ORBITAL OVERLAP AS A TOOL IN COMPUTATIONAL DESIGN OF MOLECULES

by

STEPHANIE ING JONES

Bachelors of Science, 2014

California State University, Los Angeles

Los Angeles, California

Submitted to the graduate faculty of the

College of Science and Engineering

Texas Christian University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May, 2019

Acknowledgement

I'm extremely grateful to Dr. Benjamin Janesko. The completion of my work here would not have been possible without his support. His unwavering guidance, insightful suggestions, constructive criticism, and extensive knowledge helped me grow in my abilities as a chemist and I could never thank him enough for his efforts.

I would like to express my deepest appreciation to my committee- Drs. Eric Simanek, Onofrio Annunziata, and Thomas Cundari. Their invaluable advice on my research has challenged me to think beyond the scope of my knowledge.

I'm deeply indebted to my colleague, Arshad Mehmood, who I have done collaborative research with. His work ethic and ever-patient character eased the difficulty of tackling the exhaustive tools/theories used in computational chemistry. He helped me grow as an individual and has taught me a lot.

I'd also like to extend my gratitude to the other members of the Janesko Group- Drs. Mark Bridle, Emil Proynov and Andrew Mahler, for helping in my understanding of quantum chemistry.

I'm extremely grateful for my family, who supported me throughout this journey. My parents, who encouraged me to pursue a PhD and my father in law, Dr. Jack Yu, who helped me grow in my knowledge in biochemistry.

I cannot begin to express my thanks to my husband, Robert, whose relentless support and profound belief in my abilities made me who I am today. From reading multiple drafts of all my work, to listening to my practice presentations multiple times-I couldn't have done this without you. Thank you for your continuous encouragement, I love you.

Finally, I would like to thank TCU Department of Chemistry and Biochemistry for their generous funding.

Author's Declaration

This is an original work, except where references have been made. No part of this work has been previously submitted as part of a requirement for an academic degree.

Part of this dissertation appeared in the following publications:

1. Janesko, B. G.; Jones, S. I., Quantifying the delocalization of surface and bulk F-centers. *Surface Science* **2017,** *659*, 9-15.

2. Mehmood, A.; Jones, S. I.; Tao, P.; Janesko, B. G., An Orbital-Overlap Complement to Ligand and Binding Site Electrostatic Potential Maps. *Journal of Chemical Information and Modeling* **2018**, *58*, 1836-1846.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

AMBER	Assisted model building with energy refinement
B3LYP	Becke's 3 parameter global hybrid incorporating Lee-Yang-Parr correlation
BSSE	Basis set superposition error
CADD	Computer aided drug discovery
CENP-E	Centromere-associated protein E
DESTINI	Deep structural inference for proteins
DFT	Density functional theory
EDR	Electron delocalization range
EFC	Effective core potential
FGE	Formylglycine-Generating Enzyme
GGA	Generalized gradient approximation
GGA	Generalized gradient approximation
GLA	γ-carboxyglutamic acid
H	Hydrogen
HF	Hartree-Fock
HOMO	Highest occupied molecular orbital

HSAB	Hard/soft Lewis acids/bases
hVKOR	Human vitamin K epoxide reductase
IC50	Half maximal inhibitory concentration
IL	Ionic liquid
KO	Vitamin K 2,3-epoxide
LDA	Local-density approximation
LSDA	Local spin-density approximation
LUMO	Lowest unoccupied molecular orbital
MeCOH	Acetaldehyde
MeCOMe	Acetone
MeOH	Methanol
MeOMe	Dimethyl ether
MESP	Molecular surface electrostatic potential
MD	Molecular dynamics
MM	Molecular mechanics
m-MeOPh	m-Methoxybenzanol
MP	Møller–Plesset perturbation theory

MP2	Møller–Plesset perturbation theory
NNA	Non-nuclear Attractor
NPT	Isothermal-isobaric esemble
o-MeOPh	o-Methoxybenzanol
ONIOM	Our own n-layered integrated molecular orbital and molecular mechanics
PBE	Perdew, Burke, and Ernzerhof gradient-correlation functional
PC	Point charge
Ph	Phenyl
QM	Quantum mechanics
QTAIM	Quantum theory of atoms in molecules
RHF	Restricted Hartree-Fock
RMSD	Root-mean-square deviation
ROHF	Restricted open-shell Hartree-Fock
SOMO	Singly occupied molecular orbital
STO	Slater type orbitals
synVKOR	synechococcus vitamin K epoxide reductase
TBG	Thyroxine Binding Globulin

TM	Transmembrane
TMA	Trimethylamine
TMAO	Trimethylamine N-oxide
trx	Thioredoxin
UEG	Uniform electron gas
UFF	Universal force field
UHF	Unrestricted Hartree-Fock
VDE	Vertical detachment energy
VKOR	Vitamin K epoxide reductase
VKR	Vitamin K reductase
VMD	Visual molecular dynamics
WIE	Warfarin inhibition efficacy
WT	Wild type
XC	Exchange-correlation

Chapter 1: Introduction

1.1 Applications of Computational Chemistry

The use of computational chemistry is to solve chemical problems using computer simulations. It can quantify chemical phenomena such as delocalization, aromaticity, and atomic partial charge that are conceptually useful, but difficult to measure directly. It can also investigate reaction kinetics, reaction rates, and selectivities via structures and energy transition states and reaction paths. It can quantify substituent and modification effects and use these to design hypothetical optimum structures. Here, we apply computational chemistry to explore the many chemical occurrences in experimental observations.

Some examples illustrating computational chemistry's value include the following. It can quantify the appearance of σ-holes in halogen bonding.[1-2] This feature is explained with electrostatics and polarization with dispersion, where σ-holes are the reason for the electron accepting property in halogens. It can also assign atomic partial charges for a wide variety of chemicals. One of these methods is the Hirshfeld population analysis, which involves a clear partitioning of electron density.[3] Computation can also be applied in drug development.

1.2 Computational Chemistry in Drug Development

One of the many advantages in using computational chemistry is to develop novel drugs by minimizing the number of ligands that need to be screened in experimental assays. Computer-

aided drug discovery not only ensures the best possible lead compounds, but it lowers the cost and time it takes for a drug to reach the market.[4] This tool can identify lead drug molecules for testing and the bioavailability of the possible drug molecules. An example that used computer-aided drug discovery was on the development of an inhibitor for proteasome. The discovery was that the addition of a triphenylphosphine group into the base molecule pyridazinone causes an inhibitory effects.[5] Many pharmaceutical drugs have been discovered using computer-aided drug discovery, and have reached the consumer market today.[6-8]

Another example is when Hirayama and co-workers developed a CENP-E inhibitor[9-10] using a combination of high-throughput screening, structure-activity relationship measurements, and homology model docking lead to their lead compound **6a** (Figure 1.1). CENP-E is a mitotic spindle motor protein and a promising target for cancer therapies. The authors produced a computed electrostatic potential map by performing a structure activity relationship analysis. They discovered that the in vitro activity was correlated with a neutral electrostatic potential on the aromatic ring moiety, this feature is highlighted in the region with black boxes in Figure 1.1. Compound **1e** was later synthesized and found to have a higher in vivo activity than compound **6a**. Further improvement on compound **1e** through structure activity relationship analysis produced species **1j** and **1h**, which lead to the production of a potent 5-methoxyimidazo[1,2-*a*]pyridine derivative.

Unfortunately, electrostatic potentials alone provide an incomplete picture of chemistry.[11] To add to this, an analysis of Lewis acid and base properties can help with understanding the effects of the sulfur, bromine and nitrogen groups in these CENP-E inhibitor. Lewis acids and bases depend on both charge and chemical softness: strong acids tend to be

hard, strong bases tend to be soft, and acids & bases of comparable strength prefer hard-hard and soft-soft interactions over hard-soft interactions (HSAB principle[12-13]). The missing piece of information not contained in electrostatic potentials can be quantified from many different perspectives like kinetic energies,[14] polarizability,[15] Fukui functions,[16-17] the QTAIM,[18] steric energies[19] and the electron localization function[20-21] etc. All these effects are connected to orbital overlap. While quantum chemistry calculations can capture orbital overlap effects, CADD methods focusing on charges and MESP discard this important information.

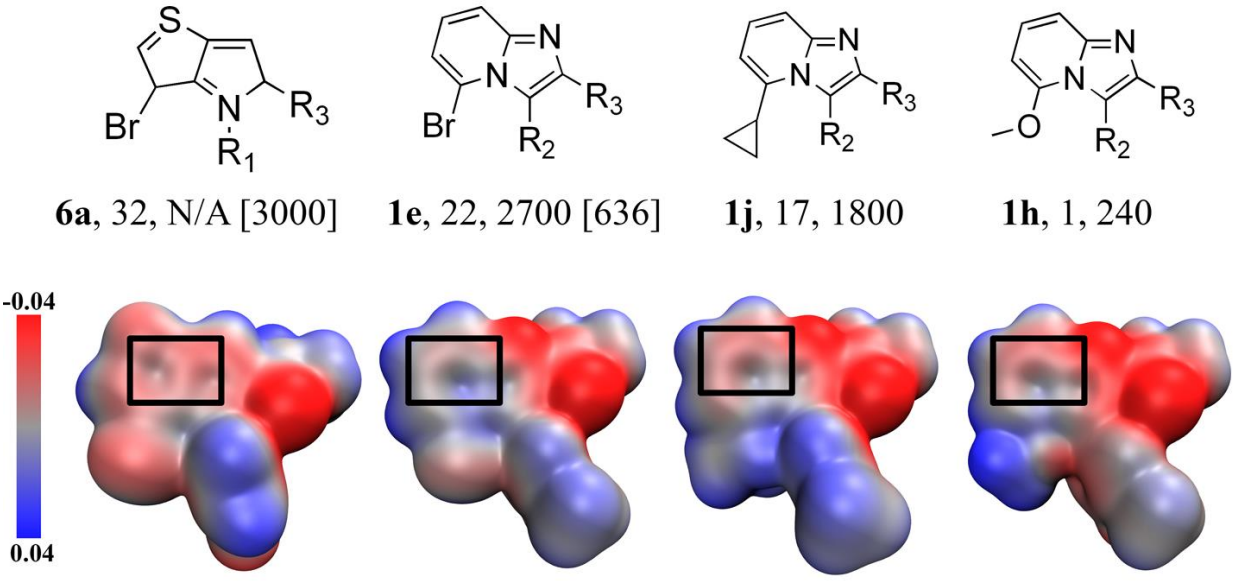


Figure 1.1: (top) Structure, in vitro CENP-E IC_{50} values (in nM), and in vivo HeLa cell proliferation values (in nM) [p-HH3 EC_{50} (nM)] of CENP-E inhibitors **6a**, **1e**, **1j**, and **1h** from refs 9 and 10(numbering follows the experimental references). R_1 = *p*-fluorobenzene; R_2 = *m*-methoxy-*p*-fluorobenzene; R_3 = $C(=O)N(C_2H_4NMe_2)PhCl_2$. (bottom) Predicted electrostatic potentials of the fused-ring regions where the neutral fused-ring electrostatic potential (black boxes) was previously shown to be correlated with the in vitro activity.[9]

1.3 Orbital Overlap Distance

The orbital overlap distance function $D(\vec{r})$ [11, 22] provides a novel way to extract information about orbital overlap from quantum chemistry calculations. Distance $D(\vec{r})$ measures the characteristic size of the orbital lobes that contribute to a computed wavefunction at point r.[22] The orbital overlap distance $D(\vec{r})$ is based on the electron delocalization range function, $EDR(\vec{r}, d)$,[23] where the EDR quantifies the extent to which the molecular orbitals around point r overlap with a the "test orbital" of width d centered at r. Chemically hard and tightly bound regions on the system tend to have $D(\vec{r})$ smaller than softer, loosely bound regions. $D(\vec{r})$ combined with computed electrostatic potentials has been shown to provide a more complete description of aromaticity, nucleophilicity, allotrope stability, substituent effects, and the reactivity of diverse systems. [11]

All the properties of an N-electron system can be expressed using the wavefunction $\psi(\vec{r}, \vec{r}_2 \ldots \vec{r}_N)$, where it defines all the occupied orbitals while spin dependence is suppressed to keep the wavefunction succinct.

$$\gamma(\vec{r}, \vec{r}') = N \int d^3\vec{r}_2 \ldots \int d^3\vec{r}_N \psi(\vec{r}, \vec{r}_2 \ldots \vec{r}_N)\psi^*(\vec{r}', \vec{r}_2 \ldots \vec{r}_N)$$

$$\gamma(\vec{r}, \vec{r}') = \sum_i n_i \phi_i(\vec{r})\phi_i(\vec{r}')$$

Equation 1.1

$\gamma(\vec{r},\vec{r}')$ and $\rho(\vec{r})$ has units of length^{-3} and normalization $\int d^3\vec{r}'|\,\gamma(\vec{r},\vec{r}')|^2 = \rho(\vec{r})$. The one-particle density matrix in Equation 1.1 describes the probability of an electron delocalizing between points $\vec{r}$ and $\vec{r}'$, where N is the number of electrons in the system. Off-diagonal elements of the matrix quantify electron delocalization. When $\vec{r}$ and $\vec{r}'$ are on different atoms, γ > 0 tend to correspond to a bonding interaction while γ < 0 expresses an antibonding interaction. Diagonal elements $\lim_{\vec{r}'\to\vec{r}} \gamma(\vec{r},\vec{r}') = \rho(r)$ give the probability density for finding the electron at point r. Unlike molecular orbitals, the one-particle density matrix is uniquely defined in many electron systems.[24]

We define the electron delocalization range $EDR(\vec{r},d)$ to quantify the extent an electron at point $\vec{r}$ delocalizes over length scale d. We obtain the $EDR(\vec{r},d)$ by contracting the density matrix with a test function of the distance of delocalization $|\vec{r}-\vec{r}'|$:

$$EDR(\vec{r};d) = \int d^3\vec{r}'\, g(\vec{r},\vec{r}')\gamma(\vec{r},\vec{r}')$$

Equation 1.2

$$g(\vec{r},\vec{r}') = \left(\frac{2}{\pi d^2}\right)^{3/4} \rho^{-1/2}(\vec{r})\exp\left(-\frac{|\vec{r}-\vec{r}'|^2}{d^2}\right)$$

Equation 1.3

$g(\vec{r},\vec{r}')$ is unitless and $\gamma(\vec{r},\vec{r}')$ has units of in length^{-3}(inverse volume). Integrating the product of $\gamma(\vec{r},\vec{r}')$ and $g(\vec{r},\vec{r}')$ in all space gives the unitless overlap, $EDR(\vec{r},d)$, between all the occupied molecular orbitals and the test function $g(\vec{r},\vec{r}')$. Global descriptors of the EDR can be expressed by evaluating the density weighted average:

$$\langle EDR(d)\rangle = \int d^3\vec{r}'\,\rho(\vec{r})EDR(\vec{r},d)$$

Equation 1.4

$$D_{max} = argmax_d\langle EDR(d;N+1)\rangle - \langle EDR(d;N)\rangle$$

Equation 1.5

D_{max} shows the characteristic orbital size of an N+1 electron system's most weakly bound electron. This value is obtained by finding the largest change in density weighed average for the odd electron between the charged and neutral species.

$$D(r) = argmax_d EDR(\vec{r};d)$$

Equation 1.6

$D(\vec{r})$ is the distance d which maximizes $EDR(\vec{r};d)$ at point r. It quantifies the best overlap size of the test function orbital with that of the occupied molecular orbital in the system. An example of the application of orbital overlap $D(\vec{r})$ is shown in Figure 1.2. Here, the atomic orbitals of hydrogen, helium, and oxygen atoms are shown with green and red surfaces and the EDR test functions centered at points r are shown with blue surfaces. The atomic orbital "surfaces" were taken to be at 0.001 e/bohr^3 density isosurface,[25] where this cutoff encompasses 96% of the electronic charge[26] and is considered standard in calculations of electrostatic potential.[27] Orbital overlap distance $D(\vec{r})$ provides a characteristic length, where it is the value of the test function width d that maximizes the test function's overlap with the atomic orbitals. The test function's overlap with a compact helium atom is maximized at relatively small $D(\vec{r})$=3.1 bohr. The test function's overlap with the more diffuse lithium 2s orbital is maximized at large

$D(\vec{r})$=6.5 bohr. The overlap between oxygen atom 2s and 2p orbitals and the EDR test function is the largest due to the test function overlapping with one of the lobes on the oxygen p orbital, where a large distance is a large puffy orbital and a short distance is a compact orbital.

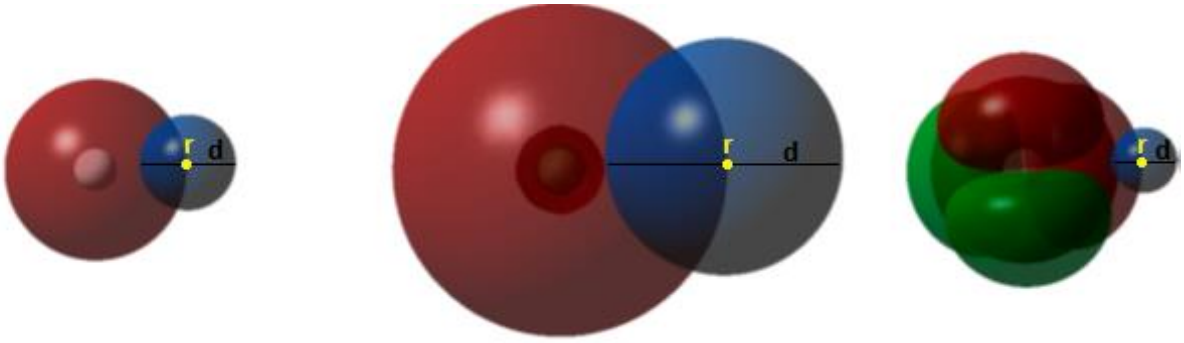


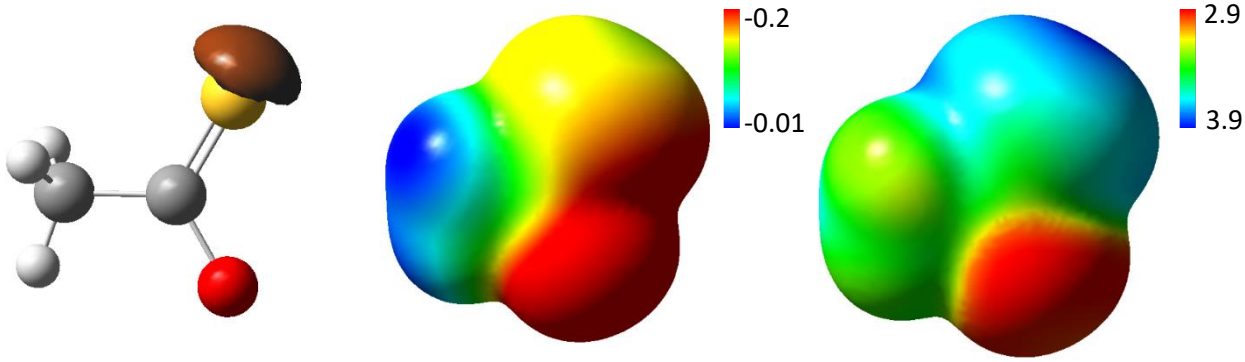
Figure 1.2: Calculation of the orbital overlap distance at a point r on the surface of helium (left), lithium (middle), and oxygen (right) atoms. The atomic orbitals (He 1s, Li 2s, O 2s and 2p) are shown in red and green. The EDR test function centered at yellow point r is shown in blue. The test function's width d is chosen to maximize its overlap with the occupied orbitals. Orbital overlap values are He, $D(\vec{r})$ = 3.1 bohr; Li, $D(\vec{r})$ = 6.5 bohr; and O, $D(\vec{r})$ = 2.6 bohr.

$$D_A = \int d^3\vec{r}\rho(\vec{r})D(\vec{r})w_A(\vec{r})$$

Equation 1.7

We can also compute the atomic overlap distance, D_A, which gives the average size of the molecular orbital lobes around atom A. This is computed using the Hirshfeld partitioning[3] contracted to the orbital overlap. $w_A(\vec{r})$ is the Hirshfeld weight, i.e., the extent to which piont r is assigned to atom A in a molecule.

To demonstrate the different values we can obtain using the $EDR(\vec{r};d)$, we have analyzed the molecule thioacetate anion, to give the following images and quantities in Figure 1.3. EDR(r;D_{max}) shows that the most weakly bound electron is localized on the sulfur atom and D_{max} provides the orbital size of that electron being 2.54 bohr. The electrostatic potential map gives the charge distribution of the molecule in space, where there is a clear indication that the thioacetate's oxygen has the most negatively charged region. The orbital overlap distance map indicates the smaller, more compact orbital region is on the oxygen atom and the largest orbital size is on the sulfur. D_s and D_o are the atomic overlap distances of the sulfur and oxygen atoms, respectively. Q_s and Q_o are the corresponding Hirschfeld charges of the sulfur and oxygen atoms, respectively. These values indicate that the oxygen's orbital is more compact and negative than sulfur.



EDR(r;D_{max}) Electrostatic Potential Overlap Distance $D(\vec{r})$

D_{max}	VDE	D_s/Q_s	D_o/Q_o
2.54 bohr	2.10 eV	1.90 bohr/-0.51 e	1.38 bohr/-0.58 e

Figure 1.3: EDR(r;D_{max}), Electrostatic potential, and $D(\vec{r})$ maps; D_{max} between the N-1 and N molecues; VDE of the electron; and atomic overlap distance (D_A) and Hirschfeld charge (Q_A) of the sulfur and oxygen atoms in the molecule.

Orbital overlap distance and EDR have been shown to give quantitively relevant predictions. EDR has been used to quantify the delocalization of the electron in a variety of molecules[23] to give relevant information on the hard and soft regions of molecule surfaces,[22] stretched polar covalent bonds,[28] and solvated electrons' delocalization.[29-30] Atom-averaged orbital overlap distances rationalized why PhS− is a better nucleophile than PhO− in S_N2 reactions with MeI, even though the partial charge of oxygen on PhO− is more negative.[31] Based on all of these examples, orbital overlap distance has shown to be a great addition to study a molecule's electronic behavior.

This work aims to apply the orbital overlap tools on various systems. The many applications of the orbital overlap can be used for studying electron localization of color centers and biological systems. It can also potentially help design an alternative drug based on the computational results found.

1.4 Electron Delocalization in Anionic Defects

Chapter 2 deals with anionic defects. F-centers in alkali halides and alkaline-earth oxides have been studied extensively for many decades.[32-40] F(*Fabre*) centers are defects in ionic crystals where an anion is replaced by one or more trapped electrons. These trapped electrons contain distinctive optical, electric, and magnetic properties[41] where it can be used in optoelectronic devices.[42-46] F-centers at surfaces and other anionic defects are also relevant in catalysis.[47-53] The localization of electrons in F centers[54-55] motivates the application of theoretical tools for quantifying this electron's orbital size.

Here, we significantly extend a previous study[23] and apply computational chemistry to quantify the localization of a trapped electron in surface and bulk F-centers.[54-55] Electronic structure calculations simulating surface and bulk F-centers have a long history,[38, 56] where modern simulations often use either a single F-center surrounded by a finite cluster of a few surrounding atoms,[49-51, 57-60] or periodic supercell models of infinite defect arrays.[53, 61-62] Many of these studies are motivated to produce energetic properties such as ionization or excitation energies, properties which can be quite sensitive to simulation details.[62] Fewer studies consider our focus, the size of the trapped electrons. Previous studies of trapped electrons' localization consider either the frontier orbital containing the trapped electron[63-65] or the spin polarization of the electron density.[54] Both approaches have limitations: single orbitals are not appropriate for strong coupling where MO theory breaks down,[24, 66] and spin density analysis is not applicable to closed-shell systems. Bader and Platts went beyond these limitations, using the QTAIM[67] to characterize the structure of bulk F-centers in LiF.[60] They produced electron density maps of the three systems: $Li_{14}F_{13}$, $Li_{14}F_{12}^{+}$, $Li_{14}F_{12}^{+}2$ to confirm that the trapped electron yields a NNA. NNAs were also found in sodium electrosodalite[68] and defect magnesium oxide[69] F-centers. Studies of the electron localization function[21] also provides insight into localization of F-centers electrons.[61] However, none of these methods directly measures the orbital size of the trapped electron.

We apply the EDR as an interpretive tool to quantify the size of the trapped electrons on surface and bulk F-centers . We studied three different systems of isolated F centers: LiF, NaCl, and KBr, where these systems are studied using two different models. Model 1 is a box of length L, where L is two times the lattice constant and model 2 is a cube of ions surrounding the defect site. The purpose of this study is to confirm the defect site scales to the size of the box, and that

the ions surrounding the box minimize the delocalization length of the trapped electron. We also found that bulk F-centers' characteristic delocalization length D_{max} scales with defect size. This study is followed by magnesium oxide edge vacancy, surface vacancy, and reverse corner stabilized with a hydrogen atom. The reverse corner site stabilizes the extra electron by adding a hydrogen atom, where adsorbing an N_2 molecule onto this site transfers the extra electron onto the molecule's LUMO.

1.5 Orbital Overlap in Proteins

Chapter three examines the EDR to quantify orbital overlap on protein molecules. Typical analysis of protein molecules involves electrostatic and hydrophobic interaction maps,[9-10, 70-71] but the addition of the orbital overlap distance would be a useful tool to rationalize noncovalent interactions in protein active sites. In this study, we consider how a combination of orbital overlap distance function $D(\vec{r})$ and molecular electrostatic potential can rationalize noncovalent interactions in a protein's active site. We consider a biologically relevant molecule, cysteine, along with four diverse examples of Lewis acid-base chemistry in active sites: Avidin, GolB, TBG, and FGE.

The study of cystine considered three atoms of interest: oxygen, nitrogen, and sulfur. We found that the orbital overlap distance for the sulfur atom was large while nitrogen had a small and compact orbital. This result is consistent with HSAB theory, where sulfur is a softer Lewis base and nitrogen is a harder Lewis base. A highlighted feature studied here is the metal binding protein, FGE, which has binding pockets for one Cu^+ and two Ca^{2+} ions. The crystal structure used in this study was crystallized using Ag^+ instead of Cu^+. The studies found that the Ag^+ binding

site has a less negative electrostatic potential compared to the two Ca^{2+} ions. The orbital overlap distance studies found that Ag^{+} has a larger overlap distance than Ca^{2+}, which is consistent with HSAB principle. These findings can potentially be used to develop new ligands that may bind to these proteins and motivate the use of the orbital overlap distance for protein binding site analysis.

1.6 Receptor Design for Trimethylamine N-oxide

Chapter four deals with a preliminary computational design project. Choline, betaine, and L-carnitine, which are found in meats and eggs, have been shown to catabolize into TMAO by bacteria in the digestive tract. TMAO has been linked to whole-body cholesterol metabolism, chronic kidney disease, vascular inflammation, and atherosclerosis.[72] Our goal is to design a receptor for TMAO, where we can track its pathway in the body. The receptor built is designed from the crystal structure of the interaction between TMAO and binding protein TorT.

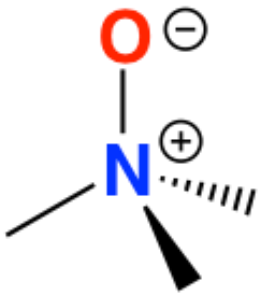


Figure 1.4: Isolated TMAO molecule.

TMAO has an unusual three hydrogen bond acceptor property at the oxygen terminus (Figure 1.4). We use this to build a receptor off a triazine with three linker arms to participate in hydrogen bonding to the oxygen atom. We tested the hydrogen bonding distances for TMAO, dimethyl ether, acetone, and acetaldehyde. The hydrogen bonding test involved, one, two, and three methanol molecules binding to the molecules studied. TMAO was found to not only

support three hydrogen bonds on its' oxygen terminus (while the other molecules could only support two), but TMAO had also a stronger and shorter hydrogen bond distance to the alcohol than the other molecules studied.

1.7 Mechanistic Studies in Lignin Model Compounds

Chapter five details an early project on lignin. In the pulping and biorefinery processes, lignin is disposed as a waste stream after lignocellulose's other components (hemicellulose and cellulose) are extracted.[73-74] The most efficient method to extract lignin is with phosphoric acid and acetone, but once lignin is separated, depolymerization is difficult due to a variety of linkages in the structure.[75-76]

This chapter is on the mechanistic study of hydrolysis of lignin model compounds. We studied substituent effects of β-o-4 linkages on lignin models along two reaction pathways: S_N2 and E_1. We found that the model with R_1=m-MeOPh, R_2=MeOH, and R_3=o-MeOPh matches the real lignin compound in structure and it underwent a smaller reaction barrier through an S_N2 pathway.

The difference between the two pathways involves the production of a cationic intermediate on different atoms. The stability of the intermediate depends on its neighboring substituent. For example, intermediate A produces a carbocation which shares a bond with substituent R_1. This would mean that the stability of intermediate A depends on the electronic effects of R_1. Intermediate E, which is the first intermediate for the S_N2 reaction, produces a

cation on an oxygen atom attached to R_3. This intermediate's stability depends on the substituent R_3 electronic effects.

1.8 The Study of Approved Drugs for Vitamin K Epoxide Reductase

Chapter six combines two ideas using orbital overlap distance and other tools to design new drugs. hVKOR structure has not been identified due to difficulty in crystallization of a membrane bound protein. This study predicts hVKOR structure and the docking site for warfarin, a known drug used as an anticoagulant. We docked warfarin into many cavities found using FINDSITE^{COMB} program. We aim to use these findings to understand the mechanism of the type of inhibition warfarin has on VKOR. Both the electrostatic potential and orbital overlap are used together to analyze the strength of the binding warfarin has on the predicted binding pocket. These results can later help design a new drug type molecule that can better bind to VKOR.

Chapter 2: F centers in Anionic Defects

F (Farbe) centers, also named color centers, are defects in ionic crystals where a trapped electron replaces an anion. These defects often absorb visible light (Figure 2.1), giving a color that is a function of the size of the defect. Anionic defects in alkali halides are of particular interest due to their optical, electric, and magnetic properties.[41, 77] Alkaline earth oxides are popular as catalyst support materials[78] and their anionic defects can help control catalyst properties.[52]

Figure 2.1: NaCl, KCl, and KBr after irradiation with a Tesla coil[79], showing presence of color centers.

2.1 F-Center Analyses

Most analyses of trapped electrons are based on the HOMO from a HF or DFT wavefunction.[64-65] HOMOs are not uniquely defined in many-electron wavefunctions.[24] HF

HOMOs do not include correlation effects[80] and DFT HOMOs correspond to a single Slater determinant describing noninteracting electrons.[81] Single Slater determinant wavefunctions provide an incomplete picture of strongly correlated systems, such as the dissociation of a covalent bond.[82] To overcome these issues of the non-uniqueness of HOMOs, the electron delocalization range $EDR(\vec{r}, d)$ is developed to quantify delocalization by contracting the one-particle density matrix with a Gaussian test function (details are in Chapter 1.3).

We use the electron delocalization range $EDR(\vec{r}, d)$ to visualize and quantify the characteristic distance d over which an electron at point r delocalizes. In solvated electrons, the D_{max} (Equation 1.5) characterizes the "size" of the solvated electron in anionic water cluster $(H_2O)_N^-$. Previous work showed that D_{max} in these systems is correlated with the radius of gyration of the singly occupied molecular orbital.[64] We find that in alkali halide F-centers, the trapped electrons' delocalization scales with the effective size of the cavity. Calculations on edge and surface defects in MgO confirm the large delocalization of electrons at surface defects. Application to N_2 activation by electrons at MgO surface illustrates how the trapped electron localizes into the absorbed molecule's LUMO.

2.2 Particle in a Box and F-Centers in Alkali Halides Comparative Studies

We begin by considering the EDR of isolated F-center defects in alkali halides. Three different systems of isolated F centers are studied: LiF, NaCl, and KBr. These systems are investigated using two different models. Model 1 is a box of length L, where L is two times the lattice constant (visualized in Figure 2.2, left). Model 2 is a cube of ions surrounding the defect

site (visualized in Figure 2.2, right). Model 2 of the LiF cluster has been used previously by Bader and Platts.[60] We obtained the same VDE as Bader and Platts, confirming that the model is identical (shown in Table 2.1).

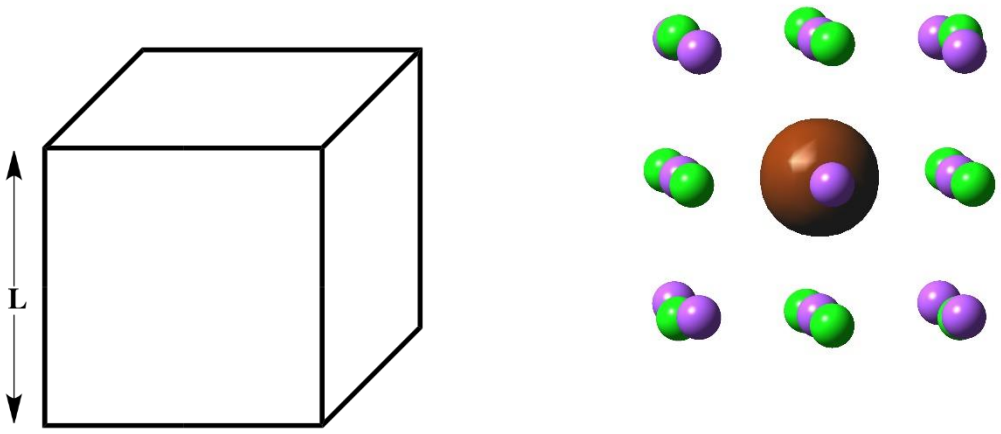


Figure 2.2: A simple model for an F-center defect (left) and $EDR(\vec{r}; D_{max})$ of NaCl cluster (right).

	LiF	**NaCl**	**KBr**
VDE (eV)	4.88	4.65	3.92

Table 2.1: VDE for LiF, NaCl, KBr F center defect clusters.

Table 2.2 provides the box length L for the three systems, absorption wavelength λ,[83] model 1 λ computed as the HOMO-LUMO gap of the isolated electron, and models 1 and 2 maximum delocalization of a trapped electron D_{max}. L increases as LiF<NaCl<KBr, consistent with

the atomic radii. Model 1 gave larger D_{max}, which then results in smaller absorption wavelengths than in the real system. In model 2, the effective cavity size is reduced because the surrounding atoms protrude into the cavity. Figure 2.2 (right) shows isosurface $EDR(\vec{r}; D_{max})$ = 0.8 of the NaCl cluster.

	LiF	**NaCl**	**KBr**
L (bohr)	7.5	10.6	12.4
Absorption λ (nm)	243	448	601
Model 1 λ (nm)	179	350	479
Model 1 D_{max} (bohr)	6.4	9.0	10.4
Model 2 D_{max} (bohr)	5.6	8.4	9.2

Table 2.2: Box length L, experimental and theoretical λ, and delocalization lengths D_{max} for models 1 and 2.

EDR studies complement existing studies of correlation effects in solvated electrons.[84] Table 2.3 gives the values of D_{max} with the use of different electronic structure approximations with the same basis set for NaCl. In the HF method, electron correlation effects are approximated by taking the average repulsive potential felt by one electron from the other electrons in the system. The calculations using both ROHF and UHF gave the same D_{max}, where the delocalization of the trapped electron remains the same even when the methods treat paired electrons differently. Electron correlation with the use of MP2 gave a larger D_{max} than that of the HF

method. A DFT calculation using the LSDA exchange-correlation functional gives a relatively large D_{max}. This is consistent with the LSDA's known tendency to over-delocalize charge and spin. The PBE uses GGA, where the electron density gradient is an ingredient in the exchange-correlation approximation. PBE is generally less delocalizing than LSDA, where our results are consistent to this known fact. B3LYP uses hybrid exchange functional where the odd electron is localized more than PBE and LSDA methods.

Method	NaCl
UHF	8.4
ROHF	8.4
MP2	8.6
LSDA	9.4
PBE	9.0
B3LYP	9.0

Table 2.3: Value of D_{max} for NaCl cluster with the use of different methods and basis set of 6-31G(d).

Bulk F-centers experience the effects of all surrounding ions and not just the first shell of ions. We model this interaction by replacing ions outside of the defect with PCs. PCs are embedded around NaCl defected cluster to quantify the effect of VDE and D_{max} as we increase the number of PC shells as shown in Table 2.4. PC shells represent the number of layered point

chargers surrounding the cluster. As we increase the number of point chargers around the cluster, we notice a decrease in VDE, which is due to the stabilization of the trapped electron in the cavity. The D_{max} changes very little with the addition of point charges, which supports that the use of cluster model systems is enough to represent bulk F-centers.

N_s	D_{max} (bohr)	VDE (eV)
0	4.2	5.32
2	4.2	5.17
4	4.2	4.87
8	4.2	4.64
16	4.2	4.50
22	4.2	4.46

Table 2.4: D_{max} and VDE for an F-center in NaCl as functions of (N_s) PC shells about the NaCl quantum-mechanical region.

2.3 Changing Defect Size

We have also studied the expansion and contraction of the two model clusters as shown in Figure 2.3. The motivation behind this study is that we expect the delocalization to be a function of the cavity size and of the chemistry of the cavity walls. Figure 2.3 plots D_{max} vs. cluster size L, for rigid expansion of the cluster. The line represents the results from Model 1. The vertical lines on the graph matching the cluster color give the experimental crystal structure cation-cation

distance for each of these alkali halide clusters. As expected, the D_{max} decreased as we contracted the clusters and increased as we expanded the clusters. In the limit of large L, the D_{max} for each of the clusters studied converges to the same value. In Model 2, we model the orbital containing the trapped electron using linear combinations of the atomic orbitals ("basis functions") of the removed anion and basis functions on the other atoms. The basis functions of F, Cl, and Br are different, so the trapped electron is different when the other atoms are far away. At Larger L, Model 1 gave a lower D_{max} than the clusters in Model 2. This is due to the trapped electron in model 2 delocalizing onto its neighboring atoms.

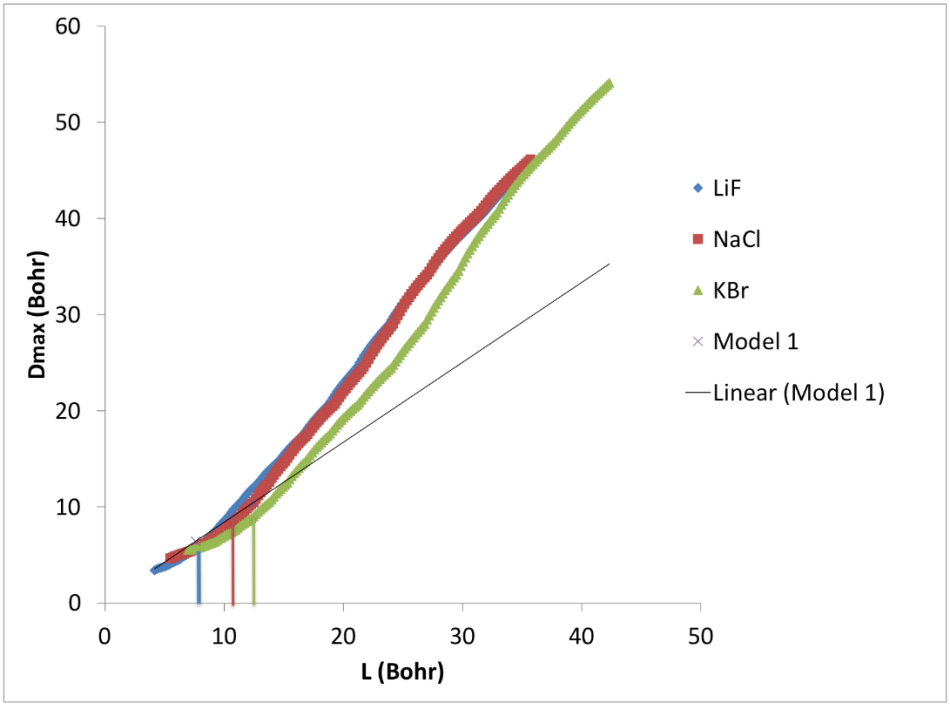


Figure 2.3: Delocalization as a function of cavity size for geometry-frozen, stretched and squeezed LiF, NaCl, KBr cavities and Model 1 using Hartree-Fock/6-31+G(d,p).

2.4 Anionic Defects in MgO

Chemistry happens on the surface of metal oxides when highly reactive sites are present.[85] Highly reactive sites include atomic vacancies (as in bulk F centers), as well as atomic steps and kinks on the surface. We have considered trapped electrons in vacancies and an atomic step on the surface of MgO clusters. Three different systems are studied for a trapped electron: edge vacancy, surface vacancy, and a reverse corner site stabilized by a hydrogen atom. These systems are studied by using Model 2, where ions surround the defect site.

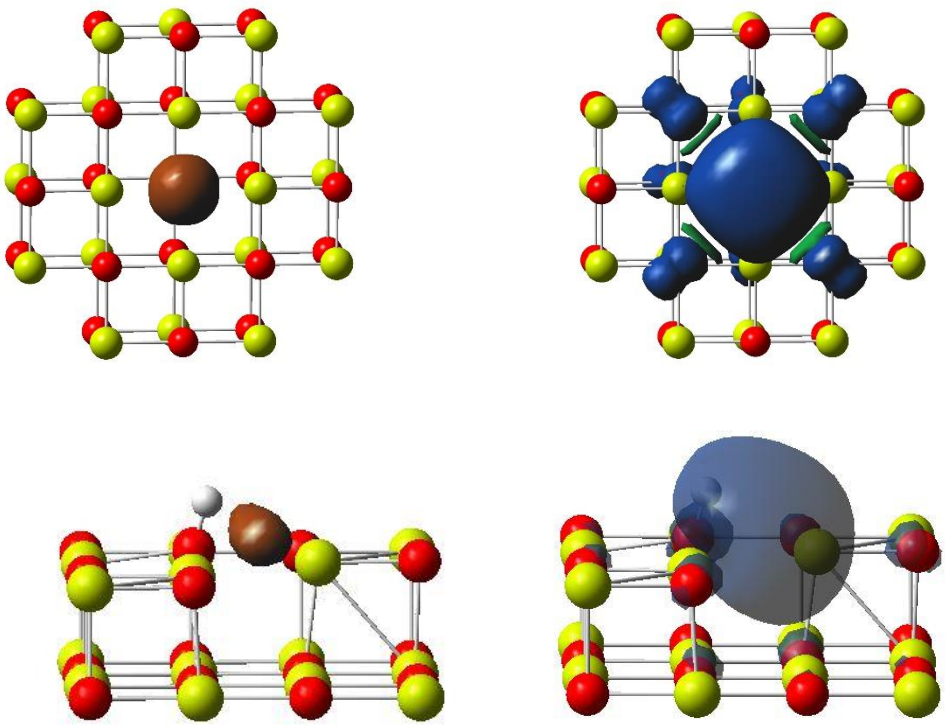


Figure 2.4: D_{max} (brown isosurface) and spin density (blue isosurface) for anionic oxygen vacancy in surface defect (two left) and reverse corner (two right).

Table 2.5 provides D_{max} and VDE values of these three types of clusters considered. The D_{max} is smallest for the reverse corner defect, as the extra electron is stabilized by the adjacent H(+). D_{max} is largest for the edge defect, as there are few ions around the trapped electron. Figures 2.4 provide the $EDR(\vec{r}; D_{max})$ and spin density of the surface vacancy, edge vacancy, and the hydrogenated reverse corner site. The spin density is the total electron density of alpha spin electrons subtracted by the beta spin electrons. In other words, spin density gives the location of the alpha spin odd electron. The isosurface images are transparent for the reverse corner sites to better see the positions of the other atoms in the clusters. Yellow spheres: Mg and red spheres: O.

	Edge Vacancy	**Surface Vacancy**	**Reverse Corner**
D_{max} (bohr)	6.0	5.8	5.2
VDE (eV)	11.66	11.54	7.50

Table 2.5: D_{max} and VDE for MgO cluster defects.

2.5 N_2 Activation by MgO Anionic Defect

Chemically, any catalyst that can activate small molecules such as N_2, O_2, and CO_2 is useful for practical chemistry. Pacchioni and coworkers[57, 86] suggested that an adsorbed hydrogen on an MgO reverse corner site can further interact with an N_2 molecule, forming the N_2^- species stabilized at the site. The odd electron localized by the hydrogen atom is stabilized at the π^* orbital. This electron then localizes into the adsorbed molecule's LUMO. We analyzed the two

orientations of N_2 adsorption and calculated their D_{max}.[86] Table 2.6 provides the values of N_2 bond length, D_{max}, and N_2 charge before and after their adsorption to the reverse corner site. We found that not only does the N_2 bond length stretch with interaction with the MgO site, but also their D_{max} increases due to some of the electron density dispersing onto the MgO cluster. This is proven further by analyzing the Mulliken population of these species where the adsorbed N_2 experiences about half of the charge that an N_2^- would have. Figure 2.5 shows the $EDR(\vec{r}; D_{max})$ and spin density of the two orientations of N_2 interacting with the MgO reverse site. Yellow spheres: Mg and red spheres: O.

	N-N Bond (bohr)	**D_{max} (bohr)**	**N_2 Charge**
Isolated N_2	2.2	--	0
Isolated N_2^-	2.4	1.0	-1.0
MgO(H^+)(e^-)-N_2 (a)	2.3	1.6	-0.59
MgO(H^+)(e^-)-N_2 (b)	2.3	1.6	-0.52

Table 2.6: Bond length, D_{max}, and Mulliken charge of N_2 on reverse corner site.

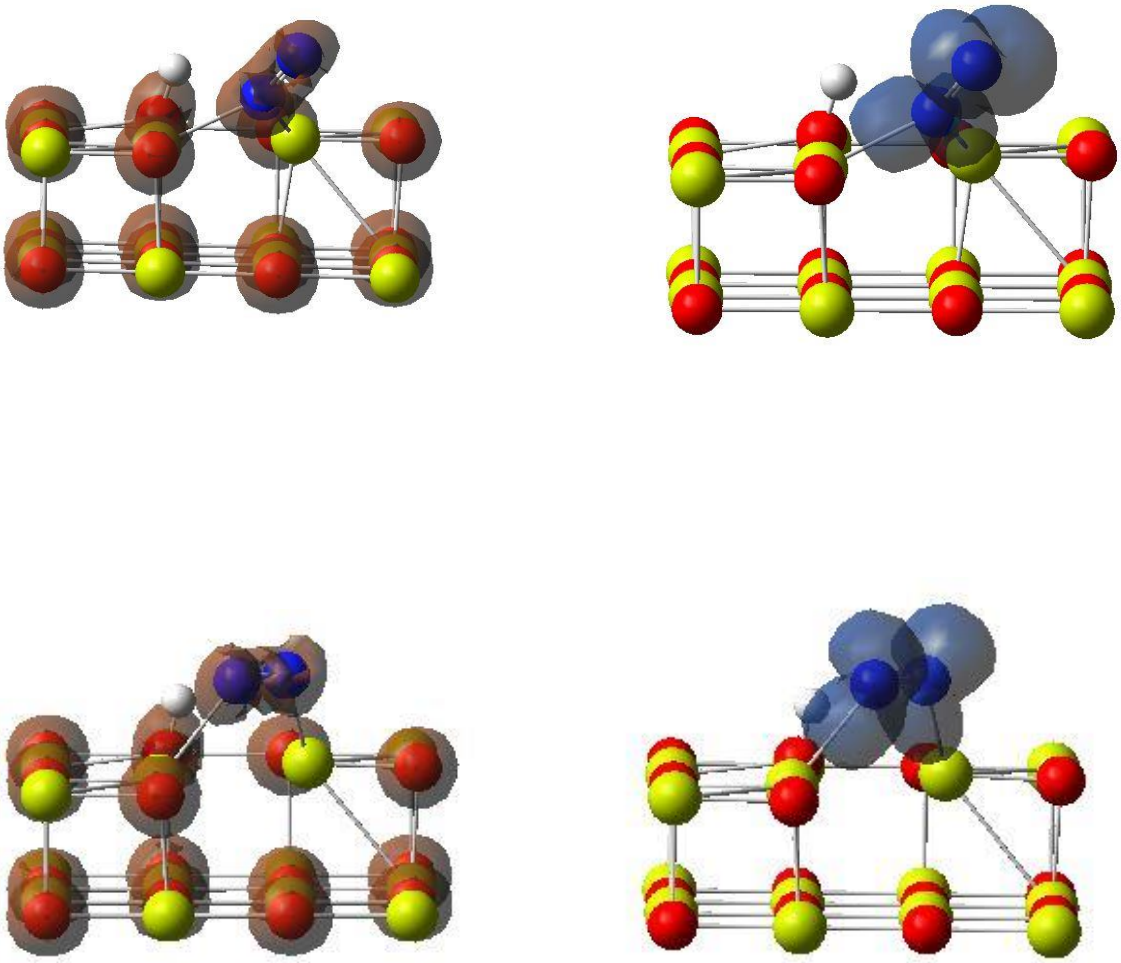


Figure 2.5: $EDR(\vec{r}; D_{max})$ (brown isosurface) and spin density (blue isosurface) for two orientations of N_2 interacting with reverse corner site.

2.6 Future Directions

Since the EDR of MgO has been studied, an alternative crystals (CaO) may also be explored as a comparative study. These results may lead to answers regarding the difference in reactivity as catalyst support materials.[87] EDR studies on doped alkaline-earth oxides with a transition

metal can also be studied, since these types of molecules can lead to charge transfer from the transition metal to the species adsorbed on the oxide surface.[88] We may also explore the EDR of reactions on surfaces, such as the adsorption and dissociation of ammonia on silicon.[89]

2.7 Computational Details

All of the computational work done uses the development version of the Gaussian suite of programs. [90]

The EDR is evaluated using HF, second order MP2, and 16 generalized Kohn-Sham DFT. DFT calculations use various approximate XC functionals which include LSDA,[91] PBE,[92] and B3LYP.[93-95] Molecular calculations on open-shell systems are performed spin-unrestricted unless noted otherwise. EDR is evaluated using molecular calculations described previously.[23, 96] $\langle EDR(d) \rangle$ is evaluated by numerical integrations of equation 4 using standard DFT numerical integration grid.[97]

Other details on individual calculations are as follows. The calculation on bulk alkali halide clusters were initially geometry optimized with an open shell defect following Bader and Platts[60] using unrestricted HF with basis set 6-31+G(d). PC embedding on an NaCl cluster models the long-range Coulombic interactions and ensures that the Madelung potential is contained within the center of the cluster. PC= ±1 are placed around the cluster. PC shells represent the number of layered point charges surrounding the cluster. We only studied the even number of PC shells since the odd number of shells gave a total cluster charge of -2 while the even numbers gave a

cluster charge of 0. The expansion/contraction study of these clusters uses a geometry frozen cluster at different values of L. A “Ghost” atom is added to the cluster center.

The calculations on surface and edge defects on MgO clusters used unrestricted HF/6- 31+G(d). Surface and edge defects, with cluster sizes Mg14O + 13 and Mg21O + 20 respectively, are embedded in an array of PC= ±2 with dimensions of 13 x 13 x 4.[54] MgO reverse corner studies use B3LYP and an Mg14O14 cluster with dimensions of 2 x 4 x 5. This cluster has a 2 x 2 site removed from the corner and replaced with a hydrogen atom. The 17 hydrogen atom is attached to the oxygen at the center of the removed site at 45◦ . An array of PC= ±2 with dimensions 6 x 11 x 11 surrounds this cluster. The interface between the cluster and PC replaces the PC= +2 with semi-local ECF[98] to contain the oxygen’s electron density within the cluster region, which is denoted as Mg∗ . The final cluster size (excluding the PC added) is Mg14O14Mg∗ 14(H+) with a charge of 72 and multiplicity of 2. The adsorption of N2 onto the surface of the cluster occurs in two orientations,[86] and the cluster including the PC and N2 is further geometry optimized using B3LYP/genECP. Atomic charges are obtained from Mulliken population analysis.

Images of calculated molecular geometries use a “ball-and-stick” model of chemical bonds as a mere guide to the eye. VDE is the energy needed to eject an electron from the cluster species without changing the geometry. The VDE is the difference in energy between the open and closed shell systems and D_{max} is obtained using equation 5.

Chapter 3: Orbital Overlap Distance - an Added Tool to Electrostatic Potential Studies in Docking

3.1 The Approach to Drug Screening Methods

Today, almost all drug screening techniques rely on molecular dynamics or Monte Carlo simulations.[99] CADD have been used to develop important therapeutic molecules, [100-103] inhibitors for DNA gyrase,[102, 104] HIV-1 protease,[103, 105] dopamine D3 receptor agonist,[106] anti-glaucoma agent,[107] antibiotics,[108] and development of anti-cancer drugs like Erlotinib,[109-111] Sorafenib,[112-113] Lapatinib,[114-115] Abiraterone.[116-118] Today, most CADD studies rely on "ball-and-spring" MM simulations,[119] which poorly describe [120-121] the fundamentally quantum-mechanical interactions. Although the convergence rate of these techniques is high,[122-123] accuracy of specific binding interactions is not completely defined. A quantum mechanical approach such as DFT can capture the effects[124] of nonstandard bonding motifs,[125] proton transfer,[126] metal coordination,[126] and noncovalent interactions,[124-125, 127-129] including aromatic interactions,[124] but this method is too computationally expensive when it comes to drug, target, and solvent system simulations.

3.2 Combining the Orbital Overlap Distance and Electrostatic Potential

The terms that are not addressed in drug discovery are HSAB. A Lewis acid is an electron acceptor while a Lewis base is an electron donor. Hard “species” are small and weakly polarizable atoms while soft “species” are big and strongly polarizable atoms. Our previously introduced

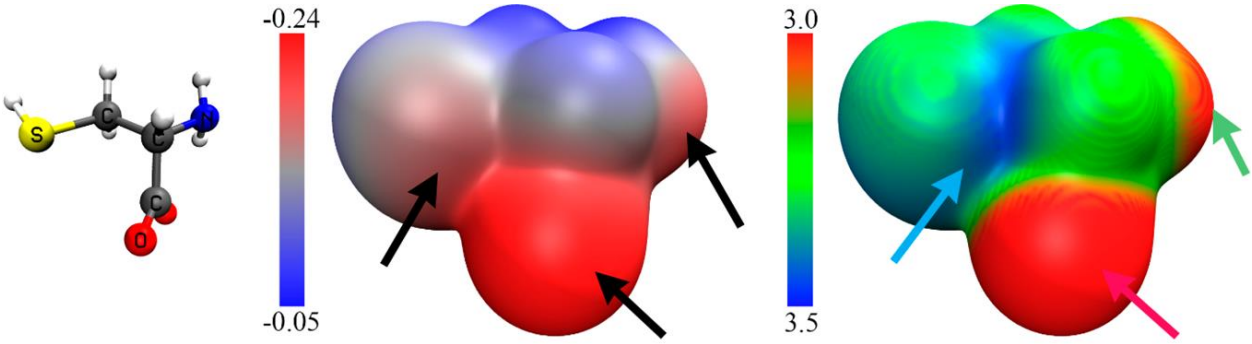
tool[22] (Orbital Overlap Distance) can address these new terms where it can be used alongside electrostatic potential studies. Orbital overlap distance measures the characteristic size of the molecular orbitals and determines the hard regions (electrons are more diffuse) and soft regions (electrons are more compact) in different regions of the protein, which aligns with the hard/soft regions of the drug.

Here, we apply the orbital overlap distance and electrostatic potential to representative problems in biological and medicinal chemistry. In this study, we also show how the combination of orbital overlap distance function $D(\vec{r})$ and molecular electrostatic potential can rationalize noncovalent interactions in protein active sites. We consider four diverse examples of Lewis acid-base chemistry in active sites, including metalloregulatory proteins and the streptavidin-biotin interaction. These results motivate use of the overlap distance in future applications of quantum chemistry to CADD.

3.3 Orbital Overlap and Electrostatic Potential Studies of Cysteine

We begin by applying the orbital overlap distance and electrostatic potential of molecular surfaces to a novel and biologically relevant molecule. [22] Figure 3.1 is of a cysteine molecule that has been proven to chelate in many different coordination sites[130] which involve the distinct Lewis bases oxygen, nitrogen, and sulfur. This figure provides the computed electronic structure, electrostatic potential map, and orbital overlap distance map of a deprotonated cysteine in continuum water solvent. In the electrostatic potential map, oxygen, nitrogen, and sulfur are clearly distinguished by their charge, but there isn't much of a difference between the charges of

nitrogen and sulfur. In the orbital overlap distance map, there is a clear distinction between the large orbital lobe of sulfur from the smaller, more compact lobe, of nitrogen. This is due to sulfur being a softer Lewis base while nitrogen is a harder Lewis base. The table accompanied to Figure 3.1 provides the values of the electrostatic potential and overlap distance at the points labeled with arrows.



Site	Value of surface electrostatic potential (au)	Value of surface orbital overlap distance D(r) (bohr)
O	-0.25	3.00
N	-0.18	3.15
S	-0.15	3.45

Figure 3.1: (top, left to right) Structure, molecular electrostatic potential map, and overlap distance $D(\vec{r})$ map of deprotonated cysteine. The electrostatic potential map displays the three Lewis basic sites (black arrows) in the amino acid. The overlap distance differentiates the diffuse and chemically soft sulfur base (blue arrow) from the compact and chemically hard oxygen base (red arrow) and nitrogen base (green arrow). (bottom) Electrostatic potential and overlap distance values at the three points labeled with arrows.

3.4 Avidin-Biotin Complex

The first species studied is the Avidin-biotin complex. Avidin proteins have been used for many decades as an essential biotechnological tool for analysis [131] and drug development [132] due to their extraordinarily high affinity[133-134] for biotin through noncovalent interactions and ultrahigh stability of avidin-biotin complex. [134-135] The crystal structure reveals hydrophobic interactions of biotin with Trp70, Phe72, Phe79, and Trp97 from one avidin monomer and Trp-110 from an adjacent avidin monomer. The hydrophilic interactions involve a carbonyl behaving as a hydrogen bond acceptor with Asn12, Ser16, and Tyr33; nitrogen hydrogen bond donation with Thr35 and Asn118; sulfur interacting with Thr77; a carboxylate oxygen hydrogen bonding with N-H of Ala39 and Thr40; and a side chain of Thr38 and the other carboxylate oxygen behaving as a hydrogen bond acceptor with Ser73 and Ser75.[134] This complex was chosen because biotin has a chemically soft region (sulfur) and a chemically hard region (amines). The combination of MESP and $D(\vec{r})$ elaborates the selectivity of binding pocket for biotin.

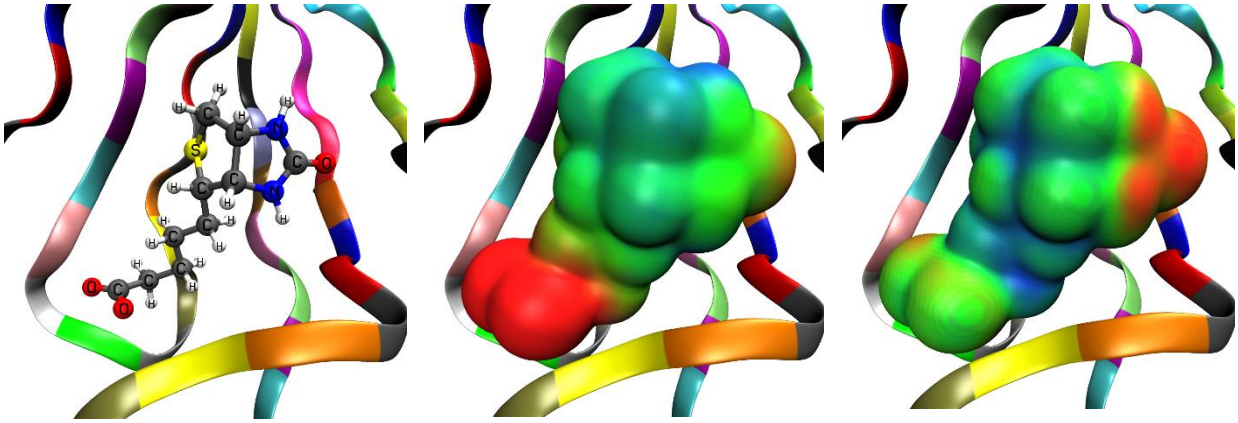


Figure 3.2: Biotin in the strepavidin biding pocket. (Left) atomic structure. (Middle) Electrostatic potential from -0.15 au (red) to 0.02 au (blue). (Right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.6 bohr (blue).

The biotin molecule contains both hydrogen bond donors and acceptors which interact with the avidin binding pocket. Figure 3.2 provides the structure, MESP, and overlap distance of bound biotin. The MESP distinguishes the negative (red) valeryl carboxylate and ureido ring from the positive (blue) ureido NH and less charged (green-yellow) hydrophobic regions. The overlap distance shows that both negative and positive regions are small and compact (red-green), while the hydrophobic regions and sulfur are more diffuse (green-blue). This shows that the overlap distance distinguishes the diffuse, soft Lewis base site of the from the compact, hard Lewis base site of the nitrogen portion of the ureido ring.

Figure 3.3 shows the MESP and overlap distance of the streptavidin binding pocket. Regions A and B show the Lewis-acidic OH groups Ser73, Ser75 and Ser16, Tyr33 respectively. Regions C and E show the Lewis-basic oxygens of Tyr33, Asn118 and Thr35 respectively. Region D shows the Thr77 hydroxyl group. The active site's Lewis-acidic (A and B) and Lewis-basic (C and E) sites are aligned with biotin's hydrogen bond donors and acceptors. The overlap distance $D(\vec{r})$ plot in Figure 3.2 shows that the localized, chemically "hard" Lewis acids and bases interacting with the ureido ring are clearly distinguished from the chemically "soft" Lewis acids interacting with the tetrahydrothiophene sulfur and the hydrophobic backbone. The relative comparison of $D(\vec{r})$ molecular surface plots in Figures 3.1 and 3.2 clearly highlights the interactions between chemically hard and soft Lewis acids and bases of biotin with corresponding regions of avidin which are responsible for selective biotin complexation.

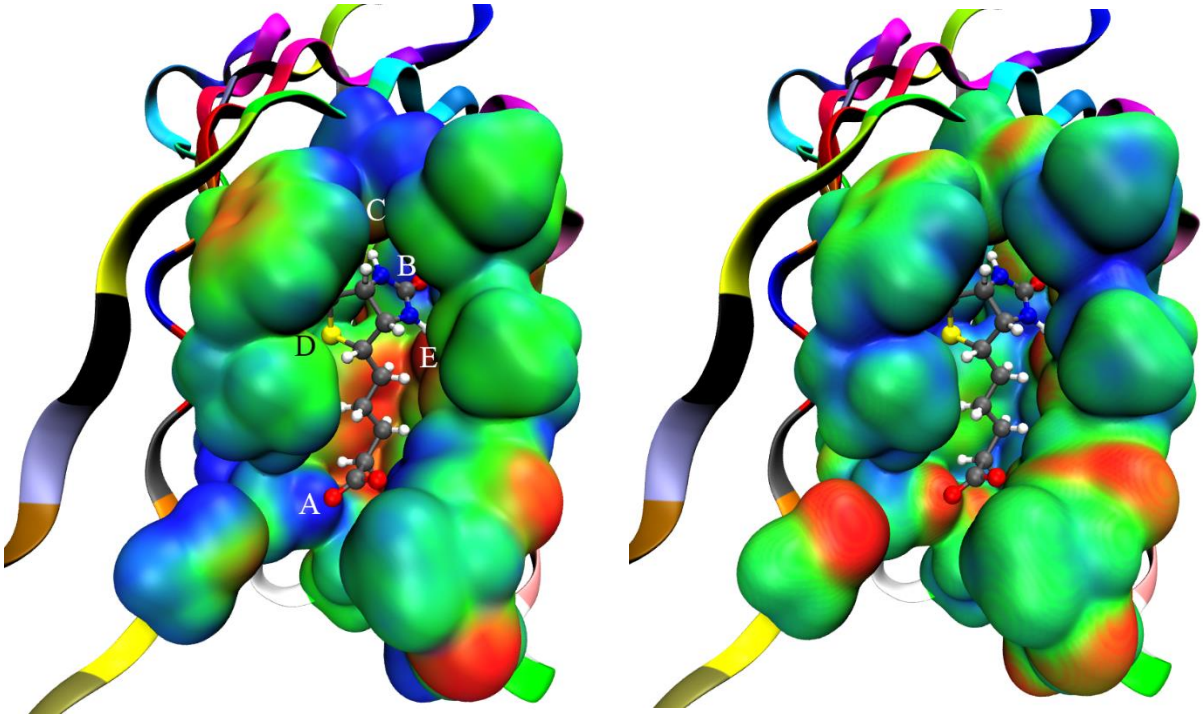


Figure 3.3: Pocket atoms of avidin at high-level layer: (left) electrostatic potential from -0.05 au (red) to 0.05 au (blue); (right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.6 bohr (blue) blue. The biotin molecule in the pocket (low-level layer) is shown using ball-and-stick model. This orientation of the complex is selected as it highlights the chemically most interesting regions.

3.5 Gold-Binding Protein GolB

We next studied the gold-specific protein GolB and gold ion interaction. Bacteria such as Salmonella typhimurium, Salmonella enterica and Cupriavidus Metallidurans use gold selective MerR-type transcriptional regulators and metallochaperones to detect the presence of toxic gold ions.[136-139] These metalloregulatory proteins are able to distinguish Au^+ from Cu^+ or Ag^+.[137-138] This high selectivity for Au^+ ions disables them to effect the functioning of other

metalloregulatory proteins especially the Cu^+ trafficking proteins.[136] The putative gold-chaperone, GolB from Salmonella typhimurium is a typical example which selectively binds soft Lewis acid Au^+ with much higher affinity than even the moderately soft Lewis acid Cu^+. [140] Many have studied the gold-thiol interaction[141-144] to understand the strength and structural details, but Zhao et al.[136] were able to crystallize this type of interaction in the molecular level between the GolB protein and an Au^+. GolB uses a conserved binding site involving the thiolates of Cys10 and Cys13. Though the rupture forces for Cu-S bond in GolB binding pocket, estimated using AFM-based SMFS methods, are comparable with those of Au-S bond, there is a significant difference in the binding affinities of Au^+ and Cu^+.[136] The Au-S interaction is a useful extension to our study since gold is known to be a very soft atom.

Figure 3.4 shows how combining the MESP and surface overlap distance $D(\vec{r})$ highlight the chemistry of the gold-binding pocket. The MESP on the Cys10 and Cys13 thiolates (Figure 3.4, left) is relatively negative (red), consistent with binding to a Lewis acid. More importantly, the overlap distance (Figure 3.4, right) is particularly large on the thiolates, consistent with selective coordination to soft Lewis acids. The combination of surface MESP and $D(\vec{r})$ clearly shows the binding pocket's soft-base character. The overlap distance and MESP also quantify the different character of Cu^+ (MESP = 0.304 au, $D(\vec{r})$ = 1.789 bohr) and Au^+ (MESP = 0.270 au, $D(\vec{r})$ = 2.030 bohr) ligands. The overlap distance clearly shows that the binding site's selectivity to Au+ is consistent with its softer, weaker Lewis acid character.

Wei *et al.* have shown experimentally that this conserved binding site of GolB has a much higher affinity for Au^+ than Cu^+.[140] Gas phase calculations of both ions at ωB97X-D/ cc-pVDZ level of theory using effective core potential compliment these findings. Though Cu^+ (MESP = 0.304 au,

$D(\vec{r})$ = 1.789 bohr) is more positive compared to Au^+ (MESP = 0.270 au, $D(\vec{r})$ = 2.030 bohr) but small overlap distance makes it a hard Lewis acid compared to Au^+, which is a preferred candidate for this binding site due to its large overlap distance i.e. soft Lewis acid.

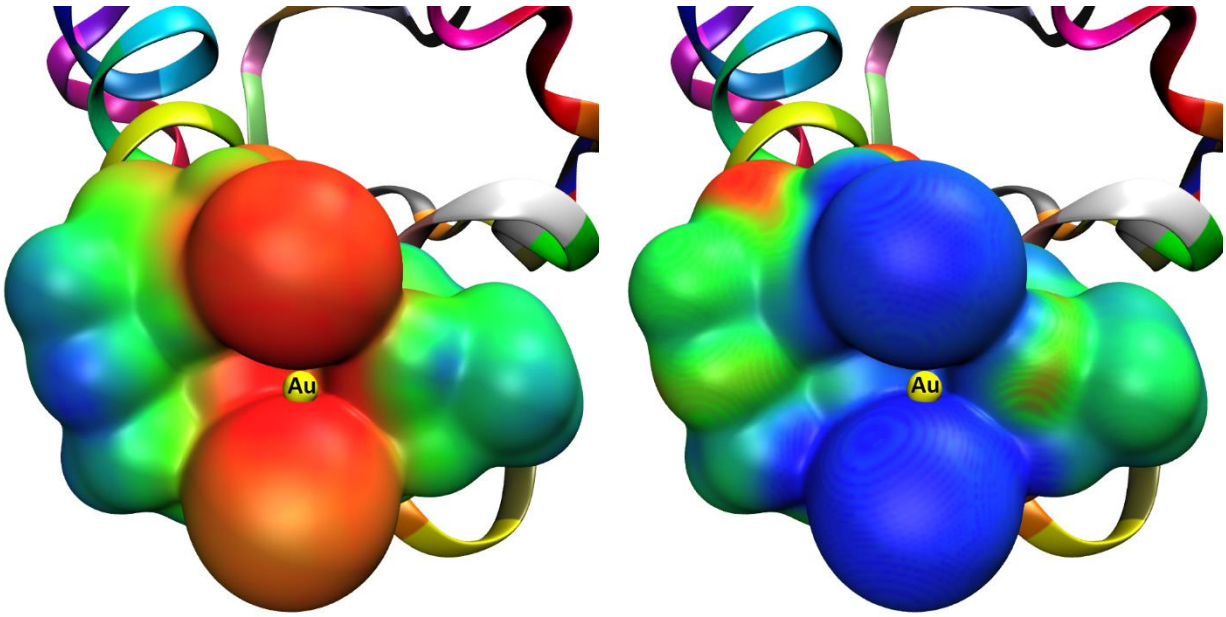


Figure 3.4: Au^+ binding site of GolB: (left) electrostatic potential from -0.3 au (red) to -0.06 au (blue); (right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.6 bohr (blue). Bound Au^+ is shown as a sphere. The binding site electron density is simulated with QM/MM embedding, and the full protein structure is shown as a ribbon diagram.

3.6 Thyroxine Binding Globulin Complex

Our next study is of thyroxine hormone binding to TBG. Thyroxine hormone regulates cellular oxygen consumption and the metabolism of body and brain.[145-146] The concentrations of thyroxine in the tissues and its controlled release are fundamental because its higher

concentration leads to hyperactivity and its deficiency develops dormancy.[145, 147] In the blood, thyroxine is carried by TBG, which binds it with a very high affinity.[148] Thyroxine combines chemically hard Lewis acids and bases with soft iodine-substituted phenyl groups.

Figure 3.5 shows the structure, MESP, and overlap distance of bound thyroxine. The MESP is dominated by the charged peptide region at left. The iodo-substituted phenyl is relatively un-charged. The overlap distance is very large on the iodo substituents and smaller on the chemically hard peptide region. It is interesting to note that aminopropionate group (MESP = -0.047 au, $D(\vec{r})$ = 3.00 bohr) and carboxylic oxygen (MESP = -0.047 au, $D(\vec{r})$ = 2.76 bohr) have equal charge but different overlap distance which shows that former is a soft Lewis-basic sites compared to latter. The oxygen atom (MESP = -0.025 au, $D(\vec{r})$ = 2.89 bohr) bridging the tyrosine and phenolic groups is less negative compared to phenolic oxygen (MESP = -0.042 au, $D(\vec{r})$ = 2.79 bohr) but its large overlap distance shows that it is more diffuse and soft Lewis-basic sites.

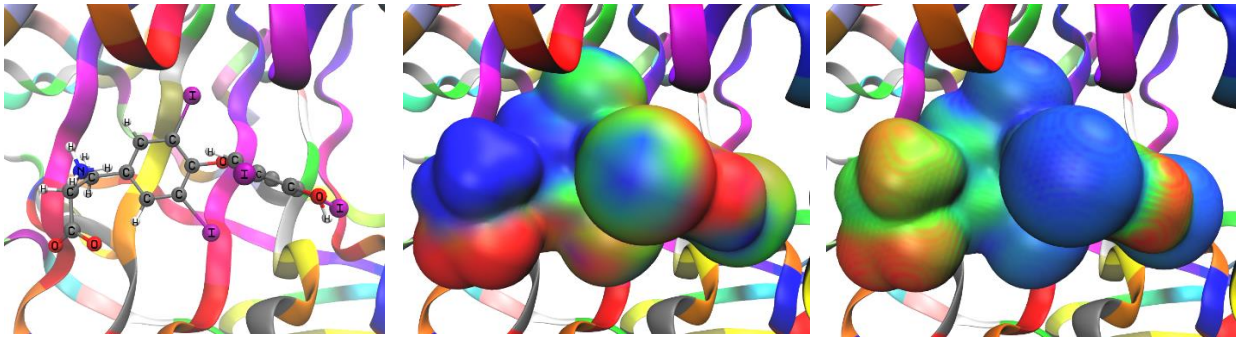


Figure 3.5: Graphic depictions of thyroxine high-level layer: (left) electrostatic potential from -0.05 au (red) to 0.05 au (blue); (right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.6 bohr (blue). The low-level layer is not shown for clarity.

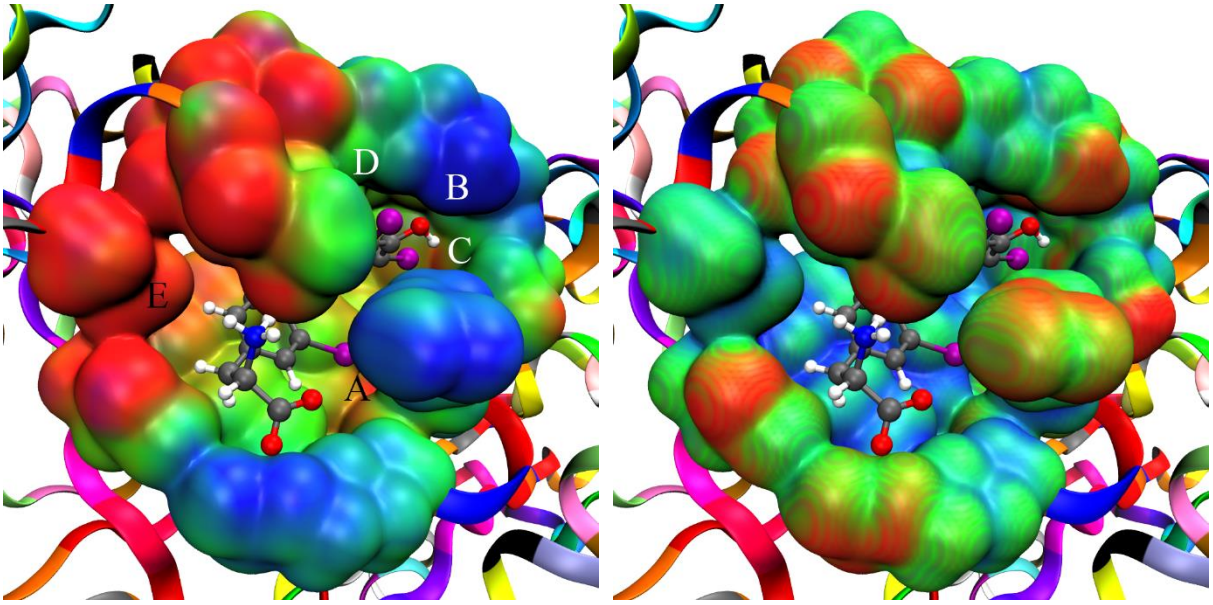


Figure 3.6: Pocket atoms of TBG at high-level layer: (left) electrostatic potential from -0.05 au (red) to 0.05 au (blue); (right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.6 bohr (blue). The biotin molecule in the pocket (low-level layer) is shown using ball-and-stick model.

Fig. 3.6 shows the structure and MESP of the TBG binding site. The Lewis-acidic and Lewis-basic regions adjacent to thyroxine's peptide region (A) and phenol OH (B) are relatively chemically hard, with small overlap distance. In contrast, the hydrophobic pocket containing the iodo-substituted phenyl (C-D) is relatively chemically soft.

3.7 Metal Binding: Formylglycine-Generating Enzyme

Our next example shows how the overlap distance and MESP can distinguish the selectivities of two binding sites on a single protein. FGE, recognized as a powerful tool in protein engineering[149-154], is a unique copper protein which catalyzes the oxygen dependent conversion

of specific cysteine residues of arylsulfatases and alkaline phosphatases on client proteins to formylglycine [151, 155]. Insufficient FGE in human cells causes multiple sulfatase deficiency, a rare and fatal disease [156-157]. FGE has one high-affinity[151] binding site for Cu^{+} [158] and two binding sites for Ca^{2+} [151] (Ca1 and Ca2 in Fig. 3-4). Copper has been reported to increase the in vitro activity of FGE up to 20 fold [155]. The crystal structure used in the present studies, reported by Meury et al.[151], was crystallized using Ag^{+} instead of Cu^{+} to avoid oxidation related problems. Ag^{+} binding pocket constitutes the sulfhydryl groups of the catalytic active site in two cysteines Cys269 and Cys274 [151]. This binding site can also hold Cd^{2+} with some specific conformational changes at the active site [151]. The binding site of Ca1 consists of carboxylate group of Asp202 and hydroxyl groups of Ile189, Asn188 and Tyr204 in addition to a coordinated water molecule whereas the hydroxyl groups of Asn222, Val223, Gly225 and Val227 constitute the binding site of Ca2. Surface MESP and $D(\vec{r})$ captures the hard-soft acid-base selectivity of binding pockets for relatively hard Lewis acid Ag^{+}, Cu^{+} and $Cd^{2+}$and relatively soft Lewis acid Ca^{2+}.

The top of Figure 3.7 shows MESP surfaces of FGE Ag^{+} and Ca^{2+} binding pockets. The binding pocket of Ag^{+} is less negative i.e. weak Lewis basic site due to the involvement of less electronegative sulfur atoms as compared to the binding pockets of Ca^{2+} which contain all electronegative oxygen atoms. The most negative MESP values in the binding pockets are -0.009 au for Ag^{+} and -0.41 au for Ca^{2+}.

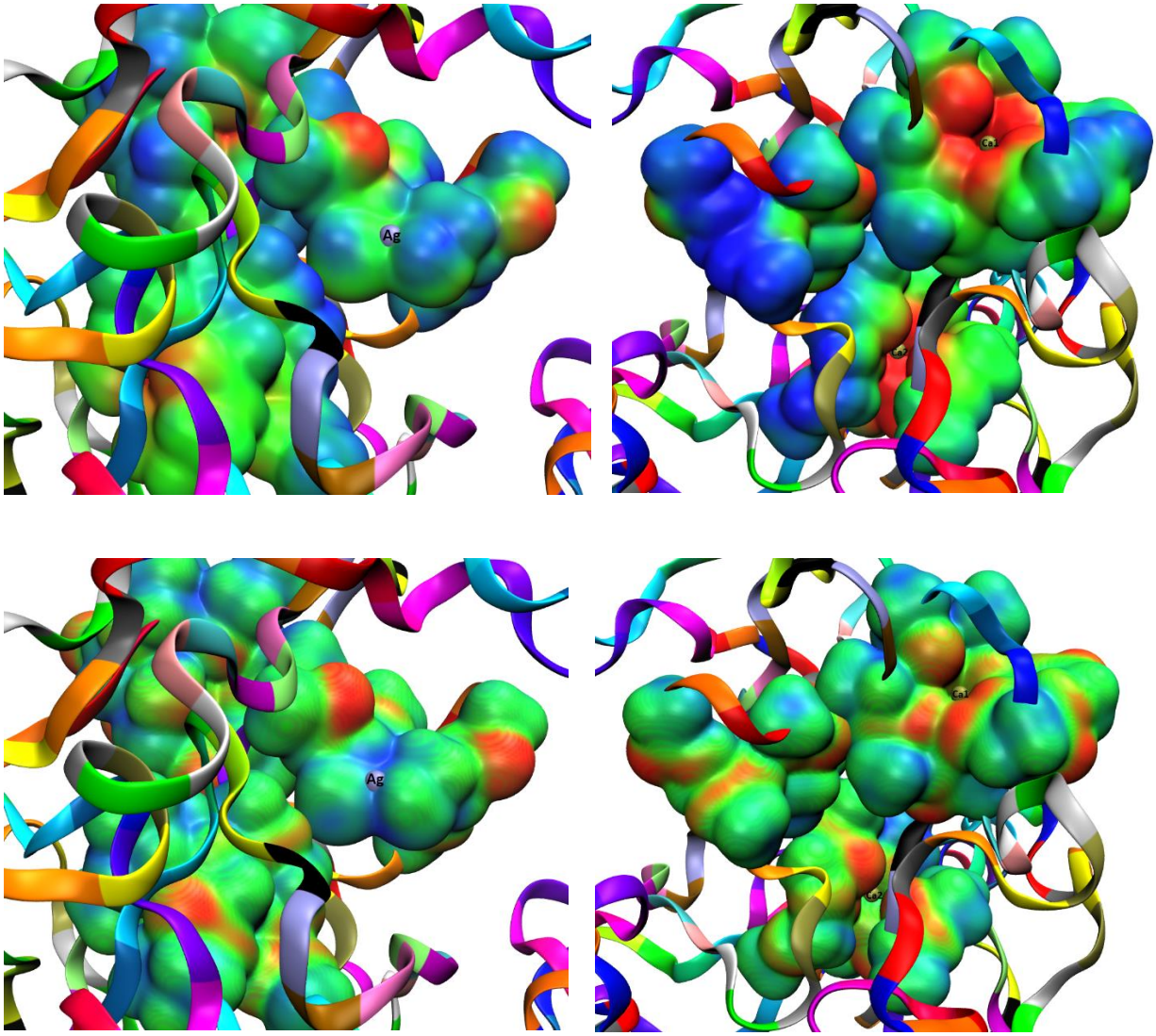


Figure 3.7: (top) Electrostatic potential and (bottom) overlap distance of (left) Ag^{+} and (right) Ca^{2+} binding sites of FGE. (Left) Front view showing Ag^{+} pocket, MESP from -0.06 (red) to +0.04 (blue). (Right) Back view showing Ca^{2+} binding pockets, MESP from -0.30 (red) to -0.10 (blue) and -0.20 (red) to 0.00 (blue) for Ca1 and Ca2 respectively. $D(\vec{r})$ from 2.8 (red) to 3.6 (blue).

The bottom of Figure 3.7 shows the corresponding overlap distances. The binding pocket of Ag+ has an overlap distance much larger than that of Ca2+. The largest $D(\vec{r})$ values in the

binding pockets are 3.5 bohr for Ag^+ and 2.9 bohr for Ca+, consistent with different hardness & softness. Again, the overlap distance and MESP also quantify the different character of the ligands. Cu^+ (MESP = 0.304 au, $D(\vec{r})$ = 1.789 bohr) and Ag^+ (MESP = 0.283 au, $D(\vec{r})$ = 1.893 bohr) are very similar, consistent with their competition for the soft binding site on FGE. This is consistent with experimental findings that Ag^+ is an excellent mimic of Cu^+ in copper transporting ATPases,[159-160] copper sensing transcription factors,[161-163] and copper chaperones[164-165]. The bivalent Cd^{2+} (MESP = 0.625 au, $D(\vec{r})$ = 1.778 bohr is clearly a harder and stronger Lewis acid, consistent with its binding to the other FGE binding site.

3.8 Computational Details

The calculations in Figures 3.1 used the ωB97X-D XC functional[166] and the aug-ccpVTZ- PP basis set.The X-ray based crystal structures of systems, Avidin-Biotin complex (PDB ID 2AVI) [134], TBG (PDB ID 2CEO)[148], gold-specific binding protein GolB (PDB ID 4Y2I) [136] and FGE (PDB ID 5NXL) [167] were obtained from the Protein Databank and employed as the starting points for the present study. Prior to quantum mechanical modeling, all these initial structures underwent a similar preparatory sequence consisting of: adding hydrogen atoms, setting MM parameters for ligands and metal sites, adding counter ions, solvating the systems with a octahedron water box, optimizing the initial coordinates and equilibrating the system to room temperature using MD simulations under NPT conditions. All these steps were carried out using AmberTools.[168] For each system, the final frame of the equilibrated MD trajectories was assumed as a model configurations of the proteins and was used to isolate the protein-ligand/metal coordinates for further quantum mechanical calculations.

All the two-layer ONIOM[169] calculations use Gaussian 09 [90] suite of programs. These involves a combination of high-level QM and low-level MM calculations. Long-range corrected hybrid density functional ωB97X-D[170] which includes empirical dispersion was employed to describe the higher QM layer in combination with the 6-311G(d,p) basis set involving ECP for Iodine atom, whereas the AMBER[171] molecular mechanics were used to treat the low-level layer of the protein molecules. The missing atom type parameters for molecular mechanics were incorporated using *amber=hardfirst* keyword in Gaussian ONIOM calculations. For avidin-biotin and thyroxine-binding globulin complexes the calculations were carried out in two steps. In first step the high-level layer only included the ligand atoms and the rest of protein residues were treated as low-level layer, whereas in the second step, the binding pocket of the protein was treated as high-level layer and the remainder constitutes the low-level layer. For GolB and formylglycine-generating enzyme only the second step involving the binding pocket at high-level layer was performed for which, initially, all the atoms from the residues within a 6Å sphere of the metal or the ligand were included in high-level layer and then the residues beyond 6Å were truncated at chemically reasonable positions. The Gaussian formatted checkpoint files were used to obtain electron density, MESP and orbital overlap distance $D(\vec{r})$ Gaussian cube files using Multiwfn program[172]. For the calculations of orbital overlap distance, an even-tempered grid of 50 exponents was used starting from 2.50 bohr^{-2} and with an increment of 1.50 bohr^{-2}. For the studied systems, the calculated MESP and orbital overlap distance $D(\vec{r})$ cube files were projected over 0.001 (e/bohr^3) molecular electron density surface. This particular isosurface of molecular electron density encompasses approximately 96% of the electronic charge of a molecule[173] and is considered as standard in calculations of electrostatic potential[174]. In the plots of molecular

surfaces, the color scale changes in the direction of red → orange → yellow → green → blue. For MESP density isosurface, red and blue colors shows most negative and most positive values respectively whereas the overlap distance $D(\vec{r})$ increases in the direction of red to blue. In the molecular geometry and density isosurface plots of all systems, the low-level layer of protein is depicted as ribbon diagram colored by residue ID using VMD [175] version 1.9.3.

Chapter 4: Towards Computational Design of Molecules for Noncovalent Interactions: Artificial TMAO Receptors

4.1 TMAO in marine species

TMAO is known to accumulate in marine organisms' tissues at high concentrations to protect against protein-destabilizing effects of urea. One example is the marine elasmobranch which includes sharks, rays, skates, and sawfish. This class of species uses TMAO to protect it from fluctuating temperatures, salinity, high urea and hydrostatic pressure. The effects of urea buildup inhibit cellular functions and destabilize many macromolecular structures, which are counteracted by TMAO.[176] For organisms that use urea as an osmolyte and buoyancy factor, TMAO has been shown to restore enzyme function of proteins and restore them to their native state.[177-178] For deep sea creatures, TMAO and betaine counteract hydrostatic pressure on proteins.[179] TMAO is metabolized to mono-, di-, and trimethylamine, which is also a precursor to marine aerosols and nitrous oxide.[176]

Although the effects of TMAO have been known to marine biologists, the positive correlation between TMAO and increased risk of cardiovascular diseases for humans only recently emerged.[180]

4.2 TMAO in the Body

TMAO is a small molecule present in the blood that participates in changes in whole-body cholesterol metabolism, chronic kidney disease, vascular inflammation, and atherosclerosis.[72] Bacteria in the digestive tract turn choline (a byproduct of lecithin phospholipids found in meat and eggs), betaine, and L-carnitine (mostly found in red meat) into TMA. The bloodstream then absorbs TMA and the liver metabolizes it to produce TMAO using flavin monooxygenases (FMO1 and FMO3).[181-182] TMAO is then used by the body as an osmolyte or cleared by the kidney.[72, 180, 183-185] Studies have shown that the production of TMA/TMAO is through intestinal microbiota,[72] where several families of bacteria are involved.[180, 186-187] Increasing levels of TMAO in the body is still an ongoing debate, where in a crossover feeding trial showed consumption of eggs, beef, and fish increased the blood and urine levels of TMAO.[188] Contradictory to this study, a trial involving 271 people consuming meat, eggs, or fish showed that their diet does not correlate with their levels of TMAO, choline, or betaine.[189]

Our goal is to build a receptor that will bind TMAO and track its pathway in the body. Our proposed receptor design is inspired by the interaction of TMAO with TMAO binding protein TorT (Figure 4.1, right). The oxygen on TMAO hydrogen binds to Trp45, Tyr71, and an Asp42-coordinated water molecule. The aromatic amino acids Tyr44, Trp140, and Tyr252 surround the quaternary amine on TMAO. Unlike most molecules, TMAO can accept up to three hydrogen bonds at its oxygen terminus, which provides a “handle” for selective detection.

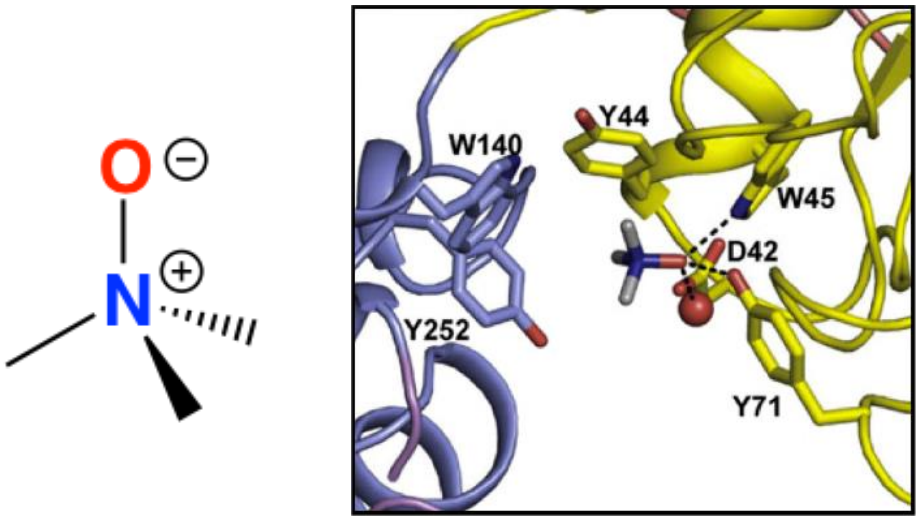


Figure 4.1: (left) Isolated TMAO molecule and (right) crystallographic binding site of TMAO binding protein TorT.[190]

4.3 Computational Study

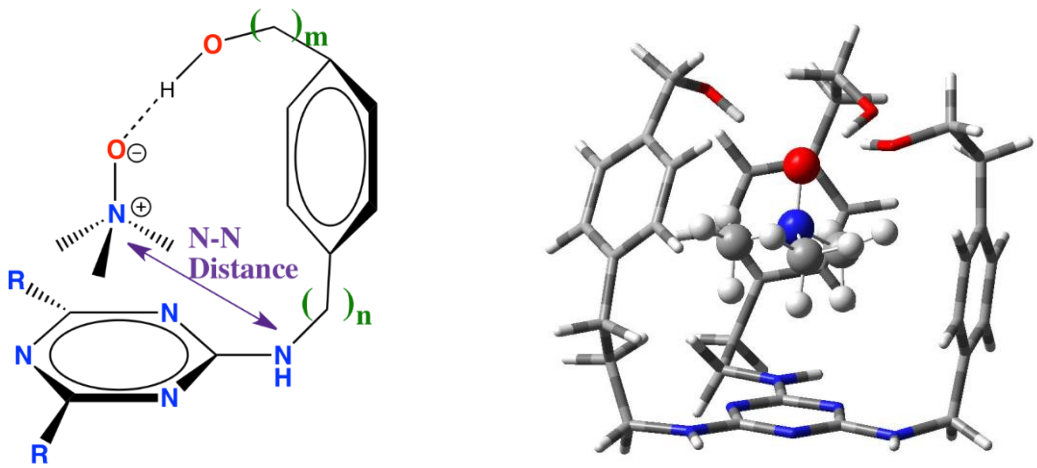


Figure 4.2: TMAO receptor model (left) and optimized geometry of receptor and TMAO complex (right). Red spheres: O and blue spheres: N.

We chose to build on the Simanek group's experience with triazine dendrimers, proposing the receptor in Figure 4.2 (left). We first considered TMAO binding to our proposed receptor's components: melamine, benzene, ethanol, and other similar molecules. Tables 4.1-4 show our most important results, confirming that TMAO, unlike most molecules, accepts three hydrogen bonds to one oxygen atom (see Figures 4.3-5). These tables also compare the binding energy of methanol to three molecules that can have hydrogen bonding interactions. The TMAO molecule not only has a stronger and shorter hydrogen bond distance to the methanol molecules, but it can also form three hydrogen bonding interactions. No other molecule in the test set three hydrogen bonds to a single atom. Based on these results, a receptor containing three hydrogen bonding donors would be more selective towards TMAO.

Geometry optimization of an isolated melamine (triazine with three primary amine linkers) and TMAO resulted in melamine arranging off-center to TMAO. The results show that the nitrogen on TMAO and the three nitrogen atoms on the branched amine groups on melamine gave three different N-N distances (labeled in purple in Figure 4.2, left). This tells us that each linker on melamine will potentially have various lengths to favorably hydrogen bind with the oxygen on TMAO. To test this hypothesis, we adjusted the three linker lengths by changing the number of methyl groups (n and m labeled in green in Figure 4.2, left) from one to three. While freezing triazine and TMAO to its optimized geometry and varying n and m, we tested each hydrogen on the amine (labeled left/right) to give binding energies. The best linker lengths that gave the most favorable binding energies for each amine are (right) n=3, m=1; (left) n=2, m=2; and (left) n=1, m=2. Figure 4.2(right) shows the geometry optimized receptor with the best linker lengths interacting with TMAO with a binding energy of -31.1 kcal/mole. A comparative study of

a receptor with (right) n=2, m=2; (left) n=2, m=2; and (left) n=2, m=2 gave a binding energy of -21.5 kcal/mole. This test confirms that the receptor shown in Figure 4.2 (right) potentially yields the lowest binding energy.

MeOH	TMAO Vacuum	MeOMe Vacuum	MeCOMe Vacuum	MeCOH Vacuum
1	-13.6	-6.4	-9.2	-7.2
2	-12.3	-5.2	-6.6	-6.2
3	-10.8	--	--	--

Table 4.1: Binding energies (in kcal/mole) of 1, 2, and 3 methanol(s) hydrogen bonding to various hydrogen bond acceptors in vacuum.

MeOH	TMAO Water	MeOMe Water	MeCOMe Water	MeCOH Water
1	-7.6	-7.1	-7.1	-5.4
2	-7.6	-5.6	-5.6	-3.8
3	-6.0	-6.0	--	-6.5

Table 4.2: Binding energies (in kcal/mole) of 1, 2, and 3 methanol(s) hydrogen bonding to various hydrogen bond acceptors in water.

MeOH	TMAO	MeOMe	MeCOMe	MeCOH
1	1.71	1.92	1.91	1.94
2	1.75	1.95	1.94	1.95
	1.75	1.96	1.95	1.97
3	1.77	1.90	1.86	1.88
	1.78	1.91	1.93	1.95
	1.79	3.12	3.00	3.56

Table 4.3: Bond distances (in Angstroms) between 1, 2, and 3 methanol(s) and various hydrogen bond acceptors in vacuum.

MeOH	TMAO	MeOMe	MeCOMe	MeCOH
1	1.67	1.83	1.83	1.85
2	1.70	1.84	1.86	1.87
	1.71	1.85	1.93	1.88
3	1.72	1.93	1.86	1.89
	1.73	1.95	1.90	1.90
	1.73	2.04	3.03	2.22

Table 4.4: Bond distances (in Angstroms) between 1, 2, and 3 methanol(s) and various hydrogen bond acceptors in water.

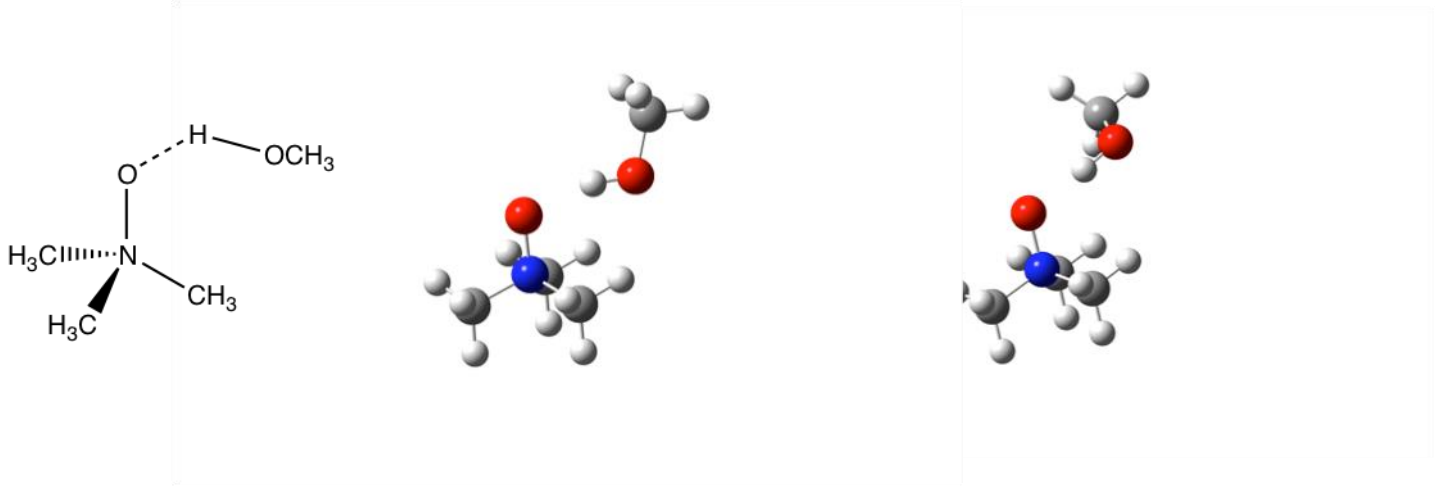


Figure 4.3: TMAO hydrogen bonding with one methanol molecule. Geometry optimization in vacuum (center) and in water (right).

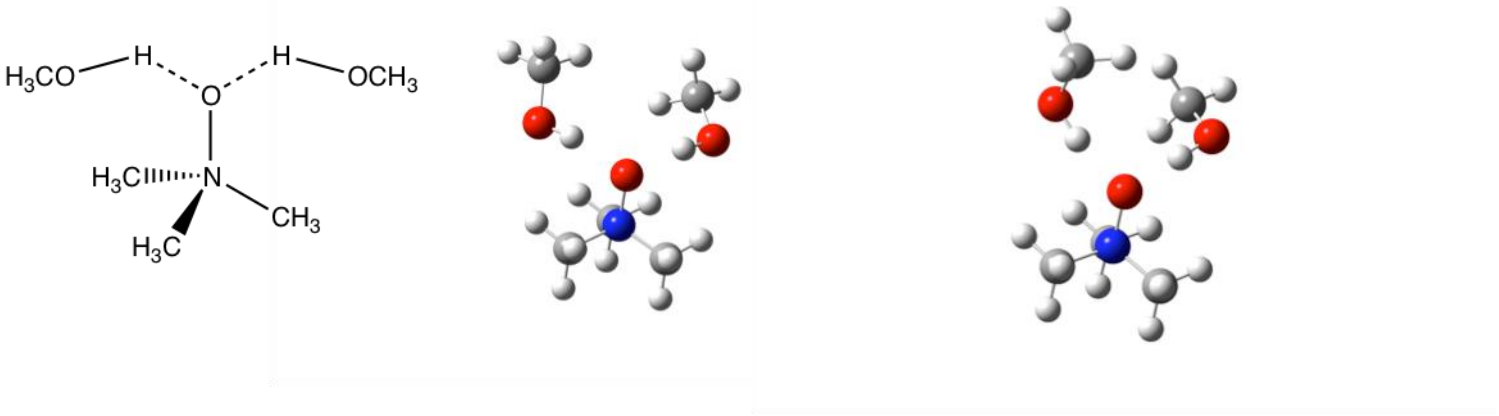


Figure 4.4: TMAO hydrogen bonding with two methanol molecules. Geometry optimization in vacuum (center) and in water (right).

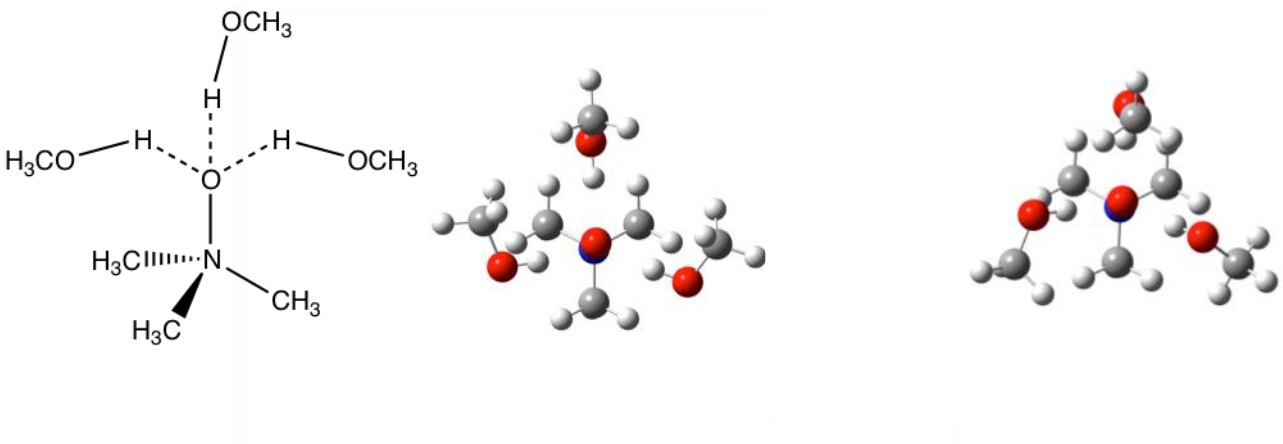


Figure 4.5: TMAO hydrogen bonding with three methanol molecules. Geometry optimization in vacuum (center) and in water (right).

4.4 Computational Details

All the computational work done uses the development version of the Gaussian suite of programs. [90] TMAO receptor studies use DFT WB97XD/6-311++G(2d,2p). This method and basis set are used because of its capability of measuring dispersion and non-covalent interactions. BSSE[191] is used in this study due to basis function overlap between atoms. Binding energies are calculated by taking the difference between the energy of the complex and the sum of energies of the separately geometry optimized TMAO and receptor. A negative binding energy means that the interaction is favorable.

Chapter 5: Computational Study of Acid Catalyzed Hydrolysis on Lignin Models with β-o-4 Linkages

5.1 Lignin Background

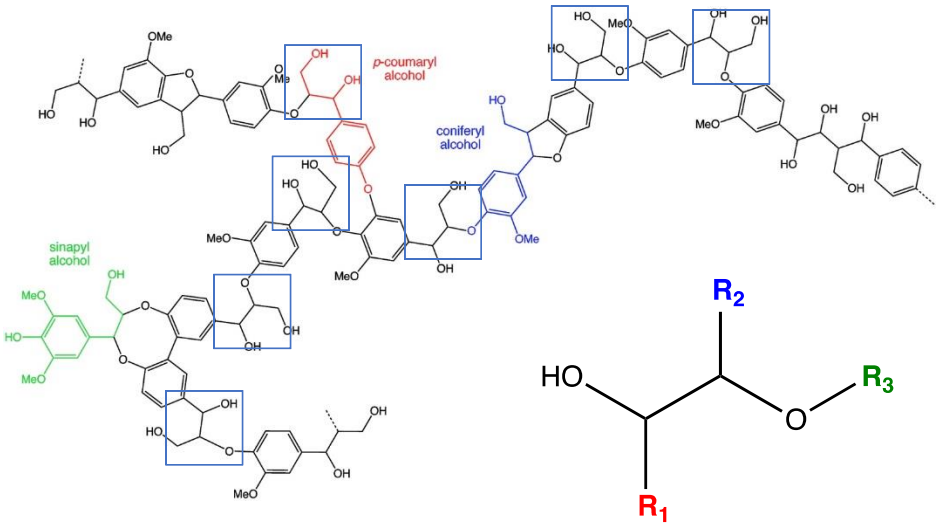


Figure 5.1: Representative structure of lignin (top), where the boxed regions are β-o-4 linkages. β-o-4 linkage model compound, where R_1, R_2, and R_3 are varied (bottom).

Lignocellulose is the fourth largest source of energy in the world (in the form of heat and power) and is produced at about 170 billion metric tons annually.[192-193] It consists of three main components: hemicellulose, cellulose, and lignin. Hemicellulose and cellulose are currently being used as biofuels and important chemicals, but lignin is a waste stream in biorefinery and pulping processes.[73-74] However, lignin is used in some production to make phenolic resins, polyurethane foams and emulsifying agents,[194] automotive brakes and tires, wood panel products,[195] but the

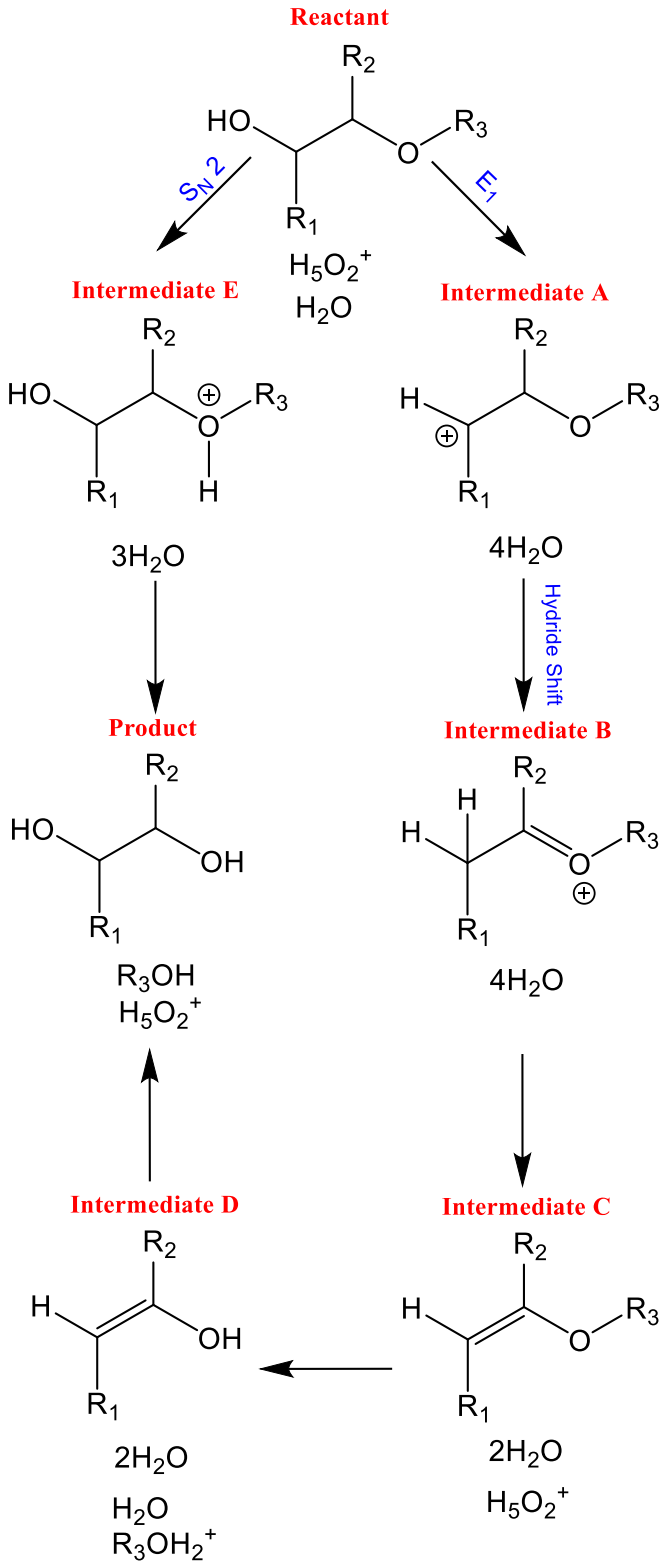


Figure 5.2: E_1 and dehydration (right scheme) vs protonation and nucleophilic substitution (left scheme) of β-o-4 linkages.

high-value products from lignin are unknown due to the difficulty on separating its heterogeneous molecular structure and lack of efficient depolymerization methods.[75-76] The most efficient method to separate lignin from cellulose is with phosphoric acid and acetone mixture,[196-199] but this technique typically leads to the decomposition of its aromatic compounds. Lignin is also a renewable source and is composed of high-volume aromatic compounds for chemical industry that would not compete with food production, which then do not cause social and ecological consequences.[200]

Lignin is primarily a complex cross-linked macromolecule that adds strength and rigidity to cell walls and its composition and its content in biomass differ between the types of plants.[201-203] Lignin polymer has mainly aromatic molecules[192] primarily composed of the tree units:

sinapyl (3,5-dimethyoxy 4-hydroxycinnamyl), coniferyl (3-methoxy 4-hydroxycinnamyl), and p-coumaryl(4-hydroxycinnamyl) alcohols, The composition of lignin varies with even the same plant species due to factors such as growing environment, area, and analysis methods.[204]

The β-o-4 is the most abundant ether linkage in lignin[205] consists of 48-57% of inter-subunit linkages in soft wood lignin and 43-65% in hardwood lignin.[192, 206] A β-o-4 linkage is a bond formed between the beta carbon of the aliphatic side chain and the oxygen atom attached to the C4 position of the aromatic moiety as shown in figure 5.1, bottom. R_1, R_2, and R_3 represent type of substituents and are labeled in order in Table 5.1. For example, p-phenol-MeOH-Ph is where R_1=para-phenol, R_2=methanol, and R_3=phenol. Typically, the process to separate lignocellulose into cellulose, hemicellulose, and lignin is through the use of organic solvents and acids.[207] In these processes, lignin β-o-4 linkages are broken due to the liability of alkyl aryl ethers in acidic conditions.[208]

There have been many mechanistic studies of model compounds bearing β-o-4 linkages, where the use of model compounds would help elucidate the mechanism in breaking down lignin without decomposing the aromatic compounds.[209] Model compounds 2-phenoxy-1-phenylethanol,[210-212] 1-phenyl-2-phenoxy-1,3-propanediol[212] and 1-(4-hydroxyphenyl)-2-phenoxy-1,3-propanediol[212] match our studies here of Ph-H-Ph, Ph-MeOH-Ph and p-Ph-MeOH-Ph, respectively. Typically, experimental studies of β-o-4 linkages use much more complex molecules involving muti-substituted phenyl rings[209, 212-218] to focus on keeping the aromatic moiety intact.

A previous computational study proposes an E_2 dehydration (rather than E_1) followed by an S_N2 with the use of ionic liquids.[219] The E_1 pathway is limited by the formation of a free carbocation, rather than a complexed carbocation through the E_2 pathway. Experimentalists have shown that some ILs are a promising solvent that can dissolve both cellulose[220-221] and lignocellulose composites.[222] An ionic liquid is defined as an organic salt that is liquid below 100 C°.[223] Previously, it was shown that the extended pi-systems of benzimidazolium ionic liquids give a stronger interaction with lignin,[224] which supports the reason for dissolution of lignin in ILs.

We aim to study the substituent effect of model compounds with β-o-4 linkages to give a better understanding of how lignin undergoes hydrolysis. We use a continuum solvent of dichloroethane,[225] which has a dielectric constant that is comparable to many ILs. Previous studies have confirmed that using continuum solvent models can provide a chemically reasonable picture of reactions in complex solvents such as ionic liquids.[225-227] A previous mechanistic study of the role of substituents in the cleavage of β-o-4 has been done previously,[228] but here we look at smaller molecules as well as the larger structures. Our approach here is more simplistic: which oxygen containing group in the β-o-4 linkage would protonate and initiate the dehydration mechanism. This gave us the following pathways which were previously suggested:[219] an E_1 followed by a dehydration and a protonation and nucleophilic substitution (S_N2) shown in Figure 5.2.

5.2 Overall Results

We begin by considering the E_1 and S_N2 pathways for lignin hydrolysis. Both pathways involve producing a cationic intermediate, but on different atoms. The E_1 pathway produces a cation on the carbon attached to R_1, where the stability would depend on the substituent R_1. This is formed due to the alcohol group being protonated and then eliminated by dehydration. Intermediate B is a hydride shift from the previous intermediate, stabilizing the carbocation and forming an oxygen cation. The proton is then pulled off by a water molecule, followed by electrons shifting to the neighboring bond and stabilizing the oxygen cation in intermediate C. Intermediate D is formed when the oxygen is protonated and a water molecule undergoes a nucleophilic substitution on the carbon attached to R_2, pushing off the protonated oxygen along with R_3. Then, the final product is a simple hydrolysis across the double bond. The S_N2 pathway is a protonation on the other oxygen atom, forming a cation, which is similar to intermediate C to D, but without a double bond.

	A*	B*	C*	D*	E*	Product*	Path**
H-H-H	N/A	5.2	4.5	4.5	9.1	0	N/A
MeOH-H-H	N/A	2.4	4.7	4.7	11.3	0	N/A
Ph-H-H	15.1	7.7	2.5	2.5	9.3	0	S_N2
H-H-Ph	N/A	6.7	6.8	10.0	15.7	3.8	N/A
Ph-H-Ph	15.4	8.9	3.7	7.0	11.7	3.0	S_N2
Ph-MeOH-Ph	13.7	5.1	2.9	2.0	10.3	0.9	S_N2
m-MeOPh-H-Ph	15.0	9.1	3.9	7.1	17.2	3.0	E_1
Ph-H-o-MeOPh	15.7	7.3	2.9	2.1	12.3	-4.0	S_N2
m-MeOPh-MeOH-o-MeOPh	15.2	6.5	5.7	4.7	11.8	-2.5	S_N2
p-phenol-MeOH-Ph	7.8	6.4	3.9	2.9	11.4	1.7	E_1
p-MeOPh-MeOH-Ph	6.4	5.2	3.0	2.5	12.5	1.7	E_1
m-phenol-MeOH-Ph	16.4	6.4	3.3	2.4	11.9	1.4	S_N2
m-MeOPh-MeOH-Ph	15.4	7.3	4.5	6.0	12.9	1.5	S_N2
p-phenol-H-Ph	6.7	6.5	2.1	5.1	11.6	1.6	E_1
p-MeOPh-H-Ph	6.0	7.6	3.9	7.4	16.7	3.0	E_1
m-phenol-H-Ph	16.1	9.4	3.4	6.2	13.9	3.0	E_1

* All relative energies are reported in kcal/mol

**Predicted lowest energy pathway

Table 5.1: Relative energies of all the intermediates and the products in kcal/mol. Substituents R_1, R_2, and R_3 are labeled in the order it is written in the table. Note that all species with hydrogen on R_3 give relative energies of 0 to produce the product, since the reactant and product are the same. Species with N/A relative energies for intermediate A converged directly to intermediate B.

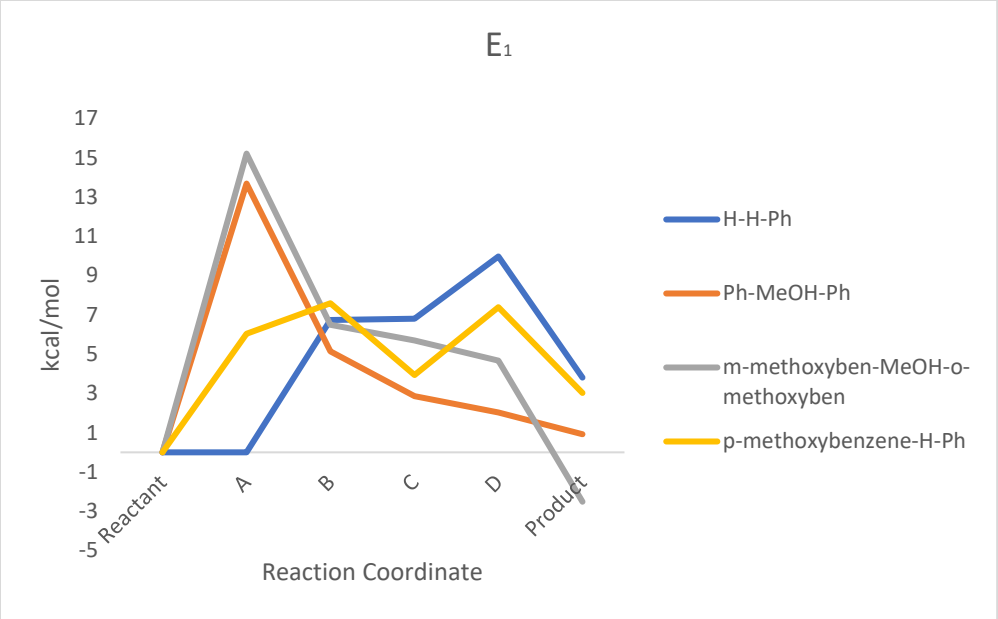


Figure 5.3: Reaction coordinate of some E_1 reactions from reactant to product for representative species studied.

Figure 5.4: Reaction coordinate of some S_N2 reaction from reactant to product for representative species studied.

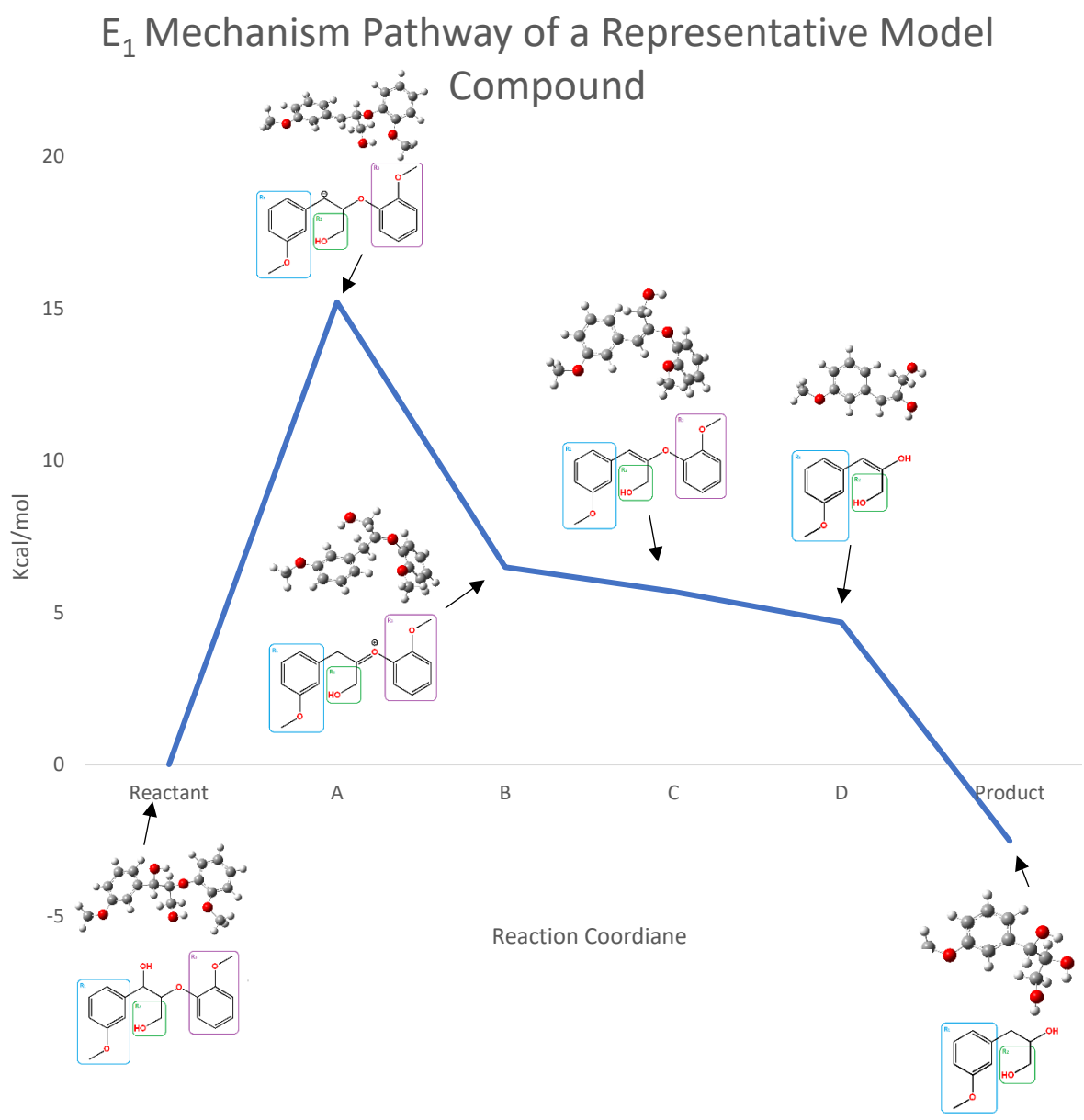


Figure 5.5: Model compound m-MeOPh-H-MeOPh E_1 relative energy pathway. The figure includes the computed structures of the reactant, product, and intermediates A, B, C, and D. Substituents R_1, R_2, and R_3 are labeled in the drawn figures.

5.3 Production of Intermediate A

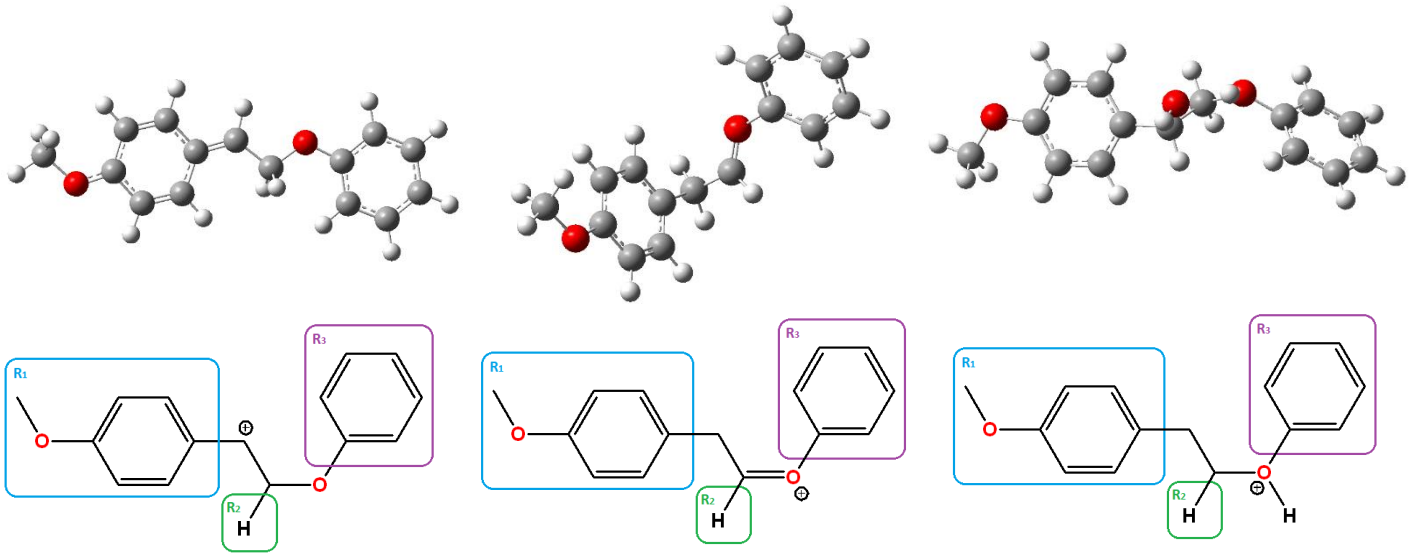
H-H-Ph, H-H-H, and MeOH-H-H did not produce an intermediate A, but the calculation converged directly to intermediate B, which is the hydride shift. p-MeOPh-H-Ph, p-MeOPh-MeOH-Ph, p-phenol-H-Ph, and p-phenol-MeOH-Ph are the most stable species as intermediate A at about 7 Kcal/mol in reference to its reactant. Out of these four molecules, p-methoxybenzanol containing substrates are the most stable intermediate due to an extended conjugation between the phenyl ring and the methoxy group as seen in figure 5.6, left. p-methoxybenzanol is also a more electron-donating group than p-phenol, giving more stability to the carbocation.

The next most stable specie is the Ph-MeOH-Ph at about 14 Kcal/mol. This molecule produces comparable stability to the unsubstituted phenyl ring, which does not stabilize nor destabilize the carbocation. The molecules that have a substituent on the meta position, m-MeOPh-H-Ph, m-MeOPh-MeOH-o-MeOPh, m-phenol-MeOH-Ph, m-MeOPH-MeOH-Ph, m-phenol-H-Ph, are less stable due to substituents being orto-para directing, not meta.

5.4 Production of Intermediate B

The shift from intermediate A to B is energetically favorable for all the molecules studied except from p-MeOPh-H-Ph. In intermediate A, this molecule participates in extended conjugation from its methoxy group to the carbocation, causing stability. Intermediate B's conjugation is only isolated in the ring and becomes energetically unfavorable due to the oxygen atom with a formal charge +1. This is seen in figure 5.6.

MeOH-H-H has the most stable intermediate at 2.4 Kcal/mol due to intramolecular hydrogen bonding as seen in figure 5.7. The rest has generally about the same reaction energy, while m-phenol-H-Ph, m-MeOPh-H-ph, and Ph-H-Ph are the most unstable species due to R2 being H and not MeOH.



Relative Energy (kcal/mol)		
Intermediate A	**Intermediate B**	**Intermediate E**
6.0	7.6	16.7

Figure 5.6: p-MeOPh-H-Ph intermediates A (left) and B (middle), and E (right) where intermediate A is more stable than B due to extended conjugation with methoxy group and R_1 stabilizes the carbocation over R_3, giving a favorable path through E_1.

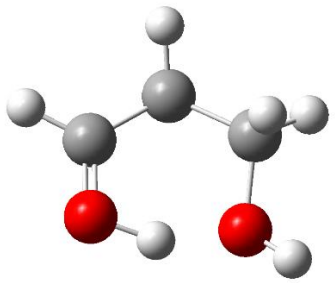


Figure 5.7: MeOH-H-H intramolecular hydrogen bonding as intermediate B.

5.5 Production of Intermediate C

In general, intermediate C is more stable than B since it does not have any charges associated with it. MeOH-H-H rose in reaction energy since the intramolecular hydrogen bonding distance went from 2.109A to 2.574A, causing a less stable intermediate C.

5.6 Production of Intermediate D

Ph-H-o-MeOPh has the lowest reaction energy (-0.46 kcal/mol) due to o-methoxybenzanol intramolecular hydrogen bonding. H-H-Ph is the most unstable molecule since it does not provide an extended conjugation, while the other species studied does. The species that decreased in relative energy are Ph-MeOH-Ph, Ph-H-o-MeOPh, m-MeOPh-EtOH-o-MeOPh, p-phenol-MeOH-Ph, p-MeOPh-MeOH-Ph, and m-phenol-MeOH-Ph. These molecules have intramolecular hydrogen bonding between R_2 = methanol and when R_3 is replaced with a hydrogen, which gives stability when forming this intermediate.

5.7 Overall Reaction Thermodynamics

The product and reactant are identical for species with R_3= hydrogen, giving a relative energy for the formation of product to be 0. For most of the species, the final product is energetically uphill except for the specie with -o-MeOPh attached on R3. These molecules have the lowest reaction energy due to o-methoxybenzanol intramolecular hydrogen bonding as seen in figure 5.8. This is followed by species with R2 being MeOH, since this substituent provides an intramolecular hydrogen bonding which ultimately stabilizes the product as seen in figure 5.9.

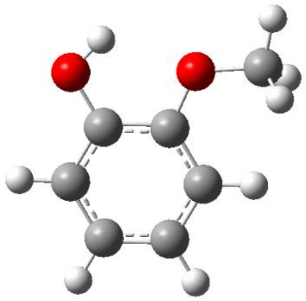


Figure 5.8: Intramolecular hydrogen bonding of R3 being o-methoxybenzanol.

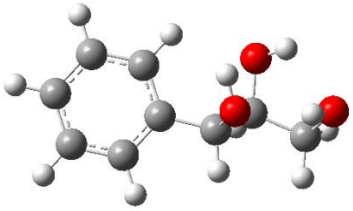


Figure 5.9: Example of intramolecular hydrogen bonding of Ph-MeOH-Ph product with R2 being MeOH.

5.8 Production of Intermediate E

m-phenol-H-Ph is the most unstable species, which is followed by other species with R2 and R3 being H and Ph, respectively. The difference between m-phenol-H-Ph and Ph-MeOH-Ph is the intramolecular hydrogen bonding of 1.74 A and 1.66A, respectively. Since Ph-MeOH-Ph has a stronger intramolecular hydrogen bond, it is thus more stable. Ph-H-H and H-H-H are the most stable intermediate.

5.9 Conclusion

We studied substituent effects on hydrolysis of a β-o-4 linkage and found that the molecules with R_3 = o-methoxybenzanol are the most energetically favorable to produce the product, where both m-MeOPh-MeOH-o-MeOPh and Ph-H-o-MeOPh undergo a smaller reaction barrier through the S_N2 pathway. Out of all the molecules we studied here, m-MeOPh-MeOH-o-MeOPh best matches a real lignin model compound.

Although we have arrived at these conclusions about the pathway a β-o-4 linkage on lignin, we still have not uncovered everything. Lignin β-o-4 linkages have more substituents than what was studied here, and most have multiple substituents on each phenyl ring. This can alter the results quite drastically depending on the substituents added and to which position. A way to include this is by changing the model compound studied, adding more substituents. Another study that can also be explored is the treatment of the extra proton in intermediate D and product. It could be tested whether the proton would be more stable complexed with two water molecules or on the ligand complexed with one water molecule.

5.10 Computational Details

All of the computational work done uses the development version of the Gaussian suite of programs. [90]

Calculations were carried out using Becke's three-parameter global hybrid incorporating B3LYP[93-95] with 6-31+G(d,p) basis set. Energy minima were fully optimized with a continuum solvent in dichloroethane. Frequency calculations were done to ensure that the molecules tested were in fact intermediates, not transition states. The water molecules are treated as shown on the scheme, where a hydronium or ligand with water complex is used to stabilize the proton through hydrogen bonding.

Two forms of solvent models can be used in computational chemistry: explicit and implicit models. Explicit solvent models can provide a spatial description of the interactions of certain solutions with the molecule studied, but oftentimes, these methods can be very computationally demanding due to the increase in the size of the system studied. Implicit solvent models are often more time efficient but fail to describe the direct interactions of solvent-solute. Previously, Cramer and coworkers have extended their SMD[229] continuum solvent model to ionic liquids.[230] The studies here use the SMD continuum solvent model, which is based on the charge density of the solute. We use a continuum solvent of dichloroethane,[225] which has a dielectric constant that is comparable to many ILs.

Chapter 6: An in silico Approach to Developing Alternative Drugs to Warfarin

6.1 Warfarin and VKOR

Warfarin (brand named Coumadin) is a medication used as an anticoagulant drug to prevent blood clots. It is commonly used to treat deep vein thrombosis, pulmonary embolism, and to prevent strokes and heart attacks. The recommended dosage for each patient varies depending on age, gender, weight, and co-medications; and its therapeutic window is very narrow, where over-prescribing causes bleeding or purple toe syndrome and under-prescribing increases the risk of blood clots.[231-232] Often, patients are required to be monitored by a doctor when they are prescribed warfarin.[232]

Warfarin binds to VKOR enzyme and inhibits two of the three steps in the vitamin K cycle (Figure 6.1).[233-236] It inhibits VKR, resulting in depletion of the reduced form of vitamin K (KH2)[233] and also inhibits VKOR by increasing circulatory KO relative to vitamin K.[237] The comparison of the structures of KO vs warfarin is shown in Figure 6.2. VKOR is essential in the vitamin K cycle, where it recycles vitamin K to be used for carboxylation of glutamic residues within the vitamin K-dependent coagulation factors.[238] Once the glutamic acids on the coagulation factors are carboxylated to Gla, they participate in blood coagulation by binding to calcium ions, resulting in conformational changes in the Gla domain, thus exposing the hydrophobic residue.[232] This

hydrophobic residue is then inserted into the phospholipid layer of the platelet membrane, causing coagulation factor activation.[239-240]

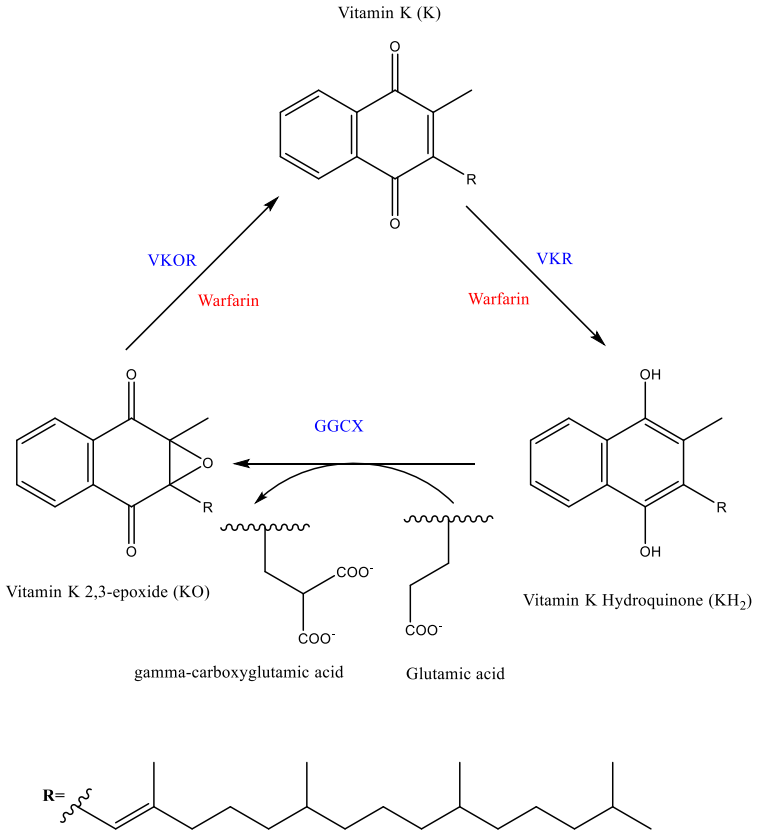


Figure 6.1: Vitamin K cycle

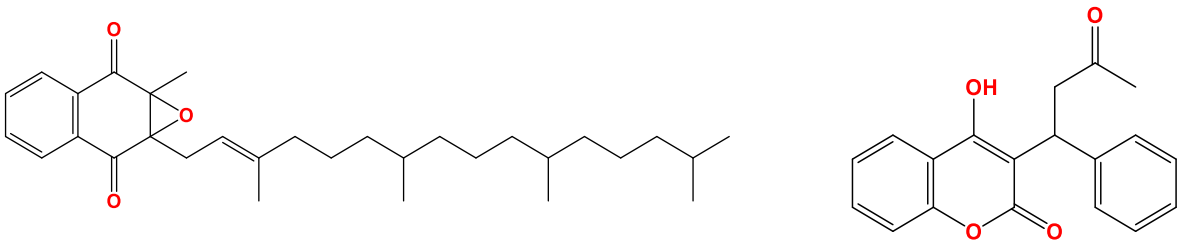


Figure 6.2: Vitamin K 2,3-epoxide (left) and warfarin (right) structures.

To this day, hVKOR has not been crystallized, and there have been ongoing disagreements on its structure and warfarin's binding site. The disagreements stem from whether there are three or four transmembrane helices in hVKOR structure. [241-243] More recently, Tie et al.[244] suggested a 3 TM structure of hVKOR based on their immunofluorescence results and performed a MD simulation on a suggested structure. Oldenburg[245] however suggested a 4 TM structure, where the loop cysteines are arranged at an appropriate location to suggest electron transfer to the center of the protein.[246-247] Our studies here aim to validate the structure of hVKOR by using computational chemistry.

6.2 hVKOR Mutations

One factor that affects patients' optimal warfarin dosage is mutation in VKOR[248] (patients can also have warfarin resistance due to increased metabolic clearance, decrease absorption, or abnormal pharmacodynamics).[231] There are twenty-eight reported single mutations in hVKOR,[249] where we will study eleven of those mutations here (Figure 6.3). The effects of these mutations vary- Tyr139Phe,[250-252] Leu120Glu, Leu128Glu/Ser/Arg,[251, 253-254] and Val29Leu have shown to decrease the affinity of warfarin to the active site of the enzyme; Arg58Gly, Trp59Arg, and Asp36Tyr can disturb the ½-helix as a lid;[247, 255] and Arg35Pro, Val45Ala, and Val66Met could possibly cause a conformational change to the active site.[247, 256]

6.3 hVKOR Mutation *in vitro* Studies

There has only been one *in silico* study of hVKOR mutant binding with warfarin and KO,[257] but not a study to compare all of the found hVKOR mutations. We have compiled data of eleven *in vitro* confirmed mutations in hVKOR (Figure 6.4, right), where four of these mutations are mapped close to the binding site while the other seven are close to the ½-helix. The *in vitro* studies gave the relative VKOR activity of the WT and each of the mutants with and without warfarin at different concentrations. The WT relative activity with no warfarin was set to 100%.

VKOR's activity depends on the concentration of warfarin added to the enzyme, where mutation studies would provide a better understanding of which amino acids are involved when binding warfarin. We have scaled the relative activities of mutant VKOR to WT to find the WIE. WIE is the difference between the mutant studied and WT when the mutant's relative activity without warfarin is scaled to WT without warfarin. These values are reported on Table 6.1 where the higher the WIE is, the less effective warfarin is at inhibiting VKOR activity. The order of WIE from most to least effective is Trp59Arg, Arg58Gly, Val45Ala, Arg35Pro, WT, Leu128Arg, Val29Leu, Leu128Ser, Leu128Gln, Arg33Pro, Leu120Gln, Tyr139Phe. Tyr139Phe and Leu120Gln lower VKOR activity by 71% and 41%, respectively. These amino acids are likely bound to warfarin, where they are in the center of the protein.

The impact of the enzymes on VKOR activity in order of most to least active are Tyr139Phe, WT, Val29Leu, Leu128GLn, Arg33Pro, Arg35Pro, Val45Ala, Arg58Gly, Leu120Gln, Leu128Ser, Trp59Arg, and Leu128Arg. Tyr139Phe is more likely a favorable mutation for binding to KO, but as stated previously, warfarin inhibition lowers with this mutation.

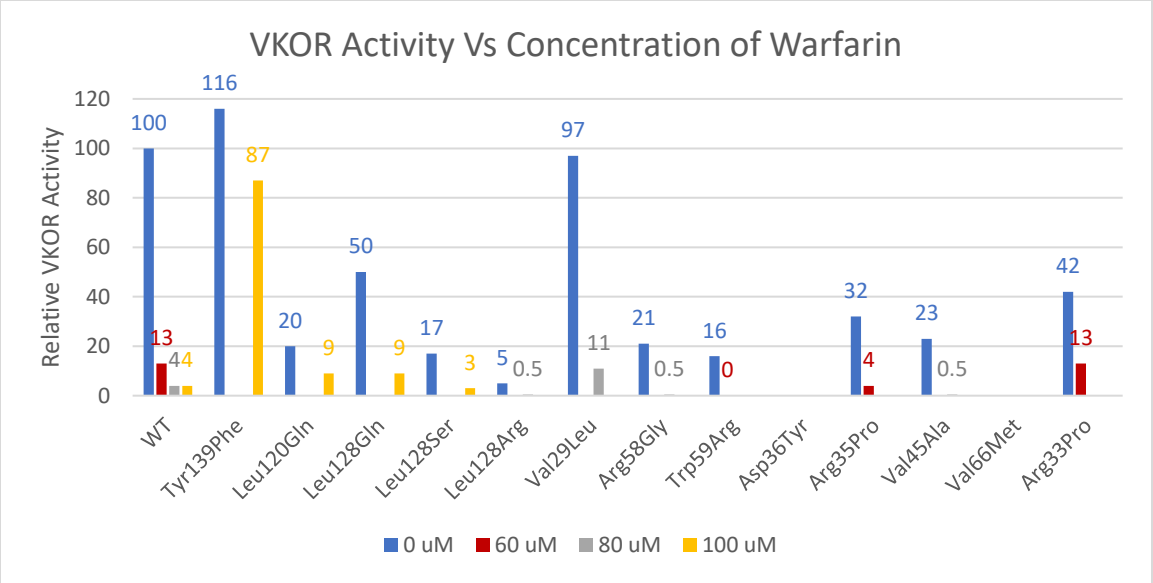


Figure 6.3: Literature reported relative activities for wild-type and mutant VKOR, measured at four different warfarin concentrations 0 uM, 60uM,[21] 80uM,[20] and 100uM[18] or no warfarin.

	Y139F	L120G	L128G	L128S	L128R	V29L	R58G	W59R	R35P	V45A	R33P
WIE	0.25	0.55	0.82	0.82	0.90	0.87	0.98	1	0.88	0.98	0.69

Table 6.1: WIE on VKOR based on the mutations studied. The lower the WIE is, the more effective warfarin is at inhibiting VKOR activity.

6.4 hVKOR Model Studies

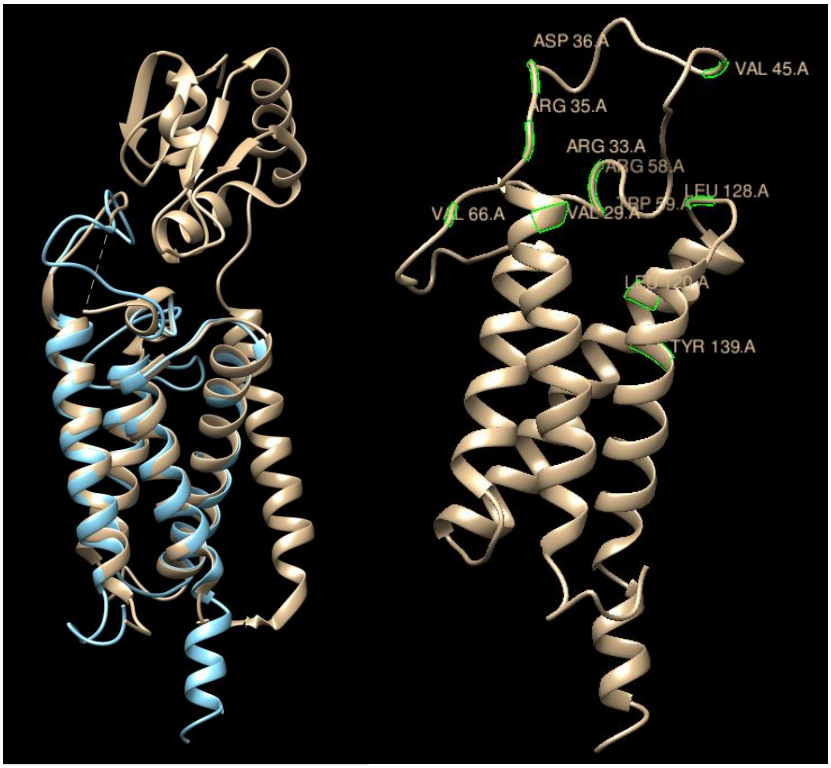


Figure 6.4: (left) Overlap between *syn*VKOR (yellow) and hVKOR homologue Model 1(blue). (right) Model 1 adopted from Oldenburg2017. [245] The highlighted regions in the structure are of the amino acids that will be studied using mutation. RMSD of the overlap between *syn*VKOR (yellow) and hVKOR homologue Model 1(blue) is 0.666 Angstrom.

The homologue of bacterial VKOR, *syn*VKOR, has been found (Figure 6.4, left),[247] but the structure of hVKOR is still being investigated due to difficulty in crystallization. The challenge today is that the mechanism of action and binding regions of warfarin and k vitamers remain elusive.

The First step to this study is to validate the hVKOR homology model built by Oldenburg and coworkers.[245] We chose this structure to compare to since this is the most recently studied hVKOR model over another model.[258] We have determined an alternative homology model using DESTINI (Deep Structural Inference for Proteins),[259] a deep learning protein structure predictor.

6.5 Warfarin Docking Studies

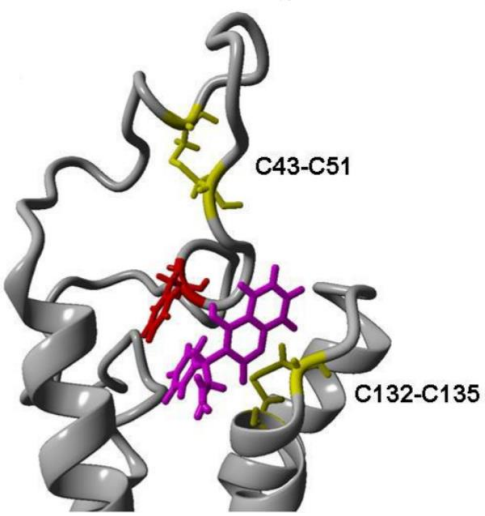


Figure 6.5: Warfarin docked onto hVKOR image from Oldenburg and coworkers result; Ph55 in red.[3]

Oldenburg suggested that there are multiple binding modes for warfarin to their model, where mutations are found within these modes clustering between three interfaces.[260] They suggest that the most preferred binding site is due to Phe55 (Figure 6.5), but in the *syn*VKOR sequence, Phe is not conserved. Rather, a Leu takes that position in the bacterial crystal structure, which would result in a difference in binding affinities of KO to VKORs.

We used our homology model and searched for potential binding pockets using FINDSITE$^{comb2.0}$.[261] We docked warfarin into these pockets using Autodock Vina,[262] compared the electrostatic potential and overlap distance of both warfarin and the binding pockets. We found that binding site 1 is the most probable docking site for warfarin since the *in vitro* data collected supports that prediction.

6.6 Chemistry of Warfarin

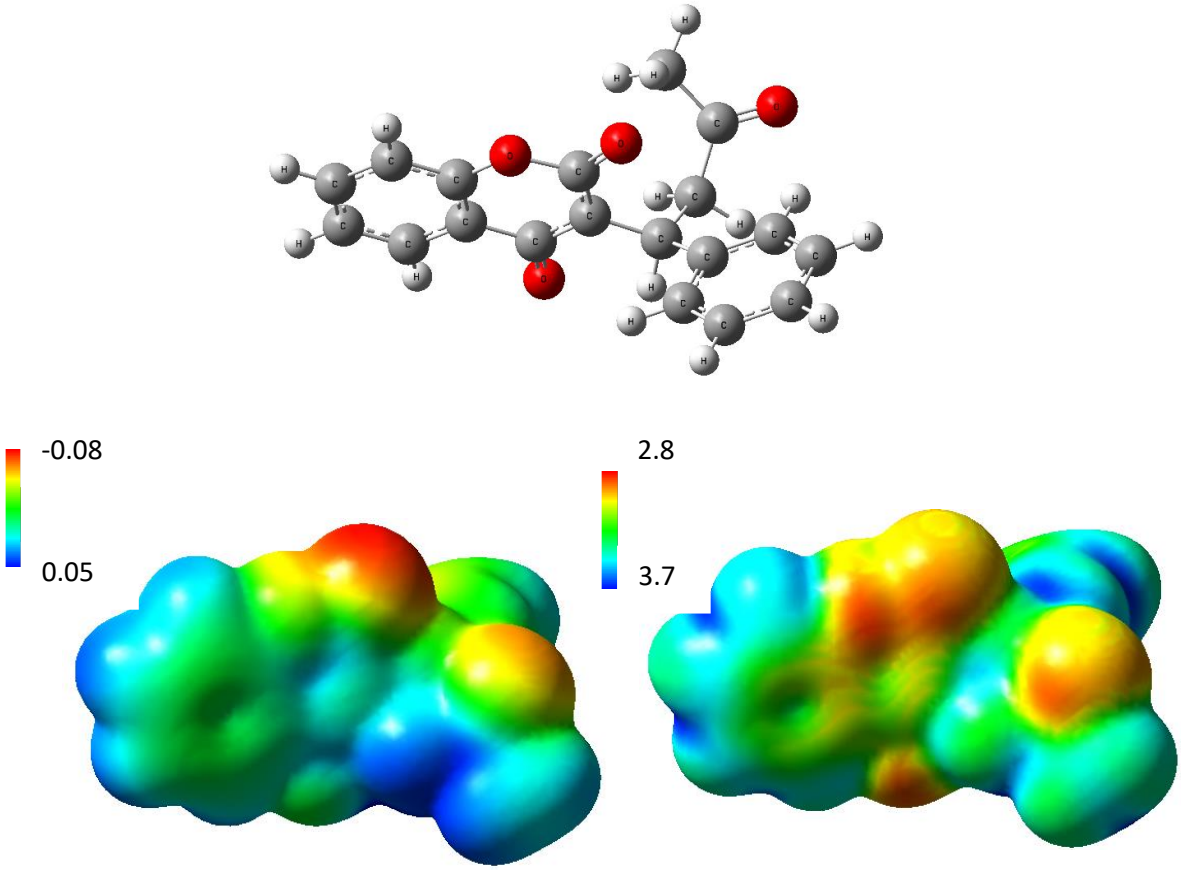


Figure 6.6: (top) Ground state structure, (left) surface electrostatic potential, and (right) surface overlap distance of s-warfarin.

S-Warfarin is 2-5 times more potent than R-isomer in producing an anticoagulant response.[232, 263-264] Due to warfarin's Pka being close to 5,[265] it is more likely to deprotonated in a biological media. We used the deprotonated form of warfarin to analyze its electrostatic potential and overlap distance maps as seen in Figure 6.6. The most negative charge is located on the carbonyl attached to the ring. This site appears to draw the electron density away from

its neighboring oxygen, causing what is shown a less negative charge in that region. This carbonyl more likely participates in electron donation more so than the other oxygen containing groups of the molecule. The isolated carbonyl group attached to the alkyl chain also experiences a slight negative charge, but not as negative as the carbonyl on the ring. This site may also participate in hydrogen bonding. The overlap distance shows that all the oxygen atoms in the system have the largest orbitals, where the electrons are more likely extended enough to participate in electron donation.

6.7 hVKOR Competing Homology Modeling

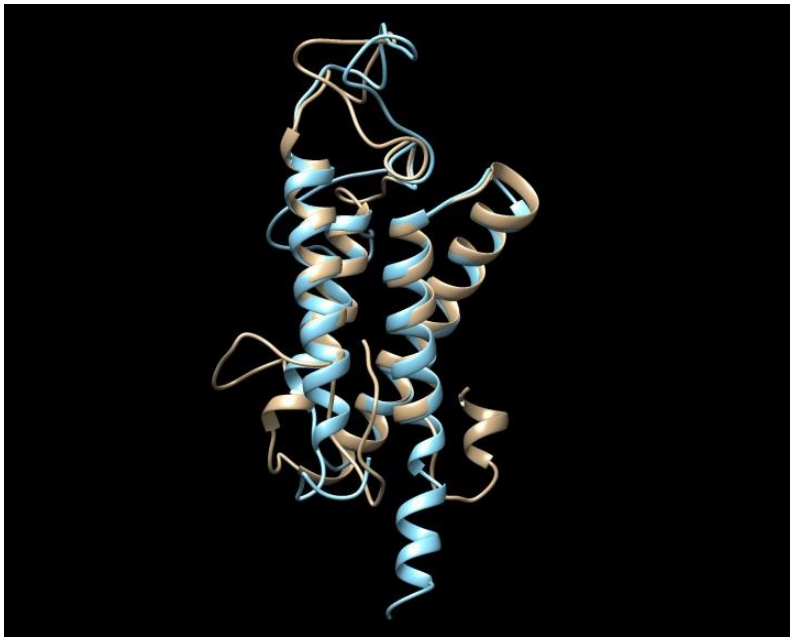


Figure 6.7: Overlap between Oldenburg's model[3] (blue) to DESTINI's model (yellow) of VKOR. RMSD between the two structures is 0.730 Angstrom.

Many have argued that a 3 TM structure is the more likely conformation of this enzyme,[241-243] but a 4TM structure is more plausible due to the placement of important residues. The residues that align well in the 4 TM structure include the CXXC motif, conserved serine/threonine in the ½-helix, and 2 cysteines in the 1/2-segment.[246-247] This topology of the residues allows the transfer of reducing equivalents from the trx-like redox partner, to the loop cysteines, to the active site cystines, and finally to the quinone in VKOR.[266] Another indication for hVKOR having 4 TM helices is that it aligns the structure to that of *syn*VKOR, where the warfarin resistant residues map to the same mutations as one another. Our results confirm hVKOR is a 4 TM structure.

Our hVKOR model compares to Oldenburg's model[3] as seen in Figure 6.7. We used DESTINI to produce our homology model, which produced a similar structure. One key difference is the start and end of the protein, where our prediction gives more random coils than helices compared to Oldenburg's model. Our prediction shows that the C-terminus of the protein comes back up again, which compares to the *syn*VKOR's fifth TM helix. The alpha helical regions of the proteins match almost identically to one another, with a RMSD value of 0.730 Angstrom between the two structures. This RMSD value is fairly small, which may be due to the similarity of YASARA[267] (use by Oldenburg) and DESTINI protein modeling through a template based approach. The advantage of DESTINI over YASARA is that it includes an extra factor to the homology modeling, which is through contact predictions. DESTINI can later model each of the mutant structures to test their binding with KO, warfarin, and other potential drugs

6.8 hVKOR Predicted Binding Sites

Figure 6.8: Binding predictions of warfarin to hVKOR homology model. The highlighted green and labeled amino acids are the mutants studied here. Binding sites are ordered based on the FINDSITE^{comb} predicted ligand binding affinity. The rank is ordered from highest to lowest binding from the color of warfarin: pastel blue on site 1, purple on site 2, green on site 3, orange on site 4, and white on site 5.

There have been ongoing disagreements on how warfarin inhibits VKOR, where studies have shown that the inhibition is either competitive or non-competitive. Experimental results have determined that warfarin binds non-competitively to VKOR,[248, 268-271] but the crystal structure of the enzyme[247] shows only one site for occupying either KO or warfarin.[272] Not only that, but the two structures of *syn*VKOR and hVKOR homology model[3] are similar, where there is

a high chance of them binding to the same region. Oldenburg and coworkers[245] challenge the current literature that warfarin binding to VKOR is noncompetitive, even while the mechanism of action of VKAs and both K vitamers are unknown. They suggest that K, KO and warfarin bind to the same region in VKOR by *in silico* studies.

Our results through FINDSITE^{comb} found five cavities where potential ligands may bind. Figure 6.8 shows the hVKOR homology model docked with warfarin on these five sites using Autodock Vina. The highlighted green regions labeled amino acids are the mutations studied here. The binding sites are docked with warfarin and are labeled from most to least favorable are pastel blue on site 1, purple on site 2, green on site 3, orange on site 4, and white on site 5. The Autodock Vina docking scores are reported in Table 6.2. The results from the FINDSITE^{comb} and Autodock Vina do not match exactly (trend from most to least favorable site). FINDSITE^{comb} ranks their binding sites according to comparative homology receptor and potential ligand screening, which does not directly use warfarin to rank these sites, but multiple drugs including warfarin. Based on the *in vitro* studies analyzed previously, warfarin activity was drastically lowered when Tyr139Phe and Leu120Gln mutations have taken place. Binding site 1 involves these amino acids in its proximity, which may mean that binding site 1 is the most probable location for warfarin to bind to. ONIOM calculations were also done to analyze the stability of warfarin docking into these five sites, where the results confirm that FINDSITE^{comb} found reasonable “pockets” for warfarin to dock to.

Our predicted binding site 1 involves hydrogen bonding interactions of Ser117, Glu67, and Asn142 to warfarin’s ester carbonyl in the naphthalene ring, carbonyl in the naphthalene ring, and ketone on the side chain respectively. There are also noncovalent interactions involving

Leu120, Thr138, and Leu70 to the oxygen containing side of the naphthalene ring, the other side of the naphthalene ring, and to the phenyl ring respectively.

Site 1	Site 2	Site 3	Site 4	Site 5
-8.4	-3.0	-6.5	-5.4	-4.2

Table 6.2: Docking scores of every position warfarin is docked in kcal/mol.

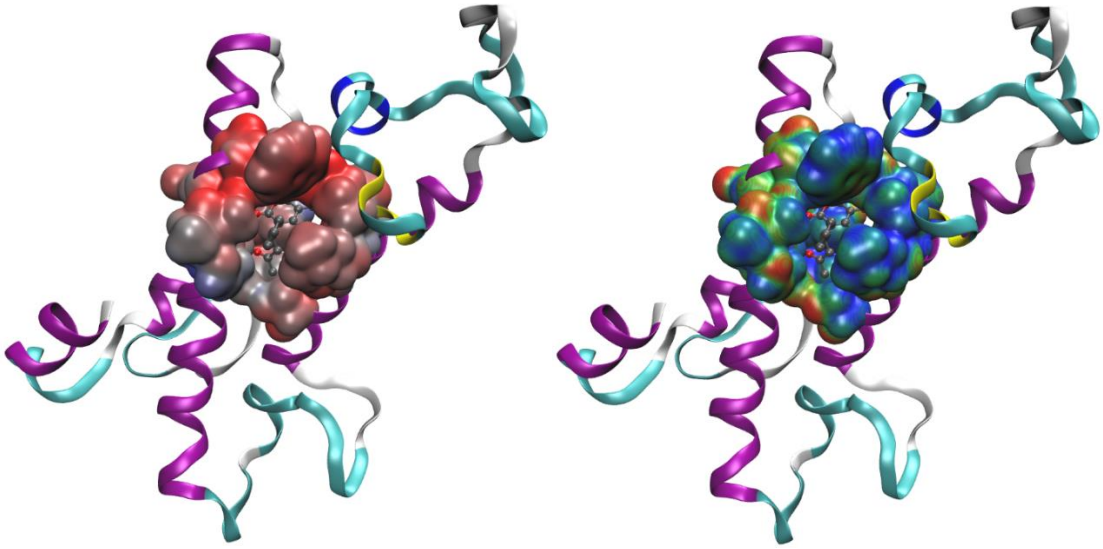


Figure 6.9: Warfarin docked onto binding site 1 of hVKOR homology model. (left) electrostatic potential from -0.15 au (red) to 0.10 au (blue); (right) overlap distance $D(\vec{r})$ from 2.8 bohr (red) to 3.4 bohr (blue). The warfarin molecule in the pocket (low-level layer) is shown using ball-and-stick model. This orientation of the complex is selected as it highlights the chemically most interesting regions.

The electrostatic potential and overlap distance are shown on Figure 6.9 for binding site 1 with warfarin bounded to that cavity. When the electrostatic potential map is analyzed, there was no surprise in the results since the regions that correspond with a positive electrostatic potential interact with the oxygen atoms on warfarin. Phe139 has a negative electrostatic charge on that site due to the amino acid having an alcohol group, but that interaction is through hydrogen bonding with the ester on the ring on warfarin.

The overlap distance provides an additional picture of how the ligand interacts with the protein. Figure 6.10 shows an in-focus view of the ester's carbonyl on warfarin's ring interacting with a smaller overlap distance on VKOR homology model. The other oxygen atoms on warfarin interact with a larger overlap distance, which is why we are not focusing on those sites. The smaller overlap distance for this site is due to serine being in this region, which hydrogen bonds to the ester on warfarin.

6.9 hVKOR Mutation Predicted Effects on Binding Site Mutations

The first type of mutation we examine is the amino acids located at the proposed binding site according to Oldenburg and coworkers.[3] For Tyr139Phe, an alcohol group is removed from the para position of a phenyl group. This removes any potential hydrogen bonding donation from this position. This mutation causes VKOR to become more active than the WT for both with and without warfarin present. The activity of the enzyme can be explained with two reasons: removing the alcohol group opens up space for the enzyme to bind more favorably to KO and it also reduces the binding interaction with warfarin since the drug may require hydrogen bonding

interactions.[273] Mutation of Tyr139Phe/Cys/Ser has been shown to produce a byproduct of 3-OH-vit K_1, where Tyr139 can be involved in the protonation of the hydroxyl group.[274]

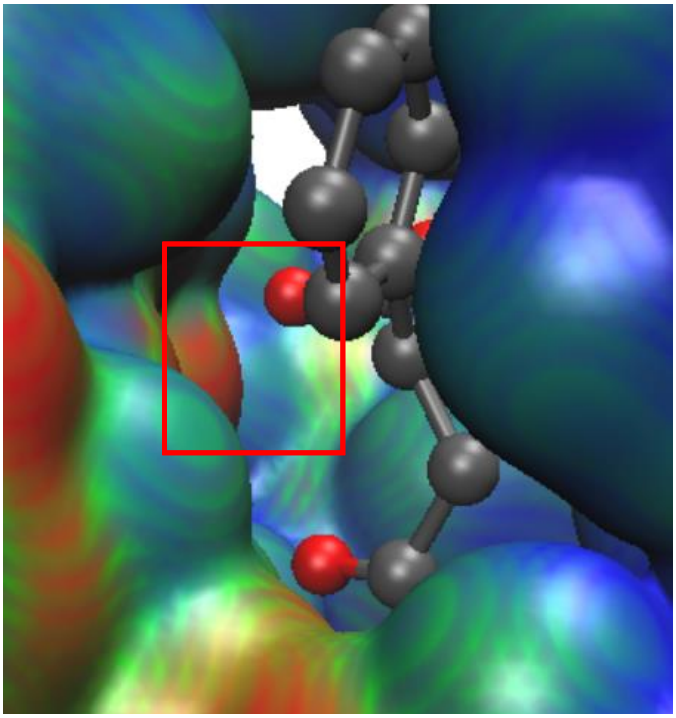


Figure 6.10: An in-focused and highlighted region of warfarin bound to site 1 on hVKOR model. The red box highlights the carbonyl on warfarin interacting with a region in the protein that has a smaller overlap distance $D(\vec{r})$. This specific site has a small orbital overlap, which is a serine that hydrogen bonds to warfarin.

The next type of mutation in the binding site is Leu120Gln. This mutation is where an isobutyl group is replaced by a propaneamide. The mutation causes a large decrease in relative activity when there is no warfarin,[251] but the change in relative activity is much smaller when we compare it to the WT with 100uM warfarin. This mutation introduces a larger chain and

hydrogen bonding into the enzyme. Hydrophobic interactions in this position may be required for appropriate VKOR activity and warfarin inhibition.

The next mutation considered is Leu128/Ser/Arg. Leu128Gln VKOR activity is reduced to 50% without the presence of warfarin, but its activity is still higher than Leu120Gln, meaning the binding site may not be as crowded as the substitution on 120. A higher activity can also mean that the introduction of hydrogen bonding may have increased the activity. Leu128Ser has an even lower relative VKOR activity compared to a Gln mutation. Replacing the isobutyl group with a methyl alcohol could have removed the hydrophobic interactions and caused a decrease in activity. It was also found that warfarin was less tightly bound and VKOR inhibition was reversible.[275] Leu128Arg had an even lower relative activity, where the addition of propyl guanidine could be too sterically hindered as compared to the other substitutions. This mutation caused the lowest VKOR activity without warfarin out of all the mutations studied. Patient studies have also been found to need to increase their VKA requirement to maintain normal clotting times.[254]

The last mutation in the binding site is Val29Leu. This mutation is the replacement of an isopropyl group with an isobutyl group. The activity is slightly lowered with this mutation, both with and without warfarin, which may have to do with steric hindrance, but not too significantly where it alters the activity.

6.10 hVKOR Mutation Predicted Effects on Conformational Change Mutations

The next few mutations we examine have been shown to disturb the ½-helix, where it acts as a lid for the active site.[247, 255] The following mutations would presumably cause the

calculated conformation to vary substantially compared to the WT. Arg58Gly is where a propyl guanidine is replaced by a hydrogen. The activities both with and without warfarin are decreased, where removing guanidine removed possible hydrogen bonding guanidine contributed to, which may cause the ½-helix to no longer fold over as a lid.[247, 255] Patients have been studied with this mutation where their required dosage of warfarin is 220-250mg/week versus a normal dosage of 10-60mg/week.[253-254]

Trp59Arg replaces a methyl indole with a propyl guanidium. Methyl indole is a nonpolar aromatic amino acid, while propyl guanidium is polar and has multiple locations for hydrogen bond donation. The relative activities are lowered substantially compared to the WT values, which may have to do with changing the polarity of the side chain.

The next few mutations studied are located away from the binding site. The first mutation is Arg35Pro, where a propyl guanidine is replaced with a pyrrolidine side chain. This side chain is nonpolar and much shorter than the guanidine side chain, which can be the reason for the decrease in activity since guanidine is a polar group. Patient studies have shown that this mutation causes warfarin to bind less tightly and causes the inhibition of VKOR to be reversible.[276-277]

Val45Ala is where an isopropyl side chain is replaced by a methyl group. The relative activity here surprisingly is much lower than the WT, considering that the only change is very slightly to the side chain. What may be happening is that the exact size of the isopropyl group is needed for a hydrophobic interaction. Patient studies have shown that an increase of warfarin is required to maintain normal clotting times.[254]

The last mutation studied is Arg33Pro. This mutation is the replacement of a propyl guanidium with a pyrrolidine. VKOR activity is cut by half when no warfarin is involved while introduction of warfarin gives the same relative activity as WT. This would mean that warfarin binding may not be altered, while KO binding is lowered.

6.11 Future Directions

The next step to this study is to generate structures of mutant hVKOR to compare to WT hVKOR model. We will then run FINDSITE^{comb} to scan cavities in the mutant hVKOR models and compare these cavities to our WT hVKOR model's binding site 1. The next step is to dock warfarin to these sites and quantify the computed binding affinities for each of the mutant hVKOR docked with warfarin. We will compare these quantities to the *in vitro* experimental results' relative activities on Table 6.1.

We can develop new VKAs based on the results acquired from this study to improve the efficacy of the drug on mutated hVKOR. This improvement will help those patients with warfarin resistance through hVKOR mutations by lowering the dosage of the new VKA needed to acquire the desired results.

The other drugs that can be developed through these studies are superwarfarins. Superwarfarins have been shown to be very effective as a rodenticide in small doses, with a high partition coefficient and a substantially low IC_{50}.[278] A drug that is considered a superwarfarin may not be an ideal medicine due to exceedingly long biological half-lives being greater than 20 days while warfarin has a half-life of 15-48 hours.[279] IC_{50} values of published warfarin

derivatives[280] can be compared to their relative binding with hVKOR. This study can potentially expand the knowledge of how changing the substituents of warfarin changes scoring values, leading to a structure that can be more favorable than warfarin as an inhibitor.

6.12 Computational Details

Electrostatic potentials and the orbital overlap distance were evaluated using the Gaussian suite of programs.[90] Electrostatic potential of warfarin-s was computed in vacuum.[70] warfarin and hVKOR homology model "surfaces" were taken to be the 0.001 e/bohr^3 density isosurface,[281] which encompasses approximately 96% of the electronic charge.[26] Electrostatic potentials and overlap distances of the hVKOR homology model is visualized using VMD version 1.9.3.[175]

The ground-state structure of deprotonated s-warfarin was determined using B3LYP/6-311+G(d,p)[282] method and basis set using Gaussian09 suite of programs (Figure 6.6).[90] hVKOR structural prediction was done using DESTINI,[259] the binding site pockets predictions were done using FINDSITE^{comb},[3] and warfarin docking orientations were predicted using Autodock Vina[262] (Vina) scoring functions. Autodock Vina uses a rigid receptor and flexible ligand docking algorithm to better predict favorable binding interactions between the rigid side chains and the flexible ligand.

ONIOM optimization calculations of warfarin binding stability to the five sites studied used HF/3-21G basis set for warfarin and UFF[283] on the protein. These calculations were used to confirm that the binding sites predicted by FINDSITE^{comb} can serve as docking sites for warfarin.

The calculations to generate the images on Figure 6.9 are as follows. We used the DESTINI homology structure and bound warfarin to binding site 1 (found by FINDSITE^{comb}) using Autodock Vina. We used the AmberTools[284] package to add hydrogen atoms at neutral pH. The ONIOM high level layer for the protein site was selected by first including all atoms 8 Angstrom from center of the ligand bound then by terminating at single bond links. The dangling bonds were terminated using H atoms. The calculations were carried out using B3LYP/6-31G(d,p) for High level layer and UFF[283] for low level layer.

Appendix A

Electronic Structure Theory

The goal for most quantum chemical calculations is to find the solution to the time-independent Schrodinger equation:

$$\widehat{H}\psi\left(\vec{x}_1, \vec{x}_2 \ldots \vec{x}_{N,} \vec{R}_1, \vec{R}_2 \ldots \vec{R}_M\right) = E_i\psi(\vec{x}_1, \vec{x}_2 \ldots \vec{x}_N, \vec{R}_1, \vec{R}_2 \ldots \vec{R}_M)$$

Equation A.1

$$\widehat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}$$

Equation A.2

Where $\widehat{H}$ is the Hamiltonian operator corresponding to the total energy in the system and ∇_A^2 is the Laplacian operator that is a sum of differential operators.

$$\nabla_A^2 = \frac{\partial^2}{\partial x_A^2} + \frac{\partial^2}{\partial y_A^2} + \frac{\partial^2}{\partial z_A^2}$$

$\vec{x}_i$ and $\vec{R}_i$ are the spatial and spin coordinates of the electron and spatial coordinates of the nuclei respectively. $r_{iA}, r_{ij}, and\ R_{AB}$ are the distances between each of the electron and nucleus, two electrons, and two nuclei. M_A is the mass of the nucleus and Z_M is the charge of the nucleus. $\widehat{H}$ is a differential operator for both the kinetic and potential energy between the electrons and nuclei in the molecule. The operator acts on the wave function $\psi_i(\vec{r}_1, \vec{r}_2 \ldots \vec{r}_N)$, which in turn solves the total energy and recovers the wave function. The wave function contains all the information of the quantum system studied.

In order to save on computational cost, the Hamiltonian operator can be simplified using the Born-Oppenheimer approximation, which reduces the Hamiltonian to the electronic term.

$$\widehat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} = \hat{T} + \hat{v}_{Ne} + \hat{v}_{ee}$$

Equation A.3

In this approximation, the nuclei move much slower in comparison to the electrons. In other words, the nuclei are considered fixed in comparison to the movement of electrons, thus giving a total kinetic energy of zero and potential energy to be constant for nucleus-nucleus repulsion.

Many Electron Wave Function

$$\psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N) = \chi_1(\vec{x}_1)\chi_2(\vec{x}_2) \dots \chi_N(\vec{x}_N)$$

Equation A.4

Electrons in the wave function are indistinguishable, where exchanging the coordinates of any two or more electrons does not change the observable.

$$\psi(\vec{x}_2, \vec{x}_1 \dots \vec{x}_N) = P\chi_1(\vec{x}_2)\chi_2(\vec{x}_1) \dots \chi_N(\vec{x}_N)$$

Equation A.5

P is a permutation in the electron coordinates. There are only two types of particles in nature that can describe the wave function as being symmetric (Bosons) or antisymmetric (Fermions). Electrons are Fermions with spin=1/2, so the exchange of the two electrons will cause a sign change in the wave function. This is based on the Pauli exclusion principle, where the set of two electrons may not have the same quantum numbers. This gives rise to the Slater Determinant:

$$\psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N) = \phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & \dots & \chi_N(\vec{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & \dots & \chi_N(\vec{x}_N) \end{vmatrix}$$

Equation A.6

The slater equation preserves the antisymmetry of electrons, where the exchange of the spatial coordinates of two electrons will cause a permutation.

For example, a system of two electrons in two orbitals

$$\psi(\vec{x}_1, \vec{x}_2) = \phi_{SD} = \frac{1}{\sqrt{2!}} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) \end{vmatrix} = \frac{1}{2}(\chi_1(\vec{x}_1)\chi_2(\vec{x}_2) - \chi_1(\vec{x}_2)\chi_2(\vec{x}_1))$$

Equation A.7

Hartree-Fock Approximation

The HF approximation uses the variational principle, which systematically approaches the wave function of the ground state. It does this by minimizing E_0 with respect of the wave function, resulting in E_{HF}.

$$E_{HF} = \min_{\phi_{SD} \to N} E[\phi_{SD}]$$

Equation A.8

The Fock operator is a one-electron operator, where the first two terms are the kinetic and potential energy. $V_{HF}(i)$ is the Hartree-Fock potential, which is the average repulsion felt by the i'th electron from the other N-1 electrons.

$$\hat{f}_i = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{r_{iA}} + v_{HF}(i)$$

Equation A.9

$$E_{HF} = \langle \phi_{SD} | \hat{H} | \phi_{SD} \rangle = \sum_i^N (i|\hat{h}|i) + \frac{1}{2}\sum_i^N \sum_j^N (ii|jj) - (ij|ji)$$

$$(i|\hat{h}|i) = \int \chi_i^*(\vec{x}_1) \left\{ -\frac{1}{2}\nabla^2 - \sum_A^M \frac{Z_A}{r_{1A}} \right\} \chi_i(\vec{x}_1) d\vec{x}_1$$

$$(ii|jj) = \int\int |\chi_i(\vec{x}_1)|^2 \frac{1}{r_{12}} |\chi_j(\vec{x}_2)|^2 d\vec{x}_1 d\vec{x}_2$$

$$(ij|ji) = \int\int \chi_i(\vec{x}_1)\chi_j^*(\vec{x}_1) \frac{1}{r_{12}} \chi_j(\vec{x}_2)\chi_i^*(\vec{x}_2) d\vec{x}_1 d\vec{x}_2$$

Equations A.10

The operator $\left(i\middle|\hat{h}\middle|i\right)$ defines the contribution of the kinetic and electron-nucleus attraction, $(ii|jj)$ is the Coulomb integral, and $(ij|ji)$ is the exchange integral. The Coulomb integral provides the average Coulomb repulsion felt by the first electron from a second electron in the second integral. The exchange integral is an artifact of the antisymmetry of the Slater determinant and applies to all fermions.

Restricted and Unrestricted Hartree-Fock

RHF method treats closed shell systems very well, where two spin orbitals will have the same orbital energy. For open-shelled systems such as radicals, excited state atoms/molecules, and the hydrogen atom, RHF needs to be generalized to the restricted open-shell case, where the highest energy spatial orbitals are singly occupied and the others are doubly occupied. Therefore, ROHF and UHF would apply here. ROHF follows the RHF method in treating the paired electrons as being in the same orbital, but the unpaired electrons are in a different orbital. UHF puts all of the electrons in a different orbitals, creating different orbital energies for each electron. For example, if a lithium atom is evaluated using UHF and ROHF, ROHF will have two degenerate 1s orbitals and one 2s orbital while UHF will generate a different orbital for each of the three electrons.

One of the downfalls to HF theory is that electron correlation effects are approximated in the potential energy v_{HF}. It takes the average repulsive potential experienced by one electron due to the other electrons in the system, taking the electron-electron repulsion in an average way.

Møller-Plesset

MP is a post-Hartree-Fock method that improves the electron correlation by using perturbation theory.

$$\hat{H} = \hat{H}_0 + \lambda V$$

Equation A.11

Where $\hat{H}_0$ is the HF Hamiltonian operator and λ is an ordering parameter multiplied with the pertubation V. By expanding the exact eigenfunction and eigenvalues in a Taylor series in λ, giving the following relations.

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \cdots$$

$$\psi = \psi^{(0)} + \lambda \psi^{(1)} + \lambda^2 \psi^{(2)} + \cdots$$

Equation A.12

By substituting Eq. A.12 into Eq. A11 and equating coefficients of λ^n

$$\widehat{H}_0|\psi^{(0)}\rangle = E^{(0)}|\psi^{(0)}\rangle, n = 0$$

$$\widehat{H}_0|\psi^{(1)} + V|\psi^{(0)}\rangle = E^{(0)}|\psi^{(1)}\rangle + E^{(1)}|\psi^{(0)}\rangle, n = 1$$

$$\widehat{H}_0|\psi^{(2)} + V|\psi^{(1)}\rangle = E^{(0)}|\psi^{(2)}\rangle + E^{(1)}|\psi^{(1)}\rangle + E^{(2)}|\psi^{(0)}\rangle, n = 2$$

$$\widehat{H}_0|\psi^{(3)} + V|\psi^{(2)}\rangle = E^{(0)}|\psi^{(3)}\rangle + E^{(1)}|\psi^{(2)}\rangle + E^{(2)}|\psi^{(1)}\rangle + E^{(3)}|\psi^{(0)}\rangle, n = 3$$

$$\vdots$$

Equation A.13

Then multiplying with $\psi^{(0)}$ and applying the orthogonality relation

$$E^{(0)} = \langle\psi^{(0)}|\widehat{H}_0|\psi^{(0)}\rangle$$

$$E^{(1)} = \langle\psi^{(0)}|V|\psi^{(0)}\rangle$$

$$E^{(2)} = \langle\psi^{(0)}|V|\psi^{(1)}\rangle$$

$$E^{(3)} = \langle\psi^{(0)}|V|\psi^{(2)}\rangle$$

$$\vdots$$

Equation A.14

The MP energy is corrected through the perturbation theory, resulting in the following

$$\hat{H} = \hat{H}_0 + V$$

$$V = \sum_{i<j} r_{ij}^{-1} - \sum_i v_{HF}$$

Equation A.15

$\hat{H}_0$ is the HF Hamiltonian operator and v_{HF} is the HF potential energy. When applying the first order energy from the perturbation equations, we get the corrected (MP_1) energy.

$$E = E^{(0)} + E^{(1)} = E^{(0)} + \langle \psi^{(0)} | V | \psi^{(0)} \rangle = E^{(0)} - \frac{1}{2} \sum_{ab} \langle ab || ab \rangle$$

Equation A.16

where

$$\langle ab||ab\rangle = \langle ab|ab\rangle - \langle ab|ba\rangle$$

$$= \int\int \chi_a^*(\vec{x}_1)\chi_b^*(\vec{x}_2)\frac{1}{r_{12}}[\chi_a^*(\vec{x}_1)\chi_b^*(\vec{x}_2) - \chi_b^*(\vec{x}_1)\chi_a^*(\vec{x}_2)]d\vec{x}_1 d\vec{x}_2$$

Following the same procedures, the second order corrected (MP_2) energy is

$$E = E^{(0)} + E^{(1)} + E^{(2)} = E^{(0)} + \langle \psi^{(0)} | V | \psi^{(0)} \rangle + \langle \psi^{(0)} | V | \psi^{(1)} \rangle$$

$$\langle \psi^{(0)} | V | \psi^{(1)} \rangle = \frac{1}{4} \sum_{abrs} \frac{\langle ab||rs \rangle^2}{\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s}$$

Equation A.17

Where ε_a and ε_b are the occupied spin orbital energy while ε_r and ε_s are the unoccupied spin orbitals.

Density Functional Theory (DFT)

The wave function is not an observable, meaning it can not be measured experimentally. Squaring the wave function gives the probability that the electrons in the system are found simultaneously in the desired volume elements.

$$|\psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N)|^2 d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N$$

Equation A.18

we get the electron density

$$\rho(r_1) = N \int \dots \int |\psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N)|^2 d\vec{s}_1 d\vec{x}_2 \dots d\vec{x}_N$$

Equation A.19

Where integrating the electron density gives the number of electrons in the system

$$\int \rho(\vec{r}_1)\, d\vec{r}_1 = N$$

Equation A.20

The electron density gives the probability of finding any of the electrons within their designated volume element with arbitrary spin. The x coordinates provides the spatial coordinates r_i and the spin coordinate s_i. Unlike the wave function, the electron density is an observable, where it can be measured experimentally such as in X-ray diffraction.

All modern day DFT is based on the two theorems from Hohenberg-Kohn. The first theorem states that "the external potential $V_{ext}(\vec{r})$ is (to within a constant) a unique function of $\rho(\vec{r})$; since, in turn $V_{ext}(\vec{r})$ fixes $\widehat{H}$ we see that the full many particle ground state is a unique functional of $\rho(\vec{r})$." This results to the following equation where the ground state energy is a functional of the ground state electron density.

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + F_{HK}[\rho_0]$$

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] = \langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle$$

Equation A.21

$E_{Ne}[\rho_0]$ is the external potential from the nuclei and electron, $T[\rho]$ is the functional for kinetic energy, and $E_{ee}[\rho]$ is the functional for electron-electron interaction.

The second theorem states that the functional $F_{HK}[\rho]$ delivers the ground state energy of the system if the density is the true ground state density ρ_0. This theorem uses the variational principle with trial density to define its own Hamiltonian, then wave function.

$$\langle\tilde{\psi}|\hat{H}|\tilde{\psi}\rangle = T[\tilde{\rho}] + V_{ee}[\tilde{\rho}] + \int \tilde{\rho}(\vec{r})V_{ext}d\vec{r} + F_{HK} = E[\tilde{\rho}] \geq E_0[\rho_0] = \langle\psi|\hat{H}|\psi\rangle$$

Equation A.22

Local Spin Density Approximation

A class of approximations using DFT method is the LDA, which uses a UEG to model the density at each point in space. The use of UEG model for density is useful for bulk transition metals and simple metals such a sodium, but this method can not model the real system of typical atoms and molecules. The benefit in this approach is that the exchange and correlation energies are known.

$$E_{XC}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{XC}\big(\rho(\vec{r})\big)d\vec{r}$$

$$E_{X}^{LDA}[\rho(\vec{r})] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \rho(\vec{r})^{4/3}d\vec{r}$$

Equation A.23

$\varepsilon_{XC}\big(\rho(\vec{r})\big)$ is the exchange-correlation energy per particle in the UEG of density $\rho(\vec{r})$. It is weighed with the probability, $\rho(\vec{r})$, of an electron being in that space. E_X^{LDA} is the exchange-energy density of the UEG.

For the unrestricted case of open-shelled systems, the electron density is replaced with spin densities. The following expression is for the LSDA:

$$E_{XC}^{LSDA}[\rho_\alpha, \rho_\beta] = \int \rho(\vec{r})\varepsilon_{XC}\left(\rho_\alpha(\vec{r}), \rho_\beta(\vec{r})\right)d\vec{r}$$

Equation A.24

Generalized Gradient Approximation

LDA is further improved by considering the gradient of the electron density, $\nabla\rho$, through GGA. Simple GGA functional approximations have proven to give reasonable accuracy for electron correlation.

$$E_{XC}^{GGA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{XC}^{GGA}\big(\rho(\vec{r}), \nabla\rho(\vec{r})\big)d\vec{r}$$

Equation A.25

PBE improved the GGA upon the LSD description of molecules.

$$E_{XC}^{PBE}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_X\big(\rho(\vec{r})\big)F_{XC}(r_s, \zeta, s)d\vec{r}$$

$$F_X(s) = 1 + \kappa - \frac{\kappa}{1 + \frac{\mu s^2}{\kappa}}$$

$$s = \frac{|\nabla\rho(\vec{r})|}{2k_F\rho(\vec{r})}$$

Equation A.26

Where μ=0.235 and κ=0.804, which are exact constraints and k_F is a Fermi wave vector equaling to $(3\pi^2\rho(\vec{r}))^{1/3}$.

Hybrid Functional

Hybrid functional incorporate the exact exchange from HF theory. It uses Kohn-Sham orbitals instead of the gradient density. A popular method is the B3LYP hybrid functional, which uses the three parameter exchange correlation functional mixing the exact exchange of HF

exchange correlation and Lee, Yang, and Parr correlation functional that recovers the dynamic electron correlation.

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSDA} + aE_{XC}^{\lambda=0} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LSDA}$$

Equation A.27

Where a=0.20, b=0.72, and c=0.81, $\lambda = 0$ is an interaction free system.

Basis Sets

Basis sets are used in quantum chemistry to describe molecular orbitals, where they are composed of finite set of functions. The higher the number of basis functions used on the system gives a more accurate result that describes the molecular orbitals. The downfall to increasing the number of basis functions is that the computational cost becomes longer. Basis functions are typically centered on atoms, but they have been placed in bonds and lone pairs. The first basis functions used were STOs, since they resembled the eigenfunctions in the hydrogen atom.

$$S_{nlm}^{\zeta}(r,\vartheta,\varphi) = Nr^{n-1}e^{-\zeta r}Y_l^m(\vartheta,\varphi)$$

Equation A.28

Where N is the normalization factor and $e^{-\zeta r}$ is the decay of the wave function. $Y_l^m(\vartheta, \varphi)$ are the spherical harmonics of the system with n, l, and m being the principal, orbital angular momentum, and magnetic quantum numbers respectively.

STOs are approximated as linear combination of Gaussian type orbitals (GTOs):

$$G_{ijk}^{\alpha,R}(r) = N_{ijk}^{\alpha}(x - R_1)^i(y - R_2)^j(z - R_3)^k e^{-\alpha(r-R)^2}$$

Equation A.29

With N_{ijk}^{α} being the normalization factor, R is the center, α is the exponent, and ijk are the cartesian coordinates. The summation of the exponents, $i + j + k = L$, give the orbital type being s (L=0), p (L=1), or d (L=2). GTOs are not orbitals, they are simpler functions which are often called gaussian primitives. It is obtained through quantum calculations and the exponents are typically varied until the lowest total energy of the atom is found.

Basis Set Superposition Error

BSSE is present in all molecular calculation that use finite basis sets, but the effects are much larger in weakly bound complexes. BSSE becomes larger as atoms from two molecules approach one another, their basis functions overlap, where the monomer "borrows" functions

from their neighboring molecule. The corrected relative interaction energy with counterpoise correction then becomes:

$$E_{AB}^{cc} = E_{AB}(G, AB) + V_{AB}^{cc} + V_{AB}(G)$$

$$V_{AB}^{cc}(G) = E_{AB}(G, AB) - E_A(G, AB) - E_B(G, AB)$$

$$V_{AB}(G) = E_{AB}(G, AB) - E_A(A) - E_B(B)$$

Equation A.30

Where $V_{AB}^{cc}(G)$ and $V_{AB}(G)$ are the interaction energies with and without counterpoise correction respectively. $E_{A/B}(G, AB)$ are the total energies of the monomers A/B computed with the dimer basis set at geometry G and the geometry of the dimer. This means that when calculating one of the monomers, the basis set of the other monomer is present, excluding the nuclei of the other monomer. $E_{A/B}(A/B)$ is the total energy of the monomer A/B with their monomer basis set.

References

1. Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P., Halogen bonding: the σ-hole. *Journal of molecular modeling* **2007,** *13* (2), 291-296.
2. Politzer, P.; Murray, J. S.; Clark, T., Halogen bonding and other σ-hole interactions: a perspective. *Physical Chemistry Chemical Physics* **2013,** *15* (27), 11178-11189.
3. Hirshfeld, F. L., Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta* **1977,** *44* (2), 129-138.
4. Owens, P. K.; Raddad, E.; Miller, J. W.; Stille, J. R.; Olovich, K. G.; Smith, N. V.; Jones, R. S.; Scherer, J. C., A decade of innovation in pharmaceutical R&D: the Chorus model. *Nature Reviews Drug Discovery* **2015,** *14* (1), 17.
5. Yang, L.; Wang, W.; Sun, Q.; Xu, F.; Niu, Y.; Wang, C.; Liang, L.; Xu, P., Development of novel proteasome inhibitors based on phthalazinone scaffold. *Bioorganic & medicinal chemistry letters* **2016,** *26* (12), 2801-2805.
6. Talele, T. T.; Khedkar, S. A.; Rigby, A. C., Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Current topics in medicinal chemistry* **2010,** *10* (1), 127-141.
7. Clark, D. E., What has computer-aided molecular design ever done for drug discovery? *Expert opinion on drug discovery* **2006,** *1* (2), 103-110.
8. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery* **2004,** *3* (11), 935.
9. Hirayama, T.; Okaniwa, M.; Imada, T.; Ohashi, A.; Ohori, M.; Iwai, K.; Mori, K.; Kawamoto, T.; Yokota, A.; Tanaka, T., Synthetic studies of centromere-associated protein-E (CENP-E) inhibitors: 1. Exploration of fused bicyclic core scaffolds using electrostatic potential map. *Bioorganic & medicinal chemistry* **2013,** *21* (17), 5488-5502.
10. Hirayama, T.; Okaniwa, M.; Banno, H.; Kakei, H.; Ohashi, A.; Iwai, K.; Ohori, M.; Mori, K.; Gotou, M.; Kawamoto, T., Synthetic Studies on Centromere-Associated Protein-E (CENP-E) Inhibitors: 2. Application of Electrostatic Potential Map (EPM) and Structure-Based Modeling to Imidazo [1, 2-a] pyridine Derivatives as Anti-Tumor Agents. *Journal of medicinal chemistry* **2015,** *58* (20), 8036-8053.
11. Mehmood, A.; Janesko, B. G., An Orbital-Overlap Complement to Atomic Partial Charge. *Angew. Chem. Int. Ed.* **2017,** *56* (24), 6878-6881.
12. Pearson, R. G., Hard and Soft Acids and Bases. *J. Am. Chem. Soc.* **1963,** *85* (22), 3533-3539.
13. Pearson, R. G., Acids and Bases. *Science* **1966,** *151* (3707), 172-177.
14. Ayers, P. W.; Parr, R. G.; Nagy, A., Local kinetic energy and local temperature in the density-functional theory of electronic structure. *Int. J. Quantum Chem* **2002,** *90* (1), 309-326.
15. Calaminici, P.; Jug, K.; Köster, A. M., Density functional calculations of molecular polarizabilities and hyperpolarizabilities. *J. Chem. Phys.* **1998,** *109* (18), 7756-7763.
16. Fukui, K., Role of Frontier Orbitals in Chemical Reactions. *Science* **1982,** *218* (4574), 747-754.

17. Beker, W.; Szarek, P.; Komorowski, L.; Lipiński, J., Reactivity Patterns of Imidazole, Oxazole, and Thiazole As Reflected by the Polarization Justified Fukui Functions. *J. Phys. Chem. A* **2013,** *117* (7), 1596-1600.
18. Bader, R. F. W., *Atoms in Molecules: A Quantum Theory*. Clarendon Press: 1994.
19. Liu, S., Steric effect: A quantitative description from density functional theory. *J. Chem. Phys.* **2007,** *126* (24), 244103.
20. Silvi, B.; Savin, A., Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994,** *371*, 683.
21. Becke, A. D.; Edgecombe, K. E., A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990,** *92* (9), 5397-5403.
22. Janesko, B. G.; Wiberg, K. B.; Scalmani, G.; Frisch, M. J., Electron delocalization range in atoms and on molecular surfaces. *Journal of chemical theory and computation* **2016,** *12* (7), 3185-3194.
23. Janesko, B. G.; Scalmani, G.; Frisch, M. J., How far do electrons delocalize? *The Journal of chemical physics* **2014,** *141* (14), 144104.
24. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *The Journal of Chemical Physics* **1985,** *83* (2), 735-746.
25. Bader, R.; Henneker, W. H.; Cade, P. E., Molecular charge distributions and chemical binding. *The Journal of Chemical Physics* **1967,** *46* (9), 3341-3363.
26. Bader, R. F.; Carroll, M. T.; Cheeseman, J. R.; Chang, C., Properties of atoms in molecules: atomic volumes. *Journal of the American Chemical Society* **1987,** *109* (26), 7968-7979.
27. Kolar, M. H.; Hobza, P., Computer modeling of halogen bonds and other σ-hole interactions. *Chemical reviews* **2016,** *116* (9), 5155-5187.
28. Mehmood, A.; Janesko, B. G., The electron delocalization range in stretched bonds. *International Journal of Quantum Chemistry* **2016,** *116* (23), 1783-1795.
29. Janesko, B. G.; Scalmani, G.; Frisch, M. J., Quantifying solvated electrons' delocalization. *Physical Chemistry Chemical Physics* **2015,** *17* (28), 18305-18317.
30. Deng, Z.; Martyna, G. J.; Klein, M. L., Electronic states in metal-ammonia solutions. *Physical review letters* **1993,** *71* (2), 267.
31. Mehmood, A.; Janesko, B. G., An Orbital-Overlap Complement to Atomic Partial Charge. *Angewandte Chemie International Edition* **2017,** *56* (24), 6878-6881.
32. Itoh, N., Creation of lattice defects by electronic excitation in alkali halides. *Advances in Physics* **1982,** *31* (5), 491-551.
33. Itoh, N.; Tanimura, K., Formation of interstitial-vacancy pairs by electronic excitation in pure ionic crystals. *Journal of Physics and Chemistry of Solids* **1990,** *51* (7), 717-735.
34. Williams, R.; Song, K., The self-trapped exciton. *Journal of Physics and Chemistry of Solids* **1990,** *51* (7), 679-716.
35. Song, K.; Williams, R. T., Alkali halides. In *Self-Trapped Excitons*, Springer: 1993; pp 123-219.
36. Lord, N. W., Electron-nuclear double resonance of F centers in lithium fluoride. *Physical Review Letters* **1958,** *1* (5), 170.
37. Seitz, F., Color centers in alkali halide crystals. II. *Reviews of modern physics* **1954,** *26* (1), 7.

38. Seitz, F., Color centers in alkali halide crystals. *Reviews of modern physics* **1946,** *18* (3), 384.
39. Doyle, W. T., Optical absorption by F centers in alkali halides. *Physical Review* **1958,** *111* (4), 1072.
40. Holton, W. C.; Blum, H., Paramagnetic resonance of F centers in alkali halides. *Physical Review* **1962,** *125* (1), 89.
41. Stoneham, A. M., *Theory of defects in solids: electronic structure of defects in insulators and semiconductors*. Oxford University Press: 2001.
42. Yu, S.-J.; Kang, M.-W.; Chang, H.-C.; Chen, K.-M.; Yu, Y.-C., Bright fluorescent nanodiamonds: no photobleaching and low cytotoxicity. *Journal of the American Chemical Society* **2005,** *127* (50), 17604-17605.
43. Falk, A. L.; Klimov, P. V.; Buckley, B. B.; Ivády, V.; Abrikosov, I. A.; Calusine, G.; Koehl, W. F.; Gali, Á.; Awschalom, D. D., Electrically and mechanically tunable electron spins in silicon carbide color centers. *Physical review letters* **2014,** *112* (18), 187601.
44. Jacques, V.; Wu, E.; Grosshans, F.; Treussart, F.; Grangier, P.; Aspect, A.; Roch, J.-F., Experimental realization of Wheeler's delayed-choice gedanken experiment. *Science* **2007,** *315* (5814), 966-968.
45. Koehl, W. F.; Buckley, B. B.; Heremans, F. J.; Calusine, G.; Awschalom, D. D., Room temperature coherent control of defect spin qubits in silicon carbide. *Nature* **2011,** *479* (7371), 84.
46. Bassett, L. C.; Heremans, F. J.; Christle, D. J.; Yale, C. G.; Burkard, G.; Buckley, B. B.; Awschalom, D. D., Ultrafast optical control of orbital and spin dynamics in a solid-state defect. *Science* **2014,** *345* (6202), 1333-1337.
47. Yoon, B.; Häkkinen, H.; Landman, U.; Wörz, A. S.; Antonietti, J.-M.; Abbet, S.; Judai, K.; Heiz, U., Charging effects on bonding and catalyzed oxidation of CO on Au8 clusters on MgO. *Science* **2005,** *307* (5708), 403-407.
48. Pacchioni, G.; Sicolo, S.; Valentin, C. D.; Chiesa, M.; Giamello, E., A route toward the generation of thermally stable Au cluster anions supported on the MgO surface. *Journal of the American Chemical Society* **2008,** *130* (27), 8690-8695.
49. Eid, K. M.; Ammar, H., Adsorption of SO2 on Li atoms deposited on MgO (1 0 0) surface: DFT calculations. *Applied Surface Science* **2011,** *257* (14), 6049-6058.
50. Eid, K. M.; Ammar, H., A density functional study of NO2 adsorption on perfect and defective MgO (1 0 0) and Li/MgO (1 0 0) surfaces. *Applied Surface Science* **2012,** *258* (19), 7689-7698.
51. Ammar, H.; Eid, K. M., NO2 Interaction with Au Atom Adsorbed on Perfect and Defective MgO (100) Surfaces: Density Functional Theory Calculations. *Journal of nanoscience and nanotechnology* **2013,** *13* (10), 6660-6671.
52. Jensen, M. B.; Pettersson, L. G.; Swang, O.; Olsbye, U., CO2 sorption on MgO and CaO surfaces: A comparative quantum chemical cluster study. *The Journal of Physical Chemistry B* **2005,** *109* (35), 16774-16781.
53. Rinke, P.; Schleife, A.; Kioupakis, E.; Janotti, A.; Rödl, C.; Bechstedt, F.; Scheffler, M.; Van de Walle, C. G., First-principles optical spectra for F centers in MgO. *Physical review letters* **2012,** *108* (12), 126404.

54. Ferrari, A. M.; Pacchioni, G., Electronic structure of F and V centers on the MgO surface. *The Journal of Physical Chemistry* **1995,** *99* (46), 17010-17018.
55. Sharma, R.; Stoneham, A., Theory of the F+ s surface centre in MgO. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1976,** *72*, 913-919.
56. Eglitis, R., Ab initio calculations of SrTiO 3, BaTiO 3, PbTiO 3, CaTiO 3, SrZrO 3, PbZrO 3 and BaZrO 3 (001),(011) and (111) surfaces as well as F centers, polarons, KTN solid solutions and Nb impurities therein. *International journal of modern physics B* **2014,** *28* (17), 1430009.
57. Ricci, D.; Di Valentin, C.; Pacchioni, G.; Sushko, P. V.; Shluger, A. L.; Giamello, E., Paramagnetic defect centers at the MgO surface. An alternative model to oxygen vacancies. *Journal of the American Chemical Society* **2003,** *125* (3), 738-747.
58. Mallia, G.; Orlando, R.; Roetti, C.; Ugliengo, P.; Dovesi, R., F center in LiF: A quantum mechanical ab initio investigation of the hyperfine interaction between the unpaired electron at the vacancy and its first seven neighbors. *Physical Review B* **2001,** *63* (23), 235102.
59. Ewig, C. S.; Tellinghuisen, J.; Mendenhall, M. H., Ab initio F-center electron topology in LiF. *Chemical physics letters* **1992,** *188* (5-6), 501-509.
60. Bader, R.; Platts, J., Characterization of an F-center in an alkali halide cluster. *The Journal of chemical physics* **1997,** *107* (20), 8545-8553.
61. Carrasco, J.; Lopez, N.; Illas, F., On the convergence of isolated neutral oxygen vacancy and divacancy properties in metal oxides using supercell models. *The Journal of chemical physics* **2005,** *122* (22), 224705.
62. Karsai, F.; Tiwald, P.; Laskowski, R.; Tran, F.; Koller, D.; Gräfe, S.; Burgdörfer, J.; Wirtz, L.; Blaha, P., F center in lithium fluoride revisited: Comparison of solid-state physics and quantum-chemistry approaches. *Physical Review B* **2014,** *89* (12), 125429.
63. Chen, G.; Liu, Z.; Gong, X., Ab initio study on structural and electronic properties of Ba n O m clusters. *The Journal of chemical physics* **2004,** *120* (17), 8020-8024.
64. Williams, C. F.; Herbert, J. M., Influence of Structure on Electron Correlation Effects and Electron− Water Dispersion Interactions in Anionic Water Clusters. *The Journal of Physical Chemistry A* **2008,** *112* (27), 6171-6178.
65. Zurek, E.; Edwards, P. P.; Hoffmann, R., A molecular perspective on lithium–ammonia solutions. *Angewandte Chemie International Edition* **2009,** *48* (44), 8198-8232.
66. Gräfenstein, J.; Cremer, D., An efficient algorithm for the density-functional theory treatment of dispersion interactions. *The Journal of chemical physics* **2009,** *130* (12), 124105.
67. Bader, R., Atoms in Molecules-A Quantum Theory, Oxford University Press, Oxford, 1990; R. **1998**.
68. Madsen, G.; Gatti, C.; Iversen, B.; Damjanovic, L.; Stucky, G.; Srdanov, V., F center in sodium electrosodalite as a physical manifestation of a non-nuclear attractor in the electron density. *Physical Review B* **1999,** *59* (19), 12359.
69. Mori-Sanchez, P.; Recio, J.; Silvi, B.; Sousa, C.; Pendás, A. M.; Luaña, V.; Illas, F., Rigorous characterization of oxygen vacancies in ionic oxides. *Physical Review B* **2002,** *66* (7), 075103.

70. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** *268* (5214), 1144-1149.
71. Zhu, X.; Yu, W.; McBride, R.; Li, Y.; Chen, L.-M.; Donis, R. O.; Tong, S.; Paulson, J. C.; Wilson, I. A., Hemagglutinin homologue from H17N10 bat influenza virus exhibits divergent receptor-binding and pH-dependent fusion activities. *Proceedings of the National Academy of Sciences* **2013,** *110* (4), 1458-1463.
72. Tang, W. W.; Wang, Z.; Levison, B. S.; Koeth, R. A.; Britt, E. B.; Fu, X.; Wu, Y.; Hazen, S. L., Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *New England Journal of Medicine* **2013,** *368* (17), 1575-1584.
73. Mohan, S. V.; Karthikeyan, J., Removal of lignin and tannin colour from aqueous solution by adsorption onto activated charcoal. *Environmental Pollution* **1997,** *97* (1-2), 183-187.
74. Zhang, Q.; Chuang, K. T., Adsorption of organic pollutants from effluents of a Kraft pulp mill on activated carbon and polymer resin. *Advances in Environmental Research* **2001,** *5* (3), 251-258.
75. Behling, R.; Valange, S.; Chatel, G., Heterogeneous catalytic oxidation for lignin valorization into valuable chemicals: what results? What limitations? What trends? *Green chemistry* **2016,** *18* (7), 1839-1854.
76. Duval, A.; Lawoko, M., A review on lignin-based polymeric, micro-and nano-structured materials. *Reactive and Functional Polymers* **2014,** *85*, 78-96.
77. Markham, J. J., *F-centers in Alkali Halides*. Academic Pr: 1966; Vol. 8.
78. Downing, C.; Sokol, A.; Catlow, C., The reactivity of CO 2 on the MgO (100) surface. *Physical Chemistry Chemical Physics* **2014,** *16* (1), 184-195.
79. Michael Condren Christian Brothers University, F-Centers. https://archive.education.mrsec.wisc.edu/background/F_center/index.htm.
80. Wiberg, K. B.; Hadad, C. M.; LePage, T. J.; Breneman, C. M.; Frisch, M. J., Analysis of the effect of electron correlation on charge density distributions. *The Journal of Physical Chemistry* **1992,** *96* (2), 671-679.
81. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** *136* (3B), B864-B871.
82. Cremer, D., Density functional theory: coverage of dynamic and non-dynamic electron correlation effects. *Molecular Physics* **2001,** *99* (23), 1899-1940.
83. Fowler, W. B., Physics of color centers. **1968**.
84. Sommerfeld, T.; Bhattarai, B.; Vysotskiy, V. P.; Cederbaum, L. S., Correlation-bound anions of NaCl clusters. *The Journal of chemical physics* **2010,** *133* (11), 114301.
85. Henrich, V. E.; Cox, P. A., *The surface science of metal oxides*. Cambridge university press: 1996.
86. Ricci, D.; Pacchioni, G.; Sushko, P. V.; Shluger, A. L., Reactivity of (H+)(e−) color centers at the MgO surface: formation of O2− and N2− radical anions. *Surface science* **2003,** *542* (3), 293-306.
87. Sun, Z.; Wang, J.; Du, W.; Lu, G.; Li, P.; Song, X.; Yu, J., Density functional theory study on the thermodynamics and mechanism of carbon dioxide capture by CaO and CaO regeneration. *RSC Advances* **2016,** *6* (45), 39460-39468.

88. Prada, S.; Giordano, L.; Pacchioni, G., Charging of gold atoms on doped MgO and CaO: identifying the key parameters by DFT calculations. *The Journal of Physical Chemistry C* **2013,** *117* (19), 9943-9951.
89. Sniatynsky, R.; Janesko, B. G.; El-Mellouhi, F.; Brothers, E. N., Application of screened hybrid density functional theory to ammonia decomposition on silicon. *The Journal of Physical Chemistry C* **2012,** *116* (50), 26396-26404.
90. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
91. Vosko, S. H.; Wilk, L.; Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics* **1980,** *58* (8), 1200-1211.
92. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized gradient approximation made simple. *Physical review letters* **1996,** *77* (18), 3865.
93. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* **1988,** *38* (6), 3098.
94. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988,** *37* (2), 785.
95. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics* **1993,** *98* (7), 5648-5652.
96. Janesko, B. G.; Scalmani, G.; Frisch, M. J., Practical auxiliary basis implementation of Rung 3.5 functionals. *The Journal of chemical physics* **2014,** *141* (3), 034103.
97. Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J., Achieving linear scaling in exchange-correlation density functional quadratures. *Chemical physics letters* **1996,** *257* (3-4), 213-223.
98. Wadt, W. R.; Hay, P. J., Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *The Journal of Chemical Physics* **1985,** *82* (1), 284-298.
99. Ilatovskiy, A. V.; Abagyan, R.; Kufareva, I., Quantum mechanics approaches to drug research in the era of structural chemogenomics. *International journal of quantum chemistry* **2013,** *113* (12), 1669-1675.
100. Prada-Gracia, D.; Huerta-Yépez, S.; Moreno-Vargas, L. M., Application of computational methods for anticancer drug discovery, design, and optimization. *Boletín Médico del Hospital Infantil de México* **2016,** *73* (6), 411-423.

101. Kapetanovic, I. M., COMPUTER-AIDED DRUG DISCOVERY AND DEVELOPMENT (CADDD): in silico-chemico-biological approach. *Chem. Biol. Interact.* **2008,** *171* (2), 165-176.
102. Jorgensen, W. L., The Many Roles of Computation in Drug Discovery. *Science* **2004,** *303* (5665), 1813-1818.
103. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Computational Methods in Drug Discovery. *Pharmacological Reviews* **2014,** *66* (1), 334-395.
104. Boehm, H.-J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F., Novel Inhibitors of DNA Gyrase: 3D Structure Based Biased Needle Screening, Hit Validation by Biophysical Methods, and 3D Guided Optimization. A Promising Alternative to Random Screening. *J. Med. Chem.* **2000,** *43* (14), 2664-2674.
105. Meagher, K. L.; Carlson, H. A., Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case. *J. Am. Chem. Soc.* **2004,** *126* (41), 13276-13281.
106. Varady, J.; Wu, X.; Fang, X.; Min, J.; Hu, Z.; Levant, B.; Wang, S., Molecular Modeling of the Three-Dimensional Structure of Dopamine 3 (D3) Subtype Receptor: Discovery of Novel and Potent D3 Ligands through a Hybrid Pharmacophore- and Structure-Based Database Searching Approach. *J. Med. Chem.* **2003,** *46* (21), 4377-4392.
107. Baldwin, J. J.; Ponticello, G. S.; Anderson, P. S.; Christy, M. E.; Murcko, M. A.; Randall, W. C.; Schwam, H.; Sugrue, M. F.; Gautheron, P., Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. *J. Med. Chem.* **1989,** *32* (12), 2510-2513.
108. Olsen, L.; Jost, S.; Adolph, H.-W.; Pettersson, I.; Hemmingsen, L.; Jørgensen, F. S., New leads of metallo-β-lactamase inhibitors from structure-based pharmacophore design. *Biorg. Med. Chem.* **2006,** *14* (8), 2627-2635.
109. Pollack, V. A.; Savage, D. M.; Baker, D. A.; Tsaparikos, K. E.; Sloan, D. E.; Moyer, J. D.; Barbacci, E. G.; Pustilnik, L. R.; Smolarek, T. A.; Davis, J. A.; Vaidya, M. P.; Arnold, L. D.; Doty, J. L.; Iwata, K. K.; Morin, M. J., Inhibition of epidermal growth factor receptor-associated tyrosine phosphorylation in human carcinomas with CP-358,774: Dynamics of receptor inhibition in situ and antitumor effects in athymic mice. *J. Pharmacol. Exp. Ther.* **1999,** *291* (2), 739-748.
110. Ng, S. S. W.; Tsao, M. S.; Nicklee, T.; Hedley, D. W., Effects of the epidermal growth factor receptor inhibitor OSI-774, Tarceva, on downstream signaling pathways and apoptosis in human pancreatic adenocarcinoma. *Molecular Cancer Therapeutics* **2002,** *1* (10), 777-783.
111. Bulgaru, A. M.; Mani, S.; Goel, S.; Perez-Soler, R., Erlotinib (Tarceva®): A promising drug targeting epidermal growth factor receptor tyrosine kinase. *Expert Review of Anticancer Therapy* **2003,** *3* (3), 269-279.
112. Heim, M.; Sharifi, M.; Hilger, R. A.; Scheulen, M. E.; Seeber, S.; Strumberg, D., Antitumor effect and potentiation or reduction in cytotoxic drug activity in human colon carcinoma cells by the Raf kinase inhibitor (RKI) BAY 43-9006. *International journal of clinical pharmacology and therapeutics* **2003,** *41* (12), 616-617.
113. Ahmad, T.; Eisen, T., Kinase Inhibition with BAY 43–9006 in Renal Cell Carcinoma. *Clinical Cancer Research* **2004,** *10* (18), 6388S-6392S.

114. Xia, W.; Liu, L. H.; Ho, P.; Spector, N. L., Truncated ErbB2 receptor (p95ErbB2) is regulated by heregulin through heterodimer formation with ErbB3 yet remains sensitive to the dual EGFR/ErbB2 kinase inhibitor GW572016. *Oncogene* **2004,** *23* (3), 646-653.
115. Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L., A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): Relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Research* **2004,** *64* (18), 6652-6659.
116. Jarman, M.; Elaine Barrie, S.; Liera, J. M., The 16,17-double bond is needed for irreversible inhibition of human cytochrome P45017α by abiraterone (17-(3-Pyridyl)androsta-5,16-dien-3β-ol) and related steroidal inhibitors. *J. Med. Chem.* **1998,** *41* (27), 5375-5381.
117. O'Donnell, A.; Judson, I.; Dowsett, M.; Raynaud, F.; Dearnaley, D.; Mason, M.; Harland, S.; Robbins, A.; Halbert, G.; Nutley, B.; Jarman, M., Hormonal impact of the 17α-hydroxylase/C17,20-lyase inhibitor abiraterone acetate (CB7630) in patients with prostate cancer. *British Journal of Cancer* **2004,** *90* (12), 2317-2325.
118. Jagusch, C.; Negri, M.; Hille, U. E.; Hu, Q.; Bartels, M.; Jahn-Hoffmann, K.; Mendieta, M. A. E. P. B.; Rodenwaldt, B.; Müller-Vieira, U.; Schmidt, D.; Lauterbach, T.; Recanatini, M.; Cavalli, A.; Hartmann, R. W., Synthesis, biological evaluation and molecular modelling studies of methyleneimidazole substituted biaryls as inhibitors of human 17α-hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure. *Bioorg. Med. Chem.* **2008,** *16* (4), 1992-2010.
119. Mobley, D. L., Let's get honest about sampling. *J. Comput. Aided Mol. Des.* **2012,** *26* (1), 93-95.
120. Paton, R. S.; Goodman, J. M., Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *Journal of Chemical Information and Modeling* **2009,** *49* (4), 944-955.
121. Yilmazer, N. D.; Heitel, P.; Schwabe, T.; Korth, M., Benchmark of electronic structure methods for protein–ligand interactions based on high-level reference data. *Journal of Theoretical and Computational Chemistry* **2015,** *14* (01), 1540001.
122. Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V., Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. *The journal of physical chemistry B* **2012,** *116* (29), 8375-8382.
123. Abagyan, R. A.; Totrov, M., Ab InitioFolding of Peptides by the Optimal-Bias Monte Carlo Minimization Procedure. *Journal of computational physics* **1999,** *151* (1), 402-421.
124. Wheeler, R. A.; Spellmeyer, D. C., *Annual Reports in Computational Chemistry, Vol. 6*. Elsevier: 2010.
125. Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R., Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π–π, CH/π, and SH/π interactions. *J. Comput. Chem.* **2009,** *30* (14), 2187-2193.
126. Ehrlich, S.; Göller, A. H.; Grimme, S., Towards full Quantum-Mechanics-based Protein–Ligand Binding Affinities. *ChemPhysChem* **2017,** *18* (8), 898-905.
127. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E., Prediction of Protein−Ligand Interactions. Docking and Scoring: Successes and Gaps. *J. Med. Chem.* **2006,** *49* (20), 5851-5855.

128. Yan, Y.; Wang, W.; Sun, Z.; Zhang, J. Z. H.; Ji, C., Protein–Ligand Empirical Interaction Components for Virtual Screening. *Journal of Chemical Information and Modeling* **2017,** *57* (8), 1793-1806.
129. Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R., Forging the Basis for Developing Protein–Ligand Interaction Scoring Functions. *Acc. Chem. Res.* **2017,** *50* (2), 302-309.
130. Norbury, A.; Sinha, A., The co-ordination of ambidentate ligands. *Quarterly Reviews, Chemical Society* **1970,** *24* (1), 69-94.
131. Bayer, E. A.; Wilchek, M., Application of avidin—biotin technology to affinity-based separations. *J. Chromatogr. A* **1990,** *510*, 3-11.
132. Wilchek, M.; Bayer, E. A., Avidin-biotin technology ten years on: has it lived up to its expectations? *Trends Biochem. Sci* **1989,** *14* (10), 408-412.
133. Green, N. M., Avidin. 1. The use of [(14)C]biotin for kinetic studies and for assay. *Biochem. J* **1963,** *89* (3), 585-591.
134. Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. L., Three-dimensional structures of avidin and the avidin-biotin complex. *PNAS* **1993,** *90* (11), 5076-5080.
135. Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L.; Chen, W.; Moghaddam, S.; Gilson, M. K.; Kim, K.; Inoue, Y., A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy–entropy compensation. *PNAS* **2007,** *104* (52), 20737-20742.
136. Wei, W.; Sun, Y.; Zhu, M.; Liu, X.; Sun, P.; Wang, F.; Gui, Q.; Meng, W.; Cao, Y.; Zhao, J., Structural Insights and the Surprisingly Low Mechanical Stability of the Au–S Bond in the Gold-Specific Protein GolB. *Journal of the American Chemical Society* **2015,** *137* (49), 15358-15361.
137. Checa, S. K.; Soncini, F. C., Bacterial gold sensing and resistance. *BioMetals* **2011,** *24* (3), 419-427.
138. Checa, S. K.; Espariz, M.; Audero, M. E. P.; Botta, P. E.; Spinelli, S. V.; Soncini, F. C., Bacterial sensing of and resistance to gold salts. *Mol. Microbiol.* **2007,** *63* (5), 1307-1318.
139. Jian, X.; Wasinger, E. C.; Lockard, J. V.; Chen, L. X.; He, C., Highly Sensitive and Selective Gold(I) Recognition by a Metalloregulator in Ralstonia metallidurans. *J. Am. Chem. Soc.* **2009,** *131* (31), 10869-10871.
140. Wei, W.; Zhu, T.; Wang, Y.; Yang, H.; Hao, Z.; Chen, P. R.; Zhao, J., Engineering a gold-specific regulon for cell-based visual detection and recovery of gold. *Chemical Science* **2012,** *3* (6), 1780-1784.
141. Häkkinen, H., The gold-sulfur interface at the nanoscale. *Nature chemistry* **2012,** *4* (6), 443-455.
142. Grandbois, M.; Beyer, M.; Rief, M.; Clausen-Schaumann, H.; Gaub, H. E., How strong is a covalent bond? *Science* **1999,** *283* (5408), 1727-1730.
143. Hollinger, M. A., Toxicological aspects of topical silver pharmaceuticals. *Critical reviews in toxicology* **1996,** *26* (3), 255-260.
144. Xue, Y.; Li, X.; Li, H.; Zhang, W., Quantifying thiol–gold interactions towards the efficient strength control. *Nature communications* **2014,** *5*.
145. Qi, X.; Chan, W. L.; Read, R. J.; Zhou, A.; Carrell, R. W., Temperature-responsive release of thyroxine and its environmental adaptation in Australians. *Proceedings of the Royal Society B: Biological Sciences* **2014,** *281* (1779).

146. Cheng, S.-Y.; Leonard, J. L.; Davis, P. J., Molecular Aspects of Thyroid Hormone Actions. *Endocrine Reviews* **2010,** *31* (2), 139-170.
147. Robbins, J., New Ideas in Thyroxine-Binding Globulin Biology. *The Journal of Clinical Endocrinology & Metabolism* **2000,** *85* (11), 3994-3995.
148. Zhou, A.; Wei, Z.; Read, R. J.; Carrell, R. W., Structural mechanism for the carriage and release of thyroxine in the blood. *PNAS* **2006,** *103* (36), 13321-13326.
149. Appel, M. J.; Bertozzi, C. R., Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chemical Biology* **2015,** *10* (1), 72-84.
150. Hudak, J. E.; Barfield, R. M.; de Hart, G. W.; Grob, P.; Nogales, E.; Bertozzi, C. R.; Rabuka, D., Synthesis of Heterobifunctional Protein Fusions Using Copper-Free Click Chemistry and the Aldehyde Tag. *Angew. Chem. Int. Ed.* **2012,** *51* (17), 4161-4165.
151. Meury, M.; Knop, M.; Seebeck, F. P., Structural Basis for Copper–Oxygen Mediated C−H Bond Activation by the Formylglycine-Generating Enzyme. *Angew. Chem. Int. Ed.* **2017,** *56* (28), 8115-8119.
152. Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L.-X.; Bertozzi, C. R., Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjugate Chem.* **2014,** *25* (4), 788-795.
153. Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R., Introducing genetically encoded aldehydes into proteins. *Nature Chemical Biology* **2007,** *3*, 321.
154. Rush, J. S.; Bertozzi, C. R., New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008,** *130* (37), 12240-12241.
155. Knop, M.; Engi, P.; Lemnaru, R.; Seebeck, F. P., In Vitro Reconstitution of Formylglycine-Generating Enzymes Requires Copper(I). *ChemBioChem* **2015,** *16* (15), 2147-2150.
156. Schmidt, B.; Selmer, T.; Ingendoh, A.; Figurat, K. v., A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **1995,** *82* (2), 271-278.
157. Dierks, T.; Dickmanns, A.; Preusser-Kunze, A.; Schmidt, B.; Mariappan, M.; von Figura, K.; Ficner, R.; Rudolph, M. G., Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme. *Cell* **2005,** *121* (4), 541-552.
158. Knop, M.; Dang, T. Q.; Jeschke, G.; Seebeck, F. P., Copper is a Cofactor of the Formylglycine-Generating Enzyme. *ChemBioChem* **2017,** *18* (2), 161-165.
159. Solioz, M.; Odermatt, A., Copper and Silver Transport by CopB-ATPase in Membrane Vesicles of Enterococcus hirae. *J. Biol. Chem.* **1995,** *270* (16), 9217-9221.
160. Petris, M. J.; Mercer, J. F.; Culvenor, J. G.; Lockhart, P.; Gleeson, P. A.; Camakaris, J., Ligand-regulated transport of the Menkes copper P-type ATPase efflux pump from the Golgi apparatus to the plasma membrane: a novel mechanism of regulated trafficking. *The EMBO Journal* **1996,** *15* (22), 6084-6095.
161. Changela, A.; Chen, K.; Xue, Y.; Holschen, J.; Outten, C. E.; O'Halloran, T. V.; Mondragón, A., Molecular Basis of Metal-Ion Selectivity and Zeptomolar Sensitivity by CueR. *Science* **2003,** *301* (5638), 1383-1387.

162. Philips, S. J.; Canalizo-Hernandez, M.; Yildirim, I.; Schatz, G. C.; Mondragón, A.; O'Halloran, T. V., Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* **2015,** *349* (6250), 877-881.
163. Rensing, C.; Fan, B.; Sharma, R.; Mitra, B.; Rosen, B. P., CopA: An Escherichia coli Cu(I)-translocating P-type ATPase. *PNAS* **2000,** *97* (2), 652-656.
164. Lu, Z. H.; Solioz, M., Copper-induced Proteolysis of the CopZ Copper Chaperone ofEnterococcus hirae. *J. Biol. Chem.* **2001,** *276* (51), 47822-47827.
165. Gitschier, J.; Moffat, B.; Reilly, D.; Wood, W. I.; Fairbrother, W. J., Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. *Nature Structural Biology* **1998,** *5*, 47.
166. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008,** *10* (44), 6615-6620.
167. Ajees, A. A.; Marapakala, K.; Packianathan, C.; Sankaran, B.; Rosen, B. P., Structure of an As (III) S-adenosylmethionine methyltransferase: insights into the mechanism of arsenic biotransformation. *Biochemistry* **2012,** *51* (27), 5476-5485.
168. D.A. Case, D. S. C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman *AMBER 2017*, AMBER 2017, University of California, San Francisco, 2017.
169. Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J., Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006,** *2* (3), 815-826.
170. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *PCCP* **2008,** *10* (44), 6615-6620.
171. 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.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
172. Lu, T.; Chen, F., Multiwfn: a multifunctional wavefunction analyzer. *Journal of computational chemistry* **2012,** *33* (5), 580-592.
173. Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C., Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987,** *109* (26), 7968-7979.
174. Kolář, M. H.; Hobza, P., Computer Modeling of Halogen Bonds and Other σ-Hole Interactions. *Chem. Rev.* **2016,** *116* (9), 5155-5187.
175. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *Journal of molecular graphics* **1996,** *14* (1), 33-38.
176. Lidbury, I.; Murrell, J. C.; Chen, Y., Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proceedings of the National Academy of Sciences* **2014,** *111* (7), 2710-2715.

177. Lin, T.-Y.; Timasheff, S. N., Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry* **1994,** *33* (42), 12695-12701.
178. Yancey, P. H.; Siebenaller, J. F., Trimethylamine oxide stabilizes teleost and mammalian lactate dehydrogenases against inactivation by hydrostatic pressure and trypsinolysis. *Journal of Experimental Biology* **1999,** *202* (24), 3597-3603.
179. Yancey, P. H.; Blake, W. R.; Conley, J., Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **2002,** *133* (3), 667-676.
180. Koeth, R. A.; Wang, Z.; Levison, B. S.; Buffa, J. A.; Org, E.; Sheehy, B. T.; Britt, E. B.; Fu, X.; Wu, Y.; Li, L., Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature medicine* **2013,** *19* (5), 576.
181. Anthoni, U.; Christophersen, C.; Gram, L.; Nielsen, N. H.; Nielsen, P., Poisonings from flesh of the Greenland shark Somniosus microcephalus may be due to trimethylamine. *Toxicon* **1991,** *29* (10), 1205-1212.
182. Marzo, A.; Curti, S., L-Carnitine moiety assay: an up-to-date reappraisal covering the commonest methods for various applications. *Journal of Chromatography B: Biomedical Sciences and Applications* **1997,** *702* (1-2), 1-20.
183. Hsu, C.-N.; Lu, P.-C.; Lo, M.-H.; Lin, I.; Chang-Chien, G.-P.; Lin, S.; Tain, Y.-L., Gut Microbiota-Dependent Trimethylamine N-Oxide Pathway Associated with Cardiovascular Risk in Children with Early-Stage Chronic Kidney Disease. *International journal of molecular sciences* **2018,** *19* (12), 3699.
184. Tang, W. W.; Wang, Z.; Kennedy, D. J.; Wu, Y.; Buffa, J. A.; Agatisa-Boyle, B.; Li, X. S.; Levison, B. S.; Hazen, S. L., Gut Microbiota-Dependent Trimethylamine N-Oxide (TMAO) Pathway Contributes to Both Development of Renal Insufficiency and Mortality Risk in Chronic Kidney DiseaseNovelty and Significance. *Circulation research* **2015,** *116* (3), 448-455.
185. Ierardi, E.; Sorrentino, C.; Principi, M.; Giorgio, F.; Losurdo, G.; Di Leo, A., Intestinal microbial metabolism of phosphatidylcholine: a novel insight in the cardiovascular risk scenario. *Hepatobiliary surgery and nutrition* **2015,** *4* (4), 289.
186. Craciun, S.; Balskus, E. P., Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proceedings of the National Academy of Sciences* **2012,** *109* (52), 21307-21312.
187. Zhu, Y.; Jameson, E.; Crosatti, M.; Schäfer, H.; Rajakumar, K.; Bugg, T. D.; Chen, Y., Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *Proceedings of the National Academy of Sciences* **2014**, 201316569.
188. Cho, C. E.; Taesuwan, S.; Malysheva, O. V.; Bender, E.; Tulchinsky, N. F.; Yan, J.; Sutter, J. L.; Caudill, M. A., Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: A randomized controlled trial. *Molecular nutrition & food research* **2017,** *61* (1), 1600324.
189. Rohrmann, S.; Linseisen, J.; Allenspach, M.; von Eckardstein, A.; Müller, D., Plasma Concentrations of Trimethylamine-N-oxide Are Directly Associated with Dairy Food Consumption and Low-Grade Inflammation in a German Adult Population, 2. *The Journal of nutrition* **2015,** *146* (2), 283-289.

190. Moore, J. O.; Hendrickson, W. A., An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure* **2012,** *20* (4), 729-741.
191. Boys, S. F.; Bernardi, F. d., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970,** *19* (4), 553-566.
192. Li, C.; Zhao, X.; Wang, A.; Huber, G. W.; Zhang, T., Catalytic transformation of lignin for the production of chemicals and fuels. *Chemical Reviews* **2015,** *115* (21), 11559-11624.
193. Amidon, T. E.; Liu, S., Water-based woody biorefinery. *Biotechnology advances* **2009,** *27* (5), 542-550.
194. Hatakeyama, H.; Kosugi, R.; Hatakeyama, T., Thermal properties of lignin-and molasses-based polyurethane foams. *Journal of Thermal Analysis and Calorimetry* **2008,** *92* (2), 419-424.
195. Vazquez, G.; Rodríguez-Bona, C.; Freire, S.; Gonzalez-Alvarez, J.; Antorrena, G., Acetosolv pine lignin as copolymer in resins for manufacture of exterior grade plywoods. *Bioresource technology* **1999,** *70* (2), 209-214.
196. Zhang, Y.-H. P.; Cui, J.; Lynd, L. R.; Kuang, L. R., A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **2006,** *7* (2), 644-648.
197. Zhang, Y. H. P.; Ding, S. Y.; Mielenz, J. R.; Cui, J. B.; Elander, R. T.; Laser, M.; Himmel, M. E.; McMillan, J. R.; Lynd, L. R., Fractionating recalcitrant lignocellulose at modest reaction conditions. *Biotechnology and bioengineering* **2007,** *97* (2), 214-223.
198. Moxley, G.; Zhu, Z.; Zhang, Y.-H. P., Efficient sugar release by the cellulose solvent-based lignocellulose fractionation technology and enzymatic cellulose hydrolysis. *Journal of agricultural and food chemistry* **2008,** *56* (17), 7885-7890.
199. Kim, J.; Mazza, G., Optimization of phosphoric acid catalyzed fractionation and enzymatic digestibility of flax shives. *Industrial crops and products* **2008,** *28* (3), 346-355.
200. Gomiero, T.; Paoletti, M. G.; Pimentel, D., Biofuels: efficiency, ethics, and limits to human appropriation of ecosystem services. *Journal of agricultural and environmental ethics* **2010,** *23* (5), 403-434.
201. Lewin, M.; Goldstein, I. S., *Wood structure and composition*. CRC Press: 1991.
202. Wen, J.-L.; Xue, B.-L.; Xu, F.; Sun, R.-C.; Pinkert, A., Unmasking the structural features and property of lignin from bamboo. *Industrial Crops and Products* **2013,** *42*, 332-343.
203. Yue, F.; Lu, F.; Sun, R.; Ralph, J., Synthesis and Characterization of New 5-Linked Pinoresinol Lignin Models. *Chemistry–A European Journal* **2012,** *18* (51), 16402-16410.
204. Manara, P.; Zabaniotou, A.; Vanderghem, C.; Richel, A., Lignin extraction from Mediterranean agro-wastes: Impact of pretreatment conditions on lignin chemical structure and thermal degradation behavior. *Catalysis Today* **2014,** *223*, 25-34.
205. Nimz, H., Beech lignin—proposal of a constitutional scheme. *Angewandte Chemie International Edition in English* **1974,** *13* (5), 313-321.
206. Chatel, G.; Rogers, R. D., Oxidation of lignin using ionic liquids□ An innovative strategy to produce renewable chemicals. *ACS Sustainable Chemistry & Engineering* **2013,** *2* (3), 322-339.

207. Li, M.-F.; Sun, S.-N.; Xu, F.; Sun, R.-C., Organosolv fractionation of lignocelluloses for fuels, chemicals and materials: A biorefinery processing perspective. In *Biomass conversion*, Springer: 2012; pp 341-379.
208. Dence, C. W., The determination of lignin. In *Methods in lignin chemistry*, Springer: 1992; pp 33-61.
209. Yang, L.; Seshan, K.; Li, Y., A review on thermal chemical reactions of lignin model compounds. *Catalysis Today* **2017,** *298*, 276-297.
210. Yao, S. G.; Meier, M. S.; Pace III, R. B.; Crocker, M., A comparison of the oxidation of lignin model compounds in conventional and ionic liquid solvents and application to the oxidation of lignin. *RSC Advances* **2016,** *6* (106), 104742-104753.
211. Deng, W.; Zhang, H.; Wu, X.; Li, R.; Zhang, Q.; Wang, Y., Oxidative conversion of lignin and lignin model compounds catalyzed by CeO 2-supported Pd nanoparticles. *Green chemistry* **2015,** *17* (11), 5009-5018.
212. Sturgeon, M. R.; Kim, S.; Lawrence, K.; Paton, R. S.; Chmely, S. C.; Nimlos, M.; Foust, T. D.; Beckham, G. T., A mechanistic investigation of acid-catalyzed cleavage of aryl-ether linkages: implications for lignin depolymerization in acidic environments. *ACS Sustainable Chemistry & Engineering* **2013,** *2* (3), 472-485.
213. Jennings, J. A.; Parkin, S.; Munson, E.; Delaney, S. P.; Calahan, J. L.; Isaacs, M.; Hong, K.; Crocker, M., Regioselective Baeyer–Villiger oxidation of lignin model compounds with tin beta zeolite catalyst and hydrogen peroxide. *RSC Advances* **2017,** *7* (42), 25987-25997.
214. Gierer, J.; Pettersson, I., Studies on the condensation of lignins in alkaline media. Part II. The formation of stilbene and arylcoumaran structures through neighbouring group participation reactions. *Canadian Journal of Chemistry* **1977,** *55* (4), 593-599.
215. GIERER, J.-F.; LINDEBERG, O., Studies on the Condensation of Lignins in Alkaline Media. Part III.* The Formation of Stilbenes, Aryl. *Acta Chem. Scand. B* **1979,** *33* (8).
216. Imai, T.; Yokoyama, T.; Matsumoto, Y., Revisiting the mechanism of β-O-4 bond cleavage during acidolysis of lignin IV: Dependence of acidolysis reaction on the type of acid. *Journal of wood science* **2011,** *57* (3), 219-225.
217. Yokoyama, T.; Matsumoto, Y., Revisiting the mechanism of β-O-4 bond cleavage during acidolysis of lignin. Part 2: detailed reaction mechanism of a non-phenolic C6-C2 type model compound. *Journal of Wood Chemistry and Technology* **2010,** *30* (3), 269-282.
218. Yokoyama, T., Revisiting the mechanism of β-O-4 bond cleavage during acidolysis of lignin. Part 6: A review. *Journal of Wood Chemistry and Technology* **2015,** *35* (1), 27-42.
219. Janesko, B. G., Acid-catalyzed hydrolysis of lignin β-O-4 linkages in ionic liquid solvents: a computational mechanistic study. *Physical Chemistry Chemical Physics* **2014,** *16* (11), 5423-5433.
220. Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D., Dissolution of cellose with ionic liquids. *Journal of the American chemical society* **2002,** *124* (18), 4974-4975.
221. Wu, J.; Zhang, J.; Zhang, H.; He, J.; Ren, Q.; Guo, M., Homogeneous acetylation of cellulose in a new ionic liquid. *Biomacromolecules* **2004,** *5* (2), 266-268.
222. Fort, D. A.; Remsing, R. C.; Swatloski, R. P.; Moyna, P.; Moyna, G.; Rogers, R. D., Can ionic liquids dissolve wood? Processing and analysis of lignocellulosic materials with 1-n-butyl-3-methylimidazolium chloride. *Green Chemistry* **2007,** *9* (1), 63-69.

223. Plechkova, N. V.; Seddon, K. R., Applications of ionic liquids in the chemical industry. *Chemical Society Reviews* **2008,** *37* (1), 123-150.
224. Janesko, B. G., Modeling interactions between lignocellulose and ionic liquids using DFT-D. *Physical Chemistry Chemical Physics* **2011,** *13* (23), 11393-11401.
225. Chiappe, C.; Pomelli, C. S., Computational studies on organic reactivity in ionic liquids. *Physical Chemistry Chemical Physics* **2013,** *15* (2), 412-423.
226. Cramer, C. J.; Truhlar, D. G., Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chemical Reviews* **1999,** *99* (8), 2161-2200.
227. Bessac, F.; Maseras, F., DFT modeling of reactivity in an ionic liquid: How many ion pairs? *Journal of computational chemistry* **2008,** *29* (6), 892-899.
228. Fraile, J. M.; García, J. I.; Hormigón, Z.; Mayoral, J. A.; Saavedra, C. J.; Salvatella, L., Role of Substituents in the Solid Acid-Catalyzed Cleavage of the β-O-4 Linkage in Lignin Models. *ACS Sustainable Chemistry & Engineering* **2017,** *6* (2), 1837-1847.
229. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B* **2009,** *113* (18), 6378-6396.
230. Bernales, V. S.; Marenich, A. V.; Contreras, R.; Cramer, C. J.; Truhlar, D. G., Quantum mechanical continuum solvation models for ionic liquids. *The Journal of Physical Chemistry B* **2012,** *116* (30), 9122-9129.
231. Moore, G.; Thomson, G.; Harrington, D. J., *Warfarin Monitoring: Standard Practice and Beyond*. Nova Science Publishers: 2012.
232. Lam, Y.-W. F.; Cavallari, L. H., *Pharmacogenomics: Challenges and Opportunities in Therapeutic Implementation*. Academic Press: 2013.
233. Tie, J.-K.; Jin, D.-Y.; Straight, D. L.; Stafford, D. W., Functional study of the vitamin K cycle in mammalian cells. *Blood* **2011,** *117* (10), 2967-2974.
234. Tie, J. K.; Stafford, D. W., Structural and functional insights into enzymes of the vitamin K cycle. *Journal of Thrombosis and Haemostasis* **2016,** *14* (2), 236-247.
235. Silverman, R. B., Chemical model studies for the mechanism of vitamin K epoxide reductase. *Journal of the American Chemical Society* **1981,** *103* (19), 5939-5941.
236. Torkashvand, M.; Gholivand, M. B.; Taherpour, A.; Boochani, A.; Akhtar, A., Introduction of a carbon paste electrode based on nickel carbide for investigation of interaction between warfarin and vitamin K1. *Journal of Pharmaceutical and Biomedical Analysis* **2017,** *139* (Supplement C), 156-164.
237. Bell, R. G.; Matschiner, J. T., Warfarin and the inhibition of vitamin K activity by an oxide metabolite. *Nature* **1972,** *237* (5349), 32-33.
238. Berkner, K. L., The vitamin K–dependent carboxylase. *Annu. Rev. Nutr.* **2005,** *25*, 127-149.
239. Furie, B.; Bouchard, B. A.; Furie, B. C., Vitamin K-dependent biosynthesis of γ-carboxyglutamic acid. *Blood* **1999,** *93* (6), 1798-1808.
240. Mann, K.; Nesheim, M.; Church, W.; Haley, P.; Krishnaswamy, S., Surface-dependent reactions of the vitamin K-dependent enzyme complexes. *Blood* **1990,** *76* (1), 1-16.
241. Tie, J.-K.; Nicchitta, C.; Von Heijne, G.; Stafford, D. W., Membrane topology mapping of vitamin K epoxide reductase by in vitro translation/cotranslocation. *Journal of Biological Chemistry* **2005,** *280* (16), 16410-16416.

242. Tie, J.-K.; Jin, D.-Y.; Stafford, D. W., Human vitamin K epoxide reductase and its bacterial homologue have different membrane topologies and reaction mechanisms. *Journal of Biological Chemistry* **2012,** *287* (41), 33945-33955.
243. Sinhadri, B. C. S.; Jin, D.-Y.; Stafford, D. W.; Tie, J.-K., Vitamin K epoxide reductase and its paralogous enzyme have different structures and functions. *Scientific reports* **2017,** *7* (1), 17632.
244. Wu, S.; Chen, X.; Jin, D.-Y.; Stafford, D. W.; Pedersen, L. G.; Tie, J.-K., Warfarin and vitamin K epoxide reductase: a molecular accounting for observed inhibition. *Blood* **2018,** *132* (6), 647-657.
245. Czogalla, K. J.; Biswas, A.; Höning, K.; Hornung, V.; Liphardt, K.; Watzka, M.; Oldenburg, J., Warfarin and vitamin K compete for binding to Phe55 in human VKOR. *Nature structural & molecular biology* **2017,** *24* (1), 77-85.
246. Goodstadt, L.; Ponting, C. P., Vitamin K epoxide reductase: homology, active site and catalytic mechanism. *Trends in biochemical sciences* **2004,** *29* (6), 289-292.
247. Li, W.; Schulman, S.; Dutton, R. J.; Boyd, D.; Beckwith, J.; Rapoport, T. A., Structure of a bacterial homolog of vitamin K epoxide reductase. *Nature* **2010,** *463* (7280), 507.
248. Hodroge, A.; Longin-Sauvageon, C.; Fourel, I.; Benoit, E.; Lattard, V., Biochemical characterization of spontaneous mutants of rat VKORC1 involved in the resistance to antivitamin K anticoagulants. *Arch Biochem Biophys* **2011,** *515* (1-2), 14-20.
249. Hodroge, A.; Matagrin, B.; Moreau, C.; Fourel, I.; Hammed, A.; Benoit, E.; Lattard, V., VKORC1 mutations detected in patients resistant to vitamin K antagonists are not all associated with a resistant VKOR activity. *Journal of Thrombosis and Haemostasis* **2012,** *10* (12), 2535-2543.
250. Tie, J. K.; Stafford, D. W., Structure and function of vitamin K epoxide reductase. *Vitamins & Hormones* **2008,** *78*, 103-130.
251. Pelz, H.-J.; Rost, S.; Hünerberg, M.; Fregin, A.; Heiberg, A.-C.; Baert, K.; MacNicoll, A. D.; Prescott, C. V.; Walker, A.-S.; Oldenburg, J., The genetic basis of resistance to anticoagulants in rodents. *Genetics* **2005,** *170* (4), 1839-1847.
252. Rost, S.; Fregin, A.; Hünerberg, M.; Bevans, C. G.; Müller, C. R.; Oldenburg, J., Site-directed mutagenesis of coumarin-type anticoagulant-sensitive VKORC1: evidence that highly conserved amino acids define structural requirements of enzymatic activity and inhibition by warfarin. *Thrombosis and haemostasis* **2005,** *94* (4), 780-786.
253. Rost, S.; Fregin, A.; Ivaskevicius, V.; Conzelmann, E.; Hörtnagel, K.; Pelz, H.-J.; Lappegard, K.; Seifried, E.; Scharrer, I.; Tuddenham, E. G., Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* **2004,** *427* (6974), 537.
254. Rost, S.; Pelz, H.-J.; Menzel, S.; MacNicoll, A. D.; León, V.; Song, K.-J.; Jäkel, T.; Oldenburg, J.; Müller, C. R., Novel mutations in the VKORC1 gene of wild rats and mice–a response to 50 years of selection pressure by warfarin? *Bmc Genetics* **2009,** *10* (1), 4.
255. Liu, S.; Cheng, W.; Grider, R. F.; Shen, G.; Li, W., Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer. *Nature communications* **2014,** *5*, 3110.

256. D'ambrosio, R.; D'andrea, G.; Cafolla, A.; Faillace, F.; Margaglione, M., A new vitamin K epoxide reductase complex subunit-1 (VKORC1) mutation in a patient with decreased stability of CYP2C9 enzyme. *Journal of Thrombosis and Haemostasis* **2007,** *5* (1), 191-193.
257. Lewis, B. C.; Nair, P. C.; Heran, S. S.; Somogyi, A. A.; Bowden, J. J.; Doogue, M. P.; Miners, J. O., Warfarin resistance associated with genetic polymorphism of VKORC1: linking clinical response to molecular mechanism using computational modeling. *Pharmacogenetics and genomics* **2016,** *26* (1), 44-50.
258. Chatron, N.; Chalmond, B.; Trouvé, A.; Benoit, E.; Caruel, H.; Lattard, V.; Tchertanov, L., Identification of the functional states of human vitamin K epoxide reductase from molecular dynamics simulations. *RSC Advances* **2017,** *7* (82), 52071-52090.
259. Gao, M.; Zhou, H.; Skolnick, J., DESTINI: A deep-learning approach to contact-driven protein structure prediction. *Scientific reports* **2019,** *9* (1), 3514.
260. Czogalla, K. J.; Biswas, A.; Wendeln, A.-C.; Westhofen, P.; Müller, C. R.; Watzka, M.; Oldenburg, J., Human VKORC1 mutations cause variable degrees of 4-hydroxycoumarin resistance and affect putative warfarin binding interfaces. *Blood* **2013**, blood-2013-05-501692.
261. Zhou, H.; Cao, H.; Skolnick, J., FINDSITEcomb2.0: A New Approach for Virtual Ligand Screening of Proteins and Virtual Target Screening of Biomolecules. *Journal of Chemical Information and Modeling* **2018,** *58* (11), 2343-2354.
262. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010,** *31* (2), 455-461.
263. Hirsh, J.; Fuster, V.; Ansell, J.; Halperin, J. L., American Heart Association/American College of Cardiology foundation guide to warfarin therapy. *Circulation* **2003,** *107* (12), 1692-1711.
264. Shaik, A. N.; Grater, R.; Lulla, M.; Williams, D. A.; Gan, L. L.; Bohnert, T.; LeDuc, B. W., Comparison of enzyme kinetics of warfarin analyzed by LC–MS/MS QTrap and differential mobility spectrometry. *Journal of Chromatography B* **2016,** *1008*, 164-173.
265. Valente, E.; Lingafelter, E.; Porter, W.; Trager, W., Structure of warfarin in solution. *Journal of Medicinal Chemistry* **1977,** *20* (11), 1489-1493.
266. Rishavy, M. A.; Usubalieva, A.; Hallgren, K. W.; Berkner, K. L., Novel insight into the mechanism of the vitamin K oxidoreductase (VKOR) electron relay through Cys43 and Cys51 reduces VKOR to allow vitamin K reduction and facilitation of vitamin K-dependent protein carboxylation. *Journal of Biological Chemistry* **2011,** *286* (9), 7267-7278.
267. Krieger, E.; Vriend, G., YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **2014,** *30* (20), 2981-2982.
268. Lasseur, R.; Longin-Sauvageon, C.; Videmann, B.; Billeret, M.; Berny, P.; Benoit, E., Warfarin resistance in a French strain of rats. *Journal of biochemical and molecular toxicology* **2006,** *19* (6), 379-385.
269. Bevans, C. G.; Krettler, C.; Reinhart, C.; Tran, H.; Kossmann, K.; Watzka, M.; Oldenburg, J., Determination of the warfarin inhibition constant Ki for vitamin K 2,3-epoxide reductase complex subunit-1 (VKORC1) using an in vitro DTT-driven assay. *Biochim Biophys Acta* **2013,** *1830* (8), 4202-10.

270. Silverman, R. B.; Oliver, J. S., 2-(Fluoromethyl)-3-phytyl-1, 4-naphthoquinone and its 2, 3-epoxide. Inhibition of vitamin K epoxide reductase. *Journal of medicinal chemistry* **1989,** *32* (9), 2138-2141.
271. Fregin, A.; Czogalla, K. J.; Gansler, J.; Rost, S.; Taverna, M.; Watzka, M.; Bevans, C. G.; MüLler, C. R.; Oldenburg, J., A new cell culture-based assay quantifies vitamin K 2,3-epoxide reductase complex subunit 1 function and reveals warfarin resistance phenotypes not shown by the dithiothreitol-driven VKOR assay. *Journal of Thrombosis and Haemostasis* **2013,** *11* (5), 872-880.
272. Jin, D.-Y.; Tie, J.-K.; Stafford, D. W., The conversion of vitamin K epoxide to vitamin K quinone and vitamin K quinone to vitamin K hydroquinone uses the same active site cysteines. *Biochemistry* **2007,** *46* (24), 7279-7283.
273. Rost, S.; Fregin, A.; Hünerberg, M.; Bevans, C. G.; Müller, C. R.; Oldenburg, J., Site-directed mutagenesis of coumarin-type anticoagulant-sensitive VKORC1. *Thrombosis and haemostasis* **2005,** *93* (04), 780-786.
274. Matagrin, B.; Hodroge, A.; Montagut-Romans, A.; Andru, J.; Fourel, I.; Besse, S.; Benoit, E.; Lattard, V., New insights into the catalytic mechanism of vitamin K epoxide reductase (VKORC1)–The catalytic properties of the major mutations of rVKORC1 explain the biological cost associated to mutations. *FEBS Open Bio* **2013,** *3* (1), 144-150.
275. MacNicoll, A.; Dean, R., Warfarin inhibition of hepatic vitamin K epoxide reductase activity in warfarin-susceptible and-resistant house mice (Mus domesticus). *Pesticide Biochemistry and Physiology* **1992,** *44* (1), 68-73.
276. Thijssen, H., Warfarin resistance: vitamin K epoxide reductase of Scottish resistance genes is not irreversibly blocked by warfarin. *Biochemical pharmacology* **1987,** *36* (17), 2753-2757.
277. Misenheimer, T. M.; Suttie, J., Warfarin resistance in a Chicago strain of rats. *Biochemical pharmacology* **1990,** *40* (9), 2079-2084.
278. Feinstein, D. L.; Akpa, B. S.; Ayee, M. A.; Boullerne, A. I.; Braun, D.; Brodsky, S. V.; Gidalevitz, D.; Hauck, Z.; Kalinin, S.; Kowal, K., The emerging threat of superwarfarins: history, detection, mechanisms, and countermeasures. *Annals of the New York Academy of Sciences* **2016,** *1374* (1), 111-122.
279. O'Reilly, R. A.; Aggeler, P. M.; Leong, L. S., Studies on the coumarin anticoagulant drugs: the pharmacodynamics of warfarin in man. *The Journal of clinical investigation* **1963,** *42* (10), 1542-1551.
280. Gebauer, M., Synthesis and structure–activity relationships of novel warfarin derivatives. *Bioorganic & medicinal chemistry* **2007,** *15* (6), 2414-2420.
281. Bader, R.; Henneker, W.; Cade, P. E., Molecular charge distributions and chemical binding. *The Journal of Chemical Physics* **1967,** *46* (9), 3341-3363.
282. Davis, C. H.; Deerfield, D.; Wymore, T.; Stafford, D. W.; Pedersen, L. G., A quantum chemical study of the mechanism of action of Vitamin K epoxide reductase (VKOR): II. Transition states. *Journal of Molecular Graphics and Modelling* **2007,** *26* (2), 401-408.
283. Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W. A.; Skiff, W. M., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society* **1992,** *114* (25), 10024-10035.

284. Case, D.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N., AMBER17. 2017. *San Francisco: University of California* **2018**.

VITA

Stephanie Ing Jones

2800 South University Drive

Fort Worth, TX 76129

Sid W. Richardson Building, Room 427

E-mail: stephanie.jones@tcu.edu

EDUCATION

- **Texas Christian University**……**August 2014-2019**
  PhD in Computational Chemistry

- **California State University, Los Angeles**……**September 2012-2014**
  B.S. in Chemistry

RESEARCH EXPERIENCE

- **Georgia Institute of Technology**……**January 2019-Present**
  Visiting graduate student under Dr. Jeffrey Skolnick (computational biology)
  - Researched drug docking studies of warfarin to vitamin K epoxide reductase (VKOR) enzyme and the development of alternative drugs to warfarin

- **Texas Christian University**……**August 2014-Present**
  Graduate student under Dr. Benjamin Janesko (computational chemistry) and Dr. Eric Simanek (organic chemistry)
  - Synthesized and computationally built a receptor to bind to trimethylamine N-oxide (TMAO) in the body. Calculations predicted that the receptor relied on specified noncovalent interactions as well as several hydrogen bonding interactions in the oxygen terminus. Used calculations to propose a receptor by constructing a substituted triazine and testing its predicted selectivity toward TMAO.
  - Quantified electron delocalization using the Electron Delocalization Range (EDR) tool on F center cluster models, edge and surface defects.
  -Studied acid-catalyzed hydrolysis mechanisms of lignin model compounds with β-O-4 linkages: E1 elimination and dehydration versus protonation and nucleophilic substitution.
  - Combined electrostatic potential and Orbital Overlap Distance $D(\vec{r})$ tool developed from EDR to provide a picture of chemically "hard" and "soft" regions in enzymes. These regions were shown to bind with their respective "hard" and "soft" sections on the ligand.

- **California State University, Los Angeles**……..……..……**June 2012-2014**
Intern under Dr. Carlos Gutierrez (synthetic organic chemistry)

- Provided assistance researching selective deprotection of tris (N-T-BOCserine) Trilactone to provide a motivated approach to the early "Trojan Horse" delivery strategies.

- Extensive training and comprehension in the operation of various machinery and techniques such as Bruker 400 MHx Nuclear Magnetic Resonance (NMR), Rotary Evaporators, Melting Point Apparatus, Vacuum Pumps and Manifolds, Solvent Stills, Melting Point Apparatus, Gas Chromatography, Column Chromatography, Fourier Transform Infrared (FTIR), and Flash Chromatography Separation.

- Thorough understanding of all laboratory safety requirements and protocols in accordance with CSULA regulations.

PRESENTATIONS

- Jones, S. I. & Janesko, B. G. (2017). *Acid Catalyzed Hydrolysis of Lignin Model Compounds with β-O-4 Linkages.* American Chemical Society- 50th Annual Meeting in Miniature of Dallas-Fort Worth, Texas Christian University, Fort Worth, TX.
  -Awarded 4th place in physical/computational chemistry oral presentation.
- Jones, S. I. & Janesko, B. G. (2016). *Quantifying the Delocalization of Trapped Electrons in Crystal Defects.* Southwest Regional Meeting (SWRM), Galveston, TX.
- Jones, S. I. & Janesko, B. G. (2016). *Electron Delocalization in Anionic Defects.* Student Research Symposium (SRS), Texas Christian University, Fort Worth, TX.
- Jones, S. I. & Janesko, B. G. (2016). *Electron Delocalization in Anionic Defects.* 26th Austin Symposium of Molecular Structure and Dynamics (ASMD@D), Dallas, TX.
- Jones, S.; Godinez, C., & Gutierrez, C. (2014). *The Calculation-Guided Synthesis of Unsymmetric Derivatives of E. Coli Siderophore Enterobactin.* Emerging Researchers National (ERN) Conference, Washington D. C.
- Jones, S.; Godinez, C., & Gutierrez, C. (2013). *The Calculation-Guided Synthesis of Unsymmetric Derivatives of E. Coli Siderophore Enterobactin.* Annual Biomedical Research Conference for Minority Students (ABRCMS), Nashville, TN.
- Jones, S.; Godinez, C., & Gutierrez, C. (2013). *The Calculation-Guided Synthesis of Unsymmetric Derivatives of E. Coli Siderophore Enterobactin.* Minority Opportunity in Research (MORE) Programs Retreat, Lake Arrowhead, CA.
- Jones, S.; Amador, R., Godinez, C., & Gutierrez, C. (2012). *Enterobactin Protection with Di-tert-butyl Dicarbonate*. Minority Opportunity in Research (MORE) Programs Retreat, Laguna Beach, CA

- Jones, S.; Amador, R., Godinez, C., & Gutierrez, C. (2012). *The Synthesis and Deprotection of Tris(TritylSerine) Trilactone.* Minority Opportunity in Research (MORE) Programs Chemistry Poster Presentation, CSULA, CA.

PUBLICATIONS

- Mehmood, A.; **Jones, S. I.**; Tao, P.; Janesko, B. G. **"**An Orbital-Overlap Complement to Ligand Binding Site Electrostatic Potential Maps" *Journal of Chemical Information and Modeling* **2018**, 58, 1836.
- Janesko, B. G.; **Jones, S. I.** "Quantifying the Delocalization of Surface and Bulk F-centers" *Surface Science* **2017**, 659, 9.

AWARDS, RECOGNITION, AND CERTIFICATES

- **National Science Foundation (NSF) Grant**……**2016-2017**

- Division of Materials Research (DMR) Grant # 1305343

- **Louis Stokes Alliance for Minority Participation (LSAMP)**……**January 2013-2014**

- Undergraduate Fellowship, supported by the National Science Foundation (NSF).

- **Minority Biomedical Research Support Program-Research Initiative for Scientific Enhancement (MBRS-RISE)**……**September 2012-2014**

- Undergraduate Fellowship, supported by National Institute of Health (NIH) and National Institute of General Medical Sciences (NIGMS).

- **The National Tutoring Association**……**January 2012**

- Underwent training and became certified at basic level of tutoring.

TEACHING EXPERIENCE

- **Teaching Assistant……2014-2017**

- Taught general, organic, and physical chemistry labs. Responsibilities included lecturing, chemical preparations, inventory, and equipment upkeep of spectrophotometers, infrared spectroscopy, refractometer and pH meters.

• **CRSP (College Re-Admittance Success Program) Mentor**……..**2012**
- Mentoring students that have been on academic probation. Helping students learn study skills, good habits, and guidance to success.

• **Math Tutor**……..**2011**
- Tutoring students that required additional help in math at the Teaching and Learning Center (TLC) at Pasadena City College (PCC).

COMMUNITY AND PROFESSIONAL SERVICES

• **Georgia Institute of Technology**……..**April 2019**
- Spring Undergraduate Research Symposium judge.

• **Oak Crest Institute of Science, Pasadena, California**……..**August 2012**
- Public Policy Intern: Attended research conferences, wrote newsletter articles, and interviewed researchers.

• **Faith Community Church**……..**2007-2008**
- Math Tutor to students from elementary to high school.

ABSTRACT

ORBITAL OVERLAP AS A TOOL IN COMPUTATIONAL DESIGN OF MOLECULES

by

Stephanie Ing Jones, Ph.D.

Department of Chemistry and Biochemistry

Texas Christian University

Dissertation Advisor: Benjamin G. Janesko, Associate Professor

There are two different aspects to this project, design and tool, which are later combined and applied to a final project to study protein binding sites.

The first design aspects focus on building a receptor that will bind to trimethylamine N-oxide (TMAO) in the body. Studies have shown that higher levels of TMAO increases the risk of cardiovascular diseases (CVD) and atherosclerosis. Here, a receptor is computationally built based on the binding between TorT and TMAO.

The second design project is on lignin. In the pulping and biorefinery processes, lignin is disposed as a waste stream after lignocellulose's other components (hemicellulose and cellulose) are extracted. Substituent effects of β-o-4 linkages were studied on lignin models along two reaction pathways: SN2 and E1.

The last design project is on warfarin, which is the medicine of choice when it comes to treating blood clots since 1955. One of the many reasons for dosage variance is due to the makeup of each patient's vitamin K epoxide reductase (VKOR) enzyme, which has been shown to

mutate. To this day, human VKOR (hVKOR) has not been crystallized, and there have been ongoing disagreements on its structure and warfarin's binding site. This study predicts hVKOR structure and the docking site for warfarin.

The tool used here is the orbital overlap distance function $D(\vec{r})$, which quantifies the size of molecular orbitals of system being studied. Chemically hard species with tightly bound electrons tend to have a smaller $D(\vec{r})$ than a soft species with loosely bound electrons. Here, the orbital overlap distance is tested on F centers, which are singly occupied electron systems.

Orbital overlap distance function is also applied onto protein binding sites and ligands. Typical analysis of protein molecules involves electrostatic and hydrophobic interaction maps, but the addition of the orbital overlap distance would be a useful tool to rationalize noncovalent interactions in a protein's active site. In this study, a combination of $D(\vec{r})$ and molecular electrostatic potential are used to rationalize noncovalent interactions in a protein's active site. These studies can then be applied to designing an alternative drug to warfarin; where $D(\vec{r})$ can justify whether warfarin and vitamin K 2,3-epoxide would bind to the predicted binding site.