A Computational Investigation of the Intrinsic Electric Field of ATP Synthase

By
Youji Cheng, B.Sc. (Hons.)

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Applied Science

December 2019, Halifax, Nova Scotia, Canada

Approved: Dr. Chérif F. Matta – Supervisor
Professor, Dept. of Chemistry & Physics
Mount Saint Vincent University, Halifax, Nova Scotia, Canada
&
Adjunct Professor, Dept. of Chemistry
Saint Mary's University, Halifax, Nova Scotia, Canada

Approved: Dr. Cory Pye – Committee Member
Associate Professor, Dept. of Chemistry
Saint Mary's University, Halifax, Nova Scotia, Canada

Approved: Dr. Paul Muir – Committee Member
Professor, Dept. of Mathematics & Computing Science
Saint Mary's University, Halifax, Nova Scotia, Canada

Approved: Dr. Katherine Darvesh – External Examiner
Professor, Dept. of Chemistry & Physics
Mount Saint Vincent University, Halifax, Nova Scotia, Canada

Date: December 23, 2019

Acknowledgements

I would like to thank the support from my family. When I was six years old, my mother, a chemistry teacher, demonstrated me an ammonia fountain experiment. My father, who devoted his entire career to marine engineering, would take me to his shipyard, and I explored industrial cranes. My parents opened up a new world for me, and I have been curious ever since. I am fortunate to be the first person in my family to pursue graduate school, because the political movement in China kept my parents away from further education back in the days. I thank them for their financial and emotional support in the past two years. In addition, I would like to thank Peter and Laura Mack Ledaire, who have really become my family since I moved to Halifax.

There are many mentors and academic collaborators I need to thank. I first became interested in computational chemistry due to an initial introduction by Professor Khashayar Ghandi. I also own the most gratitude to Professor Chérif F. Matta for his supervision, mentorship, and intellectual discussion - he generated the brilliant research ideas upon which this thesis is based. I want to thank Professor Cory Pye and Professor Paul Muir for advising me as committee members. Professor Porter Scobey offered valuable training on Python programming and scientific computing. Mr. William Fiset assisted with strategizing the algorithm of the Python script for the PrES program. Mr. Connor Tannahill assisted me with optimizing the Python script. Dr. Ross Dickson helped me debug PrES. I enjoyed technical discussion with Professor Lou Massa, with whom I co-authored a journal paper, and he inspired me with new ideas and future directions.

I would like to acknowledge support from fellow research group members Mr. Lázaro A. Monteserín Castanedo and Mr. Cyrus A. Toussi. The final visualization of results was achieved with their assistance and collaboration.

Additional financial support and computing service were provided by Saint Mary's University, National Sciences and Engineering Research Council of Canada (NSERC) through my supervisor's Discovery Grant, Mount Saint Vincent University, Compute Canada, and Advanced Research Computing in Atlantic Canada (ACENET).

Youji Cheng

November 2019

To my parents:

Xintai Cheng
&
Yan Qiu

for their understanding and support

Table of Content

Chapter 5. Conclusion

List of Figures

Page

List of Tables

Page

List of Equations

A Computational Investigation of the Intrinsic Electric Field of ATP Synthase

Youji Cheng

Abstract

Mitchell's chemiosmotic theory stipulates that an electrochemical gradient of protons across the inner mitochondrial membrane provides the free energy for the biosynthesis of adenosine 5'-triphosphate (ATP). This proton gradient generates an electric field of ~ 10^7 $V \cdot m^{-1}$. It is hypothesized that ATP synthase, the enzyme catalyzing ATP synthesis, itself generates a field/potential comparable to that of the proton gradient. A programme, PrES, was developed to compute the intrinsic field/potential of ATP synthase (or that of any protein). The calculated field/potential by PrES is based on AMBER charges and is found to be of sufficient accuracy within the biological context. The intrinsic field of ATP synthase is found to be of strengths 10^6 - 10^8 $V \cdot m^{-1}$ around the protein and is associated with a voltage of ~ 70 mV between its extremities. This potential difference adds a ΔG of around -7 kJ (per mol of protons), a significant correction as hypothesized.

December, 2019

Chapter 1. An Introduction to Mitochondrial ATP Synthase

"Is there some special magic about life, essential to making molecular machinery work?"

- K. Eric Drexler, *Engines of Creation,* 1986

Adenosine 5'-tri-phosphate (ATP) is an important molecule in all forms of life, as it is the energy currency in all cells. For adult human beings, one can produce around 10^{21} molecules of ATP every second.[1] ATP synthase is a protein enzyme located in mitochondria – a critical organelle for all cells, and it facilitates the synthesis of ATP as the enzyme functions as a "molecular machine". Peter Mitchell demonstrated that ATP synthesis is driven by the chemical and electrical gradient generated by H^+ along the inner-mitochondrial membrane.[2-5] Dysfunction of ATP synthase plays a role in a wide range of degenerative diseases such as prostate cancer and Leigh syndrome.[6, 7] Thus, the function of ATP synthase is a crucial topic in biological and medical research. Based on available literature to date, this chapter introduces factors that drive ATP synthesis, followed by some preliminary ideas complementing current knowledge. In addition, metabolic stress induced by these driving forces will be discussed.

1.1 The Oxidative Phosphorylation

The mitochondrion is the energy powerhouse in most cells. It exhibits a "double-bag" structure with two layers of membranes (Figure 1.1). Many fundamental metabolic processes take place on the inner-membrane of mitochondrion. In particular, the electron transport chain (ETC) and the ATP synthesis together make up a metabolic pathway known as oxidative phosphorylation. The ETC utilizes respiratory oxygen (O_2) to store

energy to the phosphate group in ATP, yielding metabolic water (Figure 1.2). As a part of oxidative phosphorylation, ATP synthase is embedded in the inner-membrane of mitochondrion.

In addition to ATP synthase, the ETC depends on a series of proteins embedded in the inner-mitochondrial membrane: protein complex I, II, III, and IV (Figure 1.2). Energy carrier (reduced) co-enzyme molecules, nicotinamide adenine dinucleotide (NADH) and flavine adenine dinucleotide ($FADH_2$), are produced in another metabolic pathway known as the tricarboxylic acid cycle. NADH and $FADH_2$ carry energy in the form of hydrogen equivalents from metabolic processes to the ETC. The ETC is a chain of coupled redox reactions, all occurring within the inner mitochondrial membrane. Eventually, electrons are transported from NADH and $FADH_2$ to form molecular oxygen (O_2) with increasing standard reduction potential along protein complex I to IV. O_2 possesses the highest standard reduction potential ($E°'$) in the series, and O_2 accepts the electrons and the associated protons to form metabolic water (Figure 1.2).

Thus, electrons are channelled following a specific order of increasing standard reduction potential (Figure 1.2). This is tantamount to an order of increasingly exergonic redox reactions, since the released standard free energy $\Delta G°' = -nFE°'$, where F is Faraday's constant. Thus, the electrons are passed from NADH and $FADH_2$ to the protein complexes I to IV in the ETC (Figure 1.2), and the released $\Delta G°'$ contributes to the proton (H^+) translocation from the mitochondrial matrix to the inter-membrane space (Figure 1.2).[8] For each electron passed along, three protons are translocated, and the coupled energy on the ETC maintains a high concentration gradient of H^+ across the inner-membrane as a result.[2]

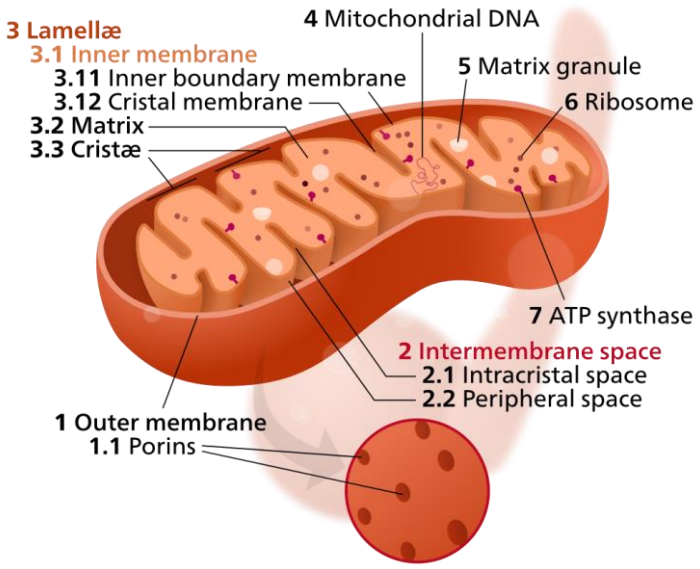


Figure 1.1 Basic structure of a mitochondrion.
(Wikimedia Foundation – CC BY 4.0)

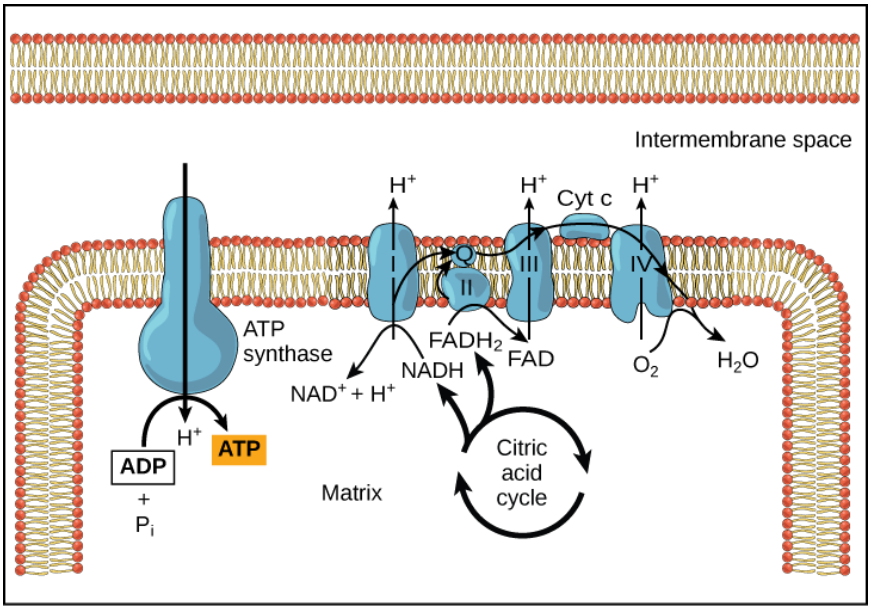


Figure 1.2 Three protons are translocated from the matrix to intermembrane space for each electron transport from Complex I to Complex IV during Oxidative Phosphorylation. Metabolic water forms at the Complex IV. (OpenStax CNX, 2019 – CC BY 4.0, ref 8)

According to Mitchell's chemiosmotic theory,[2-5] the total free energy that drives the proton translocation is given by Equation 1.1:

$$\Delta G = RT \ln ([H^+]_{out} - [H^+]_{in}) + zF\Delta\psi \qquad (1.1)$$

where R is the universal gas constant, T is temperature, $[H^+]_{out}$ is the concentration of protons in the inter-membrane space, $[H^+]_{in}$ is the concentration of protons in the mitochondrial matrix, z is the charge number of a cation X^{z+} ($z = 1$ for H^+), F is the Faraday constant, and $\Delta\psi$ is the electric potential difference inside the mitochondrial matrix relative to the inter-membrane space. The term "$RT \ln ([H^+]_{out} - [H^+]_{in})$" accounts for the energy contribution from proton concentration (chemical gradient) to the total free energy, and the term "$zF\Delta\psi$" accounts for the energy contribution of the electrostatic potential difference generated from the H^+ gradient to the total free energy.

The protons enter the mitochondrial matrix via a channel within the ATP synthase. Driven by the proton motive force generated by H^+, ATP synthase catalyzes the formation of adenosine 5'-triphosphate (ATP), where adenosine 5'-diphosphate (ADP) binds to an inorganic phosphate group (Pi). The free energy released by proton translocation drives an endergonic reaction given by Equation 1.2: [9]

$$ADP + Pi \rightarrow ATP, \qquad \Delta G^{\circ\prime} = +30.9 \text{ kJ.mol}^{-1} \qquad (1.2)$$

Thus, ATP eventually captures the energy from all nutrients we eat. At physiological pH, ATP stores energy in two terminal pyrophosphate bonds (Figure 1.4). In particular, the energy is stored as electrostatic repulsion between the negatively charged phosphate groups. These charges are partly neutralized by complexation with Mg^{2+} ions, and the ATP molecule is trapped kinetically into stability even if it is thermodynamically unstable. The hydrolysis of ATP to ADP releases back the +30.9

kJ.mol^{-1} necessary for its synthesis at standard biochemical conditions. Mitochondrial ATP synthase catalyzes the above reaction that synthesizes ATP. Once formed and exported out of the mitochondrion, ATP is then used to drive endergonic biochemical reactions, such as the reactions leading to the muscle contraction via myosin protein, or the active pumping of Na^+ and K^+ ions in nerve cells against concentration gradients during nerve conduction. This is why ATP is generally considered as the energy currency of all living cells.

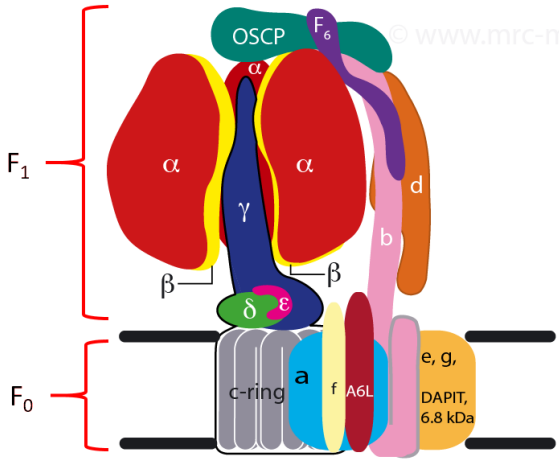


Figure 1.3 Subunit composition of mitochondrial ATP synthase.
(Walker, 2019 – reproduced with permission, ref 10)

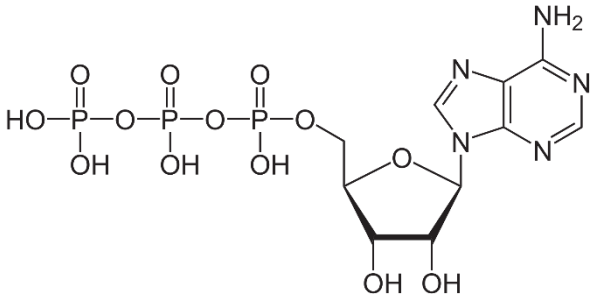


Figure 1.4 Chemical structure of adenosine 5'-tri-phosphate (ATP).

1.2 ATP Synthesis

Structurally, the ATP synthase is a protein consisting a F_o and a F_1 unit. The F_o unit is embedded in the inner-membrane of mitochondrion, acting as an ion channel for proton (H^+) translocation. Similar to a motor, the F_o unit rotates as protons pass through it. The F_o unit is linked to the F_1 unit via a rotor made of the γ, δ, and ε subunit (Figure 1.3),[10] and the rotation of the F_o drives the conformational changes on the F_1 to synthesize ATP molecules (Figure 1.5).[11] First, a loose conformation (L) binds with a phosphate group (Pi) and an ADP molecule. Then, the β subunit turns to a tight (T) conformation. Next, an inorganic phosphate unit (Pi) binds with ADP to form an ATP molecule. As the conformation changes, an ATP from the previous synthetic cycle is released as the T conformation turns to an O conformation. [12]

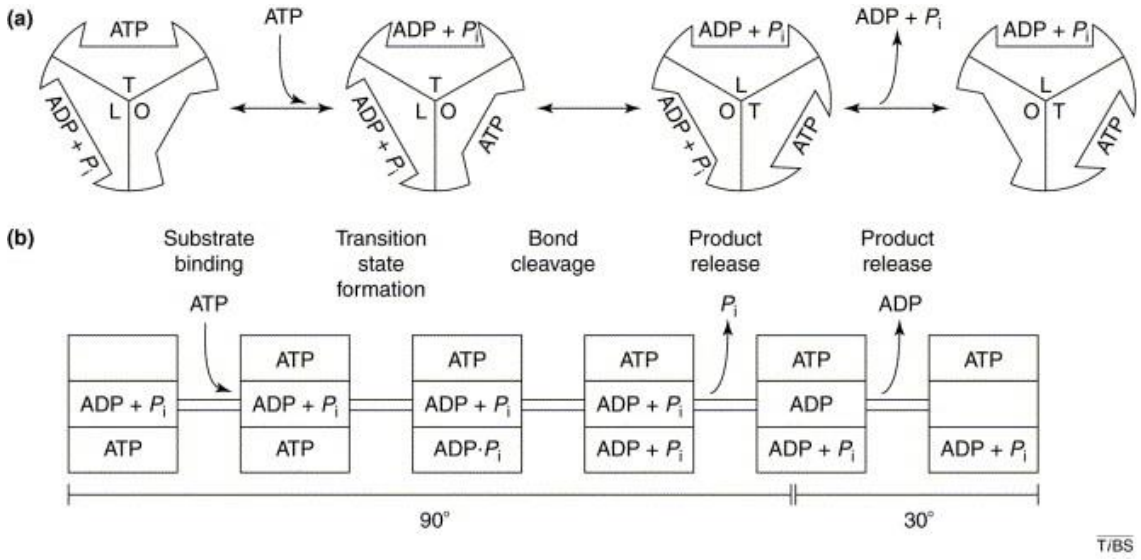


Figure 1.5 The conformational changes of the β subunit in the F_1 unit of ATP synthase. (Capaldi and Aggeler, 2002 – reproduced with permission, ref 12)

1.3 Chemical Gradient

The F_o unit of the ATP synthase includes subunits a, b, and c. In particular, 10 to 15 c subunits (or c rings), depending on the species, form the F_o unit. The number of H^+ translocated per rotation is directly proportional to the number of c subunits in the ring, meaning that the averaged rotation speed differs among different species.[13] Although it is still unclear why there is a diverse c ring stoichiometry among various species, the c ring stoichiometry has important implications for the rotation of F_o.

As shown in Figure 1.2, the subunit a and the subunit c are embedded in the inner-membrane, and they are linked to subunit b located in the mitochondrial matrix. A strong chemical gradient is generated by the concentration of H^+ ions in the inter-membrane space (Figure 1.6). An early study suggests that the protonation and deprotonation on c rings "paddles" the F_o rotation.[14] An H^+ ion enters c rings of the F_o by protonating an arginine (Arg 210) on the c rings, and it leaves the rotating structure via a deprotonation process on an aspartic acid (Asp 61). As such, the chemical gradient is a driving force for the rotation of F_o unit.

The rotation of F_o is regulated and reversible. As discussed in section 1.2 of this chapter, the torque generated by F_o twisting is transferred to the rotor (the γ, δ, and ε subunit, Figure 1.3), resulting in the conformational changes of F_1. Partially embedded in the F_1, a reversely rotating γ subunit drives the enzymatic reaction towards ATP synthesis, which lowers the binding affinity of F_1 to ATP molecules in order to release ATP.[15] The rotatory function of the γ subunit mainly depends on an alpha-helix structure on the γ subunit, because the alpha helix accounts for around 80% of the γ subunit's mass.[16]

1.4 Electrical and Proton Concentration Gradients

Mitchell's chemiosmosis theory rests on the indirect coupling between the energetically reduced coenzymes and the ATP synthesis. This coupling is achieved via the proton gradient across the inner-membrane. As mentioned earlier, the electrostatic potential of the H^+ ions ($zF\Delta\psi$) in the inter-membrane space is a principal driving force for proton translocation. The electrostatic potential is built up by the unbalanced charges from H^+ across the membrane (Figure 1.6).[5, 9]

Charged residues on the subunit a (Figure 1.3) make up two H^+ channels, each H^+ channel stretches half of the membrane that drives the rotation of c subunits.[17] An electric field is associated with the electrostatic potential difference across the membrane, by which protons are translocated along the electrical gradient (Figure 1.6). The H^+ translocation process is exergonic. A minimum electrostatic potential of approximately 169 mV is needed for oxidative phosphorylation to occur,[18] and this trans-membrane potential can reach a maximum of around 180 mV.[19]

The rotation of the c subunits is driven by the torque that primarily comes from coupling of the electric field to the protonated (Arg 210) and deprotonated (Asp 61) sites.[14] In summary, components of the electric field across the inner membrane plays a critical role in oxidative phosphorylation and overall metabolism in biological systems.

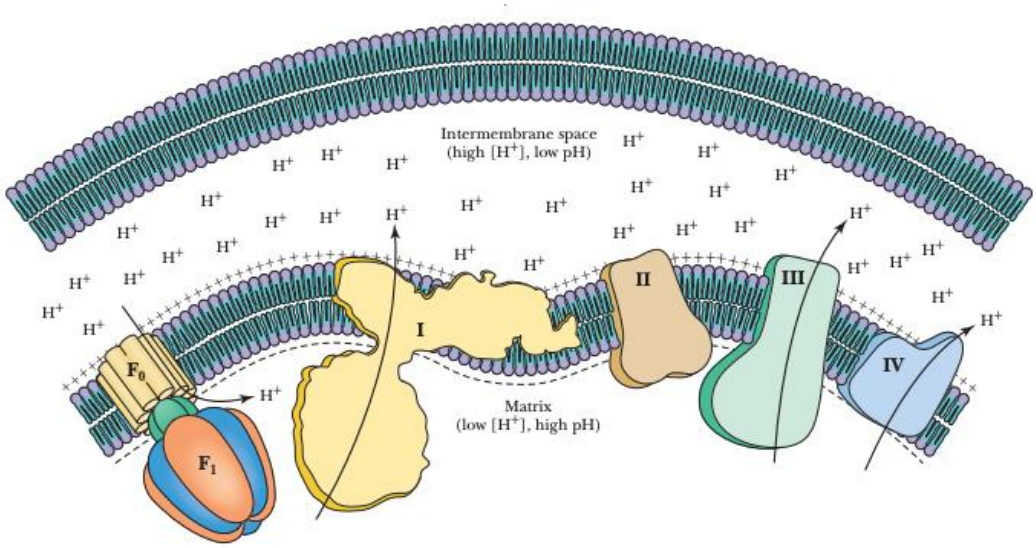


Figure 1.6 The chemical and electric gradient of proton (H^+) across the inner mitochondrial membrane. (Garrett and Grisham, 2010 – reproduced with permission, ref 9)

1.5 ATP Synthase as a Maxwell's Demon

As an H^+ ion approaches the ion channel, ATP synthase has to recognize it as either a "proton" or "non-proton" before it can select H^+ to enter its ion channel. The situation is similar to Maxwell's demon, an intelligent creature that Maxwell introduced in his book on thermodynamics,[20] in which he describes a situation that appears to violate the 2nd law of thermodynamics. Maxwell's paradox describes an intelligent being that can "select" and sort hot (fast) and cold (slow) molecules, separated by a partition. Such an intelligent being can create free energy for free, essentially creating a perpetual machine. Maxwell knew that this was impossible, however, the resolution came 65 years later in Leo Szilard's work and was further elaborated by Landauer and others.[21] The resolution of the Maxwell paradox relies on the fact that sorting molecules must involve first an act of observation, which requires the communication through an interrogation signal with the molecules prior

to sorting them. The recoil of the photons returning to the demon will soon raise its temperature sufficiently for it to melt and stop working. Thus, there is no paradox and no violation of the 2nd law of thermodynamics. What needs to happen is that after every sorting act the demon must dissipate a minimum energy equal to or greater than $k_BT\ln 2$. This has been referred in the literature as "the cost of sorting". It is simply unavoidable and is independent of the mechanism.

ATP synthase operates as a Maxwell's demon, because the enzyme's primary function is to synthesize ATP, and it picks protons out of a noisy background to pass them to the other side of a membrane just like a Maxwell's demon. The enzyme must dissipate at least $k_BT\ln 2$ for every proton it picks. In the calculation of thermodynamic efficiency of ATP synthase, if one includes this unavoidable energetic cost to sort H^+ in addition to the necessary work of 30.9 kJ/mol (Equation 1.2), it brings up the thermodynamic efficiency of ATP synthase from around 60% to nearly 90%.[22, 23]

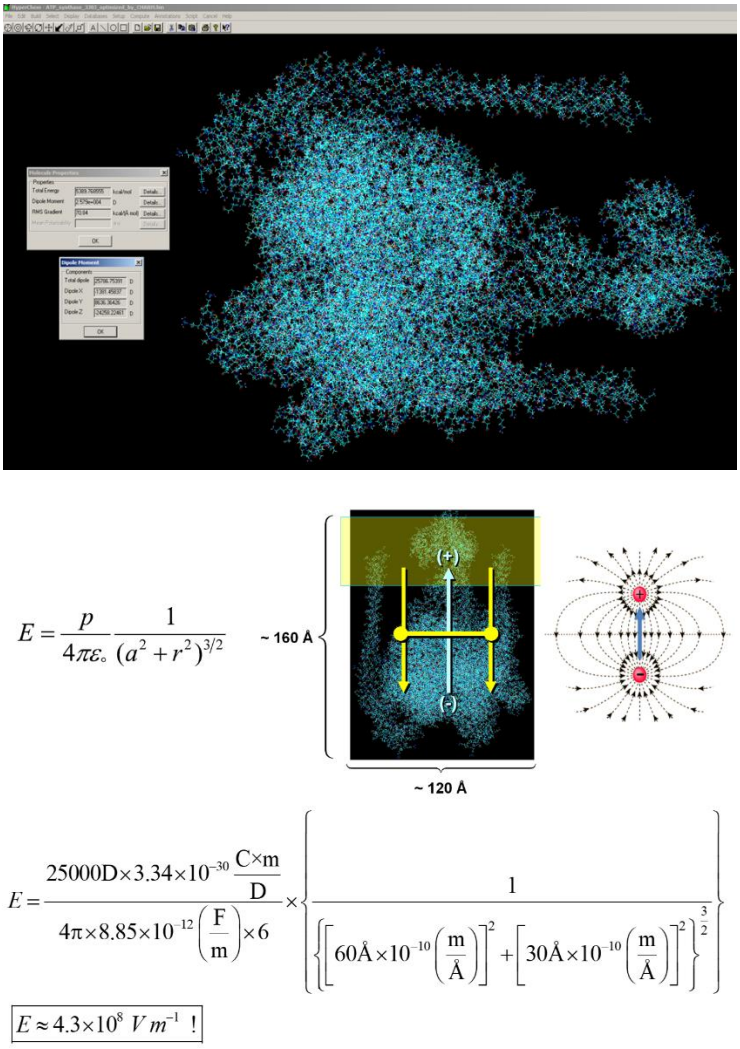
The sorting of protons is primarily driven by Coulombic forces.[22, 23] Therefore, understanding the electrical properties of ATP synthase, a Maxwell's demon, will advance our knowledge into the model of operation of this molecular machine and its thermodynamic efficiency.

1.6 Intrinsic Electric Fields of Proteins

An early study has explored the intrinsic electric field of immobilized protein. Zabusky and Deem employed papain and bovine serum albumin (BSA) protein layer (thickness: 15-100μm) to study the relationship between the intrinsic electric field of BSA protein and proton diffusion in BSA.[24] They found that the intrinsic protein electric

field speeds up the proton diffusion. It should be noted that the experiment was carried out in immobilized protein.

The inner mitochondrial membrane (thickness: ~5 nm) [25] is a fluid environment, and it is much thinner than the BSA protein layer (5 nm << 15-100μm). Nonetheless, these early results imply that the intrinsic protein electric field may be critical to the H^+ mobility near giant protein molecules. Unpublished results by Matta indicate that ATP synthase has a dipole moment of approximately 25,000 debyes, which can give rise a dipolar field with strength of ~10^8 $V \cdot m^{-1}$ in some regions surrounding ATP synthase (Figure 1.7). This was achieved by subjecting the X-ray crystallographic structure of the protein to a single point molecular mechanics (MM) calculation. The MM calculation was carried out by the HyperChem software to only calculate its dipole moment (and not the entire electric field). If one assumes that the poles of the dipole lie on the principal molecular axis with a separation of 120 Å, and given that the positive pole is near the F_O unit and the negative pole near the F_1 unit, then a dipolar field of 2.579×10^4 debye generates an electric field of approximately 4.3×10^4 $V \cdot m^{-1}$ at a distance of 30 Å from the midpoint of the dipole, as indicated graphically in Figure 1.7 (using the expression for the field of an electric dipole and a dielectric constant of 6 for the protein interior). These preliminary results, calling for refinement, come to a central point in this thesis.



$$E = \frac{p}{4\pi\varepsilon_\circ}\frac{1}{(a^2+r^2)^{3/2}}$$

$$E = \frac{25000\text{D}\times 3.34\times 10^{-30}\,\frac{\text{C}\times\text{m}}{\text{D}}}{4\pi\times 8.85\times 10^{-12}\left(\frac{\text{F}}{\text{m}}\right)\times 6}\times\left\{\frac{1}{\left\{\left[60\text{Å}\times 10^{-10}\left(\frac{\text{m}}{\text{Å}}\right)\right]^2+\left[30\text{Å}\times 10^{-10}\left(\frac{\text{m}}{\text{Å}}\right)\right]^2\right\}^{\frac{3}{2}}}\right\}$$

$$\boxed{E\approx 4.3\times 10^{8}\ V\,m^{-1}\ !}$$

Figure 1.7 (*Top*) Screen shot of the HyperChem results for the molecular dipole moment of ATP Synthase generated from force-field charges. **(*Bottom*)** Assuming that the separation of the two poles is 100 Å, the position of evaluation of the dipolar field (yellow arrows) are 60Å on either side of the principal axis, the field is of ~ 10^8 V·m^{-1}. The faint yellow box near the top is the approximate location of the inner mitochondrion membrane in relation to the ATP Synthase molecule. (Matta, 2019 – reproduced with permission, ref 34).

As well, Boxer and co-workers have explored the external electric field effect on biochemical reaction kinetics including reaction rate constant,[26] electron transfer kinetics,[27, 28] and its application on infrared spectroscopy and vibrational stark effect.[29] Further experiments have studied how external electric field affect the dynamics,[30]

structure, and redox potential[31] of cytochrome c protein. External electric field can induce re-arrangement of hydrogen bond networks and slow down the redox process in cytochrome c protein. In the inner mitochondrial membrane, cytochrome c facilitates the electron transfer process between complex III and complex IV (Figure 1.2). As ATP synthase is embedded in the inner membrane along with cytochrome c, the electric field originating from ATP synthase itself can affect the functions of cytochrome c. Moreover, free radical formation is affected when external electric field of 10^6 $V \cdot m^{-1}$ or more is applied to biochemical reactions. Free radicals are atom, molecules, or ions that possess unpaired electrons.

Free radicals are damaging to biological systems, causing programmed cell death (apoptosis) and several degenerative diseases. A strong intrinsic electric field of ATP synthase may partly explain rather damaging metabolic processes in the inner mitochondrial membrane. Despite its importance, limited literature has reported the effects of the intrinsic electric fields of proteins on biochemical reactions. Current effort and challenge will be discussed in Chapter 2 of this thesis.

Nussbaum and Grodzinsky formulated a theoretical model for H^+ transport through the proton channel in ATP synthase. Driven by electromechanical force, protonation and deprotonation occur at fixed charged groups in immobilized protein.[32] As H^+ passes through ATP synthase via certain carboxylic acid residues (arginine and aspartic acid), the intrinsic electric field of ATP synthase may be a missing factor in this protonation-deprotonation mechanism. If the electric field aligns with the direction of H^+ translocation, it assists the translocation; if the field is against the translocation direction,

it will hinder the H^+ translocation. This view will be diagrammed in detail in Chapter 3 (Figure 3.1).

The hydrated cytochrome c protein exhibits an electric field greater than 7 x 10^7 $V \cdot m^{-1}$ on the protein surface.[33] Recalling section 1.4 of this chapter, the H^+ gradient possesses an electrostatic potential of at least 169mV across the inner mitochondrial membrane, and the thickness of mitochondrial membrane is around 5nm. We can calculate the electric field generated by the H^+ gradient across the membrane:

$$membrane\ thickness\ =\ \ 5\ nm\ \ = 5\ \times 10^{-9}\ m$$

$$electrostatic\ potential\ (H^+\ gradient) = 169\ mV = 1.69\ \times\ 10^{-1}\ V$$

$$electric\ field\ (H^+\ gradient)\ \approx \frac{electrostatic\ potential\ (H^+\ gradient)}{membrane\ thickness}$$

$$= \frac{1.69\ \times 10^{-1}\ V}{5\ \times\ 10^{-9}\ m}$$

$$= 3.38\ \times\ 10^{7}\ V \cdot m^{-1}$$

As such, the electric field generated by the H^+ gradient across the membrane is of an order of 10^7 $V \cdot m^{-1}$, which is comparable in order of magnitude to the electric field of hydrated cytochrome c protein.

ATP synthase is embedded in a highly mobile and aqueous environment with surrounding lipid bilayer. Since the size of ATP synthase is much larger than the cytochrome c protein, it can be expected that ATP synthase protein will exhibit an intrinsic electric field $\geq 10^7$ $V \cdot m^{-1}$. Preliminary calculations performed by Matta indicated that the ATP synthase exhibits a dipolar electric field of $10^7 - 10^8$ $V \cdot m^{-1}$.[34] This project is designed to test this hypothesis.

If the hypothesis is validated, then the electric field generated by the protein itself adds vectorially to the field generated by the H^+ gradient (3.38 x 10^7 $V \cdot m^{-1}$) across the

inner mitochondrial membrane. This is significant for two reasons. First, the intrinsic electric field is an electrical property of protein enzyme itself, and the ATP synthase enzyme not only catalyzes the reaction but may also contribute to the standard free energy ($\Delta G°'$) of the reaction. No previous literature has reported this phenomenon, and this will open a door for studies on the intrinsic protein electric field. Second, the conventional chemiosmotic theory only considers the free energy contributed by the H^+, whereas the new evidence considers additional free energy contributed by ATP synthase. To summarize, the free energy for the ATP synthesis reaction can be expressed as: [34]

$$\Delta G = 2.3nRT\Delta\text{pH} + nF\Delta\Psi_{H^+} \quad (\pm nF\Delta\Psi_{ATP\ Synthase}) \qquad (1.3)$$

In addition, recent studies reveal that protein enzymes utilize highly specific electrostatic field to assist enzymatic catalysis. For example, dipole moments from residues on histone deacetylase 8 (HDAC8, EC.3.5.1.98) may affect the catalytic power at the active site of the enzyme.[35] As well, the intrinsic electrostatics of monoamine oxidase A (MAO-A, EC.1.4.3.4) seems to be the driving force for its catalytic function.[36] In order to lower the computational expenses, the researchers had to apply computational manipulation to MAO-A, such as excluding point charges and cutting the enzyme off by distance. Hybrid quantum mechanics/molecular mechanics (QM/MM) is a common method to explore electrostatic property of enzymes active sites,[37] but it is not an ideal tool for our purpose, since we seek to calculate the overall electrostatics of the protein, rather than a refined answer for the active site. The research question is: what is the magnitude and direction of the electric field around a gigantic biomolecule made of approximately 64,000 atoms? The only option is to rely on charges extracted from a

molecular mechanics (MM) force field that has been parametrized for proteins, and this approach will be further introduced in Chapter 2.

1.7 "Short Circuit" through Mitochondrial Inner-Membrane

The membrane structure is fundamental in maintaining the chemical and electrical gradient across the inner mitochondrial membrane. Mitochondria can be subject to electroporation effect, where an electric field directly enhances the permeability of the inner mitochondrial membrane.[38] The enhanced permeability disrupts the built-up chemical and electrical gradient, leading to proton leaks and a "short circuit" of the electron transport chain known as "mitochondrial permeability transition pore". The disruption results in dysfunction of ATP synthase and cell apoptosis, and it is considered as a possible cause of various degenerative diseases as mentioned in the beginning of this chapter.[1]

Since an strong electric field can enhance mitochondrial permeability, would the intrinsic electric field of ATP synthase be significant enough to play a role in mitochondrial permeability disruption? Some propose that the transition pore forms at c subunits of ATP synthase. A high Ca^{2+} concentration in the mitochondrial matrix enlarge the size of c subunits. The enlarged site forms a voltage-sensitive channel that depolarizes the inner membrane.[39, 40] Others argue that these enlarged sites are smaller channels, they have much lower conductance (Figure 1.8), and the pores locate elsewhere on the inner-membrane.[41] In either case, the intrinsic electric field of ATP Synthase plays a role in chain of reactions along the inner mitochondrial membrane.

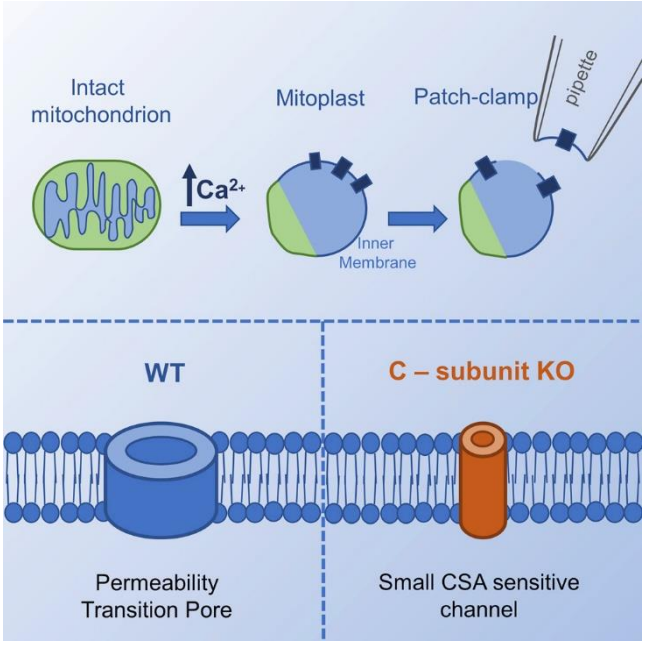


Figure 1.8 Channel at c-subunit (bottom right) has much smaller conductance, because it is much smaller compared to the permeability transition pore (bottom left).
(Neginskaya et al, 2019 – CC BY 4.0, ref 41)

1.8 Summary

ATP production is important to biological metabolism. The synthesis of ATP relies on the rotation and conformational changes of a complex molecular motor known as ATP synthase. The intrinsic electric field of ATP synthase may play a fundamental role in metabolism, contributing a missing factor in the Mitchell's chemiosmotic theory, which can disrupt the inner mitochondrial membrane in conditions such as degenerative diseases. A computational study of intrinsic electrical properties of protein enzymes is hence indicated. If the field/potential is within an order of magnitude of the natural chemiosmotic field/potential, this would constitute a significant finding previously overlooked.

Since only an order-of-magnitude answer is necessary and in view of the large size of ATP synthase, molecular mechanics and force field will be introduced in Chapter 2, and the force field charges will be used to estimate the electric field and electrostatic of ATP synthase. Thus, in Chapter 2 and Chapter 3, AMBER charges, a molecular mechanics force field particularly designed for reproducing protein structures, will be used to evaluate the field and potential over a 3D grid surrounding ATP synthase. It is expected that ATP synthase exhibits an intrinsic protein electric field around $10^6 – 10^8$ $V \cdot m^{-1}$, in which case this would be a large perturbation of the chemiosmotic electric field generated by the H^+ gradient. Finally yet importantly, the strong intrinsic electric field can trigger electroporation effects on the mitochondrial inner-membrane phospholipid bilayer, causing serious biological consequences such as cell apoptosis.

Chapter 2. PrES: A New Tool Approximating Protein Electrostatics

"All models are wrong, but some are useful."

- George E. P. Box, *Robustness in Statistics,* 1979

Recall that the purpose of this project is to find an order of magnitude approximation for the intrinsic electric field of ATP synthase. This chapter discusses computational methods available to answer this question. In particular, a new computational tool, PrES (Protein Electrostatics), is introduced, including the pros and cons of PrES compared with other tools. Furthermore, a benchmark PrES calculation is performed on a smaller molecule, demonstrating the level of accuracy of PrES in comparison with a standard density functional theory (DFT) result. In this benchmark example, PrES calculation shows results of the same order of magnitude as results from DFT calculations. This means that the accuracy of PrES calculation satisfies our aim to find an order of magnitude answer.

The PrES program is based on molecular mechanics. Quantum mechanical and molecular mechanical methods will be briefly discussed in this chapter, with an emphasis on the inevitability of using molecular mechanics charges in PrES. An in-depth analysis of quantum mechanical and molecular mechanical methods is beyond the scope of this thesis. In our benchmark example, PrES calculated the electrostatic potential of met-enkephalins – a small peptide molecule, and DFT performed electrostatic potential calculations on the same molecule.[42]

Although quantum chemistry methods such as DFT can calculate electrostatics of chemical molecules, one has to solve the Schrödinger equation and calculate wave functions one way or another.[42, 43] If we use quantum calculations to approximate the overall electrostatics of a large biomolecule such as ATP synthase, the amount of computation is so costly that no current computers may be able to process.[44, 45] Therefore, one has to turn to molecular mechanics for solutions.

The PrES program is a new tool to approximate protein electrostatics based on molecular mechanical methods. This tool will be described in section 2.2 of this chapter, and its application on ATP synthase will be described in chapter 3 to answer the research question in this project. Molecular mechanics methods represent a trade-off between the accuracy of calculations and lower computational expenses. To validate the accuracy, we use PrES to calculate the electrostatic potential of a met-enkephalins molecule, and then the results are compared with quantum calculation performed by DFT. This benchmark calculation will be presented in section 2.5 of this chapter.

2.1 Molecular Mechanics and AMBER Force Field

Molecular mechanics methods avoid involving a wave function or the total electron density in the approximation process. Unlike quantum chemical calculations, which describe each electronic motion fully, molecular mechanics (MM) methods treat a molecule as a collection of charged point-masses connected by springs. MM methods assign empirical potential functions to describe the various term such as atom-atom bond stretching, angular and torsion energy terms, and electrostatic energy terms based on atomic charges assigned to the various atoms. Experimental data or high-level quantum

chemical calculations parametrize a set of mathematical equations including these various terms in what is known as a "force field", and the force fields then are used to optimize the molecular geometry through the minimization of the energy.[46]

In this project, force field charges are obtained from Assisted Model Building with Energy Refinement (AMBER), a molecular mechanics software package. AMBER parameters are first obtained from experimental measurements (X-ray diffraction, microwave, neutron diffraction, etc.), and then refined by calculations based on:[47]

$$E_{total} = \sum_{bond} K_t\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] + \sum_{i<1}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{DR_{ij}}\right] + \sum_{H-bond}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (2.1)$$

In Equation 2.1, each term accounts for contribution to the overall molecular motion from chemical bond, bond angles, dihedral potentials, van der Waals potential, and hydrogen bond, respectively.

In addition, the H-bonding effect is introduced in AMBER using a supplementary potential term. Furthermore, a dielectric constant dependent on interatomic distance accounts for intramolecular electrostatics in water media; atoms with distance greater than 9.0 Å are excluded in intramolecular electrostatics calculation.[47]

In general, the electrostatic term in molecular mechanics (MM) force fields (including AMBER) is computed from a simple Coulombic energy expression between atomic charges centered on the nuclear positions. Thus, the electrostatic field (or potential) originating from the real charge distribution (including a positive contribution of a sum over discrete point-like nuclei and a negative contribution from the continuous distribution of electron density) is condensed and approximated by a set of discrete

atomic charges. To explicate the conceptual step undertaken by a force field in assigning atomic charges, one can write, for example, the following expression for the electric field evaluated at position $\mathbf{r_0}$ in the exact and approximate manner as follows:

$$\mathbf{E}(\mathbf{r_0})^{\text{exact}}_{\text{quantum}} = +\frac{e}{4\pi\varepsilon_0}\sum_{i=1}^{N}\frac{Z_i}{|\mathbf{r_i}-\mathbf{r_0}|^2}\,\hat{r}_{i,0} - \frac{e}{4\pi\varepsilon_0}\iiint_{-\infty}^{+\infty}\frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r_0}|^2}\,d^3\mathbf{r}$$

$$\approx \mathbf{E}(\mathbf{r_0})^{approximate}_{MM-AMBER} = \frac{e}{4\pi\varepsilon_0}\sum_{i=1}^{N}\frac{Q_i^{MM-AMBER}}{|\mathbf{r_i}-\mathbf{r_0}|^2}\,\hat{r}_{i,0} \quad (2.2)$$

where Q_i and Z_i are in multiples of the elementary charge (of magnitude e), and where the positions $\mathbf{r_i} = \mathbf{r_0}$ or $\mathbf{r} = \mathbf{r_0}$ are excluded to avoid non-physical singularities.

The approximation embodied in Equation 2.2 uses atomic charges centered on nuclear positions to evaluate the approximate electrostatic potential of the biological macromolecule, and this bypasses the expensive quantum mechanical calculation necessary to evaluate the electron density $\rho(\mathbf{r})$ term in Equation 2.2. These charges are obtained by fitting the rigorously calculated electrostatic potentials of numerous small molecules to a set of discrete atomic charges placed at the position of the nuclei. Quantum mechanical calculations, such as DFT (with an appropriate basis set), pre-calculated the electrostatics of small molecules including 20 amino acids and/or building blocks of nucleic acids). The atomic charges are then adjusted, in a least square sense, to reproduce the quantum mechanical electrostatic potential as closely as possible. Finally, transferability is assumed whereby the charges calculated for the separate amino acids are transported to the corresponding atoms in the macromolecule based atomic types. Older version of the AMBER force field suppressed hydrogen atoms – it calculates an effective charge for the entire group containing the hydrogen atoms, for example, $–CH_3$ is considered as a single point charge. A modern version used in this thesis places charges

on all atoms including hydrogens. This charge assignment was achieved by the "MM Charge" feature available in GaussView 5.0 software.

An electric field is a vector field expressing the force that a unit electric charge experiences at a point in space near an electric charge that generates the field. An electric field, hence, has the dimensions of force per unit charge. Through a simple dimensional analysis, electric field is equivalent to voltage per length – which, in SI units, is expressed as $N \cdot C^{-1}$ or $V \cdot m^{-1}$. A positive charge is, by convention, a source of an electric field while a negative charge is a sink of field lines. A proton H^+ possesses a positive charge, thus it is a source of electric field while an anion acts as a sink for the field.

The force between two charged bodies is obtained from Coulomb's law which defines that a particle with electric charge q_1 at position $\mathbf{r_1}$ imposes a force $\mathbf{F}$ on position $\mathbf{r_2}$ with charge q_2 (or vice versa by Newton's third law):

$$\mathbf{F} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{|\mathbf{r_1} - \mathbf{r_2}|^2}\,\hat{r}_{2,1} \qquad (2.3)$$

where ε_0 is the permittivity constant of free space, and $\hat{r}_{2,1}$ is the unit vector from position $\mathbf{r_2}$ to $\mathbf{r_1}$. Then, the electric field at position $\mathbf{r_2}$ can be expressed as electrical force $\mathbf{F}$ per unit charge at $\mathbf{r_2}$:

$$\mathbf{E}(\mathbf{r_2}) = \frac{\mathbf{F}}{q_2} = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{|\mathbf{r_1} - \mathbf{r_2}|^2}\,\hat{r}_{2,1} \qquad (2.4)$$

To calculate the overall electric field of a large protein molecule such as ATP synthase, one must consider positions and electric charges on all atoms in the molecule. At position r_0 in space, one can "feel" the electric field from all charges nearby. Thus, the calculation sums up all atoms and atomic charges at $\mathbf{r_i}$, and the intrinsic electric field is a collective physical property from all charges in this molecule:

$$\mathbf{E}(\mathbf{r_0}) = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{N}\frac{q_i}{|\mathbf{r_i} - \mathbf{r_0}|^2}\,\hat{r}_{i,0} \qquad (2.5)$$

Note that the position of interest cannot be the same as the atom nucleic position (i.e. $i \neq 0$), because the electric field value approaches infinity (i.e. $\mathbf{E(r_0)} \rightarrow \infty$) when the distance between the two positions approaches zero (i.e. $|\mathbf{r_i} - \mathbf{r_0}| \rightarrow 0$). But this is not an issue outside of the protein structure.

As any conservative vector field, the electric field can be derived from an underlying scalar field, in this case termed the electrostatic potential (φ) – up to an arbitrary constant:

$$\mathbf{E} = -\nabla(\varphi + constant) \qquad (2.6)$$

The electrostatic potential is the derivative of the field **E** in the gradient operator – it has the distance denominator raised only to the first power. Thus, neglecting the constant, the potential is written as:

$$\varphi(\mathbf{r}_2) = \frac{Energy}{q_2} = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{|\mathbf{r}_1 - \mathbf{r}_2|} \qquad (2.7)$$

where the electrostatic potential is a scalar, and it is the electrical energy that a unit electric charge experiences at a point near the electric charge giving rise to the potential. The potential, thus, has the dimensions of energy per unit charge. Hence, the potential is equivalent to voltage - which is expressed in $J.C^{-1}$ or V in SI units. An isolated positive charge is surrounded by a positive potential, since it takes energy to move a positive unit charge from infinity to a point in its vicinity. A negative charge is surrounded by a negative potential, since energy is released as the unit positive charge approaches the negative charge. The potential at infinity is considered to be equal to zero, hence it is reasonable to eliminate the constant in the Equation 2.6 above. Similar to Equation 2.5,

the below Equation 2.8 describes the electrostatic potential that arises from a collection of discrete charges. The electrostatic potential φ ($\mathbf{r_0}$) is defined as the work required to move charge q_i from infinity to spatial position $\mathbf{r_0}$:

$$\varphi(\mathbf{r_0}) = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{N}\frac{q_i}{|\mathbf{r_i} - \mathbf{r_0}|} \qquad (2.8)$$

To compute the field and the potential surrounding a protein, one needs to affix the AMBER charges to the atomic position determined by X-ray diffraction. Then, Equation 2.5 and Equation 2.8 can be applied at an arbitrary point in the space. Ideally, one should evaluate the field and the potential at grid points surrounding the protein of interest, and section 2.3 and section 2.4 will describe how this is done.

Here we need experimental data to provide atomic positions and dress them with AMBER point charges. To this end, the x-ray diffraction (XRD) crystallographic structure of a protein molecule can serve as a starting point to compute the electric field of the protein. XRD can map out the structure of a macromolecule. In this project, the XRD structure of ATP synthase from *Paracoccus denitrificans* is obtained from the Protein Data Bank (PDB ID: 5DN6). The experimental data provides the spatial coordinates (x, y, z) of each atom on ATP synthase with a resolution of 4.0 Å.[48]

To compute the electric field based on Equation 2.5, one needs not only the Cartesian coordinate at (x, y, z) of each atoms, but also the electric charges at those positions. As mentioned earlier, force field charges are obtained from AMBER - a molecular mechanics software package.[49] The AMBER force field extracts physical parameters of bio-molecules from experimental measurements, and it produces high-quality force field results for protein and nucleic acid molecules.[47]

Chemistry At Harvard Macromolecular Mechanics (CHARMM) is another simulation tool used to obtain force fields based on protein electrostatics.[46, 49, 50] CHARMM force fields can assign partial atomic charges by an analogy algorithm described by Vanommeslaeghe and co-workers.[51] However, CHARMM is designed for much smaller molecules. In a previous study on bio-molecule conformation, CHARMM produced much more missing parts of molecular conformation compared to AMBER.[49] This can lead to a large portion of incomplete charge assignment on large protein molecules such as ATP synthase, so CHARMM is unsuitable for assigning force field charges in this project.

Therefore, AMBER is selected in this project as the force field to obtain partial atomic charges of ATP synthase. The atomic charge obtained by AMBER is called AMBER charge in this project.

2.2 Review of Available Software

Recall that this study is to calculate an order of magnitude answer of the intrinsic electric field of ATP synthase. This task is challenging due to the molecular size of the ATP synthase protein. A common strategy is to combine the Poisson equation with the Coulomb equation to first obtain the electrostatic potential $\varphi(\mathbf{r})$, then one can use the position-dependent dielectric constant $\varepsilon(\mathbf{r})$ and the charge density distribution $\rho(\mathbf{r})$ to differentiate electrostatic potential $\varphi(\mathbf{r})$ to obtain the local electric field: [52, 53]

$$\nabla\varphi(\mathbf{r}) = \frac{-4\pi\rho(\mathbf{r})}{\nabla\varepsilon(\mathbf{r})\ kT} \qquad (2.9)$$

The DelPhi software, developed by Alexov group, applies Equation 2.9 to compute electric field, and the electric field calculation depends on the electrostatic

potential calculation.[54, 55] Furthermore, the DelPhiForce program can compute electrostatic force. First, DelPhi uses the Poisson-Boltzmann equation to approximate the charge density of a molecule; then, DelPhi estimates the electrostatic potential of the molecule based on the charge density; finally, DelPhi takes derivatives (negative gradient) of the electrostatic potential at each grid point to obtain the electric field.

There are two limitations associated with DelPhi. For one, DelPhi has high computational costs associated with its generation of the electric field from the electrostatic potential because for a protein molecule, the program has to calculate multi-variable derivatives at x, y, and z dimension for a large number of atoms. For another, DelPhi does not exclude grid point near or at the atomic positions. According to Equation 2.9, the distance between the measurement position $\mathbf{r_0}$ and the atomic position $\mathbf{r_i}$ will approach zero when we compute the value at or near a nucleus, and the defined value of electric field at r_0 approaches infinity, making the approximation less accurate.

2.3 Protein Electrostatic (PrES) Program

Since conventional computational tools such as DelPhi are limited and computationally expensive. In this section, we introduce the Protein Electrostatic program (PrES). The AMBER force field is applied to assign force field charges for input files of any chemical molecules.

PrES is a new computational tool to quickly obtain a reasonable estimate for the intrinsic electric field of any protein molecule (Figure 2.1). PrES program also overcomes the limitations of existing tools described in section 2.3. The algorithm used in PrES and a benchmark example of PrES calculation are presented in this section.

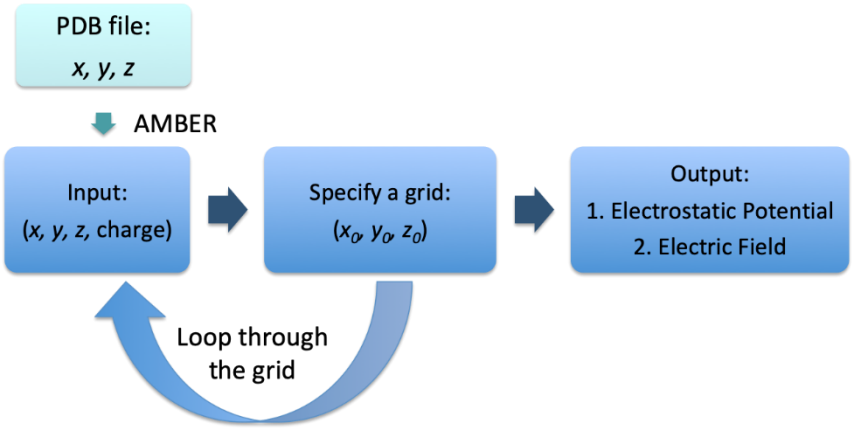


Figure 2.1 Algorithm Flow Chart of PrES Program.

As shown in Figure 2.1, an input file has to be prepared for a PrES calculation. Typically, an input file contains five columns: atom type, atomic position x, y, z, and the AMBER charge values for each atom. To prepare an input file, one should first obtain a Protein Data Bank (PDB) file with the Cartesian coordinate (x, y, z) of each atom. Then, AMBER charge values are obtained using GaussView 5.0. Cornell and co-workers have described how to obtain AMBER electrostatics in the Gaussian and GaussView programs.[56] Further, the PrES will ask users to specify the grid, either a particular grid point or a grid with numerous grid points in code, and this is to specify where the electric fields should be calculated. It should be noted here that, if a grid point (x_0, y_0, z_0) is within 10^{-3} Å close to an atomic position (x, y, z), PrES will automatically exclude this grid point. This is to avoid introducing an infinity, i.e., division by zero, in Equation 2.5 and Equation 2.8. Finally, PrES produces an output file with the magnitude of the electric field at each grid point.

PrES calculates electric field and electrostatic potential independently. The electric field calculation is based on Equation 2.4, and the electrostatic potential calculation is similarly based on the definition of electrostatic potential as given by Equation 2.8 above.

As such, PrES calculates the electrostatic potential independently from (x, y, z, AMBER charge). Note that PrES assumes that each AMBER charge represents the real atomic charge accurately. Although this approach significantly reduces the computational expenses compared to the DelPhi program described above, the accuracy does depend on the AMBER charge values used to directly represent the atomic charge values.

2.4 An Example of PrES Calculation: Test of Accuracy

A benchmark calculation was carried out on the met-enkephalins peptide. This is to (1) show the reader how to carry out a typical PrES calculation and (2) test of the accuracy of PrES. The met-enkephalins peptide was selected due to its simplicity and short length. A peptide is a basic subunit of a protein molecule. An NMR-characterized met-enkephalins structure is used (PDB ID: 1PLW).[57] Our benchmark calculation is compared with a calculation obtained using DFT calculations.

A snapshot of the input file of this PrES calculation is shown in Table 2.1, and the results obtained from the PrES output file are presented in Table 2.2. To compare with the PrES results, the electrostatic potential values were calculated by DFT method at eight vertices points of a box surrounding a met-enkephalins molecule. Figure 2.2 shows where those eight points are located near the met-enkephalins molecule. The DFT

method was carried out by the Gaussian 09 program, and the numerical results are also presented in Table 2.2. The results of two methods can be compared in Table 2.2.

As Table 2.2 shows, the PrES program yields electrostatic potential values of met-enkephalins on a reasonable order of magnitude as the values from a DFT calculation. The B3LYP method with 6-31G** basis set was used in the DFT calculation, as this is an appropriate approach to calculate molecular electrostatics based on literature.[44, 46] The outliers in Table 2.2 may be due to the distance from those 8 corner points to the molecule, as shown in Figure 2.2. While calculating potentials at the vertices of a smaller box may improve the results, this benchmark calculation already demonstrates a reasonable accuracy level of the PrES program. This level of accuracy satisfies our research need of quickly calculating the overall electrostatics of large biomolecules.

In addition, PrES only takes a few seconds to yield its results, whereas Gaussian 09 program requires approximately 1 hour of CPU time to yield the DFT results shown in Table 2.2. Although DFT calculations may produce highly accurate results, the computational expense would be enormous in order to calculate the electric field for large protein such as ATP synthase.

				AMBER
Atom Type	x	y	z	charge
…	…	…	…	…
O	1.689	5.595	1.440	0.6123
C	-0.622	3.693	3.070	-0.5713
…	…	…	…	…

Table 2.1 A snapshot of PrES input file.

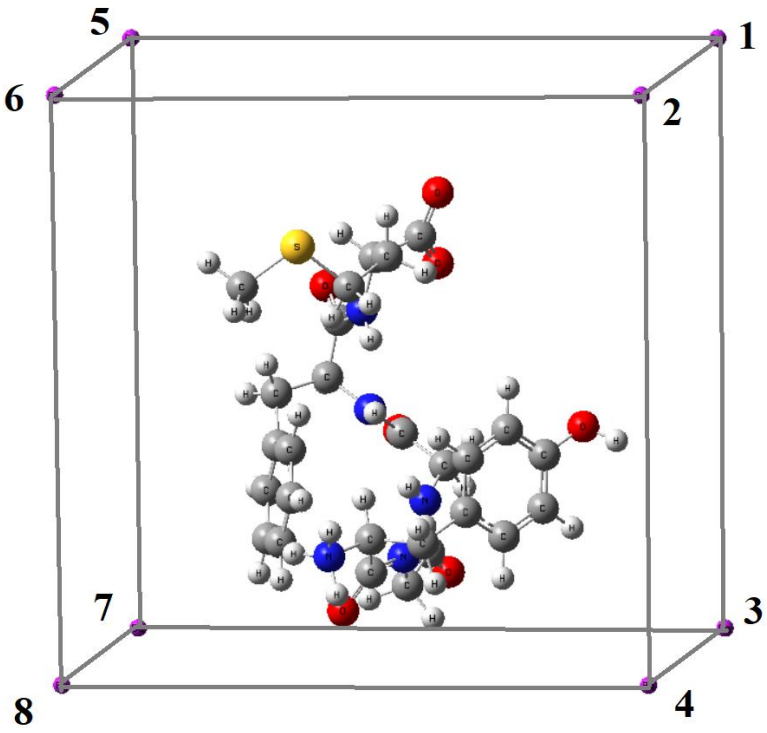


Figure 2.2 Met-enkephalins in a box with 8 points at which the electrostatic potential has been evaluated by both DFT and PrES labeled in the same order as listed in Table 2.2. The image was produced by GaussView 5.0 program.

Point No.	DFT Calculation (B3LYP/6-31G**)	PrES Calculation
1	-2.13×10^{-2}	-3.01×10^{-2}
2	-1.00×10^{-4}	-6.49×10^{-4}
3	-2.79×10^{-3}	-6.02×10^{-3}
4	2.09×10^{-2}	2.70×10^{-2}
5	-1.83×10^{-2}	-2.44×10^{-2}
6	2.25×10^{-4}	2.61×10^{-3}
7	-6.33×10^{-4}	-1.53×10^{-3}
8	2.32×10^{-2}	3.46×10^{-2}

Table 2.2 Comparing the results of DFT vs. PrES calculations of the electrostatic potential (in atomic units) at 8 points around met-enkephalin at the corners of the box in Figure 2.2. The DFT calculation was carried out by Gaussian 09 program. The CPU times are 3-5 seconds for the PrES calculation and approximately 1 hour for the DFT calculation on the same processor.

2.5 Summary

A PrES calculation can quickly approximate the intrinsic electric field and electrostatic potential for peptides and protein, or to that effect, any (bio)polymer, as long as reasonable point charges can be assigned to the atoms composing it. A benchmark PrES calculation shows that PrES satisfies our interest in obtaining qualitatively acceptable results for the intrinsic electrostatic potential of ATP synthase. This benchmarking was achieved by comparing the electrostatic potential, obtained from PrES with that obtained from a DFT calculation at 8 points surrounding a penta-peptide of met-enkephalins, at the same geometry.

Care has been taken in PrES to exclude the contribution of a given atom to the electric field/potential if these physical fields are to be estimated at precisely the nuclear position of that very atom. Thus, if a grid point (x_0, y_0, z_0) happens to be located at an atomic position (x_i, y_i, z_i), PrES will automatically exclude this grid point from the calculation. This will make PrES results meet a more authentic physical representation of an electric field and an electrostatic potential.

The next chapter will introduce how PrES is applied to solve the main research question, that is, to calculate the intrinsic protein electric field of ATP synthase. In particular, PrES will calculate electric field at a grid across the ATP synthase protein.

Chapter 3. Intrinsic Electric Field of ATP Synthase

3.1 Hypothesis

As discussed in Chapter 1, oxidative phosphorylation occurs at the inner-membrane of mitochondria, where an H^+ gradient is generated between the inner and outer mitochondrial membrane, and then the energy from the H^+ gradient is converted to chemical energy in ATP. Thus, the synthesis of ATP is an endothermic process. The total free energy ΔG is given by:

$$\Delta G = \Delta G_{chem.} + \Delta G_{elec.-H^+} \qquad (3.1)$$

where $\Delta G_{chem.}$ is the free energy provided by the chemical gradient of H^+, and $\Delta G_{elec.-H^+}$ is the free energy provided by the electrochemical gradient of H^+. As shown in Chapter 1, the electric field generated by the H^+ gradient is of the order of 10^7 V·m^{-1}.

The intrinsic electric field of immobilized protein can speed up H^+ diffusion.[58] However, the phospholipid bilayer of the inner mitochondrial membrane is a mobilized environment; as H^+ passes through ATP synthase, can the electric field of the protein contribute additional energy to the H^+ translocation? Since the H^+ gradient generates a field with a 10^7 V·m^{-1} order of magnitude, the intrinsic electric field of ATP synthase has to exhibit a comparable order of magnitude (10^6 – 10^8 V·m^{-1}) to interfere significantly with the H^+-generated electric field.

Furthermore, the electric field is a vector field. Hence the projection of the intrinsic field of ATP synthase can reinforce or weaken the chemiosmotic electric field depending on its direction and magnitude whether two electric fields are parallel or anti-parallel. Thus, the additional electric field of the protein may assist or hinder the H^+ translocation as shown in Figure 3.1.

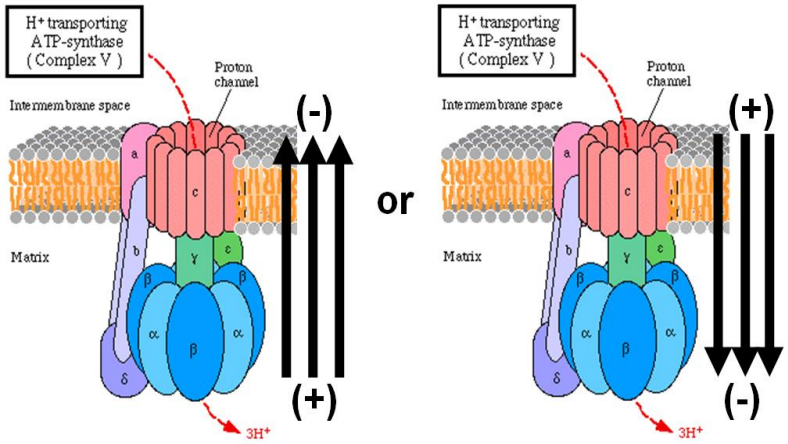


Figure 3.1 Effect of an intrinsic electric field parallel or antiparallel to the direction of proton translocation through the membrane on the electrostatic force. Protons are shown to move from the intermembrane space to the mitochondrial matrix. (Matta, 2015 – reproduced with permission, ref 34)

If the intrinsic electric field of ATP synthase indeed plays a role in H^+ translocation, the protein not only catalyzes ATP synthesis, but it also contributes to the energetics of chemiosmotic theory, and we may add a third term to complement Mitchell's equation shown in Equation 3.1:[34]

$$\Delta G = \Delta G_{chem.} + \Delta G_{elec.\text{-}H^+} \quad (\pm \Delta G_{elec.\text{-}ATP\ Synthase}) \qquad (3.2)$$

$$\Delta G = 2.3nRT\Delta\text{pH} + nF\Delta\Psi_{H^+} \quad (\pm nF\Delta\Psi_{ATP\ Synthase}) \qquad (3.3)$$

3.2 Intrinsic Protein Electric Field

Proteins are biomolecules formed by one or more amino acid chains with more than 30 residues. When the net electric charge of a protein molecule is zero, it is a result of the atomic charge canceling each other. Atomic point charges collectively generate an intrinsic electric field, as long as the flux in the field on any Gaussian surface surrounding

the protein is equal to the total enclosed charge divided by the permittivity of the medium. Whether or not the protein is neutral, it generates an intrinsic electric field. If the total charge is zero, only the total flux is zero, but locally there is still regions that act a source and other as sink of field lines, and intrinsic electrical energy in protein is stored in those atomic point charges.[59]

Recent work has shown that the protein intrinsic electric field can regulate dipolar molecules passing through protein channels. The study finds that the porins protein channel of *E. Coli* exhibits an intrinsic electric field of 40 $mV \cdot Å^{-1}$, or 4 x 10^8 $V \cdot m^{-1}$. This strong intrinsic field filters dipolar molecules such as water by dipole alignment in the restricted region. Furthermore, the local field changes direction when escaping from the restricted region, imposing an additional energy barrier.[60] A proton (H^+) is charged positively, and a strong intrinsic electric field of ATP synthase "filters" H^+ in the same manner.

Patthy and Thész found that highly charged protein side-chains, especially arginine residues, generate electric field for ion binding.[61] As mentioned in Chapter 1, at the entry to the ATP synthase H^+ channel, arginine residues play a critical role in the protonation/deprotonation process. Is the entry to the H^+ highly charged? As this may also affect the H^+ translocation processes.

Further, the intrinsic protein electric field plays roles in enzymatic catalysis[62, 63] and protein-metal binding selectivity.[64] As mentioned in the Chapter 2, there is no computational tool that can quickly and directly calculate intrinsic protein electric field. Most tools first approximate electrostatic potential and then obtain the electric field by

taking a negative gradient at each grid point. QM/MM method can study refined protein electrostatics at enzyme active site, but it doesn't study overall protein electrostatics.

3.3 Protein Electrostatics (PrES) Program

As mentioned in Chapter 2, the electric field defined as:

$$\mathbf{E}(\mathbf{r_0}) = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{N}\frac{q_i}{|\mathbf{r_i} - \mathbf{r_0}|^2}\,\hat{r}_{i,0} \qquad (2.5)$$

where an atom with electric charge q_i at position $\mathbf{r_i}$ imposes an electric field $\mathbf{E(r_0)}$ on position $\mathbf{r_0}$, ε_0 is the permittivity constant of free space, and $\hat{r}_{i,0}$ is the unit vector from atomic position $\mathbf{r_i}$ to the grid point $\mathbf{r_0}$. PrES uses molecular mechanics charges and is coded in Python, and has been introduced in Chapter 2 of thesis. PrES approximates the magnitude of the electric field $\vec{E}$ on a grid by incorporating an AMBER charge on all atoms and their atomic position (x_i, y_i, z_i). Furthermore, PrES can approximate the electrostatics potential of the protein.

The PrES program is meant for a quick calculation of the protein electric field and/or protein electrostatic potential, yielding an order of magnitude approximation of both values. A benchmark PrES calculation has been shown in Chapter 2. PrES is suitable to perform our desired calculation on protein molecules. The PrES results are presented and interpreted below in Section 3.4.

3.4 PrES Calculation of ATP Synthase

As mentioned earlier, the goal of this project is to calculate the intrinsic electric field of ATP synthase, and to determine if the electric field may contribute additional free

energy to the H^+ translocation across the inner mitochondrial membrane. To this end, a PrES calculation was carried out in a box on an ATP synthase molecule from *P. denitrificans* (Figure 3.2), using a 4.0 Å-resolution crystal structure of ATP synthase (PDB ID: 5DN6) determined by Morales-Rios and co-workers.[48] This crystal structure has the highest resolution available.

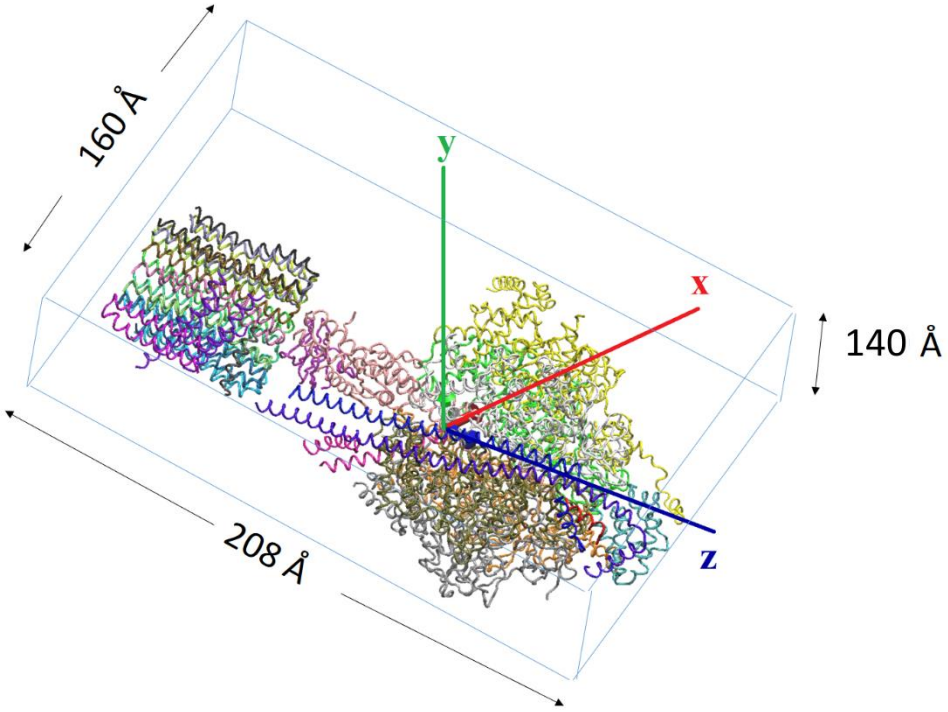


Figure 3.2 ATP Synthase in a box with a Cartesian coordinate of the pdb file. The image was produced by the UCSF Chimera program. (Matta, 2019 – reproduced with permission)

As shown in Figure 3.2, ATP synthase is a large protein with approximate dimension of 140 Å x 160 Å x 208 Å. (x_i, y_i, z_i) coordinates of each atom of this molecule were exported to a PDB file and processed by GaussView 5.0 software. GaussView 5.0 enabled us to obtain AMBER electric charge values in order to prepare a PrES input file. Then, the input file was processed by the PrES program, and the

magnitude of electric field at selected grid point inside the box was calculated. The PrES results are presented in Table 3.1.

				Electric Field ($V \cdot m^{-1}$)	
	x(Å)	y(Å)	z(Å)	In vacuum	In water medium
Inside the box:	5	5	5	4.15×10^8	5.46×10^6
	10	10	10	3.69×10^8	4.85×10^6
	15	15	15	3.87×10^8	5.09×10^6
	20	20	20	1.27×10^9	1.67×10^7
	25	25	25	1.67×10^{10}	2.20×10^8
	50	50	50	1.34×10^9	1.76×10^7
10 Å Outside:	174	75	168	4.44×10^8	5.84×10^6
25 Å Outside:	189	90	183	3.20×10^8	4.21×10^6
100 Å Outside:	264	165	258	1.04×10^8	1.37×10^6

Table 3.1 Magnitude of intrinsic electric field of ATP synthase at selected grid points calculated by PrES program. The electric field magnitude in water medium is adjusted by relative permittivity constant of aqueous medium ($\kappa = 76$).

Table 3.1 shows the result of our preliminary calculation. The box shown in Figure 4 represents the approximate dimension of the protein (140 Å x 160 Å x 208 Å). In Table 3.1, the electric field magnitude was first calculated in vacuum, and then it is adjusted to aqueous medium by the relative permittivity constant $\kappa = 76$. The intrinsic electric field is between 10^6 and 10^8 $V \cdot m^{-1}$ inside and near this box. Recall that the H^+ gradient exhibits an electric field of around 10^7 $V \cdot m^{-1}$, this suggests that the intrinsic field of ATP synthase is indeed of a magnitude comparable with that of the H^+ gradient across inner mitochondrial membrane. The intrinsic electric field of ATP synthase appears to be so strong that it should contribute additional free energy to the H^+ translocation, either facilitating or hindering the H^+ translocation. This also suggests that the electric field

exerted by ATP synthase may affect structure and functions of other proteins (such as cytochrome c) embedded in the inner mitochondrial membrane.

According to Table 3.1, certain regions inside the protein molecule exert an electric field as high as 2.20×10^8 $V \cdot m^{-1}$. From previous research by Patthy and Thész, protein backbone dipoles such as α-helices were shown to exert strong local electric fields, which facilitates protein binding.[61] The electric field strength decreases towards the surface of the protein, and it drops further outside the box. The protein electric field strength is, of course, a function of protein structure and residue sequences.

Nonetheless, the electric field 10 - 100 Å outside the box is found to be around 10^8 $V \cdot m^{-1}$ for the sampling points in Table 3.1. Chapter 1 illustrated that protons go through the channel via a protonation/deprotonation process at the two carboxylic acid residues: arginine and aspartic acid. What is the electric field strength near these two carboxylic acid residues? Does a strong field strength attract H^+ at the entrance of the proton channel? For certain residues with a strong local electric field, they may play unknown roles in the ADP + Pi → ATP reaction by assisting or hindering the H^+ translocation.

3.5 Mapping the Field and Charge Distribution

PyMOL is a visualization software for macromolecules.[65, 66] With the same PDB structure, PyMOL demonstrates a net charge of -3.000 a.u. by sampling 198 points across the ATP synthase molecule. This qualitatively shows that the ATP synthase molecule possesses a net negative charge. In an environment with high H^+ concentration, a

negatively charged ATP synthase can naturally attract H^+ to its ion channel. However, this may also create an energy barrier for H^+ to exit the ion channel.

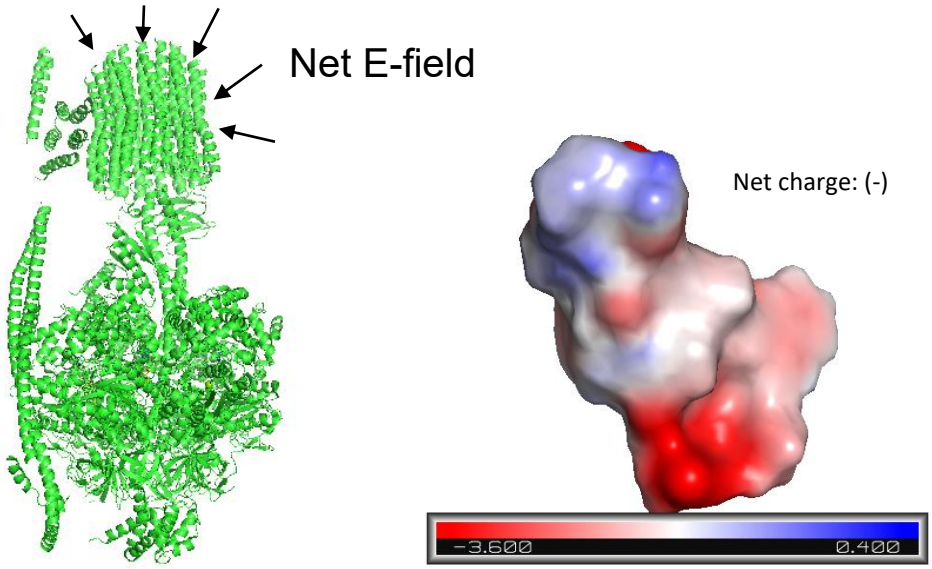


Figure 3.3 ***(Left)***A net negative charge enforces the overall electric field vector to point towards ATP synthase. ***(Right)*** As visualized by PyMOL software, the net charge of ATP synthase is negative, despite that the entrance of ATP synthase is charged positively.

Evidence by Elston et al. shows an energy barrier in the F_o unit of ATP synthase, whereby the local electric field works against the rotor motions in F_o.[67] They further explained the biological consequences of this energy penalty in ATP synthase. For one, the electric field keeps the arginine residues protonated and sustains the H^+ supply during the H^+ translocation process. For another, the electric field enables the H^+ to be passed to the stator part of ATP synthase (subunit b, d, e, g, figure 1.3), which improves the effectiveness of rotor's diffusion.

The top of the cylindrical part of ATP synthase is positively charged (Figure 3.3, right). The cylindrical part of ATP synthase is the where F_o is located. H^+ in the inter-

mitochondrial membrane space enters ATP synthase by protonating arginine residues on F_o. As a result, the top of F_o, where H^+ enters ATP synthase, maintains a high pKa value.[67] This explains a positively charged area in an overall negatively charged protein molecule.

Last but most importantly, PyMOL was used to visualize and identify the top and the bottom of ATP synthase. The identified coordinates were used to approximate the electrostatic potential between the top and the bottom. As shown in Figure 3.4, the potential difference is ~ 70 mV. As mentioned in Chapter 1, a potential difference between 160 mV and 180 mV is generated by H^+, and this is comparable to that of the potential difference generated by ATP synthase (~ 70 mV). Indeed, the potential generated by the H^+ is comparable to that generated by protein intrinsic electric field.

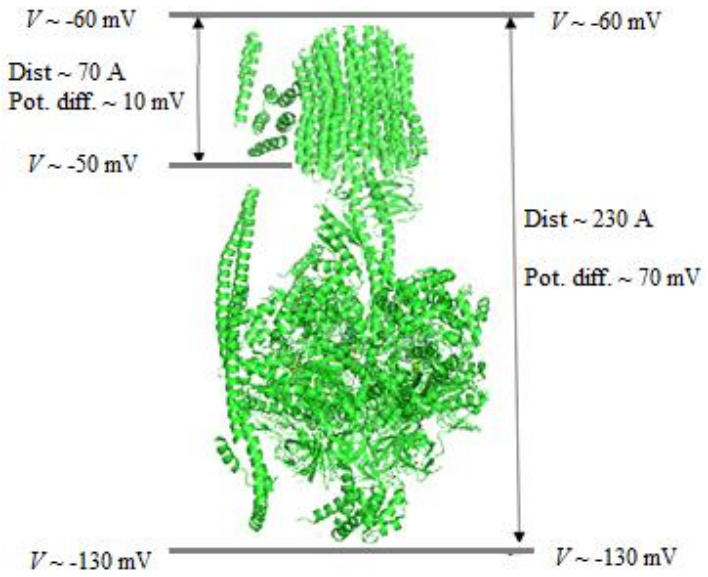


Figure 3.4 Approximate voltage differences between key locations of ATP synthase in an aqueous medium with a dielectric constant $\kappa = 76$. The positions of the top and bottom are defined as 6 Å away along the vertical axis from the furthest atoms in the *b*2 and δ subunit's respectively, while the intermediate position is the midpoint of the gap. PyMOL was used to visualize and identify the key positions of ATP synthase.

PyMOL was used to visualize and identify the key representative locations in the vicinity of ATP synthase. The identified coordinates were used to approximate the electrostatic voltage between the top and the bottom. In Figure 3.3, the protein is oriented so that the top represents the F_o subunits that crosses the bilayer of the membrane, while the larger bottom part is primarily the α and β subunits.

H^+ enters and exits from a channel in the subunit c in F_o. A proton will thus go from a potential of *ca.* -60 mV to a higher potential of only -50 mV when it exits the channel and into the matrix. This would translate into an energy given by Equation 3.3 presented in section 3.1:

$$\Delta G = -nF\Delta\varphi$$

$$= -1 \times 96.5 \text{ (kJ / (V} \times \text{equiv.weight))} \times (-0.01 \text{ V})$$

$$\approx 1 \text{ kJ/mol} (\approx 0.23 \text{ kcal/mol})$$

where *F* is Faraday's constant, and where mol refers to a mol of proton crossing the channel of the F_o subunit.

But the proton then will encounter a downhill gradient due to the intrinsic electric field of the protein until it reaches the "bottom" of the molecule well into the mitochondrial matrix. Now the proton is going with the gradient, hence it is an exergonic process. The energy released is estimated, this time, with a favorable electrostatic potential difference of 80 mV:

$$\Delta G = -96.5 \times (0.08 \text{ V})$$

$$\approx -8 \text{ kJ/mol} (\approx -2 \text{ kcal/mol})$$

If transferred under chemiosmotic voltage alone, i.e. ignoring the role of ATP synthase in the translocation, how do these energies compare with protons? We repeat the calculation with the potential of the membrane ~ 170 mV:

$$\Delta G = - 96.5 \times (0.17 \text{ V})$$

$$\approx -16 \text{ kJ/mol} (\approx -4 \text{ kcal/mol})$$

The intrinsic field of the protein at the entry and exit point will diminish the available free energy by approximately 1 kJ per mole of proton, that is ((-16+1)/16) ≈ 6%, and it operates as a barrier slowing the entry of the protons through the F_o unit. A downhill release of energy follows, for per mole of protons passing ATP synthase protein, it releases 8 kJ – this is about half of the 16 kJ per mole of energy in the "naked" chemiosmotic electric field without consideration of the protein intrinsic field.

Therefore, a full image of the H^+ translocation process emerges as a chemical reaction with a "transition state". The transition state is while the proton is within the channel crossing the membrane, with a "forward barrier" of 1 kJ/mol, a "backward barrier" of 8 kJ/mol, and a clearly exergonic overall reaction with a ΔG of approximately -7 kJ/mol (Figure 3.5).

If one realizes that a large number of protons are constantly crossing ATP synthases, and that there are possibly 100,000 ATP synthase in a mitochondrion, 500 to 5000 mitochondria in a cell, and trillions of cells in a biological organism, our calculation appears to be more than a correction to the assumed numbers so far in the literature.

The rate of protons passing through ATP synthase is believed to be of the order of 1000/sec per ATP synthase molecule. Hence, the electric field of this molecule can be a significant source of regulation on the metabolism rate and on its overall energetic

efficiency. The exact details of this contribution should be worked out in the future given its central importance to intermediate metabolism.

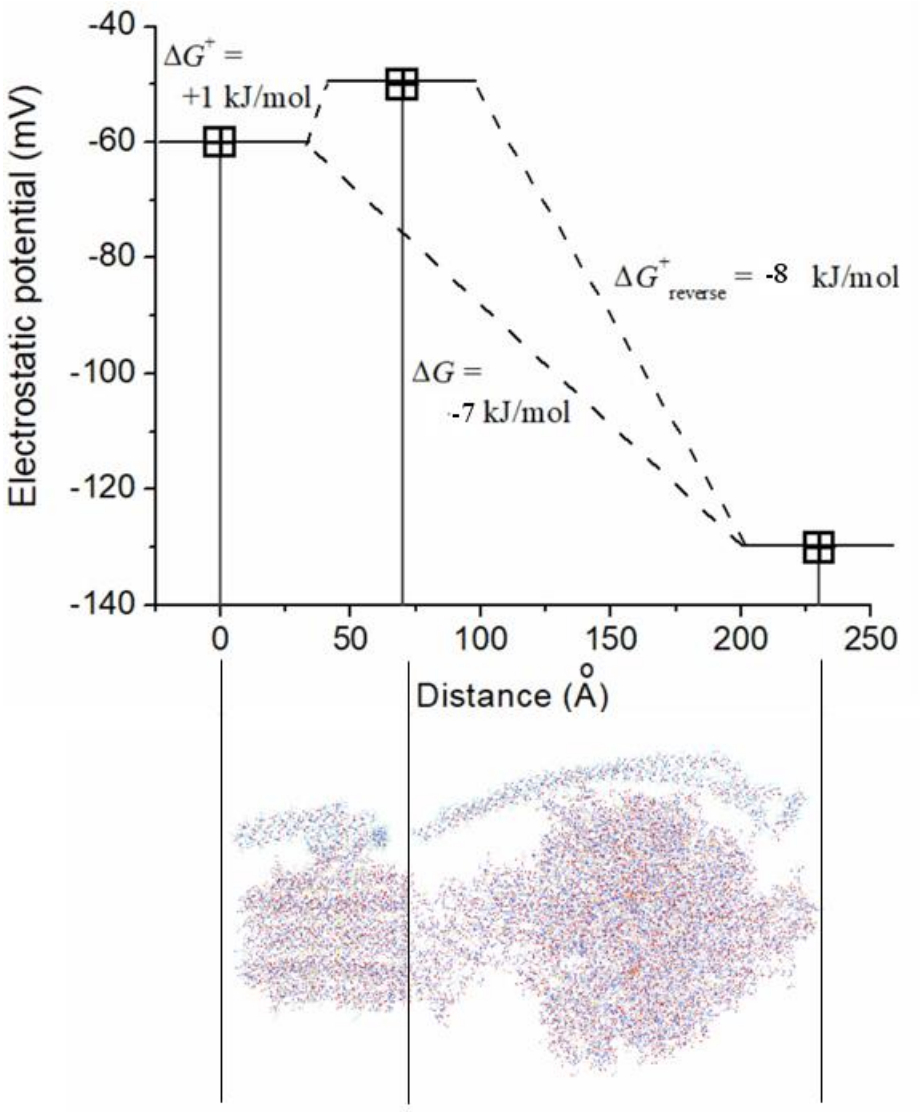


Figure 3.5 Energy profile of a proton (in kJ/mol) in aqueous medium (dielectric constant $\kappa = 76$) as the proton is translocated from the intermembrane gap (left, 0 or negative distances) through the F_0 unit of ATP synthase (from 0 to ~70 Å, the unit spanning/crossing the membrane) and ending-up in a location near the end of the protein within the mitochondrial matrix (~ 230 Å). (Matta, 2019 – reproduced with permission)

3.6 Summary

Our approximation suggests that ATP synthase possesses an intrinsic electric field as strong as 10^8 - 10^{10} $V \cdot m^{-1}$. In particular, local electric field in the F_o unit may be as strong as 10^{10} $V \cdot m^{-1}$. The local field near F_o works against the rotor motions in ATP synthase, creating energy barrier for H^+ to exit the ion channel. However, this energy penalty sustains the H^+ supply in the H^+ translocation process, and it keeps moving the translocation process forward.

A strong intrinsic electric field may lead to biological consequences to mitochondria. As mentioned in Chapter 1, it may enhance the permeability of the inner mitochondrial membrane. As the ATP synthase is embedded in the inner mitochondrial membrane, a strong field from the ATP synthase can disrupt the membrane. This suggests a possible answer for the hotly debated issue on the mitochondrial permeability pore: the protein electric field activates an electroporation process that open up mitochondrial permeability, and this is a critical question for further investigation.

In addition, the PrES program can quickly calculate the electrostatic potential and the intrinsic electric field around large protein molecules. The PrES program can be a useful tool for us to study the electrostatic properties of other protein enzymes. Although PrES is a useful tool to answer our proposed research question, its potential should be further investigated. As well, DelPhi can be used to study the electrostatics of ATP synthase, and the DelPhi results should be compared and discussed with the PrES calculations.

Chapter 4. Future Works

4.1 Applying Kernel Energy Method (KEM) to ATP Synthase

The Kernel Energy Method (KEM) is an energy partitioning method designed to predict the properties of large molecules such as ATP synthase. In the event of calculating a property of a large molecule, the KEM first cuts the molecule into a number of fragments known as "kernels", then it computes the property of each kernel, and in the end, an equation sums up the property to calculate the molecular property:

$$P_{KEM2} = \sum_{a-1}^{m-1} \sum_{b=a+1}^{m} P_{ab} - (m-2) \sum_{c=1}^{m} P_c \qquad (4.1)$$

where P_{ab} is the property calculated for double kernel ab and P_c is the property calculated for the single kernel c. The reader can refer to Appendix 2 for a description of KEM and the definition of double kernel and single kernel.

The KEM has demonstrated high quality results for large molecules such insulin,[68] tRNA,[69] graphene,[70] and ribosome.[71] In a way, the KEM is a quantum method, where quantum mechanics is applied to calculations on kernel fragments. The KEM performs well across a variety of quantum mechanical methods as DFT, Hatree-Fock, and semi-empirical AM1.[72]

The KEM has also been combined with the Quantum Theory of Atoms In Molecules (QTAIM) to study the origins of its accuracy. The author of this thesis has described the research combining KEM and QTAIM in a journal publication attached in Appendix 2.[73]

The electrostatic potential (ESP) of ATP synthase (or of any large molecule to that matter) can be obtained from KEM. KEM can approximate many QTAIM properties, for

example, electron densities at bond and ring critical points (BCPs, and RCPs) and localization-delocalization matrices of very large molecules.[74-76] KEM can also deliver response properties such as changes in properties induced by exposure to an external electric field.[77]

It has been proposed that KEM can deliver the electrostatic potential of a large molecule numerically from cubes of kernel and double kernel electron densities summed point-by-point according to Equation 4.1. This approximate KEM density can then be used with the known crystallographic structure to give the approximate electrostatic potential:

$$V_{KEM}(\mathbf{r}) = \sum_{\substack{i=1 \\ R_i \neq r}}^{M} \frac{Z_i}{|\mathbf{R_i} - \mathbf{r}|} - \int \frac{\rho_{KEM}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \qquad (4.2)$$

One can also conceive the approximation of the electrostatic potential $V_{\mathrm{KEM}}(\mathrm{r})$ from a point-by-point KEM summation according to equation 4.1 of cubes of electrostatic potential:

$$V_{KEM}(\mathbf{r}) = \sum_{a-1}^{m-1} \sum_{b=a+1}^{m} V_{ab} \; - (m-2) \sum_{c=1}^{m} V_c \qquad (4.3)$$

Huang and co-workers have demonstrated how to use the KEM to deliver QTAIM electrostatic charge for atoms.[78] In this project, the QTAIM charge produced by the KEM can replace the AMBER charges in a PrES input file. In fact, the QTAIM charge obtained by the KEM demonstrates higher accuracy.[78] A PrES calculation can be performed based on QTAIM charges to obtain the intrinsic protein electric field of ATP synthase. A major issue with this approach is that the QTAIM delivers a series of atomic multipoles to summarize the charge distribution within an atomic basin (all of which are necessary to yield accurate electrostatics),[79] and the atomic point charges delivered by QTAIM may

not be as accurate as AMBER force field charges (which are by design optimized to reproduce the protein's electrostatics).

Alternatively, KEM can re-construct electrostatic potential of ATP synthase, and then the electric field can be obtained by taking the gradient of the potential.[76] Unlike PrES calculation, this approach falls into the same category as DelPhi calculations, in which it obtains the protein electric field indirectly by differentiating pre-constructed electrostatic potential in a grid. Reader can refer to Chapter 2 for the limitations of DelPhi calculation.

As shown in the journal publication below, the accuracy of KEM relies on the cancelation of errors among the atoms and among the kernels.[73] A complication associated with this process lies in the nature of the KEM, because a large number of kernel fragments are needed to obtain KEM results for ATP synthase with high accuracy. Regardless, a combination of KEM and PrES calculation is possible and should be considered in the future, and it may further improve the accuracy of PrES calculations.

4.2 Rich Biochemistry of Electron Transport Chain

This study only looked at the electric field of mitochondrial ATP synthase from a bacteria species *P. denitrificans*. Mitochondrial ATP synthase from different biological species can exhibit (possibly major) differences in structures. For example, there are 12 cylindrical c subunits in human mitochondrial ATP synthase, whereas there are only 10 c subunits in yeast mitochondrial ATP synthase.[80] The c subunits are a critical components of the F_o unit in ATP synthase, and the average rotation speed of F_o unit varies across difference species.

Future works should compare the intrinsic electric field of ATP synthase from difference biological species. It can be expected that the F_o rotation speed may be subject to the electric field strength, because local electric field acts against the F_o rotation.

Furthermore, the inner mitochondrial membrane consists of a series of protein molecules responsible for the electron transport chain. Do these proteins also exhibit strong intrinsic electric fields? Benchmark studies of field strengths should be carried out for ETC protein complexes I - IV. Although these proteins are not as large as ATP synthase, it is still reasonable to expect that their intrinsic electric field strength is as strong as 10^7 $V \cdot m^{-1}$ based on the results presented in Chapter 3.

The molecular structure of complex I (PDB ID: 5LC5),[81] complex II (PDB ID: 1YQ3),[82] complex III (PDB ID: 3TGU),[83] and complex IV (PDB ID: 5Z62) [84] have been reported in the past. To calculate the strength of the intrinsic electric field of these protein complexes, one should refer to the PrES calculation described in Chapter 2 and Chapter 3 of this thesis.

If the electric field strength of each protein is as strong as 10^7 $V \cdot m^{-1}$, local electric field on the inner mitochondrial membrane may also lead to electroporation and disruption of the membrane structure. This may be a key in solving the mystery formation of the mitochondrial permeability transition pore.

Chapter 5. Conclusion

Adenosine 5'-tri-phosphate (ATP) is fundamental to metabolism as a universal biological "energy currency" of all living systems. The biosynthesis of ATP relies on a protein enzyme known as ATP synthase. ATP synthase is embedded in the inner mitochondrial membrane. According to Mitchell's chemiosmotic theory, the H^+ ion concentration gradient and the H^+ electrochemical gradient across the inner membrane are the two driving forces of ATP biosynthesis. The H^+ gradient is so strong that it generates an external electric field of around 10^7 V·m^{-1} to the inner mitochondrial membrane (compared to the electric field between clouds giving rise to lightning).

5.1 A New Role of Protein Intrinsic Electric Field

The ATP synthase is a large protein composed of around 64,000 atoms. In addition to the external electric field that H^+ generates, atomic charges on ATP synthase generate an intrinsic electric field (and an associated potential difference). This study investigates the intrinsic electric field of ATP synthase. Our calculations find that ATP synthase generates an intrinsic protein electric field around $10^6 - 10^8$ V·m^{-1}, which is within an order of magnitude as that of the H^+ gradient. This reveals a possible unknown role of the intrinsic electric field of ATP synthase: The intrinsic electric field may be an additional force that drives H^+ translocation. Hence, this study predicts an energetic term in addition to the conventional Mitchell's equation:

$$\Delta G = 2.3nRT\Delta\text{pH} + nF\Delta\Psi_{H^+} \quad (\pm nF\Delta\Psi_{ATP\ Synthase}) \qquad (3.3)$$

where the conventional term "$2.3nRT\Delta$pH" accounts for the energy contribution from the H^+ concentration to the total free energy, and the conventional term "$zF\Delta\psi$" accounts for

the energy contribution of the electrostatic potential difference generated from the H^+ gradient to the total free energy. Based on our results, the proposed "$nF\Delta\Psi_{ATP\ Synthase}$" contributes additional free energy that reinforces the H^+ translocation process. Overall, the proposed term adds 7 kJ/mol (of protons) of free energy to the conventional Mitchell's equation. However, this 7 kJ/mol of free energy comes with an kinetic barrier that will be discussed in the next section.

More importantly, the intrinsic electric field is a property of ATP synthase itself, and it affects the overall thermodynamics of the reaction that ATP synthase catalyzes. This study is only a starting point. In the future, it is worth to study intrinsic electric field of a wide range of protein enzymes. It can be expected that, due to the intrinsic electric field effect, protein enzymes of similar sizes (to ATP synthase) may also affect the thermodynamics and not only the kinetics of the biochemical reaction(s) it catalyzes. In particular, future studies should investigate the intrinsic electric field of the mitochondrial protein complexes I - IV, which will draw a more complete picture for the biochemical energy transduction.

5.2 An Energy Barrier in the Mitochondrial H^+ Translocation

ATP synthase operates as a motor: the rotation of the F_o unit of ATP synthase drives the H^+ translocation forward. To pass through the ATP synthase ion channel, however, a H^+ ion first encounters a local electric field that works against the F_o rotation. This local electric field raises a small energy barrier of ~ 1 kJ/mol. This is possibly due to the protonation process at the H^+ binding site, and the energy barrier regulates the H^+ translocation and ensures an appropriate level of H^+ supply. Then, the same H^+ ion

undergoes a cascade of exergonic process, releasing around 9 kJ/mol of free energy. This way, a proportion of energy from the H^+ gradient is transferred to the torque that drives the F_o rotation. Finally, as we already knew, the F_o rotation translates to the conformational changes of F_1. How the F_1 conformation synthesizes ATP is briefly introduced in section 1.2 of Chapter 1.

Previous studies by Rastogi et al. and Elston et al. have qualitatively mentioned an energy barrier in the F_o rotation and the H^+ translocation process.[11, 65] Our study further quantifies an energy barrier of ~ 1 kJ/mol by theoretical calculations, closing a knowledge gap of how H^+ drives the F_o rotation in ATP synthase. The energy barrier may play a role in regulating ATP synthesis, and further study is required to explore this role.

The protonation/deprotonation at Arg 210 and Asp 61 residues are the H^+ entrance and exit on the ATP synthase ion channel. Do these critical residues form transition states with H^+ binding at the energy barrier? Further studies can use quantum calculation to draw a more rigorous picture of how the protonation process contributes to the energy barrier.

5.3 Hot Mitochondrion

As discussed in the previous section, the local electric field works against the F_o rotation, creating an energy barrier for the H^+ translocation. This means that, when H^+ passes through ATP synthase, the local electrical force acts against the H^+ ions. An anti-parallel local electric field will lower the thermodynamic efficiency of ATP synthase – a natural molecular machine. Recent studies have revealed that a mitochondrion might be hotter than its surroundings.[85] (The reader can refer to a submitted journal article in

Appendix 3). A hotter environment can reduce the viscosity of the membrane and, hence, enhance the effect of the intrinsic electric field of the protein on electroporation (discussed below).

5.4 Electroporation and Biological Consequences

A series of large protein molecules, including ATP synthase, are embedded in the inner mitochondrial membrane. The local intrinsic electric field of these protein enzymes can disrupt the proper function of the inner-membrane phospholipid bilayer. This may have biological consequences such as a "short circuit" situation mentioned in section 1.7 of Chapter 1.

In particular, the intrinsic electric field may enhance the permeability of the inner-membrane phospholipid bilayer to protons via electroporation.[35] This suggests a possible answer for the mystery formation of the mitochondrial permeability pore. The protein electric field activates an electroporation process that open up mitochondrial permeability, and this is a critical question for further investigation. Enhanced permeability, whether or not it is a permeability pore, causes dysfunction of ATP synthase and cell programmed death (apoptosis). As mentioned above, the dysfunction can lead to a wide range of degenerative diseases such as prostate cancer and Leigh syndrome.[2, 3]

5.5 PrES is a Useful Tool in Protein Electrostatics Calculation

To reduce the computational cost of studying the protein electrostatics in this study, a new computational tool PrES is developed based on the AMBER molecular

mechanical force field. PrES calculates the electric field and electrostatic potential of proteins, yielding results within an order of magnitude.

When the PrES calculations are applied on ATP synthase protein, the magnitude of the protein intrinsic electric field and the electrostatic potential were quickly quantified. Electrostatics are important in studying properties and functions of other proteins and similar bio-molecules such as bio-polymers. In the future, PrES can serve as a general utility tool to point the general direction for researchers to study a wide range of large biomolecules.

In addition, the PrES program demonstrates certain advantages over current computational tools to study protein electrostatics. First, it reduces computational costs by directly calculating the electric field from force field charges, rather than from a pre-approximated electrostatic potential. Second, PrES excludes the grid point near or at the atomic positions. This avoids the value of the electric field or electrostatic potential to approach infinity, making the approximation more accurate.

5.6 Applying Kernel Energy Method (KEM) to ATP Synthase

The study briefly discussed the application of the Kernel Energy Method (KEM) in predicting the properties of large biomolecules such as ATP synthase. Because the KEM is a quantum method, the KEM has demonstrated high quality results for large biomolecules,[66,67,68,69] where quantum mechanics is applied to calculations on kernel fragments in these biomolecules.

KEM can yield the electrostatic potential of ATP synthase, as well as many other large biomolecules. For example, electron densities at bond and ring critical points very

large biomolecules can be predicted by KEM.[72-74] KEM can also deliver response properties such as changes in properties induced by exposure to an external electric field.[75]

As mentioned in Chapter 2, the electric charges delivered by the AMBER force field raises concerns on the accuracy of PrES calculation, especially when calculating electrostatic potential of grid points far away from a molecule. In the case of ATP synthase, KEM may be combined with the Quantum Theory of Atoms In Molecules (QTAIM) to deliver highly accurate electric charges.[76] In other words, the QTAIM charges produced by the KEM can replace the AMBER charges in a PrES input file.

A PrES calculation can be performed based on QTAIM charges to obtain the intrinsic protein electric field of ATP synthase. A major challenge, however, is that the atomic charge distribution provided by QTAIM are not optimal for reproducing the electrostatic potential, being only the first term in a multipolar expansion,[77] and the atomic point charges delivered by QTAIM may not be as accurate as AMBER force field charges. Future calculations should investigate if additional QTAIM multipoles could be introduced (e.g. atomic dipolar polarization in addition to the point charges) for a more accurate reproduction of macromolecular electrostatics using an extension of the PrES programme.

Alternatively, electrostatic potential of ATP synthase can be re-constructed by KEM, and then the electric field can be obtained by taking the gradient of the potential.[74] However, this approach of obtaining the protein electric field by differentiating the pre-constructed electrostatic potential in a grid can have many numerical issues and may be too inaccurate.

References

[1] Mathews, C.K., van Holde, K.E., Appling, D.R. and Anthony-Cahill, S.J. (2013). *Biochemistry, 4th ed.* Pearson, Toronto Canada. Ch. 15, pp.625-649.

[2] Mitchell, P. (1961). *Nature*, *191*(4784), pp.144-148.

[3] Mitchell, P. and Moyle, J. (1965). *Nature*, *208*(5006), pp.147-151.

[4] Mitchell, P. (2011). *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1807*(12), pp.1507-1538.

[5] Nicholls, D. G. (1982). *Bioenergetics: An Introduction to the Chemiosmotic Theory*.

[6] Arnold, R.S., Sun, C.Q., Richards, J.C., Grigoriev, G., Coleman, I.M., Nelson, P.S., Hsieh, C.L., Lee, J.K., Xu, Z., Rogatko, A. and Osunkoya, A.O. (2009). *The Prostate*, *69*(1), pp.1-11.

[7] Sgarbi, G., Baracca, A., Lenaz, G., Valentino, L. M., Carelli, V. and Solaini, G. (2006). *Biochemical Journal*, *395*(3), pp.493-500.

[8] OpenStax CNX. (Online). Oxidative Phosphorylation. *Biology*. cnx.org/contents/GFy_h8cu@9.85:7oTVAgrZ@7/Oxidative-Phosphorylation (accessed on May 13, 2019. CC BY 4.0)

[9] Garrett, R. and Grisham, C. (2010). *Biochemistry*,*4th ed.* Brooks Cole Cengage Learning, Boston USA. Ch. 20, pp.592-661

[10] Walker, J. E. (Online). *Subunit Composition of ATP Synthase.* www.mrc-mbu.cam.ac.uk/projects/2679/subunit-composition (accessed on Jan 31, 2019).

[11] Walker, J. E. (Online). *The Structure and Function of ATP Synthases.* www.mrc-mbu.cam.ac.uk/projects/2680/structure-and-function-atp-synthases (accessed on Feb 1, 2019)

[12] Capaldi, R.A. and Aggeler, R. (2002). *Trends in Biochemical Sciences*, *27*(3), pp.154-160.

[13] Von Ballmoos, C., Wiedenmann, A. and Dimroth, P. (2009). *Annual Review of Biochemistry*, *78*, pp.649-672.

[14] Rastogi, V. K. and Girvin, M. E. (1999). *Nature*, *402*(6759), pp.263-268.

[15] Iko, Y., Tabata, K. V., Sakakihara, S., Nakashima, T. and Noji, H. (2009). *FEBS Letters*, *583*(19), pp.3187-3191.

[16] Mnatsakanyan, N., Hook, J. A., Quisenberry, L. and Weber, J. (2009*). Journal of Biological Chemistry, 284*(39), pp.26519-26525.

[17] Klusch, N., Murphy, B. J., Mills, D. J., Yildiz, Ö. and Kühlbrandt, W. (2017). *Elife*, *6*, e33274.

[18] Miller Jr, J. H., Rajapakshe, K. I., Infante, H. L. and Claycomb, J. R. (2013). *PLoS One*, *8*(9), e74978.

[19] Quintanilha, A. T. and Packer, L. (1977). *FEBS Letters*, *78*(2), pp.161-165.

[20] Maxwell, J. C. (1872). *Theory of Heat.* Longmans, Green, and Co., London.

[21] Landauer, R. (1961). *IBM Journal of Research and Development*, *5*(3), pp.183-191.

[22] Matta, C. F. and Massa, L. (2015). *Biochemistry*, *54*(34), pp. 5376-5378.

[23] Matta, C.F. and Massa, L., (2017). *The Journal of Physical Chemistry A*, *121*(47), pp.9131-9135.

[24] Zabusky, N. J. and Deem, G. S. (1979). *Biophysical Journal*, *25*(1), pp.1-15.

[25] Song, D. H., Park, J., Maurer, L. L., Lu, W., Philbert, M. A. and Sastry, A. M. (2013). *Physical Review E*, *88*(6), pp.062723.

[26] Franzen, S., Goldstein, R. F. and Boxer, S. G. (1990). *Journal of Physical Chemistry*, *94*(12), pp.5135-5149.

[27] Lockhart, D. J., Kirmaier, C., Holten, D. and Boxer, S. G. (1990). *Journal of Physical Chemistry*, *94*(18), pp.6987-6995.

[28] Franzen, S. and Boxer, S. G. (1993). *The Journal of Physical Chemistry*, *97*(23), pp.6304-6318.

[29] Suydam, I. T. and Boxer, S. G. (2003). *Biochemistry*, *42*(41), pp.12050-12055.

[30] Ly, H.K., Wisitruangsakul, N., Sezer, M., Feng, J.J., Kranich, A., Weidinger, I.M., Zebger, I., Murgida, D.H. and Hildebrandt, P. (2011). *Journal of Electroanalytical Chemistry*, *660*(2), pp.367-376.

[31] Murgida, D.H. and Hildebrandt, P. (2001). *The Journal of Physical Chemistry B*, *105*(8), pp.1578-1586.

[32] Nussbaum, J. H. and Grodzinsky, A. J. (1981). *Journal of Membrane Science*, *8*(2), pp.193-219.

[33] Seyedi, S. and Matyushov, D. V. (2018). *Chemical Physics Letters*, *713*, pp.210-214.

[34] Matta, C. F. (2015). *Personal Communication.* (unpublished results)

[35] Morgenstern, A., Jaszai, M., Eberhart, M. E. and Alexandrova, A. N. (2017). *Chemical Science*, *8*(7), pp.5010-5018.

[36] Prah, A., Frančišković, E., Mavri, J. and Stare, J. (2019). *ACS Catalysis*, *9*(2), pp.1231-1240.

[37] Kulik, H. J. (2018). *Physical Chemistry Chemical Physics*, *20*(31), pp.20650-20660.

[38] Napotnik, T. B., Wu, Y. H., Gundersen, M. A., Miklavčič, D. and Vernier, P. T. (2012). *Bioelectromagnetics*, *33*(3), pp.257-264.

[39] Bonora, M., Bononi, A., De Marchi, E., Giorgi, C., Lebiedzinska, M., Marchi, S., Patergnani, S., Rimessi, A., Suski, J.M., Wojtala, A. and Wieckowski, M.R. (2013). *Cell Cycle*, *12*(4), pp.674-683.

[40] Alavian, K.N., Beutner, G., Lazrove, E., Sacchetti, S., Park, H.A., Licznerski, P., Li, H., Nabili, P., Hockensmith, K., Graham, M. and Porter, G.A. (2014). *Proceedings of the National Academy of Sciences*, *111*(29), pp.10580-10585.

[41] Neginskaya, M. A., Solesio, M. E., Berezhnaya, E. V., Amodeo, G. F., Mnatsakanyan, N., Jonas, E. A. and Pavlov, E. V. (2019). *Cell Reports*, *26*(1), pp.11-17 (CC BY 4.0)

[42] Cramer, C. J. (2003). *Essentials of Computational Chemistry.* John Wiley & Sons, Ltd. Ch. 4, pp.106-107; Ch. 8, pp.252-297.

[43] Matta, C. F. (2002). *Applications of the Quantum Theory of Atoms in Molecules to Chemical and Biochemical Problems.* McMaster University. Ch. 1, pp.1-27.

[44] McQuarrie, D. A. (2008). *Quantum Chemistry*. University Science Books. Ch. 3, pp.97; Ch. 7, pp. 359-360; Ch. 8, pp.381-403.

[45] Gatti, C., MacDougall, P. J. and Bader, R. F. (1988). *The Journal of Chemical Physics*, *88*(6), pp.3792-3804.

[46] David, C. Y. (2001). *Computational Chemistry.* John Wiley & Sons, Inc. Ch. 6, pp.49-66.

[47] Doucet, J. P. and Weber, J. (1996). *Computer-aided Molecular Design: Theory and Applications*. Academic Press. Ch. 5, pp.149-151.

[48] Morales-Rios, E., Montgomery, M. G., Leslie, A. G. and Walker, J. E. (2015). *Proceedings of the National Academy of Sciences*, *112*(43), pp.13231-13236.

[49] Kaminský, J. and Jensen, F. (2007). *Journal of Chemical Theory and Computation*, *3*(5), pp.1774-1788.

[50] Patel, S., Mackerell Jr, A. D. and Brooks III, C. L. (2004). *Journal of Computational Chemistry*, *25*(12), pp.1504-1514.

[51] Vanommeslaeghe, K., Raman, E. P. and MacKerell Jr, A. D. (2012). *Journal of Chemical Information and Modeling*, *52*(12), pp.3155-3168.

[52] Laberge, M., Vanderkooi, J.M. and Sharp, K.A. (1996). *The Journal of Physical Chemistry*, *100*(25), pp.10793-10801.

[53] Laberge, M. (1998). *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1386*(2), pp.305-330.

[54] Kim, E., Sharp, A. and Nicholls, A. (Online) *DelPhi v. 4 - The New Macromolecular Electrostatics Modeling Package*. pp.6-15 (accessed August 20, 2019).

[55] Li, L., Chakravorty, A. and Alexov, E. (2017). *Journal of Computational Chemistry*, *38*(9), pp.584-593.

[56] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995). *Journal of the American Chemical Society*, *117*(19), pp.5179-5197.

[57] Marcotte, I., Separovic, F., Auger, M. and Gagné, S.M., (2004). *Biophysical Journal*, 86(3), pp.1587-1600.

[58] Zabusky, N. J. and Deem, G. S. (1979). *Biophysical Journal*, *25*(1), pp.1-15.

[59] Laberge, M. (1998). *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1386*(2), pp.305-330.

[60] Acosta-Gutierrez, S., Scorciapino, M. A., Bodrenko, I. and Ceccarelli, M. (2015). *The Journal of Physical Chemistry Letters*, *6*(10), pp.1807-1812.

[61] Patthy, L. and Thész, J. (1980). *European Journal of Biochemistry*, *105*(2), pp. 387-393.

[62] Morgenstern, A., Jaszai, M., Eberhart, M. E. and Alexandrova, A. N. (2017). *Chemical Science*, *8*(7), pp.5010-5018.

[63] Prah, A., Frančišković, E., Mavri, J. and Stare, J. (2019). *ACS Catalysis*, *9*(2), pp.1231-1240.

[64] Dudev, T., Ilieva, S. and Doudeva, L. (2018). *Physical Chemistry Chemical Physics*, *20*(38), pp.24633-24640.

[65] Seeliger, D. and de Groot, B. L. (2010). *Journal of Computer-aided Molecular Design*, *24*(5), pp.417-422.

[66] Lerner, M. G. and Carlson, H. A. (2006). *APBS Plugin for PyMOL.* University of Michigan, Ann Arbor.

[67] Elston, T., Wang, H. and Oster, G. (1998). *Nature*, *391*(6666), pp.510-513.

[68] Huang, L., Massa, L. and Karle, J. (2005). *Proceedings of the National Academy of Sciences*, *102*(36), pp.12690-12693.

[69] Huang, L., Massa, L. and Karle, J. (2006). *Proceedings of the National Academy of Sciences*, *103*(5), pp.1233-1237.

[70] Huang, L., Bohorquez, H. J., Matta, C. F. and Massa, L. (2011). *International Journal of Quantum Chemistry*, *111*(15), pp.4150-4157.

[71] Huang, L., Krupkin, M., Bashan, A., Yonath, A. and Massa, L. (2013). *Proceedings of the National Academy of Sciences*, *110*(37), pp.14900-14905.

[72] Huang, L., Massa, L. and Karle, J. (2006). *International Journal of Quantum Chemistry*, *106*(2), pp.447-457.

[73] Massa, L., Keith, T., Cheng, Y. and Matta, C. F. (2019). *Chemical Physics Letters*, *734*, pp.136650.

[74] Matta, C. F. (2018). *Computational and Theoretical Chemistry*. *1124*, pp.1-14.

[75] Matta, C. F. (2014). *Journal of Computational Chemistry* 35, pp.1165-1198.

[76] Timm, M.J., Matta, C.F., Massa, L. and Huang, L. (2014). *The Journal of Physical Chemistry A*, *118*(47), pp.11304-11316.

[77] Huang, L., Massa, L. and Matta, C.F. (2014). *Carbon*, *76*, pp.310-320.

[78] Huang, L., Matta, C. and Massa, L. (2015). *Structural Chemistry*, *26*(5-6), pp.1433-1442.

[79] Polkosnik, W., Matta, C. F., Huang, L. and Massa, L. *International Journal of Quantum Chemistry*, e25986.

[80] Von Ballmoos, C., Wiedenmann, A. and Dimroth, P. (2009). *Annual Review of Biochemistry*, *78*, pp.649-672.

[81] Zhu, J., Vinothkumar, K. R. and Hirst, J. (2016). *Nature*, *536*(7616), pp.354-358.

[82] Huang, L.S., Sun, G., Cobessi, D., Wang, A.C., Shen, J.T., Tung, E.Y., Anderson, V.E. and Berry, E.A., (2006). *Journal of Biological Chemistry*, *281*(9), pp.5965-5972.

[83] Hao, G.F., Wang, F., Li, H., Zhu, X.L., Yang, W.C., Huang, L.S., Wu, J.W., Berry, E.A. and Yang, G.F., (2012). *Journal of the American Chemical Society*, *134*(27), pp.11168-11176.

[84] Zong, S., Wu, M., Gu, J., Liu, T., Guo, R. and Yang, M. (2018). *Cell Research*, *28*(10), pp.1026-1034.

[85] Fahimi, P., Nasr, M. A., Castanedo, L. A. M., Cheng, Y., Toussi, C. A. and Matta, C.F. (2019). *Biophysical Bulletin.* (submitted)

Appendix 1. Supplementary information: the Code of PrES Program (Author: Cheng, Y., Tannahill, C., Muir, P., Matta, C. F.)

This code calculates the electric field across a grid across the protein molecule. The grid is automatically generated by this program. The code is written is Python.

```
headline = '''
Input file must have 5 columns with the following:
column 1) The atom symbol
column 2) The atom X coordinate
column 3) The atom Y coordinate
column 4) The atom Z coordinate
column 5) The atom's AMBER charge
'''
import sys
import time
import math
import numpy as np
import itertools



# Here are the constants
epsilon0 = 8.8541878176e-12 #permittivity of free space
e = 1.60217733e-19
coeff_const = 1 / (4 * math.pi * epsilon0 )
meters_per_angstrom = 1e-10
NA = 6.022140857e+23 #Avogadro's number

INF  = 100000000000000000
NINF = -100000000000000000



def parseLine(line):
    split = line.strip().split(",")
    parsed_variables = [split[0]]
    for i in range(1, len(split)):
        column = split[i]
        if column:
            parsed_variables.append(float(column))
        else:
            parsed_variables.append(0)
    return parsed_variables
```

```
def getBoundingBox(in_data):
    x_min, x_max = INF, NINF
    y_min, y_max = INF, NINF
    z_min, z_max = INF, NINF


    # x
    x_min = np.amin(in_data[:,0])
    x_max = np.amax(in_data[:,0])

    # y
    y_min = np.amin(in_data[:,1])
    y_max = np.amax(in_data[:,1])

    # z
    z_min = np.amin(in_data[:,2])
    z_max = np.amax(in_data[:,2])

    boundingBox = {
        'x_min': x_min,
        'x_max': x_max,
        'y_min': y_min,
        'y_max': y_max,
        'z_min': z_min,
        'z_max': z_max
    }

    return boundingBox

def readInFile(input_file_name):
    # Get the number of rows in the input file.
    num_lines = sum(1 for line in open(input_file_name, 'rt'))

    # Create array of appropriate size
    data = np.zeros(shape=(num_lines, 4))

    with open(input_file_name, 'rt') as f:
        for i, line in enumerate(f.readlines()):
            _, x, y, z, amber = parseLine(line)
            data[i,:] = x*meters_per_angstrom, y*meters_per_angstrom,
z*meters_per_angstrom, amber
```

```
    return data


def getUserXYZ():
    #ask users to specify the number of grid on axis
    print("Specify number of grid elements in each direction")
    nx = input("nx = ")
    ny = input("ny = ")
    nz = input("nz = ")

    #calculate distance R between the specified point and each atom
    return (int(nx), int(ny), int(nz))


def generateGrid(nx, ny, nz, boundingBox):

    # Greate grid generating iterator.
    gridIter = itertools.product(
        np.linspace(boundingBox['x_min'], boundingBox['x_max'], nx), \
        np.linspace(boundingBox["y_min"], boundingBox["y_max"], ny), \
        np.linspace(boundingBox["z_min"], boundingBox["z_max"], nz))

    return gridIter


'''
pointEfield function computes the Efield contribution from each atom
args:
point: an (x, y, z) tuple representing a point in the bounding Box
in_data: an array of the xyz coordinate and amber charges
the function returns the Efield (Ex, Ey, Ez) contributed by each atom
'''

def pointEfield(point, in_data):

    print('point {}'.format(point))

    # Take slices of the array.
    x = in_data[:,0]
    y = in_data[:,1]
    z = in_data[:,2]
```

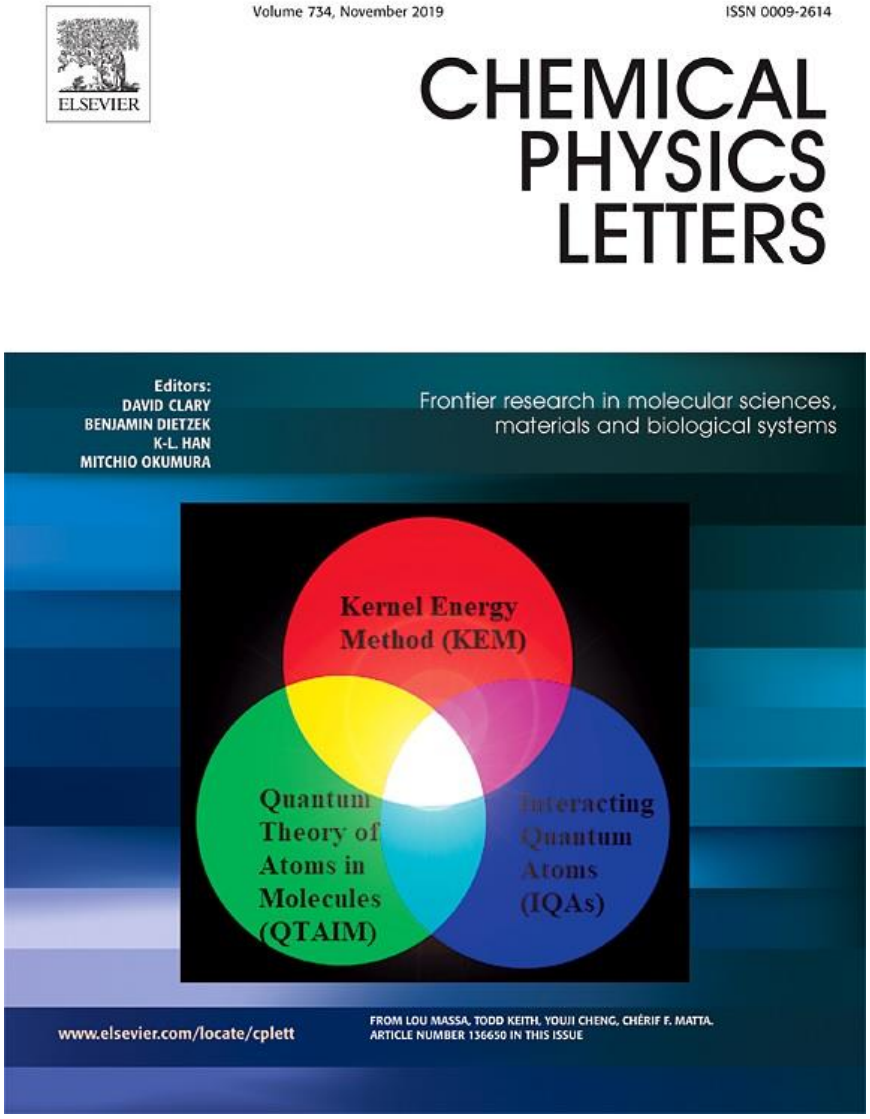
```
    amber = in_data[:,3]


    # Scaling by e to create charges
    charge =  e * amber

    # Take the L2 norm elementwise
        R = np.sqrt(np.square(x) + np.square(y) + np.square(z))

    # The grid point position should not be equal to the atomic position

    # Sum of electric field contributed by all atomic charges
    Ex = coeff_const * np.sum((charge * (x - point[0]))/(R**3))
    Ey = coeff_const * np.sum((charge * (y - point[1]))/(R**3))
    Ez = coeff_const * np.sum((charge * (z - point[2]))/(R**3))

    return (Ex, Ey, Ez)


def computeEfield(input_file_name, output_file_name):


    #store data in array "in_data"
    in_data = readInFile(input_file_name)
    output_file = open(output_file_name, 'w')
    boundingBox = getBoundingBox(in_data)
    for k, v in boundingBox.items():
        print(k, v)

    #generate grid
    nx, ny, nz = getUserXYZ()
    print("genering the grid")
    start_time = time.time()
    grid  = generateGrid(nx, ny, nz, boundingBox)
    print("done")

    print("Going through the grid points")
    # modify amber with charge
    # in_data[:,3] = e*in_data[:,3]
    # threshold to eliminate grid point too close to nuclei is 10^-3 or 0.001
    for gridPoint in grid:
        Ex, Ey, Ez = pointEfield(gridPoint, in_data)
        output_file.write(str(Ex) + "   " + str(Ey) + "   " + str(Ez) + '\n')
```

```
    output_file.close()

    print('--- time {0} ---'.format(time.time() - start_time))


def printHelp():
    print(headline)
    print("Please provide the name of an input file on the commandline")
    print("and also provide an output file name.")
    print("")
    print("Example:")
    print("python main.py <myinputfile.csv> <myoutfile.csv>")
    print("")


# Program entry point.
if __name__ == '__main__':

    # Make sure user provided input file.
    if len(sys.argv) != 3:
        printHelp()
        sys.exit()


    input_file_name = sys.argv[1]
    output_file_name = sys.argv[2]
    print("")
    print("Input file: " + input_file_name)
    print("Output file: " + output_file_name)
    print("")

    computeEfield(input_file_name, output_file_name)
    print("check output file to find (Ex, Ey, Ez) in SI unit")
```

Appendix 2. Applying Kernel Energy Method to ATP Synthase

Massa, L., Keith, T., Cheng, Y. and Matta, C. F. (2019). *Chem. Phys. Lett.*, 734, 136650. (Reference [71] – reproduced with permission).

Youji Cheng contributed to the Gaussian calculations, preparation of the figure, and was the presenter of a contributed poster: Cheng, Y.; Keith, T. A.; Massa, L.; Matta, C. F. "Interpretation of the Kernel Energy Method (KEM) using the Theory of Interacting Quantum Atoms (IQA)", International Union of Crystallography (IUCr)'s Sagamore XIX Conference on Quantum Crystallography (QCr) (Sagamore - 2018), Halifax, NS, Canada (8-13 July 2018).



Chemical Physics Letters 734 (2019) 136650

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Research paper

The kernel energy method applied to quantum theory of atoms in molecules – energies of interacting quantum atoms

Lou Massa[a], Todd Keith[b], Youji Cheng[c,d], Chérif F. Matta[c,d,e,f,*]

[a] Hunter College and the Graduate School, City University of New York, New York, NY 10065, USA
[b] Semichem, Inc., Shawnee, KS 66216, USA
[c] Dept. of Chemistry & Physics, Mount Saint Vincent University, Halifax, NS B3M 2J6, Canada
[d] Dept. of Chemistry, Saint Mary's University, Halifax, NS B3H 3C3, Canada
[e] Dept. of Chemistry, Dalhousie University, Halifax, NS B3H 4J3, Canada
[f] Dép. de Chimie, Université Laval, Québec, QC G1V 0A6, Canada

HIGHLIGHTS

- The accuracy of the "kernel energy method" is due to an atom-level error cancellation.
- The energies of the kernels can be an order of magnitude more accurate than those of the composing atoms.
- IQA provides an insightful decomposition of the errors of the KEM approximation.

ARTICLE INFO

Keywords:
Interacting quantum atoms (IQA)
Quantum theory of atoms in molecules (QTAIM)
Kernel energy method (KEM)
Molecular fragmentation for linear scaling quantum chemistry
Quantum density matrices

ABSTRACT

The Kernel Energy Method (KEM) has been widely applied to predict the properties of large molecules from kernels fragments in a fast and reliable manner. Kernels are typically made of many atoms raising the question: *Is the accuracy of KEM traceable to the atomic level?* An exact atomic decomposition of the molecular energy within the Quantum Theory of Atoms in Molecules (QTAIM), termed "Interacting Quantum Atoms (IQA)", performed on a triglycyl model demonstrates that this is *not* the case. The kernel energies are found to be up to an order of magnitude more accurate than the underlying IQA atomic energies. The accuracy of KEM energies is, thus, the result of a subtle cancellation of errors among the different atoms composing the kernels.

1. Introduction

The Kernel Energy Method (KEM) is a many-body expansion-based fragmentation approximation meant to compute the properties of large molecules relatively quickly and accurately. KEM can be applied to individual physical properties or to density matrices to approximate their *ab initio* counterparts within a given model chemistry [1–6]. The method has been shown to approximate properties of several large molecules accurately [4,7–13] including difficult cases such as extended aromatics [14,15] even when an external intense homogenous static electric field is applied [16].

KEM's speed compared to direct full-molecule calculation is due to its requiring the application of quantum chemical calculations only on the defined kernels (fragments) and kernel combinations (just kernel pairs in the most common case of 2nd order KEM), hence leading to a reduction of the scaling from the *n*th power of a large number (the full molecule calculation) to the sum of the *n*th powers of much smaller numbers. Further gains in speed can result from the fact that calculations on separate kernels and kernel combinations can easily be run on separate processors or nodes, for a naturally parallelizable method.

KEM has been reviewed numerous times [1–17], here we remind the reader with the basic tenets of KEM to 2nd order (KEM2). The KEM2 approximation of property P (scalar or vector-component, intensive or extensive) for a large molecule is written as,

$$P_{\mathrm{KEM2}} = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} P_{ab} - (m-2) \sum_{c=1}^{m} P_c \tag{1}$$

where P_{ab} is the property calculated for double kernel ab and P_c is the property calculated for the single kernel c.

The accuracy of the KEM2 approximation suggests a corresponding accuracy for the underlying density matrices, hence one may write [7]:

* Corresponding author.
E-mail address: cherif.matta@msvu.ca (C.F. Matta).

https://doi.org/10.1016/j.cplett.2019.136650
Received 15 July 2019; Received in revised form 31 July 2019; Accepted 31 July 2019
Available online 03 August 2019
0009-2614/

$$\rho_2 = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{2(ab)} - (m-2) \sum_{c=1}^{m} \rho_{2(c)} \quad (2)$$

$$\rho_1 = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{1(ab)} - (m-2) \sum_{c=1}^{m} \rho_{1(c)} \quad (3)$$

$$\rho = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} \rho_{(ab)} - (m-2) \sum_{c=1}^{m} \rho_{(c)} \quad (4)$$

whereby ρ_2 and ρ_1 are the second- and first-order reduced density matrices and ρ the electron density. With ρ_2, ρ_1, and ρ, all quantum properties of interest of the large molecule can be calculated from the kernels. Further, by imposing *N*-representability, the resulting energies can be made to satisfy the variational theorem [18].

KEM2 usually works well because it accounts for the most important contributions to the properties of a system from a many-body expansion standpoint: the unperturbed fragment properties and their changes due to pair-wise interactions with other fragments.

An important question is: to what extent is KEM2 accurate at the atomic level, i.e., for the atoms in a molecule? Can the accuracy of KEM2 as a computational approximation for a molecule be traced to the accuracy of atoms composing the molecule? To answer this question one needs to define atoms in molecules and their properties. Within Bader's Quantum Theory of Atoms in Molecules (QTAIM) [19–21], a rigorous partitioning of the total energy into atomic self- and diatomic interaction-terms is termed the Interacting Quantum Atoms (IQA) approach, pioneered by Martín Pendás and coworkers [22–26]. In this paper, we focus on this partitioning of the total energy in terms of IQAs to determine the accuracy of KEM2 applied on an atom-by-atom basis.

IQA typically uses the same definition of atoms in molecules as QTAIM (*i.e.*, bounded by zero-flux surfaces in the electron density gradient), and partitions the molecular energy into atomic self-energies and diatomic interaction energies. Thus the total energy *E* is written as:

$$E = \sum_{\mathrm{A}} E^{\mathrm{A}}_{self} + \frac{1}{2} \sum_{\mathrm{A}} \sum_{\mathrm{B} \neq \mathrm{A}} E^{\mathrm{AB}}_{int} \quad (5)$$

where, an atom A's self-energy (also known as intra-atomic energy) contains all of the electronic kinetic energy within the atomic basin, the interaction energies between the electrons within this atom and its nucleus, and the electron-electron interaction within that atom, *i.e.*:

$$E^{\mathrm{A}}_{self} = T^{\mathrm{A}} + V^{\mathrm{AA}}_{en} + V^{\mathrm{AA}}_{ee} \quad (6)$$

where the interaction energy between a pair of atoms A and B in the molecule contains the remaining energy terms:

$$E^{\mathrm{AB}}_{int} = V^{\mathrm{AB}}_{nn} + V^{\mathrm{AB}}_{en} + V^{\mathrm{AB}}_{ne} + V^{\mathrm{AB}}_{ee}, (\mathrm{A} \neq \mathrm{B}) \quad (7)$$

where the notation is clarified by the example V^{AB}_{ne} which denotes the potential energy of interaction between the nucleus (subscripted *n*) of atom A (superscripted) and the electrons (subscripted *e*) contained in the basin of atom B.

This IQA partitioning satisfies the following summation properties yielding an "*additive atomic energy*" of the atom (say atom A) as "self" + 1/2 of interaction terms:

$$E^{\mathrm{A}}_{add} = E^{\mathrm{A}}_{self} + \frac{1}{2} \sum_{\mathrm{B} \neq \mathrm{A}} E^{\mathrm{AB}}_{int} \quad (8)$$

the sum of which is *E*, the total energy of the molecule. IQA thus expresses the total energy of the molecule in terms of atomic self-energy terms and a *complete network* of interaction channels between every pair of atoms in the molecule. The set of IQA self and interatomic energies can be expressed as a symmetric matrix in which the sum of any row or corresponding column is an additive IQA atomic energy and the sum of all matrix elements is the total energy [27,28].

In this paper, a simple tripeptide, triglycine, will be used as a test case to investigate the accuracy of KEM2 for molecular energy as well as the additive IQA atomic energies.

2. Method

The energy of triglycine Gly_1-Gly_2-Gly_3 in both the direct and the KEM2 calculation is calculated in terms of atomic contributions with QTAIM and IQA [22–26] to see how accurate KEM2 is for atoms in molecules in comparison to the molecule as a whole. Following Eq. (1), a second-order Kernel Energy Method (KEM2) delivers the following approximate total energy of triglycine:

$$E_{\mathrm{KEM2}} = E(\mathrm{DK}_{12}) + E(\mathrm{DK}_{13}) + E(\mathrm{DK}_{23}) - E(\mathrm{SK}_1) - E(\mathrm{SK}_2) - E(\mathrm{SK}_3) \quad (9)$$

in which DK and SK refer to double- and single-kernels, respectively.

Meanwhile, the energy decomposition according to IQA (Eq. (8)) is calculated from the following explicit expression:

$$E^{\mathrm{A}}_{add} = \underbrace{T^{\mathrm{A}} + V^{\mathrm{AA}}_{en} + V^{\mathrm{AA}}_{ee}}_{E^{\mathrm{A}}_{self}} + \frac{1}{2} \underbrace{\left(\sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{en} + \sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{ne} + \sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{ee} + \sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{nn} \right)}_{\sum_{\mathrm{B} \neq \mathrm{A}} E^{\mathrm{AB}}_{int}} \quad (10)$$

that provides the total IQA energy of an atom in a molecule (which is additive in the sense that their sum yields the molecular total energy, exactly), where:

$$E^{\mathrm{A}}_{self} \equiv T^{\mathrm{A}} + V^{\mathrm{AA}}_{en} + \underbrace{V^{\mathrm{AA}}_{ee\ Coul.} + V^{\mathrm{AA}}_{ee\ X}}_{V^{\mathrm{AA}}_{ee}} \quad (11)$$

and where

$$\sum_{\mathrm{B} \neq \mathrm{A}} E^{\mathrm{AB}}_{int} \equiv \sum_{\mathrm{B} \neq \mathrm{A}} (V^{\mathrm{AB}}_{en} + V^{\mathrm{AB}}_{ne} + \underbrace{V^{\mathrm{AB}}_{ee\ Coul.} + V^{\mathrm{AB}}_{ee\ X}}_{V^{\mathrm{AB}}_{ee}} + V^{\mathrm{AB}}_{nn}) \quad (12)$$

Through an appropriate grouping of terms, the expression above for the additive IQA energies can be re-expressed as:

$$E^{\mathrm{A}}_{add} \equiv E^{\mathrm{A}}_{self} + \frac{1}{2} \sum_{\mathrm{B} \neq \mathrm{A}} E^{\mathrm{AB}}_{int} = T^{\mathrm{A}} + V^{\mathrm{A}}_{ee\ Coul.} + V^{\mathrm{A}}_{ee\ X} \quad (13)$$

where $V^{\mathrm{A}}_{Coul.}$ is the total Coulombic potential energy contribution of a given atom to the molecular total energy defined as:

$$V^{\mathrm{A}}_{Coul.} \equiv V^{\mathrm{AA}}_{ee\ Coul.} + \frac{1}{2} (V^{\mathrm{A,Mol.}}_{en\ Coul.} + V^{\mathrm{A,Mol.}}_{ne\ Coul.} + V^{\mathrm{A,Mol.}}_{nn\ Coul.}) \quad (14)$$

where $V^{\mathrm{AA}}_{ee\ Coul.}$ is the two-electron Coulomb energy contribution of atom A defined as:

$$V^{\mathrm{A}}_{ee\ Coul.} \equiv V^{\mathrm{AA}}_{ee\ Coul.} + \frac{1}{2} \sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{ee\ Coul.} \quad (15)$$

and where $V^{\mathrm{A}}_{ee\ X}$ is the total exchange energy contribution of atom A to the molecular total energy, defined as:

$$V^{\mathrm{A}}_{ee\ X} \equiv V^{\mathrm{AA}}_{ee\ X} + \frac{1}{2} \sum_{\mathrm{B} \neq \mathrm{A}} V^{\mathrm{AB}}_{ee\ X} \quad (16)$$

Quantum chemical calculations (geometry optimization and generation of the wavefunctions) were performed at the Hartree-Fock (HF) level using a 6-311G(*d*,*p*) basis set, a level of theory denoted as HF/6-311G(*d*,*p*)//HF/6-311G(*d*,*p*). Electronic structure calculations and geometry optimizations were performed using the Gaussian software [29]. The resulting wavefunctions were then subjected to QTAIM and IQA calculations using the AIMAll package [30].

3. Results

The KEM2 error in any given property term *P* is defined by the difference of that value obtained from a calculation on the full molecule and that estimated from the KEM formula (Eq. (1)) as:

$$\Delta P = P(\mathrm{KEM2}) - P(\mathrm{exact}) \quad (17)$$

So, for example, the KEM2 errors in the additive IQA atomic energies are defined as:

Table 1
KEM2 errors (ΔP) in atomic energy contributions (in kcal/mol) and charges (in atomic units (a.u.)). See Fig. 1 for atom locations and Eqs. (10)–(17) for the definition of the various quantities listed in this table.

Object	ΔE^A_{add}	ΔT^A	$\Delta V^A_{ee\ Coul.}$	$\Delta V^A_{ee\ X}$	ΔE^A_{self}	ΔE^A_{int}	Δq^A(1)
O1	−0.22	−0.94	0.31	0.41	−0.41	0.19	0.0014
O2	−1.77	−1.10	−1.49	0.82	−0.26	−1.52	0.0029
O3	2.28	−0.34	2.29	0.32	−0.19	2.47	0.0011
O4	1.79	0.95	1.45	−0.60	0.43	1.37	−0.0022
N5	−0.04	−0.65	−0.10	0.71	−1.54	1.49	0.0034
N6	5.39	1.29	3.91	0.19	0.56	4.83	0.0013
N7	−1.31	−0.91	−0.92	0.52	−1.06	−0.25	0.0022
C8	0.09	0.71	−0.32	−0.30	−0.29	0.38	−0.0007
C9	−2.64	2.32	−3.03	−1.93	−2.45	−0.20	−0.0056
C10	−0.92	−1.05	−0.97	1.11	0.65	−1.57	0.0039
C11	2.75	−1.55	3.39	0.91	1.51	1.24	0.0031
C12	0.79	−0.06	0.81	0.04	−0.32	1.10	−0.0001
C13	−3.28	1.34	−3.69	−0.94	−1.60	−1.68	−0.0034
H14	0.03	0.32	−0.12	−0.17	0.15	−0.13	−0.0003
H15	0.01	0.37	−0.15	−0.21	0.17	−0.16	−0.0005
H16	−0.05	0.11	−0.09	−0.07	−0.10	0.06	−0.0003
H17	−1.41	0.38	−1.54	−0.25	−0.22	−1.19	−0.0012
H18	−0.33	1.00	−0.57	−0.76	−0.09	−0.24	−0.0036
H19	−0.38	0.58	−0.51	−0.45	−0.06	−0.31	−0.0022
H20	0.00	−0.02	−0.01	0.03	0.02	−0.02	0.0001
H21	0.00	−0.03	0.00	0.03	0.02	−0.02	0.0001
H22	0.48	−0.32	0.58	0.22	0.15	0.33	0.0010
H23	0.48	−0.31	0.57	0.22	0.15	0.33	0.0010
H24	−1.09	0.55	−1.34	−0.30	−0.38	−0.71	−0.0014
K1(2)	2.05	4.71	−0.97	−1.69	−0.91	2.96	−0.0077
K2(3)	−2.82	2.24	−3.50	−1.55	−4.45	1.64	−0.0026
K3(4)	1.40	−4.31	2.93	2.78	0.20	1.20	0.0103
H caps	0.26	0.24	0.14	−0.12	−1.00	1.26	0.0000
$\Sigma\Delta P$(5)	0.89	2.88	−1.42	−0.57	−6.16	7.05	0.0000
$\Sigma\|\Delta P\|$(6)	27.53	17.2	28.16	11.51	12.78	21.79	0.0430

(1) q^A = net charge of atom A
(2) K1 = O3 + O4 + N6 + C10 + C13 + H17 + H18 + H19 + H24.
(3) K2 = O1 + N5 + C8 + C9 + H14 + H15 + H16
(4) K3 = O2 + N7 + C11 + C12 + H20 + H21 + H22 + H23
(5) Sum includes the hydrogen caps.
(6) Sum does not include the hydrogen caps.

$$\Delta E^A_{add} = E^A_{add}(\text{KEM2}) - E^A_{add}(\text{exact}) \quad (18)$$

which are listed in Table 1 and displayed graphically in Fig. 1 for the triglycyle system where they are expressed in kcal/mol. In Fig. 1, each glycine residue constitutes a kernel where the lines across the amide bonds indicate the different kernels.

The total energy obtained from the direct calculation ("exact") and that obtained from the KEM2 approximation (Eq. (9)) differ by 0.89 kcal/mol, *i.e.*, KEM2 delivers the energy of this system within chemical accuracy (exact total energy = −696.6518 hartrees, KEM2 total energy = −696.6504 hartrees).

The sum of the atomic IQA energies calculated directly (on the full intact molecule) compared to those calculated by KEM2 differ by 0.06 kcal/mol (–696.6504 and −696.6505 hartrees, respectively), indicating accurate numerical integrations. This is notable given the wider range of KEM2 atomic errors seen in Fig. 1, that reach extremes of +5.39 kcal/mol for N6 and −2.64 kcal/mol for C9, yet the sum of these energies from both sets of calculations agree with less than a 0.1 kcal/mol global error. So the errors in the individual atomic energies can be up to a full order of magnitude larger than the error in the KEM2 sum of these atomic energies. A glance at the bottom of Table 1 reveals that the sum of all the self atomic errors (ΔE^A_{self}) is −6.16 kcal/mol (indicating a net stabilization error introduced in the KEM approximation) while the sum of the interaction errors (ΔE^A_{int}) is +7.05 kcal/mol (*i.e.*, a net destabilization error introduced by KEM). The two global errors clearly nearly cancel and one obtains a total error of just +0.89 kcal/mol (i.e. a net destabilization error such that KEM does not violate the variational principle *in this case*. However, ***KEM is not guaranteed to be variational*** and in those cases which do violate the variational principle *it can be made variational* as described in Ref. [18]. When KEM is made variational the magnitude of the total KEM error is further reduced, at least in the cases that were studied so far). There is thus a clear cancellation of atomic KEM2 errors that underlies the chemical accuracy of the KEM2 energy for this molecule. This phenomenon is commonplace; that is relatively large local errors cancel one another to yield a smaller global error. We now delve into the origin of this error compensation.

Table 1 lists the errors in KEM2 IQA atomic energy components relative to their corresponding values obtained directly from the full molecular wavefunction of triglycine. The table lists the atomic KEM2 errors in the additive IQA energy (ΔE^A_{add}), the atomic kinetic energy (ΔT^A), the atomic Coulomb contribution to the energy ($\Delta V^A_{ee\ Coul.}$), and the atomic exchange contribution to the energy ($\Delta V^A_{ee\ X}$). The table also

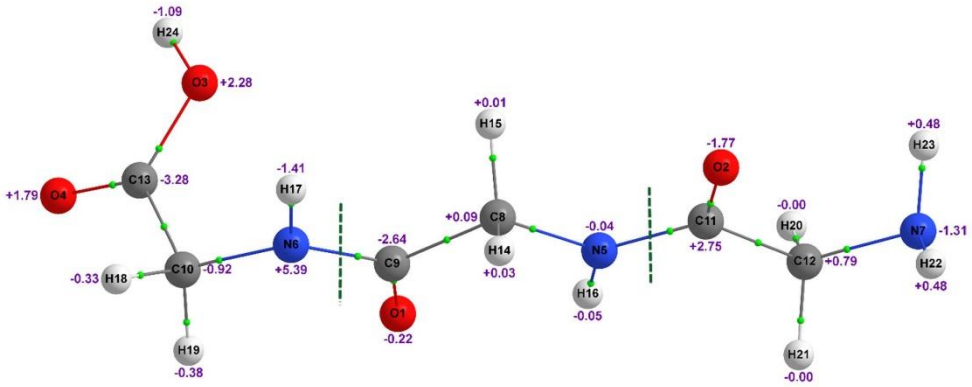


Fig 1. QTAIM molecular graph of triglycine, calculated at HF/6-311G(d,p)//6-311G(d,p). The dashed vertical lines indicate the 3 glycine fragments: K1 (left), K2 (middle) and K3 (right). Errors in atomic IQA additive energies calculated using KEM2 relative to the full system are shown in kcal/mol. (For colour figure, the reader is referred to the web version of this article.)

lists the intra-atomic and inter-atomic contributions to the atomic KEM2 energy errors, ΔE^{A}_{self} and ΔE^{A}_{int}, respectively. It is noteworthy that there is a non-negligible range of KEM2 errors for the atomic and kernel IQA energies and their contributions compared to the molecular KEM2 energy error.

The last two rows of Table 1 list the sums of the errors over the entire system ($\Sigma\Delta P$) and the corresponding sums of the unsigned errors (the sums of the magnitudes of the errors) ($\Sigma|\Delta P|$). These two rows reveal an interesting pattern. This pattern can be qualitatively summarized in that the quantities that are proper to the atom itself are better reproduced than those explicitly dependent on the remainder of the atoms in the system. Hence, atomic charges, obtained though the integration of the electron density over the atomic basin, exhibit a total absolute cumulative error of $0.043/24 < 0.002$ atomic units per atom, a high level of accuracy as previously observed [7]. The sum of the magnitudes of the errors in the interaction energy term (ΔE^{A}_{int}) is of 21.8 kcal/mol, almost twice the total unsigned error in the self-energies (ΔE^{A}_{self}) which amounts to only 12.8 kcal/mol.

Another observation is the small error introduced by the hydrogen caps themselves. Remarkably, the KEM2 master formula (Eq. (1)) results in the nearly complete cancellation of the energies of these caps, which in total introduce an overall error of 0.26 kcal/mol. We note in passing that K2 (the central kernel) is stabilized as a whole (its $\Delta E^{A}_{add} = -2.8$ kcal/mol < 0) while the two peripheral kernels (K1 and K2) are destabilized (by 2 and 1.4 kcal/mol, respectively). The bulk of the stabilization of K2 is traceable to its self-energy (E^{A}_{self}) which is decreased by 4.45 kcal/mol compared to the atoms of this kernel in the intact molecule as opposed to its interaction energy which is increased by 1.64 kcal/mol only.

4. Conclusion

This paper accords with previously published results which indicate that KEM2 delivers reliable approximations for the properties of large molecules. What the calculations of the present paper indicate is that cancellation of errors among the atoms and among the kernels leads to the favorable results associated with KEM2 for molecules. In other words, KEM2 is accurate not just because the individual KEM2 atomic energy errors are sufficiently small but because there is a near cancellation of these often larger atomic energy errors. The KEM energies at the atomic level are up to an order of magnitude less accurate than the overall KEM energies, and hence, the accuracy of KEM is not reducible to an atomic basis but rather emerges as a result of a subtle cancellation of errors.

This paper is clear about one important question, *viz.*, since KEM2 has been tested extensively in the past and found to be quite accurate, wherein does that accuracy arise? The accuracy at the level of the kernels is found to exceed that at the atomic level, reflecting a cancelation of errors within the KEM2 summation. That is in itself a new numerical fact, previously not known, leaving open why such error cancelation must occur.

The KEM approximation has been considered here only to the second order (KEM2) on a system with three kernels. It is expected that similar results would be obtained for KEM2 applied to larger systems with more kernels. For systems with more than three kernels, it is expected that KEM3 would probably lead to more accurate reproduction of both atomic IQA energies and molecular energies. There is clearly a pathway to further exploration.

Acknowledgements

L.M. was funded by the U.S. Naval Research Laboratory (Project 47203-00 01) and by the Professional Staff Congress, City University of New York (CUNY) (63842-00 41). C.F.M acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI), and Mount Saint Vincent University for funding. This work has, in part, been presented as a poster at the Sagamore 2018 Conference on Quantum Crystallography held in Halifax, Nova Scotia, Canada, 8-13 July 2018.

References

[1] L. Massa, L. Huang, J. Karle, J. Int.Quant. Chem.: Quant. Chem. Symp. 29 (1995) 371.
[2] L. Huang, L. Massa, J. Karle, J. Int.Quant. Chem.: Quant. Chem. Symp. 30 (1996) 479.
[3] L. Huang, L. Massa, J. Karle, IBM J. Res. Dev. 45 (2001) 409.
[4] L. Huang, L. Massa, J. Karle, Quantum Biochemistry: Electronic Structure and Biological Activity, in: C.F. Matta (Ed.), Wiley-VCH, Weinheim, 2010, p. 3.
[5] L. Massa, C.F. Matta, J. Comput. Chem. 39 (2018) 1021.
[6] A. Genoni, L. Bucinskż, N. Claiser, J. Contreras-Garcia, B. Dittrich, P.M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J.-M. Gillet, D. Jayatilaka, P. Macchi, A.Ų. Madsen, L. Massa, C.F. Matta, K.M. Merz Jr., P. Nakashima, H. Ott, U. Ryde, W. Scherer, K. Schwarz, M. Sierka, S. Grabowsky, Chem. Eur. J. 24 (2018) 10881.
[7] L. Huang, C.F. Matta, L. Massa, Struct. Chem. 26 (2015) 1433.
[8] L. Huang, L. Massa, M. Krupkin, A. Bashan, A. Yonath, Proc. Natl. Acad. Sci. USA 110 (2013) 14900.
[9] L. Huang, L. Massa, J. Karle, Proc. Natl. Acad. Sci. USA 106 (2009) 1731.
[10] L. Huang, L. Massa, J. Karle, Proc. Natl. Acad. Sci. USA 103 (2006) 1233.
[11] L. Huang, L. Massa, J. Karle, Proc. Natl. Acad. Sci. USA 102 (2005) 12690.
[12] L. Huang, L. Massa, J. Karle, Biochemistry 44 (2005) 16747.
[13] L. Huang, L. Massa, J. Karle, Int. J. Quant. Chem. 103 (2005) 808.
[14] M.J. Timm, C.F. Matta, L. Massa, L. Huang, J. Phys. Chem. A 118 (2014) 11304.
[15] L. Huang, H. Bohorquez, C.F. Matta, L. Massa, Int. J. Quant. Chem. 111 (2011) 4150.
[16] L. Huang, L. Massa, C.F. Matta, Carbon 76 (2014) 310.
[17] L. Huang, L. Massa, J. Karle, Proc. Natl. Acad. Sci. USA 105 (2008) 1849.
[18] W. Polkosnik, L. Massa, J. Comput. Chem. 39 (2018) 1038.
[19] R.F.W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, Oxford, U.K., 1990.
[20] P.L.A. Popelier, Atoms in Molecules: An Introduction, Prentice Hall, London, 2000.
[21] C.F. Matta, R.J. Boyd (Eds.), The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design, Wiley-VCH, Weinheim, 2007.
[22] M.A. Blanco, A. Martin Pendas, E. Francisco, J. Chem. Theor. Comput. 1 (2005) 1096.
[23] Á. Martín Pendás, M.A. Blanco, E. Francisco, J. Chem. Phys. 125 (2006) 184112.
[24] Á. Martín Pendás, E. Francisco, M.A. Blanco, J. Phys. Chem. A 110 (2006) 12864.
[25] E. Francisco, Á. Martín Pendás, M.A. Blanco, J. Chem. Theory Comput. 2 (2006) 90.
[26] Á. Martín Pendás, E. Francisco, M.A. Blanco, C. Gatti, Chem. Eur. J. 13 (2007) 9362.
[27] C.F. Matta, Comput.Theor. Chem. 1124 (2018) 1.
[28] C.F. Matta, J. Comput. Chem. 35 (2014) 1165.
[29] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision B.01, Gaussian Inc., Wallingford CT, 2010.
[30] Todd A. Keith, AIMAll (Version 17.11.14), TK Gristmill Software, Overland Park KS, USA, 2019.

Appendix 3. Hot Mitochondria

Fahimi, P., Nasr, M. A., Castanedo, L. A. M., Cheng, Y., Toussi, C. A. and Matta, C.F. (2019) *Biophys. Bull.* (submitted, reference [85])

Youji Cheng's contributed to the general area of this article and was the presenter of an oral contribution: Cheng, Y.; Matta, C. F. "A Possible Fundamental Role of the Intrinsic Electric Field of ATP Synthase", 102nd Canadian Chemistry Conference & Exhibition (CSC 2019), Québec City, QC, Canada, (3 - 7 June 2019).

Invited Paper:

A Note on the Consequences of a Hot Mitochondrion: Some Recent Developments and Open Questions[1]

Peyman Fahimi,[a,b] Mohamed A. Nasr,[a,c] Lazaro A. M. Castanedo,[a,d] Youji Cheng,[a,d] Cyrus A. Toussi,[a,e] Chérif F. Matta,*[a-d,f]

[a] *Department of Chemistry and Physics, Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M2J6.* [b] *Dép. de chimie, Université Laval, Québec, Québec, Canada G1V 0A6.* [c] *Center of Excellence for Stem Cells and Regenerative Medicine (CESC), Zewail City of Science and Technology, 6th of October City 12588, Egypt.* [d] *Department of Chemistry, Saint Mary's University, Halifax, Nova Scotia, Canada B3H3C3.* [e] *Department of biomedical engineering, Hakim Sabzevari University, Sabzevar, Iran.* [f] *Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H4J3.*

** Tel.: +1-(902)-457-6142, Fax: +1-(902)-457-6134, E-mail: cherif.matta@msvu.ca.*

Abstract

The recent suggestion that the mitochondrion might possibly be hotter than its surrounding demands some reconsideration of several aspects of mitochondrial biochemistry and biophysics. Some of these repercussions of a higher mitochondrial temperature open new questions that beg for answers. A few examples of these consequences and new open questions are very briefly highlighted.

Keywords: Hot mitochondrion, mitochondrial biophysics, ATP synthase, Maxwell demon, chemiosmotic theory.

Since Philip Siekevitz coined the term, the mitochondrion has been recognized as the "powerhouse of the cell" [1]. A thermal power-plant's efficiency is given by the celebrated Kelvin's formula:

$$\eta = 1 - \left(\frac{T_{\text{cold}}}{T_{\text{hot}}} \right). \tag{1}$$

However, the living cell is considered to operate under isothermal conditions (at a temperature of 37 °C in humans), even though this view may require minor revision. For an isothermal system, the efficiency can be described by the more general equation:

[1] This communication is patterned after the presentation of the corresponding author (C.F.M.) entitled "*Controlled Thermogenesis in the Mitochondrion*" presented at the *6th International Conference on Nanobiophysics: Fundamental and Applied Aspects* (NBP-2019) held at the Institute of Physics of the National Academy of Sciences of Ukraine in Kyiv (Ukraine), 1-4 October 2019.

$$\eta = \frac{w}{\Delta G(T)} = \frac{w}{\Delta H - T\Delta S}, \qquad (2)$$

that is, the ratio of useful work (or energy containing product such as ATP) over the energy content of ingested foodstuffs.

Mitochondrial energy efficiency is important in characterizing and understanding a number of diseases (e.g. in insulin resistance of cells that results in diabetes mellitus [2], or in the well-known switching of metabolism to glycolysis in cancerous cells known as the Warburg effect) and in understanding thermodynamic aspects of biological evolution [3]. Thermodynamic efficiency clearly depends on temperature through ΔG's dependence on the temperature of the surrounding thermal bath in isothermal conditions (Eq. (2)) or on the difference in the temperatures of two compartments in the case of a heat engine when a temperature gradient exists (Eq. (1)). Hence, knowledge of the temperature of operation of the mitochondrion and its surroundings is crucial for a proper understanding of its thermodynamic efficiency.

Using a molecular thermometer (mito thermo yellow (MTY)) that distributes preferentially inside the mitochondrial matrix [4], Chrétien *et al.* have recently shown that the mitochondrion operates at temperatures that are higher than their immediate cellular surroundings, temperatures that can perhaps fetch as high as 50 °C [5]. Doubts on the interpretation of the experimental data have been cast by Lane [6] to the effect that it may be just a matter of degree, that is, the mitochondrion is probably hotter than the rest of the cell but perhaps not as hot as Chrétien claims. Much more dramatically, and based on an order of magnitude estimation of the power consumption of a growing bacterial cell and the thermal conductivity of water, Sear challenges the Chrétien finding and arrives at a maximum estimated energy flux of 1 W/m^2 which translates into a temperature gradient of a maximum of 1 K/m, that is, a maximum gradient of the order of a μK across the dimensions of the cell [7]. Sear's challenge is to essentially falsify the proposition of any notable temperature difference between cell compartments since at the most these would be in the micro-Kelvin range. However, the fact remains as to how can one explain away Chrétien's results? And also the fact that other experimenters have come to similar conclusions, albeit not as dramatic as Chrétien's 10 °C gradient. For instance, Shen *et al.* use rhodamine B methyl ester dye as a mitochondrial thermometer to establish that the natural anti-inflammatory autacoid prostaglandin E_2 (PGE_2) can decrease the intracellular temperature of hepatocytes [8]. These workers argue that

Chrétien's findings are consistent with theirs since PGE_2 downregulates metabolism and, hence, heat production in the hepatocites [8]. Using fluorescent thermometry, Okabe *et al.* report a 1 °C gradient between the cell nucleus and the cytosol [9,10]. The presumed existence of (significant) intracellular temperature gradients motivated researchers to develop molecular thermometers that can penetrate mammalian cells and yeast (see for example Ref. [11]).

If indeed the mitochondrion is hotter than its surroundings, even by a few degrees, then some of our understanding of mitochondrial biochemistry and biophysics may possibly need to be slightly revisited. We insist on the adjective "slight" since even in the most extreme case of ~ 10 °C temperature gradient, this would translates through Eq. (1) into a theoretical maximum of only ~ 4% revision of the efficiency, while this number drops to a maximum of ~ 0.3% for a 1 °C gradient which is, at best, a perturbation that is much less than any hoped-for precision of intracellular conditions. One has also to consider the local random fluctuation of temperature in very small confined spaces in the spatial dimensions [12] as well as in the temporal dimension [13].

Hence, and more explicitly, such a revision may include a reconsideration of the explicit dependence of the first term in the Gibbs energy expression of the chemiosmotic energy (Eq. (1)) on T and also the indirect effect of the temperature on the second (electric term):

$$\Delta G = \underbrace{2.3nRT\Delta\mathrm{pH}}_{\Delta G_{\mathrm{chem.}}} + \underbrace{n\mathscr{F}\Delta\psi}_{\Delta G_{\mathrm{elec.}}}, \tag{3}$$

where $\Delta\mathrm{pH} = \mathrm{pH}_{\mathrm{in}} - \mathrm{pH}_{\mathrm{out}}$, and Faraday's constant $\mathscr{F}$=96.485 $\mathrm{kJ.V^{-1}.mol^{-1}}$, which yields for a $\Delta\mathrm{pH}$ = 1 unit and 37 °C:

$$\Delta G = 2.3\,RT + \mathscr{F}(0.25\,\mathrm{V}) \approx \underset{(20\%)}{6\,\mathrm{kJ}} + \underset{(80\%)}{24\,\mathrm{kJ}} \approx 30\,\mathrm{kJ.mol^{-1}}, \tag{4}$$

where mol here refers to a mole of protons.

Now if we assume a temperature of 320 K (i.e., 47 °C, ten degrees higher than normal body temperature), and focusing only on the chemical term (ignoring the potential term since we do not know the functional dependence of the potential on the temperature at this point), the change in ΔG per proton would be then:

$$\Delta\Delta G_{\mathrm{chem.}} = \left(\Delta G_{320\,\mathrm{K}} - \Delta G_{310\,\mathrm{K}}\right)_{\mathrm{chem.}} = 2.3\times 1\times R\times 10\times 1 \approx 0.2\,\mathrm{kJ.mol^{-1}}, \tag{5}$$

which is a very small correction of the order of 3% of the chemical (minor) term. At this

point we have no estimates of the effect of temperature on the dominant electrical term. As for temperature, pH, and concentrations, and in view of the small physical space of the mitochondrial "capacitor", one must also account for the local and temporal fluctuation of the voltage and its associated electric field across the inner mitochondrial membrane [14,15]. Furthermore, the obligatory energy dissipation by ATP synthase, a Maxwell demon as it acts as a molecular sorting machine [16-18], should be mildly revised as well whereby the minimum energy of $k_B T \ln 2$ per proton crossing it should be increase by 10 $k_B \ln 2$ (which represents a correction of again of ~ 3% at the most). Finally, perhaps one can also consider a minor aspect of a "heat engine" operation in the mitochondrion whereby one can add an energy term proportional to $[1-(T_{\text{cold}} / T_{\text{hot}})]$.

None of the above considerations touch upon another crucial aspect that would be far more (exponentially) sensitive to a temperature difference and that is kinetics aspects. In fact just from simple Arrhenius law, for a reaction with an activation energy of, say, 50 kJ/mol, the ratios of the rate constants' in the two extreme cases of 1°C temperature gradient (37 → 38 °C) and 10 °C gradient (37 → 47 °C) would be bracketed by:

$$\sim 1.06 (\text{at } 38^{\circ}\text{C}) \leq \left[\frac{k_2}{k_1} = e^{\frac{E_a}{R}\left(\frac{1}{T_1}-\frac{1}{T_2}\right)} \right] \leq \sim 1.83 \text{ (at } 47^{\circ}\text{C)}, \qquad (6)$$

implying an almost doubling of reactions (with E_a = 50 kJ/mol) in the mitochondrion if the gradient at the higher end and a 6% rise in the rate of this reaction for a 1°C gradient.

Nasr *et al.* have recently examined some of the biological consequences of a possible "hot mitochondrion" [3]. Among the problems that such a higher temperature mitochondrion is representing is that of the instability of mitochondrial macromolecules, particularly proteins and nucleic acids, at elevated temperatures as they can approach or reach their melting temperatures. This is further complicated by the known increase in reactive oxygen species (ROS) production at these higher temperatures (See Ref. [3] and references therein). Nasr *et al.* suggest that heat shock proteins (Hsps), abundant in the mitochondria and primarily thought to act as molecular chaperones to ensure the proper folding of imported proteins, may also act according to their conventional role as stabilizers of protein structures at higher temperatures. These Hsps can also have a secondary role as protectants against the temperature-dependent increase in ROS levels. These propositions are corroborated by the presence of other compatible solutes (known thermoprotectants and stabilizers of protein structures) such as di-*myo*-inoitol-1,1'-

phosphate, and also by the richness of at least some mitochondrial nucleic acids encoding for Hsps in purine tracks compared to their cytosolic counterparts [3]. Purine tracks are long segments of a nucleic acid that have purine bases stacked one on the top of the next, and - given the larger aromatic rings of purines compared to pyrimidines – this imparts added stability to these nucleic acids presumably to confront the hotter mitochondrial environment [3]. The very presence of these Hsps and their proposed role as structural stabilizers against the insults of the higher temperature argues in favor of the proposition that the mitochondrion appears to be indeed "hotter" than its surroundings, but the open question appears to be "by how much?"

Hotter or not, the fact remains that temperature must be maintained within a range that allows for the proper functioning of the mitochondrion. So the next question enquires on how the mitochondrion is capable of maintaining its thermal homeostasis? How does it regulate its temperature within an acceptable range? The authors are proposing a feedback mechanism for the regulation of the temperature of the mitochondrion [19].

The proposed control loop is based on the realization that the rate limiting step of the electron transport chain (ETC) is the lateral diffusion of ubiquinone/ol (in its various oxidation states, collectively referred to as "UQ" species) within the inner mitochondrial membrane [20,21]. An increase in the rate of heat production will be associated with a rise in the temperature of the mitochondrion, the higher temperature will decrease the viscosity of the membrane which, in turn, increases the diffusion coefficients of UQ. The faster mobility of UQ accelerates the electron transport chain, and as a result enhances the rate of proton pumping into the inter-membrane gap. The higher gradient created as a result will be associated with a stronger (chemiosmotic) electric field perpendicular to the membrane (and larger voltage across the inner mitochondrial membrane). As any capacitor, this electric field exerts a pressure P over the capacitor dielectric medium (here the phospholipid membrane itself) according to:

$$P = \frac{\varepsilon}{2} E^2, \tag{7}$$

where ε is the dielectric constant of the membrane and E is the magnitude of the electric field normal to the membrane.

This increase in the pressure on the membrane liquid crystal, in turn, increases its viscosity which reduces the diffusion coefficient of UQ putting the breaks on the further

pumping of protons and of heat generation. Naturally, the loop operates in the opposite mode if the heat production is lower than optimal. This positive/negative feedback mechanism establishes the thermal homeostasis needed to maintain the mitochondrial temperature within working bounds.

Acknowledgements

The authors thank Professor Galina I. Dovbeshko (*National Academy of Sciences of Ukraine*), Professor Stephen L. Bearne (*Dalhousie University*), Professor Nagwa El-Badri (*Zewail City of Science and Technology*), and Professor Lou Massa (*City University of New York*) for helpful discussions and suggestions. The authors are grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI), Mount Saint Vincent University, Université Laval, and Saint Mary's University for financial and material support.

References

[1] Siekevitz P. Powerhouse of the cell. *Sci. Am.* 197(1), 131-40 (1957).

[2] Crescenzo R, Bianco F, Mazzoli A, Giacco A, Liverini G, Iossa S. Mitochondrial efficiency and insulin resistance. *Front. Physiol.* 5, Article 512 (pp.1-5) (2014).

[3] Nasr MA, Dovbeshko GI, Bearne SL, El-Badri N, Matta CF. Heat shock proteins in the "hot" mitochondrion: Identity and putative roles. *BioEssays* 41, 1900055 (pp. 1-6) (2019).

[4] Arai S, Suzuki M, Park SJ, Yoo JS, Wang L, Kang N-Y, Ha H-H, Chang Y-T. Mitochondria-targeted fluorescent thermometer monitors intracellular temperature gradient. *Chem. Commun.* 51, 8044-8047 (2015).

[5] Chrétien D, Bénit P, Ha H-H, Keipert S, El-Khoury R, Chang Y-T, Jastroch M, Jacobs HT, Rustin P, Rak M. Mitochondria are physiologically maintained at close to 50°C. *PLOS Biology* 16(1) e2003992, 1-17 (2018).

[6] Lane N. Hot mitochondria? *PLOS Biology* (doi.org/10.1371/journal.pbio.2005113), pp.1-6 (2018).

[7] Sear RP. Diffusiophoresis in cells: A general nonequilibrium, nonmotor mechanism for the metabolism-dependent transport of particles in cells. *Phys. Rev. Lett.* 122, 128101 (2019).

[8] Shen L, Xie T-R, Yang R-Z, Chen Y, Kang J-S. Application of a dye-based mitochondrion-thermometry to determine the receptor downstream of prostaglandin

E2 involved in the regulation of hepatocyte metabolism. *Scient. Rep.* 8, Article 13065 (2018).

[9] Okabe K, Inada N, Gota C, Harada Y, Funatsu T, Uchiyama S. Intracellular temperature mapping with a fluorescent polymeric thermometer and fluorescence lifetime imaging microscopy. *Nature Commun.* 3, Article 705 (2012).

[10] Hayashi T , Fukuda N, Uchiyama S, Inada N. A cell-permeable fluorescent polymeric thermometer for intracellular temperature mapping in mammalian cell lines. *PLoS ONE* 10, Article e0117677 (2015).

[11] Tsuji T, Yoshida S, Yoshida A, Uchiyama S. Cationic fluorescent polymeric thermometers with the ability to enter yeast and mammalian cells for practical intracellular temperature measurements. *Anal. Chem.* 85, 9815-9823 (2013).

[12] Johnson HA. Thermal noise and biological information. *Quarter. Rev. Biol.* 62, 141-152 (1987).

[13] Hickman J , Mishin Y. Temperature fluctuations in canonical systems: Insights from molecular dynamics simulations. *Phys. Rev. B* 94, Article 184311 (pp. 10) (2016).

[14] Procopio J, Fornes J. Fluctuations of the proton-electromotive force across the inner mitochondrial membrane. *Phys. Rev. E.* 55, 6285-6288 (1997).

[15] Procopio J, Fornes JA. Fluctuation-dissipation theorem imposes high-voltage fluctuations in biological ionic channels. *Phys. Rev. E.* 51, 829-831 (1995).

[16] Matta CF, Massa L. Information theory and the thermodynamic efficiency of biological sorting systems: Case studies of the kidney and of mitochondrial ATP-synthase. *Chapter 1 in: Sustained Energy for Enhanced Human Functions and Activity, Bagchi, D. (Ed.), Academic Press - An imprint of Elsevier (London)* 3-29 (2017).

[17] Matta CF, Massa L. Chapter 1: Information Theory and the Thermodynamic Efficiency of Biological Sorting Systems: Case Studies of the Kidney and of Mitochondrial ATP-Synthase. *Sustained Energy for Enhanced Human Functions and Activity*; Elsevier, The Netherlands (2017) , pp 3-29.

[18] Matta CF, Massa L. Energy equivalence of information in the mitochondrion and the thermodynamic efficiency of ATP synthase. *Biochemistry* 54, 5376-5378 (2015).

[19] Fahimi P, Castanedo LAM, Nguyen-Dang T-T, Matta CF. Coupled electrical-thermal feedback control of the inner mitochondrial proton gradient. A hypoythesis. *J. Theor. Biol.* (2019).

[20] Mathai JC , Sauna ZE, John O, Sitaramam V. Rate-limiting step ine lectron transport: Osmotically sensitive diffusion of quinones through voids in the bilayer. *J. Biol. Chem.* 268, 15442-15454 (1993).

[21] Moncelli MR, Herrero R, Becucci L, Guidelli R. Kinetics of electron and proton transfer to ubiquinone-10 and from ubiquinol-10 in a self-assembled phosphatidylcholine monolayer. *Biochim. Biophys. Acta* 1364, 373-384 (1998).