**An Investigation of The Applications of Localization Delocalization Matrices on The Prediction of Reaction Characteristics**

Lauryn Mason

Submitted 29 April 2024

Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

the Degree of Bachelor of Science

with Honours in Chemistry.



Approved: Dr. Kai E. O. Ylijoki
Supervisor

Approved: Dr Jane Ferguson
Reader

**Abstract**

An Investigation of The Applications of Localization Delocalization Matrices on The Prediction of Reaction Characteristics

Lauryn Mason

Submitted 29 April 2024

The localization delocalization matrix (LDM) is a complete chemical graph that ties the quantum theory of atoms in molecules (AIM) with chemical graph theory (CGT), allowing for quantitative structure activity relationships (QSAR) to be developed. The utility of the LDM for mechanistic analysis has been explored, and the mean topological contribution of the Frobenius distance between mechanisms has been isolated as a potential metric for analyzing mechanistic similarity. The mechanisms of two crossed dimerization reactions of substituted 1,3,5,7-cyclooctatetraenes have been computationally explored using density functional theory (DFT).

**Acknowledgements**

I would like to extend my sincere thanks to Dr. Kai Ylijoki for working with me on this project and for taking the time to help me understand when things were going wrong. I would not have been able to complete this project without his help, often times outside of normal working hours.

I would like to thank Dr. Chris Frazee for letting me bounce ideas off of him while I was TA-ing for his lab section.

I would like to thank my good personal friends, Alicia Piasecki and Molly Munroe for kindly lending their perspectives as academics outside of the field of Chemistry, and for allowing me to bore them with quantum mechanics so that they could help me with this project.

Lastly, I would like to thank my dog Lincoln, for his endless moral support.

## Table of Contents

## List of Symbols and Abbreviations

| | |
|---|---|
| DFT | density functional theory |
| $\rho(\mathbf{r})$ | electron density distribution function |
| $v(\mathbf{r})$ | external potential function |
| LSDA | local spin density approximation |
| ωB97X-D | long-range corrected hybrid density functional with damped atom-atom dispersion corrections |
| PES | potential energy surface |
| DOF | degrees of freedom |
| IRC | intrinsic reaction coordinate |
| CGT | chemical graph theory |
| QSPR | quantitative structure property relationship |
| QSAR | quantitative structure activity relationship |
| AIM | quantum theory of atoms in molecules |
| Ψ | quantum mechanical state function |
| NCP | nuclear critical point |
| BCP | bond critical point |
| RCP | ring critical point |
| CCP | cage critical point |
| LI | localization index |

| | |
|---|---|
| $\Lambda(\Omega_i)$ | localization index of i-th atomic basin |
| DI | delocalization index |
| $\delta(\Omega_i, \Omega_j)$ | delocalization index between i-th and j-th atomic basins |
| $\Omega_i$ | i-th atomic basin |
| PH | Poincaré-Hopf |
| LDM | localization delocalization matrix |
| COT | 1,3,5,7-cyclooctatetraene |
| INT | intermediate |
| TS | transition state |
| $R^2$ | coefficient of determination |
| PM3 | parametric method 3 |
| HOMO | highest occupied molecular orbital |
| LUMO | lowest unoccupied molecular orbital |

## List of Figures, Tables, and Schemes

# 1. Introduction

## *1.1. Computational Background*

### *1.1.1. Density Functional Theory (DFT) and Potential Energy Surface Elucidation*

Density Functional Theory (DFT) is a profoundly useful tool for computing molecular energy and structure that yields good results with relatively short computational time. It is based on the work conducted by Hohenberg and Kohn in 1963, where they showed that there exists some functional F that relates electron density $\rho(\mathbf{r})$ to external potential $v(\mathbf{r})$, which can then be used to construct a molecular Hamiltonian, and therefore, by the Schrödinger equation, molecular energy can be determined.[1] Kohn and Sham would later develop a system of self-consistent equations analogous to the Hartree-Fock equations that could be used to solve for the electron density $\rho(\mathbf{r})$.[2] This approach to solving for the energy of a molecular system naturally incorporates the electron correlation energy that is neglected in Hartree-Fock theory, but results in the exchange correlation potential, which must be solved for empirically.[3] This exchange correlation energy is the only term that introduces error into DFT, and thus the accuracy of any DFT result is correlated to the accuracy of approximation of the exchange correlation potential.[4]

A number of exchange correlation functionals have been developed to attempt to approximate this energy. For example, the work conducted herein utilizes extensions to the Becke hybrid exchange-correlation functional, which was originally described in 1992. Becke's approach used the exact Hartree-Fock exchange energy coupled with the local spin-density approximation (LSDA) exchange correlation energy, which showed good results when compared to experimental data.[5] A number of corrections to the Becke functional have been made, and the work here utilizes the ωB97X-D hybrid functional, which is a Becke-based functional that

includes long-range correlation and dispersion terms, and shows good agreement with experimental results.[6] This functional was chosen for this work to allow comparison with previous results, as it was selected by Castanedo *et al.* in their work on the localization delocalization matrix.[7]

The potential energy surface (PES) is a useful concept that arises from the Born-Oppenheimer approximation. It is essentially the function E(**DOF**), where E is chemical potential energy and **DOF** is the vector coordinate that includes the degrees of freedom of a molecule, *i.e.,* all geometric information. The PES is commonly thought of as the n-dimensional extension of the reaction coordinate taught in any introductory organic chemistry class. In this schema, a reactant or a product is a PES minimum (*i.e.,* a point at which the PES gradient is 0 and the PES second derivative matrix, also known as the Hessian matrix, has all positive eigenvalues), and a transition state is a first order saddle point (*i.e.,* a point at which the PES Hessian matrix has one and only one negative eigenvalue).[8] PES second derivative matrices with more than one negative eigenvalue correspond to neither an intermediate or a transition state, and so are chemically irrelevant. PES minima and first order saddle points are computed by geometry optimization algorithms,[9] and the path between transition states and minima, known as the intrinsic reaction coordinate (IRC), can be elucidated using the Hratchian predictor-corrector method.[10] In practice, these algorithms are implemented in a software package. This usage of DFT is widespread throughout the literature for developing understanding of novel reaction mechanisms.[3]

### *1.1.2. Chemical Graph Theory and QSAR/QSPR*

Chemical graph theory (CGT) is the application of the field of graph theory in mathematics to chemistry. It is essentially the study of the connectivity of molecules and how that affects structure reactivity relationships. A molecule is broken down into its graph, which is a set of vertices representing atoms and edges representing bonds. Graphs are usually made excluding hydrogen, as it does not usually contribute much useful information about reactivity.[11] Additionally, so-called chemical pseudographs can also be made and analyzed, which contain extra information about the molecule, such as multiple bonds and lone pairs. For example, Figure 1.1.2.1 shows a pseudograph of 3-chloroprop-1-ene, which contains extra bonding and lone-pair information. A graph can be converted into a matrix form in a number of ways. Commonly, the adjacency matrix is used, which is defined as the matrix where the i,j-th entry is the number of connections between the i-th and j-th atom, and the i,i-th entry is the number of self-connections. Figure 1.1.2.1 also shows the adjacency matrices for the graph and pseudograph of chloroprop-1-ene.

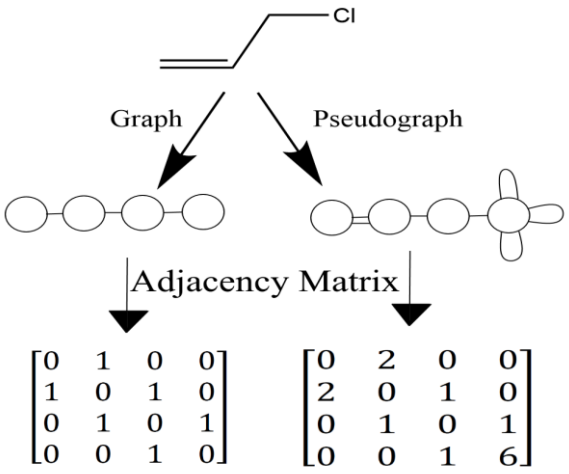



Figure 1.1.2.1. The graph and pseudograph of 3-chloroprop-1-ene.

Metrics of the graph, such as eigenvalues, eigenvector components, determinants, norms, and others are correlated to experimental results and properties of the corresponding chemicals.[11] These kinds of relationships are known as a quantitative structural-property relationships (QSPR), or a quantitative structural-activity relationships (QSAR). Note that not all QSPR or QSAR studies are studies in CGT and are only considered to be within CGT if the chemical graph or some property thereof is used as an independent variable in the development of the relationship. There is a trend in science by which data generation is outpacing data analysis, and the broad aim of CGT is to attempt to increase the rate of data analysis so that we "avoid drowning in this rising sea".[11]

Indeed, many useful applications have risen from chemical graph theory. In 2001, Basak *et al* used a CGT QSAR approach to predict mutagenicity of aromatic amines.[12] Öberg *et al* used an *ab initio* Hartree-Fock QSAR model to predict non-linear optical properties (second and third harmonic generation) of a selection of organic compounds.[13] In one study by Pogliani, they developed CGT models to quite accurately predict minimum anesthetic concentrations of trifluoromethylethanes, antagonism of adrenaline by 2-bromo-2-phenethylamines, inhibition of the influenza virus by benzimidazoles, and the boiling points of alcohols and primary amines.[14]

The work presented in this thesis is a synthesis of CGT with quantum chemistry, with the aim to develop novel parameters for quantitative study and prediction, with a focus on novel parameters for rapid study of chemical reaction mechanisms.

### *1.1.3. The Theory of Atoms in Molecules (AIM)*

The theory of atoms in molecules (AIM) was developed by Bader over many years.[15] It aims to rigorously link the quantum mechanical state function ψ with the intuitive chemical understanding of the atom as it exists within a molecule. Classically, quantum theory attempts to derive the state function of the entire system, which in the chemical application is usually a molecule, and then predict observables by applying the corresponding operator onto the state function. The issue is that this approach loses the chemical understanding of the molecule as a system of functional groups and reactive centers, and instead considers it as a single total system. AIM was developed in an attempt to regain this intuition and implant it in a rigorous theoretical foundation.[15]

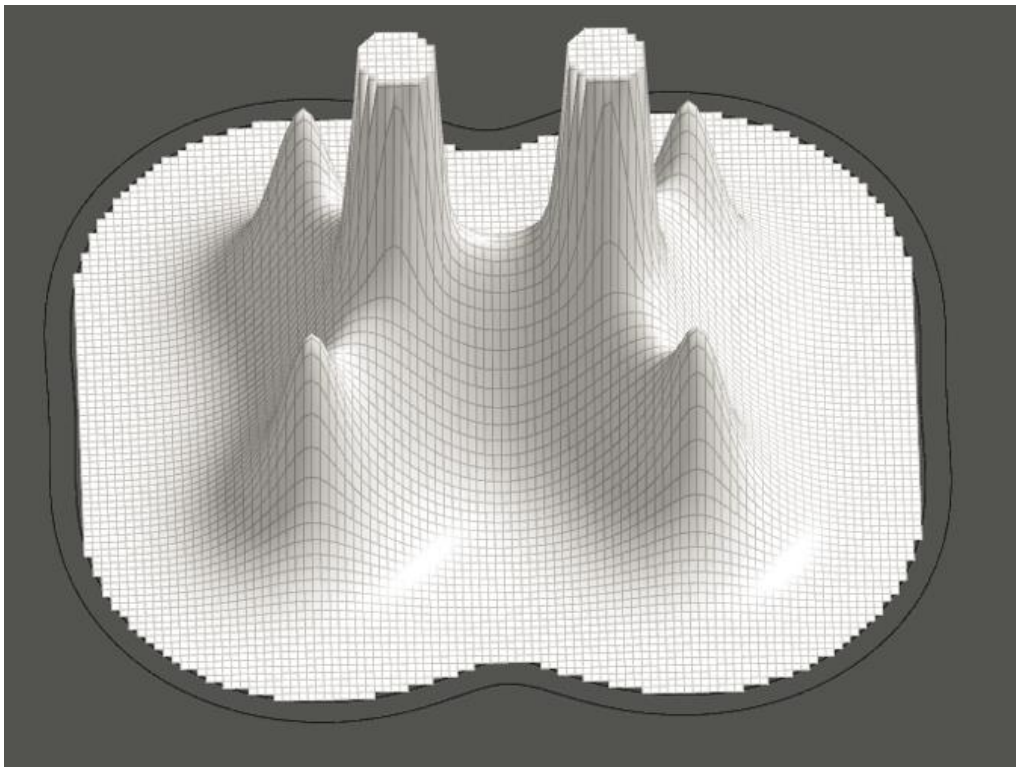


Figure 1.1.3.1. Electron density distribution function, $\rho(\mathbf{r})$, of ethylene ($C_2H_4$).

Essentially, AIM is a theory of the topology of the electron density function $\rho(\mathbf{r})$. Bader showed that atoms and bonds can be recovered from $\rho(\mathbf{r})$, or rather the best way to define a quantum subsystem of the molecule that recovers chemical intuition is to consider critical points of the gradient vector space of $\rho(\mathbf{r})$ (that is where the gradient vector equals the zero vector, $\nabla\rho=\mathbf{0}$).[15] Figure 1.1.3.1. shows the electron density function of ethylene as a visualization aid.

The electron density function $\rho(\mathbf{r})$ shows local maxima at the positions of the nuclei,[15] called nuclear attractor critical points (NCP). There are four small maxima and two large maxima on Figure 1.1.3.1, which represent the hydrogen and carbon nuclei respectively. There exists a set of gradient trajectories that converge to the maximum, which Bader called a basin.[15] Bader defined atoms as the nuclear critical point and its associated basin.[15] A type of first-order saddle point called a (3,-1) critical point, also called a bond critical point (BCP) is found only between nuclei.[15] Five such critical points can be seen on Figure 1.1.3.1, and they lay between each atomic basin. The line between the (3,-1) critical points and the corresponding nuclei are called atomic interaction lines, and contain information pertaining to bonding.[15] Additionally, two other types of critical points are defined by AIM: the ring critical point (RCP) which is found at the center of a ring, and the cage critical point (CCP), which is found in bicyclic and caged molecules at the center of the cage.[15]

Importantly, the number of electrons localized on a single atom, and the number of electrons shared between atoms can be computed as a result of AIM quite virtuously, and are defined symbolically as the localization index (LI), $\Lambda(\Omega_i)$, and the delocalization index (DI), $\delta(\Omega_i,\Omega_j)$ respectively, where $\Omega_i$ is the i-th atomic basin.[16-18]

### *1.1.4. Poincaré-Hopf Theorem*

The Poincaré-Hopf (PH) theorem is a result from topology that is best intuitively understood as an extension of the so-called Hairy Ball theorem, which states that if you had a golf ball covered in hair, there is no way to comb the ball such that all the hair lies flat.[16] We can generalize the hairy ball to be a vector space on a sphere, which makes the point where the hair sticks up a maximum in the vector space. Essentially, the PH theorem places limitations on the number and types of critical points in a vector space. In the application to AIM, the theorem says that the following equation must hold:[20]

$$n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP} = 1$$

where n denotes the number of the corresponding type of critical point. Inequality in this equation usually means that the AIM analysis was not completed appropriately, and that a critical point has not yet been found. In this way, AIM results can be validated, since the equation will most likely only give equality if the complete set of critical points has been found.

## *1.2. The Localization Delocalization Matrix (LDM)*

### *1.2.1. Background and Definition*

The localization delocalization matrix (LDM) in its modern form was defined by Matta in 2014 as a novel QSAR parameter for modelling chemical and biological properties.[21] Matta described it as a "natural extension of the graph-theoretical approach [that] merge[s] it with the topological analysis of the electron density."[21] In essence, it is a molecular graph that contains not only connectivity information, but a plethora of other molecular information including a complete description of the electronic topology, bond strengths, atomic charges, and even proton

NMR coupling constants.[18] It is defined as follows: the diagonal elements contain the i-th LI, i.e., $ldm_{ii} = \Lambda(\Omega_i)$, and the off-diagonal elements contain half the i-th and j-th DI, i.e., $ldm_{ij} = \delta(\Omega_i,\Omega_j)/2$, $i \neq j$, see equation 1.[18,19,21-24]

$$LDM := \begin{bmatrix} \Lambda(\Omega_1) & \delta(\Omega_1,\Omega_2)/2 & \cdots & \delta(\Omega_1,\Omega_n)/2 \\ \delta(\Omega_2,\Omega_1)/2 & \Lambda(\Omega_2) & \cdots & \delta(\Omega_2,\Omega_n)/2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta(\Omega_n,\Omega_1)/2 & \delta(\Omega_n,\Omega_2)/2 & \cdots & \Lambda(\Omega_n) \end{bmatrix} (1)$$

An important characteristic of the LDM is that the sum of the values across the i-th column and the sum of the values across the i-th row are both equal to the number of electrons within the i-th atomic basin, which can be stated mathematically as follows:

$$\sum_{x=1}^{n} ldm_{ix} = \sum_{x=1}^{n} ldm_{xi} = N(\Omega_i) \ (2)$$

where $N(\Omega_i)$ is the number of electrons localized on atomic basin i.

An additional summation relation of the LDM gives that the sum of all values in the LDM is equal to the total number of electrons in the system:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} ldm_{ij} = N \ (3)$$

It is for this reason that the delocalization indexes are halved. They are accounted for twice: once in the ij-th entry, and once in the ji-th entry.

One parameter that appears quite frequently in this thesis, as well as other work on LDMs, is the Frobenius distance between LDMs. It defined as the Frobenius norm of the

distance matrix. The distance matrix C=A-B is defined as the matrix that contains values $c_{ij}=a_{ij}-b_{ij}$, *i.e.,* all of the corresponding matrix entries are simply subtracted. The Frobenius distance between A and B is then simply the Frobenius norm of C, which can be represented symbolically as follows:

$$\left|\left|A-B\right|\right|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|c_{ij}\right|^2} \quad (4)$$

where A and B are LDMs, and $c_{ij}$ is the ij-th element of C:=A-B. The Frobenius distance appears frequently because it is a single number measure of the magnitude of difference between two matrices, which in the case of LDMs, can be used to quantify the difference in electron topology between two molecules.

### *1.2.2. Previous Applications*

The LDM has seen a number of applications in recent literature. Matta initially conceived of the LDM as a QSAR tool for drug design, and as such, there has been work conducted in using the LDM for prediction of organic molecular properties.[21-22] Cook used LDM eigenvectors for the prediction of toxicity of a number of anti-mosquito compounds, which showed a linear relationship between an optimized linear combination of eigenvector components and experimental toxicity.[24] Matta has previously extracted proton NMR coupling constants from LDMs.[18] Sumar *et al* used LDM Frobenius distances to accurately predict the pKa and maximum absorption wavelength of a series of substituted benzoic acids.[23]

### *1.3. Dimerization of 1,3,5,7-Cyclooctatetraene (COT)*

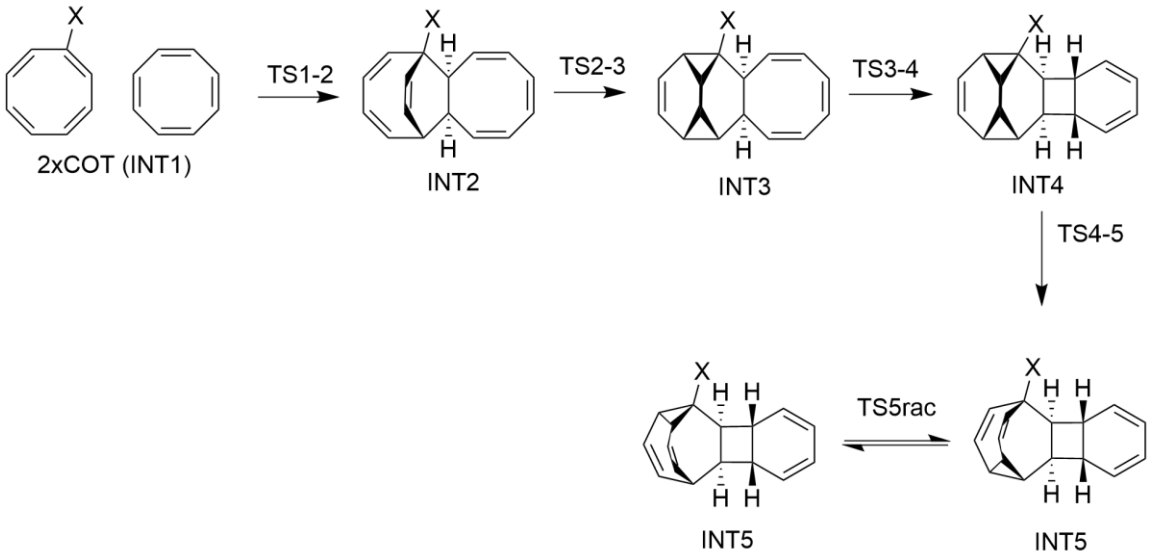
#### *1.3.1. Dimerization of Unsubstituted 1,3,5,7-Cyclooctatetraene (COT)*

1,3,5,7-Cyclooctatetraene (COT), also known as [8]annulene, is a nonaromatic cyclic unsaturated hydrocarbon. It is known to undergo a thermal dimerization reaction at 100°C to yield an interesting product **INT5** (see scheme 1.3.1.1) in approximately 40% yield, which is chiral and undergoes rapid racemization. This reaction was first reported by W.O. Jones in 1953,[25] and the structure of the product was first confirmed by Schröder in 1964.[26]

Computational studies of the mechanism of this reaction suggest that it proceeds through four transition states: first, an intermolecular *endo* Diels-Alder reaction occurs between two equivalents of COT yielding dimer **INT2** which is known to be isolable;[25-26] second, an intramolecular Diels-Alder reaction yields **INT3**; third, a disrotatory 6π electrocyclic ring closure yields **INT4**; and finally, an interesting sigmatropic rearrangement yields **INT5**, which racemizes readily (see scheme 1.3.1.1).[7] The rate determining step of this reaction is thought to be the sigmatropic rearrangement, but the initial intermolecular Diels-Alder reaction also has quite a large energetic barrier.[7]

This reaction has some important properties that make it useful for exploring the applications of the LDM. Firstly, it has 100% atom economy, which is beneficial as it allows for consistent unambiguous atom labelling throughout the entire duration of the reaction mechanism, which is advantageous because the LDM suffers from labelling ambiguity that can result in some values extracted from the matrix to change as labels are switched, such as matrix determinants. Secondly, the reaction mechanism has a large number of stationary points, which is advantageous

because each stationary point represents one LDM. Therefore, many data points can be extracted from one reaction, allowing better elucidation of trends, and better precision of results.




Scheme 1.3.1.1. COT dimerization mechanism proposed by Castanedo *et al.* (X=H)[7] and monohalogenated reaction mechanism studied in this work (X=F, Cl, Br). Labels given are consistent with previous work and used throughout this thesis.

An interesting trend was discovered in the LDM data of this reaction: it was found that the Frobenius distances between each stationary point and **INT1** correlated linearly ($R^2$=0.9795) to the intrinsic reaction coordinates,[7] which suggests that the set of LDMs for the reaction mechanism completely encodes the reaction mechanism, or rather that the magnitude of the change of the electron density function throughout the mechanism is proportional to progress throughout the reaction. Indeed, it is an interesting result, and confirmation of its generality would be useful.

### *1.3.2. Dimerization of Substituted 1,3,5,7-Cyclooctatetraenes (COTs)*

Little work exists on the dimerization of COT and the resulting products. Even fewer results are known for the dimerization of substituted COTs. Although substituted COTs are occasionally used as transition metal and *f*-block ligands,[27] the thermal dimerization of substituted COTs has not yet been reported experimentally. Nevertheless, the set of reactions remains of theoretical interest.

The crossed reaction of monohalogenated COT and unsubstituted COT was selected for the exploration into the utility of the LDM for the analysis of reaction mechanisms contained herein for several important reasons. Firstly, the set of reactions maintains all of the benefits of the unsubstituted dimerization as described above. Halogen substituents were chosen because they contribute a large change in electronics while still being monoatomic, which allows for the maintenance of consistent labelling, while still providing interesting differences between each reaction. The crossed reaction was chosen to keep the number of heteroatoms in the total system small, as large quantities of heteroatoms contribute significantly to computational cost, as cost scales rapidly with the total number of electrons. The project started with the goal of analyzing every monohalogenated dimer shown in scheme 1, but due to time constraints, only the *fluoro* and *bromo* compounds were successfully analyzed. A discussion of unsuccessful computations is given in section 3.4.

The reaction is predicted to proceed initially *via* a Diels-Alder reaction, and so the position of the halogen atom on the diene and the electron-donating alkyl groups on the dienophile would make this first step an inverse electron demand Diels-Alder reaction.

Positioning of the halogen in the vinylic position of the diene allows mesomeric electron donating effects, and inductive electron withdrawing effects. One might postulate that the heavier halogens will be stronger mesomeric donating groups, and weaker inductive withdrawing groups, as they are more polarizable and less electronegative, and so it could be predicted that the reactivity of this initial step will slow as the size of the halogen is increased.

The goal of this research was to explore novel parameters extracted from LDM data to predict and model chemical reaction mechanisms, which was accomplished by theoretical analysis of monohalogenated crossed COT dimerizations as shown in Scheme 1.3.1.1. Computations were attempted for fluorinated, brominated, and chlorinated COT crossed with nonhalogenated COT dimerizations.

## 2. Experimental

### *2.1. Stationary Point Optimization*

Unsubstituted COT dimerization stationary point geometries were obtained from the supplementary information in citation 7, and modified using the Spartan'18 software package.[28] Halogen atoms were added to the structure in the position shown in scheme 1, and the structures were reoptimized using the semiempirical PM3[29-31] level of theory built into Spartan'18, and then again reoptimized using DFT with the ωB97X-D[6]/6-311+G(d,p)[32-37] functional and basis set. Vibrational analysis was conducted, and intermediates were confirmed to have no imaginary vibrational frequencies, and transition states were confirmed to have exactly one imaginary frequency corresponding to the correct reaction coordinate. Geometries from Spartan'18 were exported and wavefunctions and vibrational analysis was recomputed and reanalyzed with the same functional and basis set in Gaussian 16 for compatibility with other software.[38] Geometric coordinates, LDMs, and IRC log files are available at request to Dr. Kai Ylijoki. Computed thermodynamics are given in Table 7.1 in the Appendix.

### *2.2. Intrinsic Reaction Coordinate Computation*

IRCs were computed in Gaussian 16 using the Hratchian predictor-corrector method built into Gaussian starting from previously optimized stationary points.[10,38] The maximum number of points was increased such that the computation proceeded to the minimum, allowing for the total reaction path to be mapped. IRCs for each reaction were stitched together by setting INT1 equal to the origin, then plotting each IRC point relative to this.

### *2.3. Generation of Localization Delocalization Matrices*

Reoptimized geometries from Gaussian were exported as formatted checkpoint files, and AIM parameters were evaluated with the AIMAll software package.[39] AIMAll results were validated according to the PH relation. LDMs were generated using the AIMLDM software package.[40] Frobenius distances were computed manually using Microsoft Excel.[41]

## 3. Results and Discussion

### *3.1. Reaction Mechanism of Fluorinated Crossed COT Dimerization*

All stationary points of the crossed dimerization of fluoro-COT with COT shown in scheme 1.3.1.1 were successfully computed and confirmed to have the correct number of imaginary frequencies. Like the unsubstituted case, the reaction proceeds initially *via endo* Diels-Alder reaction, followed by intramolecular Diels-Alder reaction, then disrotatory 6π electrocyclization, then sigmatropic rearrangement, and finally reversible sigmatropic isomerization between constitutional isomers (the racemization step in the unsubstituted case becomes a constitutional isomerization due to the asymmetry of the monohalogenated product) (see scheme 1.3.1.1). The relative energies of each species are given in figure 3.1.1. Note that although the final transition state, **TS5rac**, is now a constitutional isomerization, it is still labelled as a racemization to keep labels consistent with previous work.

As predicted in section 1.3.2, the energy barrier of the initial Diels-Alder reaction step is lower than the energy reported for the unsubstituted initial Diels-Alder reaction by approximately 8 kJ/mol, which follows the prediction that the fluorine will act mostly as an inductive electron withdrawing group.

Interestingly, the intramolecular Diels-Alder (**TS2-3**) reaction barrier is approximately 3.5 kJ/mol higher than the reported value for the unsubstituted case. This can be rationalized as the fluorine is now in an inductive electron withdrawing position between both the diene and the dienophile, and so the HOMO of both might become lower in energy, and the LUMO of both

might become higher in energy, thus decreasing reactivity. The electrocyclization (**TS3-4**) shows a destabilization of approximately 2.4 kJ/mol compared to the unsubstituted case.

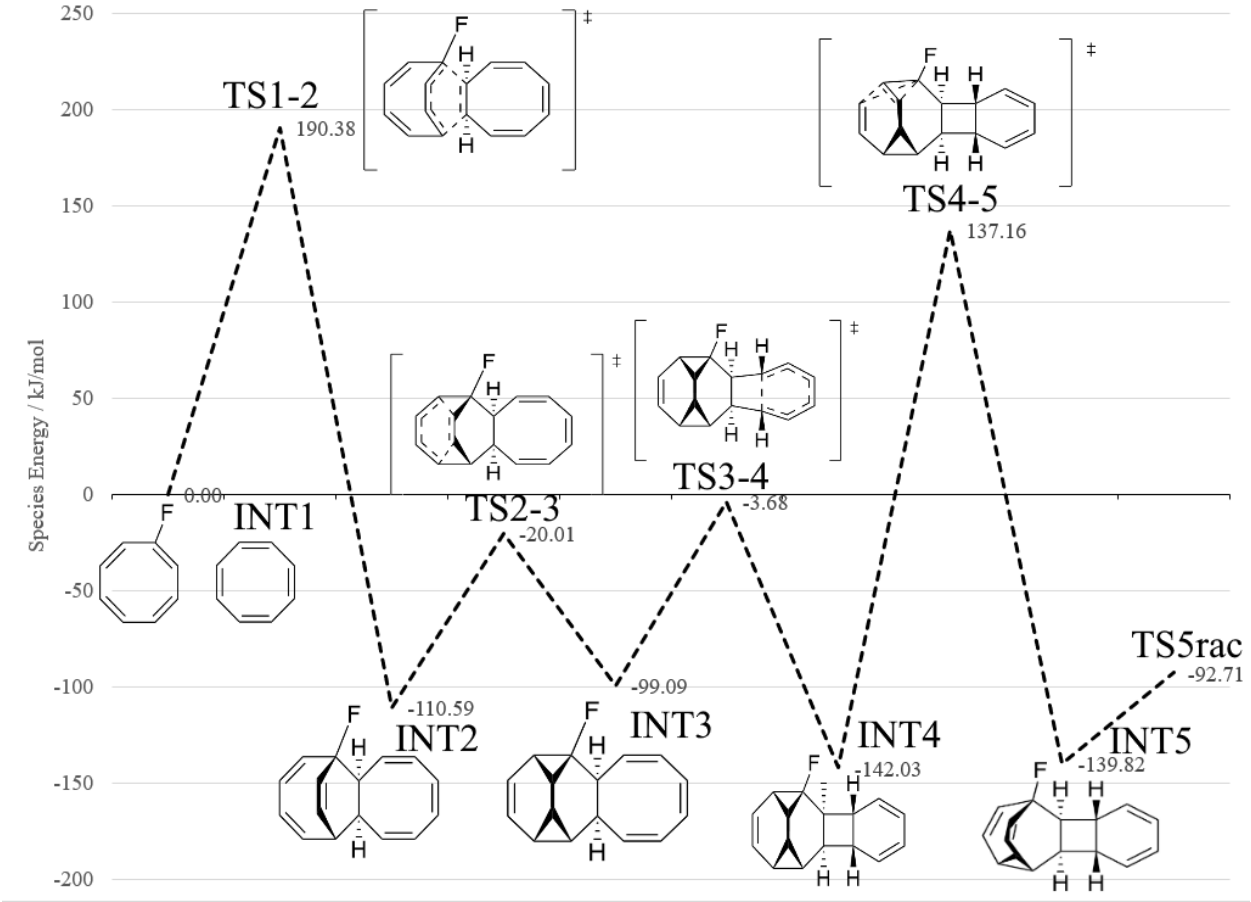



Figure 3.1.1. Relative total energies along stationary points in the fluorinated reaction mechanism.

The sigmatropic rearrangement (**TS4-5**) has an energy barrier approximately 4.4 kJ/mol higher in energy than the unsubstituted case, which is intuitive – the σ-bond in the cyclopropane ring bonded to the fluorine donates its electrons in this step, and so when the inductive withdrawing group is added, the bond becomes more electron deficient and so less reactive. This reaction is thought to proceed *via* a diradical mechanism, but either a closed- or open-shell

mechanism must have the bond electrons nearest the fluorine react as a nucleophile (see Fig. 3.1.2).

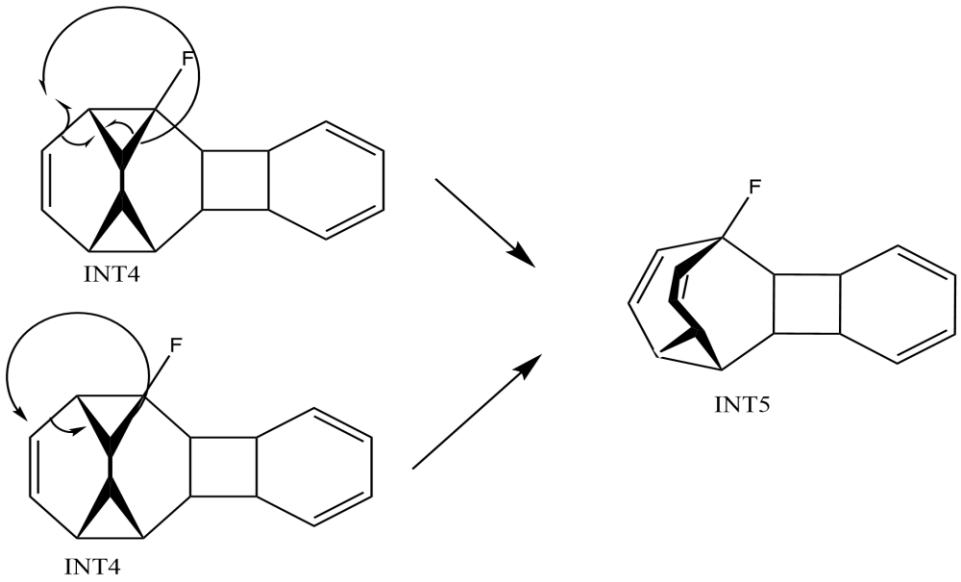



Figure 3.1.2. Arrow pushing diagrams of the fluorinated sigmatropic rearrangement (**TS4-5**).

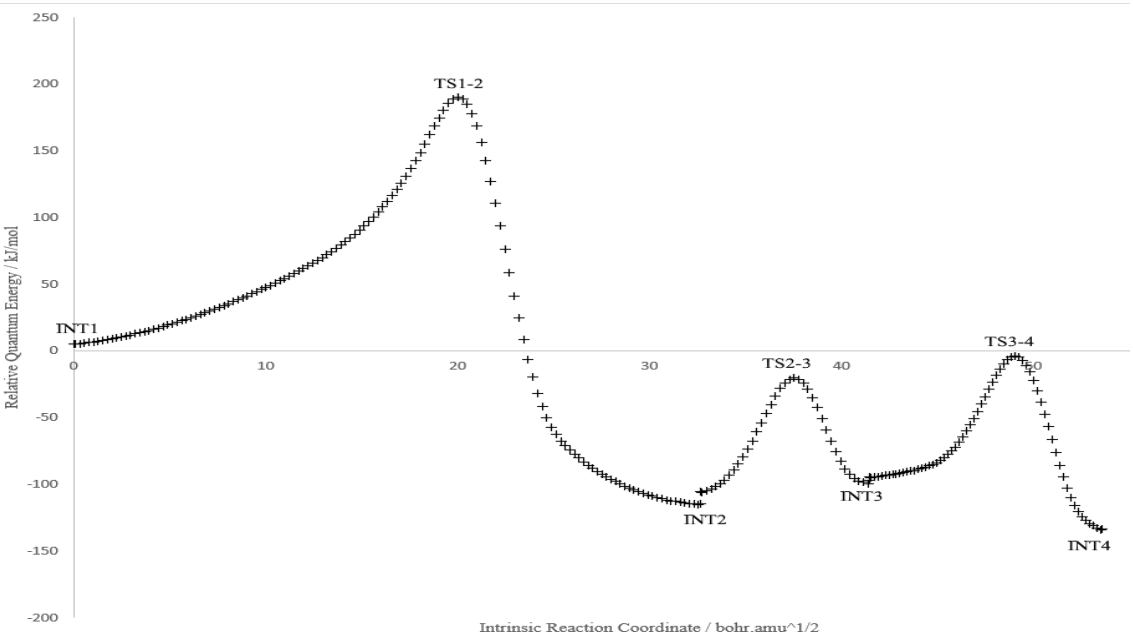



Figure 3.1.3. Stitched IRC plot for crossed fluorinated COT dimerization. Energies taken relative to **INT1**. Discontinuities are due to stitching method.

Intrinsic reaction coordinates were computed for the fluorinated case up to **INT4**. The IRC computation for **TS4-5** resulted in problems, which are discussed in Section 3.4. The stitched IRC plot for the fluorinated case is shown in Figure 3.1.3. This plot represents a slice of the PES in the direction of reaction progress.

Frobenius distances taken along the reaction coordinate up to **INT4** show linear correlation similar to the nonhalogenated reaction (see Fig. 3.1.4). The coefficient of determination, $R^2$, is slightly closer to unity ($R^2$=0.9984 versus $R^2$=0.9795), and the slope is slightly higher at 0.0414 bohr$^{-1}$ amu$^{-1/2}$ versus the literature unsubstituted result of 0.03764 bohr$^{-1}$ amu$^{-1/2}$.[7] The increased slope likely corresponds to an increased rate of change in electron topology in the fluorinated reaction, which is a fascinating and thought-provoking preliminary result, as it suggests that there is some factor which influences the rate at which electron topologies change with respect to nuclear coordinates, and that this rate is not constant for all reaction mechanisms.

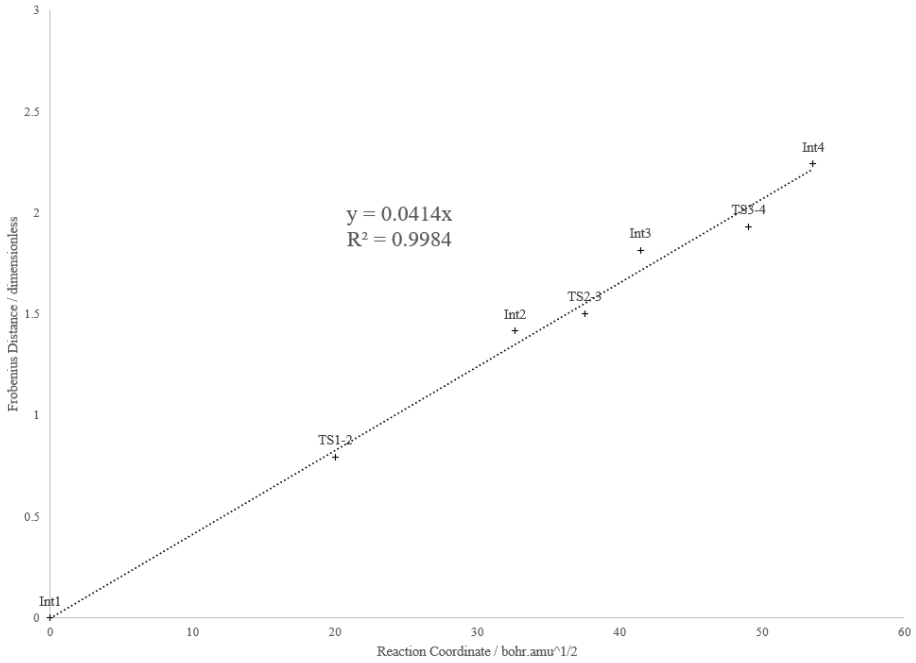



Figure 3.1.4. Frobenius distances along reaction coordinate for the fluorinated mechanism.

*3.2. Reaction Mechanism of Brominated Crossed COT Dimerization*

The stationary points for the dimerization mechanism of bromo-COT with COT shown in scheme 1.3.1.1 were successfully computed. The mechanism shows the same four steps as above, as well as the isomerization step. Relative total energies are given in Figure 3.2.1. As predicted in Section 1.3.2, the energetic barrier for the initial Diels-Alder reaction (**TS1-2**) is approximately 6 kJ/mol higher than the unsubstituted reaction, and approximately 14 kJ/mol higher than the fluorinated reaction. Following from the explanation in Section 1.3.2., this suggests that in this case, bromine has a stronger mesomeric electron donating effect than fluorine, and a weaker inductive electron withdrawing effect.

The transition states of the brominated intramolecular Diels-Alder (**TS2-3**) and electrocyclization (**TS3-4**) reactions both show stabilization of approximately 20 kJ/mol compared to the unsubstituted case. Stabilization is the reverse of the result for the fluorinated mechanism, which showed **TS2-3** and **TS3-4** increasing in energy by a small amount.

Again, the sigmatropic rearrangement step (**TS4-5**) is higher in energy than the unsubstituted reaction; this time by nearly 58 kJ/mol, which can be explained, as in Section 3.1, the mechanism requires the σ-bond nearest the halogen to donate its electrons, thus reducing the reactivity as a function of the halogen's inductive electron withdrawing capacity. The reduced rate of this step relative to the equivalent fluorinated step is likely due to reduced ring strain in the cyclopropane ring of **INT4** as a result of the increased polarizability of bromine compared to fluorine and hydrogen allowing less rigidity. The effect on ring strain is confirmed by the increased C-(C-X)-

C bond angle in the three-membered ring of the transition state being increased from 49.01° in the fluorinated species to 56.68° in the brominated species.

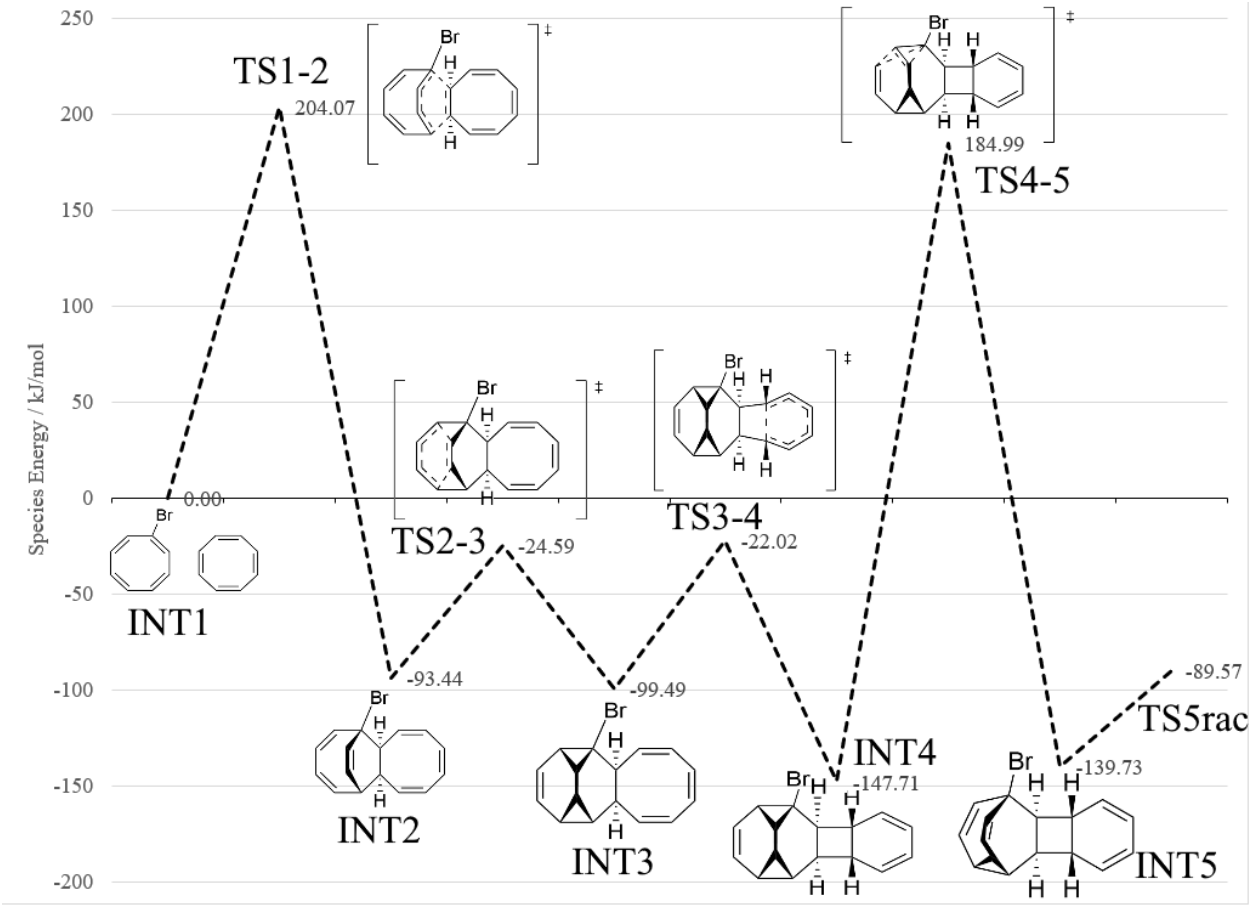



Figure 3.2.1. Relative total energies along stationary points in the brominated mechanism.

IRC computations for the crossed bromo-COT dimerization proved quite problematic, and as such the longest continuous IRC that could be stitched together only ranged from **INT1** to **INT2**. The reader is reminded that problematic computations are discussed in Section 3.4. Nonetheless, data can be extracted from this case, although the results are not solidified quite as well as those for the fluorinated reaction discussed above. Further studies are required to validate

these results to the same level of confidence as the fluorinated reaction. The stitched IRC plot is shown in Figure 3.2.2.

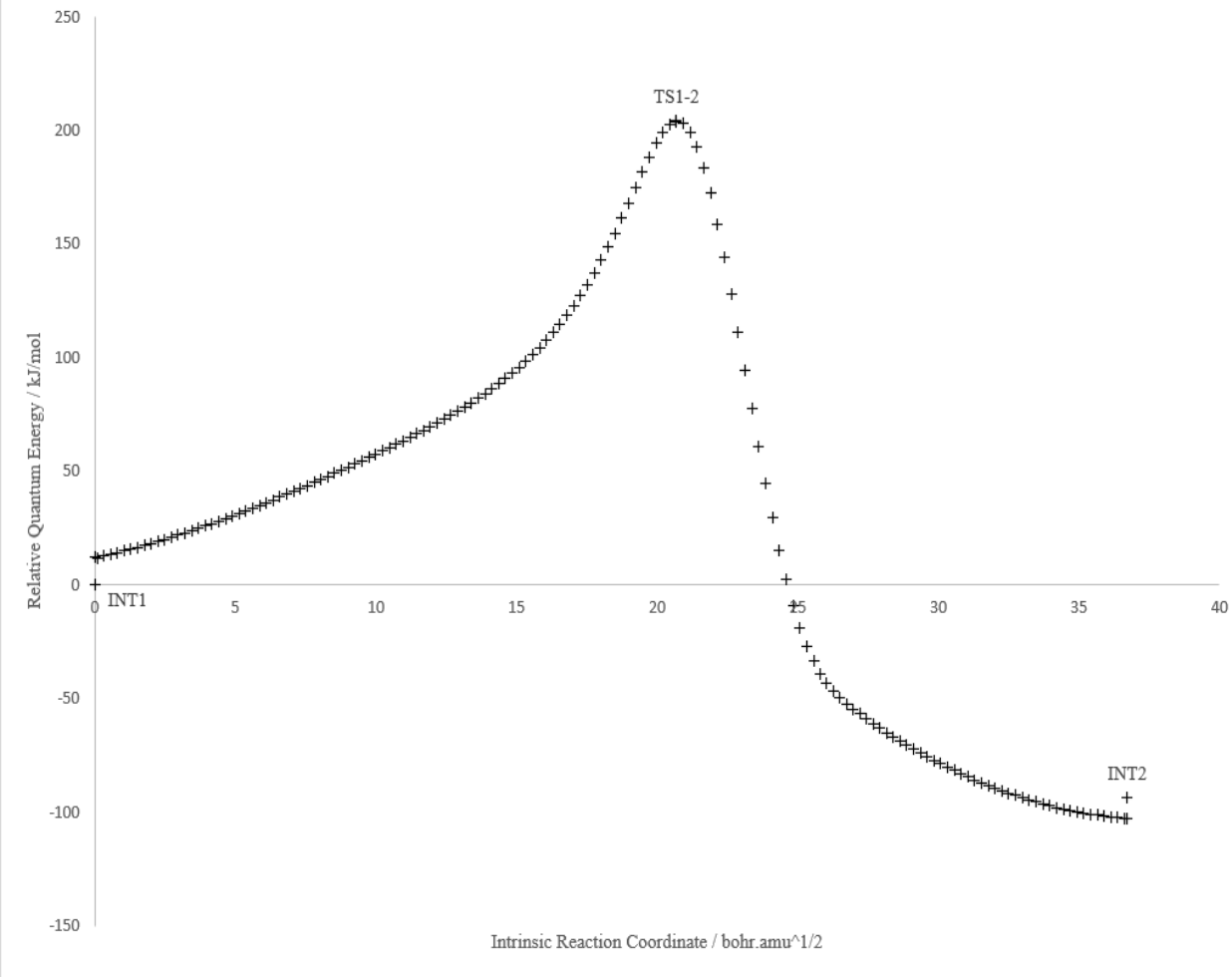



Figure 3.2.2. Stitched IRC plot for crossed brominated COT dimerization. Energies taken relative to **INT1**. Discontinuities are due to the stitching method.

Again, the plot of Frobenius distances across the IRC was linear, extending this trend further (see fig. 3.2.3). The coefficient of determination was smaller than the fluorinated case ($R^2$=0.9905), but still larger than the nonhalogenated case. Interestingly, the slope was less than the nonhalogenated slope at 0.0370 bohr$^{-1}$ amu$^{-1/2}$ versus 0.03764 bohr$^{-1}$ amu$^{-1/2}$,[7] which implies

that the electron topology is changing slower in the brominated case, and is in contrast to the fluorinated case, which showed more rapid change. As of yet, no data has been presented in the literature or in this work that demonstrates non-linearity of this relationship. The linearity of these results supports the LDM's effectiveness as a measure of the similarity between chemical species.

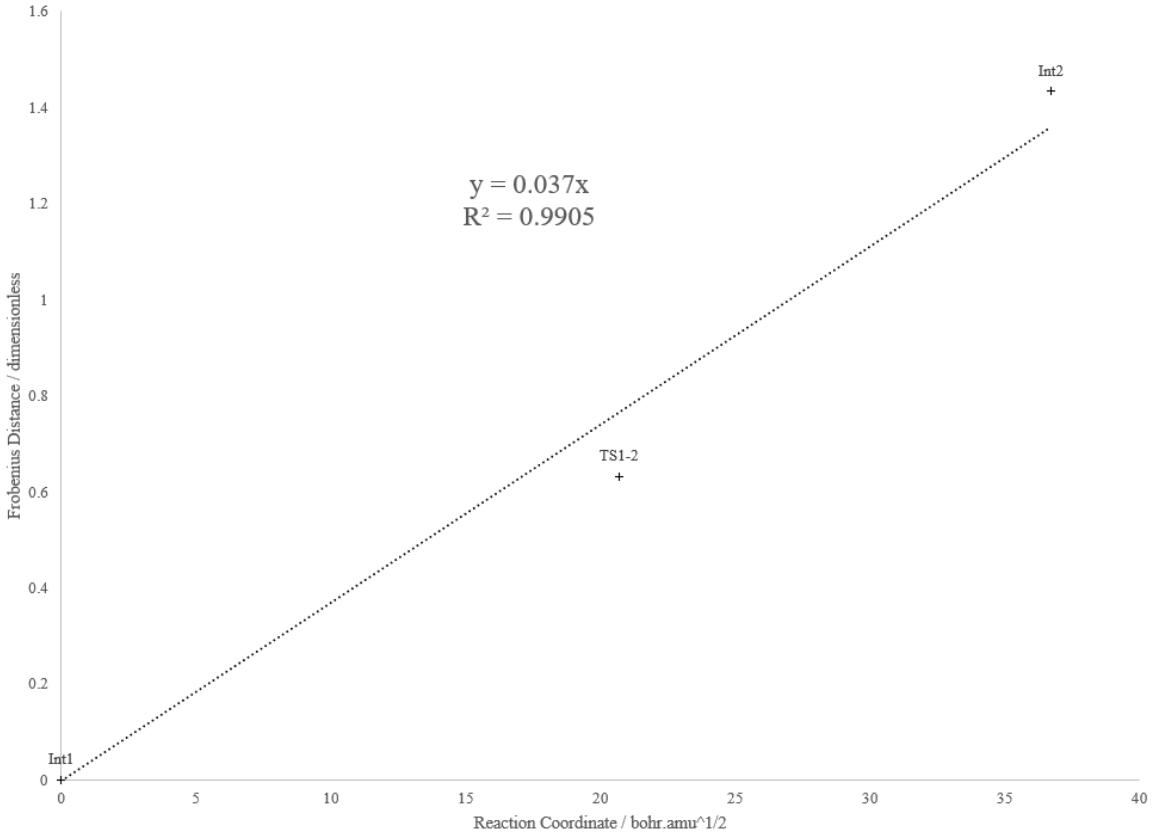



Figure 3.2.3. Frobenius distances along IRC for the brominated mechanism.

*3.3. LDM Analysis and Periodic Trends*

In an attempt to use LDMs to find some metric to totally quantify a reaction mechanism, Frobenius distances going down the halogen group were taken across stationary points of the

mechanism (i.e., $||INT1_F\text{-}INT1_H||$). These are shown in Figure 3.3.1. They varied throughout the course of the reaction, but remained near a value approximately equal to the difference in electrons between the corresponding halogen and hydrogen (i.e., 8 for fluorine, 34 for bromine). This makes intuitive sense, as there will be an increased contribution of core electrons for elements further down the periodic table, and so the Frobenius difference between species should correspond to the difference in number of electrons, plus a topological contribution.

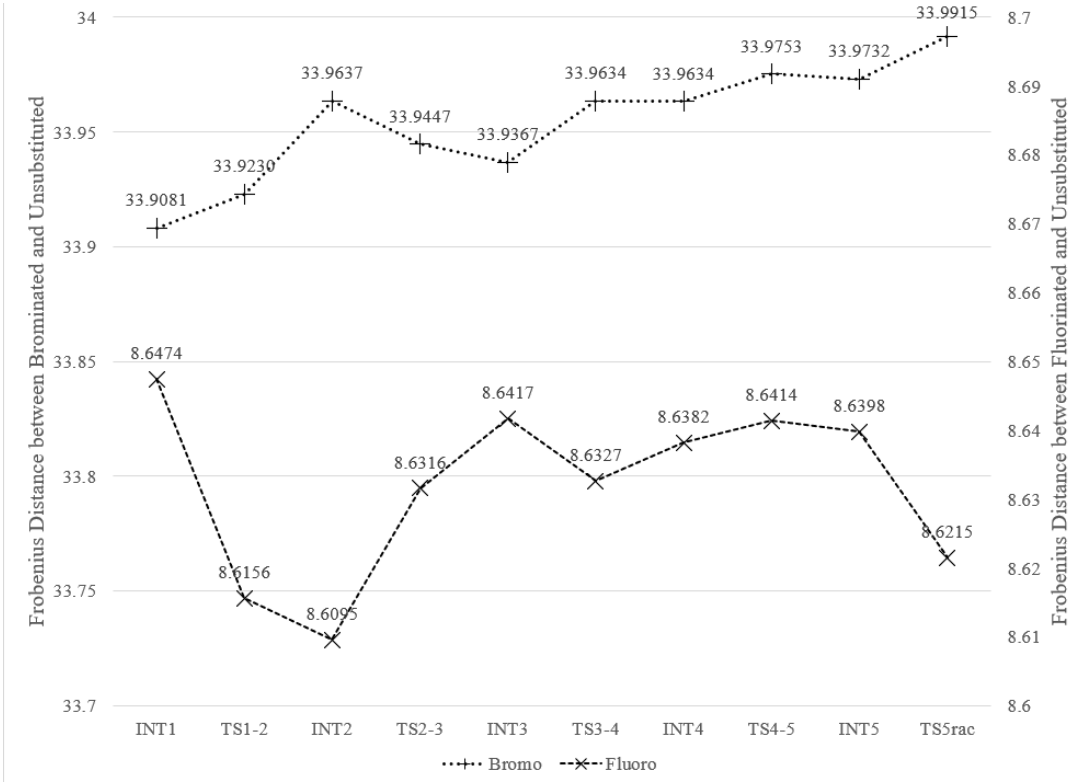



Figure 3.3.1. Frobenius distances down the periodic table (i.e., $||INT1_F - INT1_H||$, $||INT1_{Br} - INT1_H||$)

This topological contribution should be mathematically isolable simply by subtracting the corresponding number of electrons from the Frobenius distance. This was completed, and the results are shown in Figure 3.3.2. The topological contribution for the fluorinated reaction shows

a mean of 0.6319 over the entire course of the reaction, with a standard deviation of 0.0125; the topological contribution for the brominated reaction has a mean of -0.0457 and a much larger standard deviation of 0.0257. One might speculate that the increased spread of the topological contribution of the brominated species is due to increased polarizability of bromine, which allows for enhanced mechanistic freedom, *i.e.,* it can more easily change its magnitude of electron donating or withdrawing, depending on whichever is thermodynamically stable.

The mean of topological contributions being closer to zero for the brominated case seems to imply that the brominated reaction mechanism is more similar to the nonhalogenated reaction mechanism than the fluorinated reaction mechanism is to the nonhalogenated. This is likely not an energetic similarity, because the brominated energetics showed much larger differences than the fluorinated energetics, and so it seems to be purely a similarity in the electron topology.

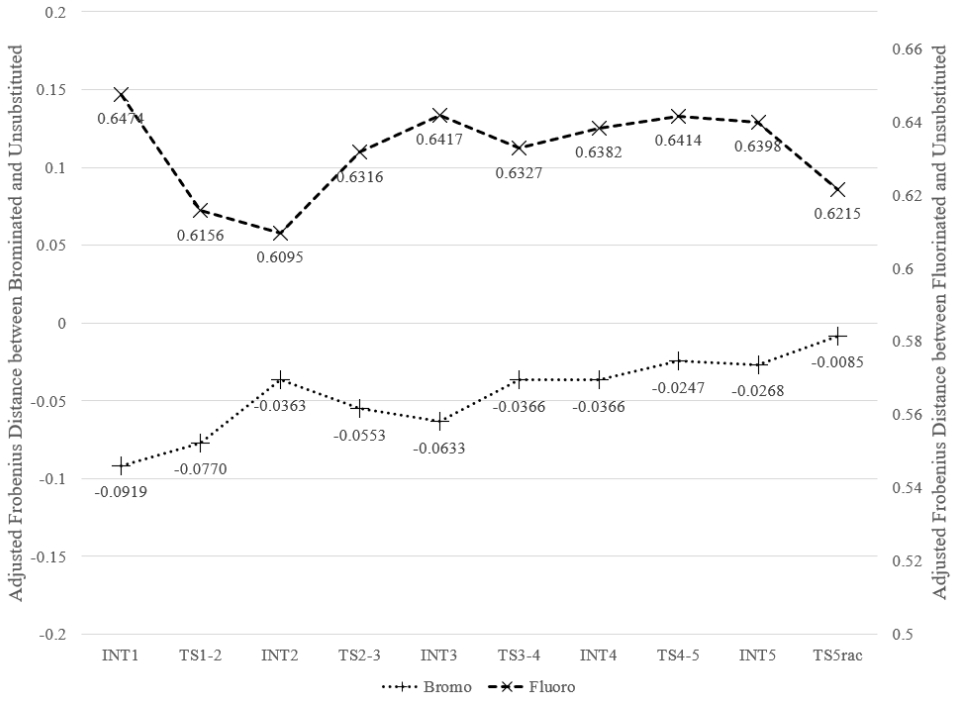



Figure 3.3.2. Frobenius distances down the periodic table, minus electronic contributions.

This similarity between the topology of the complete mechanism of the brominated dimerization and the complete mechanism of the nonhalogenated dimerization is symmetrical to the similarity in the rate of change of the individual mechanistic electron topologies shown in Section 3.2. In fact, when the mean topological contribution is plotted versus rate of topological change over reaction mechanism, a linear relation is obtained (see Fig. 3.3.3).

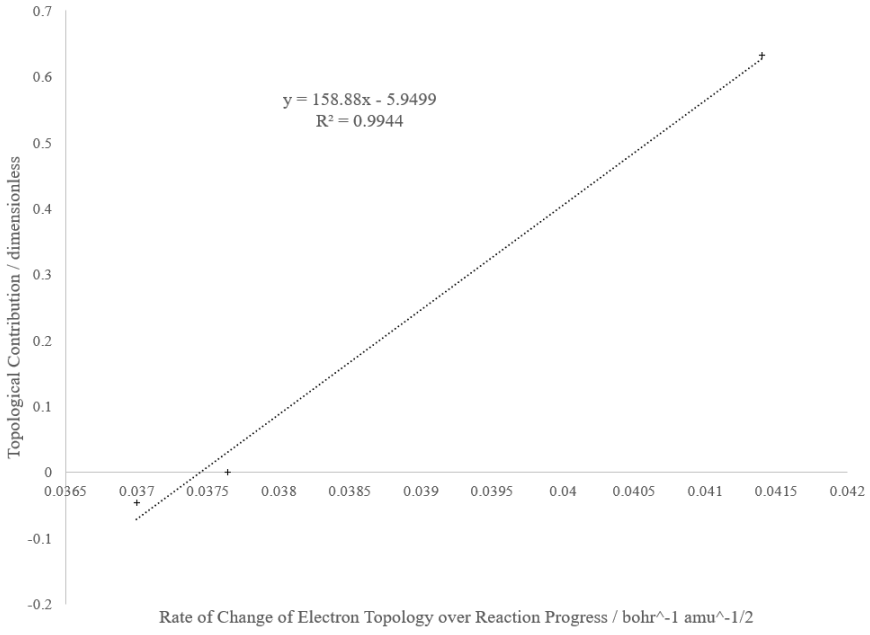



Figure 3.3.3. Linear relationship between mean topological contribution and rate of change of electron topology over reaction coordinate.

This may suggest that this mean topological contribution is an effective measure of the similarity of reaction mechanisms, but further work is required.

Rate of change of electron topology over reaction coordinate is a measure of the tendency for the electrons to change as a function of nuclear change (recall that intrinsic reaction coordinate is a function of nuclear coordinates). The mean topological contribution is a measure of the total

electronic topological difference between two differently substituted reaction mechanisms. Their linear correlation implies that substituent effects play a role in the rate of change of electrons as a function of change in nuclei. This may seem intuitive to a synthetic chemist, and here, this intuition is quantified.

The results described in this section suggest that in this set of reactions, the brominated reaction mechanism is more similar to the unsubstituted mechanism than the fluorinated mechanism is to the unsubstituted mechanism. In Section 3.2, it was suggested that bromine is acting primarily as a mesomeric electron donating group, and a weak inductive electron withdrawing group. In Section 3.1, it was suggested that fluorine is acting primarily as a strong inductive electron withdrawing group and a weak mesomeric electron donating group. Indeed, the largest magnitude of topological difference between bromine and hydrogen seems to be in the initial Diels-Alder step (**TS1-2**), where mesomeric effects are predicted to cause the largest change. The inductive electron withdrawing capability of fluorine might be able to affect every step approximately an equal amount, which is reflected in the small spread of the topological contributions for that mechanism. Notably, the smallest magnitude of topological difference between fluorine and hydrogen is also shown in the initial Diels-Alder, further solidifying this trend, as the mesomeric electron donating effects in this step bring the mechanism closer to parity with the nonhalogenated one.

This trend suggests that electron donating groups show a negative difference in topological contribution, and electron withdrawing groups show a positive difference in topological contribution, which suggests a utility of the topological contribution for predicting electron donating/withdrawing capabilities of substituents, as well as mechanistic similarity, but more

work needs to be conducted to further examine substituent effects on the topological contribution.

### *3.4. Chlorinated Crossed COT Dimerization and Other Problematic Computations*

Geometry optimizations of the chlorinated crossed COT dimerization stationary points were attempted at least three separate times. On every occasion, one of three errors were encountered in Spartan'18: the geometry would fail to converge, the geometry would converge to a ridiculous structure with inappropriate bond lengths and angles, or the geometry would converge but the frequency analysis would fail by showing a large number of imaginary frequencies with very high magnitudes.

Several methods were used in an attempt to solve the geometry optimization properly: the nonhalogenated structures were modified to add the chlorine, then optimized at the PM3 level of theory, then reoptimized at the level of theory used for the remainder of this work (this is the method used for all other successful geometry optimizations); Fluorinated nonhalogenated structures were modified, then processed in a similar way; and chlorinated structures were constructed from scratch and then optimized. All of these methods resulted in the same three failure modes. A colleague of the author allegedly solved these optimizations using Spartan '20, implying that a software bug in Spartan '18 was the problem. However, due to time constraints, this could not be verified, and further computations were not completed, so they are not presented as results in this work.

Another set of problematic computations were the IRCs. The brominated IRCs tended to cause more problems than the fluorinated IRCs, which is likely owing to the increased

computational complexity of bromine. In total, five IRC computations were unable to be completed: fluorinated **TS4-5** in reverse, brominated **TS5rac** in reverse, brominated **TS4-5** in reverse, brominated **TS3-4** in reverse, and brominated **TS2-3** in reverse. It is likely simply coincidental that all of the problematic IRCs are the reverse, as forward and reverse is assigned arbitrarily by Gaussian. As such, all of these corresponding transition states were only validated by IRC in the forward direction.

Three IRC failure modes were most common: convergence to the forward path despite reverse/forward keywords; declaration by the IRC algorithm that the transition state was a PES minimum, despite being confirmed to be a first order saddle point; and recorrector non-convergence. Several strategies were employed in attempts to solve these problems. Vibrational analyses were computed separately, and then read into the IRC computation using the "irc=rcfc" keyword in Gaussian, so that no sign flipping occurred resulting in forward path computation. Recorrector steps were increased using the "irc=maxcycle" keyword. Stepsize was varied from 0.05 bohr $amu^{1/2}$ to 0.5 bohr $amu^{1/2}$. Disabling recorrection usually resulted in failure by the second mode, but sometimes by the first mode. Unfortunately, none of these strategies solved the IRC problems, and these five IRC computations remain uncompleted and in progress.

## 4. Conclusion

The reaction mechanisms of the crossed bromo-COT and fluoro-COT dimerizations are predicted to proceed very similarly to the nonhalogenated COT dimerization. The mean topological contribution has been isolated as a potential measure of reaction mechanism similarity and is correlated to the rate of change of electron topology of reaction mechanisms. It is advantageous because it requires only that two mechanisms share the same number of atoms to be compared. Some evidence exists that suggests that the mean topological contribution is related to electron withdrawing/donating effects, but more work must be completed to further elucidate the substituent effects on mean topological contribution. The mean topological contribution is a significant potential candidate for rapid analysis of chemical reaction mechanisms.

## 5. Future Work

More work must be conducted to further elucidate the utility of the mean topological contribution. This could include further studies of substituent effects on the dimerization of COT and should also include studies into other reaction mechanisms. Experimental confirmation of halogenated COT dimerizations would allow for enhanced understanding of the dynamics of this reaction, and possibly reveal novel utility of the mean topological contribution. An example of this would be a Hammett type study into the substituent effects.

## 6. References


1. Hohenberg, P.; Kohn, W. Inhomogenous Electron Gas. *Phys. Rev. B*. **1964**, 136, 3, 864-871.

2. Kohn, W.; Sham, L.J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev. A.* **1965**, 140, 4, 1133-1138.

3. Geerlings, P.; De Proft, F.; Langanaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, 103, 1793-1873.

4. Kohn, W.; Becke, A.D.; Parr, R.G. Density Functional Theory of Electronic Structure. *J. Phys. Chem*. **1996**, 100, 12974-12980.

5. Becke, A.D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, 98, 2, 1372-1377.

6. Chai, J.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys*. **2008**, 44, 10, 6615-6620.

7. Castanedo, L.A.M.; Matta, C.F.; Ylijoki, K.E.O. The reaction path of cyclooctatetraene dimerization revisited. *Int. J. Quantum Chem.* **2022**, 122, 7, e26866. https://doi.org/10.1002/qua.26866 (accessed 30 September 2023).

8. Cramer, C.J. Essentials of Computational Chemistry - Theories and Models, 2nd ed. John Wiley & Sons, Ltd., 2004; pp 6.

9. Cramer, C.J. Essentials of Computational Chemistry - Theories and Models, 2nd ed. John Wiley & Sons, Ltd., 2004; pp 40.

10. Hratchian, H.P.; Schlegel, H.B. Accurate reaction paths using a Hessian based predictor-corrector integrator. *J. Chem. Phys*. **2004**, 120, 21, 9918-9924.

11. Garcia-Domenech, R.; Galvez, J.; de Julian-Ortiz, J.V.; Pogliani, L. Some New Trends in Chemical Graph Theory. *Chem. Rev.* **2008**, 108, 1126-1169.

12. Basak, S.C.; Mills, D.R.; Balaban, A.T.; Gute, B.D. Prediction of Mutagenicity of Aromatic and Heteroaromatic Amines from Structure: A Hierarchical QSAR Approach. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 671-678.

13. Oberg, K.; Berglund, A.; Edlund, U.; Eliasson, B. Prediction of Nonlinear Optical Responses of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 811-814.

14. Pogliani, L. How Far Are Molecular Connectivity Descriptors from Is Molecular Pseudoconnectivity Descriptors?. *J. Chem. Inf. Comput Sci*. **2001**, 41, 836-847.

15. Bader, R.F.W. Atoms in Molecules. *Acc. Chem. Res*. **1985**, 18, 9-15.

16. Lee, K. The Poincare-Hopf Theorem and the Hairy Ball Consequence. http://simonrs.com/eulercircle/diffgeo/ken-poincarehopf.pdf (accessed 30 March 2024)

17. Fradera, X.; Austen, M.A.; Bader, R.F.W. The Lewis Model and Beyond. *J. Phys. Chem. A*. **1999**, 103, 304-314.

18. Matta, C.F.; Molecules as networks: A localization-delocalization matrices approach. *Comput. Theor. Chem.* **2018**, 1124, 1-14.

19. Sumar, I. Chemical Applications of Electron Localization-Delocalization Matrices (LDMs) with an emphasis on predicting molecular properties. M.Sc. Thesis, Saint Mary's University, Halifax, Nova Scotia, September 12, 2016.

20. Matta, C.F.; Boyd, R.J. An Introduction to the Quantum Theory of Atoms in Molecules. In The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design; Wiley Online Library, 2007. https://doi.org/10.1002/9783527610709.ch1

21. Matta, C.F. Modelling Biophysical and Biological Properties from the Characteristics of the Molecular Electron Density, Electron Localization and Delocalization Matrices, and the Electrostatic Potential. *J. Comput. Chem*. **2014**, 35, 1165-1198.

22. Matta, C.F. Localization-delocalization matrices and electron density-weighted adjacency matrices: new electronic fingerprinting tools for medicinal computational chemistry. *Future Med. Chem*. **2014**, 6, 13, 1475-1479.

23. Sumar, I.; Ayers, P.W.; Matta, C.F. Electron localization-delocalizaton matrices in the prediction of pKa's and UV-wavelengths of maximum absorbance of p-benzoic acids and the definition of super-atoms in molecules. *Chem. Phys. Lett*. **2014**, 612, 190-197.

24. Cook, R.L. Principal components of localization-delocalization matrices: new descriptors for modelling biological activities of organic compounds. Part I: mosquito insecticides and repellants. *Struct. Chem*. **2017**, 28, 1525-1535.

25. Jones, W.O. cycloOctatetraene Derivatives. Part I. Dimers of cycloOctatetraene. J. *Chem. Soc.* **1953**, 0, 2036-2040.

26. Schroder, G. Die Eigenschaften zweier dimerer Cyclooctateraene vom Schmp. 53 und 76°. *Chem. Ber.* **1964**, 97, 11, 3131-3139.

27. Roesky, P.W. Substituted Cyclooctatetraenes as Ligands in f-Metal Chemistry. *Eur J. Inorg. Chem.* **2001**, 7, 1653-1660.

28. Spartan'20 for Windows, Macintosh and Linux. https://www.wavefun.com/spartan (accessed 30 September 2023).

29. Stewart, J.J.P. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **1989**, 10, 2, 209-220

30. Stewart, J.J.P. Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem.* **1989**, 10, 2, 221-264.

31. Stewart, J.J.P. Optimization of parameters for semiempirical methods. III. Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J. Comput. Chem.* **1991**, 12, 3, 320-341.

32. Clark, T.; Chandrasekhar, J.; Spitznagel, G.W.; Von Rague Schleyer, P. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first row elements, Li-F. *J. Comput. Chem.* **1983**, 4, 3, 294-301.

33. Francl, M.M.; Pietro, W.J.; Hehre, W.J.; Binkley, S.; Gordon, M.S.; DeFrees, D.J.; Pople, J.A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second row elements. *J. Chem. Phys*. **1982**, 77, 3654-3665.

34. Krishnan, R.; Binkley, J.S.; Seeger, R.; Pople, J.A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys*. **1980**, 72, 650-654.

35. McLean, A.D.; Chandler, G.S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11-18. *J. Chem. Phys*. **1980**, 72, 5639-5648.

36. Spitznagel, G.W.; Clark, T.; von Rague Schleyer, P.; Hehre, W.J. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl. *J. Comput. Chem.* **1987**, 8, 8, 1109-1116.

37. Curtiss, L.A.; McGrath, M.P.; Blaudeau, J.P.; Davis, N.E.; Binning, R.C.; Radom, L. Extension of Gaussian-2 theory to molecules containg third-row atoms Ga-Kr. *J. Chem. Phys*. **1995**, 103, 6104-6113.

38. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.V.; Bloino, J.; Janesko, B.G.; Gomperts, R.; Mennucci, B.; Hratchian, H.P.; Ortiz, J.V.; Izmaylov, A.F.; Sonnenberg, J.L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A. J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16, Revision C.01. https://gaussian.com/gaussian16/ (accessed 30 September 2023).

39. Keith, T.A. AIMALL. https://aim.tkgristmill.com/ (accessed 4 October 2023).

40. Matta, C.F.; Ayers, P.W.; Cook, R.; Sumar, I. AIMLDM: A program to generate and analyze electron localization-delocalization matrices (LDMs). *Comput. Theor. Chem*. **2015**, 1070, 55-67.

41. Microsoft Corporation. Microsoft Excel, version 2024. office.microsoft.com/excel (accessed 4 March 2024).

## 7. Appendix

Table 7.1. Computed thermodynamics.

| Species | Electronic Energy / hartrees | Thermal Energy Corrected / hartrees | Thermal Enthalpy Corrected / hartrees | Thermal Free Energy Corrected / hartrees |
|---|---|---|---|---|
| Fluoro-INT1 | -718.356039 | -718.079893 | -718.078949 | -718.142914 |
| Fluoro-TS1-2 | -718.283526 | -718.008743 | -718.007798 | -718.063747 |
| Fluoro-INT2 | -718.39816 | -718.118445 | -718.117501 | -718.169714 |
| Fluoro-TS2-3 | -718.363659 | -718.085899 | -718.04954 | -718.136025 |
| Fluoro-INT3 | -718.39378 | -718.113858 | -718.112914 | -718.163942 |
| Fluoro-TS3-4 | -718.357441 | -718.079206 | -718.078262 | -718.12813 |
| Fluoro-INT4 | -718.410137 | -718.129675 | -718.128731 | -718.179667 |
| Fluoro-TS4-5 | -718.303797 | -718.013815 | -718.012871 | -718.064321 |
| Fluoro-INT5 | -718.409294 | -718.128781 | -718.127836 | -718.179258 |
| Fluoro-TS5rac | -718.391349 | -718.112845 | -718.111901 | -718.163385 |
| Bromo-INT1 | -3192.687982 | -3192.414119 | -3192.413175 | -3192.475429 |
| Bromo-TS1-2 | -3192.610257 | -3192.336535 | -3192.335591 | -3192.393345 |
| Bromo-INT2 | -3192.723571 | -3192.444809 | -3192.443865 | -3192.498406 |
| Bromo-TS2-3 | -3192.697348 | -3192.420629 | -3192.419684 | -3192.472898 |
| Bromo-INT3 | -3192.725876 | -3192.44668 | -3192.445736 | -3192.499023 |
| Bromo-TS3-4 | -3192.696368 | -3192.419575 | -3192.418631 | -3192.470676 |
| Bromo-INT4 | -3192.744243 | -3192.464904 | -3192.46396 | -3192.517008 |
| Bromo-TS4-5 | -3192.617523 | -3192.345727 | -3192.344783 | -3192.398472 |
| Bromo-INT5 | -3192.741202 | -3192.461667 | -3192.460723 | -3192.514905 |
| Bromo-TS5rac | -3192.722099 | -3192.444735 | -3192.443791 | -3192.497937 |