopolymer of life. If RNA was synthesized first, then it must have played a primordial role in the biosynthesis of other important biomolecules: proteins, lipids and DNA. This is the main premise of the "RNA world" hypothesis.

The building blocks of **n**ucleic **a**cids (NAs) contain three main components: a nucleobase (generically named **R**ecognition **U**nit (RU)), a sugar (**T**rifunctional **C**onnector (TC)) and a phosphate group (**I**onized **L**inker (IL)). A nucleoside contains a sugar + base and a nucleotide is a nucleoside + phosphate. Most nucleosides(tides) exhibit a *β*- (the base in the C1' is in *cis* position

to the phosphate in the C5' of the sugar) instead of an *α*-configuration (the base in the C1' is in *trans* position to the phosphate in the C5' of the sugar) at the C1' of the sugar either in DNA or RNA. The bases are N-glycosylated on the C1' position of the sugar that can be D-ribofuranose in RNA or D-2'-deoxyribofuranose in DNA (**Figure 1.1a**).

Today's NAs contain mainly 5 nitrogenous bases, 2 purines: **A**denine (A) and **G**uanine (G) and 3 pyrimidines: **C**ytosine (C), **T**hymine (T) (in DNA) and **U**racil (U) instead of T (in RNA). The phosphate ion connects adjacent nucleosides usually between the C3' and C5' positions. The building blocks usually create antiparallel or parallel strands that can form a double helix by **W**atson & **C**rick (WC) complementary base pairing [18] (**Figure 1.1b**).

1.2.1 Prebiotic synthesis of nucleic acids

If we accept that RNA came first, then, other questions arise. For example, how did this molecule originate in the first place and what basis early prebiotic conditions favored the selection and assembly of particular components for the building blocks of nucleic acids?

One of the first models for the prebiotic synthesis of nucleic acids proposed that these biomolecules were the product of prebiotic and geochemical reactions. This is known as the "classic model", "drying pool" or "drying lagoon". According to the classic model, regular cycles of dehydration and rehydration can promote polymerization reactions. In this way through consecutives condensation reactions RUs can bind to the TCs to produce the nucleosides and these ones react with ILs to obtain the corresponding nucleotides. The nucleotides can bind to each other through 3'-5' phosphodiester bonds to produce the oligonucleotides and finally, two antiparallel strands can assemble with each other through WC complementary base pairing to create a double helix [11, 23] (**Figure 1.2**).

1.2.2 Challenges for the prebiotic synthesis of the building blocks of nucleic acids

Within the framework of the "classic model", Orgel and coworkers explored the potential for the formation of the glycosidic bond between the purines (A, G, inosine and xanthine) and ribose by drying and heating the reactants at 100 ℃ in the presence of acidic catalysts for 2hrs. Only A was found to couple with ribose to produce *α*- and *β*-furanosyl nucleosides within ≈ 2-
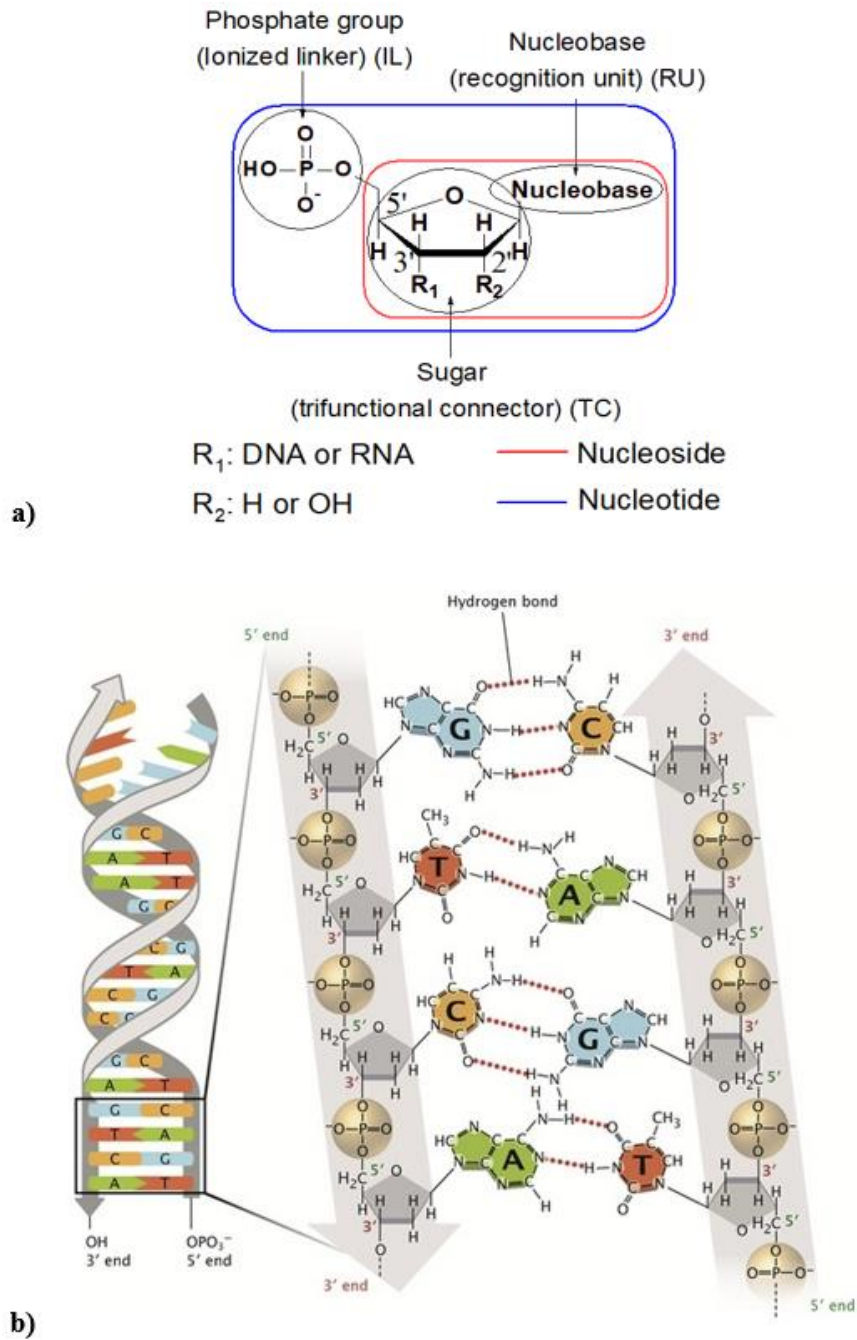
Figure 1.1 **a)** Chemical nature of today's building blocks of nucleic acids (nucleosides and nucleotides). **b)** 3D structure for a DNA double helix where A: adenine, G: guanine, C: cytosine, T: thymine (taken from [22] and reproduced with permission of Nature Education).
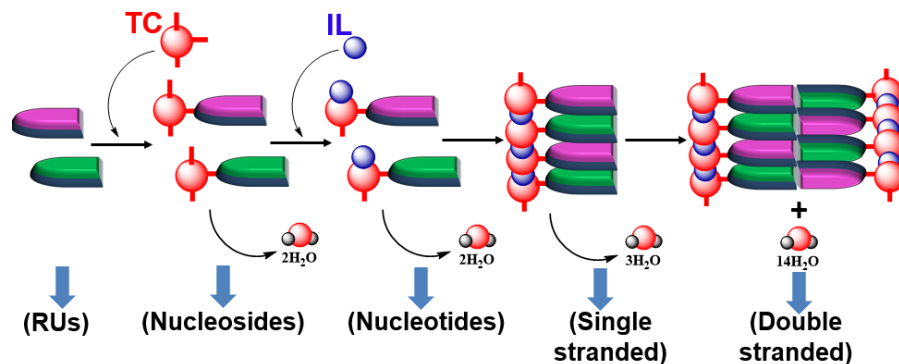
Figure 1.2   "Classic model" for the prebiotic synthesis of the building blocks of nucleic acids and the single and double stranded nucleic acid helixes [11] (taken from [11] and reprinted with permission of Chemistry & Biology).

10% of yield. The low solubility of G proved to be a severe limitation to the guanosine nucleoside formation. Other attempts on synthesizing U and C nucleosides using heating-drying reactions were unsuccessful [24].

The relatively low yield obtained by Orgel and coworkers for the synthesis of nucleosides was attributed to the instability of the glycosidic bond in aqueous environment giving rise to what is known as the "nucleoside problem" (a special case of the more general "water problem") [25, 26, 11, 27]. This means that the glycosidic bond between canonical nucleobases and the sugar ribose or 2'-deoxyribose is thermodynamically unstable in aqueous solution.

It is well known that the N-glycosylation is better achieved when the anomeric C1' of the sugar moiety is previously activated with a better leaving group than water, but these synthetic methods include the use of hydrolytically unstable activating agents and reaction conditions that may not be similar to the environment of the prebiotic earth [28].

Other challenges on the prebiotic synthesis of the building blocks of nucleic acids following a "classic" approach also include the following:

✓ The inefficiency of the phosphorylation of nucleosides and the polymerizing of the nucleotides through phosphodiester bonds in water is due to the thermodynamical instability of the phosphodiester bonds in aqueous solution. This limitation is also part of "the water problem" [29]. The formation of the nucleotides (phosphorylated nucleosides)

and the phosphodiester bond implies that the phosphate group acts as an electrophile and there is the release of a water molecule in the process, but the phosphate group deprotonates, except in acidic conditions (pH ≤ 2) and this decreases its electrophilicity increasing the electronic repulsion and making the phosphodiester bond formation more difficult [30].

✓ The low release of phosphate by nature due to the inertness by the most relevant minerals (apatite, whitlockite, brushite) and the poor regioselectivity for the R_1-O5'-P-O3'-R_2 phosphodiester bond formation in plausible prebiotic conditions [31].

✓ The nucleobases found in life today and their corresponding free nucleosides/nucleotides do not self-assemble through WC hydrogen bonding in the presence of aqueous solvents. This challenge has been named as ''the paradox of base pairing'' [32, 33].

✓ The prebiotic synthesis of D-ribose or D-2'-deoxyribose is still also a challenge for prebiotic chemistry. The synthesis of ribose was first proposed by Butlerow through the formose reaction in 1861 [34]. The formose reaction includes a series of polymerizations of formaldehyde (H_2CO) to obtain **C**arbo**H**ydrate**s** (CHs) but it presents a series of problems: 1) ribose is obtained as an intermediate product among a large number of other sugars with more or less carbon atoms and structural isomers with the same number of carbon atoms than ribose. This happens because chemical compounds containing carbonyl groups (R-C=O) can react through an enolization and an aldol addition to create complex polymers (asphalts). Hence, this is known as "the asphalt problem" [16]. 2). As a result the yields for ribose are usually ≤ 1%, 3) the kinetic control of this reaction is crucial, 4) the reaction requires high concentrations of formaldehyde (≥ 0.1M) but it has been estimated that the concentration of this molecule in the prebiotic oceans must has been around 10^{-3} M [35, 36].

1.2.3 Alternative solutions for the challenges associated with the prebiotic synthesis of the building blocks of nucleic acids

Tetrahydroxyborate ($[B(OH)_4]^-$) minerals can sequester aldopentose and pentulose (ribose, ribulose and xylulose) from a prebiotic mixture of carbohydrates [37]. From these sugars, ribose forms the most stable $[B(OH)_2]^-$-complexes. This may be possible due the capacity of $[B(OH)_4]^-$ to bind to 1,2-diol units of molecules especially in a *cis* position in a ring. Ribose contains two C1'-

OH and C2'-OH in the same direction in the hemiacetal ring and can create additional hydrogen bonds between C3'-OH of ribose and $[B(OH)_4]^-$ that further stabilize the complex. Additional studies have shown that ribose can also be synthesized in the presence of borate [16, 38].

Since the chemical synthesis of nucleic acids in prebiotic conditions has encountered many challenges (see section 1.2.2) an alternative has been to attribute today's nucleic acids chemical nature to "evolution" [17]. The evolutionary theory for the prebiotic origin of nucleic acids proposes that DNA as we know it today may have been a descendant of ancestral predecessors. Probably a proto-RNA (first nucleic acids ancestor) with different components for the building blocks than ribose, the canonical nucleobases and phosphate were easier synthesized in the environmental conditions of the early earth [11], then through a series of chemical transformations this proto-RNA evolved into a series of pre-RNAs [33] that eventually produced todays RNA and DNA molecular structures (**Figure 1.3**).

Following this idea, different alternative synthetic routes and non-canonical components for the building blocks of nucleic acids have been proposed in the literature.

For example, the challenges collectively referred as "the water problem" led Orgel and coworkers to synthesize cytidine α-anomers from a mixture of cyanamide and pre-existing D-ribose-5'-phosphate, D-ribose and D-arabinose. Finally the α-anomers were photoanomerized to β using **U**ltra**V**iolet (UV) light [39]. These experiments are known as the "ribose centric model" [11].

Following this idea, Sutherland and coworkers have proposed numerous routes for the synthesis of cytosine containing nucleosides and nucleotides [40, 41, 23], but still the results obtained by these authors have not reached yields $\geq 60\%$ due to competitive photodecomposition of the final products and the formation of U from the deamination of C. These synthetic methods also introduced the idea that the phosphate group may have not been initially in the 5' position of the ribose.

In 2000, Hud and Anet proposed the "polymer fusion model" as a way to overcome the "paradox of base pairing". The model argues that first, certain RUs can create stacked arrays and organize in supramolecular or rosette assemblies (polymers). This polymer can then be connected by a TC-IL backbone [42]. This model proposes a plausible route for the formation of NA polymers in prebiotic conditions which is dependent on RUs that can create rosette assemblies through complementary base pairing and introduces the idea that nucleic acids can be synthesized through the glycosylation of the sugar-phosphate backbone to the nucleobases already paired.
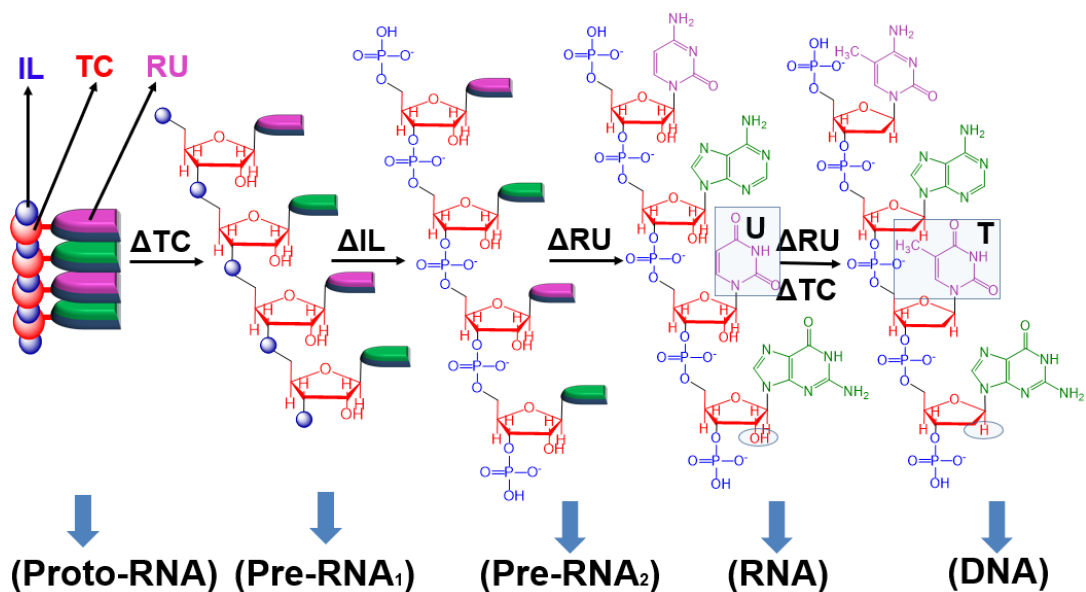
Figure 1.3 Evolutionary theory for the prebiotic emergence of nucleic acids [11] (taken from [11] and reprinted with permission of Chemistry & Biology).

Theoretical modeling of the ribose-phosphate backbone by Šponer and coworkers has shown that the formation of the nucleic acid polymers depends on the geometric properties of the molecular structure of the backbone and its components [43]. The idea that phosphate may have created bonds with ribose before the RUs has inspired many alternative synthetic studies involving the synthesis of the building blocks from a bicyclic or acyclic sugar-phosphate derivative [44, 45, 46, 47, 48, 49].

The polymer fusion model has a main disadvantage: it lacks of an explanation for how the components of the building blocks of nucleic acids may have emerged in prebiotic conditions.

Even when many models have been proposed for the prebiotic synthesis of RNA based on the "RNA world" paradigm, a plausible route for the prebiotic synthesis of DNA is missing, even when this polymer is essential for life because it contains the genetic information [3]. We cannot dismiss the possibility that RNA and DNA may have emerged simultaneously and that maybe the components of both nucleic acids coexisted in a prebiotic pool or lagoon.

1.2.3.1 Non-canonical RUs

If we consider that RNA and DNA emerged as a result of evolutionary selection, then this means that maybe the components of the building blocks of "proto- and pre-RNAs" TCs, RUs and ILs were different from the ones present in today's nucleic acids. From many RUs candidates that have been tested in order to overcome the "paradox of base pairing" and the "nucleoside problem" four non-canonical nucleobases have shown potential as the RUs for the first proto- and pre-RNAs. These bases are the triazines: **M**ela**M**ine (MM), **C**yanuric **A**cid (CA) and the pyrimidines: 2, 4, 6-**T**ri**A**mino**P**yrimidine (TAP), **B**arbituric **A**cid (BA) (**Figure 1.4a**). They can pair spontaneously through complementary hydrogen bonding similar to the 5 canonical nucleobases to create rosette or hexad (six-sided polygons) assemblies in aqueous environment (**Figure 1.4b**).

Cafferty and coworkers studied in 2018 a library of 91 nitrogenous bases that includes the 5 canonical nucleobases and 86 other heterocycles (**Figure 1.5**). These other bases contain in their structures exocyclic H, amine ($-NH_2$) and carbonyl ($C=O$) groups [52]. The authors evaluated this chemical space based mainly on 6 criteria: 1) their capacity to create complementary base pairing through at least two hydrogen bonds, similar to the canonical bases, 2) their capacity to create π stacked assemblies, 3) good chemical and photostability, 4) their ability to react with ribose in aqueous environment, 5) the ability to synthesize them in prebiotic conditions and 6) their capacity to absorb UV radiation. Cafferty *et al.* found that TAP, MM, CA and BA fulfilled these six requirements.

There have been many other studies that show how TAP, MM, BA and CA can create polymeric assemblies in water. For example, in a publication from 2016 [51] Cafferty and coworkers proved the formation of 2nm hexads between 5-ribofuranosyl-**C**-**B**arbiturate-5'-**M**ono**P**hosphate (C-BAMP) and N- ribofuranosyl-**M**elamine-5'-**M**ono**P**hosphate (MMP) or BA with MM in aqueous solution (see **Figure 4b**). An important remark of this study is the fact that the *β*-anomer of the nucleotides favored the formation of the supramolecular assemblies over the *α*-counterpart. The polymers were characterized using **A**tomic **F**orce **M**icroscopy (AFM) and proton **N**uclear **M**agnetic **R**esonance (^{1}H-NMR) spectroscopy.

In other studies, Cafferty *et al.* proved that TAP and CA derivatives create fibers that are linear π stacking assemblies of hexads 1.6 - 2 nm thick. The topography of the fibers was obtained
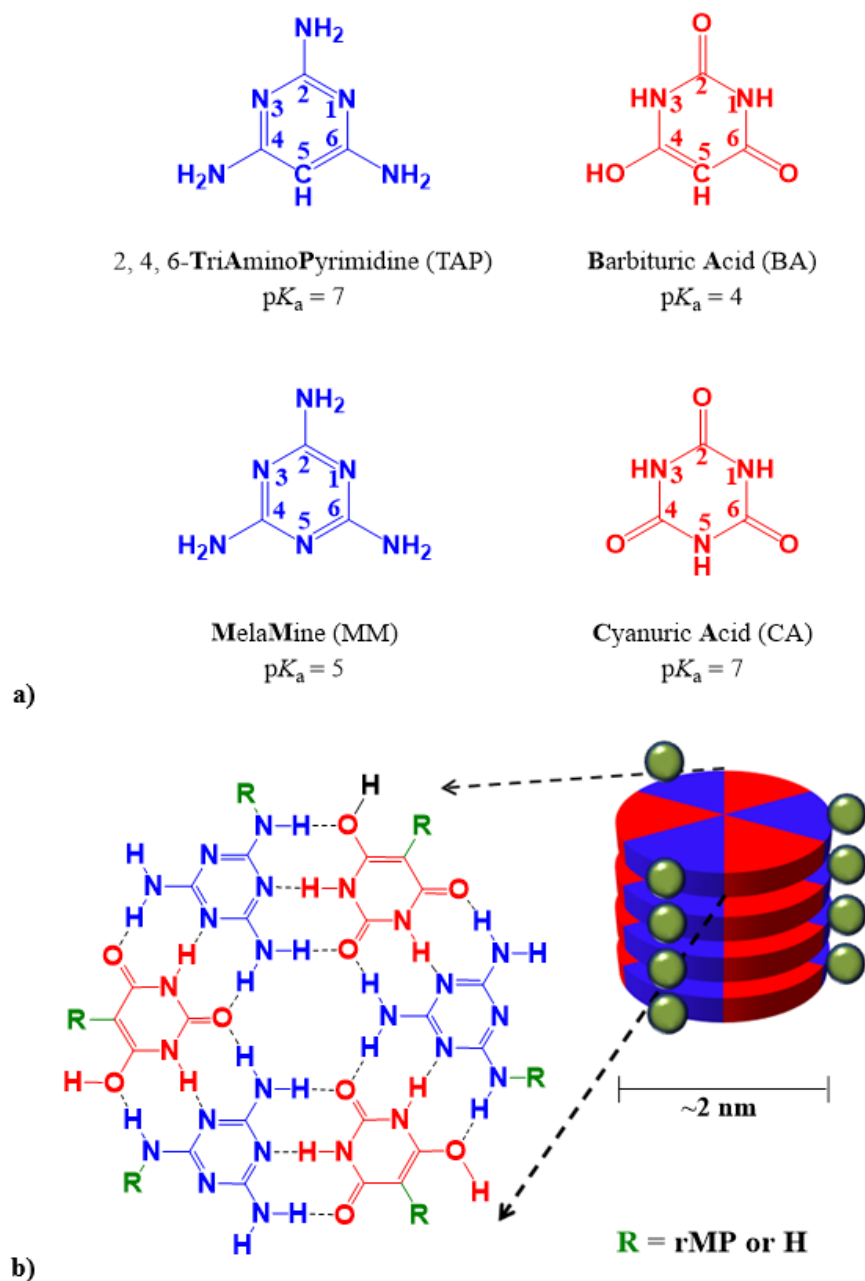
Figure 1.4 **a)** Molecular structure for **M**ela**M**ine (MM), **C**yanuric **A**cid (CA), 2, 4, 6-**T**ri**A**mino**P**yrimidine (TAP) and **B**arbituric **A**cid (BA) with their corresponding pKa [50]. **b)** Formation of a hexad in aqueous solution between the 5-ribofuranosyl-C-**BA**rbiturate-5'-**M**ono**P**hosphate (C-BAMP) or **B**arbituric **A**cid (BA) (in red) with N-ribofuranosyl-**M**elamine-5'-**M**ono**P**hosphate (MMP) or **M**ela**M**ine (MM) (in blue). The green group R or green spheres represent H for BA and MM and **r**ibofuranosyl-5'-**M**ono**P**hosphate (rMP) for C-BAMP and MMP [51].

using atomic force microscopy. They conjugated TAP with succinic acid through an amide bond (TAPAS) and CA linked with hexanoic acid or butyric acid (CyCo6) to enhance both non-canonical bases aqueous solubility. They find that TAPAS can create macromolecular assemblies with CA and CyCo6 with TAP [36, 53].

Cafferty and Hud have also explored the formation of complementary base pairing between the non-canonical nucleobases TAP and BA and the relevance of these interactions in the replication of a proto-RNA. These two molecules can assemble spontaneously in circular hexads and the aqueous exposure of the hexads is thermodynamically unfavored (13 kcal/mol) [50, 32].

On another study, Li and coworkers created supramolecular assemblies that contained adenine with CA and **A**denosine 5'-**M**ono**P**hosphate (AMP) and 2, 6-**D**i**A**mino**P**urine (DAP) with CA and a CA derivative with a short acid tail (butanoic acid) (CACo4) respectively. All the nucleobases were proved to assemble with the modified CACo4. The data was tested using UV absorption and NMR techniques [54].

Addressing the "nucleoside problem" numerous studies have reported that these four non-canonical nucleobases TAP, MM, BA and CA can spontaneously create N- and C-glycosidic bonds with different types of sugars or phosphate-sugars derivatives. For example, when TAP is dried with ribose at 35ºC the corresponding β-C-nucleoside was obtained as the majority product with a yield of at least 20%. The β-N-, α-N-, β-C and α-C-nucleosides with the furanose or pyranose forms of ribose were obtained as minor products with a combined yield of 60-90% [55].
In the same paper from 2016 Cafferty and coworkers [51] mixed BA and MM with ribofuranose-5'-monophosphate at 20ºC for 24h at different pH ranges and obtained a yield of ~80% and ~33-55% for the anomeric mixture (β- and α-configurations) of the BA (C-glycosylated) and MM nucleotides respectively.

Mungi and coworkers studied the reaction between ribose 5'-monophosphate and BA heating the mixture at 90ºC for 3hrs. They obtained the corresponding β-C-nucleotide as the main product in a yield of 40-85% depending on the initial conditions. When the authors performed the polymerization of the resulting nucleotides through dehydration-rehydration cycles at low pH and high temperature, a mixture of the α- and β-polynucleotides was obtained [56].
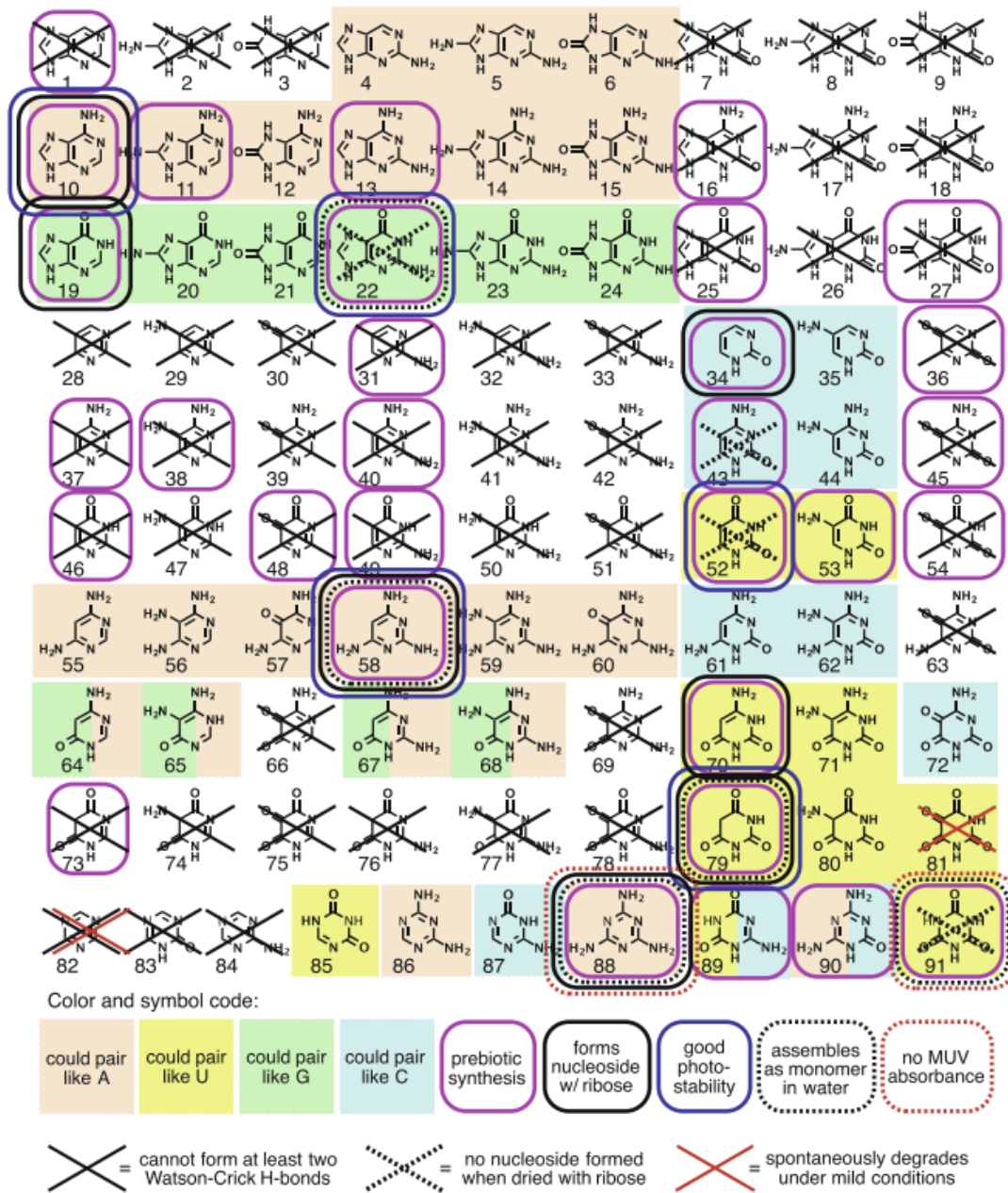
Figure 1.5 Set of nitrogenous bases tested by Cafferty and coworkers (taken from [52] and reprinted with permission of Springer Nature).

1.2.3.2 Other trifunctional connectors (TCs)

As previously mentioned, the synthesis of ribose through the formose reaction faces the "asphalt problem". In addition to this the mutarotation of ribose in water produces an equilibrium

mixture between the 6 member ring (6-MR) pyranose isomers: β (~ 59%), α (~ 20%), the 5 member rings (5-MR) ribofuranose: β (~ 13%), α (~ 7%) and the open chain (~ 1%) [57, 58]. In order to avoid the challenges associated with the formose reaction the prebiotic chemists have looked at alternative TCs.

Nucleic acids with non-standard sugars are denominated "*Xeno*-**N**ucleic **A**cids" (XNA) [59].

Different TCs have been employed instead of the usual 5-**M**ember **R**ing (5-MR) 2' - deoxy or ribofuranose, e.g. 6-**M**ember **R**ing (6-MR) sugars (**p**yranosil-RNA {(p)-RNA}) [60, 61], **t**hreose (in **T**hreosyl **N**ucleic **A**cids (TNAs)) [11, 62, 63], glycerol (in **G**lycerol **N**ucleic **A**cids (GNA)) [11, 64, 65, 66, 67, 68, 69] and **P**eptide **N**ucleic **A**cids (PNAs) - a variety of XNA where the sugar-phosphate is replaced by a peptide backbone [70, 71, 72, 73, 74, 75, 67] (**Figure 1.6** for an example on different XNAs).

The question on whether the first proto-nucleic acids had a TC different than ribofuranose becomes even more intriguing since there are examples of naturally occurring nucleosides(tides) with TC that differ by some extent from ribofuranose, e.g., salicin (a natural nucleoside with anti-inflammatory properties that contains a glucose instead of ribose and is isolated from willow bark) [76].

In 2018, Fialho [77] and coworkers reported the results for the glycosylation of TAP with a set of sugars that included 5 aldopentoses, one ketopentose, 8 aldohexoses, one ketohexose and two aldotetroses. All sugars generate 5- and 6-MR nucleosides except ribulose, erythrose, ribose 5'-monophosphate (rMP) and threose. The mixture of TAP+TC in 1:1 molar ratio was heated at 85 $^{\circ}$C for 24 hrs at pH=1 or pH=7. All products were analyzed by **L**iquid **C**hromatography-**M**ass **S**pectrometry (LC-MS) and ^{1}H-NMR. All TCs glycosylated to TAP in different yields, e.g., at pH = 7 ribose (31%), rMP (8%) and threose (2%). The low yield associated with the threose-TAP nucleosides was related to the higher degradation rate of threose compared to other sugars. This study proposes the idea that if a proto-RNA with non-canonical nucleobases emerged as the first biomolecule then it is possible that this ancestral polymer may also have non-canonical TCs.

The extensive literature on alternative TCs that can glycosylate with RUs, phosphorylate and generate macromolecular assemblies similar to today's nucleic acids led the Eschenmoser's group to hypothesize that ribose was later selected by nature as the present TC due to its ability to offer a balance between stability and flexibility of the double helix in comparison with other sugars.
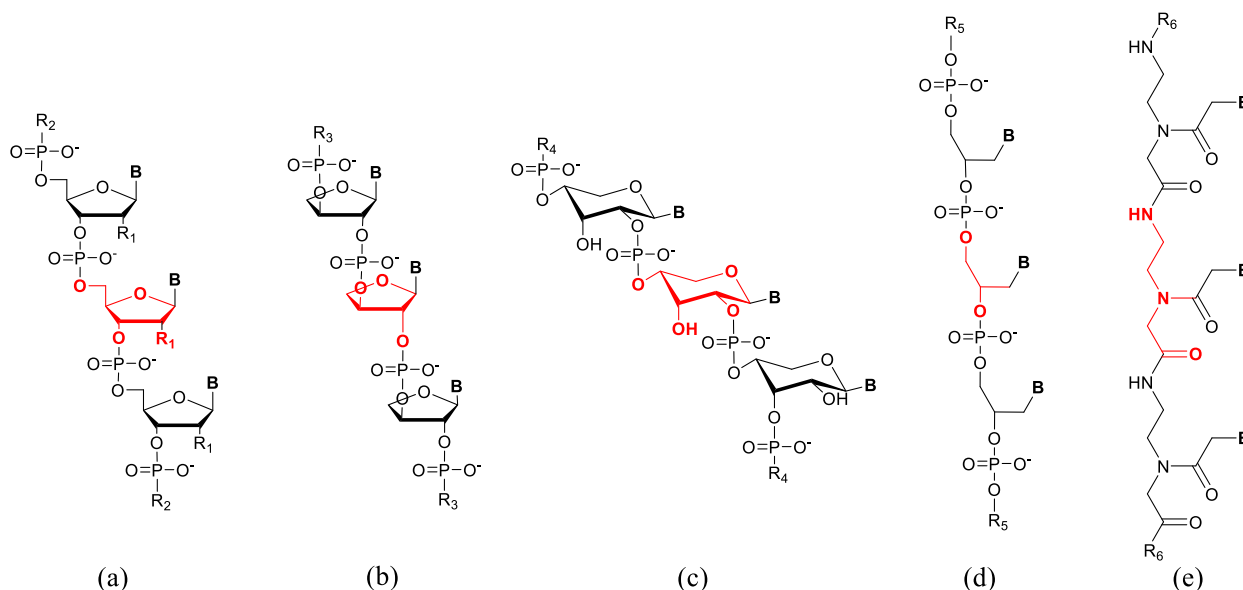
R₁: H or OH, R₂: DNA or RNA, R₃: TNA, R₄: p-RNA, R₅: GNA, R₆: PNA

Figure 1.6   Molecular structure of single stranded of (a) DNA/RNA, (b) S-**T**hreose **N**ucleic **A**cids (TNA), (c) **p**yranosil nucleic acid ((*p*)-RNA), (d) **G**lycerol **N**ucleic **A**cid (GNA) and (e) N-(2-**A**mino**E**thyl)**G**lycine (AEG) peptide nucleic acid (aegPNA) (taken from [17] and reprinted with permission of Taylor & Francis).

These authors hypothesize that this characteristic may have also pressured evolution to select the 5-MR ribofuranose instead of the 6-MR ribopyranose since the 6-MR creates stronger complementary base pairing [78, 79].

In the next sections, a brief description of the literature available on the prebiotic plausibility of different XNAs will be provided.

Pyranosil (p)-RNAs

The Eschenmoser group has studied extensively the ability of nucleotides containing pyranose ring sugars to create double stranded polynucleotides and hybrid (*p*)-RNA:RNA helices.

As an example of these studies; in a paper from 1993 Pitsch and coworkers formulated the question on why nature selected 5-MR instead of 6-MR sugars [79]. The authors analyzed the thermodynamic properties of an antiparallel octamer of A:U with C4'-C2' phosphodiester bonds

they found that the complementary base pairing of (*p*)-RNA is stronger than its RNA analog. This observation was related to the geometry of the (*p*)-RNA backbone in which the base pair distance was approximately 4.5Å shorter than the base distance in other homo-RNAs. The ΔG° of formation for the double helix of a ribopyranosil-A: ribopyranosil-U tetramer (A_4:U_4) is -6.2 kcal/mol meanwhile for the furanosyl analog is almost half (-3.3 kcal/mol). *The higher thermodynamic stability of the (p)-RNA double helix may have posed a limitation for its survival through natural selection due to its reduced suitability for replication.*

On the other hand, in this study all the mono- and polynucleotides used by the authors were synthesized using a non-prebiotic method that involved protecting groups in a Vorbrüggen-Hilbert-Johnson glycosylation.

Pitsch and coworkers also studied in 1999 [80] the polymer properties of a group of RNA derivatives in which the TC-phosphate backbone contained sugars different from ribose. Their study included the analysis of nucleic acids containing 2', 3'-dideoxy-D-glucose (homo-DNA) which could create double stranded-like nucleic acids but cannot pair with a single strand of RNA.

A similar finding was reported in 2003 [81] where Pitsch *et al.* proposed that (*p*)-RNA could also create double helixes but not chimeric double stranded (*p*)-RNA:RNA polymers [17].

Threose nucleic acids (TNAs)

Supporting the idea of prebiotic-threose nucleic acids the Eschenmoser group [63] proposed in 2000 that *α*-threofuranose could create different *α*-L-threo furanosyl-(3'-PO_4^--2') antiparallel double stranded helices and could also cross-pair with RNA and DNA strands. The measured melting points (T_m) from **U**ltra**V**iolet (UV) spectroscopy and ΔG° (kcal/mol) for the formation of the TNA double strands are comparable to the values measured for RNA and DNA double stranded oligonucleotides.

The **8**-**T**-oligonucleotide (T_8) TNA showed more resistance to hydrolytic degradation of the phosphodiester bond at pH = 8 and 35°C with a half-life time of months compared to the DNA-type analogs, meanwhile for ribopyranosil-T_8 the half-life time was 4 days and for ribofuranosyl U_8 was ½ day. *In this study the corresponding nucleosides were synthesized using non-prebiotic conditions and reagents.*

Eschenmoser and coworkers [63] proposed that, in contrast to (*p*)-DNAs, TNA strands have the capacity to create hybrid double helices with RNA and that this property may have given this XNA an evolutionary advantage allowing it to exchange information with RNA and be used as template in the non-enzymatic formation of RNA sequences. This report is historically relevant because it demonstrates there are exceptions for the rule that stipulates that a nucleic acid backbone needs to have 6-bonds per phosphodiester unit to generate a stable duplex. TNA contains 5 bonds instead of 6 and still can generate stable duplexes in a similar fashion to DNA and RNA.

Glycerol nucleic acids (GNAs)

Meggers and coworkers [65, 82] have published that a simpler nucleic acid containing a backbone of propylene glycol (glycerol derivative) instead of ribose exhibited WC-like complementary base pairing and could create antiparallel double helices. In the polymer, propylene glycol contains two stereoisomers R and S at the C^2. The authors found that GNA strands are stable and resistant to hydrolysis at pH = 7 and 298 K. This means that the terminal OH groups of the GNA strands do not attack vicinal phosphate groups causing the strand to break through a transesterification mechanism. Furthermore, the authors conclude that the GNAs were also resistant to alkaline hydrolysis at pH ≈ 13.

Meggers *et al.* compared the thermostability of different 17-mer oligonucleotides of DNA, DNA-GNA hybrids and GNA double helices. The T_m of the DNA double helices was measured to be 47 °C and as they introduced single and tri-GNA-nucleotides the stability of the double helix decreased. This translates in a suboptimal pairing between the GNA and DNA strands. Interestingly, when they created a GNA duplex with a similar sequence to the one in DNA the thermostability increased to 71 °C. This result showed that GNAs could create more stable double helixes than their DNA analogs. This stability is in agreement with the thermodynamic stability of the duplexes with $\Delta\Delta G° \approx$ - 2.6 to - 5.9 kcal/mol.

The GNA duplexes created antiparallel strands with WC complementary base pairing similar to DNA, with 1:1 base proportion. GNA was slightly less susceptible to infidelity in the pairing (changes in A:T, G:C pairing). For GNA the mismatches caused a decrease in thermal stability of -6 to -18 °C meanwhile for DNA was -10 to -23 °C. R-GNA and S-GNA can create SS- and RR-homo-duplexes but cannot create SR- or RS-GNA double helices.

Similarly, S-GNA can pair with RNA but not with DNA or TNA [83]. R-GNA cannot pair with either DNA or RNA.

It is important to remark that in these studies the propylene glycol nucleosides derivatives (purines and pyrimidines phosphoramidites) were once again synthesized in a non-prebiotic manner using a method initially developed by Isis Pharmaceuticals that started from pure commercially available R- or S-glycidol [84, 85] and the oligonucleotides were obtained using **C**ontrolled **P**ore **G**lass (CPG) synthesis supports. *Hence, the question of whether the prebiotic synthesis of GNA-nucleosides(tides) following a "classic" synthesis is possible appears to have not yet been addressed*.

In a paper published in 2011 Hernández-Rodríguez and coworkers [64] described the organic synthesis, oligomerization and base pairing of a nucleic acid containing a glyceric acid with C2'-C3' phosphodiester backbone with either 2,4,5-triaminopyrimidine, 4-oxo-2,5-diaminopyrimidine or 2,4-dioxo-5-aminopyrimidine as RUs.

The authors found using temperature-dependent UV spectrometry that for the glyceric acid hexamers containing 2,4-dioxo-5-aminopyrimidine cross-pairing with hexamers of A-, dA-olygonucleotides and dodecamers of 2,6-triaminopurine is more versatile than the analog backbone derived from glycerol.

This glyceric acid backbone shows the following thermal stability order: DNA (UV-T_m = 63.6 °C) > RNA (UV-T_m = 44.4 °C) > TNA (UV-T_m = 24.0 °C). The pairing with DNA was stronger than the one for a (dA:dT) duplex hexadecamer (UV-T_m = 55.3 °C) and (A:dT) (UV-T_m = 59.3 °C). For the 4-Oxo-4,5-aminopyrimidine pentamers paired with a strand of DNA containing 2'-deoxythymidine the authors found a similar trend.

It is interesting to notice the similarity of the bases used in this study with the non-canonical TAP and BA.

Peptide nucleic acids (PNAs)

The building blocks of PNAs contain a backbone with amino acids (aa) instead of ribose. There are different types of PNAs:

1) aegPNAs discovered in 1991 by Nielsen and coworkers [69]. These PNAs have a polypeptide backbone containing repetitive units of N-(2-**A**mino**E**thyl)**G**lycine (AEG) bound through a carbonyl methylene group to the nucleobases.

2) Dipeptide **PNAs** (bPNAs) or nucleopeptides. They contain α-aa bound to the nucleobases through ethylene or poly-ethylene units.

3) **T**hioester **PNAs** (tPNAs) . tPNAs bind the nucleobases to a 59-cysteine dipeptide through thioester bonds.

4) **A**lanyl **N**ucleic **A**cids (ANAs). They were discovered in 1990 and contain alternative L- and D-alanines bound to the nucleobases through ethylene units.

5) αPNAs. This type of PNA has been obtained and their cross-pairing with DNA has been studied. αPNAs contain a L-serine-RU building block every 4 α-aa [70].

PNAs have become attractive as potential prebiotic TCs for the first nucleic acids due to many reasons. Some of these are mentioned below.

The experiments by Miller-Urey [14, 12, 13] have produced molecules that are precursors of aa-N^1-pyrimidines and aa-N^9-purines. These reactants include α, γ-diaminobutyric acid, glycolic acid, carbonic acid, purine and pyrimidine nucleobases and L-glycine.

The sugar-phosphate backbone in DNA/RNA is replaced in aegPNAs by a peptide bond which is more resistant to hydrolysis in aqueous solution, than the phosphodiester bond [70].

Many building blocks of the PNAs are natural products which mean that they may have been present in the first nucleic acids and then were replaced by the DNA and RNA building blocks through evolutionary selection, leaving only traces of them in nature. Examples of these include N-(2-aminoethyl)glycine, the component of aegPNAs that have been found in cyanobacteria strains [86]. It is considered that this bacteria emerged around 3.5 Gyr. Cyanobacteria can survive extreme environments as for example the geothermal vents, which is one of the proposed scenarios for the origins of life. Willardiine is an alanine-containing dipeptide with uracil as RU and has been isolated from the seeds of the *Acacia willardiana* [87]. Lupinic acid (β-[6-(4-hydroxy-3-methylbut-*trans*-2-enylamino)purin-9-yl]alanine) has been isolated from *Lupinus angustifolius' seeds* [88].

Many precursors of aa have been found in cosmic scenarios. More than 80 aa, of them 8 natural, have been identified in meteorites (e.g., diamino butyric acid in Murchisons meteorite [89]). These findings support the idea that maybe many of the PNAs' components came from outer space.

Other studies have shown that the ionization of icy mixtures similar to the ones in different types of interstellar clouds using differences sources or radiation can produce a vast spectrum of compounds, amount them aa, e.g., a study by Muñoz and coworkers [90] has identify the formation of 16 natural and non-natural aa by irradiating a $H_2O:CH_3OH:NH_3:CO:CO_2$ (2:1:1:1:1) ice mixture at 12 K. N-Ethyl-glycine, glycine and the D- and L- alanine, proline, valine, aspartic acid and serine were amount the aa identified from the mixture.

The prebiotic synthesis of the aegPNAs' components have been proposed in different studies, e.g., Nelson and coworkers [91] studied the prebiotic synthesis AEGs either by electric discharge experiments of a mixture of CH_4 (g), N_2(g), NH_3(g) and H_2O and by polymerization of NH_4CN. Even when the yields in % obtained for AEG through both methods was low (10^{-5}-10^{-6} %) the authors hypothesize that the production of this aa could have been possible in the early Earth due to the high solubility of its monohydrochloride and that AEG can be obtained through the Strecker synthesis, from the reaction of **E**thylen**D**iamine (ED) with HCN and H_2CO. ED can be obtained from a mixture of CH_4 and NH_3 with 33% of yield from the polymerization of NH_4CN and 10% H_2CO in 0.05% in the presence of UV radiation [92]. Nelson *et al*., estimate that low concentrations of the Strecker mixture (~10^{-6}M) can produce up to 33% of AEGs. Adenine-N^9-**Ac**etyl (Ac) and guanine-N^9-Ac were obtained from polymerization of HCN in the presence of glycine in 0.0062% and 0.011% respectively meanwhile, cytosine-N^1-Ac and Uracil-N^1-Ac were produced from the reaction between hydantoic acid with cyanoacetaldehyde at 100°C in 18% and 1.8%.

Nelson and coworkers have also found [91] that AEGs can polymerize in dry conditions at 100°C better than α-aa at higher temperatures .

PNAs single strands can create parallel and antiparallel double helixes through WC-like complementary base pairing with either DNA or RNA, being the anti-parallel preferred. PNA 8-mer strands containing only T and C can create PNA_2-DNA/RNA complexes using both WC and Hoogsteen complementary base pairing in a preferential parallel fashion. Additionally, PNAs with only pyrimidines can also pair to double stranded DNAs. The lack of charge in the aegPNAs backbone facilitates their cross-pair with DNAs or RNAs due to the absence of electrostatic repulsion forces involved [70, 93, 68, 94, 69].

The carbon atoms of the aegPNAs backbone lack chirality. This property may have given them an evolutionary advantage in the early Earth since, in contrast with today's nucleic acids, they

are not subjected to **E**nantiomeric **C**ross-**I**nhibition (ECI). ECI refers to the exclusive relationship between DNA/RNA biochemical functions and the D-enantiomerism of the ribofuranose or 2'-deoxyribofuranose. For example, studies by Joyce and coworkers [95] show that the template directed polymerization of DNA containing D-ribose can be inhibited when D-ribose-phosphate is replaced by the L-enantiomer in the building blocks of the D-homochiral template [74, 67].

1.2.3.3 Was phosphate the first ionized linker (IL)?

The functions of the phosphate group as IL are well known, but the plausibility of the prebiotic phosphorylation of nucleosides and nucleotides poses a challenge due to the "water problem" and the inertness of prebiotic minerals containing phosphorus (P), among other limitations [31].

If we consider RNA and DNA as a product of evolution then, perhaps P was not part of the proto-RNA [96].

Arsenic (As) has received attention as an attractive candidate to replace P in the first nucleic acids [96]. This idea is mainly based in three facts:

1. As and P have similar chemical properties and valence. In fact one of the secondary mechanisms of poisoning by As is its ability to replace phosphorus. As can replace P in bones and hairs which is used in forensic chemistry to establish As poisoning as the cause of fatality.
2. There are enzymes that can catalyze not only phosphate esters formation but also, arsenate esters, such as hexokinase [96].
3. Wolfe-Simon *et al*. has reported that the GFAJ-1 strain of the bacterium *Halomonadaceae* uses As instead of P to grow [97]. The authors of this paper found a predominance of As-based DNA over the P- counterpart in genomic samples extracted from the bacteria cells.

1.2.4 Stereoselectivity of the building blocks of nucleic acids

We may turn now to a fundamental question related to the stereoselectivity of modern building blocks. In today's nucleic acids, most (>90%) nucleosides(tides) exhibit a *β-* over and *α-* configuration at the C1' of the sugar (**Figure 1.7**).
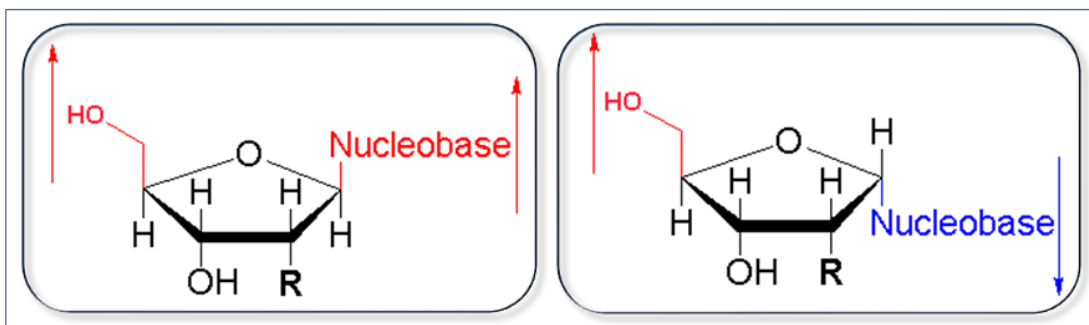
Figure 1.7 Orientation of the nucleobase at the C1' of the furanose with respect to the hydroxymethyl group at the C4' for the (left): β- and (right): α-anomers of the canonical ribonucleosides. The substituent R can be either H in 2'-deoxynucleosides (in DNA) or OH in ribonucleosides (in RNA) (taken from [98] and reprinted with permission of RSC Advances).

α-Nucleosides(tides) can be found as lesions induced by free radicals in natural DNA [99] or in natural sources [98, 100], *e.g.* α-nicotinamide adenine dinucleotide in *Azotobacter vinelandii* [101] and α-adenosine in *Propionibacterium shermanii* [102].

Another evidence for the possibility of α-NAs in a prebiotic pool, are the numerous reports on the laboratory synthesis of WC complementary α-α and α-β double stranded polynucleotides (see Figure 1.8 for some molecular models obtained from data in the literature) [103, 104, 105, 106, 107]. These double strands have multiple applications in medicine as anti-tumor, anti-bacterial and anti-malarial [98]. Paoletti *et al.* in 1989 studied the experimental synthesis of a α-β complex between the α-D(CCTTCC) hexanucleotide and its complementary β-D(GGAAGG). When the authors compared the formation of this complex with its natural β analog (β-D(CCTTCC) + β-D(GGAAGG)) they found that the formation of the non-natural complex was only 1 kcal/mol more favored than its natural counterpart [105], an energetic difference which is in the experimental error - in other words they are equally stable.

It is well known that 6-MR hexopyranoses, (e.g. galactopyranose, glucopyranose) adopt two preferential axial A_1 or $^1C_4$ and equatorial E_1 or $^4C_1$ chair conformations in equilibrium for their respective β- and α-anomers (see **Figure 1.9a**). The displacement of this equilibrium depends on the contribution of different stereoelectronic factors [108, 109, 110].
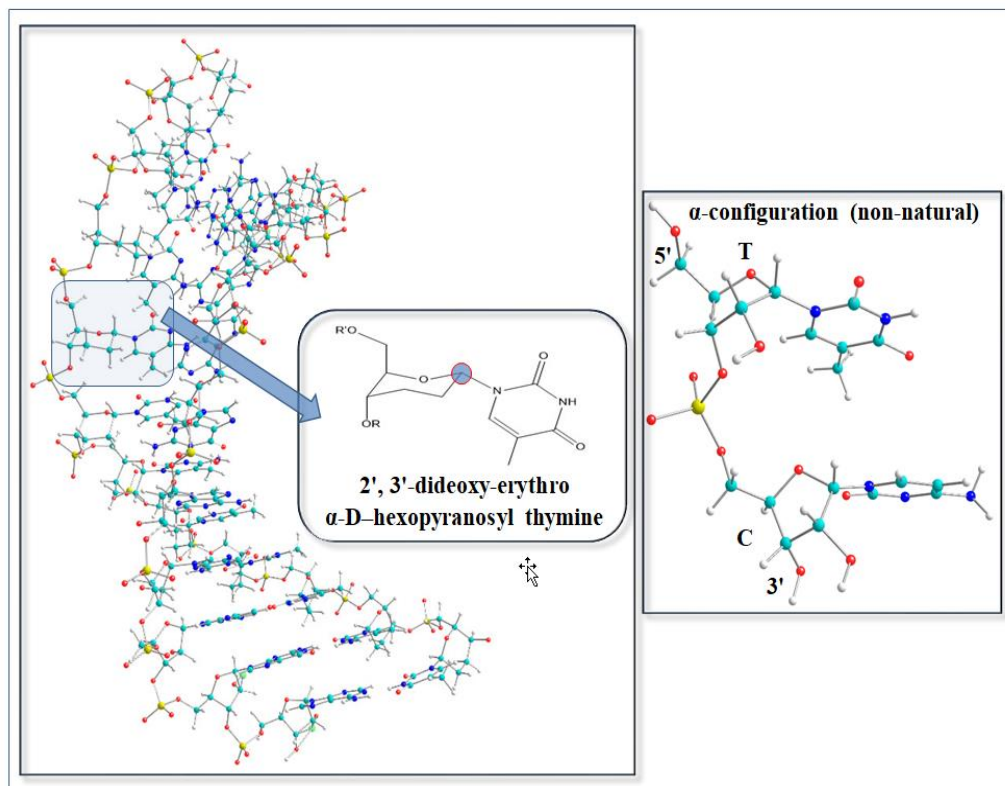
Figure 1.8 (Left): PDB structure for the WC base pairing between two strands; one with an α-homo-DNA (oligo (2', 3'-dideoxy-erythro α-D-hexopyranosyl thymine)$_{10}$) and the other with a β-RNA (oligo β-(riboadenosine)$_{10}$) sequence provided by Dr. Froeyen, M [103]. (Right): model constructed using the values for the torsion angles provided in Lancelot, G et al. [106] for an α-ribofuranosyl thymidine (5'-P-3') dinucleotide.

Following the same principle, Thibaudeau and coworkers (see pp. 22-46 in [111] and [112, 113]) proposed that the conformational changes of the 5-MR ribofuranose is an important factor to take into account when analyzing the stability of the β- and α-anomers of the building blocks of nucleic acids (see **Figure 1.9b**).

The furanose rings have different conformers that can be described by the generalized puckering parameters, or puckering coordinates, of sugar rings (changes in the endocyclic torsion angles of the furanose/pyranose ring) proposed by **C**remer and **P**ople (CP) in 1975 [115].
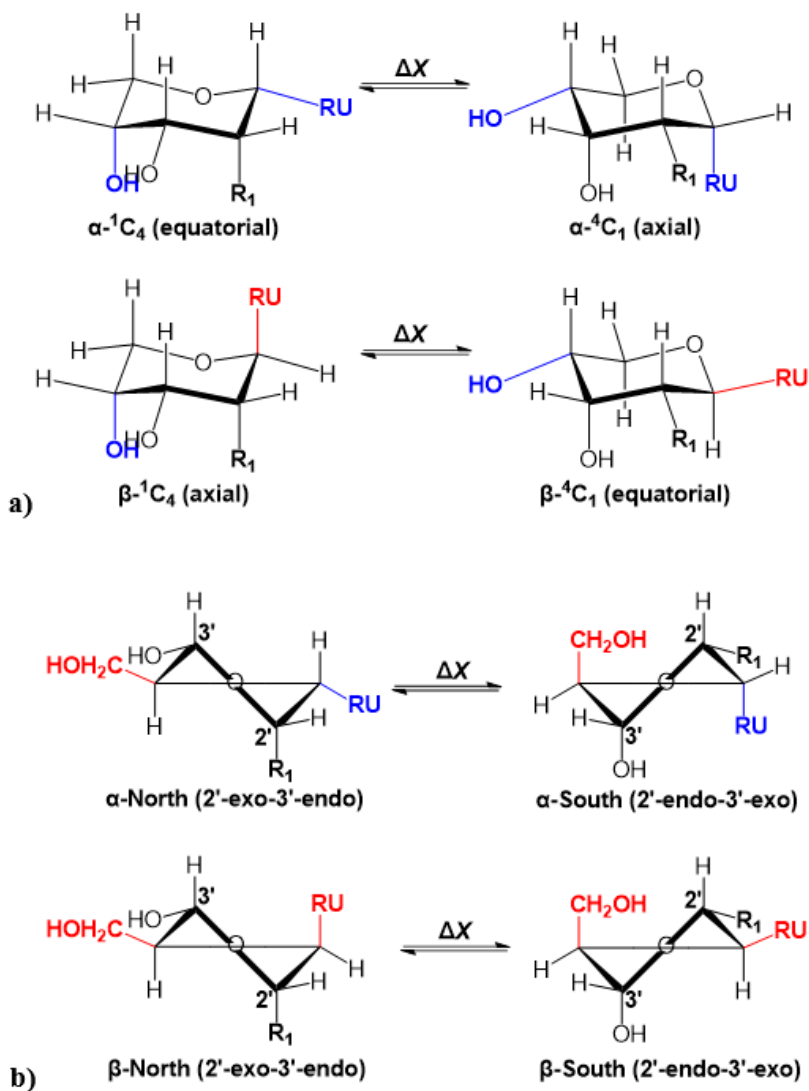
Figure 1.9 **a)** Equilibrium between the E1 ($^4C_1$) and A1 ($^1C_4$) conformations for a nucleoside with a 6-member ring sugar [114], **b)** equilibrium between the $^3T_2$ or (C2'-exo-C3'-endo) and $^2T_3$ or (C2'-endo-C3'-exo) conformations for a nucleoside with a 5-MR sugar (see pp. 8-46 in [111] and [112, 113]. ΔX: change in enthalpy ($\Delta H°$) or Gibbs free energy ($\Delta G°$) (R_1: H for 2dRibf and Tho, OH for Ribf) (taken from [112] and reprinted with permission of Elsevier).

In CP formulation the positions of the different atoms in the ring of size N are defined by the sets of cartesian coordinates x_j, y_j, z_j with position vector $\overrightarrow{R_j}$ that has origin (0, 0, 0)

$\left\{ \sum\limits_{j=1}^{N} R_j = 0 \right\}$ (**Figure 1.10**). x_j, y_j, z_j can be obtained from X-ray diffraction or theoretical data.

The next step involves determining the average mean plane defined by the vector ($\vec{n}$) (see equation 1.3) of the ring that is normal to the two vectors $\vec{R}'$ and $\vec{R}''$ (equations 1.1 and 1.2) (**Figure 1.10**):

$$\vec{R}' = \sum_{j=1}^{N} \vec{R}_j \sin\left(\frac{2\pi(j-1)}{N}\right) \tag{1.1}$$

$$\vec{R}'' = \sum_{j=1}^{N} \vec{R}_j \cos\left(\frac{2\pi(j-1)}{N}\right) \tag{1.2}$$

$$\vec{n} = \frac{\vec{R}' \times \vec{R}''}{\left|\vec{R}' \times \vec{R}''\right|} \tag{1.3}$$

The vector $\vec{n}$ is normal to the plane and defines the displacements from the mean plane (z_j) for each atom of the ring from can be calculated by using equation 1.4 (**Figure 1.10**):

$$z_j = \vec{R}_j \bullet \vec{n} \tag{1.4}$$

If the number of atoms N in the ring is odd then the puckering is defined by two puckering parameters: puckering amplitude (q_m) and phase angle (ϕ_m) that can be calculated by equations 1.5 - 1.8:

$$q_m \sin\phi_m = -|\sqrt{2/N}| \sum_{j=1}^{N} z_j \sin\left(\frac{2\pi m(j-1)}{N}\right) \tag{1.5}$$

$$q_m \cos\phi_m = |\sqrt{2/N}| \sum_{j=1}^{N} z_j \cos\left(\frac{2\pi m(j-1)}{N}\right) \tag{1.6}$$
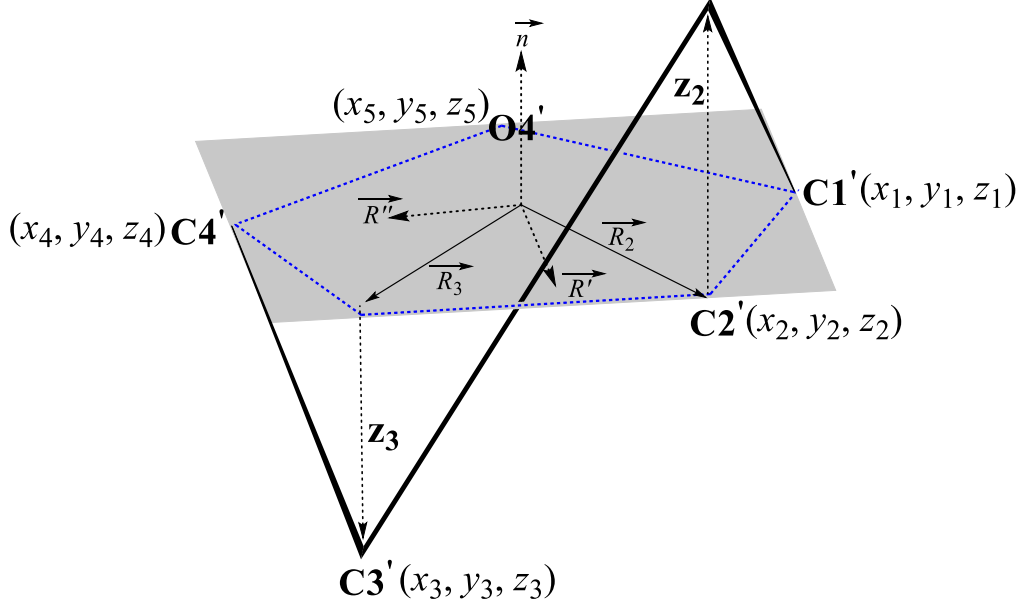
Figure 1.10   Representation of the position vectors $\vec{R}_j$, the mean square plane vectors $\vec{R}'$, $\vec{R}''$, $\vec{n}$ and the displacements z_j for the $^2T_3$ puckering of the 5-MR of the ribofuranose.

$$q_m = \sqrt{(q_m \sin \phi_m)^2 + (q_m \cos \phi_m)^2} \tag{1.7}$$

$$\phi_m = \tan^{-1} \left(\frac{q_m \sin \phi_m}{q_m \cos \phi_m} \right) \tag{1.8}$$

where m can be equal to 2, 3, …, $\frac{(N-1)}{2}$. In case for a 5-MR there ring conformation will be defined by q_2 and ϕ_2.

If N is even, m takes values up to $\frac{N}{2} - 1$ in equations 1.5 - 1.6 and there is a third amplitude $q_{\frac{N}{2}}$ that is calculated as:

$$q_{\frac{N}{2}} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} z_j (-1)^{j-1}} \tag{1.9}$$

In the case of a 6-MR the puckering parameters will be q_2, q_3 and ϕ_2. From these parameters the total puckering amplitude (Q) is:

$$Q = \sqrt{\sum_{j=1}^{N} z_j^2} = \sqrt{\sum_{m} q_j^m} \tag{1.10}$$

and for the case of the 6-MR the CP puckering coordinates can be converted to polar coordinates by determining a third phase angle theta (θ):

$$\theta = \cos^{-1}\left(\frac{q_3}{Q}\right) \tag{1.11}$$

Additionally, Thibaudeau et al. analyzed the enthalpic contribution to the free energy from the pseudorotation of the N-sugar compared to the S-sugar at different pK_a determining that the North-conformation possessed more flexibility and hence, was able to change to different conditions (pH and temperature) making it a better candidate for evolutionary selection.

The displacement of the equilibrium towards one preferential conformation is influenced by different stereo-electronic factors, e.g., the anomeric effects of the N-aglycone and the gauche effect) (see pp. 22-46 in [111] and [112, 113]).

A paper published in 2022 by Castanedo and Matta [116] explores the role of thermodynamics as a driver of evolutionary selection for the chemical structure of today's nucleic acids in implicit solvation models and vacuum-gas phase using computational modeling. This paper estimates the changes in different thermodynamical parameters for the selection of the *β*- over the *α*- anomers of D-**r**ibofuranose (r), D-2'-**d**eoxy**r**ibofuranose (dr), corresponding sugar monophosphates, nucleosides and nucleotides.

The results obtained by Castanedo and Matta [116] suggest not energetic difference beyond the intrinsic error of the computational methods between the *β*- and *α*-anomers of the D-**r**ibofuranose and D-2'-**d**eoxy**r**ibofuranose. A slightly thermodynamic advantage is estimated that favors the selection of the *β*- over the *α*- for some nucleosides and nucleotides (see Tables 2, 3 and 4 in paper).

1.2.5   Why did nature select T for DNA and U for RNA?

It is well known that usually **t**hymine (T) is present in DNA and **u**racil (U) in RNA. U differs from T by lacking the C5-methyl group. Deamination of C to U in DNA turns C:G base pairs in the U:G mutated pair. Present-day organisms contain a repair mechanism that fixes this mutation following three steps: 1) **U**-**N**-**G**lycosylase (UNG) removes any U detected to leave an abasic site, 2) the previous step is complemented by base excision that creates a gap in the opposite G, 3) the gap is filled with the C nucleotide by DNA polymerase..

Poole and coworkers [117] proposed that if it is assumed that early DNAs contained a mixture of T and U eventually evolution replaced U by T as a way to allow repair proto-enzymes to fix and eliminate U obtained from deamination of cytosine. If U is not replaced then the repair enzymes could be incapable to recognizing the mismatched base pairs because they would consider U as naturally occurring base. In the same paper the authors propose that evolution "tinkered" the role of key proto-enzymes to eventually replace U by T. In this direction a primitive form of UNG and uracil DNA-glycosylase (MUG) (a related enzyme to UNG) must have repaired the mismatched U:G and any unrepair pairs.

On the other hand, thymidylate synthase (an enzyme that synthesizes T from U) must have played a primordial role in the tunning of the genetic material.

 If we consider that proto-nucleic acids emerged in the absence of enzymes and that uracil-DNAs (U-DNA) can be also found in viral and bacterial genomes [118] the question on how and why U was replaced by T in DNA remains unsolved.

In a paper published in 1969 in the Journal of Theoretical Biology [119] Lesk presented plausible reasons for the advantageous replacement of U by T in DNA. The main argument presented by the author for this biological fact is the greater resistance of T to photochemical mutation compared with U. T nucleotides are the target for damaging radiation, so their higher resistance to this kind of mutation could have been a selection force that pushed nature to choose one base over the other in order to preserve the genetic information. Also, U but not T forms a stable photo hydration product. This means that the dimerization of T can be partially photo reversed by irradiation at relatively longer wavelengths (less energetic) but this process is less effective for U. These arguments reference UV light as a factor to consider in the evolutionary selection of T for DNA and U for RNA.

1.3 Computational chemistry for the modeling of the building blocks of nucleic acids

Computational chemistry or (theoretical chemistry) is the discipline that studies the construction of molecular models, predicts or explains their structural properties and obtains information from chemical and biological phenomena using the mathematical calculations implemented in the computational tools [120, 121].

1.3.1 Classification of the different methods

Computational chemistry has a wide range of methods and tools which can be divided into two broad categories based on their theoretical foundation:

- Classical methods, based on the laws of classical physics, do not consider the electrons explicitly. These methods simulate the presence of the electrons using parameterization from quantum calculations and experimental data. Usually, they consider the atoms as balls and the chemical bonds as springs. Because of this approximation, these methods do not allow interpretations or physical generalization of enzymatic reactions and the formation and/or breaking of chemical bonds [122]. Examples of these methods are **M**olecular **M**echanics (MM) and classical **M**olecular **D**ynamics (MD) [123, 124, 125, 126, 127, 128].

- **Q**uantum **M**echanical (QM) methods seek a solution of the time-independent non-relativistic Schrodinger equation [129] to describe the molecular systems through a direct treatment of the electronic structure. The quantum methods are further subdivided into several classes: semi-empirical, *ab initio* and the **D**ensity **F**unctional **T**heory (DFT) methods that use the electron density as its starting point. The DFT methods calculate the energy and system's properties from the information contained in the electron density [120]. There are also advanced *ab initio* molecular dynamics methods, such as those developed by Car and Parinello [130, 131, 132], whereby the forces are estimated at every

time step from quantum mechanical calculations. *Ab initio* MD methods are, of course, much more expensive than their classical counterparts alluded to above.

Quantum biochemistry and pharmacology emerged as a field that uses the concepts and methods of quantum mechanics to study the electronic structure and properties of biological systems and their involvement in biochemical and biophysical processes that are characteristic of living organisms [133, 134, 135].

Using any of these methods it is possible to calculate and explore the **P**otential **E**nergy hyper**S**urface (PES) of a given reaction.

The selection and use of the correct method and the interpretation of the theoretical results depend on the system at hand [37, 136, 121]. To study the (bio)chemical properties of NAs both types (classical and quantum mechanical) have been used depending on the size of the molecular system and the problem to be solved, but if we want to study the chemical properties of the building blocks of nucleic acids, QM would be the right choice. Using QM methods, it is possible to obtain more precise results about the structural information at a nano or pico scale and allows for a deeper understanding of the electronic effects because the electrons are considered explicitly.

Erwin Schrödinger, in 1944, was among the firsts to suggest that QM could solve the mysteries of the origins of life [137]. Since then, the world has seen an accelerated development of quantum and computational chemistry, molecular modeling, and MD (both with classical fields or using quantum mechanically calculated forces (known as ab initio MD)). Thanks to these methods, the *in silico* modeling of the prebiotic chemistry of ancestral NAs [138] is now at our fingertips. The same can be said about the theoretical investigations aiming to determine the physical chemical properties of contemporary DNAs and RNAs sequences [17, 139, 37, 140, 136, 141, 43].

Hence, in the next sections the QM methods will be described in more detail.

1.3.2 QM methods

The microscopic systems that present a corpuscular and undulatory behavior obey the laws of quantum mechanics. These laws were discovered by Heisenberg, Born and Jordan in 1925 and Schrödinger in 1926 [129].

The QM methods describe the molecules in terms of the interactions between the nuclei and the electrons of their atoms. In theory, QM can predict exactly any property of an atom or molecule. In practice, the QM equations have only been solved by finding their exact solutions for the mono electronic atoms or ions. For this reason, a large number of approximations (variational and/or perturbative) have emerged to find reasonable solutions for multielectron systems [142].

The majority of the QM methods are based on the time-independent non-relativistic Schrödinger equation, within the Born-Oppenheimer approximation:

$$\hat{H}\psi(r, R) = E\psi(r, R) \tag{1.12}$$

where $\hat{H}$ is the Hamiltonian operator for a system of nuclei and electrons with special coordinates R and r respectively, expressed in general units as:

$$\hat{H} = -\sum_{i=1}^{N}\frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_{\alpha=1}^{M}\frac{\hbar^2}{2m_\alpha}\nabla_\alpha^2 - \underbrace{\sum_{i=1}^{N}\sum_{\alpha=1}^{M}\frac{1}{4\pi\epsilon_0}*\frac{Z_\alpha|e|}{|R_\alpha - r_i|}}_{\text{electron-nuclear attraction}} + \underbrace{\sum_{i=1}^{N}\sum_{i>j}^{N}\frac{1}{4\pi\epsilon_0}*\frac{|e|^2}{r_{ij}}}_{\text{electron-electron repulsion}} + \underbrace{\sum_{\alpha=1}^{M}\sum_{\beta>\alpha}^{M}\frac{1}{4\pi\epsilon_0}*\frac{Z_\alpha|e|^2}{R_{\alpha\beta}}}_{\text{nuclear-nuclear repulsion}} \tag{1.13}$$

$$\underbrace{\phantom{-\sum_{i=1}^{N}\frac{\hbar^2}{2m_e}\nabla_i^2}}_{\substack{\text{electronic}\\\text{K.E}}} \underbrace{\phantom{\sum_{\alpha=1}^{M}\frac{\hbar^2}{2m_\alpha}\nabla_\alpha^2}}_{\substack{\text{nuclear}\\\text{K.E}}}$$

where K.E represents the **k**inetic **e**nergy, N is the number of electrons, $\hbar$ is the Dirac constant, m_e is the atomic mass of the electron, ∇_i^2 is the Laplacian of the electron i, ∇_α^2 is the Laplacian of the nucleus α, m_α is the mass of the nucleus, M is the number of nucleus in the molecule, Z_α is the nuclear charge, e is the electron charge, the $|R_\alpha - r_i|$ term represents the distance between the nucleus α and the electron i, r_{ij} is the distance between two electrons i and j and $R_{\alpha\beta}$ is the distance between two nuclei α and β.

Which upon adoption of **a**tomic **u**nits (a.u.) (mass of the electron = elementary charge = $\hbar$ = electric constant $\frac{1}{4\pi\epsilon_0} = 1$), equation (2.13) has the simpler appearance:

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{M}\frac{1}{2m_\alpha}\nabla_\alpha^2 - \sum_{i=1}^{N}\sum_{\alpha=1}^{M}\frac{Z_\alpha}{|R_\alpha - r_i|} + \sum_{i=1}^{N}\sum_{i>j}^{N}\frac{1}{r_{ij}} + \sum_{\alpha=1}^{M}\sum_{\beta>\alpha}^{M}\frac{Z_\alpha}{|R_\alpha - R_\beta|} \tag{1.14}$$

The wave function (which is a real-space representation of the more general state vector) $\Psi(r, R)$ depends *explicitly* on the spatial coordinates of electrons (r) and nuclei (R) and the spin "coordinates" components of the electrons (ω) [143]. This wave function contains all the information that can be known about the system. Such information is "extracted" from the wave function through the application of linear Hermitian operators. The solutions of the equation 1.1, which applies one such operator (the Hamiltonian operator), are the numerical values or eigenvalues or **e**nergies (*E*) that represent the energy of the stationary state described by ψ [142].

The application of the **variational principle** is extremely useful. It has a primordial role in all quantum-chemical applications to systematically approach the exact wave function of the ground state ψ. For these states the expected value from $\hat{H}$ for a well-behaved arbitrary wave function is always greater than the lowest eigenvalue E_0 of $\hat{H}$ corresponding to the exact solution of the Schrödinger equation [144].

The quantum kinetic energy operator (in general units) is:

$$T = -\frac{\hbar^2}{2m}\nabla^2 \tag{1.15}$$

where ∇^2, the Laplacian, is defined (in Cartesian coordinates) as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.16}$$

Since in the K.E operator, mass appears in the denominator, the heavier the particle the lesser its kinetic energy, or in classical terms the slower its movement. Since even the lightest nucleus, that of a hydrogen atom, namely a proton, is 1,838 times heavier than the electron, it is an excellent approximation to consider that nuclei move at a much slower timescale than nuclei.

This is the essence of the **B**orn-**O**ppenheimer (B-O) or adiabatic approximation [145]. With this approximation, a considerable simplification is made to solving the molecular problem by separating the nuclear problem from the electronic one. Hence, it is possible to simplify the

mathematical expression for $\widehat{H}$, ignoring the components related with the kinetic energy of the nuclei and considering constant the potential energy of the nucleus-nucleus repulsion.

This is tantamount to writing the total wavefunction as a product of an electronic wavefunction and a nuclear one ($\psi_{\text{overall}} \approx \psi_{\text{nuclear}}\psi_{\text{electronic}}$), that is, mathematically separating the mathematical problem into two independent ones. In doing so, the Schrödinger equation of interest to us, namely the one describing the motion of the electrons, is expressed in function of the electronic Hamiltonian $\widehat{H}_{\text{e}}$ (the first, third, and fourth terms in expressions (a) or (b) above):

$$\widehat{H}_{\text{e}}\,\psi_{\text{e}}(\mathrm{r},\mathrm{R}) = E_{\text{e}}\psi_{\text{e}}(\mathrm{r},\mathrm{R}) \tag{1.17}$$

where the "electronic" Hamiltonian delivering the purely electronic energy (E_{e}) is now defined as:

$$\widehat{H}_{\text{e}} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{\alpha=1}^{M}\frac{Z_\alpha}{|R_\alpha - r_i|} + \sum_{i=1}^{N}\sum_{i>j}^{N}\frac{1}{r_{ij}} \tag{1.18}$$

in which the nuclear-nuclear kinetic energy term has been eliminated and so is the final term in equations (1.13) and (1.14) describing the nuclear-nuclear repulsion term.

That latter term, $\sum_{a=1}^{M}\sum_{\beta>a}^{M}\frac{Z_a}{R_{a\beta}}$ (in a.u.), is constant for a given geometry and is added at the end of the calculation to the electronic energy obtained from 1.17 to obtain the total molecular energy (simply referred to as "total energy"):

$$\widehat{H}\psi = \left(-\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{\alpha=1}^{M}\frac{Z_\alpha}{|R_\alpha - r_i|} + \sum_{i=1}^{N}\sum_{i>j}^{N}\frac{1}{r_{ij}} + \sum_{\alpha=1}^{M}\sum_{\beta>\alpha}^{M}\frac{Z_\alpha}{R_{\alpha\beta}}\right)\psi = E\psi \tag{1.19}$$

(Note that any eigenfunction ψ - a many electron wavefunction of $\widehat{H}_e$ is also an eigenfunction of $\widehat{H}_e$ since the two operators differ only by a constant term, that is, the nuclear-nuclear repulsion term).

Equation 1.17 cannot be solved for multielectron systems because of the presence in the Hamiltonian of the electron-electron repulsion term. One of the great simplifications to solve this problem is the molecular orbital approximation where the many electron wavefunction is expressed as a product of one-electron functions called "orbitals" as done in the Hartree-Fock approximation.

This latter approximation accounts for the average electron-electron repulsion but does not account for the instantaneous correlations of electrons due to their coulombic repulsion. This is the base for the **H**artree-**F**ock (HF) method [144]. In seeking a variational solution to the HF problem one imposes anti-symmetry on the form of an acceptable wavefunction to satisfy Pauli antisymmetry principle for an ensemble of fermions such as electron. This, in the single determinant case, is imposed by restricting the form of an acceptable solution to what has become to be known as a Slater determinant, that is:

$$
\begin{aligned}
\psi(x_1, x_2, \dots, x_N) &= \frac{1}{\sqrt{N!}}
\begin{vmatrix}
x_1(\mathrm{x}_1) & x_2(\mathrm{x}_1) & \cdots & x_N(\mathrm{x}_1) \\
x_1(\mathrm{x}_2) & x_2(\mathrm{x}_2) & \cdots & x_N(\mathrm{x}_2) \\
\vdots & \vdots & \ddots & \vdots \\
x_1(\mathrm{x}_N) & x_2(\mathrm{x}_N) & \cdots & x_N(\mathrm{x}_N)
\end{vmatrix} \\
&\equiv |x_1, x_2, \dots, x_N\rangle \\
&\equiv |1, 2, \dots, N\rangle
\end{aligned}
\tag{1.20}
$$

where N is the total number of particles, $\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_N$ are the coordinates of particles with wave functions $x_1, x_2, \dots x_N$.

The solutions for the HF equations are found using an iterative procedure known as **S**elf-**C**onsistent **F**ield (SCF) and considering the basis sets orthonormal. In general, this method uses initially a test orbitals set to solve the HF equations. The set of resulting orbitals is used then in the new iteration until the value for the energies for the initial and resulting orbitals is less than a threshold value previously defined. This value is generally very low (~ 0) [146, 147].

The iterative procedure developed by HF has been the starting point to obtain the energy of the system in all the post-HF and DFT calculations.

1.3.3   Electronic correlation and self-consistent field calculations

The electrons repel each other in any multi-electronic system such as an atom, a molecule, or a crystal. The electronic movements are correlated in the way that the probability to find two electrons in the same point of space must be null. There are two types of correlation, one due to the tendency of same-spin electrons to avoid one another (the exchange, Fermi, or Pauli correlation [148]) and this type is entirely accounted for in Hartree-Fock theory by the imposition of the Slater

determinant form on the acceptable solutions, and the other is missed in HF theory and that is the instantaneous correlation due to the coulombic electrostatic repulsion between electrons since they are all negatively charged (see pp. 19-27 in [149]). The Hartree-Fock approximation considers this latter type of (Coulombic) correlation only in an average way where each electron is moving in an orbital (occupying a region of space specified by a one -electron function) in the average field generated by the remaining $(N - 1)$ electrons.

The correlation energy (E_{corr}) for a given electronic state is the difference between the exact energy (E_0) and that obtained from a given calculation. If this calculation is of the Hartree-Fock type, say, then calling the energy it delivers is E_{HF}, the correlation energy is defined [150, 151]:

$$E_{corr} = E_0 - E_{HF} \tag{1.21}$$

As a result, the electrons are considered to be closer than they may actually be and the E_{HF} is estimated to be higher than the exact energy E_0 and the correlation energy is always negative. In order to correct this energy it is important to use other methods known as "**p**ost-**H**artree-**F**ock" (post-HF) that consider the electronic correlation. These methods can be perturbational [152], or variational [147]. An elegant way of accounting for coulombic correlation is by using the variational **D**ensity **F**unctional **T**heory (DFT).

1.3.4   Density functional theory (DFT)

QM uses the wave function as the central variable to access all the information that can be obtained from a particular system. But this approach has a significant disadvantage; it cannot yield the exact wave function for a multielectron system.

There is an alternative solution for this problem. It has been proved that the electronic density $\rho_0(r)$ can be measured experimentally using **X**-**R**ay **D**iffraction (XRD). This technique provides all the necessary information to write the Hamiltonian for the molecular system in function of $\rho_0(r)$. This is the DFT method's main core idea.

There are additional advantages to use DFT: 1) they consider the electronic correlation as implicitly in the theoretical calculations, 2) offer and excellent balance between computational cost

and the precision for the predictions compared to another more expensive methods with similar exactitude as for example the post-HF [153, 154].

This theory is based on the **H**ohenberg-**K**ohn (HK) theorems (see pp. 33-39 in [149] and [155]). The first theorem has been given an *ad absurdum* mathematical proof in 1964 and states that for electronic systems in a non-degenerate ground state, the electron density $\rho_0(\vec{r})$ determines an external potential $V_{\text{ext}}(\vec{r})$. This statement proves that the $\rho_0(\vec{r})$ is also the multi-electronic Hamiltonian.

The electronic energy for the fundamental state E_0 is a "functional" of the $\rho_0(\vec{r})$ function, where the energy of the ground state can be written as a functional of the electron density as:

$$E_0[\rho_0] = T[\rho_0] + E_{ne}[\rho_0] + E_{ee}[\rho_0] \tag{1.22}$$

Where $T[\rho_0]$ refers to the kinetic energy of electrons, E_{ne} is the potential energy of the nuclei-electrons attraction and E_{ee} is the potential energy of the electronic repulsion.

HK postulated in their second theorem that the lowest energy that can be obtained for the ground state of a molecular system is the one that corresponds to the real electronic density of this state [156].

The DFT methods have been successful because they have implemented the **K**ohn and **S**ham (KS) formalism (see pp. 41-59 in [149]) and [157, 158]. This approximation offers a practical way to obtain E_0 from ρ_0. It assumes that the kinetic energy of the electrons can be calculated using a basis set (KS orbitals) that deliver a fictitious non-interacting density which is equal to the exact electron density. The Hamiltonian $\left(\widehat{H}_{\text{KS}}\right)$ is then reformulated according to the equation 2.23 (expressed in a.u.):

$$\widehat{H}_{\text{KS}} = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 + V_{\text{S}}(\vec{r}_i) \tag{1.23}$$

where the first term inside the summation represents the kinetic energy operator for the electrons and $V_{\text{S}}(\vec{r}_i)$ is the external potential over the nuclei. (The external potential is named as such since it is the potential external to the system of electrons itself, in this case the nuclear potential in addition to the potential associated with any externally applied electrostatic or electromagnetic field(s)).

The $\hat{H}_{KS}$ operator describes a system of independent electrons by definition and hence there is no description of the interelectronic interactions. The connection between this artificial system and the real one is established choosing the effective potential $V_S(\vec{r}_i)$ in a way that the simulated electronic density is exactly equal to the ρ_0 of the real system. In this way the solutions for the equation 1.23 are going to be:

$$E[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.24}$$

where, $T_S[\rho]$ can be calculated exactly for the system of non-interacting particles, $J[\rho]$ is the coulombic potential from the electron-electron interactions, $E_{ne}[\rho]$ is the nuclei-electrons interaction term and $E_{xc}[\rho]$ is the correlation-exchange term, which is unknown. This term includes all the properties that cannot be calculated exactly and is developed using a series of different approximations. This search for the best xc functional remains one of the most active research areas of DFT. Once this term is established the energy is minimized keeping the Kohn-Sham orbitals orthogonal.

One of the most active research fields in DFT is the search of the best expression for $E_{xc}[\rho]$ that can predict more accurate results.

Many approximations have been used to define this term and one of the more promising is the use of the **H**ybrid-**G**eneralized **G**radient **A**pproximation (H-GGA). The functionals that are implemented using this approximation contain the conventional interchange-correlation energy from the **G**eneralized **G**radient **A**pproximation (GGA) methods with a percentage of Hartree-Fock exchange [156].

In the present one of the most popular hybrid functionals is the B3LYP. It was formulated by Becke and Lee-Yang-Parr in the late 80s [159, 160, 161, 162]. This functional is very similar to the Becke, 3-parameters (B3) hybrid exchange functional but the correlation functional by Perdew-Wang (PW91) is replaced by the LYP functional. The empirical parameters are the same as for B3.

The expression for B3LYP is:

$$E_{XC}^{B3LYP} = (1-\alpha)E_X^{LSDA} + aE_{XC}^{HF} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LSDA} \tag{1.25}$$

where E_X^{LSDA} is the exchange energy from the **L**ocal-**S**pin-**D**ensity **A**pproximation (LSDA), E_{XC}^{HF} is the exchange-correlation from HF, E_X^{B88} is the exchange energy from the Becke-88 functional, E_C^{LYP} is the correlation energy from the LYP functional, E_C^{LSDA} is the exchange of energy from the LSDA. $(1-\alpha)$, a, b, c and $(1-c)$ are the weight factors for each of the components or terms in the equation 1.25.

B3LYP has become so popular because of the number of properties that is able to accurately predict. Despite its success it is important to remark that these are not exact calculations. One of the B3LYP disadvantages is its inefficiency in describing **v**an **d**er **W**aals (vdW) or middle range interactions (~2-5 Å), limiting its application for the theoretical study of biological systems where the dispersive energies do not play a determinant role [163, 164]. Despite this disadvantage, this functional predicts efficiently geometries and energies with a major electrostatic contribution, e.g., (strong) hydrogen bonds [165, 166].

Specifically related to the computational modeling of the building blocks of nucleic acids and prebiotic chemistry reactions using DFT functionals, Šponer and coworkers modeled at the DFT (B3LYP/6-31++G (d,p)) level of theory the synthetic pathway for the β-cytidine proposed by Powner and coworkers in 2009 [23]. The reactions were also carried out using the dielectric **C**onductor **P**olarized **C**ontinuum **M**odel (CIEFPCM) to reproduce the dielectric constant of water in the environment of the different molecules. The authors estimate the standard free energies of Gibbs (ΔG°) for each step of the reaction sequence and compared the overall ΔG° with the ΔG° obtained for the classical pathway for the synthesis of cytidine. Finally, the authors analyze the catalytic role of the phosphate ions in the kinetics of the alternative route studied [167].

More recently Kaur and coworkers have studied using computational tools (MD and DFT) the β- and α-ribonucleosides of the non-canonical nucleobases 2,4,6-**T**ri**a**mino**p**yrimidine (TAP), **C**yanuric **A**cid (CA) and **B**arbituric **A**cid (BA), **M**ela**m**ine (MM) respectively [168, 169]. They also analyze the complementary base pairing TAP:CA and BA:MM, their π-stacking energies and de-glycosylation barriers.

Based on their modeling results Kaur and coworkers proposed that the strength for the hydrogen bonds created during the pairing of TAP with CA is comparable to the canonical base pairing. The stacking of these non-canonical bases is on the other hand weaker compared to the canonical stacking, suggesting that the enhanced stacking may have served as a driving force in

the evolution of nucleic acids. The deglycosylation barriers suggested that the glycosidic bond in the TAP nucleosides is stronger compared to the canonical nucleosides. For the case of cyanuric acid the opposite result was obtained. Finally the packing of the TAP-CA-containing helices suggested that this type of pre-RNA could have been shielded from water allowing its evolution and self-replication.

Similarly, in their paper published in 2017 [169] Kaur and coworkers found that the strength of the hydrogen bonds between BA and MM make them good candidates as components for the building blocks of ancestor NAs. On the other hand, it was determined that the stacking interactions were stronger when BA or MM were combined with a canonical nucleobase than when the stacking was between each other. These results suggest the possibility of the existence of a pre- or proto-RNA with canonical and non-canonical nucleobases. The authors found larger deglycosylation barriers for the C-C glycosidic bond of BA-ribonucleosides compared to the case of canonical nucleobases. This result is a theoretical indicator that the stability of this bond could have influenced the synthesis of NAs containing BA in extreme environments. For melamine lower deglycosylation barriers are reported.

1.3.5   Semiempirical methods

Semiempirical methods implement a set of approximations with the end goal of decreasing the computational cost making the calculations faster compared to the *ab-initio* methods. The use of these methods also allows the theoretical study of molecular systems with a greater number of atoms [170].

Some semiempirical methods use the **Z**ero **D**ifferential **O**verlap approximation (ZDO). ZDO establishes that the common spaces of the **A**tomic **O**rbitals (AO) for different atoms are so small that the values for the corresponding integrals are dismissed. This approximation reduces or makes null the repulsion integrals [146].

The semiempirical methods only consider the valence electrons explicitly. The internal electrons are calculated modifying the nuclear charge as a resultant charge (core or Z' charge) or introducing functions that model the core-electrons repulsion. The atomic basis set is considered implicitly. If in some cases it is necessary to make the basis set explicit then it is considered as a minimum Slater basis set [171].

The advantages of using these methods are the following: 1) there is not **B**asis **S**et **S**uperposition **E**rror (BSSE), due to the built-in orthonormality of the virtual basis used, 2) they consider the electronic correlation, not included in the HF methods, 3) they allow an accurate description of long range interactions and big size molecular systems, which is too expensive for *ab-initio* methods [172].

The biggest limitation for the use of *a posteriori* parametrization is that there is no systematic way to improve the calculations because the levels of theory do not advance systematically [173].

The semiempirical methods have different levels of theory. One of them is the **N**eglect of **D**iatomic **D**ifferential **O**verlap (NDDO) in which monocentric superpositions are considered but the bicentric integrals are neglected [170, 174]. Among the methods that have been developed using NDDO are AM1 (**A**ustin **M**odel 1) [175, 176, 177, 178, 179], PM6-D (**P**arametric **M**ethod 6 **D**ispersive) [175, 176, 180, 177] and PM7 (**P**arametric **M**ethod 7) [175].

The semiempirical Hamiltonian PM7 has further improvements upon PM6 as it adds explicit terms to describe non-covalent interactions, based on the ideas from the PM6-DH2, -DH+, and -D3H4 corrections done by Hobza and coworkers [176]. It has aimed at giving better results for molecules different than those in the training set. PM7 additionally corrects two minor errors in the NDDO formalism. The **A**verage **U**nsigned **E**rrors (AUE) in bond lengths decreases by about 5% and that in enthalpies of formation (ΔH_f) by about 10% in comparison to PM6 for organic solids [176].

1.3.6 Generation of random molecular arrangements & conformations: The multiple minimum hypersurface (MMH) methodology

The **M**ultiple **M**inimum **H**ypersurface (MMH) method [181, 182, 183] allows the evaluation of the collective environmental effects over a given molecule (or cluster of molecules) and for its/their systematic variation of conformation. This is implemented in the software GRANADA or its modification GRANADAROT [181]. The final goal is to obtain the energies and thermodynamical properties for the complexation of *n* molecules or for the *n* conformers using statistical thermodynamics [184, 183].

This methodology was initially used to theoretically study solvation processes of molecules through the obtaining and analysis of their solvation energies for the most stable solute-water clusters [185, 183].

MMH generates a random set of molecular arrangements for the interacting molecules or generates a random set of conformers for a given flexible molecule changing selected internal degrees of freedom, e.g. torsion angles. The random molecular arrangements are called "cells". The cells are rearranged to minimize the internal energy. The energies of a set of local minima (stable arrangements or conformers) can be converted into a representative canonical ensemble with fixed number of particles, volume, and temperature by the use of Maxwell-Boltzmann statistics [183]. In this way MMH performs a random exploration of the PES through which the most stable molecular structures are selected. The optimization of the cells is performed using semiempirical methods [186, 187].

Statistical thermodynamics links the physical-chemical properties of the "macro-" and the "micro-world" using the partition function. This function represents the statistical summation of all the states accessible by the studied system and from it one can obtain thermodynamical properties (e.g., internal energy, entropy, and association energy) [183].

The reference state is assumed to have the same number of non-interacting molecules. In this way the association process is considered as isothermic, meaning that all the other degrees of freedom (rotational, vibrational and translational) are fixed identically in the complexes or conformers as in the individual molecule(s). The summation of the total energies for the monomers is considered the reference value [182]. From standard statistical mechanics, the partition function for the ensemble of a given complex or flexible molecule reads as:

$$Z = \sum_{i=1}^{n} e^{-\frac{E_i}{k_{\mathrm{B}}T}} \qquad (1.26)$$

where i is the label of a given configuration out of the n unique ones, E_i is its corresponding total energy at a given computational method, and where the probability of formation of the ith complex is given by:

$$p(i) = \frac{e^{-\frac{E_i}{k_B T}}}{Z} \tag{1.27}$$

The Helmholtz free energy (A) and the entropy (S) are then obtained straightforwardly:

$$A = -RT \ln Z \tag{1.28}$$

$$S = R\ln Z + \frac{E_i}{T} \tag{1.29}$$

whereby k_B (Boltzmann constant) is replaced by R (gas constant) in J.K^{-1}.mol^{-1} to switch from molecular to molar ($R = k_B N_A$), N_A is the Avogadro constant and T is the temperature.

As a result, two types of degenerate situations can be found:

✓ Cells with the same energy and the same molecular structure. This type of degeneracy does not provide information to the partition function [182, 183].

✓ Cells with the same energy but different molecular structures. This degeneracy is relevant for the partition function [187].

Since these two types of degeneracies can occur in an optimization, the Q3 (program implemented in GRANADA that performs the statistical thermodynamics) uses the Tanimoto index ($S_T(A, B)$) [188, 189] that is defined by equation 1.30 for two molecules A and B with different fingerprints:

$$S_T(A, B) = \frac{c}{[a + b - c]} \tag{1.30}$$

where for molecule A with a number of *on* bits {amount of 1 in a fingerprint} and molecule B with b number of *on* bits, $c = a \cap b$ {number of shared bits between both A and B}.

The Tanimoto index is useful to analyze the similarities between the different molecular assemblies and neglects degeneracies that do not contribute to the partition function. In this way pair structural similarities are calculated, comparing the molecular complexes or conformers with an energy lower than 0.001 eV. When these pairs are located Q3 fixes a limiting value to

discriminate if the cells are geometrically similar or not. The most commonly used value for the Tanimoto index is usually 0.85. This threshold was initially determined by Patterson and coworkers [190] in which the Tanimoto index was evaluated for 20 molecular databases and there was a proven correlation between compound similarity and biological activity when $S_T(A, B) \geq 0.85$. Since then that value has been generally quoted/used in the literature.

In this way the molecular assemblies can be discriminated as follows:

- T > 0.85 the molecular assemblies are considered geometrically similar.
- T < 0.85 the molecular assemblies are geometrically different. This is the case of degeneracy of second type that contributes to the partition function.

The MMH has been successfully used for the theoretical evaluation of different biochemical systems as for example: molecular clusters of water [182], amylose-organometallic compounds molecular association [187] and solvation of brassinosteroids [185].

1.3.7   Solvation models in theoretical calculations

Solvents play an important role in numerous chemical and biological processes, e.g., enzymatic reactions, transmembrane transport, binding of ligands in the active site of receptors amount others.

Computational methods analyze solvation effects considering the solvent molecules as [191, 192, 193]:

- Implicit: reduces the complexity of the system by considering the discrete solvent as a continuum which tends to reduce the computational cost of the theoretical calculation. Examples of these methods include the **G**eneralized **B**orn (GB) implicit-solvent model used in Molecular Dynamics and MM force field modeling and the **P**olarized **C**ontinuum **M**odel (IEFPCM), which is one of the most popular **Q**uantum **M**echanics **S**elf-**C**onsistent reaction **F**ield methods (QM-SCF).
- Explicit: considers the solvent molecules explicitly in the model and computes all possible solute-solvent and solvent-solvent molecular interactions. Examples of these methods include the **P**article **M**esh **E**wald (PME) that considers the solute-solvent system as an infinite crystal.

The modeling of solvation using either type of approximation depends on the scientific problem at hand.

From one side the models that treat solvation implicitly are less expensive computationally because they treat the solvent as a continuum dielectric. This approximation can be useful in QM methods since the size of the system is reduced to exclusively the solute [193], e.g., some studies using implicit solvation in MM simulations have shown a reduction in computational times that makes the calculation ~ 2-20 times faster [194].

At the same time this approximation introduces a disadvantage making these methods not suitable if the goal is to model, e.g., intramolecular interactions in water molecules clusters and intermolecular interactions between solvent and the solute [37, 136].

On the other hand, the methods that treat solvation explicitly can give more accurate descriptions of the non-covalent interactions that dominate solvation but make the calculations increasingly expensive, making them difficult to carry-on for QM calculations [191].

1.3.7.1 Continuum implicit solvation methods

The solvation methods that model the solvent as a continuum dielectric were created in the 1980's by Tomasi's group [195, 196, 197, 198, 199, 200]. These methods assume that the solute (molecule to be modeled) is positioned in an artificial "cavity" that is defined by adding to the vdW radius of a molecule a probe or sphere of certain radius, e.g., 1.4 Å for the water radius. The main idea is to place this solute around a continuous dielectric medium to model the effects of polarization by the solute on the "invisible solvent" and vice versa. This is also called the self-consistent reaction-field (SCRF) problem (**Figure 1.11**).

The dielectric constant is defined as the electric permittivity of a solute in a given solvent related to vacuum. This value ranges from ≈ 2 for organic solvents to 78 for water.

The main idea behind the continuum dielectric formalism is that the polarization of the medium (solvent) by the charge distribution of the solute in the cavity can be quantified. This quantification is given by the dipole density $P(\vec{r})$ that replaces the electrostatic field $E(\vec{r})$ from vacuum:

$$D(r) = \varepsilon(\vec{r}) * E(\vec{r}) \tag{1.31}$$
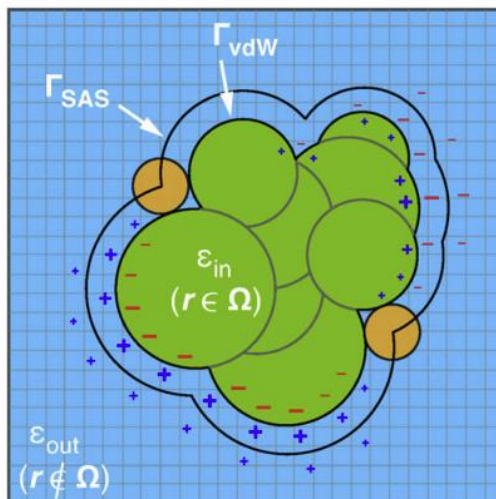
Figure 1.11 Solvent cavity (green) in solvent environment (blue). Ω: interior of cavity, Γ_{SAS}: Solvent Accessible Surface (SAS) radius, Γ_{vdW}: vdW radius and the orange sphere represents the probe that is used to add the Γ_{SAS} to the Γ_{vdW} (taken from [201] and reprinted with permission of Wiley Interdisciplinary Reviews).

where $\vec{r}$ is the atoms coordinates, $\varepsilon(\vec{r})$ is the electric permittivity that accounts for the amount of polarization P induced by the electrostatic field E and $D(\vec{r})$ is the electric displacement field.

Since $\varepsilon(\vec{r})$ is a scalar function it can take different values depending on what region of the system is evaluated. For instance, inside the cavity $\varepsilon = 1$ because all electrostatic interactions are included in the Hamiltonian operator. In the region outside the cavity permittivity is equal to the permittivity of the solvent ε_s.

From the first Maxwell's equation (Gauss's law, electric) [202], the relationship between the charge density $\rho(\vec{r})$ and the gradient of the displacement field $D(\vec{r})$ through equation:

$$\nabla D(\vec{r}) = 4\pi\rho(\vec{r}) \tag{1.32}$$

Substituting the expression of $D(r)$ in equation 1.31) in the previous equation 1.32:

$$\nabla\left(\varepsilon(\vec{r}) * E(\vec{r})\right) = 4\pi\rho(\vec{r}) \tag{1.33}$$

$E(\vec{r})$ is equivalent to minus the gradient of the electrostatic potential $-\nabla V(\vec{r})$:

$$E(\vec{r}) = -\nabla V(\vec{r}) \tag{1.34}$$

where:

$$\nabla V(\vec{r}) = \frac{\partial V}{\partial x} x + \frac{\partial V}{\partial y} y + \frac{\partial V}{\partial z} z \tag{1.35}$$

$E(\vec{r})$ can be replaced by $-\nabla V(\vec{r})$ in equation 1.33 to obtain:

$$\nabla \left[\varepsilon(\vec{r}) * \nabla V(\vec{r}) \right] = -4\pi\rho(\vec{r}) \tag{1.36}$$

Equation 1.36 is the generalized form of the Poisson-Boltzmann's equation [203]. For practical purposes equation (1.36) is converted to its ordinary form:

$$\nabla \left[\varepsilon * \nabla^2 V(\vec{r}) \right] = -4\pi\rho(\vec{r}) \tag{1.37}$$

where based on **Figure 1.11** the permittivity $\varepsilon(r)$ can take the following values depending the region where is evaluated [201]:

$$\varepsilon(\vec{r}) = \varepsilon = \begin{cases} \varepsilon_{\text{in}} = 1, & r \in \Omega \\ \varepsilon_{\text{out}} = \varepsilon_{\text{sol}}, & r \notin \Omega \end{cases} \tag{1.38}$$

In equation 1.36 $V(\vec{r})$ is considered for the cavity + solute and the environment (continuum solvent):

$$V(\vec{r}) = V^\rho(\vec{r}) + V_{\text{RF}}(\vec{r}) \tag{1.39}$$

where $V^\rho(\vec{r})$ is the electrostatic potential determined by the electron density distribution of the solute and $V_{\text{RF}}(\vec{r})$ is the electrostatic potential from the "reaction field" due to the polarization of the continuum solvent by the charge distribution in the solute.

If the charge density $\rho(\vec{r})$ and $V(\vec{r})$ are known it is possible to calculate the electrostatic solvation energy or polarization energy $E_{sol}(\vec{r})$ from either half of the "reaction field" contribution $V_{RF}(\vec{r})$ that has an associated charge density $\rho_{pol}(\vec{r})$ or from the electrostatic potential $V^{\rho}(\vec{r})$ and its associated charge density distribution $\rho(\vec{r})$ for the solute in the cavity:

$$E_{sol}(\vec{r}) = \frac{1}{2} * \int V_{RF}(\vec{r}) * \rho_{pol}(\vec{r}) = \frac{1}{2} * \int V^{\rho}(\vec{r}) * \rho(\vec{r}) \tag{1.40}$$

Using the bracket formalism the total energy for the system that can be obtained from the Schrodinger's equation is no longer just the energy in vacuum but the summation of the energy obtained from vacuum plus the electrostatic solvation energy $E_{sol}(\vec{r})$:

$$E[\psi] = \left\langle \psi \left| \hat{H} \right| \psi \right\rangle + E_{sol}(\vec{r})[\varepsilon, \rho] \tag{1.41}$$

where $\hat{H}$ is the Hamiltonian operator working on the wavefunction ψ and $E_{sol}(\vec{r})[\varepsilon, \rho]$ is the solvation energy that depends on the cartesian coordinates of the solute atoms and on the permittivity ε and the electron density distribution ρ for the solute.

1.4   Conclusions

This chapter has explored the available scientific literature on the prebiotic origins of the building blocks of NAs and state of the art of computational and theoretical methods to model/simulate the thermodynamic properties of nucleosides(tides) and the advantages and disadvantages for each of these computational methods.

The experimental synthesis of the canonical (found in today's NAs) nucleotides in aqueous solution has faced a number of challenges since the first experiments published by Orgel and coworkers [24]. These challenges have been collectively referred to as the "water problem" and the "paradox of base pairing". Additionally, the prebiotic synthesis of D-ribofuranose represents another challenge for the synthetic chemists. The poor regioselectivity for the glycosidic bond formation and poor nucleophilic and electrophilic character of the sugar and base respectively have been investigated [50].

The previous challenges have put in question the widely accepted "RNA world" hypothesis [15, 16, 17]. In this direction, alternatives that propose different synthetic routes and the use of suitable replacements for the sugar and bases have been proposed. These solutions suggest that DNA instead of having originated from RNA directly in prebiotic conditions evolved from a primitive proto- and pre-RNA with different components to the ones found in today's genetic material [11].

Additionally, the specific β-stereochemistry of the N-glycosidic bond in the building blocks represents an interesting feature of the chemical nature of the building blocks of nucleic acids. This topic has been justified on the basis of the sugar conformational changes or puckering for both anomers at different pH (see pp. 8-20 in [111]).

How Nature selected such a specific structure for today's building blocks of the DNA and RNA still remains as an open question. Maybe there are many answers and not an absolute explanation.

Computational tools are ideal in prebiotic chemistry since they can save time and resources in the laboratory and give additional insights on the thermodynamics for the selection of specific building blocks over others to be incorporated in the NAs.

In the next chapters we will deep dive in the thermodynamic properties of the molecular structure for a number of candidates for components of nucleic acids and artificial condensation reactions.

1.5   References

[1]   N. Kitadai and S. Maruyama, "Origins of building blocks of life: A review," *Geoscience Frontiers,* pp. 1117-1153, 2018.

[2]   J. M. Dohm and S. Maruyama, "Habitable Trinity," *Geoscience Frontiers,* vol. 6, pp. 95-101, 2015.

[3]   R. Laos and S. Benner, "The first self-replicating molecule and the origin of life," *Revista de Quimica PUCP,* vol. 28, pp. 1-2, 2014.

[4]   W. Martin, J. Baross, D. Kelley and M. J. Russell, "Hydrothermal vents and the origin of life," *Microbiology,* vol. 6, pp. 805-813, 2008.

[5]   G. Wachtershauser, "From volcanic origins of chemoautotrophic life to bacteria, Archaea and Eukarya," *Phylosophical Transactions of the Royal Society B,* vol. 361, pp. 1787-1808, 2006.

[6]   F. C. Dobbs and K. A. Selph , "Thermophilic bacterial activity in a deep-sea sediment from the Pacific ocean," *Aquatic Microbial Ecology,* vol. 13, pp. 209-212, 1997.

[7]   S. A. Sandford, M. Nuevo, P. P. Bera and J. T. Lee, "Prebiotic astrochemistry and the formation of molecules of astrobiological interest in interestellar clouds and protostellar disks," *Chemical Reviews,* vol. 120, pp. 4616-4659, 2020.

[8]   F. Trixler, "Quantum tunneling to the origin and evolution of life," *Current Organic Chemistry,* vol. 17, pp. 1758-1770, 2013.

[9]   A. S. Burton, J. C. Stern, J. E. Elsila, D. P. Glavini and J. P. Dworkin, "Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites," *Chemical Society Reviews,* vol. 41, pp. 5459-5472, 2012.

[10] A. Lazcano, "Oparin and the origin of life: a historical reassesssment of the heterotrophic theory," *Journal of Molecular Evolution,* vol. 83, pp. 214-222, 2016.

[11] N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, "The origin of RNA and "my grandfather's axe"," *Chemistry & Biology,* vol. 20, pp. 466-474, 2013.

[12] S. L. Miller, "Production of some organic compounds under possible primitive earth conditions," *Journal of the American Chemical Society,* vol. 77, pp. 2351-2361, 1955.

[13] S. L. Miller, "Production of amino acids under possible primitive earth conditions," *Science,* vol. 117, pp. 528-529, 1953.

[14] T. M. McCollom, "Miller-Urey and beyond: what have we learned about prebiotic organic synthesis reactions in the past 60 years?," *Annual Review of Earth and Planetary Sciences,* vol. 41, pp. 207-229, 2013.

[15] S. Ayukawa, T. Enomoto and D. Kiga, "RNA world," in *Astrobiology: From the Origins of Life to the Search for*, Singapore, Springer Nature, 2019, pp. 77-90.

[16] M. Neveu, H. J. Kim and S. A. Benner, "The ''strong'' RNA world hypothesis: fifty years old," *Astrobiology,* vol. 13, pp. 391-403, 2013.

[17] L. E. Orgel, "Prebiotic chemistry and the origin of the RNA world," *Critical Reviews in Biochemistry and Molecular Biology,* vol. 39, pp. 99-123, 2004.

[18] D. L. Nelson and M. M. Cox, "Nucleotides and nucleic acids," in *Lehninger principles of biochemistry*, New York, W.H. Freeman & Company, 2013, pp. 281-313.

[19] L. Huang, M. Krupkin, A. Bashan, A. Yonath and L. Massa, "Protoribosome by quantum kernel energy method," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 110, pp. 14900-14905, 2013.

[20] L. Massa, C. F. Matta, A. Yonath and J. Karle, "Quantum transition state for peptide bond formation in the ribosome," in *Quantum biochemistry: electronic structure and biological activity*, Weinheim, Wiley-VCH, 2010, pp. 501-515.

[21] A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath and J. Karle, "The transition state for the formation of the peptide bond in the ribosome," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, pp. 13327-13332, 2006.

[22] L. Pray, "Discovery of DNA Structure and Function: Watson and Crick," *Nature Education,* vol. 100, pp. 1-6, 2008.

[23] M. W. Powner, B. Gerland and J. D. Sutherland, "Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions," *Nature Chemistry,* vol. 459, pp. 239-242, 2009.

[24] W. D. Fuller, "Studies in prebiotic synthesis VI. Synthesis of purine nucleosides," *Journal of Molecular Biology,* vol. 67, pp. 25-33, 1972.

[25] A. do Nascimento Vieira, K. do Nascimento Vieira, W. F. Martin and M. Preiner, "The ambivalent role of water at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[26] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei and S. A. Benner, "Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: Borate as a multifaceted problem solver in prebiotic chemistry," *Angewandte Chemie International Edition,* vol. 55, pp. 15816-15820, 2016.

[27] G. F. Joyce and L. E. Orgel, "Prospects for understanding the origin of the RNA world," in *The RNA World*, 2 ed., New York, Cold Spring Harbor Press, 1999, pp. 49-78.

[28] A. V. Demchenko, "General aspects of the glycosidic bond formation," in *Handbook of glycosylation*, A. V. Demchenko, Ed., Federal Republic of Germany, Wiley-VCH Verlag GmbH & Co, 2008, pp. 1-19.

[29] S. A. Benner, H. J. Kim and M. A. Carringa, "Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides and RNA," *Accounts of Chemical Research,* vol. 45, p. 2025–2034, 2012.

[30] M. Ma and D. Bong, "Determinants of cyanuric acid and melamine assembly in water," *Langmuir,* vol. 27, pp. 8841-8853, 2011.

[31] M. Gull, "Prebiotic phosphorylation reactions on the early earth," *Challenges,* vol. 5, pp. 193-212, 2014.

[32] B. J. Cafferty and N. V. Hud, "Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA," *Israel Journal of Chemistry,* vol. 55, pp. 891-905, 2015.

[33] A. E. Engelhart and N. V. Hud, "Primitive genetic polymers," *Perspectives in Biology,* vol. 2, article #a002196 (pp. 1-21), 2010.

[34] A. Butlerow, "Formation synthetique d'une substance sucree," *Comptes Rendus de l'Academie des Sciences,* vol. 63, pp. 145-147, 1861.

[35] N. Kitadai and S. Maruyama, "Origins of building blocks of life: a review," *Geoscience Frontiers,* vol. 9, pp. 1117-1153, 2018.

[36] B. J. Cafferty and N. V. Hud, "Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology," *Current Opinion in Chemical Biology,* vol. 22, pp. 146-157, 2014.

[37] J. E. Šponer, R. Szabla, R. W. Góra, A. M. Saitta, F. Pietrucci, F. Saija, E. Di Mauro, R. Saladino, M. Ferus, S. Civišh and J. Šponer, "Prebiotic synthesis of nucleic acids and their building blocks at the atomic level – merging models and mechanisms from advanced computations and experiments," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20047-20066, 2016.

[38] A. Ricardo, M. A. Carrigan, A. N. Olcott and A. Benner, "Borate minerals stabilize ribose," *Science,* vol. 303, pp. 196, 2004.

[39] R. A. Sanchez and L. E. Orgel, "Studies in prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides," *Journal of Molecular Biology,* vol. 47, pp. 531-543, 1970.

[40] J. Xu, M. Tsanakopoulou, C. J. Magnani, R. Szabla, J. E. Šponer and J. D. Sutherland, "A prebiotically plausible synthesis of pyrimidine *β*-ribonucleosides and their phosphate derivatives involving photoanomerization," *Nature Chemistry,* vol. 9, pp. 303-309, 2016.

[41] D. Ritson and J. D. Sutherland, "Prebiotic synthesis of simple sugars by photoredox system chemistry," *Nature Chemistry,* vol. 4, pp. 895-899, 2012.

[42] N. V. Hud and F. A. L. Anet, "Intercalation-mediated synthesis and replication: a new approach to the origin of life," *Journal of Theoretical Biology,* vol. 205, pp. 543-562, 2000.

[43] J. Šponer, A. Mladek, J. E. Šponer, D. Svozil, M. Zgarbova, P. Banas, P. Jurecka and M. Otyepka, "The DNA and RNA sugar-phosphate backbone emerges as the key player. An

overview of quantum-chemical, structural biology and simulation studies," *Physical Chemistry Chemical Physics,* vol. 14, pp. 15257-15277, 2012.

[44] H.-J. Kim and J. Kim, "A prebiotic synthesis of canonical pyrimidine and purine ribonucleotides," *Astrobiology,* vol. 19, pp. 669-674, 2019.

[45] M. Gull, B. J. Cafferty, N. V. Hud and M. A. Pasek, "Silicate-promoted phosphorylation of glycerol in non-aqueous solvents: A prebiotically plausible route to organophosphates," *Life,* vol. 7, article #life7030029 (pp. 1-10), 2017.

[46] H. S. Bernhardt and R. K. Sandwick, "Purine biosynthetic Intermediate-containing ribose-phosphate polymers as evolutionary precursors to RNA," *Journal of Molecular Evolution,* vol. 79, pp. 91-104, 2014.

[47] M. A. Crowe and J. D. Sutherland, "Reaction of cytidine nucleotides with cyanoacetylene: Support for the intermediacy of nucleoside-2', 3'-cyclic phosphates in the prebiotic synthesis of RNA," *ChemBioChem,* vol. 7, pp. 951-956, 2006.

[48] A. A. Ingar, R. W. A. Luke, B. R. Hayter and J. D. Sutherland, "Synthesis of cytidine ribonucleotides by stepwise assembly of the heterocycle on a sugar phosphate," *ChemBioChem,* vol. 4, pp. 1127-1135, 2003.

[49] G. Zubai and T. Mui, "Prebiotic synthesis of nucleotides,"*Origins of Life and Evolution of Biospheres,* vol. 31, pp. 87-102, 2001.

[50] D. M. Fialho, *Physical organic principles governing the spontaneous prebiotic emergence of proto-nucleic acids,* Georgia: Georgia Tech Library, 2019, pp. 1-3.

[51] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy and N. V. Hud, "Spontaneous formation and base pairing of plausible prebiotic nucleotides in water," *Nature Communications,* vol. 7, article #11328 (pp. 1-8), 2016.

[52] B. J. Cafferty, D. M. Fialho and N. V. Hud, "Searching for possible ancestors of RNA: The self-assembly hypothesis for the origin of proto-RNA," in *Prebiotic chemistry and chemical*

evolution of nucleic Acids, nucleic acids and molecular biology, vol. 35, C. Menor-Salván, Ed., Philadelphia, Pennsylvania: Springer, 2018, pp. 143-174.

[53] B. J. Cafferty, I. Gállego, M. C. Chen, K. I. Farley, R. Eritja and N. V. Hud, "Efficient self-assembly in water of Long noncovalent polymers by nucleobase analogues," *Journal of the American Chemical Society,* vol. 135, pp. 2447-2450, 2013.

[54] C. Li, B. J. Cafferty, S. C. Karunakaran, G. B. Schuster and N. V. Hud, "Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: implications for the possible origins of RNA," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20091-20096, 2016.

[55] M. C. Chen, B. J. Cafferty, I. Mamajanov, I. Gállego, J. Khanam, R. Krishnamurthy and N. V. Hud, "Spontaneous prebiotic formation of a *β*-ribofuranoside that self assembles with a complementary heterocycle," *Journal of the American Chemical Society,* vol. 136, pp. 5640-5646, 2014.

[56] C. V. Mungi, S. K. Singh, J. Chugh and S. Rajamani, "Synthesis of barbituric acid containing nucleotide and its implications for the origins of primitive informational polymer," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20091-20096, 2016.

[57] J. Sutherland, "Ribonucleotides," *Cold Spring Harbor Perspectives in Biology,* vol. 2, article # a005439 (pp. 1-13), 2010.

[58] K. N. Drew, J. Zajicek, G. Bondo, B. Bose and A. S. Serianni, "13C-labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy," *Carbohydrate Research,* vol. 307, pp. 199-209, 1998.

[59] M. Schmidt, "Xenobiology: a new form of life as the ultimate biosafety tool," *BioEssays,* vol. 32, pp. 322-331, 2010.

[60] E. Lescrinier, M. Froeyen and P. Herdewijn, "Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids," *Nucleic Acids Research,* vol. 31, pp. 2975-2989, 2003.

[61] P. Gu, G. Schepers, J. Rozenski , A. Aerschot and P. Herdewijn, "Base pairing properties of D- and L-Cyclohexene nucleic acids (CeNA)," *Oligonucleotides,* vol. 13, pp. 479-489, 2003.

[62] L. Orgel, "A simpler nucleic acid," *Science,* vol. 290, p. 1306, 2000.

[63] K. U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, "Chemical etiology of nucleic acid structure: the *α*-threofuranosyl - (3' —>2') oligonucleotide system," *Science,* vol. 290, pp. 1347-1351, 2000.

[64] M. Hernández-Rodríguez, J. Xie, Y. M. Osornio and R. Krishnamurthy, "Mapping the landscape of potentially primordial informational oligomers: (30/20) - D - phosphoglyceric acid linked acyclic oligonucleotides tagged with 2, 4 - disubstituted 5 - aminopyrimidines as recognition elements," *Chemistry: An Asian Journal,* vol. 6, pp. 1252-1262, 2011.

[65] E. Meggers and L. Zhang, "Synthesis and properties of the simplified nucleic acid glycol nucleic acid," *Accounts of Chemical Research,* vol. 43, pp. 1092-1102, 2010.

[66] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

[67] P. E. Nielsen, "Peptide nucleic acid (PNA). Implications for the origin of the genetic material and the homochirality of life," *American Institute of Physics Conference Proceedings,* vol. 379, pp. 55-61, 1996.

[68] P. E. Nielsen, "Peptide nucleic acid (PNA): a model structure for the primordial genetic material?," *Origins of Life and Evolution of the Biosphere,* vol. 23, pp. 323-327, 1993.

[69] P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide," *Science,* vol. 254, pp. 1497-1500, 1991.

[70] M. Frenkel-Pinter, M. Samanta, G. Ashkenasy and L. J. Leman, "Prebiotic peptides: molecular hubs in the origin of life," *Chemical Reviews,* vol. 120, pp. 4707-4765, 2020.

[71] J. Wu, Q. Meng, H. Ren, H. Wang, J. Wu and Q. Wang, "Recent advances in peptide nucleic acid for cancer bionanotechnology," *Acta Pharmacologica Sinica,* vol. 38, pp. 798-805, 2017.

[72] E. J. Lee, H. K. Lim, Y. S. Cho and S. S. Hah, "Peptide nucleic acids are an additional class of aptamers," *RSC Advances,* vol. 3, pp. 5828-5831, 2013.

[73] e. Mateo-Martí and C.-M. Pradie, "A novel type of nucleic acid-based biosensors: the use of PNA probes, associated with surface science and electrochemical detection techniques," in *Intelligent and biosensors*, Rijeka, InTech, 2010, pp. 323-344.

[74] P. E. Nielsen, "Peptide nucleic acids and the origin of life," *Chemistry & Biodiversity,* vol. 4, pp. 1996-2002, 2007.

[75] V. Menchise, G. De Simone, T. Tedeschi, R. Corradini, S. Sforza, R. Marchelli, D. Capasso, M. Saviano and C. Pedone, "Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, pp. 12021-12026, 2003.

[76] D. M. Fialho, T. Y. Roche and N. V. Hud, "Prebiotic syntheses of noncanonical nucleosides and nucleotides," *Chemical Reviews,* vol. 120, pp. 4806-4830, 2020.

[77] D. M. Fialho, K. C. Clarke, M. K. Moore, G. B. Schuster, R. Krishnamurthy and N. V. Hud, "Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides," *Organic and Biomolecular Chemistry,* vol. 16, pp. 1263-1271, 2018.

[78] A. Eschenmoser, "Chemical etiology of nucleic acid structure," *Science,* vol. 284, pp. 2118-2124, 1999.

[79] S. Pitsch, S. Wendeborn, B. Jaun and A. Eschenmoser, "Why pentose-and not hexose-nucleic acids?. Part VII. Pyranosyl-RNA ('p-RNA')," *Preliminary Communication,* vol. 76, pp. 2161-2183, 1993.

[80] R. Micura, R. Kudick, S. Pitsch and A. Eschenmoser , "Opposite orientation of backbone inclination in pyranosyl-RNA and homo-DNA," *Angewandte Chemie International Edition,* vol. 38, pp. 680-683, 1999.

[81] S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculca, W. Norbert, R. Micura, M. Stanek, B. Jaun and A. Eschenmoser, "Pentopyranosyl oligonucleotide systems (9th communication). The *β*-D-ribopyranosyl-(4'→2')-oligonucleotide system (-pyranosyl-RNA'): synthesis and resumé of base-pairing properties," *Helvetica Chimica Acta,* vol. 86, pp. 4270-4363, 2003.

[82] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

[83] Y.-W. Yang, S. Zhang, E. O. McCullum and J. C. Chaput, "Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA," *Journal of Molecular Evolution,* vol. 65, p. 289–295, 2007.

[84] M. K. Schlegel and E. Meggers, "Improved phosphoramidite building blocks for the synthesis of the simplified nucleic acid GNA," *The Journal of Organic Chemistry,* vol. 74, pp. 4615-4618, 2009.

[85] P. D. Cook, O. L. Acevedo, P. W. Davis, D. J. Ecker and H. Normand, "Phosphate linked oligomers". United States of America Patent US5886177A, 1999.

[86] S. A. Banack, J. S. Metcalf, L. Jiang, D. Craighead, L. L. Ilag and P. A. Cox, "Cyanobacteria produce N-(2-aminoethyl)glycine, a backbone for peptide nucleic acids which may have been the first molecules for life on earth," *Plos One,* vol. 7, article #e49043 (pp. 1-4), 2012.

[87] R. Gmelin, "Free amino acids in seeds of *Acacia willardiana* (*Mimosaceae*). Isolation of willardiine, a new plant amino acid, probably L-uracil-[*β*-(*α*-aminopropionic acid)]," *Hoppe-Seyler's Z Physiological Chemistry,* vol. 316, pp. 164-169, 1959.

[88] j. k. MacLeod, R. E. Summons, C. W. Parker and D. S. Letham , "Lupinic acid, a purinyl amino acid and a novel metabolite of zeatin," *Journal of the Chemical Society, Chemical Communications,* pp. 809-810, 1975.

[89] U. J. Meierhenrich, G. M. Muñoz Caro, J. H. Bredehöft, , E. K. Jessberger and W. H.-P. Thiemann, "Identification of diamino acids in the Murchison meteorite," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, pp. 9182-9186, 2004.

[90] G. M. Muñoz Caro, U. J. Meierhenrich, W. A. Shutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H.-P. Thiemann, A. Brack and J. M. Greenberg, "Amino acids from ultraviolet irradiation of interstellar ice analogues," *Nature,* vol. 416, pp. 403-406, 2002.

[91] K. E. Nelson, M. Levy and S. L. Miller, "Peptide nucleic acids rather than RNA may have been the first genetic molecule," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, pp. 3868-3871, 2000.

[92] K. Ogura, C. T. Migita and T. Yamada, "Photolysis of CH_4 NH_3 H_2O mixture: formation of methylamine and ethylenediamine," *Journal of Photochemistry and Photobiology A: Chemistry,* vol. 49, pp. 53-61, 1989.

[93] P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, "DNA-like double helix formed by peptide nucleic acid," *Nature,* vol. 368, pp. 561-563, 1994.

[94] M. Egholm, O. Buchardt, P. E. Nielsen and R. Berg, "Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone," *Journal of the American Chemical Society,* vol. 114, pp. 1895-1897, 1992.

[95] G. F. Joyce, A. W. Schwartz, S. L. Miller and L. E. Orgel, "The case for an ancestral genetic system involving simple analogues of the nucleotides," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 84, pp. 4398-4402, 1987.

[96] P. R. Barry, A. A. Ajees and T. R. McDermott, "Life and death with arsenic," *Bioessays,* vol. 33, pp. 350-357, 2011.

[97] F. Wolfe-Simon, J. S. Blum, T. R. Kulp, G. W. Gordon, S. E. Hoeft, J. Pett-Ridge, J. F. Stolz, S. M. Webb, P. K. Weber, P. C. W. Davies, A. D. Anbar and R. S. Oremland, "A bacterium that can grow by using arsenic instead of phosphorus," *Science,* vol. 332, pp. 1163-1166, 2011.

[98] G. Ni, D. Yuqi, F. Tang, J. Liu, H. Zhao and Q. Chen, "Review of α-nucleosides: from discovery, synthesis to properties and potential applications," *RSC Advances,* vol. 9, pp. 14302-14320, 2019.

[99] H. Ide, H. Shimizu, Y. Kimura, S. Sakamoto, K. Makino, M. Glackin, S. S. Wallace , H. Nakamuta, M. Sasaki and S. Sugimoto, "Influence of alpha-deoxyadenosine on the stability and structure of DNA. Thermodynamic and molecular mechanics studies," *Biochemistry,* vol. 34, pp. 6947-6955, 1995.

[100] S. Latha and N. Yathindra, "Stereochemical studies on nucleic acid analogues. I. conformations of α-nucleosides and α-nucleotides: interconversion of sugar puckers via 04'-exo," *Biopolymers,* vol. 32, pp. 249-269, 1992.

[101] S. Suzuki, K. Suzuki, T. Imai, N. Suzuki and S. Okuda, "Isolation of alpha-pyridine nucleotides from Azotobacter vinelandii," *Journal of Biological chemistry,* vol. 240, pp. P554-P556, 1965.

[102] F. Dinglinger and P. Renz, "9-α-D-ribofuranosyladenin (α-adenosin), das nucleosid des corrinoidfaktors Cx aus *Propionibacterium shermanii*," *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie,* vol. 352, pp. 1157-1161, 1971.

[103] M. Froeyen, E. Lescrinier, L. Kerremans, H. Rosemeyer, F. Seela, B. Verbeure, I. Lagoja, J. Rozenski, A. V. Aerschot, R. Busson and P. Herdewijn, "α-homo-DNA and RNA form a parallel oriented non-A, non-B-type double helical structure," *Chemistry: A European Journal,* vol. 7, pp. 5183-5194, 2001.

[104] J.-L. Guesnet, F. Vovelle, N. T. Thuong and G. Lancelot, "2D NMR studies and 3D structure of the parallel-stranded duplex oligonucleotide Acrm5-alpha-d(TCTAAACTC)-beta-

d(AGATTTGAG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 29, pp. 4982-4991, 1990.

[105] J. Paoletti, D. Bazile, F. Morvan, J.-L. Imbach and C. Paoletti, "*α*-DNA VIII:thermodynamic parameters of complexes formed between the oligo-alphadeoxynucleotides: *α*-d(GGAAGG) and *α*-d(CCTTCC) and their complementary oligo-betadeoxynucleotides: *β*-d(CCTTCC) and *β*-d(GGAAGG) are different," *Nucleic Acid Research,* vol. 17, pp. 2693-2704, 1989.

[106] G. Lancelot, J.-L. Guesnet and F. Vovelle, " Solution structure of the parallel-stranded duplex oligonucleotide *α*-d(TCTAAAC)-*β*-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 28, pp. 7871-7878, 1989.

[107] G. Lancelot, J.-L. Guesnet, V. Roig and N. T. Thuong, "2D-NMR studies of the unnatural duplex *α*-d(TCTAAAC)-*β*-d(AGATTTG)," *Nucleic Acids Research,* vol. 15, pp. 7531-7547, 1987.

[108] M. Miljkovic, Carbohydrates: synthesis, mechanisms, and stereoelectronic effects, New York: Springer, 2009.

[109] S. Gerber-Lemaire and P. Vogel, "Anomeric effects in pyranosides and related acetals," in *Carbohydrate chemistry: chemical and biological approaches*, Cambridge, RSC Publishing, 2009, pp. 13-32.

[110] D. E. Levy and P. Fügedi, The organic chemistry of sugars, 1 ed., Boca Ratón: Taylor & Francis Group, 2005.

[111] C. Thibaudeau, P. Acharya and J. Chattopadhyaya, Stereoelectronic effects in nucleosides and nucleotides and their structural implications, Uppsala: Uppsala University Press, 2005.

[112] C. Thibaudeau, A. Foldesi and J. Chattopadhyaya, "The first experimental evidence for a larger medium-dependent flexibility of natural *β*-D-nucleosides compared to the *α*-D-nucleosides," *Tetrahedron,* vol. 53, pp. 14043-14072, 1997.

[113] C. Thibaudeau, J. Plavec and J. Chattopadhyaya, "Quantitation of the anomeric effect in adenosine and guanosine by comparison of the thermodynamics of the pseudorotational equilibrium of the pentofuranose moiety in N- and C-nucleosides," *Journal of the American Chemical Society,* vol. 116, pp. 8033-8031 , 1994.

[114] R. Bielski and M. Tencer, "A possible path to the RNA world: enantioselective and diastereoselective purification of ribose," *Origins of Life and Evolution of Biospheres,* vol. 37, pp. 167-175, 2007.

[115] D. Cremer and J. A. Pople, "A general definition of ring puckering coordinates," *Journal of the American Chemical Society,* vol. 97, pp. 1354-1358, 1975.

[116] L. A. Monteserín Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article#e09657 (pp. 1-12), 2022.

[117] A. Poole, D. Penny and B.-M. Sjöberg, "Confounded cytosine! Tinkering and the evolution of DNA," *Nature Reviews Molecular Cell Biology,* vol. 2, pp. 147-151, 2001.

[118] R. A. Hitzeman and A. R. Price, "Relationship of Bacillus subtilis DNA polymerase III to bacteriophage PBS2-induced DNA polymerase and to the replication of uracil-containing DNA," *Journal of Virology,* vol. 28, pp. 697-709, 1978.

[119] A. M. Lesk, "Why does DNA contain Thymine and RNA Uracil?," *Journal of Theoretical Biology,* vol. 22, pp. 537-540 , 1969.

[120] C. J. Cramer, "What are theory, computation, and modeling?," in *Esentials of computational chemistry: theories and models*, 2 ed., West Sussex, Wiley & Sons. Ltd, 2004, pp. 1-16.

[121] D. C. Young, "Introduction," in *Computational chemistry. A practical guide for applying techniques to real-world problems*, New York, John Wiley & Sons, Inc, 2001, pp. 1-5.

[122] F. Jensen, "Introduction," in *Introduction to computational chemistry*, 2 ed., New York, Wiley & Sons. Ltd, 1999, pp. 1-21.

[123] R. Salomon-Ferrer, D. A. Case and R. C. Walker, "An overview of the Amber biomolecular simulation package," *WIREs Computational Molecular Science,* vol. 3, pp. 198-210, 2013.

[124] B. Hess, D. Kutzner, D. van der Spoel and E. Lindahl, "GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation,* vol. 4, pp. 435-447, 2008.

[125] D. A. Case, I. T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, J. K. M. Merz, C. Onufriev, C. Simmerling, B. Wang and R. Woods, "The Amber biomolecular simulation programs," *Journal of Computational Chemistry,* vol. 26, pp. 1668-1688, 2005.

[126] D. Frenkel and B. Smit, Understanding molecular simulation: from algorithms to applications, New York: Academic Press, 2002.

[127] T. Schlick, Molecular modeling and simulation: an interdisciplinary guide, New York: Springer, 2002.

[128] J. -P. Doucet and J. Weber, Computer-aided molecular design: theory and applications, London: Academic Press, ltd, 1996.

[129] E. Schrödinger, "Quantisierung als eigenwertproblem," *Annalen der Physik,* vol. 384, p. 361–377, 1926.

[130] The CPMD consortium, MPI für Festkörperforschung Stuttgart and IBM Zurich Research Laboratory, "CPMD, version 3.11," Copyright IBM Corp 1990-2019, copyright MPI für Festkörperforschung Stuttgart, 2006. [Online]. Available: http://www.cpmd.org/.

[131] D. Marx and J. Hutter, Ab initio molecular dynamics: theory and implementation, in modern methods and algorithms of quantum chemistry, Forschungszentrum Jülich: NIC-Directors, 2000.

[132] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical Review Letters,* vol. 55, pp. 2471-2474, 1985.

[133] C. F. Matta, "Introductory reflections on quantum biochemistry: from context to contents," in *Quantum biochemistry. Electronic structure and biological acivity*, vol. 1, Weinheim, Wiley-VCH , 2010, pp. XI-LI.

[134] W. G. Richards, Quantum pharmacology, London: Butterworth & Co., Ltd, 1983.

[135] B. Pullman and A. Pullman, Quantum biochemistry, New York: Interscience Publishers, 1963.

[136] J. Šponer, J. E. Šponer, A. Mládek, P. Banáš, P. Jurečka and M. Otyepka, "How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists," *Methods,* vol. 64, pp. 3-11, 2013.

[137] E. Schrödinger, What is life? The physical aspect of the living cell and mind, Cambridge: Cambridge University Press, 1944.

[138] J. F. C. Mayen, "Recent results on computational molecular modeling of the origins of life," *Foundations of Computing and Decision Sciences,* vol. 45, pp. 35-46, 2020.

[139] C. Gatti, G. Macetti, R. J. Boyd and C. F. Matta, "An electron density source-function study of DNA bases pairs in their neutral and ionized ground states," *Journal of Computational Chemistry,* vol. 39, pp. 1112-1128, 2018.

[140] J. E. Šponer, P. Banáš, P. Jurečka, M. Zgarbová, P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer and M. Otyepka, "Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome," *The Journal of Physical Chemistry Letters,* vol. 5, pp. 1771-1782, 2014.

[141] A. Y. L. Sim, p. Minary and M. Levitt, "Modeling nucleic acids," *Current Opinion in Structural Biology,* vol. 22, pp. 273-278, 2012.

[142] P. Atkins and R. Friedman, "The foundations of Quantum Mechanics," in *Molecular Quantum Mechanics*, 4th ed., Oxford, Oxford University Press, 2005, pp. 9-40.

[143] L. I. Deych, "Unitary operators and quantum dynamics," in *Advanced undergraduate quantum mechanics. Methods and applications*, New York, Springer, 2018, pp. 95-111.

[144] C. J. Cramer, "Foundations of molecular orbital theory," in *Esentials of computational chemistry: theories and models*, 2nd ed., West Sussex, John Wiley & Sons. Ltd, 2004, pp. 95-120.

[145] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik,* vol. 84, pp. 457-484, 1927.

[146] D. W. Rogers, "Semiempirical calculations on larger molecules," in *Computational chemistry using the PC*, New Jersey, John Wiley & Sons Ltd, 2003, pp. 263-278.

[147] I. N. Levine, "El metodo de variaciones," in *Química cuántica*, 5ta ed., Madrid, Pearson Educación, 2001, pp. 205-231.

[148] W. Pauli, "Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren," *Zeitschrift für Physik,* vol. 3, pp. 765-783, 1925.

[149] W. Koch and M. C. Holthausen, A chemist's guide to Density Funcional Theory, 2nd ed., Weinheim: Wiley-VCH, 2001, pp. 19-27.

[150] P. O. Löwdin, "Correlation problem in many-electron quantum mechanics. 1. Review of different approaches and discussion of some current ideas," *Advances in Chemical Physics,* vol. 2, pp. 207-322, 1959.

[151] P. O. Löwdin, " Quantum Theory of Many-Particle Systems. III. Extension of the Hartree-Fock scheme to include degenerate systems and correlation effects," *Physical Review,* vol. 97, pp. 1509-1520, 1955.

[152] H. J. Werner, R. M. Frederick, P. Knowles and P. J. Knowles, "Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations," *The Journal of Chemical Physics,* vol. 118, pp. 8149-8161, 2003.

[153] E. R. Johnson, I. D. Mackie and G. A. DiLabio, "Dispersion interactions in density-functional theory," *Journal of Physical Organic Chemistry,* vol. 22, pp. 1127-1135, 2009.

[154] S. F. Sousa, P. A. Fernandes and M. J. Ramos, "General performance of density funcionals," *The Journal of Physical Chemistry A,* vol. 111, pp. 10439-10452, 2007.

[155] P. Hohenberg and W. Kohn, "Inhomogeneus electron gas," *Physical Review ,* vol. 136, pp. 1133-1141, 1964.

[156] P. Geerlings, F. De-Proft and W. Langenaeker, "Conceptual Density Functional Theory," *Chemical Reviews,* vol. 103, pp. 1793-1873, 2001.

[157] A. J. Cohen, P. Mori-Sánchez and W. Yang, "Challenges for Density Functional Theory," *Chemical Reviews,* vol. 112, pp. 289-320, 2012.

[158] W. kohn and J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review,* vol. 140, pp. 1133-1138, 1965.

[159] M. Orio, D. A. Pantazis and F. Neese, "Density Functional Theory," *Photosynthesis Research,* vol. 102, pp. 443-453, 2009.

[160] R. H. Hertwig and W. Koch, "On the parameterization of the local correlation functional. What is Becke-3-LYP?," *Chemical Physics Letters,* vol. 268, pp. 345-351, 1997.

[161] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A,* vol. 38, pp. 3098-3100, 1988.

[162] C. Lee, W. Yang and R. G. Parr, "Development of the Colle–Salvetti correlationenergy formula into a functional of the electron density," *Physical Review B,* vol. 37, pp. 785-789, 1988.

[163] Y. Zhao and D. G. Truhlar, "Density functionals with broad applicability in chemistry," *Accounts of Chemical Research,* vol. 41, pp. 157-167, 2008.

[164] J. E. Del-Bene, W. B. Person and K. Szczepaniak, "Properties of hydrogen-bonded complexes obtained from the B3LYP functional with 6-31G(d,p) and 6-31+G(d,p) basis

sets: comparison with MP2/6-31+G(d,p) results and experimental data," *The Journal of Physical Chemistry,* vol. 99, pp. 10705-10707, 1995.

[165] P. Hobza, R. Zahradnik and K. Müller-Dethlefs, "The world of non-covalent interactions," *Collection of Czechoslovak Chemical Communications,* vol. 71, pp. 443-531, 2006.

[166] U. Zimmerli, M. Parrinello and P. Koumoutsakosa, "Dispersion corrections to density functionals for water aromatic interactions," *The Journal of Chemical Physics,* vol. 120, pp. 2693-2699, 2004.

[167] J. E. Šponer, J. Šponer and M. Fuentes-Cabrera, "Prebiotic routes to nucleosides: a quantum chemical Insight into the energetics of the multistep reaction pathways," *Chemistry - A European Journal,* vol. 17, pp. 847-854, 2011.

[168] S. Kaur, P. Sharma and S. D. Wetmore, "Can cyanuric acid and 2, 4, 6-triaminopyrimidine containg ribonucleosides be components of prebiotic RNA? Insights from QM calculations and MD simulations," *ChemBioChem,* vol. 20, pp. 1415-1415, 2019.

[169] S. Kaur, P. Sharma and S. D. Wetmore, "Structural and electronic properties of barbituric acid and melamine-containing ribonucleosides as plausible components of prebiotic RNA: implications for prebiotic self-assembly," *Physical Chemistry Chemical Physics,* vol. 19, pp. 30762-30771, 2017.

[170] W. Thiel, "Semiempirical quantum–chemical methods," *Wiley Interdisciplinary Reviews: Computational Molecular Science,* vol. 4, pp. 145-157, 2013.

[171] E. Codorniu, A. Rolo and L. A. Montero, "Theoretical affinity order among flavonoids and amino acid residues: An approach to understand flavonoid-protein interactions," *Journal of Molecular Structure: THEOCHEM,* vol. 819, pp. 121-129, 2007.

[172] F. Jensen, "Electronic Structure Methods: Independent-Particle Models," in *Introduction to Computational Chemistry*, 2nd ed., West Sussex, John Wiley & Sons, 2007, pp. 80-131.

[173] E. G. Lewars, "Semiempirical calculations," in *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*, 2nd ed., New York, Springer, 2011, pp. 391-443.

[174] T. Clark, "Quo Vadis semiempirical MO-theory?," *Journal of Molecular Structure: THEOCHEM,* vol. 530, pp. 1-10, 2000.

[175] J. J. P. Stewart, "MOPAC," Stewart Computational Chemistry, 2023. [Online]. Available: http://openmopac.net/.

[176] J. J. P. Stewart, "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters," *Journal of Molecular Modeling,* vol. 19, pp. 1-32, 2013.

[177] J. J. P. Stewart, "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements," *Journal of Molecular Modeling,* vol. 13, pp. 1173-1213, 2007.

[178] J. J. Stewart, "MOPAC: a semiempirical molecular orbital program," *Journal of Computer-Aided Molecular Design ,* vol. 4, pp. 1-105, 1990.

[179] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. P. Stewart, "AM1: a new general purpose quantum mechanical model," *Journal of the American Chemical Society,* vol. 107, pp. 3902-3909, 1985.

[180] M. Korth, M. Pitoňák, J. Řezáč and P. Hobza, "Transferable H-bonding correction for semiempirical quantum-chemical methods," *Journal of Chemical Theory and Computation,* vol. 6, pp. 344-352, 2010.

[181] L. A. Montero, "GRANADA Programme: Distribución aleatoria de moléculas alrededor de un sistema Poliatómico central (Random distribution of molecules around a central polyatomic system)," 2019, http://karin.fq.uh.cu/mmh/.

[182] L. A. Montero, J. Molina and J. Fabian, "Water clusters for calculations of association energy," *International Journal of Quantum Chemistry,* vol. 79, pp. 8-16, 2000.

[183] L. A. Montero, A. M. Esteva, J. Molina, A. Zapardiel, L. Hernández, H. Márquez and A. Acosta, "A Theoretical Approach to Analytical Properties of 2,4-Diamino-5-phenylthiazole in Water Solution. Tautomerism and Dependence on pH," *Journal of the American Chemical Society,* vol. 120, pp. 12023-12033, 1998.

[184] R. Crespo-Otero, Y. Pérez-Badell, J. A. Padrón-García and L. A. Montero-Cabrera, "Exploring the potential energy surfaces of association of NO with aminoacids and related organic functional groups: the role of entropy of association," *Theoretical Chemistry Accounts,* vol. 118, pp. 649-663, 2007.

[185] C. Morera-Boado, E. Alonso-Becerr, R. González-Jonte, L. A. Montero-Cabrera and J. M. García-de-la-Vega, "A theoretical approach to the solvation of brassinosteroids," *Journal of Molecular Graphics and Modelling,* vol. 27, pp. 600-610, 2009.

[186] J. D. Holliday, N. Salim and P. Willett, "On the magnitudes of coefficient values in the calculation of chemical similarity and dissimilarity," *ACS Symposium Series,* vol. 894, pp. 77-95, 2005.

[187] E. Sánchez-García and L. A. Montero, "Análisis de semejanza en las asociaciones moleculares de la amilosa con algunos compuestos organolépticos," *Revista cubana de Física,* vol. 22, pp. 73-80, 2005.

[188] D. Bajusz, A. Rácz and K. Héberger, "Why is Tanimoto index an appropriate choice for fingreprint-based similarity calculations?," *Journal of Cheminformatics,* vol. 7, pp. Article#s13321-015-0069-3 (pp. 1-13), 2015.

[189] T. Tanimoto, "An elementary mathematical theory of classification and prediction," International Business Machine (IBM) Corporation, New York, 1957.

[190] D. E. Patterson, R. D. Cramer, A. M. Ferguson, R. D. Clark and L. E. Weinberger, "Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors," *Journal of Medicinal Chemistry,* vol. 39, pp. 3049-3059, 1996.

[191] J. Chen, Y. Shao and J. Ho, "Are explicit solvent models more accurate than implicit solvent models? A case study on the Menschutkin reaction," *The Journal of Physical Chemistry A,* vol. 123, pp. 5580-5589, 2019.

[192] J. Zhang, H. Zhang, T. Wu, Q. Wang and D. v. d. Spoel, "Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents," *Journal of Chemical Theory and Computation,* vol. 13, pp. 1034-1043, 2017.

[193] R. Anandakrishnan, A. Drozdetski, R. C. Walker and A. V. Onufriev, "Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations," *Biophysical Journal,* vol. 108, pp. 1153-1164, 2015.

[194] R. E. Amaro, X. Cheng, I. Ivanov, D. Xu and J. A. McCammon, "Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via Generalized Born molecular dynamics and end-point free energy calculations," *Journal of the American Chemical Society,* vol. 131, pp. 4702-4709, 2009.

[195] B. Mennucci, "Polarizable continuum model," *Advanced Review,* vol. 2, pp. 386-404, 2012.

[196] J. Tomasi, "Selected features of the polarizable continuum model for the representation of solvation," *WIREs Computational Molecular Science,* vol. 1, pp. 855-867, 2011.

[197] V. Barone, M. Cossi and J. Tomasi, "A new definition of cavities for the computation of solvation free energies by the polarizable continuum model," *The Journal of Chemical Physics,* vol. 107, pp. 3210-3221, 1997.

[198] M. Cossi, V. Barone, R. Cammi and J. Tomasi, "Ab initio study of solvated molecules: A new implementation of the polarizable continuum model," *Chem. Phys. Lett,* vol. 255, pp. 327-335, 1996.

[199] S. Miertuš and J. Tomasi, "Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes," *Chemical Physics,* vol. 65, pp. 239-245, 1982.

[200] S. Miertuš, E. Scrocco and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chemical physics,* vol. 55, pp. 117-129, 1981.

[201] J. M. Herbert, "Dielectric continuum methods for quantum chemistry," *Advanced Review,* vol. 11, pp. Article #e1519, pp. 1-73, 2021.

[202] J. C. Maxwell, "A dynamical theory of the electromagnetic field," *Philosophical Transactions of the Royal Society of London,* vol. 155, pp. 459-512, 1865.

[203] R. K. Wangsness, Electromagnetic fields, New York: John Wiley & Sons, 1986.

Chapter 2

On the prebiotic selection of nucleotide anomers: a computational study[1,2,3]

> "The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)"
>
> H S. Bernhardt,
> *Biology Direct,* vol. 7, pp. 1-10, 2012

This chapter reproduces the content of a scientific paper published in Heliyon in 2022 [1] The format of the original paper, including figures and tables has been rearranged to fit the general format of this PhD thesis dissertation.

Abstract

Present-day known predominance of the *β*- over the *α*-anomers in nucleosides and nucleotides emerges from a thermodynamic analysis of their assembly from their components, *i.e.* bases, sugars, and a phosphate group. Furthermore, the incorporation of uracil into RNA and thymine into DNA rather than the other way around is also predicted from the calculations. Without desregarding other factors, thermodynamics must have driven evolutionary selection of life's building blocks. In this work, based on quantum chemical calculations, we focus on the latter control as a tool for "natural selection".

---

[1] This chapter reproduces a paper published in Heliyon in 2022 [1] while the data related has been published in [74]. The format of the original paper, including figures and tables has been rearranged to fit the general format of this PhD thesis.

[2] All final DFT-optimized molecular structures reported in this chapter are available in: drive.google.com/PhD_Thesis/Optimized_Final_Structures/Chapter#2.

[3] All thermodynamic quantities are represented in kcal/mol since this unit is frequently addopted by biologist, chemists and biochemists.

2.1 Introduction

On what basis did early prebiotic conditions favor the selection and assembly of particular building blocks of nucleic acids? An aspect of this question is the subject of the present investigation. The broader aspects of this question are well-documented as, for instance, reviewed by Šponer et al. [2, 3] and by Serrano Giraldo and Zarante [4], but here we restrict ourselves to a narrower question. Specifically, the majority of contemporary natural nucleic acids, whether DNA or RNA, are polymers of nucleotides in the β-configuration at the C1′ carbon of the furanose sugar and seldom in the α-configuration (Fig. 1), but why? The question is amplified by the ease by which anomers can be synthesized and by their similar ability to form Watson-Crick double helices (**Figure 2.1**).
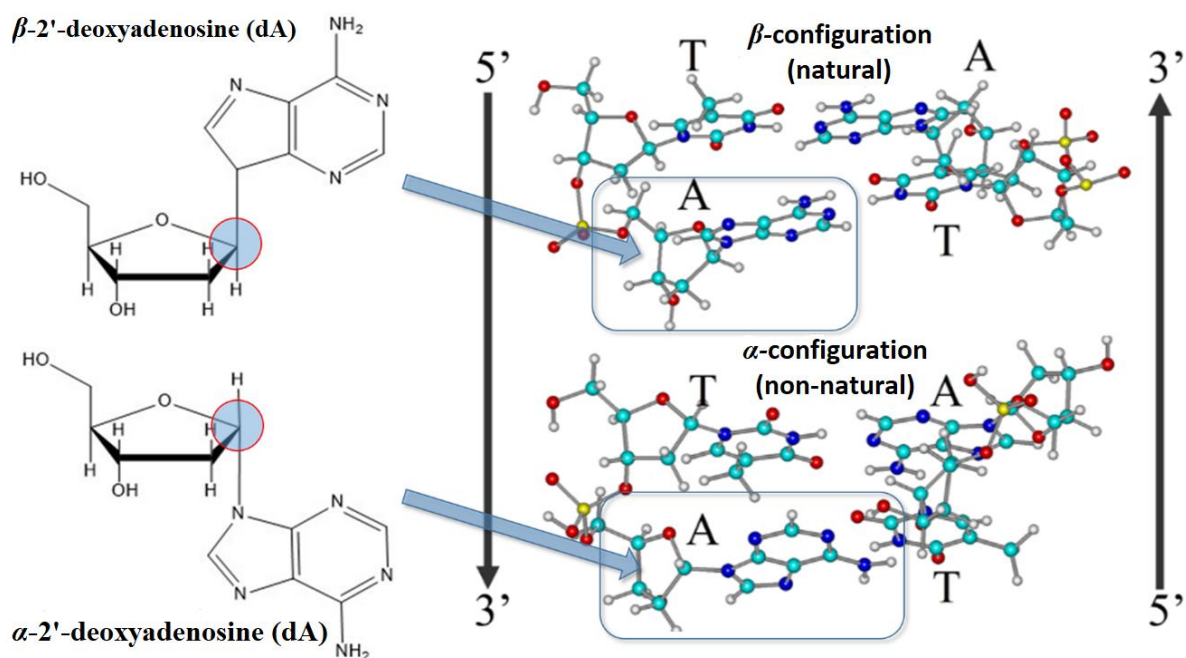
Figure 2.1 Left: (*top*) An example of a β-nucleoside (β-2'-deoxyadenosine), the form that predominates in present day nucleic acids, and (*bottom*) the corresponding α-isomer which is seldom observed. Right: (*top*) A representation of a present-day β-DNA **W**atson & **C**rick (WC) double-helix, and (*bottom*) a model constructed using molecular builders (*HyperChem / GaussView*) demonstrating the perfect *geometric* WC base pairing in the non-predominant form of α-DNA.

Another explored question is the factors that might have driven the selection of thymine for incorporation into DNA and for uracil into RNA. After all, the switching of the bases and sugars seems to be equally likely. Nucleosides, as predominantly existing in today's genetic material, are 2′-**d**eoxy**t**hymidine (dT) in DNA and **u**ridine (U) in RNA rather than as **t**hymidine (T) and 2′-**d**eoxy**u**ridine (dU). Why?

Thermodynamic and kinetic control drive chemical reactions. In this paper, the role of thermodynamics as a driver of evolutionary selection is being explored. The idea of thermodynamics-driven natural selection has been applied, for example, to explain the origin of the genetic code by Grosjean and Westhof [5] and by Klump, Völker, and Breslauer [6]. Here, this line of thought is extended to inquire whether the present-day forms of the building blocks of nucleic acids are, at least, aligned with such an energetic argument.

Clearly a natural thermodynamic selection of the β- over the α-configuration or of the "correct" choice of U/T for the proper nucleic acid category, represent restricted questions from a vast repertoire of possible ones. For example, one may wish to inquire into the evolutionary pressures that picked the present-day particular arrangement of nucleic acids in terms of a pentose sugar, a phosphate group, and a nitrogenous base - rather than - say - having a 2-aminoethylglycine as a linker, as occurs in **p**eptide **n**ucleic **a**cids (PNAs) [7], instead of the phosphodiester backbone [8]? No one knows. One can question Nature's Central Dogma (DNA↔DNA→RNA→protein) [9] and whether this is the only conceivable way to produce living systems, etc. Clearly these wider questions are of utmost importance to understand the origins of life but are vastly larger than the scope of this investigation.

In 1955, Kaplan et al. [10] reported the isolation of a compound that had the same composition as **N**icotinamide **A**denine **D**inucleotide (NAD+), the latter was termed diphosphopyridine nucleotide at the time. Subtle deviations in the properties of this compound from those anticipated for NAD+ led these authors to conclude that it is an isomer of NAD+. Indeed, the compound discovered in 1955 is the α-isomer of the NAD+ (which has a β-configuration at the glycosidic bond) [10]. A decade later, Suzuki and co-workers [11] isolated bacterial *Azotobacter vinelandii* α-NAD, α-nicotinic acid adenine dinucleotide, α-NADP, and α-nicotinic acid mononucleotide. The latter work shows that, while much less frequent, the α-form does occur indeed in living systems [11].

Paoletti et al. report the experimental synthesis of an α-β complex between a α-d(CCTTCC) hexanucleotide and its complementary β-d(GGAAGG) [12]. A comparison of the formation of this complex with its natural β analog (β-d(CCTTCC) + β-d(GGAAGG)) reveals that the formation of the non-natural form is only $\approx$1 kcal/mol more favored than its natural counterpart [12]. There are other reports of synthesis of nucleic acids containing α-nucleic acids strands [13, 14, 15, 16].

Kaur, Sharma, and Wetmore (KSW) have proposed barbituric acid and melamine [17] and **c**yanuric **a**cid (CA) and 2,4,6-**tria**mino**p**yrimidine (TAP) [18] as precursors of prebiotic RNA on the basis of quantum mechanical calculations and **m**olecular **d**ynamics (MD) simulations. In their more recent paper [18], KSW use density functional theory (DFT) calculations to obtain potential energy surfaces describing the rotation around the glycosidic bonds of β- and α-ribonucleosides of the non-canonical nucleobases TAP and CA as well as their complementary base pairing TAP:CA and stacking energies. Additionally, these authors studied the base pairing of these nucleobases with the canonical nucleobases (A, G, C, T and U) and compared the canonical 10-mer A-form of RNA duplexes 5′ -AAAAAAAAAA3' paired with 5′ -UUUUUUUUUU-3′ and 5′ -AAAGCGCAAA-3′ paired with its complementary 5′ -UUUCGCGUUU-3′ with the non-canonical duplexes 5′ -AAAXXYYAAA-3′ paired with 3′ -UUUYYXXUUU-5', where X = CA and Y = TAP. The results obtained suggest that the strength for the hydrogen bonds created in the TAP:CA pairing is comparable to the canonical base pairing. The stacking of these non-canonical bases is, on the other hand, weaker compared to the canonical stacking, suggesting that the enhanced stacking may have served as a driving force in the evolution of nucleic acids. Finally, the assembled structure of TAP-CA containing helices suggests that this type of pre-RNA could have been shielded from water allowing its evolution and self-replication. The results in that paper, [18], place TAP and CA as plausible candidates for a pre- or proto-RNA.

In their earlier study, KSW computational results on **b**arbituric **a**cid (BA) and **m**ela**m**ine (MM) suggest their plausibility as non-canonical nucleic acid bases that may have been present in the precursor of present-day nucleic acids [17]. The authors find that the strength of the hydrogen bonds between BA and MM makes them good candidates as building blocks of ancestral nucleic acids. On the other hand, they find that the stacking interactions were stronger when either BA or MM are combined with a canonical nucleobase than when the stacking was between each other. These results suggest the possibility of the existence of a pre- or proto-RNA that mixes canonical and non-canonical complementary nucleobases within one structure [17].

In their first paper, KSW report that the potential energy hypersurface of breaking the glycosidic bond is consistent with a stronger bond in TAP nucleosides compared to canonical nucleosides. Interestingly, in the case of the CA the opposite result is obtained [18]. KSW found larger deglycosylation barriers for the C-C glycosidic bond of BA-ribonucleosides compared to canonical nucleobases while the reverse occurs in the case of MM [17].

The biopolymers of life are believed to have emerged between 3.5 and 4 billion years ago [19, 20], with details still to be worked-out. For instance, which was first: Proteins or nucleic acids? The consensus is that nucleic acids were probably the first biomolecules, specifically RNA in what is commonly known as the "RNA world hypothesis" [21, 22], since RNA is both a catalyst and a repository of genetic information making it candidate for the first biopolymers [23, 24, 25].

If we accept that RNA came first, then other questions arise. For example, how did this molecule originate in the first place? It has been proposed that nucleic acids were the product of prebiotic and geochemical reactions, namely, the "drying pool", "drying lagoon", also known as the "classic model", whereby regular cycles of dehydration-re-hydration can promote the polymer formation. Given the important role of the so-called "water problem" in early evolution, whereby H_2O impedes the synthesis of nucleic acids [26, 27, 28, 29], in this work both vacuum phase and aqueous phase calculations were conducted.

Within the classic model, Orgel and coworkers explored the formation of the glycosidic bond between purines (adenine, guanine, inosine, xanthine) and ribose sugar by drying and heating the reactants in the presence of catalysts [30]. Only adenine was found to couple with the ribose to produce α- and β-furanosyl nucleosides with yields typically within $\approx$2–10%. The relatively low yield by Orgel and coworkers were attributed to the instability of the glycosidic bond in aqueous environment giving rise to what is known as the "nucleoside problem" (a special case of the more general "water problem") [26, 27, 28, 29] . Challenges including the nucleoside problem have led scientists to look for alternative synthetic routes that start, for example, from phosphorylated sugars and free bases [31, 32, 33, 34, 35, 36].

The hypothesis being tested here is that Nature's stereo-selection of the present-day canonical nucleosides/nucleotides is consistent with an energy-driven natural selection. Thus, the two anomeric forms of the nucleosides/nucleotides were studied as they occur within both DNA and RNA and compared for their thermodynamic stabilities.

In final account, since the deamination of cytosine transforms it into uracil with a consequential change in the genetic message, selection pressures may have driven the transition from uracil to thymine in DNA [37]. This question has long been raised [38] but appears to have never been resolved on the grounds of relative total molecular energies to the best of the authors' knowledge. Since there appears to be no a priori reason for Nature to have selected thymine to be incorporated exclusively in DNA and uracil in RNA (with exceptions), mixed ("wrong", or non-canonical) nucleotides have also been considered in our calculations as an available choice for natural selection.

To sum-up, we observe a preponderance of β- over the α-anomers in present-day nucleosides(tides). This work indicates that such preponderance is consistent with thermodynamic parameters calculated quantum chemically within the assumptions and approximations of the study. The known specificity of uracil to RNA and thymine to DNA is also consistent with these results. While both kinetics and thermodynamics, and the interactions with solvent molecules and ions, must have all driven together evolutionary selection of life's building blocks, in this work we restrict the question as to whether there exists an inherent preference in the building block themselves that is consistent with the observed "naturally selected" present day nucleic acids.

2.2   Computational details

The structures of each nucleoside (sugar + nitrogenous base) in the two anomeric forms (β and α) were constructed with the graphic interfaces of Hyperchem 7.0 [39] and GaussView 5.0 [40]. The initial 20 nucleosidic structures {[(2 sugars (ribose, and 2'-deoxyribose)) × (5 bases (**A**denine (A), **T**hymine (T), **G**uanine (G), **C**ytosine (C), and **U**racil (U)))] × 2 (β and α)} were subjected individually to a soft **p**otential **e**nergy hyper**s**urface (PES) scan with respect to the torsion angle that governs the N-glycosidic bond between the base and the sugar. "Soft scan" means that the only constrain is the angle being scanned stepwise while all other geometrical parameters are allowed to relax in response to that angle.

The above-mentioned PES scans were performed in the following way: *Z*-matrices for the ribofuranose and 2'-deoxyribofuranose sugar, each in both the β- and α-configuration, were read into the programme Granadarot [41, 42]. The Granadarot

algorithm was used to create, for the ribofuranose, 1,000 different conformers by randomly varying the five dihedral angles that involve all the sugar's hydroxyl group (4 angles of the H-O-C-C type, in addition to the O-C5'-C4'-C3' angle – also known as β angle (not to be confused with the anomeric label)). For the 2'-deoxyribofuranose, a similar procedure was used to also generate 1,000 conformers with the difference that now there are only 3 angles of the H-O-C-C type. The number of initial random conformers (1,000) strikes a balance between a good sampling of the PES and computational costs.

For each of these initial randomized sugar structures (4,000 in total), the geometries were then optimized at the semiempirical PM7 level of quantum chemical theory, a method that includes empirical corrections for dispersion and hydrogen bonding interactions [43, 44]. By including these corrections, PM7 has an important advantage over other semiempirical methods that generally do not account for dispersion and hydrogen bonding explicitly. The PM7 geometry optimizations were conducted through a gradient minimization of the total energy using the **M**olecular **O**rbital **PAC**kage (MOPAC)2016 package [45] until the forces on the nuclei were negligible.

The geometry optimizations at the semiempirical level were performed twice: once in vacuum and a second time with the **CO**nductor-like **S**creening **MO**del (COSMO) continuum solvation model [46, 47].

Several of the sets of 1,000 optimizations described above converge to one and the same respective final geometry (Tanimoto index ≤ 0.85). The number n of final unique optimized geometries is: *in vacuum:* β-ribofuranose ($n = 28$), α-ribofuranose ($n = 34$), β-2'-deoxyribofuranose ($n = 42$), α-2'-deoxyribofuranose ($n = 16$); *in solvent:* β-ribofuranose ($n = 110$), α-ribofuranose ($n = 84$), β-2'-deoxyribofuranose ($n = 78$), α-2'-deoxyribofuranose ($n = 65$).

For each of one of these 8 systems, the n' conformers that collectively contribute at least 50% to the partition function (Z) were kept for refinement with more accurate calculations and the rest of the conformers with minor contributions were discarded. This reduced the number of investigated conformers to: *in vacuum:* β-ribofuranose ($n' = 4$, contributing 59.5% of Z), α-ribofuranose ($n' = 2$, contributing 60.4% of Z), β-2'-deoxyribofuranose ($n' = 3$, contributing 54.6% of Z), α-2'-deoxyribofuranose ($n' = 1$, contributing 51.0% of Z); *in solvent:* β-ribofuranose ($n' = 12$, contributing 51.4% of Z), α-

ribofuranose ($n' = 12$, contributing 51.6% of Z), β-2'-deoxyribofuranose ($n' = 7$, contributing 53.0% of Z), α-2'-deoxyribofuranose ($n' = 7$, contributing 51.9% of Z).

The geometry of every sugar structure of the n' that survived the initial screening, was re-optimized without constraints at the **D**ensity **f**unctional level of **t**heory (DFT) [48, 49, 50] using the B3LYP/6-31G(*d,p*) functional [51, 52] / basis set [53] combination.

Aqueous solvation was accounted for in the DFT calculations using the **I**ntegral **E**quation **F**ormalism variant of the "*Polarizable Continuum Model*" (IEFPCM) [54, 55, 56, 57, 58] implemented in Gaussian 16 [59], the software package used in all DFT calculations in this work.

The so-called "water problem" [60] describes the consensus understanding that the primordial soup may have been non-polar in nature or, at least, had a controlled exposure to water (see for example Ref. [61] and literature cited therein). Hence, the primary results to be considered and discussed here are those in the vacuum phase as a surrogate for non-polar environment. However, solvation modeling has been included for completion and to also test the effects of this very "water problem" since aqueous medium predominates in contemporary living systems.

Solvation remains an open problem for quantum chemical calculations. One has to pick from the dichotomy of *explicit solvation* or the modeling of its effects by placing the solvent in a cavity within a bulk dielectric continuum; *implicit solvation* [55, 62, 56, 63, 64, 65, 66]. Explicit solvation is best to describe strong localized interactions between the solute molecule and immediate solvation shell molecules, while continuum solvation modeling is better tuned to capture the long-range averaged effects of solvation by the solvent bulk. Explicit solvation, ideally, entails a gradual addition of solvent molecules until the convergence of some relevant parameters, which is impractical here in view of the large number of studied systems. Hence, the second best option, that is, the continuum modeling, has been applied in this work.

Every geometry optimization in this work has been followed by a harmonic vibrational analysis and all cases were found to exhibit no imaginary frequency as required to confirm their status as local minima on the PES. For each one of the eight groups described above, the most stable structure of the sugar that resulted from the DFT optimization was saved for the next step and the rest of the structures were discarded.

The five nucleobases (A, G, C, T, U) were optimized at the DFT-B3LYP/6-31G(d, p) level of theory, in a similar procedure as the one outlined above for the sugars, in vacuum and in solvent phase. These optimized structures were then stitched to the sugars leading to 40 separate initial N-nucleoside guess geometries (5 bases × 2 sugars × 2 configurations × 2 phases/environments).

For consistency, the N-glycosidic bond, C1'-N1 in pyrimidines (Y) or C1'-N9 in purines (R) was initially set to 1.52 Å while the dihedral angle H-C1'-N1(Y)/N9(R)-Cx was set initially at -161.9° (anti). Each of the 40 nucleoside structures was then subjected to a fully counterclockwise relaxed scan around this dihedral in 6 steps each of 60°. The lowest structure from this scan was refined by subjecting it to a final fully-unconstrained optimization to obtain the final structure of the nucleosides in vacuum and in solvent. A harmonic frequency calculation was performed as usual to ensure that the final structures are indeed minima on the PES.

Finally, a mono-anionic dihydrogen phosphate group ($H_2PO_4^-$), the form dominant at pH 6.5 [3], has been subjected to an unconstrained optimization in the vacuum phase and in continuum solvent at the DFT-B3LYP/6-31G(d, p) level of theory. The optimized phosphate was then attached to the nucleosides at C5'-OH setting the initial O-C5'-C4'-C3 torsion angle to 30.9° (the standard angle for the stored structures in GaussView). A soft scan was then performed in 6 steps of 60 degrees, and again the lowest energy conformer of the nucleotide was retained for further analysis. The procedure outlined above yields 80 final optimized structures: 40 for each class (nucleoside, and nucleotide), 20 in gas- and 20 in solution-phase.

The steps described above proceed for the following order of condensation reactions: *sugar + base → nucleoside* followed by a geometry optimization of the nucleoside, and then by the reaction *nucleoside + phosphate → nucleotide* followed by a geometry optimization of the nucleotide.

Since every step of these two concatenated "reactions" is followed by a geometry optimization, a change in the order of these reactions does not necessarily yield the same geometries (and energies). Hence, the procedure described so far has been repeated except by reversing the order of the reaction, that is: *sugar + phosphate → 5'sugar-monophosphate → optimization → 5'sugar-monophosphate + base → nucleotide → optimization.*

For every given pair of α/β-anomers, the difference in their energies (relative energy) is defined:

$$\Delta X_{\beta\alpha} \equiv X_\beta - X_\alpha \qquad\qquad (2.1)$$

where ΔX denotes ΔE (the difference is of the total energies), or $\Delta E_{(ZPE)}$ (the difference in the total energies corrected for **Z**ero-**P**oint vibrational **E**nergies (ZPEs)), or ΔG^o (the difference in the Gibbs energies at standard conditions). The inclusion of solvation effects will be indicated when necessary using extra symbols.

The DFT-B3LYP/6-31G(d, p) level of theory has been chosen for this study as a reasonable compromise of accuracy and speed/feasibility. To have a sense of the magnitude of the error bars intrinsic to this level of theory would require a full-fledged benchmarking study which is outside of the scope of this work. The error for a similar level of theory, namely, DFT-B3LYP/6-31+G(d,p) has been benchmarked by Zhao and Truhlar to be around 3.6 kcal/mol [67]. Zhao and Truhlar obtained this estimate by comparing the calculated and experimental thermodynamic data for 177 main-group compounds. On that basis, we may take the intrinsic uncertainty of the method used in this work to be around $\approx$ 3-4 kcal/mol.

2.3   Results and discussion

2.3.1   Two hypothetical synthetic pathways and their Gibbs energies

The two pathways for the formation of a nucleotide are depicted in **Figure 2.2** and are *not* equivalent. That non-equivalency is due to the effect of the first condensation on geometry which leads to final products trapped at different wells on the potential energy surface of the nucleotide. In other words; (a+b) and (c+d) are *different* pathways with *different* $\Delta G^{\circ}_{\text{reaction}}$.
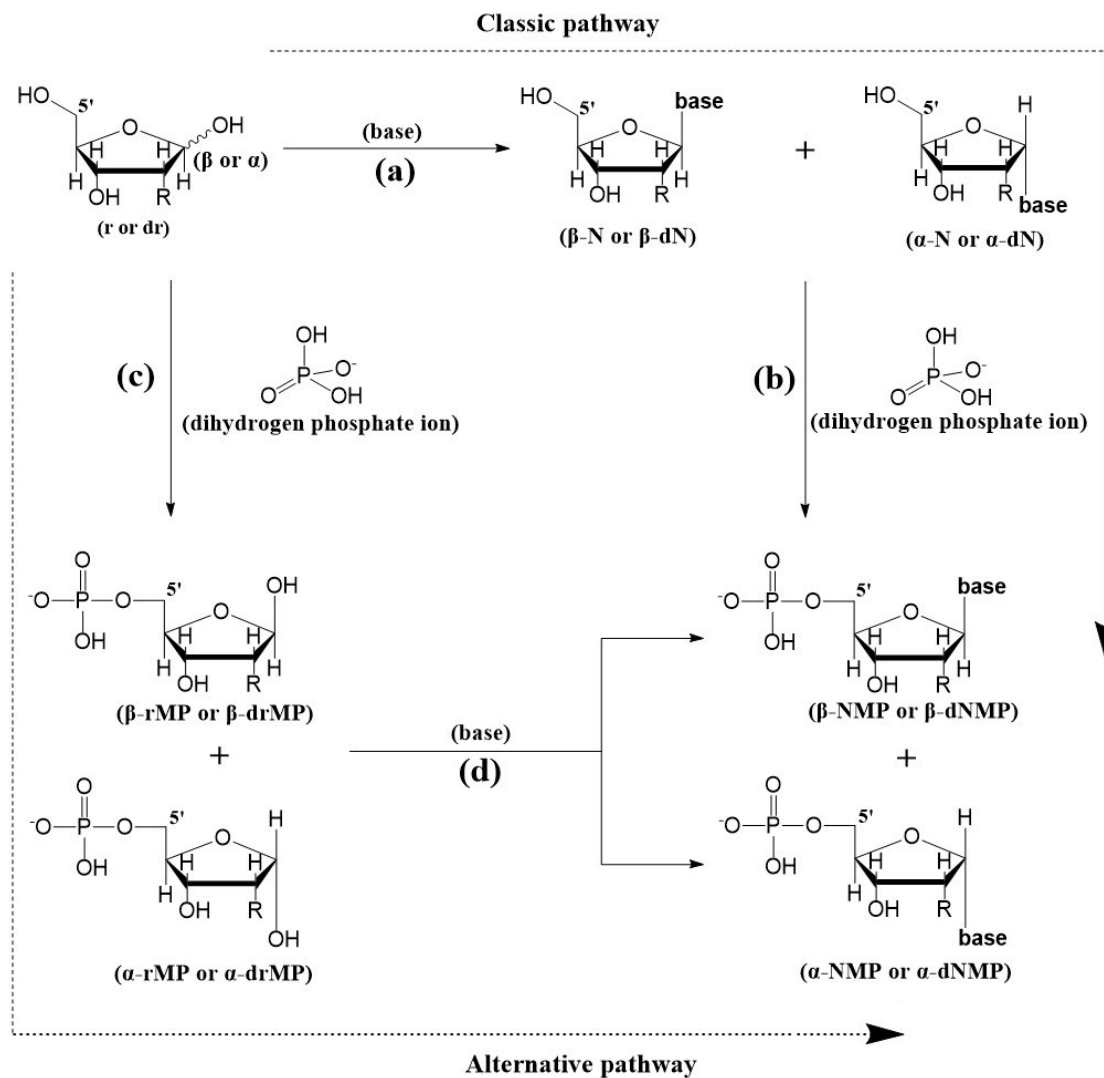
Figure 2.2 Two different pathways for constructing the *β*- and *α*-anomers of nucleosides and nucleotides. (Reaction pathways referred-to in the text and tables are labeled with lower-case letters: Classical pathway (a+b) and alternative pathway (c+d); R=H,OH for DNA and RNA, respectively).

Recapping, the two condensation sequences considered are:

(1) The condensation of a sugar with a base to obtain the N-nucleoside (Ns) followed by the condensation of the nucleoside with a dihydrogen phosphate anion ($H_2PO_4^-$) at the 5' position to obtain the nucleotide (Nt). The Gibbs free energy of this reaction is defined as (equation (2.2)):

$$\Delta G^{\circ}_{(a+b)} = \left[(G^{\circ}_{Ns} + G^{\circ}_{H_2O}) - (G^{\circ}_{sugar} + G^{\circ}_{base}) \right]_a + \left[(G^{\circ}_{Nt^{\circ}} + G^{\circ}_{H_2O}) - \left(G^{\circ}_{Ns} + G^{\circ}_{H_2PO_4^-} \right) \right]_b \qquad (2.2)$$

(2) The condensation reaction of a $H_2PO_4^-$ group with a sugar at C5' first, *followed* by a condensation of the sugar-5'-monophosphate with the base. In this case the Gibbs energy of reaction is defined as (equation (2.3)):

$$\begin{aligned}
\Delta G^{\circ}_{(c+d)} = &\left[\left(G^{\circ}_{sugar-5'-HPO_4^-} + G^{\circ}_{H_2O} \right) - \left(G^{\circ}_{sugar} + G^{\circ}_{H_2PO_4^-} \right) \right]_c \\
&+ \left[(G^{\circ}_{Nt} + G^{\circ}_{H_2O}) - \left(G^{\circ}_{sugar-5'-HPO_4^-} + G^{\circ}_{base} \right) \right]_d
\end{aligned} \qquad (2.3)$$

2.3.2 Which furanose or furanose-phosphate anomers are more stable: α or β?

As a prelude to this study, we first revisit the relative stabilities of the furanose forms of the sugars themselves. A ^{13}C-NMR study complemented with a statistical mechanics analysis by Dass et al. demonstrates a strong temperature-(gradient)-dependence of the equilibrium ratios of the various open-chain and cyclic forms of ribose sugar [68]. These authors are simulating the conditions of temperature and temperature-gradient near hydrothermal vents to answer the question of which form(s) of the ribose sugars were favored at early prebiotic times. From these authors' Fig. 2 [68], in pure aqueous solution at 25 °C, the *β-pyranose* (βP-form) is dominant with a mole fraction of ≈ 0.6, followed by the αP-form with a mole fraction of ≈ 0.2. These authors' figure indicates much lower populations for the two furanose forms under these conditions, both anomers having similar mole fractions of ≈ 0.1 each. This means that at 25 °C the pyranose form dominates largely and especially in its β-form. These figures do not change significantly when the medium simulates Hadean waters [68].

Interestingly, Fig. 2 of [68] features break-even points of the eight curves. Beyond a temperature of ≈ 130 °C, the population is inverted with an increasing dominance of the furanose form, starting with a small excess of $\approx 1.5 \times$ favoring the β form at the beginning, with a gap between the α and the β populations that widens as the temperature increases (whether in pure or Hadean water) reaching a β/α ratio of ≈ 2 at ≈ 130 °C [68]. It is inferred, in conclusion, that the early hot atmosphere may have been the driver for the selection of the β-furanose form that remained to this day after the temperatures have cooled down.

This proclivity to select the β-furanose form can be enhanced at lower temperature in the presence of large temperature gradients as it occurs near hydrothermal vents for example. At room temperature, however, the data of Dass et al. show a slight but not substantial propensity for the β- over α-furanose, whether in pure aqueous solution or one that simulate Hadean medium [68].

Azofra et al.'s [69] DFT exploration of the potential energy landscape generated thousands of rotamers of (deoxy)ribopyranose, (deoxy)ribofuranose in their open chain and β- and α-anomeric forms. In their study, these authors report results based on DFT vacuum-phase calculations with both the B3LYP/6-311++G(d, p) and M06-2X/6-311++G(d, p) chemical models [69]. The authors also find a dominance of the pyranose form over the furanose from 0 K to room temperature (298 K), in agreement with the more recent experimental results of Dass et al. [68]. Meanwhile, within the small proportion of furanose, in the case of ribofuranose, the B3LYP/M06-2X functional predicts a higher Boltzmann populations of α-ribofuranose (3.4/0.2% (298 K) and 1.5/0.1% (0K)) than the β-ribofuranose (0.6/0.0% (298 K) and 0.2/0.0 (0K)). A similar trend is also found for the α-2′-deoxyribofuranose forms, in which case the respective Boltzmann populations (B3LYP/M06-2X) are (6.2/0.8% (298 K), 2.4/0.3% (0K)) compared with β-2'-deoxyribofuranose (1.2/0.1% (298 K) and 0.4/0.0 (0K)). One of the main results of Azofra et al.'s [69] study is that the pyranose form is more populated than the furanose form at room (and lower) temperatures, but the results using both DFT functionals indicate a slight advantage, within the furanose population, for the α-anomer. However, the differences in energies and their corresponding Boltzmann's populations are probably within the error bars of the DFT calculations. Hence, we may conclude that these studies do not indicate a clear advantage of one furanose anomer over the other at room temperature.

Cocinero et al. performed a combined experimental/computational study (rotational FT-MW spectroscopy and three levels of theory: MP2/6-311++G(d, p), B3LYP//6-311++G(d, p), and M06-2X//6-311++G(d, p)) that again shows the almost exclusive dominance of the pyranose form in the gas-phase at room temperature [70]. However, these authors also demonstrate that aqueous solvation increases the propensity for the furanose form at room temperature compared to the gas-phase [70].

Our results listed in **Table 2.1** are consistent with the findings of Azofra et al.'s [69] suggesting a borderline preference of the α- over the β-furanose anomer at room temperature,

whether in solvent or in the vacuum phase, and for either ribose or deoxyribose (see first three entries of **Table 2.1**, especially the third, for differences in Gibbs energy which are less than *ca.* 3 kcal/mol). In fact these results are consistent with all those mentioned above by other workers since they indicate an inconclusive advantage of one form or another.

        Table 2.1 also lists the effect of adding the phosphate group at position 5' of the sugar. *The phosphate group has a significant effect whereby the marginal preference of the α- over the β- forms of the free sugars is now reversed.* As can be seen from the entries in the Table, the sugar monophosphates are slightly more stable in the β-forms, whether in gas- or solution- phase and whether ribose or deoxyribose (differences in Gibbs energy are all above a kcal/mol, approximately 2 - 3 kcal/mol for RNA and 4 - 5 kcal/mol in DNA, in vacuum, and with smaller

Table 2.1 Differences in energies between the most stable β- and α-anomers for the sugars 2'-deoxy (d) or (r)ibose in vacuum and in aqueous environment in kcal/mol (equation 2.1). Included differences are between: The total energies without ($\Delta E_{\beta\alpha}$) and with zero-point vibrational correction (ZPE) ($\Delta E_{\beta\alpha(\mathrm{ZPE})}$), and Gibbs energies ($\Delta G^{\circ}_{\beta\alpha}$) at STP conditions. The listed results are from DFT (B3LYP/6-31G(*d,p*)) calculations. The integral equation formalism of the polarizable continuum model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	DNA			RNA		
	$S^{(1)}$	vac.$^{(2)}$	solv.$^{(3)}$	$S^{(1)}$	vac.$^{(2)}$	solv.$^{(3)}$
$\Delta E_{\beta\alpha}$	dr	1.1	4.1	r	2.8	3.8
$\Delta E_{\beta\alpha(\mathrm{ZPE})}$		1.2	3.7		2.7	2.8
$\Delta G^{0}_{\beta\alpha}$		1.5	3.2		2.8	1.9
$\Delta E_{\beta\alpha}$	drMP	-5.1	-2.4	rMP	-3.0	0.8
$\Delta E_{\beta\alpha(\mathrm{ZPE})}$		-4.9	-2.5		-3.2	-0.3
$\Delta G^{0}_{\beta\alpha}$		-3.9	-1.1		-1.8	-0.8

(1) S = unspecified sugar or 5'-monophosphate (MP) sugar. [2] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\mathrm{ZPE})}$, and in $\Delta G^{\circ}_{\beta\alpha}$ $\Delta G^{0}_{\beta\alpha}$ in vacuum at the DFT level. [3] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\mathrm{ZPE})}$, and in $\Delta G^{\circ}_{\beta\alpha}$ in solvent.

magnitudes (but still negative values) in solution) (**Table 2.1**). (See figures A1 - A4 in the Appendices section).

2.3.3 Which nucleoside anomer is more stable: α or β?

Ball-and-stick representations of the optimized geometries of all studied structures and their Gibbs energies of inter-conversion can be found in the Appendices (figures A.5 - A.14).

The 20 differences in energies between the β- and α-anomers for each of the nucleosides obtained from equation (1) are listed in **Table 2.2**. For most cases, the difference in stabilities of the anomers falls within the probable error bars of the theoretical method (DFT-B3LYP/6-31G(d, p)), that is, approximately 3 - 4 kcal/mol. In the case of DNA, all differences in Gibbs energies are < 2 kcal/mol, whether in vacuum or in solution and for all five nucleosides. One notices that, in all cases, solvation magnifies the relative stability of the α-anomer by ≈ 2 kcal/mol. As for RNA in vacuum, Gibbs energies suggest a slight relative stability of the β-anomer of G over the α-form (by ≈ 2 kcal/mol), while the reverse is true for the rest, with the α-form being more stable by ≈ 2 kcal/mol for A, T, and U, and by ≈ 4 kcal/mol for C.

In solution phase, and judging from the relative ΔG values, the α-forms of the purines are more stable than their β counterparts by ≈ 2 kcal/mol while for pyrimidines the differences between the two forms are negligible (below chemical accuracy of 1 kcal/mol), **Table 2.2**. These energy differences between the anomers are within an order of magnitude of the thermal energy at room temperature $K_B T$ (298 K) ≈ 0.5 kcal/mol, and that at 70 C ($K_B T$ (343 K) ≈ 0.7 kcal/mol), a temperature believed to have prevailed during the Archean eon when the first forms of primordial life may have emerged [71].

The calculated energy differences listed in **Table 2.2**, whether Gibbs or total energies, fall within the probable error bars of the method and hence, while indicative, cannot be considered definitive. The consistency of the trend in **Table 2.2** may allow us to conclude that there appears to be a small thermodynamic advantage for the α-anomer over the β-anomer in solution in both nucleic acids. In vacuum phase, in the case of DNA, the Gibbs energy differences are below chemical accuracy and hence the two forms are iso-energetic. Meanwhile, for RNA in vacuum generally the α-anomer is more stable (slightly for A, T, and U, and more notably for C) except for G for which the β-anomer has a slight advantage.

125

Table 2.2 Differences between the energies of the most stable β- and α-anomers of the 2' deoxy (d) or ribonucleosides in vacuum and in aqueous environment (equation 2.1). Included differences are between: The total energies without ($\Delta E_{\beta\alpha}$) and with zero-point vibrational correction (ZPE) ($\Delta E_{\beta\alpha(\text{ZPE})}$), and Gibbs energies ($\Delta G^\circ_{\beta\alpha}$). All results are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The integral equation formalism of the polarizable continuum model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	DNA			RNA		
	$N^{(1)}$	vac.$^{(2)}$	solv.$^{(3)}$	$N^{(1)}$	vac.$^{(2)}$	solv.$^{(3)}$
			purines			
$\Delta E_{\beta\alpha}$	dA	1.2	0.6	A	1.7	3.9
$\Delta E_{\beta\alpha(\text{ZPE})}$		0.8	0.7		1.3	3.2
$\Delta G^0_{\beta\alpha}$		1.0	1.4		1.5	2.1
$\Delta E_{\beta\alpha}$	dG	-0.9	0.3	G	-1.8	3.6
$\Delta E_{\beta\alpha(\text{ZPE})}$		-1.1	0.7		-1.9	3.1
$\Delta G^0_{\beta\alpha}$		-0.5	2.0		-1.5	2.0
			pyrimidines			
$\Delta E_{\beta\alpha}$	dC	-1.4	2.4	C	4.9	-0.1
$\Delta E_{\beta\alpha(\text{ZPE})}$		-1.1	2.2		4.6	-0.5
$\Delta G^0_{\beta\alpha}$		-0.2	1.8		4.3	-0.8
$\Delta E_{\beta\alpha}$	dT	-0.4	2.8	T	2.4	1.6
$\Delta E_{\beta\alpha(\text{ZPE})}$		-0.3	2.5		1.9	1.0
$\Delta G^0_{\beta\alpha}$		0.3	1.9		1.9	0.1
$\Delta E_{\beta\alpha}$	dU	0.0	2.9	U	2.3	1.5
$\Delta E_{\beta\alpha(\text{ZPE})}$		0.1	2.6		1.8	1.2
$\Delta G^0_{\beta\alpha}$		0.7	2.0		1.8	0.8

[1] N = unspecified nucleoside. [2] Difference in total energies in vacuum at semiempirical level. [3] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^0_{\beta\alpha}$ in vacuum at the DFT level. [4] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^0_{\beta\alpha}$ in solvent.

2.3.4 Which nucleotide anomer is more stable, in what conditions?

βα

Tables 2.3 and **2.4** give the relative (Gibbs) energies of the α and β- anomers obtained *via* the "classical" pathway ((a + b) – **Table 2.3**, **Figures A.15 - A.24**) and the alternative pathway ((c + d) – **Table 2.4**, **Figures A.25 - A.34**). The sequences of the two-step additions

defining the two pathways are represented in **Figure 2.2**. The differences in energies were obtained by equations of the form of equation (1). Since all individual energies are negative, entries in these tables that are negative indicate that the β anomer is more stable and *vice versa*.

Table 2.3 Differences between the energies of the most stable β- and α-anomers of the 2' deoxy (d) or (rib)onucleotides in vacuum and in aqueous environment (equation 2.1) as given by the reaction pathway sequence (a) and (b) of **Figure 2.1**. Included differences are between: The total energies without ($\Delta E_{\beta\alpha}$) and with zero-point vibrational correction (ZPE) ($\Delta E_{\beta\alpha(\text{ZPE})}$), and Gibbs energies ($\Delta G^{\circ}_{\beta\alpha}$). All results are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The Integral Equation Formalism of the Polarizable Continuum Model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	DNA			RNA		
	NMP[1]	vac.[2]	solv.[3]	NMP[1]	vac.[2]	solv.[3]
	purines					
$\Delta E_{\beta\alpha}$	dAMP	-7.9	2.0	AMP	-8.0	-0.6
$\Delta E_{\beta\alpha(\text{ZPE})}$		-7.9	2.2		-8.9	-0.2
$\Delta G^{0}_{\beta\alpha}$		-7.7	2.4		-9.9	1.5
$\Delta E_{\beta\alpha}$	dGMP	-18.6	1.8	GMP	-5.7	-2.8
$\Delta E_{\beta\alpha(\text{ZPE})}$		-18.7	1.1		-6.6	-2.3
$\Delta G^{0}_{\beta\alpha}$		-17.1	1.8		-6.7	-0.8
	pyrimidines					
$\Delta E_{\beta\alpha}$	dCMP	4.7	-0.7	CMP	-8.0	-4.6
$\Delta E_{\beta\alpha(\text{ZPE})}$		4.2	-0.9		-8.6	-4.5
$\Delta G^{0}_{\beta\alpha}$		3.3	-0.6		-9.0	-3.2
$\Delta E_{\beta\alpha}$	dTMP	5.9	-1.9	TMP	-1.5	-4.7
$\Delta E_{\beta\alpha(\text{ZPE})}$		6.0	-1.1		-2.2	-4.5
$\Delta G^{0}_{\beta\alpha}$		6.4	0.3		-2.5	-2.8
$\Delta E_{\beta\alpha}$	dUMP	6.1	-1.3	UMP	-10.7	-4.4
$\Delta E_{\beta\alpha(\text{ZPE})}$		6.1	-1.3		-11.6	-4.0
$\Delta G^{0}_{\beta\alpha}$		6.6	-0.7		-12.5	-2.2

[1] NMP = unspecified nucleoside 5'-monophosphate (nucleotide). [2] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^{0}_{\beta\alpha}$ in vacuum at the DFT level. [3] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^{0}_{\beta\alpha}$ in solvent.

Table 2.4 Differences between the energies of the most stable β- and α-anomers of the 2' deoxy (d) or (rib)onucleotides in vacuum and in aqueous environment (equation 2.1) as given by the reaction pathway sequence (c) and (d) of **Figure 2.1**. Included differences are between: The total energies without ($\Delta E_{\beta\alpha}$) and with zero-point vibrational correction (ZPE) ($\Delta E_{\beta\alpha(\text{ZPE})}$)), and Gibbs energies ($\Delta G^{\circ}_{\beta\alpha}$). All results are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The Integral Equation Formalism of the Polarizable Continuum Model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	DNA			RNA		
	NMP[1]	vac.[2]	solv.[3]	NMP[1]	vac.[2]	solv.[3]
	purines					
$\Delta E_{\beta\alpha}$	dAMP	4.3	1.4	AMP	1.6	-0.5
$\Delta E_{\beta\alpha(\text{ZPE})}$		4.2	1.5		1.8	-0.8
$\Delta G^{0}_{\beta\alpha}$		4.3	1.6		2.6	-0.8
$\Delta E_{\beta\alpha}$	dGMP	-3.5	-2.1	GMP	0.2	2.3
$\Delta E_{\beta\alpha(\text{ZPE})}$		-4.5	-1.6		0.0	1.0
$\Delta G^{0}_{\beta\alpha}$		-4.3	1.2		0.9	-0.2
	pyrimidines					
$\Delta E_{\beta\alpha}$	dCMP	-12.6	0.2	CMP	0.0	0.7
$\Delta E_{\beta\alpha(\text{ZPE})}$		-12.4	0.1		-0.7	0.0
$\Delta G^{0}_{\beta\alpha}$		-12.8	0.7		-1.6	-0.3
$\Delta E_{\beta\alpha}$	dTMP	-12.2	0.0	TMP	6.7	1.3
$\Delta E_{\beta\alpha(\text{ZPE})}$		-11.5	0.3		6.8	0.7
$\Delta G^{0}_{\beta\alpha}$		-10.6	1.7		7.5	0.1
$\Delta E_{\beta\alpha}$	dUMP	-10.0	1.1	UMP	9.3	-1.1
$\Delta E_{\beta\alpha(\text{ZPE})}$		-10.1	1.1		9.9	-1.8
$\Delta G^{0}_{\beta\alpha}$		-10.9	1.3		11.1	-1.9

[1] NMP = unspecified nucleoside 5'-monophosphate (nucleotide). [2] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^{0}_{\beta\alpha}$ in vacuum at the DFT level. [3] Differences in $\Delta E_{\beta\alpha}$, $\Delta E_{\beta\alpha(\text{ZPE})}$, and in $\Delta G^{0}_{\beta\alpha}$ in solvent.

Assuming RNA preceded DNA chronologically, a glance at **Table 2.3** (nucleotides formed *via* pathway (a+b)) suggests that β-anomers are favored across the board in vacuum/non-polar medium. In this case, the relative Gibbs energies, ordered in decreasing magnitudes, are

$\Delta G^{\circ}_{\beta\alpha}$(UMP) ≈ -13 kcal/mol, $\Delta G^{\circ}_{\beta\alpha}$(AMP) ≈ -10 kcal/mol, $\Delta G^{\circ}_{\beta\alpha}$(CMP) ≈ -9 kcal/mol, $\Delta G^{\circ}_{\beta\alpha}$(GMP) ≈ -7 kcal/mol, $\Delta G^{\circ}_{\beta\alpha}$(TMP) ≈ -3 kcal/mol. Interestingly, the least remarkable difference is that of the rarely seen nucleotide, that is, TMP (favouring the β-form by only 3 kcal/mol) as opposed to the one that actually occur in today's RNA, that is, UMP which exhibits, in fact, the highest differential stability favoring the β-anomer (by 13 kcal/mol). Coincidence? There is no way to tell for certain, but suggestive it is.

Continuum solvation reduces the clear advantage of the β-anomer over their α-counterparts to the level of computational noise, yet with consistency (except for a small reversal ≈ 2 kcal/mol for AMP) (**Table 2.3**). As emphasized above, those results including solvation can only be taken as a qualitative guide given the lack of localized solvent-solute interaction(s) that may stabilize or destabilize the system.

As for DNA, the vacuum calculations suggest a considerable advantage for the canonical forms of purines and the reverse for pyrimidines. The advantage of the *β*-form is particularly marked in the case of dGMP by ≈17 kcal/mol of more stability (lower energy) compared to the *α*-form. In aqueous solution, the results are inconclusive being, probably, within computational uncertainties (as discussed above).

Moving to the path (c + d) (**Figure 2.2**), **Table 2.4** suggests that, in vacuum (or in non-polar medium), the two forms of the RNA nucleotides have indistinguishable stabilities within the probable error bars of the method except for TMP and UMP. In these latter two cases, the *α* form is considerably more stable with ≈ 8 and 11 kcal/mol, respectively. These observations are inconsistent with today's state of affairs on three grounds: (*i*) AMP and GMP are predicted to have energies marginally favoring their *α*-forms, and, more importantly (*ii*) TMP and UMP are much more stable in their *α*-form especially UMP. Since these contradict what is being observed – that pathway is less likely to have been Nature's choice leaving the other pathway (a + b) as a more probable scenario.

The addition of a phosphate group on the sugar first *then* the base last, in the (c+d) pathway, creates ample opportunity for the highly anionic oxygens of the phosphate to hydrogen bond with the sugar's hydroxyl groups (see figures A.25 - A.34). In the classical pathway, (a+b), when the base is added first on the sugar, the acidic hydrogens of the base in the β form – being in closer proximity to the phosphate group (both are on the same face of the sugar mean plane) can form

hydrogen bonds with the latter (see figures A.15 - A.24). This hydrogen bonding locks the phosphate group on that side of the mean plane of the sugar and competes with its capacity to form more hydrogen bonds with the sugar hydroxyl groups.

2.3.5 The order of addition of the three components of nucleotides matters

The hypothesis being tested here is captured by the following question. Suppose a series of net reactions were available in prebiotic times that lead to the formation of the first nucleotides, whether those of DNA or of RNA. *Is there a particular order of addition that is more energetically favorable?* In other words, which one of the following net reactions, fleshed-out in **Figure 2.2**, is energetically more favorable, *i.e.* leads to a more negative ΔG:

- *Reaction (a+b):*

$$\text{sugar} + \text{base} \xrightarrow{\text{(a)}} \text{Ns}_{(a)} + H_2O$$

$$\underline{\text{Ns}_{(a)} + H_2PO_4^- \xrightarrow{\text{(b)}} \text{Nt}_{(a+b)} + H_2O}$$

$$\text{sugar} + \text{base} + H_2PO_4^- \xrightarrow{\text{(a+b)}} \text{Nt}_{(a+b)} + 2H_2O, \qquad \Delta G_{(a+b)}$$

- *Reaction (c+d):*

$$\text{sugar} + H_2PO_4^- \xrightarrow{\text{(c)}} \text{5'-SMP}_{(c)} + H_2O$$

$$\underline{\text{5'-SMP}_{(c)} + \text{base} \xrightarrow{\text{(d)}} \text{Nt}_{(c+d)} + H_2O}$$
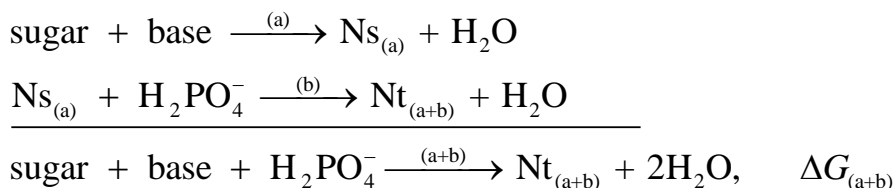
$$\text{sugar} + \text{base} + H_2PO_4^- \xrightarrow{\text{(c+d)}} \text{Nt}_{(c+d)} + 2H_2O, \qquad \Delta G_{(c+d)}$$
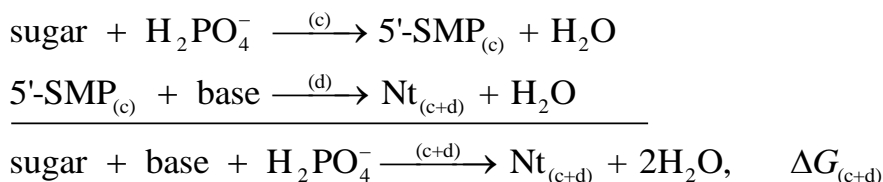
where Ns = nucleoside and Nt = nucleotide in either the α- or the β-anomeric form, the sugar could be either ribose or deoxyribose, and 5'-SMP = 5'-sugar monophosphate. The subscripts in bracket indicate the addition sequence. There is a total of 2 (pathways) × 2 (sugars) × 5 (bases) × 2 (anomers) × 2 (solvation conditions) = 80 "reactions" in total the Gibbs energies of which are summarized in **Table 2.5** (see also **Figures 2.3** and **2.4**). The Gibbs energies of the two reaction pathways are defined by equations (2.2) and (2.3). It is important to remind the reader that since the optimized geometry, and hence the energy, of the final product depends on the sequence, we have *different* "products" (local minima) ($\text{Nt}_{(a+b)} \neq \text{Nt}_{(c+d)}$).

Table 2.5 Gibbs (ΔG°) energies at standard pressure and temperature in kcal/mol for a hypothetical condensation leading to the 5 canonical β-ribonucleosides 5'-monophosphate (NMPs) (nucleotides) and their α-counterparts in vacuum and in aqueous environment. The Gibbs energies of the two reaction pathways are defined by equations 2.2 and 2.3. "Reaction" pathways are labeled according to **Figure 2.2**. (From DFT calculations at the B3LYP/6-31G(d,p) level of theory, with aqueous solvation modeled with the IEFPCM model).

vacuum — *classic path*

NMP[1]	ΔG_a^0 α	β	ΔG_b^0 α	β	ΔG_{a+b}^0 α	β	$\Delta\Delta G_{a+b}^0$ [2]
DNA							
purines							
dAMP	2.5	2.1	-6.2	-14.9	-3.7	-12.8	-9.1
dGMP	4.6	2.6	-8.8	-25.4	-4.3	-22.8	-18.6
pyrimidines							
dCMP	6.5	4.8	-14.8	-11.3	-8.3	-6.4	1.9
dTMP	8.1	6.9	-19.7	-13.7	-11.7	-6.7	4.9
dUMP	7.8	7.0	-20.2	-14.2	-12.3	-7.2	5.2
RNA							
purines							
AMP	2.3	1.0	-6.9	-18.3	-4.6	-17.3	-12.6
GMP	6.2	1.9	-9.8	-15.1	-3.6	-13.1	-9.5
pyrimidines							
CMP	1.9	3.4	-5.4	-18.8	-3.5	-15.3	-11.8
TMP	6.7	5.7	-8.9	-13.2	-2.2	-7.5	-5.3
UMP	6.8	5.8	-9.5	-23.8	-2.7	-18.0	-15.2

solvent — *classic path*

NMP[1]	ΔG_a^0 α	β	ΔG_b^0 α	β	ΔG_{a+b}^0 α	β	$\Delta\Delta G_{a+b}^0$ [2]
DNA							
purines							
dAMP	4.5	2.6	0.6	1.6	5.0	4.2	-0.8
dGMP	4.3	3.0	0.6	0.4	4.9	3.4	-1.5
pyrimidines							
dCMP	5.9	4.5	2.6	0.1	8.5	4.6	-3.9
dTMP	6.3	4.9	2.1	0.5	8.3	5.4	-2.9
dUMP	6.3	5.0	2.2	-0.5	8.5	4.6	-4.0
RNA							
purines							
AMP	3.7	3.9	2.4	1.8	6.1	5.7	-0.4
GMP	4.3	4.4	3.4	0.6	7.7	4.9	-2.8
pyrimidines							
CMP	3.5	2.7	3.1	0.8	8.6	3.5	-5.1
TMP	7.7	5.9	2.8	-0.1	10.5	5.8	-4.7
UMP	6.9	5.8	2.6	-0.4	9.5	5.4	-4.1

alternative path (vacuum)

NMP[1]	ΔG_c^0 α	β	ΔG_d^0 α	β	ΔG_{c+d}^0 α	β	$\Delta\Delta G_{c+d}^0$ [2]
DNA							
purines							
dAMP	-3.1	-8.5	-3.7	4.6	-6.8	-3.9	2.9
dGMP	-3.1	-8.5	-3.2	-3.5	-6.3	-12.0	-5.7
pyrimidines							
dCMP	-3.1	-8.5	4.7	-4.2	1.5	-12.7	-14.3
dTMP	-3.1	-8.5	1.4	-5.2	-1.7	-13.7	-12.0
dUMP	-3.1	-8.5	0.7	-6.3	-2.4	-14.8	-12.4
RNA							
purines							
AMP	-5.2	-9.8	0.8	5.1	-4.4	-4.6	-0.2
GMP	-5.2	-9.8	-7.1	-4.4	-12.3	-14.1	-1.8
pyrimidines							
CMP	-5.2	-9.8	2.8	3.0	-2.4	-6.8	-4.4
TMP	-5.2	-9.8	-8.4	1.0	-13.6	-8.8	4.8
UMP	-5.2	-9.8	-9.2	3.7	-14.4	-6.0	8.4

alternative path (solvent)

NMP[1]	ΔG_c^0 α	β	ΔG_d^0 α	β	ΔG_{c+d}^0 α	β	$\Delta\Delta G_{c+d}^0$ [2]
DNA							
purines							
dAMP	2.0	-2.3	5.0	7.6	7.0	5.3	-1.7
dGMP	2.0	-2.3	4.6	6.9	6.6	4.6	-2.0
pyrimidines							
dCMP	2.0	-2.3	6.2	7.9	8.2	5.6	-2.5
dTMP	2.0	-2.3	6.7	9.4	8.7	7.1	-1.6
dUMP	2.0	-2.3	6.4	8.7	8.4	6.4	-1.9
RNA							
purines							
AMP	1.8	-0.9	5.7	5.7	7.5	4.8	-2.8
GMP	1.8	-0.9	6.0	6.6	7.9	5.7	-2.1
pyrimidines							
CMP	1.8	-0.9	5.5	6.0	7.3	5.1	-2.2
TMP	1.8	-0.9	8.4	9.4	10.3	8.5	-1.8
UMP	1.8	-0.9	8.1	7.0	9.9	6.1	-3.8

(1) NMP = unspecified nucleoside 5'-monophosphate (nucleotide). (2) The $\Delta\Delta$ values are the Gibbs energies of reaction along a given two-step pathway for the β anomer minus the same pathway but for the α

anomer. For example, from the pathways labeled in **Figure 2.2**, $\Delta\Delta G^0_{a+b}$ is the difference of $\left(\Delta G^0_{a+b}\right)_\beta - \left(\Delta G^0_{a+b}\right)_\alpha$. In turn, $\left(\Delta G^0_{a+b}\right)_\beta$, for instance, is the sum of the Gibbs energies for the condensation reaction along the pathway (a) then (b) yielding the nucleotide. Expressed symbolically, $\left(\Delta G^0_{a+b}\right)_\beta = \left(\Delta G^0_a\right)_\beta + \left(\Delta G^0_b\right)_\beta$. Hence, in general: $\Delta\Delta G^0_{i+j} = \left(\Delta G^0_{i+j}\right)_\beta - \left(\Delta G^0_{i+j}\right)_\alpha$, where $i = a, c,$ and $j = b, d.$
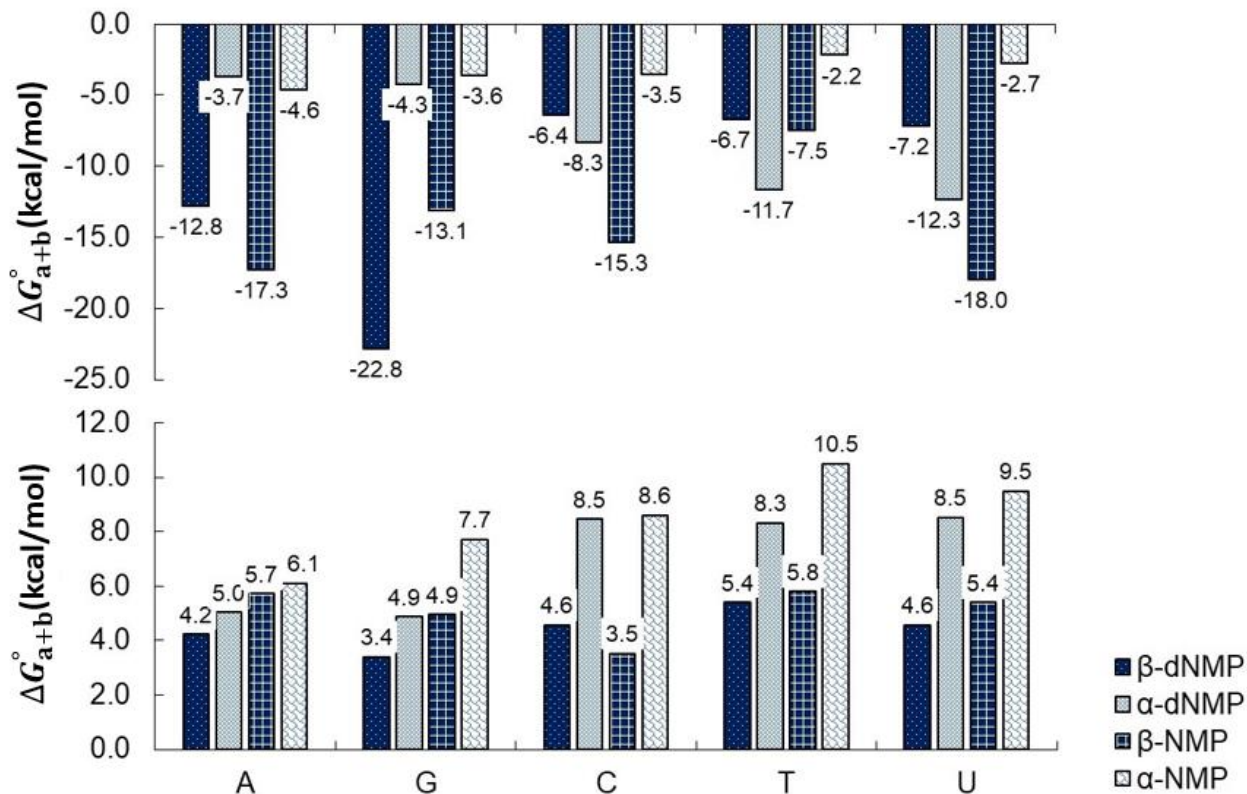
Figure 2.3 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic pathway (pathway (a + b), **Figure 2.2**) leading to the 5 canonical β-nucleotides and their α-counterparts. The Gibbs energies of these reactions are defined by equation 2.2. (Top) ΔG°, B3LYP/6-31G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-31G(d, p) in aqueous medium using the IEFPCM solvation model.
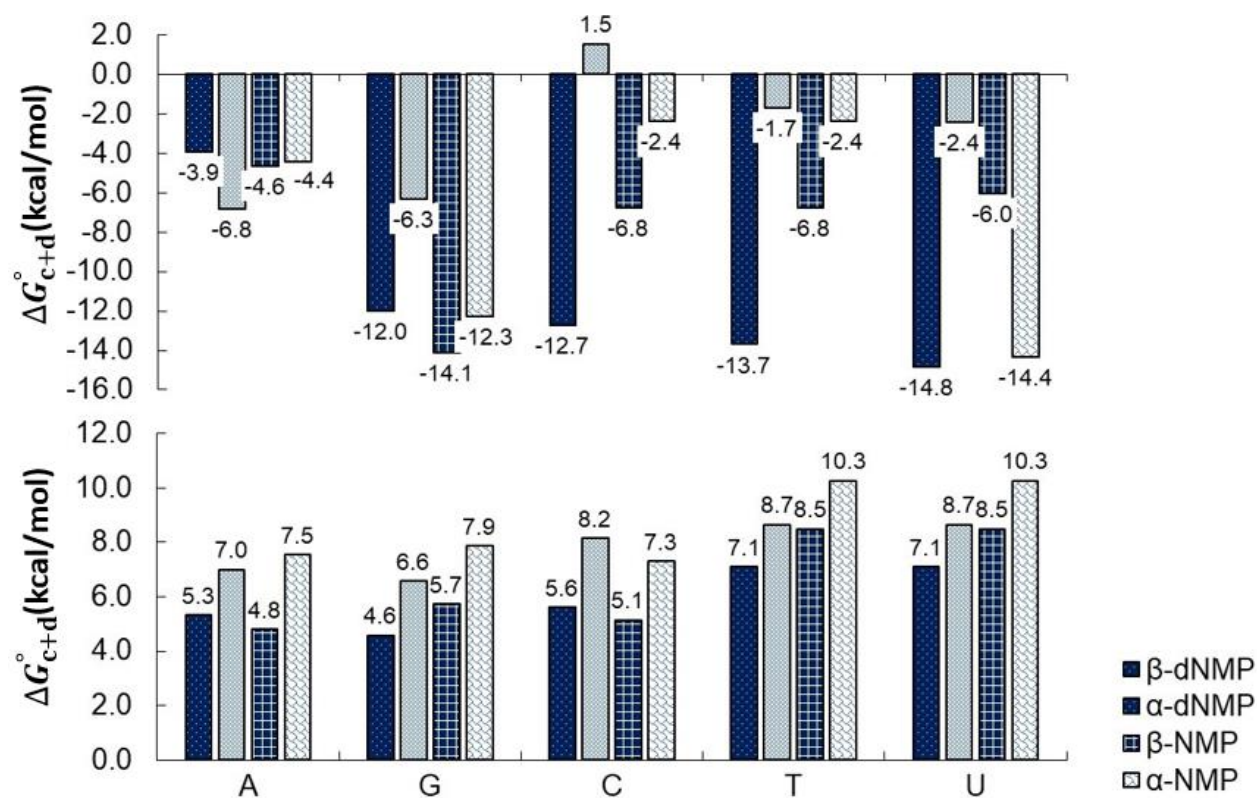
Figure 2.4   Comparison of Gibbs energies of reaction (ΔG°) at 298∘K for the alternative pathway (pathway (c + d), **Figure 2.2**) leading to the 5 canonical β-nucleotides and their α-counterparts. The Gibbs energies of these reactions are defined by equation 2.3. (Top) ΔG°, B3LYP/6-31G (*d*, *p*) in vacuum, (Bottom) ΔG°, B3LYP/6-31G(*d*, *p*) in aqueous medium using the IEFPCM solvation model.

A glance at **Table 2.5** suggests that the pathway (a+b) for the β-anomer is the most favored (more exergonic) in vacuum yet both pathways are endergonic in the continuum solvation model used. Hence the following discussion will focus on pathway (a+b) with the vacuum-phase results examined first. As can be seen from **Table 2.5** and **Figure 2.3**, the condensation reaction between a base and a sugar (reaction (a)) is not favored thermodynamically either in vacuum or aqueous solution. However, the next coupled step in this pathway, step (b), is sufficiently exergonic to drive the entire reaction to competition with negative free energy falling in magnitude within the range 11 kcal/mol $< |\Delta G| <$ 24 kcal/mol in the case of the β-anomers, and to a lesser extent in the case of the α-anomers in which case the magnitudes of the energies of reaction are 6 kcal/mol $< |\Delta G| <$ 20

kcal/mol.

It is further noticed from **Table 2.5** that step (b) reactions are generally more exergonic for the β-anomers of RNA than for their DNA counterparts. This second step, (b), is also exergonic for the α-anomers. This step, for most α-nucleotides, is less exergonic (and significantly so) than the corresponding β-nucleotides except for the deoxyribonucleotides of the pyrimidine bases.

The overall reaction energies strongly favor the (a + b) pathways for all ribonucleotides in vacuum and for both anomers. The overall β-pathways are typically doubly or triply more exergonic than the α- pathways (as can be seen from **Table 2.5** and **Figure 2.3**) except in the case of the pyrimidines-deoxy-ribonucleotides. The differential Gibbs energies between the α- and β-pathways, $\Delta\Delta G$(a + b), that captures the effect of α/β isomerization on the overall reaction Gibbs energies are more exergonic for the β-reactions except for the pyrimidines in DNA.

In aqueous medium, an examination at the overall energies of the (a + b) pathways shows that solvation flips all the vacuum-phase spontaneous reactions to non-spontaneous ones (**Table 5** and **Figure 2.3**). This is in line with the "water problem" [28] which seems to support a non-polar primordial soup. Interestingly, the reactions of the α-anomers are all more endergonic than those with the β-reactions, again suggesting that – in this case – the β-reaction is "less forbidden", so to speak, than the α-counterpart.

In vacuum, for the alternative (c + d) pathway, the first step, namely (c), the condensation of the phosphate and the sugar is exergonic across the board, especially for the β-form (**Table 2.5**, and **Figure 2.4**). Meanwhile when the sugar is substituted at its 5′ position by the phosphate group, the energies of step (d) do not suggest particularly strong trends. The overall reaction energy, though, still indicates that all are spontaneous but to lesser extents than the classical pathway (as mentioned above). Continuum solvation generally flattens the magnitudes of all reactions in the alternative pathways. In this case, step (c) is converted to a non-spontaneous reaction for the α-anomers and marginally exergonic for all the β-anomers. The next step (d) in water is unfavorable in all cases leading to an overall endergonic (c + d) pathway in all cases (as in the classical pathway), with – on average - a marginally less favorable reaction in aqueous medium for both anomers.

It is concluded that the classical pathway for the β-anomers (the anomer which prevails in today's nucleic acids) emerges, again, as the generally favored thermodynamic choice.

2.3.6 Sugar exchange reactions between U and T nucleosides and nucleotides

Lesk posed the question of "Why does DNA contain Thymine and RNA Uracil?" in the title of an important paper that appeared as early as 1969 [38]. Lesk suggested several reasons for the choice including the suggestion of a slight thermodynamic advantage of the dominant forms over the minor forms. This point completes the present study which is addressed in a similar fashion as the manner as above.

From two bases (U, T), two sugars (ribose and 2′ -deoxyribose), and two configurations (α, β) there are 8 possibilities (see **Figure 2.5**). The top panel of this figure represents the (canonical) nucleosides that predominate in the genetic material of all contemporary living organisms, that is to say, thymine on deoxyribose (dT) and uracil on ribose (U). The bottom panel of **Figure 2.5** presents those that occur infrequently in the genetic material, i.e., T and dU.

Table 2.6 compares the calculated energies of the thiamine and uracil nucleosides in hypothetical sugar exchange reactions, defined by "chemical reactions" (reaction 2.4) and (reaction 2.5) below, (where, for clarity, "r" is added to denote ribose sugar):

$$\beta - rT + \beta - dU \; \rightarrow \; \beta - rT + \beta - dU \tag{2.4}$$

and

$$\alpha - rT + \alpha - dU \; \rightarrow \; \alpha - rT + \alpha - dU \tag{2.5}$$
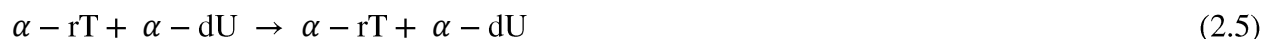
in both the vacuum phase and solvent phase.

These reactions switch the pair of bases from their sugars in their canonical nucleosides to the sugars in their non-canonical ones delivering the energies of sugar double exchange "reactions". The energies listed in **Table 2.6** indicate a consistent lack of any significant energy difference upon affecting this transformation in all types of calculations and energies in the cases of the
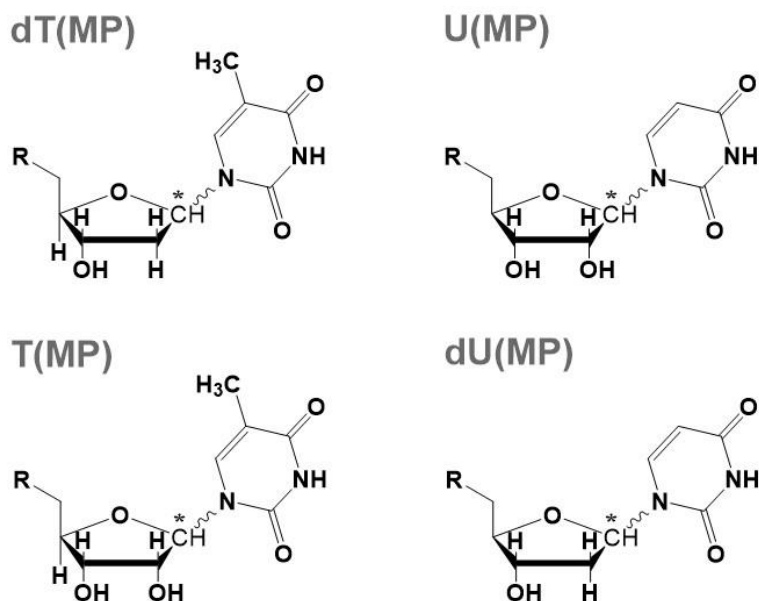
Figure 2.5   Natural and un-natural nucleosides. (Top) the predominant form occurring in nature, that is, 2′-deoxythymidine (dT) and uridine (U) occurring in DNA and RNA respectively. (Bottom) The minor forms, i.e., thymidine (T) and 2′-deoxyuridine (dU) which are not normally incorporated in RNA and DNA, respectively. The sugar exchange (or swapping) is written as "chemical reactions" in equations 2.4 and 2.5. Also see text and **Table 2.6**. (The star (∗) denotes the anomeric center (C1′) of the sugar.)

nucleosides(the first two rows in **Table 2.6**). Thus, for the nucleosides, all differences in the Table (whether of uncorrected total energy differences, ZPE-corrected energy differences, or Gibbs energy differences are below chemical accuracy) suggest essentially equal stability of the "correct" and "wrong" nucleosides. *Thus, there is no clear thermodynamic advantage of attaching one base on one particular sugar, whether in the α- or in the β-configurations, in the vacuum or solution phase*.

Next, the effect of attaching a phosphate group at position 5′ on the relative stabilities of the canonical and non-canonical nucleotides is explored. As mentioned above, we have two distinct step-wise additions of the three components of the nucleotides as illustrated in **Figure 2.2**. The energies of the "sugar exchange reactions" for both pathways can be found in **Table 2.6**. From the

Table 2.6 Differences between the energies of the canonical (predominant) nucleosides and their minor counterparts (**Figure 2.5**) in vacuum and in aqueous environment (energies of the canonical form minus that of the minor form). Included differences are between: The total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies ΔG°. All energies are in kcal/mol and are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The sugar exchange (or swapping) is written as "chemical reactions" in equations 2.4 and 2.5. The Integral Equation Formalism of the Polarizable Continuum Model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

Compared systems[1]	vac.			solv.		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\beta dT+\beta U-\beta T-\beta dU)$	-0.0	-0.0	-0.0	0.3	0.1	-0.3
$\Delta X(\alpha dT+\alpha U-\alpha T-\alpha dU)$	0.4	0.4	0.4	0.5	0.0	-0.8
$\Delta X_{a+b}(\beta dTMP+\beta UMP-\beta TMP-\beta dUMP)$	-9.3	-9.5	-10.0	0.01	0.5	0.4
$\Delta X_{a+b}(\alpha dTMP+\alpha UMP-\alpha TMP-\alpha dUMP)$	0.1	0.1	0.1	0.3	-0.2	-1.2
$\Delta X_{c+d}(\beta dTMP+\beta UMP-\beta TMP-\beta dUMP)$	0.3	1.7	3.9	-3.9	-3.5	-1.7
$\Delta X_{c+d}(\alpha dTMP+\alpha UMP-\alpha TMP-\alpha dUMP)$	0.0	0.0	-0.1	-0.5	-0.2	-0.1

[1]$\Delta X = \Delta E, \Delta E_{(ZPE)}, \Delta G^0$.

listed entries in this Table, the most remarkable one is that of the exchange of the sugar in the β-forms in vacuum (or non-polar medium), clearly favouring the canonical form by up to 10 kcal/mol along the (a + b) pathway (adding the base first then the phosphate). This again reinforces our suggestion that the classical pathway is favoured and the thermodynamic advantage of the β-anomeric form existing in contemporary nucleic acids. This advantage of the 10 kcal/mol of the "correct" over the "wrong" pairing of base with sugar is reduced to noise below chemical accuracy in aqueous solution.

The alternative pathway (c + d) (phosphate first then base) leads to a reversal of the energetic advantage of the β-forms in vacuum in favor of minor form by $\approx$ 4 kcal/mol in terms of ΔG°. In aqueous solution, the (c + d) pathway slightly favors the canonical forms by $\Delta G^{\circ} \approx 2$ kcal/mol. See **Table 2.6**.

From these considerations it may be inferred that only when the phosphate is among the "reactants" there exists an advantage of the canonical forms: (i) by $\approx$ 14 kcal/mol given the order of addition is base first then phosphate in a non-polar medium, or (ii) by $\approx$ 2 kcal/mol in aqueous environment for the reverse order of addition, that is, phosphate first.

2.4 Conclusions

The calculations suggest a slight thermodynamic advantage (-20 kcal/mol $\leq \Delta E_{\beta\alpha} <$ -4 kcal/mol) that favors the selection of the β- over the α-anomers. This is aligned with the concept of an evolutionary "energetic" selection of the fittest. Calculations accounting for implicit solvation in aqueous medium renders either pathways (a + b) or (c + d) thermodynamically unfavorable (**Figures 2.3** and **2.4**). This last observation is consistent with the well-known "water problem".

The present work suggests an order of combination of the three nucleotide components: The condensation of the base with the sugar is first followed by the condensation of the phosphate at the 5′, second. That is, the "classical pathway" emerges as the natural choice for the sequential addition of these components of nucleotides. As mentioned already, the addition of the two first reactants changes the geometries sufficiently to result in different geometries (and energies) when the third reactant is then added as the last condensation.

The final question addressed in this work is whether Nature's choice of incorporating U in RNA and T in DNA is consistent with a thermodynamic explanation. The results suggest an affirmative answer. Indeed, a comparison of the canonical *vs*. the non-canonical nucleotides of these two bases, (i) reinforces the conclusion that the classical pathway is favored, and (ii) indicates an advantage of the canonical pairs compared to the no-canonical pairs when gauged by the "sugar exchange reactions". Furthermore, this thermodynamic advantage exists only for the β-anomers (10 kcal/mol) and vanishes in the case of the α-anomers.

Thermodynamics have been invoked as a driver behind the syntax of the genetic code as seen today [6, 5]. Three decades ago, an editorial in *Nature* had the intriguing title "[*i*]*s Darwinism a thermodynamic necessity*?" [72]. That editorial highlights a paper by Torres in which Darwinian "*fitness*" has been formulated in thermodynamic terms [73]. Torres addresses the logical fallacy of *circulus in probando* (circle in proving, commonly known as circular reasoning) of the concept of *survival of the fittest* [73]. The fallacy is

centered on that fittest is, *by definition*, the ability to be a survivor. These earlier works provide the context of the present one underscoring thermodynamics' role in driving the natural selection of today's canonical nucleotides.

This research addresses the question of why Nature selected the building blocks of nucleic acids as we know them today? The answer is sought in thermodynamic terms, with the underlying hypothesis that free energies are a factor that may have driven particular evolutionary choices. Other factors may have contributed at a particular selection of a molecular form. Such factors may include, for example, kinetics, catalysis including self-catalysis, interaction with light, etc. Our purpose here is much more modest and restricted as the questions addressed suggest.

Other factors that were not considered in this work is the effect of the inclusion of ions as well as of explicit solvation on the energy ordering. For now, the question this paper is addressing is, again, more modest and, that is, in the absence of these additional and relevant factors, *what is the energy ordering of the nucleotides in their isolated forms with and without continuum solvation?*

2.5   References

[1]   L. A. M. Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article#e09657 (pp. 1-12), 2022.

[2]   J. E. Šponer, R. Szabla, R. W. Góra, A. M. Saitta, F. Pietrucci, F. Saija, E. Di Mauro, R. Saladino, M. Ferus, S. Civišh and J. Šponer, "Prebiotic synthesis of nucleic acids and their building blocks at the atomic level - merging models and mechanisms from advanced computations and experiments," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20047-20066, 2016.

[3]   J. E. Šponer, J. Šponer and M. Fuentes-Cabrera, "Prebiotic routes to nucleosides: a quantum chemical Insight into the energetics of the multistep reaction pathways," *Chemistry - A European Journal,* vol. 17, pp. 847-854, 2011.

[4]   J. Serrano Giraldo and I. Zarante, "El origen del ADN: un recorrido por las hipótesis sobre su evolución quimica (The origin of DNA: a journey through the hypothesis on its chemical evolution)," *Anales de Química de la RSEQ,* vol. 117, pp. 274-282, 2021.

[5] H. Grosjean and E. Westhof, "An integrated, structure- and energy-based view of the genetic code," *Nucleic Acids Research,* vol. 44, pp. 8020-8040, 2016.

[6] H. H. Klump, J. Völker and K. J. Breslauer, "Energy mapping of the genetic code and genomic domains: Implications for code evolution and molecular Darwinism," *Quarterly Reviews of Biophysics,* vol. 53, article # e14 (pp. 1-8), 2020.

[7] F. Pellestor and P. Paulasova, "The peptide nucleic acids (PNAs), powerful tools for molecular genetics and cytogenetics," *European Journal of human Genetics,* vol. 12, pp. 694-700, 2004.

[8] S. A. Banack, J. S. Metcalf, L. Jiang, D. Craighead, L. L. Ilag and P. A. Cox, "Cyanobacteria produce N-(2-aminoethyl)glycine, a backbone for peptide nucleic acids which may have been the first molecules for life on earth," *Plos One,* vol. 7, article #e49043 (pp. 1-4), 2012.

[9] F. Crick, "Central dogma of molecular biology," *Nature,* vol. 227, pp. 561-563, 1970.

[10] N. O. Kaplan, M. M. Ciotti, F. E. Stolzenbach and N. R. Bachur, "Isolation of a DPN isomer containing nicotinamide riboside in the α linkage," *Journal of the American Chemical Society,* vol. 77, pp. 815-816, 1955.

[11] S. Suzuki, K. Suzuki, T. Imai, N. Suzuki and S. Okuda, "Isolation of alpha-pyridine nucleotides from *Azotobacter vinelandii*," *Journal of Biological chemistry,* vol. 240, pp. P554-P556, 1965.

[12] J. Paoletti, D. Bazile, F. Morvan, J.-L. Imbach and C. Paoletti, "α-DNA VIII:thermodynamic parameters of complexes formed between the oligo-alphadeoxynucleotides: α-d(GGAAGG) and α-d(CCTTCC) and their complementary oligo-betadeoxynucleotides: β-d(CCTTCC) and β-d(GGAAGG) are different," *Nucleic Acid Research,* vol. 17, pp. 2693-2704, 1989.

[13] M. Froeyen, E. Lescrinier, L. Kerremans, H. Rosemeyer, F. Seela, B. Verbeure, I. Lagoja, J. Rozenski, A. V. Aerschot, R. Busson and P. Herdewijn, *Chemistry: A European Journal,* vol. 7, pp. 5183-5194, 2001.

[14] J.-L. Guesnet, F. Vovelle, N. T. Thuong and G. Lancelot, "2D NMR studies and 3D structure of the parallel-stranded duplex oligonucleotide Acrm5-alpha-d(TCTAAACTC)-beta-d(AGATTTGAG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 29, pp. 4982-4991, 1990.

[15] G. Lancelot, J.-L. Guesnet and F. Vovelle, " Solution structure of the parallel-stranded duplex oligonucleotide *α*-d(TCTAAAC)-*β*-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 28, pp. 7871-7878, 1989.

[16] G. Lancelot, J.-L. Guesnet, V. Roig and N. T. Thuong, "2D-NMR studies of the unnatural duplex *α*-d(TCTAAAC)-*β*-d(AGATTTG)," *Nucleic Acids Research,* vol. 15, pp. 7531-7547, 1987.

[17] S. Kaur, P. Sharma and S. D. Wetmore, "Structural and electronic properties of barbituric acid and melamine-containing ribonucleosides as plausible components of prebiotic RNA: implications for prebiotic self-assembly," *Physical Chemistry Chemical Physics,* vol. 19, pp. 30762-30771, 2017.

[18] S. Kaur, P. Sharma and S. D. Wetmore, "Can cyanuric acid and 2, 4, 6-triaminopyrimidine containg ribonucleosides be components of prebiotic RNA? Insights from QM calculations and MD simulations," *ChemBioChem,* vol. 20, pp. 1415-1415, 2019.

[19] N. Kitadai and S. Maruyama, "Origins of building blocks of life: A review," *Geoscience Frontiers,* pp. 1117-1153, 2018.

[20] J. D. Sutherland and J. N. Whitfield, "Prebiotic chemistry: a bioorganic perspective," *Tetrahedron,* vol. 53, pp. 11493-11527, 1997.

[21] T. R. Cech, "The RNA worlds in context," *Cold Spring Harbor Perspectives in Biology,* vol. 4, article #a006742 (pp. 1-5), 2012.

[22] R. F. Gesteland, T. R. Cech and J. F. Atkins, The RNA world, New York: Spring Harbor Laboratory Press, 1999.

[23] S. Ayukawa, T. Enomoto and D. Kiga, "RNA world," in *Astrobiology: From the Origins of Life to the Search for*, Singapore, Springer Nature, 2019, pp. 77-90.

[24] M. Neveu, H. J. Kim and S. A. Benner, "The ''strong'' RNA world hypothesis: fifty years old," *Astrobiology,* vol. 13, pp. 391-403, 2013.

[25] L. E. Orgel, "Prebiotic chemistry and the origin of the RNA world," *Critical Reviews in Biochemistry and Molecular Biology,* vol. 39, pp. 99-123, 2004.

[26] A. do Nascimento Vieira, K. Kleinermanns, W. F. Martin and M. Preiner, "The ambivalent role of waer at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[27] N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, "The origin of RNA and "my grandfather's axe"," *Chemistry & Biology,* vol. 20, pp. 466-474, 2013.

[28] G. F. Joyce and L. E. Orgel, "Prospects for understanding the origin of the RNA world," in *The RNA world*, New York, Cold Spring Harbor Press, 1999, pp. 49-78.

[29] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei and S. A. Benner, "Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: Borate as a multifaceted problem solver in prebiotic chemistry," *Angewandte Chemie International Edition,* vol. 55, pp. 15816-15820, 2016.

[30] W. D. Fuller, "Studies in prebiotic synthesis VI. Synthesis of purine nucleosides," *Journal of Molecular Biology,* vol. 67, pp. 25-33, 1972.

[31] H. S. Bernhardt and R. K. Sandwick, "Purine biosynthetic Intermediate-containing ribose-phosphate polymers as evolutionary precursors to RNA," *Journal of Molecular Evolution,* vol. 79, pp. 91-104, 2014.

[32] M. A. Crowe and J. D. Sutherland, "Reaction of cytidine nucleotides with cyanoacetylene: Support for the intermediacy of nucleoside-2', 3'-cyclic phosphates in the prebiotic synthesis of RNA," *ChemBioChem,* vol. 7, pp. 951-956, 2006.

[33] M. Gull, B. J. Cafferty, N. V. Hud and M. A. Pasek, "Silicate-promoted phosphorylation of glycerol in non-aqueous solvents: A prebiotically plausible route to organophosphates," *Life,* vol. 7, article #life7030029, 1-10, 2017.

[34] A. A. Ingar, R. W. A. Luke, B. R. Hayter and J. D. Sutherland, " Synthesis of cytidine ribonucleotides by stepwise assembly of the heterocycle on a sugar phosphate," *ChemBioChem,* vol. 4, pp. 1127-1135, 2003.

[35] H.-J. Kim and J. Kim, "A prebiotic synthesis of canonical pyrimidine and purine ribonucleotides," *Astrobiology,* vol. 19, pp. 669-674, 2019.

[36] G. Zubai and T. Mui, "Prebiotic synthesis of nucleotides," *Origins of Life and Evolution of Biospheres,* vol. 31, pp. 87-102, 2001.

[37] A. Poole, D. Penny and B.-M. Sjöberg, "Confounded cytosine! Tinkering and the evolution of DNA," *Nature Reviews Molecular Cell Biology,* vol. 2, pp. 147-151, 2001.

[38] A. M. Lesk, "Why does DNA contain Thymine and RNA Uracil?," *Journal of Theoretical Biology,* vol. 22, pp. 537-540 , 1969.

[39] H. Inc, hyperChem release 7.0 for Windows: reference manual, Waterloo: Hypercube, Inc, 2002.

[40] R. Dennington, T. A. Keith and J. M. Millam, GaussView, version 5, Shawnee Mission: Semichem Inc, 2009.

[41] L. A. Montero, "GRANADA Programme: Distribución aleatoria de moléculas alrededor de un sistema Poliatómico central (Random distribution of molecules around a central polyatomic system)," Faculty of Chemistry, University of Havana, 2019. [Online]. Available: http://karin.fq.uh.cu/mmh/.

[42] L. A. Montero, A. M. Esteva, J. Molina, A. Zapardiel, L. Hernández, H. Márquez and A. Acosta, "A Theoretical Approach to Analytical Properties of 2,4-Diamino-5-phenylthiazole in Water Solution. Tautomerism and Dependence on pH," *Journal of the American Chemical Society,* vol. 120, pp. 12023-12033, 1998.

[43] J. J. P. Stewart, "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters," *Journal of Molecular Modeling,* vol. 19, pp. 1-32, 2013.

[44] J. J. P. Stewart, "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements," *Journal of Molecular Modeling,* vol. 13, pp. 1173-1213, 2007.

[45] J. J. P. Stewart, "MOPAC," Stewart Computational Chemistry, 2023. [Online]. Available: http://openmopac.net/.

[46] A. Klamt, "The COSMO and COSMO-RS solvation models," *Advance Review,* vol. 1, pp. 699-709, 2011.

[47] A. Klamt and G. Schüümann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *Journal of the Chemical Society, Perkin Transactions 2,* vol. 2, pp. 799-805, 1993.

[48] D. S. Sholl and J. A. Steckel, Density Functional Theory. A practical introduction, New Jersey: Wiley, 2023.

[49] W. Koch and M. Holthausen, "The Kohn-Sham Approach," in *A chemist's guide to Density Functional Theory*, 2nd ed., Weinheim, Wiley-VCH and John Wiley & Sons, 2001, pp. 41-59.

[50] R. G. Parr and W. Yang, Density Functional Theory of atoms and molecules, Oxford: Oxford University Press, 1989.

[51] A. Becke, "Density-functional thermochemistry. 3. The role of exact exchange," *The Journal of Chemical Physics,* vol. 98, pp. 5648-5652, 1993.

[52] C. Lee, W. Yang and R. G. Parr, "Development of the Colle–Salvetti correlationenergy formula into a functional of the electron density," *Physical Review B,* vol. 37, p. 785–789, 1988.

[53] W. J. Hehre, L. Radom, J. A. Pople and P. Schleyer, Ab initio molecular orbital theory, New York: John Wiley & sons Ltd, 1986.

[54] J. Tomasi, "Selected features of the polarizable continuum model for the representation of solvation," *WIREs Computational Molecular Science,* vol. 1, pp. 855-867, 2011.

[55] J. Tomasi, C. Cappelli, B. Mennucci and R. Cammi , "From molecular electrostatic potentials to solvations models and ending with biomolecular photophysical processes," in *Quantum Biochemistry: electronic structure and biological activity*, Weinheim, Wiley-VCH, 2010, pp. 131-170.

[56] J. Tomasi, B. Mennucci and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical Reviews,* vol. 105, pp. 2999-3093, 2005.

[57] S. Miertuš and J. Tomasi, "Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes," *Chemical Physics,* vol. 65, pp. 239-245, 1982.

[58] S. Miertuš, E. Scrocco and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chemical physics,* vol. 55, pp. 117-129, 1981.

[59] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji , X. Li, M. Caricato, A. V. Marenich, J.

Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian , J. V. Ortiz, A. F. Izmaylov , J. L. Sonnenberg , D. Williams-Young , F. Ding, F. Lipparini , F. Egidi, J. Goings , B. Peng , A. Petrone , T. Henderson , D. Ranasinghe , V. G. Zakrzewski , J. Gao, N. Rega, G. Zheng , W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida , T. Nakajima , Y. Honda , O. Kitao , H. Nakai , T. Vreven , K. Throssell , J. A. J. Montgomery , J. E. Peralta , F. Ogliaro , M. J. Bearpark , J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari , A. P. Rendell , J. C. Burant, S. S. Iyengar , J. Tomasi , M. Cossi , J. M. Millam , M. Klene , C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, revision C.01, Wallingford CT: Gaussian Inc, 2019.

[60] G. F. Joyce and L. E. Orgel, "Prospects for understanding the origin of the RNA world," in *The RNA World*, 2 ed., New York, Cold Spring Harbor Press, 1999, pp. 49-78.

[61] A. do Nascimento Vieira, K. Kleinermanns, W. F. Martin and M. Preiner, "The ambivalent role of water at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[62] A. V. Marenich, C. J. Cramer and D. G. Truhlar, "Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions," *The Journal of Physical Chemistry B,* vol. 113, pp. 6378-6396, 2009.

[63] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: equilibria, structure, spectra, and dynamics," *Chemical Reviews,* vol. 99, pp. 2161-2200, 1999.

[64] J. B. Foresman and A. Frisch, Exploring chemistry with electronic structure methods, Pittsburgh: Gaussian, Inc, 1996.

[65] C. J. Cramer and D. G. Truhlar, "Continuum solvation models: classical and quantum mechanical implementations," in *Reviews in computational chemistry*, New York, VCH Publishers, 1995, pp. 1-72.

[66] J. Tomasi and M. Persico, "Molecular interactions in solution: An overview of methods based on continuous distributions of solvent," *Chemical Reviews,* vol. 94, pp. 2027-2094, 1994.

[67] Y. Zhao and D. G. Truhlar, "Density functionals with broad applicability in chemistry," *Accounts of Chemical Research,* vol. 41, pp. 157-167, 2008.

[68] A. V. Dass, T. Georgelin , F. Westall , F. Foucher , P. De Los Rios , D. M. Busiello , S. Liang and F. Piazza , "Equilibrium and non-equilibrium furanose selection in the ribose isomerisation network," *Nature Communications,* vol. 12, article # 2749 (pp.1-10), 2021.

[69] L. M. Azofra, M. M. Quesada-Moreno, I. Alkorta, J. R. Avilés-Moreno, J. J. López-González and J. Elguero , "Carbohydrates in the gas phase: conformational preference of D-ribose and 2-deoxy-D-ribose," *New Journal of Chemistry,* vol. 38, pp. 529-538, 2014.

[70] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J.-U. Grabow, J. A. Fernández and F. Castaño, "Ribose found in the gas phase," *Angewandte Chemie International Edition,* vol. 51, pp. 3119-3124, 2011.

[71] A. K. Garcia, J. W. Schopf, S. Yokobori, S. Akanuma and A. Yamagishi, "Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 114, pp. 4619-4624, 2017.

[72] J. Maddox, "Is Darwinism a thermodynamic necessity?," *Nature,* vol. 350, p. 653, 1991.

[73] J.-L. Torres, "Natural selection and thermodynamic optimality," *Il Nuovo Cimento,* vol. 13, pp. 177-185, 1991.

Appendices

Ball-and-stick representations of the optimized geometries (B3LYP/6-31G(d,p)), in vacuum and in solvent (IEFPCM continuum solvation model), along with the Gibbs energies of the elementary transformations considered in this work.
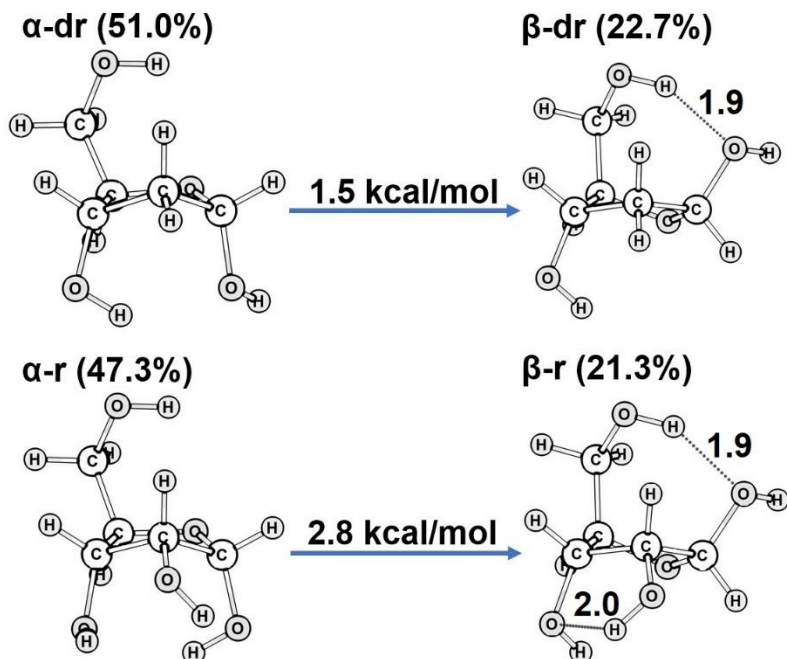
A.1   Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-2'-deoxy (d) and (r)ibose in vacuum. (***Top***) D-2'-deoxyribose (dr). (***Bottom***) D-ribose (r). The energy quoted in kcal/mol is the Gibbs energy of the *β*-form minus the Gibbs energy of the *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(d,p). See text and **Table 2.1**. (X%) represents the relative population in % for each conformer obtained at the PM7 level.
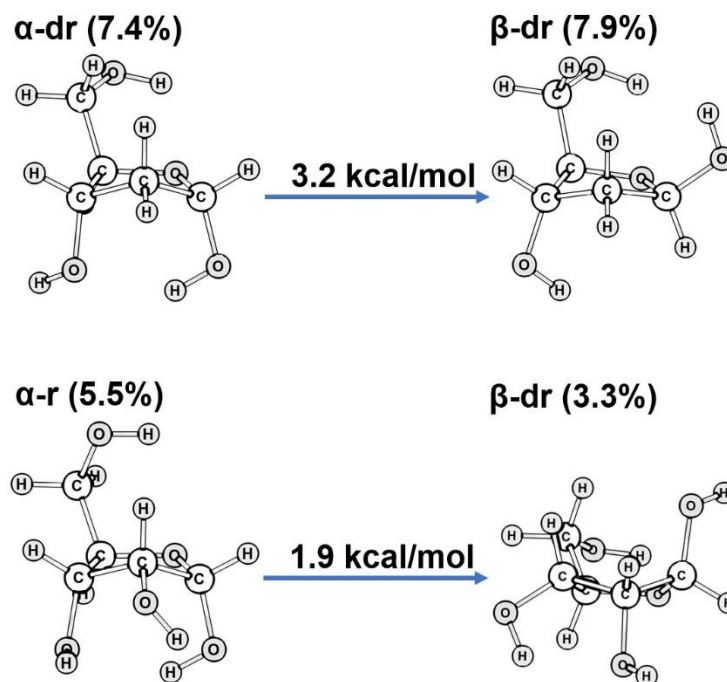
A.2 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-2'-deoxy (d) and (r)ibose obtained using the IEFPCM model for the aqueous solvation. (***Top***) D-2'-deoxyribose (dr). (***Bottom***) D-ribose (r). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.1**. (X%) represents the relative population in % for each conformer obtained at the PM7 level.
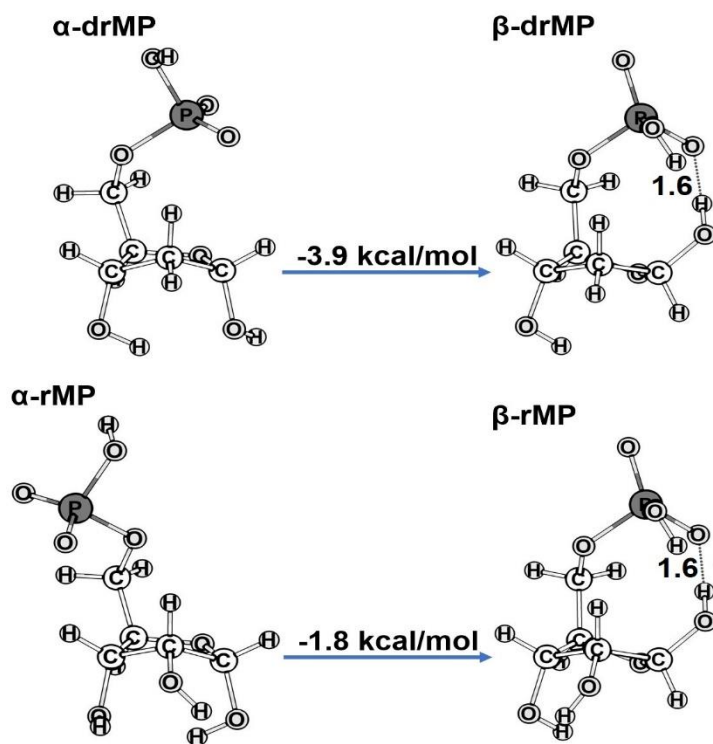
A.3 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α- 5'-monophosphate(MP) sugar in vacuum. (***Top***) D-2'-deoxyribose-5'-monophosphate (drMP). (***Bottom***) D-ribose-5'-monophosphate (rMP). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(d,p). See text and **Table 2.1**.
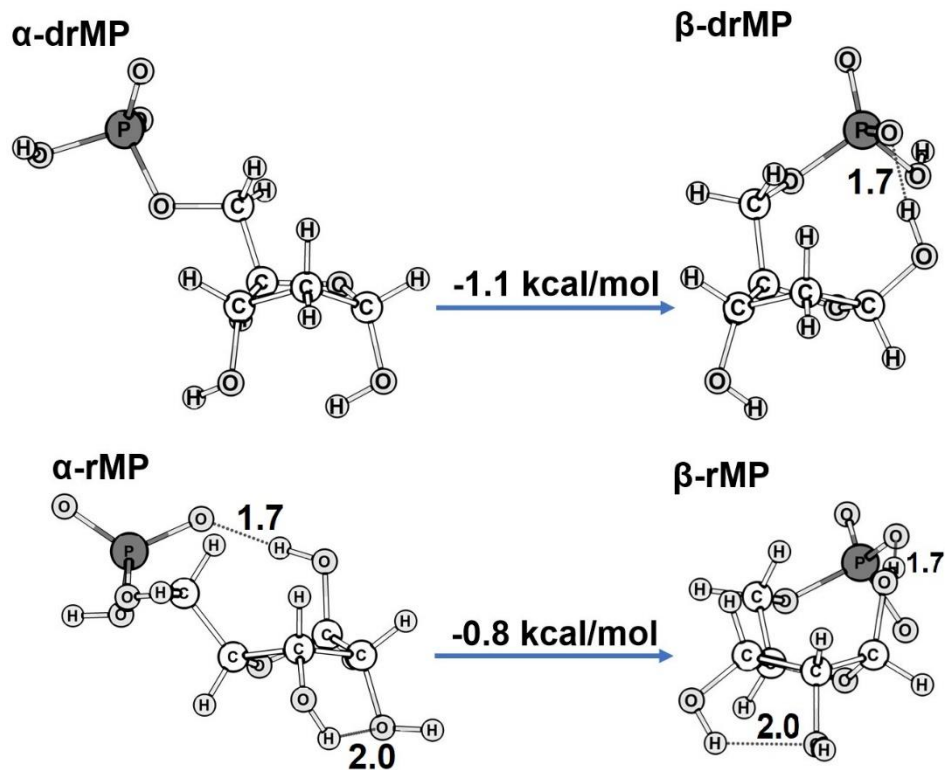
A.4 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α- 5'-monophosphate(MP) sugar obtained using the IEFPCM model for the aqueous solvation. (***Top***) D-2'-deoxyribose-5'-monophosphate (drMP). (***Bottom***) D-ribose-5'-monophosphate (rMP). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.1**.
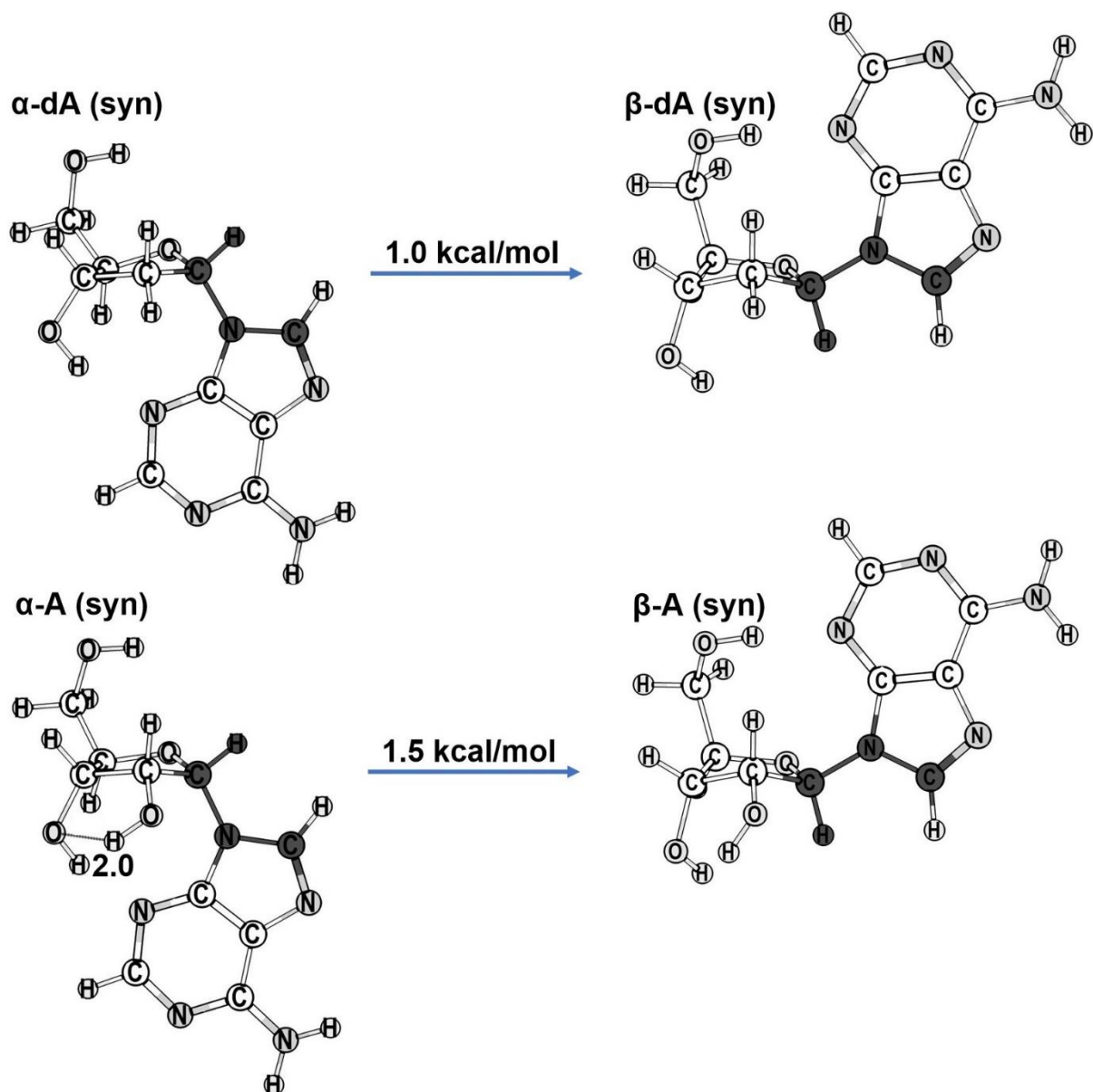
A.5 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of **A**denine (A) in vacuum. (***Top***) 2'-deoxy**A**denosine (dA). (***Bottom***) **A**denosine (A). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
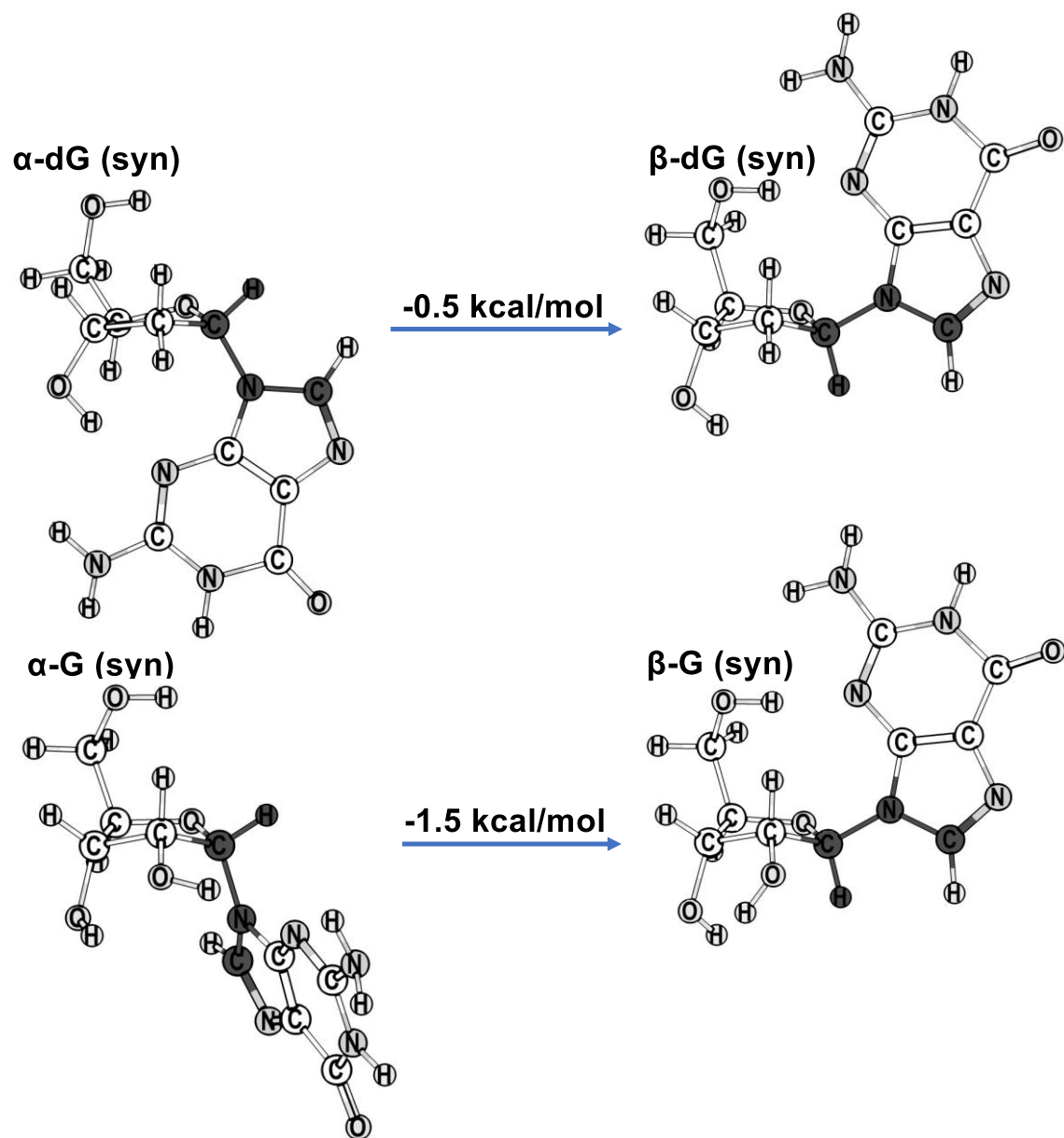
A.6 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of guanine (G) in vacuum. (***Top***) 2'-deoxyguanosine (dG). (***Bottom***) Guanosine (G). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
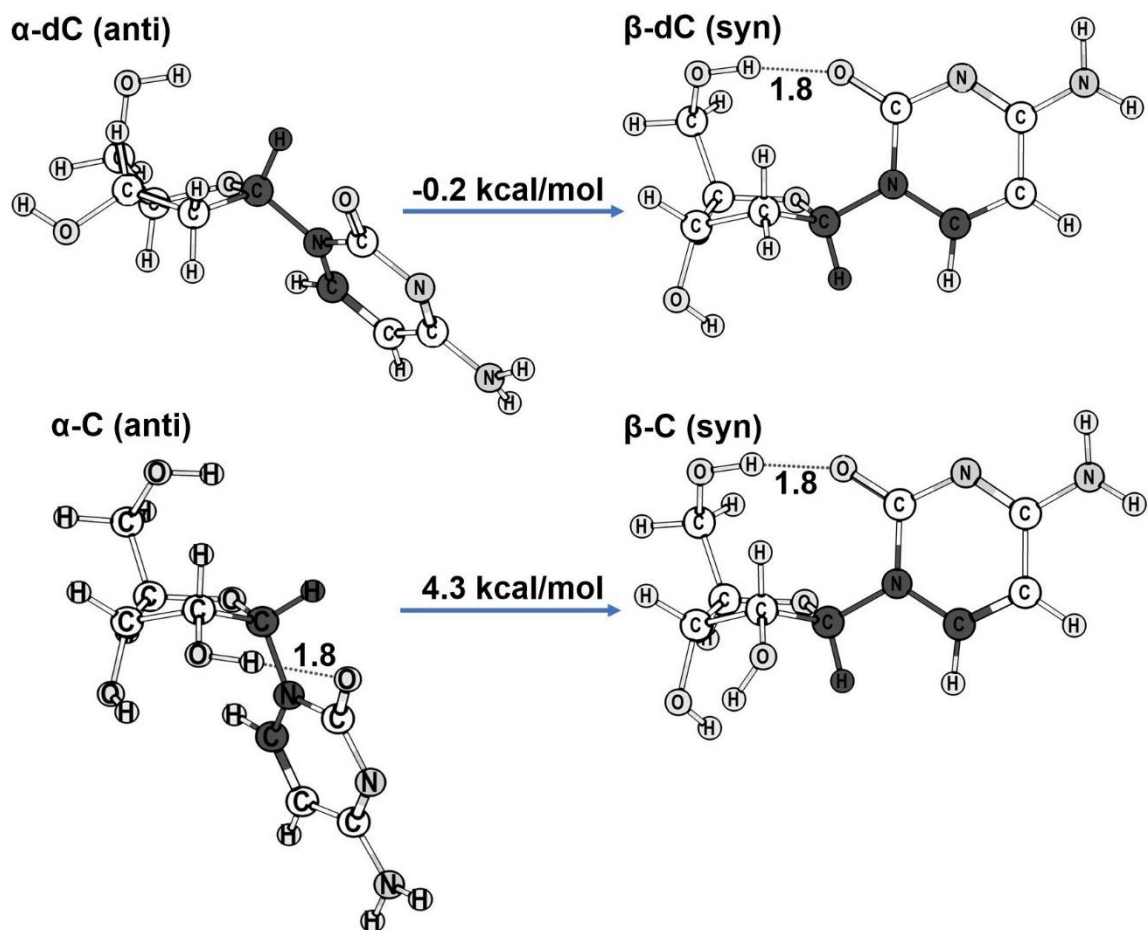
A.7 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of citosine (C) in vacuum. (**Top**) 2'-deoxycytidine (dC). (**Bottom**) Cytidine (C). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
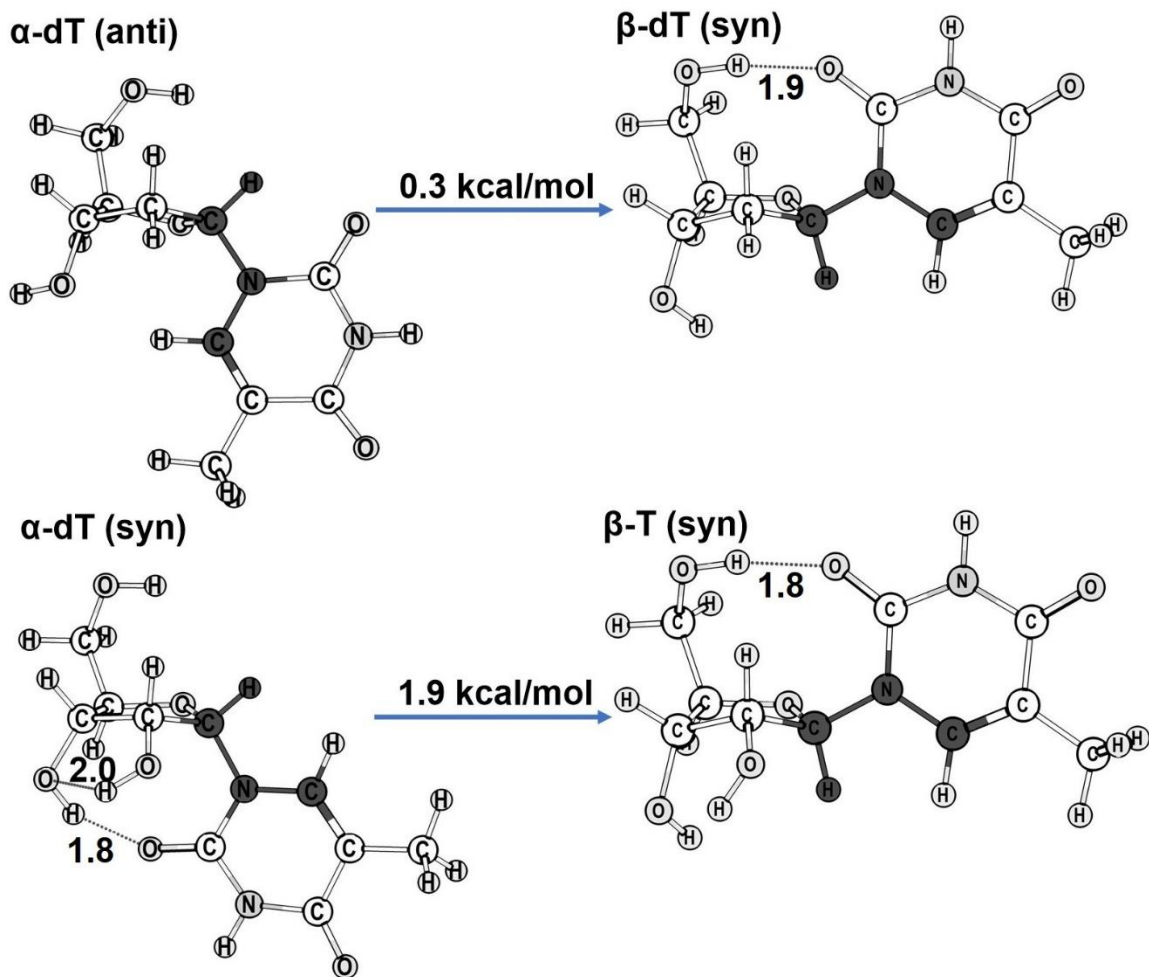
A.8 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleosides of thymine (T) in vacuum. (***Top***) 2'-deoxythymidine (dT). (***Bottom***) Thymidine (T). The energy quoted in kcal/mol is the Gibbs energy of the *β*-form minus the Gibbs energy of the *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
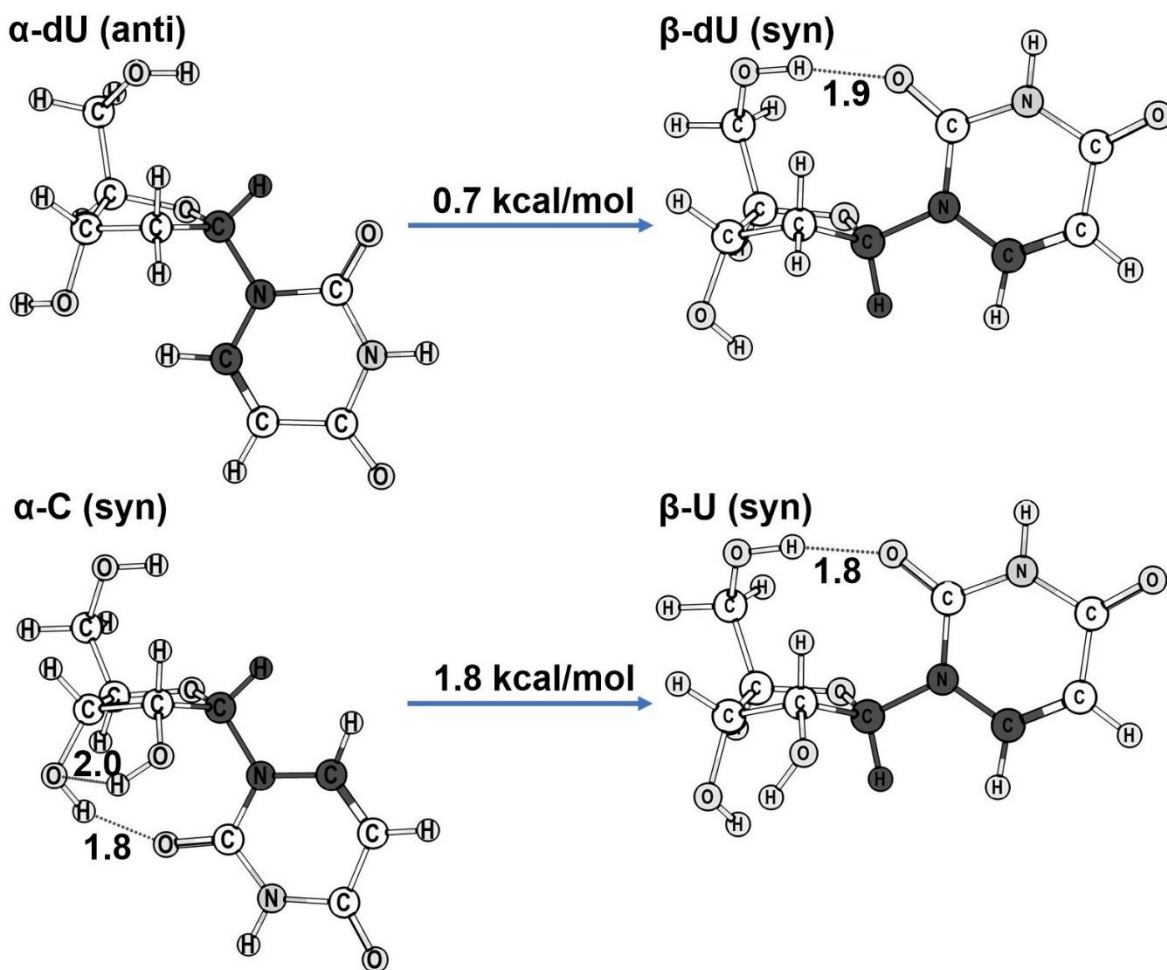
A.9 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of uracil (U) in vacuum. (***Top***) 2'-deoxyuridine (dG). (***Bottom***) Uridine (U). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.

α-dA (anti)   1.4 kcal/mol →   β-dA (syn)

α-A (syn)   2.1 kcal/mol →   β-A (syn)
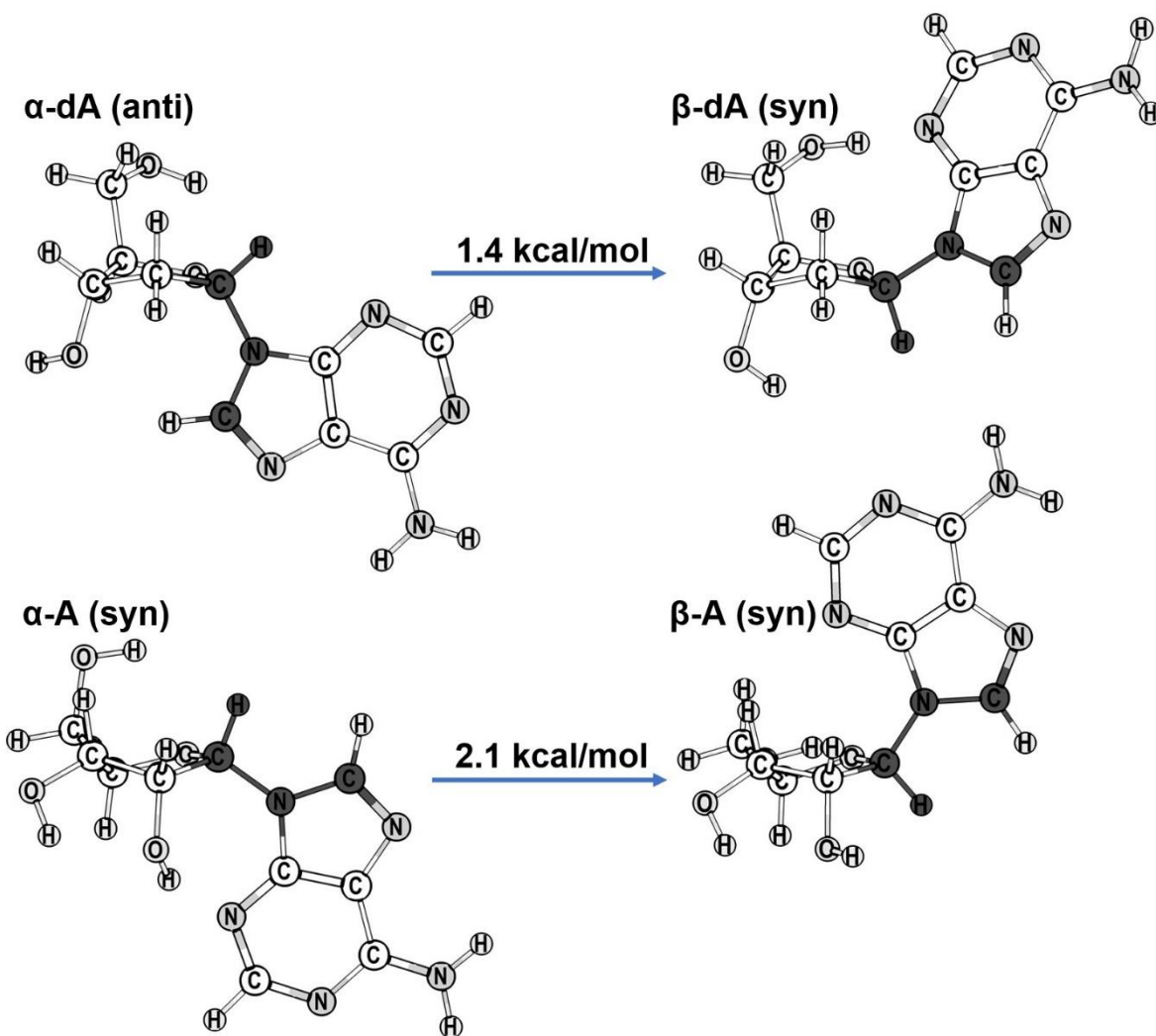
A.10   Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleosides of **A**denine (A) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyadenosine (dA). (***Bottom***) Adenosine (A). The energy quoted in kcal/mol is the Gibbs energy of the *β*-form minus the Gibbs energy of the *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
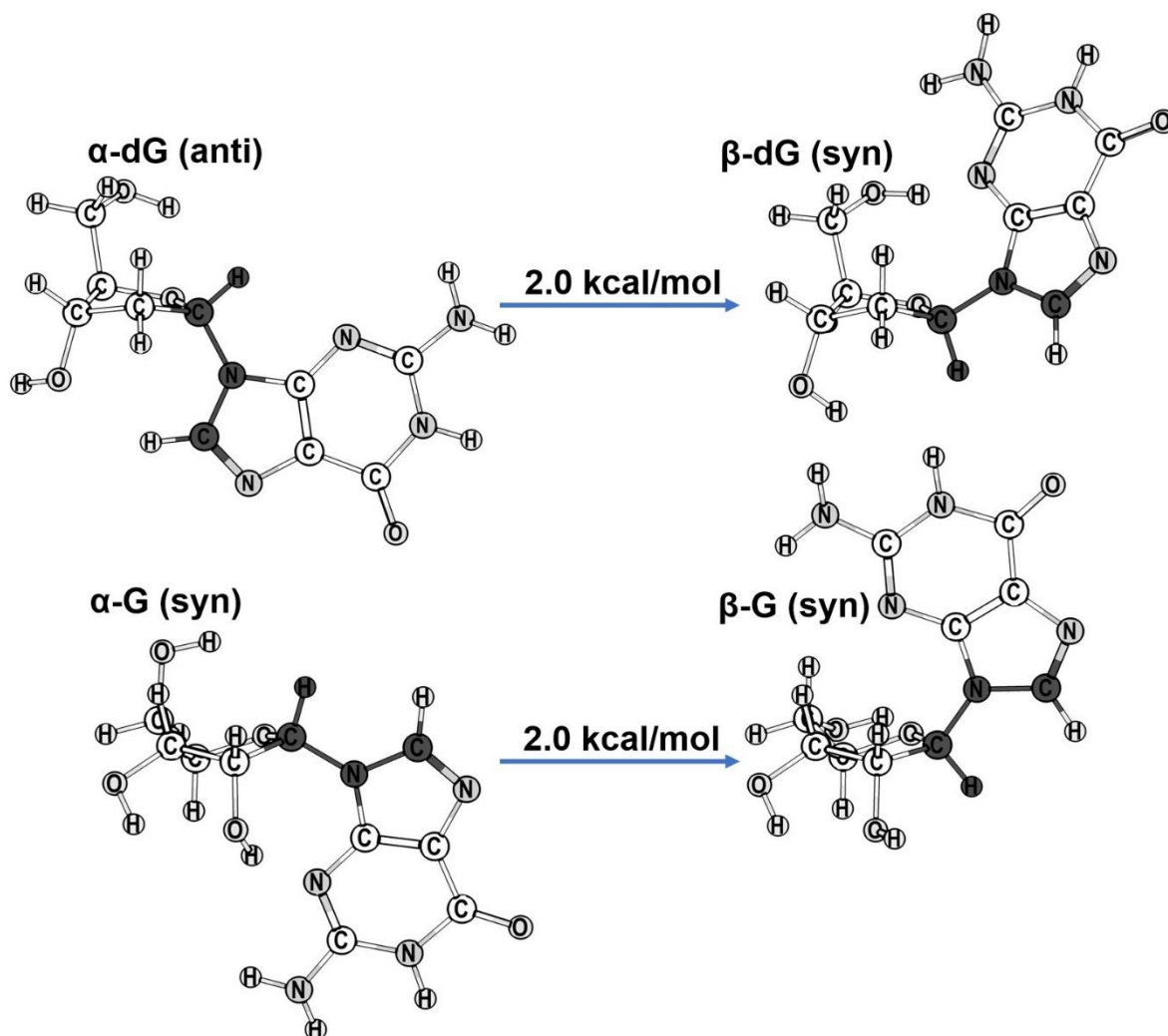
A.11 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleosides of guanine (G) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyguanosine (dG). (***Bottom***) Guanosine (G). The energy quoted in kcal/mol is the Gibbs energy of the *β*-form minus the Gibbs energy of the *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
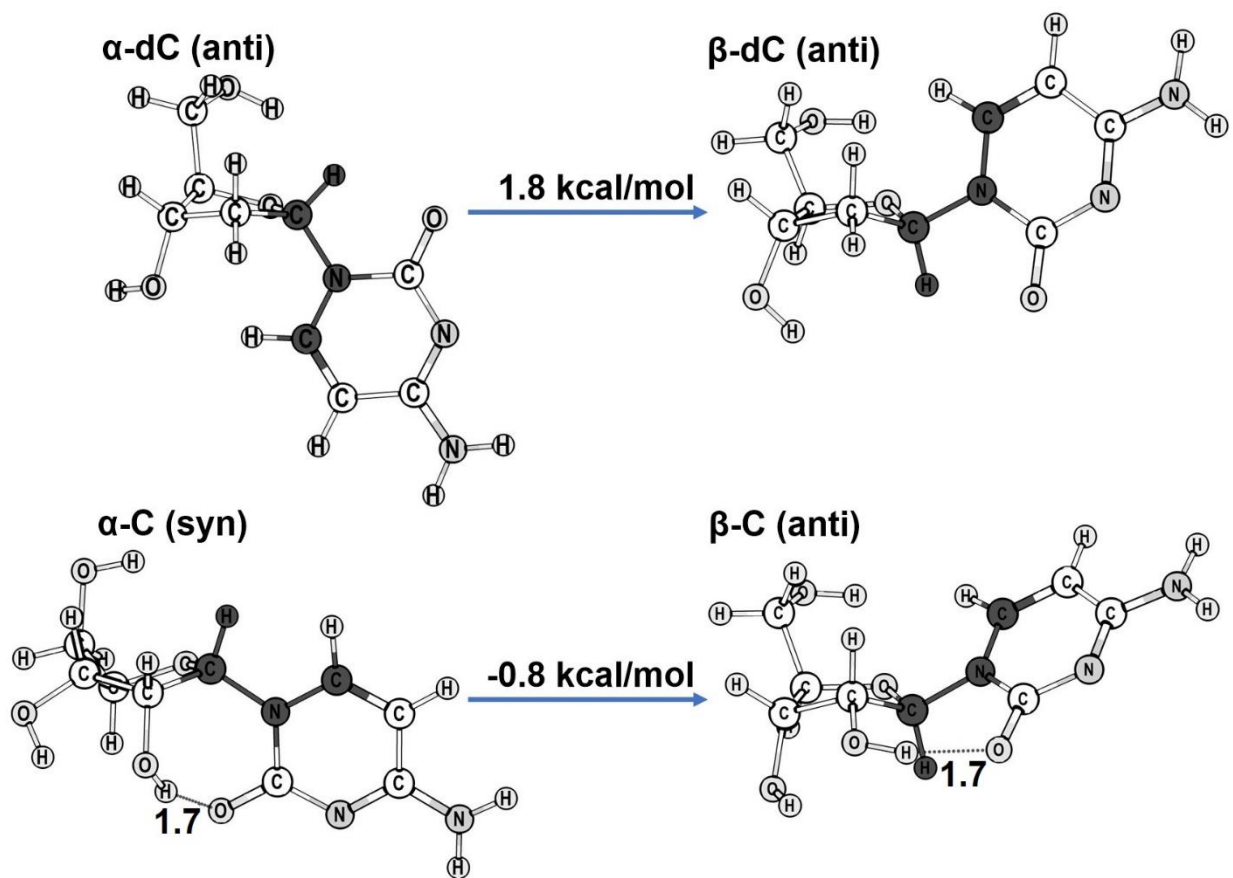
A.12 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of cytosine (C) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxycytidine (dC). (***Bottom***) Cytidine (C). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
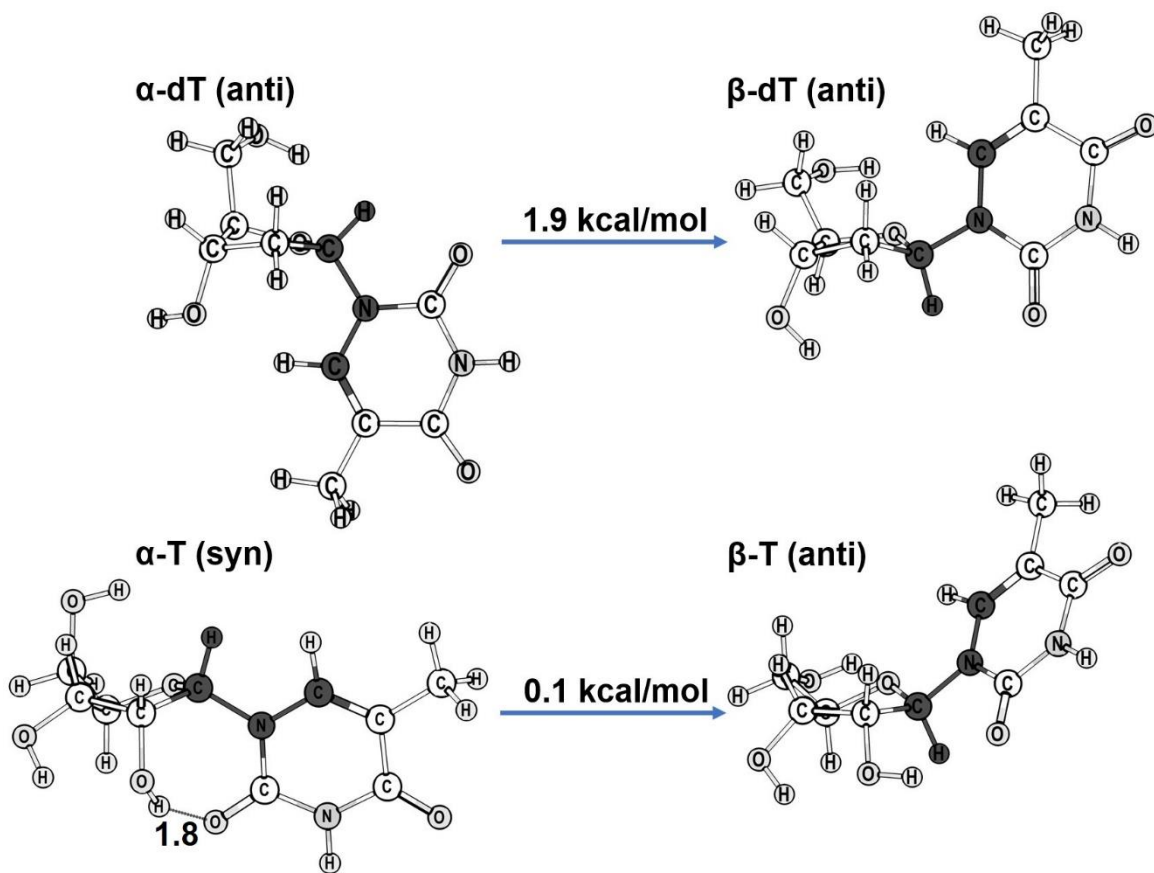
A.13 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleosides of thymine (T) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxythymidine (dT). (***Bottom***) Thymidine (T). The energy quoted in kcal/mol is the Gibbs energy of the *β*-form minus the Gibbs energy of the *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
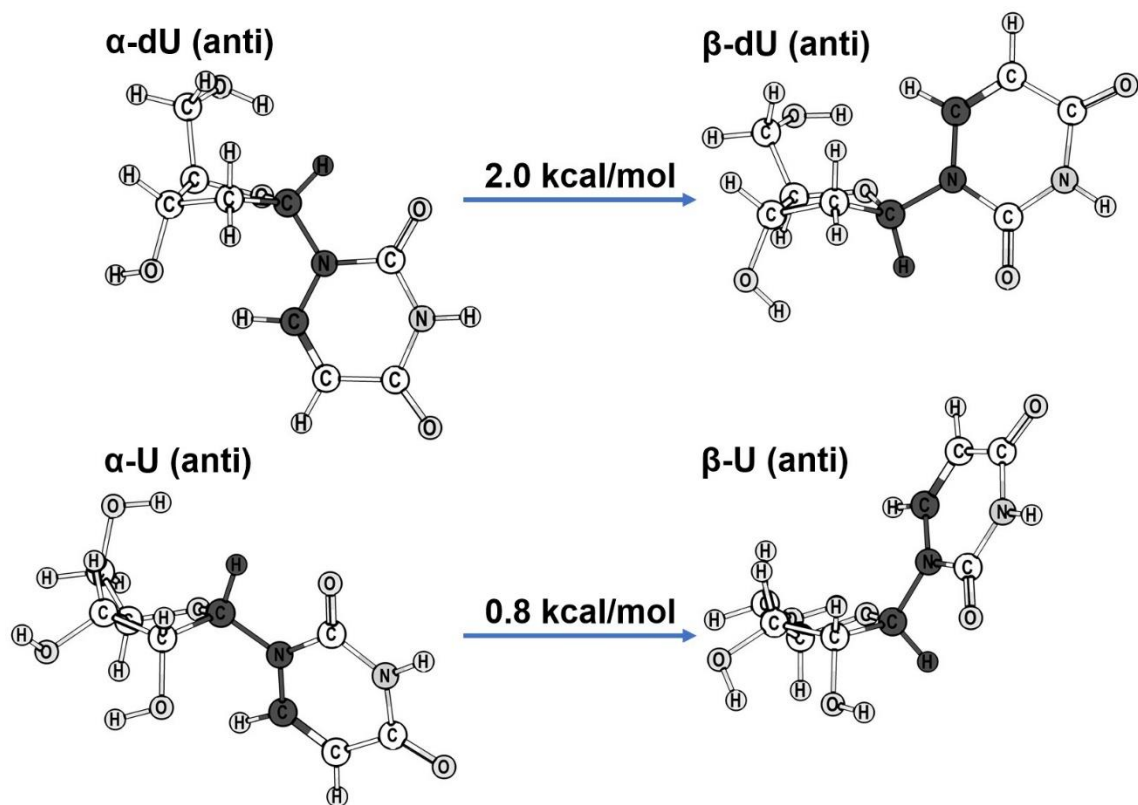
A.14 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleosides of uracil (U) using the IEFPCM model for the water effect. (***Top***) 2'-deoxyuridine (dU). (***Bottom***) Uridine (U). The energy quoted in kcal/mol is the Gibbs energy of the β-form minus the Gibbs energy of the α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.2**. The atoms involved in the torsion angle rotated in the PES are in bold.
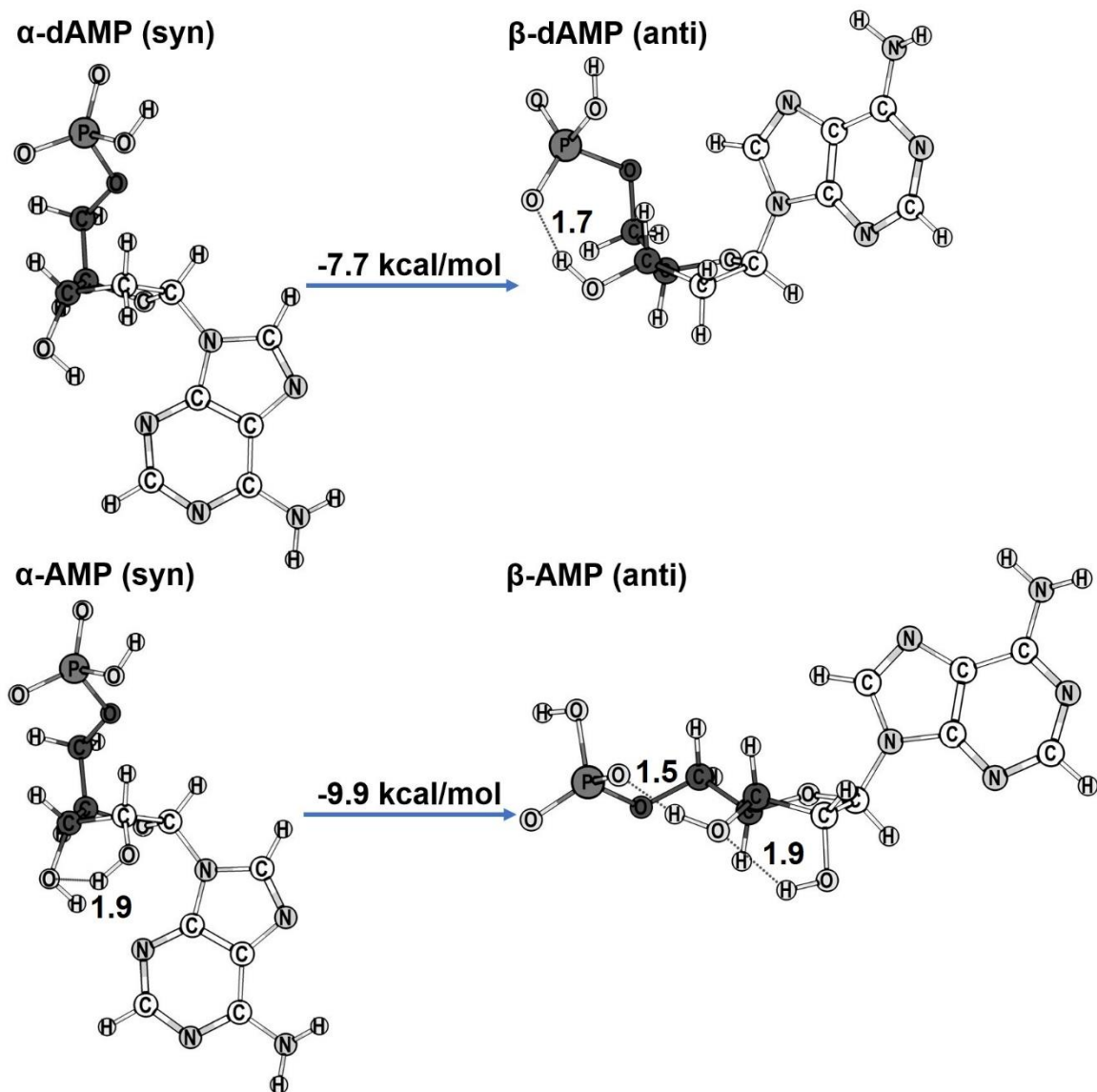
A.15 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of **A**denine (A) in vacuum obtained for the classic pathway (pathway (a+b), **Figure 2.2**). (*Top*) 2'-deoxyadenosine-5'-monophosphate (dAMP). (***Bottom***) Adenosine-5'-monophosphate (AMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
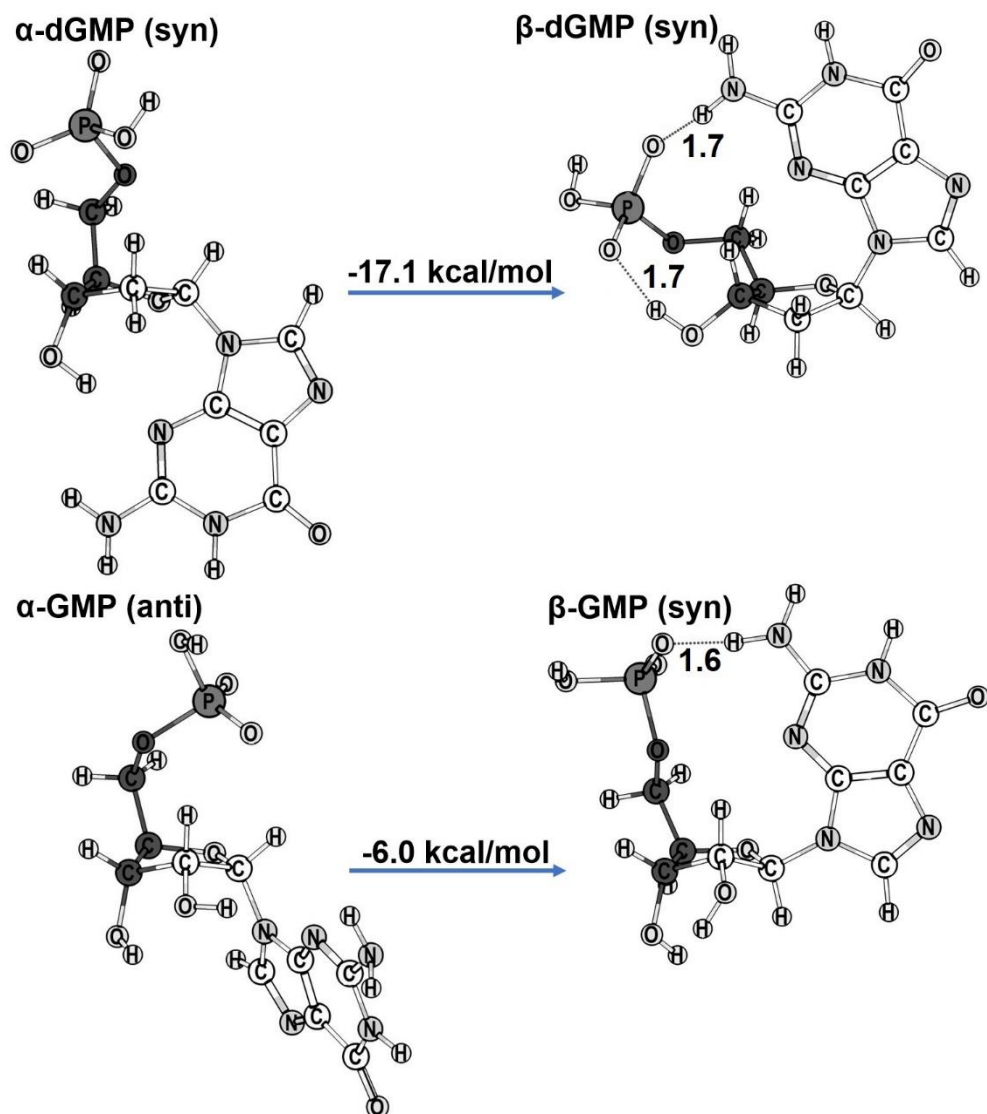
A.16 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied
β- and *α*-nucleotides of guanine (G) in vacuum obtained for the classic pathway (pathway
(a+b), **Figure 2.2**). (***Top)*** 2'-deoxyguanosine-5'-monophosphate (dGMP). (***Bottom)***
Guanosine-5'-monophosphate (GMP). The energy quoted in kcal/mol is the total energy of
the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-
31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the
PES are in bold.
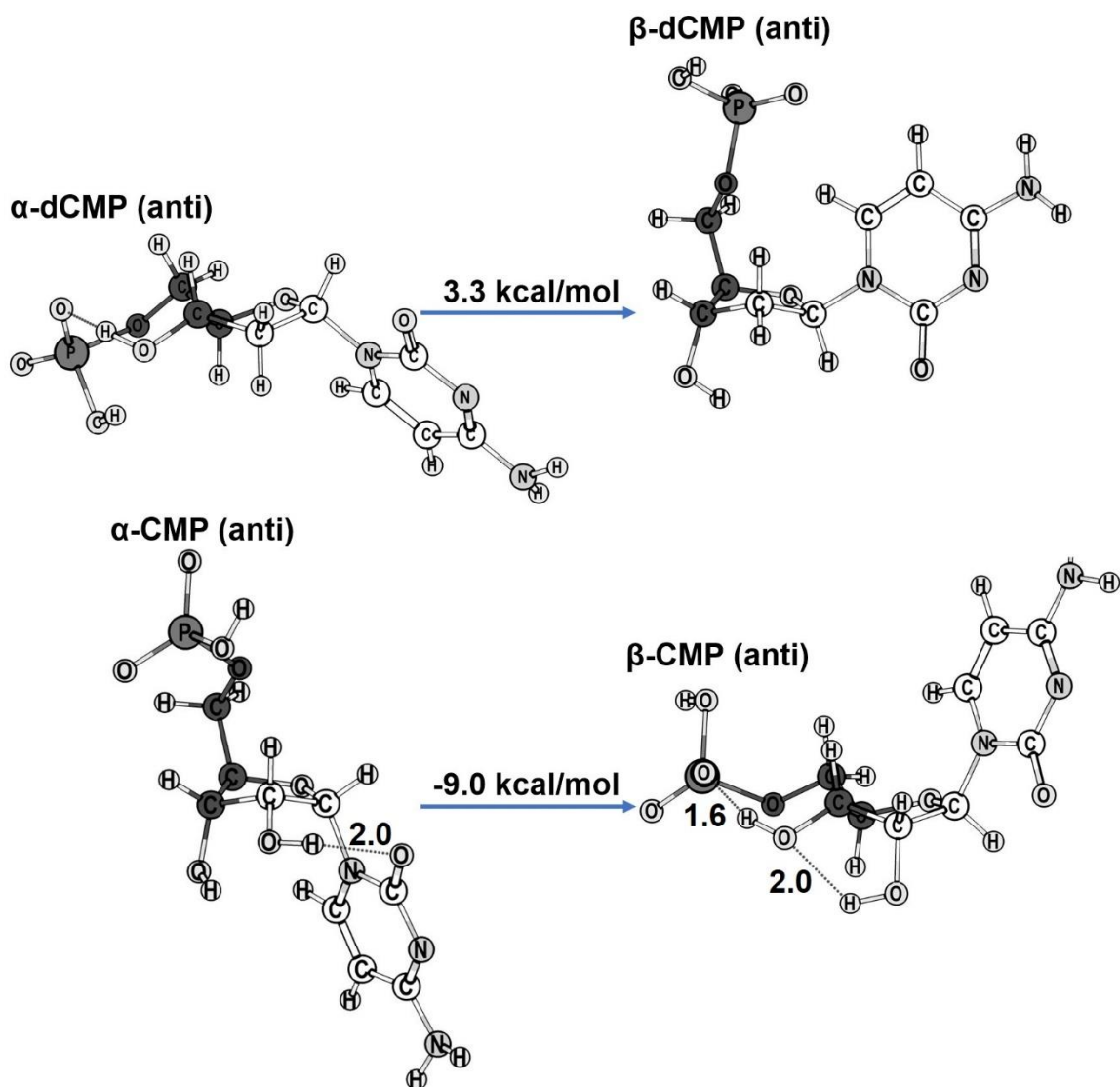
A.17 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of cytosine (C) in vacuum obtained for the classic pathway (pathway (a+b), **Figure 2.2**). (*Top*) 2'-deoxycytidine-5'-monophosphate (dCMP). (*Bottom*) Cytidine-5'-monophosphate (CMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
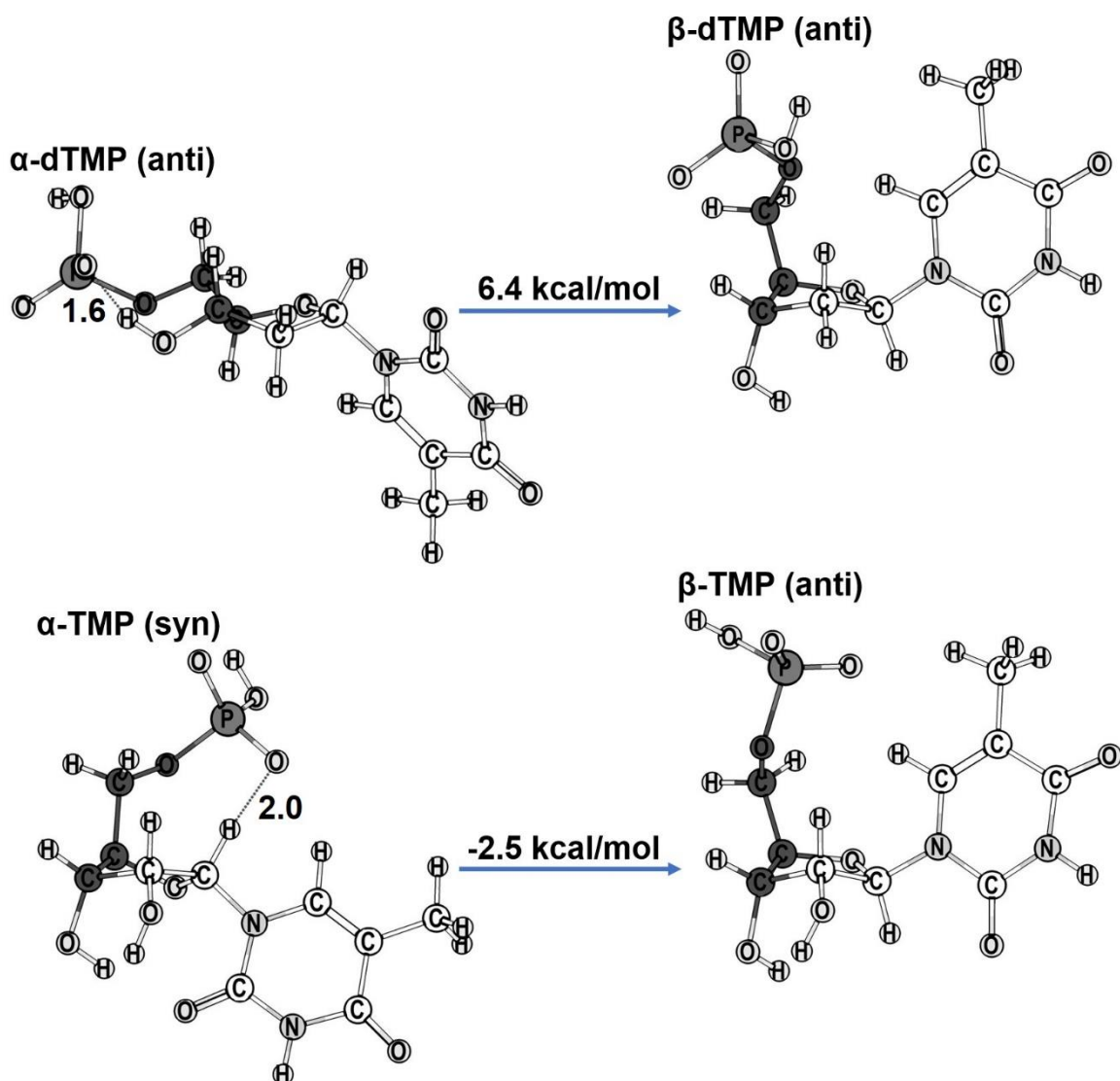
A.18   Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied
β- and α-nucleotides of thymine (T) in vacuum obtained for the classic pathway (pathway
(a+b), **Figure 2.2**). (***Top***) 2'-deoxythymidine-5'-monophosphate (dTMP). (***Bottom***)
Thymidine-5'-monophosphate (TMP). The energy quoted in kcal/mol is the total energy of
the β-form minus the total energy of α-form (equation 2.1) obtained at the DFT-B3LYP/6-
31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the
PES are in bold.
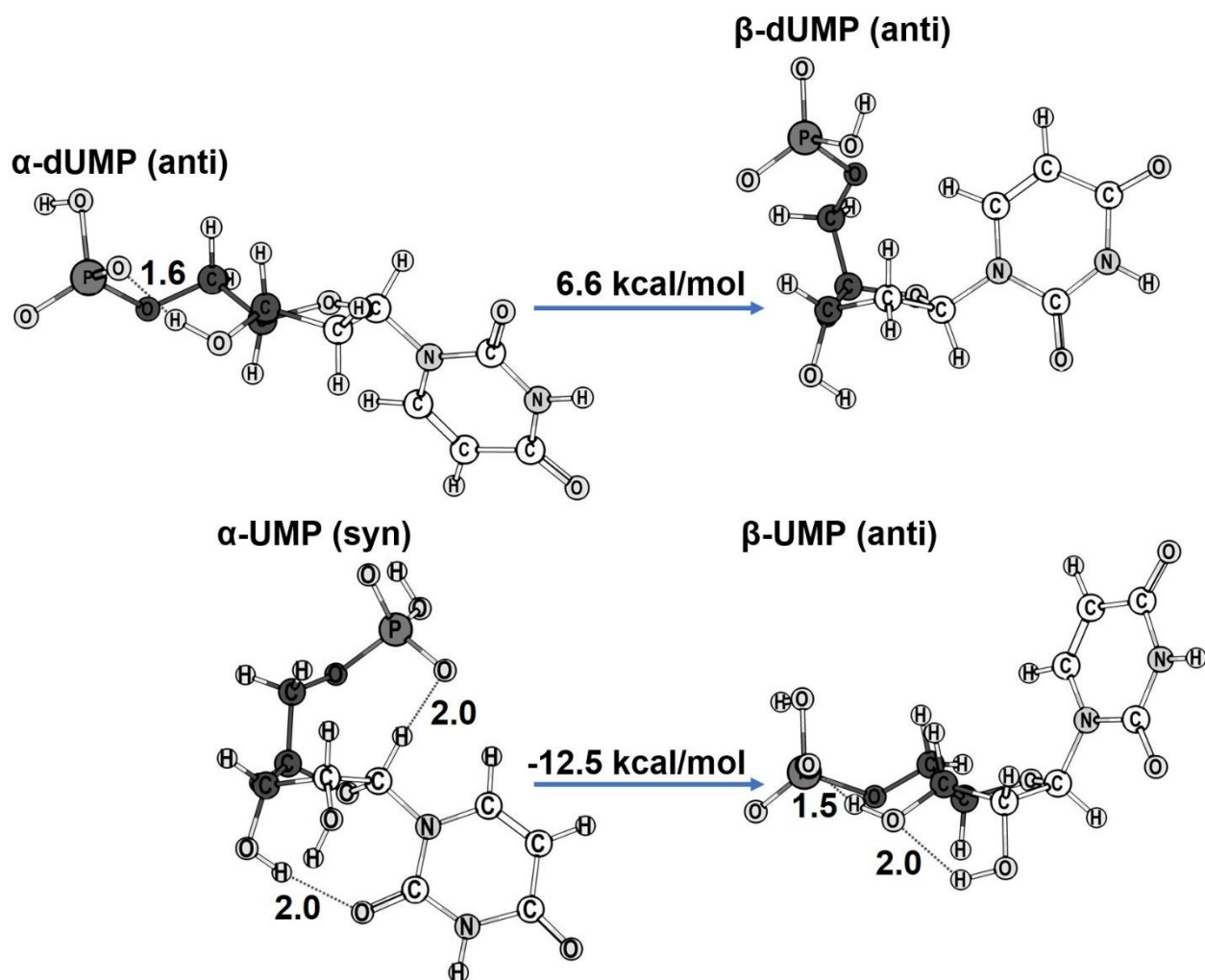
A.19 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of uracil (U) in vacuum obtained for the classic pathway (pathway (a+b), **Figure 2.2**). (***Top***) 2'-deoxyuridine-5'-monophosphate (dUMP). (***Bottom***) Uridine-5'-monophosphate (UMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
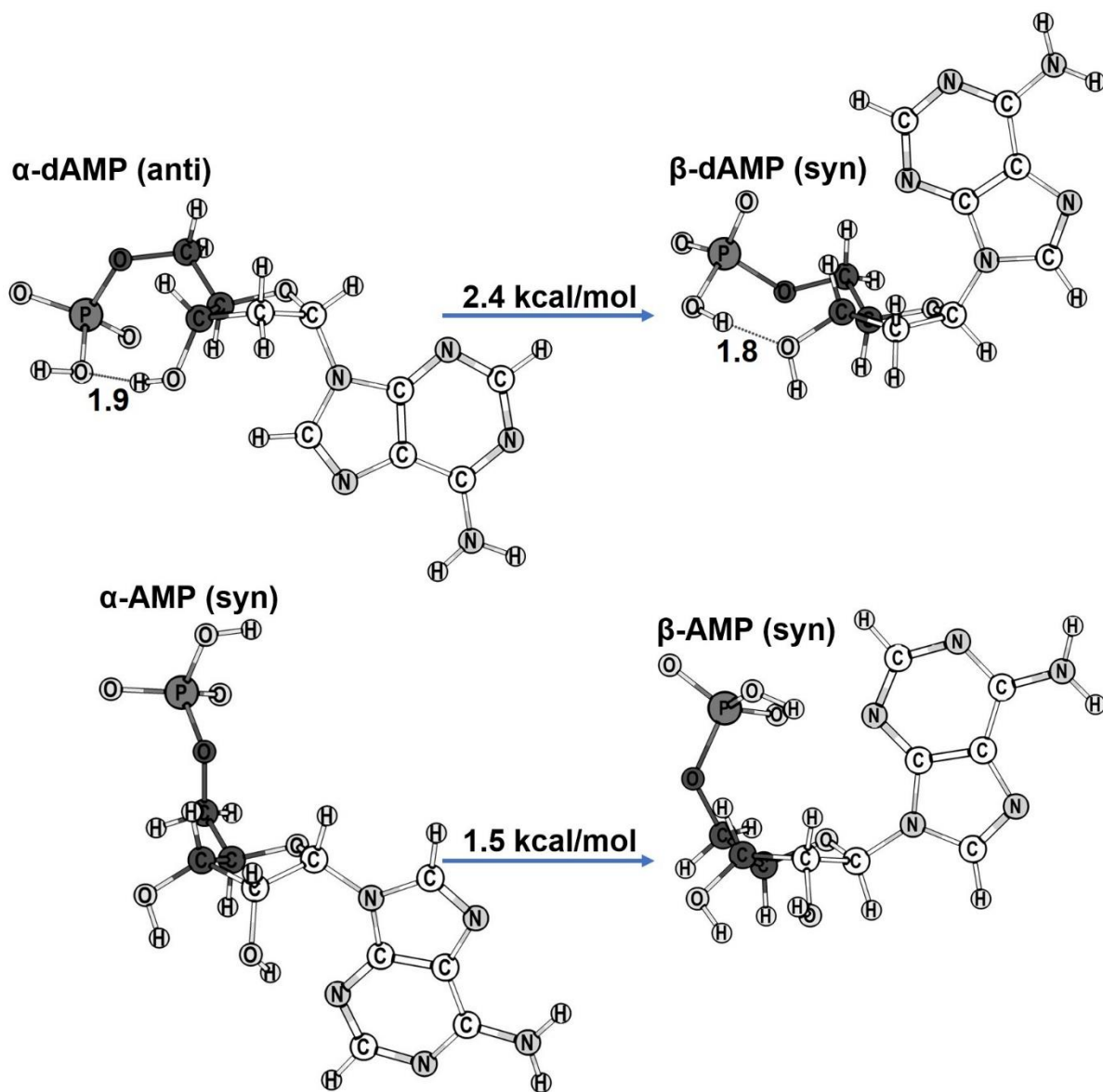
A.20 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of **A**denine (A) for the classic pathway (pathway (a+b), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyadenosine-5'-monophosphate (dAMP). (***Bottom***) Adenosine-5'-monophosphate (AMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
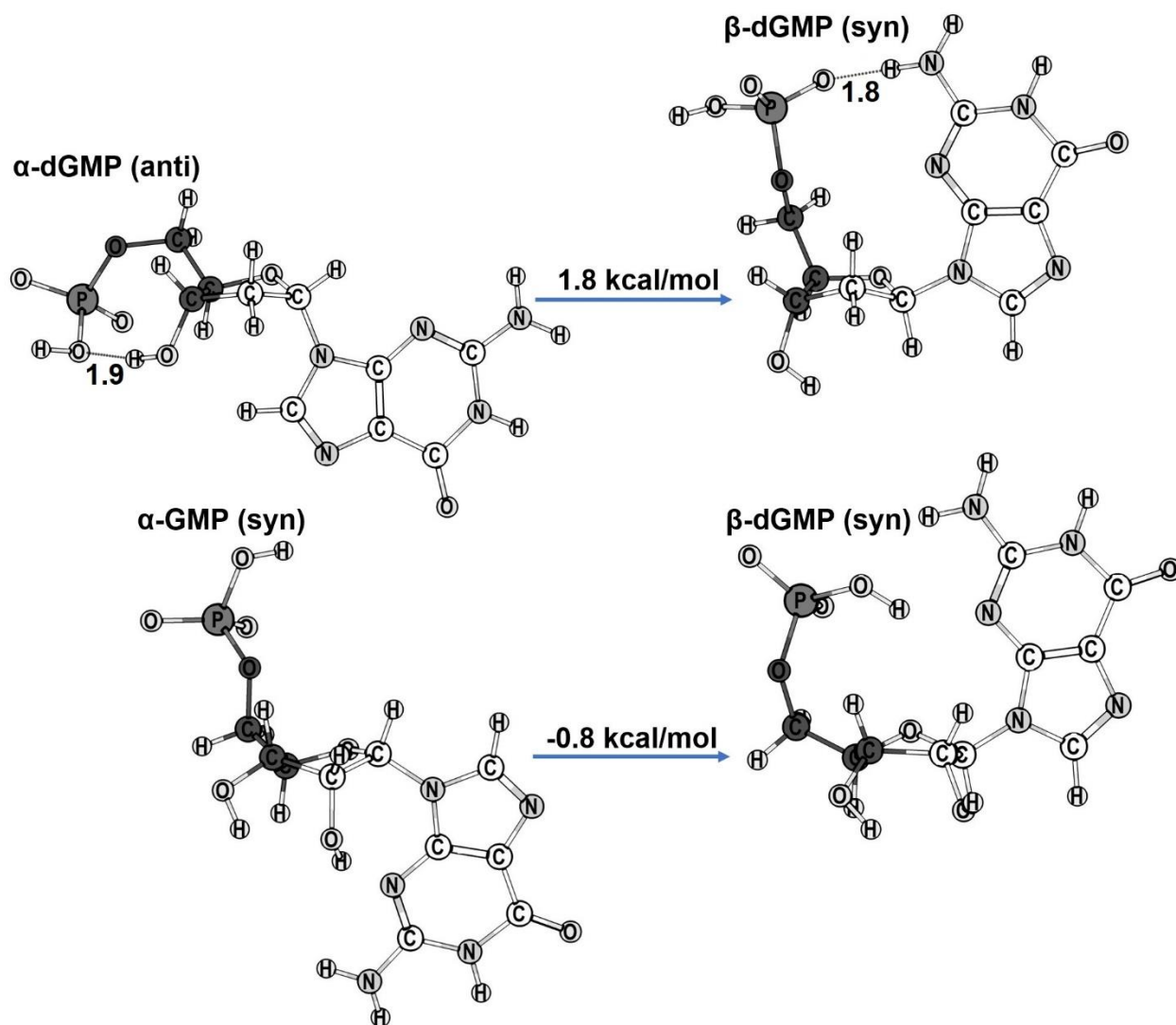
A.21   Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied β- and α-nucleotides of guanine (G) for the classic pathway (pathway (a+b), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyguanosine-5'-monophosphate (dGMP). (***Bottom***) Guanosine-5'-monophosphate (GMP). The energy quoted in kcal/mol is the total energy of the β-form minus the total energy of α-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
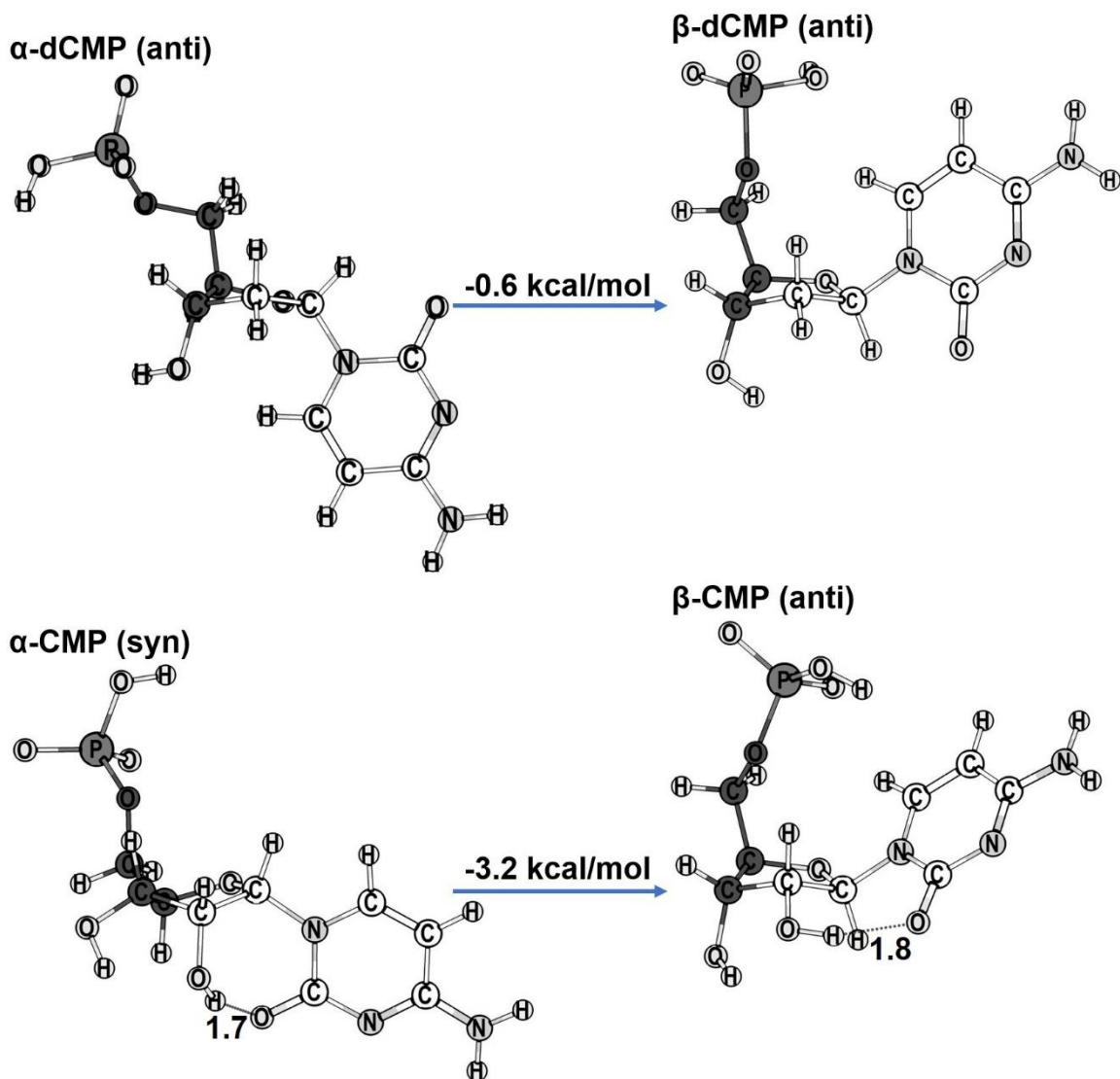
A.22 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of cytosine (C) for the classic pathway (pathway (a+b), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxycytidine-5'-monophosphate (dCMP). (***Bottom***) Cytidine-5'-monophosphate (CMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
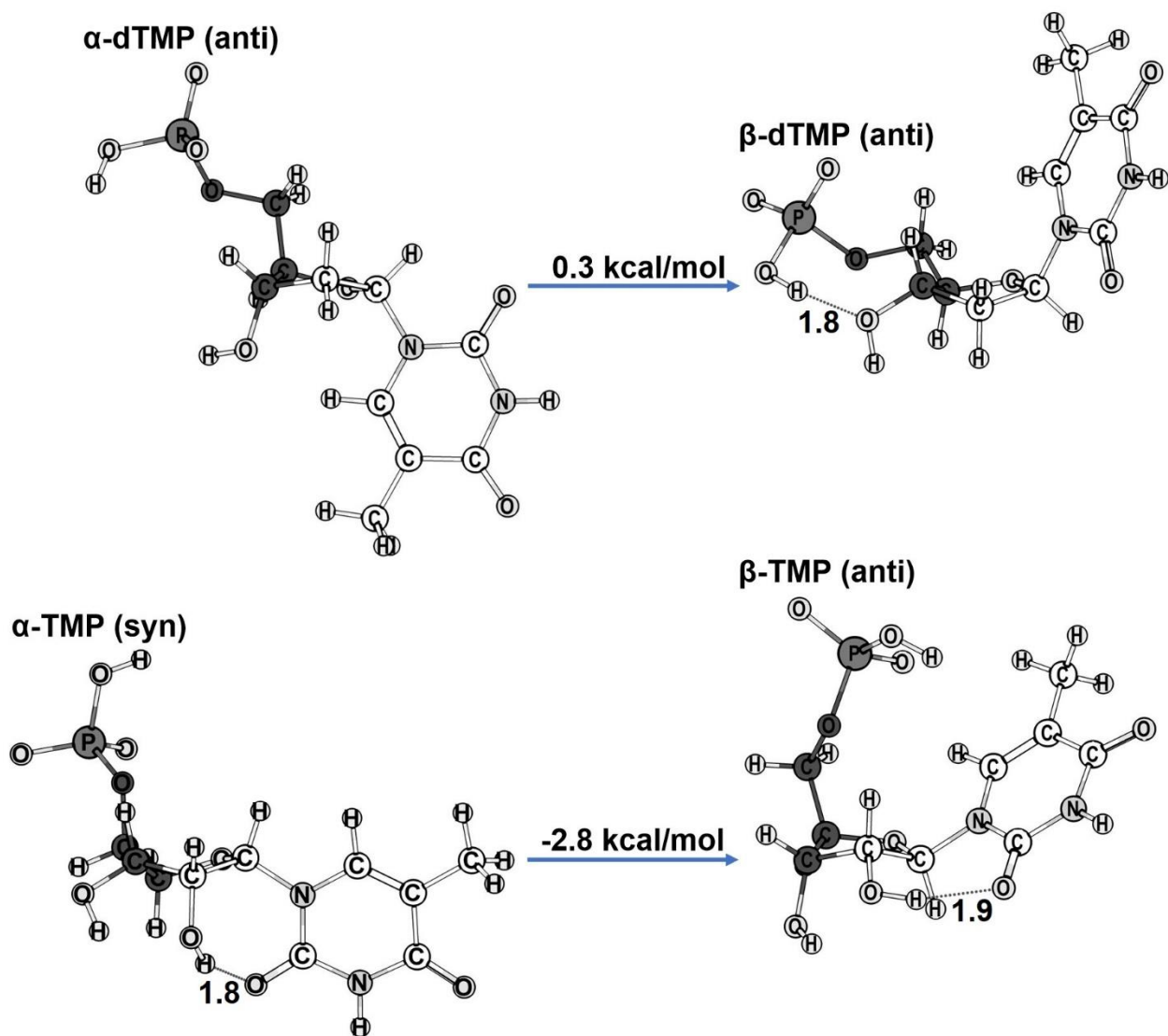
A.23 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of thymine (T) for the classic pathway (pathway (a+b), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxythymidine-5'-monophosphate (dTMP). (**Bottom**) Thymidine-5'-monophosphate (TMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
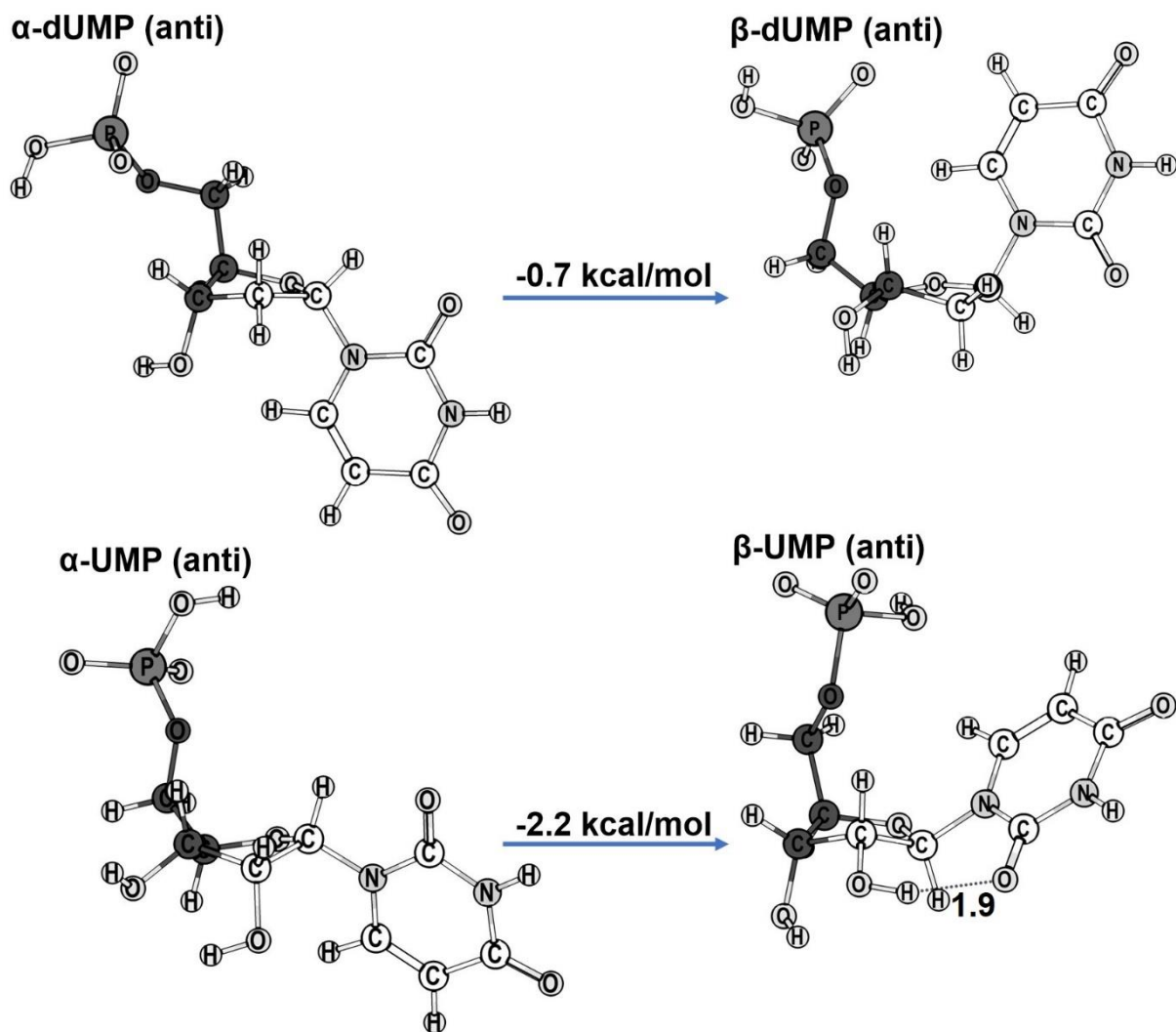
A.24 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of uracil (U) for the classic pathway (pathway (a+b), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyuridine-5'-monophosphate (dUMP). (***Bottom***) Uridine-5'-monophosphate (UMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.3**. The atoms involved in the torsion angle rotated in the PES are in bold.
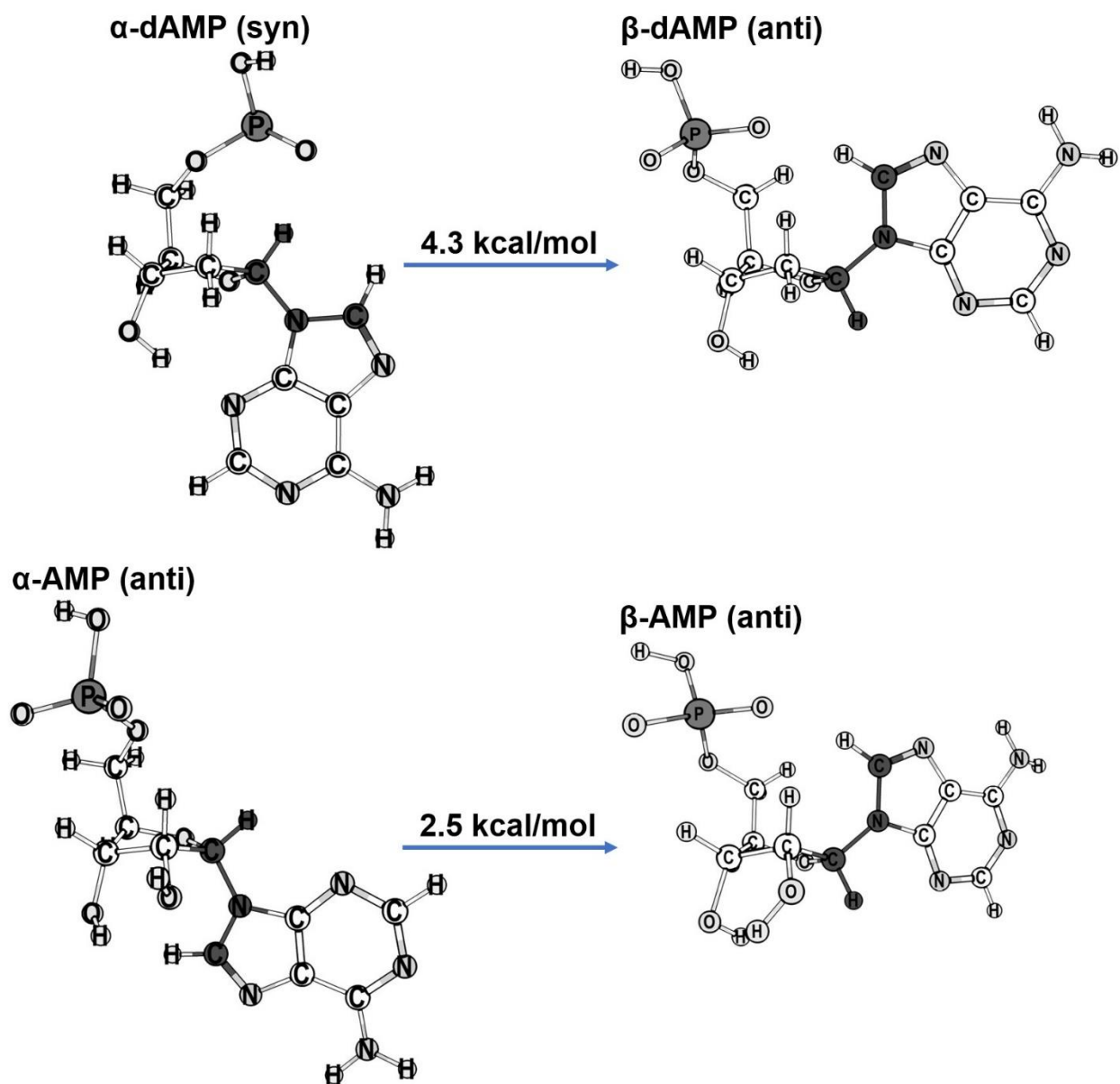
A.25 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of **A**denine (A) in vacuum for the alternative pathway (pathway (c+d), **Figure 2.2**). (***Top***) 2'-deoxyadenosine-5'-monophosphate (dAMP). (***Bottom)*** Adenosine-5'-monophosphate (AMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
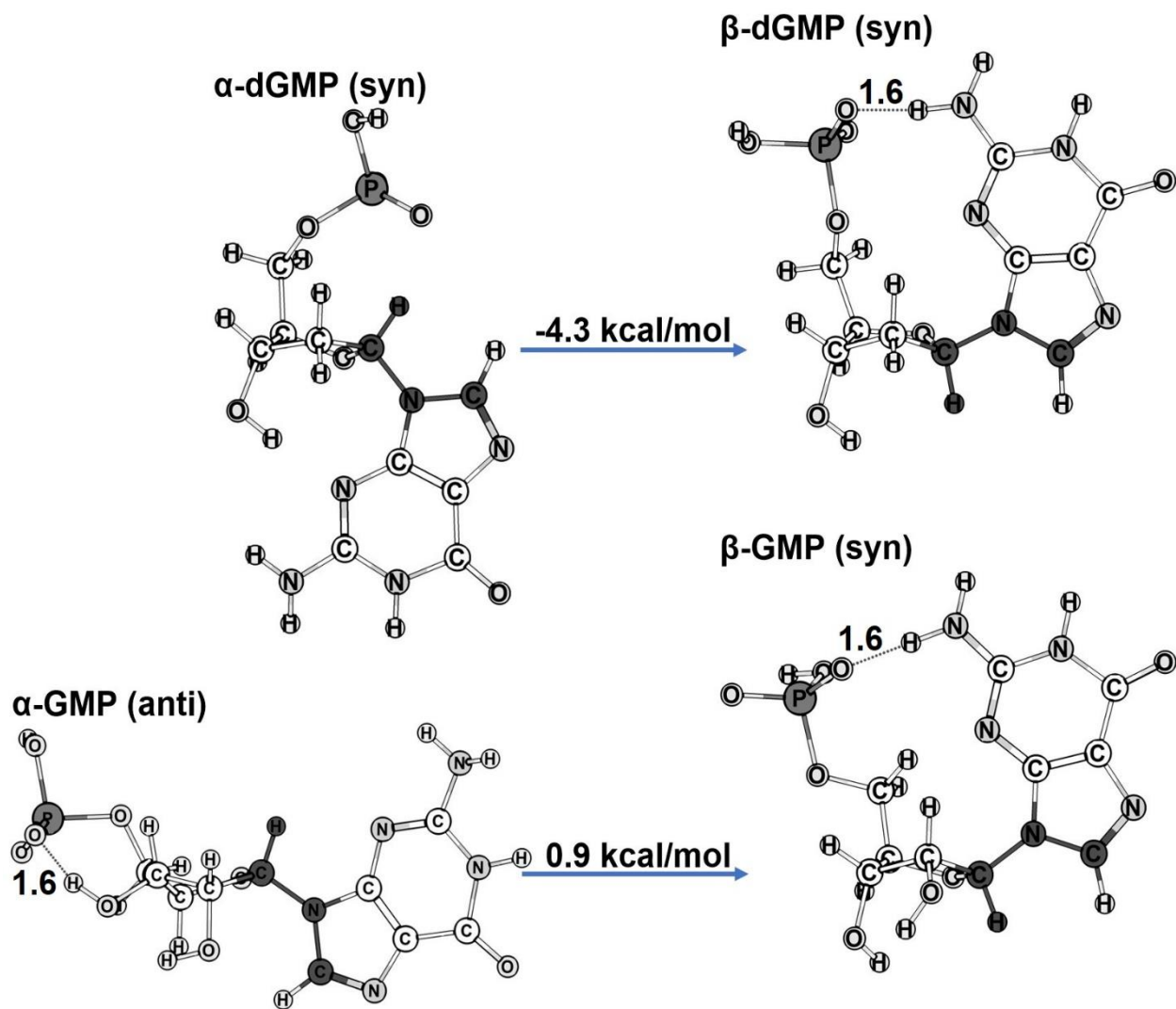
A.26 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of guanine (G) in vacuum for the alternative pathway (pathway (c+d), **Figure 2.2**). (*Top*) 2'-deoxyguanosine-5'-monophosphate (dGMP). (*Bottom*) Guanosine-5'-monophosphate (GMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
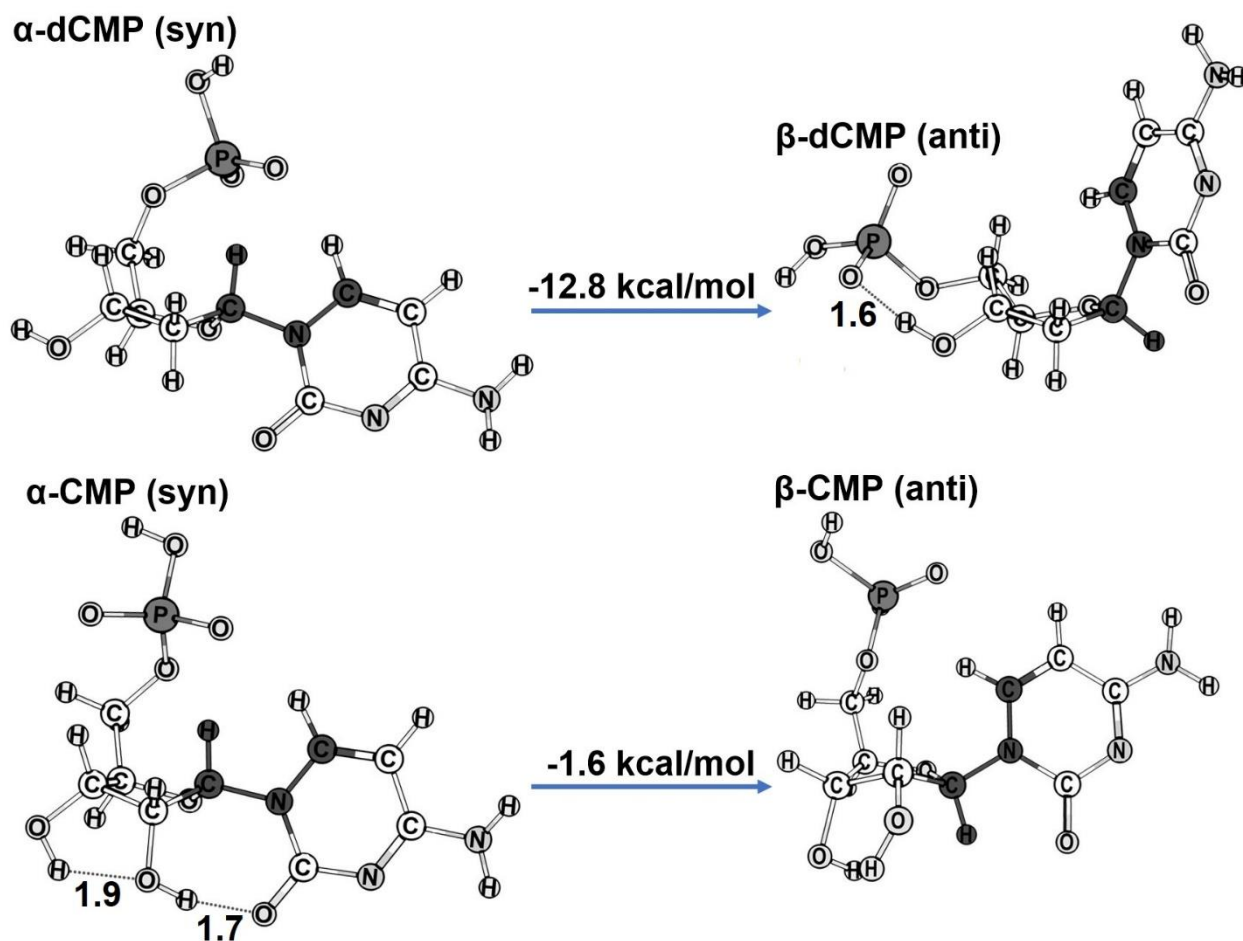
A.27 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of cytosine (C) in vacuum for the alternative pathway (pathway (c+d), **Figure 2.2**). (***Top***) 2'-deoxycytidine-5'-monophosphate (dCMP). (***Bottom***) Cytidine-5'-monophosphate (CMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
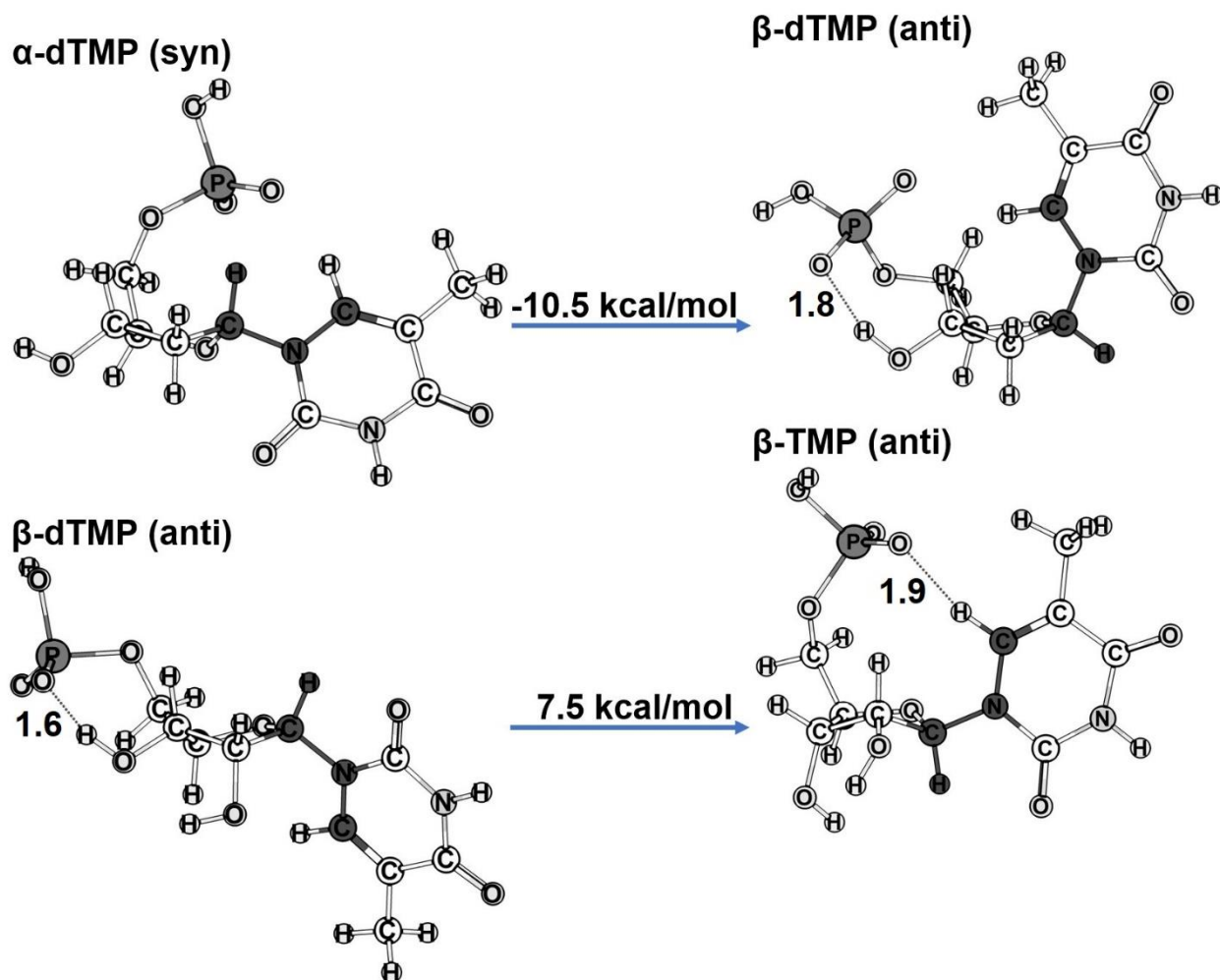
A.28 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of thymine (T) in vacuum for the alternative pathway (pathway (c+d), **Figure 2.2**). (***Top***) 2'-deoxythymidine-5'-monophosphate (dTMP). (***Bottom)*** Thymidine-5'-monophosphate (TMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.

α-dUMP (syn)   **β-dUMP (anti)**

−10.9 kcal/mol

α-UMP (anti)   **β-UMP (anti)**

11.1 kcal/mol

1.6

1.6
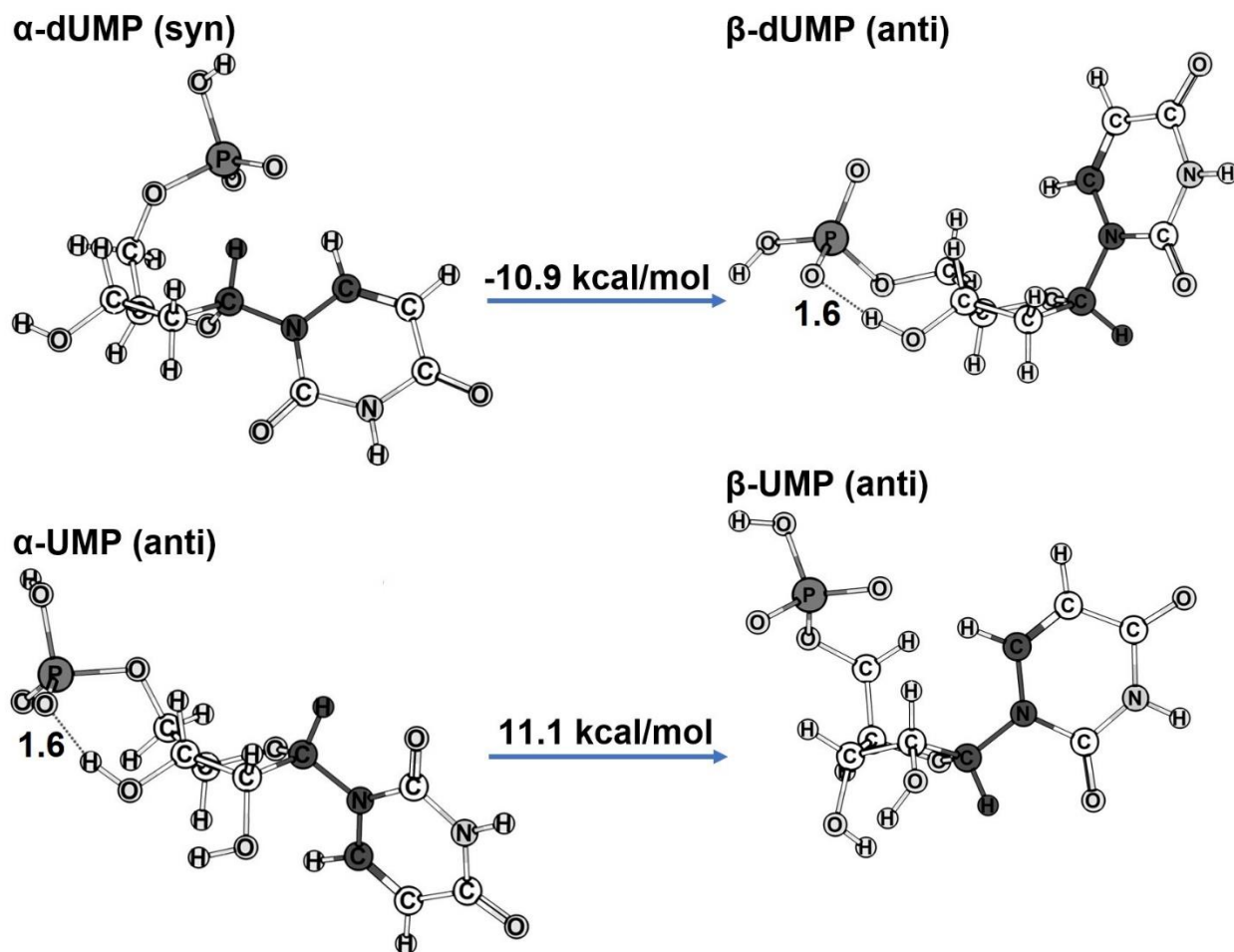
A.29 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of uracil (U) in vacuum for the alternative pathway (pathway (c+d), **Figure 2.2**). (***Top***) 2'-deoxyuridine-5'-monophosphate (dUMP). (***Bottom***) Uridine-5'-monophosphate (UMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
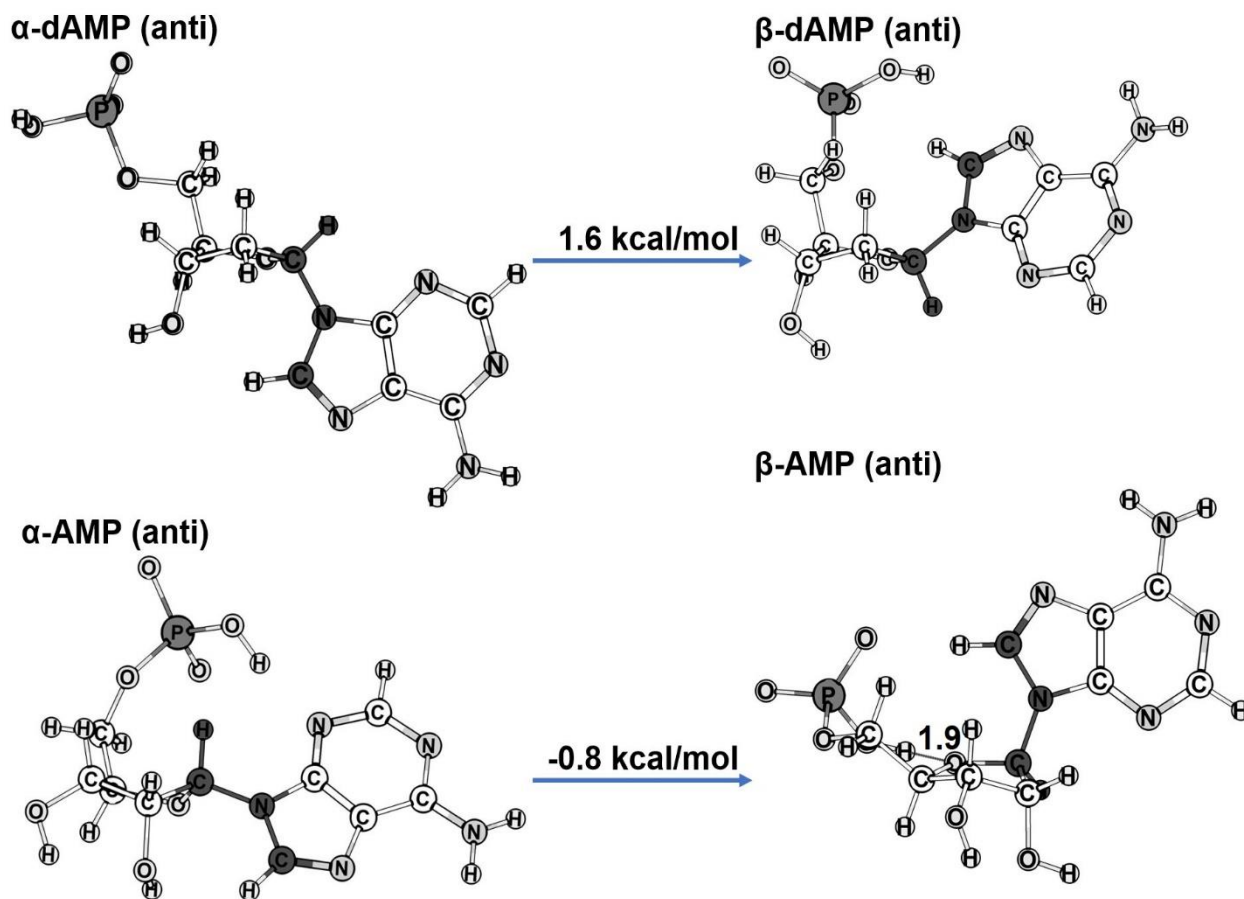
A.30 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of **A**denine (A) for the alternative pathway (pathway (c+d), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyadenosine-5'-monophosphate (dAMP). (***Bottom)*** Adenosine-5'-monophosphate (AMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
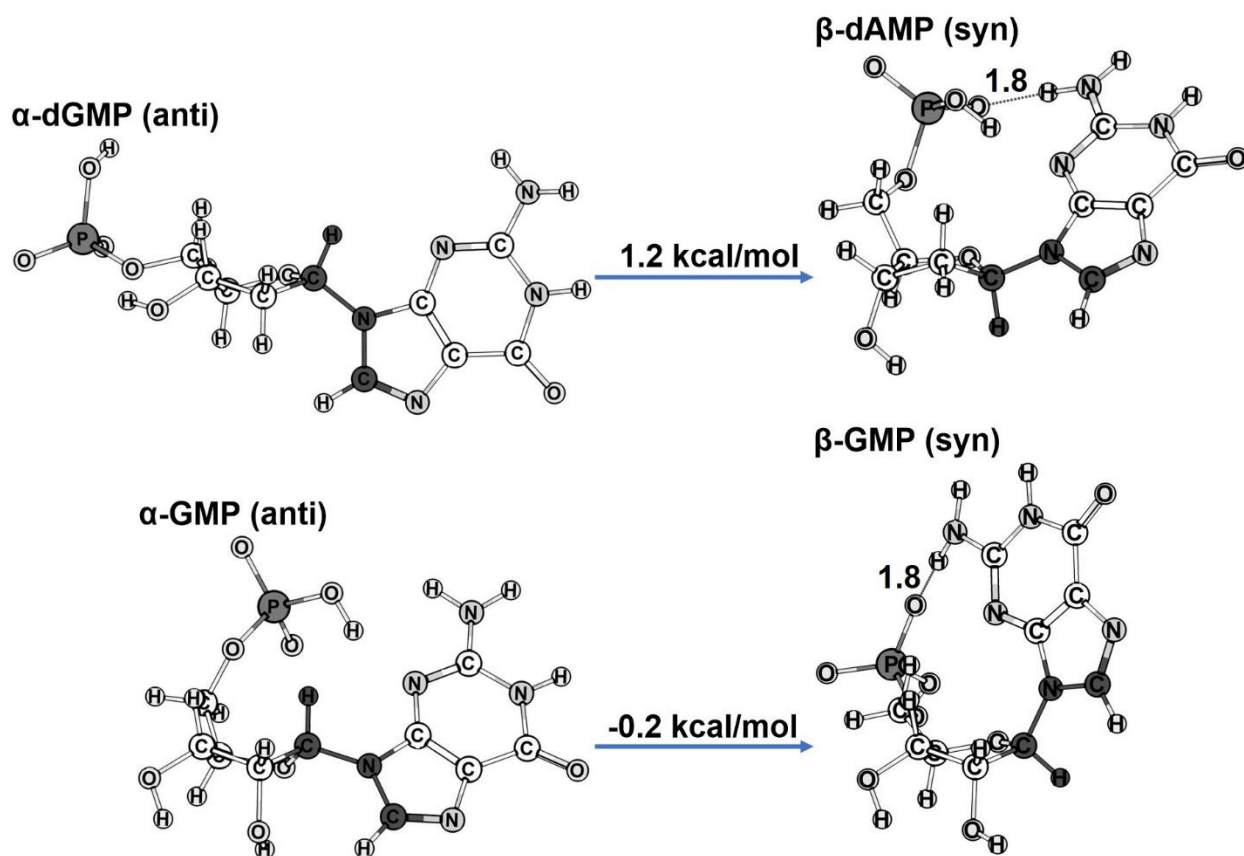
A.31 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of guanine (G) for the alternative pathway (pathway (c+d), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyguanosine-5'-monophosphate (dGMP). (***Bottom)*** Guanosine-5'-monophosphate (GMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
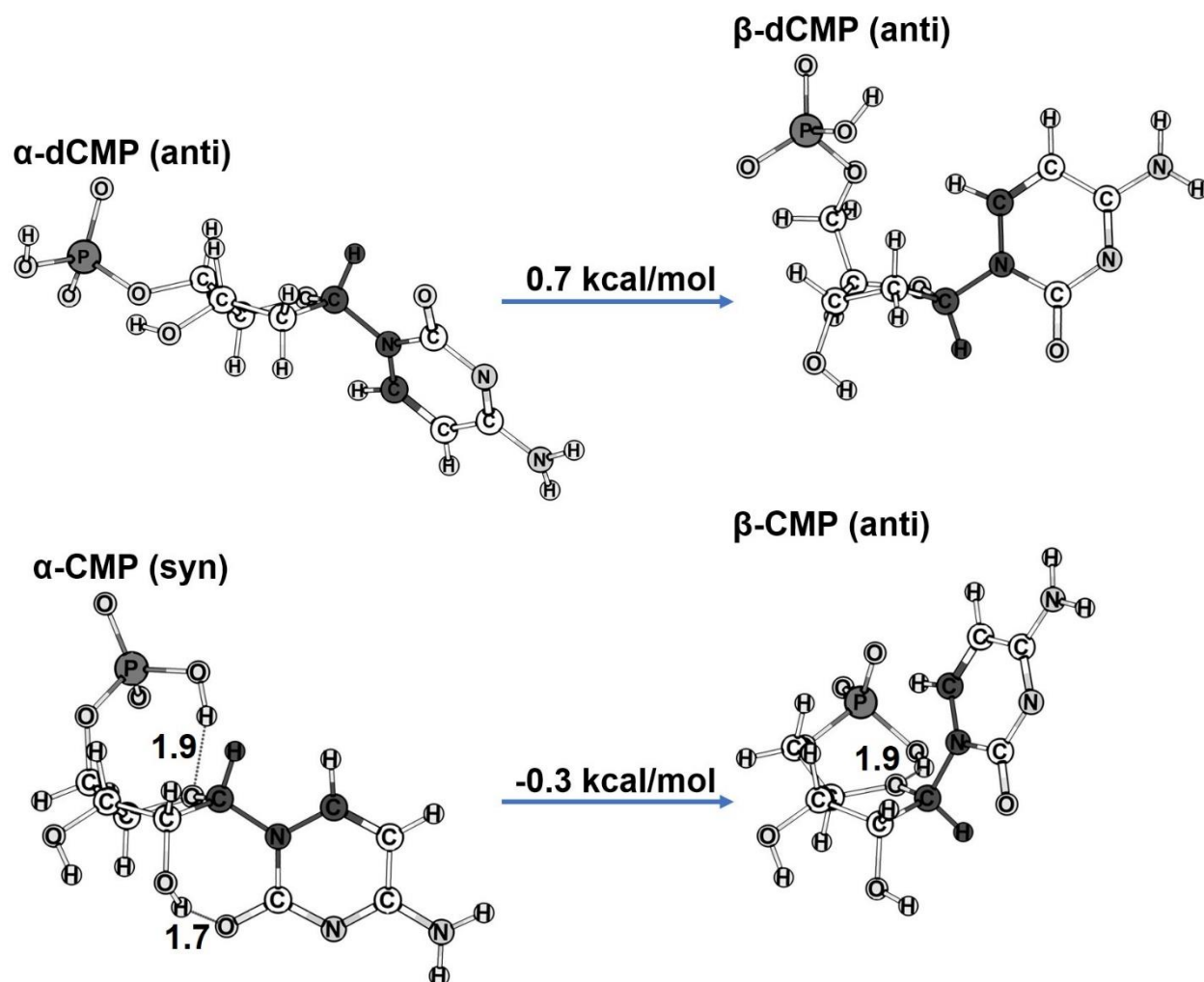
A.32 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of cytosine (C) for the alternative pathway (pathway (c+d), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxycytidine-5'-monophosphate (dCMP). (***Bottom***) Cytidine-5'-monophosphate (CMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
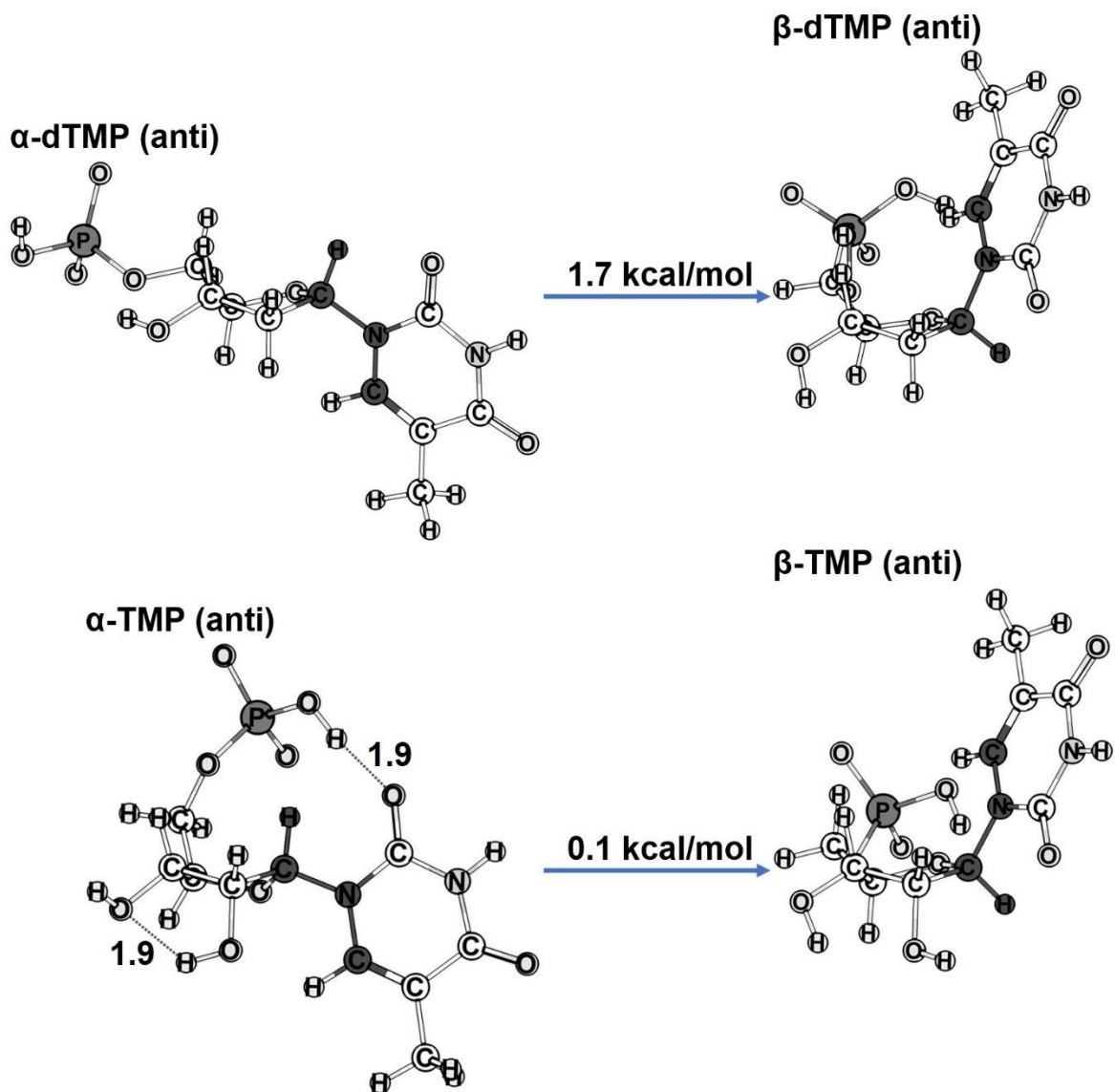
A.33 Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of thymine (T) for the alternative pathway (pathway (c+d), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxythymidine-5'-monophosphate (dTMP). (***Bottom***) Thymidine-5'-monophosphate (TMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d,p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.
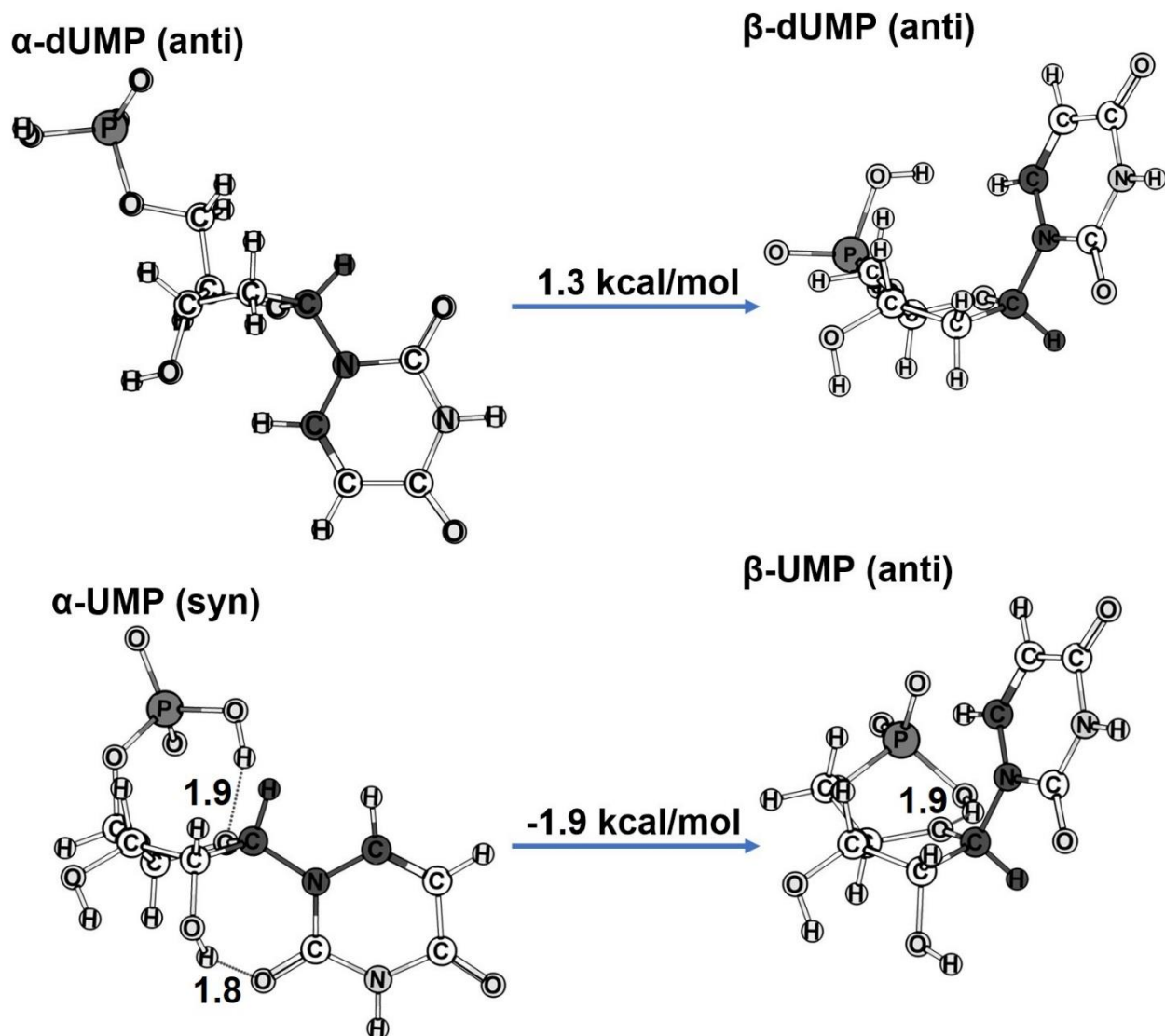
A.34   Display of the optimized geometries with bond lengths (in ångströms (Å)) for the studied *β*- and *α*-nucleotides of uracil (U) for the alternative pathway (pathway (c+d), **Figure 2.2**) obtained using the IEFPCM model for the aqueous solvation. (***Top***) 2'-deoxyuridine-5'-monophosphate (dUMP). (***Bottom***) Uridine-5'-monophosphate (UMP). The energy quoted in kcal/mol is the total energy of the *β*-form minus the total energy of *α*-form (equation 2.1) obtained at the DFT-B3LYP/6-31G(*d*,*p*). See text and **Table 2.4**. The atoms involved in the torsion angle rotated in the PES are in bold.

Chapter 3

Thermodynamic basis for the emergence of the proto-nucleosides: a computational assessment[4,5,6]

"Knowing "why" (an idea) is more important than learning "what" (the fact)"

James D. Watson, 2009

"Avoid Boring People: Lessons from a Life in Science", p.34, Vintage

Abstract

This chapter explores the thermodynamic plausibility for the prebiotic formation of β-anomers of the nucleosides over their α-counterparts through a simple condensation reaction between its components in vacuum and implicit solvation. Different, trifunctional connectors, sugar ring conformations and recognition units are considered. For the most part it has been found that there is not a definitive thermodynamic preference for one anomer over the other. Since the "classic" synthesis of nucleosides containing glycerol or N-(2-**A**mino**E**thyl)-**G**lycine (AEG) is thermodynamically favored it is also proposed that these nucleosides may have been part of an ancestral nucleic acid that may have preceded today's DNA and RNA. Finally, the correct 2'-deoxythymidine and uridine are favored over the 2'-deoxyuridine and thymidine (usually not found in nature) for nucleosides with a 6-**M**ember **R**ing (6-MR) 2'-deoxyribopyranose or ribopyranose.

---

[4] See https://github.com/mattas-research-group/scripts_PhD_thesis_Lazaro for a list of all the scripts written in bash and python to generate the nucleosides, post-process and analyze all the results and generate diagrams and tables.
[5] All final DFT-optimized molecular structures reported in this chapter are available in:
drive.google.com/PhD_Thesis/Optimized_Final_Structures/Chapter#3.
[6] All thermodynamic quantities are represented in kJ/mol since this unit is addopted by the International System of Units (SI).

3.1   Introduction

The question on how the building blocks of today's nucleic acids emerged in the early earth billions of years ago still remains one of the most intriguing enigmas for the prebiotic chemists.

A widely accepted theory for the origins of life considers that RNA was the first biopolymer to emerge. This theory is known as the "RNA world" hypothesis and proposes that RNA came first because this biomolecule has simultaneously catalytic activity **r**ibosomal RNA (rRNA) and can store the genetic information **m**essenger RNA (mRNA) [1, 2, 3].

Initial attempts on synthetizing nucleosides by Orgel and coworkers [4] following a simpler "classic model" soon showed that the formation of N-glycosidic bonds between the canonical nitrogenous bases (generically named in the literature as **R**ecognition **U**nits or RUs [5]) and ribose (a **T**rifunctional **C**onnector or TC [5]) through a condensation reaction in dehydrated conditions is rather thermodynamically unfavored in aqueous solution. Orgel attempted to glycosylate ribose with adenine, guanine, inosine and xanthine heating the reaction at 100 ℃ with and without the presence of catalysts but only obtained the a mixture of the β- and α-ribofuranosyl adenosine in ~ 2-10% of yield. Guanosine was not obtained due to the low solubility of guanine in water. Similar unsuccessful attempts to glycosylate uracil and cytosine with ribose were also reported [4].

The inability of canonical bases to create nucleosides with ribose in aqueous solution has been referred as "the nucleoside problem" (a special case of "the water problem") [6, 7, 5, 8].

Hence the classic model for the prebiotic synthesis of nucleosides imply many challenges for the chemists, alternative models have emerged in an attempt to explain how these building blocks were selected by nature, synthesized and assembled in today's nucleic acids. Some of these models include the "ribose centric model" [5] and the "polymer fusion model" [9].

On the other hand, maybe the first nucleic acids had different TCs, RUs and **I**onized **L**inkers (ILs) for their building blocks than today's ribose, the five canonical nucleobases and the phosphate group. This hypothesis has proposed that the contemporaneous nucleic acids emerged as a product of evolutionary selection from a proto- and pre-RNAs in which non-canonical (alternative) TCs and RUs glycosylated and assembled easier in prebiotic conditions [5]. Then through evolutionary changes these components became what is today β-ribofuranose, **A**denine (A), **G**uanine (G), **C**ytosine (C), **T**hymine (T) and **U**racil (U) and phosphate [10].

Kolb and Miller (KM) [11] reported prebiotic synthesis of nucleosides containing a non-canonical base (urazole or 1, 2, 4-triazolidine-3,5-dione) and ribose. KM obtained a mixture of α- and β-configurations of the urazole nucleosides with the ribose in the F- and P-forms. The β-anomer with ribose in the pyranose form was predominant with a 53% yield.

Following this idea Cafferty and coworkers [12] explored a library of 91 RUs that included the 5 canonical A, G, C, T, U for the prebiotic plausibility based on 5 criteria:

1. Can create complementary base pairing with at least two hydrogen bonds and π stacking interactions in water.
2. It is chemically photo-stable.
3. Can glycosylate with ribose in aqueous environment.
4. Could be synthesized in prebiotic conditions.
5. Is a good UV-chromophore.

Three non-canonical nucleobases: the triazine **M**elamine (MM) and the pyrimidines 2,4,6-**T**ri**A**mino**P**yrimidine (TAP) and **B**arbituric **A**cid (BA) fulfilled all 5 criteria.

These 3 bases and the triazine **C**yanuric **A**cid (CA) can create hexads (linear π stacking assemblies in the form of 6 sided polygon) in aqueous solution [13, 14, 15, 16, 17, 18]

Numerous studies have showed the ability of these RUs to overcome the "glycoside problem" by creating N- and C-glycosides with ribose, e.g., a mixture of β-N-, α-N-, β-C and α-C-nucleosides between TAP-ribofuranose and TAP-ribopyranose with a yield of ~33-55% can be obtained when heating the reactants for 24h at 35 ºC. The β-C-ribofuranose nucleoside was predominant by achieving a 20% yield [19].

Now let's address the question on whether ribose was present or not in the first proto-nucleic acids.

The most accepted synthetic method for obtaining ribose in prebiotic conditions is the "formose reaction". This reaction was first proposed by Butlerow [20] and includes the polymerization of formaldehyde (H_2CO) to obtain sugars. This method presents a major challenge: "the asphalt problem" in which the simpler sugars can polymerize (create asphalts) through a series of competitive enolizations and aldol additions between the carbonyl groups. As a result ribose is minor product with only ~2% of yield. Additionally, ribofuranose with a 5-**M**ember **R**ing (5-MR) is the least of the products from the mutarotation of ribose with a yield around ~13% for the β- and ~7% for the α-anomer [21, 22].

In order to overcome the different challenges associated with considering ribose as the TC present in the first nucleic acids an alternative solution can be to consider that the first TCs were different from today's ribofuranose.

The nucleic acids with a TC different from ribose are named as "**X**eno-**N**ucleic **A**cids" (XNA) and they have been used in Biotechnology in different applications, e.g., as biomarkers for cancer, AIDS and hepatitis [23].

Nucleic acids analogs with a TC-phosphate backbone in which the TC is either a 6-**M**ember **R**ing (6-MR) sugar, e.g., (p)-RNA) [24, 25], threose (TNAs) [27, 92, 93], glycerol (GNA) [5, 28, 29, 30, 31, 32, 33] and **P**eptide **N**ucleic **A**cids (PNAs) – where the sugar-phosphate is replaced by a peptide backbone [34, 35, 36, 37, 38, 39, 31] have been proposed in the literature as plausible candidates for a proto-nucleic acid.

The question of how evolution selected phosphate as part of today's nucleic acids has also faced numerous roadblocks. These include the thermodynamic instability of phosphodiester bonds in aqueous solution as part of the "water problem" [40], the chemical inertness of minerals containing phosphorus (P) and the poor regioselectivity of the phosphorylation of ribose [41]. In the search for an alternative solution arsenic (As) has been proposed in the literature as a viable candidate [42] based on its similar chemical nature to P and a report by Wolfe-Simon *et al*. that reported a cyanobacteria that can metabolize As to multiply [43]. But a question remains: why and how Nature replaced As which is toxic for humans by P?

We may now turn to another fundamental question related to the chemical nature of today's nucleic acids. The building blocks of contemporaneous DNA and RNA contain a β- instead of an α-configuration at the anomeric C1' position of the sugar D-ribofuranose (**Figure 3.1**). But why not α?

Numerous experimental reports have suggested that α-strands of DNA can create **W**atson & **C**rick (WC) complementary α - α and α - β double stranded polynucleotides [111, 14, 12, 15, 16].

It is well known that 6-MR sugars (hexopyranoses), e.g. galactopyranose, glucopyranose, adopt two preferential axial A_1 or $^1C_4$ and equatorial E_1 or $^4C_1$ chair conformations in equilibrium for their respective β- and α-anomers (**Figure 3.2a**). The displacement of this equilibrium depends on the contribution of different stereoelectronic factors (see Chapter #1 in [50]) and [51, 52, 53]. Following the same principle, Thibaudeau and coworkers (see Chapter #2 of [50] and [54, 55])
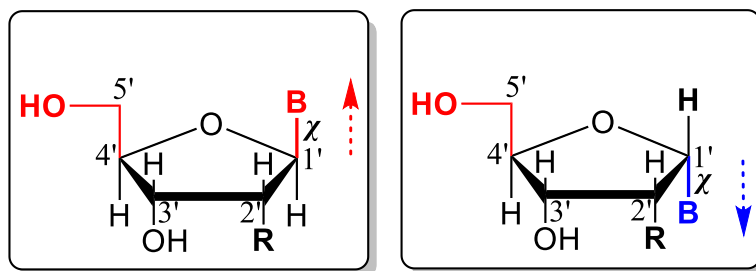
Figure 3.1 Orientation of the nucleobase at the C1' of the furanose with respect to the hydroxymethyl group at the C4' for the (left): *β*- and (right): *α*-anomers of the canonical ribonucleosides. The substituent R can be either H in 2'-deoxynucleosides (in DNA) or OH in ribonucleosides (in RNA) (taken from [44] and reprinted with permission of RSC Advances).

proposed that the conformational changes of the 5-MR ribofuranose is an important factor to take into account when analyzing the stability of the *β*- and *α*-anomers of the building blocks of nucleic acids (**Figure 3.2b**).

The furanose ring has different conformers that can be described by the generalized puckering parameters or puckering coordinates, of sugar rings (changes in the endocyclic torsion angles of the furanose/pyranose ring) proposed by **C**remer and **P**ople (CP) in 1975 [56].

A recent publication by Castanedo *et al* 2022 [57] explored the role of thermodynamics as a driver of evolutionary selection for the chemical structure of today's nucleic acids in implicit solvation models and vacuum-gas phase using computational modeling.

This paper estimated the energetic difference between the *β*- and the *α*-anomers of D-ribofuranose, D-2'-deoxyribofuranose, corresponding sugar monophosphates, nucleosides and nucleotides and the thermodynamic changes for a sugar exchange reaction between U and T in the corresponding nucleosides/nucleotides in the two configurations (*α*, *β*) in an attempt to explain if thermodynamic may have guided the natural selection of T for DNA and U for RNA.

Additionally, this paper also estimated the thermodynamic feasibility for the synthesis of the five canonical nucleotides through a classic and alternative pathway in which the order of the reactants was changed.

The results presented in this publication suggested no (or marginal, at best) energetic difference beyond the intrinsic error of the computational methods between the anomers of the
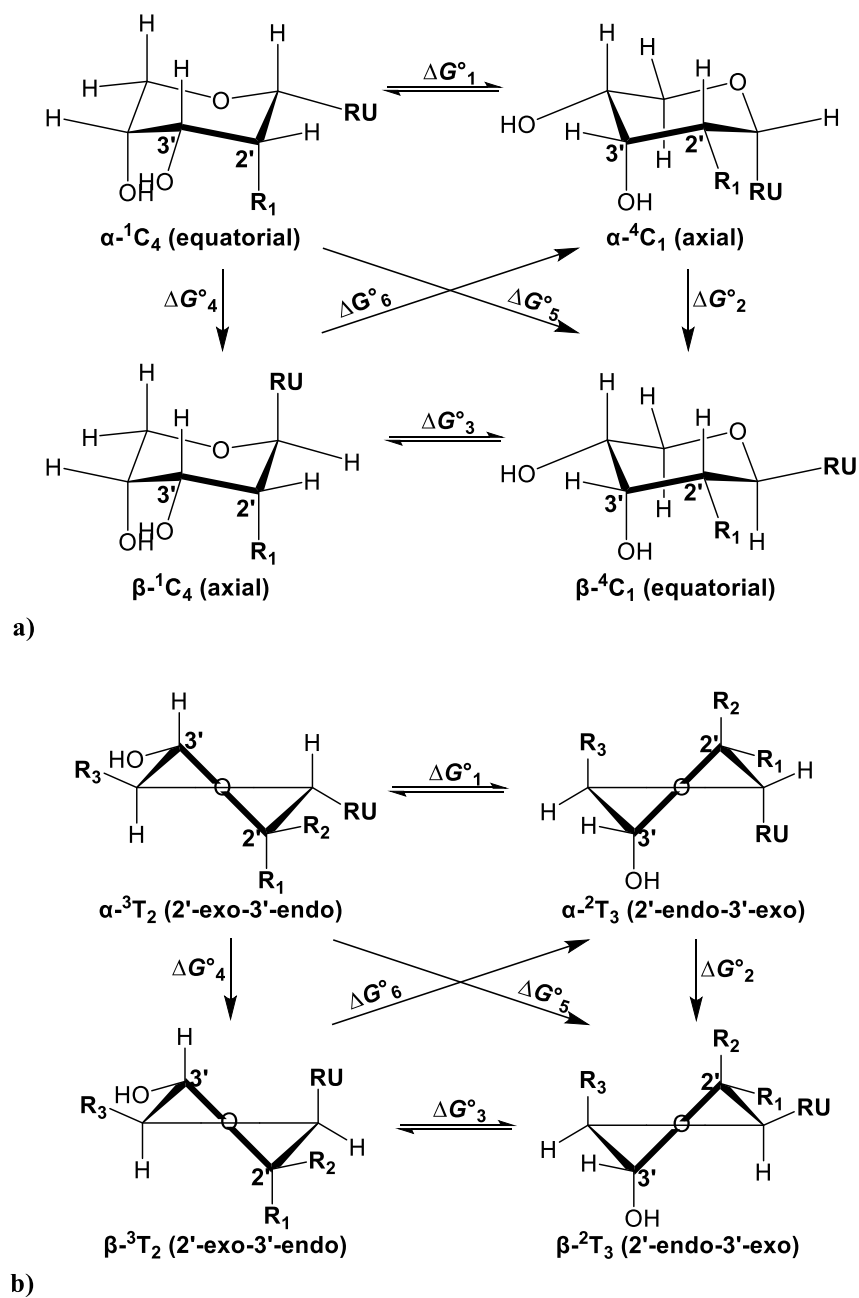
Figure 3.2 **a)** Equilibrium between the E_1 ($^4C_1$) and A_1 ($^1C_4$) conformations for a nucleoside with a 6-MR sugar (see pp. 8-20 in [50] and [54, 55]), **b)** equilibrium between the $^3T_2$ (2'-exo-3'-endo) and $^2T_3$ (2'-endo-3'-exo) conformations for a nucleoside with a 5-MR sugar (R_1: H for 2dRibf and Tho, OH for Ribf. R_2: H for 2dRibf, Ribf and OH for Tho. R_3: CH_2OH for 2dRibf and Ribf and H for Tho). (see pp. 22-46 in [50] and [54, 55]) (modified from [54] and reprinted with permission of Elsevier).

D-ribofuranose (r) and D-2'-deoxyribofuranose (dr). A slightly thermodynamic advantage that favors the selection of the β- over the α-anomers for some nucleosides is estimated. The analysis of the energy changes for the sugar exchange reaction of T and U: (*i*) reinforces the conclusion that the classical pathway is favoured and (*ii*) indicates an advantage of the canonical pairs compared to the no-canonical pairs when gauged by the "sugar exchange reactions" for the β-anomer in vacuum and vanishes in the case of the α-anomers. The classic pathway in which two condensation reactions give the nucleotides as product was estimated overall as the preferred artificial synthetic route in vacuum. Non synthesis of nucleotides through either pathway was estimated to be favored in aqueous solution. This result is a validation of the "water problem" using quantum computational chemistry.

This chapter will explore the influence of the sugar ring conformation in the pseudorotational equilibrium and the different anomer-exchange reactions between two preferential conformations for furanose and pyranose rings in their corresponding α- and β-anomers. For this, the different combinations between canonical and non-canonical TCs and RUs will be considered. The main question to be answered with this study will be the following: does Thibaudeau and coworkers' hypothesis (see Chapter #1 and #2 of [50]) apply as well to non-canonical TCs and nucleosides?

Additionally, the following questions will be addressed on the thermodynamic basis using QM and computational modeling:

1) Can thermodynamics explain why the sugar ring of ribose and 2'-deoxyribose chose a 5-MR (furanose) instead of a 6-MR (pyranose)?

2) Is the synthesis of canonical and/or non-canonical nucleosides possible following the "classic model"?

3) Is the "water problem" really a problem?

4) Are the canonical bases conformation "*syn*" and "*anti*" in nucleic acid biopolymers still predominant in free canonical and non-canonical nucleosides?

5) Is the sugar puckering of canonical and non-canonical nucleosides comparable?

6) Can thermodynamics explain the preference of T in DNA and U in RNA? What would be the preference if the ribose or 2'-deoxyribose had instead of a **F**uranose form (F-form) a **P**yranose form (P-form)?

3.2   Computational methods

The chemical structures of all the canonical and non-canonical components (TCs and RUs) and nucleosides from DNA, RNA and TNA were modified from the nucleotides contained in the pdb 1BNA [58]. For the case of **p**yranosyl-RNA {(*p*)-RNA} and **p**yranosyl-DNA {(*p*)-DNA} building blocks and components the sequence provided by Dr. Froeyen, M [111] was used and for the aegPNA the pdb 1PNN [59] was modified accordingly. The graphic interfaces UCSF Chimera [60], Hyperchem Release 7.0 [61] and GaussView 5.0 [62]. were used to create the 3D structures.

All possible combinations between each TC with each RU and IL represented in **Figure 3.3** were explored.

The keto form of BA is predominant in solid state, meanwhile in aqueous solution BA is predominant in its enol form through a keto-enol equilibrium [63]. Hence we have used in all our modeling calculations the enol form of BA.

Additionally, for the sugars in their furanose and pyranose form each RUs was positioned in the *β*- and *α*-configuration at the anomeric C1' of the TC.

The sugar ring conformations $^2T_3$ or 2'-exo-3'-endo, $^3T_2$ or 2'-endo-3'-exo and E_1 ($^4C_1$), A_1 ($^1C_4$) were considered for the 5-MR (threofuranose, 2'-deoxyribofuranose, ribofuranose) and 6-MR sugars (2'-deoxyribopyranose and ribopyranose), respectively since accordingly to Thibaudeau and coworkers (see Chapter #2 of [50] and [54, 55]) is an important factor to include in the modeling. Hence, in order to preserve these conformations the coordinates of the carbon atoms and the hemiacetalic oxygen in the furanose and pyranose rings were kept frozen during the PM7 and DFT calculations while relaxing the rest of the atoms coordinates.

Each of the TCs (considering each of the *β*- and *α*-anomer and the different sugar ring conformations) and RUs with rotatable bonds (see red arrows in **Figure 3.3**) were subjected to a **p**otential **e**nergy **s**can (PES) by reading their *Z*-matrices in the programme GRANADAROT [64, 65, 66]. This algorithm was used to generate for each of the component 1000 different conformers by randomly changing each rotatable bond.

The number of conformers represent a balance between good sampling of the conformational space and computational time.

Each of the initial (29,000 = {6000 for RUs} + {23000 for TCs}) randomly generated conformers were optimized with at the semiempirical Parametric Method 7 (PM7) (**P**arametric
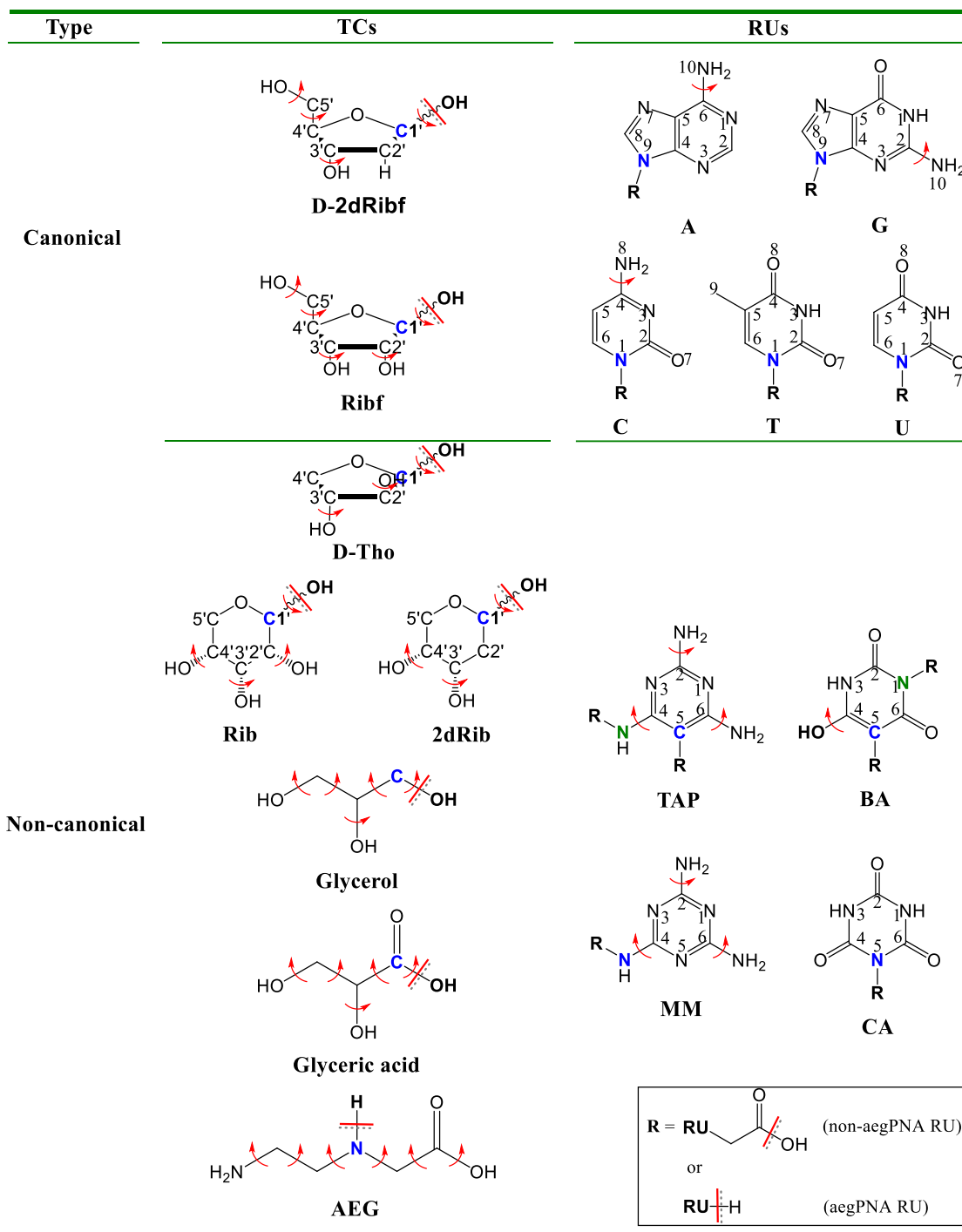
Figure 3.3 Canonical and non-canonical recognition units (RUs) and **T**rifunctional **C**onnectors (TCs) considered in the modeling of the nucleosides. A: adenine, G: guanine, C: cytosine, T: thymine, U: uracil, TAP: 2, 4, 6 - triaminopyrimidine, BA: barbituric acid in enol form, MM: melamine, CA: cyanuric acid, D-2dRibf: D-2'-

deoxyRibofuranose (*β*-anomer present in DNA), D-Ribf: D-Ribofuranose (*β*-anomer present in RNA), D-Tho: D-Threose (L-enantiomer is present in TNA), D-2dRib: D-2'-deoxyRibopyranose (present in *p*-DNA), D-Rib: D-Ribopyranose (present in pRNA), Glycerol and glyceric acid (present in GNA) and AEG: N-(2-AminoEthyl)-Glycine (present in PNA). ┅┅: Bonds been broken during the condensation reactions. ⤙: Rotatable bonds changed during modeling of the potential energy surface at the semiempirical PM7. The R group is H for the RUs in non-PNAs nucleosides and acetyl (-CH_2-COOH) for acetyl derivatives of Rus in the PNAs nucleosides. The blue and green C and N represent reactive centers in the condensation reaction. Notice that in the case of TAP and BA they can either create N- or C^5-nucleosides.

Method 7) [67] level of **QM** through a gradient minimization of energy until convergence using the **M**olecular **O**rbital **PAC**kage (MOPAC) 2016 package [67]. This semiempirical method includes empirical corrections for dispersive and hydrogen bonding interactions which are crucial in the modeling of biological molecules subject to our interest.

The geometry optimization at the PM7 level were conducted through the minimization of the energy gradient until a value ~ 0.0 kcal/mol/Å were the forces on the nuclei were negligible and the calculation was performed in vacuum and using the **CO**nductor-like **S**creening **MO**del (COSMO) solvation model [68, 69] that considers continuum implicit hydration effects.

Some of the 1000 optimized structures in the different sets converged to the same final geometry. For each set of final *n* unique optimized geometries the final *n'* conformers that collectively contribute at least 50% to the partition function *Z* were kept for refinement at a more accurate QM level and the rest of the conformers with minor contributions were discarded.

For each TC, RU and acetylated RU with rotatable bonds in vacuum and solvation the numbers *n*, *n'* and their contribution to *Z* are summarized in **A1-A2** (in the Appendices section).

The resultant *n*' conformers were subjected to a fully-relaxed geometry optimization using the B3LYP DFT functional [70, 71, 72, 73] with the 6-311++(d, p) basis set [74], as implemented in the Gaussian 16 package [75].

The B3LYP functional has been widely used for the computational modeling of the building blocks of nucleic acids [76, 77, 78].

The DFT-B3LYP/6-311++G(*d, p*) level of theory has been chosen for this study as a reasonable compromise of accuracy and speed/feasibility and also as a way to reduce the over estimation of energies due the **B**asis **S**et **S**uperposition **E**rror (BSSE). The error bars for a similar level of theory, namely, DFT-B3LYP/6-31+G(*d, p*) have been benchmarked by Zhao and Truhlar (ZT) to be around 15.0 kJ/mol (3.6 kcal/mol) [79]. ZT obtained this estimate by comparing the calculated and experimental thermodynamic data for 177 main-group compounds) [79]. On another study by ZT [80] it was reported an average mean unassigned error of 13.8 kJ/mol (3.29 kcal/mol) for the estimation of the interaction energies from the same B3LYP/6-31+G(*d, p*) for 22 hydrogen-bonded complexes. Finally, Rao and coworkers [81] evaluated 11 DFT functionals in the accuracy of the estimation of hydrogen bonding energies and relative energies of a conformational scan for 14 systems of biological interest that included the amino acids glycine, proline, and serine. In this study it was estimated that the DFT-B3LYP/6-31++G(2*d*, 2*p*) was the most accurate for predicting conformational energies and the mean absolute deviation of the predicted binding energies for this method was 6.1 kJ/mol (1.5 kcal/mol). On that basis, we may take the intrinsic uncertainty of the method used in this work to be around $\approx$ 8-17 kJ/mol (2-4 kcal/mol).

In total 506 nucleosides are designed. The total number of nucleosides included 440 nucleosides with a sugar as TC {(11 RUs) x 2 environments (vacuum and implicit solvation) x 20 sugar ring conformations (12 [6 anomers (β and α]) for the 5-MR and 8 [4 anomers (β and α] for the 6-MR sugars respectively)} and 66 nucleosides with a non-sugar TC {(11 RUs) x 2 environments x (3 TCs)}.

For consistency, the C- and N-glycosidic bond between each canonical and non-canonical RUs and TCs, was initially set to 1.52 Å while the dihedral angle that contains the glycosidic bond was set initially to -179.5º. Each of the 506 nucleoside structures was then subjected to a fully relaxed scan around this dihedral from 0-360º in 6 steps of 60º. The lowest energy structure (minimum of PES) from each scan was refined by subjecting it to a final fully unconstrained optimization to obtain the final structure of the nucleosides

in vacuum and in solvent. A harmonic frequency calculation was performed as usual to ensure that the final structures are indeed minima on the PES.

The thermodynamics of the pseudorotational equilibrium of the TCs with 5-MR and 6-MR and their corresponding nucleosides was analyzed by estimating the corresponding $\Delta G°$ for the equilibriums 1 and 3 of **Figure 3.2**. Additionally, all possible anomeric exchange reactions in the same figure were analyzed. The $\Delta G°$ for each conversion step in **Figure 3.2** was calculated as follows:

$$\Delta X_1 = X\,(\alpha\text{-}^4C_1) - X\,(\alpha\text{-}^1C_4) \text{ or } X\,(\alpha\text{-}^2T_3) - X\,(\alpha\text{-}^3T_2) \tag{3.1}$$

$$\Delta X_2 = X\,(\beta\text{-}^4C_1) - X\,(\alpha\text{-}^4C_1) \text{ or } X\,(\beta\text{-}^2T_3) - X\,(\alpha\text{-}^2T_3) \tag{3.2}$$

$$\Delta X_3 = X\,(\beta\text{-}^4C_1) - X\,(\beta\text{-}^1C_4) \text{ or } X\,(\beta\text{-}^2T_3) - X\,(\beta\text{-}^3T_2) \tag{3.3}$$

$$\Delta X_4 = X\,(\beta\text{-}^1C_4) - X\,(\alpha\text{-}^1C_4) \text{ or } X\,(\beta\text{-}^3T_2) - X\,(\alpha\text{-}^3T_2) \tag{3.4}$$

$$\Delta X_5 = X\,(\beta\text{-}^4C_1) - X\,(\alpha\text{-}^1C_4) \text{ or } X\,(\beta\text{-}^2T_3) - X\,(\alpha\text{-}^3T_2) \tag{3.5}$$

$$\Delta X_6 = X\,(\alpha\text{-}^4C_1) - X\,(\beta\text{-}^1C_4) \text{ or } X\,(\alpha\text{-}^2T_3) - X\,(\beta\text{-}^3T_2) \tag{3.6}$$

The relative position of each nucleobase with respect to the TC was analyzed by estimating the torsion angle around the glycosidic bond χ.

In the case of nucleosides with 5-MR the and 6-MR the final sugar ring conformation or ring puckering after the DFT optimization was analyzed by estimating the **C**remer-**P**ople (CP) generalized puckering parameters [56] and using circular statistics for the calculation of the circular mean and standard deviation. These parameters are the phase angle phi (ϕ_2) and the total puckering amplitude (Q) for furanose rings and the phase angles {phi (ϕ) and theta (θ)} and the radial Q for pyranoses. An in-house bash-Python script that used the Ring programme [82] was implemented for this task.

The thermodynamic feasibility for the synthesis of the canonical and non-canonical nucleosides following the "classic model" was analyzed by estimating the $\Delta G°$ for the condensation reaction of $TC\text{-}\mathbf{OH} + \mathbf{H}\text{-}RU \rightarrow nucleoside + H_2O$ for non-PNA nucleosides and $AEG\text{-}\mathbf{H} + \mathbf{HO}OC\text{-}CH_2\text{-}RU \rightarrow nucleoside + H_2O$ for the PNA nucleosides.

Aqueous solvation has been accounted for in the DFT calculations using an **I**ntegral **E**quation **F**ormalism variant of the "***Polarizable Continuum Model***" (IEFPCM) [83, 84, 85, 86, 87, 88] implemented in Gaussian 16 [75], the software package used in all DFT calculations in this work.

If we accept the "water problem" [8] this implies that the complex prebiotic chemistry that produced the first building blocks of ancestral NAs took place in a non-polar or non-aqueous) environment or at least the reactants had controlled exposure to water (see for example Ref. [89] and literature cited therein). Still we decided to analyze the effect of solvent in this study as a surrogate to complete the analysis and use the evaluation of this "water problem" as a validation test for the accuracy of the theoretical calculations.

Every geometry optimization at the DFT level in this work for each of the final TCs, RUs and nucleosides has been followed by a harmonic vibrational analysis finding in all cases no imaginary frequency which demonstrates that the structures are local minima on the PES.

In the next sections the different TCs will be referred to by using the following abbreviations: D-2dRibf: 2'-deoxyribofuranose (β-anomer present in DNA), D-Ribf: D-ribofuranose (β-anomer present in DNA), D-Tho: D-threose (L-enantiomer in TNA), 2dRib: D-2'-deoxyribopyranose (present in p-DNA), Rib: ribopyranose (present in pRNA), Glycerol and glyceric acid (present in GNA) and AEG: N-(2-aminoethyl)-glycine (present in PNA).

$^{2}T_{3}$ (C2'-endo-C3'-exo) and $^{3}T_{2}$ (C2'-exo-C3'-endo) (**Figure 3.2b**) refer to the two preferential 5-MR puckering conformations of the furanoses 2dRibf, Ribf, Tho and associated nucleosides, meanwhile $^{1}C_{4}$ and $^{4}C_{1}$ (**Figure 3.2a**) refer to the two preferential 6-MR conformations of the pyranoses 2dRib, Rib and corresponding nucleosides. The 2, 4, 6-triaminopyrimidine (TAP) and the barbituric acid (BA) can create either N- or C-glycosidic bonds [40, 16], hence in the next sections of this chapter TAP-C^{5} refers to non-canonical nucleosides with a C-glycosylated- and TAP-N refers to the nucleosides with a N-glycosylated-TAP respectively. Similarly for BA; BA-C^{5} refers to the nucleosides with the BA C-glycosylated and BA-N refers to the nucleosides with a TC-N-BA.

3.3 Results and discussion

3.4.1 Which sugar ring conformation and anomer is more stable in ribose?

The estimation of the thermodynamic differences between the sugar rings of difference size for ribose addresses the question on: why Nature chose the F-form over a P-form for the 2'-deoxyribose in DNA and ribose in RNA?

As a first step on finding an answer for this question let's examine previous studies on the comparison of the relative stability of both F- and P-forms.

When β- or α-ribose is in aqueous solution it can exist not only as the aldopentose open chain but also in its F- and P-form through an equilibrium called "mutarotation" [90]. The major product from the mutarotation of ribose in aqueous solution is the β-anomer of the pyranose form with ~59%, followed by the α-anomer with ~20%. The β-furanose forms accounts for ~13% of the mixture, while the α-furanose only represents ~7%. In mutarotation both pyranose and furanose rings can interconvert to each other through an equilibrium that goes through the open aldopentose chain which only represents < 1% [21, 22].

Dass and coworkers (Dcw) [91] used ^{13}C-NMR and statistical mechanics analysis to study the significance of the dependence between the equilibrium of the open chain and cyclic forms of ribose with the temperature gradient. This study simulated which form of ribose is predominant in an environment that simulates the hydrothermal deep vents in terms of temperature and temperature gradient. The deep vents are considered as one of the possible scenarios for the origins of life in the early Earth [92, 93, 94].

Dcw found (see Fig. 2 of [91]) that at room temperature and in water solution the β-P-form is predominant (mole fraction $\approx$ 0.6). The two β- and α-F-forms were obtained in the least quantities with a molar fraction of $\approx$ 0.1. These results suggest that at **N**ormal **P**ressure and **T**emperature (NPT) conditions the 6-MR prevails over the 5-MR, mostly in its β-form and this tendency does not change much when the authors repeated their experiments in a saline solution potentially similar to the Hadean ocean.

Meanwhile, the results of this study seem to suggest that pyranose is favored over furanose a detail examination of Fig. 2 of this paper [91] reveals that beyond $\approx$ 130 ºC the furanose ring tends to be more favored with a directly proportional increase of a gap

between the *α*- and *β*-anomer, in favor of the first one with respect to the increase of temperature, with a *β/α* ratio of ≈ 2 at ≈ 130 °C, either in aqueous or aqueous solution. This result suggests that a hot environment like the one probably in the prebiotic Earth could have shifted the mutarotation equilibrium in favor of the *β*-furanose form (observed in today's nucleic acids) over the pyranose ring.

On another study Azofra *et al.* (Acw) [95] modeled in vacuum at the B3LYP/6-311++G(*d*, *p*) and M06-2X/6-311++G(*d*, *p*) DFT levels of theory the PES of 2'-deoxyribose in its open chain, P- and F-forms by generating thousands of rotamers. Acw also considered the cyclic forms in their *β*- and *α*-configurations. Similarly to the previous referred study by Das *et al.* [91] these authors also found a preference of the 2'-deoxyribose for its pyranose form having the highest Boltzmann populations in % at 0 K and 298 K.

Cocinero *et al.* [96] also found the almost exclusive preference of ribose for its *β*-P form. In this paper Cocinero combined experimental and theoretical studies by characterizing the different molecular arrangements of ribose in gas-phase by using rotational **F**ourier **T**ransform - **M**icro**W**ave (FT-MW) spectroscopy and the DFT MP2/6-311++G(*d*, *p*), B3LYP//6-311++G(*d*, *p*), and M06-2X//6-311++G(*d*, *p*) levels of theory.

The differences in free energies between the 5-MR: 2dRibf, Ribf and the 6-MR: 2dRib and Rib ($\Delta G°_{\text{5MR-6MR}}$) are presented in **Table 3.1** and **Figure 3.4**.

All the energies were calculated as:

$$\Delta G°_{\text{5MR}-\text{6MR}} = G°_{\text{5MR}} - G°_{\text{6MR}}$$ (3.7)

A glance at the heatmap matrices represented in **Figure 3.4** shows that overall *the P-form is estimated to be more energetically favored ($\Delta G° > 0$) than the F-form in either vacuum or continuum solvation*.

The higher differences are obtained for the 2dRibf in vacuum with $G°$ (*β*-South or *β*-2'endo) - $G°$ ($^1C_4$-*α*) = 22.82 kJ/mol and for $G°$ (*β*-South or *β*-2'endo) - $G°$ ($^1C_4$-*β*) = 20.73 kJ/mol.

If we average each $\Delta G°$ *for each* column in the heatmaps we obtain the mean $\Delta G°$ in kJ/mol for each anomeric form of each F-form conformation *α*-$^2T_3$, *β*-$^2T_3$, *α*-$^3T_2$ and *β*-$^3T_2$. These energies are the following: 2dRibf in vacuum (-1.0, 18.6, 8.5 and 16.5 kJ/mol), 2dRibf in water (2.7, 11.6, 9.4 and 9.2 kJ/mol), Ribf in vacuum (6.29, 9.33, 6.25 and 13.3 kJ/mol) and Ribf in water (3.8, 17.0,

Table 3.1 Differences in kJ/mol for the total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies (ΔG°) at NPT conditions between the different furanose and pyranose ring conformations of 2'-deoxyribose and ribose in vacuum (numbers in black) and in implicit solvation using the IEFPCM model (numbers in dark blue).

		Ring conf.	DNA 2dRibf $\alpha\text{-}^2T_3$	$\beta\text{-}^2T_3$	$\alpha\text{-}^3T_2$	$\beta\text{-}^3T_2$
ΔE		$\alpha\text{-}^1C_4$	-1.7, **2.3**	20.8, **18.1**	10.2, **14.4**	18.7, **16.9**
$\Delta E_{(ZPE)}$			-1.7, **2.6**	22.4, **19.6**	11.6, **16.2**	20.2, **18.5**
ΔG°			-3.1, **0.6**	16.5, **13.7**	6.4, **9.8**	14.4, **12.7**
ΔE		$\beta\text{-}^1C_4$	3.4, **2.5**	25.9, **18.2**	15.3, **14.6**	23.8, **17.1**
$\Delta E_{(ZPE)}$	**2dRib**		4.0, **3.1**	28.0, **20.1**	17.3, **16.7**	25.9, **19.0**
ΔG°			1.2, **0.0**	20.7, **13.2**	10.7, **9.3**	18.6, **12.1**
ΔE		$\alpha\text{-}^4C_1$	6.9, **3.8**	29.4, **19.5**	18.8, **15.9**	27.3, **18.3**
$\Delta E_{(ZPE)}$			8.5, **5.6**	32.5, 22.6	21.8, **19.1**	30.4, **21.5**
ΔG°			3.3, **-0.4**	22.8, 12.8	12.8, **8.9**	20.7, **11.8**
ΔE		$\beta\text{-}^4C_1$	-4.1, **-0.6**	18.4, **15.2**	7.8, **11.5**	16.4, **14.0**
$\Delta E_{(ZPE)}$			-4.4, **-0.4**	19.7, **16.6**	8.9, **13.2**	17.5, **15.5**
ΔG°			-5.3, **-2.5**	14.2, **10.7**	4.1, **6.7**	12.1, **9.6**
			RNA Ribf			
ΔE		$\alpha\text{-}^1C_4$	8.5, **12.5**	13.3, **26.2**	10.4, **12.9**	19.5, **22.4**
$\Delta E_{(ZPE)}$			9.5, **13.9**	15.6, **28.1**	12.5, **14.1**	22.4, **24.7**
ΔG°			5.8, **8.7**	8.8, **21.8**	5.7, **9.0**	12.8, **17.0**
ΔE		$\beta\text{-}^1C_4$	9.9, **9.4**	14.7, **23.1**	11.9, **9.8**	20.9, **19.3**
$\Delta E_{(ZPE)}$	**Rib**		10.8, **10.5**	16.9, **24.7**	13.8, **10.7**	23.8, **21.3**
ΔG°			7.0, **5.7**	10.1, **18.8**	7.0, **6.0**	14.0, **14.0**
ΔE		$\alpha\text{-}^4C_1$	7.9, **6.5**	12.7, **20.3**	9.8, **6.9**	18.9, **16.4**
$\Delta E_{(ZPE)}$			8.9, **8.0**	14.9, **22.2**	11.9, **8.2**	21.8, **18.8**
ΔG°			4.9, **2.5**	7.9, **15.6**	4.8, **2.8**	11.9, **10.8**
ΔE		$\beta\text{-}^4C_1$	9.1, **0.3**	14.0, **14.1**	11.1, **0.7**	20.1, **10.2**
$\Delta E_{(ZPE)}$			9.1, **0.6**	15.2, **14.8**	12.2, **0.8**	22.1, **11.4**
ΔG°			7.5, **-1.6**	10.6, **11.5**	7.5, **-1.3**	14.6, **6.7**
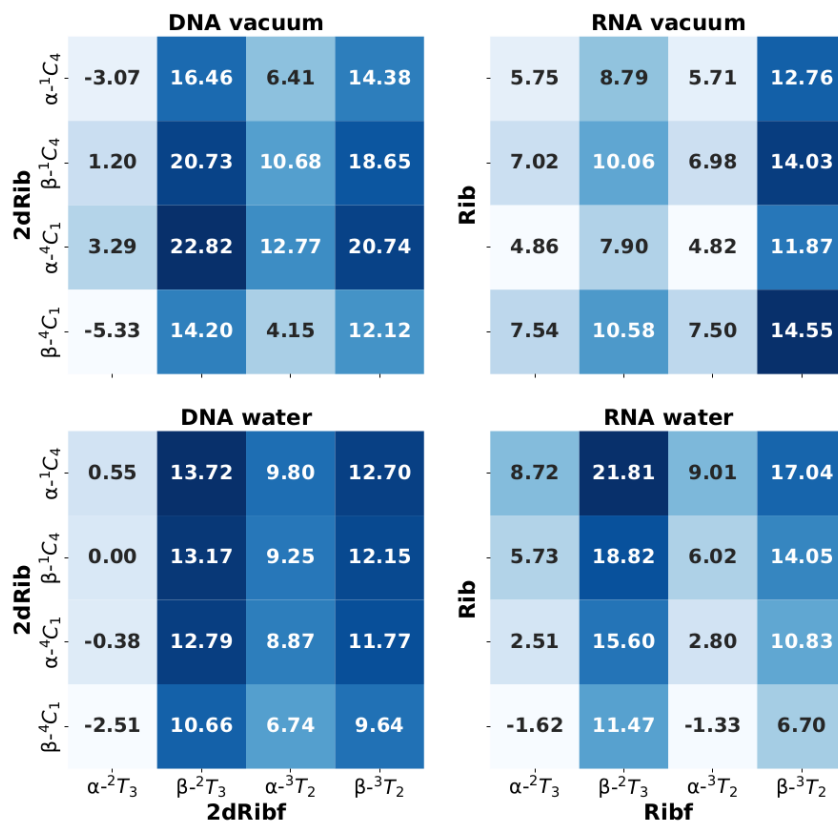
Figure 3.4 Changes in the Gibbs energies (ΔG°) at STP conditions in (kJ/mol). ΔG° was estimated at the DFT-B3LYP/6-311++G(d,p) level of calculation. $\Delta G^{\circ} = G^{\circ}$ (each 5-MR conformation) - G° (each 6-MR conformation). The 5-MR conformations are α-$^2T_3$, β-$^2T_3$, α-$^3T_2$ and β-$^3T_2$. The 6-MR conformations are α-$^1C_4$, β-$^1C_4$, α-$^4C_1$ and β-$^4C_1$. 2dRib: 2'-deoxyribopyranose, 2dRibf: 2'-deoxyribofuranose, Rib: ribopyranose and Ribf: ribofuranose.

4.1 and 12.2 kJ/mol). The major average differences disfavoring the 5-MR are consistently for the β-$^2T_3$.

The mean ΔG° in kJ/mol for the corresponding P-form ring conformations α-$^1C_4$, β-$^1C_4$, α-$^4C_1$ and β-$^4C_1$ are obtained by averaging the ΔG° of each row in the heatmaps: 2dRib in vacuum (8.5, 12.8, 14.9 and 6.3 kJ/mol), 2dRib in water (9.2, 8.6, 8.3 and 6.1 kJ/mol), Rib in vacuum (8.5, 9.5, 7.4 and 10.0 kJ/mol) and finally, Rib in water (14.1, 11.2, 7.9 and 3.8 kJ/mol).

Overall, the P-form seems to be more thermodynamically favored than the F-form and these results are in certain agreement with Das *et al*., 2021 [91], Azofra *et al*., 2014 [95] and Cocinero *et al*., 2011 [96].

Nevertheless, the energetic differences are within the range of the intrinsic error for the theoretical calculations (8-16 kJ/mol), which can be translated into the possibility for both sugar rings to have coexisted in the two α- and β- conformations. Hence, the reasons on why Nature chose one sugar ring over the other may lay beyond an analysis of the thermodynamic stability of the hemiacetal ring in either configuration based on the level of DFT theory used in this study.

In a paper from 1993 titled "Why pentose-and not hexose-nucleic acids?. Part VII. pyranosyl-RNA ('(*p*)-RNA')" [97] Pitsch and coworkers (Pcw) explored the reasons why the ribose 5-MR may have been preferred by Nature over the 6-MR to be part of today's NAs by analyzing the thermodynamic stability of a double stranded antiparallel octamer containing A:U as complementary bases and a D-ribopyranose-C4'-C2'-phosphate backbone and found that the stability of this (*p*)-RNA was 2.9 Kcal/mol more stable than its furanosyl analog. This thermodynamic property may have put (*p*)-RNA in disadvantage to participate in replication-translation processes and be able to carry the genetic information.

The results by Pcw do not exclude the possibility of existence of a proto-nucleic acid containing a ribopyranosyl-phosphate backbone that eventually through evolution selection was replace by today's sugar-phosphate backbone in DNA and RNA.

3.4.2 Pseudo rotational equilibrium and anomer-exchange reactions for nucleosides containing 5-MR and 6-MR sugars

The ring conformation of 5-MR furanose sugars impacts the biological properties and functions of glycoconjugates, e.g., cellular recognition.

Thibaudeau and coworkers (Tcw) (see Chapter #2 of [50] and [54, 55]) reported that the control of the pseudorotational equilibrium of sugars may be considered as an important factor in comparing the stability of β- and α-anomers of canonical ribofuranosyl-nucleosides.

Hence, we have decided to include this effect in our computational analysis for the thermodynamical stability of both anomers.

The pseudorotation concept refers to the transition between the ring conformations or puckering forms for the cyclopentane [98]. Pseudorotation allows the cyclopentane ring to stabilize from its less favored planar form with a steric energy of 22 kJ/mol (Chapter #2 of [50]).

Additionally, the furanose form exhibits more internal flexibility than the P-form. For instance, the activation energies between the two chair conformations of cyclohexane is ≈ 42 kJ/mol (see Chapter #2 of [50] and [99]), meanwhile, the energetic barrier to interconvert the two preferential puckering conformations $^2T_3$ and $^3T_2$ of 5-MR saturated rings is lower (e.g., < 20 - 25 kJ/mol for purine nucleosides (Chapter #2 of [50])).

This means that differently from the hexopyranose ring furanoses can coexist in a wild range of conformations. *This intrinsic flexibility may have given ribofuranose an evolutionary advantage over the ribopyranose ring*.

Wang and Woods (WW) in 2016 [100] proved the validity of the two-state model proposed by Altona and Sundaralingam [101] which states that the preferential ring conformations for furanose sugars are the $^2T_3$ and $^3T_2$. WW conducted 300 ns **M**olecular **D**ynamics (MD) simulations of C5'-methylated(Me) furanose sugars; β- and α-anomers of Me-C1'-D-ribofuranose, Me-C1'-D-arabinofuranose, Me-C1'-D-lyxofuranose, Me-C1'-D-xylofuranose and the C2'-deoxy-β-D-ribofuranose and obtained the **N**uclear **M**agnetic **R**esonance (NMR) *J*-coupling constants from quantum HF/6-31G(d) and B3LYP/cc-pVTZ DFT calculations on the conformations from the simulation. For most of pentofuranoses, local minima were found for the $^2T_3$ and $^3T_2$ conformations. The preferable sugar ring conformations were also analyzed by plotting the distribution of the ring puckering against the ring conformations across the simulations. An exception to the two-state model was found for the α-D-arabinofuranose in which there is the possibility of multiple conformers through the ^0E puckering conformation.

Another computational study conducted by Szczepaniak and Moc (ZM) on ribose and 2'-deoxyribose was published in 2014 [102]. In this investigation the authors explored the preferential sugar ring puckering of the β- and α-anomers of the cyclic P- and F-forms of both sugars and also the conformations of their open chains in vacuum. They combined **M**olecular **M**echanics (MM) simulations to generate and minimize many conformers and then re-optimized at the QM MP2/6-311++G(d, p) and M06-2X/6-311++G(d, p) levels the ones with relative energies ≤ 62 kJ/mol. Finally, the free energies of Gibbs (ΔG°) in kJ/mol at 298 °K were estimated.

ZM found for the D-ribose that from 38 structures sorted in decreasing order of their $\Delta G°$ obtained from MM, the 7 more stable structures at the QM level corresponded to 4 α- and 3 β-isomers of the ${}^1C_4$ and ${}^4C_1$ pyranose ring conformations. The local minima corresponded to the β-${}^1C_4$ pyranose conformer. The 16th conformer with an energy difference of 11.53 kJ/ mol at M06-2X/6-311++G (*d*, *p*) and 8.74 kJ/mol at MP2/6- 311++G(d,p) corresponded to the α-${}^2T_1$ twist conformation of the ribofuranose. Meanwhile, from these structures the one with the highest energy difference (less stable) from the local minima corresponded to an open chain conformer with a difference of 30.13 and 20.23 kJ/mol respectively for the M06-2X/6-311++G(d,p) and MP2/6- 311++G(d,p) levels of calculation.

For the 2'-D-deoxyribose the local minima from the 37 top ranked conformations corresponded to an α-${}^4C_1$ pyranose conformation, followed by the β-${}^1C_4$ with only 5.6 kJ/mol difference. Again, similarly to the case of D-ribose the 7 top more stable conformers corresponded to a mixture of either α- or β- anomers of ${}^4C_1$ and ${}^1C_4$ 6-MR conformations. The most stable furanose conformer corresponded to the twist α-${}^2T_1$ with 12.25 kJ/mol difference to the local minima. The most stable conformation from the open chain rotamers was ranked as 37 with an energy difference of 33.83 and 25.00 kJ/mol for M06-2X/6-311++G(d,p) and MP2/6- 311++G(d,p) respectively.

Another study by De Leeuw *et al*. (DLcw) [103] analyzed the conformational populations of the β-anomers of nucleosides containing D-ribofuranose, 2'-D-deoxyribofuranose and arabinofuranose. In total 174 crystal structures were analyzed and it was estimated that the bulk of conformations were found in the ${}^2T_3$ and ${}^3T_2$ regions. For ribofuranose the number of conformers were almost equally distributed between both regions, meanwhile for the 2'-D-deoxyribofuranose there was observed a preference for the ${}^2T_3$ region with a proportion of 3:1 conformers between ${}^2T_3$ and ${}^3T_2$ and some conformations laying in the East region around the ${}^0E$ conformation.

Other studies [50, 54] have shown the preferential ${}^2T_3$ and ${}^3T_2$ conformations for the D-ribofuranose, 2'-D-deoxyribofuranose sugars present in either A-, B- and Z-DNAs and RNAs.

Based on the evidence from the literature the preferential ${}^2T_3$ and ${}^3T_2$ conformations for 5-MRs and the chair conformations ${}^4C_1$ and ${}^1C_4$ for the 6-MRs were set as initial ring puckering conformations to theoretically estimate the thermodynamics of the pseudorotational equilibrium and anomer-exchange reactions of the α- and β-configurations of furanose, pyranose TCs and their corresponding canonical and non-canonical nucleosides.

The differences in corrected Gibbs energies (ΔG°) at 298 °K were analyzed for TCs and associated nucleosides for all the reactions presented in **Figure 3.2**.

In total 12 reactions in vacuum and implicit solvation, respectively {4 pseudorotational equilibriums (two for furanoses and two pyranoses) and 8 anomer-exchange reactions (4 for each sugar ring)} were analyzed:

Pseudorotational equilibriums:

5-MR
$$\alpha\text{-}^3T_2 \;\xrightleftharpoons{\;\Delta G^{\circ}_1\;}\; \alpha\text{-}^2T_3 \qquad \Delta G^{\circ}_1 < 0 \; (^2T_3) \tag{3.8}$$
$$\beta\text{-}^3T_2 \;\xrightleftharpoons{\;\Delta G^{\circ}_3\;}\; \beta\text{-}^2T_3 \qquad \Delta G^{\circ}_3 > 0 \; (^3T_2) \tag{3.9}$$

6-MR
$$\alpha\text{-}^1C_4 \;\xrightleftharpoons{\;\Delta G^{\circ}_1\;}\; \alpha\text{-}^4C_1 \qquad \Delta G^{\circ}_1 < 0 \; (^4C_1) \tag{3.10}$$
$$\beta\text{-}^1C_4 \;\xrightleftharpoons{\;\Delta G^{\circ}_3\;}\; \beta\text{-}^4C_1 \qquad \Delta G^{\circ}_3 > 0 \; (^1C_4) \tag{3.11}$$

Anomer-exchange reactions:

5-MR
$$\alpha\text{-}^2T_3 \;\xrightleftharpoons{\;\Delta G^{\circ}_2\;}\; \beta\text{-}^2T_3 \tag{3.12}$$
$$\alpha\text{-}^3T_2 \;\xrightleftharpoons{\;\Delta G^{\circ}_4\;}\; \beta\text{-}^3T_2 \tag{3.13}$$
$$\alpha\text{-}^3T_2 \;\xrightleftharpoons{\;\Delta G^{\circ}_5\;}\; \beta\text{-}^2T_3 \tag{3.14}$$
$$\beta\text{-}^3T_2 \;\xrightleftharpoons{\;\Delta G^{\circ}_6\;}\; \alpha\text{-}^2T_3 \tag{3.15}$$

$$\Delta G^{\circ}_2 < 0 \; (\beta), \quad \Delta G^{\circ}_2 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_4 < 0 \; (\beta), \quad \Delta G^{\circ}_4 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_5 < 0 \; (\beta), \quad \Delta G^{\circ}_5 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_6 < 0 \; (\alpha), \quad \Delta G^{\circ}_6 > 0 \; (\beta)$$

6-MR
$$\alpha\text{-}^4C_1 \;\xrightleftharpoons{\;\Delta G^{\circ}_2\;}\; \beta\text{-}^4C_1 \tag{3.16}$$
$$\alpha\text{-}^1C_4 \;\xrightleftharpoons{\;\Delta G^{\circ}_4\;}\; \beta\text{-}^1C_4 \tag{3.17}$$
$$\alpha\text{-}^1C_4 \;\xrightleftharpoons{\;\Delta G^{\circ}_5\;}\; \beta\text{-}^4C_1 \tag{3.18}$$
$$\beta\text{-}^1C_4 \;\xrightleftharpoons{\;\Delta G^{\circ}_6\;}\; \alpha\text{-}^4C_1 \tag{3.19}$$

$$\Delta G^{\circ}_2 < 0 \; (\beta), \quad \Delta G^{\circ}_2 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_4 < 0 \; (\beta), \quad \Delta G^{\circ}_4 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_5 < 0 \; (\beta), \quad \Delta G^{\circ}_5 > 0 \; (\alpha)$$
$$\Delta G^{\circ}_6 < 0 \; (\alpha), \quad \Delta G^{\circ}_6 > 0 \; (\beta)$$
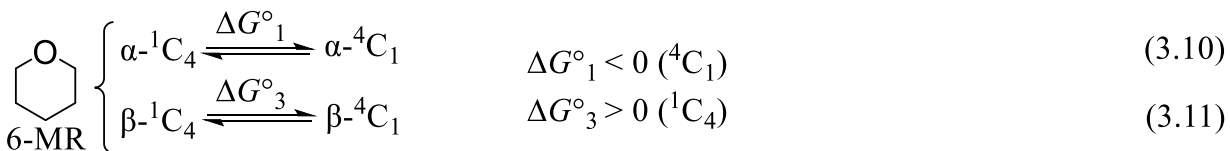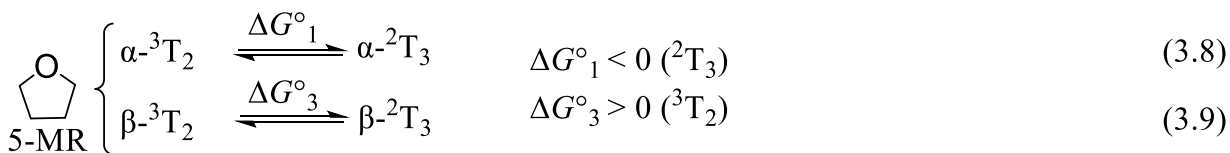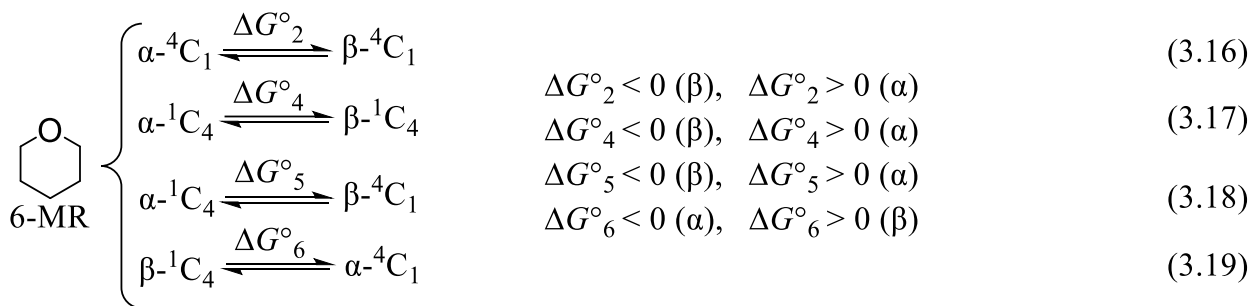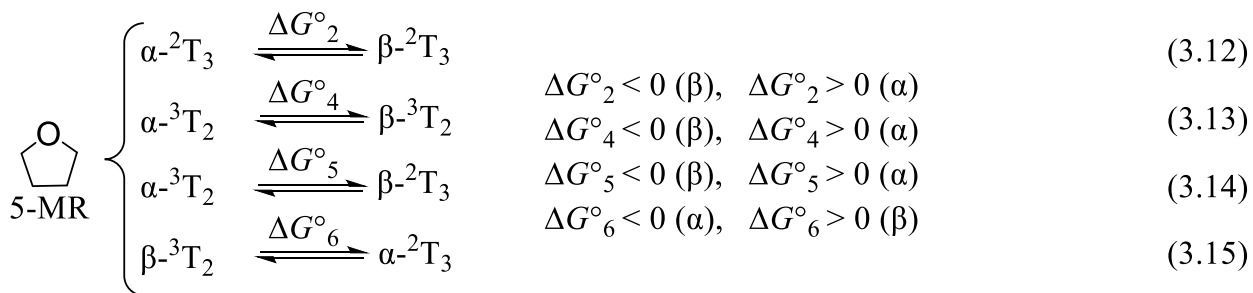
Schemes 3.8-3.19, **Figures 3.5-3.10** and **Tables A3-A4** represent the chemical transformations, the values in kJ/mol for the different ΔG° and the preferential ring puckering and anomer depending on the sign of ΔG°.

Pseudorotational equilibrium and anomer-exchange reactions for sugar-like TCs

Figure 3.5 and **Table A3** shows the ΔG_1° and ΔG_3° for the pseudorotational equilibrium of the α- and β-anomers of each TC with a furanose or pyranose ring. Overall, for the furanoses 2dRibf, Ribf and Tho either in vacuum or water the $^2T_3$ conformation is more favored for the α-, meanwhile the $^3T_2$ is more favored for the β-configuration. The major differences observed are for the Tho in which $\Delta G_3^{\circ} = 34.5$ kJ/mol. The rest of Gibbs energies differences are in the intrinsic error for the method. For the α-Ribf there is not preference between $^3T_2$ and $^2T_3$.

For the pyranose 2dRib in vacuum $^4C_1$ is favored for the α-, meanwhile $^1C_4$ is more favored for the β-anomer. For the rest of the 2dRib and Rib TCs there are not significant differences between the ring puckering for the different anomers. Overall, the energetic differences between the different ring puckering conformations decrease when going from vacuum to implicit solvation, with the exception of Tho. These results and the ones presented in section 3.4.1 are somehow in agreement with SC results [102].

The difference in stabilities between the β- and α-anomers from the different anomer-exchange reactions can be analyzed from the ΔG_2°, ΔG_4°, ΔG_5° and ΔG_6° signs and magnitudes (equations 3.2, 3.4, 3.5 and 3.6). In most of cases the magnitude of these ΔG° is in the intrinsic error for B3LYP/6-311++G (d, p) ($\approx$ 17 kJ/mol). The only exceptions are the ΔG_2° (19.5 kJ/mol) and ΔG_6° (17.4 kJ/mol) favoring the α- and β-anomers respectively for the 2dRibf in vacuum and for Tho in aqueous solution the ΔG_2° (46.4 kJ/mol) and ΔG_5° (38.0 kJ/mol) also favoring the α-anomer.
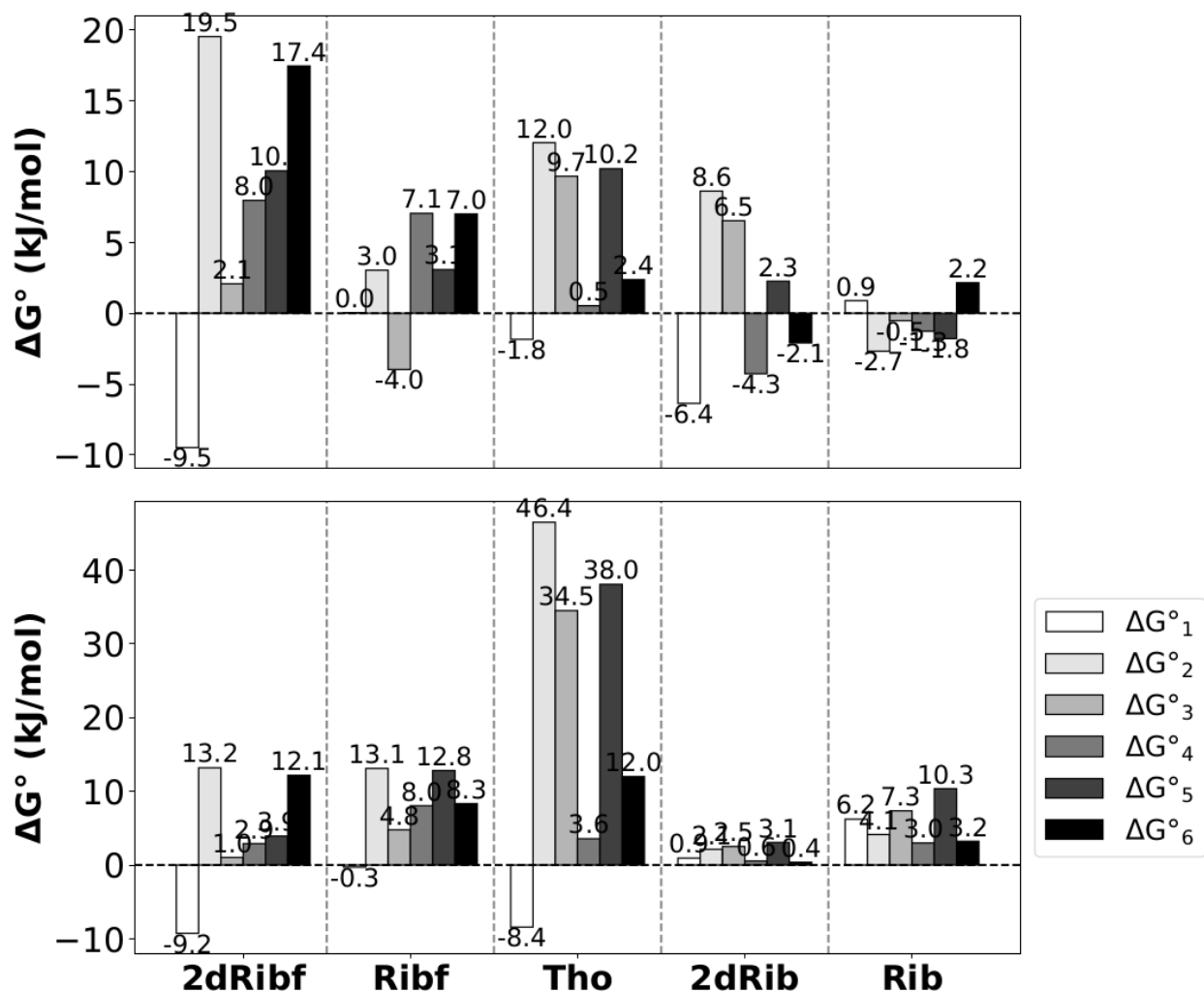
Figure 3.5 Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^{3}T_2$ and $^{2}T_3$) of furanose rings of 2dRibf: 2'-deoxyribofuranose, Ribf: ribofuranose, Tho: threose and the conformations ($^{1}C_4$ and $^{4}C_1$) of pyranose rings 2dRib: 2'-deoxyribopyranose and Rib: ribopyranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-31G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-31G(d, p) in aqueous medium using the IEFPCM solvation model.

Pseudorotational equilibrium and anomer-exchange reactions of canonical and non-canonical nucleosides

Figure 3.6 and **Table A4** represents the ΔG_1° and ΔG_3° (equations 3.1 and 3.3) for the 2dRibf canonical and non-canonical nucleosides. A glance at this graph shows that for the canonical nucleosides there are small differences that favours the $^3T_2$ conformation in both α- and β- when having A in vacuum and water. For the G, C, T, U, BA-C^5 in vacuum and water the equilibrium for the β- anomer favours more the $^3T_2$ conformation, meanwhile there is not a preferable sugar ring conformation for the α-anomer. A similar trend is observed for the non-canonical nucleosides in vacuum with the exception of TAP-C^5 in which the equilibrium seems to favor the $^3T_2$ for the α- and the $^2T_3$ for the β-anomer, respectively. There is no difference between either conformation in the corresponding anomers for the non-canonical RUs in aqueous solution or in the case of the T and U nucleosides.

The bars in **Figure 3.6** corresponding to the anomer-exchange reactions (ΔG_2°, ΔG_4°, ΔG_5° and ΔG_6°) show that overall there are not differences between the α- and β-anomers of the nucleosides with 2dRibf either in vacuum or aqueous solution beyond the intrinsic error of B3LYP/6-311++G (d, p) ($\approx$ 17 kJ/mol). Exceptions are observed for A containing nucleosides where the α-$^3T_2$ configuration is favored in both vacuum and implicit solvation for reaction #5 ($\Delta G_5^\circ \approx 20$ kJ/mol). Similarly, the α-$^2T_3$ and the α-$^3T_2$ are favored for cytidine and MM nucleosides in both environments for the reactions # 2 and #4 ($\approx$ 20.2 kJ/mol) and reactions #2 and #5 ($\approx$ 21.6 kJ/mol) respectively.

In the case of Ribf nucleosides, **Figure 3.7** shows that for the nucleosides containing canonical bases in vacuum or aqueous solution the pseudorotational equilibrium there is no clear preference for one specific sugar ring conformation over another in the case of either anomer. These differences are all under 1.8 kcal/mol which is still in the error for the DFT calculation. For the non-canonical nucleosides in vacuum or implicit solvation there are also not marked differences that may favour one preferential conformation over the other.

No significant difference in the stabilities of the different anomers were predicted when examining the free energies of the anomer-exchange reactions. The only exception was found for
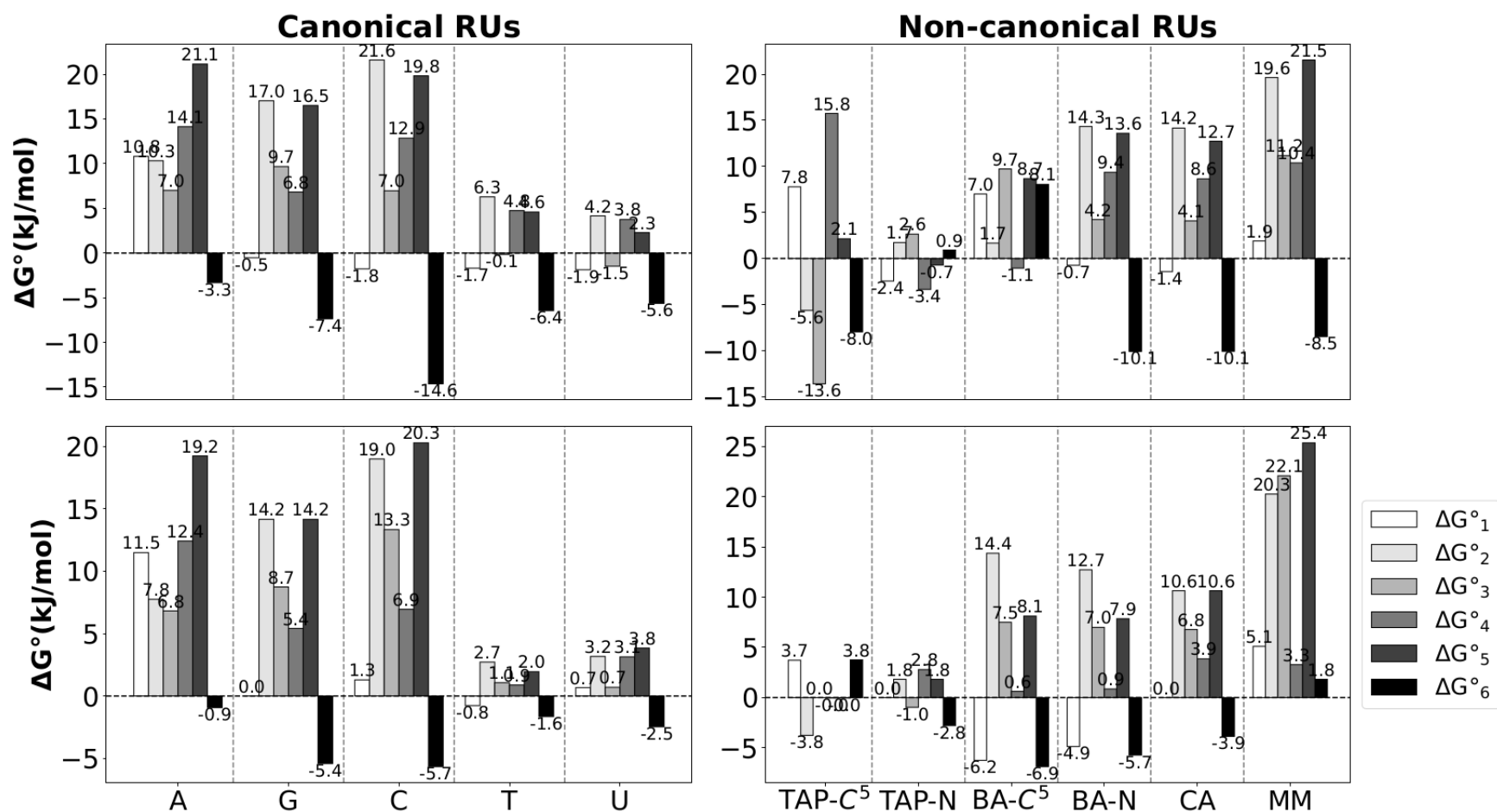
Figure 3.6 Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^2T_3$ and $^3T_2$) of the 5-MR in 2dRibf: 2'-deoxyribofuranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
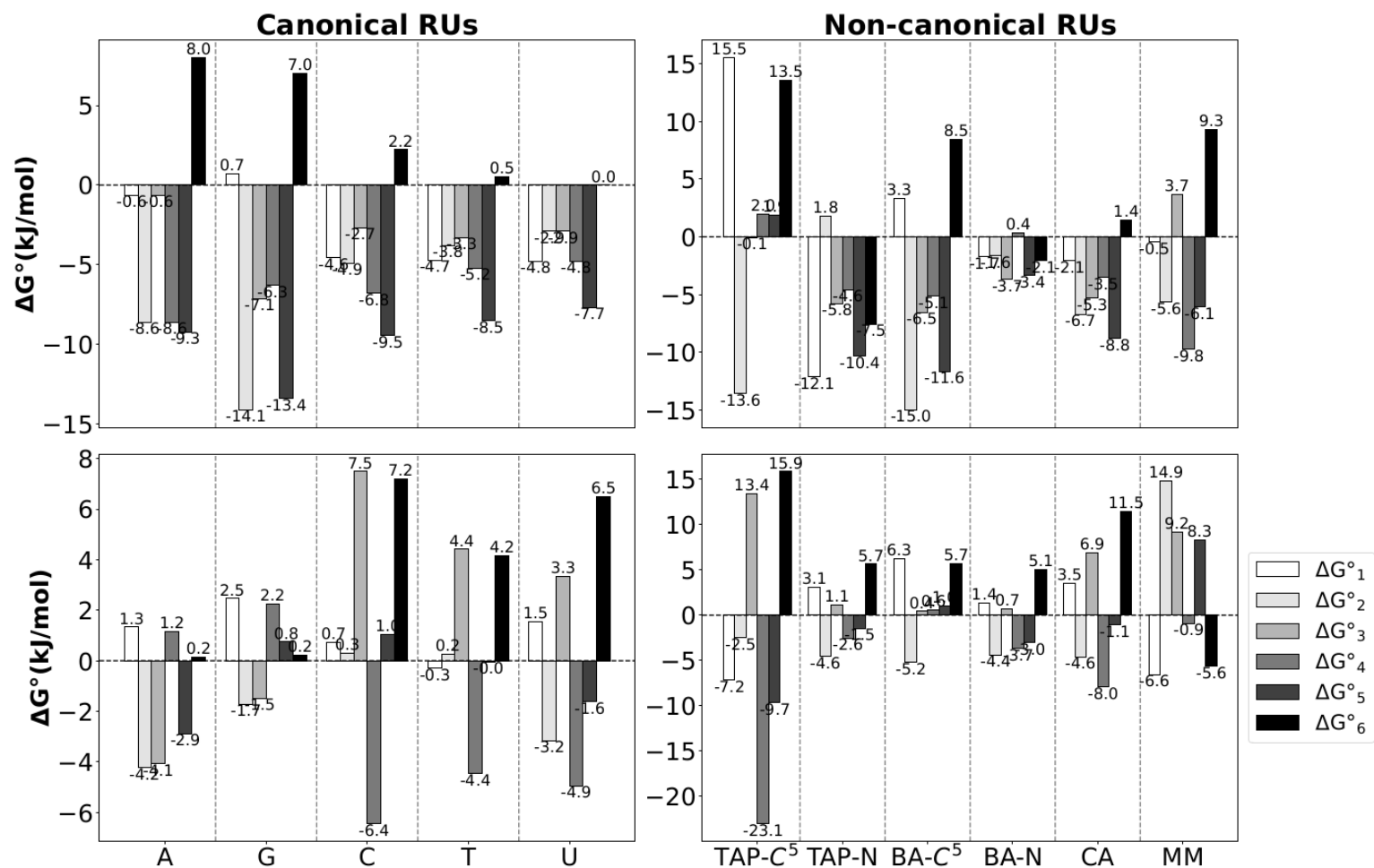
Figure 3.7 Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^{2}T_{3}$ and $^{3}T_{2}$) of the 5-MR in Ribf: ribofuranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

the TAP-C^5 in aqueous environment, in which reaction #4 favors the β-$^3T_2$ with a $\Delta G_4^\circ = -23.1$ KJ/mol.

An inspection of **Figure 3.8** corresponding to the Tho-related canonical and non-canonical nucleosides reveals that in general there are not significant differences in the ΔG_1° and ΔG_3° that favours one sugar ring conformation over the other for either anomer in vacuum or water with the exceptions of the α-anomer of A ($\Delta G_1^\circ = 27.5$ kJ/mol) and TAP-C^5 ($\Delta G_1^\circ = 19.2$ kJ/mol) nucleosides favoring in both cases the $^3T_2$ puckering.

An analysis of the signs and magnitude of ΔG_2°, ΔG_4°, ΔG_5° and ΔG_6° shows that the difference in stabilities between the α- and β-anomers for the different anomer-exchange reactions does not overcome the intrinsic error for B3LYP/6-311++G (d, p). The only exceptions were found for the reaction # 5 of canonical pyrimidine nucleosides in vacuum favoring α, for reactions #4 and #5 in the TAP-C^5 in implicit solvation and for reaction #5 of the MM nucleosides in vacuum. In all these cases the α-$^3T_2$ was more favored with differences in free energy $\approx$ 17 - 23 kJ/mol.

For the case of the 2dRib nucleosides (**Figure 3.9**) containing canonical bases it is observed that in vacuum the only cases in which the pseudorotational equilibrium shifts is for the β-anomer with a noticeable change in ΔG_3° for C, T and U containing nucleosides favouring the β-$^4C_1$.

For the canonical nucleosides in implicit solvation the T and U nucleosides favor the α-anomer in the $^1C_4$ conformation with ΔG_1° in the order of 22.8 kJ/mol and 20.5 kJ/mol respectively, meanwhile the β-anomer is favored in the $^4C_1$ conformation for the C, T and U nucleosides with values of ΔG_3° of 18.5, 19.0 and 18.6 kJ/mol.

For the pseudorotation equilibrium of the non-canonical nucleosides in vacuum it is observed that for the TAP-C^5 nucleosides the differences in ΔG° are higher favoring the α-$^1C_4$ ($\Delta G_1^\circ = 25.1$ kJ/mol) and the β-$^4C_1$ ($\Delta G_3^\circ = $ -23.3 kJ/mol). For the case of the β-anomers of BA-C^5 the β-$^4C_1$ is favored with $\Delta G_3^\circ = $ -26.5 kJ/mol. When implicit solvation is included for the case of the ΔG_1° in the TAP-C^5, TAP-N, BA-C^5, CA, and MM nucleosides the α-$^1C_4$ is favored ($\approx$ 17 - 34 kJ/mol), meanwhile for the β-anomer of TAP-C^5 the $^1C_4$ puckering is preferred (-29.5 kJ/mol).

The values of the delta free energies for the different anomer-exchange reactions show that in general there are not differences between the different 2dRib nucleosides' configurations.

Exceptions were found for the T and U nucleosides in both environments for reaction #4 in which α-$^1C_4$ was preferred with differences $\approx$ 17 - 21 kJ/mol. In the case of the non-canonical
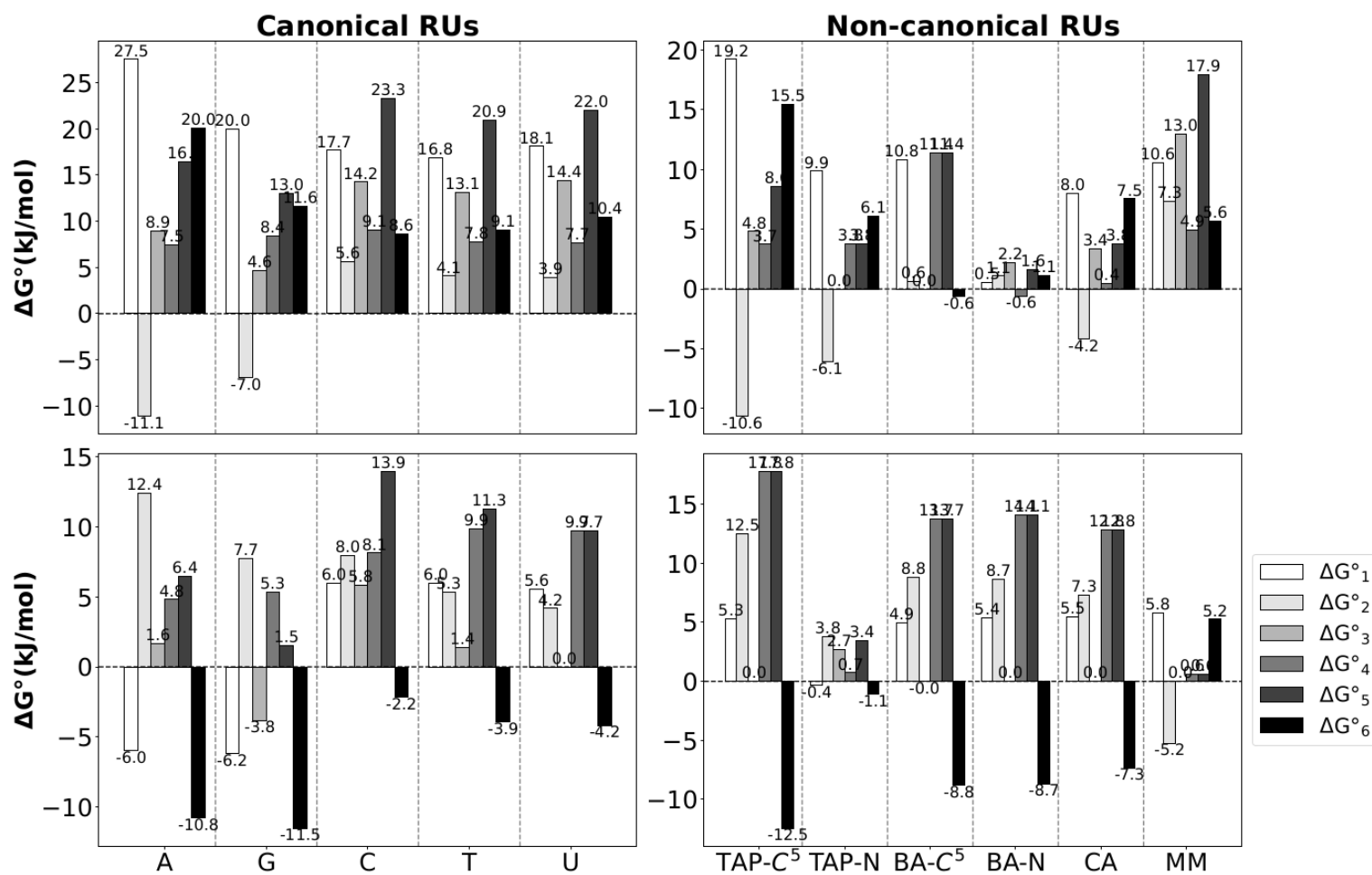
Figure 3.8   Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^2T_3$ and $^3T_2$) of the 5-MR in Tho: threofuranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
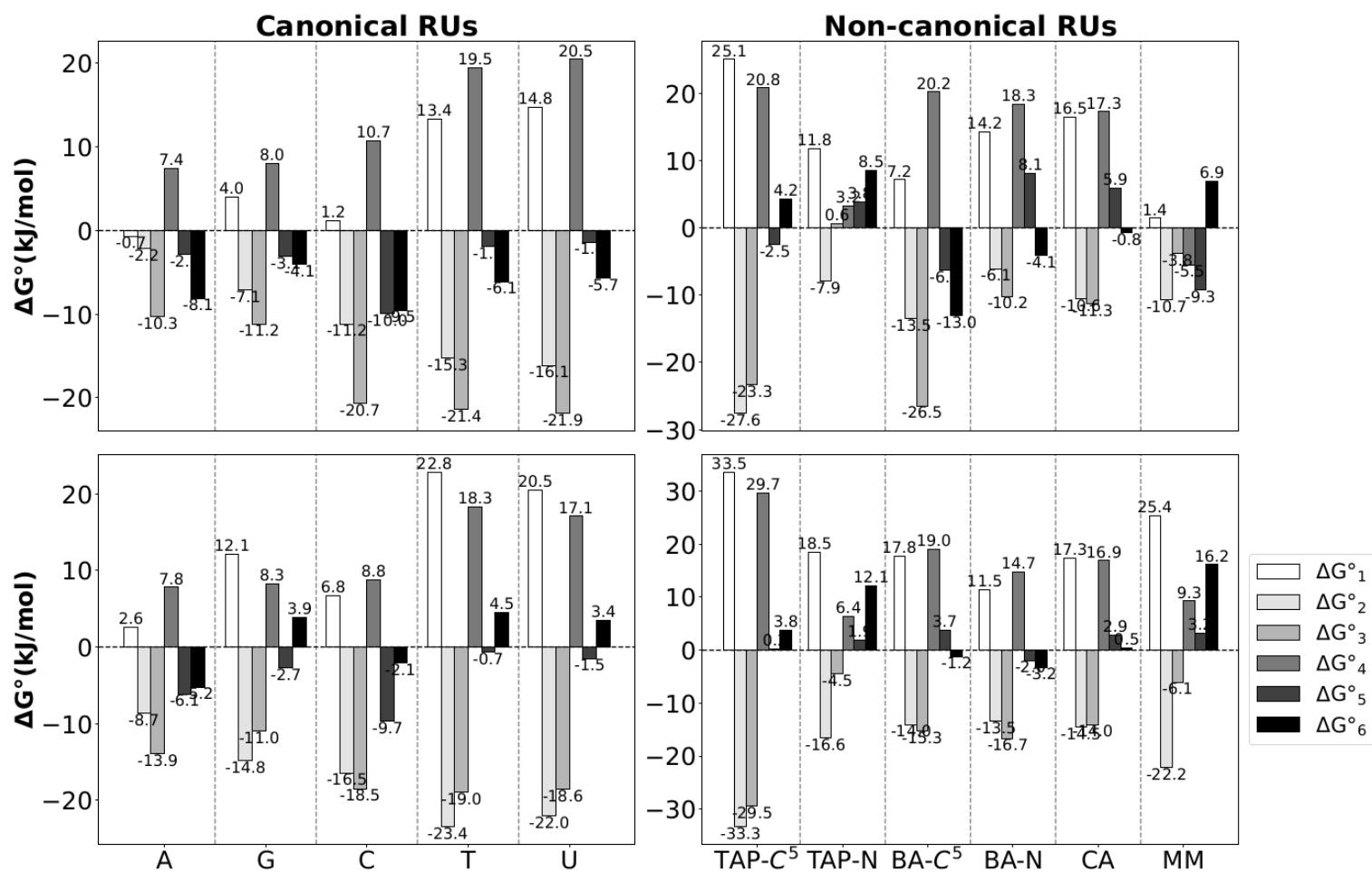
Figure 3.9 Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^2T_3$ and $^3T_2$) of the 6-MR in 2dRib: 2'-deoxyribopyranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

2dRib nucleosides the TAP-C^5 nucleosides were more favored in the β-$^4C_1$ and α-$^1C_4$ anomer-puckering for reactions #2 (-27.6 kJ/mol in vacuum and -33.3 kJ/mol in the IEFPCM model), #4 (20.8 kJ/mol in vacuum and 29.7 kJ/mol in aqueous environment) respectively. The BA-C^5 nucleoside was favored in the α-$^1C_4$ for reaction #4 in both environments ($\approx$ 19.5 kJ/mol). Similarly, BA-N and CA nucleosides were also favored in the α-$^1C_4$ for reaction #4 in vacuum (18.3 and 17.3 kJ/mol respectively).

For the case of the Rib nucleosides (**Figure 3.10**) in both vacuum and solvent the pseudorotational equilibrium prefers for the β-anomer the $^4C_1$ conformation and for the α-anomer the $^1C_4$ conformation. Most of these energies are in the intrinsic error of the DFT method used with the exceptions of the β-anomer of the 5 canonical bases in vacuum and for C, T and U in both environments, TAP-C^5 in vacuum and water, BA-C^5 in vacuum, BA-N in water and the α-anomer of T and U in vacuum and water, C in water, TAP-C^5 and MM in both environments.

An analysis of the different anomer-exchange reactions revealed that for all canonical nucleosides reaction #2 favors the β-$^4C_1$ 6-MR puckering in both environments ($\approx$ -18 - -35 kJ/mol). Opposite to this, reaction #4 favors the α-$^1C_4$ for the pyrimidine nucleosides in both environments ($\approx$ 20 - 25 kJ/mol). In the case of the non-canonical Rib nucleosides reaction #2 favors again the β-$^4C_1$ in both environments for TAP-C^5, BA-C^5, CA and MM nucleosides (between $\approx$ -18 and -38 kJ/mol) and reaction #4 favors α-$^1C_4$ for BA-C^5 nucleosides in vacuum (19.5 kJ/mol). The rest of the energetic differences are negligible.

Overall it can be noticed that most of the differences in free energies estimated from this study for the equilibrium of pseudorotation of the different nucleosides does not obey in general the same patterns predicted by Thibaudeau *et al*. (see Chapter #2 of [50] and [54, 55]) with the exception of the 2dRibf nucleosides. These authors used in their study as a TC D-2', 3'-dideoxynucleosides to prevent hydrogen bonds between hydroxyl groups in the sugar ring that would change the conformation of the sugar ring.

In our study we reproduce to a certain extent the results obtained from the previous study conducted by Castanedo *et al*. [57] in which no difference between α- and β-anomers of the canonical nucleosides was observed using lower DFT- B3LYP/6-31G (*d*, *p*) level of theory.
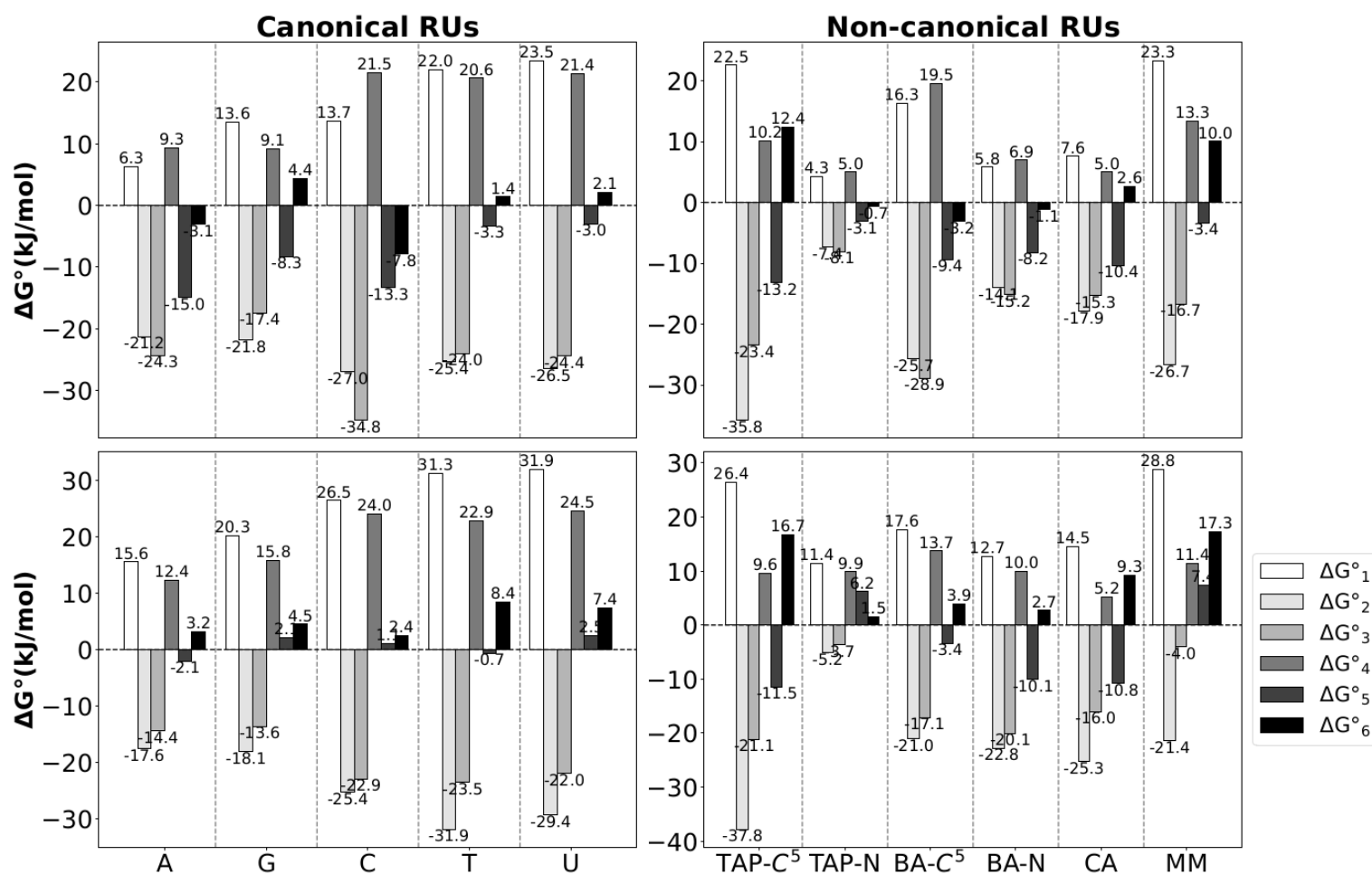
Figure 3.10 Comparison of Gibbs energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions between the conformations ($^{2}T_{3}$ and $^{3}T_{2}$) of the 6-MR in Rib: ribopyranose. The Gibbs energies of these reactions are defined by equations 3.1-3.6. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

Even when in this study the different nucleosides started from predefined sugar ring conformations the addition of a RU and the presence of hydroxyl groups may create intramolecular interactions that can modify the final sugar ring conformations after DFT geometric optimization eliminating any differences in the pseudorotational equilibrium. Still, the fact that an initial sugar ring conformation was imposed it may make the pseudorotational equilibrium to shift in some cases as have been described before.

When analyzing the averages of the ΔG_1° and ΔG_3° (3.2 and 3.1 kJ/mol for F-forms and 15.8 and -16.8 kJ/mol for P-forms) for furanose and pyranose nucleosides it can be noticed that in general these thermodynamic parameters have higher differences for the case of the P-form nucleosides in comparison to the F-form counterparts. This is suggestive of the concept of "flexibility" presented by Thibaudeau *et al*. (see Chapter #2 of [50] and [54, 55]) in the context of the difference of thermodynamic parameters for the equilibrium of pseudorotation at different pH values. These authors proposed that the *β*-anomer of nucleosides could had an evolutionary advantage with respect to its *α*-counterpart due to its flexibility or capacity to shift the pseudorotational equilibrium when environmental conditions. e.g., pH. In our case we refer to this concept to provide a justification on why Nature chose a F-form and not a P-form for the 2dRibf and Ribf of today's NAs. *We propose that a propensity by the 5-MR of r2dRibf and Ribf to populate either sugar ring conformation based on the free energies for the pseudorotational equilibrium may have posed an advantage to this sugar ring over the pyranose counterpart*.

When analyzing the differences in stabilities between the anomer-exchange reactions between the different anomers in can be noticed that even when overall the difference in stabilities between both anomers can be considered in the error of B3LYP/6-311++G (*d*, *p*) some reactions favor one specific anomer, e.g., *α*-anomer for some nucleosides with the furanoses 2dRibf, Ribf and Tho and the *β*-anomers for some nucleosides containing the pyranoses 2dRib and Rib. Nevertheless, the results presented in this section are in certain agreement with the results presented by Castanedo and Matta (CM) in 2022 [57] in which not considerable energetic differences were predicted between the anomers of canonical nucleosides in either vacuum or implicit solvation.

3.4.3 Thermodynamics of the synthesis of canonical and non-canonical nucleosides through a classic condensation reaction of its components

Beyond analyzing the difference in stabilities between the different anomers of canonical and non-canonical nucleosides it is important to assess the plausibility of their respective prebiotic synthesis. In this way it may be possible, hopefully to justify based on theoretical calculations if a thermodynamic can justify the emergence of proto-nucleosides previous to the emergence of the building blocks of today's DNA and RNA.

In this direction, theoretical modeling of the "classic" synthesis (condensation reaction between an electrophile {TC} and a nucleophile {RU}) for the prebiotic emergence of canonical and non-canonical nucleosides seems to be a reasonable starting point.

It is well addressed in the literature that using the classic model implies facing the "nucleoside problem" as specific case of the "water problem" [6, 7, 5, 8]. The nucleoside problem refers to the inability of canonical bases to N-glycosylate with ribose in aqueous environment due to the thermodynamic instability of the glycosidic bond. This challenge was first found by the experimental works of Orgel and coworkers (Ocw) [4, 104]. Despite the challenges of the synthesis, Ocw obtained the adenosine in some yield (1-5%). Interestingly, these authors also obtained N-nucleosides with the adenine exocyclic NH_2 in 50-70% of yield [4].

In order to overcome the "nucleoside problem" different alternatives have been proposed that go from *e.g.*, using different synthetic routes to the searching for non-canonical components (non-canonical is referred to components non present in the building blocks from today's NAs). In this section we explore the latest strategy.

We have filtered the literature for TCs that can create non-conventional xeno-NAs backbones, e.g., threose, glycerol, glyceric acid, AEG and RUs that have been proven to be successful in circumventing the "glycoside problem". These RUs are 2, 4, 6 triaminopyrimidine (TAP), barbituric acid (BA) and melamine (MM).

Additionally the base cyanuric acid (CA), is also tested since its nucleotide can create hydrogels (stacking hexads) through complementary base pairing with BA nucleotides [13] and with adenosine 5'-monophosphate (AMP) and 2, 6-diaminopurine (DAP) [18] similar to how the canonical nucleobases create hydrogen bonds in the DNA double helix.

TAP is a pyrimidine nucleobase that contains three exocyclic amino groups. This base

exhibits 4 nucleophilic sites: the three exocyclic NH_2 groups in positions 2, 4, 6 and the endocyclic C5. The amino groups in position 4 and 6 are equivalent by symmetry and exhibit a better nucleophilic character than the one in position 2 because of its depleted negative charge due to the electron withdrawing of the two ortho NH_2 in position 4 and 6. Additionally, the prebiotic synthesis of these base has been previously proposed [105, 106, 107].

TAP can glycosylate with different TCs depending on the reaction conditions [16, 108]. This property is enhanced by its higher solubility in water compared to canonical bases [19].

For example Chen and coworkers [19] were the first to test the glycosylation of TAP with ribose by heating the mixture in aqueous solution at 35-95 ºC and pH = 8. A complex mixture of products was obtained. An important result from this study is that after 10 days of reaction a combined yield of 60% was obtained for the mixture of TAP-ribose and 90% was obtained for the mixture at a higher temperature. Even without dehydration the reaction was successful in water at temperatures of 35-65 ºC and even 4 ºC. The major product ($\approx 20\%$ yield) was isolated through **H**igh **P**erformance **L**iquid **C**hromatography (HPLC) and analyzed using **M**ass **S**pectrometry (MS) and NMR and the β-ribofuranose-C^5-TAP nucleoside was detected as preferential product. The mixture also included other N-glycosylated nucleosides which were found through hydrolysis of the ribose-TAP mixture.

Studies by Cafferty and coworkers [12] reproduced these results obtaining at similar conditions (35 ºC and 10 days of reaction) a complex mixture of TAP-nucleosides that included pyranose and furanose rings for the ribose, β- and α-anomers and C^5- and N-nucleosides identified from EM-HPLC studies after basic hydrolysis in ammonium hydroxide for 4, 20 and 44h. This mixture represented a combined yield of 60%. The presence of N-nucleosides with exocyclic NH_2 of TAP is somehow related with the findings by Fuller *et al*. [4].

Even when N-nucleosides are predominant in today's DNA and RNA, C-nucleosides can also be found in Nature, e.g., pseudouridine which is a product from the post-translation of RNA. For example, rRNA from mammals presents around 100 pseudouridines per ribosome [109, 19, 110]. Most of these C-nucleosides have anticancer and antibiotic activity [109, 110].

Fialho expanded on the work by Cafferty *et al*. [12] by testing the glycosylation of TAP with 17 different sugars that included hexoses, pentoses and tetroses. The reactions were conducted for 24h at pH = 1 or 7 at 85 °C in a mixture TAP+sugar (1:1) ratio. The products were analyzed by UV-LC/MS and ^{1}H-NMR and it was proven that all glycosylation reactions proceeded to a

certain extent. The sugar that exhibited the higher yield of glycosylation was the pentose arabinose with 61% and 55% at pH = 1 and 7, respectively. Ribonucleosides were obtained in a 28% and 31% of yield at pH = 1 and 7 respectively. The hexose glucose reacted in 31 and 44% of yield at pH = 1 and 7 obtaining the β-C-nucleosides as the main product and threose reacted with a 3% and 2% of yield at pH = 1 and 7. Fialho argues that the low efficiency in the glycosylation of threose may be related to its high degradation rate at the experimental conditions [40, 16, 108].

Barbituric acid (BA) is a strong acid ($pK_a = 4$) that can react with ribose in aqueous solution due to the nucleophilic character of its barbiturate ion [16]. BA' reaction has been tested with glucosamine at pH = 7 in a (1:2) BA-glucosamine ratio, for 10 hrs and 50 ℃. The major C-β-nucleoside in basic catalysis has been produced [16, 111]. The C-glycoside of the enol form of BA can create stacked hexads with MM when dissolved in water [13].

Melamine (MM) contains a heterocyclic ring with three N that makes it more electron withdrawing, hence it is less nucleophilic than TAP and BA. Additionally, MM is less soluble in water at pH ≥ 5. Its solubility at 65 ℃ is 0.2M [16]. MM can co-precipitate in aqueous solution with phosphates, sulfates and carboxylates [16].

Cafferty and coworkers [13] showed that MM can glycosylate with ribose in aqueous solution when the reaction last for 24hrs at 20 ℃ despise the MM disadvantages described above. Both β- and α-anomers with the 5-MR and 6-MR of the ribose were detected for the MM ribosides by using EM-HPLC chromatography and measuring the ^{1}H-NMR coupling constants for the anomeric H.

Fialho [16] explored the reactivity of 80 reactions between 8 nucleophilic bases that included the canonical adenine, uracil, TAP, BA, CA and MM with a set of 10 electrophiles that included the aldehydes D-, L-glyceraldehyde, D-ribose, one anhydride, one imide, 3 esters, 2 thioesters and one Michael acceptor. These electrophiles were selected as they fulfilled two criteria to be considered good prebiotic electrophiles: i) resistant to hydrolysis or they can be obtained from reactants also hydrolytic resistant, ii) can be easily synthesized in prebiotic conditions. All nucleobases tested with the exception of uracil and CA were able to react with D-ribose and at least one product was detected through UV-**L**iquid **C**hromatography (LC)/MS.

Table A5 from the Appendices section shows the ΔG° reaction for the formation of all 506 nucleosides. $\Delta G^{\circ} < 0$ suggests that the condensation reaction may be thermodynamically favored meanwhile $\Delta G^{\circ} > 0$ indicates otherwise.

A glance at **Figure 3.11** suggests that for the 2dRibf canonical nucleosides most of the ΔG^0 are within the intrinsic error for the method, with some negative values. The lowest ΔG^0 value corresponds to the α-$^3T_2$ of adenosine in aqueous solution (ΔG° = -18.8 kJ/mol). The formation of the β-$^3T_2$ of adenosine is also favored even when the values of ΔG° = -9.4 kJ/mol. Similarly, the formation of some non-canonical nucleosides is not favored in either environment, although for the TAP-C^5, TAP-N and BA-C^5 the β-anomers were more favored than the α-counterpart with the lowest values for the delta free energies of the BA-C^5 in vacuum with ΔG° = -35.7 and -43.3 kJ/mol for the β-$^2T_3$ and β-$^3T_2$ respectively. The less favored reactions are predicted in both environments to be for the CA nucleosides ($\approx$ 8 - 18 kJ/mol) and the BA-N ($\approx$ 6 - 17 kJ/mol).

Figure 3.12 represent the values of ΔG° for the classic formation of the Ribf nucleosides. For the case of the canonical nucleosides in vacuum the formation of both β-$^2T_3$ and β-$^3T_2$ is estimated as more favorable than the α anomers. The cytidine has the more negative values with ΔG° of -16.5 and -17.8 kJ/mol for the $^2T_3$ and β-$^3T_2$ respectively. A similar picture is observed in the case of the canonical nucleosides in aqueous environment with the exceptions of the thymidine and uridine in which no reaction is favored. The nucleoside that is more favored in aqueous solution is the β-$^2T_3$ from adenosine with ΔG° = -18.6 kJ/mol.

In the case of the synthesis of the Ribf non-canonical nucleosides in vacuum even when the β-$^2T_3$ and β-$^3T_2$ are more favored to be obtained for the TAP-C^5 and the BA-C^5, in the first case the energies are almost negligible. All anomers in all sugar ring conformations are predicted to be synthesized for the BA-C^5 with values that overcome the error of the DFT method: (ΔG° = -32.7 kJ/mol for the α-$^2T_3$; ΔG^0 = -50.7 kJ/mol for the β-$^2T_3$; ΔG° = -36.7 kJ/mol for the α-$^3T_2$ and ΔG^0 = -48.1 kJ/mol for the β-$^3T_2$). The formation of the BA-N and CA is predicted as unfavorable with values beyond the DFT intrinsic error (BA-N: $\approx$11 - 18 kJ/mol and CA: $\approx$10 - 23 kJ/mol).

For the non-canonical Ribf nucleosides in implicit solvation a similar trend is observed. In this case the β-anomers of both furanose ring conformations are predicted to be thermodynamically favored while the α-anomers are not for the TAP-C^5, BA-C^5, BA-N and MM. For TAP-C^5 the lowest value is for β-$^3T_2$ with a ΔG° = -15.2 kJ/mol. The BA-C^5 nucleoside has the lowest ΔG° values across all non-canonical nucleosides in water for the β-anomers: (ΔG° = -26.5 kJ/mol for the β-$^2T_3$ and ΔG° = -22.1 kJ/mol for the β-$^3T_2$). Again, the formation of BA-N and CA nucleosides was estimated as non favored with ΔG° $\approx$ 2.6 - 26.8 kJ/mol. The formation of MM nucleosides
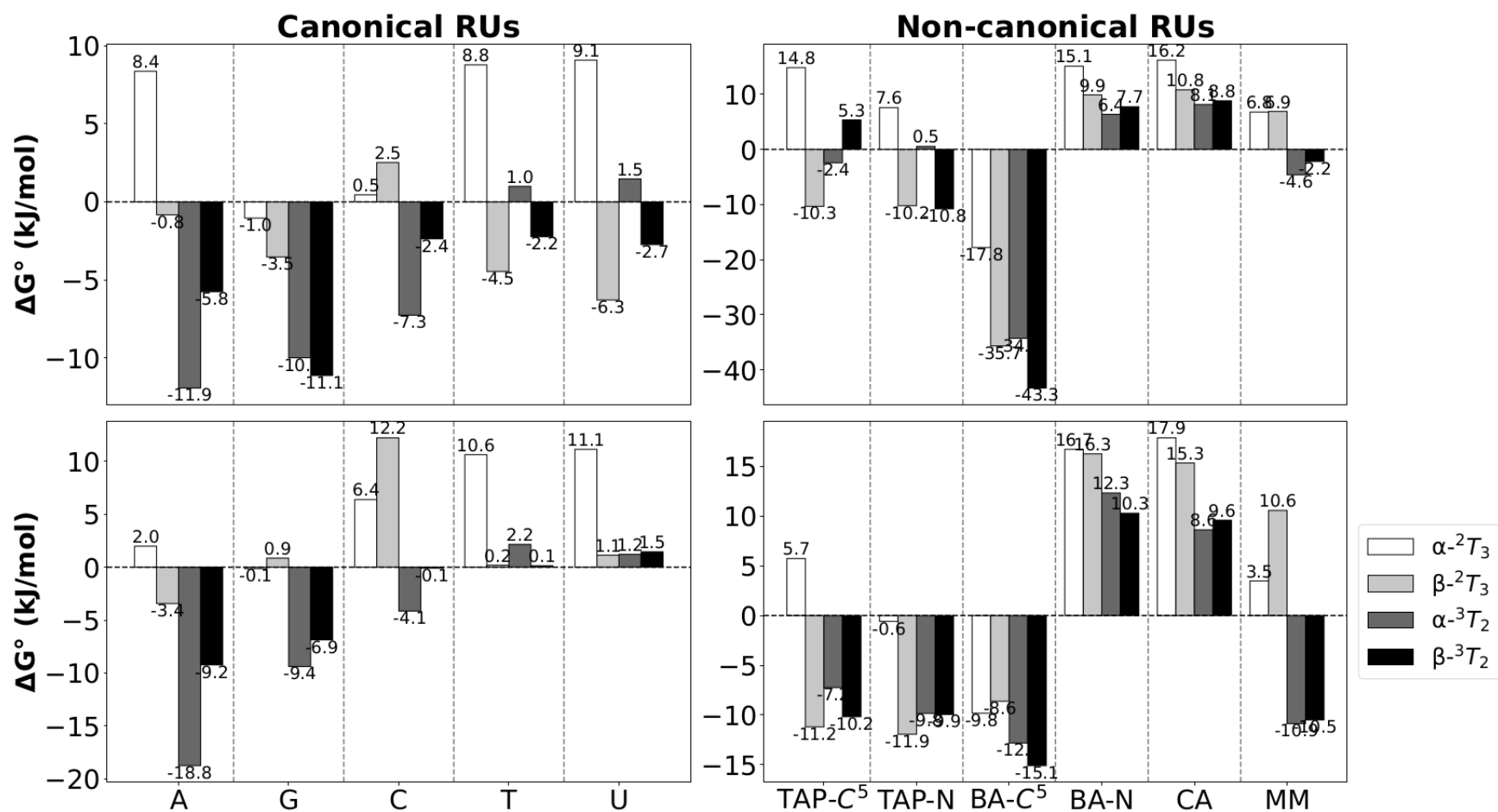
Figure 3.11 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical β- and α-counterparts of 2dRibf nucleosides. Each bar color represents the initial puckering conformations for the 5-MR of the sugar. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
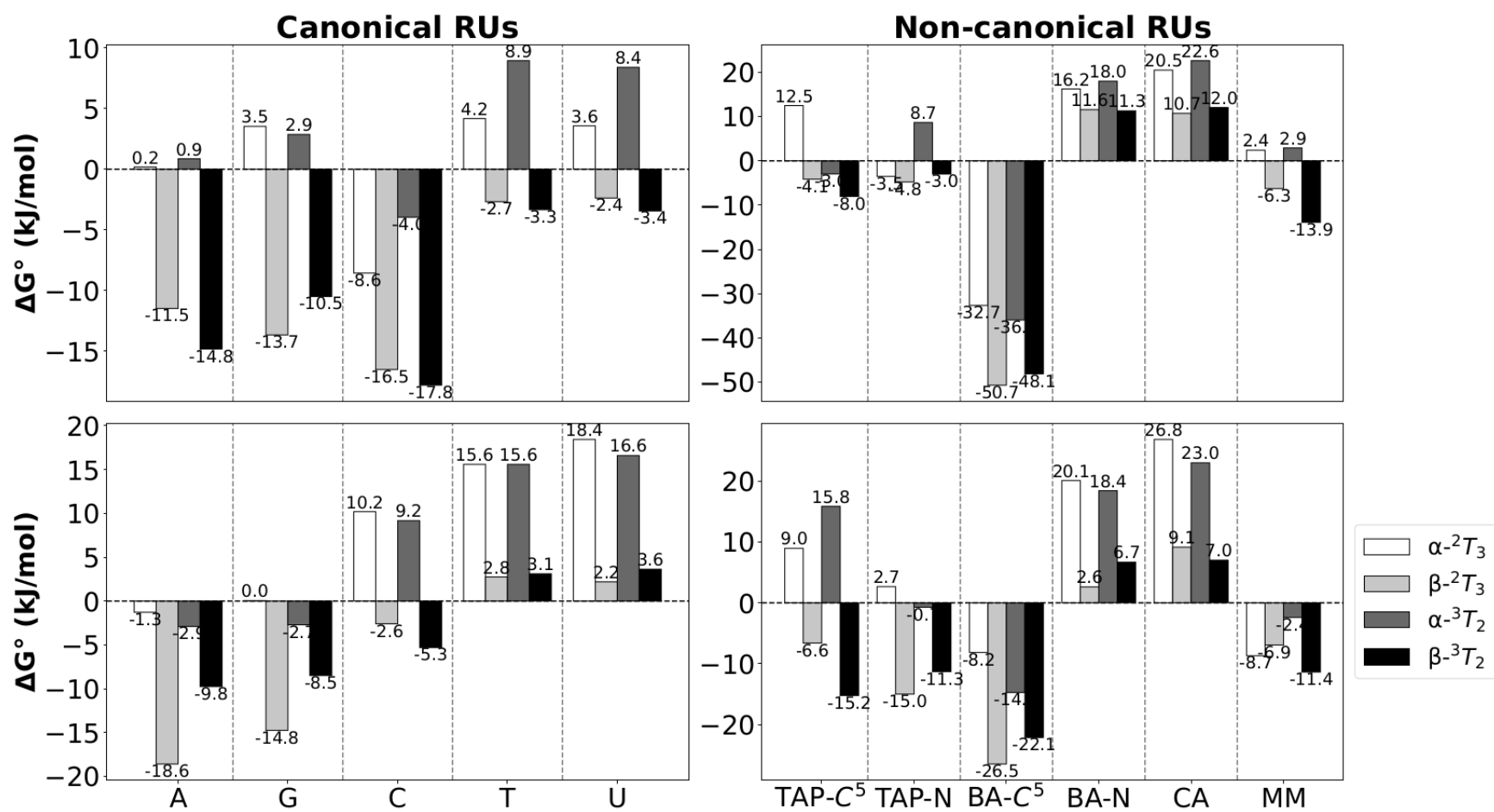
Figure 3.12 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical β- and α-counterparts of Ribf nucleosides. Each bar color represents the initial puckering conformations for the 5-MR of the sugar. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

was favored but the energies were within the error for the estimation of ΔG° at the B3LYP/6-311++G (d, p) level.

Since the condensation reaction of nucleobases with ribose has been widely reported in the literature, we can compare our results with some of these reports. For the case of the canonical nucleosides it can be noticed that our results are in certain agreement with the studies by Fuller *et al*. [4, 104] on the condensation reaction between different bases and ribose in aqueous solution with and without catalysis. In these two papers the authors only obtained β-glycosides in a low yield for the case of the reaction between A or G and ribose. No product was observed when reacting a mixture of pyrimidines with the ribose in aqueous solution. Similarly in our study the formation of β-nucleosides for A and G is predicted to be thermodynamically favored. Our results also constitute a computational corroboration of the well known "nucleoside problem".

In the case of the non-canonical nucleosides our results are in agreement to a certain extend with: 1) the results obtained by Chen *et al*. [19] for the glycosylation of TAP with ribose where the β-ribofuranose-C^5-TAP was the major product, 2) the studies by Cafferty and coworkers [12, 13, 14] for the synthesis of TAP, BA, CA and MM ribosides and 3) the work by Fialho [112, 16, 108] in which the CA seemed not to react with the ribose, meanwhile β-nucleosides were obtained for the reaction of TAP, BA and MM in different yields depending on the reaction conditions.

In one of the works by Cafferty and coworkers a library of 91 bases were tested for their capacity to: i) create complementary base pairing, ii) generate stacking assemblies, iii) be chemically and photostable, iv) react with ribose in aqueous environment, v) be obtained in prebiotic conditions and vi) absorb UV radiation. They found that only TAP, BA and MM fulfilled the 6 criteria.

In our case the β-anomers of the TAP-C^5, TAP-N, BA-C^5 and MM nucleosides with the sugar ring conformations $^2T_3$ and $^3T_2$ were favored in implicit solvation. This result suggests that at least TAP, BA and MM can react with ribose in water and circumvent the "nucleoside problem". Hence, the modeling of the Ribf nucleosides may constitute a validation case to corroborate the predictivity of the ΔG° of reaction from the DFT-B3LYP/6-311++G (d, p) level.

Figure 3.13 shows the bar graphs corresponding to estimated ΔG° of reaction for the formation of the canonical and non-canonical nucleosides containing Tho in vacuum and water environments. A glance at this figure reveals that all bases may react to a certain extent with Tho in aqueous solution. In vacuum the BA-C^5 nucleosides exhibit the more negative ΔG° with a ΔG°

= -37.1 kJ/mol for the α-$^2T_3$, a ΔG° = -48.5 kJ/mol for the β-$^2T_3$, a ΔG° = -49.7 kJ/mol for the α-$^3T_2$ and ΔG° = -38.8 kJ/mol for the β-$^3T_2$). The formation of the BA-N and CA nucleosides is estimated to not be favored with a $\Delta G^\circ \approx 11$ - 30 kJ/mol.

In aqueous solution the β-$^2T_3$ anomers of all canonical and non-canonical nucleosides seems to be thermodynamically favored with all values beyond the intrinsic error for the DFT method and the highest values correspond to the BA-C^5 nucleoside with a ΔG^0 = -52.1 kJ/mol. Notice that the only reference found in the literature for the reaction between threose with a nucleobase in aqueous solution is the studies by Fialho [40] and Fialho *et al.* [108] in which a low yield was obtained for the synthesis in dehydrating conditions between threose and TAP due to the rapid decomposition of Tho.

The thermodynamic plausibility for the formation of 2dRib nucleosides can be assessed by inspecting **Figure 3.14**. For the case of the canonical bases in both environments most of condensation reactions are unfavorable or their ΔG° negligible. The less favored reactions are for the pyrimidine nucleosides with $\Delta G^\circ \approx 1.8$ - 26.2 kJ/mol in vacuum and $\Delta G^\circ \approx 2.3$ - 28.3 kJ/mol in aqueous environment. In the case of the non-canonical nucleosides some of the energies are within 17 kJ/mol (error of computational calculations) but in vacuum the most favorable nucleosides are all the BA-C^5 with the most negative value (ΔG° = -43.8 kJ/mol) for the β-$^4C_1$. The same anomer-conformation is also favored for the TAP-C^5 but only by -9.2 kJ/mol. In the absence of solvation again the BA-N and CA nucleosides where the less favored with $\Delta G^\circ > 30$ kJ/mol.

When implicit solvation is considered the only $\Delta G^\circ \geq 17$ or ≤ -17 kJ/mol are observed for the unfavorable formation of BA-N (≈ 16 - 35.2 kJ/mol) and CA (≈ 17 - 34 kJ/mol) nucleosides. For the remaining nucleotides the β-$^4C_1$ anomer is favored even when the free energies for the condensation reaction are < -12 kJ/mol.

With the exception of the BA-C^5 nucleosides overall the formation of the α-anomers is less favored than the β-counterparts.
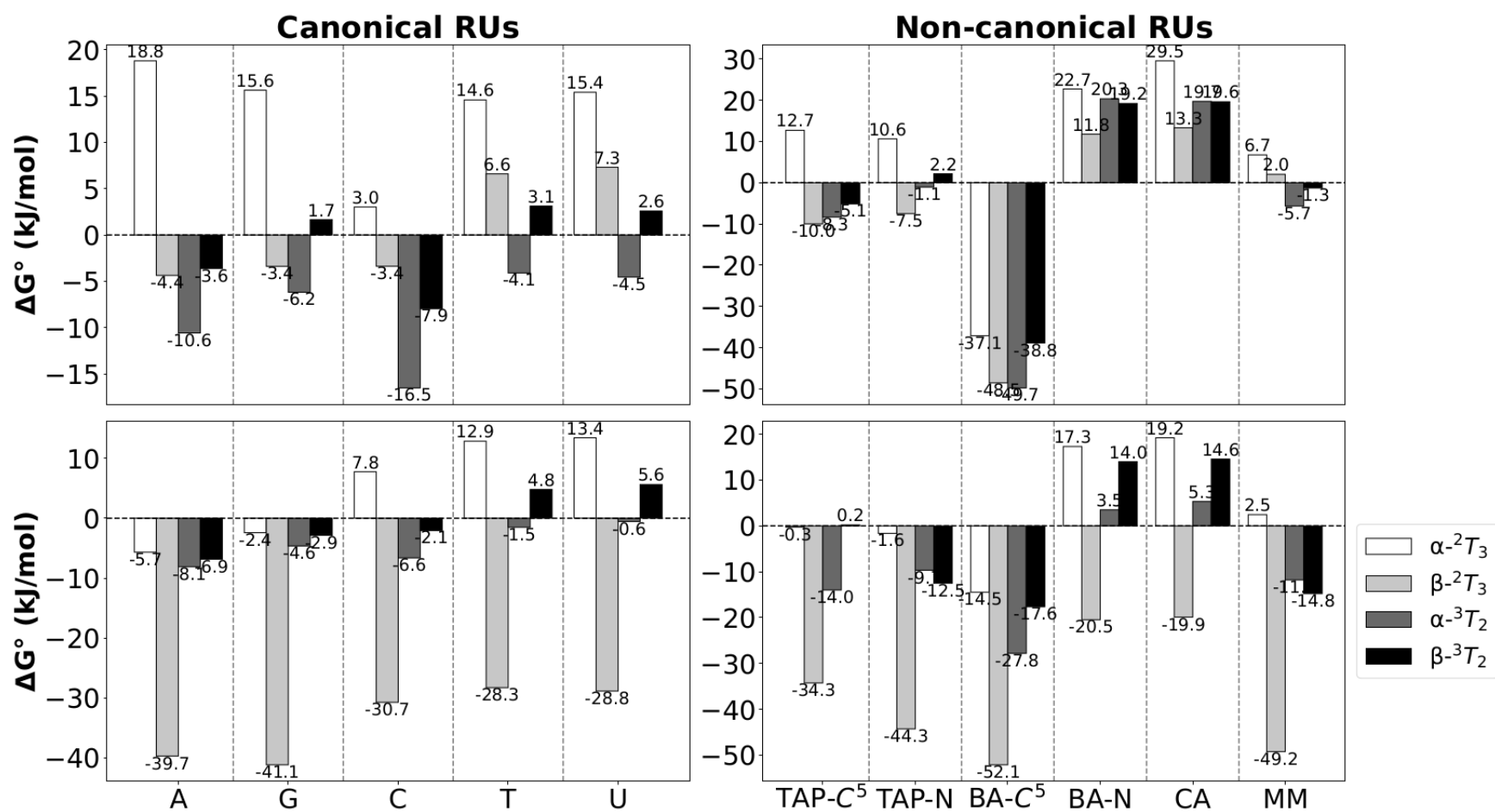
Figure 3.13 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical β- and α-counterparts of Tho nucleosides. Each bar color represents the initial puckering conformations for the 5-MR of the sugar. (Top) ΔG°, B3LYP/6-311++G (*d*,*p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d*, *p*) in aqueous medium using the IEFPCM solvation model.
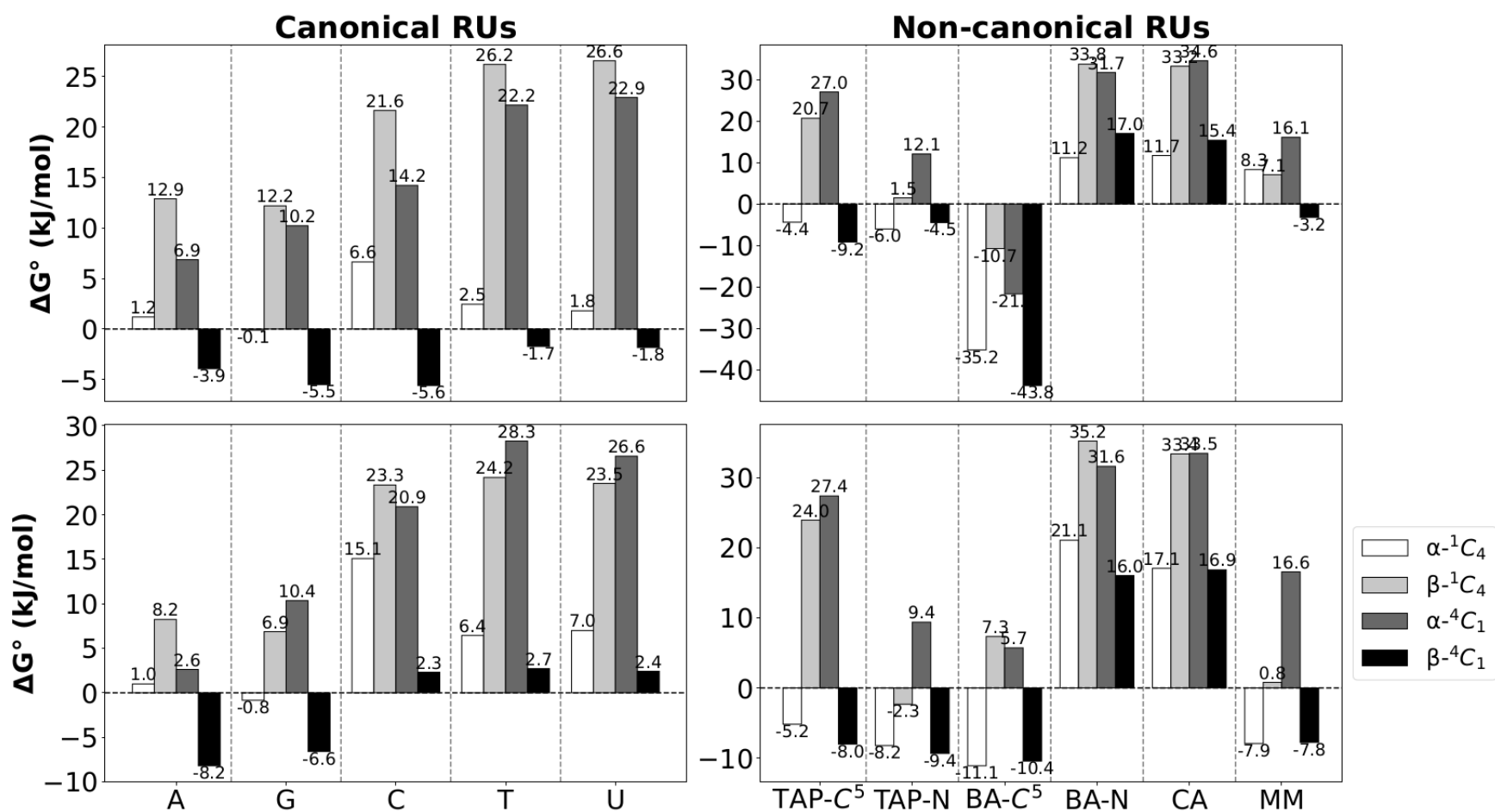
Figure 3.14 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical β- and α-counterparts of 2dRib nucleosides. Each bar color represents the initial puckering conformations for the 5-MR of the sugar. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

Analyzing the ΔG^0 for the condensation reaction between ribopyranose and the canonical and non-canonical bases (**Figure 3.15**) significant positive values for the ΔG° of formation of canonical pyrimidines, BA-N and CA nucleosides are observed in either environment. Most of these values overpass the intrinsic error of the DFT method used with the picks for the α-$^4C_1$ anomer of the CA nucleosides in aqueous environment (43.7 kJ/mol). The formation of all BA-C^5 nucleosides in vacuum is favored with a ΔG° of -43.2 kJ/mol for the β-$^4C_1$ anomer.

A comparison of the formation of 2dRib/Ribf nucleosides with respect to the formation of 2dRib/Rib nucleosides shows that in most cases the formation of the nucleosides from the furanose form is more favored than the nucleosides with a pyranose ring for the same sugar, despise the fact that the pyranose form it is overall more stable than the furanose form (see results from section 3.4.1).

In a previous paper published in 2022 by Castanedo *et al.* [57] it was analyzed the plausibility of the similar classic synthesis for the β- and α-anomers of canonical 2dRibf and Ribf nucleosides but at a lower DFT-B3LYP/6-31G (*d*, *p*) level in vacuum and using IFIEFPCM. In this paper it was found that overall the synthesis of neither of them could be predicted as thermodynamically favored or unfavored in either environment since most of the energies were within the intrinsic error of the method used (3 - 4 kcal/mol). Meanwhile, in this study the artificial synthesis of some A, G and C nucleosides is predicted to be favored through the condensation reaction of the components when considering different puckering conformations of the sugar and a higher DFT level of calculation.

Maybe the only two other theoretical works on the computational modeling of TAP, MM, BA and CA nucleosides(tides) are the works by Kaur and coworkers (Kcw) [76, 77]. In the publication from 2019 Kcw studied theoretically using MD, MM and DFT-B3LYP the β- and α-ribonucleosides of the non-canonical TAP-C^5 and CA, their complementary base pairing TAP-C^5:CA, stacking energies and deglycosylation barriers in vacuum and implicit solvation using IEFPCM. Kcw found from the deglycosylation profiles obtained for the most stable β- and α-anomers of both TAP ribonucleosides anomers at the B3LYP/6-31G (*d*, *p*) that the glycosidic bond in the TAP nucleosides (relative dissociation energy $\{E_r\} \approx 350 - 447$ kJ/mol) is stronger than the one in the canonical ribonucleosides ($E_r \approx 222 - 258$ kJ/mol) in both environments.

Meanwhile, an opposite result was obtained for both CA ribonucleosides anomers ($E_r \approx 155 - 193$ kJ/mol {vacuum} and 207-267 kJ/mol {IFIEFPCM}), leading the authors to propose
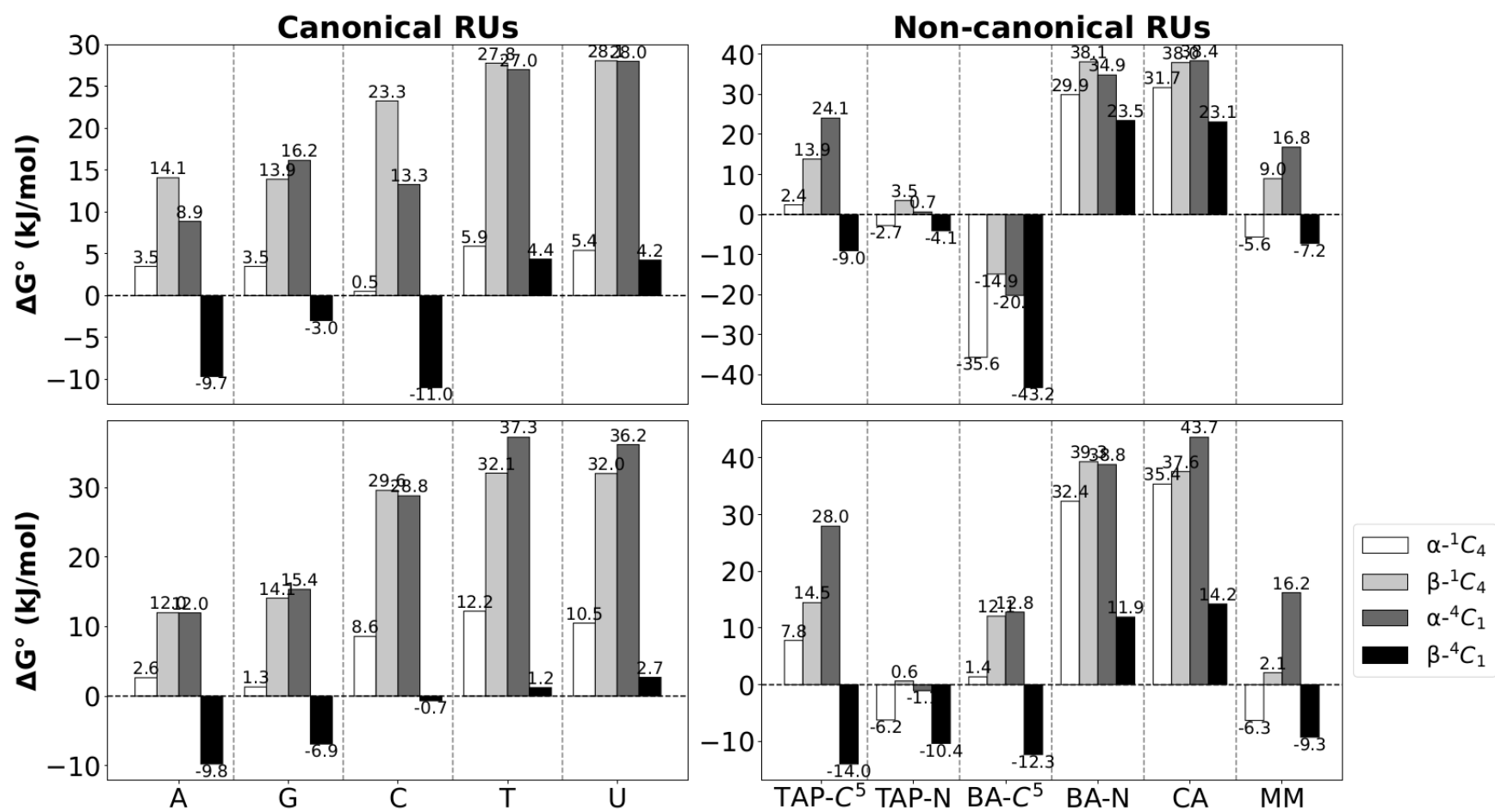
Figure 3.15 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical β- and α-counterparts of Rib nucleosides. Each bar color represents the initial puckering conformations for the 5-MR of the sugar. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

that maybe TAP-C^5 ribonucleosides were present in more hydrolytic prebiotic environments, while CA ribonucleosides could have been in less hostile conditions.

A similar study was conducted by Kcw in 2017 for the BA-C^5 in its keto form and MM ribonucleosides. When the authors analyzed the energetic barriers to dissociate the glycosidic bond between the two non-canonical nucleobases and ribose in the *β-* and *α*-configurations using the same DFT levels from the paper in 2019. They found that for BA-C^5 ($E_r \approx 317$ - 451 kJ/mol at a glycosidic bond length of 3.0 - 4.1Å) the cleavage of the glycosidic bond was less energetically favored than for canonical ribonucleosides. The glycosidic bond in MM ribonucleosides is predicted to be more resistant to dissociation in the *β*-configuration and overall, even when less stable than the one in BA-C^5 ribonucleotides it can be stronger than the glycosidic bond in canonical ribonucleosides. This could make both non-canonical ribonucleosides suitable for hydrolytic prebiotic scenarios.

The present study complements the results obtained by both papers from Kcw [76, 77]. In our case we decided to expand the analysis by considering the sugar puckering of both furanose and pyranose forms of ribose and also analyzed the C- and N-glycosides for the TAP and BA bases. In this study instead of analyzing the stability of the glycosidic bond we estimated the thermodynamic feasibility for the formation of the given ribonucleosides and compared them with the formation of the canonical nucleosides, predicting that *overall the formation of the β-anomer of the BA-C^5 ribonucleosides is the most favored in both environments compared to the canonical nucleosides. For the case of the TAP-C^5 and MM a similar trend is observed even when in some cases the energies are within the error of the calculations. In general, the classic synthesis of the CA ribonucleosides is estimated as the most unfavorable, even less favored than the one for some canonical nucleosides*. These results are in certain agreement with the results from Kcw [76, 77].

When analyzing the thermodynamic plausibility for the formation of canonical and non-canonical nucleosides with a non-sugar TCs the followed is observed:

i)   Glycerol can condensate with all RUs in both environments (see **Figure 3.16**), been the nucleosides with A and G the most favored in both environments for the canonical bases (-27.1 kJ/mol , -25.7 kJ/mol {vacuum} and -28.6 kJ/mol and -24.8 kJ/mol {water}) and for the non-canonical bases: the TAP-C^5 (-19.1 kJ/mol {vacuum}, -32.5 kJ/mol {water}) and BA-C^5 (-55.6 kJ/mol {vacuum}, -42.8 kJ/mol {water}).

ii) The formation of glyceric acid nucleosides (**Figure 3.17**) is unfavored for all canonical and non-canonical RUs with highly positive values in both environments that reach the 94.9 kj/mol (21 kcal/mol) for the CA. The only exception is similarly to other TCs the formation of BA-C^5 nucleosides in vacuum but the energy is only -9.1 kJ/mol which is within the error of the DFT method used.

iii) The condensation reaction of the acetylated derivatives of canonical and non-canonical nucleobases with AEG (**Figure 3.18**) is thermodynamically favored in aqueous environment for all cases but not in vacuum. The most favored reactions are for the AEG-A (-19.4 kJ/mol), AEG-C (-22.3 kJ/mol) and AEG-N^1-BA (-17.9 kJ/mol) nucleosides.

These three last results support the idea that maybe the first TCs may have been different than ribose or not even a sugar. There have been theories proposing the possible existence of a "**R**ibo**N**ucleo**P**rotein (RNP) world" instead of an "RNA world" [113]. This topic was widely discussed in an special issue of *Life* with 17 publications dedicated to discuss the origins and evolution of RNA and the "RNA world" hypothesis. Carter and coworkers [114] discussed this hypothesis on the basis of a coexistence between both polynucleotides (as storage of information) and polypeptides (as cofactors with catalytic activity) to complement the formation of RNA in the evolutionary process. Also, in support of the RNP theory Smith *et al*. [115] address the fact that some proteins can store information and catalyze the synthesis of other proteins. Wächtershäuser [116] argued against the "Strong RNA World" hypothesis analyzing reactions that require metals as catalyzers and hypotheses that until this type of chemistry was well defined in the prebiotic world non of the processes invoked by the "RNA world" theory could have happened. Additionally, it has been reported that the components of aegRNA can be synthesized in prebiotic conditions [117].

AEG nucleosides can create PNA polymers and there is not need for a phosphate group, since the monomers can polymerize through peptide bonds which are also favorable in potential prebiotic conditions avoiding possible repulsive forces of electrostatic nature and the single PNA strands can cross-pair with DNA and RNA [34, 118, 32, 119, 33].

Taking into consideration this evidence our results for the AEG nucleosides further support the idea of aegPNAs or GNA building blocks as precursors of today's nucleosides for an aqueous
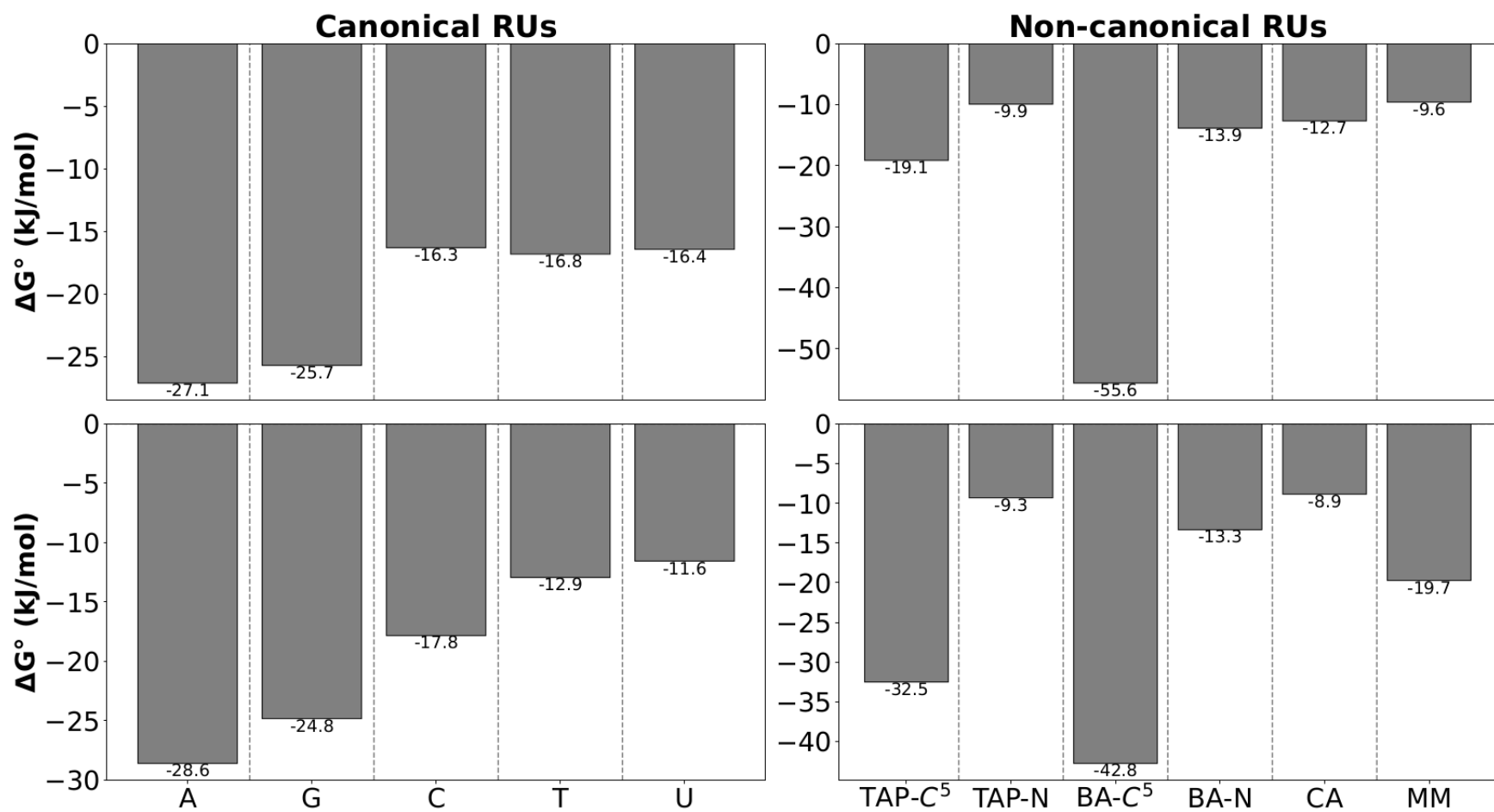
Figure 3.16 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical glycerol nucleosides. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
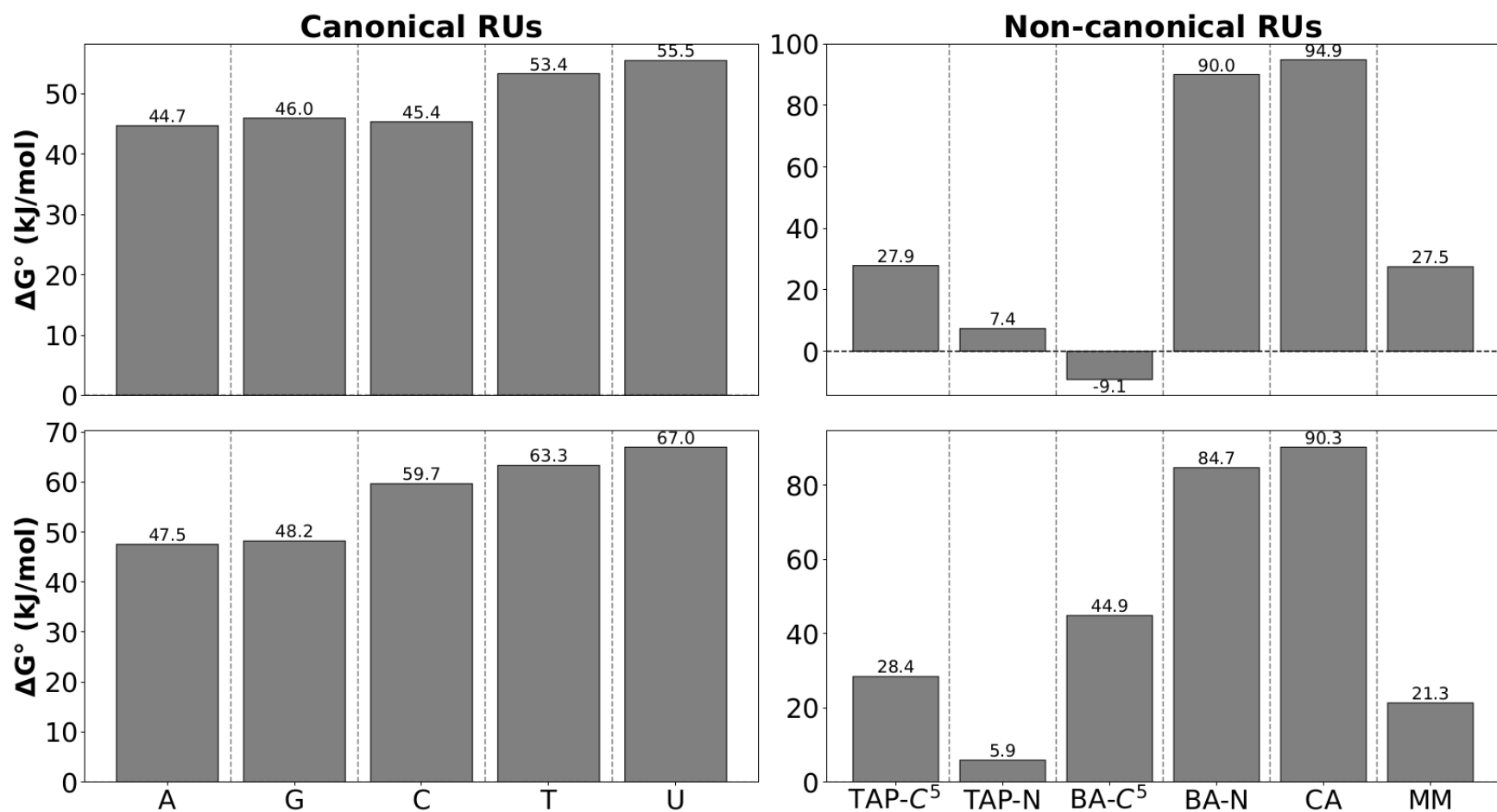
Figure 3.17 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 5 canonical and 6 non-canonical glyceric acid nucleosides. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d,p) in aqueous medium using the IEFPCM solvation model.
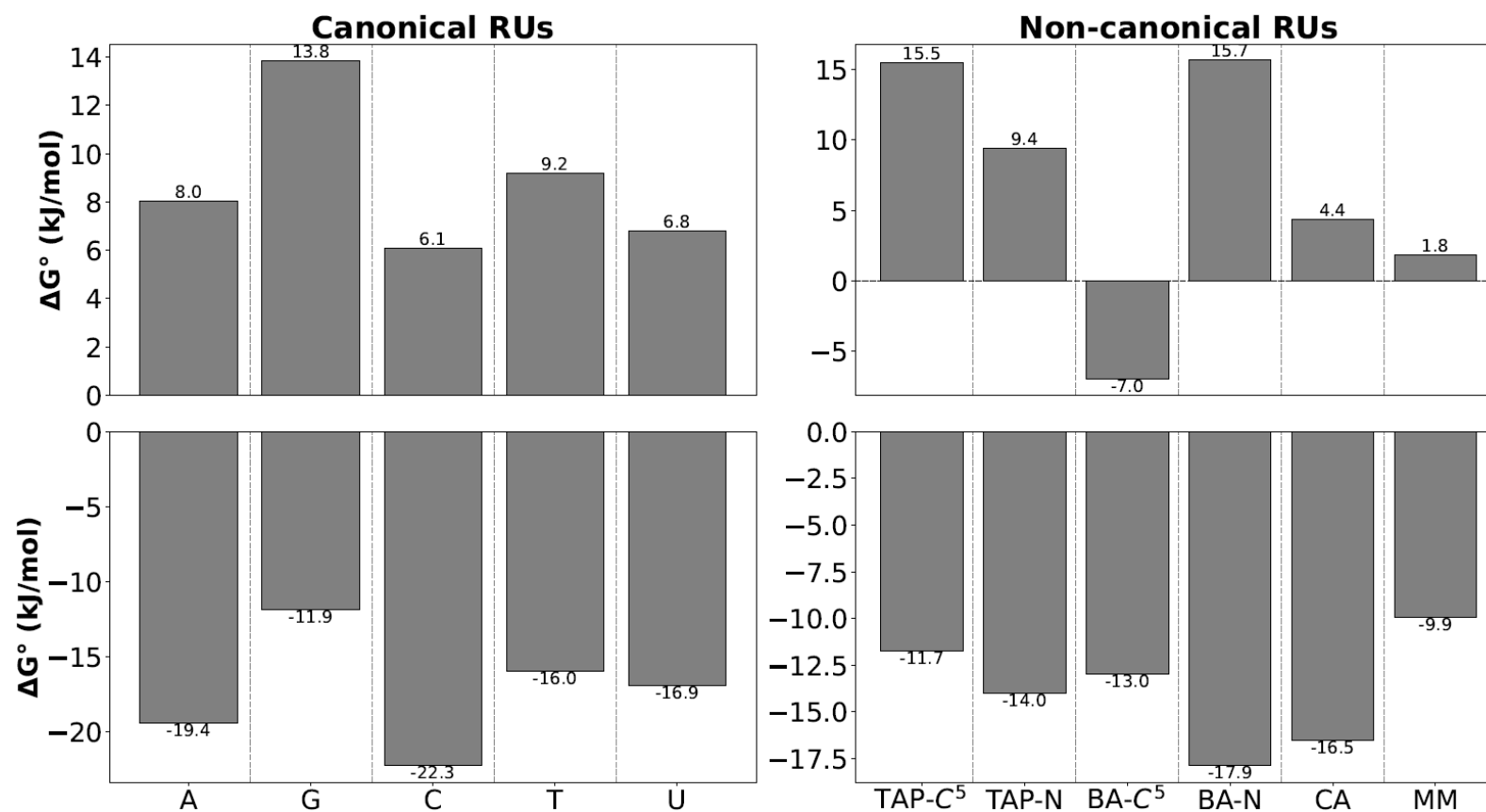
Figure 3.18 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis, leading to the 11 PNA nucleosides between AEG and the 5 canonical acetylated bases (RUs-Ac): A (adenine-N^9-Ac), G(guanine-N^9-Ac), C (cytosine-N^1-Ac), T (thymine-N^1-Ac), U (uracil-N^1-Ac) and 6 non-canonical RUs-Ac: TAP-C^5 (2, 4, 6-triaminopyrimidine-C^5-Ac), TAP-N (2, 4, 6-triaminopyrimidine-N^4-Ac), BA-C^5 (barbituric acid-C^5-Ac), BA-N (barbituric acid-N^1-Ac), CA (cyanuric acid-N^5-Ac) and MM (melamine-N-Ac). (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

229

environment that may have disfavor the thermodynamic synthesis of ribose nucleosides.

If this hypothesis can be proven experimentally, then an obvious next question emerges: how the building blocks of PNA and GNA transitioned to today's nucleosides(tides)? One accepted theory is that their strands can participate in template-directed reactions since they both can create hybrid double stranded helixes with DNA and/or RNA [29, 120] [34, 118, 32, 119, 33].

Another possibility could be that single AEG or glycerol nucleosides(tides) could have served as scaffold for the glycosylation and phosphorylation of the canonical components of today's nucleotides. Supporting this idea, Nature has left us a clue hidden in our own genetic material: the "**t**ransfer-RNA (tRNA)". tRNA can create polypeptides by transferring an amino acid bonded to a C3'-terminal of one of the tRNA strands to the nascent polypeptide chain in the peptidyl transferase center of ribosomal RNA (rRNA) [121, 122, 123, 124].

A recent study by Suárez-Marina and coworkers [125] has shown that canonical nucleosides and nucleotides can be synthesized through condensation reactions of their components in dehydrating conditions in the presence of glycine. Their study showed that glycine can react with the canonical bases and direct the glycosylation to the correct site with the ribose-5'-monophosphate.

Hirakawa and coworkers (Hcw) [126] have been able to synthesize ribose-5'-phosphate in the presence of urea, borate (BO_3^{-3}) and phosphate ions in heating conditions at 80 ºC for 24hrs. The product was obtained in a 22% yield after removing by acid hydrolysis the excess of reactants, the borate and urea. Hcw cleverly combined the use of borate to overcome the "asphalt problem" [2] derived from the formose reaction for the synthesis of ribose. The use of BO_3^{-3} was initially proposed by Ricardo and coworkers [2, 127] that proved that pentoses can be synthesized and stabilized in the presence of borates.

What if it can be possible to combine all these methods in the presence of a AEG or glycerol nucleoside as scaffold to further assist the condensation reactions between ribose, canonical bases and phosphate and stabilize the product by π-π stacking interactions? Suggestive as this hypothesis may be it will require further validation in the laboratory and theoretical calculations.

3.4.4 Circular statistics analysis on the conformation of the RU around the glycosidic bond

The canonical bases can adopt two preferential conformations around the torsion angle χ that involves the glycosidic bond in the nucleosides(tides) present in A-, B-, Z-DNAs and RNAs of biological importance. These conformations are *syn* ($\chi = 90^0$ - 270^0) where for canonical bases the N3 of purines or the H6 of pyrimidines occupy the same plane of the C5' in the sugar or *anti* ($\chi = 0^0$ - 90^0 and 270^0 - 360^0) where the previous atoms occupy the opposite plane [128, 129].

Some previous experimental and theoretical studies have shown that the *anti*-conformation prevails since can be more energetically favored in the building blocks of most A- and B-DNAs double helixes [130, 131], *e.g.*, the *syn* conformation can be thermodynamically disadvantageous by a marginal 5.0 kJ/mol for free purine nucleotides [130] and in secondary double helixes by 10.0 kJ/mol [132].

Sokoloski and coworkers (Scw) [133] analyzed in their paper from 2011 the base conformation in 51 RNAs of biological interest to understand the importance of *syn*-conformations in the stabilization of the 3D structure and folding of NAs. The authors argued that building blocks with *syn* conformation play an important role in the stabilization and folding of complex structures (triplex, quadruplex, loops, bulges) of NAs since it can create a more compact structure.

In their analysis, Scw excluded crystal structures with a resolution > 3 Å and their found that for all the 51 RNA structures (7311 nucleobases) 4.2% had a *syn* conformation, meanwhile 93.8 % had an *anti*. From those the *anti*-conformations prevailed in stacking (87%) and base pairing (81%) interactions, meanwhile *syn* conformations were predominant in tertiary stacking (74%) and tertiary base-pairing (93%).

Figures 3.19-3.23 represent the circular frequency histograms (rose diagrams) for the distribution of the torsion angle χ that defines the rotation of the RUs around the TCs in canonical, non-canonical and both types of nucleosides combined (see **Table A6** for the value of χ in the different nucleosides).

An analysis of the RUs' conformations around the α- and β-configuration of the 2dRibf (**Figure 3.19**) show that the canonical bases in vacuum predominate in the *syn* conformation in both β- and α-configurations, meanwhile there is an equal distribution of *syn*:*anti* (1:1) in implicit
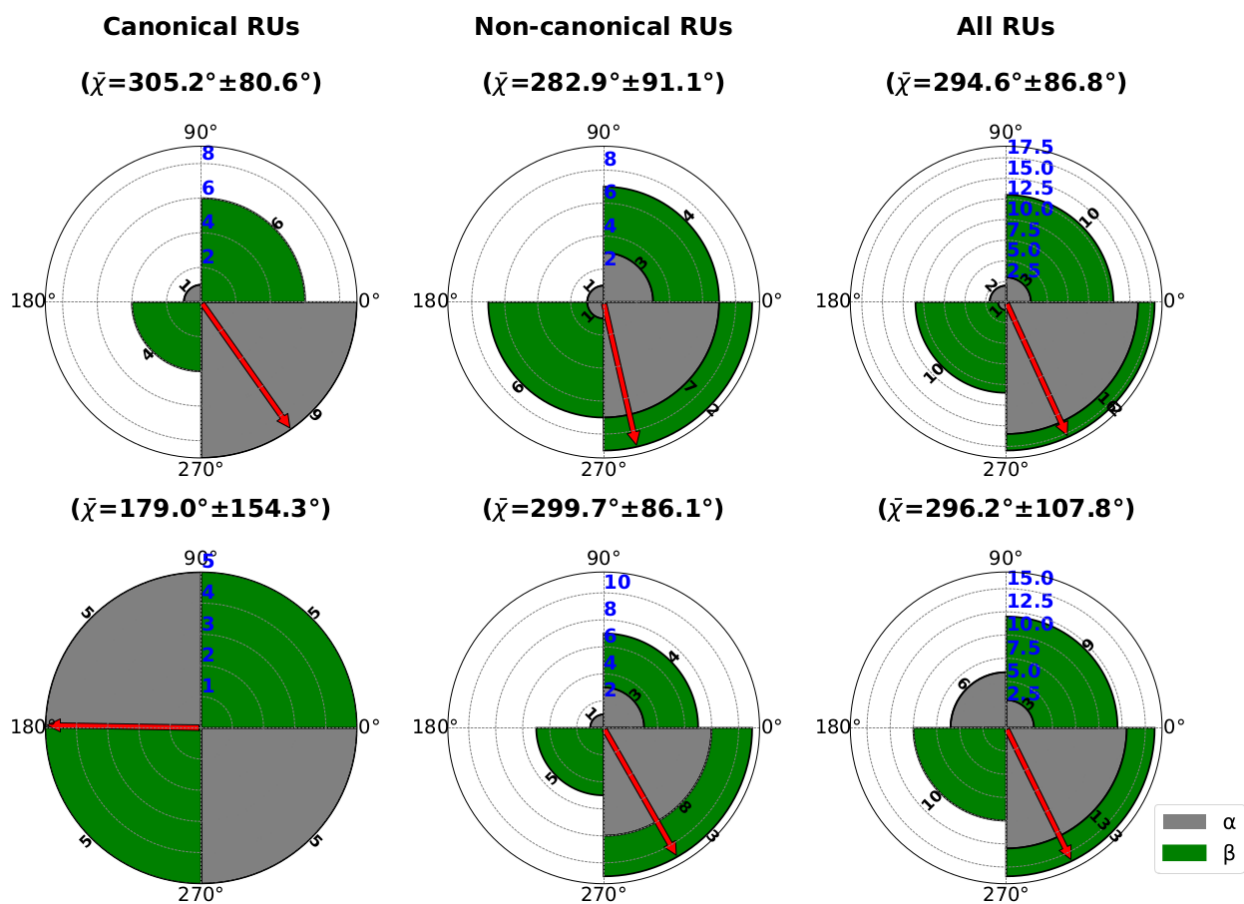
Figure 3.19 Rose diagrams (circular frequency histograms) for the torsion angle χ in the TC-RU glycosidic bond that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around 2dRibf. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(*d*, *p*) in vacuum, (Bottom) B3LYP/6-311++G(*d*, *p*) in aqueous medium using the IEFPCM solvation model.

water. The non-canonical nucleosides have a higher frequency of a *syn*-conformation in both environments and anomeric forms and overall, all the nucleosides also prefer the *syn*-conformation.

Figure 3.20 shows that both anomers of the canonical Ribf nucleosides in vacuum prefer an *anti*-conformation, while the opposite is observed in aqueous environment. For the non-
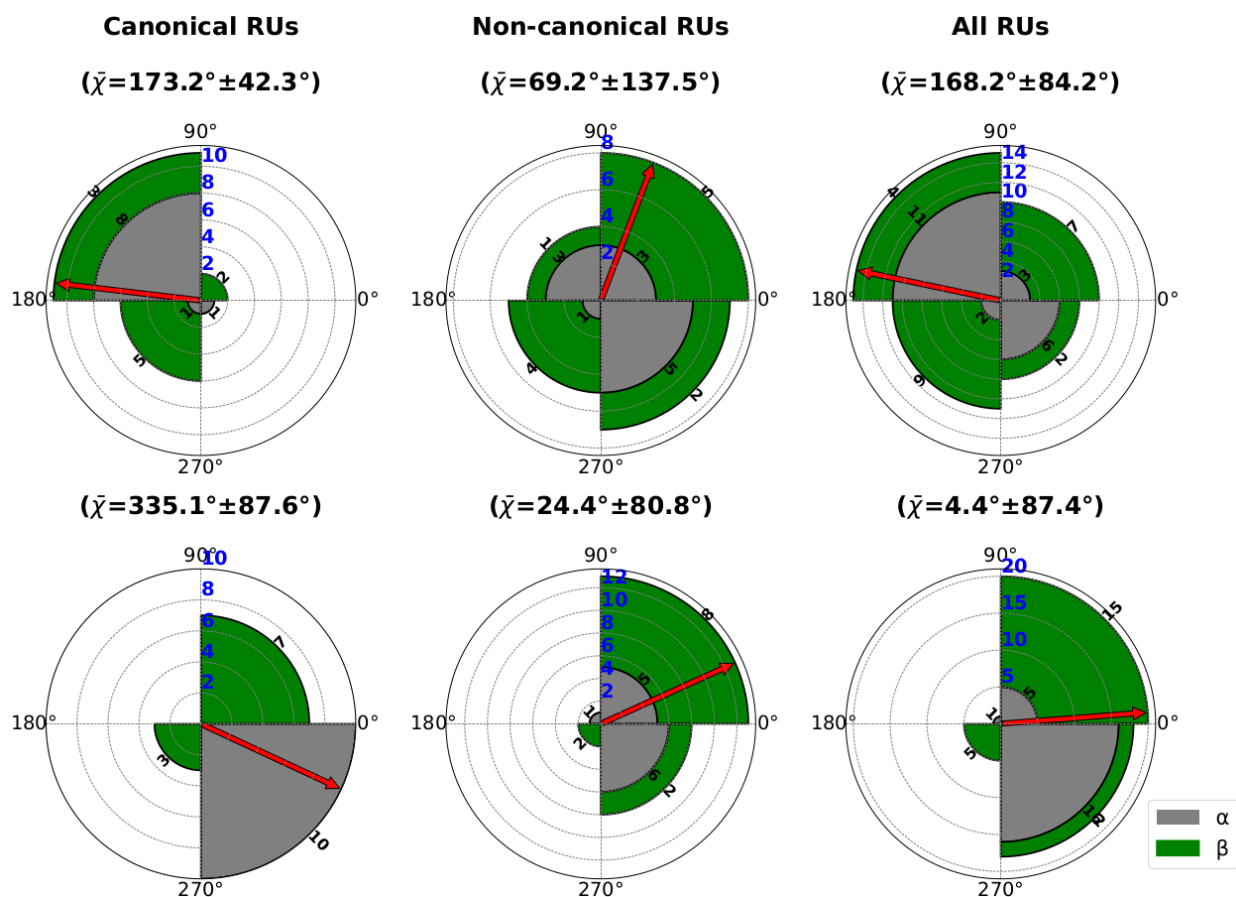
Figure 3.20 Rose diagrams (circular frequency histograms) for the torsion angle χ in the TC-RU glycosidic bond that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around Ribf. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

canonical nucleosides the *syn* conformers are more frequent in both vacuum and implicit solvation and overall all RUs in the Ribf nucleosides are more prompted to have an *anti*-conformation in vacuum and *syn* in aqueous environment.

In the case of the Tho nucleosides (**Figure 3.21**) both anomers of the canonical bases tend to be in a *syn* position with more frequency in vacuum, meanwhile in implicit solvation the *β*-anomers are more frequent in the *anti*-conformation and the *α*-anomers are more frequent in the *syn* conformation. The non-canonical nucleosides in vacuum have more *anti*-conformations for the *β*-anomers and *syn* for the *α*-configuration. In implicit solvation both anomers have more *syn*-conformations. In general, all nucleosides have a higher count of *anti*-conformations for both anomers in vacuum and in aqueous solution more *syn* conformations for the *α*-anomer and equal distribution of *syn:anti* for the *β*-counterpart.

In the case of nucleosides with 2dRib (**Figure 3.22**) most of the canonical bases in vacuum have a preferential *anti*-conformation for both anomers, meanwhile in water both anomers are almost equally distributed in both conformation. The *β*-anomers of the non-canonical bases in vacuum are more frequent in the *anti*-conformation, meanwhile the *α*-counterparts prefer the *syn*-conformation. In water both anomers are more frequent in the *syn*-conformation. In general, all *β*-nucleosides in vacuum have a higher probability for an *anti*-conformation and the *α*-counterparts for the *syn* conformation. In aqueous environment there is equal preference for either RU conformation in the *β*-anomer and the *α*-configuration prefers the *syn*.

The canonical nucleosides containing Rib (**Figure 3.23**) in vacuum prefer the *anti*-conformation in both anomeric configurations and both environments. The *β*-non-canonical nucleosides in vacuum are more frequently found in the *syn* position related to the Rib, meanwhile *anti* is more frequent for the *α*-anomers. In implicit solvation there are equal distributions of *syn* and *anti* for both anomers. For all nucleosides the frequency of both anomers in the *anti*-conformers in both environments are higher than the *syn* counterparts.

Our data agrees to a certain extent with the statistical analysis from Scw [133] regarding the canonical Ribf nucleosides in vacuum. It is important to outline that the general knowledge on the preferential *anti* conformation of the base around the sugar ribose has been a generalization from the analysis of nucleotides (not nucleosides) as monomers or as part of the polymeric structures of nucleic acids. Nevertheless, Scw proved that there are structural factors that justify the existence of the *syn* conformation, mostly in the folding of highly complex RNAs.
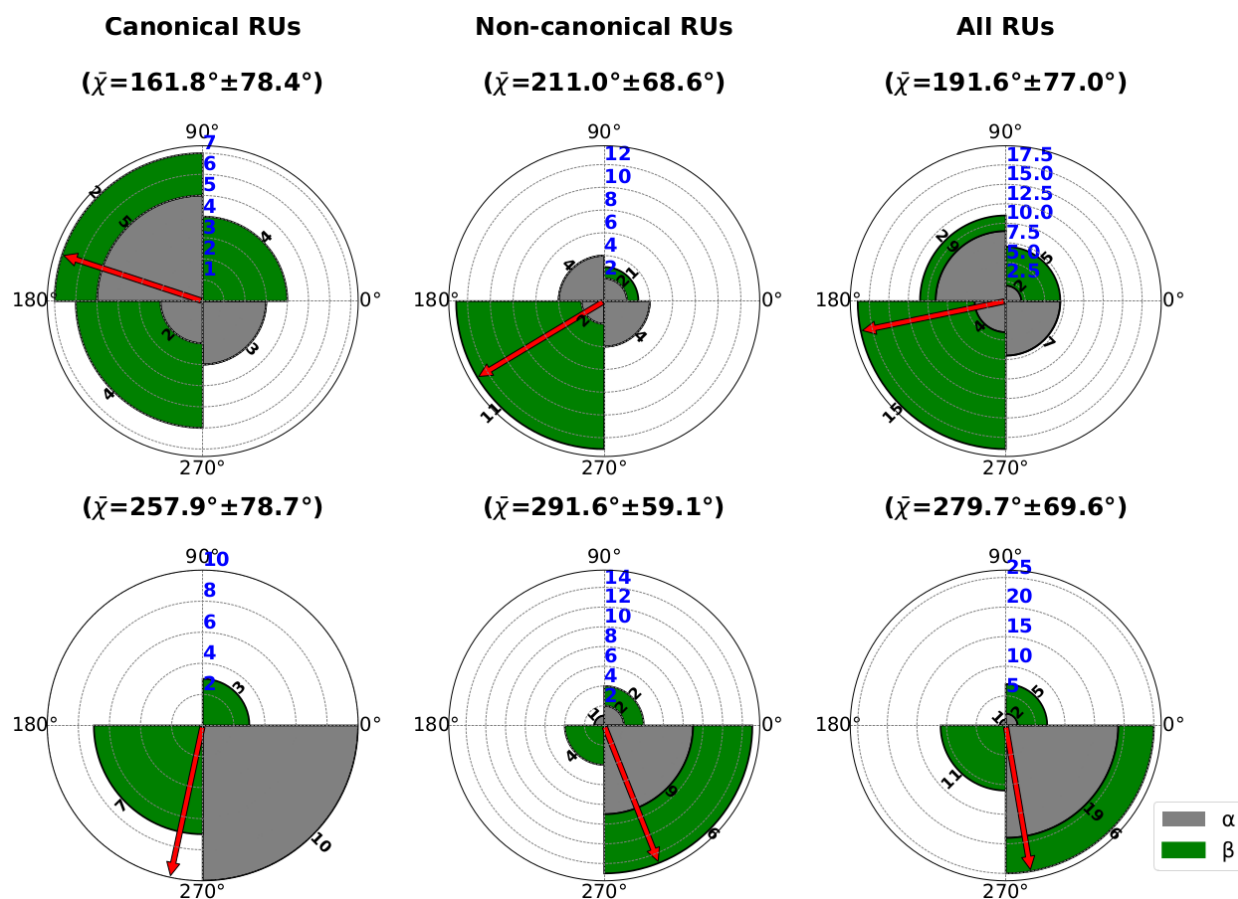
Figure 3.21 Rose diagrams (circular frequency histograms) for the torsion angle χ in the TC-RU glycosidic bond that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N[5], TAP-N, BA-C[5], BA-N, CA and MM) RUs around Tho. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
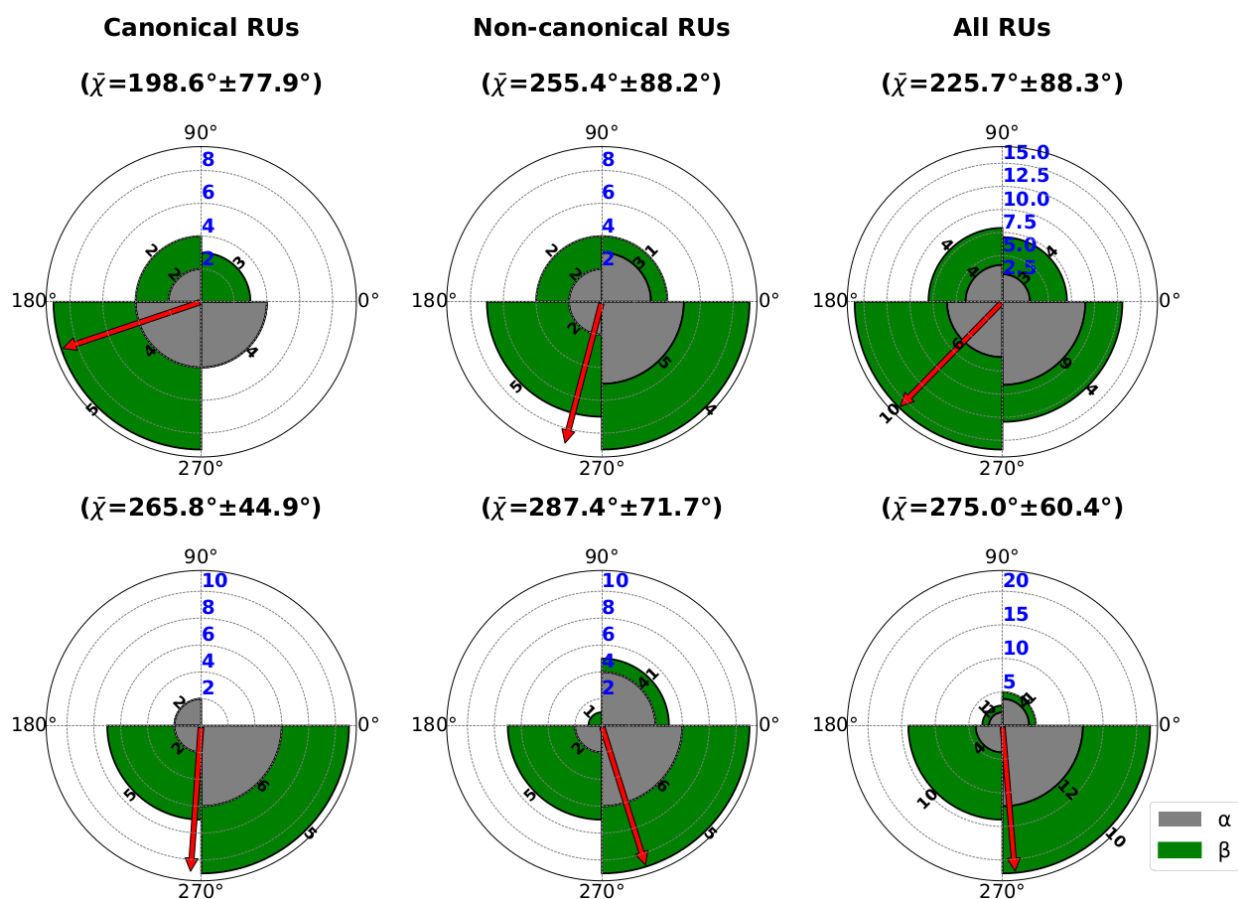
Figure 3.22 Rose diagrams (circular frequency histograms) for the torsion angle χ in the TC-RU glycosidic bond that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N[5], TAP-N, BA-C[5], BA-N, CA and MM) RUs around 2dRib. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
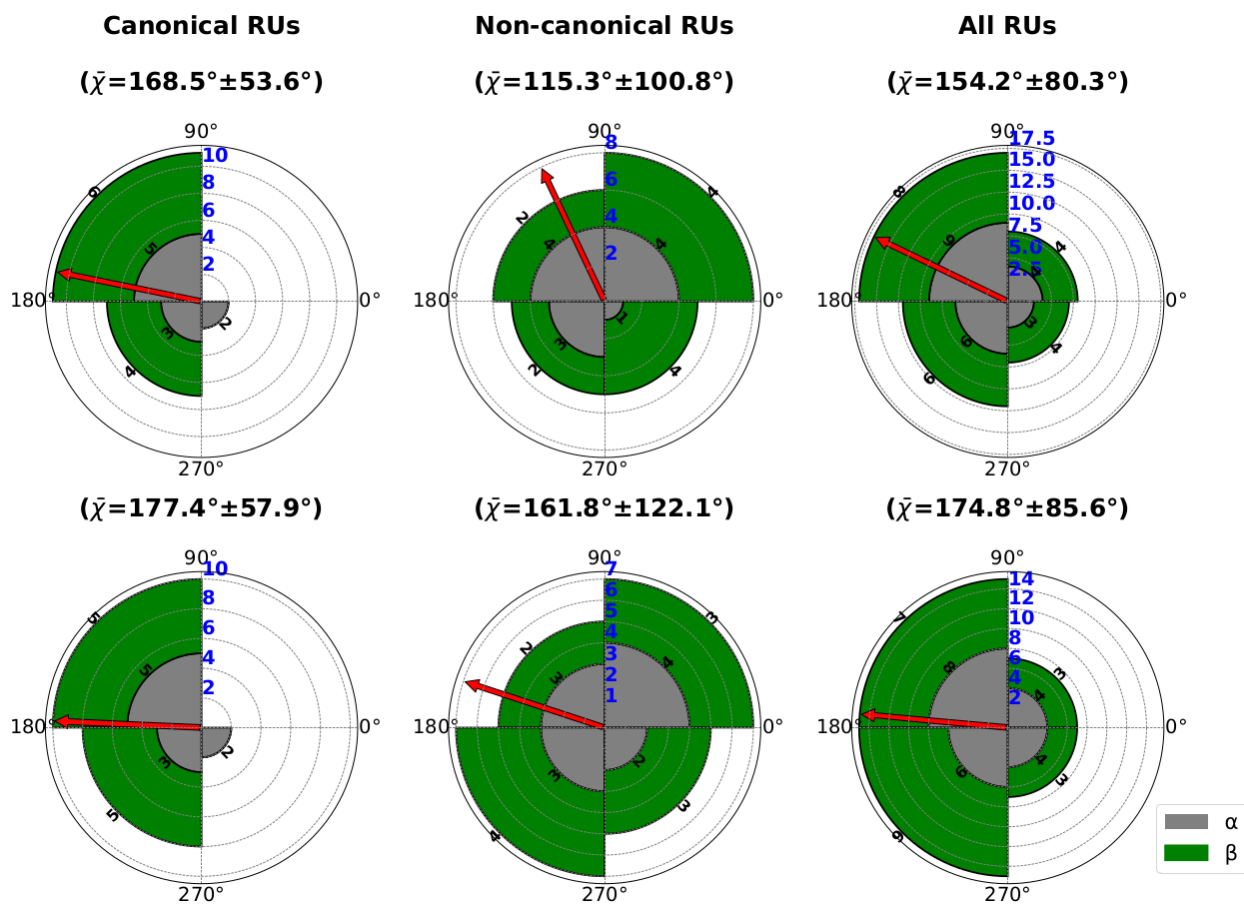
Figure 3.23 Rose diagrams (circular frequency histograms) for the torsion angle χ in the TC-RU glycosidic bond that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N[5], TAP-N, BA-C[5], BA-N, CA and MM) RUs around Rib. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

With respect to the non-canonical Ribf nucleosides we must compare our results to the ones obtained by Kaur and coworkers [76, 77]. The authors analyzed the torsion profiles for the rotation of the α- and β-anomers of the TAP-C^5 glycosylated, BA-C^5, CA and MM around the Ribf by rotating the torsion angle from 0º to 360º every 10º. After full-optimizations of the local minima of each PES they found that most global minima had an anti-conformation around the Ribf. In our study we used the same atom definitions for the glycosidic bond but exploring the surface only by 60º steps and optimizing at a higher DFT level. We found that in our cases there is a slight predominance of the *syn*-conformation for both anomers of the non-canonical Ribf nucleosides. Additionally, in our case we considered different initial puckering conformations $^2T_3$ and $^3T_2$ for the Ribf.

3.4.5 Sugar puckering parameters in furanose and pyranose TCs and associated nucleosides

The concept of puckering was first introduced by Kilpatrick *et al*. [98]. In the Kilpatrick's method the puckering conformations of pentagonal rings were determined by estimating the out-of-plane displacements for the ring atoms from the least square plane, but this description is only mathematical and does not consider the irregularities of a 5-MR sugar and the effects of axial and equatorial substituents.

Altona and Sundaralingam (AS) [101] circumvented the generalized character of Kilpatrick's model by obtaining two analytical expressions for a phase angle (*P*) and maximum puckering amplitude (ψ_m). *P* and ψ_m were deduced from 5 endocyclic angles in a set of β-nucleosides and nucleotides. Although, the Altona's method has been widely used in the literature for analyzing the puckering of ribose [100, 134] it has two main disadvantages: it depends on which atoms and in which order are selected as part of the torsion angles, besides it was only developed for 5-MRs [135].

Cremer-**P**ople (CP) developed in 1976 [56] their generalized puckering parameters that relied on the cartesian coordinates (x_i, y_i and z_i) of the ring atoms. These parameters are the phase angle (ϕ_2) and the radial q_2 for 5-MRs and ϕ_2, q_2 and q_3 for 6-MR. q_2 and q_3 can be transformed in the total puckering amplitude (*Q*). For pyranose rings the CP coordinates can be interconverted in

the polar coordinates: zenithal and azimuthal angles (ϕ_2 and θ) and the radial total puckering amplitude (Q) [56].

The CP puckering parameters define 20 (10 envelopes and 10 twists) puckering conformations for 5-MRs and 38 6-MR forms; 2 chairs, 6 boats, 6 twist/skew boats, 12 half-chairs and 12 envelopes [136].

The phase angle ϕ_2 defines the sugar puckering around the 5-MR of furanose sugars. **Figures 3.25-3.27** show the rose diagrams for the distribution of ϕ_2 in canonical and non-canonical nucleosides in their α- and β-configurations. The diagram of **Figure 3.24** shows the ϕ_2 values and the ideal twist conformations are in the vicinity of $90^{\rm o}$ for the $^2T_3$ and around $180^{\rm o}$ for the $^3T_2$ (see **Table A7** for the ϕ_2 and Q values in each nucleoside).

Figure 3.25 shows that for the 2dRibf nucleosides with canonical bases in vacuum the β-anomers have similar distributions between $0^{\rm o}$ (^0E) and $108^{\rm o}$ ($^0T_1$) and $288^{\rm o}$ - $360^{\rm o}$ with an unclear preference for a specific puckering. The α-anomers show a preference for a puckering between $0^{\rm o}$ (^0E) and $18^{\rm o}$ ($^0T_1$). For the same nucleosides in implicit solvation the β-anomers have conformations around different ϕ_2 values including $90^{\rm o}$ ($^2T_3$) - $108^{\rm o}$ (E_3). The α-anomers also have sugar puckering around different ϕ_2 values.

The non-canonical β-nucleosides in vacuum have different puckering but in implicit solvation there is a slight preference for $90^{\rm o}$ ($^2T_3$) - $108^{\rm o}$ (E_3). The α-anomers have different puckering in vacuum but in water there are 5 structures with puckering conformations around $126^{\rm o}$ ($^4T_3$) - $144^{\rm o}$ (^4E) and 4 nucleosides between $288^{\rm o}$ (^3E) - $306^{\rm o}$ ($^3T_4$). Overall, all α-nucleosides in vacuum have a preference for the $0^{\rm o}$ (^0E) and $18^{\rm o}$ ($^0T_1$) puckering, meanwhile the β-counterparts have a less clear preference. In aqueous environment the β-anomers have a preference for the $90^{\rm o}$ ($^2T_3$) - $108^{\rm o}$ (E_3) range with 7 structures, meanwhile the α-counterparts have 8 nucleosides in the $126^{\rm o}$ ($^4T_3$) - $144^{\rm o}$ (^4E).

A similar analysis of **Figure 3.26** shows the preferential puckering of the canonical and non-canonical β- and α-anomers of Ribf nucleosides. The β-anomers of canonical nucleosides in vacuum have a preference for $54^{\rm o}$ ($^2T_1$) - $72^{\rm o}$ (^2E) puckering, while the α-anomer has different puckering around $0^{\rm o}$ - $90^{\rm o}$ and $288^{\rm o}$ - $324^{\rm o}$. In aqueous environment the β-anomer has 4 nucleosides in the $54^{\rm o}$ ($^2T_1$) - $72^{\rm o}$ (^2E) range meanwhile the α-anomer preferred puckering between $270^{\rm o}$ ($^3T_2$) - $306^{\rm o}$ ($^3T_4$).
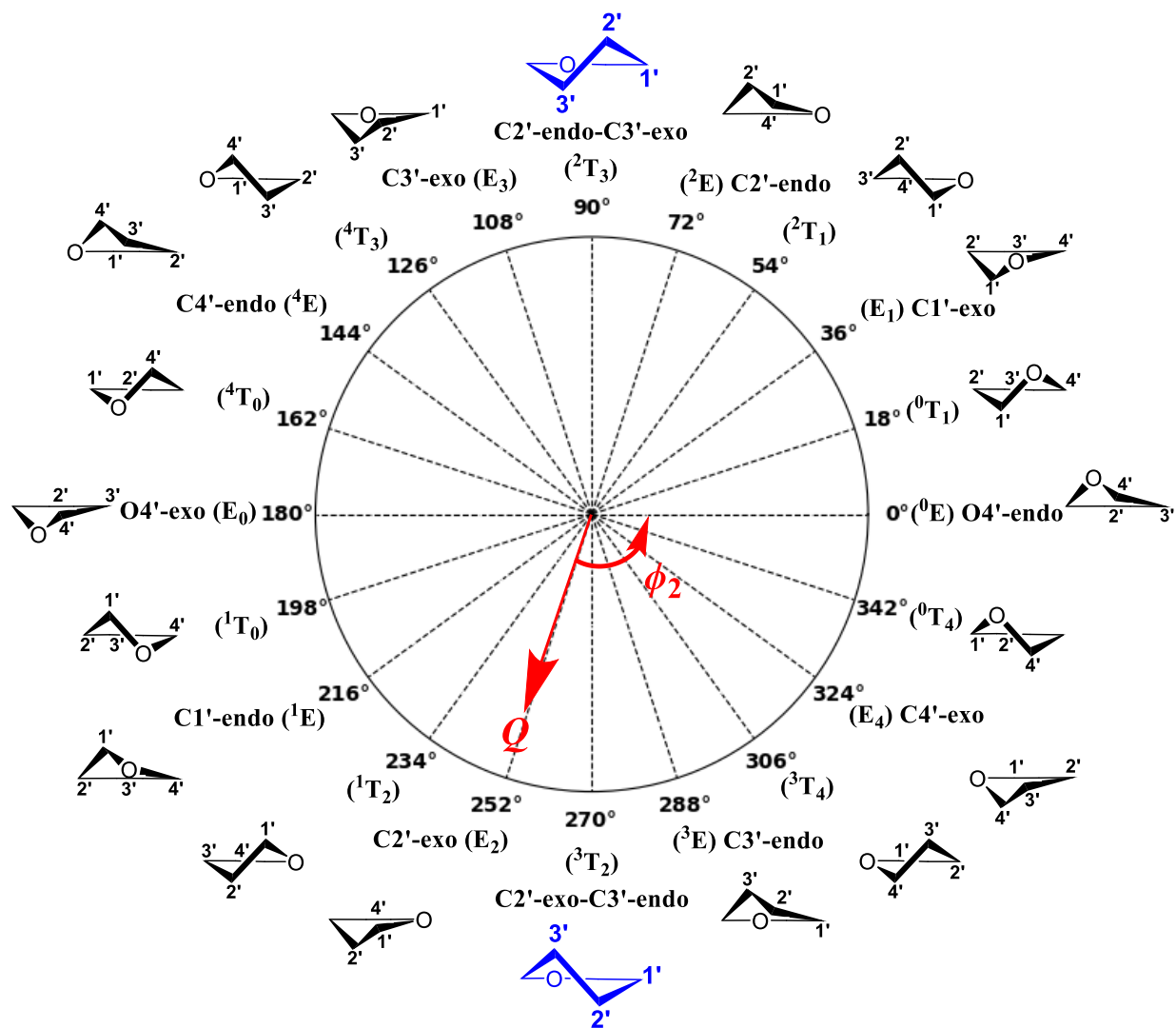
Figure 3.24 Sugar ring conformations for the furanose (5-MR) sugars [137, 138, 139] accordingly to the Cremer-Pople (CP) phase angle ϕ_2. The preferential conformations $^2T_3$ (C2'-endo-C3'-exo) and $^3T_2$ (C2'-exo-C3'-endo) are colored in blue. E: east, T: twist.
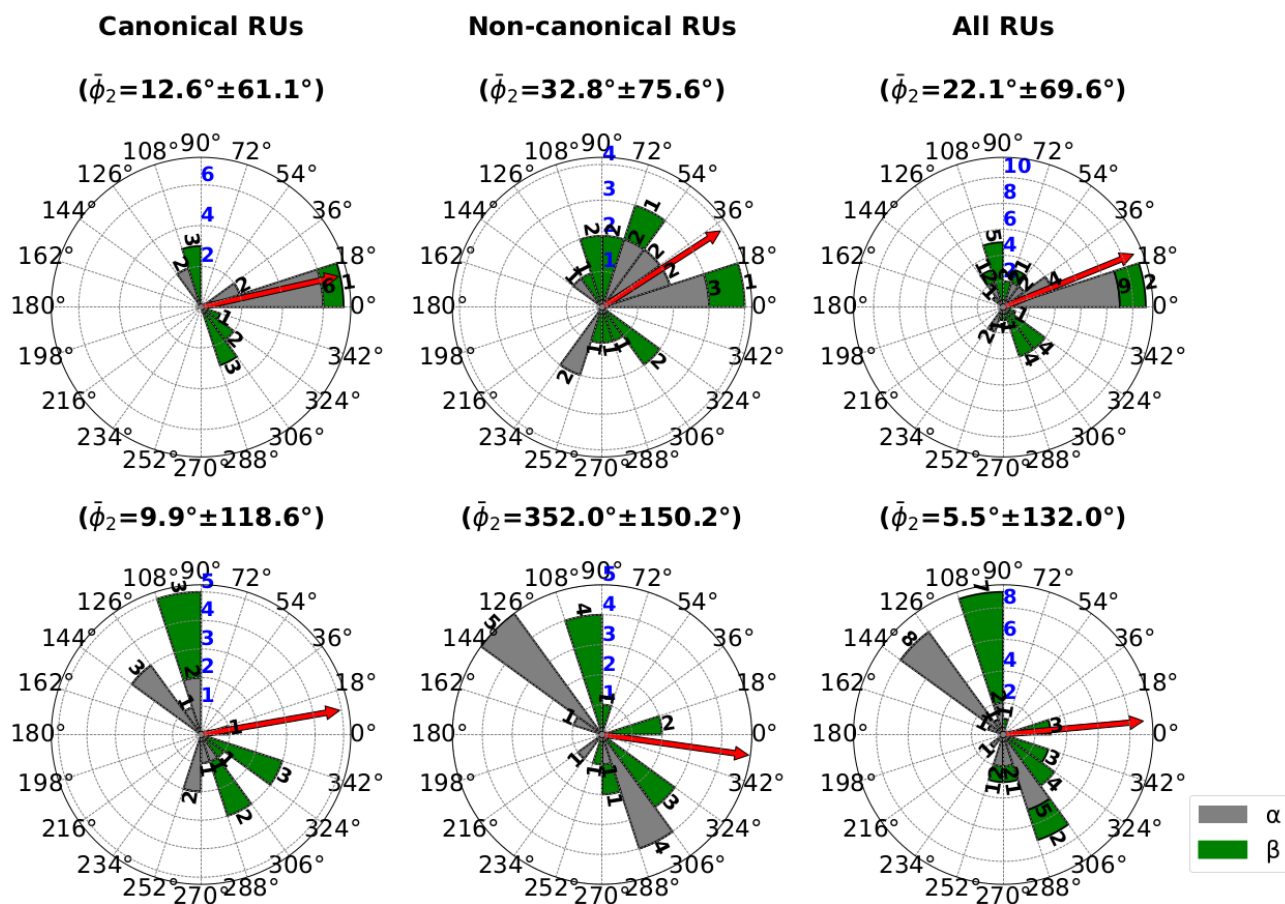
Figure 3.25 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR puckering conformation for the 2dRibf. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue ones represent the α-counterparts for the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

241
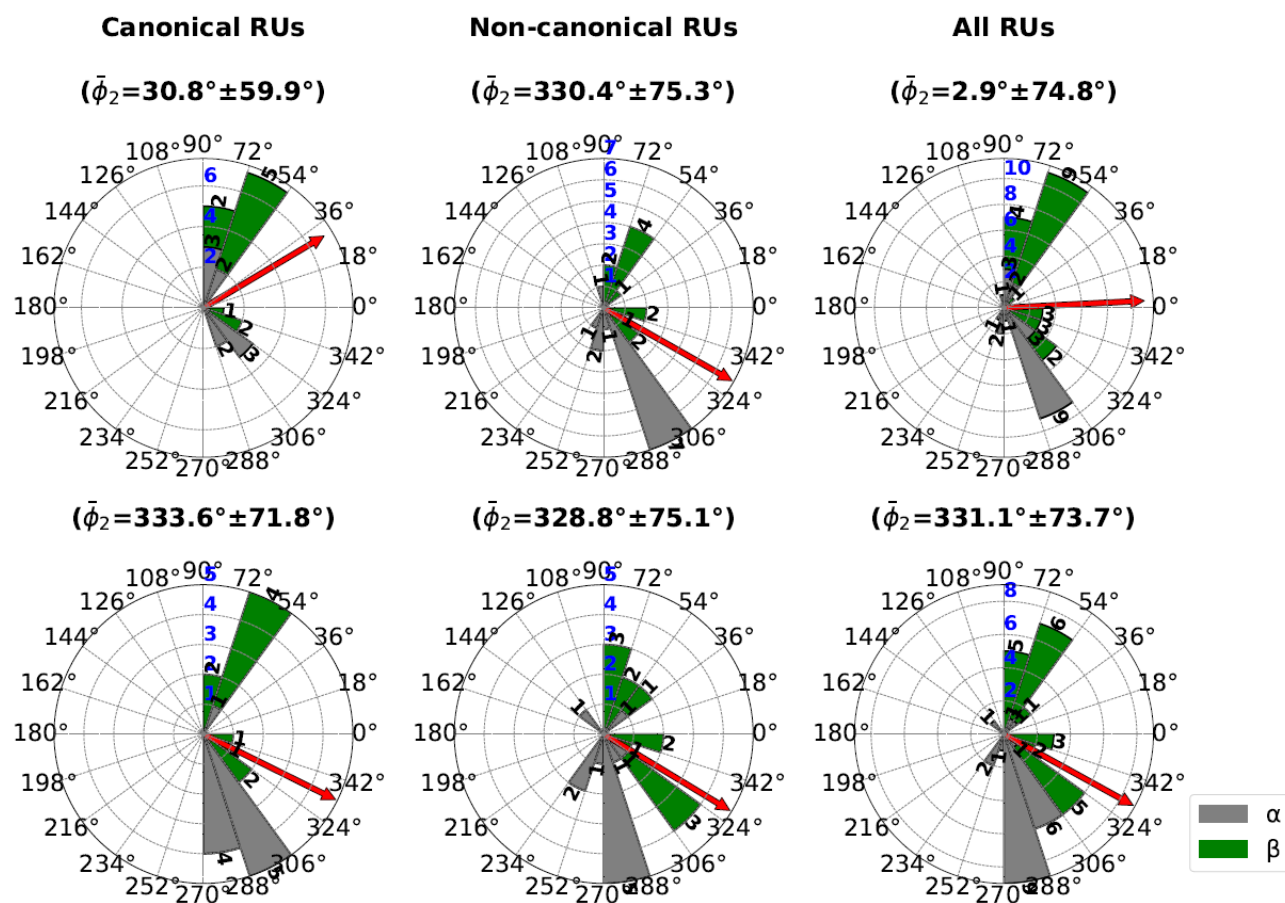
Figure 3.26 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR puckering conformation for the Ribf. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue ones represent the α-counterparts for the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

In vacuum, the β-anomers of the non-canonical nucleosides have a distribution of puckerings between 0° (^{0}E) - 90° (^{2}T$_3$) and 306° (^{3}T$_4$) - 360° (^{0}E), meanwhile the α-anomers have a clear preference for the 288° (^{3}E) - 306° (^{3}T$_4$). In implicit solvation, the β-anomers populate the puckering in the 0° (^{0}E) - 90° (^{2}T$_3$) and 306° (^{3}T$_4$) - 360° (^{0}E). The α-anomers have a preference for sugar ring conformations between 270° (^{3}T$_2$) and 288° (^{3}E).

Taking a look at all the nucleosides in vacuum the β-anomers have 9 nucleosides with sugar puckering between 54° (^{2}T$_1$) - 72° (^{2}E) and the α-anomers have 9 conformers between 288° (^{3}E) - 306° (^{3}T$_4$). Considering implicit solvation the β-anomers followed a similar preference to the ones in vacuum and the α-anomers preferred the puckering between 270° (^{3}T$_2$) and 306° (^{3}T$_4$).

Figure 3.27 shows that the β-anomers in vacuum for either canonical, non-canonical and all Tho nucleosides preferred two main puckering ranges: 108° (E$_3$) - 126° (^{4}T$_3$) and 270° (^{3}T$_2$) - 288° (^{3}E). The α-anomers in vacuum have in all cases different distributions of sugar ring puckering with the highest counts between 270° (^{3}T$_2$) - 306° (^{3}T$_4$).

In aqueous solution the β-anomers of the Tho canonical nucleosides have different distributions in the 126° (^{4}T$_3$) - 288° (^{3}E), meanwhile the α-anomers have 8 structures between 126° (^{4}T$_3$) - 144° (^{4}E).

In the case of the non-canonical nucleosides optimized using the IEFPCM model the β-anomers have 4 structures in the 108° (E$_3$) - 126° (^{4}T$_3$) and the 306° (^{3}T$_4$) - 324° (E$_4$) respectively , meanwhile the α-anomers have the higher count between the 126° (^{4}T$_3$) - 144° (^{4}E) sugar puckering conformations.

For all the nucleosides the β-anomers have different distributions meanwhile the α-counterparts are more frequent to have the puckering of the threose ring between the 126° (^{4}T$_3$) - 144° (^{4}E) conformations.

Figure 3.28 shows that the preferential conformations for 6-MR are the chairs ^{1}C$_4$ and ^{4}C$_1$.

Figures 3.29-3.30 show the 2D kernel density probability surfaces as a function of the phase angles ϕ and θ for the puckering of the P-forms of the canonical and non-canonical 2dRib and Rib nucleosides. Areas in darker blue represent concentration (high Kernel density probability) of similar P-form puckering for the sugar in the different nucleosides (see **Table A8** for ϕ, θ and Q values).

A glance at **Figure 3.29** reveals that the β-anomer of all nucleosides have a high probability around the ^{4}C$_1$ ($\phi = 0^o$ and $\theta = 0^o$) and around the ^{3}H$_4$ ($\varphi_2 = 30^o$ and $\theta = 135^o$) - ^{5}E ($\varphi_2 = 120^o$ and
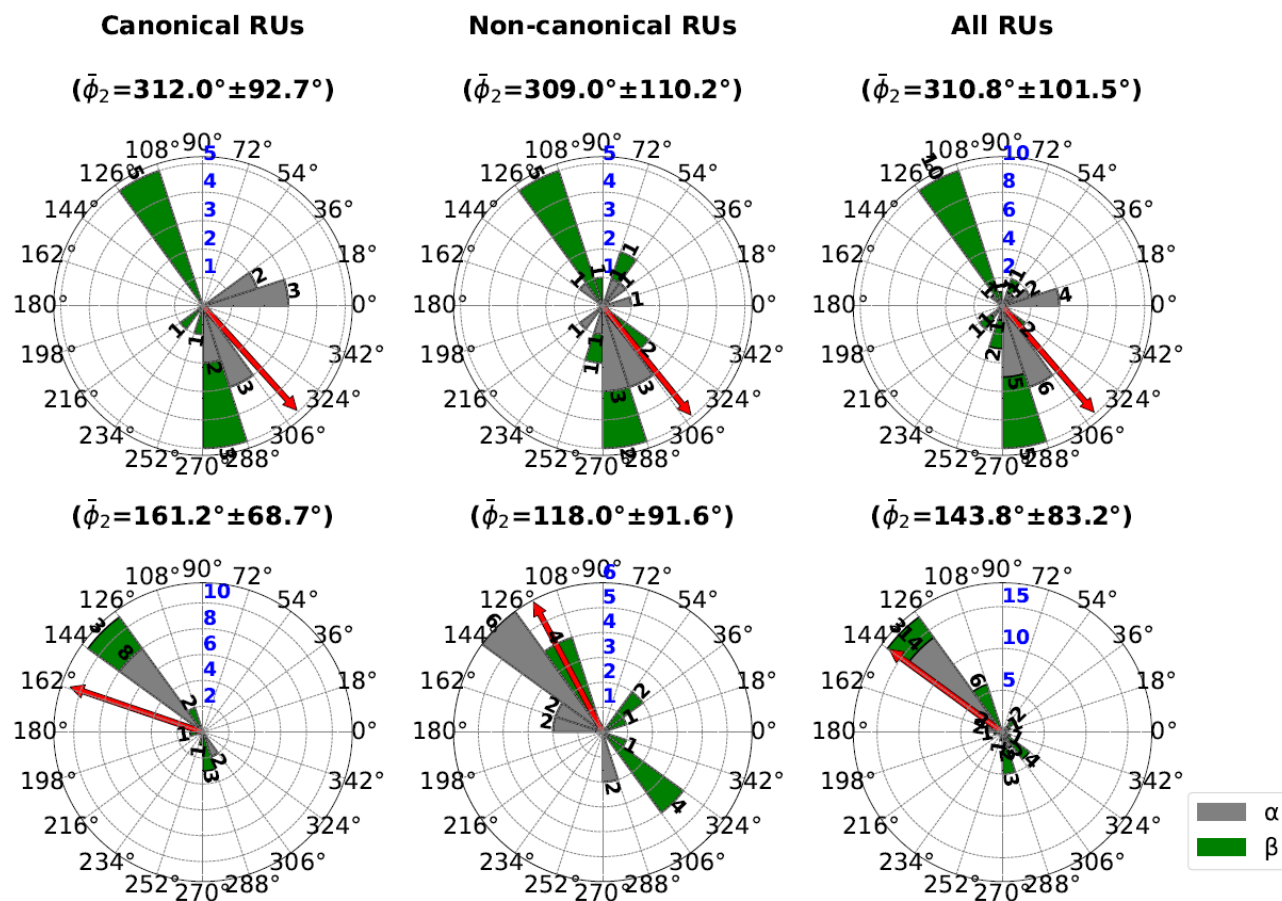
Figure 3.27 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR puckering conformation for the Tho. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue ones represent the α-counterparts for the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
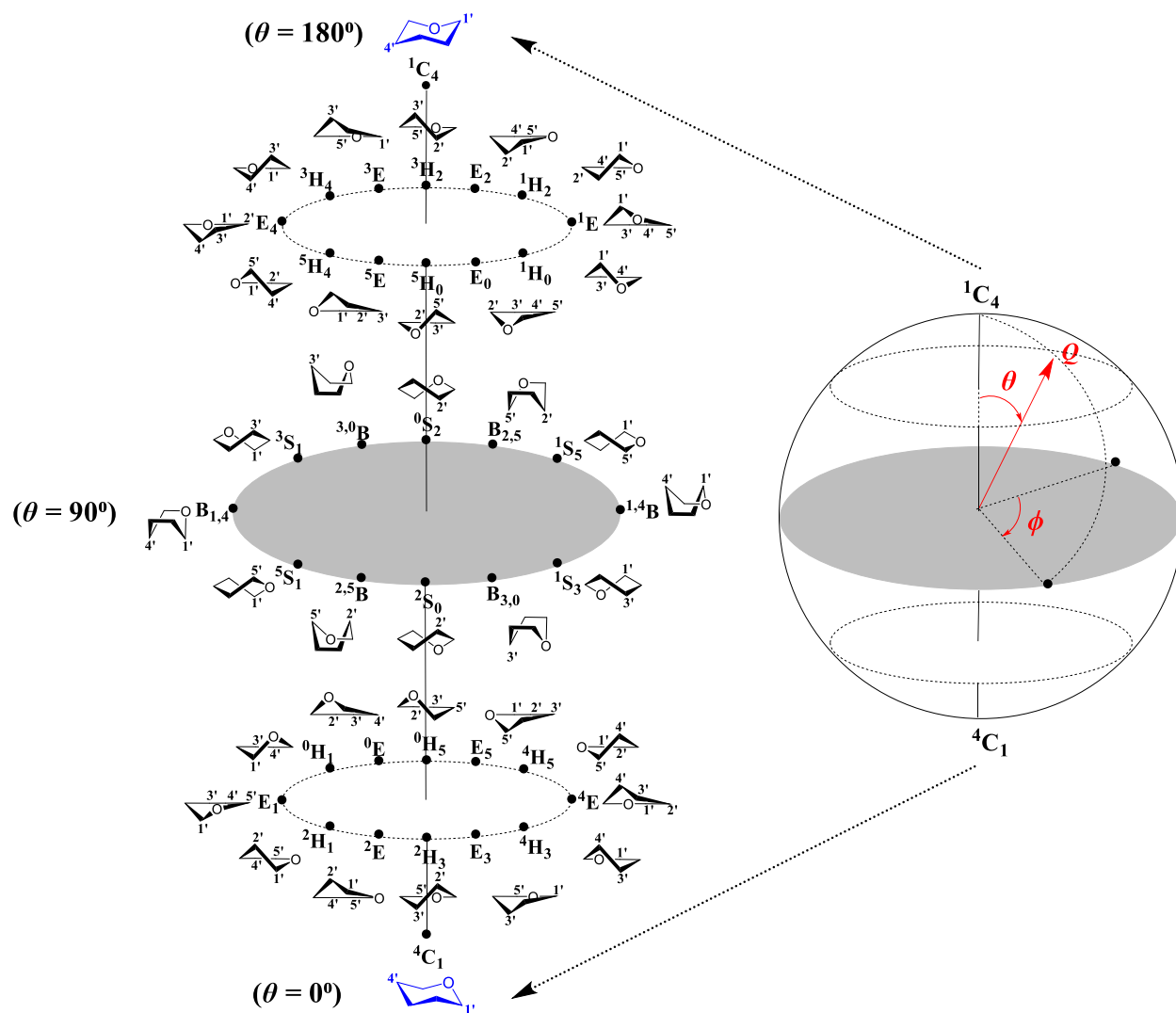
Figure 3.28 Sugar ring conformations for the pyranose (6-MR) sugars accordingly to the CP polar coordinates ϕ (zenithal angle), θ (azimuthal angle) and Q (total puckering amplitude) [139]. The preferential chair conformations $^4C_1$ and $^1C_4$ are colored in blue. C: chair, B: boat, S: twist/skew boat, H: half-chair, E: envelope.
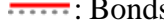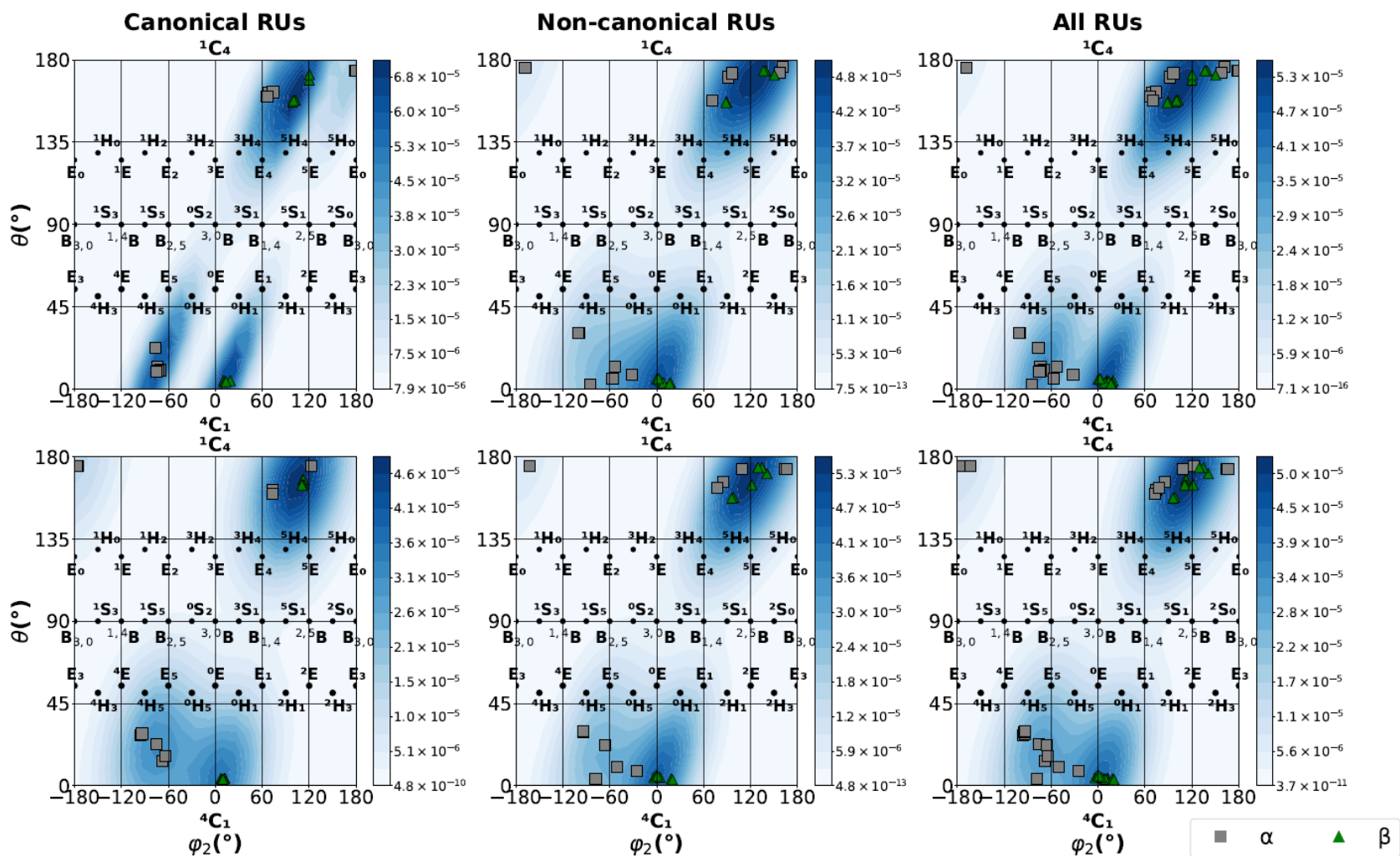
Figure 3.29 Mercator projection of the kernel density probability and scatter plot for the distribution of the values for the phase angles ϕ and θ in (°) for the puckering of the 6-MR β- and α-2dRib. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
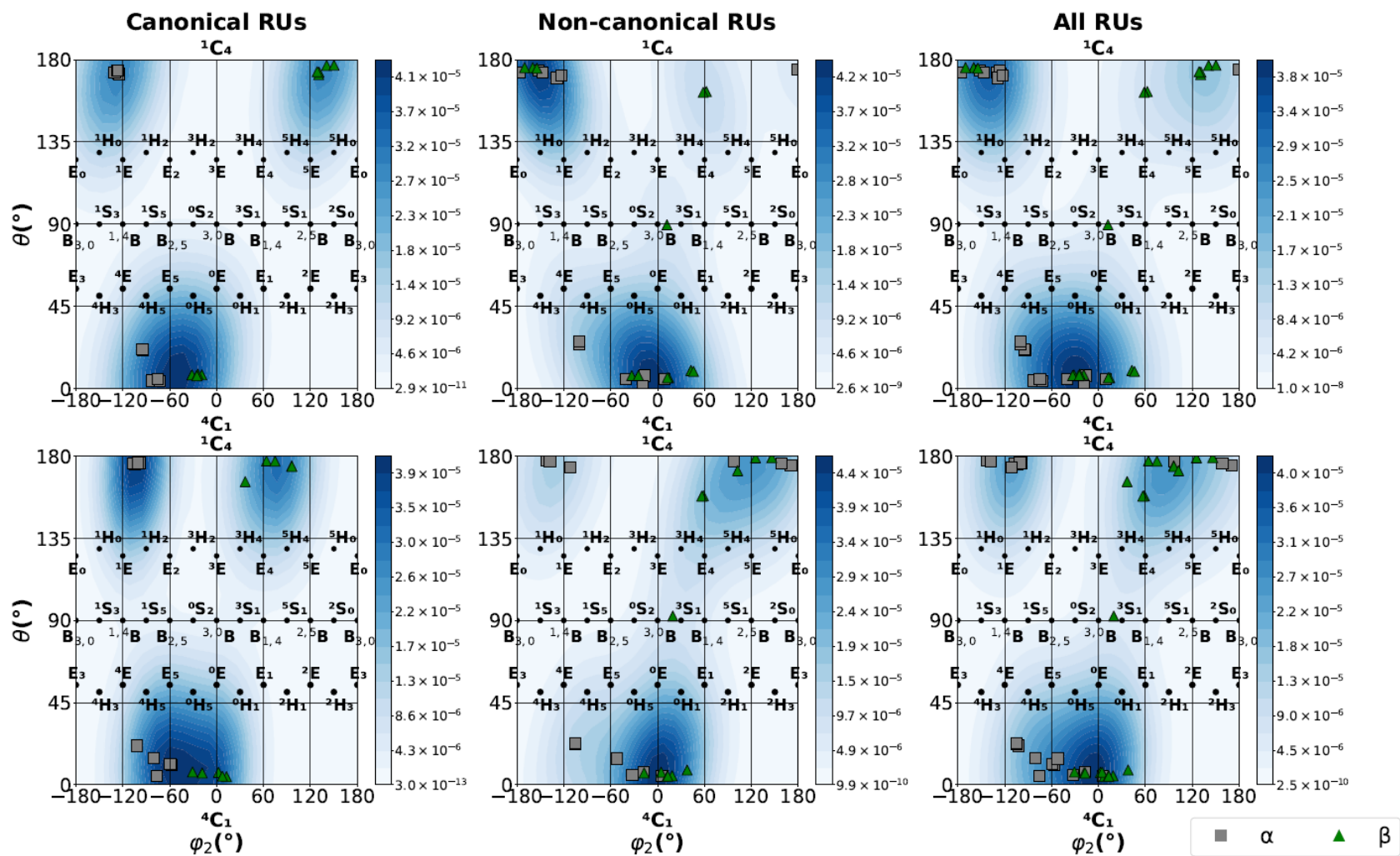
Figure 3.30 Mercator projection of the kernel density probability and scatter plot for the distribution of the values for the phase angles ϕ and θ in (°) for the puckering of the 6-MR β- and α-Rib. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

$\theta = 135^o$) in both environments. $^4C_1$ is one of the preferential puckering conformations of 6-MRs. In the case of the α-anomers they are mostly found in the $^3H_4$ ($\varphi_2 = 30^o$ and $\theta = 135^o$) - ^5E ($\varphi_2 = 120^o$ and $\theta = 135^o$) and in the ^4E ($\varphi_2 = -120^o$ and $\theta = 45^o$) - $^0H_5$ ($\varphi_2 = -30^o$ and $\theta = 45^o$) puckering conformations.

Figure 3.30 shows a similar plot to the previous one but for the canonical and non-canonical Rib nucleosides. In this case both anomers of all nucleosides in both environments populate the $^4C_1$ region. The α-anomers also can populate in a certain extend the $^1H_0$ ($\varphi_2 = -150^o$ and $\theta = 135^o$) - E_2 ($\varphi_2 = -60^o$ and $\theta = 135^o$) ring conformations.

3.4.6 Computational modeling of the selection of T for DNA and U for RNA

A well known difference between DNA and RNA is the substitution of T in the first one by U in the ribonucleic acids [128]. The genetic misplace of U in DNA is common since the DNA polymerase enzyme responsible for the replication of DNA does not discriminate between T and U [140].

The substitution of T by U in DNA is a genetic error that can appear in DNA as a product of the deamination (type of genetic damage) of CMP or by replacing T by U during the replication process [140, 141]. The first mechanism can generate U:G complementary base pairing which is stable. There are repairing and prevention mechanisms to correct this error.

The incorporation of U in DNA can be prevented by dUTP nucleotidohydrolase (dUTPase) or repaired by uracil-N-glycosylase (UDG) [140].

Uracil-excision involves the participation of UDG that deletes U from the genetic code [142]. If this error is not fixed eventually it will become a genetic mutation after two replications [143].

dUTPase on the other hand, hydrolyzes dUTP into dUMP and pyrophosphate to reduce the availability of dUTP for DNA polymerase in the replication process [141].

Even when the previously discussed repair and prevention mechanism exist there are bacteria and viruses that contain a genetic code with U-DNA, e.g., in the **P**hosphate **B**uffered **S**aline (PBS) from *Bacillus subtilis* [144]. This may suggest the possibility that maybe the first DNAs contained a mixture of U and T nucleotides.

Poole and coworkers (Pcw) [145] discussed the replacement of U by T in DNA on the basis

of the evolutionary substitution of one base by the other by proto-enzymes. Pcw argued that this may have been a way for Nature to discriminate and repair cytosine deamination, otherwise if the genetic code would have U it would has been impossible to separate U that comes from the transformation of C.

Lesk and coworkers (Lcw) [146] further discussed this matter on the basis of the higher photostability and resistance to photodamage by UV radiation for T compared to U. Hence, Nature may have chosen T to preserve the genetic information.

In addition to the Pcw and Lcw studies, the following question still remains: could we explain the selection of T by DNA and U by RNA based in the thermodynamics (energetic stability) of the T and U nucleosides? A confirmation of this hypothesis could mean that beyond considering UV-light as a selection factor in early times and if everything started with RNA in a protein-free prebiotic scenario the energetic stability of the building blocks of T and U may have contributed to their incorporation later on in the biopolymeric nucleic acids.

In order to answer the previous posed question thermodynamic quantities are estimated at the B3LYP/6-311++G (d, p) level for a sugar-exchange reaction between the β- and α-anomers of the "correct" uridine (rU) and D-2'-(deoxy)thymidine (dT) and the "wrong" D-2'-(deoxy)uridine (dU) and thymidine (rT) in the furanoses 2dRibf, Ribf and pyranoses 2dRib and Rib (see **Figure 3.31**).

Similar to the theoretical study published in 2022 [57] two different types of sugar-exchange reactions are considered:

$$\beta - \mathrm{rT} + \beta - \mathrm{dU} \rightarrow \beta - \mathrm{dT} + \beta - \mathrm{rU} \tag{3.20}$$

and

$$\alpha - \mathrm{rT} + \alpha - \mathrm{dU} \rightarrow \alpha - \mathrm{dT} + \alpha - \mathrm{rU} \tag{3.21}$$

In addition to this, all possible combinations that include the initial sugar puckering conformations $^{2}T_{3}$ and $^{3}T_{2}$ for the furanoses and $^{1}C_{4}$ and $^{4}C_{1}$ for the pyranoses are considered. Hence, 64 sugar-exchange reactions were modeled in vacuum and implicit solvation respectively.

Table A9 in the Appendices section shows the estimated thermodynamic parameters for the different reactions and it can be noticed that for the nucleosides with the sugar in the F-form conformations all ΔG° values are within the intrinsic error for the DFT method used not given a clear picture on which nucleosides pair is preferred. Meanwhile for the T and U nucleosides with a P-form some exceptions are found with $\Delta G^{\circ} < 0$ (correct pair favored) and $\Delta G^{\circ} > 0$ (incorrect pairs preferred). In vacuum the most negative values for ΔG° are obtained in the β-configuration for the reaction: β-$^4C_1$-dT + β-$^4C_1$-rU → β-$^1C_4$-rT + β-$^1C_4$-dU (ΔG° = -45.8 kJ/mol in vacuum) and the most positive ΔG° is obtained also for the β-configuration for the sugar exchange reaction: β-$^1C_4$-dT + β-$^1C_4$-rU → β-$^4C_1$-rT + β-$^4C_1$-dU (ΔG° = 45.7 kJ/mol). In aqueous environment the most negative values were obtained for the α-configuration and the reaction: α-$^1C_4$-dT + α-$^1C_4$-rU → α-$^4C_1$-rT + α-$^4C_1$- dU (-54.1 kJ/mol), meanwhile the most positive values were observed for the reaction α-$^4C_1$-dT + α-$^4C_1$-rU → α-$^1C_4$-rT + α-$^1C_4$-dU (52.4 kJ/mol).

These results show that when considering the sugar ring puckering the sugar exchange reaction favours the "correct" dT, rU in the β-anomer in vacuum and the α-configuration in implicit solvation. *If there were proto-nucleosides with specific puckering for the pyranose rings in the prebiotic Earth they could have led through mutarotation to the correct dT and rU with furanose rings that eventually were incorporated in the RNA and DNA.*
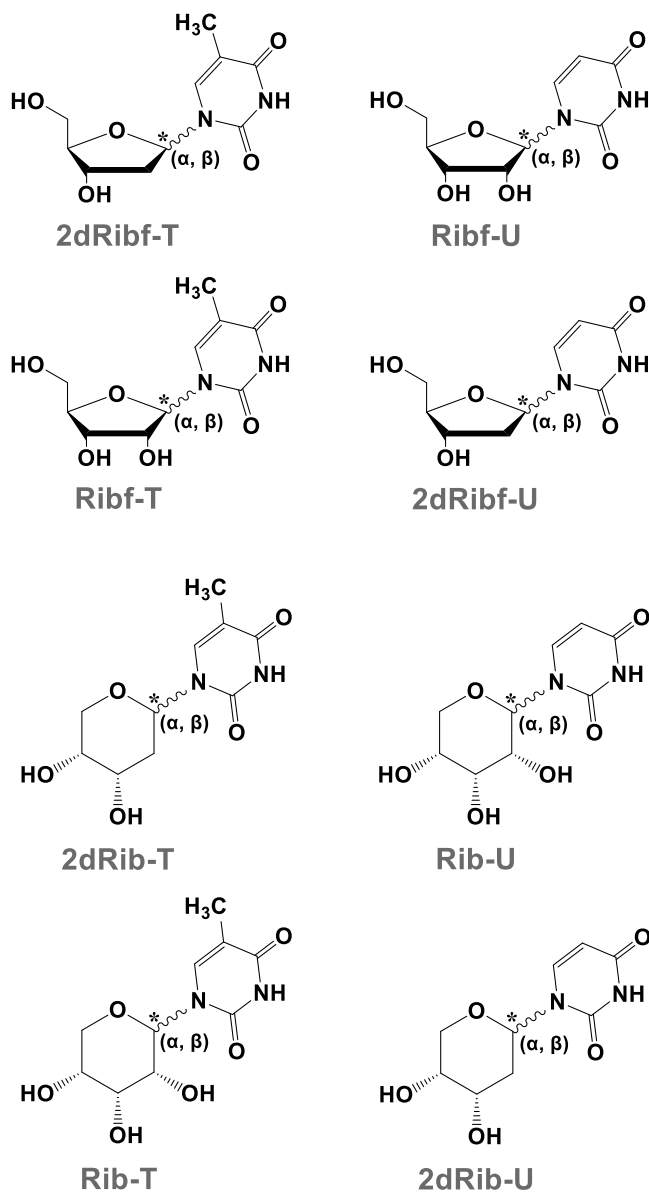
Figure 3.31 Natural and un-natural T, U nucleosides. The predominant form occurring in nature is 2'-deoxythymidine (dRibf-T) and uridine (Ribf-U) occurring in DNA and RNA respectively. The minor forms, *i.e.*, thymidine (Ribf-T) 2'-deoxyuridine (dRibf-U) which are not normally incorporated in RNA and DNA, respectively. The four figures at the bottom represent the nucleosides but with the pyranose ring (2dRib-T, Rib-U, Rib-T and 2dRib-U), also un-natural. See text and **Table A9**. (The star (*) denotes the anomeric center (C1') of the sugar).

3.4 Conclusions

In the present chapter the thermodynamic plausibility for the selection of the β- over the α-anomers of canonical and non-canonical nucleosides has been studied using computational modeling at the DFT-B3LYP/6-311++G (d, p) level of theory.

When comparing the thermodynamic stabilities of each 5-MR in their different puckering conformations with their pyranose-counterparts it is predicted that overall, the pyranose ring is more stable in both the β- and the α-anomer even when different puckering conformations were considered.

To compare the stabilities of both β- and α-anomers the pseudorotational equilibrium of the canonical and non-canonical nucleosides and the different anomer-exchange reactions were modeled. It can be concluded that overall there are not significant differences between both anomers or preferential displacement of the pseudorotational equilibrium for the nucleosides with the furanoses 2dRibf, Ribf and Tho. For the case of 2dRib and Rib there is a more clear preference of puckering in the pseudorotation for the β-$^4C_1$ and the α-$^1C_4$ conformations. In the case of the Rib nucleosides can be noticed that most of the nucleosides are more stable in their β-$^4C_1$ when comparing them with the α-counterpart.

In general, the differences in energies for the pseudorotational equilibrium of both furanoses and pyranoses-containing nucleosides suggested a slight flexibility in favor of the furanoses which could have provided these nucleosides with an evolutionary advantage.

When analyzing if the different nucleosides could be obtain through a "classic" prebiotic synthesis it was noticed that for sugar-containing nucleosides it is predicted that the formation of the β-anomer of some canonical and mostly non-canonical nucleosides containing TAP-C^4, BA-C^4 and MM is favored.

When analyzing the synthesis of non-sugar containing nucleosides it is observed that the formation of nucleosides containing glycerol and AEG are estimated as the most favorable to be synthesized across the board. This result suggests the possibility that maybe a "RNP or GNA world" preceded DNA and RNA reinforcing the idea of a proto-RNA with TCs different from ribofuranose or 2'-deoxyribofuranose.

When analyzing the preferential conformations of the different nucleosides it is concluded that most of their β- and α-anomers prefer the *syn* conformation around the furanose sugars,

meanwhile when a 2dRib is present in vacuum the *β*-anomer prefers the *anti*-conformation meanwhile the *α*-anomer is found more frequently in the *syn-* conformation. The opposite trend is found for the Rib.

An analysis using circular statistics of the distribution of the different sugar puckering conformations across the different nucleosides did not reveal a clear picture on which are the preferential puckering conformations in either canonical or non-canonical nucleosides.

The estimation of the thermodynamics for a sugar-exchange reaction between the right (dT and rU) and wrong (rT and dU) nucleosides revealed that when a 6-MR sugar is considered the combination of specific sugar puckering conformations favored the correct and incorrect pairs for the *β*-anomer in vacuum and for the *α*-configuration in aqueous solution. This suggests the idea that maybe the selection of T for DNA and U for RNA came from proto-building blocks with a P-form.

It is important to clarify that the results presented in this chapter complement the theoretical studies previously published in the literature and constitute the initial step for a more complete theoretical understanding of the forces, specifically thermodynamics that may have determined the chemical nature of today's nucleic acids. Nevertheless we cannot disregard the relevance of other factors. Amount these factors are kinetics [147, 148, 149], pH [150], the catalytic effect of ions like Mg^+ [151] and BO_3^{-3} [152, 2, 127, 153, 4, 104], different solvents, e.g., formamide [154], urea [126, 106] and deep eutectic solvents [155]. Including solvation explicitly may also be tested since water molecules may create important interactions with the solutes that could influence the thermodynamics and kinetics of the formation of the building blocks of nucleic acids.

Nevertheless, the present study has shed a light on a more broader question on how Nature chose the building blocks of today's nucleic acids and provides theoretical evidence that supports the idea that the first nucleic acids may have been different than the contemporaneous ones with a variety of TCs, sugar puckering conformations and that the TAP and BA are suitable bases to be considered in the formation of potential prebiotic C-nucleosides.

3.5 References

[1] S. Ayukawa, T. Enomoto and D. Kiga, "RNA world," in *Astrobiology: From the Origins of Life to the Search for*, Singapore, Springer Nature, 2019, pp. 77-90.

[2] M. Neveu, H. J. Kim and S. A. Benner, "The ''strong'' RNA world hypothesis: fifty years old," *Astrobiology,* vol. 13, pp. 391-403, 2013.

[3] L. E. Orgel, "Prebiotic chemistry and the origin of the RNA world," *Critical Reviews in Biochemistry and Molecular Biology,* vol. 39, pp. 99-123, 2004.

[4] W. D. Fuller, "Studies in prebiotic synthesis VI. Synthesis of purine nucleosides," *Journal of Molecular Biology,* vol. 67, pp. 25-33, 1972.

[5] N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, "The origin of RNA and "my grandfather's axe"," *Chemistry & Biology,* vol. 20, pp. 466-474, 2013.

[6] A. do Nascimento Vieira, K. do Nascimento Vieira, W. F. Martin and M. Preiner, "The ambivalent role of water at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[7] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei and S. A. Benner, "Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: Borate as a multifaceted problem solver in prebiotic chemistry," *Angewandte Chemie International Edition,* vol. 55, pp. 15816-15820, 2016.

[8] G. F. Joyce and L. E. Orgel, "Prospects for understanding the origin of the RNA world," in *The RNA World*, 2 ed., New York, Cold Spring Harbor Press, 1999, pp. 49-78.

[9] N. V. Hud and F. A. L. Anet, "Intercalation-mediated synthesis and replication: a new approach to the origin of life," *Journal of Theoretical Biology,* vol. 205, pp. 543-562, 2000.

[10] A. E. Engelhart and N. V. Hud, "Primitive genetic polymers," *Perspectives in Biology,* vol. 2, article #a002196 (pp. 1-21), 2010.

[11] V. M. Kolb, J. P. Dworkin and S. L. Miller, "Alternative bases in the RNA world: the prebiotic synthesis of urazole and its ribosides," *Journal of Molecular Evolution,* vol. 38, pp. 549-557, 1994.

[12] B. J. Cafferty, D. M. Fialho and N. V. Hud, "Searching for possible ancestors of RNA: The self-assembly hypothesis for the origin of proto-RNA," in *Prebiotic chemistry and chemical evolution of nucleic Acids, nucleic acids and molecular biology*, vol. 35, C. Menor-Salván, Ed., Philadelphia, Pennsylvania: Springer, 2018, pp. 143-174.

[13] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy and N. V. Hud, "Spontaneous formation and base pairing of plausible prebiotic nucleotides in water," *Nature Communications,* vol. 7, article #11328, (pp. 1-8), 2016.

[14] B. J. Cafferty and N. V. Hud, "Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology," *Current Opinion in Chemical Biology,* vol. 22, pp. 146-157, 2014.

[15] B. J. Cafferty, I. Gállego, M. C. Chen, K. I. Farley, R. Eritja and N. V. Hud, "Efficient self-assembly in water of Long noncovalent polymers by nucleobase analogues," *Journal of the American Chemical Society,* vol. 135, pp. 2447-2450, 2013.

[16] D. M. Fialho, *Physical organic principles governing the spontaneous prebiotic emergence of proto-nucleic acids,* Georgia: Georgia Tech Library, 2019, pp. 1-3.

[17] B. J. Cafferty and N. V. Hud, "Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA," *Israel Journal of Chemistry,* vol. 55, pp. 891-905, 2015.

[18] C. Li, B. J. Cafferty, S. C. Karunakaran, G. B. Schuster and N. V. Hud, "Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: implications for the possible origins of RNA," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20091-20096, 2016.

[19] M. C. Chen, B. J. Cafferty, I. Mamajanov, I. Gállego, J. Khanam, R. Krishnamurthy and N. V. Hud, "Spontaneous prebiotic formation of a β-ribofuranoside that self assembles with a

complementary heterocycle," *Journal of the American Chemical Society,* vol. 136, pp. 5640-5646, 2014.

[20] A. Butlerow, "Formation synthetique d'une substance sucree," *Comptes Rendus de l'Academie des Sciences,* vol. 63, pp. 145-147, 1861.

[21] J. Sutherland, "Ribonucleotides," *Cold Spring Harbor Perspectives in Biology,* vol. 2, article # a005439 (pp. 1 - 13), 2010.

[22] K. N. Drew, J. Zajicek, G. Bondo, B. Bose and A. S. Serianni, "13C-labeled aldopentoses: detection and quantitation of cyclic and acyclic forms by heteronuclear 1D and 2D NMR spectroscopy," *Carbohydrate Research,* vol. 307, pp. 199-209, 1998.

[23] K. Sefah, Z. Yang, K. M. Bradley, S. Hoshika, E. Jiménez, L. Zhang, G. Zhu, S. Shanker, F. Yu, D. Turek, W. Tan and S. A. Benner, "In vitro selection with artificial expanded genetic information systems," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 111, pp. 1449-1454, 2014.

[24] E. Lescrinier, M. Froeyen and P. Herdewijn, "Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids," *Nucleic Acids Research,* vol. 31, pp. 2975-2989, 2003.

[25] P. Gu, G. Schepers, J. Rozenski , A. Aerschot and P. Herdewijn, "Base pairing properties of D- and L-Cyclohexene nucleic acids (CeNA)," *Oligonucleotides,* vol. 13, pp. 479-489, 2003.

[26] L. Orgel, "A simpler nucleic acid," *Science,* vol. 290, pp. 1306, 2000.

[27] K. U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, "Chemical etiology of nucleic acid structure: the α-threofuranosyl - (3' —>2') oligonucleotide system," *Science,* vol. 290, pp. 1347-1351, 2000.

[28] M. Hernández-Rodríguez, J. Xie, Y. M. Osornio and R. Krishnamurthy, "Mapping the landscape of potentially primordial informational oligomers: (30/20) - D - phosphoglyceric

acid linked acyclic oligonucleotides tagged with 2, 4 - disubstituted 5 - aminopyrimidines as recognition elements," *Chemistry: An Asian Journal,* vol. 6, pp. 1252-1262, 2011.

[29] E. Meggers and L. Zhang, "Synthesis and properties of the simplified nucleic acid glycol nucleic acid," *Accounts of Chemical Research,* vol. 43, pp. 1092-1102, 2010.

[30] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

[31] P. E. Nielsen, "Peptide Nucleic Acid (PNA). Implications for the origin of the genetic material and the homochirality of life," *AIP Conference Proceedings,* vol. 379, pp. 55-61, 1996.

[32] P. E. Nielsen, "Peptide nucleic acid (PNA): a model structure for the primordial genetic material?," *Origins of Life and Evolution of the Biosphere,* vol. 23, pp. 323-327, 1993.

[33] P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide," *Science,* vol. 254, pp. 1497-1500, 1991.

[34] M. Frenkel-Pinter, M. Samanta, G. Ashkenasy and L. J. Leman, "Prebiotic peptides: molecular hubs in the origin of life," *Chemical Reviews,* vol. 120, pp. 4707-4765, 2020.

[35] J. Wu, Q. Meng, H. Ren, H. Wang, J. Wu and Q. Wang, "Recent advances in peptide nucleic acid for cancer bionanotechnology," *Acta Pharmacologica Sinica,* vol. 38, pp. 798-805, 2017.

[36] E. J. Lee, H. K. Lim, Y. S. Cho and S. S. Hah, "Peptide nucleic acids are an additional class of aptamers," *RSC Advances,* vol. 3, pp. 5828-5831, 2013.

[37] E. Mateo-Martí and C.-M. Pradie, "A novel type of nucleic acid-based biosensors: the use of PNA probes, associated with surface science and electrochemical detection techniques," in *Intelligent and biosensors*, Rijeka, InTech, 2010, pp. 323-344.

[38] P. E. Nielsen, "Peptide nucleic acids and the origin of life," *Chemistry & Biodiversity,* vol. 4, pp. 1996-2002, 2007.

[39] V. Menchise, G. De Simone, T. Tedeschi, R. Corradini, S. Sforza, R. Marchelli, D. Capasso, M. Saviano and C. Pedone, "Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, pp. 12021-12026, 2003.

[40] D. M. Fialho, T. P. Roche and N. V. Hud, "Prebiotic syntheses of noncanonical nucleosides and nucleotides," *Chemical Reviews,* vol. 120, pp. 4806-4830, 2020.

[41] M. Gull, "Prebiotic phosphorylation reactions on the early earth," *Challenges,* vol. 5, pp. 193-212, 2014.

[42] P. R. Barry, A. A. Ajees and T. R. McDermott, "Life and death with arsenic," *Bioessays,* vol. 33, pp. 350-357, 2011.

[43] F. Wolfe-Simon, J. S. Blum, T. R. Kulp, G. W. Gordon, S. E. Hoeft, J. Pett-Ridge, J. F. Stolz, S. M. Webb, P. K. Weber, P. C. W. Davies, A. D. Anbar and R. S. Oremland, "A bacterium that can grow by using arsenic instead of phosphorus," *Science,* vol. 332, pp. 1163-1166, 2011.

[44] G. Ni, D. Yuqi, F. Tang, J. Liu, H. Zhao and Q. Chen, "Review of α-nucleosides: from discovery, synthesis to properties and potential applications," *RSC Advances,* vol. 9, pp. 14302-14320, 2019.

[45] M. Froeyen, E. Lescrinier, L. Kerremans, H. Rosemeyer, F. Seela, B. Verbeure, I. Lagoja, J. Rozenski, A. V. Aerschot, R. Busson and P. Herdewijn, *Chemistry: A European Journal,* vol. 7, pp. 5183-5194, 2001.

[46] J.-L. Guesnet, F. Vovelle, N. T. Thuong and G. Lancelot, "2D NMR studies and 3D structure of the parallel-stranded duplex oligonucleotide Acrm5-alpha-d(TCTAAACTC)-beta-d(AGATTTGAG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 29, pp. 4982-4991, 1990.

[47] J. Paoletti, D. Bazile, F. Morvan, J.-L. Imbach and C. Paoletti, "α-DNA VIII:thermodynamic parameters of complexes formed between the oligo-alphadeoxynucleotides: α-d(GGAAGG) and α-d(CCTTCC) and their complementary oligo-betadeoxynucleotides: β-d(CCTTCC) and β-d(GGAAGG) are different," *Nucleic Acid Research,* vol. 17, pp. 2693-2704, 1989.

[48] G. Lancelot, J.-L. Guesnet and F. Vovelle, " Solution structure of the parallel-stranded duplex oligonucleotide α-d(TCTAAAC)-β-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations," *Biochemistry,* vol. 28, pp. 7871-7878, 1989.

[49] G. Lancelot, J.-L. Guesnet, V. Roig and N. T. Thuong, "2D-NMR studies of the unnatural duplex α-d(TCTAAAC)-β-d(AGATTTG)," *Nucleic Acids Research,* vol. 15, pp. 7531-7547, 1987.

[50] C. Thibaudeau, P. Acharya and J. Chattopadhyaya, Stereoelectronic effects in nucleosides and nucleotides and their structural implications, Uppsala: Uppsala University Press, 2005.

[51] M. Miljkovic, Carbohydrates: synthesis, mechanisms, and stereoelectronic effects, 1 ed., New York: Springer, 2009.

[52] S. Gerber-Lemaire and P. Vogel, "Anomeric effects in pyranosides and related acetals," in *Carbohydrate chemistry: chemical and biological approaches*, vol. 35, A. P. Rauter and T. K. Lindhorst, Eds., Cambridge, RSC Publishing, 2009, pp. 13-32.

[53] D. E. Levy and P. Fügedi, Eds., The organic chemistry of sugars, 1 ed., Boca Ratón: Taylor & Francis Group, 2005.

[54] C. Thibaudeau, A. Foldesi and J. Chattopadhyaya, "The first experimental evidence for a larger medium-dependent flexibility of natural β-D-nucleosides compared to the α-D-nucleosides," *Tetrahedron,* vol. 53, pp. 14043-14072, 1997.

[55] C. Thibaudeau, J. Plavec and J. Chattopadhyaya, "Quantitation of the anomeric effect in adenosine and guanosine by comparison of the thermodynamics of the pseudorotational

equilibrium of the pentofuranose moiety in N- and C-nucleosides," *Journal of the American Chemical Society,* vol. 116, pp. 8033-8031 , 1994.

[56] D. Cremer and J. A. Pople, "A general definition of ring puckering coordinates," *Journal of the American Chemical Society,* vol. 97, pp. 1354-1358, 1975.

[57] L. A. M. Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article #e09657 (pp. 1-12), 2022.

[58] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, "Structure of a B-DNA dodecamer: conformation and dynamics," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 78, pp. 2179-2183, 1981.

[59] L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, "A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex," *Science,* vol. 270, pp. 1838-1841, 1995.

[60] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, "UCSF Chimera - a visualization system for exploratory research and analysis," *Journal of Computational Chemistry,* vol. 25, pp. 1605-1612, 2004.

[61] H. Inc, hyperChem release 7.0 for Windows: reference manual, Waterloo: Hypercube, Inc, 2002.

[62] R. Dennington, T. A. Keith and J. M. Millam, GaussView, version 5, Shawnee Mission: Semichem Inc, 2009.

[63] J. Lubczak and E. Mendyk, "Stable enol form of barbituric acid," *Heterocyclic communications,* vol. 14, pp. 149-154, 2008.

[64] L. A. Montero, "GRANADA Programme: Distribución aleatoria de moléculas alrededor de un sistema Poliatómico central (Random distribution of molecules around a central polyatomic system)," Faculty of Chemistry, University of Havana, 2019. [Online]. Available: http://karin.fq.uh.cu/mmh/.

[65] L. A. Montero, J. Molina and J. Fabian, "Water clusters for calculations of association energy," *International Journal of Quantum Chemistry,* vol. 79, pp. 8-16, 2000.

[66] L. A. Montero, A. M. Esteva, J. Molina, A. Zapardiel, L. Hernández, H. Márquez and A. Acosta, "A Theoretical Approach to Analytical Properties of 2,4-Diamino-5-phenylthiazole in Water Solution. Tautomerism and Dependence on pH," *Journal of the American Chemical Society,* vol. 120, pp. 12023-12033, 1998.

[67] J. J. P. Stewart, "MOPAC," Stewart Computational Chemistry, 2023. [Online]. Available: http://openmopac.net/.

[68] A. Klamt, "The COSMO and COSMO-RS solvation models," *Advance Review,* vol. 1, pp. 699-709, 2011.

[69] A. Klamt and G. Schüümann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *Journal of the Chemical Society, Perkin Transactions 2,* vol. 2, pp. 799-805, 1993.

[70] M. Orio, D. A. Pantazis and F. Neese, "Density Functional Theory," *Photosynthesis Research,* vol. 102, pp. 443-453, 2009.

[71] R. H. Hertwig and W. Koch, "On the parameterization of the local correlation functional. What is Becke-3-LYP?," *Chemical Physics Letters,* vol. 268, pp. 345-351, 1997.

[72] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A,* vol. 38, pp. 3098-3100, 1988.

[73] C. Lee, W. Yang and R. G. Parr, "Development of the Colle–Salvetti correlation energy formula into a functional of the electron density," *Physical Review B,* vol. 37, pp. 785-789, 1988.

[74] W. J. Hehre, L. Radom, J. A. Pople and P. Schleyer, Ab initio molecular orbital theory, New York: John Wiley & sons Ltd, 1986.

[75] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji , X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian , J. V. Ortiz, A. F. Izmaylov , J. L. Sonnenberg , D. Williams-Young , F. Ding, F. Lipparini , F. Egidi, J. Goings , B. Peng , A. Petrone , T. Henderson , D. Ranasinghe , V. G. Zakrzewski , J. Gao, N. Rega, G. Zheng , W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida , T. Nakajima , Y. Honda , O. Kitao , H. Nakai , T. Vreven , K. Throssell , J. A. J. Montgomery , J. E. Peralta , F. Ogliaro , M. J. Bearpark , J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari , A. P. Rendell , J. C. Burant, S. S. Iyengar , J. Tomasi , M. Cossi , J. M. Millam , M. Klene , C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, revision C.01, Wallingford CT: Gaussian Inc, 2019.

[76] S. Kaur, P. Sharma and S. D. Wetmore, "Can cyanuric acid and 2, 4, 6-triaminopyrimidine containg ribonucleosides be components of prebiotic RNA? Insights from QM calculations and MD simulations," *ChemBioChem,* vol. 20, pp. 1415-1415, 2019.

[77] S. Kaur, P. Sharma and S. D. Wetmore, "Structural and electronic properties of barbituric acid and melamine-containing ribonucleosides as plausible components of prebiotic RNA: implications for prebiotic self-assembly," *Physical Chemistry Chemical Physics,* vol. 19, pp. 30762-30771, 2017.

[78] J. E. Šponer, J. Šponer and M. Fuentes-Cabrera, "Prebiotic routes to nucleosides: a quantum chemical Insight into the energetics of the multistep reaction pathways," *Chemistry – A European Journal,* vol. 17, pp. 847-854, 2011.

[79] Y. Zhao and D. G. Truhlar, "Density functionals with broad applicability in chemistry," *Accounts of Chemical Research,* vol. 41, pp. 157-167, 2008.

[80] Y. Zhao and D. G. Truhlar, "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class

functionals and 12 other function," *Theoretical Chemistry Accounts,* vol. 120, pp. 215-241, 2008.

[81] L. Rao, H. Ke, G. Fu, X. Xu and Y. Yan, "Performance of several Density Functional Theory methods on describing hydrogen-bond interactions," *Journal of Chemical Theory and Computation,* vol. 5, pp. 86-96, 2009.

[82] L. Chan, G. R. Hutchison and G. M. Morris, "Understanding ring puckering in small molecules and cyclic peptides," *Journal of Chemical Information and Modeling,* vol. 61, pp. 743-755, 2021.

[83] B. Mennucci, "Polarizable continuum model," *Advanced Review,* vol. 2, pp. 386-404, 2012.

[84] J. Tomasi, "Selected features of the polarizable continuum model for the representation of solvation," *WIREs Computational Molecular Science,* vol. 1, pp. 855-867, 2011.

[85] V. Barone, M. Cossi and J. Tomasi, "A new definition of cavities for the computation of solvation free energies by the polarizable continuum model," *The Journal of Chemical Physics,* vol. 107, pp. 3210-3221, 1997.

[86] M. Cossi, V. Barone, R. Cammi and J. Tomasi, "Ab initio study of solvated molecules: A new implementation of the polarizable continuum model," *Chem. Phys. Lett,* vol. 255, pp. 327-335, 1996.

[87] S. Miertuš and J. Tomasi, "Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes," *Chemical Physics,* vol. 65, pp. 239-245, 1982.

[88] S. Miertuš, E. Scrocco and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chemical physics,* vol. 55, pp. 117-129, 1981.

[89] A. do Nascimento Vieira, K. Kleinermanns, W. F. Martin and M. Preiner, "The ambivalent role of water at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[90] P. M. Collins and R. J. Ferrier, "Preliminary matters - structures, shapes and sources," in *Monosaccharides: their chemistry and their roles in natural products*, West Sussex, John Wiley & Sons, 1995, pp. 4-59.

[91] A. V. Dass, T. Georgelin , F. Westall , F. Foucher , P. De Los Rios , D. M. Busiello , S. Liang and F. Piazza , "Equilibrium and non-equilibrium furanose selection in the ribose isomerisation network," *Nature Communications,* vol. 12, article # 2749 (pp. 1-10), 2021.

[92] W. Martin, J. Baross, D. Kelley and M. J. Russell, "Hydrothermal vents and the origin of life," *Microbiology,* vol. 6, pp. 805-813, 2008.

[93] G. Wachtershauser, "From volcanic origins of chemoautotrophic life to bacteria, Archaea and Eukarya," *Phylosophical Transactions of the Royal Society B,* vol. 361, pp. 1787-1808, 2006.

[94] F. C. Dobbs and K. A. Selph , "Thermophilic bacterial activity in a deep-sea sediment from the Pacific ocean," *Aquatic Microbial Ecology,* vol. 13, pp. 209-212, 1997.

[95] L. M. Azofra, M. M. Quesada-Moreno, I. Alkorta, J. R. Avilés-Moreno, J. J. López-González and J. Elguero , "Carbohydrates in the gas phase: conformational preference of D-ribose and 2-deoxy-D-ribose," *New Journal of Chemistry,* vol. 38, pp. 529-538, 2014.

[96] E. J. Cocinero, A. Lesarri, P. Écija, F. J. Basterretxea, J.-U. Grabow, J. A. Fernández and F. Castaño, "Ribose found in the gas phase," *Angewandte Chemie International Edition,* vol. 51, pp. 3119-3124, 2011.

[97] S. Pitsch, S. Wendeborn, B. Jaun and A. Eschenmoser, "Why pentose-and not hexose-nucleic acids?. Part VII. Pyranosyl-RNA ('p-RNA')," *Preliminary Communication,* vol. 76, pp. 2161-2183, 1993.

[98] J. E. Kilpatrick, K. S. Pitzer and R. Spitzer, "The thermodynamics and molecular structure of cyclopentane," *Journal of the American Chemical Society,* vol. 69, pp. 2483-2488, 1947.

[99] R. S. Squillacote, O. L. Sheridan, O. L. Chapman and F. A. L. Anet, "Spectroscopic detection of the twist-boat conformation of cyclohexane. Direct measurement of the free

energy difference between the chair and the twist-boat," *Journal of the American Chemical Society,* vol. 97, pp. 3244-3246, 1975.

[100] X. Wang and R. J. Woods, "Insights into furanose solution conformations: beyond the two-state model," *Journal of Biomolecular NMR,* vol. 64, pp. 291-305, 2016.

[101] C. Altona and M. Sundaralingam, "Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation," *Journal of the American Chemical Society,* vol. 94, pp. 8205-8212, 1972.

[102] M. Szczepaniak and J. Moc, "Conformational studies of gas-phase ribose and 2-deoxyribose by density functional, second order PT and multi-level method calculations: the pyranoses, furanoses, and open-chain structures," *Carbohydrate Research,* vol. 384, pp. 20-36, 2014.

[103] H. P. M. De Leeuw, C. A. G. Haasnoot and C. Altona , "Empirical correlations between conformational parameters in *β*-D-furanoside fragments derived from a statistical survey of crystal structures of nucleic acid constituents† full description of nucleoside molecular geometries in terms of four parameters," *Israel Journal of Chemistry ,* vol. 20, pp. 108-126, 1980.

[104] W. D. Fuller, R. A. Sanchez and L. E. Orgel, "studies in prebiotic synthesis. VII solid -state synthesis of purine nucleosides," *Journal of Molecular Evolution,* vol. 1, pp. 249-257, 1972.

[105] C. Pérez-Fernández, J. Vega, P. Rayo-Pizarro, E. Mateo-Marti and M. Ruiz-Bermejo, "Prebiotic synthesis of noncanonical nucleobases under plausible alkaline hydrothermal conditions," *Scientific Reports,* vol. 12, pp. Article#15140 (pp. 1-15), 2022.

[106] C. M. Salván, M. Bouza, D. M. Fialho, B. T. Burcar, F. M. Fernández and N. V. Hud, "Prebiotic origin of pre-RNA building blocks in a urea "warm little pond" scenario," *ChemBioChem,* vol. 21, pp. 3504-3510, 2020.

[107] P. Trinks, "Zur chemie der aminopyrimidine," ETH, Zürich, 1987.

[108] D. M. Fialho, K. C. Clarke, M. K. Moore, G. B. Schuster, R. Krishnamurthy and N. V. Hud, "Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for

the prebiotic synthesis of nucleosides," *Organic and Biomolecular Chemistry,* vol. 16, pp. 1263-1271, 2018.

[109] E. D. Clercq, "C-nucleosides to be revisited," *Journal of Medicinal Chemistry,* vol. 59, pp. 2301-2311, 2015.

[110] K. W. Wellinton and S. A. Benner, "A review: synthesis of aryl C-glycosides via the Heck coupling reaction," *Nucleosides, Nucleotides & Nucleic Acids,* vol. 25, pp. 1309-1333, 2006.

[111] M. A. Gonzalez, J. L. Jimenez Requejo, J. C. Palacios Albarran and J. A. Gabis Perez, "Facile preparation of C-glycosylbarbiturates and C-glycosylbarbituric acids," *Carbohydrate Research,* vol. 158, pp. 53-66, 1986.

[112] D. M. Fialho, T. Y. Roche and N. V. Hud, "Prebiotic syntheses of noncanonical nucleosides and nucleotides," *Chemical Reviews,* vol. 120, pp. 4806-4830, 2020.

[113] N. Lehman, "The RNA World: 4,000,000,050 years old," *Life,* vol. 5, pp. 1583-1586, 2015.

[114] C. W. J. Carter, "What RNA world? Why a peptide/RNA partnership merits renewed experimental attention," *Life,* vol. 5, pp. 294-320, 2015.

[115] J. E. Smith, A. K. Mowles, A. K. Mehta and D. G. Lynn, "Looked at life from both sides now," *Life,* vol. 4, pp. 887-902, 2014.

[116] G. Wächtershäuser, "The place of RNA in the origin and early evolution of the genetic machinery," *Life,* vol. 4, pp. 1050-1091, 2014.

[117] K. E. Nelson, M. Levy and S. L. Miller, "Peptide nucleic acids rather than RNA may have been the first genetic molecule," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, pp. 3868-3871, 2000.

[118] P. E. Nielsen, O. Buchardt, M. Egholm and B. Norden, "DNA-like double helix formed by peptide nucleic acid," *Nature,* vol. 368, pp. 561-563, 1994.

[119] M. Egholm, O. Buchardt, P. E. Nielsen and R. Berg, "Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone," *Journal of the American Chemical Society,* vol. 114, pp. 1895-1897, 1992.

[120] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

[121] T. Bose, G. Fridkin, C. Davidovich, M. Krupkin, N. Dinger, A. H. Falkovich, Y. Peleg, I. Agmon, A. Bashan and A. Yonath, "Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds," *Nucleic Acids Research,* vol. 28, p. 1815–1828, 2022.

[122] K. Świderek, S. Marti, I. Tuñón, V. Moliner and J. Bertran, "Peptide bond formation mechanism catalyzed by ribosome," *Journal of American Chemical Society,* vol. 137, pp. 12024-12034, 2015.

[123] M. Simonović and T. A. Steitz, "A structural view on the mechanism of the ribosome-catalyzed peptide bond formation," *Biochimica et Biophysica Acta,* vol. 1789, pp. 612-623, 2009.

[124] A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath and J. Karle, "The transition state for formation of the peptide bond in the ribosome," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 108, pp. 13327-13332, 2007.

[125] I. Suárez-Marina, Y. M. Abul-Haija, R. Turk-MacLeod, P. S. Gromski, G. J. T. Cooper, A. O. Olivé, S. Colón-Santos and L. Cronin, "Integrated synthesis of nucleotide and nucleosides influenced by amino acids," *Communications Chemistry,* vol. 2, article#28 (pp. 1-8), 2019.

[126] Y. Hirakawa, T. Kakegawa and Y. Furukawa, "Borate-guided ribose phosphorylation for prebiotic nucleotide synthesis," *Scientific Reports,* vol. 12, article#11828 (pp. 1-7), 2022.

[127] A. Ricardo, M. A. Carrigan, A. N. Olcott and A. Benner, "Borate minerals stabilize ribose," *Science,* vol. 303, p. 196, 2004.

[128] D. L. Nelson and M. M. Cox, "Nucleotides and nucleic acids," in *Lehninger principles of biochemistry*, New York, W.H. Freeman & Company, 2013, pp. 281-313.

[129] X. -J. Lu and W. K. Olson, "3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures," *Nucleic Acids Research,* vol. 31, pp. 5108-5121, 2003.

[130] V. A. Bloomfield, D. M. Crothers and I. J. Tinoco, Nucleic acids: structures, properties and functions, Sausalito: University Science Books, 2000.

[131] V. L. Murthy, R. Srinivasan, D. E. Draper and G. D. Rose, "A complete conformational map for RNA," *Journal of Molecular Biology,* vol. 291, pp. 313-327, 1999.

[132] D. J. Proctor, R. Kierzek and P. C. Bevilacqua, "Restricting the conformational heterogeneity of RNA by specific incorporation of 8-bromoguanosine," *Journal of the American Chemical Society,* vol. 125, pp. 2390-2391, 2003.

[133] J. E. Sokoloski, S. A. Godfrey, S. E. Dombrowski and P. C. Bevilacqua, "Prevalence of syn nucleobases in the active sites of functional RNAs," *Bioinformatics,* vol. 17, pp. 1775-1787, 2011.

[134] H. A. Taha, M. R. Richards and T. L. Lowary, "Conformational analysis of furanoside-containing mono- and oligosaccharides," *Chemical Reviews,* vol. 113, pp. 1851-1876, 2012.

[135] S. T. Rao, E. Westhof and M. Sundaralingam, "Exact method for the calculation of pseudorotation parameters P, [tau]m and their errors. A comparison of the Altona-Sundaralingam and Cremer-Pople treatment of puckering of five-membered rings," *Acta Crystallographica Section A,* vol. A37, pp. 421-425, 1981.

[136] H. Satoh and S. Manabe, "Design of chemical glycosyl donors: does changing ring conformation influence selectivity/reactivity?," *Chemical Society Reviews,* vol. 42, pp. 4297-4309, 2013.

[137] V. Kojić , M. Svirčev, S. Đokić, I. Kovačević , M. V. Rodić , B. Srećo-Zelenović , V. Popsavin and M. Popsavin , "Synthesis and antiproliferative activity of new thiazole hybrids

with [3.3.0]furofuranone or tetrahydrofuran scaffolds," *Journal of the Serbian Chemical Society,* vol. 88, pp. 467-479, 2023.

[138] S. Lyu, N. Beiranvand, M. Freindorf and E. Kraka, "The interplay of ring puckering and hydrogen bonding in deoxyribonucleosides," *The Journal of Physical Chemistry A,* vol. 123, pp. 7087-7103, 2019.

[139] H. Satoh and S. Manabe, "Design of chemical glycosyl dononrs: does changing ring conformation influence selectivity/reactivity?," *Chemical Society Reviews,* vol. 42, pp. 4297-4309, 2013.

[140] B. G. Vértessy and J. Tóth, "Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases," *Accounts of Chemical Research,* vol. 42, pp. 97-106, 2009.

[141] T. Visnes, B. Doseth, H. S. Pettersen, L. Hagen, M. M. L. Sousa, M. Akbari, M. Otterlei, B. Kavli, G. Slupphaug and H. E. Krokan, "Uracil in DNA and its processing by different DNA glycosylases," *Philosophical Transactions of the Royal Society B,* vol. 364, pp. 563-568, 2009.

[142] E. Seeberg, L. Eide and M. Bjoras, "The base excision repair pathway," *Trends in Biochemical Sciences,* vol. 20, pp. 391-397, 1995.

[143] S. Bhattacharya, "Natural antimutagens: A review.," *Research Journal of Medicine,* vol. 5, pp. 116-126, 2011.

[144] B. K. Duncan and H. R. Warner, "Metabolism of uracil-containing DNA: degradation of bacteriophage PBS2 DNA in Bacillus subtilis," *Journal of Virology,* vol. 22, pp. 835-838, 1977.

[145] A. Poole, D. Penny and B.-M. Sjöberg, "Confounded cytosine! Tinkering and the evolution of DNA," *Nature Reviews Molecular Cell Biology,* vol. 2, pp. 147-151, 2001.

[146] A. M. Lesk, "Why does DNA contain Thymine and RNA Uracil?," *Journal of Theoretical Biology,* vol. 22, pp. 537-540 , 1969.

[147] Y. A. Jeilani and M. T. Nguyen, "Autocatalysis in formose reaction and formation of RNA nucleosides," *The Journal of Physical Chemistry B,* vol. 124, pp. 11324-11336, 2020.

[148] P. Singh, A. Singh, J. Kaur and W. Holzer, "H-bond activated glycosylation of nucleobases: implications for prebiotic nucleoside synthesis," *RSC Advances,* vol. 4, pp. 3158-3161, 2014.

[149] Y. Sheng, H. D. Bean, I. Mamajanov and N. V. Hud, "Comprehensive investigation of the energetics of pyrimidine nucleoside formation in a model prebiotic reaction," *Journal of the American Chemical Society,* vol. 131, pp. 16088-16095, 2009.

[150] R. Krishnamurthy, "Role of pKa of nucleobases in the origins of chemical evolution," *Accounts of Chemical Research,* vol. 45, pp. 2035-2044, 2012.

[151] M. Gull, B. J. Cafferty, N. V. Hud and M. A. Pasek, "Silicate-promoted phosphorylation of glycerol in non-aqueous solvents: A prebiotically plausible route to organophosphates," *Life,* vol. 7, article #life7030029, (pp. 1-10), 2017.

[152] J. E. Šponer, R. Szabla, R. W. Góra, A. M. Saitta, F. Pietrucci, F. Saija, E. Di Mauro, R. Saladino, M. Ferus, S. Civišh and J. Šponer, "Prebiotic synthesis of nucleic acids and their building blocks at the atomic level – merging models and mechanisms from advanced computations and experiments," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20047-20066, 2016.

[153] N. G. Holm, "The significance of Mg in prebiotic geochemistry," *Geobiology,* vol. 10, pp. 269-279, 2012.

[154] R. Saladino, C. Crestini, S. Pino, G. Constanzo and E. Di Mauro, "Formamide and the origin of life," *Physics of Life Reviews,* vol. 9, pp. 84-104, 2012.

[155] I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, "DNA and RNA in anhydrous media: duplex, triplex, and G-quadruplex secondary structures in a deep eutectic solvent," *Angewandte Chemie International Edition,* vol. 49, pp. 6310-6314, 2010.

Appendices

Table A1. Final unique PM7 optimized conformations (n), number of conformers (n') that contributed $\geq 50\%$ to the partition function and their total contribution in % (Z) for the different α- and β-anomers of the canonical (2dRibf, Ribf) and non-canonical (2dRib, Rib, Tho, glycerol, glyceric acid and AEG: N-(2-aminoethyl)-glycine) TCs in vacuum and in implicit solvation using the COSMO model.

Component	n vac.[1]	n solv.[2]	n' vac.	n' solv.	Z vac.	Z solv.
α-$^2T_3$[9]-2dRibf[3]	19	75	2	9	57.49	51.29
α-$^3T_2$[10]-2dRibf	18	79	2	5	51.14	53.82
β-$^2T_3$[11]-2dRibf	29	90	4	11	57.58	52.91
β-$^3T_2$[12]-2dRibf	19	77	3	6	60.13	52.22
α-$^2T_3$[13]-Ribf[4]	33	160	3	13	54.40	52.22
α-$^3T_2$-Ribf	23	95	2	8	53.64	54.30
β-$^2T_3$-Ribf	35	131	5	10	51.49	53.00
β-$^3T_2$-Ribf	25	113	6	10	56.70	51.94
α-$^2T_3$-Tho[5]	10	18	1	2	88.38	65.97
α-$^3T_2$-Tho	10	27	2	3	63.14	57.93
β-$^2T_3$-Tho	6	14	1	1	72.51	55.77
β-$^3T_2$-Tho	8	16	2	2	66.62	64.04
α-$^1C_4$[14]-2dRib[6]	6	21	1	1	99.85	66.20
α-$^4C_1$[15]-2dRib	15	20	2	2	83.29	84.03
β-$^1C_4$[16]-2dRib	14	15	2	1	91.39	52.12
β-$^4C_1$[17]-2dRib	4	22	1	2	53.87	60.94
α-$^1C_4$-Rib[7]	10	14	2	3	62.65	52.86
α-$^4C_1$-Rib	18	23	1	1	83.73	54.19
β-$^1C_4$-Rib	20	20	2	4	66.96	57.29
β-$^4C_1$-Rib	11	30	2	2	56.25	57.86
Glycerol	45	97	5	9	54.02	53.94
Glyceric acid	48	115	6	9	53.81	52.02
AEG[8]	167	223	4	2	51.86	62.15

[1]Values of n, n' and Z in vacuum at the PM7 level. [2] Values of n, n' and Z in implicit solvent using the COSMO model at the PM7 level. [3]2dRibf = 2'-deoxyribofuranose. [4]Ribf = ribofuranose. [5]Tho = threofuranose. [6]2dRib = 2'-deoxyribopyranose. [7]Rib = ribopyranose. [8]AEG: N-(2-aminoethyl)glycine. [9] $^2T_3$ (South or C2'-endo-C3'-exo) and [10] $^3T_2$ (North or C2'-exo-C3'-endo) conformations for 5-MR sugars. [11]$^1C_4$ and [12]$^4C_1$ conformations for 6-MR sugars.

Table A2. Final unique PM7 optimized conformations (n), number of conformers (n') that contributed at least 50% to the partition function and their contribution in % (Z) for the canonical (A, G, C, T), non-canonical RUs (TAP, BA, MM) and their acetylated derivatives represented as RU letter-position-Ac.

Component	n vac.[1]	n solv.[2]	n' vac.	n' solv.	Z vac.	Z solv.
A[3]	1	2	1	1	100	51.48
A-N^9-Ac[12]	5	18	1	2	66.73	68.00
G[4]	1	3	1	2	100	69.52
G-N^9-Ac	12	31	2	4	60.18	61.32
C[5]	2	5	1	3	51.41	64.10
C-N^1-Ac	4	17	1	3	80.68	69.20
T[6]	3	4	2	2	68.75	53.35
T-N^1-Ac	10	19	2	3	51.90	62.61
U[7]-N^1-Ac	3	15	1	2	74.89	51.52
TAP[8]	1	3	1	2	100	68.84
TAP-C^5-Ac	4	38	1	5	99.99	53.13
TAP-N-Ac	21	87	2	6	55.73	52.94
BA[9]	3	3	1	2	58.94	78.43
BA- C^5-Ac	13	36	1	3	96.90	58.36
BA-N-Ac	13	29	1	4	79.64	55.37
MM[10]	2	4	1	2	57.74	53.40
MM-N-Ac	9	33	2	5	63.02	51.66
CA[11]-N^5-Ac	5	13	1	2	72.15	62.17

[1]Values of n, n' and Z in vacuum at the PM7 level. [2] Values of n, n' and Z in implicit solvent using the COSMO model at the PM7 level. [3]A = adenine. [4]G = guanine. [5]C = cytosine. [6]T = thymine. [7]U = uracil. [8] TAP = 2, 4, 6 triaminopyrimidine. [9]BA = barbituric acid. [10]MM = melamine. [11]CA = cyanuric acid. [12]Ac = acetyl group (-CH$_2$-COOH).

Table A3. Differences in energies between the canonical and non-canonical sugars with different sugar ring conformations in vacuum and in aqueous environment in kJ/mol (equations. 3.1-3.6) for all sugar ring/anomer exchange reactions and the equilibrium of pseudorotation (see **Figure 3.2**). Included differences are between the total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and the Gibbs energies $\Delta G°$ at NPT conditions. The listed results were obtained at the DFT-B3LYP/6-311++G(d,p) level. The IEFPCM solvation model has been used to generate the results incorporating aqueous solvation at the same level of theory.

	DNA & RNA			**p-DNA & p-RNA**			**TNA**		
	TC[1]	vac.[2]	solv.[3]	TC	vac.	solv.	TC	vac.	Solv.
ΔE_1	2dRibf	-11.9	-12.1	2dRib	-8.61	-1.41	Tho	-2.9	-10.0
$\Delta E_{1(ZPE)}$		-13.3	-13.5		-10.17	-2.96		-3.7	-10.9
$\Delta G_1°$		-9.5	-9.3		-6.36	0.93		-1.8	-8.4
ΔE_1	Ribf	-1.9	-0.4	Rib	0.6	5.97			
$\Delta E_{1(ZPE)}$		-3.0	-0.2		0.61	5.87			
$\Delta G_1°$		0.0	-0.3		0.89	6.21			
ΔE_2	2dRibf	22.5	15.8	2dRib	10.98	4.32	Tho	13.3	47.5
$\Delta E_{2(ZPE)}$		24.0	17.0		12.86	5.96		13.9	47.9
$\Delta G_2°$		19.5	13.2		8.62	2.13		12.0	46.4
ΔE_2	Ribf	4.8	13.8	Rib	-1.25	6.19			
$\Delta E_{2(ZPE)}$		6.1	14.2		-0.26	7.4			
$\Delta G_2°$		3.0	13.1		-2.68	4.13			
ΔE_3	2dRibf	2.1	1.2	2dRib	7.46	3.04	Tho	9.3	35.0
$\Delta E_{3(ZPE)}$		2.2	1.1		8.35	3.51		9.0	35.0
$\Delta G_3°$		2.1	1.0		6.53	2.51		9.7	34.5
ΔE_3	Ribf	-6.2	3.8	Rib	0.77	9.06			
$\Delta E_{3(ZPE)}$		-6.9	3.4		1.69	9.87			
$\Delta G_3°$		-4.0	4.8		-0.52	7.35			
ΔE_4	2dRibf	8.5	2.5	2dRib	-5.09	-0.13	Tho	1.1	2.5
$\Delta E_{4(ZPE)}$		8.6	2.4		-5.66	-0.51		1.2	2.1
$\Delta G_4°$		8.0	2.9		-4.27	0.55		0.5	3.6
ΔE_4	Ribf	9.1	9.5	Rib	-1.42	3.1			
$\Delta E_{4(ZPE)}$		9.9	10.6		-1.34	3.4			
$\Delta G_4°$		7.1	8.0		-1.27	2.99			
ΔE_5	2dRibf	10.6	3.7	2dRib	2.37	2.91	Tho	10.4	37.6
$\Delta E_{5(ZPE)}$		10.8	3.5		2.69	3.0		10.2	37.1
$\Delta G_5°$		10.1	3.9		2.26	3.06		10.2	38.0
ΔE_5	Ribf	2.9	13.3	Rib	-0.65	12.16			
$\Delta E_{5(ZPE)}$		3.1	14.0		0.35	13.27			

	DNA & RNA			p-DNA & p-RNA			TNA		
	TC[1]	vac.[2]	solv.[3]	TC	vac.	solv.	TC	vac.	Solv.
ΔG_5°		3.1	12.8		-1.79	10.34			
ΔE_6	2dRibf	20.4	14.6	2dRib	-3.52	-1.28	Tho	4.0	12.5
$\Delta E_{6(ZPE)}$		21.9	15.9		-4.51	-2.45		4.9	13.0
ΔG_6°		17.4	12.1		-2.09	0.38		2.4	12.0
ΔE_6	Ribf	11.0	9.9	Rib	2.02	2.87			
$\Delta E_{6(ZPE)}$		12.9	10.8		1.95	2.47			
ΔG_6°		7.0	8.3		2.16	3.22			

[1] TC = unspecified **T**rifunctional **C**onnector. [2] Differences in ΔE, $\Delta E_{(ZPE)}$, and ΔG° in vacuum at the DFT level. [3] Differences in ΔE, ΔE(ZPE), and ΔG° in solvent.

Table A4. Differences between the energies in kJ/mol for the most stable β- and α-anomers of the 2' deoxy (d) or ribonucleosides in vacuum and in aqueous environment (equations 3.1-3.6). Included differences are between: The total energies without ($\Delta E_{\beta\alpha}$) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies ΔG° at NPT conditions. All results are obtained from DFT (B3LYP/6-311++G(d,p)) calculations. The IEFPCM solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_1	A	-2.3	15.9	-1.6	1.6	28.9	-11.2	-2.0	1.3	2.1	12.1
$\Delta E_{1(ZPE)}$		-2.4	17.8	-1.7	1.7	29.9	-13.3	-2.6	0.4	1.0	10.6
ΔG°_1		-1.8	11.5	-0.6	1.3	27.5	-6.0	-0.7	2.6	6.3	15.6
ΔE_1	G	-2.4	0.0	0.6	2.1	20.5	-8.1	3.2	9.6	10.4	16.4
$\Delta E_{1(ZPE)}$		-2.6	0.0	0.7	2.0	21.1	-9.4	2.7	8.5	9.9	15.2
ΔG°_1		-1.7	0.0	0.7	2.5	20.0	-6.2	4.0	12.1	13.6	20.3
ΔE_1	C	-2.6	-9.0	-6.4	1.4	18.4	5.9	1.9	7.3	12.2	21.1
$\Delta E_{1(ZPE)}$		-2.7	-11.1	-7.0	1.5	19.0	5.9	1.7	7.1	11.3	18.0
ΔG°_1		-1.9	1.3	-4.6	0.7	17.7	6.0	1.2	6.8	13.7	26.5
ΔE_1	T	3.2	-5.1	-6.4	-0.3	16.6	6.0	10.2	20.2	19.5	28.0
$\Delta E_{1(ZPE)}$		1.7	-6.7	-6.8	-0.4	16.8	6.1	8.9	18.9	18.7	25.8
ΔG°_1		7.8	-0.8	-4.7	-0.3	16.8	6.0	13.4	22.8	22.0	31.3
ΔE_1	U	-2.6	-4.3	-6.6	1.9	17.8	5.9	11.2	18.6	21.2	28.4
$\Delta E_{1(ZPE)}$		-2.7	-6.0	-7.0	2.0	17.9	6.0	9.9	17.6	20.4	27.0
ΔG°_1		-2.4	0.7	-4.8	1.5	18.1	5.6	14.8	20.5	23.5	31.9
ΔE_1	TAP-C[5]	4.8	1.4	11.6	-6.1	14.9	4.0	21.0	28.6	21.7	24.8
$\Delta E_{1(ZPE)}$		4.0	1.0	9.8	-5.6	14.0	3.9	18.8	26.0	21.2	23.5
ΔG°_1		7.0	3.7	15.5	-7.2	19.2	5.3	25.1	33.5	22.5	26.4
ΔE_1	TAP-N	-1.7	0.0	-11.5	1.7	5.5	-0.0	8.9	16.4	5.1	11.8
$\Delta E_{1(ZPE)}$		-1.9	0.0	-11.4	0.8	4.1	0.1	7.9	15.5	6.2	12.0
ΔG°_1		-0.7	0.0	-12.1	3.1	9.9	-0.4	11.8	18.5	4.3	11.4
ΔE_1	BA-C[5]	-2.3	-11.3	2.4	5.5	11.5	4.9	3.2	15.7	13.7	14.4
$\Delta E_{1(ZPE)}$		-2.5	-13.4	1.8	5.3	12.0	5.0	1.3	14.5	12.9	12.1
ΔG°_1		-1.4	-6.2	3.3	6.3	10.8	4.9	7.2	17.8	16.3	17.6
ΔE_1	BA-N	3.1	-7.2	-2.2	1.6	-1.4	5.4	12.2	11.4	4.7	9.8
$\Delta E_{1(ZPE)}$		3.3	-8.8	-2.5	1.7	-2.5	5.5	11.6	10.9	4.1	7.9
ΔG°_1		1.9	-4.9	-1.7	1.4	0.5	5.4	14.2	11.5	5.8	12.7
ΔE_1	CA	-2.3	0.0	-2.6	7.6	8.1	5.6	14.7	14.9	6.4	11.8
$\Delta E_{1(ZPE)}$		-2.4	0.0	-2.8	9.1	8.4	5.7	14.2	14.1	5.9	10.1
ΔG°_1		-1.8	0.0	-2.1	3.5	8.0	5.5	16.5	17.3	7.6	14.5
ΔE_1	MM	-2.4	-0.8	-0.9	-1.4	8.2	5.9	4.0	22.5	21.0	26.2
$\Delta E_{1(ZPE)}$		-2.6	-2.8	-1.0	0.3	7.6	6.0	4.1	21.4	20.0	24.9
ΔG°_1		-1.7	5.1	-0.5	-6.6	10.6	5.8	1.4	25.4	23.3	28.8

	RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_2	A	7.5	2.3	-10.6	-6.3	-13.7	15.3	2.7	-3.8	-17.8	-14.5
$\Delta E_{2(ZPE)}$		6.9	0.6	-10.4	-6.4	-15.3	16.3	4.5	-1.6	-17.4	-13.5
ΔG_2°		10.3	7.8	-8.6	-4.2	-11.1	12.4	-2.2	-8.7	-21.2	-17.6
ΔE_2	G	14.6	9.5	-15.8	-3.4	-8.5	8.1	-3.3	-11.3	-19.5	-14.5
$\Delta E_{2(ZPE)}$		14.8	8.3	-15.7	-3.5	-9.7	8.2	-1.5	-9.7	-19.2	-13.9
ΔG_2°		17.0	14.2	-14.1	-1.7	-7.0	7.7	-7.1	-14.8	-21.8	-18.1
ΔE_2	C	21.6	17.6	-1.9	-1.2	4.3	8.9	-10.1	-12.6	-25.8	-23.2
$\Delta E_{2(ZPE)}$		22.4	17.9	-1.2	-1.2	3.5	8.9	-8.8	-10.7	-25.0	-21.4
ΔG_2°		21.6	19.0	-4.9	0.3	5.6	8.0	-11.2	-16.5	-27.0	-25.4
ΔE_2	T	13.2	6.9	-1.1	-1.3	3.2	8.0	-11.2	-20.0	-25.0	-26.9
$\Delta E_{2(ZPE)}$		15.0	8.4	-0.9	-1.3	2.6	8.7	-9.5	-18.1	-24.7	-23.7
ΔG_2°		6.3	2.7	-3.8	0.2	4.1	5.3	-15.3	-23.4	-25.4	-31.9
ΔE_2	U	12.7	6.9	-0.7	-4.2	3.2	6.9	-12.8	-18.6	-25.9	-26.1
$\Delta E_{2(ZPE)}$		14.5	8.3	-0.5	-4.2	2.6	7.6	-11.2	-17.1	-25.5	-24.1
ΔG_2°		4.2	3.2	-2.9	-3.2	3.9	4.2	-16.1	-22.0	-26.5	-29.4
ΔE_2	TAP-C[5]	-1.1	-1.8	-16.0	-5.2	-9.0	12.0	-23.7	-29.2	-31.3	-32.8
$\Delta E_{2(ZPE)}$		0.7	-1.2	-17.1	-6.9	-8.6	11.9	-21.4	-26.5	-28.6	-29.6
ΔG_2°		-5.6	-3.8	-13.6	-2.5	-10.6	12.5	-27.6	-33.3	-35.8	-37.8
ΔE_2	TAP-N	4.8	1.9	1.3	-4.1	-3.2	1.9	-4.8	-14.1	-6.9	-3.8
$\Delta E_{2(ZPE)}$		6.1	1.8	1.6	-3.5	-2.8	1.1	-3.3	-12.8	-7.2	-3.6
ΔG_2°		1.7	1.8	1.8	-4.6	-6.1	3.8	-7.9	-16.6	-7.4	-5.2
ΔE_2	BA-C[5]	3.9	14.5	-12.8	-5.6	0.2	9.6	-9.5	-12.9	-22.9	-19.0
$\Delta E_{2(ZPE)}$		4.8	15.1	-12.5	-5.1	-0.3	9.8	-7.6	-11.1	-22.1	-17.2
ΔG_2°		1.7	14.4	-15.0	-5.2	0.6	8.8	-13.5	-14.0	-25.7	-21.0
ΔE_2	BA-N	17.0	13.9	-0.2	-2.4	0.3	10.1	-3.8	-10.7	-13.4	-18.8
$\Delta E_{2(ZPE)}$		18.5	15.2	0.9	-1.2	-0.1	10.6	-2.3	-8.7	-12.7	-16.2
ΔG_2°		14.3	12.7	-1.6	-4.4	1.1	8.7	-6.1	-13.5	-14.1	-22.8
ΔE_2	CA	16.3	12.2	-4.6	-9.3	-6.6	8.3	-8.7	-12.9	-16.8	-22.1
$\Delta E_{2(ZPE)}$		17.6	13.3	-3.3	-10.6	-8.4	8.6	-7.5	-11.7	-16.0	-20.6
ΔG_2°		14.2	10.6	-6.7	-4.6	-4.2	7.3	-10.6	-14.5	-17.9	-25.3
ΔE_2	MM	15.3	18.7	-8.0	6.7	7.6	-1.3	-9.2	-19.8	-26.1	-20.5
$\Delta E_{2(ZPE)}$		14.7	18.7	-8.0	4.5	7.4	0.2	-8.0	-18.3	-25.6	-19.8
ΔG_2°		19.6	20.3	-5.6	14.9	7.3	-5.2	-10.7	-22.2	-26.7	-21.4
ΔE_3	A	6.3	1.4	-3.8	-10.6	5.0	2.2	-9.3	-11.2	-24.5	-15.9
$\Delta E_{3(ZPE)}$		6.4	-0.1	-3.9	-12.8	4.4	3.0	-8.7	-10.2	-25.1	-16.9
ΔG_3°		7.0	6.8	-0.6	-4.1	8.9	1.6	-10.3	-13.9	-24.3	-14.4
ΔE_3	G	9.2	4.2	-10.9	-6.7	3.6	-4.1	-8.5	-9.6	-17.0	-13.2
$\Delta E_{3(ZPE)}$		9.2	2.7	-11.1	-8.5	3.0	-4.5	-7.6	-9.3	-17.1	-13.7
ΔG_3°		9.7	8.7	-7.1	-1.5	4.6	-3.8	-11.2	-11.0	-17.4	-13.6
ΔE_3	C	6.8	7.1	-1.9	4.6	13.4	3.9	-20.5	-16.1	-36.1	-24.6
$\Delta E_{3(ZPE)}$		7.0	5.2	-2.2	4.0	13.1	3.1	-20.4	-15.4	-36.4	-25.5
ΔG_3°		7.0	13.3	-2.7	7.5	14.2	5.8	-20.7	-18.5	-34.8	-22.9
ΔE_3	T	2.3	0.4	-2.6	-0.1	11.5	2.1	-19.5	-16.8	-24.6	-21.4
$\Delta E_{3(ZPE)}$		2.7	0.3	-3.0	-1.4	10.9	2.0	-18.8	-16.0	-24.7	-20.3
ΔG_3°		-0.1	1.1	-3.3	4.4	13.1	1.4	-21.4	-19.0	-24.0	-23.5
ΔE_3	U	2.2	0.3	-2.6	0.2	12.8	0.0	-20.4	-16.3	-25.1	-21.1
$\Delta E_{3(ZPE)}$		2.6	0.2	-3.0	-0.9	12.1	0.0	-19.8	-15.6	-25.2	-20.6
ΔG_3°		-1.5	0.7	-2.9	3.3	14.4	0.0	-21.9	-18.6	-24.4	-22.0
ΔE_3	TAP-C[5]	-8.3	0.0	-4.5	7.6	6.0	0.0	-20.6	-28.3	-22.4	-20.1
$\Delta E_{3(ZPE)}$		-6.4	0.0	-6.4	5.1	6.4	0.0	-19.4	-27.4	-21.1	-18.8
ΔG_3°		-13.6	0.0	-0.1	13.4	4.8	0.0	-23.3	-29.5	-23.4	-21.1

	RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_3	TAP-N	2.5	-1.3	-8.0	0.4	0.0	2.3	2.3	-4.3	-6.6	-2.0
$\Delta E_{3(ZPE)}$		2.5	-1.1	-8.7	-0.2	0.0	2.2	3.0	-3.9	-5.9	-1.7
ΔG_3°		2.6	-1.0	-5.8	1.1	0.0	2.7	0.6	-4.5	-8.1	-3.7
ΔE_3	BA-C[5]	11.4	2.7	-6.2	-5.6	0.0	0.0	-25.2	-15.8	-28.6	-17.8
$\Delta E_{3(ZPE)}$		11.6	1.2	-6.9	-7.5	0.0	0.0	-24.3	-15.5	-28.8	-18.0
ΔG_3°		9.7	7.5	-6.5	0.4	0.0	-0.0	-26.5	-15.3	-28.9	-17.1
ΔE_3	BA-N	5.7	5.2	-6.5	-1.7	1.5	0.0	-9.0	-14.9	-15.8	-17.1
$\Delta E_{3(ZPE)}$		6.3	5.1	-7.7	-2.6	1.4	0.0	-8.4	-13.9	-15.7	-15.4
ΔG_3°		4.2	7.0	-3.7	0.7	2.2	0.0	-10.2	-16.7	-15.2	-20.1
ΔE_3	CA	5.2	5.9	-7.4	2.2	2.8	0.0	-10.3	-13.3	-15.4	-15.4
$\Delta E_{3(ZPE)}$		5.7	5.7	-8.4	0.4	2.7	0.0	-9.9	-12.9	-15.1	-15.1
ΔG_3°		4.1	6.8	-5.3	6.9	3.4	0.0	-11.3	-14.0	-15.3	-16.0
ΔE_3	MM	10.3	14.7	-0.9	5.2	10.9	0.0	-3.6	-5.9	-18.3	-6.2
$\Delta E_{3(ZPE)}$		10.2	12.9	-1.9	3.9	9.8	0.0	-3.3	-5.6	-18.6	-7.1
ΔG_3°		11.2	22.1	3.7	9.2	13.0	0.0	-3.8	-6.1	-16.7	-4.0
ΔE_4	A	13.9	16.9	-8.3	5.8	10.2	1.9	10.0	8.6	8.8	13.5
$\Delta E_{4(ZPE)}$		14.5	18.5	-8.2	8.0	10.2	-0.0	10.6	9.0	8.8	14.0
ΔG_4°		14.1	12.4	-8.6	1.2	7.5	4.8	7.4	7.8	9.3	12.4
ΔE_4	G	5.4	5.3	-4.3	5.4	8.4	4.0	8.5	7.9	7.8	15.1
$\Delta E_{4(ZPE)}$		5.6	5.5	-3.9	7.0	8.5	3.3	8.7	8.1	7.8	15.0
ΔG_4°		6.8	5.4	-6.3	2.2	8.4	5.3	8.0	8.3	9.1	15.8
ΔE_4	C	12.5	1.5	-6.4	-4.5	9.3	10.9	12.4	10.8	22.5	22.6
$\Delta E_{4(ZPE)}$		13.0	1.5	-6.1	-3.7	9.4	11.8	13.3	11.8	22.7	22.2
ΔG_4°		12.9	6.9	-6.8	-6.4	9.1	8.1	10.7	8.8	21.5	24.0
ΔE_4	T	8.4	1.4	-5.0	-1.5	8.2	12.0	18.4	16.9	19.1	22.5
$\Delta E_{4(ZPE)}$		9.7	1.4	-4.8	-0.3	8.5	12.8	18.2	16.7	18.8	22.4
ΔG_4°		4.8	0.9	-5.2	-4.4	7.8	9.9	19.5	18.3	20.6	22.9
ΔE_4	U	7.9	2.3	-4.7	-2.5	8.2	12.7	18.8	16.3	20.3	23.3
$\Delta E_{4(ZPE)}$		9.2	2.1	-4.5	-1.4	8.4	13.6	18.5	16.2	20.1	23.4
ΔG_4°		3.8	3.1	-4.8	-4.9	7.7	9.7	20.5	17.1	21.4	24.5
ΔE_4	TAP-C[5]	10.3	-0.4	0.0	-18.9	-0.1	16.0	17.9	27.7	12.8	12.1
$\Delta E_{4(ZPE)}$		8.8	-0.1	-0.9	-17.5	-0.9	15.8	16.9	27.0	13.6	12.7
ΔG_4°		15.8	-0.0	2.0	-23.1	3.7	17.8	20.8	29.7	10.2	9.6
ΔE_4	TAP-N	-0.3	3.1	-2.1	-2.8	2.3	-0.5	1.9	6.6	4.8	10.0
$\Delta E_{4(ZPE)}$		0.9	2.9	-1.1	-2.5	1.4	-1.0	1.7	6.6	4.9	10.2
ΔG_4°		-3.4	2.8	-4.6	-2.6	3.8	0.7	3.2	6.4	5.0	9.9
ΔE_4	BA-C[5]	-2.7	0.5	-4.2	5.5	11.7	14.6	18.8	18.6	19.4	13.2
$\Delta E_{4(ZPE)}$		-2.8	0.5	-3.9	7.7	11.7	14.8	18.1	18.9	19.7	12.9
ΔG_4°		-1.1	0.6	-5.1	0.6	11.4	13.7	20.2	19.0	19.5	13.7
ΔE_4	BA-N	9.6	1.5	4.0	0.9	-2.6	15.5	17.4	15.6	7.0	8.1
$\Delta E_{4(ZPE)}$		10.3	1.4	6.1	3.2	-4.0	16.1	17.7	16.1	7.1	7.1
ΔG_4°		9.4	0.9	0.4	-3.7	-0.6	14.1	18.3	14.7	6.9	10.0
ΔE_4	CA	8.9	6.3	0.3	-3.9	-1.3	13.9	16.4	15.3	5.0	5.2
$\Delta E_{4(ZPE)}$		9.5	7.6	2.2	-1.9	-2.6	14.3	16.6	15.3	5.0	4.7
ΔG_4°		8.6	3.9	-3.5	-8.0	0.4	12.8	17.3	16.9	5.0	5.2
ΔE_4	MM	8.1	3.2	-8.0	0.1	5.0	4.6	-1.6	8.7	13.3	12.0
$\Delta E_{4(ZPE)}$		7.8	3.0	-7.1	0.9	5.1	6.2	-0.5	8.6	13.0	12.2
ΔG_4°		10.4	3.3	-9.8	-0.9	4.9	0.6	-5.5	9.3	13.3	11.4
ΔE_5	A	20.2	18.3	-12.1	-4.7	15.2	4.1	0.7	-2.5	-15.7	-2.4
$\Delta E_{5(ZPE)}$		20.9	18.4	-12.2	-4.8	14.6	2.9	1.9	-1.2	-16.3	-2.9
ΔG_5°		21.1	19.2	-9.3	-2.9	16.4	6.4	-2.9	-6.1	-15.0	-2.1

	RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_5	G	14.6	9.5	-15.2	-1.3	12.0	-0.0	-0.0	-1.7	-9.1	1.8
$\Delta E_{5(ZPE)}$		14.8	8.3	-15.0	-1.5	11.5	-1.2	1.1	-1.2	-9.3	1.3
ΔG_5°		16.5	14.2	-13.4	0.8	13.0	1.5	-3.1	-2.7	-8.3	2.1
ΔE_5	C	19.3	8.6	-8.3	0.1	22.7	14.8	-8.2	-5.3	-13.6	-2.0
$\Delta E_{5(ZPE)}$		20.0	6.8	-8.3	0.3	22.4	14.9	-7.1	-3.6	-13.8	-3.4
ΔG_5°		19.8	20.3	-9.5	1.0	23.3	13.9	-10.0	-9.7	-13.3	1.1
ΔE_5	T	10.8	1.8	-7.5	-1.6	19.8	14.1	-1.0	0.2	-5.5	1.1
$\Delta E_{5(ZPE)}$		12.4	1.7	-7.7	-1.7	19.4	14.8	-0.6	0.7	-5.9	2.1
ΔG_5°		4.6	2.0	-8.5	-0.0	20.9	11.3	-1.9	-0.7	-3.3	-0.7
ΔE_5	U	10.1	2.6	-7.2	-2.3	20.9	12.7	-1.6	-0.1	-4.8	2.2
$\Delta E_{5(ZPE)}$		11.8	2.4	-7.5	-2.3	20.5	13.6	-1.3	0.5	-5.1	2.8
ΔG_5°		2.3	3.8	-7.7	-1.6	22.0	9.7	-1.4	-1.5	-3.0	2.5
ΔE_5	TAP-C[5]	2.1	-0.4	-4.5	-11.3	5.9	16.0	-2.7	-0.6	-9.6	-8.0
$\Delta E_{5(ZPE)}$		2.4	-0.1	-7.3	-12.4	5.5	15.8	-2.5	-0.5	-7.4	-6.1
ΔG_5°		2.1	-0.0	1.9	-9.7	8.6	17.8	-2.5	0.2	-13.2	-11.5
ΔE_5	TAP-N	2.2	1.9	-10.1	-2.4	2.3	1.9	4.2	2.3	-1.7	8.0
$\Delta E_{5(ZPE)}$		3.4	1.8	-9.8	-2.7	1.4	1.3	4.7	2.7	-1.0	8.5
ΔG_5°		-0.7	1.8	-10.4	-1.5	3.8	3.4	3.8	1.9	-3.1	6.2
ΔE_5	BA-C[5]	8.7	3.2	-10.5	-0.1	11.7	14.6	-6.4	2.8	-9.2	-4.6
$\Delta E_{5(ZPE)}$		8.8	1.7	-10.8	0.2	11.7	14.8	-6.2	3.4	-9.1	-5.1
ΔG_5°		8.7	8.1	-11.6	1.0	11.4	13.7	-6.3	3.7	-9.4	-3.4
ΔE_5	BA-N	15.3	6.7	-2.5	-0.8	-1.1	15.5	8.4	0.7	-8.7	-9.1
$\Delta E_{5(ZPE)}$		16.6	6.5	-1.6	0.5	-2.6	16.1	9.3	2.2	-8.6	-8.2
ΔG_5°		13.6	7.9	-3.4	-3.0	1.6	14.1	8.1	-2.0	-8.2	-10.1
ΔE_5	CA	14.1	12.2	-7.1	-1.6	1.5	13.9	6.0	2.1	-10.4	-10.2
$\Delta E_{5(ZPE)}$		15.1	13.3	-6.1	-1.5	0.0	14.3	6.8	2.5	-10.1	-10.4
ΔG_5°		12.7	10.6	-8.8	-1.1	3.8	12.8	5.9	2.9	-10.4	-10.8
ΔE_5	MM	18.4	17.9	-8.9	5.3	15.9	4.6	-5.2	2.8	-5.0	5.7
$\Delta E_{5(ZPE)}$		18.0	15.9	-9.0	4.8	15.0	6.2	-3.9	3.1	-5.5	5.2
ΔG_5°		21.5	25.4	-6.1	8.3	17.9	0.6	-9.3	3.2	-3.4	7.4
ΔE_6	A	-1.2	-0.9	6.8	-4.2	18.7	-13.1	-12.0	-7.4	-6.7	-1.4
$\Delta E_{6(ZPE)}$		-0.6	-0.7	6.5	-6.3	19.7	-13.3	-13.2	-8.5	-7.8	-3.4
ΔG_6°		-3.3	-0.9	8.0	0.2	20.0	-10.8	-8.1	-5.2	-3.1	3.2
ΔE_6	G	-5.5	-5.3	4.9	-3.3	12.1	-12.2	-5.3	1.8	2.5	1.3
$\Delta E_{6(ZPE)}$		-5.6	-5.5	4.6	-5.0	12.6	-12.7	-6.1	0.4	2.1	0.2
ΔG_6°		-7.4	-5.4	7.0	0.2	11.6	-11.5	-4.1	3.9	4.4	4.5
ΔE_6	C	-14.8	-10.5	-0.0	5.9	9.1	-4.9	-10.5	-3.5	-10.3	-1.4
$\Delta E_{6(ZPE)}$		-15.4	-12.6	-0.9	5.2	9.6	-5.9	-11.6	-4.7	-11.4	-4.2
ΔG_6°		-14.6	-5.7	2.2	7.2	8.6	-2.2	-9.5	-2.1	-7.8	2.4
ΔE_6	T	-10.8	-6.5	-1.4	1.2	8.3	-6.0	-8.2	3.3	0.4	5.5
$\Delta E_{6(ZPE)}$		-12.3	-8.1	-2.1	-0.1	8.3	-6.7	-9.3	2.1	-0.1	3.4
ΔG_6°		-6.4	-1.6	0.5	4.2	9.1	-3.9	-6.1	4.5	1.4	8.4
ΔE_6	U	-10.5	-6.6	-1.9	4.4	9.6	-6.9	-7.6	2.3	0.9	5.0
$\Delta E_{6(ZPE)}$		-11.9	-8.1	-2.5	3.3	9.5	-7.6	-8.6	1.4	0.3	3.6
ΔG_6°		-5.6	-2.5	0.0	6.5	10.4	-4.2	-5.7	3.4	2.1	7.4
ΔE_6	TAP-C[5]	-7.1	1.8	11.6	12.8	15.0	-12.0	3.1	0.9	8.9	12.8
$\Delta E_{6(ZPE)}$		-7.2	1.2	10.7	11.9	15.0	-11.9	2.0	-0.9	7.6	10.8
ΔG_6°		-8.0	3.8	13.5	15.9	15.5	-12.5	4.2	3.8	12.4	16.7
ΔE_6	TAP-N	-2.3	-3.1	-9.4	4.4	3.2	0.4	7.0	9.9	0.3	1.8
$\Delta E_{6(ZPE)}$		-3.6	-2.9	-10.3	3.3	2.8	1.1	6.3	8.9	1.2	1.9
ΔG_6°		0.9	-2.8	-7.5	5.7	6.1	-1.1	8.5	12.1	-0.7	1.5

	RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_6	BA-C^5	7.5	-11.8	6.6	0.0	-0.2	-9.6	-15.6	-2.9	-5.6	1.2
$\Delta E_{6(ZPE)}$		6.8	-14.0	5.6	-2.4	0.3	-9.8	-16.7	-4.4	-6.8	-0.8
ΔG_6°		8.1	-6.9	8.5	5.7	-0.6	-8.8	-13.0	-1.2	-3.2	3.9
ΔE_6	BA-N	-11.3	-8.7	-6.3	0.7	1.2	-10.1	-5.2	-4.2	-2.4	1.7
$\Delta E_{6(ZPE)}$		-12.2	-10.2	-8.5	-1.5	1.5	-10.6	-6.1	-5.2	-3.0	0.8
ΔG_6°		-10.1	-5.7	-2.1	5.1	1.1	-8.7	-4.1	-3.2	-1.1	2.7
ΔE_6	CA	-11.1	-6.3	-2.8	11.5	9.4	-8.3	-1.7	-0.4	1.4	6.6
$\Delta E_{6(ZPE)}$		-11.9	-7.6	-5.0	11.0	11.0	-8.6	-2.4	-1.2	0.9	5.4
ΔG_6°		-10.1	-3.9	1.4	11.5	7.5	-7.3	-0.8	0.5	2.6	9.3
ΔE_6	MM	-5.0	-4.1	7.1	-1.5	3.2	1.3	5.7	13.8	7.8	14.2
$\Delta E_{6(ZPE)}$		-4.5	-5.8	6.0	-0.6	2.5	-0.2	4.6	12.7	7.0	12.7
ΔG_6°		-8.5	1.8	9.3	-5.6	5.6	5.2	6.9	16.2	10.0	17.3

[1]TC = unspecified **T**rifunctional **C**onnector. [2] Differences in ΔE, $\Delta E_{(ZPE)}$, and ΔG° in vacuum at the DFT level. [3] Differences in ΔE, $\Delta E_{(ZPE)}$, and ΔG° in solvent using the IEFPCM model.

Table A5. Gibbs energies (ΔG°) at standard pressure and temperature in kJ/mol for a hypothetical condensation reaction that follows the "classic" model leading to the 5 canonical and 6 non-canonical β- and α-nucleosides in vacuum and in aqueous environment modeled with the IEFPCM model. The free energies were estimated at the DFT-B3LYP/6-311++G(d,p) level of theory.

RU[1]	Vacuum				Solvent			
	DNA (2dRibf[12]-RUs)							
	$\alpha\text{-}^2T_3$[18]	$\beta\text{-}^2T_3$	$\alpha\text{-}^3T_2$[19]	$\beta\text{-}^3T_2$	$\alpha\text{-}^2T_3$	$\beta\text{-}^2T_3$	$\alpha\text{-}^3T_2$	$\beta\text{-}^3T_2$
A[2]	8.4	-0.8	-11.9	-5.8	2.0	-3.4	-18.8	-9.2
G[3]	-1.0	-3.5	-10.0	-11.1	-0.1	0.9	-9.4	-6.9
C[4]	0.5	2.5	-7.3	-2.4	6.4	12.2	-4.1	-0.1
T[5]	8.8	-4.5	1.0	-2.2	10.6	0.2	2.2	0.1
U[6]	9.1	-6.3	1.5	-2.7	11.1	1.1	1.2	1.5
TAP[7]-C^5	14.8	-10.3	-2.4	5.3	5.7	-11.2	-7.2	-10.2
TAP-N	7.6	-10.2	0.5	-10.8	-0.6	-11.9	-9.8	-9.9
BA[8]-C^5	-17.8	-35.7	-34.3	-43.3	-9.8	-8.6	-12.8	-15.1
BA-N	15.1	9.9	6.4	7.7	16.7	16.3	12.3	10.3
CA[9]	16.2	10.8	8.1	8.8	17.9	15.3	8.6	9.6
MM[10]	6.8	6.9	-4.6	-2.2	3.5	10.6	-10.9	-10.5
	RNA (Ribf[13]-RUs)							
	$\alpha\text{-}^2T_3$	$\beta\text{-}^2T_3$	$\alpha\text{-}^3T_2$	$\beta\text{-}^3T_2$	$\alpha\text{-}^2T_3$	$\beta\text{-}^2T_3$	$\alpha\text{-}^3T_2$	$\beta\text{-}^3T_2$
A	0.2	-11.5	0.9	-14.8	-1.3	-18.6	-2.9	-9.8
G	3.5	-13.7	2.9	-10.5	0.0	-14.8	-2.7	-8.5
C	-8.6	-16.5	-4.0	-17.8	10.2	-2.6	9.2	-5.3
T	4.2	-2.7	8.9	-3.3	15.6	2.8	15.6	3.1
U	3.6	-2.4	8.4	-3.4	18.4	2.2	16.6	3.6
TAP-C^5	12.5	-4.1	-3.0	-8.0	9.0	-6.6	15.8	-15.2
TAP-N	-3.5	-4.8	8.7	-3.0	2.7	-15.0	-0.7	-11.3
BA-C^5	-32.7	-50.7	-36.0	-48.1	-8.2	-26.5	-14.7	-22.1
BA-N	16.2	11.6	18.0	11.3	20.1	2.6	18.4	6.7
CA	20.5	10.7	22.6	12.0	26.8	9.1	23.0	7.0
MM	2.4	-6.3	2.9	-13.9	-8.7	-6.9	-2.4	-11.4
	TNA (Tho[14]-RUs)							
	$\alpha\text{-}^2\text{T}_3$	$\beta\text{-}^2\text{T}_3$	$\alpha\text{-}^3\text{T}_2$	$\beta\text{-}^3\text{T}_2$	$\alpha\text{-}^2\text{T}_3$	$\beta\text{-}^2\text{T}_3$	$\alpha\text{-}^3\text{T}_2$	$\beta\text{-}^3\text{T}_2$
A	18.8	-4.4	-10.6	-3.6	-5.7	-39.7	-8.1	-6.9
G	15.6	-3.4	-6.2	1.7	-2.4	-41.1	-4.6	-2.9
C	3.0	-3.4	-16.5	-7.9	7.8	-30.7	-6.6	-2.1
T	14.6	6.6	-4.1	3.1	12.9	-28.3	-1.5	4.8
U	15.4	7.3	-4.5	2.6	13.4	-28.8	-0.6	5.6
TAP-C^5	12.7	-10.0	-8.3	-5.1	-0.3	-34.3	-14.0	0.2
TAP-N	10.6	-7.5	-1.1	2.2	-1.6	-44.3	-9.7	-12.5
BA-C^5	-37.1	-48.5	-49.7	-38.8	-14.5	-52.1	-27.8	-17.6

	Vacuum				Solvent			
	α-$^2T_3$	β-$^2T_3$	α-$^3T_2$	β-$^3T_2$	α-$^2T_3$	β-$^2T_3$	α-$^3T_2$	β-$^3T_2$
BA-N	22.7	11.8	20.3	19.2	17.3	-20.5	3.5	14.0
CA	29.5	13.3	19.7	19.6	19.2	-19.9	5.3	14.6
MM	6.7	2.0	-5.7	-1.3	2.5	-49.2	-11.8	-14.8

<div align="center">p-DNA (2dRib⁽¹⁵⁾-RUs)</div>

	α-$^1C_4$[20]	β-$^1C_4$	α-$^4C_1$[21]	β-$^4C_1$	α-$^1C_4$	β-$^1C_4$	α-$^4C_1$	β-$^4C_1$
A	1.2	12.9	6.9	-3.9	1.0	8.2	2.6	-8.2
G	-0.1	12.2	10.2	-5.5	-0.8	6.9	10.4	-6.6
C	6.6	21.6	14.2	-5.6	15.1	23.3	20.9	2.3
T	2.5	26.2	22.2	-1.7	6.4	24.2	28.3	2.7
U	1.8	26.6	22.9	-1.8	7.0	23.5	26.6	2.4
TAP-C^5	-4.4	20.7	27.0	-9.2	-5.2	24.0	27.4	-8.0
TAP-N	-6.0	1.5	12.1	-4.5	-8.2	-2.3	9.4	-9.4
BA-C^5	-35.2	-10.7	-21.7	-43.8	-11.1	7.3	5.7	-10.4
BA-N	11.2	33.8	31.7	17.0	21.1	35.2	31.6	16.0
CA	11.7	33.2	34.6	15.4	17.1	33.4	33.5	16.9
MM	8.3	7.1	16.1	-3.2	-7.9	0.8	16.6	-7.8

<div align="center">p-RNA (Rib⁽¹⁶⁾-RUs)</div>

	α-$^1C_4$	β-$^1C_4$	α-$^4C_1$	β-$^4C_1$	α-$^1C_4$	β-$^1C_4$	α-$^4C_1$	β-$^4C_1$
A	3.5	14.1	8.9	-9.7	2.6	12.0	12.0	-9.8
G	3.5	13.9	16.2	-3.0	1.3	14.1	15.4	-6.9
C	0.5	23.3	13.3	-11.0	8.6	29.6	28.8	-0.7
T	5.9	27.8	27.0	4.4	12.2	32.1	37.3	1.2
U	5.4	28.1	28.0	4.2	10.5	32.0	36.2	2.7
TAP-C^5	2.4	13.9	24.1	-9.0	7.8	14.5	28.0	-14.0
TAP-N	-2.7	3.5	0.7	-4.1	-6.2	0.6	-1.1	-10.4
BA-C^5	-35.6	-14.9	-20.2	-43.2	1.4	12.1	12.8	-12.3
BA-N	29.9	38.1	34.9	23.5	32.4	39.3	38.8	11.9
CA	31.7	38.0	38.4	23.1	35.4	37.6	43.7	14.2
MM	-5.6	9.0	16.8	-7.2	-6.3	2.1	16.2	-9.3

<div align="center">GNA (glycerol-RU)</div>

A			-27.1				-28.6	
G			-25.7				-24.8	
C			-16.3				-17.8	
T			-16.8				-12.9	
U			-16.4				-11.6	
TAP-C^5			-19.1				-32.5	
TAP-N			-9.9				-9.3	
BA-C^5			-55.6				-42.8	
BA-N			-13.9				-13.3	
CA			-12.7				-8.9	
MM			-9.6				-19.7	

<div align="center">GNA (glyceric acid-RU)</div>

A			44.7				47.5	
G			46.0				48.2	

	Vacuum	_Solvent_
C	45.4	59.7
T	53.4	63.3
U	55.5	67.0
TAP-C^5	27.9	28.4
TAP-N	7.4	5.9
BA-C^5	-9.1	44.9
BA-N	90.0	84.7
CA	94.9	90.3
MM	27.5	21.3
aegPNA (AEG[17]-RU)		
A-N^9-Ac[11]	8.0	-19.4
G-N^9-Ac	13.8	-11.9
C-N^1-Ac	6.1	-22.3
T-N^1-Ac	9.2	-16.0
U-N^1-Ac	6.8	-16.9
TAP-C^5-Ac	15.5	-11.7
TAP-N-Ac	9.4	-14.0
BA-C^5-Ac	-7.0	-13.0
BA-N-Ac	15.7	-17.9
CA-N^5-Ac	4.4	-16.5
MM-N-Ac	1.8	-9.9

[1]RU = unspecified **r**ecognition **u**nit. [2]A = adenine. [3]G = guanine. [4]C = cytosine. [5]T = thymine. [6]U = uracil. [7]TAP = 2, 4, 6 triaminopyrimidine. [8]BA = barbituric acid. [9]MM = melamine. [10]CA = cyanuric acid. [11]Ac = acetyl group (-CH$_2$-COOH). [12]2dRibf = 2'-deoxyribofuranose. [13]Ribf = ribofuranose. [14]Tho = threofuranose. [15]2dRib = 2'-deoxyribopyranose. [16]Rib = ribopyranose. [17]AEG: N-(2-aminoethyl)glycine. [18] $^2T_3$ (C2'-endo-C3'-exo) and [19] $^3T_2$ (C2'-exo-C3'-endo) conformations for 5-MR. [20]$^1C_4$ and [21]$^4C_1$ conformations for 6-MR.

Table A6.   Values in (°) for the torsion angle χ that governs the RU conformation around the TC in the glycosidic bond. 2dRibf: 2'-deoxyribofuranose, Ribf: ribofuranose, Tho: threose, 2dRib: 2'-deoxyribopyranose and Rib: ribopyranose. The initial sugar ring conformations are for the nucleosides with 5-MR sugars: $^2T_3$ and $^3T_2$ and for the nucleosides with 6-MR sugars: $^1C_4$ and $^4C_1$.

RU[1]	2dRibf vac.[2]	2dRibf solv.[3]	Ribf vac.[2]	Ribf solv.[3]	Tho vac.[2]	Tho solv.[3]	2dRib vac.[2]	2dRib solv.[3]	Rib vac.[2]	Rib solv.[3]
	$\alpha\text{-}^2T_3$ [4]						$^1C_4\text{-}\alpha$ [5]			
A	94.4	294	173.6	300.3	93.9	296.3	299.9	300.5	104.8	113.1
G	285.7	103.8	171.0	301.7	91.8	296.9	300.4	294.8	100.1	112.0
C	287.9	294	164.3	303.5	285.2	292.9	296.5	297.5	153.9	145.8
T	288	294	163.9	303.3	283.9	293.8	116.3	118.1	145.6	138.9
U	288.4	294.8	163.6	304.6	284.6	294.0	117.0	119.6	146.5	140.0
TAP-C⁵	66.9	295.7	56.4	60.9	177.2	335.6	315.9	318.9	237.8	65.7
TAP-N	72.9	75.1	77.1	75.9	71.9	77.0	74.2	74.9	74.4	74.5
BA-C⁵	184.3	294.3	174.3	304.6	185.9	294.9	162.7	300.2	95.2	304.9
BA-N	292.3	295	307.3	306.2	287.3	293.6	114.1	299.5	122.5	122.0
CA	293.3	295.8	119.0	299.4	298.2	294.6	299.0	297.7	122.9	127.8
MM	296.5	296.3	309.8	85.3	296.3	295.7	303.8	84.1	83.9	84.0
	$\beta\text{-}^2T_3$ [4]						$^1C_4\text{-}\beta$ [5]			
A	47.7	42.8	56.9	56.5	62.5	190.8	128.1	285.4	142.2	145.9
G	46.3	47	57.6	57.0	169.6	59.9	135.5	284.7	141.7	143.6
C	60.3	58.5	183.6	62.8	58.9	57.9	84.1	283.8	122.8	194.2
T	237.6	241.7	179.7	61.1	58.0	184.7	87.6	284.3	127.4	129.6
U	233.8	239.1	180.0	61.3	57.9	184.7	86.3	284.5	127.9	129.5
TAP-C⁵	224.8	224.9	17.9	18.8	204.4	300.7	128.5	305.1	29.0	31.0
TAP-N	285.1	284.2	279.1	279.8	238.1	279.9	284.6	285.3	283.6	284.1
BA-C⁵	180.1	50.1	168.0	58.1	183.4	57.1	142.2	103.0	143.0	112.8
BA-N	246.3	237.8	60.3	61.8	240.0	251.0	260.9	268.4	72.9	256.6
CA	248.6	248.1	59.6	60.4	240.5	251.4	262.3	268.5	71.5	72.5
MM	56.3	60.4	55.9	60.5	50.9	275.7	277.3	278.8	274.7	276.1
	$\alpha\text{-}^3T_2$ [4]						$^4C_1\text{-}\alpha$ [5]			
A	293.1	294.1	201.6	300.4	191.7	294.6	224.1	276.4	216.5	217.2
G	295.2	103.8	300.8	301.7	186.9	295.7	275.1	226.1	215.7	289.6
C	288	133.6	178.8	303.6	177.9	293.2	221.2	284.7	223.2	231.8
T	287.7	125.6	179.2	304.3	179.5	294.3	220.4	232.0	289.0	232.1
U	288.2	128.1	178.9	304.5	179.4	294.2	219.7	284.0	289.2	291.4
TAP-C⁵	312.6	313.8	332.1	55.9	140.5	156.2	48.5	48.2	52.1	226.2
TAP-N	73	75.1	80.4	79.8	81.2	77.1	69.6	69.1	77.6	78.5
BA-C⁵	165.2	297	194.3	302.1	186.2	295.1	217.2	269.6	216.5	242.7
BA-N	293.6	113.8	306.9	306.9	114.4	294.0	289.6	284.4	291.7	293.2
CA	294.1	295.8	118.5	119.2	300.1	295.0	290.5	284.7	112.3	112.7
MM	297	83.5	309.7	310.1	171.1	296.1	215.5	210.2	210.4	209.8
	$\beta\text{-}^3T_2$ [4]						$^4C_1\text{-}\beta$ [5]			
A	51.6	70.4	162.1	70.6	107.8	88.3	201.1	256.5	153.8	139.8
G	53.3	70.2	169.8	70.1	190.5	192.8	246.4	254.6	184.9	183.8

RU[1]	2dRibf		Ribf		Tho		2dRib		Rib	
	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
C	64.3	198.4	184.7	190.9	194.5	196.7	199.6	237.0	202.8	204.8
T	242	245.4	184.3	186.0	195.1	197.6	208.6	239.2	203.9	241.6
U	240.1	241.8	184.4	186.5	195.1	184.7	206.2	239.5	203.8	238.5
TAP-C[5]	0.2	224.9	205.4	31.0	220.8	300.7	222.5	223.5	47.7	229.9
TAP-N	285.7	287.8	283.0	281.3	238.1	281.3	285.5	285.2	280.1	280.0
BA-C[5]	221.2	69.9	180.7	70.9	183.4	57.1	195.5	247.5	195.2	109.6
BA-N	241.5	251.5	250.7	254.7	238.0	251.0	252.4	252.3	101.2	251.9
CA	59.2	67.2	66.8	69.3	239.0	251.4	66.0	67.1	283.4	67.1
MM	63.3	276.6	190.6	191.9	201.2	275.7	272.9	275.8	200.9	203.7

[1] RU = unspecified **R**ecognition **U**nit. [2] χ in vacuum at the DFT level. [3] χ in implicit solvent at the DFT level using the IEFPCM model. [4]initial sugar ring puckering-anomer for 5-MR. [5]initial sugar ring puckering-anomer for 6-MR.

Table A7. Values for the Cremer-Pople (CP) puckering parameters ϕ_2 (phase angle) and Q (total puckering amplitude) that determine the sugar ring conformation for 5-MR. 2dRibf: 2'-deoxyribofuranose, Ribf: ribofuranose and Tho: threose.

CP[1]	RU[2]	2dRibf		Ribf		Tho	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]	vac.[3]	solv.[4]
				$\alpha\text{-}^2T_3$[5]			
ϕ_2	A	28.5	294.7	60.1	284.7	26.6	134.9
Q		0.3	0.4	0.3	0.4	0.3	0.4
ϕ_2	G	27.0	106.9	60.7	285.7	23.8	136.2
Q		0.2	0.3	0.3	0.4	0.4	0.4
ϕ_2	C	11.1	290.3	73.7	295.5	7.6	140.5
Q		0.2	0.4	0.3	0.4	0.2	0.4
ϕ_2	T	10.7	131.7	75.0	71.9	5.7	142.2
Q		0.2	0.3	0.3	0.2	0.2	0.4
ϕ_2	U	10.8	131.4	76.6	289.1	5.5	142.0
Q		0.2	0.3	0.3	0.4	0.2	0.4
ϕ_2	TAP-C[5]	17.7	150.4	253.6	247.5	60.9	164.0
Q		0.4	0.3	0.4	0.4	0.4	0.4
ϕ_2	TAP-N	51.0	293.6	305.4	37.5	51.4	285.6
Q		0.4	0.4	0.4	0.4	0.4	0.4
ϕ_2	BA-C[5]	70.3	136.9	100.2	126.4	285.1	141.2
Q		0.3	0.3	0.3	0.3	0.4	0.4
ϕ_2	BA-N	12.8	135.6	289.5	288.3	4.5	143.9
Q		0.2	0.3	0.4	0.4	0.2	0.4
ϕ_2	CA	13.0	136.6	288.2	323.3	293.5	145.1
Q		0.1	0.3	0.4	0.2	0.4	0.4
ϕ_2	MM	44.2	128.6	283.5	276.7	134.4	137.2
Q		0.2	0.3	0.4	0.4	0.4	0.4
				$\beta\text{-}^2T_3$[5]			
ϕ_2	A	97.9	96.2	78.7	76.7	124.4	268.0
Q		0.4	0.4	0.4	0.3	0.4	0.3
ϕ_2	G	96.6	94.0	73.3	73.7	125.2	119.5
Q		0.4	0.4	0.3	0.3	0.4	0.4
ϕ_2	C	97.6	95.1	57.1	71.7	123.8	119.0
Q		0.4	0.4	0.3	0.3	0.4	0.4
ϕ_2	T	340.6	339.2	68.4	68.8	121.4	132.9
Q		0.4	0.3	0.3	0.3	0.4	0.4
ϕ_2	U	0.7	5.8	68.5	68.7	121.3	130.8
Q		0.4	0.3	0.3	0.3	0.4	0.4
ϕ_2	TAP-C[5]	9.6	12.2	86.4	87.1	95.9	48.8
Q		0.4	0.4	0.3	0.3	0.4	0.4
ϕ_2	TAP-N	124.8	101.7	52.7	37.4	285.2	18.4
Q		0.4	0.3	0.4	0.4	0.4	0.4

CP[1]	RU[2]	2dRibf		Ribf		Tho	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]	vac.[3]	solv.[4]
				$\beta\text{-}^2T_3$[5]			
ϕ_2	BA-C^5	100.1	91.7	57.8	68.6	311.5	109.3
Q		0.4	0.4	0.4	0.3	0.4	0.4
ϕ_2	BA-N	84.3	91.3	63.6	318.7	116.7	315.4
Q		0.3	0.3	0.3	0.2	0.4	0.3
ϕ_2	CA	83.7	83.3	62.5	65.7	117.3	313.4
Q		0.3	0.3	0.3	0.3	0.4	0.3
ϕ_2	MM	96.3	94.1	72.9	74.0	115.1	115.5
Q		0.4	0.4	0.3	0.3	0.4	0.4
				$\alpha\text{-}^3T_2$[5]			
ϕ_2	A	111.9	123.0	294.0	284.3	280.7	290.1
Q		0.2	0.3	0.4	0.4	0.4	0.4
ϕ_2	G	121.7	106.9	289.6	284.3	283.6	290.6
Q		0.3	0.3	0.4	0.4	0.4	0.4
ϕ_2	C	15.2	266.8	311.9	289.7	296.5	139.9
Q		0.2	0.4	0.4	0.4	0.4	0.4
ϕ_2	T	12.7	270.7	309.7	288.2	298.4	141.0
Q		0.2	0.4	0.4	0.4	0.4	0.4
ϕ_2	U	12.6	268.6	308.7	288.5	298.5	141.3
Q		0.2	0.4	0.4	0.4	0.4	0.4
ϕ_2	TAP-C^5	239.9	231.2	258.8	243.1	224.0	165.9
Q		0.4	0.4	0.4	0.4	0.4	0.4
ϕ_2	TAP-N	64.0	293.6	235.2	263.2	254.5	286.1
Q		0.4	0.4	0.4	0.4	0.4	0.4
ϕ_2	BA-C^5	249.3	283.6	293.5	275.7	279.6	140.2
Q		0.4	0.4	0.4	0.4	0.4	0.4
ϕ_2	BA-N	21.9	294.2	290.5	287.9	292.0	143.0
Q		0.1	0.3	0.4	0.4	0.4	0.4
ϕ_2	CA	19.7	136.6	289.4	287.4	291.2	144.1
Q		0.1	0.3	0.4	0.4	0.4	0.4
ϕ_2	MM	126.5	292.7	288.1	285.1	281.1	136.7
Q		0.3	0.4	0.4	0.4	0.4	0.4
				$\beta\text{-}^3T_2$[5]			
ϕ_2	A	300.2	303.3	62.6	311.4	234.0	188.2
Q		0.3	0.3	0.4	0.4	0.3	0.3
ϕ_2	G	299.2	306.8	63.1	313.4	264.5	270.7
Q		0.3	0.3	0.4	0.4	0.4	0.3
ϕ_2	C	306.7	299.3	343.4	67.2	280.8	273.1
Q		0.2	0.3	0.4	0.3	0.4	0.3
ϕ_2	T	307.3	328.1	340.1	343.1	276.2	275.6
Q		0.3	0.3	0.4	0.4	0.3	0.3
ϕ_2	U	305.0	336.7	339.2	341.8	277.5	130.8
Q		0.3	0.3	0.4	0.4	0.3	0.3

CP[1]	RU[2]	2dRibf		Ribf		Tho	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]	vac.[3]	solv.[4]
				$\beta\text{-}^3T_2$ [5]			
ϕ_2	TAP-C^5	278.1	12.2	359.8	352.3	56.3	48.8
Q		0.3	0.4	0.4	0.4	0.4	0.4
ϕ_2	TAP-N	261.7	267.3	335.5	347.4	285.2	341.1
Q		0.3	0.3	0.4	0.4	0.4	0.4
ϕ_2	BA-C^5	54.6	320.8	347.0	330.6	311.5	109.3
Q		0.3	0.3	0.4	0.4	0.4	0.4
ϕ_2	BA-N	318.1	312.3	314.3	318.4	108.2	315.4
Q		0.3	0.3	0.4	0.3	0.3	0.3
ϕ_2	CA	319.3	311.1	316.2	318.3	109.2	313.4
Q		0.3	0.3	0.4	0.3	0.3	0.3
ϕ_2	MM	305.7	278.7	62.5	74.6	262.3	115.5
Q		0.2	0.3	0.4	0.4	0.4	0.4

[1]Cremer-Pople (CP) generalized puckering parameters for 5-MR. [2]RU = unspecified **R**ecognition **U**nit. [3]CP puckering parameters in vacuum at the DFT level. [4]CP puckering parameters in implicit solvent at the DFT level using the IEFPCM model. [5]initial sugar ring puckering-anomer for 5-MR.

Table A8. Values for the polar coordinates ϕ (zenithal angle), θ (azimuthal angle) and Q (total puckering amplitude) derived from the Cremer-Pople (CP) puckering parameters that determine the sugar ring conformation for 6-MR. 2dRib: 2'-deoxyribopyranose and Rib: ribopyranose.

CP[(1)]	RU[(2)]	2dRib		Rib	
		vac.[(3)]	solv.[(4)]	vac.[(3)]	solv.[(4)]
		$^1C_4$-α[(5)]			
ϕ	A	68.7	73.6	-126.1	-102.7
θ		162.3	162.0	172.5	175.8
Q		0.5	0.5	0.6	0.6
ϕ	G	73.8	122.6	-125.1	-104.1
θ		162.4	174.7	172.2	175.7
Q		0.5	0.6	0.6	0.6
ϕ	C	66.1	72.7	-130.4	-107
θ		160.0	159.4	173.0	176.1
Q		0.5	0.5	0.6	0.6
ϕ	T	178.5	-175.6	-126.9	-98.7
θ		174.2	175.0	174.1	176.2
Q		0.6	0.6	0.6	0.6
ϕ	U	179.5	-176.4	-126.4	-100.3
θ		174.2	174.9	174.2	176.2
Q		0.6	0.6	0.6	0.6
ϕ	TAP-C^5	161.1	-162.9	-128.8	-111.7
θ		176.4	174.9	170.0	173.7
Q		0.6	0.6	0.6	0.6
ϕ	TAP-N	157.3	163.6	-152.1	-142.3
θ		173.0	173.4	174.1	177.7
Q		0.6	0.6	0.6	0.6
ϕ	BA-C^5	-168.1	84.9	-122.7	170.6
θ		175.9	166.0	171.2	174.6
Q		0.6	0.6	0.6	0.5
ϕ	BA-N	91.5	77.3	-176.0	158.3
θ		170.5	162.7	172.9	175.8
Q		0.6	0.5	0.5	0.6
ϕ	CA	96.2	108.5	179.4	97.1
θ		172.7	173.3	174.5	177.2
Q		0.6	0.6	0.5	0.6
ϕ	MM	70.6	166.0	-146.7	-137.6
θ		157.9	173.3	172.8	177.1
Q		0.5	0.6	0.6	0.6

CP[1]	RU[2]	2dRib		Rib	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]
			$^1C_4$-β[5]		
ϕ	A	120.0	112.0	140.8	63.8
θ		169.0	167.5	176.9	177.3
Q		0.5	0.5	0.5	0.5
ϕ	G	120.2	112.4	150.2	74.7
θ		172.0	167.4	177.0	177.2
Q		0.5	0.5	0.5	0.5
ϕ	C	98.5	113.1	130.6	36.6
θ		157.1	164.6	171.9	166
Q		0.5	0.5	0.5	0.5
ϕ	T	102.4	111.5	130.9	96
θ		158.0	164.5	173.2	174.5
Q		0.5	0.5	0.5	0.5
ϕ	U	100.1	110.7	128.6	96.2
θ		157.9	164.5	173.4	174.3
Q		0.5	0.5	0.5	0.5
ϕ	TAP-C^5	150.2	140.8	12.1	19.5
θ		171.8	170.7	89.6	92.4
Q		0.5	0.6	0.7	0.8
ϕ	TAP-N	139.2	134.4	-159.7	146.1
θ		173.9	173.9	175.7	179.1
Q		0.5	0.6	0.5	0.6
ϕ	BA-C^5	139.1	121.5	-170.1	102.3
θ		174.1	164.5	175.7	172
Q		0.6	0.5	0.5	0.5
ϕ	BA-N	89.5	98.1	62.4	58.6
θ		157.1	157.7	162.6	158.2
Q		0.5	0.5	0.5	0.5
ϕ	CA	88.4	95.9	58.3	56.6
θ		156.7	157.3	162.2	158
Q		0.5	0.5	0.5	0.5
ϕ	MM	136.5	129.4	-154.4	125.2
θ		174.1	174.0	175.4	179
Q		0.6	0.6	0.5	0.6
			$^4C_1$-α[5]		
ϕ	A	-73.7	-75.5	-82.7	-76.1
θ		12.1	22.7	4.4	4.9
Q		0.5	0.5	0.5	0.5
ϕ	G	-76.3	-67.8	-72.2	-80.6
θ		22.7	13.6	4.2	14.6
Q		0.5	0.5	0.5	0.5
ϕ	C	-69.8	-95.9	-75.1	-57.1
θ		10.4	27.5	5.3	10.8
Q		0.5	0.5	0.5	0.5

CP[1]	RU[2]	2dRib		Rib	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]
			$^4C_1\text{-}\alpha^{[5]}$		
ϕ	T	-73.3	-64.0	-94.1	-59.4
θ		10.0	16.5	21.1	11.5
Q		0.5	0.5	0.5	0.5
			$^4C_1\text{-}\alpha^{[5]}$		
ϕ	U	-75.1	-93.7	-95.0	-101.8
θ		9.7	28.5	21.6	21.4
Q		0.5	0.5	0.5	0.5
ϕ	TAP-C^5	-31.9	-25.2	-16.1	-17.4
θ		7.9	8.2	7.1	7.0
Q		0.6	0.6	0.6	0.6
ϕ	TAP-N	-84.6	-78.1	9.1	5.8
θ		2.5	3.6	5.3	5.0
Q		0.5	0.5	0.6	0.6
ϕ	BA-C^5	-56.5	-65.9	-18.6	-52.1
θ		6.0	22.3	1.4	14.4
Q		0.5	0.5	0.6	0.5
ϕ	BA-N	-99.5	-94.1	-100.3	-105.4
θ		30.5	29.4	24.5	22.0
Q		0.5	0.5	0.5	0.5
ϕ	CA	-100.6	-93.9	-100.0	-105.4
θ		30.6	30.0	25.6	22.8
Q		0.5	0.5	0.5	0.5
ϕ	MM	-53.1	-50.3	-40.3	-32.4
θ		12.4	10.4	5.0	5.4
Q		0.5	0.5	0.5	0.6
			$^4C_1\text{-}\beta^{[5]}$		
ϕ	A	11.2	12.2	-18.9	2.4
θ		4.0	4.0	7.6	6.8
Q		0.6	0.6	0.6	0.6
ϕ	G	10.8	11.2	-31.5	-30.8
θ		4.7	3.8	7.5	7.0
Q		0.6	0.6	0.5	0.6
ϕ	C	19.2	7.1	-24.0	-18.5
θ		4.6	4.0	7.9	6.5
Q		0.6	0.6	0.6	0.6
ϕ	T	12.2	8.6	-25.5	12.6
θ		4.1	3.6	7.2	4.5
Q		0.6	0.6	0.6	0.6
ϕ	U	13.6	9.6	-25.2	7.4
θ		4.1	3.5	7.0	4.7
Q		0.6	0.6	0.6	0.6

CP[1]	RU[2]	2dRib		Rib	
		vac.[3]	solv.[4]	vac.[3]	solv.[4]
			$^4C_1$-β[5]		
ϕ	TAP-C5	3.8	-3.9	14.2	18.6
θ		4.7	4.4	5.9	5.0
Q		0.6	0.6	0.6	0.6
ϕ	TAP-N	0.1	-0.9	12.5	4.5
θ		5.5	5.5	6.0	6.9
Q		0.6	0.6	0.6	0.6
			$^4C_1$-β[5]		
ϕ	BA-C5	17.7	4.2	-33.1	37.8
θ		3.6	4.6	7.1	8.1
Q		0.6	0.6	0.5	0.6
ϕ	BA-N	8.5	18.5	42.6	12.5
θ		3.5	3.8	9.9	4.9
Q		0.6	0.6	0.6	0.6
ϕ	CA	16.5	20.2	45.9	13.1
θ		3.0	3.3	9.4	4.8
Q		0.6	0.6	0.6	0.6
ϕ	MM	2.5	1.0	-24.6	-17.2
θ		5.8	5.5	7.2	6.5
Q		0.6	0.6	0.5	0.6

[1]Polar coordinates derived from Cremer-Pople (CP) generalized puckering parameters for 6-MR. [2]RU = unspecified **R**ecognition **U**nit. [3]CP puckering parameters in vacuum at the DFT level. [4]CP puckering parameters in implicit solvent at the DFT level using the IEFPCM model. [5]initial sugar ring puckering-anomer for 5-MR.

Table A9. Differences between the energies of the canonical (predominant) nucleosides and their minor counterparts (see **Figure 3.31**) in vacuum and in aqueous environment (energies of the canonical form minus that of the minor form). Included differences are between: The total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies ($\Delta G°$). All energies are in kJ/mol and are obtained from DFT (B3LYP/6-31G(d,p)) calculations. The IEFPCM solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

Compared systems[1]	*furanoses*					
	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	$\Delta G°$	ΔE	$\Delta E_{(ZPE)}$	$\Delta G°$
ΔX[1](α-^{2}T$_3$-dT + α-^{2}T$_3$-rU - α-^{2}T$_3$-rT - α-^{2}T$_3$-dU)	-0.9	-0.9	-0.9	2.5	2.8	2.3
ΔX(α-^{2}T$_3$-dT + α-^{2}T$_3$-rU - α-^{2}T$_3$-rT - α-^{3}T$_2$-dU)	-7.3	-7.8	-5.6	2.2	2.4	2.1
ΔX(α-^{2}T$_3$-dT + α-^{2}T$_3$-rU - α-^{3}T$_2$-rT - α-^{2}T$_3$-dU)	-3.5	-3.6	-2.8	-1.7	-3.2	3.0
ΔX(α-^{2}T$_3$-dT + α-^{2}T$_3$-rU - α-^{3}T$_2$-rT - α-^{3}T$_2$-dU)	-9.9	-10.5	-7.5	-2.0	-3.6	2.7
ΔX(α-^{2}T$_3$-dT + α-^{3}T$_2$-rU - α-^{2}T$_3$-rT - α-^{2}T$_3$-dU)	5.7	6.1	3.9	0.6	0.8	0.8
ΔX(α-^{2}T$_3$-dT + α-^{3}T$_2$-rU - α-^{2}T$_3$-rT - α-^{3}T$_2$-dU)	-0.8	-0.8	-0.8	0.4	0.4	0.5
ΔX(α-^{2}T$_3$-dT + α-^{3}T$_2$-rU - α-^{3}T$_2$-rT - α-^{2}T$_3$-dU)	3.1	3.4	2.0	-3.6	-5.1	1.5
ΔX(α-^{2}T$_3$-dT + α-^{3}T$_2$-rU - α-^{3}T$_2$-rT - α-^{3}T$_2$-dU)	-3.3	-3.4	-2.7	-3.9	-5.5	1.2
ΔX(α-^{3}T$_2$-dT + α-^{2}T$_3$-rU - α-^{2}T$_3$-rT - α-^{2}T$_3$-dU)	1.5	1.6	0.8	7.6	9.4	3.1
ΔX(α-^{3}T$_2$-dT + α-^{2}T$_3$-rU - α-^{2}T$_3$-rT - α-^{3}T$_2$-dU)	-4.9	-5.2	-3.9	7.3	9.0	2.8
ΔX(α-^{3}T$_2$-dT + α-^{2}T$_3$-rU - α-^{3}T$_2$-rT - α-^{2}T$_3$-dU)	-1.1	-1.0	-1.1	3.3	3.5	3.8
ΔX(α-^{3}T$_2$-dT + α-^{2}T$_3$-rU - α-^{3}T$_2$-rT - α-^{3}T$_2$-dU)	-7.5	-7.9	-5.8	3.0	3.1	3.5
ΔX(α-^{3}T$_2$-dT + α-^{3}T$_2$-rU - α-^{2}T$_3$-rT - α-^{2}T$_3$-dU)	8.1	8.7	5.6	5.7	7.5	1.6
ΔX(α-^{3}T$_2$-dT + α-^{3}T$_2$-rU - α-^{2}T$_3$-rT - α-^{3}T$_2$-dU)	1.6	1.8	0.8	5.4	7.1	1.3
ΔX(α-^{3}T$_2$-dT + α-^{3}T$_2$-rU - α-^{3}T$_2$-rT - α-^{2}T$_3$-dU)	5.5	6.0	3.7	1.4	1.5	2.2
ΔX(α-^{3}T$_2$-dT + α-^{3}T$_2$-rU - α-^{3}T$_2$-rT - α-^{3}T$_2$-dU)	-0.9	-0.9	-1.0	1.2	1.1	2.0
ΔX(β-^{2}T$_3$-dT + β-^{2}T$_3$-rU - β-^{2}T$_3$-rT - β-^{2}T$_3$-dU)	0.0	0.0	2.1	-0.3	-0.1	-1.5
ΔX(β-^{2}T$_3$-dT + β-^{2}T$_3$-rU - β-^{2}T$_3$-rT - β-^{3}T$_2$-dU)	-2.6	-3.0	-1.2	-0.4	-1.4	2.9
ΔX(β-^{2}T$_3$-dT + β-^{2}T$_3$-rU - β-^{3}T$_2$-rT - β-^{2}T$_3$-dU)	2.2	2.5	0.7	0.0	0.1	-0.8
ΔX(β-^{2}T$_3$-dT + β-^{2}T$_3$-rU - β-^{3}T$_2$-rT - β-^{3}T$_2$-dU)	-0.3	-0.4	-2.6	-0.1	-1.2	3.6
ΔX(β-^{2}T$_3$-dT + β-^{3}T$_2$-rU - β-^{2}T$_3$-rT - β-^{2}T$_3$-dU)	2.6	2.9	5.0	-0.5	0.8	-4.8
ΔX(β-^{2}T$_3$-dT + β-^{3}T$_2$-rU - β-^{2}T$_3$-rT - β-^{3}T$_2$-dU)	0.0	0.0	1.7	-0.6	-0.5	-0.4
ΔX(β-^{2}T$_3$-dT + β-^{3}T$_2$-rU - β-^{3}T$_2$-rT - β-^{2}T$_3$-dU)	4.8	5.5	3.6	-0.2	1.0	-4.1
ΔX(β-^{2}T$_3$-dT + β-^{3}T$_2$-rU - β-^{3}T$_2$-rT - β-^{3}T$_2$-dU)	2.2	2.5	0.3	-0.3	-0.3	0.3
ΔX(β-^{3}T$_2$-dT + β-^{2}T$_3$-rU - β-^{2}T$_3$-rT - β-^{2}T$_3$-dU)	-2.3	-2.7	2.3	-0.7	-0.4	-2.6
ΔX(β-^{3}T$_2$-dT + β-^{2}T$_3$-rU - β-^{2}T$_3$-rT - β-^{3}T$_2$-dU)	-4.9	-5.7	-1.0	-0.8	-1.7	1.8
ΔX(β-^{3}T$_2$-dT + β-^{2}T$_3$-rU - β-^{3}T$_2$-rT - β-^{2}T$_3$-dU)	-0.1	-0.1	0.8	-0.4	-0.1	-1.9

| Compared systems[1] | furanoses | | | | | |
| | vac.[2] | | | solv.[3] | | |
	ΔE	$\Delta E_{(ZPE)}$	ΔG^o	ΔE	$\Delta E_{(ZPE)}$	ΔG^o
$\Delta X(\beta\text{-}^3T_2\text{-dT} + \beta\text{-}^2T_3\text{-rU} - \beta\text{-}^3T_2\text{-rT} - \beta\text{-}^3T_2\text{-dU})$	-2.7	-3.1	-2.5	-0.5	-1.5	2.5
$\Delta X(\beta\text{-}^3T_2\text{-dT} + \beta\text{-}^3T_2\text{-rU} - \beta\text{-}^2T_3\text{-rT} - \beta\text{-}^2T_3\text{-dU})$	0.2	0.3	5.2	-0.9	0.5	-5.9
$\Delta X(\beta\text{-}^3T_2\text{-dT} + \beta\text{-}^3T_2\text{-rU} - \beta\text{-}^2T_3\text{-rT} - \beta\text{-}^3T_2\text{-dU})$	-2.3	-2.7	1.9	-1.0	-0.8	-1.5
$\Delta X(\beta\text{-}^3T_2\text{-dT} + \beta\text{-}^3T_2\text{-rU} - \beta\text{-}^3T_2\text{-rT} - \beta\text{-}^2T_3\text{-dU})$	2.5	2.8	3.7	-0.6	0.7	-5.2
$\Delta X(\beta\text{-}^3T_2\text{-dT} + \beta\text{-}^3T_2\text{-rU} - \beta\text{-}^3T_2\text{-rT} - \beta\text{-}^3T_2\text{-dU})$	-0.1	-0.1	0.4	-0.7	-0.6	-0.8
	pyranoses					
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	-0.6	-0.7	0.2	-0.8	-0.8	-2.3
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	-20.1	-19.4	-21.9	-28.8	-26.7	-33.5
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	-11.8	-10.6	-14.6	-19.4	-18.4	-22.8
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	-31.3	-29.3	-36.6	-47.4	-44.3	-54.1
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	20.6	19.8	23.7	27.5	26.2	29.6
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	1.1	1.0	1.6	-0.5	0.3	-1.6
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	9.4	9.9	8.9	9.0	8.6	9.1
$\Delta X(\alpha\text{-}^1C_4\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	-10.1	-8.9	-13.1	-19.0	-17.3	-22.1
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	9.6	8.2	13.5	19.4	18.1	20.5
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	-9.8	-10.5	-8.5	-8.6	-7.8	-10.8
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	-1.6	-1.7	-1.3	0.8	0.4	0.0
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^1C_4\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	-21.1	-20.4	-23.3	-27.2	-25.4	-31.3
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	30.8	28.6	37.0	47.7	45.0	52.4
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^1C_4\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	11.3	9.9	15.0	19.7	19.2	21.1
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^1C_4\text{-dU})$	19.6	18.8	22.3	29.2	27.4	31.9
$\Delta X(\alpha\text{-}^4C_1\text{-dT} + \alpha\text{-}^4C_1\text{-rU} - \alpha\text{-}^4C_1\text{-rT} - \alpha\text{-}^4C_1\text{-dU})$	0.1	0.0	0.2	1.2	1.6	0.6
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	0.4	0.4	-0.1	0.7	0.7	0.6
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	24.9	25.1	23.9	22.1	21.0	24.1
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	20.7	20.2	21.8	17.0	16.4	19.2
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	45.3	44.9	45.7	38.4	36.6	42.7
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	-24.7	-24.9	-24.4	-20.4	-19.9	-21.4
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	-0.2	-0.1	-0.5	1.0	0.4	2.1
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	-4.4	-5.1	-2.6	-4.1	-4.2	-2.8
$\Delta X(\beta\text{-}^1C_4\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	20.2	19.7	21.4	17.3	16.1	20.7
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	-19.1	-18.4	-21.5	-16.1	-15.3	-18.4
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	5.5	6.3	2.5	5.3	5.0	5.2
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	1.3	1.4	0.4	0.2	0.4	0.2
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^1C_4\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	25.9	26.1	24.3	21.6	20.6	23.8
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	-44.2	-43.7	-45.8	-37.2	-35.9	-40.4
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^1C_4\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	-19.6	-18.9	-21.9	-15.8	-15.6	-16.8
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^1C_4\text{-dU})$	-23.8	-23.9	-24.0	-20.9	-20.2	-21.8
$\Delta X(\beta\text{-}^4C_1\text{-dT} + \beta\text{-}^4C_1\text{-rU} - \beta\text{-}^4C_1\text{-rT} - \beta\text{-}^4C_1\text{-dU})$	0.8	0.9	0.0	0.5	0.1	1.8

[1]$\Delta X = \Delta E, \Delta E_{(ZPE)}, \Delta G^0$. [2] ΔX at B3LYP/6-311++G (d, p) in vacuum. [3] ΔX at B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM implicit solvation model.

Chapter 4

Thermodynamic basis for the emergence of proto-nucleotides containing P or As: a computational assessment[7,8,9]

"The greatest history book ever written is the one hidden in our DNA"

H S. Wells,

in "The journey of man: a genetic odyssey", London, Penguin, 2003

4.1   Introduction

The building blocks of today's nucleic acids (NAs), contain three main components: a nucleobase or **R**ecognition **U**nit (RU), a sugar or **T**rifunctional **C**onnector (TC) and a phosphate (PO_3^-) group or **I**onized **L**inker (IL). A **N**ucleo**s**ide (Ns) contains a sugar + base and a **N**ucleo**t**ide (Nt) has the nucleoside + phosphate. The bases are N-glycosylated to the C1' position of the sugar that can be D-ribofuranose (Ribf) in RNA or D-2'-deoxyribofuranose (2dRibf) in DNA. In today's nucleic acids there are mainly 5 nitrogenous bases {2 purines: **A**denine (A) and **G**uanine (G) and 3 pyrimidines: **C**ytosine (C), **T**hymine (T) and Uracil (U)}. The PO_3^- connects the nucleosides usually between the C3' and C5' positions of the 2dRibf or Ribf creating the backbone of NAs {[-(2d)Ribf-C5'-O-PO_2^--O-C3'-(2d)Ribf-]$_n$} [1]. Another interesting molecular property of the building blocks of NAs is related with the stereoselectivity of the glycosidic bond between the sugar and the base. All nucleotides exhibit a *β*- instead of an *α*-configuration at the C1' carbon of the sugar either in DNA or RNA (see **Figure 4.1**).

---

[7] See https://github.com/mattas-research-group/scripts_PhD_thesis_Lazaro/ for a list of all the scripts written in bash and python to generate the nucleotides, post-process, analyze all results and generate diagrams and tables.
[8] All final DFT-optimized molecular structures reported in this chapter are available in: drive.google.com/PhD_Thesis/Optimized_Final_Structures/Chapter#4.
[9] All thermodynamic quantities are represented in kJ/mol since this unit is addopted by the International System of Units (SI).
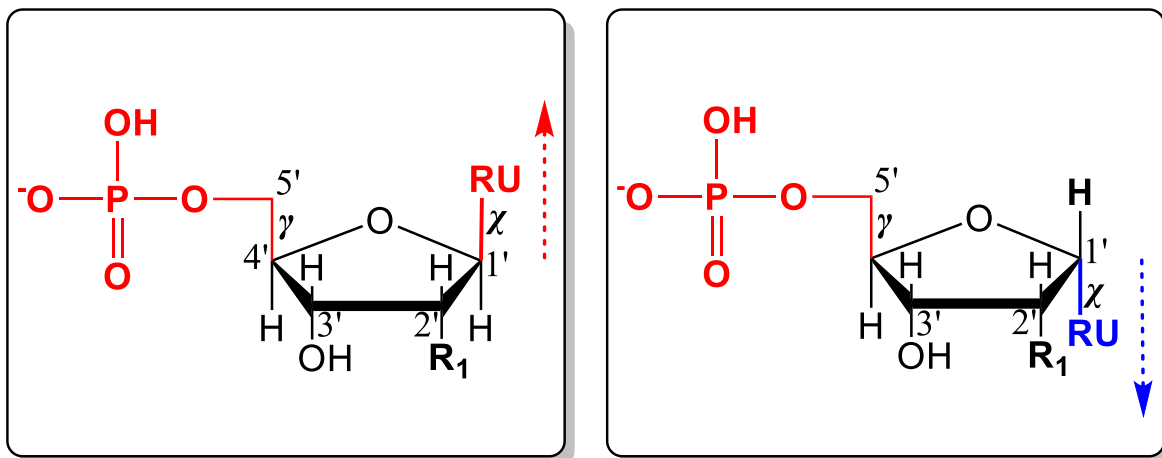
Figure 4.1 Molecular structure of nucleotides and orientation of the RU (nucleobases or recognition units) at the C1' of a 5-MR TC (trifunctional connector) with respect to the IL (**I**onized **L**inker) hydrogen-phosphate (HPO_3^-) group at the C5' for the (left): β- (found in today's NAs) and (right): α-anomers of nucleotides. R_1: H in 2'-deoxynucleotides (in DNA) or OH in ribonucleotides (in RNA).

But the question remains, what are the factors that pushed Nature to select such a molecular structure for today's nucleotides?

Can we consider thermodynamics or the energetic changes associated with the molecular structure of the nucleotides and their synthesis as a determining factor for the natural selection and/or evolution of the building blocks of NAs?

Due to the known instability of the N-glycosidic and phosphate bonds of the canonical nucleotides ("water problem") [2] and their inability to create complementary base pairing ("paradox of base pairing") [3, 4] in aqueous solution it has been proposed that DNA and RNA emerged as a product of evolution from a proto- and pre-RNA that was synthesized easier in the primitive earth from components different to today's TCs, RUs and ILs [5].

Among these alternative components the bases 2, 4, 6-**T**ri**A**mino**P**yrimidine (TAP), **B**arbituric **A**cid (BA), **M**ela**M**ine (MM) and **C**yanuric **A**cid (CA) [6, 7, 8, 9] have been chosen since they can generate π-stacking polymers and/or glycosylate with ribose dehydrating/rehydrating conditions.

Additionally, other alternative backbones have been also explored for the capacity to create **W**atson & **C**rick (WC) complementary base pairing similar to today's DNA double helix and cross-pair with single strands of DNA and RNA. These include 6 - membered ring (6 - MR) sugars (in pyranosil (p)-RNA) [10, 11], threose (in TNAs) [5, 12, 13], glycerol (in GNA) [5, 14, 15, 16, 17, 18, 19] and the arsenate ion (AsO_3^-) has been proposed as a plausible replacement for PO_3^- [20, 21, 22].

Could thermodynamics explain the plausibility for the synthesis of some of these alternative (non-canonical) nucleotides and their corresponding stereoselectivity? How is it compared with the canonical counterparts.

In summary, this study will address whether or not the β- instead of an α-configuration is preferential for alternative RUs. The role of thermodynamics in the anomeric selection and assembly of alternative non-canonical components of nucleotides and the selection of a furanose ring over a pyranose ring will also be explored.

Similar to our paper published in 2022 [23] (reproduced in this thesis as its Chapter 2) two synthetic pathways: classic and alternative pathways for the prebiotic formation of canonical and non-canonical nucleotides will be considered and the following hypothesis will be tested: did the preferential conformations of the RU around the TC and the preferential 2dRibf or Ribf sugar ring conformations were inherited from the individual monomeric blocks?

Finally, different sugar-exchange reactions will be modeled to assess if the changes in the free energy for the reaction can justify the selection of T for DNA and U for RNA.

4.2   Computational methods

All canonical and non-canonical nucleotides were designed from the nucleosides and individual components: **T**rifunctional **C**onnectors (TCs) and **R**ecognition **U**nits (RUs) modeled in Chapter 3 of this dissertation.

Two different pathways were considered in modeling the artificial prebiotic synthesis of the nucleotides:

- Classic pathway (a+b): include the condensation reaction between a TC and a RU (reaction (a) {TC-**OH** + **H**-RU → nucleoside + H_2O}) followed by a condensation reaction between

the corresponding nucleosides and the IL (reaction (b) or {nucleoside-**H** + **OH**-IL →
nucleotide + H_2O}).

- Alternative pathway (c+d): a condensation reaction between each TC with each **I**onized
Linker (IL) generates a TC-IL backbone derivative (reaction (c) or {TC-**H** + **OH**-IL → TC-
IL backbone + H_2O}). This reaction is proceeded by the condensation reaction between the
TC-IL backbone and each RU to generate the corresponding nucleotides (reaction (d) or
{TC-IL-**OH** + **H**-RU → nucleotide + H_2O}).

An in-house bash-Python scripts were used to design the 3D structures of each nucleoside,
TC-IL backbone and nucleotide by manipulating the Z-matrices of the different TCs, RUs, IL,
nucleosides and TC-IL backbones. These scripts use OpenBabel [24] in bash mode to convert files
to different formats. All possible combinations between the components represented in **Figure 4.2**
were explored and each nucleotide was considered in its β- and α-configuration.

The initial geometries of all the individual components were optimized at the DFT-
B3LYP/6-311++G(d, p) level of theory. Two initial ring puckering conformations were
considered for the furanose (F-form) ($^2T_3$ and $^3T_2$) and two for the pyranose forms (P-form) ($^1C_4$
and $^4C_1$) of the TC that were **C**arbo**H**ydrate**s** (CHs) in the corresponding α- and β-configurations
at the C1' position.

For the classic pathway 484 nucleosides that can bind an IL were modeled (see Chapter 3)
by binding each TC with each RU.

PNA nucleosides were excluded from this study since the nucleosides containing N-(2-
aminoethyl)glycine can create polymeric strands through peptide bonds without the need of an IL
[25, 26, 27, 28, 29, 30, 31].

For consistency, the N- and C-glycosidic bond lengths were initially set to 1.52 Å,
while the dihedral angle H-C1'-N(C)-Cx was set to -179.46º. These values guarantee that
there are not intramolecular steric clashes or undesirable bond formation in the building
blocks. The N- or C-glycosidic bonds included the C1'-N1 for the canonical pyrimidines
and N-glycosylated BA, C1'-N9 for the canonical A and G, C1'-C5 for C-glycosylated BA
and TAP, C1'-hexocyclic NH for N-glycosylated TAP or MM and C1'-N5 for MM.

Each of these 484 nucleosides structures (242 in vacuum and implicit solvent
respectively) were subjected to a fully relaxed scan around the dihedral χ involving the
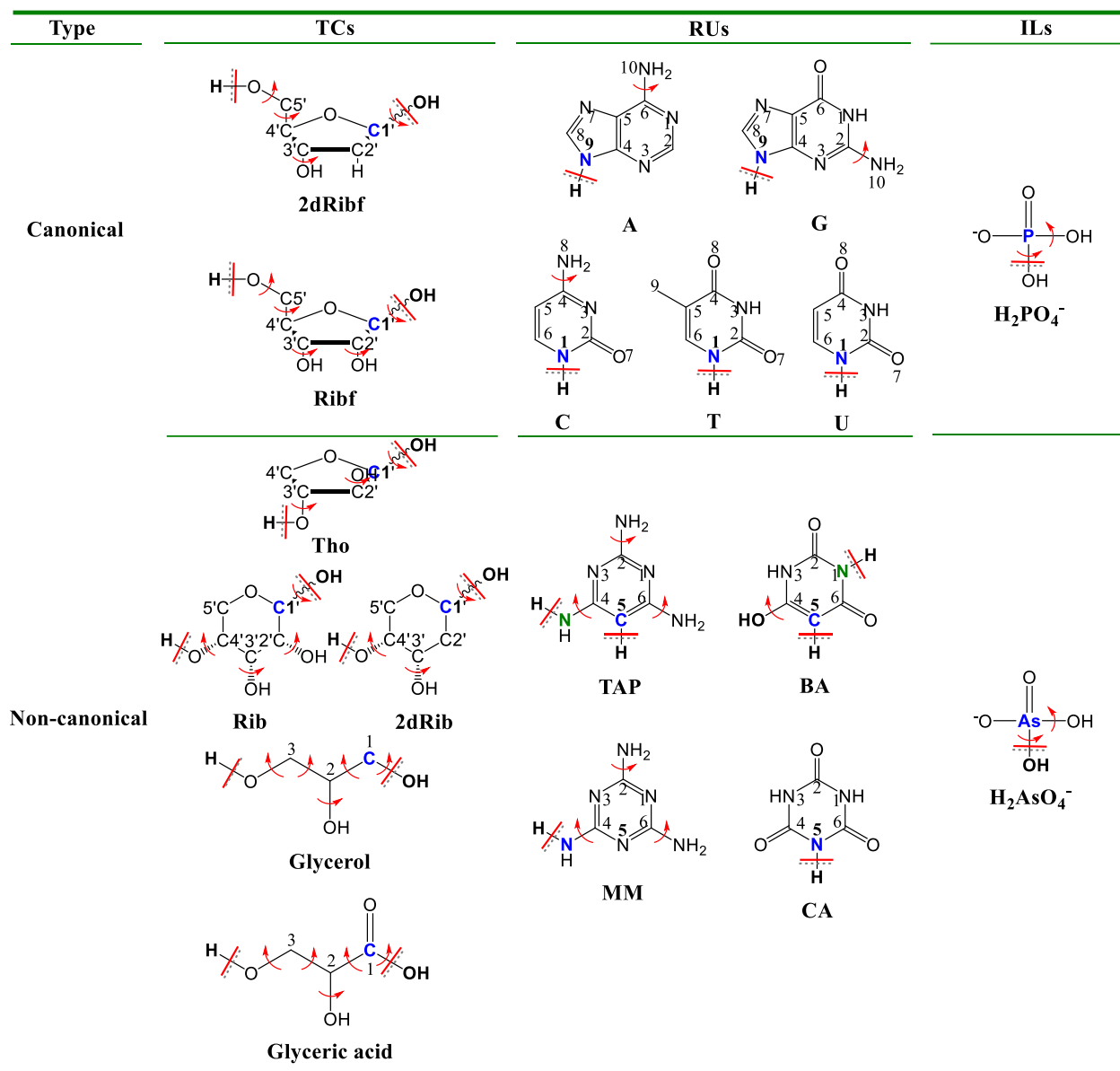
Figure 4.2 Canonical and non-canonical RUs, TCs and ILs considered in the modeling of the nucleotides. A: adenine, G: guanine, C: cytosine, T: thymine, U: uracil, TAP: 2, 4, 6 -Triaminopyrimidine, BA: barbituric acid in enol form, MM: melamine, CA: cyanuric acid, 2dRibf: 2'-deoxyribofuranose (β-anomer present in DNA), Ribf: D-ribofuranose (β-anomer present in RNA), Tho: D-threose (L-enantiomer in TNA), 2dRib: D-2'-deoxyribopyranose (present in p-DNA), Rib: ribopyranose (present in p-RNA), glycerol and glyceric acid (present in GNA), $H_2PO_4^-$: (di)hydrogen-phosphate ion, $H_2AsO_4^-$: (di)hydrogen-arsenate ion. ▬▬▬: Bonds been broken during

298

the condensation reactions. ➤ : rotatable bonds changed during modeling of the potential energy surface at the semiempirical PM7 and the DFT-B3LYP/6-311++G(d, p) levels. The blue and green C and N represent reactive centers for the glycosylation of TAP and BA that creates N- or C-nucleotides respectively.

glycosidic bond in 6 steps of 60°. The structures with the lowest energy from each of these PES were refined by subjecting each one to a final fully-unconstrained optimization at the DFT-B3LYP/6-311++G(d, p) level of theory to obtain the final structure of the nucleosides in vacuum and in implicit solvation. A harmonic frequency calculation was performed as usual to ensure that the final structures were, each, a minimum on the PES.

Each IL, either $H_2PO_4^-$ or $H_2AsO_4^-$ was bound to the C5' (2dRibf, Ribf), C4' (2dRib, Rib), C3' (Tho), C1 (glycerol or glyceric acid) of each nucleoside and a soft potential energy scan was performed by rotating the torsion angle γ that involves the HPO_3^--C5'-C4'-Cx dihedral (see **Figure 4.1**) in 6 steps of 60°. The molecular geometries with the lowest energy were fully optimized at the same B3LYP/6-311++G(d, p) in vacuum and in implicit solvation and it was verified that vibrational frequencies included no imaginary ones to guaranty that the optimized structures were local minima on the PES.

In total 968 nucleotides (484 with each IL and from those 242 in each environment) were modeled according to the protocol described above.

A similar procedure was followed for the modeling of the canonical and non-canonical nucleotides considering an alternative pathway. In this case the TC-IL backbones were first modeled by binding each IL to the C5' (2dRibf, Ribf), C4' (2dRib, Rib), C3' (Tho), C1 (glycerol or glyceric acid) generating a total of 88 TC-IL backbones {2 ILs x 2 environments (vacuum and implicit solvation) x [20 sugar ring conformations (12 [6 anomers (β and α]) for the 5-MR and 8 [4 anomers (β and α] for the 6-MR sugars respectively) + 2 non-sugar TCs]}.

Each RU was C- or N-bound to the corresponding TC-IL backbones at the C1'. A similar procedure to the one to obtain the final nucleosides in the classic pathway (a+b) was followed to obtain the final 968 canonical and non-canonical nucleotides minimized structures.

The different sugar-exchange reactions, pseudorotational equilibriums (see equations 3.1-3.6 of section 3.3 in Chapter 3), RU conformations on the glycosidic bond and puckering parameters for each nucleotide with a CH-like TC were estimated similarly to the procedure in Chapter 3.

Aqueous solvation has been implemented implicitly by using the integral equation formalism variant of the "*polarizable continuum model*" (IEFPCM) [32, 33, 34, 35, 36, 37] implemented in Gaussian 16 [38], the software package used for all DFT calculations in this work.

4.3   Results and discussion

4.3.1   Pseudorotational equilibrium and anomer-exchange reactions of canonical and non-canonical nucleotides containing 5- and 6-MR sugars

A paper published by Castanedo and Matta (CM) in 2022 [23] estimates the changes in different thermodynamical parameters for the selection of the β- over the α- anomers of the canonical nucleotides in their F-form by modeling each building block at the B3LYP/6-31G (d, p) level of theory in vacuum and using the IEFPCM to account for the effects of solvent. The energetic differences between both β- and the α-configuration were estimated as:

$$X_{\beta\alpha} = X_{\beta} - X_{\alpha} \tag{4.1}$$

where ΔX denotes ΔE (the difference is of the total energies), or $\Delta E_{(ZPE)}$ (the difference in the total energies corrected for zero-point vibrational energies (ZPEs)) and the change in the Gibbs free energy (ΔG°).

CM found that for the Nts obtained in vacuum from either classic or alternative pathways the β-anomer was favored across the board.

Thibaudeau and coworkers (Tcw) [39, 40, 41] have addressed the question on why Nature chose the β-anomers over the α-counterpart for today's nucleosides and nucleotides by analyzing the pseudorotational equilibrium between the two preferential furanose ring conformations South

or $^2T_3$ (C2'-endo-C3'-exo) and North or $^3T_2$ (C2'-exo-C3'-endo) and the different stereoelectronic effects that affect these equilibria. Some of these factors include the anomeric and gauche effect. Tcw determined that for 2',3'-di-(deoxy)nucleosides (ddNs) the α-anomer is more favored in the $^2T_3$ conformation meanwhile the β-counterpart is more favored in the $^3T_2$. Tcw measured the thermodynamic parameters for the pseudorotational equilibrium in a protonated, deprotonated and neutral state of the bases and determined that the ΔG° and ΔH° for the pseudorotation of the β-nucleosides was sensitive to the variation of the pK_a of the environment than the α-counterpart, which could mean that correct anomers is more flexible. Tcw argues that this property could have favored one anomer over the other in environments with a gradient of pH and temperature like e.g., the hydrothermal vents [42, 43, 44].

Applying the ideas from Tcw we have decided to theoretically study a total 24 reactions in vacuum and implicit solvation, respectively for each TC-IL backbone and each nucleotide from pathways (a+b) or (c+d). These reactions include 4 pseudorotational equilibrium (two for furanoses and two pyranoses) and 8 anomer-exchange reactions (4 for each sugar ring) for the nucleotides with each type of IL (see schemes 3.7-3.18 in section 3.4.2 of Chapter 3 and **Figure 4.3**).

Pseudorotational equilibrium and anomer-exchange reactions for TC-IL backbones

Figure 4.4 shows a bar graph with the ΔG^0 for the different pseudorotational equilibriums (1 and 3) and the different sugar-exchange reactions (reactions 2, 4, 5 and 6) for the TC-IL backbone molecules containing either $H_2PO_4^-$ or $H_2AsO_4^-$ in vacuum or implicit solvation.

An analysis of the equilibriums #1 and #3 shows that for the HPO_2^--O-TC in vacuum most of the energies are within the intrinsic error for the method used (≈ 17 kJ/mol). Exceptions are observed in vacuum for $HPO_2^--O-Ribf$ in which equilibrium #1 favors the α-$^3T_2$ ($\Delta G^\circ = 17.9$ kJ/mol) and for $HPO_2^--O-Ribf$ equilibrium #3 favors β-$^3T_2$ ($\Delta G^\circ = 25.1$ kJ/mol). In the case of the $HAsO_2^--O-TC$ backbones in vacuum the only exceptions are the pyranoses 2dRib (equilibrium #1 favors the α-$^3T_2$ {$\Delta G^\circ = 21.5$ kJ/mol}) and the Rib where equilibrium #1 and #3 favors the α-$^3T_2$ {$\Delta G^\circ = 17.6$ kJ/mol} and β-$^3T_2$ {$\Delta G^\circ = 20.4$ kJ/mol} respectively. In aqueous solution the only significant energies are for the equilibrium #3 of the $HAsO_2^--O-2dRibf$ in which the β-$^2T_3$ is favored.
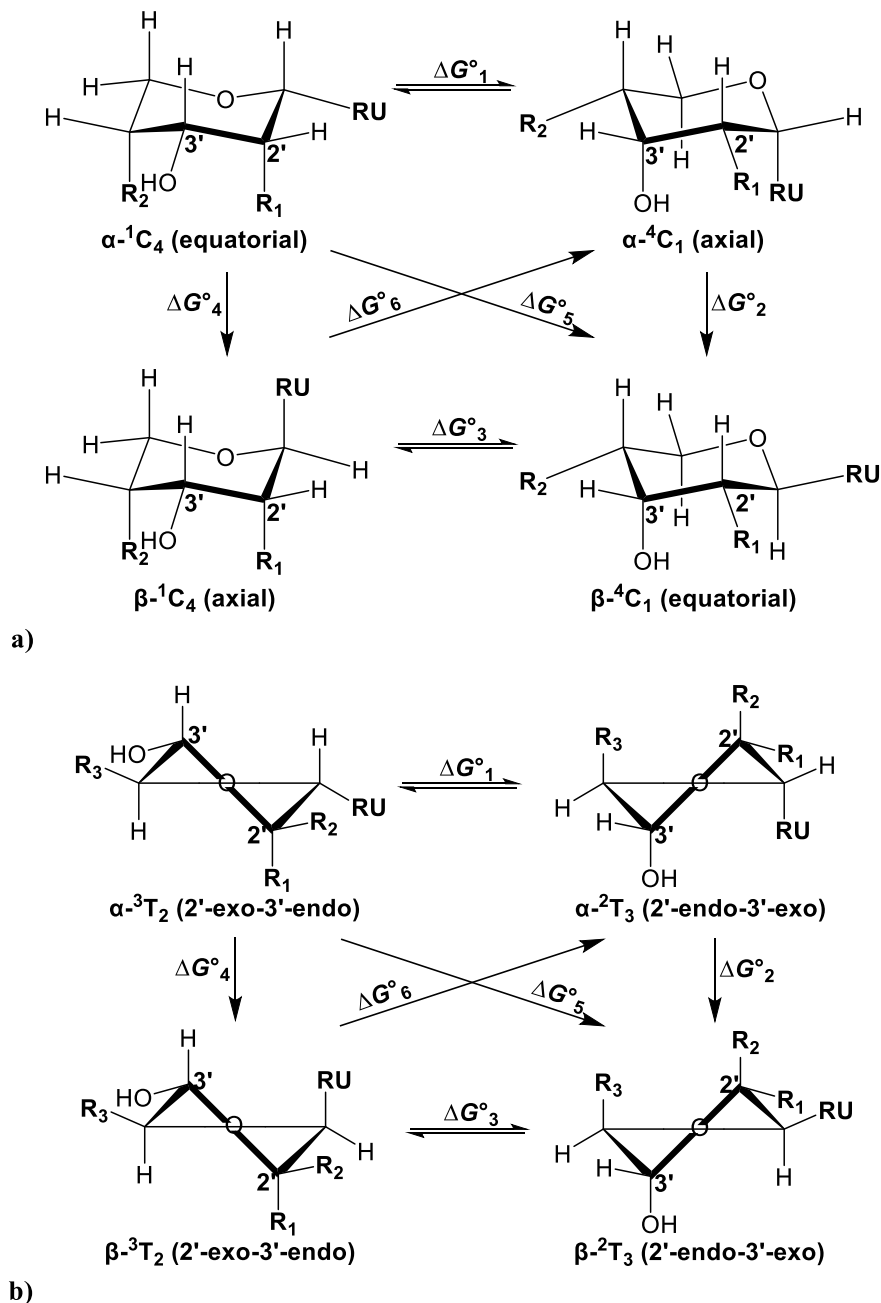
Figure 4.3 a) Equilibrium between the E_1 ($^4C_1$) and A_1 ($^1C_4$) conformations for a nucleotide with a 6-MR TC. (R_1: H for 2dRib), OH for Rib nucleotides. R_2: HPO_3^- or $HAsO_3^-$, b) equilibrium between the $^3T_2$ (2'- exo-3'-endo) and $^2T_3$ (2'-endo-3'-exo) conformations for the nucleotides with a 5-MR TC. (R_1: H for 2dRibf and Tho, OH for Ribf. R_2: H for 2dRibf, Ribf and OH for Tho. R_3: HPO_3^- or $HAsO_3^-$) [45, 40, 41] (taken from [40] and reprinted with permission of Elsevier).

302
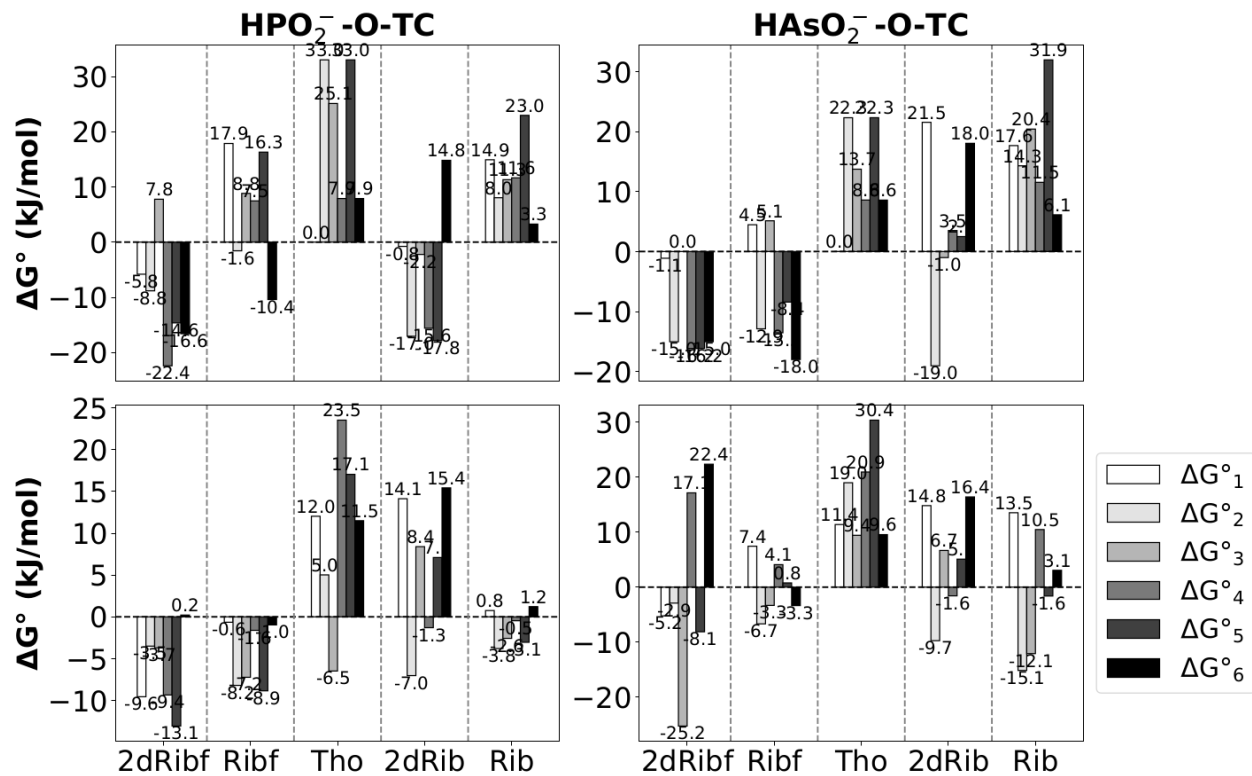
Figure 4.4   Comparison of the Gibbs free energies (ΔG°) at 298 K for the equilibrium of pseudorotation and anomeric transformation reactions for the α- and β- anomers of (left) TCs-mono-phosphate (HPO$_2^-$-O-TC) and (right) TCs-mono-arsenate (HAsO$_2^-$-O-TC) backbones. The sugar ring conformations considered are $^3T_2$ and $^2T_3$ for the 5-MR 2dRibf, Ribf, Tho and $^1C_4$ and $^4C_1$ for the 6-MR 2dRib and Rib. The anomers are α- and β-. The Gibbs energies are defined by equations 3.1-3.6 of Chapter 3. (Top) ΔG°, B3LYP/6-311++G (d, p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

An analysis of the anomer-exchange reactions for the backbone molecules containing HPO$^-_3$ shows that for the case of the HPO$^-_2$-O-Ribf in vacuum the reaction #4 favors the β-$^3T_2$ (ΔG°_4 = -22.4 kJ/mol). In the case of the HPO$^-_2$-O-Tho reaction #2 and #5 favors the α-$^2T_3$ (ΔG°_2 = 33.0 kJ/mol) and α-$^3T_2$ (ΔG°_5 = 33.0 kJ/mol) anomers. For HPO$^-_2$-O-2dRib reaction #2 and #5 favors the β-$^4C_1$ (ΔG°_2 = -17.0 kJ/mol and ΔG°_5 = -17.8 kJ/mol respectively). For the HPO$^-_2$-O-Rib the reaction #5 prefers the α-$^1C_4$.

In the case of the backbone molecules in aqueous environment the reactions #4 and #5 for HPO⁻$_2$-O-Tho favor the α-^{3}T$_2$ (ΔG_4° = 23.5 and ΔG_5° = 17.1 kJ/mol).

For the backbone molecules containing HAsO⁻$_3$ instead of HPO⁻$_3$ in vacuum the reaction #6 for HAsO⁻$_2$-O-Ribf favors the α-^{2}T$_3$ (ΔG_6° = -18.0 kJ/mol). For the HAsO⁻$_2$-O-Tho reactions #2 and #5 favor the α-^{2}T$_3$ (ΔG_2° = 22.2 kJ/mol) and the α-^{3}T$_2$ (ΔG_5° = 22.3 kJ/mol). For the HAsO⁻$_2$-O-2dRib reactions #2 favors the β-^{4}C$_1$ (ΔG_2° = -19.0 kJ/mol) and reaction #6 favors the β-^{1}C$_4$ (ΔG_6° = 18.0 kJ/mol). For the case of the HAsO⁻$_2$-O-Rib reaction #5 favors the α-^{1}C$_4$ (ΔG_5° = 31.9 kJ/mol).

For the HAsO⁻$_3$ backbone molecules in implicit solvation the only significant differences are observed for the reaction #4 and #6 of the HAsO⁻$_2$-O-2dRibf favoring the α-^{3}T$_2$ (ΔG_4° = 17.1 kJ/mol) and β-^{3}T$_2$ (ΔG_6° = 22.4 kJ/mol) respectively and for the HAsO⁻$_2$-O-Tho, reactions #2 (α-^{2}T$_3$, ΔG_2° = 19.0 kJ/mol), #4 (α-^{3}T$_2$, ΔG_4° = 20.9 kJ/mol) and #5 (α-^{3}T$_2$, ΔG_5° = 30.4 kJ/mol).

Overall, it is noticed that changing the IL from HPO⁻$_3$ to HAsO⁻$_3$ changes the energetic outcomes from the pseudorotational equilibriums and the sugar-exchange reactions. Nevertheless, in some cases both anomers cannot be discriminated thermodynamically since the free energies are within the intrinsic error of the computational method. It is also observed that for the sugar-exchange reactions in which one anomer is preferred the α-configuration is usually favored.

Pseudorotational equilibrium and anomer-exchange reactions for canonical and non-canonical nucleotides

Let's assess first how our results compared to the studies from Tcw [39, 40, 41]. Tcw predicted that overall, for ddNs and D-2'-deoxyribonucleosides (dNs) the pseudorotational equilibrium of the β-anomer is displaced towards the ^{3}T$_2$ ring conformation, meanwhile for the α-anomer the ^{3}T$_2 \rightarrow$ ^{2}T$_3$ equilibrium prefers the ^{2}T$_3$ sugar ring conformation.

The magnitude for the variation of the energies is usually higher for the β- {ΔG° = 3.1 kJ/mol} than the α-anomer {ΔG° = -1.2 kJ/mol}. Meanwhile, for the case of canonical

ribonucleotides with a PO_4^{3-} group in the C3' the gauche effects of the O2'-C2'-C1'-N1/N9 drives the pseudorotation towards the $^2T_3$ puckering for β-anomers and in the case of the α-anomers there is not preference for a specific sugar ring conformation since there is a O2'-C2'-C1'-N1/N9 gauche effect in both $^2T_3$ and $^3T_2$.

Table A3 in the Appendices section represents the % of ΔG° within the error of the method (17 kJ/mol), $\leq$ -17 kJ/mol and $\geq$ 17 kJ/mol for the different pseudorotational equilibriums and anomer-exchange reactions for canonical and non-canonical nucleotides (see Table A1 and Table A2 from the Appendices section for the specific values of each ΔG°).

Pseudorotational equilibrium #1:

In the case of the nucleotides obtained from the pathway (a+b) an analysis of the ΔG° for the pseudorotational equilibrium #1 of the canonical and non-canonical F-form nucleotides in their α-configuration with phosphate in vacuum and aqueous solution shows that most of the structures have negligible ΔG° (either sugar puckering is favored). For the case of the nucleotides with the TC in the P-form there is a higher % of structures with energies $\geq$ 17 kJ/mol favoring the α-$^1C_4$ puckering. A similar trend is observed when phosphorus is replaced by arsenic.

In the case of the nucleotides from pathway (c+d) similar trends are observed for the 5-MR and 6-MR of the nucleotides with HPO_3^- and $HAsO_3^-$.

These results suggest that for the 6-MR nucleotides with an α-configuration a predominance of the steric over the anomeric or gauche effects may force the conformation of the sugar to place the base in the most stable equatorial position in the pseudorotational equilibrium. This conformation may avoid electrostatic repulsion with the IL in the C4' and the hydroxyl group in the C2' of the sugar.

Pseudorotational equilibrium #3:

An analysis of the distributions for the pseudorotational equilibrium #3 corresponding to the β-anomers of the monophosphate nucleosides obtained from the classic pathway reveals that for the nucleotides with the sugar in the F-form in vacuum there is a slightly higher % of structures

with negligible energies (46.7%) meanwhile around 40% of structures have $\Delta G_3^{\circ} \geq 17$ kJ/mol favoring the $^3T_2$ puckering. This result is in agreement with Tcw studies [45]. In aqueous environment all energies are negligible. For the non-canonical nucleotides most of energies are within the intrinsic error of the method used. For the case of the canonical nucleotides in the P-form in either vacuum or solvation the highest % of structures have $\Delta G_3^{\circ} \geq 17$ kJ/mol favoring the β-$^1C_4$.

For the case of the canonical nucleotides containing arsenate and a 5-MR in vacuum and aqueous solution most of structures have negligible energies. In the case of the canonical nucleotides with a 6-MR there is 50% of structures with $\Delta G^{\circ} \leq 17$ (favoring the β-$^4C_1$ ring conformation) and the other 50% of structures with ΔG° within the error. Solvation makes 80% of the energies negligible. 66.7% and 83.3% of non-canonical nucleotides in vacuum and implicit solvation respectively had energies within the error.

For the case of the canonical and non-canonical nucleotides with phosphate and a TC in the F-form obtained from the alternative (c+d) pathway most of ΔG° are within 17 kJ/mol. For the case of the canonical nucleotides with a 6-member ring (6-MR) 80% and 50% of free energies in vacuum and aqueous solution contain $\Delta G^{\circ} \leq 17$ kJ/mol ($^4C_1$ favored) respectively. The rest of energies are negligible. For the canonical and non-canonical nucleotides with a 5-MR and arsenate in vacuum there is a higher % of structures with $\Delta G^{\circ} \geq 17$ kJ/mol ($^3T_2$ puckering preferable), meanwhile in aqueous solution most of energies are within the intrinsic error of DFT-B3LYP/6-311++G(d, p). For the canonical nucleotides with the TC in the P-form in either environment there is a preference for the $^4C_1$ puckering, meanwhile the non-canonical nucleotides with a 6-MR TC does not prefer either TC ring conformation in vacuum (75% with -17 kJ/mol $< \Delta G^{\circ} <$ 17 kJ/mol) and in implicit solvation there is a slight preference for β-$^4C_1$ (58.3%).

In summary, most of pseudorotational equilibriums have energies within the intrinsic error of the DFT-B3LYP/6-311++G(d, p) (-17 kJ/mol $<\approx \Delta G^{\circ} <\approx$ 17 kJ/mol) with some scarce exceptions mostly for the nucleotides with the TC in the P-form. Generally, for these cases the pseudorotational equilibrium of nucleotides obtained from the (a+b) pathway is bias towards the $^1C_4$ puckering, meanwhile for the nucleotides obtained from the alternative pathway (c+d) the equilibrium favors the $^4C_1$ sugar ring puckering.

Let's turn now to the question on if there are any thermodynamic differences between the β- and α-anomers in when the sugar ring conformations in both configurations are the same or change. In order to address this question, the % of ΔG° with a magnitude within -17-17 kJ/mol and beyond will be analyzed for the anomer-exchange reactions #2, #4, #5 and #6 (see **Figure 4.3**).

Reaction #2:

The reaction #2 represents the anomer-exchange reaction $\alpha\text{-}^2T_3 \rightarrow \beta\text{-}^2T_3$ for Nts containing the TC with a F-form and $\alpha\text{-}^4C_1 \rightarrow \beta\text{-}^4C_1$ when the Nts contain a 6-MR. When this type of reaction has a $\Delta G^{\circ} \leq$ -17 kJ/mol the correct β-configuration may be favored over the α-counterpart.

Table A3 shows that for the reaction #2 $\{\alpha\text{-}^2T_3 \rightarrow \beta\text{-}^2T_3\}$ in the pathway (a+b) of the canonical and non-canonical HPO_3^--nucleotides with a 5-MR TC *there is a higher % in vacuum of energies $\leq$ -17 kJ/mol favoring the $\beta\text{-}^2T_3$ configuration*, meanwhile in aqueous solution most of energies are inconclusive. For the nucleotides with 6-MR in both environments the $\beta\text{-}^4C_1$ seems to be favored for both canonical (50% of energies in vacuum and 90% in aqueous solution) and non-canonical nucleotides (66.7% in vacuum and 91.7% in aqueous solution). For the nucleotides with $HAsO_3^-$ a similar trend is observed to the monophosphate counterparts for either canonical and non-canonical 5- and 6-MRs. This result suggests that replacing the IL may not induce a significant change in the nucleotide's molecular geometry.

In the case of the canonical nucleotides obtained from the pathway (c+d) with a 5-MR and HPO_3^- most of the energies are negligible in both environments. In the case of the non-canonical analogs in vacuum there are 38.9% of energies favoring the $\alpha\text{-}^2T_3$ or within the intrinsic error of the method respectively, meanwhile in aqueous solution all energies are within the error of B3LYP/6-311++G(d, p). For the case of the canonical nucleotides with 6-MR obtained from (c+d) in vacuum most of energies are negligible meanwhile in aqueous solution 90% of energies favor the $\beta\text{-}^2T_3$. For non-canonical nucleotides with a 6-MR and a monophosphate a similar trend is observed.

Let's focus now in the Nts with arsenate from the alternative pathway. The Nts with a 5-MR have either similar % of energies favoring either the $\alpha\text{-}^2T_3$ or been negligible. In aqueous environment 86.7% of energies are within the error. For the non-canonical Nts a similar trend is found. For the case of the arsenate canonical nucleotides containing a 6-MR sugar in vacuum; all

energies are negligible meanwhile in aqueous solution all energies favor the β-^{2}T$_3$ anomer. In the case of the non-canonical nucleotides with the TC in the P-form 41.7% and 75% of ΔG_2° favors the β-^{2}T$_3$ in vacuum and implicit solvation respectively.

Overall, the reaction #2 in the classic pathway favors mostly the correct β-anomer, meanwhile for the alternative pathway a less conclusive result is obtained. In this case mostly nucleotides with the TC in the P-form have a higher preference for the correct β-configuration.

Reaction #4:

An analysis of the ΔG° for the anomer-exchange reaction #4 {α-^{3}T$_2 \rightarrow \beta$-^{3}T$_2$} or { α-^{1}C$_4 \rightarrow \beta$-^{1}C$_4$} shows that in the case of the nucleotides obtained through the (a+b) pathway when a canonical RU is present, the IL is monophosphate and the TC has a 5-MR in vacuum 53.3 % of reactions favor the β-^{3}T$_2$, meanwhile in aqueous solution 80% of energies are negligible. For the non-canonical counterparts, a similar trend is observed with 38.9% of reactions either favoring the β-anomer or been negligible in vacuum and in aqueous solution 88.9% of reactions has negligible energies. When the TC has a 6-MR and there is a canonical RU in either environment the energies favor the α-^{1}C$_4$ (70% in vacuum and 50% with IEFPCM). In the case of the 6-MR with non-canonical RUs the energetic changes in the reaction are negligible in both environments.

In the case of the Nts with HAsO$_3^-$ when they are obtained from the classic pathway most of reactions in either environment has negligible ΔG°.

For the nucleotides containing arsenic, either canonical or non-canonical RUs and a 5-MR sugar most of energies are not conclusive in both environments. Meanwhile, for the canonical nucleotides with a 6-MR 90% and 70% of reactions favor the incorrect anomer in vacuum and aqueous environments and for the non-canonical 6-MR nucleotides there is not clear preference since most of energies are within the error of the method.

In the case of the nucleotides obtained from the alternative (c+d) pathway with HPO$_3^-$ when the building blocks have a 5-MR and either a canonical or non-canonical base most of energies are negligible in both environments. For the 6-MR nucleotides a similar trend is observed, nevertheless in the case of the canonical ones in vacuum 80% of the reactions favor the α-^{1}C$_4$.

In the case of the Nts from (c+d) with a 5-MR and arsenate most of ΔG° are negligible with canonical or non-canonical RUs in either environment. When the 5-MR is replace by the P-form and a canonical base is present 90% and 80% of reactions favor the $\alpha\text{-}^1C_4$, meanwhile for the non-canonical counterparts most of energetic variations are negligible in both environments.

Reaction #5:

The magnitude and sign of ΔG_5° represents the change in free energy when interconverting $\alpha\text{-}^3T_2 \rightarrow \beta\text{-}^2T_3$ for 5-MR Nts or $\alpha\text{-}^1C_4 \rightarrow \beta\text{-}^4C_1$ for 6-MR Nts. In this case if the delta energies have a $-$ sign and magnitude under the intrinsic error of the method (-17 kJ/mol) then the reaction may favor the naturally observed β-anomer.

An analysis of **Table A3** shows that for the case of the nucleotides obtained from the classic (a+b) pathway, when there is a phosphate, a canonical base and a 5-MR present in vacuum there is an slight preference for the correct anomer (46.7 % of reactions), meanwhile 40% prefer the wrong configuration. In the rest of the cases most of reactions have negligible energies.

For the nucleotides with arsenate instead of phosphate a similar trend is observed.

For the phosphate Nts obtained from (c+d) most of reactions have negligible energies. Only for the case of the arsenate canonical nucleotides with a 5-MR in vacuum 46.7% of reactions prefer the $\alpha\text{-}^3T_2$.

For the case of the nucleotides obtained from the alternative (c+d) pathway a similar trend is observed to the previously described. Similarly, the only the exception to the classic Nts is found for the reaction of a base with a 5-MR TC-O-AsO$_2^-$ in vacuum. In this case 73.3% (for canonical RUs) and 38.9% (for non-canonical RUs) favor the $\alpha\text{-}^3T_2$.

Reaction #6:

The anomer-exchange reaction # 6 represents either $\beta\text{-}^3T_2 \rightarrow \alpha\text{-}^2T_3$ for nucleotides with a 5-MR or $\beta\text{-}^1C_4 \rightarrow \alpha\text{-}^4C_1$ for the nucleotides with a 6-MR. In this case the $+$ sign with a magnitude over 17 kJ/mol represents reactions that favor the correct β-anomer.

In the case of the nucleotides from the classic pathway (a+b) with a phosphate and a 5-MR in both environments the energies are within the error of the method when the RU is canonical. For non-canonical analogs in vacuum, there is 50% of reactions that favors the naturally occurring β-$^3T_2$, meanwhile in aqueous solution most of reactions does not offer a clear preference. meanwhile, for canonical nucleotides from (a+b) with a 6-MR in vacuum there are 40% of reactions that does not favor either anomer and in aqueous solution there are 60% of reactions favoring the β-$^1C_4$. And finally, for non-canonical nucleotides with a 6-MR TC and a HPO_3^- in both environments 58.3% of reactions favor the correct β-$^1C_4$.

Most reactions in both environments involving canonical (a+b) nucleotides with arsenate and a 5-MR do not exhibit a clear preference for one anomer over the other. Only for non-canonical nucleotides in vacuum, 38.9% are observed to favor the β-$^3T_2$. Arsenate nucleotides with a 6-MR exhibit a similar trend. For non-canonical Nts with the P-form TC and $HAsO_3^-$, 66.7% (vacuum) and 58.3% (in implicit solvation) favor the correct β-$^1C_4$.

For all nucleotides obtained through the alternative (c+d) pathway, most of the reactions have energies within -17 and 17 kJ/mol, hence overall there are not significant energetic differences between the β-$^3T_2$ or β-$^1C_4$ compared to the α-$^2T_3$ or α-$^4C_1$.

Overall, it can be noticed that at the exception of the anomer exchange reaction # 2 {α-$^2T_3 \rightarrow \beta$-$^2T_3$ for Nts with a 5-MR and α-$^4C_1 \rightarrow \beta$-$^4C_1$ for Nts with a 6-MR sugar} in most of reactions there is high % in which there is not a significant difference between both α- and β-configurations in different sugar ring puckering. *These observations suggest that an anomer-exchange reaction similar to #2 in which either anomer has the same $^2T_3$ or $^4C_1$ conformation (depending on the size of the sugar ring) could have favored the configuration to be selected later during evolution as part of today's RNA and DNA.*

4.4.2 Thermodynamic feasibility for the synthesis of canonical and non-canonical nucleotides from a classic and alternative pathway

In the previous section the energetic differences between the two possible β- and α-anomers of the canonical and non-canonical building blocks of nucleic acids were analyzed to model the role of thermodynamics in the selection of one anomer over the other to be present in today's informational polymers.

This section focusses on the modeling of the synthesis of canonical and non-canonical nucleotides through the condensation reactions of their components: TCs, RUs and ILs. The hypothesis to be tested is the following: did thermodynamics play also a role in the synthesis and early selection of specific building blocks over others and is the possible emergence of proto-nucleotides sustained on the thermodynamic basis.

In order to address this question two different pathways have been modeled for the synthesis of the nucleotides (see **Figure 4.5** and **Table A4** for more details from the Appendices section):

1) Classic pathway: the condensation of a TC with a RU to obtain the N-nucleoside (reaction a) is *followed* by the condensation reaction of the nucleoside (Ns) with an IL to obtain the corresponding nucleotide (Nt) (reaction b). In the process two water molecules are released. The Gibbs free energy of the reaction is estimated as:

$$\Delta G_{(a+b)}^{\circ} = \left[\left(G_{Ns}^{\circ} + G_{H_2O}^{\circ} \right) - \left(G_{TC}^{\circ} + G_{RU}^{\circ} \right) \right]_a + \left[\left(G_{Nt}^{\circ} + G_{H_2O}^{\circ} \right) - \left(G_{Ns}^{\circ} + G_{IL}^{\circ} \right) \right]_b \qquad (4.2)$$

2) Alternative pathway: an IL condense with a TC to produce a TC-IL backbone monomer (reaction c). This reaction is followed by the condensation between the TC-IL Each RU (reaction d) to obtain the corresponding nucleotides. The $\Delta G_{(c+d)}^{\circ}$ is estimated as follows:

$$\Delta G_{(c+d)}^{\circ} = \left[\left(G_{TC-IL}^{\circ} + G_{H_2O}^{\circ} \right) - \left(G_{TC}^{\circ} + G_{IL}^{\circ} \right) \right]_c + \left[\left(G_{Nt}^{\circ} + G_{H_2O}^{\circ} \right) - \left(G_{TC-IL}^{\circ} + G_{RU}^{\circ} \right) \right]_d \qquad (4.3)$$

Neither pathway obey the Hess' law [46] since the change in the order of addition for the different components may induce a different geometry for the same nucleotide and hence, different associated energies.

There is a large volume of literature on the experimental classic synthesis of primordial nucleotides and the condensation reaction of a canonical or non-canonical base upon a sugar-phosphate derivative.

For instance, Mungi and coworkers (Mcw) [47] tested the reaction between the sodium salts of BA and ribose-5'-monophosphate (RMP) by heating the reaction mixture at 90°C for 3 hrs in dehydrating-rehydrating conditions and pH $\approx$ 7-9. The products were analyzed using **h**igh **p**erformance **l**iquid **c**hromatography (HPLC), **m**ass **s**pectrometry (MS) and **n**uclear **m**agnetic
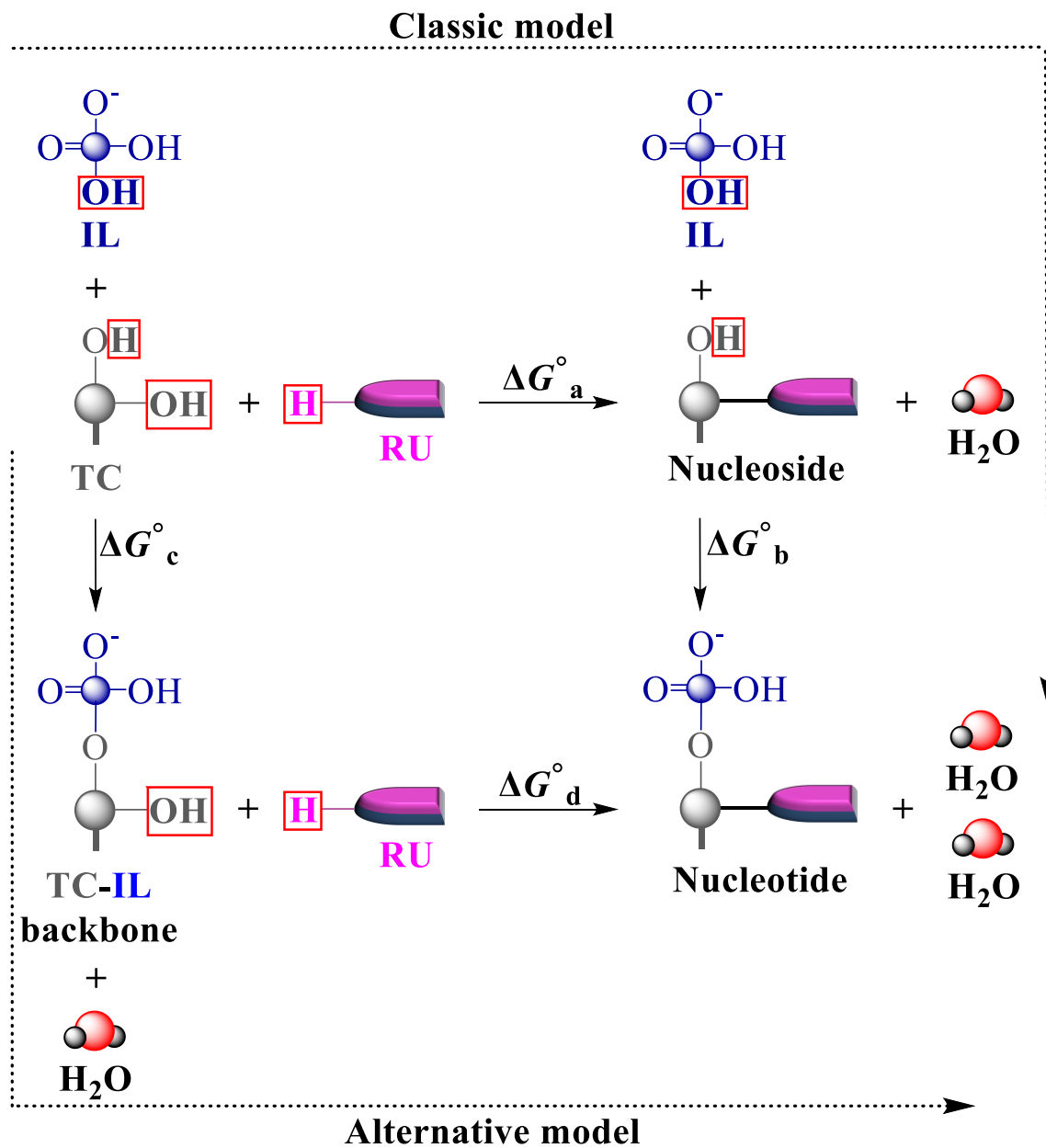
Figure 4.5 Two different models for constructing the *β*- and *α*-anomers of the nucleotides containing a sugar-like TC (2dRibf, Ribf, 2dRib, Rib and Tho) and the nucleotides with a non-sugar TCs (glycerol and glyceric acid). Reaction pathways referred to in the text and tables are labeled with lower-case letters: classic (a+b) and alternative model (c+d).

resonance (NMR) spectroscopy. Mcw obtained a mixture of β- and α-nucleotides in a combined yield of $\approx$ 40-85%. Most of nucleotides presented a C^5-glycosidic bond which is more stable than the N-counterpart at pH < 7 and high temperatures, but still the formation of N-glycosidic bonds was observed. The different anomers were identified using bidimensional NMR experiments. In subsequent steps Mcw also tested the corresponding BA nucleotides for their capacity to create polynucleotide strands in heating conditions at 90°C and pH $\approx$ 2 using POPC as catalyzer. The picks obtained in the mass spectra suggested the polymerization of the BA nucleotides. Additionally, Mcw also tried the glycosylation of BA on a RMP backbone oligomer. The authors analyzed the product mixture by HPLC and UV spectrometry observing signals suggesting the glycosylation of some BA to the ribose in the backbone. In summary, this paper suggests that BA can react with ribose following a similar synthetic route to the one been tested in this section.

Cafferty and coworkers (Ccw) [48] also reported the prebiotic synthesis of BA and MM ribonucleotides in aqueous solution from a mixture that contained each respective RU with RMP in a (1:1) proportion. Ccw heated BA and RMP at 20°C for 24 hrs at pH = 3-11 and obtained a complex mixture of β- and α-BA-C^5-RMP in a combined 80% of yield. A subsequent analysis of the purified mixture containing BA by HPLC and H^1-NMR showed that the products were C^5-glycosylated and that the β-configuration was prevalent with ($\approx$ 2:1) proportion. Similarly, for the MM its aqueous solution with RMP was heated at 65°C for 24 hrs at pH = 3-9 and a mixture of nucleotides was obtained in a combined yield of 33-55% depending on the reaction conditions. Bidimensional NMR experiments showed that in this case both anomers are produced in almost 1:1 and that MM N-glycosylate with RMP using one of the exocyclic NH_2 groups. Surprisingly, the N-glycosidic bond between MM and RMP showed to have a significant stability (half-life time = 6 months) to hydrolysis compared to the N-glycosidic bond in canonical nucleotides. This was determined by comparing the dissociation constant (K_d) of the MM ribonucleotide with respect to the ones reported by Miller and coworkers [49] obtaining that the MM nucleotides have a K_d = 3.7 μM meanwhile, e.g. the U ribonucleotide has a K_d = 700 μM.

Fialho [7] have also tested the alternative synthesis of non-canonical nucleotides starting from one mol equivalent of each free RU (BA, MM and TAP) with 1 mol of RMP. The reaction conditions changed depending on the RU used and a set of pH were tested to find the optimum acidic conditions. For instance, when combining BA and RMP in aqueous solution at pH = 9 for 24hrs at room temperature a mixture of exclusive BA-C^5-RMP was found with a combined yield

of 82% and a proportion of β:α of (7:3). Similarly, for the MM the reaction was heated at 65°C for 24 hrs at pH = 5 and the corresponding N-glycosides were obtained in 55% of yield including the β- and α-configurations in 1:1 proportion. The lowest yields of the reaction were obtained for the TAP ribonucleotides with 25% of combined yield for a 3:2 mixture of the β- and α-configurations when heating the reaction for 24hrs at 65°C in a pH = 1.

Other studies have discussed the glycosylation of canonical and non-canonical RUs onto a cyclic phosphate TC. e.g., in a special issue of *Life* dedicated to the prebiotic origins and evolution of RNA two papers by Scott and coworkers [50] and Martin *et al*. [51] discussed the synthesis of 2'-3' cyclic phosphate nucleotides and their role in the polymeric assembly of RNA.

Kim and Benner [52] tested the reaction between ribose-1'-2'-cyclic phosphate with A obtaining around 15% of yield for the β-adenosine-2'-monophosphate when heating the reaction for 18hrs at 85°C. No product was obtained when repeating the reaction with the pyrimidines U and C.

In another experiments KB combined the nucleotide containing A with borate and urea and were able to produce the β-adenosine-5'-monophosphate when heating the reaction at 90°C. When repeating the reaction with threose-1'-2'-cyclic phosphate by heating the reaction for 18 hrs at 70°C KB found that this sugar seems to be slightly more reactive than ribose since the corresponding β-U-2'-threose monophosphate was obtained in 0.4% of yield and two main products with a combined yield of 70% were obtained for the A threonucleotides. One of these products showed to have an N-glycosidic bond between the Tho and the exocyclic NH_2 of the A.

The previous studies are a sample of the abundant literature referencing that the alternative glycosylation of non-canonical RUs onto a RMP backbone monomer is experimentally feasible.

But what about the reaction between an IL and the TC? Guntha *et al*. [53] tested the phosphorylation of different sugars obtaining 87% combined yield for 1'-2'- and 2'-3'-cyclic phosphates for D-erythrose, 59% of L-2'-threose monophosphate upon reaction with diaminophosphate, 79% of combined yield for the 2'-ribofu(py)ranose monophosphate and the 3'-ribopyranose monophosphate) and 81% of yield for the phosphorylation of D-glyceraldehyde with amidotriphosphate and diaminophosphate. This paper shows that even when the reactants are not specifically the dihydrogen salt of phosphate the reactions proceed in a certain extent.

In this chapter we also model the nucleotides containing a $HAsO_3^-$ instead of a HPO_3^- group as IL. This addresses the following question: could the synthesis of canonical and non-canonical

nucleotides containing As instead of P be feasible?

The possibility of As as a proto-IL has been posed since the studies by Wolfe-Simon and coworkers (WScw) [21]. WScw performed preliminary synchroton X-ray diffraction experiments to suggest that the cyanobacteria GFAJ-1 could contain a genetic material with As instead of P. *But if As was present in the first RNAs how was it replaced by P?*

After the controversial results presented by WScw some studies have suggested that some arsenate derivatives have a higher propensity to hydrolyze than their phosphate analogs [54, 55]. The possibility of a sugar-arsenate backbone as part of the genetic polymers have been discussed by a few papers published by Mládek and coworkers (Mcw) on the basis of quantum theoretical modeling. For instance, in [56] Mcw discussed the resistance of the arsenic and phosphate esters to hydrolysis. Mcw modeled the hydrolysis mechanism and activation energy barriers (E_a) in Kcal/mol for different P and As esters that contained saturated groups with an increasing number of C atoms. This was done to test the hypothesis proposed by Baer *et al*. [57, 58]. Baer's hypothesis suggests that the activation energy (E_a) barriers are directly proportional to the size of the ester substituents since they may induce steric effects that hinder the addition of the water. Mcw used a IEFPCM model to model the effects of water implicitly and all calculations were done using the DFT- PBE1PBE/6-311++G(2d,2p) and B3LYP/6-311++G(2d,2p). Their results suggest that, overall, the As-esters had a lower energetic barrier than their P counterparts and that increasing the size of the R group in the ester did not really decrease the reactivity of these As-ester with H_2O. This result supports the idea that if the first nucleic acids had As, then eventually was replaced by phosphate in aqueous solution since the phosphodiester bond is more hydrolytically stable than the arsenate diester.

On another study [59] Mcw also modeled the geometric properties of a backbone containing two 2'-deoxyribofuranose linked by a XO_4^-, where X: P or As. They found that after the geometry optimization the bond distances between the X and the O4' and the O5' were similar for both As and P, been slightly longer for the case of the As. Also, the angle between the two O4', O5' and the X element was similar for both As and P. These results support the idea that As can create a similar flexible DNA backbone to P. They modeled all the structures at the DFT- TPSS-D/6-311++G (3df, 3pd) and used the IEFPCM to model the solvation effects.

Mcw have also modeled the hypothetical condensation reaction between the dihydrogen phosphate ($H_2PO_4^-$) or arsenate ion ($H_2AsO_4^-$) with 2dRibf obtaining 1.0 and 1.1 kcal/mol for the

both sugar monophosphates respectively. When modeling the reaction in implicit solvation the energies were 0.4 and 0.1 kcal/mol respectively.

In a recent paper [23] Castanedo and Matta (CM) estimate the changes in free energy as a measure for the plausibility of two synthetic pathways (classic and alternative) for canonical DNA and RNA nucleotides similarly to this study. The results obtained by CM suggest that the order of adding the reactants to obtain nucleosides and nucleotides matter. The sequence [base + sugar → nucleoside + water + H_2PO4^- → nucleotide + 2 H_2O] emerges as the thermodynamically favored for the synthesis of nucleotides in vacuum. Calculations accounting for implicit aqueous solvation makes either synthetic pathway thermodynamically unfavorable. This last observation is consistent with the well-known "water problem" (see Fig.3-4 and Table 5 in Castanedo and Matta's paper).

Let's discuss first the changes on the free energy for the canonical and non-canonical nucleotides obtained from both pathways containing HXO_4^- (X: P or As) and 2dRibf and Ribf, since most of the experimental evidence discussed above refers to these two CHs.

Figures 4.6 and **4.7** represent the $\Delta G^{\circ}_{(a+b)}$ for the classic synthesis of the nucleotides from 2dRibf containing HPO_4^- and $HAsO_4^-$ respectively. The synthesis of both canonical and non-canonical nucleotides is favored in vacuum environment meanwhile in aqueous environment most of energies are within the intrinsic error of the method ($\approx$ 17 kJ/mol) or the reactions are thermodynamically unfavorable. In this environment the most unfavorable reactions correspond to the BA-N and CA nucleotides. In vacuum in most of cases either β-$^2T_3$ or β-$^3T_2$ configurations are more favored than their α-counterparts. The most favored nucleotide overall contains BA C^5-glycosilated with a magnitude of $\approx$ -146 kJ/mol or -34.9 kcal/mol. Both bar graphs show similar patterns reflecting that the reactions proceed similarly in the presence of P or As.

This result is in agreement with the results from CM [23] for the canonical nucleotides.

Figures 4.8 and **4.9** represents the $\Delta G^{\circ}_{(c+d)}$ for the alternative (c+d) synthesis of the canonical and non-canonical nucleotides containing 2dRibf. The bar graphs show a similar trend like the one described for the similar nucleotides obtained from the classic pathway. There are not noticeable differences between the energetic outcome when the IL is replaced. In vacuum most of β-anomers of the canonical or non-canonical nucleotides are favored over their α-counterparts. There are some exceptions; the T nucleotides with phosphate and A, C and T with arsenate where the α-$^3T_2$ is more favored. In general, the canonical nucleotides obtained from this pathway are also

316
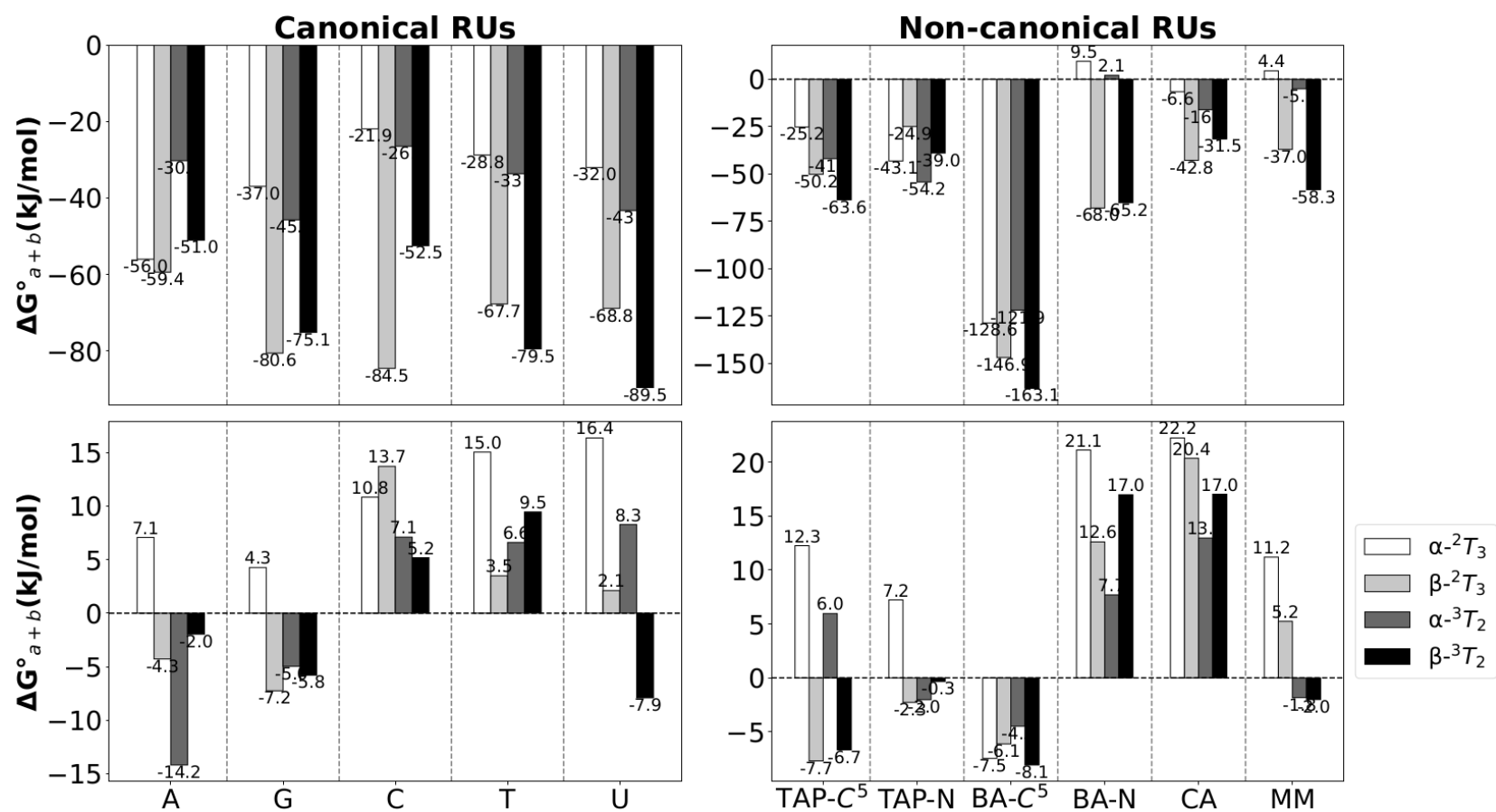
Figure 4.6 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^--2dRibf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

317
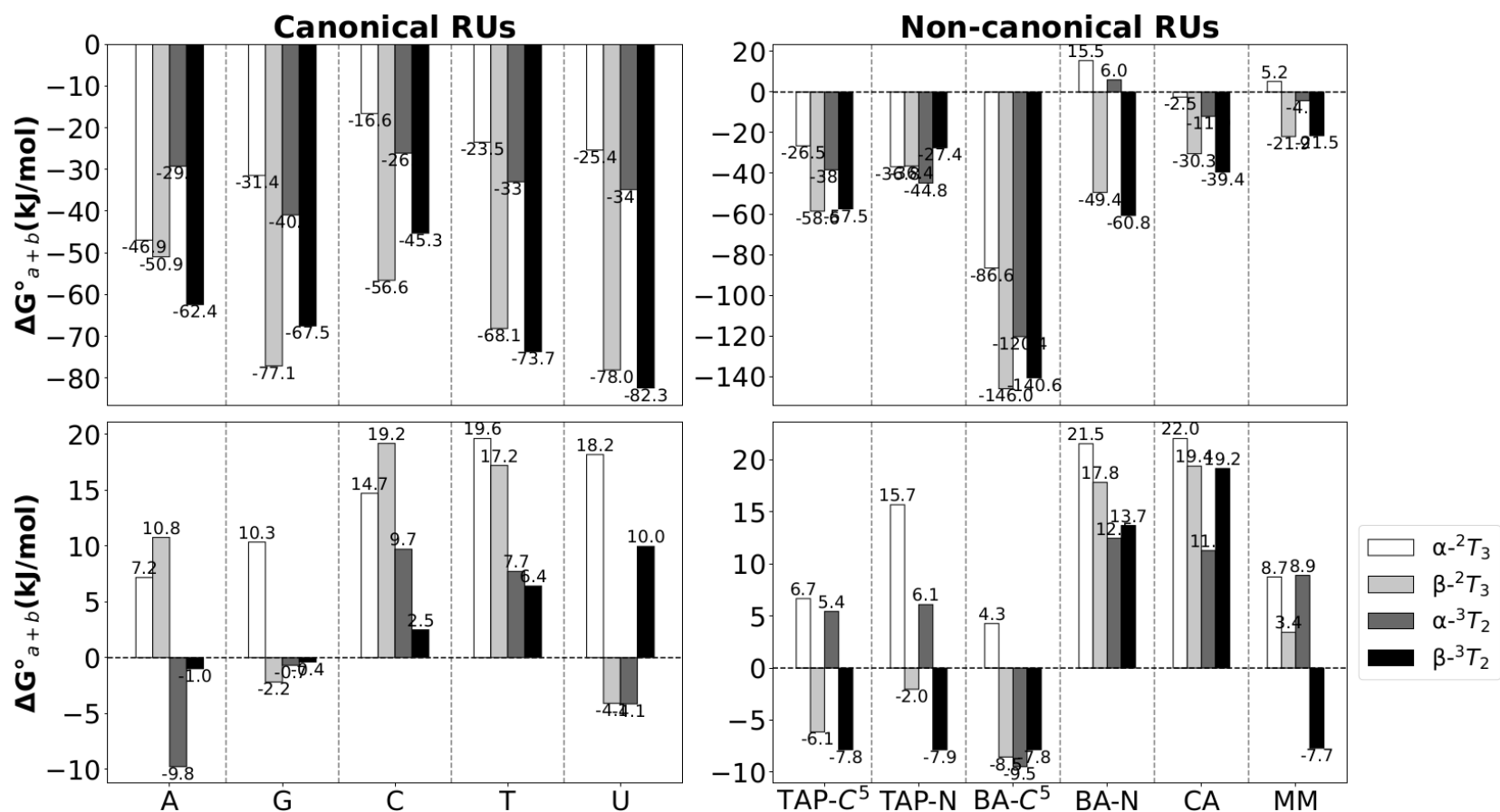
Figure 4.7 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HAsO$_3^-$-2dRibf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
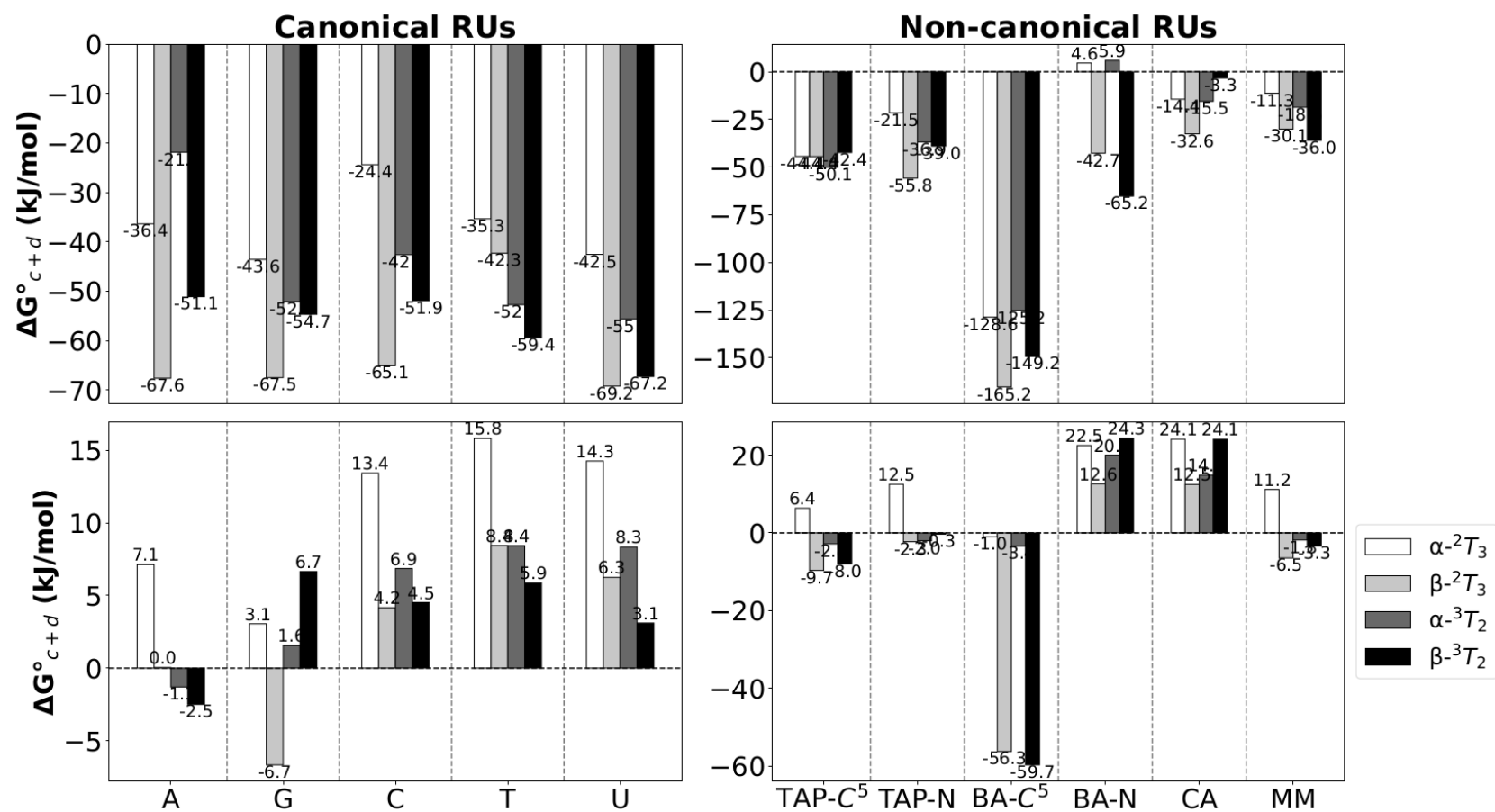
Figure 4.8 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the 2'-deoxyribofuranose mono- nucleotides phosphate (HPO_3^--2dRibf-RU). (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
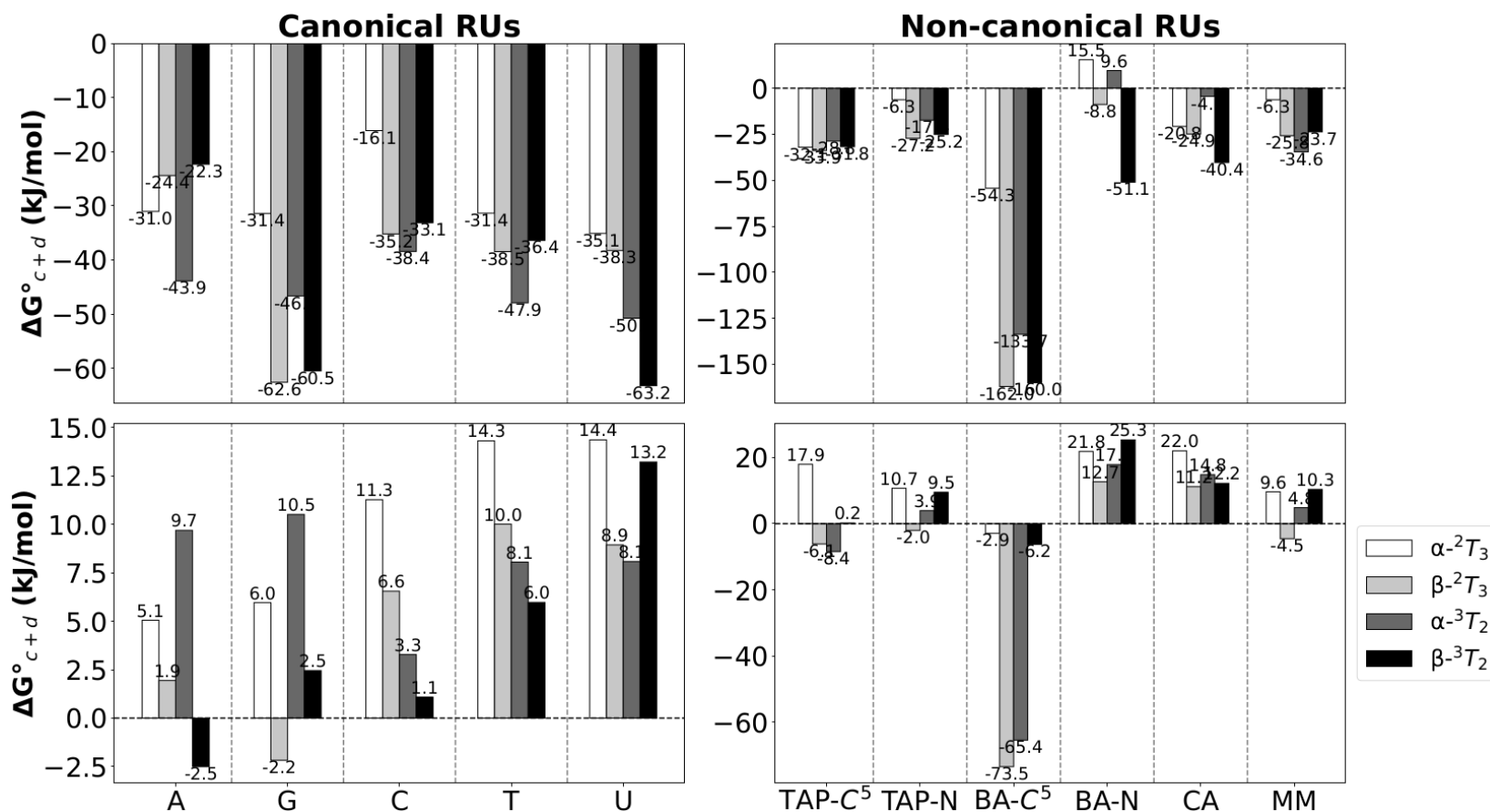
Figure 4.9 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^{-}$-2dRibf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

more favored than the non-canonical ones. In aqueous solution most of $\Delta \overset{\circ}{G}_{(c+d)}$ are negligible with the exceptions of the BA-N and CA nucleotides, which synthesis is not favored. The nucleotide with the most negative ΔG^0 contains the C-glycoside of BA. The formation of the β-configuration is significantly favored in both environments when HPO_4^- is present and for the arsenate derivatives in vacuum the β-anomer is favored but in aqueous solution there is almost equal propensity to obtain the β-$^2T_3$ and the α-$^3T_2$. The formation of the β-$^2T_3$ is around 17 kJ/mol more favored when As is present instead of P. The rest of the energetic differences are comparable for the nucleotides with both ILs.

A comparison of the energies between both pathways suggests that the classic formation of the β-configuration is in general more favored for the canonical nucleotides in vacuum with phosphate and arsenate. For the non-canonical nucleotides most of energies are comparable.

The present analysis suggests that in general: 1) the classic pathway is preferred, except for the BA-C^4 nucleotides where the alternative pathway in aqueous environment favors the formation of the nucleotide, 2) when the IL is exchange in the classic pathway it does not affect overall the preferential formation of the β-anomer, meanwhile in the alternative pathway some canonical nucleotides preference for the β- or α-configuration changes. This last result points out a possible problem when the canonical nucleotides containing $HAsO_4^-$ as IL are synthesized using an alternative pathway, 3) these results validate to an extent the "water problem" for the canonical nucleotides containing a pyrimidine and non-canonical nucleotides with CA and BA-N.

Let's analyze the nucleotides containing Ribf. In the case of the nucleotides modeled using the classic pathway (a+b) (**Figure 4.10**) with a HPO_3^- the classic synthesis of all canonical and non-canonical nucleotides modeled in vacuum is predicted to be favored to some extent. Only in the case of the A, C, T, TAP-C^5, TAP-N, BA-C^5, BA-N and MM all different β-configurations are more favored than the α-counterparts. The BA-C^4 nucleotides with either IL are the most favored.

For the synthesis of the nucleotides modeled in implicit solvation with HPO_4^- non energetic differences are conclusive, meanwhile for the case of the nucleotides with $HAsO_3^-$ (**Figure 4.11**) the only nucleotides that contains reactions energies beyond the error of the method is the $HAsO_3^-$-Ribf-C^5-BA with -18.5 kJ/mol. When the solvation effects are implicitly included the α-configurations of the canonical nucleotides with T and U and either IL are the less favored with energies in the order of $\approx$ 20 kJ/mol, meanwhile for the non-canonical nucleotides with either IL the α-configurations of the BA-N and CA nucleotides are the less favored with energies in the order
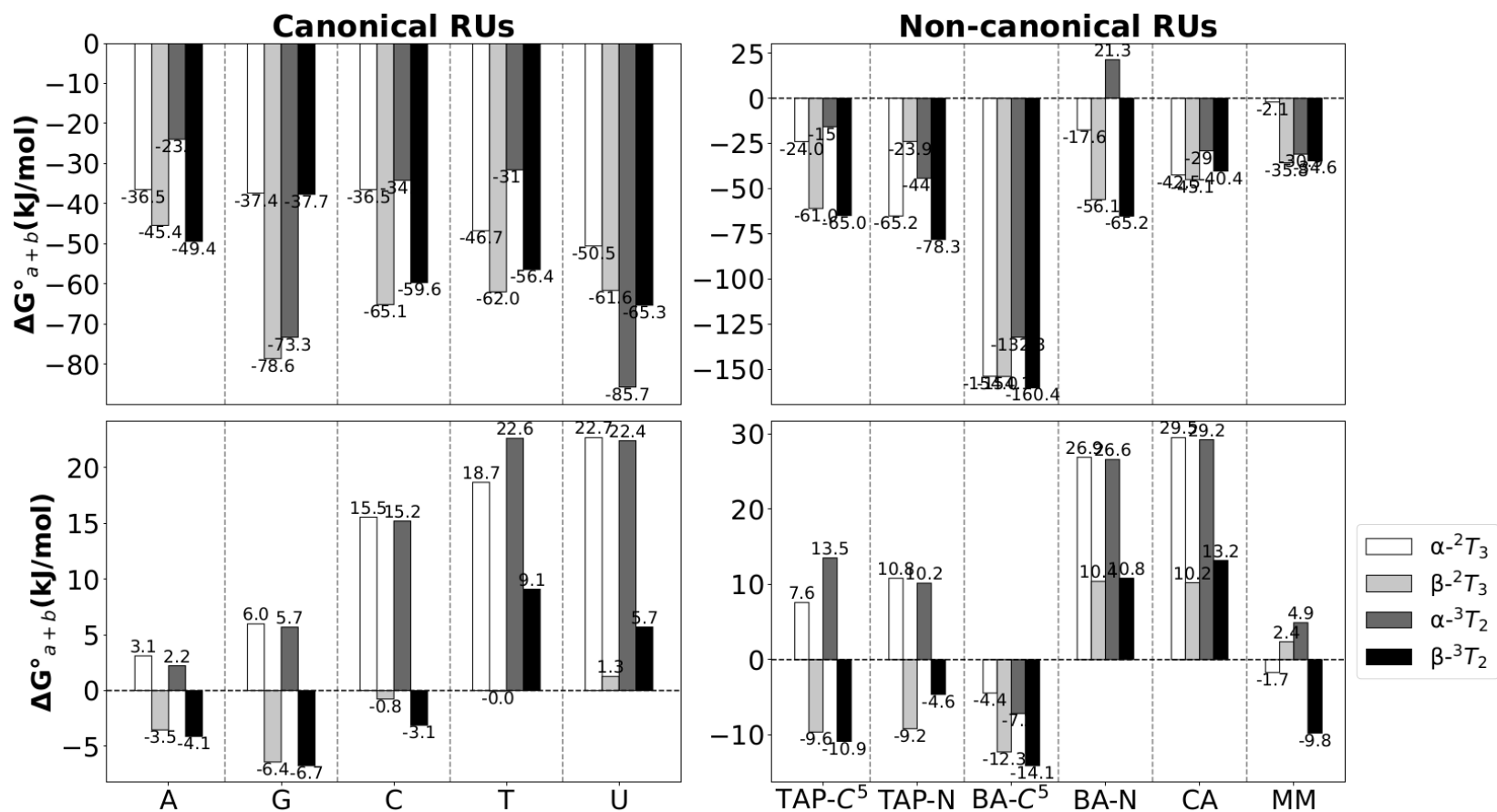
Figure 4.10 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO₃⁻-Ribf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

Figure 4.11 Comparison of Gibbs energies of reaction ($\Delta G°$) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^-$-Ribf-RU. (Top) $\Delta G°$, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) $\Delta G°$, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

of $\approx$ 30 kJ/mol.

For the case of the HPO_3^--Ribf-RU nucleotides obtained using the alternative (c+d) pathway in vacuum (**Figure 4.12** and **4.13**) the α-$^2T_3$ and α-$^3T_2$ forms of the canonical HPO_3^--Ribf-A in vacuum are favored (**Figure 4.12**). In the case of the G nucleotides the β-$^3T_2$ is the most favored, but the α-anomers can also be produced. For the case of the nucleotides containing a pyrimidine base in the case of the C nucleotides the α-$^2T_3$, β-$^2T_3$ and the α-$^3T_2$ has almost the same negative ΔG^0. For the T and U nucleotides, the β-$^3T_2$ is the most favored configuration. In the case of the non-canonical nucleotides in vacuum only TAP-C^5, BA-C^5 and BA-N contain at least a β-configuration that is favored. In aqueous solution the β-$^2T_3$ of A and G nucleotides are favored, meanwhile for the non-canonical nucleotides only the β-$^2T_3$ of TAP-C^5 and both β-$^2T_3$ and β-$^3T_2$ of the BA-C^5 nucleotides are favored. The rest of the ΔG° are negligible.

In the case of the alternative synthesis of the nucleotides with arsenate in vacuum (**Figure 4.13**) and a canonical base the α-anomers of the A and G nucleotides emerge as the most favored, meanwhile for the pyrimidine nucleotides the β-$^3T_2$ is the most favored. For the case of the non-canonical nucleotides with the exception of the TAP-N the synthesis of at least one of the β-anomers is favored. In aqueous environment most of energies are inconclusive with the only exception of the BA-C^5 nucleotide where the β-$^2T_3$ and α-$^3T_2$ have almost the same probability with a $\Delta G^\circ_{(c+d)} \approx$ -77.0 kJ/mol. The formation of the α-anomers of T, U, BA-N and CA are thermodynamically unfavored.

A comparison of both pathways for the nucleotides with an HPO_3^- shows that for the canonical nucleotides the alternative pathway tends to favor more the formation of α-configurations mostly for the purine nucleotides in vacuum. Meanwhile for the non-canonical nucleotides the alternative pathway emerges as the preferable route for the desirable β-configuration.

Analyzing the $\Delta G^\circ_{(c+d)}$ for the nucleotides modeled in aqueous environment the β-$^2T_3$ of A (-23.2 kJ/mol), G(-24.3 kJ/mol), TAP-C^5 (-22.7 kJ/mol) and both β-anomers of BA-C^5 (β-$^3T_2$ {-79 kJ/mol}, β-$^3T_2$ {-62.3 kJ/mol}) are favored. Overall, when the effects of solvation are included the synthesis of the canonical nucleotides with T, U and the non-canonical BA-N and CA are estimated as thermodynamically unfavored.

When $HAsO_3^-$ is the IL the analysis of $\Delta G^\circ_{(a+b)}$ and $\Delta G^\circ_{(c+d)}$ for both pathways show similar trends compared to the HPO_3^-. In the case of the $HAsO_3^-$-Ribf-C^5-BA from (c+d) both β-$^2T_3$ and
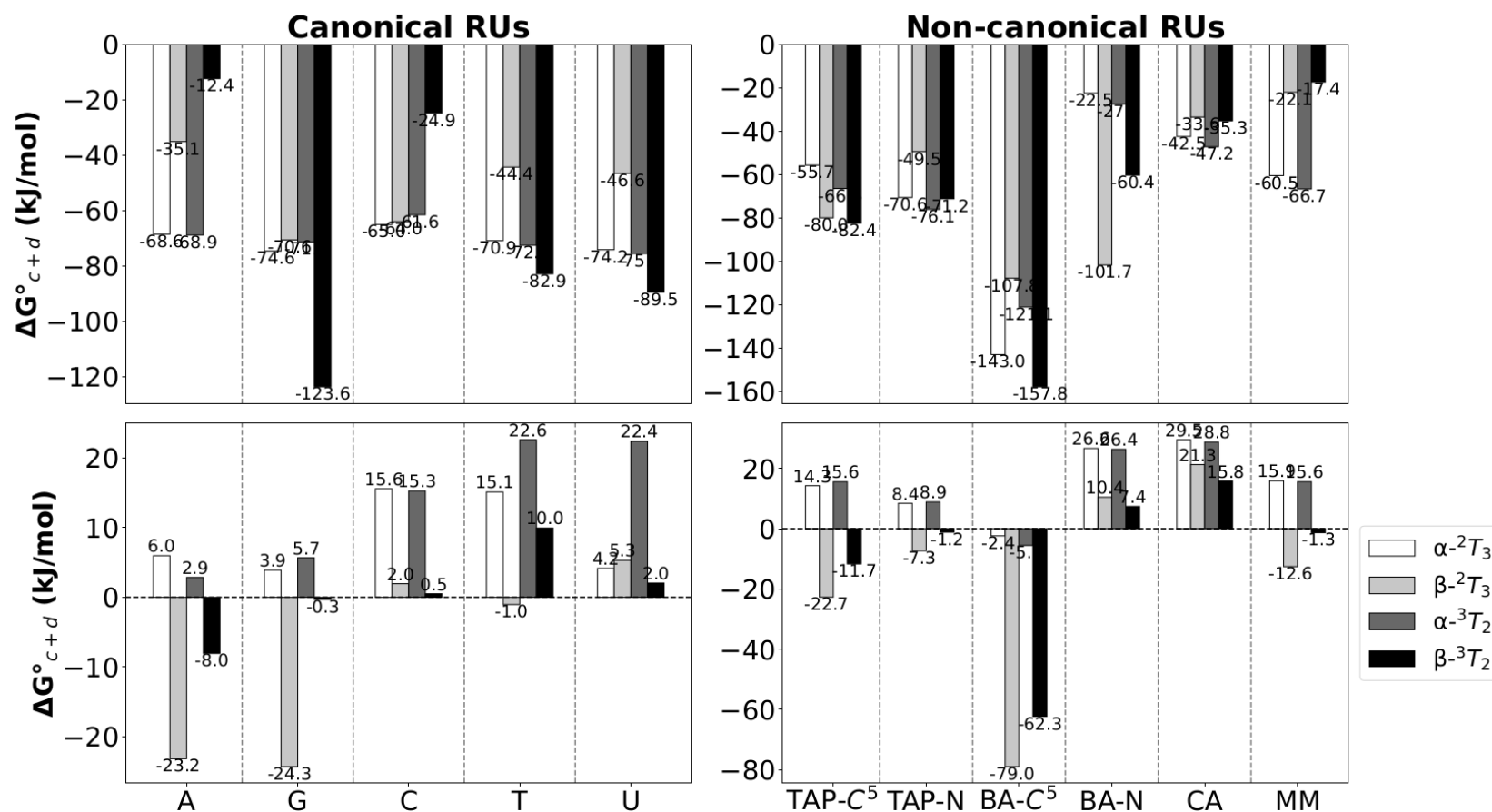
Figure 4.12 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^--Ribf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
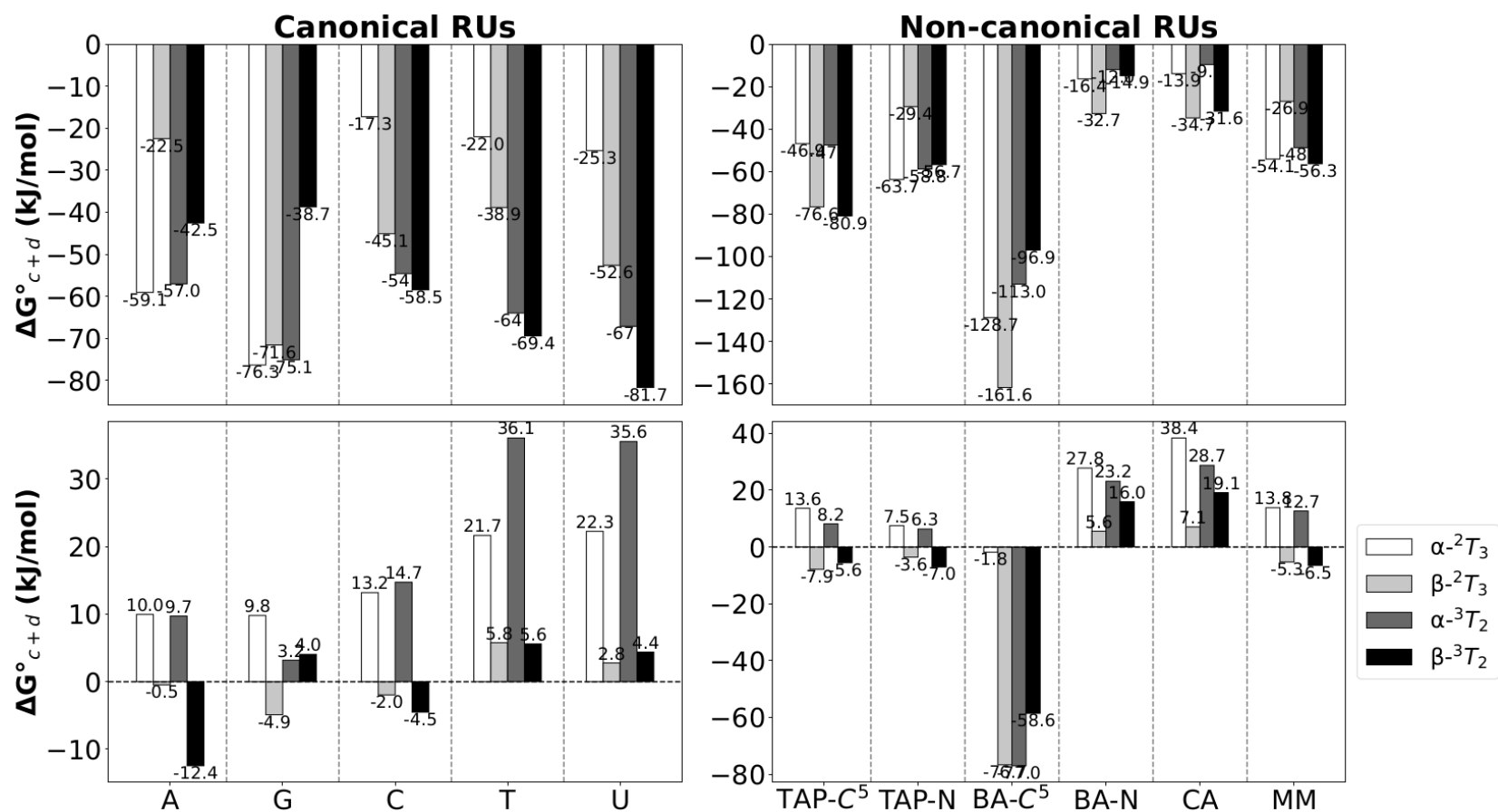
Figure 4.13 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^-$-Ribf-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

α-^{3}T$_2$ are almost equally favored in both environments.

Figures 4.14 and **4.15** represent the $\Delta G^{\circ}_{(a+b)}$ for the D-Tho nucleotides with P and As respectively. In the case of the threonucleotides with HPO$_3^-$ (**Figure 4.14**) all canonical nucleotides in vacuum are thermodynamically favored. For the Nts with a purine base the α-^{3}T$_2$ is more favored, meanwhile when the base is a pyrimidine the formation of the β-^{3}T$_2$ is more thermodynamically favored. For the non-canonical Nts in vacuum the α-^{3}T$_2$ seems to be more favorable with the exception of the BA-C^5 nucleotides in which the β-^{2}T$_3$ is more favored by -107.8 kJ/mol. In aqueous environment the β-^{2}T$_3$ of all canonical and non-canonical nucleotides emerges as the more favored, especially for the Nts with the BA C^5-glycosylated (-51.5 kJ/mol). The N-glycoside of BA is thermodynamically unfavored.

For the HAsO$_3^-$-Tho-RU Nts (**Figure 4.15**) similar trends are observed. The most favored Nt in vacuum is the β-^{2}T$_3$ of BA-C^5. In aqueous solution the synthesis of the β-^{2}T$_3$ anomer for all canonical and TAP-C^5, TAP-N, BA-C^5 and MM Nts is favored. In this environment the most favored Nts are the HAsO$_3^-$-Tho-A ($\Delta G^{\circ}_{(a+b)}$ = -43.1 kJ/mol), followed by the HAsO$_3^-$-Tho-C^5-TAP ($\Delta G^{\circ}_{(a+b)}$ = -39.7 kJ/mol) and HAsO$_3^-$-Tho-C^5-BA ($\Delta G^{\circ}_{(a+b)}$ = -36.9 kJ/mol).

Figure 4.16 shows the values of $\Delta G^{\circ}_{(c+d)}$ for the HPO$_3^-$-Tho-RU Nts obtained from the alternative (c+d). In vacuum when there is a canonical RU all anomers in all sugar ring conformations are favored. The β-^{2}T$_3$ is the most favored in vacuum or aqueous solution for all canonical nucleotides. In vacuum all canonical anomers are almost equally favored. In aqueous solution the α-^{2}T$_3$ of the G Nt is unfavored by a $\Delta G^{\circ}_{(c+d)}$ ≈ 28 - 37 kJ/mol. Overall, the synthesis of the BA-C^5 Nt is the most favored with $\Delta G^{\circ}_{(c+d)}$ = -133.0 kJ/mol.

In aqueous environment a similar trend is observed compared to the non-canonical Nts from the classic pathway is observed. For the reactions modeled using the (c+d) pathway in aqueous environment the β-^{2}T$_3$ form of all D-Tho nucleotides is the most favored. In this case the most favored β-^{2}T$_3$ are for the HPO$_3^-$-Tho-A ($\Delta G^{\circ}_{(c+d)}$ = -54.0 kJ/mol), HPO$_3^-$-Tho-A ($\Delta G^{\circ}_{(c+d)}$ = -57.8 kJ/mol), HPO$_3^-$-Tho-C^5-BA ($\Delta G^{\circ}_{(c+d)}$ = -51.5 kJ/mol), HPO$_3^-$-Tho-N-BA ($\Delta G^{\circ}_{(c+d)}$ = -45.3 kJ/mol), HPO$_3^-$-Tho-C^5-TAP ($\Delta G^{\circ}_{(c+d)}$ = -41.0 kJ/mol), HPO$_3^-$-Tho-MM ($\Delta G^{\circ}_{(c+d)}$ = -40.0 kJ/mol).
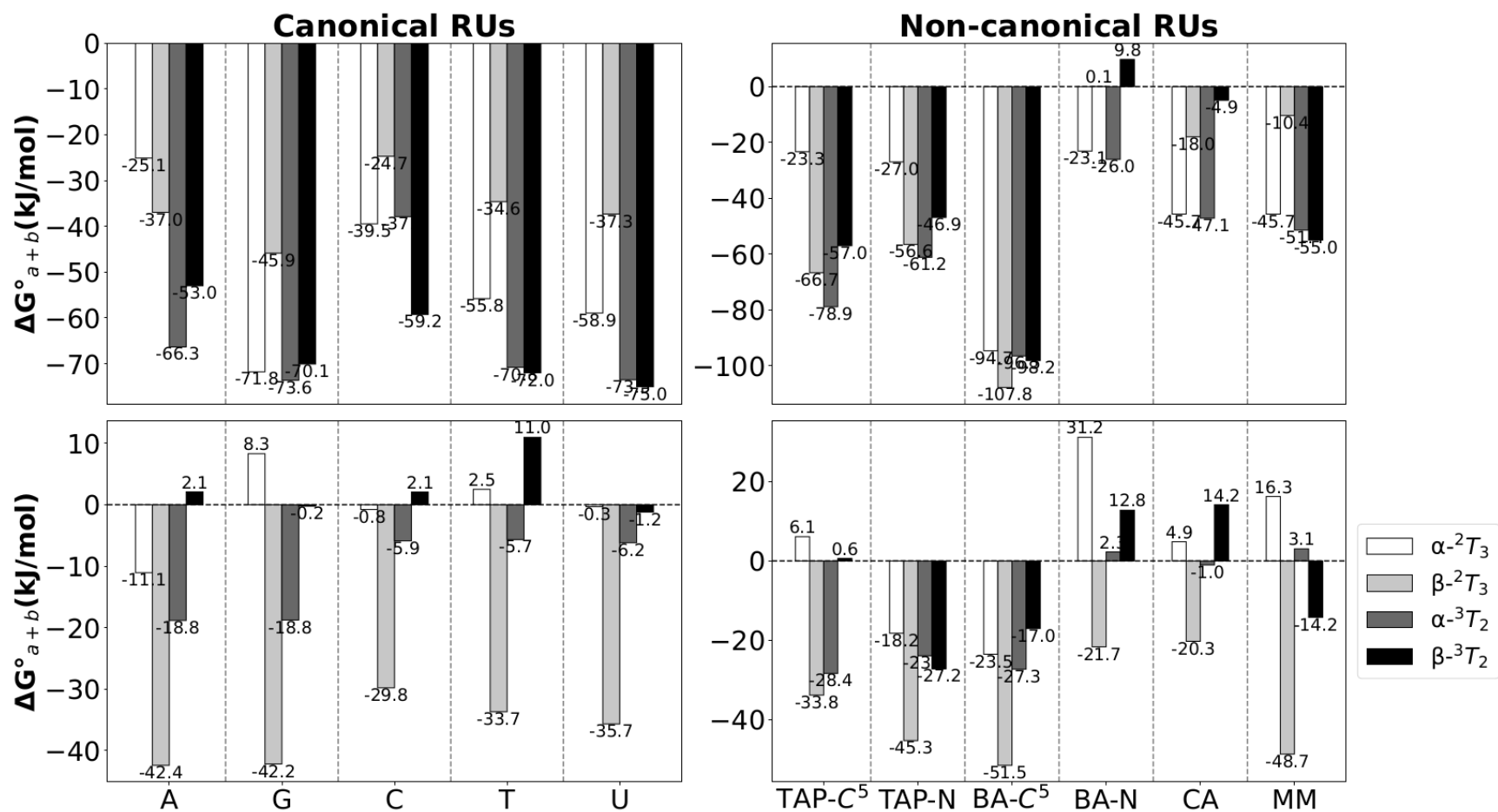
Figure 4.14 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO$_3^-$-Tho-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
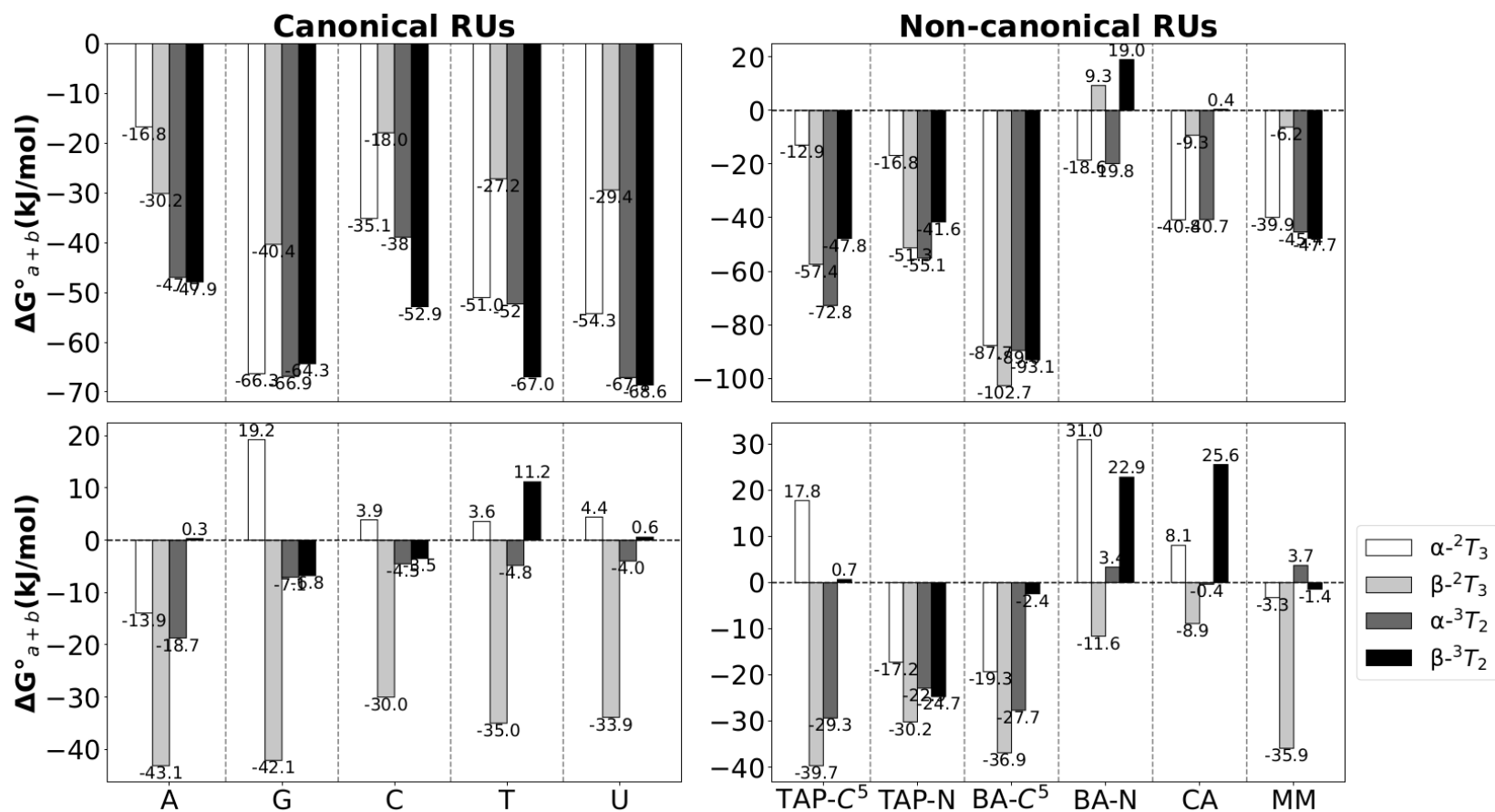
Figure 4.15 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^-$-Tho-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
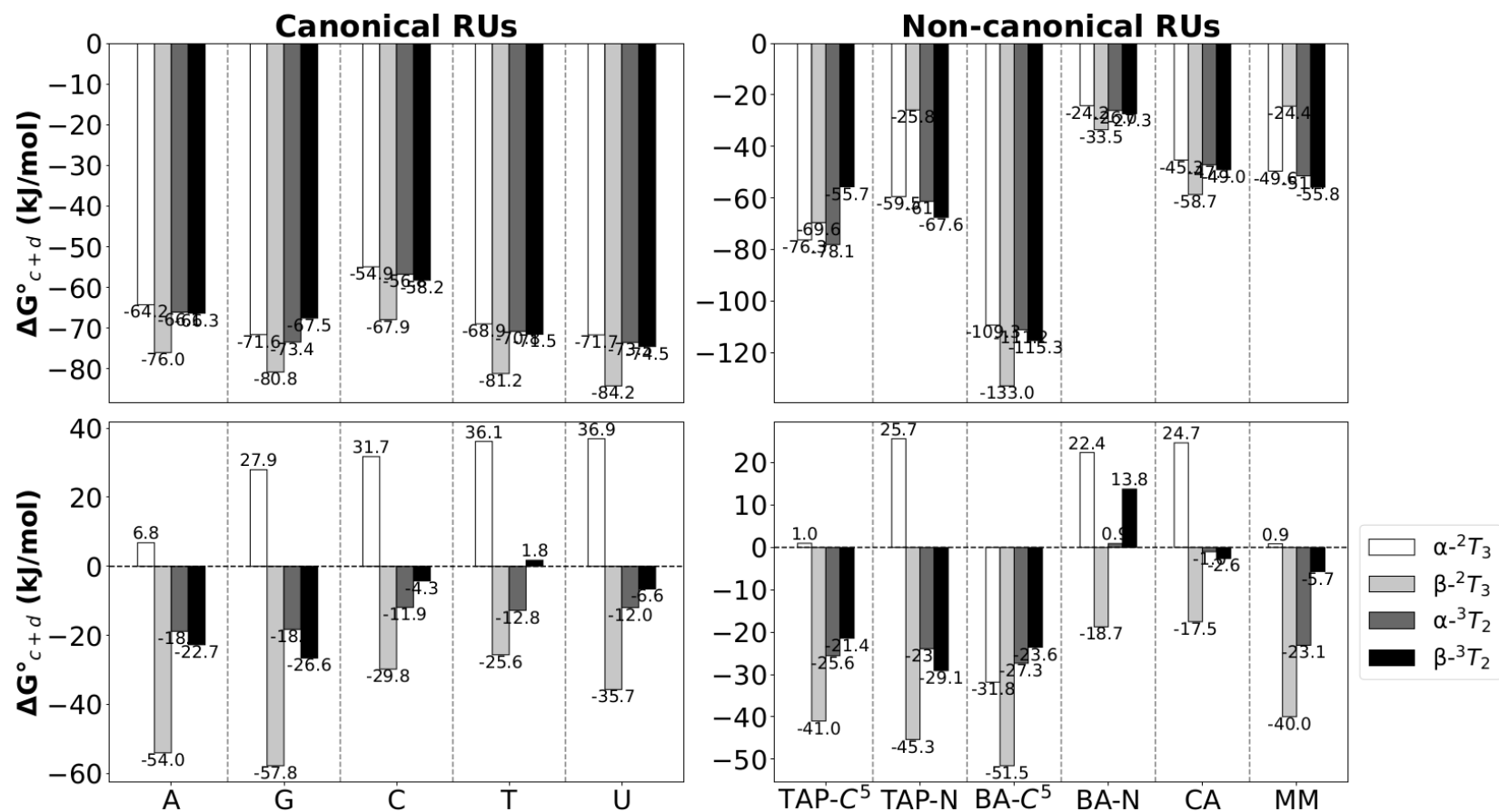
Figure 4.16 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^--Tho-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

When the hydrogen-phosphate is replaced by hydrogen-arsenate (**Figure 4.17**) similar trends are observed compared to the phosphate Nts. In this case the BA-C^5 nucleotide emerges as the overall more favored in vacuum (-131.2kJ/mol). In aqueous solution the β-$^2T_3$ is the most favored configuration for the A, G, U, TAP-C^5, TAP-N and MM. The synthesis of the α-$^2T_3$ of all canonical nucleotides and BA, CA and MM in aqueous solution is unfavored. There are three cases in which the energies jump to > 200 kJ/mol for the A, G in vacuum and the CA in aqueous environment. Upon inspection of the molecular geometries of these Nts it is noticed that the molecular structure of the nucleotides is distorted and the glycosidic bonds are broken, hence not reasonable conclusions can be achieved from these $\Delta G^{\circ}_{(c+d)}$.

Let's turn gears to the analysis of the thermodynamic feasibility for the synthesis of the canonical and non-canonical nucleotides containing the P-forms of D-2'-deoxyribose and D-ribose: 2dRib and Rib respectively.

Figure 4.18 shows the bar graphs for the $\Delta G^{\circ}_{(a+b)}$ of the HPO$_3^-$-2dRib-RU nucleotides. In the case of the canonical nucleotides in vacuum all Nts are favored, been the β-$^4C_1$ of the purine Nts the most favored, meanwhile for the pyrimidines the β-$^4C_1$ and α-$^1C_4$ are both almost equally favored. For all non-canonical nucleotides obtained in vacuum the β-$^4C_1$ is the most favored, with the exception of the BA-C^5 Nts in which the α-$^1C_4$ is the most favored. For the monophosphate Nts in aqueous environment all α-$^1C_4$ forms are unfavored for the canonical nucleotide and for TAP-C^5, BA-C^5, BA-N, CA and MM. For the CA and MM Nts all four α-$^1C_4$, β-$^1C_4$, α-$^4C_1$ and β-$^4C_1$ are unfavored.

In the case of the 2dRib Nts with arsenate obtained from the classic (a+b) route (**Figure 4.19**) similar trends are observed, which suggests that the geometries of reactants and products when P is replaced by As can be similar with a 2dRib TC.

When the synthesis of the D-2dRib nucleotides follows the alternative pathway (c+d) similar patterns to the ones observed for the classic pathway are observed.

For instance, an analysis of **Figure 4.20** shows that the canonical nucleotides containing HPO$_3^-$ obtained from the (c+d) pathway in vacuum are more favored overall than in the classic pathway.

For the case of the non-canonical nucleotides in vacuum similar energies are observed with scarce differences, e.g., the α-$^4C_1$ is favored for the TAR-C^2 from the alternative pathway. In this
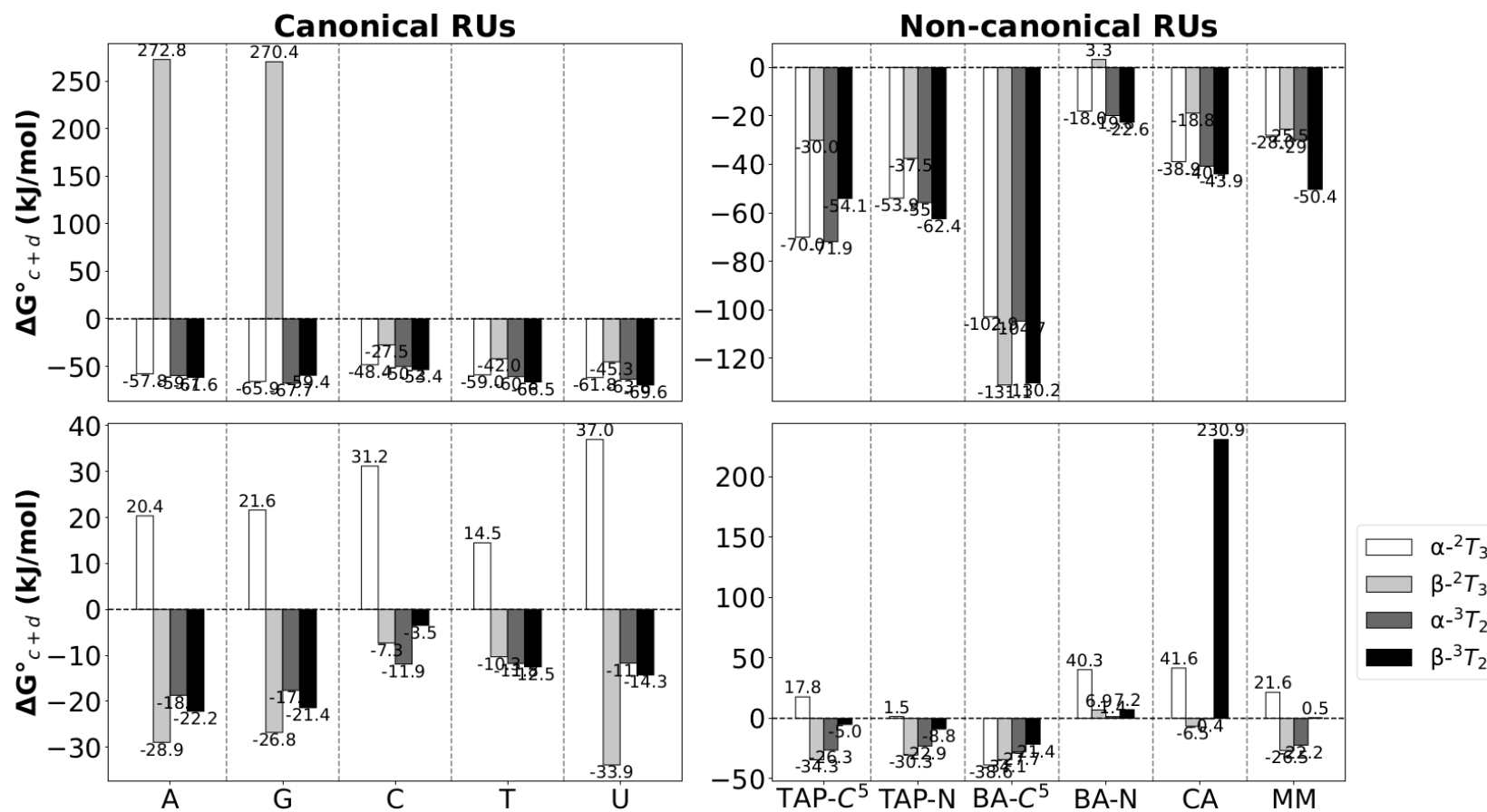
Figure 4.17 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HAsO$_3^-$-Tho-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
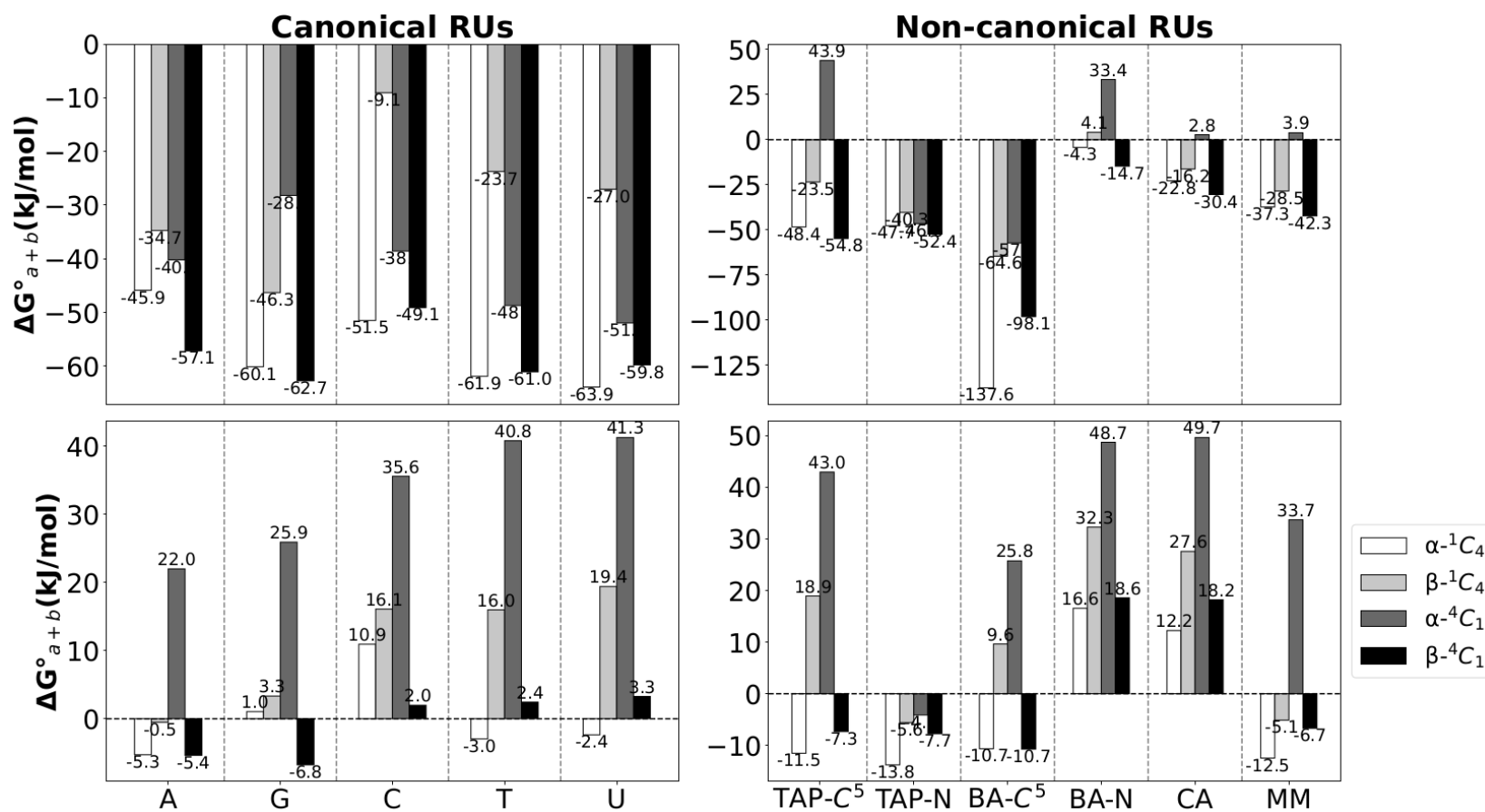
Figure 4.18 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^{-}-2dRib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
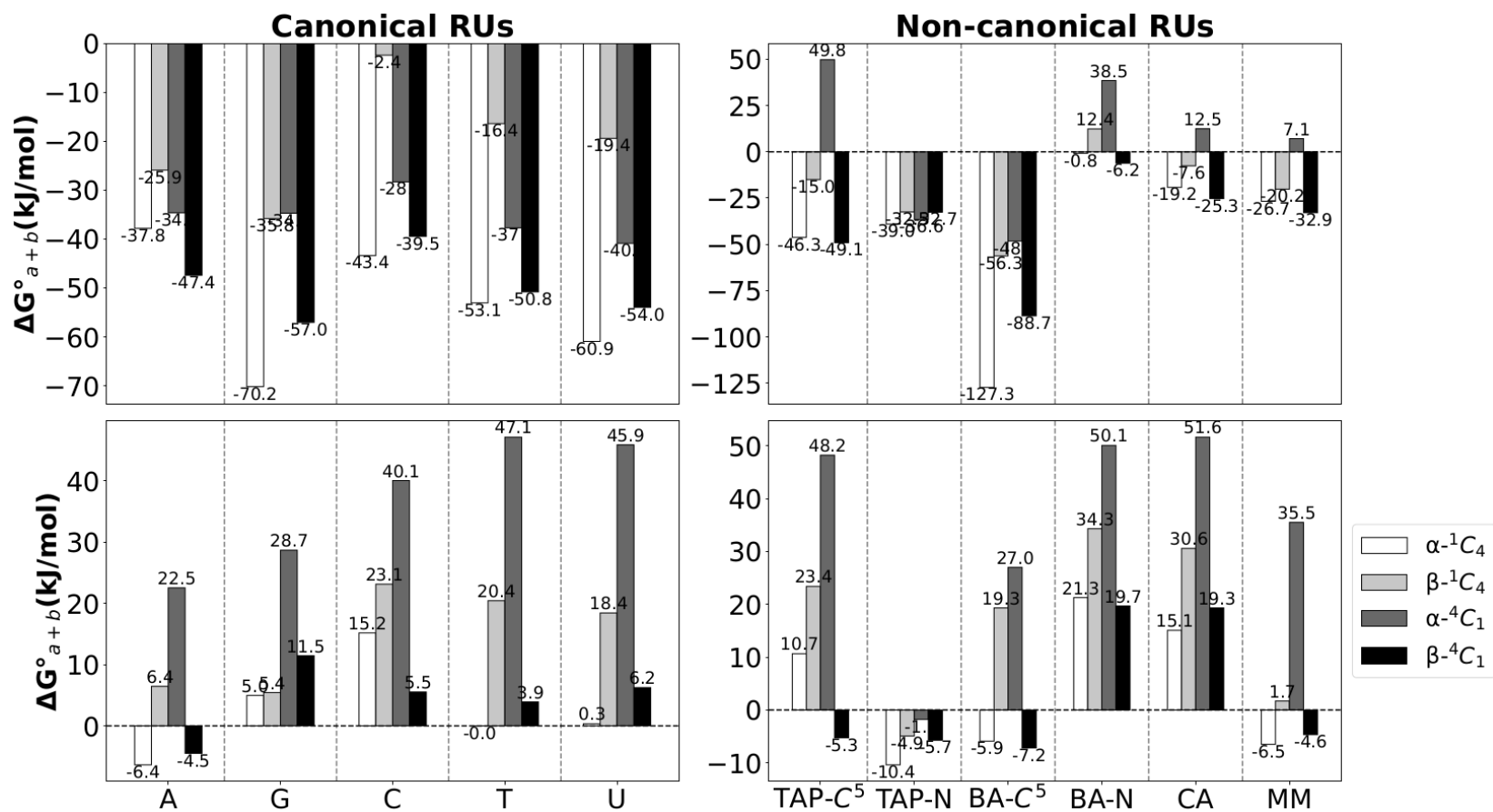
Figure 4.19 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^-$-2dRib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
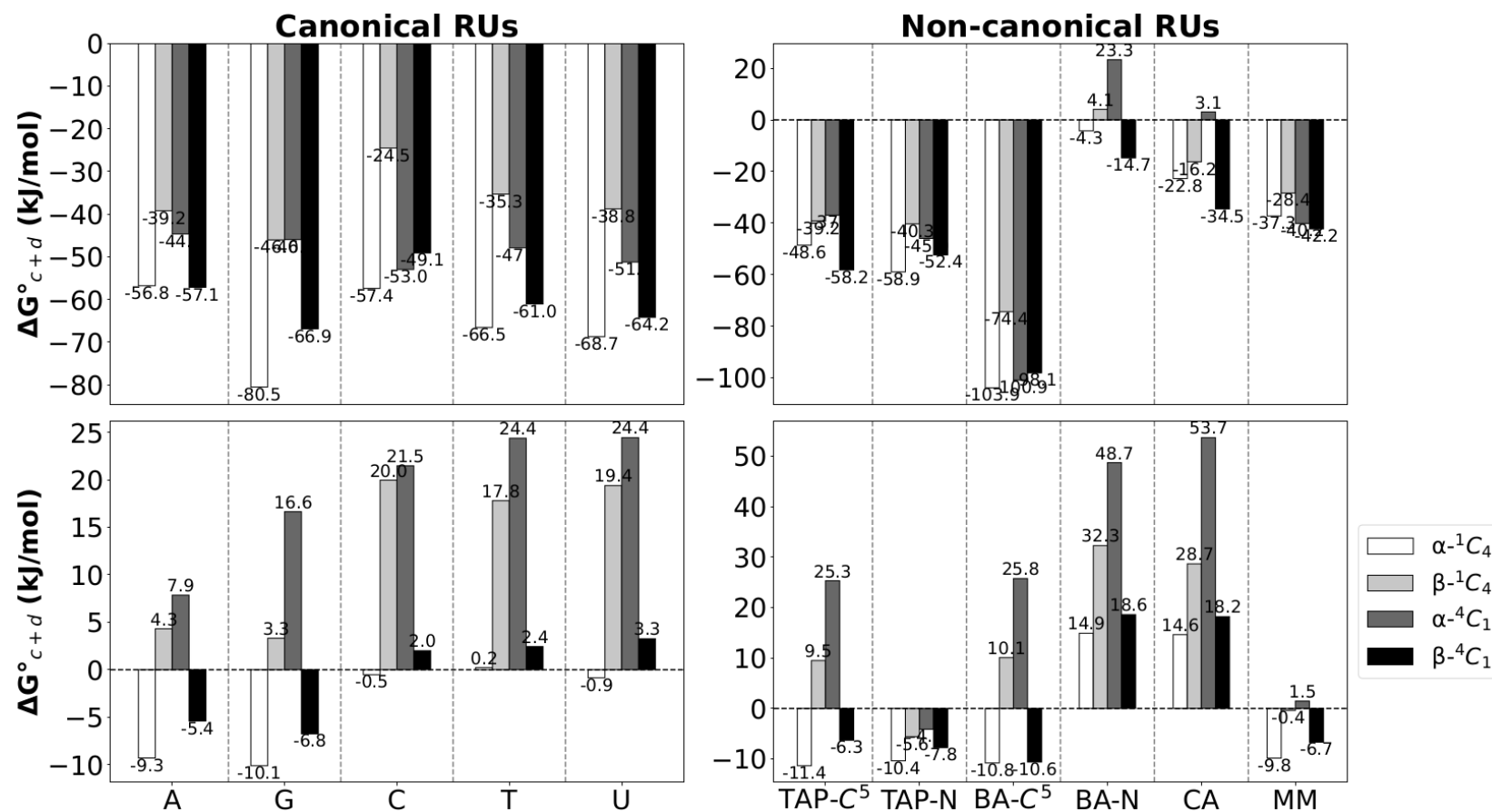
Figure 4.20 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^--2dRib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

environment the α-$^1C_4$ ($\Delta G^{\circ}_{(c+d)}$ = -103.9 kJ/mol), α-$^4C_1$ ($\Delta G^{\circ}_{(c+d)}$ = -100.9 kJ/mol) and β-$^4C_1$ ($\Delta G^{\circ}_{(c+d)}$ = -98.1 kJ/mol) forms of the BA-C^4 Nt emerge as the most favored. For the canonical nucleotides in aqueous environment the β-$^4C_1$ and α-$^4C_1$ forms are disfavored for the pyrimidine nucleotides. For the non-canonical nucleotides all nucleotides from BA-N and CA are disfavored, with the α-$^4C_1$ ($\Delta G^{\circ}_{(c+d)}$ ≈ 50.5 kJ/mol) form is the most disfavored overall.

When analyzing **Figure 4.21** for the $HAsO_3^-$-2dRib-RU obtained from the (c+d) pathway we reach similar conclusions when comparing their energies to the $\Delta G^{\circ}_{(c+d)}$ from the phosphorylated analogs. In this case for the Nts with a canonical base in vacuum in all cases the α- and β-$^1C_4$ forms are the most favored. In the case of the non-canonical nucleotides in vacuum all combinations of anomer + puckering are favored for the TAP-C^5, BA-N, BA-C^5 and MM Nts. In these cases, the α-$^1C_4$ and β-$^1C_4$ forms are the most favored. For the CA Nts the β-$^1C_4$, α-$^4C_1$ and β-$^4C_1$ are favored. In this environment the most favored nucleotide overall is the BA-C^5 Nt with a $\Delta G^{\circ}_{(c+d)}$ = -110.1 kJ/mol. In aqueous environment the α-$^1C_4$ forms of all canonical and non-canonical nucleotides is unfavored. For the non-canonical BA-N and CA all anomers+puckering combinations are unfavored.

Let's turn now to the analysis of the Nts with the ribose in the P-form. For the Nts containing a HPO_4^- modeled using the classic model (a+b) and (**Figure 4.22**) all anomers of the canonical nucleotides in vacuum are favored. For the A Nt the most favored anomer is the α-$^1C_4$. For the rest of the purine and pyrimidine nucleotides the most favored are the α-$^1C_4$ and the β-$^4C_1$. In the case of the non-canonical nucleotides in vacuum all forms from the TAP-C^5, TAP-N and BA-C^5 are favored.

All anomers from the BA-N are unfavored. The β-$^1C_4$ ($\Delta G^{\circ}_{(a+b)}$ = = -18.5 kJ/mol) form of the CA Nt and the α-$^1C_4$ (-43.0 kJ/mol) and β-$^1C_4$ (-25.7 kJ/mol) of the MM are favored. In this environment the α-$^1C_4$ anomer of the BA-C^5 is the most favored overall.

In aqueous solution all forms of the canonical and non-canonical nucleotides are unfavored or the energies are negligible. The highest $\Delta G^{\circ}_{(a+b)}$ is observed for the CA Nt with 54.6 kJ/mol.

When the nucleotides obtained from the classic pathway containing $HAsO_4^-$ (**Figure 4.23**) similar patterns are observed in comparison to the phosphate counterparts.
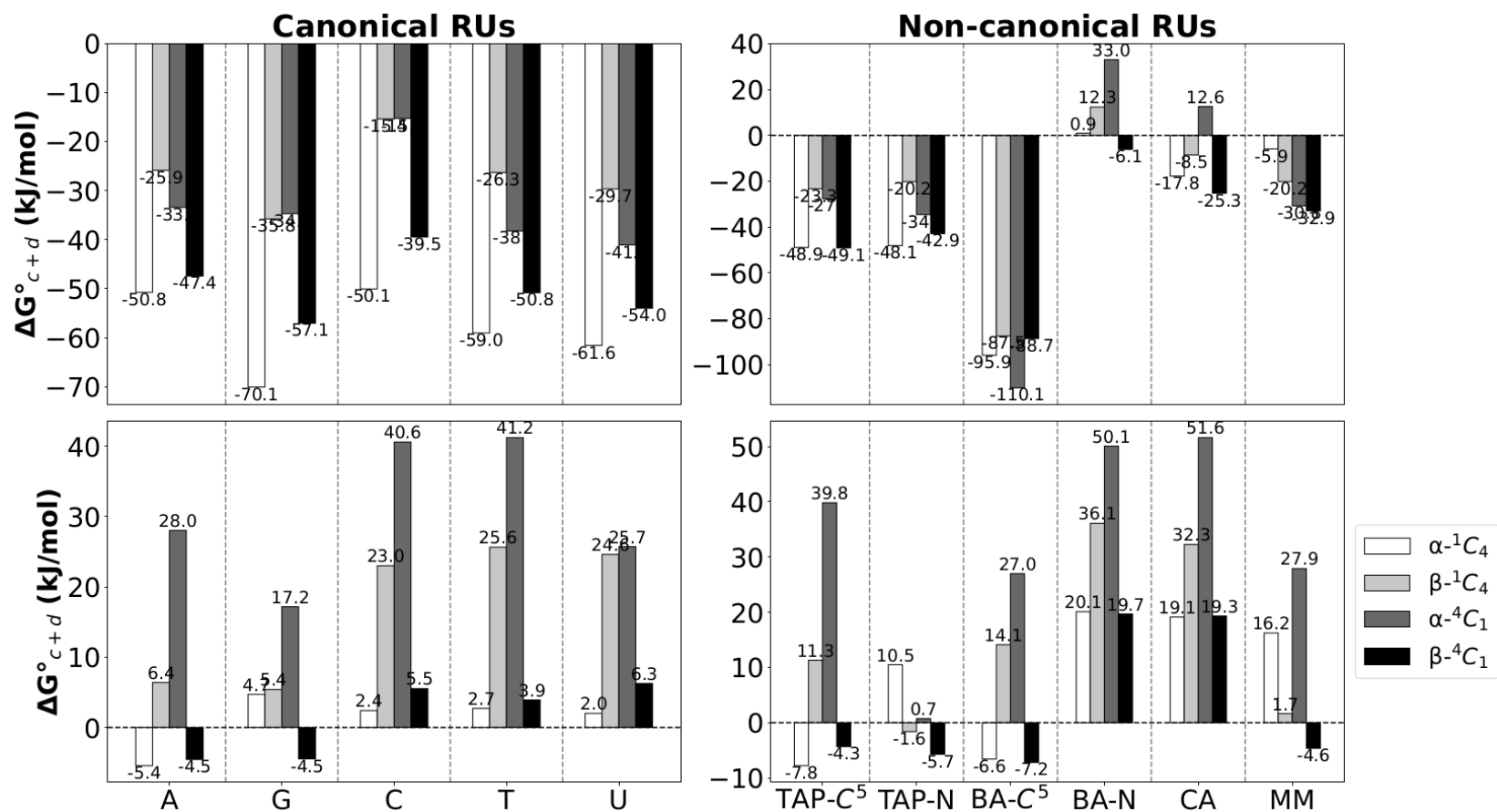
Figure 4.21 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HAsO$_3^-$-2dRib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
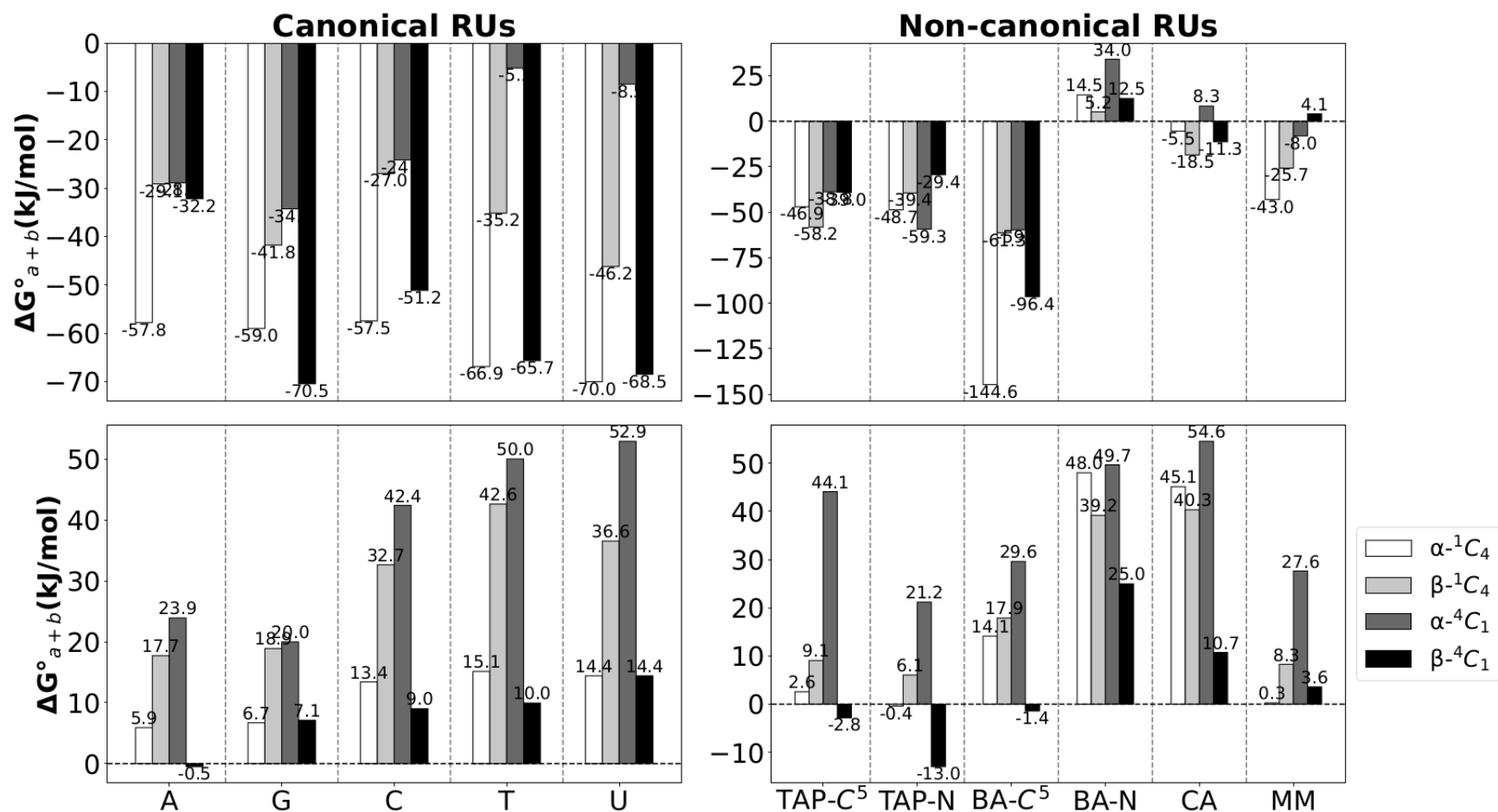
Figure 4.22 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO_3^--Rib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
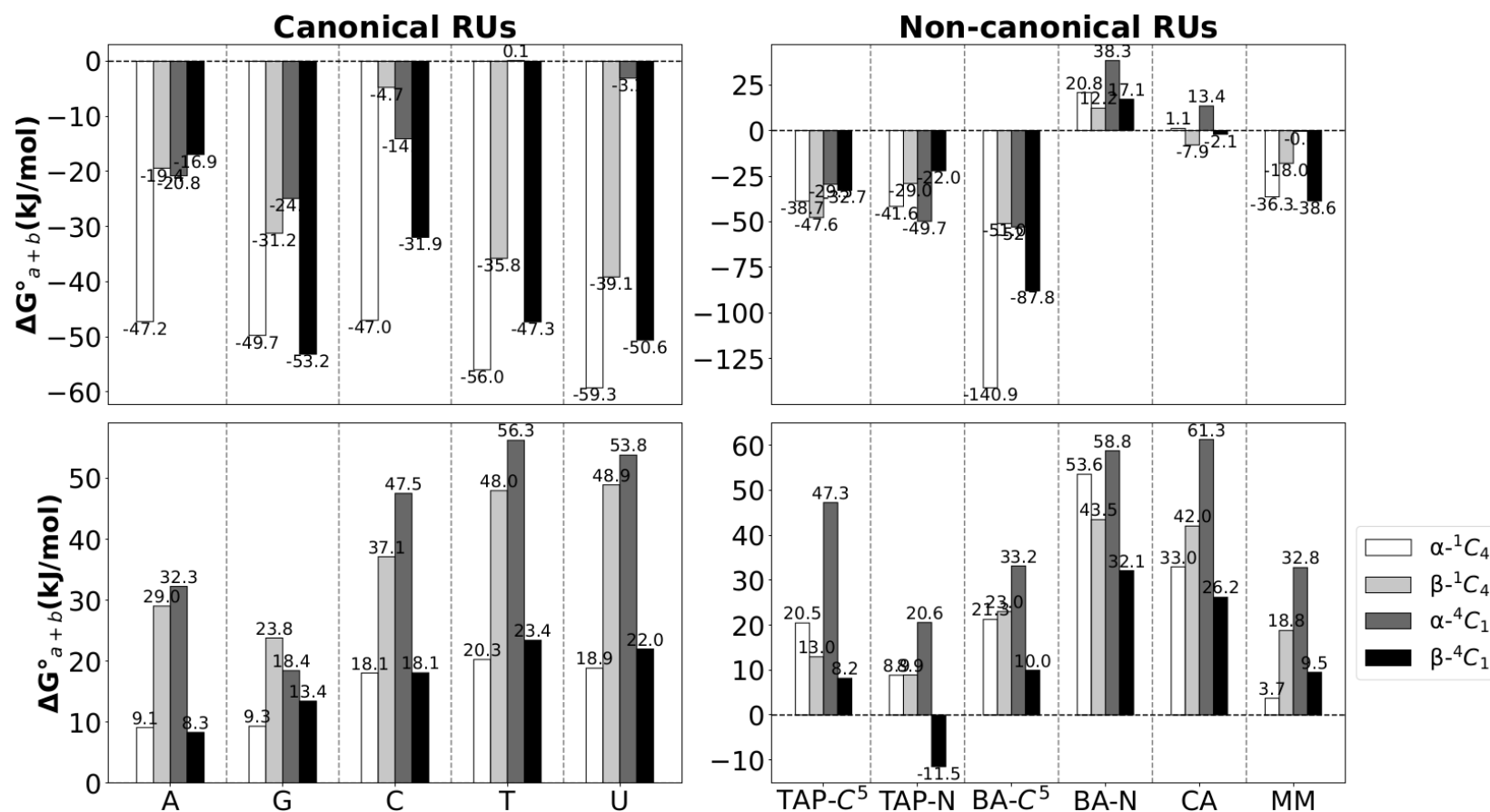
Figure 4.23 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HAsO$_3^{-}$-Rib-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

For the Rib Nts from the (c+d) pathway (**Figures 4.24** and **4.25**), similar patterns are observed but the magnitude of the energy changes significantly in some cases, e.g., the preferential forms from the canonical and non-canonical nucleotides in vacuum are more favored through (c+d) than (a+b). For the nucleotides in vacuum with a HPO_4^- the $\alpha\text{-}^1C_4$ of BA-C^5 is the most favored form ($\Delta G_{(c+d)}^\circ$ = -121.7 kJ/mol), but less than in the (a+b) path. For the case of the similar BA-C^5 Nt but with a HPO_4^- the $\alpha\text{-}^1C_4$ and $\alpha\text{-}^4C_1$ forms are the most favored. In aqueous solution the trends observed in the Nts from the classic pathway are repeated. In this case the energies for the canonical nucleotides are less positive and in the case of the non-canonical nucleotides in aqueous solution the only exception is the alternative synthesis of the $\alpha\text{-}^4C_1$ form of BA-C^5 been thermodynamically favored (-18.3 kJ/mol for HPO_3^- and -24.6 kJ/mol for $HAsO_3^-$). Again, in these cases the most unfavored synthesis are observed for the forms of the BA-N and CA nucleotides.

Let's analyze now the nucleotides synthesized from both pathways with a non-sugar TC: glyceric acid and glycerol Nts. **Figure 4.26** represents the bar graphs for the $\Delta G_{(a+b)}^\circ$ of the classic synthesis of the glyceric acid Nts with HPO_4^-. As can be noticed all canonical nucleotides in vacuum and canonical or non-canonical nucleotides in aqueous solution are thermodynamically unfavored. The only Nts that are favored in vacuum are the ones containing TAP-N ($\Delta G_{(a+b)}^\circ$ = -34.7 kJ/mol), BA-C^5 ($\Delta G_{(a+b)}^\circ$ = -44.2 kJ/mol) and the BA-N ($\Delta G_{(a+b)}^\circ$ = -21.2 kJ/mol). When the nucleotides contain arsenate instead of phosphate (**Figure 4.27**) the canonical nucleotides are not favored in either environment and the only nucleotides favored are the ones synthetized in vacuum that contain TAP-N ($\Delta G_{(a+b)}^\circ$ = -26.2 kJ/mol) and BA-C^5 ($\Delta G_{(a+b)}^\circ$ = -24.3 kJ/mol).

For nucleotides obtained from the alternative pathway (c+d), the only nucleotides which alternative synthesis is thermodynamically favored when there is a phosphate (**Figure 4.28)** are the G, TAP N-glycosylated and BA C^5-glycosilated Nts in vacuum. The rest of the energies are > 0 kJ/mol or dismissible. When the IL is arsenate (**Figure 4.29**) the G, U, TAP N-glycosylated and BA C^5-glycosilated Nts in vacuum are favored. Overall, with either IL the most favored nucleotides in vacuum contain BA-C^5 ($\Delta G_{(c+d)}^\circ$ = -93.0 kJ/mol) and ($\Delta G_{(c+d)}^\circ$ = -83.2 kJ/mol).

These results suggest that the glyceric acid appears to be a non-suitable TC for the proto-NAs. The only Nts containing glyceric acid favored are the ones containing G, U and the non-canonical TAP-N and BA-C^5 when obtained in vacuum following an alternative pathway.
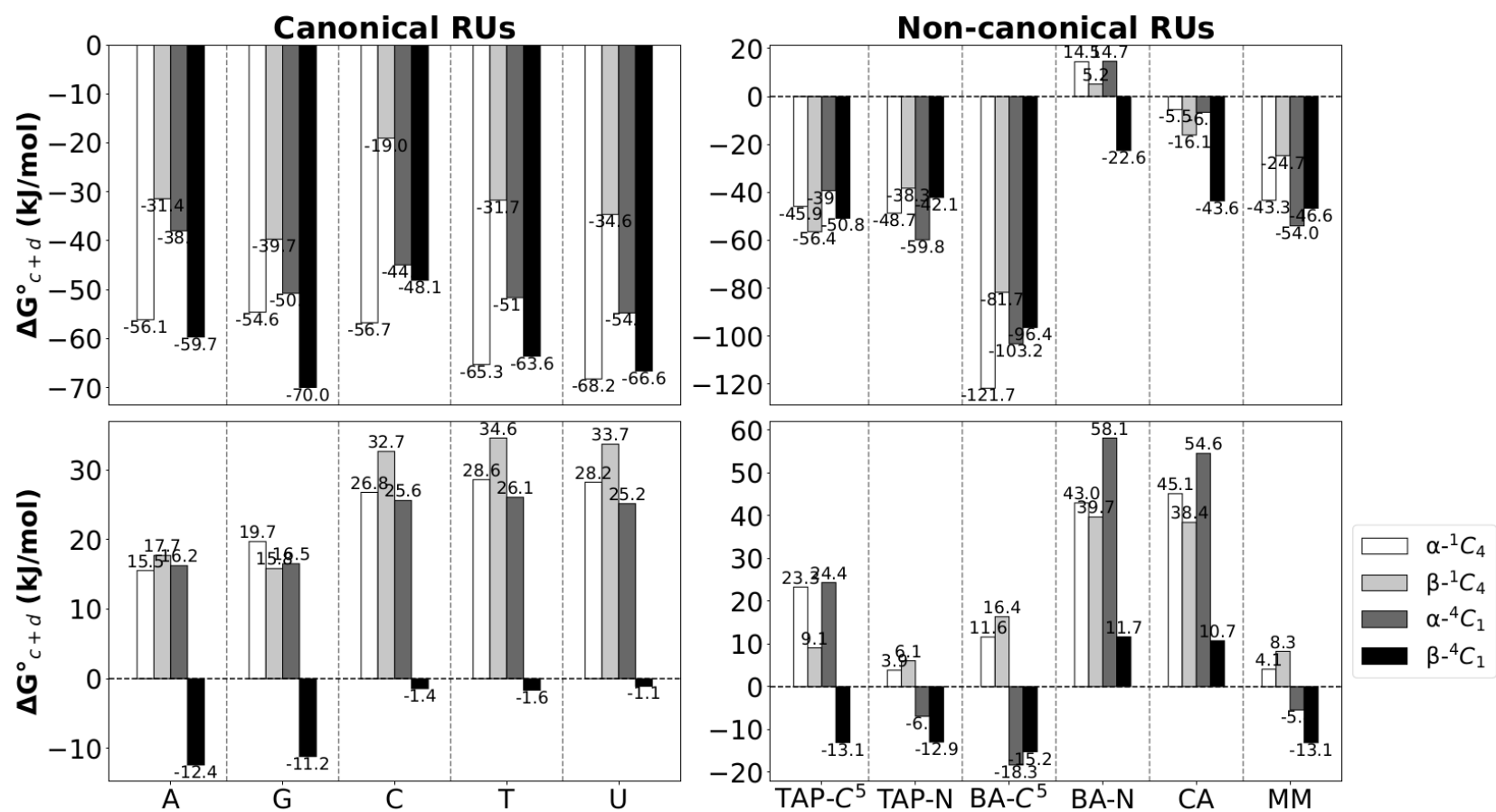
Figure 4.24 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the HPO$_3^-$-Rib-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

Figure 4.25 Comparison of Gibbs energies of reaction ($\Delta G°$) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical β- and α-counterparts of the $HAsO_3^-$-Rib-RU. (Top) $\Delta G°$, B3LYP/6-311++G (d,p) in vacuum, (Bottom) $\Delta G°$, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
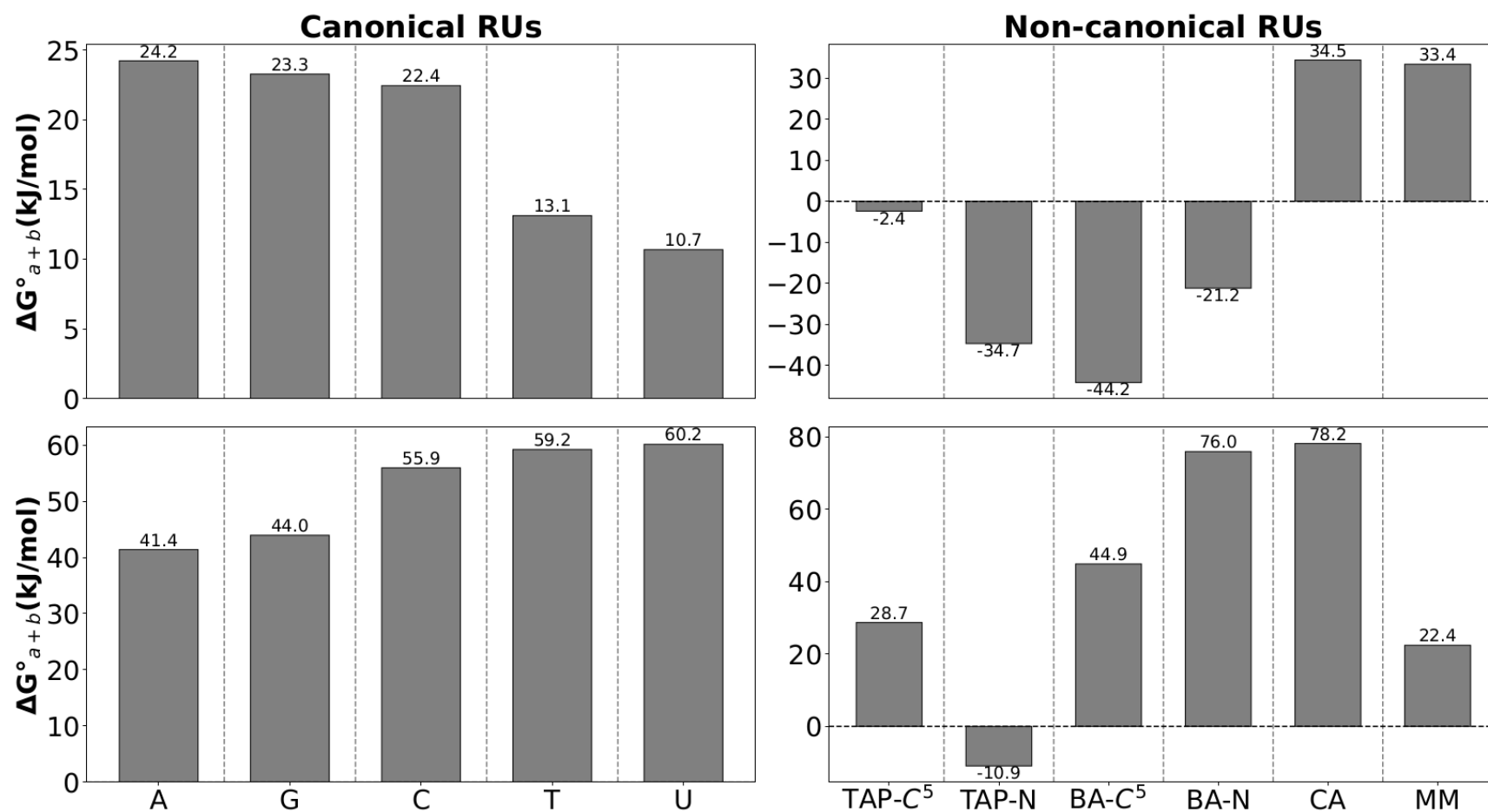
Figure 4.26 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical HPO_3^--glyceric acid-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
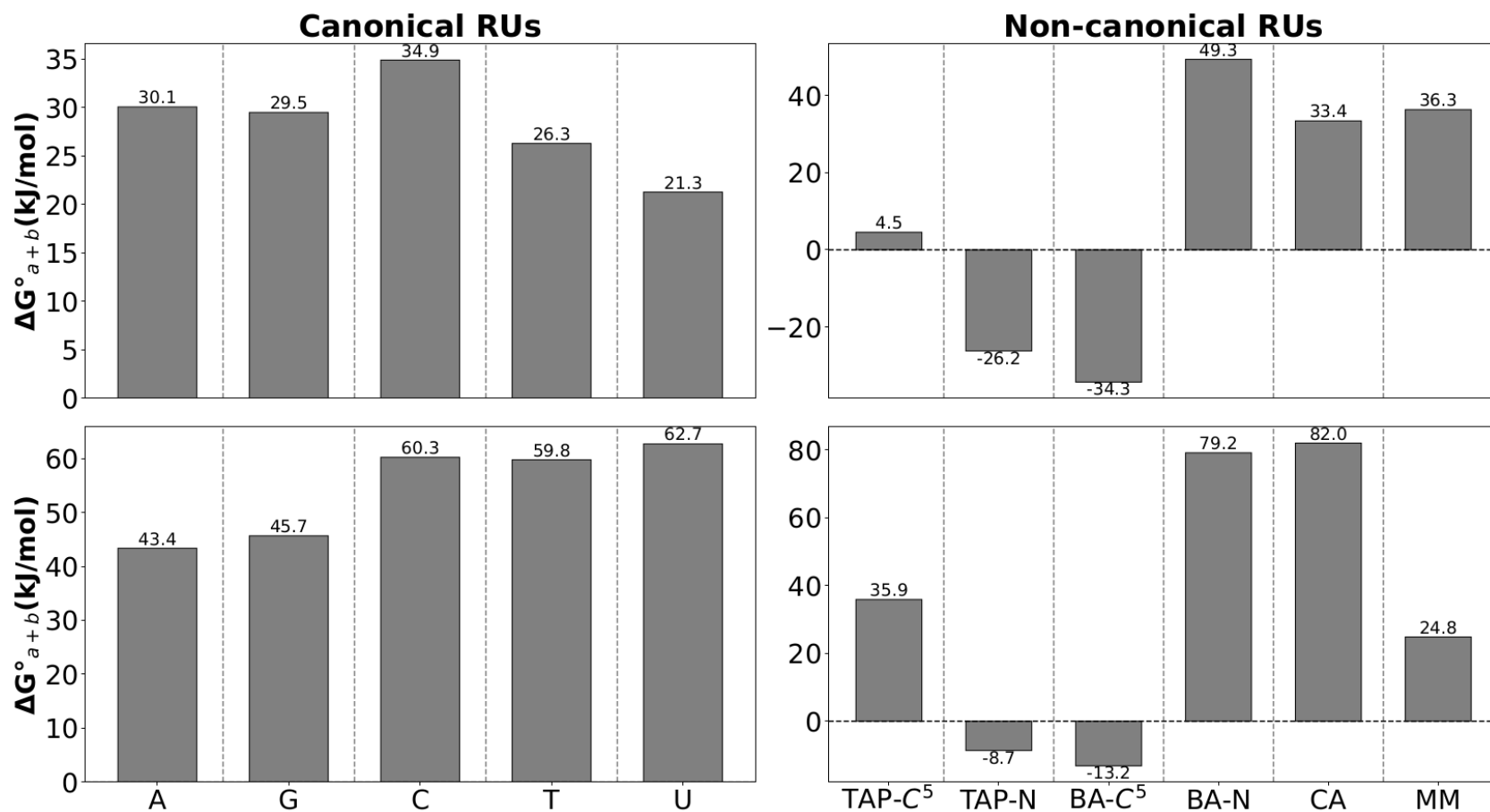
Figure 4.27 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical HAsO$_3^-$-glyceric acid-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
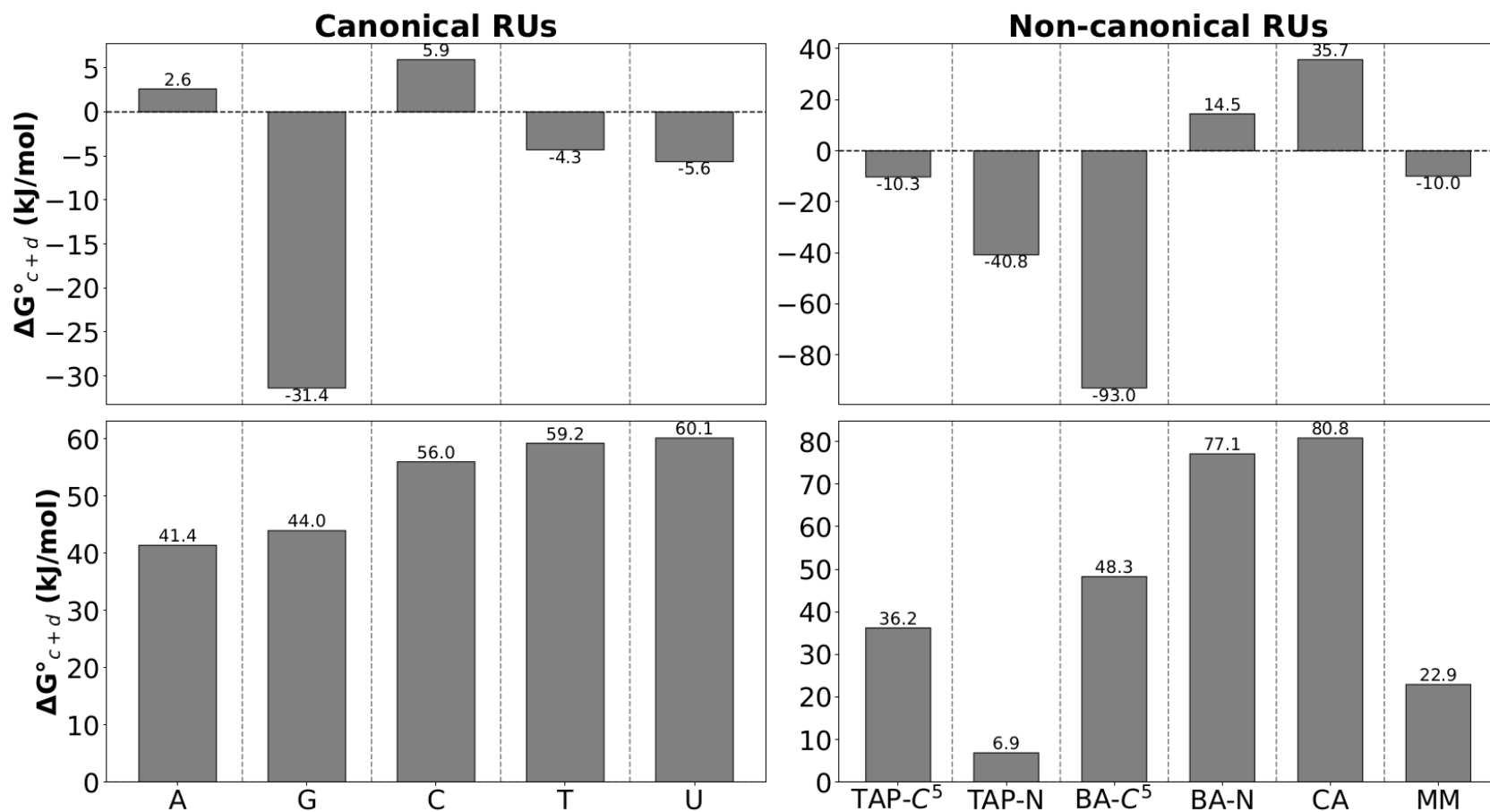
Figure 4.28 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical HPO_3^--glyceric acid-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
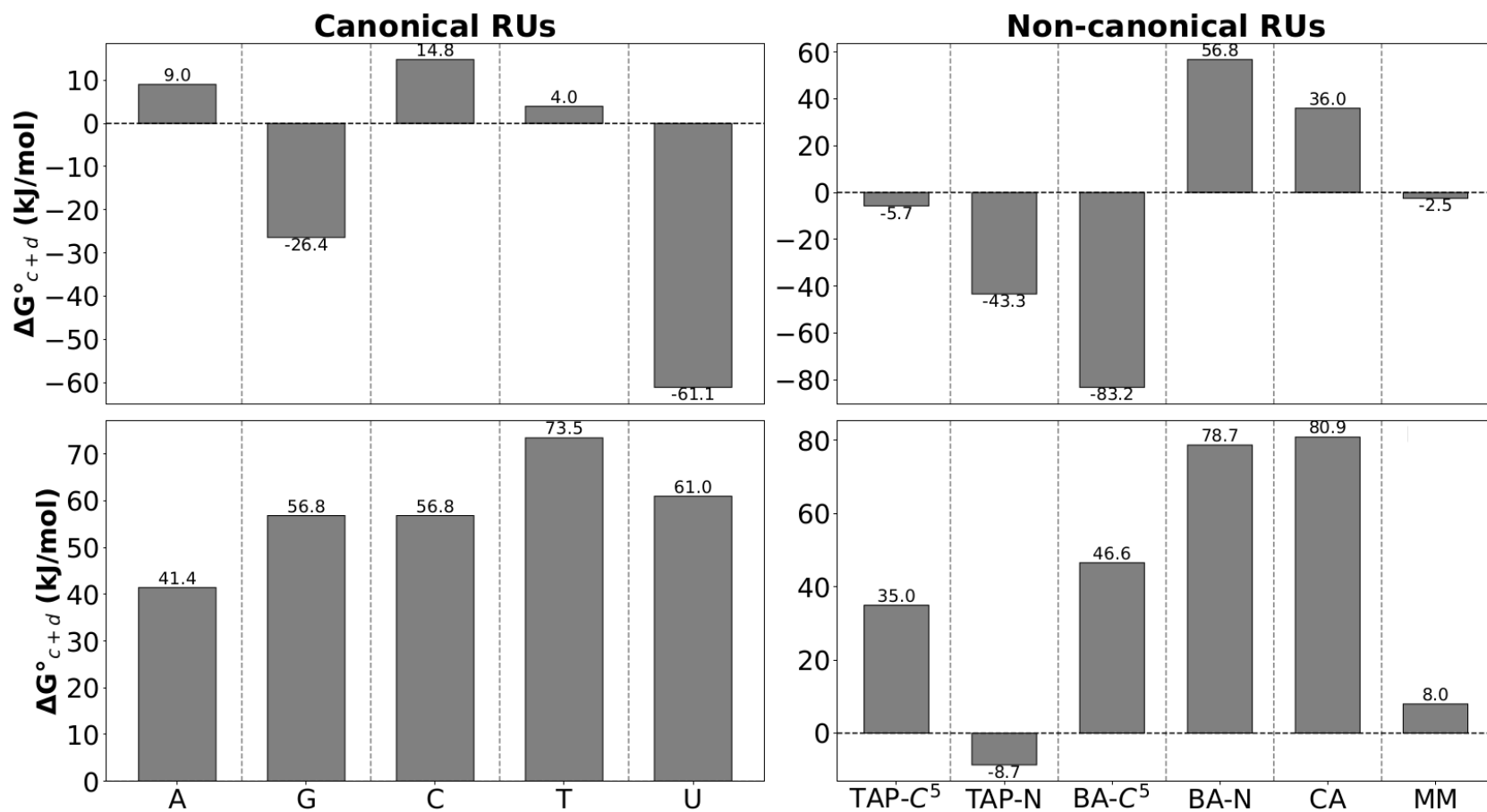
Figure 4.29 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical $HAsO_3^-$-glyceric acid-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

Figures **4.30** and **4.31** show the bar graphs for the classic synthesis of the nucleotides containing glycerol and either HPO$_4^-$ or HAsO$_4^-$ respectively. For the Nts with HPO$_4^-$ (**Figure 4.30**) all nucleotides are favored in either environment. In both environments the most favored nucleotide is the one containing BA-C^5. Some of the energies are in the intrinsic error of the method for some of the nucleotides in aqueous environment with the canonical pyrimidines and with TAP-N, BA-N and CA. When the nucleotides contain AsO$_4^-$ a similar trend is observed.

For the nucleotides obtained using an alternative pathway ((c+d) (**Figures 4.32** and **4.33**)) similar trends are observed in both environments and with both ILs but the energies are less negative overall, which means that a classic pathway tends to favor more the formation of the building blocks of GNA.

These results suggest that, possibly, if the first proto-RNAs emerged in the primitive oceans they may had a more chemically simple TC like glycerol and an array of canonical and non-canonical bases (e.g., TAP and BA C-glycosylated and MM) could have been present. Either pathway could have been followed to synthesize these GNA building blocks.

If the building blocks containing glycerol were synthesized in vacuum, then the classic pathway would have been preferable.

4.4.3 Distribution of the χ torsion angle defining the conformation of the RU around the TC

The rotation of each RU around a sugar-like TC is defined by the torsion angle χ at the glycosidic bond. The RU can be *syn* when $\chi = 0º$ - $90º$ or $270º$ - $360º$ or *anti* when $\chi = 90º$ - $270º$ with respect to the TC [60, 61].

The torsion angle χ is define by {O4'(hemiacetal O)-C1'(anomeric center in sugar)-N9-C2} and {O4'-C1'-N1-C2} for the purines and pyrimidine canonical bases respectively, meanwhile for the non-canonical bases the definitions from Kaur and coworkers [60, 61] can be used. These definitions used the sets {O4'-C1'-C5-C6} for the BA C^5-glycosylated, {O4'-C1'-N1-C6} for the BA N-glycosylated (glycosidic bond between the exocyclic NH$_2$ of BA and the anomeric center has been selected based on the NMR characterization from Mungi *et al*. [47]), {O4'-C1'-C5-C4} for the TARC glycosylated at the C5 position, {O4'-C1'-N4-C4} for the TARC N-glycosylated (the three exocyclic NH$_2$ are chemically equivalent), {O4'-C1'-N4-C4} for MM
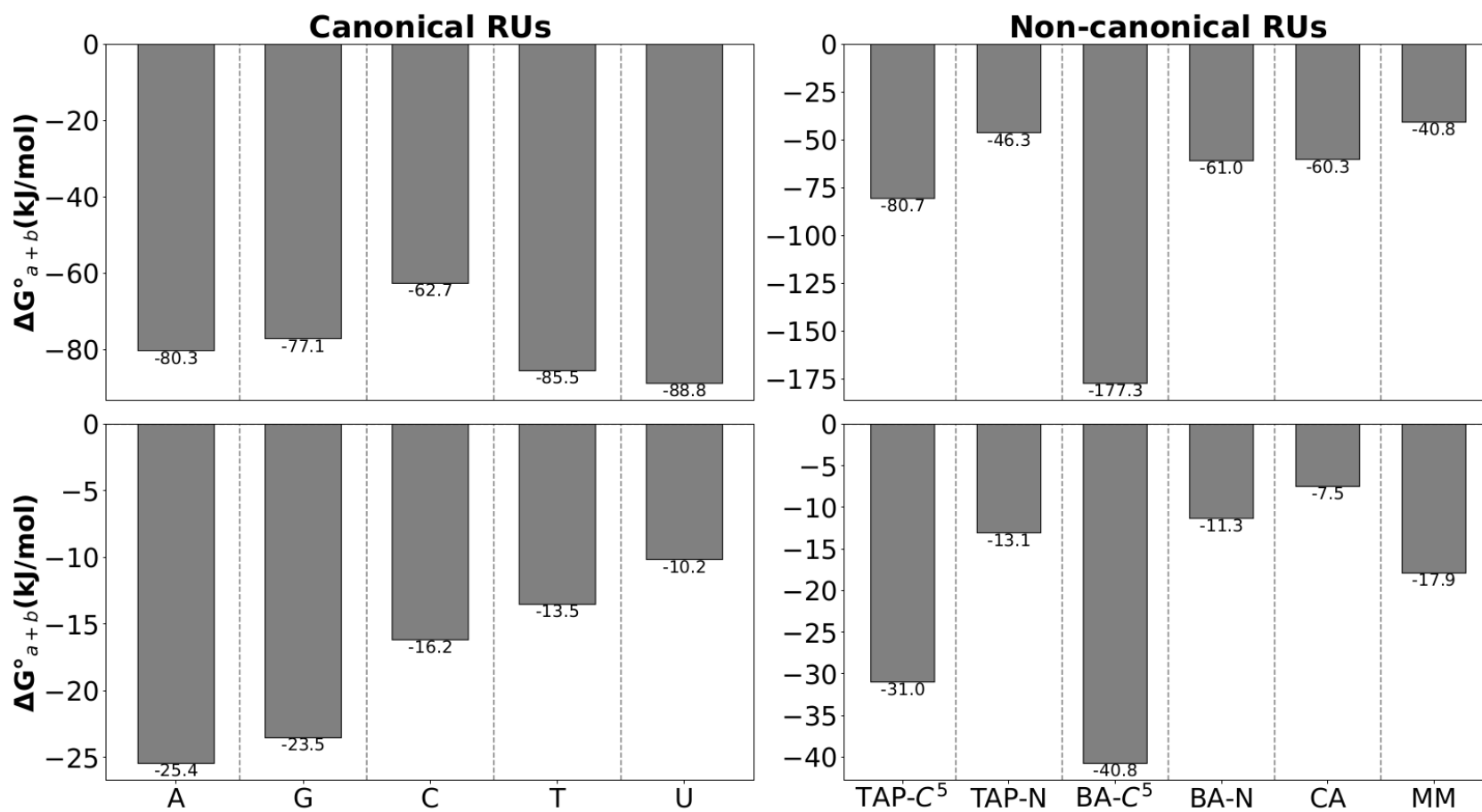
Figure 4.30 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical HPO_3^--glycerol-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
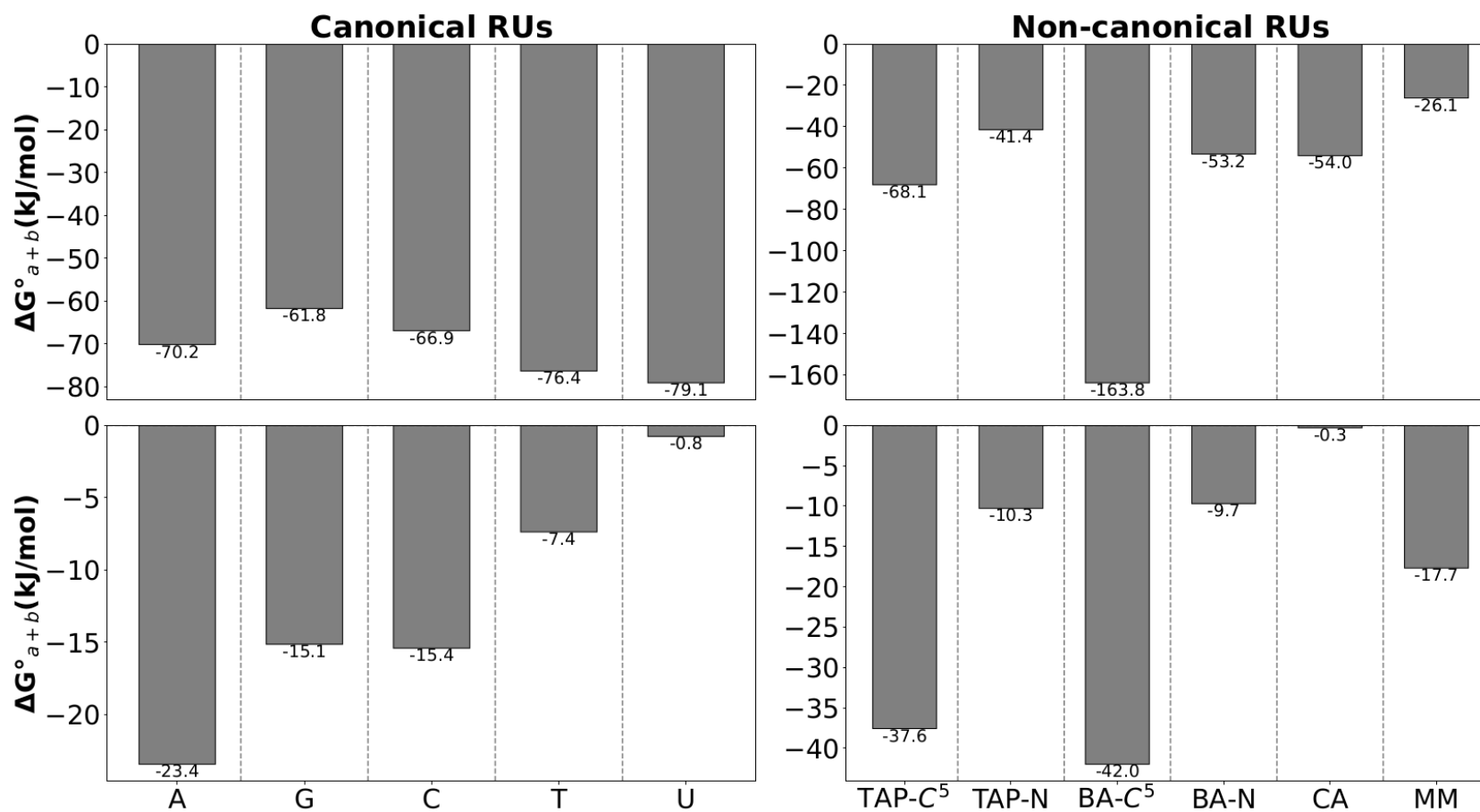
Figure 4.31 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the classic synthesis (a+b) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical $HAsO_3^-$-glycerol-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
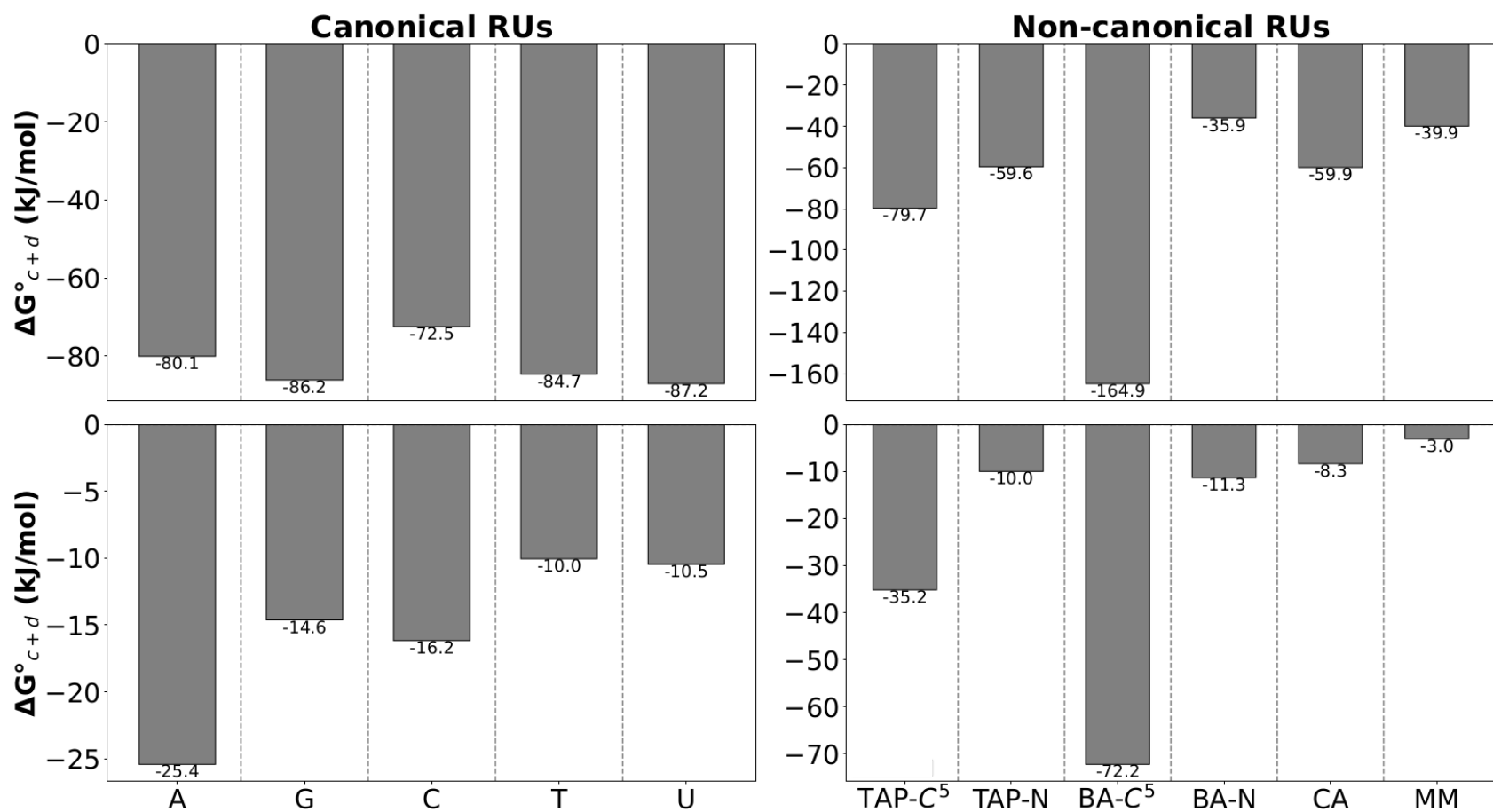
Figure 4.32 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical HPO_3^--glycerol-RU. (Top) ΔG°, B3LYP/6-311++G (*d,p*) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.
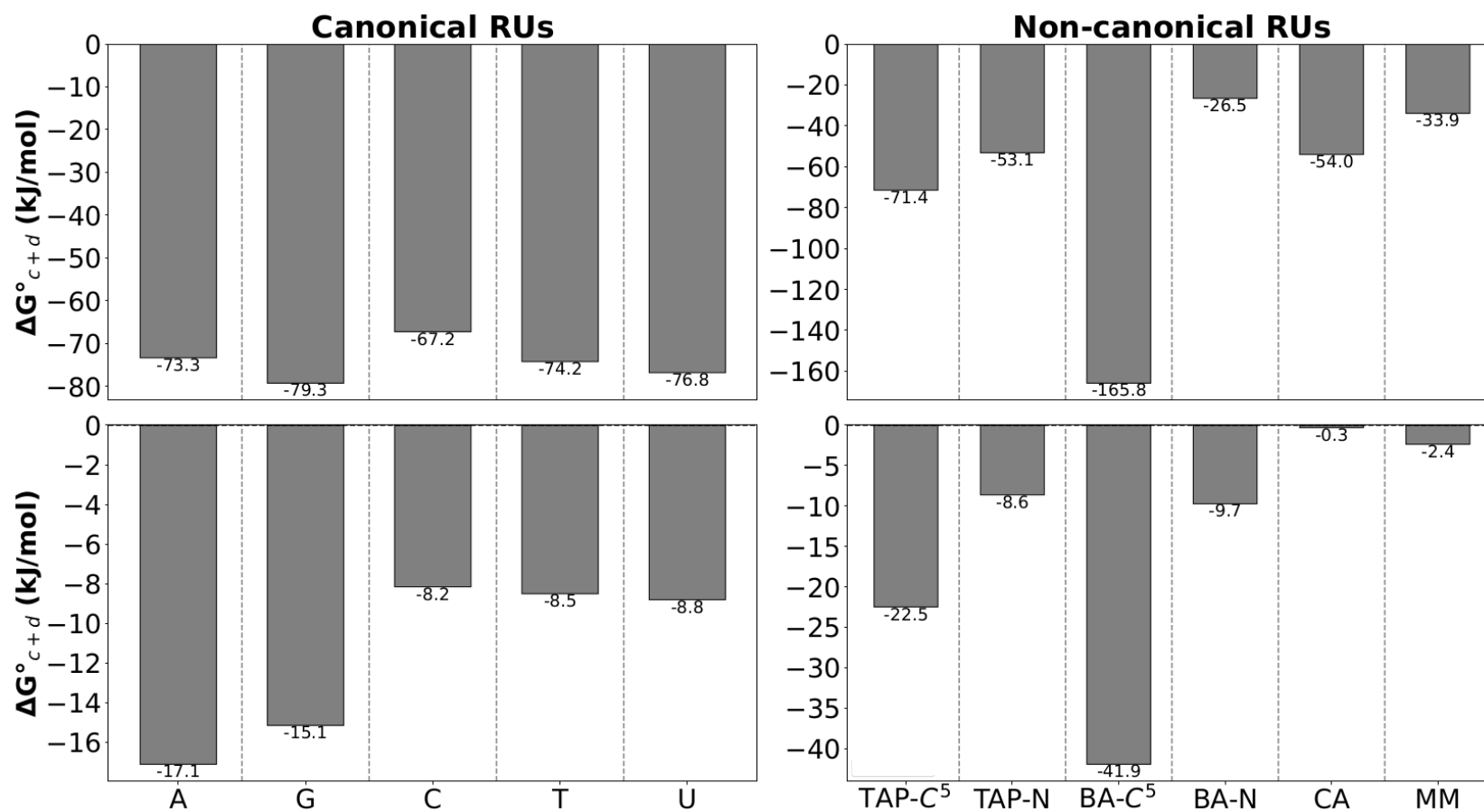
Figure 4.33 Comparison of Gibbs energies of reaction (ΔG°) at 298 K for the alternative synthesis (c+d) (see **Figure 4.5**), leading to the 5 canonical and 6 non-canonical $HAsO_3^-$-glycerol-RU. (Top) ΔG°, B3LYP/6-311++G (d,p) in vacuum, (Bottom) ΔG°, B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

and {O4'-C1'-N5-C4} for CA (see **Figure 4.34** for a schematic representation).

It has been well documented in the literature through experimental [62] and theoretical studies [63] that usually the *anti-* is preferred over the *syn*-conformation for the canonical bases in the building blocks of today's nucleic acids crystal structures [1]. This preference has been justified on the basis of the higher steric effects between the base and the sugar [64].

Kaur and coworkers (Kcw) [61] analyzed the torsion profiles for the χ angle in the *β-* and *α*-ribonucleosides containing the bases BA C-glycosylated and MM using a free and polymer (ribose with a CH_3 group at the O5') nucleoside model. They found that for the case of the *β*-anomers of the BA and MM nucleosides in either model the *anti*-conformation was preferred. Still the studies by Kcw does not address how the presence of an IL or the puckering of the ribose ring may affect the position of the RU with respect to the TC.

This section aims to fill in the gaps in the literature referring to the preferential conformations of canonical and non-canonical bases with respect to different potential prebiotic TCs-ILs backbone derivatives and the effects of starting from different sugar ring puckering.

Figures 4.35 and **4.36** shows the distribution of the χ torsion across all canonical and non-canonical nucleotides obtained from the classic (a+b) pathway with a HPO_3^- and $HAsO_3^-$ (for more details on the specific values of the torsion angle χ for each canonical and non-canonical TC see **Table A6**).

An analysis of **Figure 4.35** shows that when phosphorus is present the canonical nucleotides in vacuum prefer the *anti*-conformation in both *β-* and *α*-configurations. In the case of the non-canonical nucleotides in the *β*-configuration also *anti* is preferred, meanwhile when the base has an *α*-configuration *syn* is slightly preferred. Analyzing all the nucleotides there is an overall preference for the *anti*-conformation for the *β*-anomers and *syn* for the *α*-counterparts.

In aqueous environment the *β*-anomers of the canonical nucleotides prefer more *anti* meanwhile the *α*-counterparts prefer more *syn*. Both anomers of the non-canonical nucleotides have a higher preference for the *syn* conformation and overall, both anomers of all nucleotides tend to prefer more the *syn* conformation in implicit solvation.

When the P is replaced by As, the histograms from **Figure 4.36** show that in vacuum the canonical nucleotides similarly to the monophosphate counterparts prefer the *anti*-position in both configurations. The non-canonical nucleotides prefer slightly more *anti* in the *β-* and *syn* when there is an *α*-configuration. In general, all nucleotides prefer more *anti* for *β* and *syn* for *α*. In the
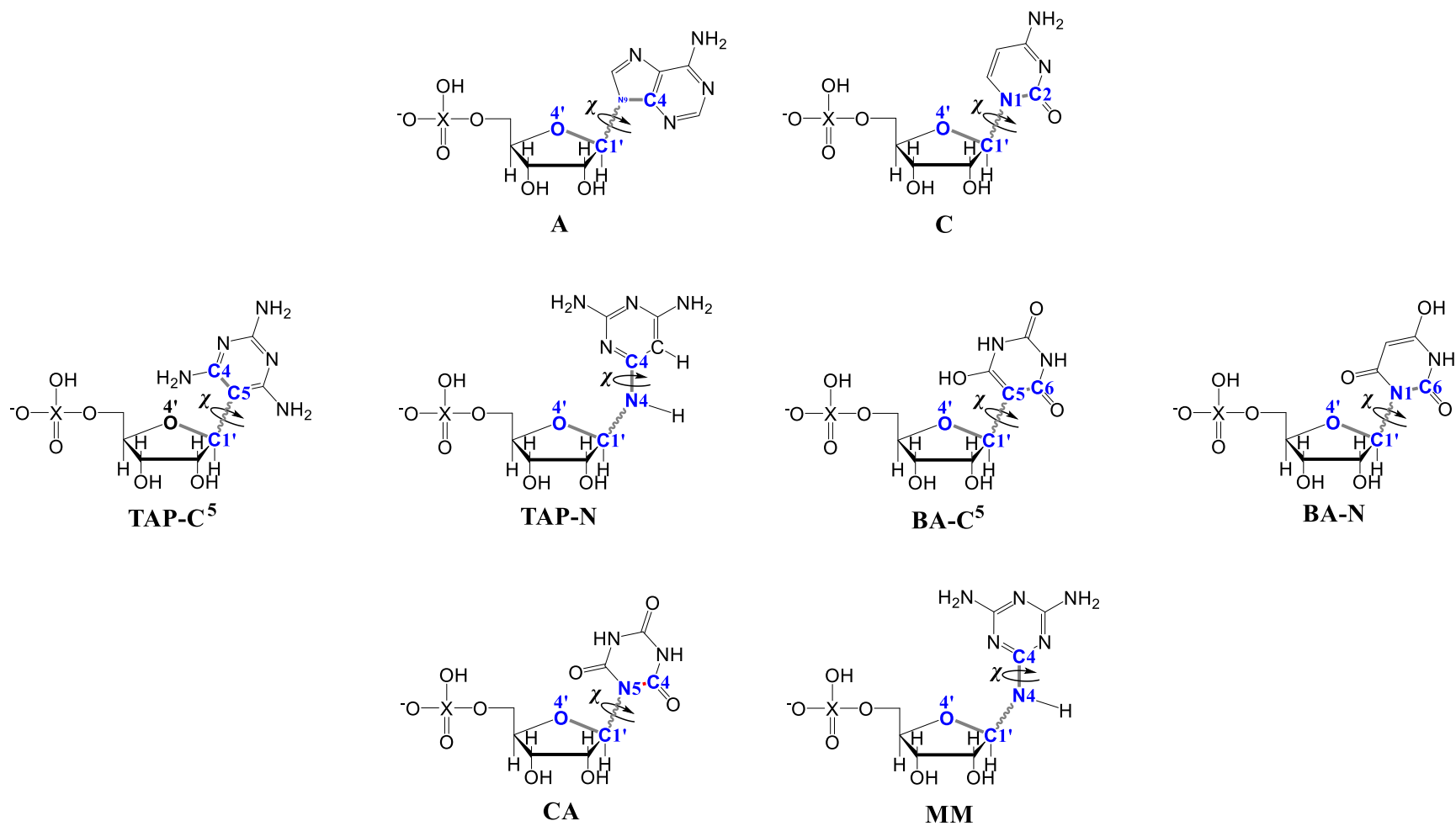
Figure 4.34 Representation of the sets of atoms (blue) and connecting bonds (gray) of the χ torsion angle at the glycosidic bond between the canonical (top) and non-canonical (middle and bottom) RUs and a Ribf-5'-XO$_3^-$, where X = P, A
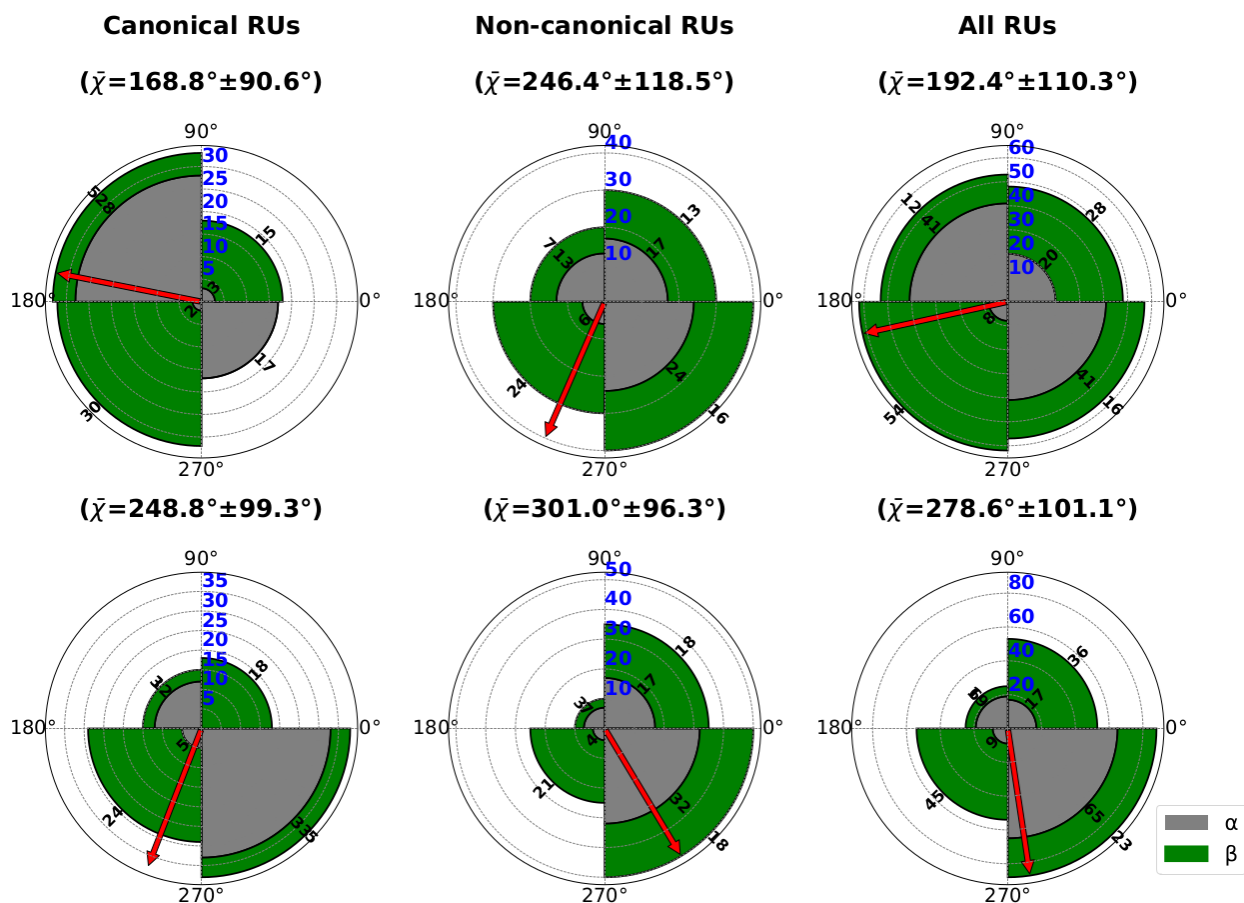
Figure 4.35 Rose diagrams (circular frequency histograms) for the torsion angle χ in the nucleotides from the classic model (a+b) that defines the conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around the HPO$_3^-$-TCs molecules. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the gray pies represent the α-counterparts of the different nucleotides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
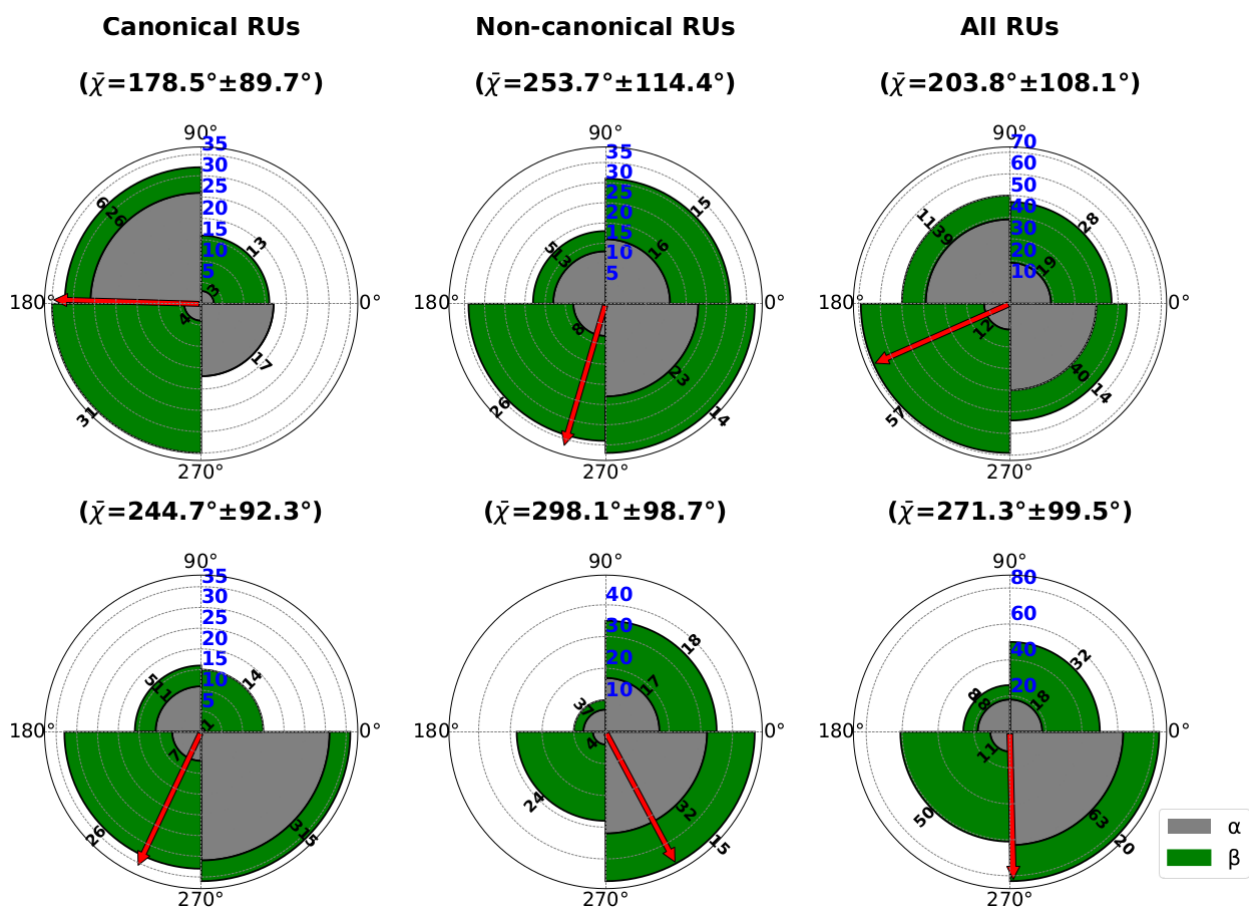
Figure 4.36 Rose diagrams (circular frequency histograms) for the torsion angle χ in the nucleotides from the classic model (a+b) that defines the conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around the $HAsO_3^-$-TCs fragments. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(*d, p*) in vacuum, (Bottom) B3LYP/6-311++G(*d, p*) in aqueous medium using the IEFPCM solvation model.

case of the nucleotides in aqueous environment with a canonical base a similar trend to the one in vacuum is found. Meanwhile, the non-canonical nucleotides tend to prefer slightly more the syn conformation for both anomers. Overall, all nucleotides tend to prefer more *anti* for the *β*-anomer and *syn* for the *α*-counterpart.

These results show the surprisingly similarity in the distributions from the classic nucleotides when having different ILs which suggests that the presence of either P or As does not impose much difference in the conformation of a canonical or non-canonical base around the TC.

There is an overall preference for the *anti*-conformation mostly in vacuum and for the canonical bases. This may suggest that maybe some of the conformational behavior observed in the nucleotides of today informational biopolymers may have been inherited from their monomeric units if they came from a classic synthesis in vacuum.

Let's move on now to the analysis of the χ distributions for the Nts obtained using an alternative (c+d) pathway.

Figures 4.37 and **4.38** represent the statistical frequency distributions of the torsion angle χ for all Nts obtained through an alternative (c+d) path (see **Table A4** for the values of each torsion angle).

Figure 4.37 shows the distributions of χ for the nucleotides modeled from (c+d) that contain a HPO_3^-. In vacuum the canonical nucleotides prefer the *anti*-conformation when having both the *β*- and *α*-configurations. Both anomers of the non-canonical nucleotides prefer more *syn*. Overall, all nucleotides are more frequent to be in the *anti* for the *β*-anomer and *syn*-position for the *α*-counterpart. When the solvation effects are implicitly included the canonical nucleotides prefer more the *anti*-conformation when they are in the *β*- and the *syn* when they are in the *α*-configuration. The non-canonical nucleotides prefer more the *syn* in both anomeric forms and all Nts prefer more the *anti* for the *β*-anomer and *syn* for the *α*-counterpart.

When P is replaced by As the rose diagrams for χ (**Figure 4.38**) show that in vacuum the canonical nucleotides prefer the *anti*-conformation when the configurations are either *β* or *α*. Both anomers of the non-canonical nucleotides prefer more the *syn* position and, overall, all Nts are more prompted to be in *anti* when the RU is *β* and *syn* when *α*. In aqueous environment either canonical either canonical, non-canonical or all nucleotides prefer more the *anti*-conformation for the *β*- configuration and *syn* for the *α*-counterpart respectively.
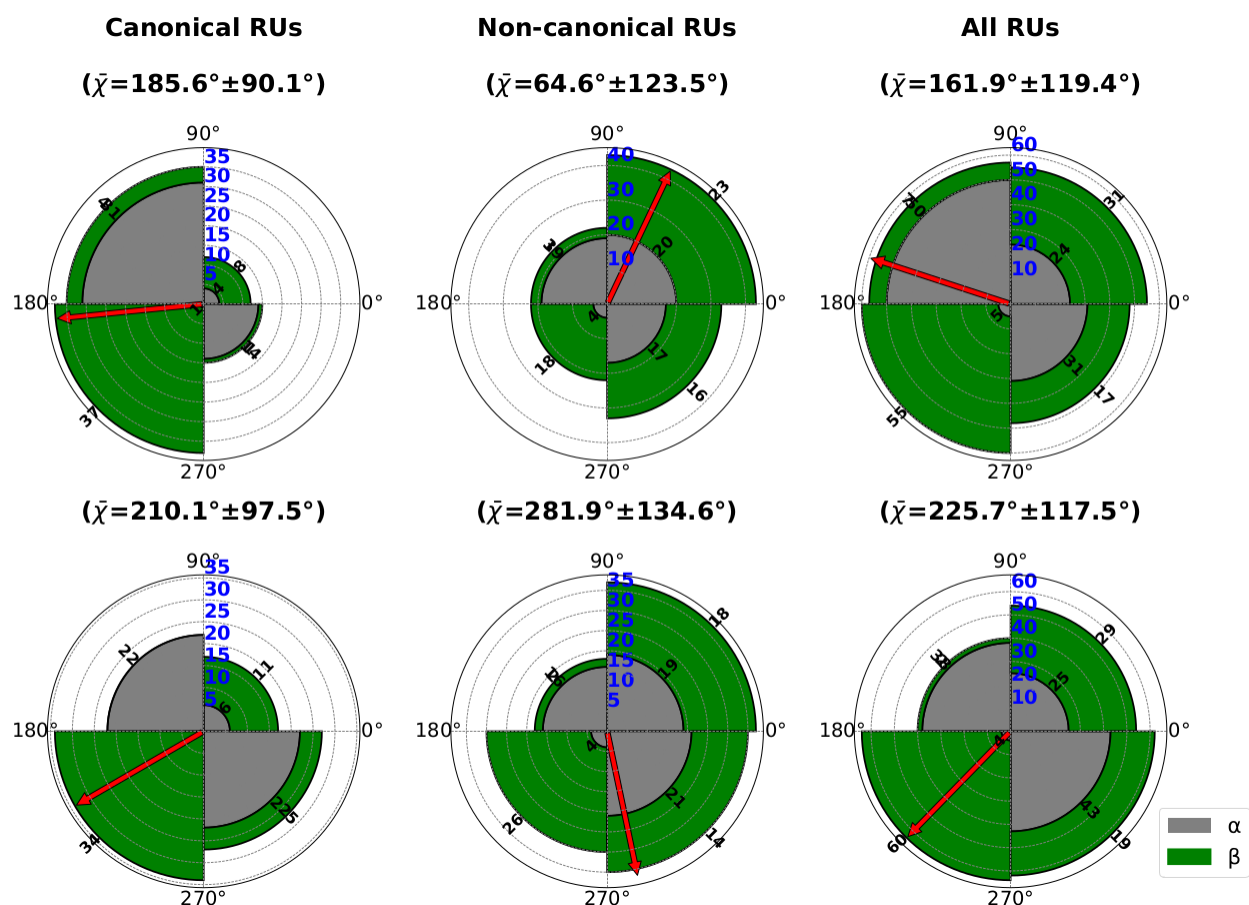
Figure 4.37 Rose diagrams (circular frequency histograms) for the torsion angle χ in the nucleotides from the alternative model (c+d) that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around the HPO_3^--TCs fragments. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
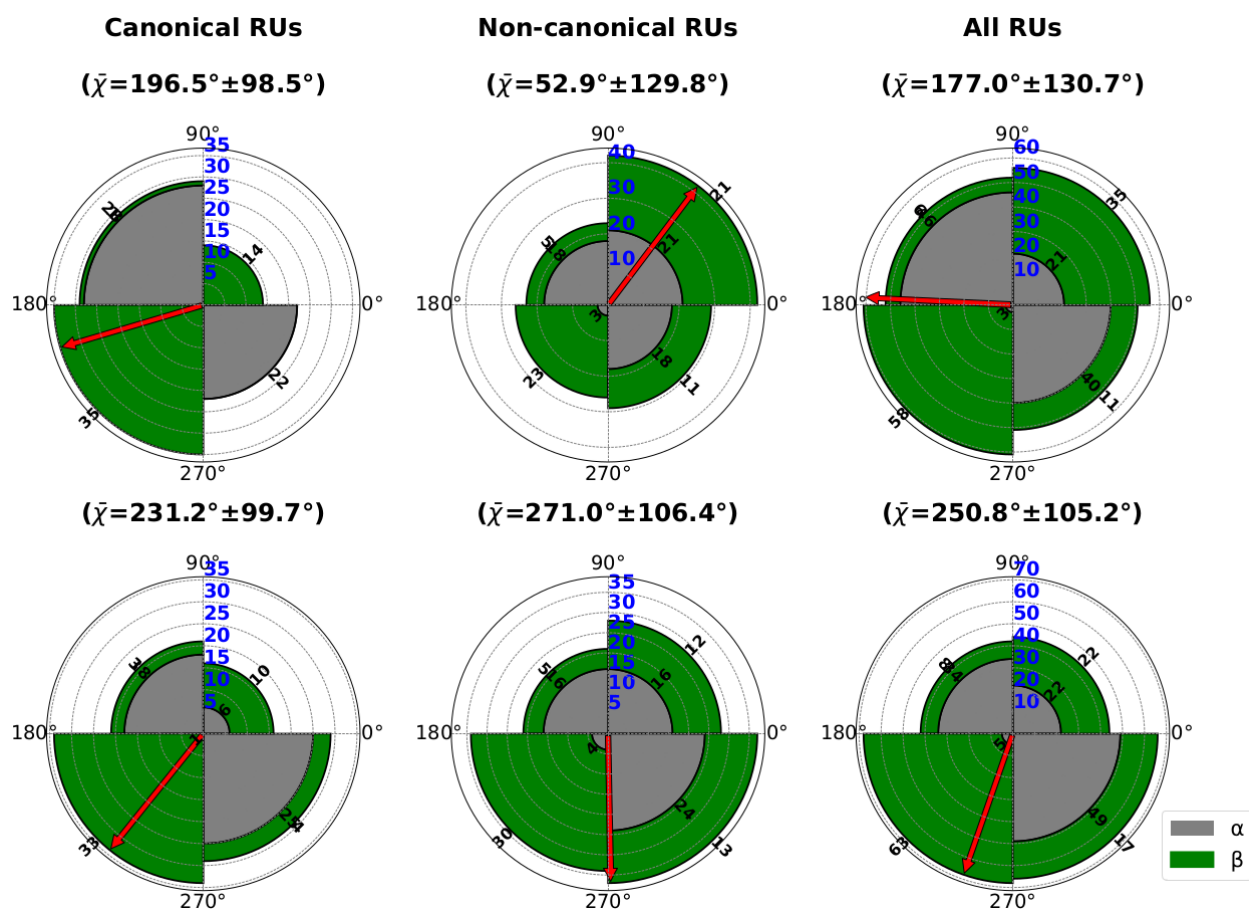
Figure 4.38 Rose diagrams (circular frequency histograms) for the torsion angle χ in the nucleotides from the alternative model (c+d) that defines the RU's conformation of the 5 canonical (A, G, C, T and U) and the 6 non-canonical (TAP-N^5, TAP-N, BA-C^5, BA-N, CA and MM) RUs around the $HAsO_3^-$-TCs fragments. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean χ in (°) for the torsion angle. The green pies represent the β-anomers and the grey pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

4.4.5 Sugar puckering parameters for the F-form and P-form of the TC in the canonical and non-canonical nucleotides

The hemiacetalic ring of ribofuranose or 2'-deoxyribofuranose is not planar. The sugar ring adopts different conformations with out of plane atoms. These conformations are known as sugar puckering and they can be twisted with two atoms in opposite sides of the plane or eclipsed with 4 atoms in the plane [22].

A way to characterize the sugar ring conformations is by using the generalized **C**remer-**P**ople (CP) puckering parameters. This mathematical methodology uses the cartesian coordinates of the ring atoms. These puckering parameters can be generalized for a ring with an odd or even number of atoms.

The CP parameters are the phase angle ϕ_2 and the total puckering amplitude Q for a 5-MR and the phase angles θ and ϕ and the radial Q for a 6-MR [65].

Figures 4.24 and **4.28** of Chapter 3 illustrate all possible ring puckering conformations for the 5- and 6-MR TCs.

The denomination endo {super index anteceding the initials T(twisted) or E(eclipsed)} refers in the puckering notation of 5-MRs to a given atom been in the same plane as the C5', meanwhile exo {sub index preceding the letters T or E} refers to the atom is in the opposed plane of the C5' position [22].

DNA contain three major forms A, B and Z. The first one has two main groups AI and AII and similarly B-DNA also has a BI and BII [22].

Svozil and coworkers (Scw) [66] did a benchmarking of the main torsion angles dominating the conformational changes of $\approx$ 8000 NAs crystallographic structures with a resolution ≤ 1.9 Å. This study proposes that both A-D(R)NA forms tend to have the sugar in a puckering with the C3'-endo [22, 66].

The two main conformers BI and BII from B-DNA contain a predominantly puckering of the sugar with the C2' endo [22].

Z-DNAs usually has ${}^{2}T_{3}$ and ${}^{3}T_{2}$ alternated puckering conformations for the 2dRibf [67].

In the case of NAs interacting with protein-ligand complexes with mixed A-B duplexes had nucleotides with their sugar in the C1'-exo/O4'-endo puckering.

The furanose rings have a predominant $^2T_3$ (C2'-endo) or $^3T_2$ (C3'-endo) in free [45, 40, 41] or polynucleotidic canonical Nts from A, B and Z-DNA [1], meanwhile the pyranoses have a preference for the $^1C_4$ or $^4C_1$ in free canonical nucleosides and nucleotides. Hence, we have set the initial puckering conformations of the furanoses 2dRibf, Ribf and Tho to the $^2T_3$ and $^3T_2$, meanwhile for the pyranoses 2dRib and Rib the $^1C_4$ or $^4C_1$ have been selected.

This analysis will allow us to hopefully find answers for the following questions: 1) does the configuration (β, α) at the anomeric center of canonical and non-canonical Nts affects the sugar ring puckering?, 2) do the free nucleotides preserve the preferential ring conformations found in the Nts in DNA and RNA or in another words may the conformations observed in today's NAs have come in certain extend from the monomeric nucleotides in vacuum or aqueous solution?, 3) changing the addition of the three components: TC, IL and RU affects the sugar ring conformations of the Nts? and finally, does changing the IL from HPO_4^- to $HAsO_4^-$ affects the sugar ring puckering in the Nts?

The next sections will analyze the statistical frequency distributions of the CP puckering parameters in 5- and 6-MR Nts but for further details on the values of these phase angles and amplitudes see **Table A7** from the Appendices section.

Figure 4.39 shows the rose diagram (circular frequency histogram) for the distribution of the phase angle ϕ_2 for the nucleotides obtained using a classic route with a HPO_4^-.

When the Nts contain a canonical base and are modeled in vacuum the β-anomers have the highest frequency between 72°(^2E) - 108°(E_3) and between 252°(E_2) and 288° (^3E). These values for the phase angle include the C2'-endo and C3'-endo puckering which is in agreement with experimental observations for free and polymeric canonical Nts. The α-anomers have a different array of conformations that goes from 90° ($^2T_3$) to 162° ($^4T_0$) (C4'-endo-O4'-exo), hence for the α-configuration the sugar ring tends to adopt a larger number of sugar ring conformations. In aqueous solution the highest population for both anomers is around 270°($^3T_2$) - 306°($^3T_4$) which involves C3'-endo puckering conformations.

For the case of the non-canonical bases similarly to the canonical counterparts when they are in the β-configuration the sugar ring populates mostly the 0°(^0E) - 108°(E_3) and the 234°($^1T_2$)-270°($^3T_2$) and these intervals include the preferentially observed C2'-endo
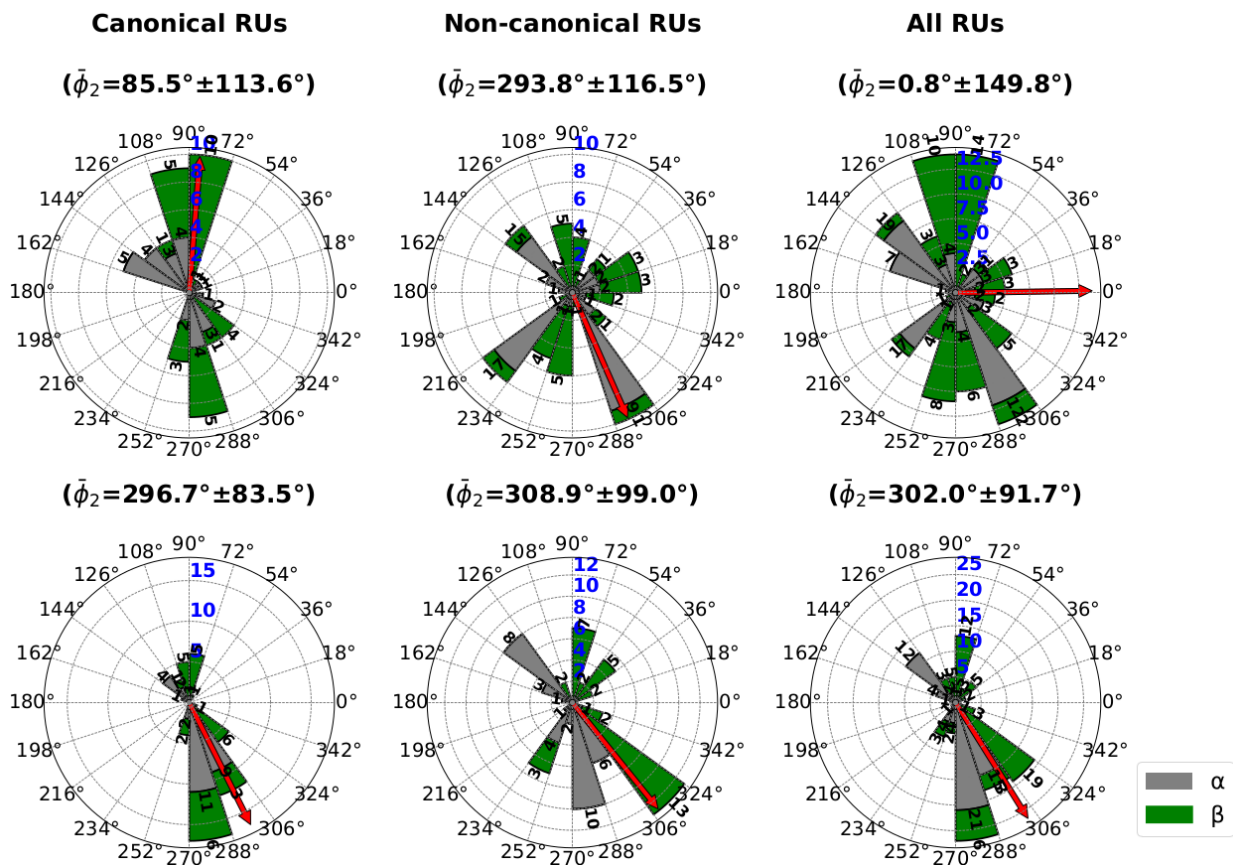
Figure 4.39 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR ring conformation of the mono-phosphate (HPO_3^-) nucleotides obtained from the classic (a+b) model. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

and C3'-endo. In aqueous environment both anomers prefer the range $270°(^3T_2)$ - $324°(E_4)$ that includes C3'-endo 5-MR conformations.

Overall, all β-nucleotides in vacuum prefer more the C2'-endo and C3'-endo, meanwhile the α-counterparts are found more frequently in the conformations between 126°-144° and 288°-306°. In aqueous environment both anomers prefer more the 270°-324° range.

This shows that to a certain extend there is an influence of the environment and the RU configuration on the puckering of the sugar. The correct anomer β tends to prefer a sugar ring conformation around the C2'-endo and C3'-endo.

Figure 4.40 shows the rose diagrams for the Nts obtained from the classic (a+b) pathway when the X element of the IL is replaced from P to As.

It can be noticed that replacing the IL does not impose major changes in the distributions of ϕ_2 for the canonical nucleotides in either environment or for either anomer. Meanwhile, for the non-canonical nucleotides it can be notice that their conformations are more spread in vacuum and in aqueous solution there is a higher preference for the β-anomers to be in the $306°(^3T_4)$ - $342°(^0T_4)$ range with a higher predominance of C4'-endo conformations. Overall, all nucleotides have similar distributions to the nucleotides with a monophosphate in both environments.

A comparison of both graphs shows that replacing the IL does not impose significant changes in the phase angle distribution for canonical nucleotides but it may change the distributions of non-canonical nucleotides. Another reason to think that maybe the HPO_4^- IL may have played a role in selecting the correct RUs for their incorporation later in the biopolymers. This result also shows that changing the IL does not impose significant changes to the sugar puckering, again supporting the idea that As can preserve the sugar puckering of the 5-MR sugars analyzed.

How does the distribution of the furanose puckering is affected if the nucleotides is modeled by using an alternative pathway in which the order of the reactants changes?

Figure 4.41 shows the rose diagrams for the mono-phosphate nucleotides obtained from the alternative (c+d) pathway.

A glance at the frequency bars from **Figure 4.41** shows that indeed when changing the order of reactants, the pattern of preferential puckering conformations tends to change for most of
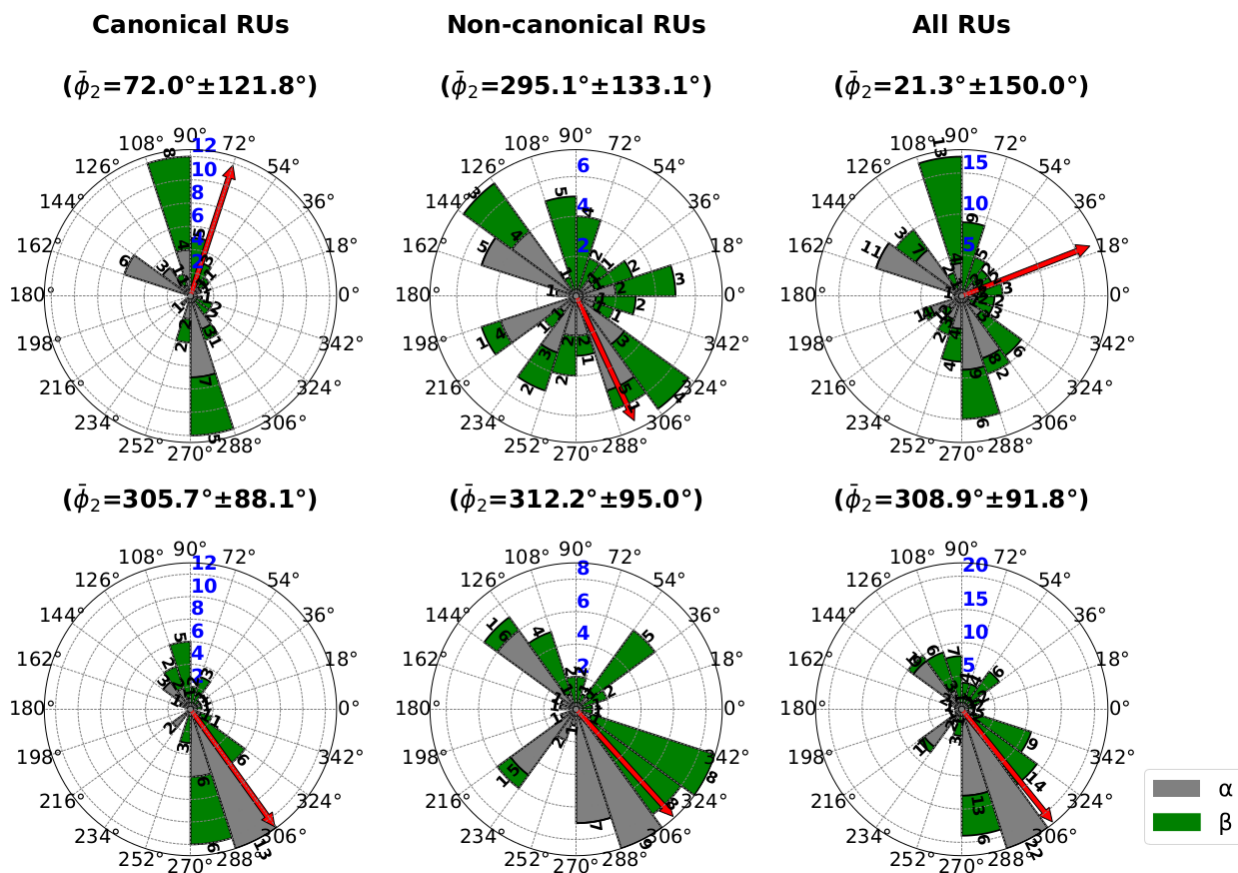
Figure 4.40 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR ring conformation of the monoarsenate ($HAsO_3^-$) nucleotides obtained from the classic (a+b) model. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
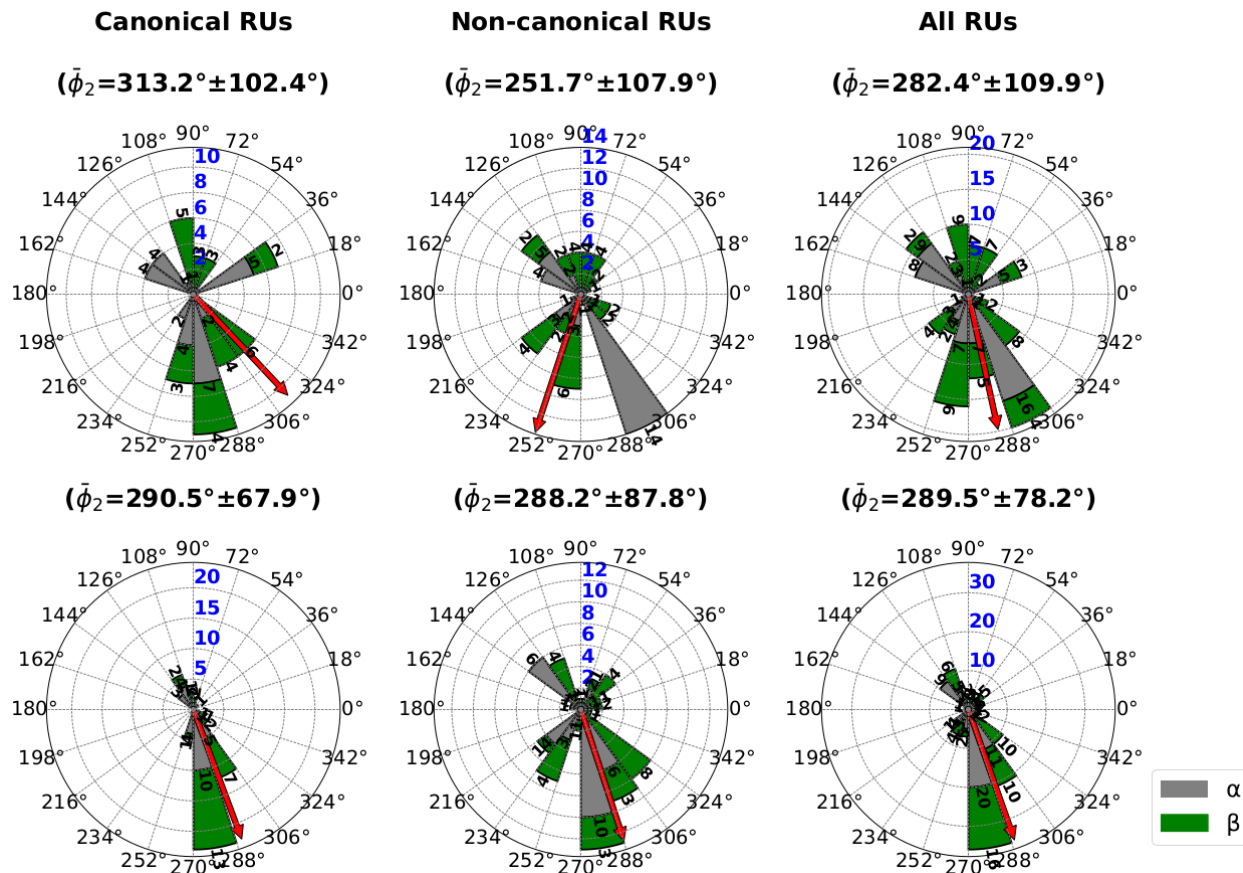
Figure 4.41 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR ring conformation of the mono-phosphate (HPO$_3^-$) nucleotides obtained from the alternative (c+d) model. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

canonical and non-canonical nucleotides in both environments, with the exception of the canonical and non-canonical nucleotides in aqueous solution. For the case of the canonical nucleotides in vacuum it is observed that the interval around $252°(E_2)$ - $324°(E_4)$ is preferred by both the α- and the β-anomers. This interval includes C3'-endo and C4'-exo puckering. The canonical nucleotides modeled in aqueous environment prefer the intervals C3'-endo $270°(^3T_2)$ - $288°(^3E)$. In the case of the non-canonical nucleotides in vacuum the β-anomers tend to prefer more the $216°$ - $270°$ interval, meanwhile there is large spread of nucleotides with small counts in the $54°(^2T_1)$ - $144°(^4E)$ range with C2'-endo conformations and the $216°(^1E)$ - $270°(^3T_2)$ interval. The α-configurations prefer almost exclusively the $288°(^3E)$ - $306°(^3T_4)$ range. When the non-canonical nucleotides are modeled in implicit solvation the interval $270°(^3T_2)$ - $324°(E_4)$ with C3'-endo conformations is preferred by both anomeric forms.

Overall, all nucleotides in vacuum have a large spread of frequencies in the $18°(^0T_1)$ - $162°(^4T_0)$ range and the $252°(E_4)$ - $306°(^3T_4)$, meanwhile in aqueous environment there is a clear preference for C3'-endo and C4'-exo conformations from $270°(^3T_2)$ to $324°(E_4)$.

Figure 4.42 shows the rose diagrams for the (c+d) nucleotides with a mono-arsenate. Changing the IL and the order for the addition of the components does not affect significantly the distributions of the nucleotides in water but it does in vacuum.

For instance, the β-anomers of the canonical nucleotides in vacuum tend to prefer more puckering conformations in the $72°(^2E)$ - $90°(^2T_3)$ and the $270°(^3T_2)$ - $324°(E_4)$, meanwhile the α-counterparts have a distribution of puckering from $126°(^4T_3)$ to $288°(^3E)$ that include more C4'-endo, C2'-endo sugar ring conformations. Both anomers of the non-canonical nucleotides in vacuum have a higher preference for the C2'-exo, C3'-endo and C4'-exo from $252°(E_2)$ to $324°(E_4)$. Overall, all nucleotides in vacuum tend to prefer more the $270°(^3T_2)$ - $342°(^0T_4)$ range.

The frequency histograms in aqueous solution show that the canonical nucleotides in the β-configuration prefer the $270°(^3T_2)$ - $306°(^3T_4)$ range with C3'-endo conformations, meanwhile the α-counterparts have a distribution of counts between $126°(^4T_3)$ - $270°(^3T_2)$. The non-canonical Nts modeled in aqueous environment have 5-MR conformations in almost all puckering intervals with 8 Nts having values of ϕ_2 in between $306°(^3T_4)$ - $324°(E_4)$. Overall, all Nts in aqueous solution prefer the $270°(^3T_2)$ - $306°(^3T_4)$ in the β-
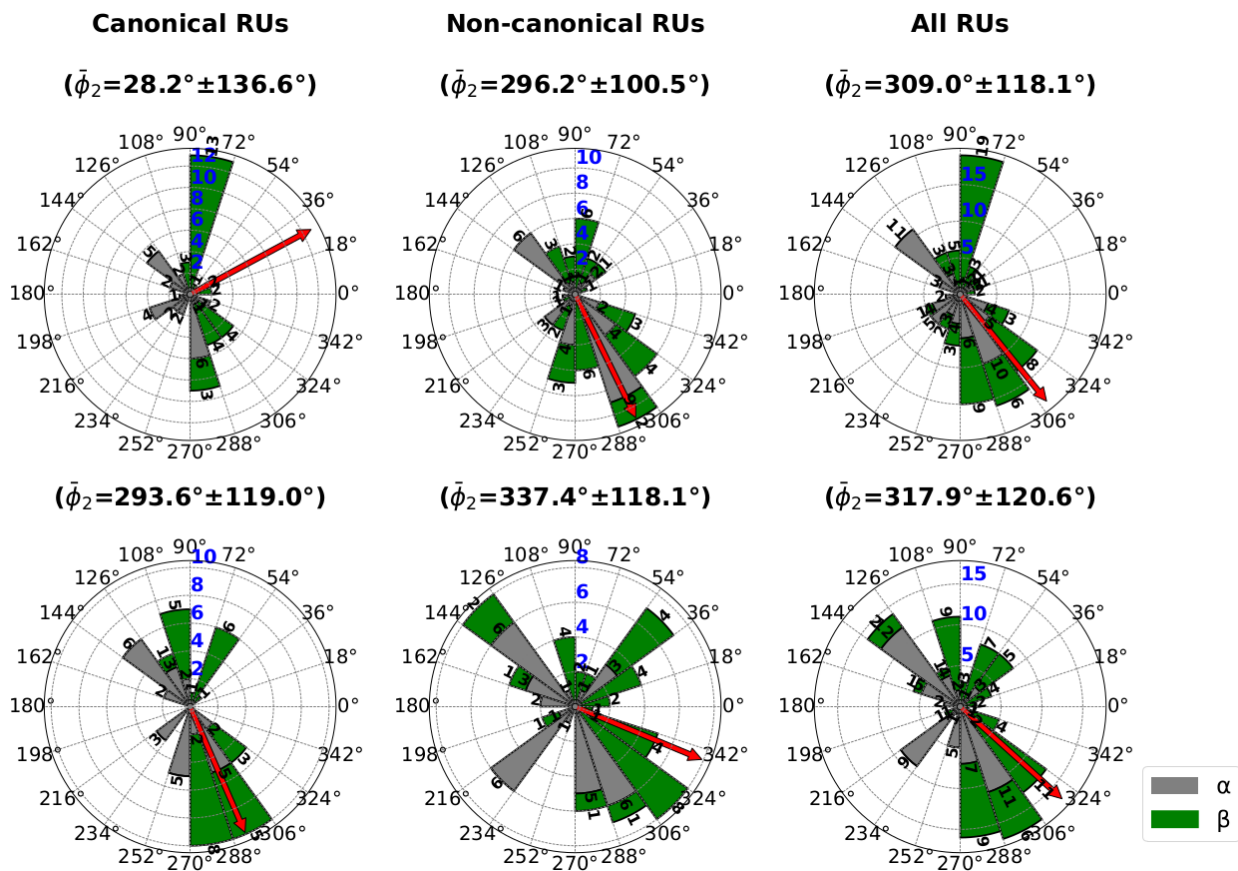
Figure 4.42 Rose diagrams (circular frequency histograms) for the phase angle ϕ_2 that defines the 5-MR ring conformation of the furanose mono-phosphate ($HAsO_3^-$) nucleotides obtained from the alternative (c+d) model. The blue vertical numbers represent the frequency scale as the radius of the different concentric circles. The red solid arrow marks the circular mean ϕ_2 in (°) for the angle. The green pies represent the β-anomers and the blue pies represent the α-counterparts of the different nucleosides. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.

configuration, meanwhile for the α-anomers the sugar ring puckering can be found in the $126°(^4T_3)$ - $306°(^3T_4)$ interval with the highest frequencies in the $126°(^4T_3)$ - $144°(^4E)$ and the $288°(^3E)$ - $306°(^3T_4)$ intervals.

Let's analyze now the Nts that contain a 6-MR sugar. As discussed previously in the case of 6-MR the three CP puckering coordinates q_2, q_3 and ϕ_2 can be transformed into a set of polar coordinates Q, ϕ and θ. In this case the phase angle θ defines if the 6-MR has a chair {C} ($\theta = 0°$ or $180°$), a boat {B} or skew boat {S} ($\theta = 90°$) and a half-chair {H} or envelope {E} ($\theta = 45°$ or $135°$). Hence, the projection of the sphere onto the 2D dimensions as a plot of θ vs ϕ is known as the Mercator projection and it is rational way to explore the distribution of the puckering conformations in a 6-MR.

Figures 4.43-**4.46** show the Mercator projection with a kernel density probability map of the distributions for either the β- and the α-anomers of Nts with a pyranose ring from the (a+b) and (c+d) pathways with both mono-phosphate and mono-arsenate ILs.

A glance at the four figures shows that the β- and some α-configurations of most Nts create one cluster close to the $^1C_4$ puckering and 1-2 clusters around the $^4C_1$ chair conformation. Very few Nts are observed to have other puckering conformation, e.g., some non-canonical nucleotides contain some β-configurations around the $^3H_4$ and some α-configurations around the $^4H_5$. The skew boat and boat puckering conformations have the lowest density probabilities.

In the P-form the sugar ring puckering is also less affected by the influence of the environment, the IL and the RU configuration, hence this is in support of the idea that, possibly, Nature has chosen 5-MR sugars instead of 6-MR since they are more prone to conformational changes when exposed to different factors.

4.4.6 Did thermodynamics drive the selection of the nucleotides containing T for DNA and U for RNA?

Can thermodynamics explain the selection of T for DNA and U for RNA? This section will explore this question upon theoretical calculations of the nucleotides obtained from the classic and alternative pathways.

In order to find an answer, and following the study reported in Ref. [23], two different sugar-
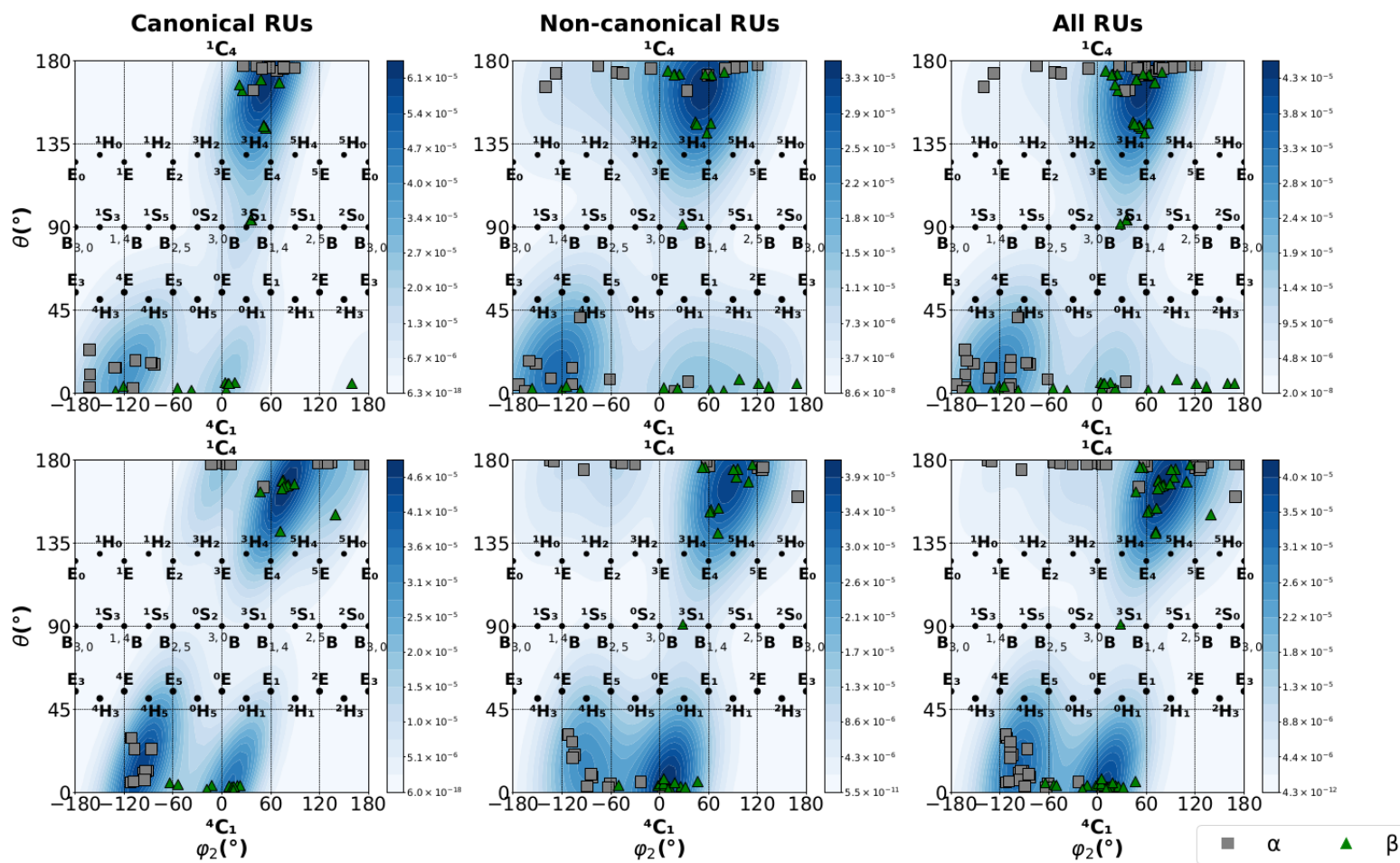
Figure 4.43 Mercator projection of the kernel density probability for the values of the phase angles ϕ and θ in (°) for the puckering of the 6-MR ring conformation of the pyranose mono-phosphate (HPO_3^-) nucleotides obtained from the classic (a+b) model. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
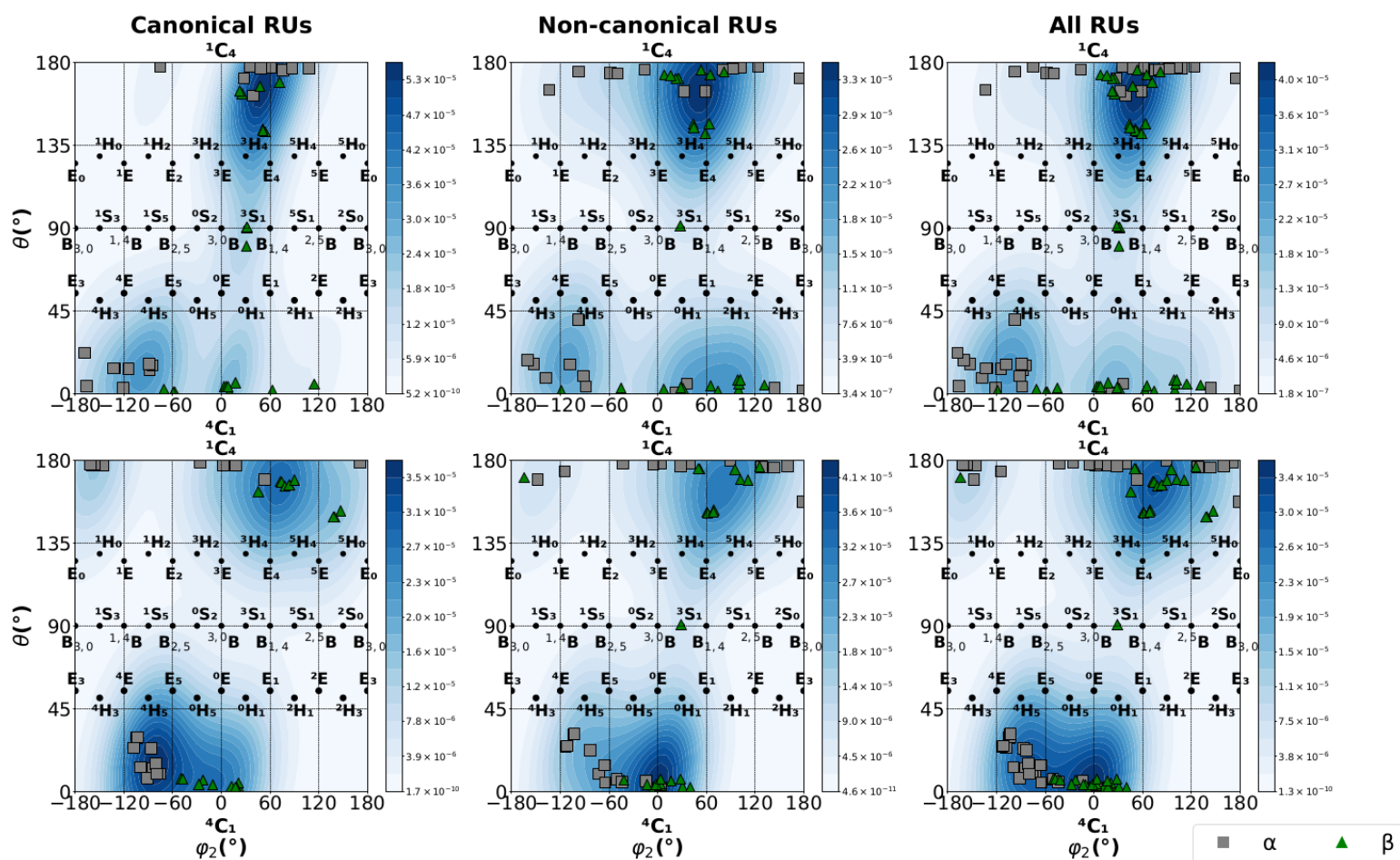
Figure 4.44 Mercator projection of the kernel density probability for the values of the phase angles ϕ and θ in (°) for the puckering of the 6-MR ring conformation of the pyranose mono-phosphate ($HAsO_3^-$) nucleotides obtained from the classic (a+b) model. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
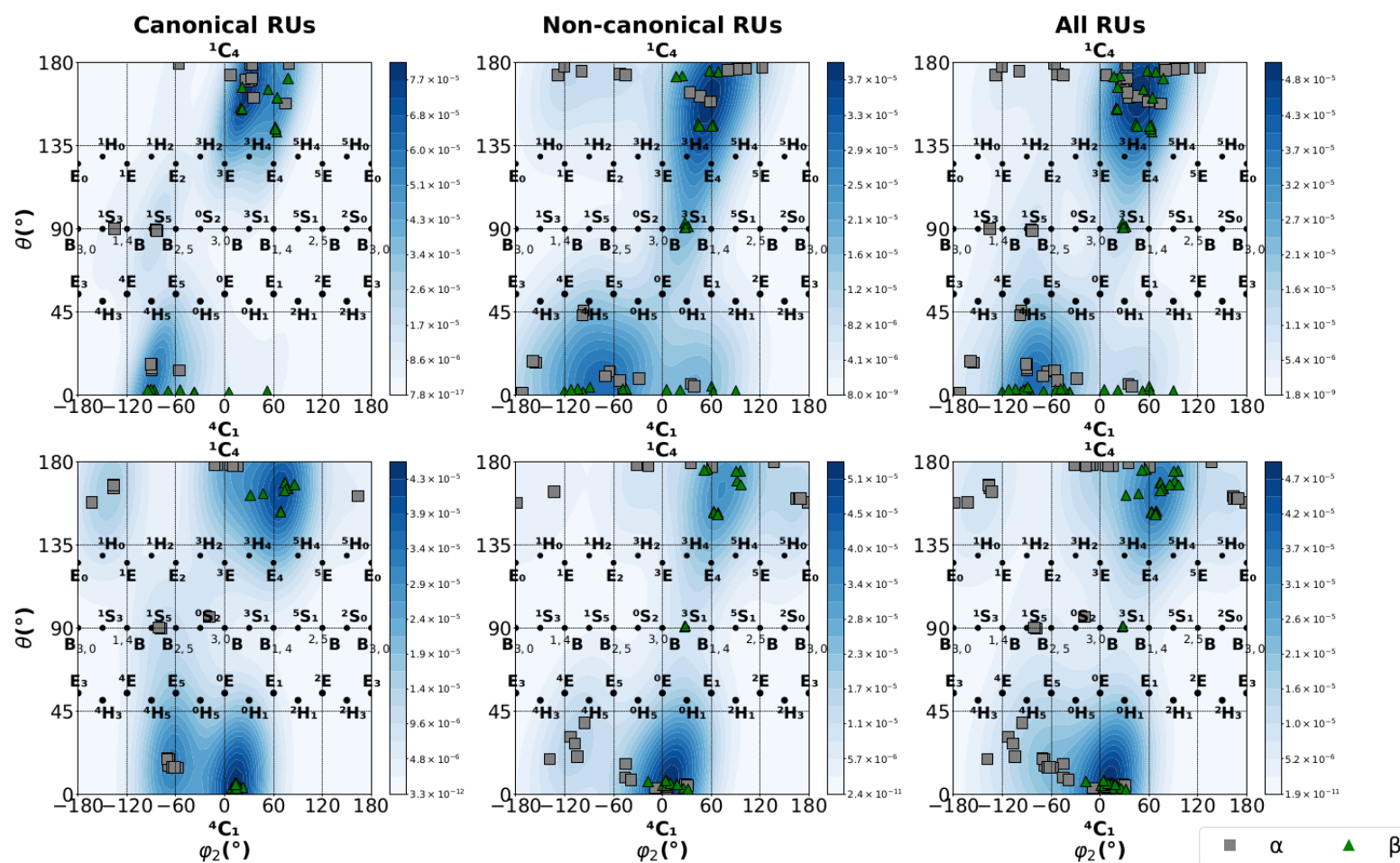
Figure 4.45 Mercator projection of the kernel density probability for the values of the phase angles ϕ and θ in (°) for the puckering of the 6-MR ring conformation of the pyranose mono-phosphate (HPO_3^-) nucleotides obtained from the classic (c+d) model. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model.
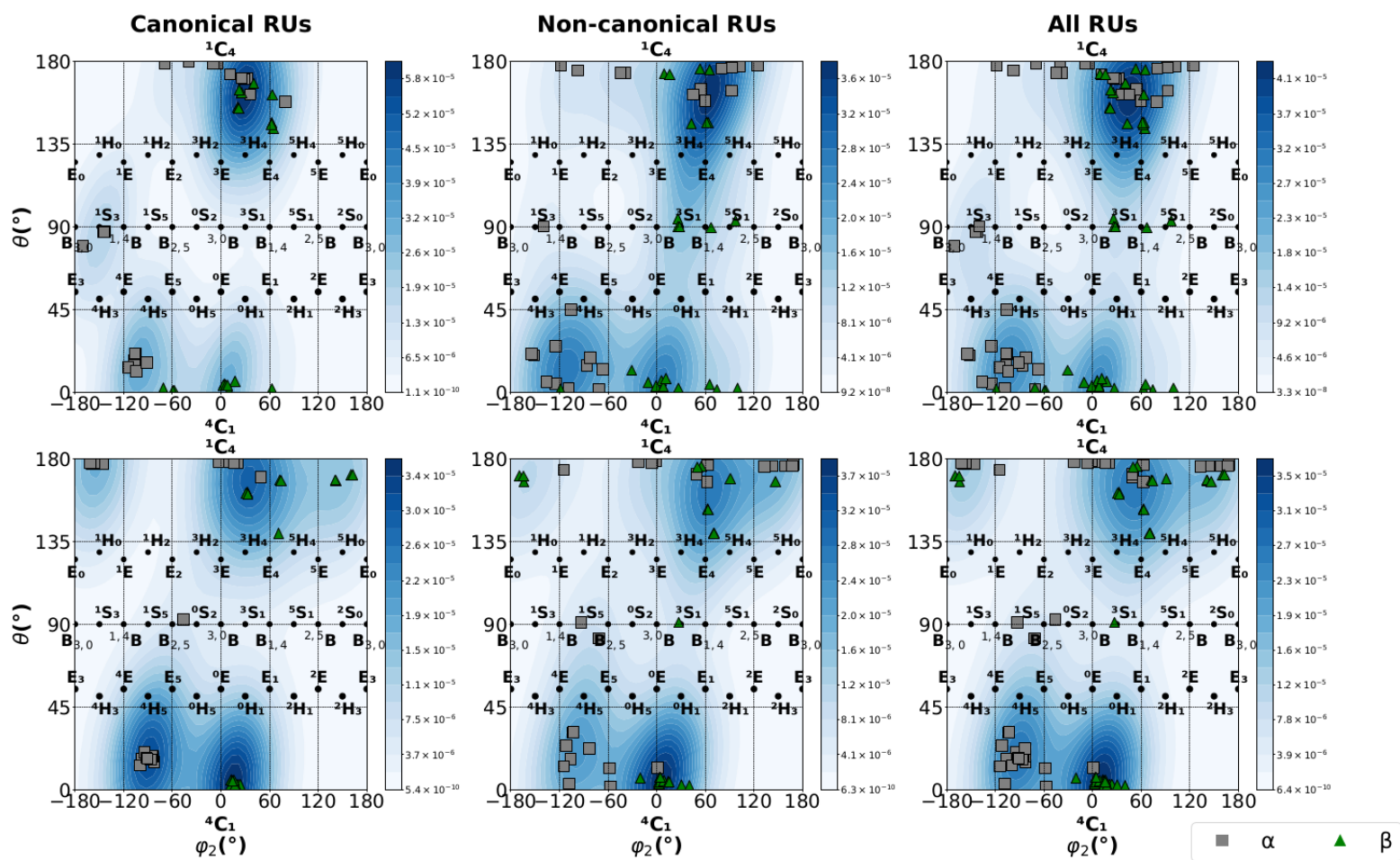
Figure 4.46 Mercator projection of the kernel density probability for the values of the phase angles ϕ and θ in (°) for the puckering of the 6-MR ring conformation of the pyranose mono-phosphate (HAsO$_3^-$) nucleotides obtained from the classic (c+d) model. (Top) B3LYP/6-311++G(d, p) in vacuum, (Bottom) B3LYP/6-311++G(d, p) in aqueous medium using the IEFPCM solvation model

exchange reactions for the β- and α-anomers of T and U Nts in the F- and P-form will be modeled, respectively:

$$\beta - rTMX \ + \ \beta - dUMX \rightarrow \ \beta - dTMX + \ \beta - rUMX \qquad\qquad (4.4)$$

and

$$\alpha - rTMX \ + \ \alpha - dUMX \rightarrow \ \alpha - dTMX + \ \alpha - rUMX \qquad\qquad (4.5)$$
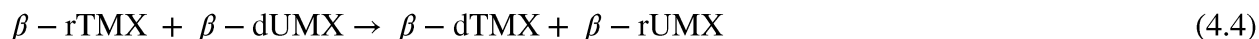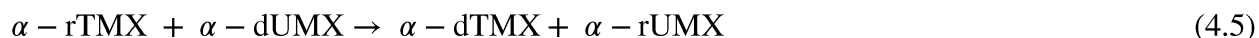
The energetic variations for the previous reactions are estimated as:

$$\Delta X = (X_{rUMP} + X_{dTMP}) - (X_{rTMP} + X_{dUMP}) \qquad\qquad (4.6)$$

where ΔX can be ΔE, $\Delta E_{(ZPE)}$ and ΔG^o.

In these reactions the products represent the "correct" pairs of D-2'-(deoxy)thymidine monophosphate (HPO$_3^-$-2dRibf(2dRib)-T or dTMP) and uridine monophosphate (HPO$_3^-$-Ribf(Rib)-T or rTMP) and the reactants are the "wrong" pair D-2'-(deoxy)uridine monophosphate (HPO$_3^-$-2dRibf(2dRib)-U or dUMP) and thymidine monophosphate (HPO$_3^-$-Ribf(Rib)-T or TMP) (see **Figure 4.47**).

We will consider all possible combinations between the different starting puckering conformations in the different furanose and pyranose sugar rings. A total of 256 reactions {128 reactions in each environment} are analyzed.

Table A8 in the Appendices section shows the values for the thermodynamic quantities for the 256 reactions. For the nucleotides from the classic (a+b) pathway in vacuum the most negative ΔG° are obtained for the reactions of the P-forms α-$^1C_4$-dTMP + α-$^1C_4$-UMP - α-$^4C_1$-TMP - α-$^4C_1$-dUMP (ΔG°=-69.3 kJ/mol), but also the reaction β-$^4C_1$-dT + β-$^4C_1$-U - β-$^1C_4$-T - β-$^1C_4$-dU favors the correct pair (ΔG°=-61.3 kJ/mol) and there is no noticeable difference between the ΔG° of these two reactions. The $\Delta G^\circ \geq 17$ kJ/mol favoring the incorrect pair are predicted for the reaction of the Nts in the P-form α-$^4C_1$-dTMP + α-$^4C_1$-UMP - α-$^1C_4$-TMP - α-$^1C_4$-dUMP (ΔG°=68.1 kJ/mol). In aqueous environment also the two reactions favoring and disfavoring the
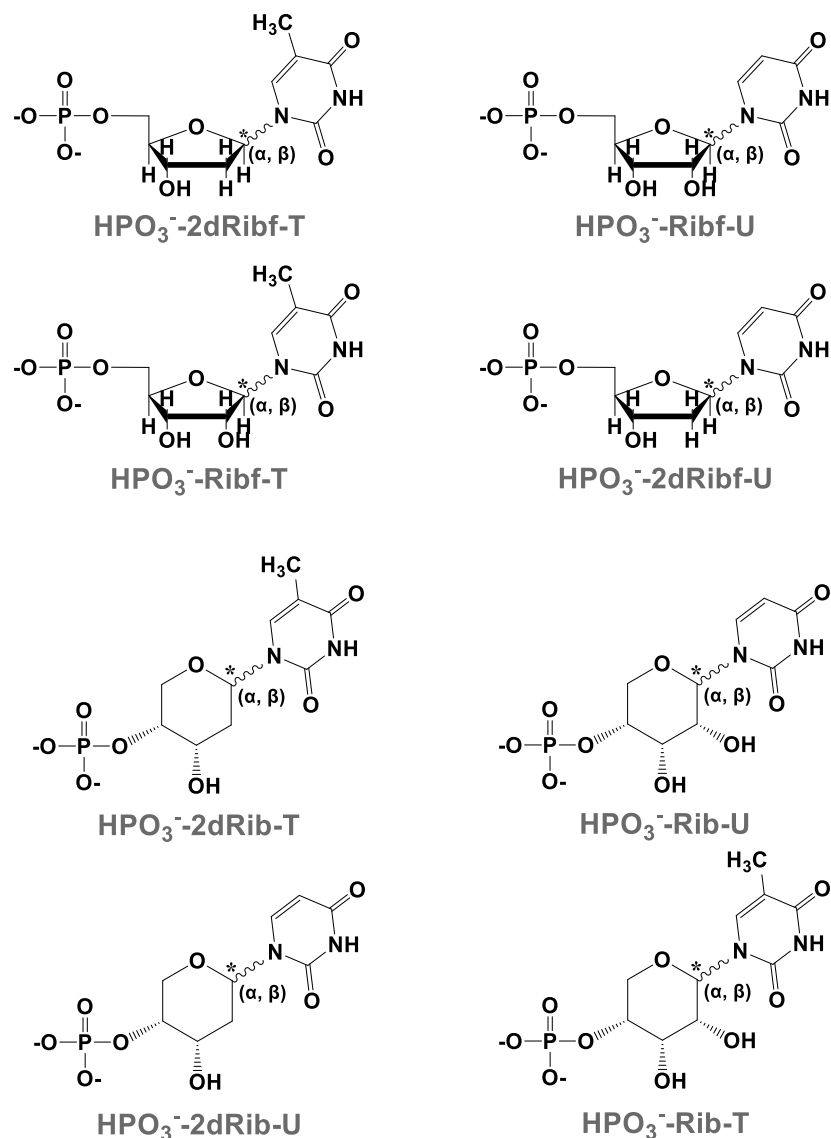
Figure 4.47 Natural and un-natural T, U nucleotides. The predominant form occurring in nature is D-2'-(deoxy)thymidine-monophosphate (HPO$_3^-$-2dRibf-T) and D-uridine-monophosphate (HPO$_3^-$-Ribf-U) occurring in DNA and RNA respectively. The unnatural forms, *i.e.* for the furanose ring, thymidine-monophosphate (HPO$_3^-$-Ribf-T) and 2'-(deoxy)uridine-monophosphate (HPO$_3^-$-dRibf-U) which are not normally incorporated in RNA and DNA, respectively. The four figures at the bottom represent the nucleotides with the pyranose ring HPO$_3^-$-2dRib-T, HPO$_3^-$-Rib-U and HPO$_3^-$-Rib-T and HPO$_3^-$-2dRib-U), also un-natural. See text and **Table A8**. The star (*) denotes the anomeric center (C1') of the sugar.

correct pair involve the α-anomer of the nucleotides with a pyranose ring α-$^1C_4$-dTMP + α-$^1C_4$-UMP - α-$^4C_1$-TMP - α-$^4C_1$-dUMP (ΔG°=-87.0 kJ/mol) and α-$^4C_1$-dT + α-$^4C_1$-U - α-$^1C_4$-T - α-$^1C_4$-dU (ΔG°=88.1 kJ/mol) respectively. Hence, the ΔG° from the nucleotides in the classic pathway suggest that these types of sugar-exchange reactions may had decided which pair would be selected by Nature for the sugar in its P-form and during the evolution of the Nts the naturally found F-form was selected.

When the Nts are obtained through the alternative (c+d) pathway in vacuum the reactions that favors or unfavor the correct pair involve the F-form and the β-configuration of the RU. These reactions are β-$^3T_2$-dTMP + β-$^3T_2$-UMP - β-$^2T_3$-TMP - β-$^2T_3$-dUMP (ΔG°=-33.5 kJ/mol) and β-$^2T_3$-dTMP + β-$^2T_3$-UMP - β-$^3T_2$-TMP - β-$^3T_2$-dUMP (ΔG°=59.3 kJ/mol). In aqueous solution the α-configuration of the F-form favors the correct dTMP and UMP pairs $\{\alpha$-$^2T_3$-dT + α-$^2T_3$-U - α-$^3T_2$-T - α-$^3T_2$-dU$\}$ (ΔG°=-20.5 kJ/mol).

When the sugar of the nucleotides contains a pyranose instead of a furanose in the alternative pathway the sugar-exchange reactions of the β-anomers in either environment favors or disfavors the correct pair. In vacuum these reactions are β-$^4C_1$-dT + β-$^4C_1$-U - β-$^1C_4$-T - β-$^1C_4$-dU (ΔG°=-51.2 kJ/mol), β-$^1C_4$-dT + β-$^1C_4$-U - β-$^4C_1$-T - β-$^4C_1$-dU (ΔG°=51.8 kJ/mol). In implicit solvation the reactions are: β-$^4C_1$-dT + β-$^4C_1$-U - β-$^1C_4$-T - β-$^1C_4$-dU (ΔG°=-42.8 kJ/mol) and β-$^1C_4$-dT + β-$^1C_4$-U - β-$^4C_1$-T - β-$^4C_1$-dU (ΔG°=40.0 kJ/mol). The rest of the energies are within the intrinsic error of the method; hence they are inconclusive.

The analysis of the energy changes for the sugar exchange reaction of the F-form of the T and U Nts in Castanedo and Matta, 2022 [23]: (*i*) reinforces the classical pathway as preferable and (*ii*) indicates an advantage of the canonical pairs compared to the no-canonical pairs when gauged by the "sugar exchange reactions" for the β-anomer in vacuum and vanishes in the case of the α-anomers (see table. 6 in Castanedo and Matta's paper).

In the present study it is observed that when the modeling of the T and U Nts starts from different sugar ring conformations and sizes there are specific combinations puckering in the β-anomer of the P-form that may favor the correct pair dTMP, UMP. This observation is valid for the Nts obtained from either pathway.

4.4   Conclusions

In this chapter and throughout this PhD thesis dissertation the role of thermodynamics in the evolutionary emergence of the building blocks of contemporary NAs from a proto- or pre-RNA has been explored using computational modeling.

The first question seeking an answer is the following: did Nature select the β-anomer of today's nucleotides instead the α-counterpart based on the thermodynamic stability and flexibility of one anomer over the other? Thibaudeau and coworkers [39, 40, 41] have proposed that the pseudorotational equilibrium between the two major puckering forms ($^2T_3$ and $^3T_2$) of ribose is an important factor to consider when comparing stabilities of both anomeric forms. Following this idea, it has been predicted at the DFT level of theory the ΔG° for the pseudorotational equilibrium of the F- and P-form of canonical and non-canonical β- and α-Nts. Additionally, 4 anomer-exchange reactions in each case have been also analyzed. These reactions involve the same or different sugar ring initial conformations. The Nts modeled have been obtained from two different pathways (sequence of condensation reactions between the TCs, RUs and ILs) and the IL contain either phosphorus or arsenic.

The results obtained are in agreement with the Tcw proposition that for the canonical Nts modeled in vacuum with a 5-MR and HPO_4^- the pseudorotation of the β-anomer favors the $^3T_2$ puckering of the sugar and for some of the β-anomers of nucleotides with 6-MR the $^4C_1$ puckering is favored. The rest of energies are within the expected error bars of the theoretical method {$\approx \pm$ 17 kJ/mol}. There were no significant differences between the different anomers when P was replaced by As.

An analysis of the anomer-exchange reactions energies shows that in the cases in which the energies are not negligible the correct β-anomer is favored for either pathway. The reaction that favored one of the anomers was mostly reaction#2 (**Figure 4.3**) for either 5- or 6-MR TCs.

An analysis of the ΔG° for a classic pathway (a+b) shows that in vacuum the synthesis of most canonical or non-canonical β-Nts is favored ($\Delta G^\circ \leq$ -17 kJ/mol), meanwhile in aqueous solution most of the reactions are not favored. This finding is in support of the "water problem".

On the other hand, for the alternative pathway (c+d) and in agreement with experimental results from the literature [6, 7, 8] the synthesis of some non-canonical nucleotides in implicit

solvation is favored, e.g., TAP-, BA C^5-glycosilated and MM. Either classic or alternative synthesis of the BA-N and CA nucleotides is unfavored across the board. BA-C^5 nucleotides are overall the most favored Nts followed by TAP-C^5 and MM Nts in a less extension. These results are in support of the idea that the first proto-RNA contained a C-glycosylated base either BA or TAP. The alternative pathway emerges as the way to synthesize non-canonical nucleotides more effectively than the classic pathway when water is present.

The synthesis in vacuum of the building blocks of nucleic acids is suggestive of a possible extraterrestrial scenario for the synthesis of these molecules, e.g., in dense or diffuse interstellar clouds [69, 70, 71]. Additionally, either artificial synthetic route gives similar results for nucleotides containing either P or As based ILs, which suggests that either mono-phosphate or mono-arsenate Ns can be synthetized in prebiotic conditions.

Either classic or alternative synthesis of the Nts containing glycerol as TC are predicted to be the most feasible thermodynamically overall. In most of cases in which either synthetic route favors the formation of a Nt is the *β*-anomer more favored than the *α*-counterpart.

There is an overall preference for the *anti*-conformation, mostly in the vacuum environment, and for canonical Nts. This may suggest that possibly some of the conformational behavior observed in the nucleotides of today informational biopolymers may have been inherited from their monomeric units if they came from a classic synthesis in vacuum. When the Nts are modeled using an alternative pathway there is a higher preference for the most frequent *anti*-conformation for the naturally observed *β*-anomer, meanwhile the *α*-counterpart is more frequent in the *syn* position in both vacuum and aqueous environment. This suggests that if some nucleotides emerged from the reaction between a RU and a TC-IL derivative, then the propensity for the *anti*-conformation increases.

There is less flexibility or number of sugar ring puckering conformations when the sugar is in the P-form instead of the F-form. In the P-form there is also less influence of the environment, the IL and the RU configuration on the change of puckering, hence this result may suggest that maybe Nature chose 5-MR 2dRibf or Ribf instead of the 6-MR 2dRib or Rib since they are more sensible to change its conformations when exposed to different factors. In either pathway, environment or IL, the canonical nucleotides tend to have the preferential $^2T_3$ or $^3T_2$ for 5-MRs and the $^4C_1$ and $^1C_4$ for the 6-MRs. For the non-canonical nucleotides, a wider range of sugar ring puckering are observed in either vacuum or aqueous

solution.

An analysis of the sugar-exchange reaction between different "correct" T for DNA and U for RNA nucleotides including the two different initial sugar conformations shows that the reactions involving a 6-MR either favors or disfavors the correct pair. In our first report [23] on the role of thermodynamics on the selection of the correct dTMP, UMP pair was shown that the β-anomer of the nucleotides obtained from the classic pathway favored the correct pair in vacuum. In this study it is shown that when the sugar ring conformation is taken into consideration the selection of T or U for the building blocks of either DNA or RNA may depend on the initial sugar ring conformation and the size of the ring. This result suggests that if the Nts emerged from a TC-IL, similarly to what the polymer fusion model proposes, then the Nts with their TC in the P-form depending on specific combinations of puckering either favored or disfavored the correct pyrimidine pairs. As it was discussed in section 4.4.2 the synthesis of pyrimidine nucleotides with a HPO_4^- is predicted to be thermodynamically plausible in vacuum using an alternative pathway and not in aqueous environment. Hence it can be concluded that the preference for a specific pyrimidine base for either DNA or RNA may have come from a proto-NA that contained a ribopyranose and/or the 2'-deoxyribopyranose in the specific $^4C_1$ puckering, establishing in that way the correct pairs and eventually the sugar interconverted into the F-form leaving the correct forms available for today's DNA and RNA.

In summary, the results presented in this chapter show that thermodynamics can support the selection of C-glycosylated non-canonical nucleotides using an alternative pathway in aqueous solution, mostly BA C^5-glycosilated. The replacement of P by As in the HXO_4^- does not seems to affect the different in stabilities of anomers, prebiotic synthesis, RU *syn-*, *anti*-conformation and sugar ring puckering. Hence, the prebiotic emergence of an As-based proto-NA cannot be disregarded. This is in agreement with the reports from Mládek and coworkers [56]. Sponer and coworkers have argued that a possible cause for the early selection of the P over As in the first ILs could be based on the kinetics of the hydrolysis of arsenate and phosphate esters [22, 56, 57, 58, 55].

In this chapter, the thermodynamic plausibility for the prebiotic emergence of nucleotides containing glycerol instead a sugar ribose or 2'-deoxyribose is also suggested. This may support the view that the first RNAs contained a simpler TC-IL backbone that was eventually replaced by today's TC as a way to transfer and preserve the genetic information. There is evidence in the

literature showing that GNA strands are resistant to hydrolysis and can cross-pair with DNA/RNA strand [15, 72].

4.5   References

[1]   D. L. Nelson and M. M. Cox, "Nucleotides and nucleic acids," in *Lehninger principles of biochemistry*, New York, W.H. Freeman & Company, 2013, pp. 281-313.

[2]   S. A. Benner, H. J. Kim and M. A. Carringa, "Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides and RNA," *Accounts of Chemical Research,* vol. 45, pp. 2025-2034, 2012.

[3]   B. J. Cafferty and N. V. Hud, "Was a Pyrimidine-Pyrimidine Base Pair the Ancestor of Watson-Crick Base Pairs? Insights from a Systematic Approach to the Origin of RNA," *Israel Journal of Chemistry,* vol. 55, pp. 891-905, 2015.

[4]   A. E. Engelhart and N. V. Hud, "Primitive genetic polymers," *Perspectives in Biology,* vol. 2, article #a002196 (pp. 1-21), 2010.

[5]   N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, "The origin of RNA and "my grandfather's axe"," *Chemistry & Biology,* vol. 20, pp. 466-474, 2013.

[6]   D. M. Fialho, T. P. Roche and N. V. Hud, "Prebiotic syntheses of noncanonical nucleosides and nucleotides," *Chemical Reviews,* vol. 120, pp. 4806-4830, 2020.

[7]   D. M. Fialho, *Physical organic principles governing the spontaneous prebiotic emergence of proto-nucleic acids,* Georgia: Georgia Tech Library, 2019, pp. 1-3.

[8]   D. M. Fialho, K. C. Clarke, M. K. Moore, G. B. Schuster, R. Krishnamurthy and N. V. Hud, "Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides," *Organic and Biomolecular Chemistry,* vol. 16, pp. 1263-1271, 2018.

[9] B. J. Cafferty, D. M. Fialho and N. V. Hud, "Searching for possible ancestors of RNA: The self-assembly hypothesis for the origin of proto-RNA," in *Prebiotic chemistry and chemical evolution of nucleic Acids, nucleic acids and molecular biology*, vol. 35, C. Menor-Salván, Ed., Philadelphia, Pennsylvania: Springer, 2018, pp. 143-174.

[10] E. Lescrinier, M. Froeyen and P. Herdewijn, "Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids," *Nucleic Acids Research,* vol. 31, pp. 2975-2989, 2003.

[11] P. Gu, G. Schepers, J. Rozenski , A. Aerschot and P. Herdewijn, "Base pairing properties of D- and L-Cyclohexene nucleic acids (CeNA)," *Oligonucleotides,* vol. 13, pp. 479-489, 2003.

[12] L. Orgel, "A simpler nucleic acid," *Science,* vol. 290, pp. 1306, 2000.

[13] K. U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, "Chemical etiology of nucleic acid structure: the α-threofuranosyl - (3' —>2') oligonucleotide system," *Science,* vol. 290, pp. 1347-1351, 2000.

[14] M. Hernández-Rodríguez, J. Xie, Y. M. Osornio and R. Krishnamurthy, "Mapping the landscape of potentially primordial informational oligomers: (30/20) - D - phosphoglyceric acid linked acyclic oligonucleotides tagged with 2, 4 - disubstituted 5 - aminopyrimidines as recognition elements," *Chemistry: An Asian Journal,* vol. 6, pp. 1252-1262, 2011.

[15] E. Meggers and L. Zhang, "Synthesis and properties of the simplified nucleic acid glycol nucleic acid," *Accounts of Chemical Research,* vol. 43, pp. 1092-1102, 2010.

[16] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

[17] P. E. Nielsen, "Peptide nucleic acid (PNA). Implications for the origin of the genetic material and the homochirality of life," *American Institute of Physics Conference Proceedings,* vol. 379, pp. 55-61, 1996.

[18] P. E. Nielsen, "Peptide nucleic acid (PNA): a model structure for the primordial genetic material?," *Origins of Life and Evolution of the Biosphere,* vol. 23, pp. 323-327, 1993.

[19] P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide," *Science,* vol. 254, pp. 1497-1500, 1991.

[20] P. R. Barry, A. A. Ajees and T. R. McDermott, "Life and death with arsenic," *Bioessays,* vol. 33, pp. 350-357, 2011.

[21] F. Wolfe-Simon, J. S. Blum, T. R. Kulp, G. W. Gordon, S. E. Hoeft, J. Pett-Ridge, J. F. Stolz, S. M. Webb, P. K. Weber, P. C. W. Davies, A. D. Anbar and R. S. Oremland, "A bacterium that can grow by using arsenic instead of phosphorus," *Science,* vol. 332, pp. 1163-1166, 2011.

[22] J. Šponer, A. Mládek, J. E. Šponer, D. Svozil, M. Zgarbová, P. Banáš, P. Jurečka and M. Otyepka, "The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies," *Physical Chemistry Chemical Physics,* vol. 14, pp. 15257-15277, 2012.

[23] L. A. M. Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article #e09657 (pp. 1-12), 2022.

[24] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, "Open Babel: an open chemical toolbox," *Journal of Cheminformatics,* vol. 3, article #33 (pp. 1-14), 2011.

[25] M. Frenkel-Pinter, M. Samanta, G. Ashkenasy and L. J. Leman, "Prebiotic peptides: molecular hubs in the origin of life," *Chemical Reviews,* vol. 120, pp. 4707-4765, 2020.

[26] J. Wu, Q. Meng, H. Ren, H. Wang, J. Wu and Q. Wang, "Recent advances in peptide nucleic acid for cancer bionanotechnology," *Acta Pharmacologica Sinica,* vol. 38, pp. 798-805, 2017.

[27] E. J. Lee, H. K. Lim, Y. S. Cho and S. S. Hah, "Peptide nucleic acids are an additional class of aptamers," *RSC Advances,* vol. 3, pp. 5828-5831, 2013.

[28] e. Mateo-Martí and C.-M. Pradie, "A novel type of nucleic acid-based biosensors: the use of PNA probes, associated with surface science and electrochemical detection techniques," in *Intelligent and biosensors*, Rijeka, InTech, 2010, pp. 323-344.

[29] P. E. Nielsen, "Peptide nucleic acids and the origin of life," *Chemistry & Biodiversity,* vol. 4, pp. 1996-2002, 2007.

[30] V. Menchise, G. De Simone, T. Tedeschi, R. Corradini, S. Sforza, R. Marchelli, D. Capasso, M. Saviano and C. Pedone, "Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, pp. 12021-12026, 2003.

[31] P. E. Nielsen, "Peptide Nucleic Acid (PNA). Implications for the origin of the genetic material and the homochirality of life," *AIP Conference Proceedings,* vol. 379, pp. 55-61, 1996.

[32] B. Mennucci, "Polarizable continuum model," *Advanced Review,* vol. 2, pp. 386-404, 2012.

[33] J. Tomasi, "Selected features of the polarizable continuum model for the representation of solvation," *WIREs Computational Molecular Science,* vol. 1, pp. 855-867, 2011.

[34] V. Barone, M. Cossi and J. Tomasi, "A new definition of cavities for the computation of solvation free energies by the polarizable continuum model," *The Journal of Chemical Physics,* vol. 107, pp. 3210-3221, 1997.

[35] M. Cossi, V. Barone, R. Cammi and J. Tomasi, "Ab initio study of solvated molecules: A new implementation of the polarizable continuum model," *Chem. Phys. Lett,* vol. 255, pp. 327-335, 1996.

[36] S. Miertuš and J. Tomasi, "Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes," *Chemical Physics,* vol. 65, pp. 239-245, 1982.

[37] S. Miertuš, E. Scrocco and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chemical physics,* vol. 55, pp. 117-129, 1981.

[38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji , X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian , J. V. Ortiz, A. F. Izmaylov , J. L. Sonnenberg , D. Williams-Young , F. Ding, F. Lipparini , F. Egidi, J. Goings , B. Peng , A. Petrone , T. Henderson , D. Ranasinghe , V. G. Zakrzewski , J. Gao, N. Rega, G. Zheng , W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida , T. Nakajima , Y. Honda , O. Kitao , H. Nakai , T. Vreven , K. Throssell , J. A. J. Montgomery , J. E. Peralta , F. Ogliaro , M. J. Bearpark , J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari , A. P. Rendell , J. C. Burant, S. S. Iyengar , J. Tomasi , M. Cossi , J. M. Millam , M. Klene , C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, revision C.01, Wallingford CT: Gaussian Inc, 2019.

[39] C. Thibaudeau, P. Acharya and J. Chattopadhyaya, Stereoelectronic effects in nucleosides and nucleotides and their structural implications, Uppsala: Uppsala University Press, 2005.

[40] C. Thibaudeau, A. Foldesi and J. Chattopadhyaya, "The first experimental evidence for a larger medium-dependent flexibility of natural β-D-nucleosides compared to the α-D-nucleosides," *Tetrahedron,* vol. 53, pp. 14043-14072, 1997.

[41] C. Thibaudeau, J. Plavec and J. Chattopadhyaya, "Quantitation of the anomeric effect in adenosine and guanosine by comparison of the thermodynamics of the pseudorotational equilibrium of the pentofuranose moiety in N- and C-nucleosides," *Journal of the American Chemical Society,* vol. 116, pp. 8033-8031 , 1994.

[42] W. Martin, J. Baross, D. Kelley and M. J. Russell, "Hydrothermal vents and the origin of life," *Microbiology,* vol. 6, pp. 805-813, 2008.

[43] G. Wachtershauser, "From volcanic origins of chemoautotrophic life to bacteria, Archaea and Eukarya," *Phylosophical Transactions of the Royal Society B,* vol. 361, pp. 1787-1808, 2006.

[44] F. C. Dobbs and K. A. Selph , "Thermophilic bacterial activity in a deep-sea sediment from the Pacific ocean," *Aquatic Microbial Ecology,* vol. 13, pp. 209-212, 1997.

[45] C. Thibaudeau, P. Acharya and J. Chattopadhyaya, "Stereoelectronic effects in nucleosides and nucleotides," in *Stereoelectronic effects in nucleosides and nucleotides and their structural implications*, 2 ed., Uppsala, Uppsala University Press, 2005, pp. 22-46.

[46] T. W. Davis, "A common misundrstanding of Hess's law," *Journal of Chemical Education,* vol. 28, pp. 584-585, 1951.

[47] C. V. Mungi, S. K. Singh, J. Chugh and S. Rajamani, "Synthesis of barbituric acid containing nucleotide and its implications for the origins of primitive informational polymers," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20144-20152, 2016.

[48] B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy and N. V. Hud, "Spontaneous formation and base pairing of plausible prebiotic nucleotides in water," *Nature Communications,* vol. 7, article #11328, (pp. 1-8), 2016.

[49] V. M. Kolb, J. P. Dworkin and S. L. Miller, "Alternative bases in the RNA world: the prebiotic synthesis of urazole and its ribosides," *Journal of Molecular Evolution,* vol. 38, pp. 549-557, 1994.

[50] W. G. Scott, A. Szöke , J. Blaustein, S. M. O'Rourke and M. P. Robertson, "RNA catalysis, thermodynamics and the origin of life," *Life,* vol. 4, pp. 131-141, 2014.

[51] L. L. Martin, P. J. Unrau and U. F. Müller, "RNA synthesis by in vitro selected ribozymes for recreating an RNA world," *Life,* vol. 5, pp. 247-268, 2015.

[52] H.-J. Kim and S. A. Benner, "Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates," *Proceedings of*

the National Academy of Sciences of the United States of America, vol. 114, pp. 11315-11320, 2017.

[53] S. Guntha, R. Krishnamurthy and A. Eschenmoser, "Regioselective α-phosphorylation of aldoses in aqueous solution," *Angewandte Chemie,* vol. 30, pp. 2281-2285, 2000.

[54] M. I. Fekry, P. A. Tipton and K. S. Gates, "Kinetic consequences of replacing the internucleotide phosphorus atoms in DNA with arsenic," *ACS Chemical Biology,* vol. 6, pp. 127-130, 2011.

[55] J. W. Long and J. Ray Jr, "Kinetics and thermodynamics of the formation of glucose arsenate. Reaction of glucose arsenate with phosphoglucomutase," *Biochemistry,* vol. 12, pp. 3932-3937, 1973.

[56] A. Mládek , J. Šponer, B. G. Sumpter, M. Fuentes-Cabrera and J. E. Šponer, "Theoretical modeling on the kinetics of the arsenate-ester hydrolysis: implications to the stability of As-DNA," *Physical Chemistry Chemical Physics,* vol. 13, pp. 10869-10871, 2011.

[57] C. D. Baer, J. O. Edwards and P. H. Rieger, "Kinetics of the hydrolysis of arsenate(V) triesters," *Inorganic Chemistry,* vol. 20, pp. 905-907, 1981.

[58] C. D. Baer, J. O. Edwards, M. J. Kaus, T. G. Richmond and P. H. Rieger, "Kinetics of an associative ligand-exchange process: alcohol exchange with arsenate(V) triesters," *Journal of the American Chemical Society,* vol. 102, pp. 5793-5798, 1980.

[59] A. Mládek , J. Šponer, B. G. Sumpter, M. Fuentes-Cabrera and J. E. Šponer, "On the geometry and electronic structure of the As-DNA backbone," *The Journal of Physical Chemistry Letters,* vol. 2, pp. 389-392, 2011.

[60] S. Kaur, P. Sharma and S. D. Wetmore, "Can cyanuric acid and 2, 4, 6-triaminopyrimidine containing ribonucleosides be components of prebiotic RNA? Insights from QM calculations and MD simulations," *ChemBioChem,* vol. 20, pp. 1415-1415, 2019.

[61] S. Kaur, P. Sharma and S. D. Wetmore, "Structural and electronic properties of barbituric acid and melamine-containing ribonucleosides as plausible components of prebiotic RNA:

implications for prebiotic self-assembly," *Physical Chemistry Chemical Physics,* vol. 19, pp. 30762-30771, 2017.

[62] A. Gelbin, B. Schneider, L. Clowney, S.-H. Hsieh, W. K. Olson and H. M. Berman, "Geometric parameters in nucleic acids: sugar and phosphate constituents," *Journal of the American Chemical Society,* vol. 118, pp. 519-529, 1996.

[63] N. Foloppe, B. Hartmann, L. Nilsson and A. D. MacKerell Jr, "Intrinsic conformational energetics associated with the glycosyl torsion in DNA: a quantum mechanical study," *Biophysical Journal,* vol. 82, pp. 1554-1569, 2002.

[64] W. Saenger, Principles of nucleic acid structure, New York: Springer-Verlag, 1984.

[65] D. Cremer and J. A. Pople, "A general definition of ring puckering coordinates," *Journal of the American Chemical Society,* vol. 97, pp. 1354-1358, 1975.

[66] D. Svozil, J. Kalina, M. Omelka and B. Schneider, "DNA conformations and their sequence preferences," *Nucleic Acids Research,* vol. 36, pp. 3690-3706, 2008.

[67] P. S. Ho and M. Carter, "DNA structure: alphabet soup for the cellular soul," in *DNA Replication-Current Advances*, Rijeka, InTech, 2011, pp. 4-28.

[68] S. A. Sandford, M. Nuevo, P. P. Bera and J. T. Lee, "Prebiotic astrochemistry and the formation of molecules of astrobiological interest in interestellar clouds and protostellar disks," *Chemical Reviews,* vol. 120, pp. 4616-4659, 2020.

[69] N. Kitadai and S. Maruyama, "Origins of building blocks of life: A review," *Geoscience Frontiers,* pp. 1117-1153, 2018.

[70] F. Trixler, "Quantum tunneling to the origin and evolution of life," *Current Organic Chemistry,* vol. 17, pp. 1758-1770, 2013.

[71] L. Zhang, A. Peritz and E. Meggers, "A simple glycol nucleic acid," *Journal of the American Chemical Society,* vol. 127, pp. 4174-4175, 2005.

Appendices

Table A1. Differences between the energies in kJ/mol for the most stable β- and α-anomers of the canonical and non-canonical nucleotides obtained using the classic (a+b) set of reactions in vacuum and in aqueous environment (equations 4.2-4.7). The energies for all sugar ring/anomer exchange reactions and the equilibrium of pseudorotation of pyranoses and furanose rings are reported (see **Figure 4.5**). Included differences are between: The total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies $\Delta G°$ at STP conditions. All results are obtained from DFT (B3LYP/6-311++G (*d*, *p*)) calculations. The polarizable continuum model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	RU[1]	2dRibf[4] vac.[2]	2dRibf[4] solv.[3]	Ribf[5] vac	Ribf[5] solv	Tho[6] vac	Tho[6] solv	2dRib[7] vac	2dRib[7] solv	Rib[8] vac	Rib[8] solv
						HPO_3^-					
ΔE_1	A	-31.3	20.0	-7.8	0.1	36.7	-0.7	2.5	23.4	28.5	28.3
$\Delta E_{1(ZPE)}$		-30.4	22.8	-7.0	0.0	37.4	-0.8	2.5	21.5	28.4	29.6
$\Delta G_1°$		-35.2	12.0	-12.5	0.6	39.4	-0.6	-0.7	28.2	29.8	24.2
ΔE_1	G	-1.5	0.0	42.0	0.0	-0.8	14.8	24.1	32.2	24.6	11.9
$\Delta E_{1(ZPE)}$		-1.7	0.0	44.2	0.0	-0.9	14.8	23.6	33.7	24.9	9.8
$\Delta G_1°$		-0.7	-0.0	36.0	0.0	0.0	18.7	25.5	25.8	25.6	19.5
ΔE_1	C	-1.0	-16.5	0.8	0.0	-4.2	-3.4	8.5	27.6	34.8	39.5
$\Delta E_{1(ZPE)}$		0.5	-20.4	1.2	0.0	-5.2	-3.3	8.8	27.7	35.1	39.9
$\Delta G_1°$		-4.9	-5.5	-2.2	0.1	-3.4	-3.3	6.6	25.6	34.2	35.2
ΔE_1	T	2.4	-12.1	-11.4	-8.6	9.4	-0.6	8.8	44.2	64.7	46.2
$\Delta E_{1(ZPE)}$		4.2	-15.3	-10.4	-10.0	8.6	-0.8	9.1	44.2	65.8	47.7
$\Delta G_1°$		-4.6	-0.8	-15.1	-4.2	13.2	-0.2	6.8	44.7	62.6	41.1
ΔE_1	U	5.7	-11.6	34.9	0.0	9.0	-3.2	7.5	46.2	64.6	44.7
$\Delta E_{1(ZPE)}$		6.3	-14.8	36.2	0.0	8.2	-3.3	7.8	46.5	65.7	44.9
$\Delta G_1°$		1.8	-1.2	35.2	-0.0	12.7	-2.5	5.6	44.5	62.4	44.7
ΔE_1	TAP-C5	9.6	-2.0	-15.8	-4.9	51.0	28.4	89.4	55.7	9.2	50.3
$\Delta E_{1(ZPE)}$		9.8	-0.9	-19.0	-4.4	51.0	29.2	90.0	54.7	9.1	51.3
$\Delta G_1°$		7.3	-3.0	-8.2	-6.2	53.7	26.1	86.0	55.4	9.1	47.7
ΔE_1	TAP-N	1.7	0.0	-17.5	-1.2	26.7	-3.2	-9.5	6.6	-7.0	29.9
$\Delta E_{1(ZPE)}$		1.6	0.0	-17.9	-2.6	26.0	-3.2	-10.5	5.6	-5.5	30.8
$\Delta G_1°$		1.6	0.0	-20.9	0.4	32.4	-2.7	-5.3	10.6	-9.7	27.8
ΔE_1	BA-C5	-20.1	-18.0	-21.7	-5.3	0.0	-3.7	72.3	38.8	85.0	29.8
$\Delta E_{1(ZPE)}$		-22.0	-20.7	-21.5	-7.4	0.0	-3.5	71.3	38.7	84.8	32.2
$\Delta G_1°$		-16.2	-12.2	-21.6	2.4	0.0	-4.6	73.9	37.3	85.9	21.7
ΔE_1	BA-N	-0.1	0.3	-40.6	0.0	-0.2	19.8	29.0	32.4	22.2	7.7
$\Delta E_{1(ZPE)}$		0.1	-0.3	-40.7	0.0	-0.5	19.8	29.0	32.5	22.8	7.9
$\Delta G_1°$		-2.1	4.2	-38.9	0.0	1.1	20.5	31.3	33.1	20.4	7.9
ΔE_1	CA	0.0	0.0	-14.4	0.0	-1.3	-3.4	16.3	38.4	17.6	24.4
$\Delta E_{1(ZPE)}$		0.0	0.0	-14.6	0.0	-1.5	-3.5	15.9	38.8	18.4	27.3
$\Delta G_1°$		0.0	0.0	-13.4	-0.0	-0.4	-2.5	19.2	38.3	14.7	15.6

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_1	MM	0.0	-2.2	34.8	-3.4	-6.1	8.4	29.8	46.1	37.0	38.9
$\Delta E_{1(ZPE)}$		0.0	-3.8	35.9	-2.0	-8.4	9.8	28.0	46.1	37.6	40.9
ΔG_1°		0.0	3.8	28.9	-6.9	3.9	4.8	34.8	47.1	35.8	33.6
ΔE_2	A	21.5	-3.5	-6.9	11.4	1.7	17.3	-6.9	-20.8	-2.8	-17.8
$\Delta E_{2(ZPE)}$		22.3	-4.9	-7.4	13.2	1.1	17.5	-6.1	-18.8	-1.9	-17.0
ΔG_2°		16.1	1.8	-5.9	6.5	0.2	15.1	-8.3	-25.2	-6.0	-20.3
ΔE_2	G	-29.6	-1.0	-43.6	-5.4	37.9	2.4	-16.6	-30.6	-34.7	5.0
$\Delta E_{2(ZPE)}$		-30.0	-0.6	-44.6	-5.9	38.5	2.7	-14.9	-30.7	-34.0	9.1
ΔG_2°		-24.1	1.7	-38.2	0.7	37.9	-4.1	-25.9	-30.5	-38.9	-8.8
ΔE_2	C	-43.9	16.3	-28.3	-4.7	27.8	19.0	1.2	-33.3	-27.1	-29.3
$\Delta E_{2(ZPE)}$		-44.4	17.4	-28.7	-3.8	28.0	18.7	2.1	-33.4	-26.1	-28.4
ΔG_2°		-43.1	16.1	-25.6	-3.2	26.8	17.4	-1.9	-31.5	-29.7	-29.3
ΔE_2	T	-24.8	5.0	-15.8	-1.4	33.8	15.1	-1.1	-36.7	-62.5	-32.1
$\Delta E_{2(ZPE)}$		-25.5	6.7	-17.0	0.7	34.1	17.1	-0.3	-36.8	-62.5	-29.5
ΔG_2°		-19.4	1.6	-12.2	-5.6	33.2	10.2	-3.7	-36.2	-63.2	-36.0
ΔE_2	U	-22.1	8.1	-10.6	-10.3	34.4	17.7	4.7	-38.8	-62.0	-27.0
$\Delta E_{2(ZPE)}$		-22.8	11.0	-11.7	-9.9	34.8	19.4	5.4	-39.2	-62.0	-23.7
ΔG_2°		-17.3	-1.1	-8.0	-8.3	33.6	11.1	0.8	-35.9	-62.6	-34.4
ΔE_2	TAP-C5	-1.0	-13.0	-29.5	-3.1	-30.9	5.9	-90.9	-48.9	2.2	-39.2
$\Delta E_{2(ZPE)}$		1.1	-14.8	-28.2	-2.7	-30.2	5.7	-90.9	-47.9	4.7	-36.7
ΔG_2°		-5.5	-6.8	-34.0	-4.1	-31.3	6.5	-90.1	-48.2	-2.9	-42.8
ΔE_2	TAP-N	35.9	-6.7	38.1	-4.3	-15.3	22.5	5.2	-0.2	30.4	-29.2
$\Delta E_{2(ZPE)}$		35.8	-9.6	37.5	-2.8	-15.4	23.5	6.1	0.4	31.4	-28.8
ΔG_2°		37.7	3.7	44.3	-6.9	-17.5	19.4	2.8	-1.5	27.2	-30.0
ΔE_2	BA-C5	11.6	24.5	4.7	9.1	-1.7	16.9	-30.9	-35.4	-34.9	-23.4
$\Delta E_{2(ZPE)}$		14.1	28.3	5.1	10.7	-2.1	16.3	-30.3	-34.8	-33.5	-22.2
ΔG_2°		1.2	14.5	2.9	5.3	-1.1	18.4	-32.2	-34.4	-39.4	-26.9
ΔE_2	BA-N	-68.1	7.3	-41.3	1.9	38.4	-2.0	-38.5	-29.9	-22.9	-11.6
$\Delta E_{2(ZPE)}$		-70.1	8.4	-42.6	4.4	39.0	-1.3	-38.4	-30.2	-22.0	-8.0
ΔG_2°		-57.9	4.7	-35.5	-3.4	35.2	-6.4	-39.5	-28.0	-24.2	-20.6
ΔE_2	CA	-25.9	16.9	2.0	-3.1	43.0	25.4	-22.7	-32.3	-22.0	-40.9
$\Delta E_{2(ZPE)}$		-28.1	19.3	2.3	-1.4	43.7	26.2	-22.5	-33.0	-20.7	-40.9
ΔG_2°		-16.7	11.3	0.4	-6.2	39.7	21.3	-24.6	-29.3	-22.3	-39.7
ΔE_2	MM	-18.6	-2.0	-32.9	6.5	55.4	-11.1	-26.2	-39.7	14.6	-19.6
$\Delta E_{2(ZPE)}$		-16.9	-4.3	-31.9	4.2	57.5	-9.0	-22.4	-40.0	17.2	-19.4
ΔG_2°		-22.0	7.2	-30.6	17.2	47.4	-18.5	-37.5	-38.3	9.4	-19.9
ΔE_3	A	-1.2	-7.8	0.0	6.4	28.6	-4.0	-16.8	-4.9	-0.9	-4.0
$\Delta E_{3(ZPE)}$		-0.9	-9.7	0.0	6.2	29.3	-2.4	-16.8	-5.0	0.5	-2.3
ΔG_3°		-6.3	-1.3	0.0	5.3	25.7	-10.1	-15.9	-2.3	-3.6	-10.9
ΔE_3	G	-2.4	-0.6	-52.1	4.5	32.5	-6.6	-12.1	-7.6	-26.0	0.7
$\Delta E_{3(ZPE)}$		-1.6	-0.6	-53.4	4.4	33.0	-6.3	-12.1	-7.3	-25.2	2.3
ΔG_3°		-3.3	-0.4	-44.9	5.1	33.9	-7.5	-9.9	-7.6	-29.2	-4.4
ΔE_3	C	-35.4	8.1	-10.3	7.2	42.6	1.3	-34.2	-11.7	-22.8	-11.9
$\Delta E_{3(ZPE)}$		-37.8	7.5	-10.5	7.5	42.5	0.4	-34.2	-11.6	-22.2	-10.8
ΔG_3°		-30.0	9.6	-9.5	7.1	44.2	2.6	-33.5	-11.6	-24.7	-16.3
ΔE_3	T	12.3	-2.1	-8.8	0.2	45.6	-5.7	-32.5	-9.8	-30.3	-13.3
$\Delta E_{3(ZPE)}$		12.4	-1.5	-8.6	1.1	45.7	-3.7	-32.7	-9.4	-30.1	-8.1
ΔG_3°		13.8	-4.9	-9.5	-4.4	47.0	-10.2	-30.8	-11.0	-31.0	-25.3
ΔE_3	U	22.2	17.6	-4.0	1.5	45.9	0.0	-26.7	-13.2	-21.1	-7.2
$\Delta E_{3(ZPE)}$		22.4	19.8	-4.5	1.8	46.1	0.0	-27.0	-12.9	-20.8	-4.1
ΔG_3°		22.8	11.0	-0.3	0.3	47.4	-0.0	-26.2	-13.6	-22.8	-14.8
ΔE_3	TAP-C5	13.3	0.0	-0.1	4.6	0.0	0.0	-24.5	-24.0	22.9	3.3
$\Delta E_{3(ZPE)}$		13.4	0.0	-0.1	3.8	0.0	0.0	-24.0	-23.6	24.8	6.9
ΔG_3°		15.5	0.0	-0.0	6.0	0.0	0.0	-24.8	-23.8	18.6	-4.5
ΔE_3	TAP-N	15.6	1.1	36.7	-0.3	0.0	19.6	-6.8	-1.8	14.1	-9.2
$\Delta E_{3(ZPE)}$		14.8	1.5	35.6	-0.6	0.0	20.8	-6.8	-1.9	15.4	-8.3
ΔG_3°		16.1	-0.9	50.4	0.2	0.0	16.4	-5.5	0.4	9.5	-11.7

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_3	BA-C^5	28.0	2.7	1.5	-0.4	0.0	-0.0	-28.7	-16.9	-32.0	-4.2
$\Delta E_{3(ZPE)}$		30.4	2.6	1.6	-2.9	0.0	0.0	-28.5	-16.3	-31.1	-1.8
ΔG_3°		18.3	3.0	2.3	6.6	0.0	-0.0	-27.0	-17.9	-35.6	-11.9
ΔE_3	BA-N	3.0	-9.0	7.0	3.2	-0.0	0.0	-14.1	-12.0	9.6	1.1
$\Delta E_{3(ZPE)}$		4.5	-10.9	7.3	2.9	0.0	0.0	-14.4	-11.8	10.9	4.3
ΔG_3°		-0.7	-3.4	5.2	4.4	-0.0	0.0	-12.3	-11.2	6.8	-6.9
ΔE_3	CA	-8.6	3.4	-1.3	1.6	-2.2	0.0	-9.5	-8.5	8.2	-19.8
$\Delta E_{3(ZPE)}$		-8.2	3.1	-0.5	1.4	-2.1	0.0	-10.0	-8.9	9.1	-18.6
ΔG_3°		-9.2	4.4	-8.7	1.8	-3.4	0.0	-7.6	-6.9	6.7	-22.3
ΔE_3	MM	28.3	5.9	-0.4	20.7	53.7	0.0	-9.0	-1.3	33.5	7.2
$\Delta E_{3(ZPE)}$		30.2	4.8	2.1	22.2	53.8	0.0	-8.9	-1.7	35.5	9.0
ΔG_3°		23.3	8.3	-5.1	16.9	54.3	0.0	-7.3	0.9	29.2	2.7
ΔE_4	A	-8.6	24.3	-14.7	5.1	9.8	20.6	12.4	7.5	26.6	14.5
$\Delta E_{4(ZPE)}$		-7.2	27.6	-14.4	7.0	9.2	19.1	13.2	7.7	26.0	14.9
ΔG_4°		-12.8	15.1	-18.4	1.7	13.9	24.5	6.9	5.3	27.4	14.8
ΔE_4	G	-28.7	-0.4	50.5	-9.9	4.6	23.8	19.6	9.2	15.9	16.3
$\Delta E_{4(ZPE)}$		-30.2	0.0	53.1	-10.3	4.5	23.9	20.8	10.3	16.1	16.6
ΔG_4°		-21.5	2.1	42.7	-4.4	4.1	22.1	9.5	2.8	16.0	15.2
ΔE_4	C	-9.4	-8.3	-17.2	-11.8	-18.9	14.3	43.9	5.9	30.5	22.1
$\Delta E_{4(ZPE)}$		-6.1	-10.5	-17.0	-11.3	-19.7	15.0	45.2	5.9	31.2	22.3
ΔG_4°		-18.0	1.0	-18.4	-10.3	-20.8	11.5	38.1	5.7	29.2	22.2
ΔE_4	T	-34.7	-5.0	-18.4	-10.2	-2.4	20.2	40.2	17.3	32.5	27.4
$\Delta E_{4(ZPE)}$		-33.7	-7.1	-18.8	-10.4	-2.9	20.0	41.5	16.8	33.4	26.3
ΔG_4°		-37.8	5.8	-17.8	-5.5	-0.7	20.2	33.9	19.5	30.4	30.5
ΔE_4	U	-38.6	-21.2	28.2	-11.8	-2.5	14.4	38.9	20.6	23.8	24.9
$\Delta E_{4(ZPE)}$		-39.0	-23.5	29.0	-11.7	-3.1	16.1	40.2	20.2	24.5	25.2
ΔG_4°		-38.3	-13.3	27.5	-8.7	-1.0	8.5	32.6	22.3	22.5	25.1
ΔE_4	TAP-C^5	-4.7	-15.0	-45.2	-12.7	20.0	34.3	23.0	30.9	-11.6	7.8
$\Delta E_{4(ZPE)}$		-2.4	-15.7	-47.1	-10.9	20.9	34.9	23.1	30.5	-11.0	7.7
ΔG_4°		-13.7	-9.7	-42.1	-16.3	22.4	32.5	20.6	31.0	-12.5	9.5
ΔE_4	TAP-N	22.0	-7.9	-16.2	-5.3	11.4	-0.2	2.6	8.2	9.4	9.9
$\Delta E_{4(ZPE)}$		22.7	-11.1	-16.0	-4.7	10.6	-0.5	2.5	7.9	10.5	10.2
ΔG_4°		23.2	4.6	-26.9	-6.8	14.8	0.3	3.1	8.7	8.0	9.4
ΔE_4	BA-C^5	-36.5	3.7	-18.6	4.2	-1.7	13.2	70.1	20.4	82.1	10.5
$\Delta E_{4(ZPE)}$		-38.3	5.0	-17.9	6.2	-2.1	12.7	69.5	20.2	82.5	11.8
ΔG_4°		-33.3	-0.7	-21.0	1.1	-1.1	13.8	68.7	20.8	82.1	6.8
ΔE_4	BA-N	-71.2	16.5	-88.9	-1.4	38.3	17.8	4.6	14.5	-10.3	-5.0
$\Delta E_{4(ZPE)}$		-74.5	19.1	-90.6	1.5	38.6	18.5	4.9	14.1	-10.1	-4.3
ΔG_4°		-59.3	12.2	-79.5	-7.8	36.3	14.1	4.1	16.3	-10.6	-5.8
ΔE_4	CA	-17.3	13.5	-11.1	-4.7	43.9	22.0	3.0	14.6	-12.6	3.4
$\Delta E_{4(ZPE)}$		-19.9	16.2	-11.8	-2.8	44.2	22.8	3.4	14.6	-11.5	5.0
ΔG_4°		-7.5	7.0	-4.2	-8.0	42.8	18.8	2.2	15.9	-14.3	-1.8
ΔE_4	MM	-46.9	-10.1	2.3	-17.6	-4.5	-2.7	12.6	7.7	18.2	12.0
$\Delta E_{4(ZPE)}$		-47.1	-12.9	2.0	-20.0	-4.6	0.8	14.6	7.8	19.3	12.4
ΔG_4°		-45.3	2.7	3.4	-6.7	-3.1	-13.7	4.6	7.9	16.0	11.0
ΔE_5	A	-9.8	16.6	-14.7	11.5	38.4	16.6	-4.5	2.6	25.7	10.5
$\Delta E_{5(ZPE)}$		-8.0	17.9	-14.4	13.2	38.5	16.7	-3.5	2.7	26.5	12.6
ΔG_5°		-19.1	13.8	-18.4	7.0	39.6	14.4	-9.0	3.0	23.8	3.9
ΔE_5	G	-31.2	-1.0	-1.6	-5.4	37.1	17.2	7.5	1.6	-10.2	17.0
$\Delta E_{5(ZPE)}$		-31.8	-0.6	-0.4	-5.9	37.5	17.5	8.7	3.0	-9.1	18.9
ΔG_5°		-24.8	1.6	-2.2	0.7	38.0	14.6	-0.4	-4.8	-13.2	10.8
ΔE_5	C	-44.9	-0.2	-27.5	-4.6	23.6	15.6	9.7	-5.7	7.7	10.2
$\Delta E_{5(ZPE)}$		-43.9	-3.0	-27.5	-3.8	22.8	15.4	10.9	-5.7	9.0	11.5
ΔG_5°		-48.0	10.5	-27.9	-3.2	23.4	14.1	4.7	-5.9	4.5	5.9
ΔE_5	T	-22.4	-7.1	-27.1	-10.0	43.2	14.5	7.7	7.5	2.2	14.1
$\Delta E_{5(ZPE)}$		-21.3	-8.6	-27.4	-9.3	42.7	16.3	8.8	7.4	3.3	18.2
ΔG_5°		-24.0	0.8	-27.3	-9.9	46.4	10.0	3.1	8.5	-0.6	5.1

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_5	U	-16.4	-3.5	24.2	-10.3	43.4	14.4	12.1	7.4	2.6	17.7
$\Delta E_{5(ZPE)}$		-16.6	-3.8	24.5	-9.9	43.0	16.1	13.2	7.3	3.8	21.2
ΔG_5°		-15.5	-2.2	27.2	-8.3	46.4	8.5	6.4	8.7	-0.3	10.3
ΔE_5	TAP-C^5	8.6	-15.0	-45.3	-8.1	20.0	34.3	-1.5	6.9	11.3	11.1
$\Delta E_{5(ZPE)}$		11.0	-15.7	-47.3	-7.1	20.9	34.9	-0.9	6.9	13.8	14.6
ΔG_5°		1.8	-9.7	-42.2	-10.3	22.4	32.5	-4.1	7.3	6.1	4.9
ΔE_5	TAP-N	37.6	-6.7	20.5	-5.5	11.4	19.3	-4.3	6.4	23.5	0.7
$\Delta E_{5(ZPE)}$		37.5	-9.6	19.5	-5.4	10.6	20.3	-4.4	6.0	26.0	1.9
ΔG_5°		39.3	3.7	23.4	-6.6	14.8	16.6	-2.4	9.1	17.5	-2.2
ΔE_5	BA-C^5	-8.5	6.4	-17.0	3.8	-1.7	13.2	41.4	3.5	50.1	6.3
$\Delta E_{5(ZPE)}$		-7.9	7.6	-16.4	3.3	-2.1	12.7	41.0	3.9	51.3	9.9
ΔG_5°		-15.0	2.3	-18.7	7.7	-1.1	13.8	41.7	3.0	46.5	-5.2
ΔE_5	BA-N	-68.2	7.5	-81.9	1.9	38.3	17.8	-9.5	2.6	-0.7	-3.9
$\Delta E_{5(ZPE)}$		-70.1	8.1	-83.4	4.4	38.6	18.6	-9.5	2.3	0.8	-0.0
ΔG_5°		-60.0	8.8	-74.4	-3.4	36.3	14.1	-8.2	5.1	-3.8	-12.7
ΔE_5	CA	-25.9	16.9	-12.4	-3.1	41.7	22.0	-6.4	6.1	-4.4	-16.4
$\Delta E_{5(ZPE)}$		-28.1	19.3	-12.3	-1.4	42.2	22.8	-6.6	5.7	-2.3	-13.6
ΔG_5°		-16.7	11.3	-12.9	-6.2	39.4	18.8	-5.4	9.0	-7.6	-24.1
ΔE_5	MM	-18.6	-4.2	2.0	3.1	49.2	-2.7	3.6	6.5	51.6	19.3
$\Delta E_{5(ZPE)}$		-16.9	-8.1	4.0	2.2	49.2	0.8	5.6	6.1	54.8	21.4
ΔG_5°		-22.0	11.0	-1.8	10.3	51.3	-13.7	-2.7	8.8	45.2	13.6
ΔE_6	A	-22.7	-4.3	6.9	-5.0	26.8	-21.3	-9.9	15.9	1.9	13.8
$\Delta E_{6(ZPE)}$		-23.2	-4.8	7.4	-7.0	28.2	-19.9	-10.7	13.8	2.4	14.7
ΔG_6°		-22.4	-3.1	5.9	-1.1	25.5	-25.1	-7.5	22.9	2.4	9.4
ΔE_6	G	27.2	0.4	-8.5	9.9	-5.4	-9.0	4.5	23.0	8.7	-4.4
$\Delta E_{6(ZPE)}$		28.5	-0.0	-8.8	10.3	-5.5	-9.0	2.8	23.4	8.8	-6.8
ΔG_6°		20.7	-2.1	-6.7	4.4	-4.1	-3.4	16.0	23.0	9.7	4.3
ΔE_6	C	8.5	-8.2	18.0	11.8	14.8	-17.7	-35.4	21.7	4.3	17.4
$\Delta E_{6(ZPE)}$		6.6	-9.9	18.2	11.3	14.4	-18.3	-36.4	21.8	3.9	17.6
ΔG_6°		13.1	-6.5	16.2	10.3	17.4	-14.8	-31.6	19.9	5.0	13.0
ΔE_6	T	37.1	-7.1	7.0	1.6	11.8	-20.8	-31.5	27.0	32.2	18.8
$\Delta E_{6(ZPE)}$		37.9	-8.3	8.4	0.4	11.5	-20.8	-32.4	27.4	32.4	21.4
ΔG_6°		33.2	-6.6	2.7	1.3	13.8	-20.4	-27.0	25.2	32.2	10.6
ΔE_6	U	44.2	9.6	6.6	11.8	11.5	-17.7	-31.4	25.6	40.9	19.7
$\Delta E_{6(ZPE)}$		45.2	8.7	7.2	11.7	11.3	-19.4	-32.4	26.4	41.2	19.7
ΔG_6°		40.1	12.1	7.8	8.7	13.7	-11.1	-27.0	22.2	39.9	19.6
ΔE_6	TAP-C^5	14.3	13.0	29.4	7.7	30.9	-5.9	66.4	24.9	20.8	42.6
$\Delta E_{6(ZPE)}$		12.2	14.8	28.1	6.5	30.2	-5.7	66.9	24.3	20.0	43.6
ΔG_6°		21.0	6.8	34.0	10.1	31.3	-6.5	65.3	24.4	21.5	38.3
ΔE_6	TAP-N	-20.3	7.9	-1.4	4.1	15.3	-2.9	-12.0	-1.6	-16.3	19.9
$\Delta E_{6(ZPE)}$		-21.1	11.1	-1.9	2.2	15.4	-2.7	-12.9	-2.4	-16.0	20.6
ΔG_6°		-21.6	-4.6	6.1	7.1	17.5	-3.0	-8.3	1.9	-17.7	18.4
ΔE_6	BA-C^5	16.3	-21.7	-3.1	-9.5	1.7	-16.9	2.2	18.5	2.9	19.2
$\Delta E_{6(ZPE)}$		16.3	-25.7	-3.5	-13.6	2.1	-16.3	1.8	18.5	2.4	20.4
ΔG_6°		17.1	-11.5	-0.6	1.3	1.1	-18.5	5.2	16.5	3.8	15.0
ΔE_6	BA-N	71.1	-16.2	48.3	1.4	-38.4	2.0	24.4	17.9	32.5	12.7
$\Delta E_{6(ZPE)}$		74.6	-19.4	49.9	-1.5	-39.0	1.3	24.1	18.4	32.8	12.2
ΔG_6°		57.2	-8.0	40.6	7.8	-35.2	6.4	27.2	16.8	31.0	13.7
ΔE_6	CA	17.3	-13.5	-3.3	4.7	-45.2	-25.4	13.3	23.8	30.2	21.1
$\Delta E_{6(ZPE)}$		19.9	-16.2	-2.8	2.8	-45.7	-26.2	12.5	24.1	29.8	22.2
ΔG_6°		7.5	-6.9	-9.2	8.0	-43.2	-21.3	17.0	22.4	29.0	17.4
ΔE_6	MM	46.9	8.0	32.5	14.2	-1.6	11.1	17.2	38.4	18.8	26.9
$\Delta E_{6(ZPE)}$		47.1	9.0	34.0	18.0	-3.8	9.0	13.5	38.3	18.3	28.5
ΔG_6°		45.3	1.0	25.5	-0.3	7.0	18.5	30.2	39.2	19.8	22.6

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
						HAsO$_3^-$					
ΔE_1	A	-21.5	7.4	-13.8	2.9	28.7	-3.7	-21.5	7.4	25.3	25.2
$\Delta E_{1(ZPE)}$		-20.2	7.2	-13.5	2.7	29.5	-3.8	-20.2	7.2	25.0	23.9
ΔG_1°		-27.2	7.7	-13.1	5.3	28.4	-3.7	-27.2	7.7	27.4	29.4
ΔE_1	G	0.0	5.6	58.7	0.0	-3.1	14.0	0.0	5.6	26.3	7.4
$\Delta E_{1(ZPE)}$		0.0	7.0	61.8	0.0	-3.5	12.9	0.0	7.0	27.6	5.9
ΔG_1°		0.0	1.8	46.3	0.0	-1.3	17.9	0.0	1.8	25.7	15.3
ΔE_1	C	0.0	-9.7	2.4	0.0	-0.6	0.0	0.0	-9.7	33.8	33.5
$\Delta E_{1(ZPE)}$		0.0	-11.8	4.8	0.0	-1.3	-0.0	0.0	-11.8	34.3	32.4
ΔG_1°		0.0	-4.2	-5.8	0.0	1.9	0.0	0.0	-4.2	33.8	35.7
ΔE_1	T	0.0	-5.8	40.4	2.2	-2.4	0.0	0.0	-5.8	58.5	41.8
$\Delta E_{1(ZPE)}$		0.0	-7.8	41.5	2.9	-3.1	0.0	0.0	-7.8	59.2	41.7
ΔG_1°		0.0	2.7	39.3	0.4	-0.6	-0.0	0.0	2.7	57.0	42.2
ΔE_1	U	0.0	12.3	35.2	7.4	6.2	0.0	0.0	12.3	58.5	41.1
$\Delta E_{1(ZPE)}$		0.0	12.2	36.3	7.8	5.1	0.0	0.0	12.2	59.1	41.6
ΔG_1°		0.0	13.1	34.1	6.8	11.0	-0.0	0.0	13.1	57.1	41.2
ΔE_1	TAP-C^5	3.2	-6.2	-17.4	-19.0	53.6	32.5	3.2	-6.2	13.7	34.9
$\Delta E_{1(ZPE)}$		3.8	-5.2	-17.0	-21.8	53.4	31.0	3.8	-5.2	14.8	35.8
ΔG_1°		2.3	-8.0	-21.5	-11.6	58.0	38.7	2.3	-8.0	10.1	33.0
ΔE_1	TAP-N	-5.0	1.3	-30.3	8.8	30.9	-3.6	-5.0	1.3	-4.1	23.4
$\Delta E_{1(ZPE)}$		-5.9	1.2	-29.7	10.8	30.7	-3.8	-5.9	1.2	-2.4	26.1
ΔG_1°		-1.5	0.3	-36.9	-3.2	36.4	-2.8	-1.5	0.3	-7.2	17.9
ΔE_1	BA-C^5	25.5	1.5	0.7	10.5	0.0	-0.0	25.5	1.5	88.7	17.0
$\Delta E_{1(ZPE)}$		24.1	-0.1	0.2	12.2	0.0	0.0	24.1	-0.1	88.8	16.0
ΔG_1°		24.3	4.5	3.5	5.3	0.0	-0.1	24.3	4.5	89.0	18.1
ΔE_1	BA-N	0.0	-0.3	-121.8	6.4	-2.5	18.7	0.0	-0.3	19.7	10.6
$\Delta E_{1(ZPE)}$		0.0	-0.3	-121.0	7.6	-3.0	18.7	0.0	-0.3	20.1	10.5
ΔG_1°		0.0	-0.2	-127.7	3.3	-0.6	19.2	0.0	-0.2	18.4	11.4
ΔE_1	CA	0.0	0.0	-20.7	8.4	-3.6	0.0	0.0	0.0	15.2	30.2
$\Delta E_{1(ZPE)}$		0.0	0.1	-24.4	9.4	-4.1	0.0	0.0	0.0	15.8	28.8
ΔG_1°		0.0	1.5	-4.4	5.7	-1.9	0.0	0.0	0.0	13.2	34.5
ΔE_1	MM	0.0	-7.7	39.2	-10.8	-7.4	-8.8	0.0	-7.7	38.1	37.0
$\Delta E_{1(ZPE)}$		0.0	-8.4	38.9	-6.8	-9.6	-7.3	0.0	-8.4	38.8	37.4
ΔG_1°		-0.0	-9.4	36.8	-23.4	3.6	-15.4	-0.0	-9.4	36.8	35.3
ΔE_2	A	19.2	17.0	-4.2	0.8	0.4	15.4	19.2	17.0	3.6	-15.1
$\Delta E_{2(ZPE)}$		19.7	17.6	-5.1	3.0	0.1	14.8	19.7	17.6	4.1	-14.2
ΔG_2°		15.5	16.8	0.1	-6.2	-1.3	17.2	15.5	16.8	1.2	-19.8
ΔE_2	G	-30.0	-9.8	-44.4	-9.2	37.3	-3.8	-30.0	-9.8	-26.1	6.6
$\Delta E_{2(ZPE)}$		-29.8	-12.5	-45.0	-10.1	38.1	-0.7	-29.8	-12.5	-25.9	8.7
ΔG_2°		-26.1	0.6	-36.5	-1.3	38.0	-14.9	-26.1	0.6	-31.0	-0.9
ΔE_2	C	-17.3	17.5	-34.2	3.1	31.1	13.5	-17.3	17.5	-17.5	-23.4
$\Delta E_{2(ZPE)}$		-16.1	18.0	-34.9	5.4	31.7	13.3	-16.1	18.0	-17.3	-22.1
ΔG_2°		-20.4	17.6	-30.8	-1.0	29.2	12.5	-20.4	17.6	-20.6	-25.3
ΔE_2	T	-23.5	7.5	-25.0	-2.0	37.6	12.4	-23.5	7.5	-48.2	-25.2
$\Delta E_{2(ZPE)}$		-22.7	6.9	-26.3	-1.5	38.3	14.2	-22.7	6.9	-48.2	-22.8
ΔG_2°		-25.1	10.8	-19.1	-1.1	35.9	7.9	-25.1	10.8	-50.1	-28.7
ΔE_2	U	-28.4	-9.5	-21.7	-8.9	38.3	12.4	-28.4	-9.5	-47.9	-24.8
$\Delta E_{2(ZPE)}$		-27.3	-9.8	-23.0	-7.8	39.0	14.1	-27.3	-9.8	-47.8	-23.2
ΔG_2°		-33.2	-9.1	-15.4	-10.2	36.9	8.2	-33.2	-9.1	-50.2	-27.7
ΔE_2	TAP-C^5	-6.5	-1.3	-26.5	5.0	-31.8	-8.3	-6.5	-1.3	-3.2	-32.1
$\Delta E_{2(ZPE)}$		-5.0	-2.1	-26.4	9.0	-30.8	-5.6	-5.0	-2.1	-0.6	-29.8
ΔG_2°		-12.6	0.4	-28.5	-7.8	-32.5	-11.0	-12.6	0.4	-5.9	-34.9
ΔE_2	TAP-N	23.4	-8.7	43.9	-6.2	-21.5	34.4	23.4	-8.7	27.7	-27.2
$\Delta E_{2(ZPE)}$		24.6	-10.4	43.7	-7.9	-22.2	35.4	24.6	-10.4	28.8	-26.7
ΔG_2°		19.9	-4.5	49.4	2.0	-22.4	33.5	19.9	-4.5	25.0	-27.9

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_2	BA-C^5	-42.7	2.4	-12.2	1.3	-3.8	20.1	-42.7	2.4	-34.9	-18.5
$\Delta E_{2(ZPE)}$		-42.7	4.2	-13.0	1.8	-4.1	17.1	-42.7	4.2	-33.5	-17.5
ΔG_2°		-39.9	0.4	-12.0	2.1	-2.9	28.9	-39.9	0.4	-37.6	-19.1
ΔE_2	BA-N	-54.8	9.0	35.7	-1.6	41.2	3.1	-54.8	9.0	-23.0	-16.9
$\Delta E_{2(ZPE)}$		-57.1	9.5	35.9	-1.1	41.6	2.1	-57.1	9.5	-21.9	-14.2
ΔG_2°		-45.3	9.5	40.0	0.1	39.9	3.9	-45.3	9.5	-23.9	-22.5
ΔE_2	CA	-11.0	12.4	-18.2	-11.3	44.6	27.5	-11.0	12.4	-17.3	-27.9
$\Delta E_{2(ZPE)}$		-11.7	13.3	-17.8	-11.1	45.1	26.4	-11.7	13.3	-15.8	-27.2
ΔG_2°		-8.4	10.5	-19.0	-9.6	43.6	29.5	-8.4	10.5	-18.2	-31.0
ΔE_2	MM	-9.0	-3.0	-25.0	17.8	53.3	12.7	-9.0	-3.0	-36.7	-20.5
$\Delta E_{2(ZPE)}$		-8.9	-5.5	-22.5	17.2	55.2	13.0	-8.9	-5.5	-35.5	-20.6
ΔG_2°		-7.6	7.9	-27.1	24.5	45.7	13.8	-7.6	7.9	-40.9	-19.2
ΔE_3	A	17.8	6.9	2.8	5.0	32.2	-7.5	17.8	6.9	2.2	-10.3
$\Delta E_{3(ZPE)}$		18.3	5.8	2.4	4.6	33.5	-6.7	18.3	5.8	2.1	-9.3
ΔG_3°		13.5	12.8	6.7	8.1	27.4	-9.0	13.5	12.8	1.9	-13.4
ΔE_3	G	-3.6	-7.6	-46.5	-13.4	29.5	2.2	-3.6	-7.6	-19.1	-2.5
$\Delta E_{3(ZPE)}$		-2.3	-9.5	-46.8	-16.2	29.9	3.3	-2.3	-9.5	-19.0	-2.4
ΔG_3°		-7.5	-0.8	-38.8	-6.8	33.6	-0.9	-7.5	-0.8	-22.5	-3.0
ΔE_3	C	-14.4	17.7	-9.4	9.5	42.8	11.2	-14.4	17.7	-25.8	-12.9
$\Delta E_{3(ZPE)}$		-16.3	17.7	-8.4	11.6	43.0	12.3	-16.3	17.7	-25.8	-13.8
ΔG_3°		-9.2	17.7	-14.0	2.9	44.6	7.9	-9.2	17.7	-27.7	-11.7
ΔE_3	T	5.5	3.9	-11.5	3.0	46.4	-5.9	5.5	3.9	-8.8	-11.8
$\Delta E_{3(ZPE)}$		5.2	2.0	-10.9	2.6	46.6	-3.6	5.2	2.0	-8.1	-8.7
ΔG_3°		7.6	11.8	-12.7	6.7	49.4	-11.7	7.6	11.8	-12.1	-17.2
ΔE_3	U	6.8	-9.3	-10.1	3.3	47.0	0.0	6.8	-9.3	-8.7	-13.5
$\Delta E_{3(ZPE)}$		6.9	-9.1	-10.9	2.8	47.3	0.0	6.9	-9.1	-7.9	-10.5
ΔG_3°		6.4	-13.0	-4.8	8.2	48.8	0.0	6.4	-13.0	-12.0	-19.6
ΔE_3	TAP-C^5	-2.0	-4.2	0.4	5.4	0.0	-9.0	-2.0	-4.2	17.2	3.4
$\Delta E_{3(ZPE)}$		-2.2	-6.3	0.1	5.2	0.0	-9.5	-2.2	-6.3	18.9	4.5
ΔG_3°		0.9	2.7	1.3	4.3	0.0	-5.9	0.9	2.7	14.4	2.6
ΔE_3	TAP-N	-5.3	6.3	1.1	-4.7	0.0	31.3	-5.3	6.3	10.0	-10.1
$\Delta E_{3(ZPE)}$		-5.2	6.1	-0.1	-7.9	0.0	32.4	-5.2	6.1	11.2	-8.9
ΔG_3°		-6.9	6.8	4.5	7.0	0.0	29.0	-6.9	6.8	6.5	-13.1
ΔE_3	BA-C^5	-3.9	-1.5	-12.2	8.2	0.0	0.0	-3.9	-1.5	-36.3	-5.9
$\Delta E_{3(ZPE)}$		-4.2	-1.4	-13.6	8.1	0.0	0.0	-4.2	-1.4	-35.4	-5.4
ΔG_3°		-3.3	0.3	-9.7	8.6	0.0	0.0	-3.3	0.3	-37.3	-5.7
ΔE_3	BA-N	16.5	6.4	4.6	6.2	0.0	0.0	16.5	6.4	6.4	-1.8
$\Delta E_{3(ZPE)}$		17.5	7.0	4.8	5.8	0.0	0.0	17.5	7.0	7.8	-0.6
ΔG_3°		13.5	5.2	2.8	8.2	0.0	0.0	13.5	5.2	4.3	-4.0
ΔE_3	CA	9.5	1.3	-4.8	-4.6	0.0	0.0	9.5	1.3	5.6	-7.9
$\Delta E_{3(ZPE)}$		9.3	1.2	-4.7	-5.2	0.0	0.0	9.3	1.2	6.1	-8.2
ΔG_3°		11.1	1.2	-5.2	-2.9	0.0	0.0	11.1	1.2	5.3	-8.5
ΔE_3	MM	-1.8	0.8	5.8	17.3	51.1	0.0	-1.8	0.8	-16.3	-0.6
$\Delta E_{3(ZPE)}$		-2.9	-1.5	8.2	17.7	51.6	0.0	-2.9	-1.5	-14.7	0.6
ΔG_3°		1.7	12.2	0.9	18.3	51.2	0.0	1.7	12.2	-21.2	-2.0
ΔE_4	A	-20.1	17.5	-20.8	-1.3	-3.1	19.3	-20.1	17.5	26.8	20.3
$\Delta E_{4(ZPE)}$		-18.8	19.1	-21.0	1.1	-3.8	17.6	-18.8	19.1	27.1	19.1
ΔG_4°		-25.2	11.7	-19.7	-9.0	-0.4	22.6	-25.2	11.7	26.6	22.9
ΔE_4	G	-26.3	3.4	60.9	4.2	4.7	8.1	-26.3	3.4	19.3	16.5
$\Delta E_{4(ZPE)}$		-27.5	4.0	63.5	6.0	4.6	8.9	-27.5	4.0	20.7	17.1
ΔG_4°		-18.6	3.2	48.5	5.5	3.1	3.9	-18.6	3.2	17.3	17.5
ΔE_4	C	-2.9	-9.8	-22.4	-6.4	-12.3	2.3	-2.9	-9.8	42.1	23.0
$\Delta E_{4(ZPE)}$		0.2	-11.5	-21.7	-6.2	-12.6	1.0	0.2	-11.5	42.8	24.1
ΔG_4°		-11.2	-4.3	-22.7	-3.9	-13.4	4.6	-11.2	-4.3	41.0	22.1
ΔE_4	T	-28.9	-2.2	26.9	-2.8	-11.3	18.3	-28.9	-2.2	19.2	28.4
$\Delta E_{4(ZPE)}$		-27.9	-2.9	26.1	-1.1	-11.4	17.8	-27.9	-2.9	19.1	27.5
ΔG_4°		-32.7	1.6	32.9	-7.3	-14.1	19.5	-32.7	1.6	19.0	30.7

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_4	U	-35.2	12.2	23.6	-4.8	-2.4	12.4	-35.2	12.2	19.2	29.8
$\Delta E_{4(ZPE)}$		-34.2	11.6	24.2	-2.8	-3.1	14.1	-34.2	11.6	19.2	28.9
ΔG_4°		-39.5	17.0	23.6	-11.6	-1.0	8.1	-39.5	17.0	18.8	33.1
ΔE_4	TAP-C^5	-1.2	-3.3	-44.3	-19.4	21.8	33.2	-1.2	-3.3	-6.6	-0.6
$\Delta E_{4(ZPE)}$		1.1	-1.1	-43.5	-18.0	22.6	34.9	1.1	-1.1	-4.6	1.4
ΔG_4°		-11.2	-10.4	-51.3	-23.6	25.5	33.6	-11.2	-10.4	-10.2	-4.5
ΔE_4	TAP-N	23.7	-13.7	12.5	7.3	9.4	-0.5	23.7	-13.7	13.6	6.3
$\Delta E_{4(ZPE)}$		23.8	-15.3	14.1	10.8	8.5	-0.9	23.8	-15.3	15.1	8.3
ΔG_4°		25.3	-11.1	8.0	-8.2	14.0	1.7	25.3	-11.1	11.3	3.0
ΔE_4	BA-C^5	-13.2	5.4	0.7	3.6	-3.8	20.1	-13.2	5.4	90.1	4.4
$\Delta E_{4(ZPE)}$		-14.4	5.5	0.8	5.9	-4.1	17.1	-14.4	5.5	90.7	3.9
ΔG_4°		-12.3	4.5	1.2	-1.2	-2.9	28.8	-12.3	4.5	88.7	4.7
ΔE_4	BA-N	-71.3	2.3	-90.8	-1.4	38.7	21.8	-71.3	2.3	-9.7	-4.5
$\Delta E_{4(ZPE)}$		-74.6	2.2	-89.9	0.8	38.6	20.7	-74.6	2.2	-9.5	-3.0
ΔG_4°		-58.8	4.1	-90.5	-4.8	39.3	23.1	-58.8	4.1	-9.8	-7.1
ΔE_4	CA	-20.4	11.1	-34.0	1.7	41.0	27.5	-20.4	11.1	-7.7	10.2
$\Delta E_{4(ZPE)}$		-21.0	11.9	-37.5	3.5	41.0	26.4	-21.0	12.1	-6.1	9.9
ΔG_4°		-19.5	10.8	-18.2	-1.0	41.7	29.5	-19.5	9.3	-10.3	12.1
ΔE_4	MM	-7.2	-11.5	8.4	-10.3	-5.2	3.8	-7.2	-11.5	17.8	17.1
$\Delta E_{4(ZPE)}$		-6.0	-12.5	8.2	-7.2	-6.0	5.7	-6.0	-12.5	18.0	16.2
ΔG_4°		-9.2	-13.7	8.7	-17.3	-1.8	-1.6	-9.2	-13.7	17.1	18.1
ΔE_5	A	-2.3	24.4	-18.0	3.7	29.1	11.8	-2.3	24.4	29.0	10.1
$\Delta E_{5(ZPE)}$		-0.5	24.8	-18.6	5.6	29.7	10.9	-0.5	24.8	29.2	9.8
ΔG_5°		-11.7	24.4	-13.0	-0.8	27.0	13.6	-11.7	24.4	28.5	9.5
ΔE_5	G	-30.0	-4.2	14.3	-9.2	34.2	10.2	-30.0	-4.2	0.2	14.0
$\Delta E_{5(ZPE)}$		-29.8	-5.5	16.7	-10.1	34.5	12.2	-29.8	-5.5	1.7	14.6
ΔG_5°		-26.1	2.4	9.8	-1.3	36.7	3.0	-26.1	2.4	-5.2	14.5
ΔE_5	C	-17.3	7.8	-31.8	3.1	30.5	13.5	-17.3	7.8	16.3	10.1
$\Delta E_{5(ZPE)}$		-16.1	6.2	-30.1	5.4	30.4	13.3	-16.1	6.2	17.0	10.3
ΔG_5°		-20.4	13.4	-36.7	-1.0	31.2	12.5	-20.4	13.4	13.3	10.4
ΔE_5	T	-23.5	1.6	15.4	0.2	35.2	12.4	-23.5	1.6	10.4	16.6
$\Delta E_{5(ZPE)}$		-22.7	-0.9	15.2	1.4	35.2	14.2	-22.7	-0.9	11.0	18.8
ΔG_5°		-25.1	13.4	20.2	-0.7	35.3	7.9	-25.1	13.4	6.9	13.5
ΔE_5	U	-28.4	2.8	13.5	-1.5	44.5	12.4	-28.4	2.8	10.6	16.4
$\Delta E_{5(ZPE)}$		-27.3	2.5	13.2	-0.0	44.2	14.1	-27.3	2.5	11.3	18.4
ΔG_5°		-33.2	4.0	18.8	-3.4	47.8	8.1	-33.2	4.0	6.9	13.5
ΔE_5	TAP-C^5	-3.2	-7.5	-43.9	-14.0	21.8	24.2	-3.2	-7.5	10.6	2.8
$\Delta E_{5(ZPE)}$		-1.2	-7.4	-43.4	-12.8	22.6	25.4	-1.2	-7.4	14.3	6.0
ΔG_5°		-10.3	-7.7	-50.0	-19.3	25.5	27.7	-10.3	-7.7	4.2	-2.0
ΔE_5	TAP-N	18.4	-7.3	13.7	2.6	9.4	30.9	18.4	-7.3	23.6	-3.7
$\Delta E_{5(ZPE)}$		18.7	-9.2	14.0	2.9	8.5	31.6	18.7	-9.2	26.3	-0.6
ΔG_5°		18.4	-4.2	12.5	-1.2	14.0	30.7	18.4	-4.2	17.8	-10.0
ΔE_5	BA-C^5	-17.2	3.9	-11.6	11.9	-3.8	20.1	-17.2	3.9	53.8	-1.5
$\Delta E_{5(ZPE)}$		-18.6	4.1	-12.8	14.0	-4.1	17.1	-18.6	4.1	55.3	-1.6
ΔG_5°		-15.6	4.8	-8.5	7.4	-2.9	28.8	-15.6	4.8	51.3	-1.0
ΔE_5	BA-N	-54.8	8.7	-86.2	4.8	38.7	21.8	-54.8	8.7	-3.3	-6.3
$\Delta E_{5(ZPE)}$		-57.1	9.2	-85.1	6.5	38.6	20.7	-57.1	9.2	-1.7	-3.7
ΔG_5°		-45.3	9.3	-87.7	3.4	39.3	23.1	-45.3	9.3	-5.5	-11.1
ΔE_5	CA	-11.0	12.4	-38.8	-2.9	41.0	27.5	-11.0	12.4	-2.1	2.3
$\Delta E_{5(ZPE)}$		-11.7	13.1	-42.2	-1.7	41.0	26.4	-11.7	13.3	-0.0	1.7
ΔG_5°		-8.4	12.0	-23.4	-3.9	41.7	29.5	-8.4	10.5	-5.0	3.6
ΔE_5	MM	-9.0	-10.8	14.2	7.0	45.9	3.8	-9.0	-10.8	1.5	16.5
$\Delta E_{5(ZPE)}$		-8.9	-13.9	16.4	10.5	45.6	5.7	-8.9	-13.9	3.3	16.8
ΔG_5°		-7.6	-1.5	9.7	1.0	49.4	-1.6	-7.6	-1.5	-4.1	16.1
ΔE_6	A	-1.5	-10.1	7.1	4.2	31.8	-23.0	-1.5	-10.1	-1.4	4.8
$\Delta E_{6(ZPE)}$		-1.4	-11.8	7.4	1.6	33.3	-21.5	-1.4	-11.8	-2.0	4.8
ΔG_6°		-2.0	-4.0	6.5	14.3	28.8	-26.2	-2.0	-4.0	0.8	6.4

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_6	G	26.3	2.2	-2.2	-4.2	-7.8	5.9	26.3	2.2	7.0	-9.1
$\Delta E_{6(ZPE)}$		27.5	3.0	-1.8	-6.0	-8.2	4.0	27.5	3.0	6.9	-11.2
ΔG_6°		18.6	-1.4	-2.2	-5.5	-4.4	14.0	18.6	-1.4	8.5	-2.1
ΔE_6	C	2.9	0.1	24.8	6.4	11.7	-2.3	2.9	0.1	-8.3	10.4
$\Delta E_{6(ZPE)}$		-0.2	-0.3	26.5	6.2	11.2	-1.0	-0.2	-0.3	-8.5	8.3
ΔG_6°		11.2	0.1	16.9	3.9	15.3	-4.6	11.2	0.1	-7.2	13.6
ΔE_6	T	28.9	-3.6	13.5	5.0	8.8	-18.3	28.9	-3.6	39.3	13.4
$\Delta E_{6(ZPE)}$		27.9	-5.0	15.4	4.0	8.3	-17.8	27.9	-5.0	40.1	14.1
ΔG_6°		32.7	1.1	6.4	7.8	13.6	-19.6	32.7	1.1	38.0	11.5
ΔE_6	U	35.2	0.2	11.7	12.2	8.6	-12.4	35.2	0.2	39.2	11.3
$\Delta E_{6(ZPE)}$		34.2	0.7	12.1	10.6	8.3	-14.1	34.2	0.7	40.0	12.7
ΔG_6°		39.5	-3.9	10.6	18.4	11.9	-8.1	39.5	-3.9	38.2	8.1
ΔE_6	TAP-C[5]	4.4	-2.9	26.9	0.4	31.8	-0.7	4.4	-2.9	20.4	35.5
$\Delta E_{6(ZPE)}$		2.8	-4.2	26.5	-3.8	30.8	-4.0	2.8	-4.2	19.4	34.4
ΔG_6°		13.6	2.4	29.8	12.1	32.5	5.1	13.6	2.4	20.3	37.5
ΔE_6	TAP-N	-28.7	15.0	-42.8	1.5	21.5	-3.1	-28.7	15.0	-17.7	17.1
$\Delta E_{6(ZPE)}$		-29.8	16.5	-43.8	0.0	22.2	-2.9	-29.8	16.5	-17.6	17.8
ΔG_6°		-26.8	11.4	-44.9	4.9	22.4	-4.5	-26.8	11.4	-18.6	14.9
ΔE_6	BA-C[5]	38.8	-3.9	0.0	6.9	3.8	-20.1	38.8	-3.9	-1.4	12.6
$\Delta E_{6(ZPE)}$		38.5	-5.6	-0.6	6.3	4.1	-17.1	38.5	-5.6	-1.9	12.1
ΔG_6°		36.6	-0.0	2.3	6.5	2.9	-28.9	36.6	-0.0	0.3	13.4
ΔE_6	BA-N	71.3	-2.6	-31.1	7.8	-41.2	-3.1	71.3	-2.6	29.4	15.2
$\Delta E_{6(ZPE)}$		74.6	-2.5	-31.1	6.9	-41.6	-2.1	74.6	-2.5	29.6	13.6
ΔG_6°		58.8	-4.3	-37.2	8.1	-39.9	-3.9	58.8	-4.3	28.2	18.5
ΔE_6	CA	20.4	-11.1	13.4	6.7	-44.6	-27.5	20.4	-11.1	22.9	20.0
$\Delta E_{6(ZPE)}$		21.0	-12.1	13.1	5.9	-45.1	-26.4	21.0	-12.1	21.9	18.9
ΔG_6°		19.5	-9.3	13.8	6.6	-43.6	-29.5	19.5	-9.3	23.5	22.5
ΔE_6	MM	7.2	3.8	30.8	-0.5	-2.1	-12.7	7.2	3.8	20.3	19.9
$\Delta E_{6(ZPE)}$		6.0	4.0	30.7	0.4	-3.7	-13.0	6.0	4.0	20.8	21.2
ΔG_6°		9.2	4.3	28.0	-6.2	5.4	-13.8	9.2	4.3	19.7	17.2

[1]TC = unspecified **T**rifunctional **C**onnector. [2] Differences in ΔE, $\Delta E_{(ZPE)}$, and ΔG° in vacuum at the DFT level. [3] Differences in ΔE, $\Delta E(ZPE)$, and ΔG° in solvent. [4]2dRibf: D-2'-deoxyribofuranose. [5]Ribf: D-ribofuranose. [6]Tho: D-threose. [7]2dRib: D-2'-deoxyribopyranose. [8]Rib: D-ribopyranose.

Table A2. Differences between the energies in kJ/mol for the most stable β- and α-anomers of the canonical and non-canonical nucleotides obtained using the alternative (c+d) set of reactions in vacuum and in aqueous environment (equations 4.2 - 4.7). The energies for all sugar ring/anomer exchange reactions and the equilibrium of pseudorotation of pyranoses and furanose rings are reported(see **Figure 4.5**). Included differences are between: The total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies $\Delta G°$ at STP conditions. All results are obtained from DFT (B3LYP/6-311++G (d, p)) calculations. The polarizable continuum model (IEFPCM) solvation model has been used to generate the results incorporating aqueous solvation at the same level of DFT theory.

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
						HPO$_3^-$					
ΔE_1	A	-25.2	-4.6	3.0	4.8	0.0	20.0	1.1	17.4	24.8	13.6
$\Delta E_{1(ZPE)}$		-26.2	-5.3	3.6	6.0	0.0	21.5	0.3	17.9	26.9	16.0
$\Delta G_1°$		-24.0	-0.8	0.3	2.8	0.0	17.2	5.8	18.1	19.0	6.9
ΔE_1	G	0.7	-8.0	-0.7	2.9	0.0	31.8	30.8	22.0	14.7	15.6
$\Delta E_{1(ZPE)}$		0.7	-8.3	-0.1	4.2	0.0	30.1	31.2	21.8	16.7	19.2
$\Delta G_1°$		-0.9	-7.8	-3.3	-2.1	0.0	37.8	28.2	27.7	4.8	3.0
ΔE_1	C	12.9	-10.6	-0.9	0.0	0.0	27.5	-2.7	24.0	14.9	7.7
$\Delta E_{1(ZPE)}$		13.5	-12.9	-0.2	0.0	0.0	26.1	-2.9	24.9	16.2	9.6
$\Delta G_1°$		8.7	-2.7	-3.4	0.0	0.0	35.2	-1.9	22.9	12.7	5.0
ΔE_1	T	11.4	-7.2	5.2	-4.9	0.0	33.7	7.8	26.9	19.8	6.8
$\Delta E_{1(ZPE)}$		11.9	-9.0	6.2	-3.5	0.0	32.8	7.1	27.8	21.5	8.6
$\Delta G_1°$		7.9	-1.9	1.7	-7.8	0.0	40.5	12.3	25.1	14.5	3.7
ΔE_1	U	10.2	-7.1	4.8	-21.3	0.0	34.4	6.8	26.2	19.5	6.3
$\Delta E_{1(ZPE)}$		11.8	-8.6	5.8	-21.4	0.0	33.4	6.1	26.8	21.3	8.1
$\Delta G_1°$		3.6	-3.3	1.5	-18.5	0.0	40.5	11.1	26.2	14.4	3.1
ΔE_1	TAP-C^5	-1.1	1.3	8.0	-1.8	0.0	21.8	3.5	36.6	9.4	10.3
$\Delta E_{1(ZPE)}$		-1.3	1.8	8.3	-2.0	0.0	22.8	2.1	36.1	10.3	11.4
$\Delta G_1°$		-3.8	-0.1	10.8	-1.6	0.0	18.1	5.3	37.6	7.5	7.2
ΔE_1	TAP-N	8.4	3.6	2.8	0.5	0.0	32.9	10.2	4.4	-7.6	-0.1
$\Delta E_{1(ZPE)}$		8.6	2.7	2.8	-0.0	0.0	31.1	10.7	3.7	-6.2	2.4
$\Delta G_1°$		5.9	5.3	5.5	-0.8	0.0	41.1	6.6	7.2	-10.2	-4.6
ΔE_1	BA-C^5	-12.4	-10.8	-24.1	2.2	0.0	-14.7	-4.6	39.2	22.8	-20.8
$\Delta E_{1(ZPE)}$		-12.0	-12.6	-23.8	2.1	0.0	-15.7	-5.7	39.4	24.6	-20.5
$\Delta G_1°$		-12.9	-6.9	-21.8	2.9	0.0	-12.9	-3.4	37.5	19.4	-23.7
ΔE_1	BA-N	-9.9	-9.3	3.0	0.0	0.0	17.7	18.0	34.3	4.5	28.9
$\Delta E_{1(ZPE)}$		-10.1	-10.6	2.9	0.0	0.0	19.6	17.5	34.6	5.9	31.5
$\Delta G_1°$		-10.8	-6.8	5.1	0.0	0.0	13.1	21.2	34.8	1.1	21.4
ΔE_1	CA	-8.6	0.0	2.7	-0.1	0.0	19.7	16.1	43.6	3.2	24.4
$\Delta E_{1(ZPE)}$		-8.9	0.0	2.6	-0.1	0.0	20.9	15.6	45.0	4.4	27.3
$\Delta G_1°$		-8.4	-0.0	4.8	0.4	0.0	17.3	19.5	40.0	-0.3	15.6
ΔE_1		0.2	-2.2	4.6	0.0	0.0	17.3	-5.8	12.3	-7.1	-0.1
$\Delta E_{1(ZPE)}$	MM	0.4	-3.8	4.9	0.0	0.0	18.0	-5.0	12.8	-5.6	2.1
$\Delta G_1°$		-2.1	3.8	6.2	0.0	0.0	15.5	-9.2	12.2	-9.8	-3.3

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_2	A	-14.4	-0.3	38.1	-20.3	-1.1	-15.9	-0.4	-12.3	-23.9	-26.0
$\Delta E_{2(ZPE)}$		-15.1	-1.7	39.8	-21.8	-1.5	-17.4	0.5	-12.6	-24.2	-26.0
ΔG_2°		-11.7	6.1	36.5	-16.0	0.3	-14.3	-3.9	-11.1	-24.4	-24.5
ΔE_2	G	-10.1	-8.2	4.1	-28.9	4.8	-28.4	-8.6	-17.2	-20.4	-26.2
$\Delta E_{2(ZPE)}$		-10.9	-11.0	5.0	-31.9	5.0	-26.5	-7.9	-17.0	-20.4	-26.5
ΔG_2°		-4.4	3.5	7.0	-15.1	2.9	-39.2	-12.3	-21.3	-21.9	-23.6
ΔE_2	C	-20.5	2.0	4.6	-5.4	-4.7	-9.1	12.1	-18.6	-2.1	-21.0
$\Delta E_{2(ZPE)}$		-20.1	2.2	4.7	-5.6	-5.4	-8.5	12.6	-19.2	-1.4	-20.6
ΔG_2°		-21.1	3.9	4.1	-0.5	-0.9	-15.1	12.5	-17.3	-5.8	-22.9
ΔE_2	T	9.5	2.1	28.9	-7.5	-2.6	-9.2	-1.6	-21.6	-14.4	-21.5
$\Delta E_{2(ZPE)}$		9.7	1.8	30.1	-8.2	-3.3	-8.6	-0.8	-22.2	-14.4	-21.0
ΔG_2°		12.6	5.8	29.6	-3.1	-0.2	-15.3	-4.5	-19.8	-14.6	-23.6
ΔE_2	U	-9.3	0.1	30.1	15.4	-2.9	-16.3	-1.4	-20.7	-14.1	-20.4
$\Delta E_{2(ZPE)}$		-9.4	-0.3	31.3	16.0	-3.6	-14.1	-0.5	-21.2	-14.1	-19.9
ΔG_2°		-7.1	5.2	30.6	14.2	-0.5	-26.1	-4.3	-19.0	-14.6	-22.1
ΔE_2	TAP-C5	16.2	-2.3	-16.8	-15.6	17.4	-2.9	-8.5	-31.0	-10.6	-28.5
$\Delta E_{2(ZPE)}$		16.3	-2.5	-15.9	-13.7	16.6	-3.9	-6.5	-30.6	-9.7	-26.4
ΔG_2°		19.4	-2.9	-21.2	-23.9	18.7	4.5	-12.6	-29.5	-14.2	-33.3
ΔE_2	TAP-N	-23.0	-10.3	23.2	-10.4	41.0	-13.5	4.9	-0.2	17.8	-1.2
$\Delta E_{2(ZPE)}$		-24.1	-12.3	24.8	-11.9	41.2	-10.8	6.0	0.4	18.1	-0.8
ΔG_2°		-14.8	-1.6	24.2	-2.7	45.7	-24.5	2.2	-1.5	15.0	-1.9
ΔE_2	BA-C5	-16.3	-47.7	39.7	-71.5	-11.8	27.9	14.0	-35.3	8.0	11.1
$\Delta E_{2(ZPE)}$		-16.3	-50.0	40.1	-73.7	-11.5	28.4	15.6	-34.8	9.2	13.4
ΔG_2°		-17.0	-42.1	38.2	-63.5	-11.6	26.7	11.5	-34.2	4.1	7.2
ΔE_2	BA-N	-38.2	1.0	-80.5	-7.3	1.1	2.2	-27.6	-29.9	-38.7	-43.5
$\Delta E_{2(ZPE)}$		-40.4	0.5	-81.6	-7.3	0.8	0.5	-27.0	-30.2	-38.5	-43.2
ΔG_2°		-27.8	3.3	-76.1	-3.1	2.7	5.3	-29.4	-28.0	-39.9	-42.4
ΔE_2	CA	1.9	-0.0	4.1	2.2	-2.7	3.1	-27.1	-39.4	-38.2	-40.9
$\Delta E_{2(ZPE)}$		1.9	-0.3	3.1	2.1	-3.0	2.0	-26.5	-40.9	-38.0	-40.9
ΔG_2°		1.4	1.5	12.0	4.9	-1.4	4.2	-29.0	-33.4	-39.6	-39.7
ΔE_2	MM	-3.4	-10.0	38.2	-17.4	36.2	4.1	9.4	-8.0	10.1	-0.6
$\Delta E_{2(ZPE)}$		-4.5	-11.4	38.4	-18.4	36.9	3.5	10.6	-8.4	10.8	-0.0
ΔG_2°		0.7	-4.5	41.5	-15.4	37.2	5.6	6.5	-6.0	4.7	-3.5
ΔE_3	A	-15.4	-3.6	-21.1	-12.7	0.0	3.4	-13.5	-7.3	-23.0	-18.3
$\Delta E_{3(ZPE)}$		-16.7	-5.2	-19.9	-14.3	0.0	3.4	-13.6	-6.9	-21.6	-17.2
ΔG_3°		-14.4	3.6	-26.7	-10.3	0.0	3.2	-11.4	-7.2	-28.8	-22.8
ΔE_3	G	-16.3	-15.5	55.3	-26.0	4.9	3.5	-17.0	-7.6	-24.2	-17.1
$\Delta E_{3(ZPE)}$		-18.0	-16.7	57.8	-28.0	5.6	3.4	-16.9	-7.3	-22.6	-16.1
ΔG_3°		-10.7	-12.3	49.0	-19.2	-3.6	3.3	-14.3	-7.6	-30.8	-19.7
ΔE_3	C	-10.2	-4.9	-38.1	-0.3	0.0	12.5	-18.8	-15.6	-25.6	-23.9
$\Delta E_{3(ZPE)}$		-10.8	-6.4	-38.3	-2.5	0.0	13.3	-18.8	-15.6	-24.2	-23.1
ΔG_3°		-11.0	0.7	-43.1	6.2	0.0	9.0	-18.0	-15.4	-29.5	-26.7
ΔE_3	T	16.6	-2.8	30.3	-7.6	0.0	10.8	-20.9	-13.4	-29.0	-25.8
$\Delta E_{3(ZPE)}$		15.9	-4.5	31.3	-7.7	0.0	11.6	-21.0	-13.1	-28.0	-24.9
ΔG_3°		19.1	3.6	34.6	-6.2	0.0	7.1	-19.2	-12.9	-32.4	-28.9
ΔE_3	U	-0.2	-3.5	38.9	5.1	0.0	12.1	-20.4	-13.2	-29.0	-25.2
$\Delta E_{3(ZPE)}$		0.7	-5.1	40.2	4.0	0.0	13.8	-20.4	-12.9	-28.0	-24.4
ΔG_3°		0.1	4.2	38.9	8.0	0.0	5.4	-18.9	-13.6	-32.5	-27.4
ΔE_3	TAP-C5	0.0	-4.4	-1.6	-5.5	-2.7	9.9	-15.0	-14.4	8.0	-10.8
$\Delta E_{3(ZPE)}$		0.0	-5.9	-1.3	-5.0	-2.4	9.4	-15.5	-14.6	8.5	-9.7
ΔG_3°		0.0	-0.6	-1.5	-6.3	-4.2	14.9	-12.5	-13.3	5.1	-14.8
ΔE_3	TAP-N	-17.2	1.1	14.5	2.3	49.7	22.9	-6.8	-1.8	1.1	-9.2
$\Delta E_{3(ZPE)}$		-17.5	1.5	15.1	2.7	50.6	24.3	-6.8	-1.9	2.5	-8.3
ΔG_3°		-14.7	-0.9	17.8	-1.4	51.4	18.2	-5.5	0.4	-4.4	-11.6
ΔE_3	BA-C5	-18.8	0.8	47.1	-13.8	-8.1	8.4	-20.2	-18.4	-9.6	-23.2
$\Delta E_{3(ZPE)}$		-20.5	-0.9	47.0	-13.6	-7.6	8.4	-20.7	-17.8	-7.9	-22.5
ΔG_3°		-13.9	4.5	46.0	-11.9	-8.0	6.5	-17.2	-18.2	-15.2	-24.2

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_3	BA-N	26.5	-8.8	-42.7	5.8	3.1	1.9	-14.1	-12.0	-23.8	-19.1
$\Delta E_{3(ZPE)}$		27.4	-8.8	-41.2	5.4	3.0	1.6	-14.4	-11.8	-22.6	-18.4
ΔG_3°		24.6	-10.7	-45.3	7.8	3.4	1.9	-12.3	-11.2	-28.2	-20.7
ΔE_3	CA	-22.2	-9.9	-12.0	5.3	0.0	21.6	-14.0	-10.0	-23.9	-18.1
$\Delta E_{3(ZPE)}$		-22.5	-10.2	-13.4	3.5	0.0	21.7	-14.3	-10.3	-22.9	-17.4
ΔG_3°		-27.2	-10.6	-2.3	10.2	0.0	19.5	-11.8	-8.0	-28.0	-20.4
ΔE_3	MM	9.5	-1.8	-4.8	-7.0	42.3	4.0	-9.4	-4.0	-15.8	-11.2
$\Delta E_{3(ZPE)}$		8.8	-2.3	-3.2	-7.6	43.3	4.6	-9.6	-3.8	-14.2	-10.2
ΔG_3°		8.0	-2.1	-8.7	-6.5	41.1	0.2	-7.3	-3.8	-22.4	-14.0
ΔE_4	A	-24.1	-1.4	62.2	-2.8	-1.1	0.7	14.2	12.3	23.8	5.8
$\Delta E_{4(ZPE)}$		-24.6	-1.8	63.3	-1.6	-1.5	0.6	14.4	12.1	24.2	7.1
ΔG_4°		-21.2	1.7	63.5	-2.9	0.3	-0.3	13.3	14.2	23.5	5.1
ΔE_4	G	6.8	-0.7	-51.9	-0.1	-0.0	-0.2	39.2	12.4	18.5	6.4
$\Delta E_{4(ZPE)}$		7.8	-2.6	-53.0	0.2	-0.6	0.2	40.1	12.1	18.9	8.8
ΔG_4°		5.4	8.0	-45.2	2.0	6.4	-4.7	30.2	14.0	13.6	-0.9
ΔE_4	C	2.6	-3.8	41.9	-5.1	-4.7	5.9	28.2	21.0	38.3	10.6
$\Delta E_{4(ZPE)}$		4.2	-4.3	42.7	-3.0	-5.4	4.3	28.4	21.3	39.0	12.1
ΔG_4°		-1.4	0.6	43.8	-6.7	-0.9	11.2	28.6	21.0	36.4	8.9
ΔE_4	T	4.3	-2.4	3.8	-4.7	-2.6	13.7	27.1	18.7	34.4	11.1
$\Delta E_{4(ZPE)}$		5.6	-2.7	5.0	-4.0	-3.3	12.6	27.3	18.7	35.1	12.5
ΔG_4°		1.3	0.3	-3.3	-4.6	-0.2	18.1	27.0	18.1	32.3	9.0
ΔE_4	U	1.1	-3.5	-4.0	-11.0	-2.9	6.0	25.8	18.7	34.4	11.0
$\Delta E_{4(ZPE)}$		1.8	-3.8	-3.0	-9.4	-3.6	5.5	26.0	18.5	35.1	12.5
ΔG_4°		-3.7	-2.3	-6.9	-12.3	-0.5	9.0	25.7	20.8	32.3	8.5
ΔE_4	TAP-C5	15.0	3.5	-7.1	-11.9	20.1	9.0	9.9	19.9	-9.2	-7.4
$\Delta E_{4(ZPE)}$		15.0	5.1	-6.3	-10.7	19.1	9.5	11.1	20.1	-7.9	-5.4
ΔG_4°		15.7	-2.3	-8.9	-19.2	23.0	7.7	5.2	21.4	-11.8	-11.2
ΔE_4	TAP-N	2.6	-7.9	11.6	-12.2	-8.7	-3.6	22.0	6.0	9.1	8.0
$\Delta E_{4(ZPE)}$		2.0	-11.1	12.5	-14.6	-9.4	-4.0	23.5	6.1	9.4	9.9
ΔG_4°		5.8	4.6	11.9	-2.0	-5.7	-1.6	14.3	5.3	9.1	5.2
ΔE_4	BA-C5	-9.9	-59.3	-31.5	-55.4	-3.7	4.8	29.7	22.3	40.5	13.5
$\Delta E_{4(ZPE)}$		-7.8	-61.7	-30.7	-57.9	-3.9	4.4	30.5	22.4	41.6	15.4
ΔG_4°		-16.1	-53.4	-29.6	-48.7	-3.6	7.3	25.2	21.4	38.7	7.7
ΔE_4	BA-N	-74.6	0.6	-34.7	-13.0	-2.0	18.0	4.6	16.4	-10.3	4.6
$\Delta E_{4(ZPE)}$		-77.9	-1.3	-37.5	-12.7	-2.2	18.5	4.9	16.2	-10.1	6.6
ΔG_4°		-63.1	7.2	-25.8	-11.0	-0.7	16.4	4.1	17.9	-10.6	-0.3
ΔE_4	CA	15.5	9.9	18.8	-3.1	-2.7	1.2	3.0	14.2	-11.2	1.7
$\Delta E_{4(ZPE)}$		15.4	9.8	19.1	-1.5	-3.0	1.3	3.4	14.4	-10.7	3.7
ΔG_4°		20.1	12.1	19.0	-5.0	-1.4	2.0	2.2	14.6	-11.9	-3.7
ΔE_4	MM	-12.6	-10.4	47.5	-10.5	-6.1	17.4	13.0	8.3	18.7	10.4
$\Delta E_{4(ZPE)}$		-13.0	-12.9	46.4	-10.8	-6.4	16.9	15.2	8.1	19.4	12.3
ΔG_4°		-9.4	1.4	56.3	-8.9	-3.9	21.0	4.6	10.0	17.3	7.2
ΔE_5	A	-39.6	-4.9	41.1	-15.5	-1.1	4.1	0.8	5.1	0.8	-12.4
$\Delta E_{5(ZPE)}$		-41.3	-7.0	43.4	-15.9	-1.5	4.0	0.8	5.2	2.6	-10.1
ΔG_5°		-35.7	5.3	36.8	-13.2	0.3	2.9	1.9	7.0	-5.3	-17.6
ΔE_5	G	-9.5	-16.2	3.4	-26.1	4.8	3.3	22.2	4.8	-5.7	-10.6
$\Delta E_{5(ZPE)}$		-10.1	-19.3	4.8	-27.7	5.0	3.7	23.3	4.8	-3.7	-7.3
ΔG_5°		-5.3	-4.3	3.8	-17.2	2.9	-1.5	15.9	6.4	-17.2	-20.6
ΔE_5	C	-7.6	-8.7	3.7	-5.4	-4.7	18.4	9.4	5.4	12.7	-13.4
$\Delta E_{5(ZPE)}$		-6.6	-10.7	4.4	-5.6	-5.4	17.6	9.7	5.7	14.8	-11.0
ΔG_5°		-12.4	1.2	0.6	-0.5	-0.9	20.1	10.6	5.6	6.9	-17.9
ΔE_5	T	20.9	-5.1	34.1	-12.4	-2.6	24.5	6.1	5.3	5.4	-14.7
$\Delta E_{5(ZPE)}$		21.6	-7.2	36.3	-11.7	-3.3	24.2	6.3	5.6	7.1	-12.4
ΔG_5°		20.5	3.9	31.3	-10.9	-0.2	25.2	7.8	5.3	-0.1	-19.9
ΔE_5	U	0.9	-7.0	34.9	-5.8	-2.9	18.1	5.4	5.5	5.4	-14.1
$\Delta E_{5(ZPE)}$		2.5	-8.9	37.2	-5.4	-3.6	19.3	5.6	5.6	7.2	-11.9
ΔG_5°		-3.5	1.8	32.0	-4.3	-0.5	14.3	6.8	7.2	-0.2	-19.0

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_5	TAP-C^5	15.0	-1.0	-8.8	-17.4	17.4	18.8	-5.1	5.5	-1.2	-18.2
$\Delta E_{5(ZPE)}$		15.0	-0.8	-7.6	-15.7	16.6	18.9	-4.4	5.5	0.6	-15.1
ΔG_5°		15.7	-2.9	-10.4	-25.5	18.7	22.6	-7.4	8.1	-6.7	-26.0
ΔE_5	TAP-N	-14.6	-6.7	26.0	-9.8	41.0	19.3	15.2	4.1	10.2	-1.2
$\Delta E_{5(ZPE)}$		-15.5	-9.6	27.6	-11.9	41.2	20.3	16.7	4.1	11.9	1.6
ΔG_5°		-8.9	3.7	29.7	-3.4	45.7	16.6	8.8	5.7	4.7	-6.5
ΔE_5	BA-C^5	-28.7	-58.5	15.6	-69.3	-11.8	13.2	9.5	3.9	30.9	-9.7
$\Delta E_{5(ZPE)}$		-28.2	-62.6	16.3	-71.6	-11.5	12.8	9.9	4.6	33.8	-7.1
ΔG_5°		-30.0	-49.0	16.4	-60.6	-11.6	13.8	8.1	3.2	23.5	-16.5
ΔE_5	BA-N	-48.1	-8.2	-77.5	-7.3	1.1	19.9	-9.5	4.5	-34.1	-14.5
$\Delta E_{5(ZPE)}$		-50.5	-10.1	-78.7	-7.3	0.8	20.2	-9.5	4.4	-32.6	-11.8
ΔG_5°		-38.6	-3.5	-71.1	-3.1	2.7	18.4	-8.2	6.8	-38.8	-21.0
ΔE_5	CA	-6.7	-0.0	6.8	2.1	-2.7	22.8	-10.9	4.2	-35.1	-16.4
$\Delta E_{5(ZPE)}$		-7.0	-0.3	5.7	2.0	-3.0	22.9	-10.9	4.1	-33.7	-13.6
ΔG_5°		-7.0	1.5	16.8	5.3	-1.4	21.5	-9.5	6.6	-39.9	-24.1
ΔE_5	MM	-3.2	-12.2	42.8	-17.4	36.2	21.4	3.6	4.3	2.9	-0.8
$\Delta E_{5(ZPE)}$		-4.2	-15.2	43.3	-18.4	36.9	21.5	5.6	4.4	5.2	2.0
ΔG_5°		-1.4	-0.7	47.6	-15.4	37.2	21.2	-2.7	6.2	-5.1	-6.8
ΔE_6	A	-1.0	-3.2	-59.2	7.6	1.1	19.3	-13.1	5.1	1.0	7.7
$\Delta E_{6(ZPE)}$		-1.6	-3.5	-59.7	7.6	1.5	20.8	-14.1	5.7	2.6	8.8
ΔG_6°		-2.7	-2.5	-63.2	5.7	-0.3	17.5	-7.5	3.9	-4.4	1.8
ΔE_6	G	-6.2	-7.3	51.2	2.9	0.0	31.9	-8.4	9.6	-3.8	9.2
$\Delta E_{6(ZPE)}$		-7.1	-5.7	52.9	3.9	0.6	29.9	-8.9	9.7	-2.2	10.4
ΔG_6°		-6.3	-15.8	42.0	-4.1	-6.4	42.5	-2.0	13.7	-8.9	3.9
ΔE_6	C	10.3	-6.8	-42.7	5.1	4.7	21.6	-30.9	3.0	-23.4	-2.9
$\Delta E_{6(ZPE)}$		9.4	-8.6	-43.0	3.0	5.4	21.8	-31.3	3.7	-22.8	-2.5
ΔG_6°		10.1	-3.2	-47.2	6.7	0.9	24.0	-30.5	1.9	-23.7	-3.8
ΔE_6	T	7.1	-4.9	1.4	-0.2	2.6	20.0	-19.3	8.2	-14.6	-4.3
$\Delta E_{6(ZPE)}$		6.2	-6.3	1.2	0.5	3.3	20.2	-20.2	9.1	-13.6	-3.9
ΔG_6°		6.6	-2.2	5.0	-3.2	0.2	22.4	-14.7	7.0	-17.8	-5.3
ΔE_6	U	9.1	-3.6	8.9	-10.3	2.9	28.4	-19.0	7.4	-14.9	-4.7
$\Delta E_{6(ZPE)}$		10.0	-4.8	8.9	-12.0	3.6	27.9	-19.9	8.3	-13.9	-4.5
ΔG_6°		7.2	-1.0	8.4	-6.2	0.5	31.5	-14.5	5.4	-17.9	-5.3
ΔE_6	TAP-C^5	-16.2	-2.2	15.2	10.1	-20.1	12.8	-6.4	16.6	18.6	17.7
$\Delta E_{6(ZPE)}$		-16.3	-3.3	14.6	8.7	-19.1	13.3	-9.0	16.0	18.2	16.8
ΔG_6°		-19.4	2.2	19.7	17.6	-23.0	10.4	0.1	16.2	19.3	18.5
ΔE_6	TAP-N	5.8	11.5	-8.8	12.7	8.7	36.5	-11.8	-1.6	-16.7	-8.0
$\Delta E_{6(ZPE)}$		6.6	13.8	-9.7	14.6	9.4	35.1	-12.8	-2.4	-15.6	-7.5
ΔG_6°		0.0	0.7	-6.4	1.3	5.7	42.8	-7.7	1.9	-19.4	-9.8
ΔE_6	BA-C^5	-2.4	48.5	7.4	57.6	3.7	-19.5	-34.3	17.0	-17.6	-34.3
$\Delta E_{6(ZPE)}$		-4.2	49.2	6.9	60.0	3.9	-20.0	-36.2	17.0	-17.0	-35.9
ΔG_6°		3.2	46.6	7.8	51.6	3.6	-20.2	-28.6	16.1	-19.3	-31.4
ΔE_6	BA-N	64.7	-9.8	37.7	13.0	2.0	-0.4	13.4	17.9	14.9	24.3
$\Delta E_{6(ZPE)}$		67.8	-9.3	40.4	12.7	2.2	1.1	12.6	18.4	15.9	24.9
ΔG_6°		52.3	-14.0	30.8	11.0	0.7	-3.4	17.1	16.8	11.7	21.7
ΔE_6	CA	-24.2	-9.9	-16.1	3.1	2.7	18.5	13.1	29.4	14.3	22.8
$\Delta E_{6(ZPE)}$		-24.4	-9.8	-16.5	1.4	3.0	19.7	12.2	30.6	15.1	23.5
ΔG_6°		-28.5	-12.1	-14.2	5.4	1.4	15.3	17.2	25.4	11.6	19.3
ΔE_6	MM	12.9	8.2	-43.0	10.5	6.1	-0.1	-18.8	4.0	-25.9	-10.6
$\Delta E_{6(ZPE)}$		13.3	9.1	-41.5	10.8	6.4	1.1	-20.2	4.6	-25.0	-10.2
ΔG_6°		7.3	2.4	-50.1	8.9	3.9	-5.4	-13.9	2.2	-27.1	-10.5
						$HAsO_3^-$					
ΔE_1	A	7.2	-6.1	-4.8	2.8	0.0	29.1	6.3	29.3	23.8	25.5
$\Delta E_{1(ZPE)}$		7.5	-4.0	-5.4	4.3	0.0	29.7	5.7	27.8	25.5	26.5
ΔG_1°		3.4	-13.9	-2.0	-0.0	0.0	30.6	11.0	34.4	18.9	22.7

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_1	G	8.2	-6.0	-1.6	7.9	0.0	21.4	27.2	5.4	-6.0	7.4
$\Delta E_{1(ZPE)}$		8.4	-3.6	-1.5	9.1	0.0	19.4	27.1	2.2	-5.4	5.9
ΔG_1°		5.8	-13.8	-1.2	6.3	0.0	30.8	29.0	13.4	-7.7	15.3
ΔE_1	C	19.0	3.6	33.9	7.5	0.0	26.2	27.4	34.4	18.0	23.3
$\Delta E_{1(ZPE)}$		19.8	5.3	33.5	11.1	0.0	24.5	27.7	33.4	19.4	24.8
ΔG_1°		12.8	-1.3	37.4	-1.8	0.0	34.7	28.4	39.1	14.6	20.0
ΔE_1	T	14.0	2.8	39.3	-8.7	0.0	20.9	11.5	34.4	24.1	21.9
$\Delta E_{1(ZPE)}$		15.8	4.6	38.6	-7.3	0.0	22.1	11.1	33.3	25.8	23.5
ΔG_1°		7.1	-3.0	42.0	-14.7	0.0	17.9	14.4	39.4	19.1	18.3
ΔE_1	U	13.1	2.6	38.4	-8.8	0.0	32.8	10.8	21.7	23.9	21.7
$\Delta E_{1(ZPE)}$		14.8	4.2	38.3	-7.5	0.0	31.7	10.3	21.2	25.6	23.3
ΔG_1°		6.2	-3.0	41.8	-13.6	0.0	40.3	14.1	24.6	19.0	18.3
ΔE_1	TAP-C^5	-9.4	23.9	2.1	8.0	0.0	28.7	13.0	41.1	31.7	37.5
$\Delta E_{1(ZPE)}$		-9.7	26.5	3.5	9.6	0.0	26.8	11.2	38.3	33.6	39.4
ΔG_1°		-12.8	17.1	0.7	5.2	0.0	35.6	14.9	48.5	25.8	33.1
ΔE_1	TAP-N	2.0	5.5	-5.5	12.5	0.0	17.6	12.0	-4.7	-3.9	4.4
$\Delta E_{1(ZPE)}$		1.8	7.0	-5.6	14.5	0.0	18.5	12.7	-4.5	-2.5	6.2
ΔG_1°		1.6	-2.5	-4.9	0.9	0.0	16.0	7.2	-8.8	-7.5	-0.6
ΔE_1	BA-C^5	73.9	65.4	-18.3	84.4	0.0	-31.7	-19.1	30.2	4.1	-36.8
$\Delta E_{1(ZPE)}$		75.1	69.7	-18.6	88.0	0.0	-36.0	-18.6	28.5	6.5	-38.5
ΔG_1°		69.9	53.3	-15.7	74.9	-0.0	-19.3	-20.6	34.5	-1.1	-32.8
ΔE_1	BA-N	-5.3	-2.6	-5.0	7.7	0.0	26.4	22.5	26.4	9.6	11.7
$\Delta E_{1(ZPE)}$		-5.4	-2.2	-5.1	9.4	0.0	25.8	22.1	24.8	11.0	12.2
ΔG_1°		-3.6	-5.3	-4.3	4.3	0.0	30.5	25.7	30.9	4.2	10.9
ΔE_1	CA	-25.7	2.4	-2.6	14.3	0.0	29.9	20.5	30.1	6.6	35.0
$\Delta E_{1(ZPE)}$		-25.9	3.9	-4.6	16.6	0.0	29.7	20.0	28.9	7.6	36.4
ΔG_1°		-25.8	-2.1	-4.2	9.3	0.0	33.6	24.0	33.4	3.6	31.6
ΔE_1	MM	25.9	2.8	-5.0	2.4	0.0	30.1	-24.5	5.8	-4.6	23.7
$\Delta E_{1(ZPE)}$		28.4	4.4	-4.5	3.4	0.0	28.8	-22.7	3.1	-3.2	22.0
ΔG_1°		18.8	-4.4	-5.3	0.8	0.0	35.3	-31.2	12.6	-7.9	30.1
ΔE_2	A	22.3	2.3	37.0	2.0	340.4	-0.1	-2.8	-25.8	-14.1	-30.4
$\Delta E_{2(ZPE)}$		22.6	0.8	37.1	2.3	341.9	-0.5	-1.9	-24.0	-14.2	-29.3
ΔG_2°		26.2	10.1	39.6	2.7	342.7	-2.8	-5.4	-30.4	-15.2	-32.2
ΔE_2	G	-15.0	-6.3	12.0	-9.3	357.3	8.6	-5.0	-7.6	3.7	-13.5
$\Delta E_{2(ZPE)}$		-14.7	-9.2	14.1	-11.2	360.6	10.8	-3.5	-3.7	4.6	-10.0
ΔG_2°		-11.6	5.0	7.8	-1.6	348.3	-2.0	-13.7	-19.5	-4.7	-24.6
ΔE_2	C	-4.7	5.0	-28.0	-13.9	29.0	7.6	-15.9	-30.1	2.7	-26.6
$\Delta E_{2(ZPE)}$		-4.7	5.0	-28.9	-17.3	28.8	7.7	-15.6	-28.9	3.1	-26.3
ΔG_2°		0.4	8.5	-24.8	-2.1	32.9	8.0	-15.6	-32.9	0.8	-27.2
ΔE_2	T	6.1	5.6	-17.4	-6.3	28.3	14.8	-2.7	-30.4	-10.3	-27.1
$\Delta E_{2(ZPE)}$		4.9	5.5	-17.7	-6.2	28.7	12.9	-1.8	-29.0	-10.3	-26.6
ΔG_2°		12.4	8.9	-13.8	-2.8	29.0	21.6	-4.0	-35.2	-10.7	-28.1
ΔE_2	U	10.3	3.0	-19.9	-8.6	27.8	-16.5	-2.7	-17.3	-10.0	-25.7
$\Delta E_{2(ZPE)}$		9.3	2.4	-19.1	-8.2	28.2	-14.4	-1.8	-16.8	-9.9	-25.3
ΔG_2°		16.4	7.8	-24.2	-6.4	28.5	-24.4	-4.3	-17.3	-10.8	-25.9
ΔE_2	TAP-C^5	13.6	-12.9	-26.3	-8.5	48.1	-2.0	-11.8	-37.0	-24.5	-51.2
$\Delta E_{2(ZPE)}$		13.7	-14.2	-27.0	-8.8	47.5	0.2	-9.8	-34.1	-24.9	-49.9
ΔG_2°		17.7	-10.9	-26.7	-8.4	52.0	-5.6	-12.9	-42.0	-23.0	-54.4
ΔE_2	TAP-N	-7.8	-8.6	32.7	-6.2	23.4	11.2	1.8	-4.3	22.8	-5.2
$\Delta E_{2(ZPE)}$		-8.5	-10.6	33.5	-7.9	23.1	10.7	2.8	-3.5	22.7	-4.8
ΔG_2°		-1.4	0.4	37.4	2.0	28.5	14.7	0.3	-4.3	22.9	-4.9
ΔE_2	BA-C^5	-92.0	-64.3	-29.8	-68.0	-17.6	60.1	29.9	-27.7	52.5	22.9
$\Delta E_{2(ZPE)}$		-93.1	-66.9	-29.7	-69.8	-18.4	63.3	30.2	-25.4	50.2	26.6
ΔG_2°		-88.2	-57.4	-29.9	-61.8	-16.1	50.9	30.1	-32.0	57.0	14.5
ΔE_2	BA-N	-5.4	2.8	-17.8	-9.1	33.5	12.8	-29.6	-24.1	-35.2	-39.2
$\Delta E_{2(ZPE)}$		-5.5	2.7	-18.1	-8.6	33.9	12.6	-28.9	-22.0	-35.5	-37.2
ΔG_2°		-4.8	4.0	-13.3	-9.2	33.3	13.1	-30.5	-28.2	-32.6	-44.3

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_2	CA	16.9	2.3	-27.5	-19.9	32.2	2.0	-28.1	-27.9	-32.5	-60.0
$\Delta E_{2(ZPE)}$		17.1	2.2	-29.7	-20.5	32.6	2.7	-27.4	-26.3	-32.4	-59.6
ΔG_2°		15.4	2.4	-17.8	-18.2	32.1	-1.7	-29.2	-30.2	-33.2	-59.8
ΔE_2	MM	-9.4	-7.9	21.1	-8.4	20.2	2.3	7.9	-19.3	15.0	-24.2
$\Delta E_{2(ZPE)}$		-12.5	-9.4	19.0	-9.4	22.0	4.1	9.0	-15.2	15.3	-20.4
ΔG_2°		-0.0	-1.0	30.2	-6.0	14.6	-1.7	6.5	-30.4	13.0	-35.3
ΔE_3	A	0.0	-1.8	9.5	9.8	343.6	28.5	-14.9	-8.8	-14.5	-27.0
$\Delta E_{3(ZPE)}$		0.0	-3.6	7.3	9.0	345.7	29.1	-14.3	-8.3	-13.2	-24.5
ΔG_3°		0.0	5.5	16.1	16.7	344.1	27.7	-15.0	-8.4	-20.3	-32.7
ΔE_3	G	0.0	-4.1	-34.2	-8.3	348.7	29.6	-15.2	-8.7	-15.9	-28.8
$\Delta E_{3(ZPE)}$		0.0	-4.0	-33.2	-10.2	352.0	30.1	-14.5	-8.5	-14.7	-26.0
ΔG_3°		-0.0	-3.6	-36.8	-4.2	339.4	29.1	-14.7	-7.4	-21.1	-35.8
ΔE_3	C	0.0	-1.8	6.4	-0.7	36.0	28.2	-17.6	-16.6	-17.9	-26.3
$\Delta E_{3(ZPE)}$		0.0	-3.8	5.7	-2.4	36.7	28.5	-17.1	-16.7	-16.7	-24.8
ΔG_3°		0.0	6.5	9.4	7.3	35.6	30.7	-17.5	-14.9	-21.3	-29.7
ΔE_3	T	0.0	-0.3	18.1	-1.3	34.7	34.7	-19.0	-18.6	-20.9	-36.1
$\Delta E_{3(ZPE)}$		0.0	-1.5	17.4	-2.2	35.5	34.9	-18.5	-18.2	-19.9	-33.6
ΔG_3°		0.0	5.1	26.6	5.0	34.1	36.6	-18.0	-19.2	-24.4	-41.8
ΔE_3	U	22.5	-10.8	28.7	-2.7	34.7	16.3	-18.4	-17.4	-20.9	-36.3
$\Delta E_{3(ZPE)}$		21.3	-12.9	29.5	-3.4	35.5	17.8	-17.9	-17.4	-19.8	-33.9
ΔG_3°		27.1	-3.3	25.1	3.2	34.0	14.9	-17.8	-15.8	-24.6	-41.0
ΔE_3	TAP-C^5	0.0	-11.9	0.4	-6.4	32.2	6.7	-14.7	-15.2	15.1	-39.6
$\Delta E_{3(ZPE)}$		0.0	-13.8	0.5	-8.7	32.5	8.5	-13.0	-15.4	15.4	-36.8
ΔG_3°		0.0	-5.3	0.3	2.5	33.7	5.2	-19.3	-13.1	13.2	-47.2
ΔE_3	TAP-N	0.0	-8.1	27.2	9.0	34.1	13.9	-17.8	-3.2	9.9	-16.0
$\Delta E_{3(ZPE)}$		0.0	-8.0	29.0	8.8	34.8	15.2	-17.9	-3.2	10.9	-13.4
ΔG_3°		0.0	-10.5	23.3	8.2	34.6	13.0	-16.1	-1.6	6.5	-21.6
ΔE_3	BA-C^5	0.0	-78.3	-73.4	-16.1	11.9	23.2	6.9	-19.7	15.3	-27.9
$\Delta E_{3(ZPE)}$		0.0	-82.2	-74.4	-16.6	12.4	23.7	6.6	-19.3	14.4	-25.3
ΔG_3°		0.0	-66.2	-68.7	-13.4	8.8	21.8	5.4	-18.8	15.8	-33.6
ΔE_3	BA-N	52.9	-10.2	-27.3	-4.0	37.7	19.8	-13.1	-13.8	-15.8	-22.1
$\Delta E_{3(ZPE)}$		55.2	-10.2	-27.7	-3.5	38.6	15.2	-13.0	-13.3	-15.0	-20.0
ΔG_3°		44.3	-11.6	-21.8	-5.6	35.6	34.2	-11.9	-13.9	-18.6	-26.5
ΔE_3	CA	26.0	0.0	-14.9	-5.6	37.1	-210.6	-11.4	-11.8	-15.5	-20.9
$\Delta E_{3(ZPE)}$		27.7	0.0	-16.7	-5.4	38.1	-212.7	-11.3	-11.8	-14.7	-19.3
ΔG_3°		17.5	0.0	-7.1	-7.3	34.8	-202.9	-10.2	-10.5	-18.6	-24.0
ΔE_3	MM	0.0	-10.8	15.2	3.3	33.7	11.6	-7.4	-5.1	-7.9	-17.4
$\Delta E_{3(ZPE)}$		0.0	-9.9	13.2	2.1	34.1	13.9	-7.2	-5.0	-6.8	-15.0
ΔG_3°		0.0	-13.9	25.4	6.0	34.6	7.5	-6.2	-3.8	-12.1	-22.5
ΔE_4	A	29.4	-2.0	22.7	-4.9	-3.2	0.4	18.4	12.2	24.2	22.0
$\Delta E_{4(ZPE)}$		30.0	0.4	24.4	-2.4	-3.8	0.0	18.2	12.1	24.6	21.6
ΔG_4°		29.5	-9.3	21.5	-14.1	-1.4	0.1	20.6	12.4	24.0	23.2
ΔE_4	G	-6.8	-8.2	44.7	6.9	8.5	0.3	37.4	6.5	13.6	22.7
$\Delta E_{4(ZPE)}$		-6.3	-8.7	45.8	8.1	8.6	0.1	38.1	7.1	13.9	21.9
ΔG_4°		-5.9	-5.2	43.4	8.9	8.9	-0.2	30.0	1.2	8.6	26.5
ΔE_4	C	14.3	10.5	-0.5	-5.7	-7.0	5.7	29.1	20.9	38.5	23.0
$\Delta E_{4(ZPE)}$		15.0	14.1	-1.1	-3.8	-8.0	3.7	29.1	21.2	39.2	23.4
ΔG_4°		13.3	0.7	3.2	-11.2	-2.7	12.0	30.4	21.1	36.7	22.4
ΔE_4	T	20.0	8.8	3.9	-13.7	-6.4	0.9	27.7	22.6	34.7	30.9
$\Delta E_{4(ZPE)}$		20.7	11.6	3.5	-11.3	-6.8	0.1	27.8	22.5	35.4	30.4
ΔG_4°		19.5	0.8	1.6	-22.5	-5.2	2.9	28.4	23.5	32.8	32.0
ΔE_4	U	0.9	16.3	-10.2	-14.6	-6.8	0.0	26.5	21.8	34.8	32.3
$\Delta E_{4(ZPE)}$		2.8	19.6	-10.3	-12.3	-7.3	-0.6	26.5	21.8	35.5	31.9
ΔG_4°		-4.5	8.0	-7.5	-23.2	-5.5	1.0	27.6	23.1	32.8	33.5
ΔE_4	TAP-C^5	4.2	22.9	-24.6	5.9	16.0	20.0	16.0	19.3	-7.9	25.9
$\Delta E_{4(ZPE)}$		4.0	26.1	-23.9	9.4	15.0	18.4	14.5	19.6	-6.7	26.2
ΔG_4°		4.9	11.5	-26.3	-5.7	18.4	24.8	21.3	19.7	-10.4	25.9

399

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_4	TAP-N	-5.8	5.0	0.0	-2.7	-10.7	14.9	31.7	-5.8	9.0	15.2
$\Delta E_{4(ZPE)}$		-6.7	4.3	-1.1	-2.2	-11.7	14.0	33.4	-4.8	9.3	14.8
ΔG_4°		0.2	8.5	9.1	-5.3	-6.1	17.6	23.6	-11.6	8.9	16.1
ΔE_4	BA-C^5	-18.1	79.4	25.2	32.5	-29.6	5.1	4.0	22.2	41.3	14.0
$\Delta E_{4(ZPE)}$		-18.0	85.1	26.1	34.8	-30.8	3.6	5.0	22.4	42.3	13.4
ΔG_4°		-18.3	62.1	23.1	26.5	-24.9	9.8	4.1	21.3	40.1	15.2
ΔE_4	BA-N	-63.7	10.4	4.6	2.6	-4.2	19.3	6.0	16.1	-9.7	-5.4
$\Delta E_{4(ZPE)}$		-66.1	10.7	4.4	4.2	-4.7	23.2	6.2	16.0	-9.5	-5.0
ΔG_4°		-52.7	10.4	4.2	0.8	-2.2	9.3	7.1	16.5	-9.8	-6.8
ΔE_4	CA	-34.8	4.8	-15.2	-0.0	-4.9	242.5	3.9	14.0	-10.4	-4.0
$\Delta E_{4(ZPE)}$		-36.5	6.1	-17.7	1.4	-5.5	245.1	3.9	14.4	-10.1	-3.8
ΔG_4°		-27.9	0.3	-14.9	-1.6	-2.7	234.8	5.0	13.7	-11.0	-4.2
ΔE_4	MM	16.4	5.7	0.9	-9.3	-13.4	20.8	-9.2	-8.4	18.3	16.9
$\Delta E_{4(ZPE)}$		15.9	4.9	1.4	-8.1	-12.1	19.0	-6.5	-7.1	18.8	16.7
ΔG_4°		18.8	8.4	-0.5	-11.1	-20.0	26.2	-18.6	-14.0	17.2	17.3
ΔE_5	A	29.4	-3.8	32.2	4.8	340.4	28.9	3.5	3.5	9.7	-5.0
$\Delta E_{5(ZPE)}$		30.0	-3.2	31.8	6.6	341.9	29.1	3.8	3.8	11.4	-2.9
ΔG_5°		29.5	-3.8	37.6	2.6	342.7	27.8	5.6	4.0	3.7	-9.5
ΔE_5	G	-6.8	-12.3	10.4	-1.4	357.3	29.9	22.2	-2.2	-2.3	-6.1
$\Delta E_{5(ZPE)}$		-6.3	-12.8	12.6	-2.1	360.6	30.2	23.6	-1.4	-0.8	-4.1
ΔG_5°		-5.9	-8.8	6.6	4.7	348.3	28.9	15.3	-6.1	-12.4	-9.3
ΔE_5	C	14.3	8.7	5.9	-6.4	29.0	33.9	11.5	4.4	20.6	-3.4
$\Delta E_{5(ZPE)}$		15.0	10.3	4.6	-6.2	28.8	32.2	12.0	4.5	22.5	-1.5
ΔG_5°		13.3	7.2	12.6	-3.9	32.9	42.6	12.8	6.2	15.4	-7.3
ΔE_5	T	20.0	8.4	21.9	-15.0	28.3	35.6	8.8	4.0	13.8	-5.2
$\Delta E_{5(ZPE)}$		20.7	10.0	20.9	-13.5	28.7	35.0	9.2	4.3	15.5	-3.1
ΔG_5°		19.5	5.9	28.2	-17.5	29.0	39.5	10.5	4.3	8.4	-9.8
ΔE_5	U	23.4	5.6	18.5	-17.4	27.8	16.3	8.1	4.4	13.9	-4.0
$\Delta E_{5(ZPE)}$		24.1	6.6	19.1	-15.7	28.2	17.2	8.6	4.4	15.7	-2.0
ΔG_5°		22.6	4.8	17.6	-20.0	28.5	15.9	9.9	7.3	8.2	-7.6
ΔE_5	TAP-C^5	4.2	11.0	-24.2	-0.5	48.1	26.7	1.3	4.1	7.2	-13.7
$\Delta E_{5(ZPE)}$		4.0	12.3	-23.4	0.8	47.5	27.0	1.5	4.2	8.8	-10.6
ΔG_5°		4.9	6.2	-26.0	-3.2	52.1	30.0	2.0	6.5	2.8	-21.3
ΔE_5	TAP-N	-5.8	-3.1	27.2	6.3	23.4	28.8	13.8	-9.0	18.9	-0.8
$\Delta E_{5(ZPE)}$		-6.7	-3.6	27.9	6.6	23.1	29.2	15.5	-8.0	20.2	1.4
ΔG_5°		0.2	-2.1	32.4	2.9	28.5	30.6	7.5	-13.2	15.4	-5.5
ΔE_5	BA-C^5	-18.1	1.1	-48.2	16.4	-17.6	28.4	10.9	2.5	56.6	-13.9
$\Delta E_{5(ZPE)}$		-18.0	2.8	-48.3	18.1	-18.4	27.3	11.6	3.1	56.7	-11.9
ΔG_5°		-18.3	-4.1	-45.6	13.1	-16.1	31.6	9.5	2.5	55.9	-18.3
ΔE_5	BA-N	-10.8	0.2	-22.7	-1.4	33.5	39.2	-7.1	2.3	-25.5	-27.5
$\Delta E_{5(ZPE)}$		-10.9	0.5	-23.3	0.8	33.9	38.4	-6.8	2.8	-24.5	-25.0
ΔG_5°		-8.4	-1.3	-17.6	-4.8	33.3	43.5	-4.8	2.7	-28.4	-33.3
ΔE_5	CA	-8.8	4.8	-30.1	-5.6	32.2	31.9	-7.5	2.2	-25.9	-24.9
$\Delta E_{5(ZPE)}$		-8.8	6.1	-34.3	-4.0	32.6	32.4	-7.4	2.6	-24.8	-23.2
ΔG_5°		-10.4	0.3	-22.0	-8.9	32.1	31.8	-5.3	3.2	-29.6	-28.2
ΔE_5	MM	16.4	-5.1	16.1	-6.0	20.2	32.5	-16.6	-13.5	10.4	-0.5
$\Delta E_{5(ZPE)}$		15.9	-5.1	14.5	-6.0	22.0	32.9	-13.7	-12.1	12.1	1.7
ΔG_5°		18.8	-5.5	24.9	-5.2	14.6	33.7	-24.7	-17.8	5.1	-5.2
ΔE_6	A	-22.3	-4.1	-27.5	7.8	3.2	28.6	-12.1	17.1	-0.4	3.5
$\Delta E_{6(ZPE)}$		-22.6	-4.4	-29.8	6.7	3.8	29.6	-12.5	15.7	1.0	4.9
ΔG_6°		-26.2	-4.6	-23.5	14.1	1.4	30.5	-9.6	22.0	-5.1	-0.5
ΔE_6	G	15.0	2.3	-46.2	1.0	-8.5	21.1	-10.1	-1.1	-19.6	-15.3
$\Delta E_{6(ZPE)}$		14.7	5.1	-47.3	1.0	-8.6	19.4	-11.0	-4.9	-19.2	-16.0
ΔG_6°		11.6	-8.6	-44.6	-2.6	-8.9	31.0	-1.0	12.1	-16.3	-11.2
ΔE_6	C	4.7	-6.8	34.4	13.1	7.0	20.5	-1.7	13.5	-20.5	0.3
$\Delta E_{6(ZPE)}$		4.7	-8.7	34.6	14.9	8.0	20.8	-1.4	12.3	-19.8	1.5
ΔG_6°		-0.4	-2.0	34.2	9.4	2.7	22.7	-2.0	18.0	-22.1	-2.5

	RU[1]	2dRibf[4]		Ribf[5]		Tho[6]		2dRib[7]		Rib[8]	
		vac.[2]	solv.[3]	vac	solv	vac	solv	vac	solv	vac	solv
ΔE_6	T	-6.1	-6.0	35.5	5.0	6.4	19.9	-16.2	11.9	-10.6	-9.0
$\Delta E_{6(ZPE)}$		-4.9	-7.0	35.1	4.0	6.8	22.0	-16.7	10.8	-9.6	-6.9
ΔG°_6		-12.4	-3.8	40.4	7.8	5.2	15.0	-14.0	16.0	-13.7	-13.7
ΔE_6	U	12.2	-13.7	48.6	5.8	6.8	32.8	-15.7	-0.2	-10.9	-10.6
$\Delta E_{6(ZPE)}$		12.1	-15.3	48.6	4.8	7.3	32.2	-16.2	-0.6	-9.9	-8.6
ΔG°_6		10.7	-11.0	49.3	9.6	5.5	39.3	-13.5	1.5	-13.8	-15.2
ΔE_6	TAP-C^5	-13.6	1.0	26.7	2.1	-16.0	8.7	-3.0	21.8	39.6	11.7
$\Delta E_{6(ZPE)}$		-13.7	0.3	27.5	0.2	-15.0	8.3	-3.2	18.7	40.3	13.1
ΔG°_6		-17.7	5.6	27.0	10.9	-18.4	10.8	-6.4	28.9	36.2	7.1
ΔE_6	TAP-N	7.8	0.5	-5.5	15.3	10.7	2.7	-19.7	1.1	-12.9	-10.8
$\Delta E_{6(ZPE)}$		8.5	2.7	-4.5	16.7	11.7	4.5	-20.7	0.3	-11.8	-8.6
ΔG°_6		1.4	-10.9	-14.0	6.2	6.1	-1.6	-16.4	2.7	-16.4	-16.7
ΔE_6	BA-C^5	92.0	-14.0	-43.6	51.9	29.6	-36.8	-23.0	8.0	-37.2	-50.8
$\Delta E_{6(ZPE)}$		93.1	-15.4	-44.7	53.2	30.8	-39.6	-23.7	6.1	-35.8	-51.9
ΔG°_6		88.2	-8.8	-38.8	48.4	24.9	-29.2	-24.7	13.2	-41.1	-48.1
ΔE_6	BA-N	58.3	-13.0	-9.5	5.1	4.2	7.0	16.4	10.3	19.3	17.1
$\Delta E_{6(ZPE)}$		60.7	-12.9	-9.5	5.1	4.7	2.7	15.9	8.8	20.5	17.2
ΔG°_6		49.1	-15.7	-8.5	3.5	2.2	21.1	18.6	14.4	14.0	17.7
ΔE_6	CA	9.0	-2.3	12.6	14.3	4.9	-212.6	16.6	16.1	17.0	39.0
$\Delta E_{6(ZPE)}$		10.6	-2.2	13.0	15.1	5.5	-215.4	16.1	14.6	17.7	40.2
ΔG°_6		2.2	-2.3	10.6	10.9	2.7	-201.2	19.0	19.7	14.6	35.8
ΔE_6	MM	9.4	-2.9	-5.9	11.7	13.4	9.3	-15.3	14.2	-22.9	6.8
$\Delta E_{6(ZPE)}$		12.5	-0.5	-5.8	11.5	12.1	9.8	-16.2	10.2	-22.1	5.4
ΔG°_6		0.0	-12.9	-4.8	12.0	20.0	9.1	-12.6	26.6	-25.1	12.9

[1] TC = unspecified **T**rifunctional **C**onnector. [2] Differences in ΔE, $\Delta E_{(ZPE)}$, and ΔG° in vacuum at the DFT level. [3] Differences in ΔE, $\Delta E(ZPE)$, and ΔG° in solvent.

Table A3. % of free energies (ΔG°) for the different pseudorotational equilibriums (ΔG_1° and ΔG_3°) and anomer-exchange reactions (ΔG_2°, ΔG_4°, ΔG_5° and ΔG_6°) within the intrinsic error of DFT-B3LYP/6-311++G(d, p) ($-17 < \Delta G^0 < 17$) and outside the error ($\Delta G^{\circ} \geq 17$ kJ/mol) and ($\Delta G^{\circ} \leq -17$ kJ/mol).

classic pathway (a+b)

HPO_3^-

5-MR[3] / *5-MR*

	Canonical						non canonical					
	$\Delta G^{\circ} \geq 17$		$\Delta G^{\circ} \leq -17$		$-17 < \Delta G^{\circ} < 17$		$\Delta G^{\circ} \geq 17$		$\Delta G^{\circ} \leq -17$		$-17 < \Delta G^{\circ} < 17$	
	vac.[1]	solv.[2]	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1° (HPO_3^-)	6.7	6.7	0.0		73.3	93.3	16.7	11.1	16.7	0.0	66.7	88.9
ΔG_2°	26.7	6.7	40.0	0.0	33.3	93.3	27.8	22.2	38.9	5.6	33.3	72.2
ΔG_3°	40.0	0.0	13.3	0.0	46.7	100	22.2	0.0	0.0	0.0	77.8	100.0
ΔG_4°	13.3	20.0	53.3	0.0	33.3	80	22.2	11.1	38.9	0.0	38.9	88.9
ΔG_5°	40.0	0.0	46.7	0.0	13.3	100	33.3	11.1	27.8	0.0	38.9	88.9
ΔG_6°	33.3	0.0	6.7	13.3	60.0	86.7	50.0	5.6	16.7	11.1	33.3	83.3

6-MR[4] / *6-MR*

	Canonical						non canonical					
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1°	60.0	100.0	0.0	0.0	40.0	0.0	66.7	75.0	0.0	0.0	33.3	25.0
ΔG_2°	0.0	0.0	50.0	90.0	50.0	10.0	8.3	0.0	66.7	91.7	25.0	8.3
ΔG_3°	0.0	0.0	70.0	10.0	30.0	90.0	16.7	0.0	25.0	25.0	58.3	75.0
ΔG_4°	70.0	50.0	0.0	0.0	30.0	50.0	25.0	16.7	0.0	0.0	75.0	83.3
ΔG_5°	10.0	0.0	0.0	0.0	90.0	100.0	33.3	0.0	0.0	8.3	66.7	91.7
ΔG_6°	20.0	60.0	30.0	0.0	50.0	40.0	58.3	58.3	8.3	0.0	33.3	41.7

$HAsO_3^-$

5-MR / *5-MR*

	$\Delta G^{\circ} \geq 17$		$\Delta G^{\circ} \leq -17$		$-17 < \Delta G^{\circ} < 17$		$\Delta G^{\circ} \geq 17$		$\Delta G^{\circ} \leq -17$		$-17 < \Delta G^{\circ} < 17$	
	Vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1°	26.7	6.7	6.7	0.0	66.7	93.3	22.2	11.1	16.7	5.6	61.1	83.3
ΔG_2°	26.7	13.3	46.7	0.0	26.7	86.7	33.3	22.2	38.9	0.0	27.8	77.8
ΔG_3°	33.3	6.7	6.7	0.0	60.0	93.3	5.6	11.1	0.0	0.0	94.4	88.9

	Canonical						non canonical					
	$\Delta G^\circ \geq 17$		$\Delta G^\circ \leq -17$		$-17 < \Delta G^\circ < 17$		$\Delta G^\circ \geq 17$		$\Delta G^\circ \leq -17$		$-17 < \Delta G^\circ < 17$	
	vac.[1]	solv.[2]	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_4°	20.0	20.0	40.0	0.0	40.0	80.0	22.2	22.2	27.8	11.1	50.0	66.7
ΔG_5°	46.7	6.7	33.3	0.0	20.0	93.3	27.8	27.8	22.2	5.6	50.0	66.7
ΔG_6°	26.7	6.7	0.0	13.3	73.3	80.0	38.9	0.0	27.8	11.1	33.3	88.9

	6-MR						*6-MR*					
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1°	60.0	90.0	0.0	0.0	40.0	10.0	66.7	83.3	0.0	0.0	33.3	16.7
ΔG_2°	0.0	0.0	40.0	80.0	60.0	20.0	8.3	0.0	75.0	91.7	16.7	8.3
ΔG_3°	0.0	0.0	50.0	20.0	50.0	80.0	0.0	0.0	33.3	16.7	66.7	83.3
ΔG_4°	90.0	70.0	0.0	0.0	10.0	30.0	33.3	16.7	0.0	0.0	66.7	83.3
ΔG_5°	10.0	0.0	0.0	0.0	90.0	100.0	25.0	0.0	0.0	0.0	75.0	100.0
ΔG_6°	20.0	40.0	30.0	0.0	50.0	60.0	66.7	58.3	8.3	0.0	25.0	41.7

alternative pathway (c+d)

HPO_3^-

	Canonical						non canonical					
	5-MR						*5-MR*					
	$\Delta G^\circ \geq 17$		$\Delta G^\circ \leq -17$		$-17 < \Delta G^\circ < 17$		$\Delta G^\circ \geq 17$		$\Delta G^\circ \leq -17$		$-17 < \Delta G^\circ < 17$	
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1°	0.0	33.3	6.7	6.7	93.3	60.0	0.0	16.7	5.6	0.0	94.4	83.3
ΔG_2°	20.0	0.0	6.7	13.3	73.3	86.7	38.9	5.6	22.2	22.2	38.9	72.2
ΔG_3°	26.7	0.0	13.3	6.7	60.0	93.3	27.8	11.1	11.1	0.0	61.1	88.9
ΔG_4°	13.3	6.7	13.3	0.0	73.3	93.3	22.2	5.6	16.7	16.7	61.1	77.8
ΔG_5°	26.7	13.3	6.7	6.7	66.7	80.0	27.8	22.2	16.7	16.7	55.6	61.1
ΔG_6°	6.7	33.3	13.3	0.0	80.0	66.7	16.7	22.2	22.2	5.6	61.1	72.2

	6-MR						*6-MR*					
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
ΔG_1°	20.0	50.0	0.0	0.0	80.0	50.0	25.0	41.7	0.0	8.3	75.0	50.0
ΔG_2°	0.0	0.0	20.0	90.0	80.0	10.0	0.0	0.0	33.3	58.3	66.7	41.7
ΔG_3°	0.0	0.0	80.0	50.0	20.0	50.0	0.0	0.0	33.3	33.3	66.7	66.7

	Canonical						non canonical					
	$\Delta G° \geq 17$		$\Delta G° \leq -17$		$-17 < \Delta G° < 17$		$\Delta G° \geq 17$		$\Delta G° \leq -17$		$-17 < \Delta G° < 17$	
	vac.[1]	solv.[2]	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
$\Delta G_4°$	80.0	30.0	0.0	0.0	20.0	70.0	25.0	25.0	0.0	0.0	75.0	75.0
$\Delta G_5°$	0.0	0.0	10.0	50.0	90.0	50.0	8.3	0.0	16.7	25.0	75.0	75.0
$\Delta G_6°$	0.0	0.0	40.0	0.0	60.0	100.0	25.0	33.3	33.3	8.3	41.7	58.3

$HAsO_3^-$

	5-MR						5-MR					
	$\Delta G° \geq 17$		$\Delta G° \leq -17$		$-17 < \Delta G° < 17$		$\Delta G° \geq 17$		$\Delta G° \leq -17$		$-17 < \Delta G° < 17$	
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
$\Delta G_1°$	20.0	33.3	0.0	0.0	80.0	66.7	11.1	38.9	5.6	5.6	83.3	55.6
$\Delta G_2°$	46.7	6.7	13.3	6.7	40.0	86.7	38.9	5.6	22.2	16.7	38.9	77.8
$\Delta G_3°$	53.3	26.7	6.7	0.0	40.0	73.3	50.0	11.1	11.1	11.1	38.9	77.8
$\Delta G_4°$	26.7	0.0	0.0	13.3	73.3	86.7	16.7	33.3	33.3	0.0	50.0	66.7
$\Delta G_5°$	73.3	26.7	0.0	13.3	26.7	60.0	38.9	33.3	27.8	0.0	33.3	66.7
$\Delta G_6°$	20.0	26.7	20.0	0.0	60.0	73.3	27.8	11.1	16.7	11.1	55.6	77.8

	6-MR						6-MR					
	vac.	solv.	vac.	solv.	vac	solv	vac	solv	vac	solv	vac	solv
$\Delta G_1°$	50.0	80.0	0.0	0.0	50.0	20.0	25.0	58.3	16.7	8.3	58.3	33.3
$\Delta G_2°$	0.0	0.0	0.0	100.0	100.0	0.0	25.0	0.0	41.7	75.0	33.3	25.0
$\Delta G_3°$	0.0	0.0	80.0	60.0	20.0	40.0	0.0	0.0	25.0	58.3	75.0	41.7
$\Delta G_4°$	90.0	80.0	0.0	0.0	10.0	20.0	33.3	33.3	8.3	0.0	58.3	66.7
$\Delta G_5°$	0.0	0.0	0.0	0.0	100.0	100.0	8.3	0.0	25.0	41.7	66.7	58.3
$\Delta G_6°$	0.0	20.0	10.0	0.0	90.0	80.0	25.0	41.7	25.0	8.3	50.0	50.0

[1] Differences in $\Delta G°$ in vacuum at the DFT level. [2] Differences in $\Delta G°$ in solvent. [3]Nucleotides containing a TC that is a sugar with the hemiacetal 5-**M**ember **R**ing (MR) or furanose form. [4]Nucleotides containing a TC that is a sugar with the hemiacetal as a 6-**M**ember **R**ing (MR) or pyranose form

Table A4. Gibbs (ΔG°) energies at standard pressure and temperature in kJ/mol for a hypothetical classic (a+b) and alternative (c+d) pathway leading to the 5 canonical and 6 non-canonical β- and α-, HPO_3^-- and $HAsO_3^-$-nucleotides in vacuum and aqueous environment (From DFT calculations at the B3LYP/6-311++G (d, p) level of theory, with aqueous solvation modeled with the IEFPCM model).

RU[1]	vac.[12]						ΔΔG°_a+b [11]	solv.[13]						ΔΔG°_a+b
	ΔG°_a		ΔG°_b		ΔG°_{a+b}			ΔG°_a		ΔG°_b		ΔG°_{a+b}		
	α	β	α	β	α	β		α	β	α	β	α	β	
							classic pathway (a+b)							
							HPO_3^-							
	DNA (2dRibf[2]-RUs)							**DNA (2dRibf-RUs)**						
	$^2T_3{}^{(7)}$							$^2T_3$						
A	8.4	-0.8	-64.4	-58.6	-56.0	-59.4	-3.4	2.0	-3.4	5.1	-0.8	7.1	-4.3	-11.4
G	-1.0	-3.5	-35.9	-77.0	-37.0	-80.6	-43.6	-0.1	0.9	4.4	-8.1	4.3	-7.2	-11.5
C	0.5	2.5	-22.4	-87.0	-21.9	-84.5	-62.6	6.4	12.2	4.4	1.5	10.8	13.7	2.9
T	8.8	-4.5	-37.5	-63.2	-28.8	-67.7	-38.9	10.6	0.2	4.4	3.3	15.0	3.5	-11.5
U	9.1	-6.3	-41.1	-62.6	-32.0	-68.8	-36.8	11.1	1.1	5.2	1.0	16.4	2.1	-14.3
TAP-C⁵	14.8	-10.3	-40.0	-39.8	-25.2	-50.2	-25.0	5.7	-11.2	6.5	3.5	12.3	-7.7	-20.0
TAP-N	7.6	-10.2	-50.7	-14.7	-43.1	-24.9	18.2	-0.6	-11.9	7.8	9.7	7.2	-2.3	-9.5
BA-C⁵	-17.8	-35.7	-110.8	-111.2	-128.6	-146.9	-18.3	-9.8	-8.6	2.4	2.5	-7.5	-6.1	1.4
BA-N	15.1	9.9	-5.6	-77.9	9.5	-68.0	-77.5	16.7	16.3	4.4	-3.7	21.1	12.6	-8.5
CA	16.2	10.8	-22.8	-53.6	-6.6	-42.8	-36.2	17.9	15.3	4.4	5.0	22.2	20.4	-1.8
MM	6.8	6.9	-2.3	-43.9	4.4	-37.0	-41.4	3.5	10.6	7.7	-5.3	11.2	5.2	-6.0
	$^3T_2{}^{(8)}$							$^3T_2$						
A	-11.9	-5.8	-18.3	-45.3	-30.3	-51.0	-20.7	-18.8	-9.2	4.6	7.3	-14.2	-2.0	12.2
G	-10.0	-11.1	-35.7	-64.0	-45.7	-75.1	-29.4	-9.4	-6.9	4.4	1.1	-5.0	-5.8	-0.8
C	-7.3	-2.4	-19.2	-50.1	-26.5	-52.5	-26.0	-4.1	-0.1	11.3	5.3	7.1	5.2	-1.9
T	1.0	-2.2	-34.7	-77.2	-33.7	-79.5	-45.8	2.2	0.1	4.5	9.3	6.6	9.5	2.9
U	1.5	-2.7	-44.7	-86.8	-43.3	-89.5	-46.2	1.2	1.5	7.1	-9.4	8.3	-7.9	-16.2
TAP-C⁵	-2.4	5.3	-39.5	-69.0	-41.9	-63.6	-21.7	-7.2	-10.2	13.2	3.5	6.0	-6.7	-12.7

RU[1]	vac.[12]							solv.[13]						
	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^\circ_{a+b}$ [11]	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^\circ_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
TAP-N	0.5	-10.8	-54.8	-28.2	-54.2	-39.0	15.2	-9.8	-9.9	7.8	9.6	-2.0	-0.3	1.7
BA-C⁵	-34.3	-43.3	-87.6	-119.8	-121.9	-163.1	-41.2	-12.8	-15.1	8.3	7.0	-4.5	-8.1	-3.6
BA-N	6.4	7.7	-4.3	-72.9	2.1	-65.2	-67.3	12.3	10.3	-4.6	6.7	7.7	17.0	9.3
CA	8.1	8.8	-24.2	-40.3	-16.1	-31.5	-15.4	8.13	9.6	-24.2	7.4	-16.1	17.0	4.0
MM	-4.6	-2.2	-0.4	-56.1	-5.0	-58.3	-53.3	69.7	-10.5	-56.7	8.5	13.0	-2.0	-0.2
	RNA (Ribf[3]-RUs)							**RNA (Ribf-RUs)**						
	$^2T_3$							$^2T_3$						
A	0.2	-11.5	-36.6	-33.9	-36.5	-45.4	-8.9	-1.3	-18.6	4.4	15.1	3.1	-3.5	-6.6
G	3.5	-13.7	-40.9	-65.0	-37.4	-78.6	-41.2	0.0	-14.8	5.9	8.4	6.0	-6.4	-12.4
C	-8.6	-16.5	-27.9	-48.6	-36.5	-65.1	-28.6	85.6	-2.6	-70.0	1.8	15.5	-0.8	-16.3
T	4.2	-2.7	-50.9	-59.3	-46.7	-62.0	-15.3	15.6	2.8	3.1	-2.8	18.7	0.0	-18.7
U	3.6	-2.4	-54.1	-59.2	-50.5	-61.6	-11.1	18.4	2.2	4.2	-0.9	22.7	1.3	-21.4
TAP-C⁵	12.5	-4.1	-36.5	-56.9	-24.0	-61.0	-37.0	9.0	-6.6	-1.4	-3.0	7.6	-9.6	-17.2
TAP-N	-3.5	-4.8	-61.7	-19.2	-65.2	-23.9	41.3	2.7	-15.0	8.2	5.8	10.8	-9.2	-20.0
BA-C⁵	-32.7	-50.7	-121.3	-103.4	-154.0	-154.1	-0.1	-8.2	-26.5	3.7	14.2	-4.4	-12.3	-7.9
BA-N	16.2	11.6	-33.8	-67.7	-17.6	-56.1	-38.5	20.1	2.6	6.8	7.8	26.9	10.4	-16.5
CA	20.5	10.7	-63.0	-55.8	-42.5	-45.1	-2.6	26.8	9.1	2.7	1.1	29.5	10.2	-19.3
MM	2.4	-6.3	-4.5	-29.5	-2.1	-35.8	-33.7	-8.7	-6.9	7.0	9.3	-1.7	2.4	4.1
	$^3T_2$							$^3T_2$						
A	0.9	-14.8	-24.8	-34.6	-23.9	-49.4	-25.5	-2.9	-9.8	5.1	5.7	2.2	-4.1	-6.3
G	2.9	-10.5	-76.2	-27.2	-73.3	-37.7	35.6	-2.7	-8.5	8.4	1.8	5.7	-6.7	-12.4
C	-4.0	-17.8	-30.2	-41.8	-34.2	-59.6	-25.4	9.2	-5.3	6.0	2.2	15.2	-3.1	-18.3
T	8.9	-3.3	-40.5	-53.1	-31.6	-56.4	-24.8	15.6	3.1	7.0	6.0	22.6	9.1	-13.5
U	8.4	-3.4	-94.1	-61.9	-85.7	-65.3	20.4	16.6	3.6	5.8	2.1	22.4	5.7	-16.7
TAP-C⁵	-3.0	-8.0	-12.8	-56.9	-15.8	-65.0	-49.2	15.8	-15.2	-2.3	4.4	13.5	-10.9	-24.4
TAP-N	8.7	-3.0	-52.9	-75.3	-44.3	-78.3	-34.0	-0.7	-11.3	10.9	6.7	10.2	-4.6	-14.8
BA-C⁵	-36.0	-48.1	-96.3	-112.2	-132.3	-160.4	-28.1	-14.7	-22.1	7.6	8.0	-7.2	-14.1	-6.9
BA-N	18.0	11.3	3.3	-76.5	21.3	-65.2	-86.5	18.4	6.7	8.2	4.1	26.6	10.8	-15.8
CA	22.6	12.0	-51.7	-52.4	-29.1	-40.4	-11.3	23.0	7.0	6.2	6.2	29.2	13.2	-16.0

RU[1]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	vac.[12]							solv.[13]						
	α	β	α	β	α	β		α	β	α	β	α	β	
MM	2.9	-13.9	-33.8	-20.7	-30.9	-34.6	-3.7	-2.4	-11.4	7.3	1.6	4.9	-9.8	-14.7
TNA (Tho[4]-RUs) $^2T_3$								**TNA (Tho-RUs)** $^2T_3$						
A	18.8	-4.4	-43.9	-32.6	-25.1	-37.0	-11.9	-5.7	-39.7	-5.4	-2.7	-11.1	-42.4	-31.4
G	15.6	-3.4	-87.4	-42.5	-71.8	-45.9	25.9	-2.4	-41.1	10.7	-1.1	8.3	-42.2	-50.5
C	3.0	-3.4	-42.5	-21.3	-39.5	-24.7	14.8	7.8	-30.7	-8.5	0.9	-0.8	-29.8	-29.0
T	14.6	6.6	-70.4	-41.2	-55.8	-34.6	21.2	12.9	-28.3	-10.3	-5.4	2.5	-33.7	-36.2
U	15.4	7.3	-74.3	-44.6	-58.9	-37.3	21.6	13.4	-28.8	-13.7	-6.8	-0.3	-35.7	-35.4
TAP-C[5]	12.7	-10.0	-36.0	-56.7	-23.3	-66.7	-43.4	-0.3	-34.3	6.4	0.4	6.1	-33.8	-40.0
TAP-N	10.6	-7.5	-37.6	-49.1	-27.0	-56.6	-29.6	-1.6	-44.3	-16.6	-1.0	-18.2	-45.3	-27.1
BA-C[5]	-37.1	-48.5	-57.6	-59.3	-94.7	-107.8	-13.1	-14.5	-52.1	-9.1	0.6	-23.5	-51.5	-28.0
BA-N	22.7	11.8	-45.8	-11.7	-23.1	0.1	23.2	17.3	-20.5	13.9	-1.2	31.2	-21.7	-52.8
CA	29.5	13.3	-75.2	-31.3	-45.7	-18.0	27.7	19.2	-19.9	-14.4	-0.4	4.9	-20.3	-25.1
MM	6.7	2.0	-52.4	-12.4	-45.7	-10.4	35.3	2.5	-49.2	13.8	0.5	16.3	-48.7	-65.0
$^3T_2$								$^3T_2$						
A	-10.6	-3.6	-55.8	-49.3	-66.3	-53.0	13.3	-8.1	-6.9	-10.7	9.0	-18.8	2.1	20.9
G	-6.2	1.7	-67.4	-71.7	-73.6	-70.1	3.5	-4.6	-2.9	-14.1	2.6	-18.8	-0.2	18.5
C	-16.5	-7.9	-21.4	-51.3	-37.9	-59.2	-21.3	-6.6	-2.1	0.7	4.1	-5.9	2.1	8.0
T	-4.1	3.1	-66.7	-75.1	-70.8	-72.0	-1.2	-1.5	4.8	-4.2	6.2	-5.7	11.0	16.6
U	-4.5	2.6	-69.0	-77.6	-73.5	-75.0	-1.5	-0.6	5.6	-5.6	-6.8	-6.2	-1.2	5.0
TAP-C[5]	-8.3	-5.1	-70.6	-51.9	-78.9	-57.0	21.9	-14.0	0.2	-14.4	0.4	-28.4	0.6	29.0
TAP-N	-1.1	2.2	-60.1	-49.1	-61.2	-46.9	14.3	-9.7	-12.5	-14.2	-14.7	-23.9	-27.2	-3.3
BA-C[5]	-49.7	-38.8	-46.8	-59.3	-96.5	-98.2	-1.7	-27.8	-17.6	0.5	0.6	-27.3	-17.0	10.3
BA-N	20.3	19.2	-46.4	-9.5	-26.0	9.8	35.8	3.5	14.0	-1.2	-1.2	2.3	12.8	10.5
CA	19.7	19.6	-66.9	-24.5	-47.1	-4.9	42.2	5.3	14.6	-6.3	-0.4	-1.0	14.2	15.2
MM	-5.7	-1.3	-45.8	-53.7	-51.4	-55.0	-3.6	-11.8	-14.8	14.8	0.5	3.1	-14.2	-17.3

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
	p-DNA (2dRib[5]-RUs)							**p-DNA (2dRib-RUs)**						
	$^1C_4$ [9]							$^1C_4$						
A	1.2	12.9	-47.0	-47.6	-45.9	-34.7	11.2	1.0	8.2	-6.3	-8.8	-5.3	-0.5	4.8
G	-0.1	12.2	-60.0	-58.5	-60.1	-46.3	13.8	-0.8	6.9	1.9	-3.5	1.0	3.3	2.3
C	6.6	21.6	-58.1	-30.7	-51.5	-9.1	42.4	15.1	23.3	-4.2	-7.3	10.9	16.1	5.2
T	2.5	26.2	-64.3	-49.9	-61.9	-23.7	38.2	6.4	24.2	-9.4	-8.2	-3.0	16.0	19.0
U	1.8	26.6	-65.7	-53.6	-63.9	-27.0	36.9	7.0	23.5	-9.4	-4.1	-2.4	19.4	21.8
TAP-C[5]	-4.4	20.7	-44.0	-44.1	-48.4	-23.5	24.9	-5.2	24.0	-6.4	-5.0	-11.5	18.9	30.4
TAP-N	-6.0	1.5	-41.6	-41.8	-47.7	-40.3	7.4	-8.2	-2.3	-5.6	-3.3	-13.8	-5.6	8.2
BA-C[5]	-35.2	-10.7	-102.4	-53.8	-137.6	-64.6	73.0	-11.1	7.3	0.5	2.3	-10.7	9.6	20.3
BA-N	11.2	33.8	-15.4	-29.6	-4.3	4.1	8.4	21.1	35.2	-4.5	-3.0	16.6	32.3	15.7
CA	11.7	33.2	-34.5	-49.5	-22.8	-16.2	6.6	17.1	33.4	-4.8	-5.8	12.2	27.6	15.4
MM	8.3	7.1	-45.6	-35.5	-37.3	-28.5	8.8	-7.9	0.8	-4.5	-5.9	-12.5	-5.1	7.4
	$^4C_1$ [10]							$^4C_1$						
A	6.9	-3.9	-47.0	-53.2	-40.2	-57.1	-16.9	2.6	-8.2	19.4	2.8	22.0	-5.4	-27.4
G	10.2	-5.5	-38.4	-57.2	-28.2	-62.7	-34.5	10.4	-6.6	15.6	-0.2	25.9	-6.8	-32.7
C	14.2	-5.6	-52.8	-43.5	-38.6	-49.1	-10.5	20.9	2.3	14.7	-0.3	35.6	2.0	-33.6
T	22.2	-1.7	-70.8	-59.3	-48.7	-61.0	-12.3	28.3	2.7	12.5	-0.3	40.8	2.4	-38.4
U	22.9	-1.8	-74.9	-57.9	-51.9	-59.8	-7.9	26.6	2.4	14.7	0.8	41.3	3.3	-38.0
TAP-C[5]	27.0	-9.2	16.9	-45.6	43.9	-54.8	-98.7	27.4	-8.0	15.6	0.7	43.0	-7.3	-50.3
TAP-N	12.1	-4.5	-58.6	-47.9	-46.6	-52.4	-5.8	9.4	-9.4	-13.5	1.6	-4.1	-7.7	-3.6
BA-C[5]	-21.7	-43.8	-35.6	-54.3	-57.3	-98.1	-40.8	5.7	-10.4	20.0	-0.3	25.8	-10.7	-36.5
BA-N	31.7	17.0	1.7	-31.7	33.4	-14.7	-48.1	31.6	16.0	17.1	2.6	48.7	18.6	-30.1
CA	34.6	15.4	-31.8	-45.8	2.8	-30.4	-33.2	33.5	16.9	16.2	1.3	49.7	18.2	-31.4
MM	16.1	-3.2	-12.2	-39.0	3.9	-42.3	-46.2	16.6	-7.8	17.1	1.1	33.7	-6.7	-40.4
	p-RNA (Rib[6]-RUs)							**p-RNA (Rib-RUs)**						
	$^1C_4$							$^1C_4$						
A	3.5	14.1	-61.3	-43.2	-57.8	-29.1	28.7	2.6	12.0	3.3	5.7	5.9	17.7	11.8
G	3.5	13.9	-62.5	-55.7	-59.0	-41.8	17.2	1.3	14.1	5.4	4.8	6.7	18.9	12.2

RU[1]	vac.[12] ΔG°a		ΔG°b		ΔG°a+b		ΔΔG°a+b [11]	solv.[13] ΔG°a		ΔG°b		ΔG°a+b		ΔΔG°a+b
	α	β	α	β	α	β		α	β	α	β	α	β	
C	0.5	23.3	-58.0	-50.3	-57.5	-27.0	30.5	8.6	29.6	4.8	3.0	13.4	32.7	19.2
T	5.9	27.8	-72.8	-63.0	-66.9	-35.2	31.7	12.2	32.1	2.9	10.5	15.1	42.6	27.5
U	5.4	28.1	-75.4	-74.3	-70.0	-46.2	23.8	10.5	32.0	3.9	4.5	14.4	36.6	22.1
TAP-C^5	2.4	13.9	-49.4	-72.0	-46.9	-58.2	-11.3	7.8	14.5	-5.2	-5.4	2.6	9.1	6.5
TAP-N	-2.7	3.5	-45.9	-42.9	-48.7	-39.4	9.3	-6.2	0.6	5.8	5.4	-0.4	6.1	6.4
BA-C^5	-35.6	-14.9	-109.0	-46.4	-144.6	-61.3	83.3	1.4	12.1	12.7	5.8	14.1	17.9	3.8
BA-N	29.9	38.1	-15.4	-33.0	14.5	5.2	-9.3	32.4	39.3	15.7	-0.1	48.0	39.2	-8.8
CA	31.7	38.0	-37.2	-56.4	-5.5	-18.5	-13.0	35.4	37.6	9.7	2.7	45.1	40.3	-4.8
MM	-5.6	9.0	-37.4	-34.7	-43.0	-25.7	17.3	-6.3	2.1	6.6	6.2	0.3	8.3	8.0
$^4C_1$								$^4C_1$						
A	8.9	-9.7	-37.8	-22.5	-28.9	-32.2	-3.3	12.0	-9.8	11.9	9.3	23.9	-0.5	-24.4
G	16.2	-3.0	-50.4	-67.5	-34.3	-70.5	-36.2	15.4	-6.9	4.6	14.0	20.0	7.1	-12.9
C	13.3	-11.0	-37.5	-40.2	-24.2	-51.2	-27.0	28.8	-0.7	13.6	9.7	42.4	9.0	-33.4
T	27.0	4.4	-32.2	-70.0	-5.2	-65.7	-60.5	37.3	1.2	12.8	8.7	50.0	10.0	-40.1
U	28.0	4.2	-36.5	-72.7	-8.5	-68.5	-60.0	36.2	2.7	16.7	11.7	52.9	14.4	-38.5
TAP-C^5	24.1	-9.0	-62.8	-30.0	-38.8	-39.0	-0.2	28.0	-14.0	16.1	11.2	44.1	-2.8	-46.9
TAP-N	0.7	-4.1	-59.9	-25.3	-59.3	-29.4	29.9	-1.1	-10.4	22.3	-2.6	21.2	-13.0	-34.1
BA-C^5	-20.2	-43.2	-39.5	-53.2	-59.6	-96.4	-36.8	12.8	-12.3	16.8	10.9	29.6	-1.4	-31.0
BA-N	34.9	23.5	-0.8	-11.0	34.0	12.5	-21.5	38.8	11.9	10.8	13.1	49.7	25.0	-24.7
CA	38.4	23.1	-30.0	-34.4	8.3	-11.3	-19.6	43.7	14.2	10.9	-3.5	54.6	10.7	-43.8
MM	16.8	-7.2	-24.9	11.2	-8.0	4.1	12.1	16.2	-9.3	11.4	12.9	27.6	3.6	-24.1
GNA (glycerol-RU)								**GNA (glycerol-RU)**						
A	-27.1		-53.2		-80.3		-	-28.6		3.2		-25.4		-
G	-25.7		-51.4		-77.1		-	-24.8		1.3		-23.5		-
C	-16.3		-46.3		-62.7		-	-17.8		1.7		-16.2		-
T	-16.8		-68.7		-85.5		-	-12.9		-0.6		-13.5		-
U	-16.4		-72.4		-88.8		-	-11.6		1.4		-10.2		-
TAP-C^5	-19.1		-61.5		-80.7		-	-32.5		1.5		-31.0		-

RU[1]	vac.[12]				solv.[13]			
	ΔG_a°	ΔG_b°	ΔG_{a+b}°	$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°	ΔG_b°	ΔG_{a+b}°	$\Delta\Delta G_{a+b}^\circ$
TAP-N	-9.9	-36.4	-46.3	-	-9.3	-3.7	-13.1	-
BA-C[5]	-55.6	-121.6	-177.3	-	-42.8	2.0	-40.8	-
BA-N	-13.9	-47.1	-61.0	-	-13.3	2.0	-11.3	-
CA	-12.7	-47.6	-60.3	-	-8.9	1.4	-7.5	-
MM	-9.6	-31.2	-40.8	-	-19.7	1.8	-17.9	-

	GNA (glyceric acid-RU)				GNA (glyceric acid-RU)			
A	44.7	-20.5	24.2	-	47.5	-6.1	41.4	-
G	46.0	-22.7	23.3	-	48.2	-4.3	44.0	-
C	45.4	-22.9	22.4	-	59.7	-3.7	55.9	-
T	53.4	-40.2	13.1	-	63.3	-4.1	59.2	-
U	55.5	-44.9	10.7	-	67.0	-6.8	60.2	-
TAP-C[5]	27.9	-30.3	-2.4	-	28.4	0.3	28.7	-
TAP-N	7.4	-42.1	-34.7	-	5.9	-16.8	-10.9	-
BA-C[5]	-9.1	-35.1	-44.2	-	44.9	0.0	44.9	-
BA-N	90.0	-111.2	-21.2	-	84.7	-8.7	76.0	-
CA	94.9	-60.4	34.5	-	90.3	-12.0	78.2	-
MM	27.5	5.9	33.4	-	21.3	1.1	22.4	-

	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
UMP	-5.2	-9.8	-9.2	3.7	-14.4	-6.0	8.4	1.8	-0.9	8.1	7.0	9.9	6.1	-3.8

	HAsO₃⁻													
	α	β	α	β	α	β		α	β	α	β	α	β	
	DNA (2dRibf-RUs)							DNA (2dRibf-RUs)						
	$^2T_3$							$^2T_3$						
A	8.4	-0.8	-55.3	-50.1	-46.9	-50.9	-4.0	2.0	-3.4	5.2	14.2	7.2	10.8	3.6
G	-1.0	-3.5	-30.4	-73.5	-31.4	-77.1	-45.6	-0.1	0.9	10.5	-3.0	10.3	-2.2	-12.5
C	0.5	2.5	-17.1	-59.1	-16.6	-56.6	-39.9	6.4	12.2	8.3	7.0	14.7	19.2	4.5
T	8.8	-4.5	-32.3	-63.6	-23.5	-68.1	-44.6	10.6	0.2	9.0	17.0	19.6	17.2	-2.4
U	9.1	-6.3	-34.4	-71.8	-25.4	-78.0	-52.7	11.1	1.1	7.1	-5.2	18.2	-4.1	-22.3
TAP-C[5]	14.8	-10.3	-41.3	-48.3	-26.5	-58.6	-32.1	5.7	-11.2	0.9	5.1	6.7	-6.1	-12.8

RU[1]	vac.[12]							solv.[13]						
	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$ [11]	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
TAP-N	7.6	-10.2	-44.4	-26.2	-36.8	-36.4	0.4	-0.6	-11.9	16.3	9.9	15.7	-2.0	-17.7
BA-C^5	-17.8	-35.7	-68.8	-110.3	-86.6	-146.0	-59.4	-9.8	-8.6	14.1	0.1	4.3	-8.5	-12.8
BA-N	15.1	9.9	0.4	-59.3	15.5	-49.4	-64.9	16.7	16.3	4.8	1.6	21.5	17.8	-3.7
CA	16.2	10.8	-18.6	-41.2	-2.5	-30.3	-27.9	17.9	15.3	4.2	4.0	22.0	19.4	-2.6
MM	6.8	6.9	-1.6	-28.8	5.2	-21.9	-27.1	3.5	10.6	5.3	-7.1	8.7	3.4	-5.3
					$^3T_2$							$^3T_2$		
A	-11.9	-5.8	-17.3	-56.6	-29.2	-62.4	-33.1	-18.8	-9.2	9.0	8.2	-9.8	-1.0	8.8
G	-10.0	-11.1	-30.9	-56.3	-40.9	-67.5	-26.6	-9.4	-6.9	8.7	6.5	-0.7	-0.4	0.3
C	-7.3	-2.4	-18.8	-42.9	-26.1	-45.3	-19.2	-4.1	-0.1	13.9	2.6	9.7	2.5	-7.2
T	1.0	-2.2	-33.9	-71.4	-33.0	-73.7	-40.7	2.2	0.1	5.6	6.3	7.7	6.4	-1.3
U	1.5	-2.7	-36.3	-79.6	-34.8	-82.3	-47.5	1.2	1.5	-5.4	8.5	-4.1	10.0	14.1
TAP-C^5	-2.4	5.3	-35.9	-62.8	-38.3	-57.5	-19.2	-7.2	-10.2	12.7	2.3	5.4	-7.8	-13.3
TAP-N	0.5	-10.8	-45.3	-16.6	-44.8	-27.4	17.4	-9.8	-9.9	15.9	2.1	6.1	-7.9	-14.0
BA-C^5	-34.3	-43.3	-86.1	-97.3	-120.4	-140.6	-20.3	-12.8	-15.1	3.4	7.3	-9.5	-7.8	1.6
BA-N	6.4	7.7	-0.4	-68.5	6.0	-60.8	-66.8	12.3	10.3	0.1	3.4	12.5	13.7	1.2
CA	8.1	8.8	-20.1	-48.2	-11.9	-39.4	-27.5	69.7	9.6	-58.4	9.6	11.8	19.2	6.4
MM	-4.6	-2.2	0.4	-19.2	-4.3	-21.5	-17.2	-10.9	-10.5	19.8	2.8	8.9	-7.7	-16.6
			RNA (Ribf-RUs)							**RNA (Ribf-RUs)**				
			$^2T_3$							$^2T_3$				
A	0.2	-11.5	-21.9	-13.1	-21.7	-24.6	-2.9	-1.3	-18.6	20.7	18.8	19.5	0.2	-19.3
G	3.5	-13.7	-34.0	-56.4	-30.5	-70.1	-39.6	0.0	-14.8	1.6	2.0	1.7	-12.8	-14.4
C	-8.6	-16.5	-16.7	-42.6	-25.2	-59.1	-33.9	85.6	-2.6	-70.5	3.5	15.0	0.9	-14.1
T	4.2	-2.7	-33.6	-48.9	-29.4	-51.6	-22.2	15.6	2.8	6.1	4.7	21.7	7.5	-14.2
U	3.6	-2.4	-36.6	-49.1	-33.0	-51.4	-18.4	18.4	2.2	11.4	4.3	29.8	6.5	-23.3
TAP-C^5	12.5	-4.1	-57.0	-71.9	-44.5	-76.1	-31.6	9.0	-6.6	-0.5	-5.8	8.4	-12.4	-20.8
TAP-N	-3.5	-4.8	-55.3	-7.7	-58.8	-12.5	46.3	2.7	-15.0	4.8	11.4	7.5	-3.6	-11.1
BA-C^5	-32.7	-50.7	-113.4	-110.5	-146.1	-161.2	-15.1	-8.2	-26.5	4.5	11.9	-3.7	-14.6	-11.0
BA-N	16.2	11.6	-107.3	-65.7	-91.1	-54.1	37.0	20.1	2.6	6.7	11.1	26.8	13.7	-13.0

RU[1]	vac.[12]							solv.[13]						
	ΔG°_{a}		ΔG°_{b}		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$ [11]	ΔG°_{a}		ΔG°_{b}		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
CA	20.5	10.7	-36.7	-49.0	-16.2	-38.3	-22.1	26.8	9.1	7.9	2.9	34.7	12.0	-22.7
MM	2.4	-6.3	1.5	-20.0	3.9	-26.2	-30.1	-8.7	-6.9	4.5	14.1	-4.2	7.2	11.4
				$^{3}T_{2}$								$^{3}T_{2}$		
A	0.9	-14.8	-9.4	-20.4	-8.6	-35.3	-26.7	-2.9	-9.8	16.8	6.6	13.9	-3.1	-17.0
G	2.9	-10.5	-79.6	-24.8	-76.8	-35.3	41.5	-2.7	-8.5	4.1	7.3	1.4	-1.2	-2.5
C	-4.0	-17.8	-15.4	-31.3	-19.4	-49.1	-29.7	9.2	-5.3	5.6	8.1	14.7	2.8	-11.9
T	8.9	-3.3	-77.6	-39.5	-68.7	-42.8	25.9	15.6	3.1	5.3	2.5	20.9	5.6	-15.4
U	8.4	-3.4	-75.5	-47.2	-67.1	-50.6	16.5	16.6	3.6	6.1	-0.5	22.7	3.1	-19.6
TAP-C⁵	-3.0	-8.0	-20.1	-73.3	-23.0	-81.3	-58.3	15.8	-15.2	3.9	3.3	19.7	-11.9	-31.7
TAP-N	8.7	-3.0	-30.5	-18.0	-21.9	-20.9	0.9	-0.7	-11.3	11.2	5.5	10.4	-5.8	-16.2
BA-C⁵	-36.0	-48.1	-113.6	-107.3	-149.6	-155.4	-5.9	-14.7	-22.1	5.5	3.7	-9.2	-18.5	-9.2
BA-N	18.0	11.3	18.7	-72.1	36.7	-60.9	-97.5	18.4	6.7	4.7	3.7	23.2	10.3	-12.8
CA	22.6	12.0	-34.4	-49.1	-11.8	-37.1	-25.3	23.0	7.0	5.7	12.8	28.7	19.8	-9.0
MM	2.9	-13.9	-35.7	-17.2	-32.8	-31.1	1.7	-2.4	-11.4	21.3	5.0	18.9	-6.4	-25.3
			TNA (Tho-RUs)							**TNA (Tho-RUs)**				
				$^{2}T_{3}$								$^{2}T_{3}$		
A	18.8	-4.4	-35.6	-25.8	-16.8	-30.2	-13.4	-5.7	-39.7	-8.3	-3.4	-13.9	-43.1	-29.2
G	15.6	-3.4	-81.9	-37.0	-66.3	-40.4	26.0	-2.4	-41.1	21.6	-1.0	19.2	-42.1	-61.3
C	3.0	-3.4	-38.2	-14.6	-35.1	-18.0	17.2	7.8	-30.7	-3.8	0.7	3.9	-30.0	-33.9
T	14.6	6.6	-65.6	-33.8	-51.0	-27.2	23.8	12.9	-28.3	-9.3	-6.7	3.6	-35.0	-38.6
U	15.4	7.3	-69.7	-36.8	-54.3	-29.4	24.8	13.4	-28.8	-9.0	-5.0	4.4	-33.9	-38.3
TAP-C⁵	12.7	-10.0	-25.6	-47.5	-12.9	-57.4	-44.5	-0.3	-34.3	18.1	-5.4	17.8	-39.7	-57.5
TAP-N	10.6	-7.5	-27.4	-43.8	-16.8	-51.3	-34.5	-1.6	-44.3	-15.6	14.1	-17.2	-30.2	-12.9
BA-C⁵	-37.1	-48.5	-50.6	-54.2	-87.7	-102.7	-15.0	-14.5	-52.1	-4.8	15.2	-19.3	-36.9	-17.6
BA-N	22.7	11.8	-41.3	-2.5	-18.6	9.3	27.8	17.3	-20.5	13.7	8.9	31.0	-11.6	-42.6
CA	29.5	13.3	-70.3	-22.6	-40.8	-9.3	31.5	19.2	-19.9	-11.2	11.0	8.1	-8.9	-17.0
MM	6.7	2.0	-46.7	-8.3	-39.9	-6.2	33.7	2.5	-49.2	-5.7	13.3	-3.3	-35.9	-32.6

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
			$^3T_2$							$^3T_2$				
A	-10.6	-3.6	-36.4	-44.3	-47.0	-47.9	-0.9	-8.1	-6.9	-10.6	7.2	-18.7	0.3	19.0
G	-6.2	1.7	-60.7	-66.0	-66.9	-64.3	2.6	-4.6	-2.9	-2.4	-3.9	-7.1	-6.8	0.3
C	-16.5	-7.9	-22.4	-44.9	-38.9	-52.9	-13.9	-6.6	-2.1	2.1	-1.4	-4.5	-3.5	1.0
T	-4.1	3.1	-48.2	-70.1	-52.3	-67.0	-14.7	-1.5	4.8	-3.3	6.4	-4.8	11.2	16.0
U	-4.5	2.6	-62.5	-71.2	-67.1	-68.6	-1.5	-0.6	5.6	-3.4	-5.0	-4.0	0.6	4.6
TAP-C^5	-8.3	-5.1	-64.4	-42.6	-72.8	-47.8	25.0	-14.0	0.2	-15.4	0.5	-29.3	0.7	30.0
TAP-N	-1.1	2.2	-54.0	-43.8	-55.1	-41.6	13.5	-9.7	-12.5	-13.2	-12.2	-22.9	-24.7	-1.8
BA-C^5	-49.7	-38.8	-39.8	-54.2	-89.6	-93.1	-3.5	-27.8	-17.6	0.1	15.2	-27.7	-2.4	25.2
BA-N	20.3	19.2	-40.2	-0.2	-19.8	19.0	38.8	3.5	14.0	-0.1	8.9	3.4	22.9	19.5
CA	19.7	19.6	-60.4	-19.2	-40.7	0.4	41.1	5.3	14.6	-5.7	11.0	-0.4	25.6	25.9
MM	-5.7	-1.3	-39.7	-46.4	-45.4	-47.7	-2.3	-11.8	-14.8	15.5	13.3	3.7	-1.4	-5.1
p-DNA (2dRib-RUs)			$^1C_4$					**p-DNA (2dRib-RUs)**		$^1C_4$				
A	1.2	12.9	-39.0	-38.8	-37.8	-25.9	11.9	1.0	8.2	-7.4	-1.8	-6.4	6.4	12.8
G	-0.1	12.2	-70.1	-48.0	-70.2	-35.8	34.3	-0.8	6.9	5.8	-1.4	5.0	5.4	0.5
C	6.6	21.6	-50.0	-24.1	-43.4	-2.4	40.9	15.1	23.3	0.1	-0.2	15.2	23.1	8.0
T	2.5	26.2	-55.5	-42.6	-53.1	-16.4	36.7	6.4	24.2	-6.4	-3.8	0.0	20.4	20.4
U	1.8	26.6	-62.7	-46.0	-60.9	-19.4	41.5	7.0	23.5	-6.7	-5.1	0.3	18.4	18.1
TAP-C^5	-4.4	20.7	-41.9	-35.7	-46.3	-15.0	31.2	-5.2	24.0	15.8	-0.6	10.7	23.4	12.7
TAP-N	-6.0	1.5	-33.0	-34.0	-39.0	-32.5	6.5	-8.2	-2.3	-2.2	-2.6	-10.4	-4.9	5.5
BA-C^5	-35.2	-10.7	-92.1	-45.5	-127.3	-56.3	71.0	-11.1	7.3	5.2	12.0	-5.9	19.3	25.2
BA-N	11.2	33.8	-12.0	-21.4	-0.8	12.4	13.2	21.1	35.2	0.2	-0.9	21.3	34.3	13.0
CA	11.7	33.2	-30.9	-40.8	-19.2	-7.6	11.6	17.1	33.4	-2.0	-2.8	15.1	30.6	15.5
MM	8.3	7.1	-35.0	-27.3	-26.7	-20.2	6.5	-7.9	0.8	1.4	0.9	-6.5	1.7	8.2
			$^4C_1$							$^4C_1$				
A	6.9	-3.9	-41.5	-43.5	-34.6	-47.4	-12.8	2.6	-8.2	19.9	3.7	22.5	-4.5	-27.0
G	10.2	-5.5	-45.0	-51.5	-34.7	-57.0	-22.3	10.4	-6.6	18.3	18.1	28.7	11.5	-17.2

413

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
C	14.2	-5.6	-42.5	-33.9	-28.3	-39.5	-11.2	20.9	2.3	19.2	3.2	40.1	5.5	-34.5
T	22.2	-1.7	-59.9	-49.1	-37.7	-50.8	-13.1	28.3	2.7	18.9	1.2	47.1	3.9	-43.2
U	22.9	-1.8	-63.8	-52.2	-40.9	-54.0	-13.1	26.6	2.4	19.3	3.8	45.9	6.2	-39.7
TAP-C^5	27.0	-9.2	22.7	-39.9	49.8	-49.1	-98.9	27.4	-8.0	20.8	2.8	48.2	-5.3	-53.5
TAP-N	12.1	-4.5	-48.7	-28.2	-36.6	-32.7	3.9	9.4	-9.4	-11.2	3.6	-1.8	-5.7	-3.9
BA-C^5	-21.7	-43.8	-26.6	-44.9	-48.3	-88.7	-40.4	5.7	-10.4	21.3	3.3	27.0	-7.2	-34.2
BA-N	31.7	17.0	6.8	-23.2	38.5	-6.2	-44.7	31.6	16.0	18.4	3.7	50.1	19.7	-30.4
CA	34.6	15.4	-22.1	-40.7	12.5	-25.3	-37.8	33.5	16.9	18.1	2.5	51.6	19.3	-32.3
MM	16.1	-3.2	-9.0	-29.7	7.1	-32.9	-40.0	16.6	-7.8	18.9	3.1	35.5	-4.6	-40.1
	***p*-RNA (Rib-RUs)** $^1C_4$							***p*-RNA (Rib-RUs)** $^1C_4$						
A	3.5	14.1	-50.7	-33.5	-47.2	-19.4	27.8	2.6	12.0	6.5	17.0	9.1	29.0	19.9
G	3.5	13.9	-53.2	-45.1	-49.7	-31.2	18.5	1.3	14.1	8.0	9.7	9.3	23.8	14.5
C	0.5	23.3	-47.5	-28.0	-47.0	-4.7	42.3	8.6	29.6	9.5	7.5	18.1	37.1	19.1
T	5.9	27.8	-61.9	-63.6	-56.0	-35.8	20.2	12.2	32.1	8.1	15.9	20.3	48.0	27.7
U	5.4	28.1	-64.7	-67.2	-59.3	-39.1	20.1	10.5	32.0	8.4	16.9	18.9	48.9	30.1
TAP-C^5	2.4	13.9	-41.1	-61.5	-38.7	-47.6	-8.9	7.8	14.5	12.7	-1.5	20.5	13.0	-7.5
TAP-N	-2.7	3.5	-38.9	-32.5	-41.6	-29.0	12.6	-6.2	0.6	15.1	8.3	8.9	8.9	0.1
BA-C^5	-35.6	-14.9	-105.3	-36.1	-140.9	-51.0	90.0	1.4	12.1	19.9	10.9	21.3	23.0	1.7
BA-N	29.9	38.1	-9.1	-25.9	20.8	12.2	-8.6	32.4	39.3	21.2	4.1	53.6	43.5	-10.1
CA	31.7	38.0	-30.6	-45.9	1.1	-7.9	-9.1	35.4	37.6	-2.4	4.4	33.0	42.0	9.1
MM	-5.6	9.0	-30.7	-26.9	-36.3	-18.0	18.4	-6.3	2.1	10.0	16.7	3.7	18.8	15.1
	$^4C_1$							$^4C_1$						
A	8.9	-9.7	-29.7	-7.3	-20.8	-16.9	3.8	12.0	-9.8	20.3	18.1	32.3	8.3	-24.0
G	16.2	-3.0	-41.1	-50.2	-24.9	-53.2	-28.3	15.4	-6.9	3.1	20.3	18.4	13.4	-5.0
C	13.3	-11.0	-27.3	-20.9	-14.1	-31.9	-17.9	28.8	-0.7	18.7	18.8	47.5	18.1	-29.4
T	27.0	4.4	-26.9	-51.7	0.1	-47.3	-47.4	37.3	1.2	19.0	22.2	56.3	23.4	-32.8
U	28.0	4.2	-31.1	-54.9	-3.1	-50.6	-47.5	36.2	2.7	17.6	19.3	53.8	22.0	-31.8

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
TAP-C[5]	24.1	-9.0	-53.5	-23.7	-29.5	-32.7	-3.3	28.0	-14.0	19.3	22.2	47.3	8.2	-39.1
TAP-N	0.7	-4.1	-50.4	-18.0	-49.7	-22.0	27.7	-1.1	-10.4	21.6	-1.1	20.6	-11.5	-32.1
BA-C[5]	-20.2	-43.2	-32.7	-44.6	-52.9	-87.8	-34.9	12.8	-12.3	20.4	22.3	33.2	10.0	-23.2
BA-N	34.9	23.5	3.4	-6.4	38.3	17.1	-21.2	38.8	11.9	19.9	20.2	58.8	32.1	-26.7
CA	38.4	23.1	-25.0	-25.2	13.4	-2.1	-15.5	43.7	14.2	17.6	12.0	61.3	26.2	-35.1
MM	16.8	-7.2	-17.2	-31.4	-0.4	-38.6	-38.2	16.2	-9.3	16.6	18.8	32.8	9.5	-23.3

GNA (glycerol-RU) $HAsO_3^-$ (vac.) / **GNA (glycerol-RU)** (solv.)

RU	ΔG_a° (vac)	ΔG_b° (vac)	ΔG_{a+b}° (vac)	$\Delta\Delta G_{a+b}^\circ$	ΔG_a° (solv)	ΔG_b° (solv)	ΔG_{a+b}° (solv)	$\Delta\Delta G_{a+b}^\circ$
A	-27.1	-43.1	-70.2	-	-28.6	5.2	-23.4	-
G	-25.7	-36.0	-61.8	-	-24.8	9.7	-15.1	-
C	-16.3	-50.6	-66.9	-	-17.8	2.4	-15.4	-
T	-16.8	-59.5	-76.4	-	-12.9	5.6	-7.4	-
U	-16.4	-62.7	-79.1	-	-11.6	10.8	-0.8	-
TAP-C[5]	-19.1	-48.9	-68.1	-	-32.5	-5.0	-37.6	-
TAP-N	-9.9	-31.5	-41.4	-	-9.3	-1.0	-10.3	-
BA-C[5]	-55.6	-108.2	-163.8	-	-42.8	0.8	-42.0	-
BA-N	-13.9	-39.3	-53.2	-	-13.3	3.6	-9.7	-
CA	-12.7	-41.3	-54.0	-	-8.9	8.6	-0.3	-
MM	-9.6	-16.5	-26.1	-	-19.7	2.1	-17.7	-

GNA (glyceric acid-RU) (vac.) / **GNA (glyceric acid-RU)** (solv.)

RU	ΔG_a° (vac)	ΔG_b° (vac)	ΔG_{a+b}° (vac)	$\Delta\Delta G_{a+b}^\circ$	ΔG_a° (solv)	ΔG_b° (solv)	ΔG_{a+b}° (solv)	$\Delta\Delta G_{a+b}^\circ$
A	44.7	-14.7	30.1	-	47.5	-4.2	43.4	-
G	46.0	-16.5	29.5	-	48.2	-2.5	45.7	-
C	45.4	-10.5	34.9	-	59.7	0.6	60.3	-
T	53.4	-27.1	26.3	-	63.3	-3.5	59.8	-
U	55.5	-34.3	21.3	-	67.0	-4.2	62.7	-
TAP-C[5]	27.9	-23.4	4.5	-	28.4	7.4	35.9	-
TAP-N	7.4	-33.6	-26.2	-	5.9	-14.5	-8.7	-
BA-C[5]	-9.1	-25.1	-34.3	-	44.9	-58.1	-13.2	-
BA-N	90.0	-40.7	49.3	-	84.7	-5.6	79.2	-

RU[1]	vac.[12] ΔG°_a	ΔG°_b	ΔG°_{a+b}	$\Delta\Delta G^\circ_{a+b}$ [11]	solv.[13] ΔG°_a	ΔG°_b	ΔG°_{a+b}	$\Delta\Delta G^\circ_{a+b}$
CA	94.9	-61.5	33.4	-	90.3	-8.2	82.0	-
MM	27.5	8.8	36.3	-	21.3	3.5	24.8	-

alternative pathway (c+d)

HPO_3^-

RU	ΔG°_a α	β	ΔG°_b α	β	ΔG°_{a+b} α	β	$\Delta\Delta G^\circ_{a+b}$	ΔG°_a α	β	ΔG°_b α	β	ΔG°_{a+b} α	β	$\Delta\Delta G^\circ_{a+b}$

DNA (2dRibf-RUs) — $^2T_3$ (left: vac); **DNA (2dRibf-RUs)** — $^2T_3$ (right: solv)

RU	α	β	α	β	α	β	ΔΔ	α	β	α	β	α	β	ΔΔ
A	-25.0	-53.4	-11.4	-14.2	-36.4	-67.6	-31.2	-1.5	-18.2	8.7	18.3	7.1	0.0	-7.1
G	-25.0	-53.4	-18.5	-14.1	-43.6	-67.5	-23.9	-1.5	-18.2	4.6	11.6	3.1	-6.7	-9.7
C	-25.0	-53.4	0.6	-11.7	-24.4	-65.1	-40.7	-1.5	-18.2	15.0	22.4	13.4	4.2	-9.3
T	-25.0	-53.4	-10.3	11.1	-35.3	-42.3	-7.0	-1.5	-18.2	17.4	26.7	15.8	8.4	-7.4
U	-25.0	-53.4	-17.5	-15.8	-42.5	-69.2	-26.6	-1.5	-18.2	15.8	24.5	14.3	6.3	-8.0
TAP-C[5]	-25.0	-53.4	-19.4	8.9	-44.4	-44.4	-0.1	-1.5	-18.2	7.9	8.6	6.4	-9.7	-16.0
TAP-N	-25.0	-53.4	3.5	-2.4	-21.5	-55.8	-34.3	-1.5	-18.2	14.1	16.0	12.5	-2.3	-14.8
BA-C[5]	-25.0	-53.4	-103.6	-111.8	-128.6	-165.2	-36.6	-1.5	-18.2	0.5	-38.1	-1.0	-56.3	-55.3
BA-N	-25.0	-53.4	29.6	10.7	4.6	-42.7	-47.3	-1.5	-18.2	24.0	30.8	22.5	12.6	-9.9
CA	-25.0	-53.4	10.6	20.8	-14.4	-32.6	-18.2	-1.5	-18.2	25.7	30.7	24.1	12.5	-11.7
MM	-25.0	-53.4	13.7	23.3	-11.3	-30.1	-18.8	-1.5	-18.2	12.7	11.8	11.2	-6.5	-17.6

$^3T_2$

RU	α	β	α	β	α	β	ΔΔ	α	β	α	β	α	β	ΔΔ
A	-28.7	-59.0	6.8	8.0	-21.9	-51.1	-29.2	-1.2	-13.5	-0.1	11.0	-1.3	-2.5	-1.2
G	-28.7	-59.0	-23.4	4.4	-52.1	-54.7	-2.6	-1.2	-13.5	2.8	20.1	1.6	6.7	5.1
C	-28.7	-59.0	-13.9	7.1	-42.6	-51.9	-9.3	-1.2	-13.5	8.1	18.0	6.9	4.5	-2.3
T	-28.7	-59.0	-24.0	-0.3	-52.7	-59.4	-6.6	-1.2	-13.5	9.7	19.4	8.4	5.9	-2.6
U	-28.7	-59.0	-26.9	-8.2	-55.6	-67.2	-11.6	-1.2	-13.5	9.6	16.6	8.3	3.1	-5.2
TAP-C[5]	-28.7	-59.0	-21.4	16.7	-50.1	-42.4	7.7	-1.2	-13.5	-1.6	5.5	-2.8	-8.0	-5.2
TAP-N	-28.7	-59.0	-8.2	20.0	-36.9	-39.0	-2.1	-1.2	-13.5	-0.8	13.2	-2.0	-0.3	1.7
BA-C[5]	-28.7	-59.0	-96.5	-90.2	-125.2	-149.2	-24.1	-1.2	-13.5	-2.2	-46.3	-3.4	-59.7	-56.3
BA-N	-28.7	-59.0	34.6	-6.1	5.9	-65.2	-71.1	-1.2	-13.5	21.3	37.8	20.0	24.3	4.3
CA	-28.7	-59.0	13.2	55.7	-15.5	-3.3	12.2	-1.2	-13.5	16.1	37.6	14.9	24.1	9.2

RU[1]	vac.[12] ΔG°_{a} α	β	ΔG°_{b} α	β	ΔG°_{a+b} α	β	$\Delta\Delta G^{\circ}_{a+b}$ [11]	solv.[13] ΔG°_{a} α	β	ΔG°_{b} α	β	ΔG°_{a+b} α	β	$\Delta\Delta G^{\circ}_{a+b}$
MM	-28.7	-59.0	10.0	23.0	-18.6	-36.0	-17.4	-1.2	-13.5	-0.6	10.1	-1.8	-3.3	-1.5

RNA (Ribf-RUs) $^2T_3$ **RNA (Ribf-RUs)** $^2T_3$

RU[1]	α	β	α	β	α	β	$\Delta\Delta G^{\circ}_{a+b}$	α	β	α	β	α	β	$\Delta\Delta G^{\circ}_{a+b}$
A	-34.4	-39.0	-34.2	3.9	-68.6	-35.1	33.4	-2.8	-24.1	8.7	0.9	6.0	-23.2	-29.1
G	-34.4	-39.0	-40.2	-31.6	-74.6	-70.6	4.0	-2.8	-24.1	6.7	-0.2	3.9	-24.3	-28.2
C	-34.4	-39.0	-30.6	-25.0	-65.0	-64.0	1.0	-2.8	-24.1	18.3	26.0	15.6	2.0	-13.6
T	-34.4	-39.0	-36.5	-5.3	-70.9	-44.4	26.5	-2.8	-24.1	17.9	23.0	15.1	-1.0	-16.1
U	-34.4	-39.0	-39.8	-7.6	-74.2	-46.6	27.5	-2.8	-24.1	6.9	29.4	4.2	5.3	1.2
TAP-C^5	-34.4	-39.0	-21.3	-41.0	-55.7	-80.0	-24.3	-2.8	-24.1	17.0	1.4	14.3	-22.7	-37.0
TAP-N	-34.4	-39.0	-36.2	-10.4	-70.6	-49.5	21.2	-2.8	-24.1	11.2	16.7	8.4	-7.3	-15.8
BA-C^5	-34.4	-39.0	-108.6	-68.8	-143.0	-107.8	35.2	-2.8	-24.1	0.4	-54.9	-2.4	-79.0	-76.6
BA-N	-34.4	-39.0	11.9	-62.7	-22.5	-101.7	-79.2	-2.8	-24.1	29.4	34.5	26.6	10.4	-16.2
CA	-34.4	-39.0	-8.1	5.4	-42.5	-33.6	8.9	-2.8	-24.1	32.3	45.3	29.5	21.3	-8.2
MM	-34.4	-39.0	-26.1	16.9	-60.5	-22.1	38.4	-2.8	-24.1	18.6	11.5	15.9	-12.6	-28.5

$^3T_2$ $^3T_2$

RU[1]	α	β	α	β	α	β	$\Delta\Delta G^{\circ}_{a+b}$	α	β	α	β	α	β	$\Delta\Delta G^{\circ}_{a+b}$
A	-52.2	-51.8	-16.7	39.4	-68.9	-12.4	56.5	-2.4	-12.1	5.3	4.0	2.9	-8.0	-10.9
G	-52.2	-51.8	-19.1	-71.8	-71.3	-123.6	-52.3	-2.4	-12.1	8.1	11.7	5.7	-0.3	-6.0
C	-52.2	-51.8	-9.3	26.9	-61.6	-24.9	36.7	-2.4	-12.1	17.7	12.6	15.3	0.5	-14.8
T	-52.2	-51.8	-20.3	-31.1	-72.6	-82.9	-10.3	-2.4	-12.1	25.0	22.0	22.6	10.0	-12.7
U	-52.2	-51.8	-23.3	-37.7	-75.6	-89.5	-13.9	-2.4	-12.1	24.8	14.1	22.4	2.0	-20.3
TAP-C^5	-52.2	-51.8	-14.2	-30.6	-66.5	-82.4	-15.9	-2.4	-12.1	18.0	0.4	15.6	-11.7	-27.2
TAP-N	-52.2	-51.8	-23.8	-19.4	-76.1	-71.2	4.9	-2.4	-12.1	11.3	10.9	8.9	-1.2	-10.1
BA-C^5	-52.2	-51.8	-68.8	-106.0	-121.1	-157.8	-36.7	-2.4	-12.1	-3.1	-50.3	-5.6	-62.3	-56.8
BA-N	-52.2	-51.8	24.7	-8.5	-27.6	-60.4	-32.8	-2.4	-12.1	28.8	19.4	26.4	7.4	-19.0
CA	-52.2	-51.8	5.0	16.5	-47.2	-35.3	12.0	-2.4	-12.1	31.2	27.8	28.8	15.8	-13.0
MM	-52.2	-51.8	-14.4	34.4	-66.7	-17.4	49.3	-2.4	-12.1	18.0	10.7	15.6	-1.3	-16.9

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	TNA (Tho-RUs)							**TNA (Tho-RUs)**						
	$^2T_3$							$^2T_3$						
	α	β	α	β	α	β		α	β	α	β	α	β	
A	-51.9	-30.9	-12.3	-45.1	-64.2	-76.0	-11.8	-0.5	-42.0	7.3	-12.0	6.8	-54.0	-60.8
G	-51.9	-30.9	-19.7	-49.8	-71.6	-80.8	-9.2	-0.5	-42.0	28.5	-15.8	27.9	-57.8	-85.7
C	-51.9	-30.9	-3.0	-37.0	-54.9	-67.9	-13.0	-0.5	-42.0	32.3	12.2	31.7	-29.8	-61.5
T	-51.9	-30.9	-17.0	-50.2	-68.9	-81.2	-12.2	-0.5	-42.0	36.7	16.3	36.1	-25.6	-61.8
U	-51.9	-30.9	-19.8	-53.3	-71.7	-84.2	-12.5	-0.5	-42.0	37.5	6.3	36.9	-35.7	-72.6
TAP-C^5	-51.9	-30.9	-24.4	-38.6	-76.3	-69.6	6.7	-0.5	-42.0	1.5	1.0	1.0	-41.0	-41.9
TAP-N	-51.9	-30.9	-7.6	5.1	-59.5	-25.8	33.7	-0.5	-42.0	26.2	-3.4	25.7	-45.3	-71.0
BA-C^5	-51.9	-30.9	-57.4	-102.0	-109.3	-133.0	-23.6	-0.5	-42.0	-31.3	-9.6	-31.8	-51.5	-19.7
BA-N	-51.9	-30.9	27.7	-2.6	-24.2	-33.5	-9.3	-0.5	-42.0	22.9	23.2	22.4	-18.7	-41.1
CA	-51.9	-30.9	6.6	-27.8	-45.3	-58.7	-13.4	-0.5	-42.0	25.2	24.4	24.7	-17.5	-42.2
MM	-51.9	-30.9	2.3	6.5	-49.6	-24.4	25.2	-0.5	-42.0	1.4	2.0	0.9	-40.0	-40.8
	$^3T_2$							$^3T_2$						
A	-53.7	-46.4	-12.3	-20.0	-66.1	-66.3	-0.2	-21.0	-1.0	2.2	-21.7	-18.8	-22.7	-3.9
G	-53.7	-46.4	-19.7	-21.2	-73.4	-67.5	5.9	-21.0	-1.0	2.7	-25.6	-18.3	-26.6	-8.3
C	-53.7	-46.4	-3.0	-11.9	-56.8	-58.2	-1.5	-21.0	-1.0	9.1	-3.2	-11.9	-4.3	7.6
T	-53.7	-46.4	-17.0	-25.1	-70.8	-71.5	-0.7	-21.0	-1.0	8.2	2.8	-12.8	1.8	14.5
U	-53.7	-46.4	-19.8	-28.2	-73.5	-74.5	-1.0	-21.0	-1.0	9.0	-5.6	-12.0	-6.6	5.4
TAP-C^5	-53.7	-46.4	-24.4	-9.3	-78.1	-55.7	22.4	-21.0	-1.0	-4.6	-20.4	-25.6	-21.4	4.1
TAP-N	-53.7	-46.4	-7.6	-21.2	-61.3	-67.6	-6.2	-21.0	-1.0	-2.9	-28.1	-23.9	-29.1	-5.2
BA-C^5	-53.7	-46.4	-57.4	-69.0	-111.2	-115.3	-4.1	-21.0	-1.0	-6.3	-22.6	-27.3	-23.6	3.8
BA-N	-53.7	-46.4	27.7	19.1	-26.0	-27.3	-1.3	-21.0	-1.0	21.9	14.8	0.9	13.8	12.9
CA	-53.7	-46.4	6.6	-2.7	-47.1	-49.0	-1.9	-21.0	-1.0	20.0	-1.6	-1.0	-2.6	-1.6
MM	-53.7	-46.4	2.3	-9.5	-51.4	-55.8	-4.4	-21.0	-1.0	-2.1	-4.7	-23.1	-5.7	17.4
	p-DNA (2dRib-RUs)							**p-DNA (2dRib-RUs)**						
	$^1C_4$							$^1C_4$						
A	-27.7	-39.0	-29.1	-0.2	-56.8	-39.2	17.6	-6.9	-8.8	-2.3	13.1	-9.3	4.3	13.6

RU[1]	vac.[12]							solv.[13]						
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
G	-27.7	-39.0	-52.8	-7.0	-80.5	-46.0	34.4	-6.9	-8.8	-3.2	12.1	-10.1	3.3	13.4
C	-27.7	-39.0	-29.7	14.5	-57.4	-24.5	32.9	-6.9	-8.8	6.4	28.8	-0.5	20.0	20.5
T	-27.7	-39.0	-38.8	3.7	-66.5	-35.3	31.2	-6.9	-8.8	7.2	26.6	0.2	17.8	17.6
U	-27.7	-39.0	-41.0	0.3	-68.7	-38.8	29.9	-6.9	-8.8	6.1	28.2	-0.9	19.4	20.3
TAP-C^5	-27.7	-39.0	-20.9	-0.1	-48.6	-39.2	9.4	-6.9	-8.8	-4.4	18.3	-11.4	9.5	20.9
TAP-N	-27.7	-39.0	-31.2	-1.3	-58.9	-40.3	18.6	-6.9	-8.8	-3.4	3.2	-10.4	-5.6	4.8
BA-C^5	-27.7	-39.0	-76.2	-35.4	-103.9	-74.4	29.5	-6.9	-8.8	-3.9	18.9	-10.8	10.1	20.9
BA-N	-27.7	-39.0	23.5	43.1	-4.3	4.1	8.3	-6.9	-8.8	21.9	41.1	14.9	32.3	17.4
CA	-27.7	-39.0	5.0	22.8	-22.8	-16.2	6.5	-6.9	-8.8	21.6	37.5	14.6	28.7	14.0
MM	-27.7	-39.0	-9.6	10.7	-37.3	-28.4	8.9	-6.9	-8.8	-2.9	8.4	-9.8	-0.4	9.4
			$^4C_1$							$^4C_1$				
A	-22.1	-47.8	-22.5	-9.3	-44.6	-57.1	-12.5	6.3	-2.9	1.6	-2.5	7.9	-5.4	-13.3
G	-22.1	-47.8	-23.8	-19.1	-46.0	-66.9	-20.9	6.3	-2.9	10.4	-3.8	16.6	-6.8	-23.4
C	-22.1	-47.8	-30.8	-1.3	-53.0	-49.1	3.9	6.3	-2.9	15.2	4.9	21.5	2.0	-19.5
T	-22.1	-47.8	-25.8	-13.2	-47.9	-61.0	-13.2	6.3	-2.9	18.1	5.3	24.4	2.4	-22.0
U	-22.1	-47.8	-29.1	-16.4	-51.2	-64.2	-13.0	6.3	-2.9	18.2	6.2	24.4	3.3	-21.2
TAP-C^5	-22.1	-47.8	-14.8	-10.4	-37.0	-58.2	-21.2	6.3	-2.9	19.1	-3.4	25.3	-6.3	-31.6
TAP-N	-22.1	-47.8	-23.8	-4.6	-45.9	-52.4	-6.4	6.3	-2.9	-10.4	-4.8	-4.1	-7.8	-3.6
BA-C^5	-22.1	-47.8	-78.8	-50.3	-100.9	-98.1	2.8	6.3	-2.9	19.5	-7.7	25.8	-10.6	-36.4
BA-N	-22.1	-47.8	45.5	33.1	23.3	-14.7	-38.1	6.3	-2.9	42.5	21.5	48.7	18.6	-30.1
CA	-22.1	-47.8	25.2	13.3	3.1	-34.5	-37.6	6.3	-2.9	47.4	21.1	53.7	18.2	-35.5
MM	-22.1	-47.8	-18.0	5.6	-40.2	-42.2	-2.1	6.3	-2.9	-4.8	-3.8	1.5	-6.7	-8.1
	p-RNA (Rib-RUs)							**p-RNA (Rib-RUs)**						
			$^1C_4$							$^1C_4$				
A	-52.2	-39.3	-3.9	7.9	-56.1	-31.4	24.7	4.6	1.1	11.0	16.6	15.5	17.7	2.2
G	-52.2	-39.3	-2.3	-0.4	-54.6	-39.7	14.9	4.6	1.1	15.1	14.7	19.7	15.8	-3.9
C	-52.2	-39.3	-4.5	20.3	-56.7	-19.0	37.7	4.6	1.1	22.2	31.5	26.8	32.7	5.9
T	-52.2	-39.3	-13.0	7.6	-65.3	-31.7	33.6	4.6	1.1	24.0	33.5	28.6	34.6	6.0

RU[1]	ΔG°$_a$		ΔG°$_b$		ΔG°$_{a+b}$		ΔΔG°$_{a+b}$ [11]	ΔG°$_a$		ΔG°$_b$		ΔG°$_{a+b}$		ΔΔG°$_{a+b}$
	vac.[12]							*solv.*[13]						
	α	β	α	β	α	β		α	β	α	β	α	β	
U	-52.2	-39.3	-16.0	4.7	-68.2	-34.6	33.6	4.6	1.1	23.7	32.6	28.2	33.7	5.5
TAP-C^5	-52.2	-39.3	6.4	-17.1	-45.9	-56.4	-10.5	4.6	1.1	18.7	8.0	23.3	9.1	-14.2
TAP-N	-52.2	-39.3	3.6	1.1	-48.7	-38.3	10.4	4.6	1.1	-0.7	4.9	3.9	6.1	2.2
BA-C^5	-52.2	-39.3	-69.4	-42.4	-121.7	-81.7	40.0	4.6	1.1	7.0	15.2	11.6	16.4	4.7
BA-N	-52.2	-39.3	66.7	44.5	14.5	5.2	-9.3	4.6	1.1	38.4	38.5	43.0	39.7	-3.3
CA	-52.2	-39.3	46.8	23.3	-5.5	-16.1	-10.6	4.6	1.1	40.5	37.3	45.1	38.4	-6.7
MM	-52.2	-39.3	9.0	14.7	-43.3	-24.7	18.6	4.6	1.1	-0.5	7.1	4.1	8.3	4.2
			$^4C_1$							$^4C_1$				
A	-38.2	-27.5	0.2	-32.2	-38.0	-59.7	-21.7	-0.9	-8.8	17.1	-3.6	16.2	-12.4	-28.6
G	-38.2	-27.5	-12.5	-42.5	-50.7	-70.0	-19.2	-0.9	-8.8	17.4	-2.4	16.5	-11.2	-27.7
C	-38.2	-27.5	-6.7	-20.6	-44.9	-48.1	-3.2	-0.9	-8.8	26.5	7.4	25.6	-1.4	-27.0
T	-38.2	-27.5	-13.4	-36.1	-51.6	-63.6	-11.9	-0.9	-8.8	26.9	7.2	26.1	-1.6	-27.7
U	-38.2	-27.5	-16.5	-39.1	-54.7	-66.6	-11.9α	-0.9	-8.8	26.0	7.7	25.2	-1.1	-26.2
TAP-C^5	-38.2	-27.5	-1.0	-23.3	-39.3	-50.8	-11.5	-0.9	-8.8	25.2	-4.3	24.4	-13.1	-37.4
TAP-N	-38.2	-27.5	-21.6	-14.6	-59.8	-42.1	17.7	-0.9	-8.8	-6.1	-4.1	-6.9	-12.9	-6.0
BA-C^5	-38.2	-27.5	-65.0	-68.9	-103.2	-96.4	6.8	-0.9	-8.8	-17.4	-6.4	-18.3	-15.2	3.1
BA-N	-38.2	-27.5	52.9	4.9	14.7	-22.6	-37.2	-0.9	-8.8	59.0	20.5	58.1	11.7	-46.5
CA	-38.2	-27.5	31.6	-16.1	-6.6	-43.6	-36.9	-0.9	-8.8	55.4	19.5	54.6	10.7	-43.8
MM	-38.2	-27.5	-15.7	-19.1	-54.0	-46.6	7.4	-0.9	-8.8	-4.6	-4.3	-5.4	-13.1	-7.7
	GNA (glycerol-RU)							**GNA (glycerol-RU)**						
A	-35.2		-44.9		-80.1		-	6.7		-32.2		-25.4		-
G	-35.2		-51.1		-86.2		-	6.7		-21.4		-14.6		-
C	-35.2		-37.4		-72.5		-	6.7		-22.9		-16.2		-
T	-35.2		-49.6		-84.7		-	6.7		-16.8		-10.0		-
U	-35.2		-52.0		-87.2		-	6.7		-17.2		-10.5		-
TAP-C^5	-35.2		-44.5		-79.7		-	6.7		-41.9		-35.2		-
TAP-N	-35.2		-24.4		-59.6		-	6.7		-16.7		-10.0		-
BA-C^5	-35.2		-129.8		-164.9		-	6.7		-79.0		-72.2		-

RU[1]	vac.[12]				solv.[13]			
	ΔG°_a	ΔG°_b	ΔG°_{a+b}	$\Delta\Delta G^{\circ}_{a+b}$ [11]	ΔG°_a	ΔG°_b	ΔG°_{a+b}	$\Delta\Delta G^{\circ}_{a+b}$
BA-N	-35.2	-0.8	-35.9	-	6.7	-18.0	-11.3	-
CA	-35.2	-24.8	-59.9	-	6.7	-15.1	-8.3	-
MM	-35.2	-4.8	-39.9	-	6.7	-9.8	-3.0	-
GNA (glyceric acid-RU)					**GNA (glyceric acid-RU)**			
A	-63.1	65.8	2.6	-	-3.8	45.2	41.4	-
G	-63.1	31.8	-31.4	-	-3.8	47.8	44.0	-
C	-63.1	69.1	5.9	-	-3.8	59.8	56.0	-
T	-63.1	58.8	-4.3	-	-3.8	63.1	59.2	-
U	-63.1	57.5	-5.6	-	-3.8	64.0	60.1	-
TAP-C^5	-63.1	52.9	-10.3	-	-3.8	40.0	36.2	-
TAP-N	-63.1	22.3	-40.8	-	-3.8	10.7	6.9	-
BA-C^5	-63.1	-29.9	-93.0	-	-3.8	52.1	48.3	-
BA-N	-63.1	77.6	14.5	-	-3.8	80.9	77.1	-
CA	-63.1	98.8	35.7	-	-3.8	84.7	80.8	-
MM	-63.1	53.2	-10.0	-	-3.8	26.8	22.9	-

HAsO$_3^-$

RU	vac.							solv.						
	α	β	α	β	α	β		α	β	α	β	α	β	
	DNA (2dRibf-RUs) $^2T_3$							**DNA (2dRibf-RUs)** $^2T_3$						
A	-15.0	-49.5	-16.1	25.1	-31.0	-24.4	6.6	-0.8	-16.8	5.8	18.8	5.1	1.9	-3.1
G	-15.0	-49.5	-16.4	-13.1	-31.4	-62.6	-31.2	-0.8	-16.8	6.7	14.7	6.0	-2.2	-8.1
C	-15.0	-49.5	-1.1	14.3	-16.1	-35.2	-19.1	-0.8	-16.8	12.0	23.4	11.3	6.6	-4.7
T	-15.0	-49.5	-16.4	11.1	-31.4	-38.5	-7.1	-0.8	-16.8	15.1	26.9	14.3	10.0	-4.3
U	-15.0	-49.5	-20.1	11.3	-35.1	-38.3	-3.1	-0.8	-16.8	15.1	25.8	14.4	8.9	-5.4
TAP-C^5	-15.0	-49.5	-17.1	15.7	-32.1	-33.9	-1.8	-0.8	-16.8	18.7	10.7	17.9	-6.1	-24.1
TAP-N	-15.0	-49.5	8.7	22.3	-6.3	-27.2	-21.0	-0.8	-16.8	11.5	14.8	10.7	-2.0	-12.7
BA-C^5	-15.0	-49.5	-39.3	-112.5	-54.3	-162.0	-107.7	-0.8	-16.8	-2.1	-56.6	-2.9	-73.5	-70.5
BA-N	-15.0	-49.5	30.5	40.7	15.5	-8.8	-24.3	-0.8	-16.8	22.6	29.5	21.8	12.7	-9.2
CA	-15.0	-49.5	-5.8	24.6	-20.8	-24.9	-4.2	-0.8	-16.8	22.8	28.0	22.0	11.2	-10.8

RU[1]	vac.[12]								solv.[13]							
	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$ [11]		ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
MM	-15.0	-49.5	8.7	23.7	-6.3	-25.8	-19.6	-0.8	-16.8	10.4	12.3	9.6	-4.5	-14.2
				${}^3T_2$								${}^3T_2$			
A	-23.3	-47.5	-20.5	25.1	-43.9	-22.3	21.5	-4.8	9.4	14.5	-11.9	9.7	-2.5	-12.2
G	-23.3	-47.5	-23.3	-13.0	-46.7	-60.5	-13.8	-4.8	9.4	15.3	-7.0	10.5	2.5	-8.1
C	-23.3	-47.5	-15.1	14.3	-38.4	-33.1	5.3	-4.8	9.4	8.1	-8.3	3.3	1.1	-2.2
T	-23.3	-47.5	-24.6	11.1	-47.9	-36.4	11.5	-4.8	9.4	12.9	-3.5	8.1	6.0	-2.1
U	-23.3	-47.5	-27.4	-15.8	-50.8	-63.2	-12.5	-4.8	9.4	12.9	3.8	8.1	13.2	5.1
TAP-C^5	-23.3	-47.5	-5.4	15.7	-28.8	-31.8	-3.0	-4.8	9.4	-3.6	-9.2	-8.4	0.2	8.6
TAP-N	-23.3	-47.5	6.0	22.3	-17.4	-25.2	-7.8	-4.8	9.4	8.7	0.1	3.9	9.5	5.6
BA-C^5	-23.3	-47.5	-110.3	-112.5	-133.7	-160.0	-26.3	-4.8	9.4	-60.6	-15.7	-65.4	-6.2	59.2
BA-N	-23.3	-47.5	32.9	-3.6	9.6	-51.1	-60.7	-4.8	9.4	22.7	15.9	17.9	25.3	7.5
CA	-23.3	-47.5	18.9	7.1	-4.5	-40.4	-35.9	-4.8	9.4	19.6	2.8	14.8	12.2	-2.6
MM	-23.3	-47.5	-11.2	23.7	-34.6	-23.7	10.9	-4.8	9.4	9.6	0.9	4.8	10.3	5.5
			RNA (Ribf-RUs)								**RNA (Ribf-RUs)**				
				${}^2T_3$								${}^2T_3$			
A	-24.6	-40.5	-34.5	18.0	-59.1	-22.5	36.6	2.6	-17.2	7.4	16.7	10.0	-0.5	-10.4
G	-24.6	-40.5	-51.7	-31.1	-76.3	-71.6	4.8	2.6	-17.2	7.2	12.3	9.8	-4.9	-14.7
C	-24.6	-40.5	7.3	-4.6	-17.3	-45.1	-27.8	2.6	-17.2	10.6	15.2	13.2	-2.0	-15.1
T	-24.6	-40.5	2.6	1.7	-22.0	-38.9	-16.8	2.6	-17.2	19.1	22.9	21.7	5.8	-15.9
U	-24.6	-40.5	-0.7	-12.1	-25.3	-52.6	-27.3	2.6	-17.2	19.7	19.9	22.3	2.8	-19.5
TAP-C^5	-24.6	-40.5	-22.3	-36.1	-46.9	-76.6	-29.8	2.6	-17.2	11.0	9.3	13.6	-7.9	-21.5
TAP-N	-24.6	-40.5	-39.1	11.1	-63.7	-29.4	34.3	2.6	-17.2	4.9	13.6	7.5	-3.6	-11.1
BA-C^5	-24.6	-40.5	-104.1	-121.1	-128.7	-161.6	-32.9	2.6	-17.2	-4.4	-59.5	-1.8	-76.7	-74.9
BA-N	-24.6	-40.5	8.2	7.8	-16.4	-32.7	-16.3	2.6	-17.2	25.2	22.7	27.8	5.6	-22.2
CA	-24.6	-40.5	10.7	5.8	-13.9	-34.7	-20.8	2.6	-17.2	35.8	24.2	38.4	7.1	-31.3
MM	-24.6	-40.5	-29.5	13.6	-54.1	-26.9	27.2	2.6	-17.2	11.2	11.9	13.8	-5.3	-19.1
				${}^3T_2$								${}^3T_2$			
A	-29.0	-49.6	-28.0	7.1	-57.0	-42.5	14.5	-5.1	-9.1	14.8	-3.4	9.7	-12.4	-22.2

RU[1]	vac.[12]							solv.[13]						
	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^\circ_{a+b}$ [11]	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^\circ_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
G	-29.0	-49.6	-46.1	10.9	-75.1	-38.7	36.4	-5.1	-9.1	8.3	13.1	3.2	4.0	0.9
C	-29.0	-49.6	-25.6	-8.9	-54.6	-58.5	-3.9	-5.1	-9.1	19.9	4.6	14.7	-4.5	-19.2
T	-29.0	-49.6	-35.0	-19.8	-64.0	-69.4	-5.4	-5.1	-9.1	41.2	14.7	36.1	5.6	-30.5
U	-29.0	-49.6	-38.1	-32.1	-67.1	-81.7	-14.6	-5.1	-9.1	40.7	13.4	35.6	4.4	-31.2
TAP-C[5]	-29.0	-49.6	-18.5	-31.3	-47.5	-80.9	-33.4	-5.1	-9.1	13.3	3.5	8.2	-5.6	-13.7
TAP-N	-29.0	-49.6	-29.7	-7.1	-58.8	-56.7	2.0	-5.1	-9.1	11.5	2.0	6.3	-7.0	-13.4
BA-C[5]	-29.0	-49.6	-84.0	-47.3	-113.0	-96.9	16.1	-5.1	-9.1	-71.9	-49.5	-77.0	-58.6	18.5
BA-N	-29.0	-49.6	17.0	34.8	-12.0	-14.9	-2.9	-5.1	-9.1	28.3	25.0	23.2	16.0	-7.2
CA	-29.0	-49.6	19.4	18.0	-9.6	-31.6	-21.9	-5.1	-9.1	33.9	28.2	28.7	19.1	-9.6
MM	-29.0	-49.6	-19.7	-6.7	-48.7	-56.3	-7.6	-5.1	-9.1	17.8	2.6	12.7	-6.5	-19.2

<div align="center">

TNA (Tho-RUs)   **TNA (Tho-RUs)**

$^2T_3$ $^2T_3$

</div>

RU	α	β	α	β	α	β		α	β	α	β	α	β	
A	-47.7	-37.5	-10.1	310.3	-57.8	272.8	330.6	0.3	-27.2	20.1	-1.7	20.4	-28.9	-49.3
G	-47.7	-37.5	-18.1	307.9	-65.9	270.4	336.3	0.3	-27.2	21.4	0.4	21.6	-26.8	-48.4
C	-47.7	-37.5	-0.6	10.0	-48.4	-27.5	20.9	0.3	-27.2	30.9	19.9	31.2	-7.3	-38.5
T	-47.7	-37.5	-11.2	-4.5	-59.0	-42.0	16.9	0.3	-27.2	14.2	16.9	14.5	-10.3	-24.8
U	-47.7	-37.5	-14.1	-7.9	-61.8	-45.3	16.5	0.3	-27.2	36.7	-6.7	37.0	-33.9	-70.8
TAP-C[5]	-47.7	-37.5	-22.3	7.4	-70.0	-30.0	40.0	0.3	-27.2	17.5	-7.1	17.8	-34.3	-52.1
TAP-N	-47.7	-37.5	-6.2	0.0	-53.9	-37.5	16.4	0.3	-27.2	1.3	-3.1	1.5	-30.3	-31.8
BA-C[5]	-47.7	-37.5	-55.2	-93.6	-102.9	-131.1	-28.2	0.3	-27.2	-38.9	-6.9	-38.6	-34.1	4.5
BA-N	-47.7	-37.5	29.7	40.8	-18.0	3.3	21.3	0.3	-27.2	40.0	34.1	40.3	6.9	-33.4
CA	-47.7	-37.5	8.9	18.7	-38.9	-18.8	20.1	0.3	-27.2	41.4	20.7	41.6	-6.5	-48.2
MM	-47.7	-37.5	19.7	12.0	-28.0	-25.5	2.5	0.3	-27.2	21.3	0.7	21.6	-26.5	-48.1

<div align="center">

$^3T_2$ $^3T_2$

</div>

RU	α	β	α	β	α	β		α	β	α	β	α	β	
A	-49.6	-41.5	-10.1	-20.1	-59.7	-61.6	-2.0	-19.5	-2.2	0.8	-20.0	-18.7	-22.2	-3.5
G	-49.6	-41.5	-18.1	-17.8	-67.7	-59.4	8.4	-19.5	-2.2	1.9	-19.2	-17.6	-21.4	-3.8
C	-49.6	-41.5	-0.6	-11.9	-50.2	-53.4	-3.2	-19.5	-2.2	7.6	-1.3	-11.9	-3.5	8.4
T	-49.6	-41.5	-11.2	-25.0	-60.8	-66.5	-5.7	-19.5	-2.2	7.8	-10.3	-11.8	-12.5	-0.7

RU[1]	vac.[12]							solv.[13]						
	ΔG°_{a}		ΔG°_{b}		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$ [11]	ΔG°_{a}		ΔG°_{b}		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
U	-49.6	-41.5	-14.1	-28.1	-63.6	-69.6	-6.0	-19.5	-2.2	7.8	-12.1	-11.7	-14.3	-2.6
TAP-C^5	-49.6	-41.5	-22.3	-12.5	-71.9	-54.1	17.8	-19.5	-2.2	-6.7	-2.9	-26.3	-5.0	21.2
TAP-N	-49.6	-41.5	-6.2	-20.9	-55.8	-62.4	-6.7	-19.5	-2.2	-3.4	-6.7	-22.9	-8.8	14.1
BA-C^5	-49.6	-41.5	-55.2	-88.7	-104.7	-130.2	-25.5	-19.5	-2.2	-8.1	-19.3	-27.7	-21.4	6.3
BA-N	-49.6	-41.5	29.8	18.9	-19.8	-22.6	-2.8	-19.5	-2.2	20.9	9.4	1.4	7.2	5.8
CA	-49.6	-41.5	8.9	-2.4	-40.7	-43.9	-3.2	-19.5	-2.2	19.2	233.0	-0.4	230.9	231.2
MM	-49.6	-41.5	19.7	-8.8	-29.9	-50.4	-20.5	-19.5	-2.2	-2.6	2.6	-22.2	0.5	22.6
p-DNA (2dRib-RUs)								**p-DNA (2dRib-RUs)**						
$^1C_4$								$^1C_4$						
A	-38.8	-31.0	-12.0	5.1	-50.8	-25.9	24.9	-3.2	-5.3	-2.2	11.7	-5.4	6.4	11.8
G	-38.8	-31.0	-31.3	-4.8	-70.1	-35.8	34.3	-3.2	-5.3	7.9	10.7	4.7	5.4	0.7
C	-38.8	-31.0	-11.3	15.6	-50.1	-15.4	34.6	-3.2	-5.3	5.6	28.3	2.4	23.0	20.6
T	-38.8	-31.0	-20.3	4.7	-59.0	-26.3	32.7	-3.2	-5.3	5.9	31.0	2.7	25.6	22.9
U	-38.8	-31.0	-22.8	1.3	-61.6	-29.7	31.9	-3.2	-5.3	5.2	29.9	2.0	24.6	22.6
TAP-C^5	-38.8	-31.0	-10.1	7.7	-48.9	-23.3	25.6	-3.2	-5.3	-4.6	16.6	-7.8	11.3	19.1
TAP-N	-38.8	-31.0	-9.3	10.8	-48.1	-20.2	27.9	-3.2	-5.3	13.7	3.7	10.5	-1.6	-12.1
BA-C^5	-38.8	-31.0	-57.2	-56.5	-95.9	-87.5	8.4	-3.2	-5.3	-3.4	19.4	-6.6	14.1	20.7
BA-N	-38.8	-31.0	39.7	43.3	0.9	12.3	11.4	-3.2	-5.3	23.3	41.4	20.1	36.1	16.0
CA	-38.8	-31.0	21.0	22.5	-17.8	-8.5	9.2	-3.2	-5.3	22.3	37.6	19.1	32.3	13.1
MM	-38.8	-31.0	32.8	10.8	-5.9	-20.2	-14.3	-3.2	-5.3	19.4	7.0	16.2	1.7	-14.6
$^4C_1$								$^4C_1$						
A	-10.9	-38.5	-22.6	-8.9	-33.4	-47.4	-14.0	10.7	-1.2	17.3	-3.4	28.0	-4.5	-32.5
G	-10.9	-38.5	-23.9	-18.5	-34.7	-57.1	-22.3	10.7	-1.2	6.5	-3.3	17.2	-4.5	-21.6
C	-10.9	-38.5	-4.4	-0.9	-15.3	-39.5	-24.2	10.7	-1.2	29.9	6.7	40.6	5.5	-35.0
T	-10.9	-38.5	-27.4	-12.3	-38.2	-50.8	-12.6	10.7	-1.2	30.5	5.1	41.2	3.9	-37.3
U	-10.9	-38.5	-30.2	-15.4	-41.1	-54.0	-12.9	10.7	-1.2	15.0	7.4	25.7	6.3	-19.5
TAP-C^5	-10.9	-38.5	-16.7	-10.6	-27.6	-49.1	-21.5	10.7	-1.2	29.1	-3.2	39.8	-4.3	-44.1
TAP-N	-10.9	-38.5	-23.7	-4.4	-34.6	-42.9	-8.3	10.7	-1.2	-10.0	-4.6	0.7	-5.7	-6.5

RU[(1)]	vac.[(12)]							solv.[(13)]						
	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$[(11)]	ΔG°_a		ΔG°_b		ΔG°_{a+b}		$\Delta\Delta G^{\circ}_{a+b}$
	α	β	α	β	α	β		α	β	α	β	α	β	
BA-C^5	-10.9	-38.5	-99.3	-50.1	-110.1	-88.7	21.5	10.7	-1.2	16.3	-6.0	27.0	-7.2	-34.2
BA-N	-10.9	-38.5	43.9	32.4	33.0	-6.1	-39.1	10.7	-1.2	39.4	20.9	50.1	19.7	-30.4
CA	-10.9	-38.5	23.4	13.3	12.6	-25.3	-37.8	10.7	-1.2	40.9	20.5	51.6	19.3	-32.3
MM	-10.9	-38.5	-19.9	5.6	-30.8	-32.9	-2.2	10.7	-1.2	17.2	-3.5	27.9	-4.6	-32.5
p-RNA (Rib-RUs) $^1C_4$								**p-RNA (Rib-RUs)** $^1C_4$						
A	-45.4	-32.6	-3.6	8.8	-49.0	-23.8	25.2	5.1	12.6	5.1	17.9	10.2	30.4	20.2
G	-45.4	-32.6	2.9	0.0	-42.5	-32.6	9.9	5.1	12.6	4.2	20.3	9.3	32.8	23.5
C	-45.4	-32.6	-3.7	21.4	-49.1	-11.2	37.9	5.1	12.6	13.0	24.9	18.1	37.5	19.5
T	-45.4	-32.6	-12.1	9.2	-57.5	-23.4	34.1	5.1	12.6	15.2	36.7	20.3	49.3	29.0
U	-45.4	-32.6	-15.2	6.1	-60.6	-26.5	34.1	5.1	12.6	13.8	36.7	18.9	49.3	30.5
TAP-C^5	-45.4	-32.6	6.7	-15.2	-38.7	-47.8	-9.1	5.1	12.6	15.1	30.6	20.2	43.2	23.0
TAP-N	-45.4	-32.6	3.8	1.2	-41.6	-31.4	10.2	5.1	12.6	-0.8	4.8	4.3	17.4	13.1
BA-C^5	-45.4	-32.6	-68.7	-40.2	-114.1	-72.8	41.3	5.1	12.6	9.3	14.1	14.4	26.7	12.3
BA-N	-45.4	-32.6	66.2	44.8	20.8	12.2	-8.6	5.1	12.6	48.4	31.1	53.5	43.7	-9.8
CA	-45.4	-32.6	46.5	24.0	1.1	-8.6	-9.8	5.1	12.6	45.5	30.8	50.6	43.4	-7.2
MM	-45.4	-32.6	9.0	14.6	-36.4	-18.0	18.4	5.1	12.6	-0.6	6.2	4.5	18.8	14.3
$^4C_1$								$^4C_1$						
A	-28.6	-11.7	-2.4	-31.9	-31.0	-43.5	-12.5	12.4	-6.8	14.3	-2.8	26.7	-9.6	-36.3
G	-28.6	-11.7	-22.5	-41.5	-51.1	-53.2	-2.1	12.4	-6.8	6.0	-3.5	18.4	-10.3	-28.8
C	-28.6	-11.7	-6.8	-20.3	-35.4	-31.9	3.5	12.4	-6.8	19.4	7.3	31.8	0.4	-31.4
T	-28.6	-11.7	-10.7	-35.6	-39.3	-47.3	-8.0	12.4	-6.8	20.0	7.0	32.4	0.2	-32.2
U	-28.6	-11.7	-13.9	-39.0	-42.5	-50.6	-8.1	12.4	-6.8	18.5	7.8	30.9	0.9	-30.0
TAP-C^5	-28.6	-11.7	14.8	-22.4	-13.8	-34.1	-20.3	12.4	-6.8	34.7	-4.6	47.1	-11.4	-58.5
TAP-N	-28.6	-11.7	-21.3	-12.7	-50.0	-24.4	25.6	12.4	-6.8	-14.9	-4.6	-2.5	-11.5	-9.0
BA-C^5	-28.6	-11.7	-87.4	-44.8	-116.1	-56.4	59.6	12.4	-6.8	-37.0	-7.4	-24.6	-14.2	10.4
BA-N	-28.6	-11.7	52.7	5.8	24.1	-5.8	-29.9	12.4	-6.8	45.8	16.7	58.2	9.9	-48.4
CA	-28.6	-11.7	32.5	-15.0	3.9	-26.7	-30.5	12.4	-6.8	63.5	18.9	76.0	12.0	-63.9

RU[1]	vac.[12]							solv.[13]							
	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$ [11]	ΔG_a°		ΔG_b°		ΔG_{a+b}°		$\Delta\Delta G_{a+b}^\circ$
	α	β	α	β	α	β		α	β	α	β	α	β	
MM	-28.6	-11.7	-16.6	-17.9	-45.2	-29.5	15.6	12.4	-6.8	16.0	-4.2	28.4	-11.0	-39.5
GNA (glycerol-RU) HAsO₃⁻								**GNA (glycerol-RU)**						
A	-25.7		-47.7		-73.3		-	8.4		-25.5		-17.1		-
G	-25.7		-53.6		-79.3		-	8.4		-23.5		-15.1		-
C	-25.7		-41.6		-67.2		-	8.4		-16.5		-8.2		-
T	-25.7		-48.6		-74.2		-	8.4		-16.9		-8.5		-
U	-25.7		-51.1		-76.8		-	8.4		-17.2		-8.8		-
TAP-C[5]	-25.7		-45.8		-71.4		-	8.4		-30.9		-22.5		-
TAP-N	-25.7		-27.5		-53.1		-	8.4		-17.0		-8.6		-
BA-C[5]	-25.7		-140.2		-165.8		-	8.4		-50.3		-41.9		-
BA-N	-25.7		-0.8		-26.5		-	8.4		-18.1		-9.7		-
CA	-25.7		-28.3		-54.0		-	8.4		-8.7		-0.3		-
MM	-25.7		-8.2		-33.9		-	8.4		-10.7		-2.4		-
GNA (glyceric acid-RU)								**GNA (glyceric acid-RU)**						
A	-49.2		58.2		9.0		-	-4.2		45.6		41.4		-
G	-49.2		22.8		-26.4		-	-4.2		61.0		56.8		-
C	-49.2		64.0		14.8		-	-4.2		61.0		56.8		-
T	-49.2		53.1		4.0		-	-4.2		77.7		73.5		-
U	-49.2		-11.9		-61.1		-	-4.2		65.2		61.0		-
TAP-C[5]	-49.2		43.5		-5.7		-	-4.2		39.2		35.0		-
TAP-N	-49.2		5.9		-43.3		-	-4.2		-4.5		-8.7		-
BA-C[5]	-49.2		-34.0		-83.2		-	-4.2		50.8		46.6		-
BA-N	-49.2		105.9		56.8		-	-4.2		82.9		78.7		-
CA	-49.2		85.1		36.0		-	-4.2		85.1		80.9		-
MM	-49.2		46.7		-2.5		-	-4.2		12.2		8.0		-

[1]RU = unspecified **r**ecognition **u**nit. [2]2dRibf: D-2'-deoxyribofuranose. [3]Ribf: D-ribofuranose. [4]Tho: D-threose. [5]2dRib: D-2'-deoxyribopyranose.

$^{(6)}$Rib: D-ribopyranose. $^{(7)}$ $^2T_3$ (South or C2'-endo-C3'-exo) and $^{(8)}$ $^3T_2$ (North or C2'-exo-C3'-endo) conformations for 5-MR sugars. $^{(9)1}C_4$ and $^{(10)4}C_1$ conformations for 6-MR sugars. $^{(11)}$ The $\Delta\Delta$ values are the Gibbs energies of reaction along a given two-step pathway for the β- minus the same pathway but for the α-anomer. For example, from the pathways labeled in **Figure 4.5**, $\Delta\Delta G^\circ_{Rx(a+b)}$ is the difference of $\left(\Delta G^\circ_{a+b}\right)_\beta - \left(\Delta G^\circ_{a+b}\right)_\alpha$. In turn, $\left(\Delta G^\circ_{a+b}\right)_\beta$, for instance, is the sum of the Gibbs energies for the condensation reaction along the pathway (a) then (b) yielding the nucleotide. Expressed symbolically, $\left(\Delta G^\circ_{a+b}\right)_\beta = \left(\Delta G^\circ_a\right)_\beta + \left(\Delta G^\circ_b\right)_\beta$. Hence, in general: $\Delta\Delta G^\circ_{i+j} = \left(\Delta G^\circ_{i+j}\right)_\beta - \left(\Delta G^\circ_{i+j}\right)_\alpha$, where i = a, c, and j = b, d. $^{(12)}$ Differences in ΔG° in vacuum at the DFT level. $^{(13)}$ Differences in ΔG° in solvent.

Table A5.

Table A5. Values in (°) for the torsion angle χ that governs the conformation of the RUs around the TC in the glycosidic bond of the nucleotides obtained through the classic (a+b) and alternative (c+d) pathways. 2dRibf: 2'-deoxyribofuranose, Ribf: ribofuranose, Tho: threose, 2dRib: 2'-deoxyribopyranose and Rib: ribopyranose.

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
					classic pathway (a+b)					
					HPO_3^-					
	2dRibf		**Ribf**		**Tho**		**2dRib**		**Rib**	
	$\alpha\text{-}^2T_3$ [4]		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$ [5]		$\alpha\text{-}^1C_4$	
A	96.7	294.0	169.5	300.7	73.7	295.4	115.4	289.2	109.9	117.0
G	296.1	103.8	168.7	302.1	105.8	287.2	291.9	293.0	107.9	111.6
C	291.0	293.4	162.9	303.9	300.8	296.1	122.6	290.0	129.1	136.7
T	293.9	294.1	139.3	303.1	300.5	297.3	118.3	119.4	124.6	133.4
U	294.2	294.3	140.7	304.9	300.6	297.5	118.5	119.6	124.8	133.3
TAP-C⁵	65.2	296.7	71.7	74.6	220.2	337.1	303.8	309.1	234.6	65.6
TAP-N	74.0	75.6	72.5	91.2	76.2	76.6	72.6	75.2	73.6	74.1
BA-C⁵	276.1	294.6	250.7	304.7	142.5	297.3	122.8	292.6	88.5	299.3
BA-N	299.0	294.9	316.3	307.1	297.0	295.7	103.4	291.1	136.8	135.6
CA	298.6	296.5	141.8	308.0	298.1	298.1	289.7	292.7	137.6	136.3
MM	297.5	295.3	306.7	86.2	281.6	290.7	302.0	84.0	87.1	84.3
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	76.5	70.2	180.2	67.5	53.9	88.1	183.2	287.5	176.5	76.4
G	76.7	72.5	74.1	71.4	161.0	56.4	171.0	286.8	124.9	111.8
C	207.4	72.8	197.1	70.9	53.3	56.1	70.6	284.6	63.8	197.6
T	215.8	209.3	198.1	70.4	51.8	220.1	69.1	285.8	63.5	93.5
U	237.2	209.6	198.3	70.4	51.7	218.9	69.1	286.2	63.7	74.4
TAP-C⁵	212.6	225.4	20.0	29.7	205.0	300.4	137.3	307.0	28.6	33.0
TAP-N	216.6	253.7	292.8	275.4	212.4	281.2	289.7	285.8	290.0	284.4
BA-C⁵	242.1	71.6	194.0	66.9	207.7	56.3	160.3	112.4	148.8	105.2
BA-N	290.4	251.2	75.7	72.1	234.9	250.0	243.8	257.7	71.9	260.3
CA	294.0	257.3	69.5	71.6	234.9	250.5	246.5	258.4	69.1	73.4

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
MM	70.7	72.4	66.1	67.4	49.2	273.9	280.0	279.0	280.6	278.5
	$\alpha\text{-}^3T_2$[4]		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^4C_1$[5]		$\alpha\text{-}^4C_1$	
A	294.3	295.0	173.1	300.7	103.0	295.6	90.6	279.2	201.6	212.9
G	296.4	103.8	295.3	302.1	105.8	296.0	312.0	217.0	207.4	297.2
C	288.3	154.4	173.6	303.9	156.2	296.6	90.5	290.3	142.3	222.8
T	289.5	127.0	173.3	304.7	114.0	297.4	89.4	220.3	333.7	219.2
U	292.4	132.2	137.5	304.9	114.2	297.6	88.8	291.0	333.8	292.9
TAP-C^5	301.9	318.2	335.0	71.9	158.1	153.4	43.0	45.9	57.1	222.8
TAP-N	73.6	75.6	74.7	73.4	71.7	76.3	72.4	71.8	80.1	77.6
BA-C^5	132.1	294.4	177.6	302.7	142.6	297.2	191.1	290.3	204.8	234.8
BA-N	298.1	114.0	317.3	307.1	111.9	297.4	340.1	292.8	294.4	294.0
CA	298.6	296.5	135.8	119.6	298.2	298.5	338.0	293.5	120.5	130.7
MM	297.5	83.7	293.5	310.3	86.0	282.7	302.7	204.9	202.9	204.3
	$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^4C_1$		$\beta\text{-}^4C_1$	
A	217.0	70.2	180.2	70.8	49.5	68.4	243.1	256.4	114.4	140.9
G	59.5	64.6	197.6	64.3	257.5	191.8	239.7	257.3	240.3	187.0
C	215.0	192.1	192.8	195.0	246.6	197.1	223.1	236.2	240.7	205.5
T	211.3	192.1	194.9	194.8	246.0	197.5	234.1	239.0	238.9	242.2
U	212.9	206.9	196.9	194.8	245.7	218.9	223.5	239.2	238.7	238.4
TAP-C^5	35.5	225.4	199.8	35.0	205.0	300.4	213.4	223.1	34.1	230.2
TAP-N	287.2	285.5	290.1	275.1	212.4	281.0	286.1	285.3	277.1	280.9
BA-C^5	243.4	69.2	155.7	71.0	207.7	56.3	205.1	246.5	215.7	110.0
BA-N	290.1	252.0	289.5	254.5	234.9	250.0	254.4	251.8	92.5	252.2
CA	91.4	67.4	92.6	69.0	235.7	250.5	66.3	66.6	276.0	66.4
MM	62.0	281.0	193.6	189.1	277.4	273.9	274.9	276.2	204.1	204.5

HAsO$_3^-$

	2dRibf		Ribf		Tho		2dRib		Rib	
	$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	98.4	294.3	171.6	257.9	74.0	295.5	115.7	289.6	110.4	113.5
G	296.2	105.7	169.8	302.1	105.9	297.2	292.8	291.5	109.3	113.6
C	288.1	293.6	163.6	303.1	300.6	296.3	123.2	290.1	129.6	140.0

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	289.6	294.3	163.3	304.1	300.4	297.3	118.5	118.2	125.1	133.4
U	289.6	294.7	162.9	304.1	300.4	297.5	122.3	119.3	125.3	134.1
TAP-C^5	105.5	304.7	56.5	106.6	220.0	336.1	300.7	304.7	235.1	61.8
TAP-N	74.2	65.9	72.8	75.3	76.2	75.8	72.8	75.0	73.6	75.8
BA-C^5	171.8	295.0	256.7	304.7	141.5	297.4	114.5	292.2	88.9	319.2
BA-N	298.3	294.9	280.3	306.6	296.8	294.9	104.3	291.6	137.2	136.5
CA	298.7	296.2	131.3	309.1	298.0	298.7	290.5	292.6	138.0	130.9
MM	297.5	295.6	303.0	84.6	281.2	298.8	301.7	83.9	87.5	83.7
	β-$^2T_3$		β-$^2T_3$		β-$^2T_3$		β-$^1C_4$		β-$^1C_4$	
A	76.5	209.6	177.1	68.8	55.1	188.8	182.7	288.0	180.2	94.2
G	75.7	55.4	77.5	63.6	161.8	56.3	176.1	286.7	126.4	121.8
C	228.6	69.4	193.7	71.1	52.7	55.9	70.4	285.4	64.0	197.7
T	237.2	240.0	195.0	71.2	51.6	220.1	68.8	285.9	246.7	93.4
U	210.1	206.1	195.0	71.3	51.5	218.0	68.9	285.7	246.6	92.5
TAP-C^5	213.5	225.8	20.4	44.8	205.0	208.1	137.3	306.7	28.8	33.2
TAP-N	283.6	257.4	266.5	261.5	212.7	278.0	289.8	285.9	289.9	284.2
BA-C^5	221.6	60.8	196.3	67.5	207.7	57.4	160.5	108.2	149.0	105.7
BA-N	289.4	247.2	75.0	70.6	236.1	249.7	243.6	256.6	72.2	259.1
CA	257.1	257.6	69.4	69.2	235.9	250.2	245.3	258.1	69.4	75.5
MM	77.6	75.5	66.8	69.3	49.3	276.6	279.7	279.4	279.7	276.4
	α-$^3T_2$		α-$^3T_2$		α-$^3T_2$		α-$^4C_1$		α-$^4C_1$	
A	294.1	294.8	203.6	258.0	101.1	295.3	90.3	278.4	201.0	217.7
G	296.2	90.0	295.8	302.1	105.4	271.9	303.9	221.3	206.4	296.6
C	288.1	161.0	256.0	303.1	110.6	296.3	91.1	287.2	140.9	227.3
T	289.6	128.6	140.1	303.9	110.7	297.3	89.8	226.6	333.6	226.2
U	289.6	126.6	140.6	304.2	114.3	297.5	89.2	287.3	333.6	292.6
TAP-C^5	303.1	308.9	295.0	63.3	157.7	153.5	44.2	47.4	57.5	224.8
TAP-N	72.1	73.9	58.0	70.7	71.3	76.5	72.5	71.7	80.4	77.8
BA-C^5	265.6	297.1	242.7	302.2	141.5	297.3	190.9	276.8	204.7	242.1
BA-N	298.3	111.9	320.4	306.9	111.7	297.0	340.0	287.9	293.6	296.5
CA	298.7	296.1	34.8	119.2	298.1	298.7	338.1	288.2	119.1	114.1
MM	297.5	82.8	294.4	310.8	90.1	283.9	301.3	207.1	200.6	207.9

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
	β-$^3T_2$		β-$^3T_2$		β-$^3T_2$		β-$^{-4}C_1$		β-$^{-4}C_1$	
A	213.0	69.8	180.9	70.6	50.0	67.4	240.1	256.8	111.9	146.6
G	72.6	66.8	167.7	67.0	255.0	190.9	238.0	261.1	241.3	185.1
C	238.3	204.2	196.9	195.5	242.0	195.7	217.2	235.6	234.5	204.2
T	212.3	212.7	196.6	188.5	244.3	197.1	232.0	239.2	235.0	242.3
U	212.0	192.8	200.2	188.0	241.4	218.0	230.6	239.2	234.3	238.3
TAP-C^5	35.4	229.3	199.3	43.4	205.0	207.4	213.6	222.3	33.3	230.1
TAP-N	282.9	291.3	284.5	281.9	212.7	280.7	286.4	285.2	275.4	281.1
BA-C^5	200.5	69.3	143.5	71.1	207.7	57.4	204.2	245.3	215.8	111.0
BA-N	289.7	252.5	289.2	254.4	236.1	249.7	254.7	251.7	91.7	251.8
CA	64.8	66.8	67.5	64.9	235.9	250.2	67.6	66.8	276.8	66.9
MM	68.0	278.4	193.4	192.3	269.4	276.6	274.5	276.2	268.2	203.6

alternative pathway (c+d)

HPO_3^-

	2dRibf		Ribf		Tho		2dRib		Rib	
	α-$^2T_3$		α-$^2T_3$		α-$^2T_3$		α-$^1C_4$		α-$^1C_4$	
A	296.0	104.8	121.4	129.1	101.6	114.1	116.8	109.0	111.5	101.4
G	296.4	102.5	116.6	126.8	292.8	271.0	294.3	111.9	303.6	313.9
C	294.1	293.2	147.3	303.9	112.9	294.0	128.4	275.9	131.7	131.0
T	295.0	294.6	139.8	142.6	114.0	294.5	122.4	274.0	126.8	148.1
U	113.4	294.5	140.8	140.4	114.2	294.9	123.0	274.6	127.0	148.2
TAP-C^5	148.2	295.5	137.6	111.6	339.7	333.9	128.0	294.8	235.1	235.0
TAP-N	74.3	72.6	72.2	74.0	72.9	166.0	300.2	77.6	73.6	75.3
BA-C^5	276.1	294.5	283.6	304.4	55.4	118.4	125.3	267.4	105.5	316.0
BA-N	297.1	295.6	139.9	306.9	111.9	112.3	103.4	266.3	136.8	135.6
CA	297.8	296.2	141.8	119.6	298.2	297.8	101.0	115.8	137.6	136.4
MM	296.4	295.3	87.6	78.9	86.0	169.8	302.0	84.6	82.4	82.3
	β-$^2T_3$		β-$^2T_3$		β-$^2T_3$		β-$^1C_4$		β-$^1C_4$	
A	241.5	221.4	250.9	215.6	258.4	238.8	272.7	287.8	72.2	76.4
G	56.2	54.1	70.1	65.5	252.8	233.7	171.2	286.8	129.4	281.5
C	200.1	192.1	199.9	65.5	244.8	56.1	201.4	195.8	71.5	197.6

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	219.3	192.1	245.4	211.0	243.7	55.2	201.3	285.7	71.1	76.5
U	247.5	191.9	245.5	244.9	243.9	218.9	200.9	286.2	70.8	76.3
TAP-C^5	118.1	226.6	24.4	44.3	196.9	213.3	41.5	226.6	28.8	211.2
TAP-N	75.3	253.7	69.2	257.4	190.5	281.2	289.7	285.8	289.6	284.4
BA-C^5	243.4	221.2	202.8	249.8	64.4	56.3	205.9	110.1	21.1	103.9
BA-N	287.2	251.2	65.4	65.2	57.1	239.1	243.8	257.8	71.9	258.7
CA	61.0	67.3	280.9	68.6	57.4	239.2	62.3	75.3	70.6	75.4
MM	240.0	249.2	64.4	76.7	269.9	49.0	277.6	279.2	279.7	278.5
	$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	104.7	293.9	101.8	300.7	101.6	295.6	89.1	81.5	122.6	117.6
G	296.2	294.6	101.6	302.1	292.8	295.5	86.6	284.6	114.3	113.4
C	290.4	137.3	103.5	303.9	112.9	123.2	127.3	82.5	255.4	89.0
T	291.7	126.3	104.4	304.7	114.0	123.0	89.8	82.2	281.2	90.9
U	291.9	132.2	103.9	304.9	114.2	123.9	89.3	81.6	281.7	89.6
TAP-C^5	310.6	304.7	56.3	62.6	339.7	333.2	227.6	58.5	42.0	226.1
TAP-N	85.4	75.6	73.7	79.8	72.9	76.2	72.5	71.8	68.2	77.8
BA-C^5	284.5	297.4	110.5	302.6	55.4	297.2	98.3	290.3	255.0	99.9
BA-N	307.3	115.3	136.9	306.9	111.9	113.3	156.8	292.8	265.0	118.8
CA	120.8	296.2	312.7	119.0	298.2	298.5	337.8	148.9	129.9	130.7
MM	299.1	83.7	85.5	78.9	86.0	86.3	83.5	81.7	85.8	87.0
	$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^4C_1$		$\beta\text{-}^4C_1$	
A	268.1	216.3	177.3	213.8	258.4	239.4	243.1	256.4	243.1	250.6
G	73.2	69.6	61.7	68.9	91.0	232.8	239.9	257.3	239.8	248.9
C	238.2	205.4	187.8	200.5	244.8	195.8	223.1	236.2	237.5	240.0
T	242.5	211.9	213.5	193.2	243.7	197.1	234.1	239.0	236.9	239.1
U	226.2	209.6	214.9	205.4	243.9	57.6	233.0	239.1	236.7	237.9
TAP-C^5	118.1	229.6	22.8	211.1	302.6	35.5	213.6	222.9	303.7	220.7
TAP-N	278.4	285.5	201.6	283.5	287.8	280.9	286.1	285.3	62.5	281.0
BA-C^5	245.8	244.5	207.2	197.3	293.5	56.2	205.1	246.5	215.7	247.8
BA-N	290.1	72.0	43.7	71.2	233.8	237.7	254.4	251.8	251.9	251.6
CA	171.7	72.5	66.4	257.8	57.4	55.6	67.5	66.6	67.4	66.4
MM	69.1	280.9	193.2	279.9	279.1	60.9	274.9	276.2	272.7	272.2

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
					$HAsO_3^-$					
	2dRibf		**Ribf**		**Tho**		**2dRib**		**Rib**	
	$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	290.5	104.1	104.7	301.3	101.9	96.7	121.6	107.8	111.3	121.4
G	296.2	103.5	316.5	303.3	273.9	298.5	289.2	287.8	303.9	120.0
C	293.2	293.1	295.8	144.6	113.0	294.0	280.5	275.5	131.6	284.6
T	114.4	294.7	292.9	304.1	112.7	123.0	278.4	274.4	126.8	281.6
U	114.7	294.5	299.6	304.0	112.8	295.2	278.8	275.0	126.9	282.1
TAP-C[5]	149.0	135.8	150.0	135.9	339.3	336.1	283.8	119.4	235.1	49.3
TAP-N	71.1	72.8	72.2	75.3	72.7	74.7	303.7	306.4	73.5	79.8
BA-C[5]	119.2	294.3	102.1	304.8	55.3	153.7	280.7	267.2	106.1	280.3
BA-N	298.3	295.1	139.7	120.0	111.7	294.8	267.7	266.3	137.2	275.7
CA	302.3	296.2	305.6	301.7	298.1	296.0	108.2	115.3	138.0	277.1
MM	93.2	295.3	87.5	80.9	282.9	165.7	284.4	292.6	85.1	84.2
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	221.6	225.9	189.7	228.4	68.3	70.4	180.3	283.3	72.0	175.9
G	59.8	55.4	77.3	55.2	74.4	69.7	171.7	283.1	183.4	130.2
C	215.4	193.0	196.9	195.5	247.6	200.4	60.0	53.2	71.9	251.4
T	222.5	224.5	196.6	191.6	244.2	57.8	55.6	47.4	71.4	294.3
U	222.1	192.8	195.4	191.6	245.1	218.0	54.7	47.8	71.3	294.7
TAP-C[5]	119.1	225.8	202.0	30.3	10.1	214.3	10.6	34.9	28.8	307.2
TAP-N	58.5	257.3	64.3	261.5	287.2	278.7	107.4	303.9	289.4	129.9
BA-C[5]	243.7	245.0	0.8	245.2	266.5	63.7	263.5	264.3	21.7	116.7
BA-N	249.5	252.5	281.6	254.4	232.2	229.9	89.5	102.4	72.2	256.1
CA	61.3	252.6	211.9	254.9	53.5	250.1	247.6	244.4	71.0	252.6
MM	246.5	251.4	193.1	253.5	277.8	272.0	282.0	283.5	279.7	284.8
	$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	294.3	76.2	121.4	253.9	101.9	295.3	91.5	72.8	125.4	112.2
G	296.1	71.1	339.4	289.0	273.9	295.6	303.9	294.4	125.4	112.2
C	289.9	293.3	147.5	115.6	113.0	123.5	122.0	76.9	304.1	296.6
T	291.2	293.7	139.6	90.3	112.7	122.9	93.3	76.2	97.9	92.6

433

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
U	291.4	294.4	140.6	90.8	112.8	122.4	92.6	83.7	138.5	94.8
TAP-C^5	288.4	106.6	54.6	264.7	339.3	333.2	228.2	55.3	138.5	94.1
TAP-N	72.4	84.7	72.8	69.6	72.7	76.5	74.2	71.9	53.4	241.7
BA-C^5	89.5	281.3	110.8	100.4	55.3	297.4	98.1	276.8	79.3	77.9
BA-N	297.8	115.2	139.4	117.2	111.7	113.2	156.9	287.9	68.0	100.5
CA	298.0	297.4	80.3	116.3	298.1	298.7	337.1	288.2	126.5	113.4
MM	89.1	75.8	85.0	69.6	282.9	85.9	86.8	296.6	318.9	134.5
	β-$^3T_2$		β-$^3T_2$		β-$^3T_2$		β-$^4C_1$		β-$^4C_1$	
A	221.6	246.1	246.3	170.9	257.5	237.9	240.1	256.8	241.3	250.0
G	59.8	27.0	23.7	70.2	87.8	231.3	238.1	257.1	241.3	249.3
C	215.4	205.3	192.8	189.7	244.0	195.7	217.2	235.6	234.4	239.7
T	222.5	213.9	239.8	188.5	243.3	224.0	232.0	239.2	235.0	238.7
U	240.8	241.5	214.4	188.3	243.2	226.7	230.6	239.2	234.3	237.4
TAP-C^5	119.1	217.9	24.4	34.8	10.5	298.5	213.6	222.6	123.2	220.9
TAP-N	58.5	61.9	61.4	292.4	287.6	256.3	285.9	285.2	62.4	281.1
BA-C^5	243.7	245.0	186.4	193.8	103.9	235.4	204.2	245.2	23.5	68.5
BA-N	242.9	255.6	253.3	65.5	233.9	102.5	254.6	251.7	251.0	251.2
CA	245.5	252.6	66.2	64.7	57.7	171.9	67.6	66.8	251.2	66.2
MM	246.5	71.8	211.9	288.0	279.0	57.3	274.5	276.2	272.7	272.5

[1] RU = unspecified **r**ecognition **u**nit. [2] χ in vacuum at the DFT level. [3] χ in implicit solvent at the DFT level using the IEFPCM model. [4] initial sugar ring puckering-anomer for 5-MR. [5] initial sugar ring puckering-anomer for 6-MR.

Table A6. Values in (°) for the torsion angle χ that governs the conformation of the RUs around the TC in the glycosidic bond of the nucleotides obtained through the classic (a+b) and alternative (c+d) pathways. 2dRibf: 2'-deoxyribofuranose, Ribf: ribofuranose, Tho: threose, 2dRib: 2'-deoxyribopyranose and Rib: ribopyranose.

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
					classic pathway (a+b)					
					HPO_3^-					
	2dRibf		**Ribf**		**Tho**		**2dRib**		**Rib**	
	$\alpha\text{-}^2T_3$ [4]		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$ [5]		$\alpha\text{-}^1C_4$	
A	96.7	294.0	169.5	300.7	73.7	295.4	115.4	289.2	109.9	117.0
G	296.1	103.8	168.7	302.1	105.8	287.2	291.9	293.0	107.9	111.6
C	291.0	293.4	162.9	303.9	300.8	296.1	122.6	290.0	129.1	136.7
T	293.9	294.1	139.3	303.1	300.5	297.3	118.3	119.4	124.6	133.4
U	294.2	294.3	140.7	304.9	300.6	297.5	118.5	119.6	124.8	133.3
TAP-C^5	65.2	296.7	71.7	74.6	220.2	337.1	303.8	309.1	234.6	65.6
TAP-N	74.0	75.6	72.5	91.2	76.2	76.6	72.6	75.2	73.6	74.1
BA-C^5	276.1	294.6	250.7	304.7	142.5	297.3	122.8	292.6	88.5	299.3
BA-N	299.0	294.9	316.3	307.1	297.0	295.7	103.4	291.1	136.8	135.6
CA	298.6	296.5	141.8	308.0	298.1	298.1	289.7	292.7	137.6	136.3
MM	297.5	295.3	306.7	86.2	281.6	290.7	302.0	84.0	87.1	84.3
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	76.5	70.2	180.2	67.5	53.9	88.1	183.2	287.5	176.5	76.4
G	76.7	72.5	74.1	71.4	161.0	56.4	171.0	286.8	124.9	111.8
C	207.4	72.8	197.1	70.9	53.3	56.1	70.6	284.6	63.8	197.6
T	215.8	209.3	198.1	70.4	51.8	220.1	69.1	285.8	63.5	93.5
U	237.2	209.6	198.3	70.4	51.7	218.9	69.1	286.2	63.7	74.4
TAP-C^5	212.6	225.4	20.0	29.7	205.0	300.4	137.3	307.0	28.6	33.0
TAP-N	216.6	253.7	292.8	275.4	212.4	281.2	289.7	285.8	290.0	284.4
BA-C^5	242.1	71.6	194.0	66.9	207.7	56.3	160.3	112.4	148.8	105.2
BA-N	290.4	251.2	75.7	72.1	234.9	250.0	243.8	257.7	71.9	260.3
CA	294.0	257.3	69.5	71.6	234.9	250.5	246.5	258.4	69.1	73.4

435

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
MM	70.7	72.4	66.1	67.4	49.2	273.9	280.0	279.0	280.6	278.5
	α-$^3T_2$[4]		α-$^3T_2$		α-$^3T_2$		α-$^4C_1$[5]		α-$^4C_1$	
A	294.3	295.0	173.1	300.7	103.0	295.6	90.6	279.2	201.6	212.9
G	296.4	103.8	295.3	302.1	105.8	296.0	312.0	217.0	207.4	297.2
C	288.3	154.4	173.6	303.9	156.2	296.6	90.5	290.3	142.3	222.8
T	289.5	127.0	173.3	304.7	114.0	297.4	89.4	220.3	333.7	219.2
U	292.4	132.2	137.5	304.9	114.2	297.6	88.8	291.0	333.8	292.9
TAP-C^5	301.9	318.2	335.0	71.9	158.1	153.4	43.0	45.9	57.1	222.8
TAP-N	73.6	75.6	74.7	73.4	71.7	76.3	72.4	71.8	80.1	77.6
BA-C^5	132.1	294.4	177.6	302.7	142.6	297.2	191.1	290.3	204.8	234.8
BA-N	298.1	114.0	317.3	307.1	111.9	297.4	340.1	292.8	294.4	294.0
CA	298.6	296.5	135.8	119.6	298.2	298.5	338.0	293.5	120.5	130.7
MM	297.5	83.7	293.5	310.3	86.0	282.7	302.7	204.9	202.9	204.3
	β-$^3T_2$		β-$^3T_2$		β-$^3T_2$		β-$^4C_1$		β-$^4C_1$	
A	217.0	70.2	180.2	70.8	49.5	68.4	243.1	256.4	114.4	140.9
G	59.5	64.6	197.6	64.3	257.5	191.8	239.7	257.3	240.3	187.0
C	215.0	192.1	192.8	195.0	246.6	197.1	223.1	236.2	240.7	205.5
T	211.3	192.1	194.9	194.8	246.0	197.5	234.1	239.0	238.9	242.2
U	212.9	206.9	196.9	194.8	245.7	218.9	223.5	239.2	238.7	238.4
TAP-C^5	35.5	225.4	199.8	35.0	205.0	300.4	213.4	223.1	34.1	230.2
TAP-N	287.2	285.5	290.1	275.1	212.4	281.0	286.1	285.3	277.1	280.9
BA-C^5	243.4	69.2	155.7	71.0	207.7	56.3	205.1	246.5	215.7	110.0
BA-N	290.1	252.0	289.5	254.5	234.9	250.0	254.4	251.8	92.5	252.2
CA	91.4	67.4	92.6	69.0	235.7	250.5	66.3	66.6	276.0	66.4
MM	62.0	281.0	193.6	189.1	277.4	273.9	274.9	276.2	204.1	204.5

$HAsO_3^-$

	2dRibf		Ribf		Tho		2dRib		Rib	
	α-$^2T_3$		α-$^2T_3$		α-$^2T_3$		α-$^1C_4$		α-$^1C_4$	
A	98.4	294.3	171.6	257.9	74.0	295.5	115.7	289.6	110.4	113.5
G	296.2	105.7	169.8	302.1	105.9	297.2	292.8	291.5	109.3	113.6
C	288.1	293.6	163.6	303.1	300.6	296.3	123.2	290.1	129.6	140.0

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	289.6	294.3	163.3	304.1	300.4	297.3	118.5	118.2	125.1	133.4
U	289.6	294.7	162.9	304.1	300.4	297.5	122.3	119.3	125.3	134.1
TAP-C^5	105.5	304.7	56.5	106.6	220.0	336.1	300.7	304.7	235.1	61.8
TAP-N	74.2	65.9	72.8	75.3	76.2	75.8	72.8	75.0	73.6	75.8
BA-C^5	171.8	295.0	256.7	304.7	141.5	297.4	114.5	292.2	88.9	319.2
BA-N	298.3	294.9	280.3	306.6	296.8	294.9	104.3	291.6	137.2	136.5
CA	298.7	296.2	131.3	309.1	298.0	298.7	290.5	292.6	138.0	130.9
MM	297.5	295.6	303.0	84.6	281.2	298.8	301.7	83.9	87.5	83.7
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	76.5	209.6	177.1	68.8	55.1	188.8	182.7	288.0	180.2	94.2
G	75.7	55.4	77.5	63.6	161.8	56.3	176.1	286.7	126.4	121.8
C	228.6	69.4	193.7	71.1	52.7	55.9	70.4	285.4	64.0	197.7
T	237.2	240.0	195.0	71.2	51.6	220.1	68.8	285.9	246.7	93.4
U	210.1	206.1	195.0	71.3	51.5	218.0	68.9	285.7	246.6	92.5
TAP-C^5	213.5	225.8	20.4	44.8	205.0	208.1	137.3	306.7	28.8	33.2
TAP-N	283.6	257.4	266.5	261.5	212.7	278.0	289.8	285.9	289.9	284.2
BA-C^5	221.6	60.8	196.3	67.5	207.7	57.4	160.5	108.2	149.0	105.7
BA-N	289.4	247.2	75.0	70.6	236.1	249.7	243.6	256.6	72.2	259.1
CA	257.1	257.6	69.4	69.2	235.9	250.2	245.3	258.1	69.4	75.5
MM	77.6	75.5	66.8	69.3	49.3	276.6	279.7	279.4	279.7	276.4
	$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	294.1	294.8	203.6	258.0	101.1	295.3	90.3	278.4	201.0	217.7
G	296.2	90.0	295.8	302.1	105.4	271.9	303.9	221.3	206.4	296.6
C	288.1	161.0	256.0	303.1	110.6	296.3	91.1	287.2	140.9	227.3
T	289.6	128.6	140.1	303.9	110.7	297.3	89.8	226.6	333.6	226.2
U	289.6	126.6	140.6	304.2	114.3	297.5	89.2	287.3	333.6	292.6
TAP-C^5	303.1	308.9	295.0	63.3	157.7	153.5	44.2	47.4	57.5	224.8
TAP-N	72.1	73.9	58.0	70.7	71.3	76.5	72.5	71.7	80.4	77.8
BA-C^5	265.6	297.1	242.7	302.2	141.5	297.3	190.9	276.8	204.7	242.1
BA-N	298.3	111.9	320.4	306.9	111.7	297.0	340.0	287.9	293.6	296.5
CA	298.7	296.1	34.8	119.2	298.1	298.7	338.1	288.2	119.1	114.1
MM	297.5	82.8	294.4	310.8	90.1	283.9	301.3	207.1	200.6	207.9

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
	$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^{-4}C_1$		$\beta\text{-}^{-4}C_1$	
A	213.0	69.8	180.9	70.6	50.0	67.4	240.1	256.8	111.9	146.6
G	72.6	66.8	167.7	67.0	255.0	190.9	238.0	261.1	241.3	185.1
C	238.3	204.2	196.9	195.5	242.0	195.7	217.2	235.6	234.5	204.2
T	212.3	212.7	196.6	188.5	244.3	197.1	232.0	239.2	235.0	242.3
U	212.0	192.8	200.2	188.0	241.4	218.0	230.6	239.2	234.3	238.3
TAP-C[5]	35.4	229.3	199.3	43.4	205.0	207.4	213.6	222.3	33.3	230.1
TAP-N	282.9	291.3	284.5	281.9	212.7	280.7	286.4	285.2	275.4	281.1
BA-C[5]	200.5	69.3	143.5	71.1	207.7	57.4	204.2	245.3	215.8	111.0
BA-N	289.7	252.5	289.2	254.4	236.1	249.7	254.7	251.7	91.7	251.8
CA	64.8	66.8	67.5	64.9	235.9	250.2	67.6	66.8	276.8	66.9
MM	68.0	278.4	193.4	192.3	269.4	276.6	274.5	276.2	268.2	203.6

alternative pathway (c+d)

HPO₃⁻

	2dRibf		Ribf		Tho		2dRib		Rib	
	$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	296.0	104.8	121.4	129.1	101.6	114.1	116.8	109.0	111.5	101.4
G	296.4	102.5	116.6	126.8	292.8	271.0	294.3	111.9	303.6	313.9
C	294.1	293.2	147.3	303.9	112.9	294.0	128.4	275.9	131.7	131.0
T	295.0	294.6	139.8	142.6	114.0	294.5	122.4	274.0	126.8	148.1
U	113.4	294.5	140.8	140.4	114.2	294.9	123.0	274.6	127.0	148.2
TAP-C[5]	148.2	295.5	137.6	111.6	339.7	333.9	128.0	294.8	235.1	235.0
TAP-N	74.3	72.6	72.2	74.0	72.9	166.0	300.2	77.6	73.6	75.3
BA-C[5]	276.1	294.5	283.6	304.4	55.4	118.4	125.3	267.4	105.5	316.0
BA-N	297.1	295.6	139.9	306.9	111.9	112.3	103.4	266.3	136.8	135.6
CA	297.8	296.2	141.8	119.6	298.2	297.8	101.0	115.8	137.6	136.4
MM	296.4	295.3	87.6	78.9	86.0	169.8	302.0	84.6	82.4	82.3
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	241.5	221.4	250.9	215.6	258.4	238.8	272.7	287.8	72.2	76.4
G	56.2	54.1	70.1	65.5	252.8	233.7	171.2	286.8	129.4	281.5
C	200.1	192.1	199.9	65.5	244.8	56.1	201.4	195.8	71.5	197.6

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	219.3	192.1	245.4	211.0	243.7	55.2	201.3	285.7	71.1	76.5
U	247.5	191.9	245.5	244.9	243.9	218.9	200.9	286.2	70.8	76.3
TAP-C^5	118.1	226.6	24.4	44.3	196.9	213.3	41.5	226.6	28.8	211.2
TAP-N	75.3	253.7	69.2	257.4	190.5	281.2	289.7	285.8	289.6	284.4
BA-C^5	243.4	221.2	202.8	249.8	64.4	56.3	205.9	110.1	21.1	103.9
BA-N	287.2	251.2	65.4	65.2	57.1	239.1	243.8	257.8	71.9	258.7
CA	61.0	67.3	280.9	68.6	57.4	239.2	62.3	75.3	70.6	75.4
MM	240.0	249.2	64.4	76.7	269.9	49.0	277.6	279.2	279.7	278.5
	α-$^3T_2$		α-$^3T_2$		α-$^3T_2$		α-$^4C_1$		α-$^4C_1$	
A	104.7	293.9	101.8	300.7	101.6	295.6	89.1	81.5	122.6	117.6
G	296.2	294.6	101.6	302.1	292.8	295.5	86.6	284.6	114.3	113.4
C	290.4	137.3	103.5	303.9	112.9	123.2	127.3	82.5	255.4	89.0
T	291.7	126.3	104.4	304.7	114.0	123.0	89.8	82.2	281.2	90.9
U	291.9	132.2	103.9	304.9	114.2	123.9	89.3	81.6	281.7	89.6
TAP-C^5	310.6	304.7	56.3	62.6	339.7	333.2	227.6	58.5	42.0	226.1
TAP-N	85.4	75.6	73.7	79.8	72.9	76.2	72.5	71.8	68.2	77.8
BA-C^5	284.5	297.4	110.5	302.6	55.4	297.2	98.3	290.3	255.0	99.9
BA-N	307.3	115.3	136.9	306.9	111.9	113.3	156.8	292.8	265.0	118.8
CA	120.8	296.2	312.7	119.0	298.2	298.5	337.8	148.9	129.9	130.7
MM	299.1	83.7	85.5	78.9	86.0	86.3	83.5	81.7	85.8	87.0
	β-$^3T_2$		β-$^3T_2$		β-$^3T_2$		β-$^4C_1$		β-$^4C_1$	
A	268.1	216.3	177.3	213.8	258.4	239.4	243.1	256.4	243.1	250.6
G	73.2	69.6	61.7	68.9	91.0	232.8	239.9	257.3	239.8	248.9
C	238.2	205.4	187.8	200.5	244.8	195.8	223.1	236.2	237.5	240.0
T	242.5	211.9	213.5	193.2	243.7	197.1	234.1	239.0	236.9	239.1
U	226.2	209.6	214.9	205.4	243.9	57.6	233.0	239.1	236.7	237.9
TAP-C^5	118.1	229.6	22.8	211.1	302.6	35.5	213.6	222.9	303.7	220.7
TAP-N	278.4	285.5	201.6	283.5	287.8	280.9	286.1	285.3	62.5	281.0
BA-C^5	245.8	244.5	207.2	197.3	293.5	56.2	205.1	246.5	215.7	247.8
BA-N	290.1	72.0	43.7	71.2	233.8	237.7	254.4	251.8	251.9	251.6
CA	171.7	72.5	66.4	257.8	57.4	55.6	67.5	66.6	67.4	66.4
MM	69.1	280.9	193.2	279.9	279.1	60.9	274.9	276.2	272.7	272.2

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
					$HAsO_3^-$					
	2dRibf		**Ribf**		**Tho**		**2dRib**		**Rib**	
	$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	290.5	104.1	104.7	301.3	101.9	96.7	121.6	107.8	111.3	121.4
G	296.2	103.5	316.5	303.3	273.9	298.5	289.2	287.8	303.9	120.0
C	293.2	293.1	295.8	144.6	113.0	294.0	280.5	275.5	131.6	284.6
T	114.4	294.7	292.9	304.1	112.7	123.0	278.4	274.4	126.8	281.6
U	114.7	294.5	299.6	304.0	112.8	295.2	278.8	275.0	126.9	282.1
TAP-C^5	149.0	135.8	150.0	135.9	339.3	336.1	283.8	119.4	235.1	49.3
TAP-N	71.1	72.8	72.2	75.3	72.7	74.7	303.7	306.4	73.5	79.8
BA-C^5	119.2	294.3	102.1	304.8	55.3	153.7	280.7	267.2	106.1	280.3
BA-N	298.3	295.1	139.7	120.0	111.7	294.8	267.7	266.3	137.2	275.7
CA	302.3	296.2	305.6	301.7	298.1	296.0	108.2	115.3	138.0	277.1
MM	93.2	295.3	87.5	80.9	282.9	165.7	284.4	292.6	85.1	84.2
	$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	221.6	225.9	189.7	228.4	68.3	70.4	180.3	283.3	72.0	175.9
G	59.8	55.4	77.3	55.2	74.4	69.7	171.7	283.1	183.4	130.2
C	215.4	193.0	196.9	195.5	247.6	200.4	60.0	53.2	71.9	251.4
T	222.5	224.5	196.6	191.6	244.2	57.8	55.6	47.4	71.4	294.3
U	222.1	192.8	195.4	191.6	245.1	218.0	54.7	47.8	71.3	294.7
TAP-C^5	119.1	225.8	202.0	30.3	10.1	214.3	10.6	34.9	28.8	307.2
TAP-N	58.5	257.3	64.3	261.5	287.2	278.7	107.4	303.9	289.4	129.9
BA-C^5	243.7	245.0	0.8	245.2	266.5	63.7	263.5	264.3	21.7	116.7
BA-N	249.5	252.5	281.6	254.4	232.2	229.9	89.5	102.4	72.2	256.1
CA	61.3	252.6	211.9	254.9	53.5	250.1	247.6	244.4	71.0	252.6
MM	246.5	251.4	193.1	253.5	277.8	272.0	282.0	283.5	279.7	284.8
	$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	294.3	76.2	121.4	253.9	101.9	295.3	91.5	72.8		
G	296.1	71.1	339.4	289.0	273.9	295.6	303.9	294.4	125.4	112.2
C	289.9	293.3	147.5	115.6	113.0	123.5	122.0	76.9	304.1	296.6
T	291.2	293.7	139.6	90.3	112.7	122.9	93.3	76.2	97.9	92.6

RU[1]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
U	291.4	294.4	140.6	90.8	112.8	122.4	92.6	83.7	138.5	94.8
TAP-C^5	288.4	106.6	54.6	264.7	339.3	333.2	228.2	55.3	138.5	94.1
TAP-N	72.4	84.7	72.8	69.6	72.7	76.5	74.2	71.9	53.4	241.7
BA-C^5	89.5	281.3	110.8	100.4	55.3	297.4	98.1	276.8	79.3	77.9
BA-N	297.8	115.2	139.4	117.2	111.7	113.2	156.9	287.9	68.0	100.5
CA	298.0	297.4	80.3	116.3	298.1	298.7	337.1	288.2	126.5	113.4
MM	89.1	75.8	85.0	69.6	282.9	85.9	86.8	296.6	318.9	134.5
	$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^4C_1$		$\beta\text{-}^4C_1$	
A	221.6	246.1	246.3	170.9	257.5	237.9	240.1	256.8	241.3	250.0
G	59.8	27.0	23.7	70.2	87.8	231.3	238.1	257.1	241.3	249.3
C	215.4	205.3	192.8	189.7	244.0	195.7	217.2	235.6	234.4	239.7
T	222.5	213.9	239.8	188.5	243.3	224.0	232.0	239.2	235.0	238.7
U	240.8	241.5	214.4	188.3	243.2	226.7	230.6	239.2	234.3	237.4
TAP-C^5	119.1	217.9	24.4	34.8	10.5	298.5	213.6	222.6	123.2	220.9
TAP-N	58.5	61.9	61.4	292.4	287.6	256.3	285.9	285.2	62.4	281.1
BA-C^5	243.7	245.0	186.4	193.8	103.9	235.4	204.2	245.2	23.5	68.5
BA-N	242.9	255.6	253.3	65.5	233.9	102.5	254.6	251.7	251.0	251.2
CA	245.5	252.6	66.2	64.7	57.7	171.9	67.6	66.8	251.2	66.2
MM	246.5	71.8	211.9	288.0	279.0	57.3	274.5	276.2	272.7	272.5

[1] RU = unspecified **r**ecognition **u**nit. [2] χ in vacuum at the DFT level. [3] χ in implicit solvent at the DFT level using the IEFPCM model. [4] initial sugar ring puckering-anomer for 5-MR. [5] initial sugar ring puckering-anomer for 6-MR.

Table A7. Values for the Cremer-Pople (CP) puckering parameters ϕ_2 (phase angle) and Q (total puckering amplitude) that determine the sugar ring conformation for 5-MR and the polar coordinates ϕ (zenithal angle), θ (azimuthal angle) and Q (total puckering amplitude) that determine the sugar ring conformation for 6-MR of canonical and non-canonical mucleotides obtained from the classic (a+b) and an alternative (c+d) pathway.

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
								classic pathway HPO_3^-				
		2dRibf		**Ribf**		**Tho**			**2dRib**		**Rib**	
		$\alpha\text{-}^2T_3$[4]		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$			$\alpha\text{-}^1C_4$ [5]		$\alpha\text{-}^1C_4$	
A	ϕ_2	29.4	293.3	105.9	281.7	49.0	290.7	ϕ	67.7	-8.2	42.5	118.8
	Q	0.3	0.4	0.3	0.4	0.3	0.4	θ	174.6	178.2	177.3	178.4
								Q	0.5	0.6	0.6	0.6
G	ϕ_2	138.5	111.2	100.2	282.2	284.0	153.9	ϕ	38.8	50.9	26.3	134.9
	Q	0.3	0.3	0.3	0.4	0.4	0.4	θ	164.3	165.3	177.7	178.8
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	143.9	137.8	95.5	287.9	291.5	294.5	ϕ	88.7	-13.3	58.8	129.8
	Q	0.3	0.3	0.3	0.4	0.4	0.4	θ	176.3	177.7	176.5	178.2
								Q	0.5	0.6	0.6	0.6
T	ϕ_2	144.4	139.3	114.5	74.1	290.7	292.5	ϕ	75.1	5.4	49.5	173.8
	Q	0.3	0.3	0.3	0.2	0.4	0.4	θ	176.1	178.0	176.0	178.0
								Q	0.5	0.6	0.6	0.6
U	ϕ_2	144.8	139.9	113.5	286.7	290.6	292.5	ϕ	76.2	10.7	49.2	169.2
	Q	0.3	0.3	0.3	0.4	0.4	0.4	θ	176.5	177.8	176.1	178.1
								Q	0.5	0.6	0.6	0.6
TAP-C⁵	ϕ_2	20.8	150.4	230.2	238.7	34.2	162.5	ϕ	-127.6	60.8	-75.2	-93.4
	Q	0.4	0.3	0.5	0.4	0.4	0.4	θ	173.4	176.9	177.7	174.9
								Q	0.5	0.6	0.6	0.6
TAP-N	ϕ_2	41.9	284.2	313.4	155.6	2.2	291.1	ϕ	-10.8	-134.6	101.6	120.8
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	176.1	179.9	176.9	175.9
								Q	0.6	0.6	0.5	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
BA-C^5	ϕ_2	135.3	137.8	228.1	129.5	224.2	285.7	ϕ	59.3	-53.4	-139.9	57.6
	Q	0.4	0.3	0.4	0.3	0.4	0.4	θ	172.3	178.9	166.0	179.3
								Q	0.5	0.6	0.6	0.6
BA-N	ϕ_2	141.8	137.3	302.2	287.5	290.9	298.8	ϕ	-52.5	-46.2	91.8	125.4
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	173.8	178.1	176.5	174.8
								Q	0.6	0.6	0.5	0.6
CA	ϕ_2	144.5	138.2	301.4	286.8	289.4	292.5	ϕ	-44.9	-30.5	80.7	169.6
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	173.5	178.1	176.1	160.1
								Q	0.6	0.6	0.5	0.5
MM	ϕ_2	139.8	130.4	217.3	269.7	329.0	161.5	ϕ	33.6	-130.0	120.6	126.7
	Q	0.4	0.3	0.4	0.3	0.3	0.4	θ	163.9	179.6	177.9	176.1
								Q	0.6	0.6	0.6	0.6
		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$			$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	ϕ_2	316.7	302.6	99.8	77.4	87.1	68.7	ϕ	24.9	75.2	47.9	73.8
	Q	0.3	0.3	0.3	0.3	0.4	0.4	θ	164.0	169.3	170.0	164.4
								Q	0.5	0.5	0.5	0.5
G	ϕ_2	110.0	105.1	67.1	101.6	92.3	102.2	ϕ	21.5	74.4	70.5	89.1
	Q	0.4	0.3	0.4	0.3	0.4	0.4	θ	167.0	168.5	168.4	166.9
								Q	0.5	0.5	0.5	0.5
C	ϕ_2	278.8	302.0	87.8	316.6	86.0	110.2	ϕ	54.0	81.7	35.2	47.2
	Q	0.4	0.3	0.3	0.3	0.5	0.4	θ	143.6	166.6	94.0	162.7
								Q	0.5	0.5	0.7	0.5
T	ϕ_2	279.6	289.1	85.5	317.4	87.4	89.6	ϕ	51.6	80.6	36.8	139.6
	Q	0.4	0.4	0.3	0.3	0.5	0.4	θ	145.2	166.1	93.8	150.5
								Q	0.5	0.5	0.7	0.5
U	ϕ_2	267.0	285.6	85.3	316.2	87.8	90.2	ϕ	51.4	76.1	35.6	71.9
	Q	0.3	0.3	0.3	0.3	0.5	0.4	θ	144.7	165.6	93.9	141.3
								Q	0.5	0.5	0.7	0.5
TAP-C^5	ϕ_2	28.0	36.3	10.3	26.3	72.5	56.7	ϕ	10.1	113.9	28.0	28.5
	Q	0.4	0.4	0.4	0.4	0.5	0.4	θ	174.8	177.3	91.6	90.9
								Q	0.6	0.6	0.8	0.8
TAP-N	ϕ_2	116.1	108.6	215.5	83.0	256.4	38.3	ϕ	24.4	55.9	63.3	95.9
	Q	0.4	0.4	0.3	0.3	0.4	0.4	θ	172.8	176.2	172.6	175.1
								Q	0.6	0.6	0.5	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
BA-C^5	ϕ_2	8.9	326.6	38.4	32.2	342.3	85.1	ϕ	18.7	94.0	79.4	109.3
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	172.7	170.6	174.0	168.4
								Q	0.6	0.5	0.6	0.5
BA-N	ϕ_2	242.3	311.7	238.7	323.2	104.3	316.2	ϕ	43.8	63.1	63.0	72.5
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	147.0	153.1	146.2	153.7
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	256.0	318.7	320.6	320.4	104.3	312.9	ϕ	45.6	61.6	57.9	71.4
	Q	0.4	0.3	0.4	0.4	0.4	0.3	θ	145.8	152.0	140.9	140.3
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	103.5	73.5	70.7	46.1	83.7	87.0	ϕ	17.4	52.2	55.9	90.5
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	172.6	176.2	172.9	174.9
								Q	0.5	0.6	0.5	0.5
		α-$^3T_2$[4]		α-$^3T_2$		α-$^3T_2$			α-$^4C_1$[5]		α-$^4C_1$	
A	ϕ_2	140.0	131.5	95.8	282.3	283.1	290.3	ϕ	-106.5	-86.4	-162.6	-112.3
	Q	0.4	0.3	0.3	0.4	0.4	0.4	θ	17.8	23.4	3.4	5.2
								Q	0.5	0.5	0.6	0.6
G	ϕ_2	139.1	111.2	346.6	282.2	282.9	288.0	ϕ	-162.1	-96.2	-109.2	-94.0
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	10.1	10.5	2.9	11.2
								Q	0.5	0.5	0.6	0.5
C	ϕ_2	146.5	283.2	326.9	287.9	109.5	292.9	ϕ	-83.2	-111.8	-162.7	-94.9
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	15.5	29.3	23.6	6.5
								Q	0.5	0.5	0.6	0.6
T	ϕ_2	147.2	269.0	324.2	287.4	264.1	292.4	ϕ	-86.9	-91.4	-130.4	-108.7
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	16.5	11.5	13.9	5.7
								Q	0.5	0.5	0.5	0.6
U	ϕ_2	145.0	268.8	272.3	286.7	263.1	292.1	ϕ	-87.4	-111.1	-131.8	-107.4
	Q	0.3	0.4	0.3	0.4	0.4	0.4	θ	16.3	29.4	13.9	23.3
								Q	0.5	0.5	0.5	0.5
TAP-C^5	ϕ_2	163.8	135.1	242.9	235.8	258.5	242.7	ϕ	-107.3	-60.7	-61.1	-63.1
	Q	0.4	0.4	0.4	0.5	0.4	0.4	θ	4.9	4.7	7.5	2.7
								Q	0.6	0.6	0.6	0.6
TAP-N	ϕ_2	38.9	284.2	221.2	226.7	303.6	290.3	ϕ	18.5	4.9	34.8	-23.1
	Q	0.4	0.4	0.5	0.5	0.4	0.4	θ	1.0	2.5	6.1	5.6
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
BA-C^5	ϕ_2	305.0	244.4	13.6	267.5	224.2	283.9	ϕ	-133.4	-104.6	-173.7	-84.5
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	8.1	20.2	4.7	9.8
								Q	0.5	0.5	0.6	0.6
BA-N	ϕ_2	142.8	135.9	230.5	287.5	290.3	294.0	ϕ	-152.1	-112.6	-97.3	-107.1
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	15.9	31.0	41.0	18.5
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	144.5	138.2	307.5	286.8	289.5	292.0	ϕ	-160.2	-112.5	-97.9	-107.3
	Q	0.3	0.3	0.2	0.4	0.4	0.4	θ	17.6	31.2	41.2	27.4
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	139.8	285.7	347.5	283.6	288.2	325.7	ϕ	-107.3	-83.2	-165.6	-89.1
	Q	0.4	0.4	0.3	0.4	0.4	0.3	θ	13.7	8.0	0.9	3.1
								Q	0.5	0.5	0.6	0.6
		β-$^3T_2$		β-$^3T_2$		β-$^3T_2$			β $^{-4}C_1$		β $^{-4}C_1$	
A	ϕ_2	278.4	307.4	99.8	312.8	256.6	329.7	ϕ	5.2	19.0	159.7	-12.5
	Q	0.4	0.3	0.3	0.4	0.3	0.4	θ	1.4	3.5	5.2	3.6
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	267.0	267.7	82.0	309.0	292.6	268.5	ϕ	-130.6	22.4	15.8	-64.1
	Q	0.2	0.2	0.3	0.3	0.3	0.3	θ	1.1	3.5	5.6	5.0
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	85.3	279.5	91.4	82.9	317.1	270.0	ϕ	-54.6	8.9	4.2	-53.9
	Q	0.3	0.3	0.3	0.3	0.4	0.3	θ	2.7	3.3	5.6	4.1
								Q	0.6	0.6	0.6	0.6
T	ϕ_2	279.6	279.9	84.8	82.5	312.0	285.7	ϕ	-37.4	12.3	7.7	14.1
	Q	0.4	0.3	0.3	0.3	0.3	0.3	θ	1.5	2.8	5.2	1.1
								Q	0.6	0.6	0.6	0.6
U	ϕ_2	280.7	287.2	96.5	87.0	313.7	90.2	ϕ	-120.9	14.6	8.6	-18.5
	Q	0.4	0.4	0.3	0.3	0.3	0.4	θ	3.6	3.0	5.1	1.5
								Q	0.6	0.6	0.6	0.6
TAP-C^5	ϕ_2	18.1	36.3	10.1	39.3	72.5	56.7	ϕ	78.7	-2.2	133.8	13.0
	Q	0.4	0.4	0.4	0.4	0.5	0.4	θ	1.3	3.4	3.4	1.8
								Q	0.6	0.6	0.6	0.6
TAP-N	ϕ_2	134.6	236.5	289.7	83.6	256.4	238.3	ϕ	22.0	0.8	168.6	5.0
	Q	0.3	0.3	0.3	0.3	0.4	0.4	θ	2.5	4.7	5.2	7.2
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
BA-C[5]	ϕ_2	84.5	321.6	20.1	331.8	342.3	85.1	ϕ	-97.4	6.6	-120.3	46.7
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	1.6	3.8	1.7	5.8
								Q	0.6	0.6	0.6	0.6
BA-N	ϕ_2	229.6	312.9	236.3	321.8	104.3	316.2	ϕ	62.4	25.8	121.3	10.7
	Q	0.3	0.3	0.3	0.3	0.4	0.3	θ	1.6	3.0	5.1	1.5
								Q	0.6	0.6	0.6	0.6
CA	ϕ_2	241.0	311.4	260.0	318.9	105.2	312.9	ϕ	-156.8	31.6	97.7	18.5
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	2.5	2.6	7.5	5.0
								Q	0.6	0.6	0.6	0.6
MM	ϕ_2	278.8	238.1	113.0	112.2	254.6	87.0	ϕ	5.1	4.7	-114.6	-50.4
	Q	0.2	0.3	0.3	0.4	0.3	0.4	θ	2.9	4.7	3.8	3.7
								Q	0.6	0.6	0.6	0.6

$HAsO_3^-$

		2dRibf		Ribf		Tho			2dRib		Rib	
		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$			$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	ϕ_2	37.7	290.1	107.2	222.1	31.1	291.7	ϕ	75.7	-25.9	35.4	171.1
	Q	0.3	0.4	0.3	0.4	0.3	0.4	θ	176.1	178.8	178.2	178.9
								Q	0.5	0.6	0.6	0.6
G	ϕ_2	138.8	109.3	120.6	283.2	284.5	152.7	ϕ	38.9	53.0	-75.1	-145.6
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	162.2	169.5	178.1	177.5
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	146.2	131.5	94.9	289.0	291.9	293.7	ϕ	108.0	2.1	60.7	-159.3
	Q	0.4	0.3	0.3	0.4	0.4	0.4	θ	177.1	177.4	177.4	177.5
								Q	0.5	0.6	0.6	0.6
T	ϕ_2	146.8	133.3	103.4	100.1	291.1	292.6	ϕ	87.3	17.4	48.3	-156.4
	Q	0.4	0.3	0.3	0.2	0.4	0.4	θ	177.2	177.3	177.0	177.9
								Q	0.5	0.6	0.6	0.6
U	ϕ_2	147.1	133.4	104.1	293.7	291.0	292.5	ϕ	27.7	17.9	48.4	-161.7
	Q	0.4	0.3	0.3	0.4	0.4	0.4	θ	171.5	177.1	177.0	177.9
								Q	0.6	0.6	0.6	0.6
TAP-C[5]	ϕ_2	156.4	228.5	242.7	222.4	28.6	175.9	ϕ	58.5	-147.7	-97.4	-114.9
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	164.4	169.6	175.2	174.1
								Q	0.6	0.5	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-N	ϕ_2	43.4	203.3	314.6	45.0	1.3	292.1	ϕ	-16.2	39.5	102.3	179.0
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	176.4	178.6	177.4	157.5
								Q	0.6	0.6	0.5	0.5
BA-C[5]	ϕ_2	276.4	139.4	236.8	125.8	203.2	284.3	ϕ	175.5	3.8	-133.6	160.0
	Q	0.4	0.4	0.4	0.3	0.4	0.4	θ	171.6	178.1	165.3	177.0
								Q	0.5	0.6	0.6	0.6
BA-N	ϕ_2	143.0	137.8	15.3	287.2	291.1	301.5	ϕ	-58.0	-5.4	92.2	143.2
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	174.4	177.6	176.7	176.0
								Q	0.6	0.6	0.5	0.6
CA	ϕ_2	144.6	139.4	315.9	278.3	289.5	292.7	ϕ	-49.6	-42.7	80.0	28.4
	Q	0.3	0.3	0.2	0.4	0.4	0.4	θ	174.1	178.5	176.4	176.9
								Q	0.6	0.6	0.5	0.6
MM	ϕ_2	139.3	130.8	203.8	285.1	331.2	291.0	ϕ	32.4	89.3	124.1	128.0
	Q	0.4	0.3	0.4	0.4	0.3	0.4	θ	164.6	179.9	178.2	176.3
								Q	0.6	0.6	0.6	0.6
		β-$^2T_3$		β-$^2T_3$		β-$^2T_3$			β-$^1C_4$		β-$^1C_4$	
A	ϕ_2	314.0	109.5	71.2	80.6	44.9	266.3	ϕ	25.5	72.8	47.7	146.9
	Q	0.3	0.3	0.4	0.3	0.4	0.3	θ	163.1	168.4	167.1	152.6
								Q	0.5	0.5	0.5	0.5
G	ϕ_2	108.9	71.7	76.4	44.2	91.7	100.2	ϕ	22.6	74.2	71.7	90.0
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	164.7	168.1	169.5	169.4
								Q	0.5	0.5	0.5	0.5
C	ϕ_2	271.0	307.7	92.2	316.9	89.6	109.6	ϕ	53.7	78.2	31.0	45.3
	Q	0.3	0.3	0.3	0.3	0.4	0.4	θ	142.4	166.0	80.3	162.9
								Q	0.5	0.5	0.7	0.5
T	ϕ_2	268.4	12.4	73.2	316.7	91.2	89.9	ϕ	51.0	84.3	31.3	139.4
	Q	0.3	0.4	0.3	0.4	0.4	0.4	θ	143.9	166.5	90.5	149.3
								Q	0.5	0.5	0.8	0.5
U	ϕ_2	280.4	287.1	71.2	317.3	91.5	90.1	ϕ	51.2	83.4	31.6	138.0
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	143.3	166.6	90.5	149.1
								Q	0.5	0.5	0.8	0.5
TAP-C[5]	ϕ_2	304.8	31.2	3.5	46.8	72.8	73.5	ϕ	8.1	125.5	28.1	28.8
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	173.3	176.7	91.4	90.7
								Q	0.6	0.6	0.8	0.8

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-N	ϕ_2	136.0	103.2	90.9	100.7	256.2	7.2	ϕ	25.5	51.7	64.8	95.2
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	171.5	175.3	173.4	175.2
								Q	0.6	0.6	0.5	0.6
BA-C[5]	ϕ_2	277.9	40.0	30.0	35.5	345.0	110.8	ϕ	22.3	101.9	81.9	111.2
	Q	0.3	0.4	0.4	0.4	0.4	0.4	θ	171.3	169.4	175.0	169.3
								Q	0.6	0.5	0.6	0.5
BA-N	ϕ_2	206.0	327.8	238.4	318.2	105.2	328.9	ϕ	43.9	60.9	63.4	69.1
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	146.7	151.7	146.9	153.0
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	308.0	317.6	320.9	318.3	104.9	325.6	ϕ	44.6	61.4	58.8	68.6
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	145.0	151.4	141.4	151.9
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	60.4	39.8	71.2	314.5	82.3	114.5	ϕ	16.0	50.1	53.4	-164.5
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	173.0	175.5	176.1	170.5
								Q	0.5	0.6	0.5	0.5
		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$			$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	ϕ_2	139.3	122.7	261.7	222.3	285.7	289.8	ϕ	-88.3	-86.8	-166.3	-91.6
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	12.8	23.4	4.1	7.0
								Q	0.5	0.5	0.6	0.6
G	ϕ_2	138.8	23.9	350.3	283.2	285.8	334.2	ϕ	-114.8	-84.7	-121.2	-99.5
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	13.6	12.7	3.0	13.0
								Q	0.5	0.5	0.6	0.5
C	ϕ_2	146.2	303.5	222.8	288.9	278.6	293.7	ϕ	-84.6	-104.8	-168.4	-75.5
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	15.3	29.1	22.3	9.6
								Q	0.5	0.5	0.5	0.6
T	ϕ_2	146.8	275.5	284.8	287.8	278.7	292.6	ϕ	-88.5	-80.7	-131.7	-79.7
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	16.3	14.8	13.8	9.5
								Q	0.5	0.5	0.5	0.6
U	ϕ_2	147.1	276.8	285.6	287.9	263.2	292.5	ϕ	-89.2	-103.1	-133.3	-108.3
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	16.2	29.4	13.9	23.7
								Q	0.5	0.5	0.5	0.5
TAP-C[5]	ϕ_2	159.8	149.5	153.8	240.9	257.0	235.7	ϕ	-88.6	-50.8	-91.4	-44.2
	Q	0.4	0.3	0.4	0.5	0.4	0.4	θ	3.9	6.6	9.7	5.0
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-N	ϕ_2	295.6	228.4	215.9	227.2	308.7	291.4	ϕ	21.7	3.2	36.1	-14.1
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	0.5	1.8	5.3	5.5
								Q	0.6	0.6	0.6	0.6
BA-C[5]	ϕ_2	263.0	290.5	246.0	268.8	203.2	284.3	ϕ	-137.9	-83.4	144.3	-65.2
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	8.6	22.3	3.1	14.3
								Q	0.5	0.5	0.6	0.6
BA-N	ϕ_2	143.0	138.6	218.0	287.9	290.7	294.5	ϕ	-152.5	-104.1	-97.3	-113.7
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	16.1	30.9	40.1	24.2
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	144.6	139.4	179.1	286.8	289.8	292.7	ϕ	-160.1	-102.7	-97.7	-111.5
	Q	0.3	0.3	2.1	0.4	0.4	0.4	θ	18.2	31.2	40.3	24.8
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	139.3	229.3	350.1	302.3	286.9	324.1	ϕ	-108.3	-72.6	179.4	-64.8
	Q	0.4	0.4	0.3	0.4	0.4	0.3	θ	15.9	9.5	1.4	4.7
								Q	0.5	0.6	0.6	0.6
		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$		$\beta\text{-}^3T_2$			$\beta^{-4}C_1$		$\beta^{-4}C_1$	
A	ϕ_2	278.5	306.6	97.2	312.9	254.9	332.1	ϕ	10.0	21.0	114.4	-22.7
	Q	0.4	0.3	0.3	0.4	0.3	0.4	θ	0.8	2.8	5.2	6.0
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	64.6	266.7	92.2	56.5	302.3	267.8	ϕ	62.9	21.0	17.4	-49.8
	Q	0.4	0.2	0.4	0.4	0.3	0.4	θ	1.9	4.4	6.0	6.9
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	92.6	286.9	75.7	71.2	333.9	272.4	ϕ	-71.0	12.6	3.2	-47.6
	Q	0.3	0.4	0.3	0.3	0.4	0.4	θ	2.3	2.6	3.9	6.7
								Q	0.6	0.6	0.6	0.6
T	ϕ_2	280.3	287.4	76.3	95.6	317.7	279.7	ϕ	-57.7	18.6	6.0	-10.5
	Q	0.4	0.3	0.3	0.3	0.3	0.3	θ	1.0	2.2	3.9	3.5
								Q	0.6	0.6	0.6	0.6
U	ϕ_2	281.3	283.4	94.5	97.7	330.6	90.1	ϕ	-59.5	17.3	8.1	-27.7
	Q	0.4	0.3	0.3	0.3	0.4	0.4	θ	0.8	2.2	3.9	3.8
								Q	0.6	0.6	0.6	0.6
TAP-C[5]	ϕ_2	17.9	39.0	16.0	54.3	72.8	37.9	ϕ	65.4	-5.0	100.5	-1.9
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	4.3	2.4	5.2	3.7
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-N	ϕ_2	137.6	140.9	129.3	349.1	256.2	229.9	ϕ	31.0	2.1	131.5	3.6
	Q	0.4	0.3	0.3	0.4	0.4	0.4	θ	2.9	4.1	4.5	6.6
								Q	0.6	0.6	0.6	0.6
BA-C[5]	ϕ_2	53.9	324.6	26.6	326.8	345.0	110.8	ϕ	-118.8	3.0	28.5	28.1
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	1.8	2.9	3.5	6.6
								Q	0.6	0.6	0.6	0.6
BA-N	ϕ_2	229.7	312.7	236.2	317.9	105.2	328.9	ϕ	74.3	30.5	98.6	-13.1
	Q	0.3	0.3	0.3	0.4	0.4	0.3	θ	1.3	2.4	7.5	3.7
								Q	0.6	0.6	0.6	0.6
CA	ϕ_2	317.0	311.5	320.9	331.9	104.9	325.6	ϕ	99.8	40.2	102.0	16.6
	Q	0.3	0.3	0.3	0.3	0.4	0.3	θ	2.2	2.1	7.6	5.5
								Q	0.6	0.6	0.6	0.6
MM	ϕ_2	73.1	311.9	108.6	82.1	335.8	114.5	ϕ	7.4	3.2	-45.0	-41.5
	Q	0.4	0.3	0.3	0.3	0.4	0.4	θ	2.4	4.0	3.0	6.0
								Q	0.6	0.6	0.6	0.6

alternative pathway

HPO$_3^-$

	CP	2dRibf		Ribf		Tho		CP	2dRib		Rib	
		α-$^2T_3$		α-$^2T_3$		α-$^2T_3$			α-$^1C_4$		α-$^1C_4$	
A	ϕ_2	139.1	109.8	287.3	100.1	284.7	121.0	ϕ	7.3	-9.4	-56.2	-136.3
	Q	0.3	0.3	0.4	0.3	0.4	0.3	θ	173.5	178.5	179.3	165.6
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	139.0	109.9	282.9	101.7	299.1	325.8	ϕ	35.7	-12.8	74.5	163.7
	Q	0.3	0.3	0.4	0.3	0.3	0.3	θ	161.1	178.5	157.8	161.4
								Q	0.6	0.6	0.5	0.5
C	ϕ_2	143.6	134.4	285.2	287.9	247.9	306.4	ϕ	32.1	6.3	78.6	-163.0
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	170.7	178.3	179.4	158.1
								Q	0.6	0.6	0.6	0.6
T	ϕ_2	144.6	135.4	286.4	108.6	264.1	304.7	ϕ	29.0	15.7	31.4	-136.5
	Q	0.3	0.3	0.4	0.3	0.4	0.3	θ	170.1	177.9	179.0	167.5
								Q	0.6	0.6	0.6	0.6
U	ϕ_2	123.6	135.9	286.2	287.3	263.2	303.9	ϕ	25.6	10.1	33.5	-136.5
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	170.9	177.9	179.0	167.3
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-C^5	ϕ_2	264.9	152.3	208.0	245.3	259.9	169.4	ϕ	-128.5	60.9	-99.1	-133.2
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	173.4	177.2	175.5	163.8
								Q	0.5	0.6	0.6	0.6
TAP-N	ϕ_2	143.3	66.6	308.8	54.1	302.0	32.8	ϕ	46.0	34.7	101.7	178.4
	Q	0.4	0.4	0.4	0.4	0.4	0.3	θ	162.0	179.7	176.9	158.1
								Q	0.6	0.6	0.5	0.5
BA-C^5	ϕ_2	135.3	136.8	228.8	125.9	230.1	135.6	ϕ	59.3	-31.4	-120.8	165.1
	Q	0.4	0.3	0.4	0.3	0.4	0.4	θ	159.1	179.1	178.0	160.3
								Q	0.6	0.6	0.6	0.5
BA-N	ϕ_2	143.5	137.5	301.4	287.6	290.3	292.4	ϕ	-52.5	-31.6	91.7	167.5
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	173.8	178.2	176.5	160.0
								Q	0.6	0.6	0.5	0.5
CA	ϕ_2	144.8	139.2	301.4	286.8	289.5	292.0	ϕ	-45.0	-17.2	80.7	169.6
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	173.5	177.8	176.1	160.1
								Q	0.6	0.6	0.5	0.5
MM	ϕ_2	138.0	130.4	300.3	225.4	288.2	287.1	ϕ	33.6	137.0	122.6	-179.2
	Q	0.3	0.3	0.4	0.5	0.4	0.4	θ	163.9	179.9	177.5	157.9
								Q	0.6	0.6	0.6	0.5
		$\beta^{-2}T_3$		$\beta^{-2}T_3$		$\beta^{-2}T_3$			$\beta^{-1}C_4$		$\beta^{-1}C_4$	
A	ϕ_2	256.9	74.8	103.0	294.2	299.5	293.4	ϕ	53.3	73.7	64.4	73.8
	Q	0.3	0.3	0.3	0.4	0.3	0.3	θ	165.5	168.8	161.2	164.4
								Q	0.5	0.5	0.5	0.5
G	ϕ_2	81.5	76.5	70.5	50.7	293.4	295.1	ϕ	21.5	74.4	77.8	85.7
	Q	0.4	0.3	0.4	0.4	0.3	0.3	θ	167.0	168.5	171.6	167.7
								Q	0.5	0.5	0.5	0.5
C	ϕ_2	32.2	279.5	282.4	320.3	319.4	110.2	ϕ	19.7	31.7	63.8	47.2
	Q	0.3	0.3	0.3	0.3	0.4	0.4	θ	155.6	162.0	142.4	162.7
								Q	0.5	0.5	0.5	0.5
T	ϕ_2	76.9	280.0	93.2	286.5	313.7	110.3	ϕ	20.6	77.9	62.3	69.0
	Q	0.4	0.3	0.3	0.3	0.3	0.4	θ	154.9	165.4	145.1	153.3
								Q	0.5	0.5	0.5	0.5
U	ϕ_2	92.3	279.6	93.1	313.2	314.7	90.2	ϕ	21.3	76.1	62.0	68.9
	Q	0.3	0.3	0.3	0.3	0.4	0.4	θ	154.9	165.6	144.2	152.8
								Q	0.5	0.5	0.5	0.5

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-C5	ϕ_2	93.3	8.7	338.3	39.9	267.9	71.4	ϕ	31.8	27.2	27.0	28.5
	Q	0.4	0.4	0.4	0.3	0.4	0.4	θ	91.4	90.8	90.8	90.9
								Q	0.8	0.8	0.8	0.8
TAP-N	ϕ_2	66.3	108.6	62.7	288.2	59.0	38.3	ϕ	24.4	55.9	68.6	95.9
	Q	0.3	0.4	0.3	0.3	0.5	0.4	θ	172.8	176.2	175.2	175.1
								Q	0.6	0.6	0.5	0.6
BA-C5	ϕ_2	84.6	282.5	86.8	308.6	121.6	85.1	ϕ	27.6	91.9	28.7	96.4
	Q	0.4	0.3	0.3	0.3	0.4	0.4	θ	93.3	169.7	90.5	167.5
								Q	0.8	0.5	0.8	0.5
BA-N	ϕ_2	134.6	311.7	64.4	319.8	257.7	109.1	ϕ	43.8	63.1	63.0	68.5
	Q	0.4	0.3	0.4	0.3	0.2	0.4	θ	147.0	153.1	146.2	152.4
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	34.6	311.0	243.3	311.3	263.3	110.2	ϕ	45.6	63.3	61.1	68.1
	Q	0.3	0.3	0.3	0.4	0.2	0.4	θ	145.8	152.6	145.9	151.5
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	112.8	110.8	72.7	49.7	48.6	98.7	ϕ	16.9	51.2	57.8	90.6
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	172.5	175.6	175.8	174.9
								Q	0.5	0.6	0.5	0.5
		α-$^3T_2$		α-$^3T_2$		α-$^3T_2$			α-$^4C_1$		α-$^4C_1$	
A	ϕ_2	101.9	296.9	22.9	281.7	284.7	290.4	ϕ	-89.9	-68.7	-135.6	-78.7
	Q	0.3	0.3	0.3	0.4	0.4	0.4	θ	13.5	15.5	90.1	89.7
								Q	0.5	0.5	0.7	0.7
G	ϕ_2	136.7	288.6	22.5	282.2	299.1	288.0	ϕ	-89.5	-19.0	-85.1	-81.1
	Q	0.3	0.3	0.3	0.4	0.3	0.4	θ	14.5	96.0	89.5	90.2
								Q	0.5	0.7	0.7	0.7
C	ϕ_2	144.6	269.0	21.7	287.9	247.9	270.4	ϕ	33.3	-69.1	-56.1	-57.5
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	171.5	19.0	13.3	14.3
								Q	0.6	0.5	0.5	0.5
T	ϕ_2	145.3	269.6	23.2	287.4	264.1	270.3	ϕ	-90.5	-70.6	-83.0	-63.9
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	16.9	19.1	89.1	14.6
								Q	0.5	0.5	0.7	0.5
U	ϕ_2	145.6	268.8	23.3	286.7	263.2	269.6	ϕ	-91.0	-70.6	-83.5	-61.9
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	16.8	18.9	89.0	14.4
								Q	0.5	0.5	0.7	0.5

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-C^5	ϕ_2	161.0	223.8	249.4	240.3	259.9	247.7	ϕ	-65.0	-45.3	-52.2	-45.5
	Q	0.4	0.4	0.4	0.5	0.4	0.4	θ	12.8	16.3	7.7	9.0
								Q	0.5	0.5	0.6	0.6
TAP-N	ϕ_2	161.1	284.2	324.1	232.8	302.0	290.3	ϕ	-48.0	4.9	34.7	28.6
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	0.3	2.5	6.1	5.3
								Q	0.6	0.6	0.6	0.6
BA-C^5	ϕ_2	144.9	282.6	246.3	266.4	230.1	283.9	ϕ	-70.1	-104.6	-28.8	-38.8
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	10.0	20.2	8.9	7.7
								Q	0.5	0.5	0.5	0.5
BA-N	ϕ_2	108.5	291.9	303.5	287.6	290.3	292.4	ϕ	-155.6	-112.6	-96.9	-95.8
	Q	0.2	0.3	0.3	0.4	0.4	0.4	θ	17.6	31.0	45.9	38.4
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	113.4	139.2	304.4	286.4	289.5	292.0	ϕ	-159.2	-138.7	-97.9	-107.3
	Q	0.2	0.3	0.3	0.4	0.4	0.4	θ	18.3	18.9	43.2	27.4
								Q	0.5	0.5	0.5	0.5
MM	ϕ_2	136.1	285.7	295.1	225.4	288.2	282.6	ϕ	-172.1	-8.5	38.8	30.9
	Q	0.3	0.4	0.3	0.5	0.4	0.4	θ	1.2	3.2	4.7	4.6
								Q	0.6	0.5	0.6	0.6
		β-$^3T_2$		β-$^3T_2$		β-$^3T_2$			β-$^4C_1$		β-$^4C_1$	
A	ϕ_2	104.7	280.2	253.4	280.8	299.5	297.7	ϕ	5.2	19.0	-86.8	13.1
	Q	0.3	0.3	0.4	0.3	0.3	0.3	θ	1.4	3.5	2.8	5.9
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	78.9	303.1	273.9	304.2	274.8	297.6	ϕ	52.3	22.5	-69.4	12.9
	Q	0.4	0.3	0.3	0.4	0.4	0.3	θ	2.1	3.5	2.0	5.8
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	56.2	284.8	261.6	284.6	319.4	272.2	ϕ	-54.7	8.9	-90.9	15.0
	Q	0.4	0.3	0.4	0.3	0.4	0.4	θ	2.7	3.3	3.0	5.8
								Q	0.6	0.6	0.6	0.6
T	ϕ_2	58.0	288.0	294.2	330.1	313.7	275.0	ϕ	-37.5	12.2	-94.0	12.7
	Q	0.4	0.3	0.4	0.4	0.3	0.3	θ	1.5	2.8	2.6	5.4
								Q	0.6	0.6	0.6	0.6
U	ϕ_2	23.5	285.6	284.3	284.3	314.7	266.7	ϕ	-37.2	14.5	-94.5	13.6
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	1.3	3.0	2.6	5.4
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
TAP-C^5	ϕ_2	93.3	37.5	340.1	349.0	219.4	332.0	ϕ	60.4	4.0	-44.2	-17.6
	Q	0.4	0.3	0.4	0.3	0.4	0.4	θ	4.6	3.5	3.2	6.8
								Q	0.6	0.6	0.6	0.6
TAP-N	ϕ_2	102.2	236.5	311.1	241.4	241.6	231.8	ϕ	22.0	0.8	-88.9	5.0
	Q	0.3	0.3	0.4	0.4	0.3	0.4	θ	2.5	4.7	4.1	7.2
								Q	0.6	0.6	0.6	0.6
BA-C^5	ϕ_2	39.5	17.4	279.6	18.9	227.8	294.4	ϕ	-97.4	6.6	-120.3	9.8
	Q	0.4	0.4	0.3	0.4	0.4	0.3	θ	1.6	3.8	1.7	5.9
								Q	0.6	0.6	0.6	0.6
BA-N	ϕ_2	229.6	306.9	82.7	318.2	263.5	275.6	ϕ	62.4	25.8	-103.9	17.6
	Q	0.3	0.3	0.3	0.4	0.2	0.3	θ	1.6	3.0	3.2	5.4
								Q	0.6	0.6	0.6	0.6
CA	ϕ_2	134.9	305.9	317.1	320.1	263.3	267.8	ϕ	90.3	31.6	-112.3	18.6
	Q	0.3	0.3	0.3	0.4	0.2	0.2	θ	2.3	2.6	2.9	5.0
								Q	0.6	0.6	0.6	0.6
MM	ϕ_2	104.7	238.1	228.2	241.4	255.7	272.9	ϕ	5.1	4.7	-48.5	4.2
	Q	0.4	0.3	0.4	0.4	0.3	0.3	θ	2.9	4.7	3.2	7.2
								Q	0.6	0.6	0.6	0.6

$$HAsO_3^-$$

		2dRibf		Ribf		Tho			2dRib		Rib	
		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$		$\alpha\text{-}^2T_3$			$\alpha\text{-}^1C_4$		$\alpha\text{-}^1C_4$	
A	ϕ_2	286.5	110.6	29.6	264.8	287.5	312.9	ϕ	11.1	-3.1	-69.6	-146.0
	Q	0.4	0.3	0.3	0.4	0.4	0.4	θ	173.1	178.5	178.9	177.4
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	138.8	110.9	322.9	267.7	329.3	315.9	ϕ	36.2	49.4	79.2	-145.6
	Q	0.3	0.3	0.4	0.4	0.3	0.3	θ	162.2	170.0	158.2	177.5
								Q	0.6	0.6	0.5	0.6
C	ϕ_2	143.4	134.1	220.5	108.7	235.3	303.8	ϕ	31.8	10.7	-39.9	-159.4
	Q	0.3	0.3	0.4	0.3	0.4	0.3	θ	171.0	178.2	179.9	177.5
								Q	0.6	0.6	0.6	0.6
T	ϕ_2	121.4	135.5	189.3	100.1	204.4	268.9	ϕ	28.3	19.9	-4.5	-156.4
	Q	0.3	0.3	0.4	0.2	0.3	0.4	θ	170.5	177.6	179.2	177.9
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
U	ϕ_2	120.9	135.9	219.3	99.0	205.6	301.3	ϕ	25.8	15.9	-9.3	-161.7
	Q	0.3	0.3	0.4	0.2	0.3	0.3	θ	171.0	177.6	179.2	177.9
								Q	0.6	0.6	0.6	0.6
TAP-C[5]	ϕ_2	266.7	215.0	184.2	217.4	258.6	175.9	ϕ	54.3	62.5	-97.4	-114.7
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	165.1	176.6	175.2	174.0
								Q	0.6	0.6	0.6	0.6
TAP-N	ϕ_2	91.2	67.4	309.2	45.0	306.0	51.9	ϕ	45.3	62.4	102.4	158.7
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	162.2	167.5	177.4	176.2
								Q	0.6	0.6	0.5	0.6
BA-C[5]	ϕ_2	133.1	137.0	216.9	125.9	228.4	167.2	ϕ	59.7	-0.8	-118.4	168.4
	Q	0.3	0.3	0.4	0.3	0.4	0.4	θ	158.7	179.0	177.8	176.9
								Q	0.6	0.6	0.6	0.6
BA-N	ϕ_2	143.0	138.0	301.5	286.1	290.7	302.6	ϕ	-38.5	-22.9	92.2	143.1
	Q	0.3	0.3	0.4	0.4	0.4	0.3	θ	174.0	178.2	176.7	176.0
								Q	0.6	0.6	0.5	0.6
CA	ϕ_2	289.3	139.4	134.4	338.3	289.8	300.2	ϕ	-44.8	-6.1	80.0	133.1
	Q	0.4	0.3	0.3	0.1	0.4	0.3	θ	173.9	177.6	176.4	175.8
								Q	0.6	0.6	0.5	0.6
MM	ϕ_2	70.4	130.7	300.1	225.7	324.6	43.7	ϕ	92.8	48.9	124.6	167.3
	Q	0.4	0.3	0.4	0.5	0.3	0.3	θ	164.1	171.6	178.1	176.2
								Q	0.5	0.6	0.6	0.6
		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$		$\beta\text{-}^2T_3$			$\beta\text{-}^1C_4$		$\beta\text{-}^1C_4$	
A	ϕ_2	74.9	71.5	83.6	70.4	7.2	318.3	ϕ	25.5	72.7	63.3	160.6
	Q	0.4	0.3	0.3	0.4	0.5	0.3	θ	163.1	168.4	161.9	171.4
								Q	0.5	0.5	0.5	0.5
G	ϕ_2	85.5	71.7	75.5	70.6	5.1	315.5	ϕ	22.6	74.1	40.6	162.9
	Q	0.4	0.3	0.4	0.4	0.5	0.4	θ	164.7	168.1	168.1	171.4
								Q	0.5	0.5	0.5	0.5
C	ϕ_2	72.2	283.5	75.7	71.2	305.4	40.5	ϕ	20.9	32.6	64.5	71.1
	Q	0.4	0.3	0.3	0.3	0.4	0.5	θ	155.2	161.8	143.5	139.6
								Q	0.5	0.5	0.5	0.5

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	ϕ_2	72.4	60.9	76.3	285.9	303.7	119.9	ϕ	21.4	30.9	62.6	141.7
	Q	0.4	0.3	0.3	0.3	0.4	0.3	θ	154.4	161.3	146.8	168.2
								Q	0.5	0.5	0.5	0.5
U	ϕ_2	71.3	283.4	99.0	285.3	304.1	90.1	ϕ	22.0	33.2	62.5	141.1
	Q	0.4	0.3	0.3	0.3	0.4	0.4	θ	154.5	160.9	145.9	168.4
								Q	0.5	0.5	0.5	0.5
TAP-C⁵	ϕ_2	92.6	31.1	339.1	23.0	275.5	39.3	ϕ	67.1	27.1	26.3	-164.0
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	89.6	91.0	90.2	167.6
								Q	0.8	0.8	0.8	0.5
TAP-N	ϕ_2	56.7	103.2	72.6	100.7	270.5	5.8	ϕ	9.1	54.9	65.3	-169.6
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	173.3	175.8	175.6	170.8
								Q	0.5	0.6	0.5	0.5
BA-C⁵	ϕ_2	84.1	80.7	41.0	77.4	335.0	336.6	ϕ	97.2	90.9	28.2	146.6
	Q	0.4	0.3	0.4	0.4	0.4	0.4	θ	93.5	169.3	90.1	167.6
								Q	0.8	0.5	0.8	0.5
BA-N	ϕ_2	312.6	312.7	202.1	317.9	283.5	50.6	ϕ	42.8	62.9	63.4	70.6
	Q	0.2	0.3	0.3	0.4	0.3	0.4	θ	146.2	152.7	146.9	139.6
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	29.2	311.6	84.0	316.9	280.7	315.6	ϕ	26.1	63.2	61.4	70.6
	Q	0.3	0.3	0.3	0.3	0.3	0.3	θ	94.7	152.2	147.0	139.3
								Q	0.8	0.5	0.5	0.5
MM	ϕ_2	108.9	106.3	108.2	103.6	277.5	8.9	ϕ	16.0	50.0	53.4	-164.5
	Q	0.4	0.4	0.3	0.4	0.4	0.4	θ	173.0	175.5	176.1	170.5
								Q	0.5	0.6	0.5	0.5
		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$		$\alpha\text{-}^3T_2$			$\alpha\text{-}^4C_1$		$\alpha\text{-}^4C_1$	
A	ϕ_2	136.4	152.4	288.6	223.2	287.5	289.8	ϕ	-106.7	-82.5	-142.8	-46.0
	Q	0.3	0.4	0.4	0.4	0.4	0.4	θ	17.4	14.9	87.4	92.5
								Q	0.5	0.5	0.8	0.7
G	ϕ_2	136.1	149.5	22.7	284.7	329.3	289.9	ϕ	-114.8	-97.9	-104.4	-99.6
	Q	0.3	0.4	0.3	0.4	0.3	0.4	θ	13.6	16.7	11.4	13.0
								Q	0.5	0.5	0.5	0.5
C	ϕ_2	143.4	137.7	285.3	288.9	235.3	267.9	ϕ	-170.3	-83.6	-91.2	-86.4
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	79.7	18.2	16.0	16.4
								Q	0.7	0.5	0.5	0.5

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	ϕ_2	144.2	139.5	285.8	229.8	204.4	271.1	ϕ	-105.5	-85.0	-144.6	-91.8
	Q	0.3	0.3	0.4	0.4	0.3	0.4	θ	21.0	18.4	87.1	17.2
								Q	0.5	0.5	0.8	0.5
U	ϕ_2	144.5	139.5	285.6	229.8	205.6	269.7	ϕ	-106.3	-94.4	-144.4	-90.8
	Q	0.3	0.3	0.4	0.5	0.3	0.4	θ	21.1	20.2	87.1	16.9
								Q	0.5	0.5	0.8	0.5
TAP-C^5	ϕ_2	150.4	152.2	269.8	222.4	258.7	241.4	ϕ	-86.4	-58.4	-139.8	-93.1
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	14.7	11.7	90.2	91.0
								Q	0.6	0.6	0.8	0.8
TAP-N	ϕ_2	123.4	150.3	314.6	222.7	306.0	291.4	ϕ	-124.5	-108.0	-71.1	-57.4
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	4.7	3.2	1.8	1.6
								Q	0.6	0.6	0.6	0.6
BA-C^5	ϕ_2	131.0	140.6	234.6	232.8	228.4	284.3	ϕ	-81.9	-83.4	-66.8	1.1
	Q	0.3	0.4	0.4	0.4	0.4	0.4	θ	18.9	22.3	12.4	11.7
								Q	0.5	0.5	0.6	0.5
BA-N	ϕ_2	143.0	292.8	302.1	287.9	290.7	293.3	ϕ	-151.5	-104.1	-105.8	-111.9
	Q	0.3	0.3	0.4	0.4	0.4	0.4	θ	20.2	30.9	44.9	23.8
								Q	0.5	0.5	0.5	0.5
CA	ϕ_2	143.2	137.9	44.4	286.8	289.8	292.7	ϕ	-154.5	-102.7	-124.6	-71.3
	Q	0.4	0.3	0.8	0.4	0.4	0.4	θ	20.8	31.2	25.0	82.2
								Q	0.5	0.5	0.5	0.7
MM	ϕ_2	47.6	150.0	304.4	226.3	324.6	283.8	ϕ	-135.9	-106.5	-108.6	-114.4
	Q	0.4	0.4	0.4	0.4	0.3	0.4	θ	5.6	16.7	2.3	12.6
								Q	0.6	0.5	0.6	0.5
		$\beta^{-3}T_2$		$\beta^{-3}T_2$		$\beta^{-3}T_2$			$\beta^{-4}C_1$		$\beta^{-4}C_1$	
A	ϕ_2	74.9	90.4	92.2	81.3	302.4	296.1	ϕ	10.0	21.0	4.4	12.9
	Q	0.4	0.3	0.3	0.3	0.3	0.3	θ	0.8	2.7	4.5	5.2
								Q	0.6	0.6	0.6	0.6
G	ϕ_2	85.5	286.0	278.4	305.7	280.8	297.1	ϕ	62.9	23.8	17.4	12.6
	Q	0.4	0.3	0.4	0.4	0.4	0.3	θ	1.9	2.8	6.0	5.0
								Q	0.6	0.6	0.6	0.6
C	ϕ_2	72.2	284.9	96.7	92.2	322.9	272.4	ϕ	-71.1	12.6	3.2	15.9
	Q	0.4	0.3	0.3	0.3	0.4	0.4	θ	2.3	2.6	3.9	5.1
								Q	0.6	0.6	0.6	0.6

RU[1]	CP[6]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	vac.[2]	solv.[3]	CP[7]	vac.[2]	solv.[3]	vac.[2]	solv.[3]
T	ϕ_2	72.4	287.3	310.5	95.6	315.3	305.2	ϕ	-57.8	18.7	6.0	14.2
	Q	0.4	0.3	0.3	0.3	0.3	0.3	θ	1.0	2.2	3.9	4.8
								Q	0.6	0.6	0.6	0.6
U	ϕ_2	86.4	297.9	285.1	97.2	316.9	309.5	ϕ	-59.4	17.4	8.1	14.5
	Q	0.3	0.3	0.4	0.3	0.4	0.3	θ	0.8	2.1	3.9	4.6
								Q	0.6	0.6	0.6	0.6
TAP-C^5	ϕ_2	92.6	35.1	339.2	351.0	272.9	213.6	ϕ	65.4	2.8	5.8	-20.3
	Q	0.4	0.4	0.4	0.4	0.4	0.4	θ	4.3	2.6	6.4	6.2
								Q	0.6	0.6	0.6	0.6
TAP-N	ϕ_2	56.7	48.0	312.7	131.5	236.8	276.6	ϕ	26.8	2.1	-10.6	3.6
	Q	0.4	0.3	0.4	0.3	0.3	0.4	θ	2.1	4.1	5.3	6.6
								Q	0.6	0.6	0.6	0.6
BA-C^5	ϕ_2	84.1	18.0	296.0	37.2	268.6	309.8	ϕ	-118.8	3.0	-30.8	8.9
	Q	0.4	0.4	0.4	0.4	0.5	0.4	θ	1.8	2.9	11.9	4.9
								Q	0.6	0.6	0.5	0.6
BA-N	ϕ_2	88.4	308.4	315.0	329.6	260.7	154.5	ϕ	74.4	30.5	-0.9	14.2
	Q	0.3	0.3	0.3	0.3	0.2	0.5	θ	1.3	2.4	3.5	4.3
								Q	0.6	0.6	0.6	0.6
CA	ϕ_2	88.8	311.6	316.4	330.4	261.1	305.0	ϕ	99.8	40.2	8.1	15.5
	Q	0.3	0.3	0.3	0.3	0.2	0.8	θ	2.2	2.1	3.1	4.0
								Q	0.6	0.6	0.6	0.6
MM	ϕ_2	108.9	67.2	288.4	134.6	248.4	328.3	ϕ	7.3	3.2	11.4	4.4
	Q	0.4	0.3	0.4	0.3	0.3	0.4	θ	2.4	4.0	7.5	6.5
								Q	0.6	0.6	0.6	0.6

(1) RU = unspecified **r**ecognition **u**nit. [2] Puckering parameters in vacuum at the DFT level. [3] Puckering parameters in implicit solvent at the DFT level using the IEFPCM model. [4] initial sugar ring puckering-anomer for 5-MR. [5] initial sugar ring puckering-anomer for 6-MR. [6] Cremer-Pople (CP) generalized puckering parameters for 5-MR. [7] Polar coordinates derived from Cremer-Pople (CP) generalized puckering parameters for 6-MR.

Table A8. Differences between the energies of the canonical dTMP, UMP pairs (predominant) nucleotides and their minor counterparts TMP and dUMP (**Figure 4.48**) in vacuum and in aqueous environment (energies of the canonical form minus that of the minor form) for the sugar exchange reaction between the nucleotides of T and U. Included differences are between: the total energies without (ΔE) and with zero-point vibrational correction (ZPE) ($\Delta E_{(ZPE)}$), and Gibbs energies (ΔG°). All energies are in kJ/mol and are obtained from the DFT-B3LYP/6-311++G (d, p) calculations. The Polarizable Continuum Model (IEFPCM) has been used to generate the results in aqueous solvation at the same level of theory.

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
			classic pathway (a+b)			
			HPO_3^-			
			furanoses			
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-1.0	-1.1	-0.6	8.8	10.4	2.7
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-12.4	-11.5	-15.7	0.2	0.4	-1.5
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	4.7	5.2	1.2	-2.8	-4.4	1.5
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-6.7	-5.2	-13.9	-11.4	-14.4	-2.7
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-35.9	-37.3	-35.8	8.8	10.4	2.7
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-47.3	-47.7	-50.9	0.2	0.4	-1.5
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-30.2	-31.0	-34.0	-2.8	-4.4	1.5
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-41.6	-41.4	-49.1	-11.4	-14.4	-2.7
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-3.4	-5.3	4.0	20.9	25.7	3.5
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-14.8	-15.7	-11.1	12.3	15.7	-0.7
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	2.3	1.0	5.8	9.3	10.9	2.3
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-9.1	-9.4	-9.3	0.7	0.9	-1.9
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-38.3	-41.5	-31.2	20.9	25.7	3.5
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-49.7	-51.9	-46.3	12.3	15.7	-0.7
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-32.6	-35.2	-29.4	9.3	10.9	2.3
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-44.0	-45.6	-44.5	0.7	0.9	-1.9

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	1.4	1.6	1.5	-3.2	-4.5	2.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-7.4	-7.0	-8.1	-3.0	-3.4	-1.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	23.6	24.0	24.3	14.4	15.2	13.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	14.8	15.3	14.7	14.7	16.4	9.4
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	5.4	6.1	1.8	-4.7	-6.4	2.4
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-3.4	-2.5	-7.8	-4.5	-5.2	-2.0
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	27.6	28.5	24.5	12.9	13.4	13.4
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	18.8	19.9	15.0	13.2	14.5	9.0
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	-10.9	-10.8	-12.3	-1.1	-3.0	7.6
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-19.7	-19.4	-21.9	-0.8	-1.9	3.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	11.3	11.6	10.4	16.6	16.8	18.7
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	2.5	3.0	0.9	16.8	17.9	14.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	-6.9	-6.3	-12.1	-2.6	-4.8	7.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-15.6	-14.9	-21.6	-2.3	-3.7	2.9
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	15.3	16.1	10.7	15.1	14.9	18.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	6.5	7.5	1.2	15.3	16.1	14.0
pyranoses						
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	-1.3	-1.4	-1.1	-0.8	-0.8	-1.3
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-66.1	-67.2	-63.7	-46.9	-48.4	-42.4
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	-8.8	-9.2	-6.7	-46.9	-47.3	-45.9
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-73.5	-75.0	-69.3	-93.1	-95.0	-87.0
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	63.3	64.3	61.3	43.9	44.1	43.4
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-1.4	-1.5	-1.3	-2.3	-3.6	2.3
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	55.8	56.6	55.7	-2.3	-2.5	-1.2
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-8.9	-9.3	-6.9	-48.5	-50.1	-42.3
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	7.4	7.7	5.7	43.5	43.4	43.3
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-57.3	-58.1	-56.9	-2.7	-4.2	2.2
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	0.0	-0.1	0.1	-2.7	-3.1	-1.2
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-64.8	-65.9	-62.5	-48.9	-50.8	-42.3

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	72.1	73.4	68.1	88.1	88.3	88.1
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	7.3	7.6	5.5	41.9	40.6	47.0
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	64.6	65.6	62.5	41.9	41.7	43.5
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMp} - \alpha\text{-}^4C_1\text{-TMp} - \alpha\text{-}^4C_1\text{-dUMP})$	-0.1	-0.2	-0.1	-4.3	-5.9	2.4
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-8.7	-9.0	-7.7	6.7	2.8	15.8
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMp} - \beta\text{-}^4C_1\text{-dUMP})$	21.6	21.1	23.3	6.6	7.6	4.1
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	18.0	18.0	18.5	19.9	15.7	29.5
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	48.3	48.2	49.5	-13.8	-9.4	-24.3
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-29.9	-29.8	-30.5	-0.5	-1.3	1.0
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	0.4	0.4	0.5	-0.6	3.5	-10.7
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-3.2	-2.7	-4.3	12.7	11.6	14.7
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	27.2	27.4	26.7	-16.4	-14.7	-20.5
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-41.3	-41.7	-38.5	-3.1	-6.6	4.8
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	-11.0	-11.6	-7.5	-3.2	-1.8	-6.9
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-14.6	-14.7	-12.3	10.1	6.3	18.4
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	15.8	15.4	18.7	-23.6	-18.7	-35.3
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-62.4	-62.5	-61.3	-10.3	-10.6	-10.0
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	-32.1	-32.4	-30.3	-10.4	-5.9	-21.7
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-35.7	-35.5	-35.1	2.9	2.3	3.6
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	-5.4	-5.3	-4.1	6.7	2.8	15.8
	alternative pathway (c+d)					
	HPO_3^-					
	furanoses					
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	0.7	-0.4	3.9	-15.7	-17.3	-9.4
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	5.9	5.7	5.7	-20.6	-20.8	-17.2
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	10.9	11.4	7.5	-22.8	-25.9	-12.7
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	16.1	17.5	9.2	-27.7	-29.4	-20.5
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-4.1	-6.3	2.5	5.6	4.1	9.1
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	1.1	-0.1	4.2	0.7	0.6	1.3

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	6.1	5.5	6.1	-1.5	-4.5	5.8
$\Delta X(\alpha\text{-}^2T_3\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	11.3	11.7	7.8	-6.4	-8.0	-2.0
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-10.7	-12.3	-4.0	-8.4	-8.4	-7.5
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-5.5	-6.2	-2.3	-13.3	-11.9	-15.3
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-0.5	-0.5	-0.4	-15.5	-17.0	-10.9
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^2T_3\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	4.7	5.7	1.3	-20.4	-20.5	-18.7
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-15.5	-18.2	-5.4	12.9	13.1	11.0
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^2T_3\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-10.3	-12.0	-3.7	8.0	9.6	3.2
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^2T_3\text{-dUMP})$	-5.3	-6.3	-1.8	5.7	4.5	7.7
$\Delta X(\alpha\text{-}^3T_2\text{-dTMP} + \alpha\text{-}^3T_2\text{-UMP} - \alpha\text{-}^3T_2\text{-TMP} - \alpha\text{-}^3T_2\text{-dUMP})$	-0.1	-0.2	-0.1	0.8	1.0	-0.1
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	20.7	19.8	24.6	9.3	9.0	8.5
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	51.0	51.1	59.2	1.6	1.3	2.3
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	20.5	20.5	24.8	5.8	3.9	12.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	50.8	51.8	59.3	-1.8	-3.9	6.5
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	-18.3	-20.4	-14.3	4.1	5.1	0.5
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	12.0	10.9	20.3	-3.5	-2.7	-5.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	-18.5	-19.7	-14.2	0.7	-0.1	4.7
$\Delta X(\beta\text{-}^2T_3\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	11.8	11.6	20.4	-7.0	-7.8	-1.6
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	4.1	3.9	5.5	12.0	13.5	5.0
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	34.4	35.2	40.0	4.4	5.8	-1.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	3.9	4.5	5.6	8.5	8.3	9.1
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^2T_3\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	34.2	35.9	40.2	0.9	0.6	2.9
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^2T_3\text{-dUMP})$	-34.8	-36.3	-33.5	6.9	9.5	-3.1
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^2T_3\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-4.5	-5.0	1.1	-0.8	1.8	-9.3
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^2T_3\text{-dUMp})$	-35.0	-35.6	-33.3	3.4	4.4	1.1
$\Delta X(\beta\text{-}^3T_2\text{-dTMP} + \beta\text{-}^3T_2\text{-UMP} - \beta\text{-}^3T_2\text{-TMP} - \beta\text{-}^3T_2\text{-dUMP})$	-4.7	-4.3	1.2	-4.2	-3.3	-5.1
			pyranoses			
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	-0.8	-0.9	-0.8	-0.2	-0.3	0.7
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-20.5	-22.4	-15.3	-7.0	-8.9	-3.0

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	-7.6	-7.0	-11.9	-26.4	-27.1	-25.5
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-27.4	-28.5	-26.4	-33.1	-35.7	-29.2
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	18.7	20.4	13.6	6.1	7.7	3.8
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-1.1	-1.1	-0.9	-0.7	-0.8	0.2
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	11.9	14.3	2.5	-20.0	-19.1	-22.4
$\Delta X(\alpha\text{-}^1C_4\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-7.9	-7.2	-12.1	-26.8	-27.6	-26.1
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMp} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	7.0	6.2	11.5	26.7	27.4	25.8
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-12.8	-15.3	-3.0	19.9	18.9	22.1
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	0.2	0.1	0.4	0.5	0.6	-0.4
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^1C_4\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-19.6	-21.4	-14.2	-6.3	-7.9	-4.1
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	26.5	27.5	25.9	33.0	35.5	28.9
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^1C_4\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	6.7	6.0	11.3	26.2	27.0	25.3
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^1C_4\text{-dUMP})$	19.7	21.4	14.8	6.8	8.7	2.7
$\Delta X(\alpha\text{-}^4C_1\text{-dTMP} + \alpha\text{-}^4C_1\text{-UMP} - \alpha\text{-}^4C_1\text{-TMP} - \alpha\text{-}^4C_1\text{-dUMP})$	-0.1	-0.1	0.2	0.0	0.2	-1.0
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	0.5	0.5	0.5	-0.3	0.0	-2.5
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	29.5	28.5	32.9	25.5	24.9	26.4
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	20.9	20.9	19.4	12.9	12.9	11.2
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	49.9	48.9	51.8	38.8	37.7	40.0
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-28.5	-27.5	-32.0	-25.5	-24.4	-29.9
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	0.5	0.5	0.4	0.4	0.4	-1.0
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-8.1	-7.0	-13.1	-12.3	-11.5	-16.3
$\Delta X(\beta\text{-}^1C_4\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	20.9	21.0	19.3	13.6	13.3	12.6
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-20.5	-20.5	-18.7	-13.7	-13.2	-15.3
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	8.6	7.5	13.7	12.1	11.7	13.5
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	0.0	0.0	0.2	-0.5	-0.3	-1.7
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^1C_4\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	29.0	28.0	32.6	25.4	24.6	27.2
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-49.4	-48.4	-51.2	-38.9	-37.6	-42.8
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^1C_4\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	-20.4	-20.4	-18.8	-13.0	-12.7	-13.9
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^1C_4\text{-dUMP})$	-29.0	-28.0	-32.3	-25.7	-24.7	-29.1

Compared systems[1]	vac.[2]			solv.[3]		
	ΔE	$\Delta E_{(ZPE)}$	ΔG°	ΔE	$\Delta E_{(ZPE)}$	ΔG°
$\Delta X(\beta\text{-}^4C_1\text{-dTMP} + \beta\text{-}^4C_1\text{-UMP} - \beta\text{-}^4C_1\text{-TMP} - \beta\text{-}^4C_1\text{-dUMP})$	0.0	0.0	0.1	0.2	0.2	-0.3

[1]$\Delta X = \Delta E$, $\Delta E_{(ZPE)}$, ΔG°. [2] ΔX at B3LYP/6-311++G (d, p) in vacuum. [3] ΔX at B3LYP/6-311++G (d, p) in aqueous medium using the IEFPCM implicit solvation model.

Chapter 5

What can we say about the "Value of Information" in Biophysics?[10]

"…the genetic code is not a book nor a part of it; rather it is a translation dictionary between two different worlds (languages), i.e. the world of nucleic acids and the world of proteins…"

Cartwright *et. al.*, 2016 [1]

This chapter reproduces the content of a scientific paper [2] that is part of a bigger manuscript submitted to Physica Scripta in 2023 [3] The original paper has been modified accordingly to adapt to the research topic of this PhD thesis.

Abstract

Herein follows a flavour of a few seemingly fertile ideas from a voluminous literature of potential import in the development of biophysics. We demonstrate how some aspects of a theory developed by engineers to address problems in communication engineering are transferable to the realm of biology. Might the specific problems of biology return the favour one day, suggesting an extension of the classical theory of communication.

5.1 Early hints for a central role of "information" in biology

Modern biology and biochemistry textbooks abound with phrases like genetic code, genetic message, genetic information, replication, transcription, and translation reflecting biology's celebrated central dogma [4], that genetic information is passed unidirectionally from DNA to RNA to protein. A mutation is a change in the genetic information or an error in the copying of this

---

[10] This Chapter is based on a text co-authored with Dr. Chérif F. Matta and Dr. Peyman Fahimi. The primary author is Dr. Chérif F. Matta. My role was to contribute to the discussions and the development of the ideas.

information, either spontaneously or as a result of interaction with radiation, mutagens, or viruses. These information-theoretic sounding phrases can be traced-back to Erwin Schrödinger's influential monograph What is Life? [5] in a section titled The Hereditary Code-Script (Chromosomes).

In 1928, Frederick Griffith discovered that dead pneumococci carry a substance he termed a transforming principle that is able to transmit heritable virulence in non-virulent strains of the live bacteria. Ironically, Schrödinger's book [5] appeared in the same year (1944) as the definitive paper by Oswald Avery, Colin MacLeod, and Maclyn McCarty establishing DNA as the transforming principle and, hence, that DNA is the physical carrier of the genes [6]. Remarkably, however, the book predates by almost a decade the first reports of the discovery of the double helical structure of the DNA polymer by James Watson and Francis Crick [7], and – simultaneously - by Rosalind Franklin, Raymond Gosling [8] (and Maurice Wilkins) – the discovery that suggested an actual implementation of a code-like mode of operation for DNA [9, 10, 11]. Just a year later, the direct correspondence between the DNA language and its protein translation was proposed by the Russian physicist George Gamow [12] (although the details of how this is achieved are now known to be different that Gamow's lock-and-key proposition).

Schrödinger concludes this section by describing chromosomes with the words: "[t]hey are law-code and executive power or, to use another simile, they are architect's plan and builder's craft in one" [5]. The brilliant experiments of Leonard Adleman in the 1990's showed how DNA can be programmed into actual software to solve the traveling salesman problem numerically in the test-tube [13].

Today, following a terminology that appears to have been coined by Michael Polanyi, the distinguished physical chemist and philosopher, DNA is often referred to as the blueprint of life. [14] But a blueprint is essentially condensed information with the potential to give rise to a physical object if executed. It need not even be complete since the code's implementation interacts with the environment in producing the resulting individual, as captured by the popular phrase "Nature and nurture".

5.2 The quantity of information stored in nucleic acids and proteins: syntax

Coincidentally, the end of the 1940s also saw the birth of Claude Shannon's (classical) "Information Theory" [15, 16], a theory originally conceived in an engineering context to optimize the transmission of information through electrical wires. It did not take long for scientists to realize the relevance of this nascent theory to the realm of biology [17, 18, 19, 20, 21, 22, 23].

The intellectual atmosphere that catalyzed this appropriation was, perhaps, epitomized by the position of Michael Polanyi, who has argued strongly against a reductionist approach to biology. Polanyi was simply not convinced of the possibility, even in principle, of reducing biology to chemistry and then to physics (classical electromagnetic theory and quantum mechanics), where each level represents a "more fundamental" underlying level of description [14, 24]. For Polanyi, a living system is analogous to a "machine" in many respects, i.e., to a "mechanism" that operates in full compliance with the laws of physics and chemistry but within "boundary conditions" that are in themselves not reducible to such laws (despite not violating them) [24].

After arguing that a watch, for example, is more than just the atoms that compose it since its design as a functioning time measuring device is not a consequence of the laws of physics, Polanyi transits to biology by the following revealing statement [24]:

> "*Now, from machines let us pass on to books and other means of communication. Nothing is said about the content of a book by its physical-chemical topography. All objects conveying information are irreducible to the terms of physics and chemistry.*"

Clearly a theory of biology should somehow incorporate aspects of Information Theory since important aspects of its essence are simply boundary conditions that cannot be reduced to the laws of physics and chemistry. Biopolymers such as DNA, RNA, or proteins are a case in point. The sequence of the monomers composing those polymers is "dictated" over millions of years of evolution by unknown environmental factors and is now a "given", intrinsic to the individual from the start of its existence. The elevated temperatures at which biological systems operate will

quickly destroy any quantum coherence of entangled quantum states [25] leaving classical information theory [15] as the appropriate framework within which to study biological information. Before proceeding further, a clarification is needed. While chemical composition is irrelevant for the intended operation of a watch, in the case of the DNA, chemical structure is indispensable for its function (otherwise, for instance, how could DNA be a good substrate for the DNA polymerase or transcriptase?). The watch-DNA analogy is only meant to underscore that the information carried by the genetic material is independent of the underlying substrate and the actual physical mechanisms are independent of the chemical composition.

Influenced by Polanyi's philosophy, Lila Gatlin wrote her classic monograph "Information Theory and the Living System" [18]. The physical transmission of information from a source (e.g. DNA) to a recipient (e.g. the ribosome and eventually a protein, via mRNA) is accompanied by "noise" which may result in loss or destruction of some information, that is, an increase in the entropy of the message. A machine such as a living cell can minimize noise by ensuring that the message to be transmitted has excess information, with effective repetition providing redundancy. Chargaff's rules [26] predating the discovery of the double helix, stipulate that the composition of DNA must have equimolar amounts of the complementary bases, so [A] = [T], and [G] = [C], where A is adenine, T is thymine, G is guanine and, C is cytosine, and where the square brackets denote molar concentrations of a given base. (DNA would come to explain this through the Watson-Crick hydrogen-bonding complementarity rules, whereby A must bind to T and G to C.) However, there are no rules regulating the proportions of the AT pair with the CG pair. Thus, in real DNA, the composition is such that the concentration of AT and GC are generally different (i.e. [AT] does not equal [GC]), and the proportion ([AT]/[GC]) characterizes the specific organism.

For a language consisting of N symbols, Shannon's average information content per symbol in the message is given by the well-known relation [15]:

$$H_1 = -K \sum_{i=1}^{N} p_i \log_2 p_i \tag{5.1}$$

where if $K = 1$ and is dimensionless, H_1 is in bits (the unit adopted here), and the subscript "1" denotes that this is the average information per symbol. In equation (5.1), if $K = k_B \ln 2$ (where k_B is the Boltzman's constant) then H_1 is in units of entropy – which actually connects physical

entropy and information.

To maximize H_1 one has to equalize the probabilities of all the symbols $\{p_i\}$. The greater the departure from equiprobability, the smaller the information content of the message (to the extreme case where one symbol has a probability of 1 and all the rest zero probability, and no information is conveyed at all). In the English language, for example, the 26 letters appear with different frequencies, with "e" being the most common (probability $\sim$ 11%) and "q" the least probable (0.2%) where normalized frequencies are considered probabilities. English, therefore, has a lower information content per symbol than an "ideal" language would have with all the letters being equiprobable ($p = 1/26$). For an organism with the unlikely equiprobability of the four nucleobases, i.e. with [A] = [T] = [G] = [C] = 25%, as in E. coli, $H_1 = -\log_2 0.25 = 2$ bits per symbol. This is the maximal information carrying capacity of a nucleic acid base.

The departure from equiprobability of base pairs means that the probabilities of each of the four individual bases in the genome differ from the ideal value of 1/4. Gatlin defined the redundancy in the genetic message due to a departure from equiprobability as [18, 19]:

$$D_1 = H_1^{\max} - H_1^{\text{actual}} \tag{5.2}$$

with the subscript "1" indicating this first "type" of redundancy.

Gatlin then defined a second type of redundancy, exhibited in the genome, the departure from independence. To illustrate what this means, let's return again to the structure of the English language where, for example, the letter "q" is followed by "u" (e.g. equal, quality, or equiprobable) - thus the appearance of a given letter depends on the previous one. This is termed a first order Markov process (although higher orders of Markov processes exist, we limit ourselves to the first order for simplicity). Such a Markov process constitutes redundancy since it decreases the freedom of choice of symbols.

In the absence of this second type of redundancy, we have [18, 19]:

$$H_n^{\text{ind.}} = -\sum_{i=1}^{N} \sum_{j=1}^{N} \cdots \sum_{n=1}^{N} p_i p_j \cdots p_n \log_2\left(p_i p_j \cdots p_n\right) = nH_1 \tag{5.3}$$

implying that the total information content of the message is nothing but n times the average information content per letter or symbol.

Generally, however, there will be departures from this independence. Limiting the discussion to a first order Markov source only (with a memory $m = 1$), where the probability of a given letter in the message depends only on the letter immediately preceding it in the sequence, the departure from independence is given by [18, 19]:

$$H_n^{\text{dep.}} = -\sum_{i=1}^{N} \sum_{j=1}^{N} \cdots \sum_{n=1}^{N} p_i p_{ij} \cdots p_{(n-1)n} \log_2 \left(p_i p_j \cdots p_{(n-1)n} \right) \qquad (5.4)$$

where p_{ij} is the probability of appearance of the jth letter given that the previous letter in the message is i.

With some manipulations, the difference of equations 5.2 and 5.3 gives [18, 19]:

$$D_2 = H_2^{\text{ind.}} - H_2^{\text{dep.}} = H_1 - H_{\text{Markov}} \qquad (5.2)$$

where:

$$H_{\text{Markov}} = -\sum_{i=1}^{N} \sum_{j=1}^{N} p_i p_{ij} \log_2 \left(p_{ij} \right) \qquad (5.6)$$

The total redundancy in a DNA sequence (due to two types of redundancies) is defined by:

$$R \equiv \frac{D_1 + D_2}{\log_2(1/4)} = 1 - \frac{H_1^{\text{actual}}}{H_1^{\text{ideal}}} \qquad (5.7)$$

where "actual" refers to the characteristic redundancy of the chosen language and ideal refers to a language using the same letters but with all letters equiprobable and independent.

Redundancy measures the constraints imposed by the structure of the language that are designed to reduce transmission of errors in a message expressed in that language. It is conceivable that one of the measures of evolutionary "fitness" is how well an organism has maximized R while keeping the genetic language sufficiently flexible to code for its enormously complex structure. Since there is an inverse correlation between the redundancy and the number of potential messages expressible in a given number of symbols, a compromise must be struck.

Gatlin noted that at different steps in the evolutionary ladder organisms achieve this (constrained) maximization of redundancy by different means. The higher the organism is in the evolutionary tree, the more it achieves a higher R by keeping D_1 relatively constant while maximizing D_2. The converse is true for lower organisms which maximize their redundancy mainly by maximizing D_1. An enormous body of literature took these ideas as its point of departure in the final decades of the last century to classify organisms, quantify differences between sequences, compare coding and non-coding regions of DNA and compare homologous sequences from different organisms [27, 20]. All these ideas that apply for nucleic acid also apply for proteins, but with an alphabet comprised of 20 amino acids, which if they were equiprobable and independent would transmit a maximum of $\log_2 20 = 4.322$ bits per amino acid.

Exciting as it may be, the application of Shannon ideas to nucleic acids and proteins is limited in a significant and fundamental way - information content is a measure of entropy, no more.

5.3 The value of information stored in nucleic acids and proteins: semantics

Mikhail Volkenstein stressed the limitations of information content/entropy, emphasizing instead how one must consider the value of information in biology, too (in contrast to only the quantity of information) [28, 29, 30, 31]. Shannon's theory quantifies the amount of information (number of bits) in a message, but says nothing about the importance of this information. Volkenstein quotes [the eminent Soviet evolutionary biologist] Ivan Schmalhausen's pertinent remark that:

"the current information theory has no techniques available to it for evaluating the quality of information, although this factor is often of decisive importance in biology. When an organism receives information from the environment, first of all it evaluates this information from the standpoint of its quality..."

as "irrefutable" [32]. This statement remains essentially true today, and it is a task for the future to construct a theory of the value of biological information starting, perhaps, from where Volkenstein left off (vide infra).

Volkenstein realized that the effect on a recipient receiving information is a measure of the value of the information. He exemplified this with a "fair traffic light", meaning one that is red and green for equal amounts of time. The emission of one bit of colour information would cause considerably greater traffic to flow on a large avenue than on a small side street. Thus identical information in the Shannon sense can have dramatically different consequences depending on the receiving system [32].

Volkenstein relates the value of information to its irreplaceability, that is, non-redundancy. He argues further that the value of the information increases gradually in the course of evolution. He gives the following intriguing definition of the (dimensionless) value of information as [30, 28]:

$$V = \log_2 \frac{P_{\text{final}}}{P_{\text{initial}}} \tag{5.8}$$

where P_{final} and P_{initial} are the probabilities of producing a given effect or outcome before and after the receipt of information by the receiving system. (See [30] and references therein for the justification of choosing this definition). A reasonable "target" for an organism is to live as long as possible, while the "goal" of DNA is eventual protein synthesis.

New information is generated every time an individual of any given species is conceived through sexual reproduction by receiving half of its genetic material from its mother and half from its father. The act of sexual reproduction includes a series of random events that are not easily traceable to the laws of physics and chemistry, e.g. the decision of a particular male and female to

mate. The selection of a mate can be regarded as a Polanyi "boundary condition" [14, 24], untraceable to (but of course not violating) the laws of physics and chemistry (vide supra).

The form of equation 5.8 allows for positive or negative values of information. Imagine, for instance, that a professor, after spending an hour in class deriving an equation, discovers a mistake at the very beginning of the derivation and closes the lecture by informing the students that the entire derivation was wrong[10]. This last piece of information invalidates all information passed on during the class, and hence, has a negative value. Value can also be a function of time. Information about an impending attack by the enemy's army is valuable (actionable) intelligence before the attack but worthless once it has happened.

Further, repetition of the message before the attack has no value - it is totally redundant.

Let us examine how this idea of redundancy plays out in the eventual translation of a DNA message in a protein coding gene into the corresponding protein, assuming equiprobability of symbols for simplicity. First, in passing, we recast the trivial matter of there being three DNA letters per amino acid in terms of information theory. This minimal number of nucleotides per amino acid emerges from the ratio of the maximum information per letter of protein divided by the minimum information per letter of DNA, i.e., $4.322/2.000 = 2.161$ which, as there are no fractional nucleotides, necessitates three nucleotides per amino acid.

Now, for a protein-coding gene containing n nucleotides, $H_1^{\text{DNA}} = n * \log_2 4 = 2n$ bits. When translated to a protein, this will correspond to $H_1^{\text{protein}} = n/3 * \log_2 20 = 1.44n$ bits, i.e. there is a compression of the information on passing from DNA $\rightarrow$ protein at even at the most basic level where all bases and amino acids are equiprobable. In other words, a redundancy of $1 - 1.44/2.00 = 0.28$ exists in the primary sequence of DNA gauged with respect to its protein translation, owing to the degeneracy of the genetic code. Hence there is an increase in the value of information at the protein level – under these idealized conditions – compared to the value in the DNA sequence.

On the other hand, non-redundant information is irreplaceable. Here is where the definition in equation 5.8 comes into play. Take for example a point mutation (i.e. a mutation that changes the nature of only one of the three symbols (x, y, z) in a codon). If this mutation results in a significant change in the hydrophobicity/philicity of the coded amino acid (measured by free

---

[10] This example is not original, it was read or heard by one of the authors (C.F.M.) who regrets that he is unable to recall the source to cite it.

energy of transfer from a polar to a non-polar medium or to the gas-phase) [33, 34, 35] then this mutation is poised to have drastic effects on the protein's overall three-dimensional structure. The value of the information replaced by this mutation is, consequently, high.

The degeneracy of the genetic code is primarily in position z, in other words, 14 synonymous codons (codons coding for the same amino acid) usually differ in the third position, and hence the z-position is the least important (least valuable) position of a given codon. Meanwhile, the middle letter, y, determines whether the coded amino acid is hydrophobic or hydrophilic [33]: it is hydrophobic if this letter is pyrimidine (C or U) in the mRNA codon and hydrophilic if it is a purine (G or A). Furthermore, the middle letter is unique for a given amino acid (except for serine in which it could be either G or C), hence a mutation in the y-position almost always changes the amino acid. Thus, this letter is the most valuable since it is likely to have the most drastic consequence on the ensuing protein structure. Nature has fine-tuned the code in such a manner that the probability of replacing a residue by one with different hydrophobicity is minimized [29].

Alternatively, one can define the value of amino acids as measured by their irreplaceability in homologous protein from different species (conserved residues are more valuable). Originally, Volkenstein relied on Dayhoff's matrices of amino acid replaceability in defining the value of a given amino acid, following Bachinsky, where the "**F**unctional **S**imilarity of **A**mino acid residues (FSA)" is defined as [29]:

$$\text{FSA} = \frac{2N_{ij}}{N_i + N_j} \tag{5.9}$$

where N_{ij} is the number of times amino acid i is replaced by amino acid j within a set of homologous proteins, and where N_{ij} are the abundance of the ith or jth amino acid in the given set, respectively. The resulting (non-symmetric) matrices are 21×21 in size (20 amino acids + a termination code). They are non-symmetric because the propensity to replace (mutate) amino acid i by j is not generally equal to the probability of replacing j by i in the course of evolution.

Using these matrices and definition in equation 5.9, Volkenstein then estimates the FSA for every possible single-point mutation of every codon of the 64 codons of the genetic code. A code x,y,z can have 9 single point mutants (since we have 4 bases, one of which is already used, so the

possible mutants are 3 per position $\times$ 3 positions). If a single point mutation of a codon coincides with the same amino acid, a silent mutation, it is arbitrarily given an FSA = 100. The nine FSAs for every codon are then averaged (and divided by a numerical constant to retain a convenient magnitude), yielding q, defined as a measure of the codon irreplaceability. The value v of a residue is greater for smaller q. As an example, say the codon AAA (for lysine), yields $q = 0.74$. The value of this codon is then $v = (q + 1/2)^{-1} = 0.81$. Proceeding in this manner for all 61 unique x, y, z sense codons, the result is a genetic code table with a numerical value assigned for every coding codon [29].

If we now average the values of the (x_i, y_i, z_i) degenerate codons (different codons coding for the same amino acid), we get the value of the coded amino acid in a protein. (See Table 9.3, p. 264, of Volume II of Ref. [29]). Accordingly, the most valuable (the most irreplaceable) amino acid is tryptophan ($v_{Trp} = 1.82$) and the least valuable is alanine ($v_{Ala} = 0.52$) [29, 36]. Curiously, we note here in passing, that the partial molar volume as well as the quantum mechanically calculated molecular volume of the hydrogen-capped Trp side-chain happen to be the largest among all 20 amino acids, while that/those of Ala are the smallest [33, 34], a coincidence perhaps, but possibly worth exploring.

The average changes in the hydrophilicities of amino acids resulting from replacements of the type x $\longrightarrow$ x$^{'}$ and y $\rightarrow$ y$^{'}$ indicate that the "least dangerous" mutation is of the type A $\leftrightarrow$ G [36]. While there is a wealth of fascinating findings that we skip in this brief essay, one that stands out is that evolutionarily older proteins such as cytochrome c, unlike much more recent ones such as hemoglobin, tend to have a higher value in species that are higher in the taxonomical tree, with humans at the very top [29].

5.5   Closing remarks

Cannarozzi et al. [37] re-evaluated some of the measures of irreplaceability described above using the much larger and more recent database of Jiménez-Montaño and He [38]. In doing so, Cannarozzi et al. [37] obtained an agreement of $\sim$ 87% in the calculated values proposed by Volkenstein who used a smaller and older database [29]. Thus it appears that Volkenstein's core ideas are essentially correct even on quantitative grounds. But the field would benefit from a revisit

using the most up-to-date and extensive data and from the formulation of a full and consistent Theory of the Value of Biological Information, a theory that can serve both biophysics and communication engineering.

Today, in 2023, our knowledge has soared to unprecedented heights. That the entire human genome has been sequenced [39] is already considered history, not to mention the sequencing of the full genomes of dozens of other species. Bioinformatics is a mature field [40, 41]. UniProt [42, 43] annotates more than 20,000 proteins and their properties and locations of their coding genes. It is well established that only 2% of the genome consists of protein coding sequences while the rest of the genome does not code for any protein (**n**on-**c**oding DNA, or ncDNA). Non-coding DNA represents the bulk of nuclear DNA (98%), and its functions in living cells – if any – remain essentially an open problem. What would be the effect of mutation on these ncDNA sequences and what is their role in the first place? Are there information theoretic differences between coding and non-coding DNA? Can information theory shed light on the function of repetitive DNA segments (half of the human genome) such as tandem repeats of trinucleotides and their roles in genetic diseases such as Huntington's disease [44]? Are there information theoretic differences between nuclear and mitochondrial DNA? What is the effect of ncRNA on the translation step and its kinetics (and hence on protein folding) and what less obvious questions remain to be considered?

Irreplaceable (high value) amino acids must be crucial for the function of the protein and, hence, obvious targets for drug design and for manipulations by site directed mutagenesis and/or in vitro directed evolution and for understanding genetic disorders and viral and bacterial development of resistance (see [200] and references therein). It is entirely possible that equation 5.9 is an over-simplification, which invites further investigation into the meaning of the value of information. Might this ultimately lead to new physical theory, or perhaps even a sub-branch of the mathematics of communication?

But the role of information theory in biology does not stop at analyzing sequences. Information itself is physical, as Landauer taught us long ago [45], and to erase it you need to expend energy. The energy to erase one bit is small ($k_\mathrm{B}T \log 2$), but if this erasure is repeated by a molecular machine at a high turnover rate, the informational cost starts to be consequential. The old paradox of the extreme inefficiency of the kidney compared to any other bodily organ can only be resolved by accounting for the information theoretic cost of recognizing ions e.g. Na+ to be

selected and sorted for excretion by the kidney [23, 22, 21]. These ideas also place a limit on the thermodynamic efficiency of a molecular machine like ATP synthase/ATPase which acts as a sorting machine - picking protons for transport parallel or antiparallel to a pH gradient, respectively, across mitochondrial inner membranes or bacterial membranes [46, 47, 48, 49].

Interesting problems that do not appear to have been explored (at least extensively) in the literature include the reformulation of the following type of engineering problems into a biological context: Packet loss (i.e. the failure of a message to reach its intended destination); bit rate (the rate of information transmission); transmission delays (the time needed for a signal to flow in its entirety through a communication channel).

Translational pausing during translation regulates the rate of information flow through the mRNA-ribosome informational system apparently to allow the nascent protein sufficient time to fold properly. How is the pausing coded in the mRNA message? It is tempting to think of the information coded in the mRNA as having a dimension greater than one where the extra dimension regulates the rate of translation.

Another issue concerns the exploration of other definitions of classical information such as the Fisher information [50], originally proposed in 1922 (before Shannon's definition). Shannon's information is a "global" measure since it involves a summation (and in the limit, an integration) over the entire message. In contrast, Fisher information involves an integration over the gradient of the probability distribution function, and hence is sensitive to and magnifies local variations in the probability distribution function [50]. Can Fisher information play a role in pinpointing hot-spots in biological messages?

In closing, we draw the attention of the reader to a 1991 commentary by John Maddox "Is Darwinism a thermodynamic necessity? " on the then recent paper by J.-L. Torres in which the former proposes a thermodynamic formulation of the ill-defined concept of Darwinian "fitness" [51]. The purpose of the highlighted paper is to translate Darwinian's "fitness" into quantitative deviations from a set of ideal thermodynamics parameters characterizing a living system. Torres has succeeded, at least in principle, in lifting the circularity of the "survival of the survivors (fittest)" [51]. Could the " value" of a nucleic acid or a protein be an alternative, or perhaps an additional or complementary, dimension to measure the fitness of a species from an evolutionary standpoint?

5.6 References

[1] J. H. E. Cartwright, S. Giannerini and D. L. González, "DNA as information: at the crossroads between biology, mathematics, physics and chemistry," *Philosophical Transactions A,* vol. 374, article #20150071 (pp. 1-9), 2016.

[2] L. A. M. Castanedo, P. Fahimi and C. F. Matta, "What can we say about the "Value of Information" in Biophysics?," *Submitted to Physica Scripta,* 2023.

[3] G. Hooft, W. D. Phillips, A. Zeilinger, R. Allen, J. Baggott, F. R. Bouchet, S. M. G. Cantanhede, L. A. M. Castanedo, A. M. Cetto, A. A. Coley, B. J. Dalton, P. Fahimi, S. Franks, A. Frano, E. S. Fry, S. Goldfarb, K. Langanke, D. Nanopoulos, C. F. Matta, C. Orzel, S. Patrick, V. A. A. Sanghai, O. Shpyrko, I. K. Schuller and S. Lidström, "The sounds of science - a symphony for many instruments and voices - Part II," *Submitted to Physica Scripta,* 2023.

[4] F. Crick, "Central dogma of molecular biology," *Nature,* vol. 227, pp. 561-563, 1970.

[5] E. Schrödinger, What is life?, Cambridge: Cambridge University Press, 1944.

[6] O. T. Avery, C. M. Macleod and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III," *The Journal of Experimental Medicine,* vol. 79, pp. 137-158, 1944.

[7] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid," *Nature,* vol. 171, pp. 737-738, 1953.

[8] R. E. Franklin and R. G. Gosling, "Molecular configuration in sodium thymonucleate," *Nature,* vol. 171, pp. 740-741, 1953.

[9] F. H. C. Crick, "The genetic code," *Scientific American,* vol. 207, pp. 66-77, 1962.

[10] M. W. Nirenberg, "The genetic code: II," *Scientific American,* vol. 208, pp. 80-95, 1963.

[11] F. H. C. Crick, "The genetic code: III," *Scientific American,* vol. 215, pp. 55-62, 1966.

[12] G. Gamov, "Possible relation between deoxyribonucleic acid and protein structure," *Nature,* vol. 173, pp. 318, 1954.

[13] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science,* vol. 266, pp. 1021-1024, 1994.

[14] M. Polanyi, "Life's irreducible structure: life mechanims and information in DNA," *Science,* vol. 160, pp. 1308-1312, 1968.

[15] C. E. Shannon and W. Weaver, The mathematical theory of communication, Urbana: The University of Illinois Press, 1964.

[16] L. Brillouin, "Science and information theory," in *Dover Publications. Inc*, New York, 2004.

[17] H. Quastler, Information theory in biology, Urbana: The University of Illinois Press, 1953.

[18] L. L. Gatlin, Information theory and the living system, New York: Columbia University Press, 1972.

[19] L. L. Gatlin, "Evolutionary indices," in *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability: darwinian, neo-darwinian, and darwinian evolution*, California, University of California Press, 1972, pp. 277-296.

[20] H. P. Yockey, Information theory, evolution and the origin of life, Cambridge: Cambridge University Press, 2005.

[21] H. A. Johnson and K. D. Knudsen, "Renal efficiency and information theory," *Nature,* vol. 206, pp. 930-931, 1965.

[22] H. A. Johnson , "Information theory in biology after 18 years," *Science,* vol. 168, pp. 1545-1550, 1970.

[23] H. A. Johnson, "Thermal noise and biological information," *The Quarterly Review of Biology,* vol. 62, pp. 141-152, 1987.

[24] M. Polanyi, "Life transcending physics and chemistry," *Chemical & Engineering News,* vol. 45, pp. 54-69, 1967.

[25] O. Lombardi, F. Holik and L. Vanni, "What is quantum information," *Studies in History and Philosophy of Modern Physics,* vol. 56, pp. 17-26, 2016.

[26] E. Vischer and E. Chargaff, "The composition of the pentose nucleic acids of yeast and pancreas," *Journal of Biological Chemistry,* vol. 176, pp. 715-734, 1948.

[27] F. Fabris, "Shannon information theory and molecular biology," *Journal of Interdisciplinary Mathematics,* vol. 12, pp. 41-87, 2009.

[28] M. V. Volkenstein and D. S. Chernavskii , "Evolution and value of information," in *Self-organization autowaves and structures far from equilibrium*, Berlin, Springer-Verlag, 1983, pp. 252-261.

[29] M. Volkenstein , General biophysics, vol. 1 and 2, New York: Academic Press Inc, 1983.

[30] M. V. Volkenstein, Physical Approaches to Biological Evolution, Berlin Heidelberg: Springer-Verlag, 1994.

[31] M. V. Volkenstein, Entropy and information, Springer Science & Business Media, 2009.

[32] M. V. Volkenstein, "Information theory and evolution," in *Cybernetics of living matter: nature, man, information*, Moscow, Mir Publishers, 1987, pp. 83-94.

[33] C. F. Matta, "Modeling biophysical and biological properties from the characteristics of the molecular electron density, electron localization and delocalization matrices, and the electrostatic potential," *Journal of Computational Chemistry,* vol. 35, pp. 1165-1198, 2014.

[34] C. F. Matta, Quantum biochemistry: electronic structure and biological activity, Weinheim: Wiley-VCH, 2010.

[35] A. Radzicka and R. Wolfenden, "Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution," *Biochemistry,* vol. 27, pp. 1664-1670, 1988.

[36] M. V. Volkenstein, "Mutations and the value of information," *Journal of Theoretical Biology,* vol. 80, pp. 155-169, 1979.

[37] A. Schneider, G. M. Cannarozzi and G. H. Gonnet, "Empirical codon substitution matrix," *BMC Bioinformatics,* vol. 6, article #134 (pp. 1-7), 2005.

[38] M. A. Jiménez-Montańo and M. He, "Irreplaceable amino acids and reduced alphabets in short-term and directed protein evolution bioinformatics research and applications," in *International Symposium on Bioinformatics Research and Applications*, 2009.

[39] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature,* vol. 409, pp. 860-921, 2001.

[40] D. W. Mount, Bioinformatics: Sequence and Genome Analysis, New York: Cold Spring Harbor Laboratory Press, 2004.

[41] C. Gibas and P. Jambeck, Developing Bioinformatics Computer Skills, Cambridge: O'Reilly, 2001.

[42] T. U. Consortium, "UniProtKB," 2020. [Online]. Available: http://www.uniprot.org/.

[43] U. Consortium, "UniProt: a hub for protein information," *Nucleic Acids Research,* vol. 43, pp. D204-212, 2015.

[44] K. Usdin, N. C. M. House and C. H. Freudenreich, "Repeat instability during DNA repair: insights from model systems," *Critical Reviews in Biochemistry and Molecular Biology,* vol. 50, pp. 142-167, 2015.

[45] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development,* vol. 5, pp. 183-191, 1961.

[46] J. N. Vigneau, P. Fahimi, M. Ebert, Y. Cheng, C. Tannahill, P. Muir and et. al, "ATP synthase: a moonlighting enzyme with unprecedented functions," *Chemical Communications* , vol. 58, pp. 2650-2653, 2022.

[47] P. Fahimi, M. A. Nasr, L. A. M. Castanedo, Y. Cheng, C. A. Toussi and C. F. Matta, "A note on the consequences of a hot mitochondrion: some recent developments and open questions," *Biohysical Bulletin,* vol. 43, pp. 14-21, 2020.

[48] C. F. Matta and L. Massa, "Notes on the energy equivalence of information," *Journal of Physical Chemistry A,* vol. 121, pp. 9131-9135, 2017.

[49] C. F. Matta and L. Massa, "Information theory and the thermodynamic efficiency of biological sorting systems: case studies of the kidney and of mitochondrial ATP-synthase," in *Sustained Energy for Enhanced Human Functions and Activity*, Elsevier, 2017, pp. 3-29.

[50] B. R. Frieden, Physics from Fisher Information: A Unification, Cambridge: Cambridge University Press, 1998.

[51] J.-L. Torres, "Natural selection and thermodynamic optimality," *Il Nuovo Cimento,* vol. 13, pp. 177-185, 1991.

Chapter 6

Knowledge transfer plan

"DNA is like a computer program but far, far more advanced than any software ever created"

Bill Gates *et al*,

in: "The Road Ahead" [1]

6.1   Applications of nucleotide and nucleoside analogs

Quantum (bio) chemistry, molecular modeling, and bioinformatics of the thermodynamic properties of canonical and non-canonical nucleotides can be used to design new genetic drugs in the prevention and the diagnosis of degenerative diseases [2, 3, 4, 5, 6, 7, 8].

Nucleo**s**ides(**t**ide**s**) (Ns(t)s) **ANA**logs (Ns(t)ANA) that are building blocks of *Xeno*-**N**ucleic **A**cids (XNA) {usually contains a phosphate group, a canonical nucleobase and a TC different from 2dRibf or Ribf} have a large array of applications in the pharmaceutical industry as therapeutic drugs. For instance, Ns(t)ANA can be used as antivirals against HIV, hepatitis B and C, cytomegalovirus, varicella-zoster, ebolavirus and more recently against SARS-Cov2 [9, 10]. Their antiviral activity is possible since they can be incorporated to the virus DNA as a mutation that is not recognized by the DNA polymerase enzyme inhibiting the replication of the virus. This phenomenon is known as "lethal mutagenesis" [9]. Additionally, Ns(t)ANA and short single or double stranded polynucleotides with lengths from 13-30 units have been used as antibacterial, {e.g., derivatives from immunicillins} [11], antitumor [9] and anticancer drugs [12, 13].

Until the date a handful of Ns(t)ANA have been reported in the literature for their therapeutic properties. By 2021 Yamada reported that by October, 2021 around 15 oligonucleotides were approved by the United States **F**ood and **D**rug **A**dministration (FDA) and around 15 were in clinical trial phase 3 [13]. Until the date, the author of this dissertation has identified around 30 NsANA from the literature as antiviral drugs and around 4 as antibacterial [9, 13, 11, 14].

Synthetic Ns(t)ANA have also been incorporated in DNA/RNA molecular cages as drug delivery systems [15] and been used for the design of short single stranded RNA or DNA

oligonucleotides known as "aptamers" [16]. Aptamers similar to antibodies can bind with high affinity to biological targets which make them excellent genetic markers for Alzheimer's Disease [17], AIDS [10], cancer, hepatitis [18].

Aptamers are selected from DNA/RNA oligonucleotide libraries usually through in vitro "**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment" (SELEX) [19]. This process consists in a combinatorial design of a group of oligonucleotide sequences, usually 10^{12}-10^{18} that are obtained from mutating the nucleotides. The bulk of candidates are tested for their affinity for a given target. The sequences that bind to the target are selected from the mixture and amplified by using **P**olymerase **C**hain **R**eaction (PCR) experiments. The process is repeated from 6-15 times to enrich a library of **oligoN**ucleo**tide**s (oligoNts) with the highest binding affinities. The folding of the oligoNts depends on the experimental conditions [10].

SELEX can be a time {weeks, months} and resources consuming process and it has a low success rate. In order to accelerate and guide the experimental work involved in SELEX, *in-silico* and bioinformatic tools are usually used. The computational search for the better binding oligoNts sequences include the design of the secondary and 3D structure of the aptamer starting from the sequence of nucleotides, docking of given aptameric structures, **M**olecular **D**ynamics (MD) [20, 21, 22] to predict binding affinities to different targets and more recently the use of **M**achine **L**earning (ML) and **D**eep learning with **N**eural **N**etworks (DNN) [23] to predict their molecular structures and properties [24, 25].

Aptamers usually contain canonical TCs, RUs and ILs but their nucleic acid chemical space can be expanded *much more* than that of the five canonical bases (A, G, C, T, U), the 2dRibf, Ribf or the phosphate ion into the "**A**rtificially **E**xpanded **G**enetic **I**nformation **S**ystem" (AEGIS) [18]. AEGIS can have different chemical modifications [26]. Some studies have reported the design of XNA aptamers containing prebiotic nucleotides similar to the ones studied in this dissertation. For example, Rangel and coworkers [27] reported an aptamer containing a TNA oligonucleotide for its enhanced binding to a food contaminant called **O**chra**T**oxin **A** (OTA).

Aptamers are typically formed from nucleobases and a sugar-phosphate(s) in the β-configuration, *but not commonly including other nucleic acid configurations such as α-configuration*. Could the incorporation of alternative anomeric forms improve the binding affinities of aptamers to specific targets? A publication Kolganova and coworkers [28] reported a modified Thrombin-binding aptamer that contained *α*-2′-deoxyguanidine monophosphate and *α*-

deoxythymidine monophosphate. Anomeric modifications of certain regions favored the aptamers' folding and further thermal stabilization of their 3D structures.

The pressing hypothesis for the future application of the research results presented in this thesis is the following: can we use prebiotic nucleosides(tides) in a similar fashion to expand the repertoire of potential therapeutic Ns(t)ANA and aptamers? Perhaps, a similar approach to therapy could be innovated using ancestral Ns(t)s. One key innovation will be to combine variations of the non-canonical Ns(t)s in their two *β*- and *α*-anomeric forms, under the hypothesis that these *non-standard aptamers could be (perhaps better?) genetic markers for different diseases.*

6.2   Possible next steps: libraries of nucleosides(tides) analogs for therapeutic uses

6.2.1   Creating libraries of new nucleotides and nucleosides analogs

How can we filter down the quasi-infinite Ns/Nt chemical space to the few most promising molecular structures?

The first steps could be to investigate if there are other potential prebiotic Ns(t)s with similar properties to the ones modeled in this thesis.

Generative **A**rtificial **I**ntelligence (GAI) or graph-based **D**eep **L**earning (DL) will be implemented by training a **D**eep **N**eural **N**etwork (DNN) to predict potential prebiotic nucleotides with similar molecular structure and quantum energies to the ones investigated in Chapter #3 and #4. Generative models and DNN have been reported in the literature for their capability to predict Ns/Nts analogs and antiviral small molecules [29, 30]. In a number of different datasets DNN models have also been proven to predict more accurately molecular descriptors than traditional descriptor-based Machine Learning (ML) [31].

To the extent of our knowledge the number of nucleotides analogs proven as therapeutics is scarce in the literature and is not abundant enough to be used in ML methods. Hence, the first step will be to expand the library of biologically active Ns/Nt analogs by generating a

comprehensive database that will be used to predict new prebiotic nucleosides(tides) structures and their molecular properties.

Hopefully, this research will allow us to find new potential Ns(t)ANA candidates that can be easily synthesized in the lab through condensation reactions of their components and that have enhanced biological activity, e.g., better inhibitors of the DNA polymerase in different viruses and bacteria.

6.1.2 The training set

A database will be built and the DNN model will be train on the results from Chapter #3 and #4. The initial training database will contain information for $\approx$ 506 nucleosides + 968 nucleotides = 1474 Ns(t)s.

After the initial training this dataset will be retrained on other known Ns(t)ANA databases. These are: i) the SARS-CoV-2 nucleosides {contains 15 nucleosides with therapeutic activity against SARS-CoV-2} [32], ii) the Synthetic Nucleosides database {with 188 nucleosides' synthetic analogs} [33] and iii) a subset of 40 bioactive nucleotides and 113 nucleosides obtained from the ChemBL database [34].

A number of descriptors will be estimated and/or added to the training databases as needed. These include:

Structural features and reaction conditions

- Smiles for TCs, RUs, ILs, Ns and corresponding Nts.
- **S**tructural information entropy (S_i) and **S**tructural **I**nformation **C**ontent (SIC) [35]. These features has been used to predict plausibility for the predominance of astrobiological molecules. The hypothesis states that molecules with lower symmetry or simpler molecular structure are more probable to exist.
- RU conformation (*syn* or *anti*).
- Sugar ring puckering parameters. These parameters are the phase angle phi (ϕ_2), theta (θ) = 0 and the total puckering amplitude (Q) for F-forms and the phase angles {phi (ϕ) and theta (θ)} and the radial Q for P-forms.

- Molecular weight.

- Van der Waals volume.

- RU's pK_a.

- Environment {vacuum or aqueous solution}

- Temperature.

- The application of the Lipinski rule of 5: To be orally-active, a molecule should have five characteristics to be "drug-like", these are: Molecular mass 500 D or less, log$P < 5$, 5 hydrogen bond donors or less, a maximum of 10 hydrogen bond acceptors, and a molar refractivity from 40 to 130 [36, 37].

Quantum descriptors

- ΔX, $\Delta X_{(a+b)}$ and $\Delta X_{(c+d)}$ {energies for classic and alternative synthesis of Ns(t)s}.

- ΔX_n (where *n*: 1, 2, 3…6){energies for pseudorotational equilibrium and anomer-exchange reactions in vacuum and aqueous solution}.

- Types of intramolecular interactions, calculated from the **Q**uantum **T**heory of **A**toms in **M**olecules (QTAIM) [38].

- Activation energies from the modeling of hydrolysis mechanism for the glycosidic bond and the TC-IL ester bond of the given glycosides.

Pharmacokinetic descriptors:

- Drug-likeness using the Lipinsky rules [36, 39] and **Q**uantitative **E**stimate of **D**rug-likeness index (QED) from Bickerton and coworkers [40].

- Log$P_{\text{octanol-water}}$ partition coefficient [41] as hydrophobicity index.

- $\Delta G_{\text{affinity}}$ for the given Ns(t)ANA against different proteins and DNA polymerases. The $\Delta G_{\text{affinity}}$ will be obtained as follows: molecular docking of the candidates onto different biological targets, e.g., DNA polymerase will be performed by using Autodock Vina [42]. The best docking poses will be refined by using **Q**uantum **M**echanics/**M**olecular **M**echanics (QM/MM) ONIOM (**O**ur own **N**-layered **I**ntegrated molecular **O**rbital and **M**olecular

mechanics) [43, 44] calculations as implemented in Gaussian 16 [45]. QM/MM is a "layering /multilevel" approach treating the atoms near the reaction center quantum mechanically (at high accuracy) and distant atoms with molecular mechanics, balancing accuracy and speed [46]. These methods (and also DFT) are implemented in *Gaussian16* [45].

The overall goal of the previous steps is to generate a model robust enough from experimental and theoretical data to predict the molecular structure of new Ns(t)ANA that fulfill four main requirements: 1) can be easily synthetized in the lab through condensation reactions of their components, 2) are stable in physiological conditions, 3) are resistant to hydrolysis and 4) can bind with good affinity to specific biological targets of interest. Simultaneously this will also allow to identify molecules from the initial training set that can be used as new therapeutics.

In order to select potential new therapeutic Ns/Nts from the training sets the following criteria will be used:

- Reaction energies $\leq$ -17 kJ/mol.
- Energies $\leq$ -17 kJ/mol for at least ΔG_2°, ΔG_4°, ΔG_5° and $\geq$ 17 kJ/mol for ΔG_6°.
- Lower or similar S_i to canonical Ns(t)s.
- Shorter or comparable excited states lifetime to canonical bases ($\tau_{\text{canonical bases}} = 0.1\text{-}2.8\text{ps}$) [47].
- Lower hydrolytic rate constants.
- Higher binding affinity to specific targets.

Potential candidates from the generative DNN models will be selected based on their fingerprint similarities to promising Ns(t)ANA from the training sets using the **T**animoto **S**imilarity index (S_T) [48, 49]. This principle follows the "guilt by association" rule [50] that proposes that molecules with similar structure have affinity for similar proteins.

The final dataset of compounds from the training and generative models can be used to find new candidates from other known databases to expand the library of potential therapeutic Ns(t)ANA. For instance, Cleaves II and coworkers (Ccw) [51] generated around 706,568 Ns/Nts analogs with the elements C, H and O and 454,422 analogs with C, H, N and O from the molecular formulas $BC_{3\text{-}7}H_{5\text{-}15}O_{2\text{-}4}$ and $BC_{3\text{-}6}H_{5\text{-}15}N_{1\text{-}2}O_{0\text{-}4}$, where B is the canonical bases A, G, C, T or U. We can use our models to compare and cluster these two large databases based on fingerprint

similarity using for example the Tanimoto index. The structures with $S_T \geq 0.85$ will be selected and further experimental testing and the rest discarded.

The work here aims at preceding and guiding experimentation - it is meant to lead future (1) laboratory detection methods for new genetic markers, If successful, such an approach can save treasure and labor by cutting through the process of trial and error. We aim to (considerably) reduce the chemical space of promising compounds, from quasi-infinite to a space of much fewer molecular structures. But how? The group of my supervisor (Dr. Matta), known for work in quantum biochemistry [319, 320, 321, 322, 323, 324, 325], has recently acquired a \$76,500 **C**anada **F**oundation for **I**nnovation (CFI) grant to add a new *QuantumCube^{TM}* [59] cluster (around Dec. 2020) to the existing computational infrastructure of the lab. This cluster will largely be dedicated 50-70% to undertake this project.

The successful candidates from the previous study will be selected and their synthesis and biological activity will be tested in laboratory experiments. Collaborators will be sought to implement the synthesis and the preliminary testing of the most promising Ns(t)ANA and follow with the in vitro and in vivo testing in animal models.

6.1.3 Using nucleosides(tides) analogs in the development of an aptamer for Alzheimer's disease

Let's select a therapeutic target for the design of specific aptamers. For instance, efforts will be made to design new and better aptamers for the detection of Alzheimer's disease creating a library of promising aptamers structures from various combinations of canonical and non-canonical nucleosides(tides) (several hundreds). These aptamers will then be tested for their binding strengths (free energies of binding) and affinities to the faulty β-amyloid peptides (e.g. Aβ1-40 and Aβ1-42 fibrils). This last step will be achieved by first visual/manual docking followed by MD simulations. The aptamers will then be ranked for their affinities to β-amyloid peptides. A fluorescent tag on the aptamer can be introduced to visualize, in practice, patches of amyloid plaques in diseased brain tissues of animal models (see e.g., [60]).

In summary, I plan to expand the repertoire of aptamers by testing (1) non-canonical sugars (e.g. 6-MR sugars and/or by the inclusion of prebiotic nucleosides(tides)), (2) PNAs, and (3) nucleic acids but in the α-configuration. Combinations of these modifications will be thoroughly

tested as well (for the first time, perhaps). This will enlarge the "aptamer space" considerably. The proposed work will, hopefully, lead experiment and the development of laboratory detection methods for genetic markers of degenerative diseases potentially saving labor and substantial costs.

6.3   References

[1]   B. Gates, N. Myhrvold and P. M. Rinearson, The Road Ahead, New York: Penguin USA, 1996.

[2]   C. F. Matta, Quantum biochemistry: electronic structure and biological activity, Weinheim: Wiley-VCH, 2010.

[3]   Y. C. Martin, Quantitative drug design: a critical introduction, Boca Raton: CRC Press - Taylor & Francis Group, 2010.

[4]   D. W. Mount, Bioinformatics: sequence and genome analysis, New York: Cold Spring Harbor Laboratory Press, 2004.

[5]   T. Schlick, Molecular modeling and simulation: an interdisciplinary guide, New Yok: Springer, 2002.

[6]   C. Gibas and P. Jambeck, Developing bioinformatics computer skills, Cambridge: O'Reilly, 2001.

[7]   R. Carbó-Dorca, D. Robert, L. Amat, X. Gironé and E. Besalú, Molecular quantum similarity in QSAR and drug design, Berlin: Springer, 2000.

[8]   W. G. Richards, "Quantum chemistry in drug design," *Pure and Applied Chemistry,* vol. 60, pp. 277-279, 1988.

[9]   A. Abdullah and A. Awadh, "Nucleotide and nucleoside-based drugs: past, present, and future," *Saudi Journal of Biological Sciences,* vol. 29, article #103481 (pp. 1-7), 2022.

[10] V. M. González, M. E. Martín, G. Fernández and A. García-Sacristán, "Use of aptamers as diagnostics tools and antiviral agents for human viruses," *Pharmaceuticals,* vol. 9, article #78 (pp. 1-34), 2016.

[11] A. A. Zenchenko, M. S. Drenichev, I. A. Il'cheva and S. N. Mikhailov, "Antiviral and antimicrobial nucleoside derivatives: structural features and mechanisms of action," *Molecular Biology,* vol. 55, pp. 786-812, 2021.

[12] R. T. Walter , E. D. Clercq and F. Eckstein, Nucleoside Analogues: Chemistry, Biology and Medical Applications, New York and London: Springer Science & Business Media, 2012.

[13] Y. Yamada, "Nucleic acid drugs-current status, issues, and expectations," *Cancers,* vol. 13, article #5002 (pp. 1-19), 2021.

[14] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu , J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Research,* vol. 46, pp. D1074-D1082, 2017.

[15] V. Linko, A. Ora and M. A. Kostiainen, "DNA nanostructures as smart drug-delivery vehicles and molecular devices," *Trends in Biotechnology,* vol. 33, pp. 586-594, 2015.

[16] T. Adachi and Y. Nakamura, "Aptamers: a review of their chemical properties and modifications for therapeutic application," *Molecules,* vol. 24, article #4339 (pp. 1-14), 2019.

[17] C. Madhuri, S. Chen, P. R. Dodd and R. N. Veedu, "Nucleic acid-based theranostics for tackling Alzheimer's disease," *Theranostics,* vol. 7, pp. 3933-3947, 2017.

[18] E. Biondi and S. A. Benner, "Artificial expanded genetic information systems for new aptamers technologies," *Biomedicines,* vol. 6, article #53 (pp. 1-13), 2018.

[19] Y. Zhang, B. Lai and M. Juhas, "Recent advances in aptamer discovery and applications," *Molecules,* vol. 24, pp. 941-963, 2019.

[20] J. W. Ponder and D. A. Case, "Force fields for protein simulations," *Advances in Protein Chemistry,* vol. 66, pp. 27-85, 2003.

[21] D. Frenkel and B. Smit, Understanding molecular simulation: from algorithms to applications, New York: Academic Press, 2002.

[22] T. Schlick, Molecular modeling and simulation: an interdisciplinary guide, New York: Springer, 2002.

[23] M. Chester, Neural networks: a tutorial, Ebglewood Cliffs: PTR Prentice Hall, 1993.

[24] S. J. Lee, J. Cho, B.-H. Lee, D. Hwang and J.-W. Park, "Design and prediction of aptamers assisted by in silico methods," *Biomedicines,* vol. 11, article #356 (pp. 1-22), 2023.

[25] Z. Chen, L. Hu, B.-T. Zhang, A. Lu, Y. Wang and Y. Yu, "Artificial intelligence in aptamer-target binding prediction," *International Journal of Molecular Sciences,* vol. 22, article #3605 (pp. 1-17), 2021.

[26] S. Ni , H. Yao, L. Wang, J. Lu, F. Jiang, A. Lu and G. Zhang, "Chemical modifications of nucleic acid aptamers for therapeutic purposes," *International Journal of Molecular Sciences,* vol. 18, article #1683 (pp. 1-21), 2017.

[27] A. E. Rangel, Z. Chen, T. M. Ayele and J. M. Heemstra, "In vitro selection of an XNA aptamer capable of small-molecule recognition," *Nucleic Acids Research,* vol. 46, pp. 8057-8068, 2018.

[28] N. A. Kolganova, A. M. Varizhuk, R. A. Novikov, V. L. Florentiev, G. E. Pozmogova, O. F. Borisova, A. K. Shchyolkina, I. P. Smirnov, D. N. Kaluzhny and E. N. Timofeev, "Anomeric DNA quadruplexes: modified thrombin aptamers," *Artificial DNA: PNA & XNA,* vol. 4, article#e28422 (pp. 1-8), 2014.

[29] J. Mao, J. Wang, A. Zeb, K.-H. Cho, H. Jin, J. Kim, O. Lee, Y. Wang and K. T. No, "Transformer-based molecular generative model for antiviral drug design," *Journal of Chemical Information and Modeling,* 2023.

[30] D. Dablain , G. Siwo and N. Chawla , "Generative AI design and exploration of nucleoside analogs," *ChemRxiv.,* vol. Cambridge: Cambridge Open Engage, pp. This content is a preprint and has not been peer-reviewed., 2021.

[31] D. Jiang, Z. Wu, C.-Y. Hsieh, G. Chen, G. Liao, B. Liao, Z. Wang, C. Shen, D. Cao, J. Wu and T. Hou, "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models," *Journal of Cheminformatics,* vol. 13, article #12 (pp. 1-23), 2021.

[32] D. C. Schultz, R. M. Johnson, K. Ayyanatha, J. Miller, K. Whig, B. Kamalia, M. Dittmar, S. Weston, H. L. Hammond, C. Dillen, L. Castellana, J. S. Lee, M. Li, E. Lee, S. Constant, M. Ferrer, C. A. Thaiss, M. B. Frieman and S. Cherry, "Pyrimidine biosynthesis inhibitors synergize with nucleoside analogs to block SARS-CoV-2 infection," *bioRxiv [Preprint],* 2021.

[33] "Selleckchem Nucleoside Analog Library," [Online]. Available: https://www.selleckchem.com/screening/Nucleoside-Analogue-Library.html). [Accessed 16 August 2023].

[34] A. Gaulton, A. Hersey, M. Nowotka, P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit and A. R. Leach, "The ChEMBL database in 2017," *Nucleic Acids Research,* vol. 45, pp. D945-D954, 2017.

[35] D. S. Sabirov, "Information entropy of interstellar and circumstellar carbon-containing molecules: molecular size aga636 structural complexity," *Computational and Theoretical Chemistry,* vol. 1097, pp. 83-91, 2016.

[36] C. A. Lipinski, "Lead- and drug-like compounds: the rule-of-five revolution," *Drug Discovery Today: Technologies,* vol. 1, pp. 337-341, 2004.

[37] C. A. Lipinski , F. Lombardo , B. W. Dominy and P. J. Feeney , "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews,* vol. 23, pp. 3-25, 1997.

[38] C. F. Matta and R. J. Boyd, The Quantum Theory of Atoms in Molecules: from solid state to DNA and drug design, Weinheim: Wiley-VCH, 2007.

[39] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews,* vol. 23, pp. 3-25, 1997.

[40] G. R. Bickerton, G. V. Paolini , J. Besnard, S. Muresan and A. L. Hopkins, "Quantifying the chemical beauty of drugs," *Nature Chemistry,* vol. 4, pp. 90-98, 2012.

[41] D. Roy and C. Patel, "Revisiting the use of quantum chemical calculations in LogPoctanol-water prediction," *Molecules,* vol. 28, article #801 (pp. 1-9), 2023.

[42] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry,* vol. 31, pp. 455-461, 2010.

[43] F. R. Clemente, T. Vreven and M. J. Frisch, "Getting the most out of ONIOM: guidelines and pitfalls," in *Quantum Biochemistry: Electronic Structure and Biological Activity*, Weinheim, Wiley-VCH, 2010, pp. 61-83.

[44] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li , Z. Ke, F. Liu, H.-B. Li, L. Ding and K. Morokuma, "The ONIOM method and its applications," *Chemical Reviews,* vol. 115, pp. 5678-5796, 2005.

[45] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji , X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian , J. V. Ortiz, A. F. Izmaylov , J. L. Sonnenberg , D. Williams-Young , F. Ding, F. Lipparini, F. Egidi, J. Goings , B. Peng , A. Petrone , T. Henderson , D. Ranasinghe , V. G. Zakrzewski , J. Gao, N. Rega,

G. Zheng , W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida , T. Nakajima , Y. Honda , O. Kitao , H. Nakai , T. Vreven , K. Throssell , J. A. J. Montgomery , J. E. Peralta , F. Ogliaro , M. J. Bearpark , J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari , A. P. Rendell , J. C. Burant, S. S. Iyengar , J. Tomasi , M. Cossi , J. M. Millam , M. Klene , C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, revision C.01, Wallingford CT: Gaussian Inc, 2019.

[46] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," *Angewandte Chemie International Edition,* vol. 48, pp. 1198-1229, 2009.

[47] A. C. Rios and Y. Tor, "On the origin of the canonical nucleobases: an assesment of selection pressures across chemical and early biological evolution," *Israel Journal of Chemistry,* vol. 53, pp. 1-15, 2013.

[48] D. Bajusz, A. Rácz and K. Héberger, "Why is Tanimoto index an appropriate choice for fingreprint-based similarity calculations?," *Journal of Cheminformatics,* vol. 7, article#s13321-015-0069-3 (pp. 1-13), 2015.

[49] T. Tanimoto, "An elementary mathematical theory of classification and prediction," International Business Machine (IBM) Corporation, New York, 1957.

[50] M. Thafar, A. B. Raies, S. Albaradei, M. Essack and V. B. Bajic, "Comparison study of computational prediction tools for drug-target binding affinities," *Frontiers in Chemistry,* vol. 7, article #782 (pp. 1-19), 2019.

[51] H. J. Cleaves II, C. Butch, P. B. Burger and J. Goodwin, "One among millions: the chemical space of nucleic acid-like molecules," *Journal of Chemical Information and Modeling,* vol. 59, no. 10, pp. 4266-4277, 2019.

[52] C. F. Matta, "Google Scholar," 2023. [Online]. Available: https://scholar.google.ca/citations?user=dewOps8AAAAJ&hl=en.

[53] C. F. Matta, "Research Website," 2023. [Online]. Available: https://www.cmatta.ca/.

[54] L. A. Monteserín Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article#e09657 (pp. 1-12), 2022.

[55] P. Fahimi, M. A. Nasr, L. A. Monteserín Castanedo, Y. Cheng, C. A. Toussi and C. F. Matta, "A note on the consequences of a hot mitochondrion: Some recent developments and open questions," *Biophysical Bulletin,* Vols. [Note: this journal does not file by volumes, only issues], no. 43, pp. 14-21, 2020.

[56] M. A. Nasr, G. I. Dovbeshko, S. L. Bearne, N. El-Badri and C. F. Matta, "Heat shock proteins in the "hot" mitochondrion: identity and putative roles," *BioEssays,* vol. 41, article #1900055, (pp. 1-6), 2019.

[57] L. A. Monteserín Castanedo, A. Sánchez Lamar, B. Morera Boado, A. de la Nuez Veulens and C. F. Matta, "Genoprotection by complexation: The case of Phyllanthus orbicularis K extract," *Computational and Theoretical Chemistry,* vol. 1164, article #112555 (pp. 1-9), 2019.

[58] A. A. Arabi and C. F. Matta, "Effects of intense electric fields on the double proton transfer in the Watson-Crick guanine-cytosine base pair," *Journal of Physical Chemistry B,* vol. 122, pp. 8631-8641, 2018.

[59] Parallel Quantum Solutions (PQS) Inc, [Online]. Available: http://www.pqs-chem.com.

[60] B. Laurijssens , F. Aujard and A. Rahman, "Animal models of Alzheimer's disease and drug development," *Drug Discovery Today: Technologies,* vol. 10, pp. e319-e327, 2013.

Chapter 7

Conclusions

> "DNA neither cares nor knows. DNA just is. And we dance to its music."
>
> Richard Dawkins., 1995 [1]

7.1 Can thermodynamics be a driver for the selection of the building blocks of proto, pre and today's nucleic acids?

If it is considered that the prebiotic synthesis of canonical Ns(t)s may have occur in a hotter prebiotic Earth billions of years ago (T ≈ 55 °C - 85 °C) [3], the formation of these building blocks should be controlled by thermodynamics. Condensation reaction between D-ribofuranose or D-2'-deoxyribofuranose and A, G, C, T or U and condensation of Ns + $H_2PO_4^-$ are usually endergonic processes ($\Delta H_{reaction}^{\circ} > 0$) in aqueous solution [3]. Following LeChâtelier's principles [4] {reactants ⇌ products} equilibrium should be displaced towards products if the concentration of reactants increases [3]. Hence, dry regions on the prebiotic earth and evaporation processes (dehydration cycles) on the surface of rain ponds and tidal pools would have been ideal scenarios to find precursors of NAs [3].

In practice different experimental reports in the literature have studied the prebiotic synthesis of the canonical building blocks of NAs from simple condensation reactions of their components obtaining low or no yields for the desired products when heating the solution in a reflux system that simulates wet-dry cycles. These experiments have given birth to the "water problem" [25, 26, 11, 27] that defines that the formation of glycosidic bonds between the dr or r and canonical bases and the formation of phosphodiester bonds between two nucleosides is thermodynamically unfavorable when water is present, together with the thermal instability at temperatures ≥ 85 °C for some of the components Ns(t)s [5, 6, 7]. For instance, the formation of the phosphodiester bonds between DNA and RNA Nts have been estimated to be ≈ 3.3 kcal/mol ([8] and see pp. 493 of [9]).

Meanwhile the canonical components face the "water problem" other alternative molecules have been proven to circumvent the water problem [52].

Erwin Schrödinger, in 1943, was among the first to suggest that quantum mechanics could solve the mysteries of the origins of life [15]. Since then, the world has seen an accelerated development of quantum and computational chemistry, molecular modeling, and molecular dynamics (both with classical fields or using quantum-mechanically calculated forces (the latter known as ab initio MD)). Thanks to these methods, the in silico *modeling of the prebiotic chemistry of ancestral nucleic acids* [16] *has now become at our fingertips. The same can be said about the theoretical investigations aiming at determining the physical chemical properties of contemporary DNAs and RNAs* [17, 18, 19, 20].

Through this dissertation the following hypothesis has been tested by using computational modeling and quantum mechanics: can thermodynamics explain the prebiotic selection and emergence of the building blocks of today's nucleic acids or an ancestral proto-RNA? The short answer is: *it could*! But first let's define a few important concepts. The use of the word "thermodynamics" refers to the energetic changes or differences in terms of free energies, zero point and corrected zero-point energies for the different anomers of canonical and non-canonical nucleosides(tides) and their classic and alternative synthesis from the different order of addition of the reactants. "Canonical" refers to building blocks with the "naturally observed" β-anomer, containing the five canonical RUs (2 purines {A, G} and 3 pyrimidines {C, T, U}), either the TCs D-ribofuranose or D-2'-deoxyribose and the hydrogen-monophosphate ion as IL. Meanwhile "non-canonical" refers to nucleosides(tides) that can be building blocks of *Xeno*-**N**ucleic **A**cids (XNA) or prebiotic alternatives with either; the "incorrect" α-configuration or any components different from the canonical ones, e.g., different bases like TAP, BA, CA, MM and/or different TCs like glycerol, glyceric acid, AEG and/or the hydrogen-arsenate ion.

Chapter 2 (based on Ref. [1]) explores the thermodynamic selection of the "naturally observed" anomer β of the canonical nucleosides(tides) finding that there were not significant differences {energies were within 2-4 kcal/mol, which is the intrinsic error for the theoretical method} between the different configurations. Additionally, it is suggested that in vacuum either classic or alternative pathway favors the formation of the correct anomers, but the classic pathway emerges as the thermodynamically favored. No Ns(t) was predicted to be obtained by either pathway in the presence of implicit solvation. Additionally, it was determined that only for the

building blocks from the classic pathway in vacuum and for the β-anomer a sugar-exchange reaction could favor the selection of T for DNA and U for RNA.

In Chapters 3 and 4 the previous study was expanded by increasing the chemical space of the TCs, RUs and ILs to be used in the design of the different building blocks. Additionally, the effect of starting from different sugar ring conformations was considered. The results presented in these two chapters reproduced some of the experimental results reported in the literature with respect to the synthesis of some non-canonical Ns and Nts, e.g., the BA-C^5 glycosylated Ns(t)s were overall predicted as the most favored thermodynamically by using either pathway. Some significant differences of stabilities between the anomeric forms of canonical and non-canonical nucleotides were observed, mostly for non-canonical Ns and Nts containing TAP-C5, BA-C^5 and MM. Significant similarities were observed for the thermodynamics of the building blocks containing P or As. The synthesis of Ns with AEG and glycerol and Nts with glycerol was predicted as the most favored both in vacuum and in aqueous environment. This is suggestive of a prebiotic synthesis of canonical Ns(t)s assisted by an aegPNA, GNA Ns(t)s scaffold. These results suggest the possibility of a proto-RNA on the thermodynamic basis.

These results suggest that the emergence of a proto-RNA could be justified on the thermodynamic basis. This proto-NAs must had aegPNA or glycerol as TCs and either the canonical A, G, C, T, U or the non-canonical TAP-C^5, BA-C^5 and MM as recognition units. For the case of the GNA building blocks either dihydrogen monophosphate or dihydrogen monoarsenate could have been incorporated as ILs.

Our results have also shown that the classic or alternative synthesis of many canonical or non-canonical Ns(t)s is favored in an anhydrous environment {vacuum}. This supports the idea that prebiotic scenarios such as the surface or water bodies and the air-water interface could have contained relatively significant amounts of these building blocks. These scenarios have been explored by Deal and coworkers [22].

Singh and coworkers [23] have described a H-bond self-activated glycosylation mechanism for the glycosylation of adenine and cytidine with D-ribose that is considered to happen on the surface of water. After 8 days of heating at 60-70 ℃ β-adenosine was obtained in 15% yield while the yield of β-cytidine was obtained 12% (see Scheme 3 in Sing paper). In this study the products were identified by using LC-MS and NMR. The reaction mechanism proposed supports the interaction of the base with the α-anomer of the sugar by the assistance of hydrogen bonds leading

to a more stable transition state that goes through an oxonium intermediate that eventually interconverts into the β-counterpart of the nucleosides.

Chapter 5 [2] addresses the question of why nucleic acids are the repositories of hereditary information using Shannon's classical "*Information Theory*" in conjunction with the concepts of "redundancy" and "value" of information. A lost in information of 0.56 n bits {n is number of Nts}when going from nucleic acids to proteins is estimated assuming ideal non-redundant scenarios prompting to think why the information flow from DNA/RNA $\rightarrow$ proteins. Additionally, a redundancy of $R = 0.28$ is necessary for the accurate transmission of information (minimizing copying errors) from nucleic acids to proteins while keeping the language sufficiently flexible to encode information with relatively high density.

While this thesis has focused primarily on the role of thermodynamics in the prebiotic selection and synthesis of Ns and Nts other factors and scenarios must also be considered in future studies. These are discussed in the next sections.

7.2   Other factors/scenarios to be considered

UV light

Given that the early Earth was showered by UV light before the formation of its stratosphere and protective ozone layer [24], the role of the UV light in the prebiotic origin of life is an important factor to take in consideration [25].

The bases A, G, C, T and U are the first line of defense against UV light in NAs. U differs from T by lacking the C5-methyl group. These bases have π-conjugated aromatic rings that allows them to absorb radiation through π- π^* transitions (pick of absorbance at 260 nm, with an absorbance range of 230-280 nm) with exceptional non-radiative decay that can translate in shorter excited-state lifetime ($\tau = 0.1$ - 2.8) compared to other non-standard nucleobases, making them excellent UV chromophores [26, 27, 28]. Additionally, Brister and coworkers [26] have also explored the photophysical and spectroscopic properties in aqueous solution of two potential ancestral nucleobases: 2, 4, 6-triaminopyrimidine (TAP) and barbituric acid (BA). The authors found that these two nucleobases absorb UV light significantly and have efficient electronic

relaxation mechanisms for dissipating as heat most of the absorbed ultraviolet energy to their aqueous environment in an ultrafast time scale.

UV radiation (mostly UV-B {280-315 nm}) [29] is an exogenous agent that can cause damage in the DNA molecule [30, 13, 31, 32]. Incidentally, this is also a commonly-known pathway leading to mutation (and hence potentially to cancer as well) [33]. From these damages the formation of **C**yclobutene **P**yrimidine **D**imers (CPD) (see left image of **Figure 7.1**) represent the 75% of the total UV induced damage [30].

It has been recently discovered that DNA can self-repair CPD in the absence of DNA damage repair enzymes (see Fig. 6). This mechanism has a relevance to explain how in an RNA world without proteins, ancestral nucleic acids could have been able to survive when exposed to intense UV radiation [34].

Summing-up, could UV light have contributed to the natural selection in prebiotic conditions of the predominant β-anomer and canonical dr, r and HPO$_4^-$ of nucleic acids? *The computational modeling of the mechanisms of formation and self-repair of CPD for dinucleotides containing combinations of canonical and non-canonical TCs, pyrimidine bases, HPO$_4^-$, HAsO$_4^-$ in the β- and α-configurations can offer useful insight on this problem.*

Role of ions

A variety of minerals and ions have been considered in the literature for their importance in the retaining of components of the building blocks of nucleic acids and their catalytic activity to facilitate condensation reactions of the components and polymerization of the building blocks of nucleic acids. This list includes borate (BO$_3^{-3}$) minerals [16, 38], clays, pyrites, silicates, oxides [37] and divalent cations magnesium (Mg^{2+}) [38]. The role of such ions can as well be tackled computationally but the combinatorial complexity rises quickly especially if UV excitations, ionization states (depending on pH) and solvation effects are to be also factored-in.

pK$_a$ and pH of the environment

The effects of pH can be modeled by energy minimizing the RUs in different ionization states.
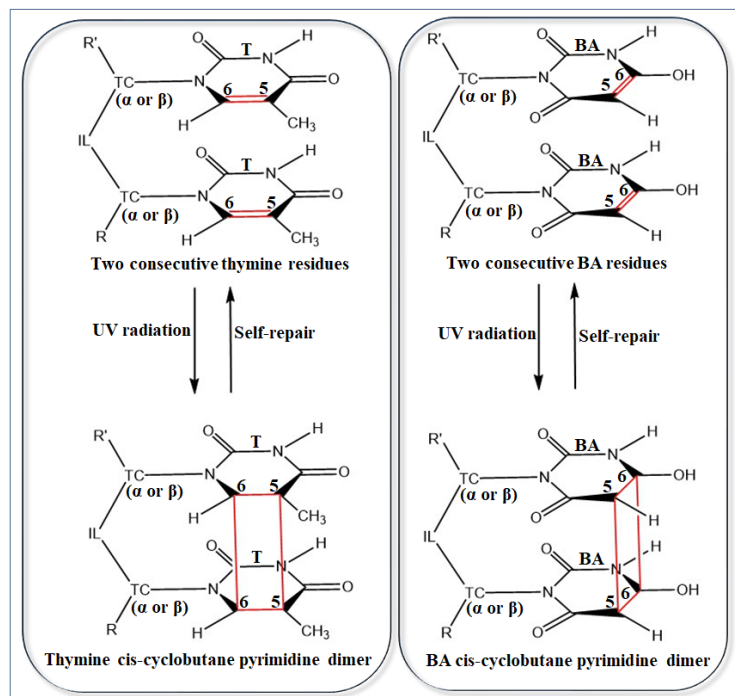
Figure 7.1 (Left): formation and self-repair of a cis - **C**yclobutane **P**yrimidine **D**imer (CPD) between two consecutive **T**hymine (T) nucleobases. (Right): hypothetical model proposed by the Candidate for the formation and self-repair of a cis-cyclobutane pyrimidine dimer between two N-glycosilated **B**arbituric **A**cid (BA) nucleobases in their enol form (predominant tautomeric form for BA in aqueous solution at pH = 7) [56]. R and R' are the rest of two different polynucleotide sequences.

For instance, Thibaudeau and coworkers (see pp. 22-46 in [111] and [112, 113]) has proposed that the preference for one specific β-anomer over the α counterpart for canonical nucleosides is related with its higher flexibility or lower differences between the $\Delta G°$ for the $^2T_3$ $\rightleftharpoons ^3T_2$ pseudorotational equilibrium in the β-configuration at different pK_a.

Previous chapters of this thesis, have referred to the prebiotic plausibility on thermodynamic basis for Ns(t)s with the non-canonical TAP-, BA-C^5 and MM. Additionally, even when CA does not seem to react with most of TCs it has been reported for its capacity to generate hexads with TAP [36, 53]. Fialho [39] refers to the difference in pK_a between the 4 non-canonical bases TAP, BA, CA and MM. Due to the different ionization states of these bases it can be inferred that only the pairs MM:BA and TAP:CA can be found at pH $\approx$ 4-5 and pH $\approx$ 6-7 respectively.

Considering this if the first proto-NAs had these RUs depending on the pH of the environment they may have only contained either BA, MM or TAP, CA.

The long standing "water problem". Alternative solvents as a viable solution

Even when water is the universal solvent and is vital for the existence of life its role in the emergence of nucleic acids constitute one of the worst "nightmares" for prebiotic chemistry ("water problem") [25, 26, 11, 27]. Taking this in consideration alternative solvents have been considered.

Formamide (H_2NCOH) has received special attention in the literature. It can be prebiotically synthesized using different routes and it can also be extensively found present in interstellar space [42]. All nucleobases and other molecules of interest for life can be synthesized from H_2NCOH (see scheme 1 in Saladino's paper). Furthermore the phosphorylation of different nucleosides in the presence of H_2NCOH is possible with different yields going from 6%-59% [31], e.g., **A**denosine 5'-**M**ono**P**hosphate (AMP) can be obtained as main product when heating at 90 °C a mixture of adenosine, formamide-water, KH_2PO_4 and hydroxyapatite [42]. Spontaneous polymerization of **c**yclic 3', 5' **G**uanosine **M**ono**P**hosphate (cGMP) in pure formamide has also been described following similar mechanism to the polymerization in water.[26] Still some limitations related to the use of formamide as solvent like the inability of create complementary base pairing [11] has motivated the search for other alternative solvents.

Polar organic solvents, e.g., ionic liquids and deep eutectic solvents have been considered a viable alternative [8], e.g., choline chloride:urea mixture has shown to assist the folding of many nucleic acids, with some structures been more stable in this solvent than in water [44].

Synthesis and polymerization of Ns(t)s assisted by lipids

In support of a "Lipid World hypothesis" Dr. Deamer group in 2008 [14, 15, 16] have reported for the first time the non-enzymatic formation of oligonucleotides with up to 100 monomers at acidic pH = 2.0 - 6.8 and temperatures ≈ 60 - 90 °C when dehydration-hydration cycles were promoted on a mixture AMP and **U**ridine **M**ono**P**hosphate (UMP) in a lipid matrix or vesicles containing **P**almitoyl-**O**leoyl**P**hosphatidyl**C**holine (POPC), **P**almitoyl-**O**leoylphosph**A**tidic **A**cid (POPA) and **L**yso**P**hosphatidyl**C**holine (LPC). Characterization of the

oligonucleotides by Rajamani and coworkers in 2017 [17] using mass spectrometry {**M**atrix-**A**ssisted **L**aser **D**esorption/**I**onization-**T**ime **O**f **F**light (MALDI-TOF)} showed depurination and depyrimidination of many sites in the oligomer. The authors hypothesized that de-glycosylation reactions happened simultaneously as the polymer has been obtained due to the acidic conditions and high temperatures at which the experiments were carried out.

Even when the results from Rajamani et. al [17] may seem discouraging, the idea that a hole prebiotic chemistry may have happened in compartmentation systems like lipid bilayers opens a door to new alternative scenarios for successful prebiotic synthesis and assembly of NAs. *Could the classic synthesis of canonical or non-canonical Ns(t)s be possible inside lipid bilayers?* This is question that I will leave open for a future investigation.

7.3   References

[1] R. Dawkins, River out of eden : a Darwinian view of life, eBook: Basic Books, 1995.

[2] A. K. Garcia, J. W. Schopf, S. Yokobori, S. Akanuma and A. Yamagishi, "Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean," *Proceedings of the National Academy of Sciences,* vol. 114, pp. 4619-4624, 2017.

[3] D. S. Ross and D. Deamer, "Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis," *Life,* vol. 6, article#28 (pp. 1-12), 2016.

[4] V. B. E. Thomsen, "LeChâtelier's principle in the sciences," *Journal of Chemical Education,* vol. 77, pp. 173-176, 2000.

[5] N. V. Hud, "Searching for lost nucleotides of the pre-RNA world with a self-refining model of early Earth," *Nature Communications,* vol. 9, article #5171 (pp. 1-4), 2018.

[6] A. do Nascimento Vieira, K. do Nascimento Vieira, W. F. Martin and M. Preiner, "The ambivalent role of water at the origins of life," *FEBS Letters,* vol. 594, pp. 2717-2733, 2020.

[7] H.-J. Kim, Y. Furukawa, T. Kakegawa, A. Bita, R. Scorei and S. A. Benner, "Evaporite borate-containing mineral ensembles make phosphate available and regiospecifically phosphorylate ribonucleosides: Borate as a multifaceted problem solver in prebiotic chemistry," *Angewandte Chemie International Edition,* vol. 55, pp. 15816-15820, 2016.

[8] N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, "The origin of RNA and "my grandfather's axe"," *Chemistry & Biology,* vol. 20, pp. 466-474, 2013.

[9] G. F. Joyce and L. E. Orgel, "Prospects for understanding the origin of the RNA world," in *The RNA World*, 2 ed., New York, Cold Spring Harbor Press, 1999, pp. 49-78.

[10] X. Guo, S. Fu, J. Ying and Y. Zhao, "Prebiotic chemistry: a review of nucleoside phosphorylation and polymerization," *Open Biology,* vol. 13, article #220234 (pp. 1-9).

[11] M. Yadav, R. Kumar and R. Krishnamurthy, "Chemistry of abiotic nucleotide synthesis," *Chemical Reviews,* vol. 120, pp. 4766-4805, 2020.

[12] A. Biscans, "Exploring the emergence of RNA nucleosides and nucleotides on the early earth," *Life,* vol. 8, article#57 (pp. 1-13), 2018.

[13] D. Nelson and M. Cox, Lehninger Principles of Biochemistry, New York: WH Freeman and Co, 2005.

[14] B. J. Cafferty, D. M. Fialho and N. V. Hud, "Searching for possible ancestors of RNA: The self-assembly hypothesis for the origin of proto-RNA," in *Prebiotic chemistry and chemical evolution of nucleic Acids, nucleic acids and molecular biology*, vol. 35, C. Menor-Salván, Ed., Philadelphia, Pennsylvania: Springer, 2018, pp. 143-174.

[15] E. Schrödinger, What is life? The physical aspect of the living cell and mind, Cambridge : Cambridge University Press, 1944.

[16] J. F. C. Mayen and J. Błażewicz, "Recent results on computational molecular modeling of the origins of life," *Foundations of Computing and Decision Sciences,* vol. 45, pp. 35-46, 2020.

[17] C. Gatti, G. Macetti, R. J. Boyd and C. F. Matta, "An electron density source-function study of DNA bases pairs in their neutral and ionized ground states," *Journal of Computational Chemistry,* vol. 39, pp. 1112-1128, 2018.

[18] Y. Xu, "Molecular modeling and thermodynamics simulation of nucleic acids. Thesis for Doctoral degree (Ph.D.)," Karolinska Institute, Stockholm, 2016.

[19] J. E. Šponer, P. Banáš, P. Jurečka, M. Zgarbová , P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer and M. Otyepka, "Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome," *The Journal of Physical Chemistry Letters,* vol. 5, pp. 1771-1782, 2014.

[20] A. Y. L. Sim, P. Minary and M. Levitt, "Modeling nucleic acids," *Current Opinion in Structural Biology,* vol. 22, pp. 273-278, 2012.

[21] L. A. Monteserín Castanedo and C. F. Matta, "On the prebiotic selection of nucleotide anomers: a computational study," *Heliyon,* vol. 8, article#e09657 (pp. 1-12), 2022.

[22] A. M. Deal, R. J. Rapf and V. Vaida, "Water-air interfaces as environments to address the water paradox in prebiotic chemistry: a physical chemistry perspective," *The Journal of Physical Chemistry A,* vol. 125, pp. 4929-4942, 2021.

[23] P. Singh, A. Singh, J. Kaur and W. Holzer, "H-bond activated glycosylation of nucleobases: implications for prebiotic nucleoside synthesis," *RSC Advances,* vol. 4, pp. 3158-3161, 2014.

[24] L. A. M. Castanedo, P. Fahimi and C. F. Matta, "What can we say about the "Value of Information" in Biophysics?," *Submitted to Physica Scripta,* 2023.

[25] K. Douglas, DNA nanoscience: from prebiotic origins to emerging nanotechnology, Boca Ratón: CRC Press, 2017.

[26] C. Doglioni, J. Pignatti and M. Coleman, "Why did life develop on the surface of the Earth in the Cambrian?," *Geoscience Frontiers,* vol. 7, pp. 865-873, 2016.

[27] M. M. Brister, M. Pollum and C. E. Crespo-Hernández, "Photochemical etiology of promising ancestors of the RNA nucleobases," *Photochemical etiology of promising,* vol. 18, pp. 20097-20103, 2016.

[28] A. A. Beckstead, Y. Zhang, M. S. Vries de and B. Kohler, "Life in the light: nucleic acid photoproperties as a legacy of chemical evolution," *Physical Chemistry Chemical Physics,* vol. 18, pp. 24228-24238, 2016.

[29] A. C. Rios and Y. Tor, "On the origin of the canonical nucleobases: an assessment of selection pressures across chemical and early biological evolution," *Israel Journal of Chemistry,* vol. 53, pp. 469-483, 2013.

[30] R. P. Rastogi , R. A. Kumar, M. B. Tyagi and R. P. Sinha, "Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair," *Journal of Nucleic Acids,* article #592980 (pp. 1-32), 2010.

[31] A. P. Schuch and C. F. M. Menck, "The genotoxic effects of DNA lesions induced by artificial UV-radiation and sunlight," *Journal of Photochemistry and Photobiology B: Biology,* vol. 99, pp. 111-116, 2010.

[32] J. -L. Ravanat, T. Douki and J. Cadet , "UV damage to nucleic acid components," in *Sun Protection in Man*, Amsterdam, Elsevier Science B. V, 2001, pp. 207-230.

[33] G. P. Pfeifer, "Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment," *Photochemistry and Photobiology,* vol. 65, pp. 270-283, 1997.

[34] S. Bhattacharya, "Natural antimutagens: a review," *Research Journal of Medicine,* vol. 5, pp. 116-126, 2011.

[35] M. R. Holman, T. Ito and S. E. Rokita, "Self-repair of thymine dimer in duplex DNA," *Journal of the American Chemical Society,* vol. 129, pp. 6-7, 2007.

[36] C. V. Mungi, S. K. Singh, J. Chugh and S. Rajamani, " Synthesis of barbituric acid containing nucleotide and its implications for the origins of primitive informational polymer," *Physical Chemistry Chemical Physics,* vol. 18, pp. 20091-20096, 2016.

[37] M. Neveu, H. J. Kim and S. A. Benner, "The ''strong'' RNA world hypothesis: fifty years old," *Astrobiology,* vol. 13, pp. 391-403, 2013.

[38] A. Ricardo, M. A. Carrigan, A. N. Olcott and A. Benner, "Borate minerals stabilize ribose," *Science,* vol. 303, p. 196, 2004.

[39] S. A. Benner, H. J. Kim and E. Biondi, "Mineral-organic interactions in prebiotic synthesis: the discontinuous synthesis model for the formation of RNA in naturally complex geological environments," in *Prebiotic Chemistry and Chemical Evolution of Nucleic Acids*, Electronic version, Springer, 2018, pp. 31-83.

[40] N. G. Holm, "The significance of Mg in prebiotic geochemistry," *Geobiology,* vol. 10, pp. 269-279, 2012.

[41] C. Thibaudeau, P. Acharya and J. Chattopadhyaya, Stereoelectronic effects in nucleosides and nucleotides and their structural implications, Uppsala: Uppsala University Press, 2005.

[42] C. Thibaudeau, A. Foldesi and J. Chattopadhyaya, "The first experimental evidence for a larger medium-dependent flexibility of natural β-D-nucleosides compared to the α-D-nucleosides," *Tetrahedron,* vol. 53, pp. 14043-14072, 1997.

[43] C. Thibaudeau, J. Plavec and J. Chattopadhyaya, "Quantitation of the anomeric effect in adenosine and guanosine by comparison of the thermodynamics of the pseudorotational equilibrium of the pentofuranose moiety in N- and C-nucleosides," *Journal of the American Chemical Society,* vol. 116, pp. 8033-8031 , 1994.

[44] B. J. Cafferty and N. V. Hud, "Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology," *Current Opinion in Chemical Biology,* vol. 22, pp. 146-157, 2014.

[45] B. J. Cafferty, I. Gállego, M. C. Chen, K. I. Farley, R. Eritja and N. V. Hud, "Efficient self-assembly in water of Long noncovalent polymers by nucleobase analogues," *Journal of the American Chemical Society,* vol. 135, pp. 2447-2450, 2013.

[46] D. M. Fialho, *Physical organic principles governing the spontaneous prebiotic emergence of proto-nucleic acids,* Georgia: Georgia Tech Library, 2019, pp. 1-3.

[47] R. Saladino, C. Crestini, S. Pino , G. Constanzo and E. Di Mauro, "Formamide and the origin of life," *Physics of Life Reviews,* vol. 9, pp. 84-104, 2012.

[48] M. Gull, "Prebiotic phosphorylation reactions on the early earth," *Challenges,* vol. 5, pp. 193-212, 2014.

[49] I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, "DNA and RNA in anhydrous media: duplex, triplex, and G-quadruplex secondary structures in a deep eutectic solvent," *Angewandte Chemie International Edition,* vol. 49, pp. 6310-6314, 2010.

[50] F. Olasagasti and S. Rajamani, "Lipid-assisted polymerization of nucleotides," *Life,* vol. 9, pp. article#83 (pp. 1-10), 2019.

[51] V. DeGuzman, W. Vercoutere, H. Shenasa and D. Deamer, "Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization," *Journal of Molecular Evolution,* vol. 78, pp. 251-262, 2014.

[52] S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti and D. Deamer , "Lipid-assisted synthesis of RNA-like plymers from mononucleotides," *Origins of Life and Evolution of Biospheres,* vol. 38, pp. 57-74, 2008.

[53] C. V. Mungi and S. Rajamani, "Characterization of RNA-like oligomers from lipid-assisted nonenzymatic synthesis: implications for origin of informational molecules on early earth," *Life,* vol. 5, pp. 65-84, 2015.