University of Alberta

Host/Guest Chemistry:
From Rings and Metals to Proteins and Drugs

by

Melissa May Gajewski

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

©Melissa May Gajewski
Fall 2011
Edmonton, Alberta

To

My mother Susan, my father Richard, & my brother Michael

Abstract

The sequestering of alkali and alkaline earth cations by host structures (ethylenediamine tetraacetate, 18-crown-6, and hexaaza[18]annulene) was accomplished using density functional theory and either all electron or model core potential basis sets. The sequestering of alkali and alkaline earth cations by host structures (ethylenediamine tetraacetate, 18-crown-6, and hexaaza[18]annulene) was accomplished using density functional theory and either all electron or model core potential basis sets. These calculations allow for an understanding of the geometric position in which the cation bind to the host structure and the stabilization energy gained when forming the host/guest structures. Some solvation effects were taken into account for these calculations by either using polarizable continuum model or the $\Delta\Delta E$ equation described in Chapter 4. The knowledge gained from these types of studies could aid in choosing the correct host molecule to remove specific cation(s) in industrial, commercial, and/or laboratory procedures.

Calculating a slice of the potential energy surface for noscapine was accomplished using quantum mechanics using semi-empirical method and rotating the molecule along a dihedral angle. Energetic barriers in the rotation along this dihedral angle can provide an understanding of the possible binding conformations.

A matrix of binding affinities of different isotypes of tubulin (ten isotypes or mutants and 1TUB from the PDB) to derivatives of peloruside (twenty five derivatives) and laulimalide (eighteen derivatives) was created. The binding affinity calculations were accomplished using molecular mechanics, molecular dynamics, and quantum mechanics methods. Several de novo drug designs have predicted improvement in tubulin binding affinities.

Acknowledgements

While working at the University of Alberta on my PhD projects, I spent the majority of my time working within the group of Mariusz Klobukowski. I must first thank Mariusz for being a superb supervisor, who inspired me to challenge myself in all aspects of my life and research. I feel privileged to have had one of the best supervisors. I would also like to extend acknowledgments to all of the members past and present in the group, who have been there to answer any questions or to just be there as friends and confidants: John Lo, Jonathan Mane, Tao (Toby) Zeng, Evan Kelly, Amelia Fitzsimons, Jesse Kadosh and Ahmed Taha Ayoub. My sincerest gratitude to Toby, who has been an exceptional friend throughout my entire PhD.

My co-supervisor, Jack Tuszynski, is someone who has a million ideas and always encourages open discussion. It has been a privilege to work with him and participate in many stimulating discussions. Thank you to the entire Tuszynski group members: Aditya Saha, Abdorreza Samarbakhsh, Md Ashrafuzzaman, Chih-Yuan Tseng, Douglas Friesen, Gerry Leenders, Ishwar Hosamani, Jonathan Mane, Jur Massop, Khaled Barakat, Laleh Alisaraie, Marc St. George, Philip Winter, Travis Craddock, Rolando Perez, Holly Freedman, Torin Huzil, and Tyler Luchko, for all the great conversations we have had.

Thank you to all my friends and peers at the University of Alberta, who have help me throughout my time at the University of Alberta. Most importantly, I would like to address my friends and family. Lindsey Krawchuk has been my best friend my whole life and her love and support has helped me immensely during my PhD and I am so grateful to have her in my life. I have always been able to rely on all of my friends for whatever support I require. I am indebted to you all and from the bottom of my heart thank you for being there for me. My family: Susan, Richard, and Michael, you have been the foundation that has allowed me to be where I am today. There is nothing that I cannot accomplish because you are there for me. My Thesis is only one example of the many accomplishments that you have helped me with. Please know that I am eternally grateful.

Thank you all for everything you have done to help and guide me through my PhD.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Symbols

6aa	Hexaaza[18]annulene
18c6	18-crown-6
ΔE	Change in energy
ΔG	Change in Gibbs free energy
μ_i	Chemical potential
AE	All-electron
AM1	Austin Model 1
AMBER	Assisted model building with energy refinement
B3LYP	Becke exchange functional and
	Lee, Yang, and Parr correlation functional
BA	Binding affinity
BO	Born-Oppenheimer
BSSE	Basis set superposition error
CAS	Complet active space
CC	Coupled cluster
CI	Configuration interaction
CNDO	Complete neglect of diatomic overlap
COSMO	Conductor-like screening model
CPU	Central processing unit
DFT	Density functional theory
DNA	Deoxyribonucleic acid
DZP	Double zeta polarization
E	Energy
ECP	Effective core potential
GAMESS	General atomic and molecular electronic structure system
GDP	Guanosine diphosphate
GP86	Gill exchange functional and
	Perdew correlation functional
GTP	Guanosine triphosphate
$\hat{H}$	Hamiltonian operator
HF	Hartree Fock
HOMO	Highest occupied molecular orbital
ILA	Isolaulimalide

KS	Kohn-Sham
LAU	Laulimalide
LCAO	Linear combination of atomic orbitals
LUMO	Lowest unoccupied molecular orbital
MCP	Model core potential
MD	Molecular dynamics
MM	Molecular mechanics
MM-GBSA	Molecular mechanics generalized Born surface area
MNDO	Modified neglect of differential overlap
MO	Molecular orbital
MOLDEN	Pre- and post processing of molecular and electronic structure
MOPAC	Molecular orbital package
MacMolPlt	graphics for plotting 3-D molecular structures and normal modes
N	Number of particles or atoms
NBO	Natural bond orbital
NDDO	Neglect of diatomic differential overlap
NPA	Natural population analysis
P	Pressure
PCM	Polarizable continuum model
PEL	Peloruside
PES	Potential energy surface
pol β	Polymerase beta
PDB	Protein Data Bank
PT	Perturbation theory
QM	Quantum mechanics
RHF	Restricted Hartree Fock
RM1	Recife Model 1
RNA	Ribonucleic acid
SCF	Self-consistent field
SE	Stabilization energy
STO	Slater-type orbital
SVall	Split-valence all electron
T	Temperature
TZP	Triple zeta polarization
V	Volume
VMD	Visual molecular dynamics
WFT	Wave function theory
ZPE	Zero point energy
$EDTA^{4-}$	Ethylenediamine tetraacetate
H_4EDTA	Ethylenediamine tetraacetic acid

Chapter 1

Introduction

1.1 General Overview

Computational chemistry has become a fundamental tool in modern chemistry. Computer programs and algorithms have been extensively developed over the past eighty years and, along with major computer hardware advancements, have allowed for many applications of computational chemistry to be exploited. Such applications include molecular minimization or optimization, frequency calculations, thermodynamics, and molecular dynamics. Using computational chemistry it is possible to obtain insight into chemistry that would either be too expensive or too difficult to do in the laboratory. There are two main classes of computational chemistry, based on quantum mechanics and classical mechanics. Quantum mechanics (QM) method development has been divided into two types which are Hartree-Fock (HF) and post-HF, or within the scope of this Thesis wave function theory (WFT) and density functional theory (DFT). Background information on the quantum mechanics methods used is given in Chapter 2. Classical mechanics employs methods such as molecular mechanics (MM) and molecular dynamics (MD). These methods are utilized for larger molecules when the more accurate quantum mechanical methods are still unattainable within the current capabilities of computer hardware. Details of molecular mechanical and molecular dynamical meth-

ods are given in Chapter 3.

The utilization of many computational programs (GAMESS [1], Gaussian 03 [2], MOPAC [3], and AMBER [4]), as well as many visualization programs (MOLDEN [5], MacMol-Plt [6], VMD [7], and Mercury [8]) was necessary in order to study the chemistry discussed in this Thesis and due to the nature of the problems that are being discussed many levels of theory were used. Two main types of host/guest problems are addressed in this Thesis. The first type of host/guest chemistry investigated involves ligands/metals interactions and in these studies quantum mechanical methods were used. The second type of host/guest chemistry that was investigated probes the interaction between proteins and drugs. The anti-cancer drugs were studied at both the quantum mechanical and molecular mechanical level, while the interactions between the protein (α/β-tubulin dimer) and the drugs were calculated using molecular mechanics/dynamics methods due to the large molecular size of the α/β-tubulin dimer.

The ligands/metal sections will focus on calculations performed using both the tetradentate chelating ligand, ethylenediamine tetraacetate ($C_{10}H_{12}N_2O_8^{4-}$, Figure 1.1a) and the ring structures, 1,4,7,10,13,16-hexaoxacyclooctadecane (for simplicity the hexaoxa complex will be referred to as 18-crown-6, $C_{12}H_{12}N_6$, Figure 1.1b) and 1,4,7,10,13,1-hexaazacyclooctadeca-1,3,5,7,9,11,13,15,16-nonaene (for simplicity the hexaaza complex will be referred to as hexaaza[18]annulene, $C_{12}H_{12}N_6$, Figure 1.1c). The protein/drug sections of this Thesis will focus on two distinct projects: the first involves calculations of a slice of the potential energy surface of noscapine (Figure 1.2a) enantiomers and derivatives and the second project calculates the change in Gibbs free energy (ΔG) of laulimalide (Figure 1.2b), peloruside (Figure 1.2c), and both known and de novo designed derivatives (derivatives are depicted in Appendices E and G, (see Supplementary Data)).

(a)
EDTA^{4-}

(b)
18-crown-6

(c)
Hexaaza[18]annulene
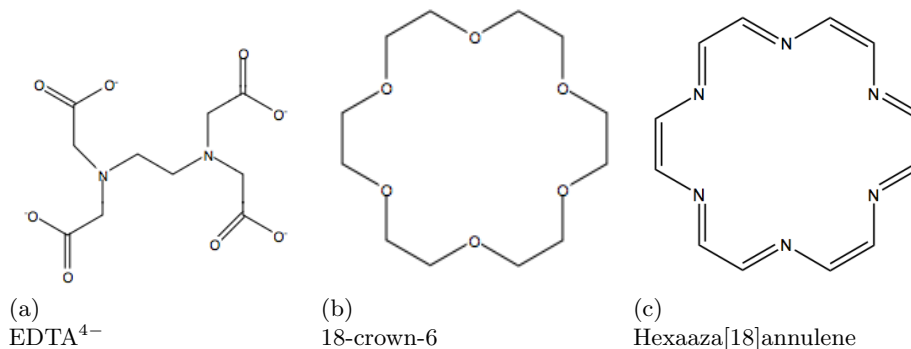
Figure 1.1: Cation capturing ligands

(a)
Noscapine

(b)
Laulimalide

(c)
Peloruside
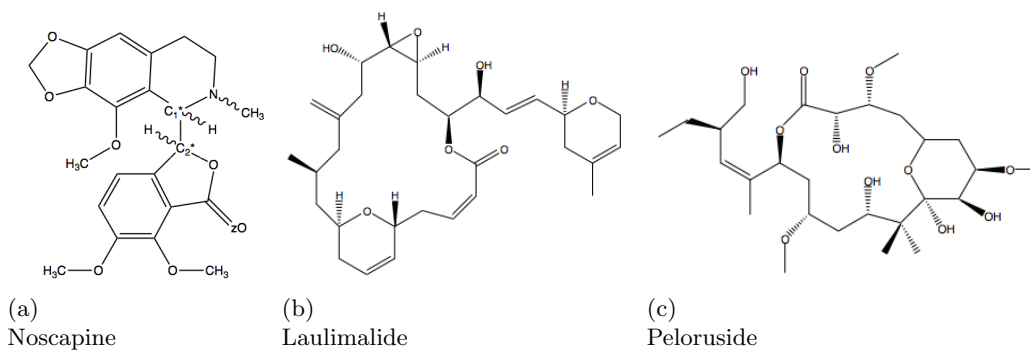
Figure 1.2: Antimitotic drugs

Additionally, deoxyribonucleic acid (DNA) polymerase beta (pol β) inhibitors were re-viewed. For modern cancer therapies, such as ionizing radiation and certain classes of chemotherapy drugs target DNA, base excision repair is a fundamental pathway respon-sible for the elimination of damaged DNA. Pol β is an essential enzyme to achieve base excision repair and is considered to be a potential therapeutic target. This work was done with Khaled Barakat and is not included within this Thesis. The interested reader is referred to a paper to be published entitled "DNA polymerase beta (pol β) inhibitors: a comprehensive overview", which has been submitted to *Current Pharmaceutical Biotech-nology.*

1.2   Host/Guest Chemistry

Host/guest chemistry describes molecular complexes that are composed of two or more components; these components are held together by forces other than full covalent bonds. These noncovalent interactions include ionic bonding, hydrogen bonds, van der Waals forces and hydrophobic interactions. The noncovalent bonding allows for molecular recognition of the guest molecule to the host molecule. Host/guest chemistry has many beneficial properties and these have been exploited in the laboratory, industry and even within our bodies. Some examples: sequestering heavy metal cations, buffering solutions, titrations, industrial cleaners, conformational changes (such as in the case of hemoglobin which allows for oxygen transport), and many more [9]. The host component is usually considered to be the larger molecule and it interacts with the smaller guest molecule/atoms .

1.3   Project Summary

The first project that will be discussed uses DFT to study the complexation reaction between $EDTA^{4-}$ and alkali/alkaline earth cations M (where M = Na^+, K^+, Rb^+, Mg^{2+}, Ca^{2+}, and Sr^{2+}). Full details of this project are given in Chapter 4 and the optimized coordinates are given in Appendix A, (see Supplementary Data). This project investigated the binding affinities of several cations with the tetradentate $EDTA^{4-}$ molecule. The DFT calculations for $MEDTA^{n-4}$ and $M(H_2O)_6{}^{n+}$ (where n is the charge of the cation) were carried out at both all-electron (AE) and the model core potential (MCP) pseudopotential levels using two DFT functionals. Results from these four methods were compared and it was found that the ions bind with $EDTA^{4-}$ with selectivity of $Mg^{2+}>Ca^{2+}>Sr^{2+}>Na^+>K^+>Rb^+$. This work was published as an invited paper in the Canadian Journal of Chemistry in 2009, [10] and was presented at the 16th Cana-

dian Symposium on Theoretical Chemistry in St. Johns, Newfoundland (2007) and at the Warburg International Summer School and Workshop on Computational Systems Biology in Berlin, Germany (2007).

The second project pertains to DFT calculations for M-6aa and M-18c6 (where M = Rb^+, Cs^+, Sr^{2+}, and Ba^{2+}). Full details of this project are given in Chapter 5 and the optimized coordinates are given in Appendix B, (see Supplementary Data). The DFT studies of the electronic structure and geometry of 18-crown-6 (18c6), hexaaza[18]annulene (6aa), $6H_2O$ and their complexes with Rb^+, Cs^+, Sr^{2+}, and Ba^{2+} were performed using MCP pseudopotential parametrization. Results from this study show that the ions bind more strongly to 6aa than to the 18c6 moiety and the difference is more pronounced for cations with smaller radii. This work was published in Inorganica Chimica Acta in 2008 [11] and was presented at the 6th Canadian Computational Chemistry Conference in Vancouver, British Columbia (2006).

In order to make the switch from the chelating agents $EDTA^{4-}$, 18-crown-6, and hexaaza[18]annulene to anti-cancer drugs, it is necessary to provide the reader with some background information about the cell, specifically microtubules. This background is imperative in order to discuss the results of the project pertaining to noscapine, peloruside, and laulimalide. Therefore, background information of anti-cancer drugs and cancer cells will be briefly discussed in Chapter 6.

The third project focuses on studies of geometric conformations of the antitussive (cough suppressant) and antimitotic (prevents or disrupts mitosis) agent noscapine. Noscapine, due to the presence of chiral carbons (shown in Figure 1.2a as C_1^* and C_2^*), has four possible enantiomers and under acidic conditions the protonation of the nitrogen would generate the proton at either the axial or equatorial position. Therefore, there are eight

systems that are considered in this investigation, SS-ax, SS-eq, SR-ax, SR-eq, RS-ax, RS-eq, RR-ax, and RR-eq. For each enantiomer, five derivatives were generated, which produced forty molecules whose energies were optimized for several fixed values of the dihedral angle C-C_2^*-C_1^*-N. In this study the reparameterization of the Austin Model 1 (AM1) [12] semiempirical molecular orbital method, the Recife Model 1 (RM1) [13] was used. This study calculated a slice of the potential energy surface along a dihedral angle (C-C_2^*-C_1^*-N) that connects the two ring structures. Full details of this project are given in Chapter 7 and the optimized coordinates for a dihedral angle of 180° are given in Appendix C, (see Supplementary Data). This work is part of a larger project and will be published when all components of the project have been completed.

The last computational project involves two new antimitotic drugs peloruside and laulimalide and full details of this project are given in Chapter 8 and in Appendices E and G, (see Supplementary Data). Additionally, optimized coordinates are given in Appendices D and F, (see Supplementary Data). Both peloruside and laulimalide have shown to be promising antimitotic drugs for future clinical use. Unfortunately, both peloruside and laulimalide present some unique problems as antimitotic agents in their parent form. For example, peloruside is extremely difficult and expensive to synthesize, and the epoxide ring on laulimalide opens up in slightly acidic conditions (not unlike the conditions found in a tumor, where the pH drops from approximately 7.3 in a healthy cell down to 6.5 in a tumor cell) to form isolaulimalide, which is far less potent than laulimalide. The drawbacks of the parent compounds create a need for new derivatives to be developed. Computationally developing and testing new derivatives enables a broader range of drugs to be tested without the large cost of synthesis impeding the search for a better antimitotic agent. Molecular dynamics calculations were done on ten isotypes or mutants and 1TUB [14] from the PDB of α/β-tubulin (1TUB, βI, βIIa, βIIb, βIII, βIVa, βIVb, βV, βVI, βVII, and βVIII) with peloruside A and B, laulimalide, isolaulimalide,

and both known and de novo designed derivatives. The values of ΔG were calculated for all forty-three compounds with the ten isotypes or mutants and 1TUB from the PDB of α/β-tubulin to form a binding affinity matrix. These forty-three drug molecules show different selectivities to the α/β-tubulin isotypes and these selectivities can be exploited to kill cancerous cells that overproduce one or more of the isotypes. This work was presented at the 91st Canadian Society of Chemistry Conference in Edmonton, Alberta (2008), the 7th Canadian Computational Chemistry Conference in Halifax, Nova Scotia (2008), the Alberta Cancer Research Institute in Banff, Alberta (2009), and the 17th Canadian Symposium on Theoretical Chemistry in Edmonton, Alberta (2010). This work will be published in the future.

Finally, Chapter 9 of this Thesis includes a discussion of future work and general conclusions. Additional details and results are collected in a separate booklet Supplementary Data; for the peloruside and laulimalide projects additional details can be found in Appendices E and G, and optimized cartesian coordinates for all of the projects can be found in Appendices A, B, C, D, and F, (see Supplementary Data). All necessary information pertaining to the implementation of the semi-empirical parameters of the RM1 method into the GAMESS program is outlined in Appendix H, (see Supplementary Data). The Supplementary Data preprint is available upon request from Melissa Gajewski (mmg8@ualberta.ca) or Mariusz Klobukowski (mariusz.klobukowski@ualberta.ca).

The majority of the calculations to determine the protein/drug binding were done on the PharmaMatrix cluster. According to the PharmaMatrix website (https://www.pharmamatrix.ualberta.ca/), the cluster "consists of 128 nodes of 64 Hewlett Packard ProLiant BL2x220c G5 double density blades. Every node in the PhamaMatrix cluster has 8GB of RAM and two Intel Xeon E5420 quad-core processors running at 2.5Ghz for a total capacity of 1024 cores and 1TB of RAM. All nodes are interconnected

by 4X DDR InfiniBand and gigabit Ethernet. Storage for research data is provided by a Sun Microsystems X4540 with 13TB of online storage, and backups are done with a Sun SL48 tape library and Bacula."

Chapter 2

Quantum Mechanical Methods

The present Chapter summarizes the most important concepts, equations, and derivations that were used in the remainder of the Thesis. The details described in this section use derivations of formulas that can be found in many quantum chemistry text books. The references that were primarily used in this section include books by Szabo and Ostlund [15], McQuarrie and Simon [16], Schatz and Ratner [17], Cramer [18], and Thiel [19].

2.1   Hartree-Fock: an all electron method

A major challenge in quantum chemistry is determining the exact solution to the Schrödinger equation

$$\hat{H}\Psi(\{R_A\}, \{r_i\}) = E\Psi(\{R_A\}, \{r_i\}), \tag{2.1}$$

which would give the wavefunction $(\Psi(R_A, r_i))$ and the energies (E). The Hamiltonian operator $(\hat{H})$, given in Equation (2.2) in atomic units, describes nuclei and electrons in terms of their positions, $\{R_A\}$ and $\{r_i\}$, respectively.

$$\hat{H} = -\underbrace{\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2}_{\hat{T}_N} - \underbrace{\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2}_{\hat{T}_e} - \underbrace{\sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}}}_{\hat{V}_{Ne}} + \underbrace{\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}}}_{\hat{V}_{ee}} + \underbrace{\sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}}_{\hat{V}_{NN}} \tag{2.2}$$

Equation (2.2) contains 5 terms; $\hat{T}_N$ and $\hat{T}_e$ are the kinetic energy operators for the nuclei and electrons, respectively, where M_A represents the mass of the nucleus A. $\hat{V}_{Ne}$ represents the Coulomb attraction between the electrons and the nuclei, where Z_A is the nuclear charge and r_{iA} is the distance between electron i and nucleus A. $\hat{V}_{ee}$ and $\hat{V}_{NN}$ are the Coulomb repulsion between electrons i and j separated by distance r_{ij} and nuclei A and B separated by distance R_{AB}.

The Schrödinger equation is very difficult to solve and its solutions can only be obtained for very few problems. Approximations must be made in order to simplify Equation (2.1). The first approximation is known as the Born-Oppenheimer (BO) approximation, according to which the true molecular wavefunction can be approximated as

$$\Psi(R_A, r_i) = \Psi_{el}(r_i; R_A)\Psi_N(R_A), \tag{2.3}$$

where the electronic (Ψ_{el}) and nuclear (Ψ_N) wavefunctions are separated. The BO approximation is acceptable because electrons are significantly lighter than nuclei and thus move more rapidly than the nuclei. In other words, during the fast electronic motion, the nuclear motion is relatively slow and the nuclear positions do not change significantly. Thus, the nuclear motion contribution is negligible and the nuclear positions can be considered as stationary. Within the BO approximation $\hat{T}_N$ can be neglected and $\hat{V}_{NN}$ is considered to be constant, since the positions of the nuclei are fixed in space. Therefore, the Schrödinger equation can be rewritten as the electronic Schrödinger equation,

$$(\hat{H}_{el} + \hat{V}_{NN})\Psi_{el} = E_{total}\Psi_{el}, \tag{2.4}$$

where

$$\hat{H}_{el} = \sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{2.5}$$

$$\hat{V}_{NN} = \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.6}$$

$$E_{total} = E_{el} + V_{NN}. \tag{2.7}$$

By neglecting V_{NN} in Equation (2.2), a purely electronic Schrödinger equation is obtained,

$$\hat{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{2.8}$$

Ψ_{el} is the electronic wavefunction which describes the motion of N electrons in a field of M nuclei. Both Ψ_{el} and E_{el} depend on the nuclear positions:

$$\Psi_{el} = \Psi_{el}(\{R_A\}) \tag{2.9}$$

$$E_{el} = E_{el}(\{R_A\}). \tag{2.10}$$

Even with the simplified electronic Hamiltonian operator given in Equation (2.5), the electronic Schrödinger equation is still very difficult to solve and exact solutions only exist for systems that have only one electron. One needs to introduce the second approximation leading to the Hartree-Fock (HF) method. This method allows for the transformation of the N-electron equation into N one-electron equations.

For a closed-shell molecular system, the ground-state Hartree-Fock wavefunction is given

by the Slater determinant based on N spinorbitals, $\psi_i(\mathbf{x}_j)$

$$\Psi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_1(\mathbf{x}_2) & \dots & \psi_1(\mathbf{x}_N) \\ \psi_2(\mathbf{x}_1) & \psi_2(\mathbf{x}_2) & \dots & \psi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{x}_1) & \psi_N(\mathbf{x}_2) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}$$

where $\mathbf{x}$ contains the space component (r), and the spin coordinate (σ). In the determinantal representation (Ψ_0) each column refers to the jth electron described by different spinorbitals, and each row refers to the ith spinorbital describing each electron. A spinorbital $\psi_i(\mathbf{x}_j)$ is a product of both space (ϕ) and spin functions $(\alpha$ and $\beta)$,

$$\psi_i(\mathbf{x}_j) = \phi_i(r_j) \times \begin{cases} \alpha(\sigma_j) \\ \\ \beta(\sigma_j) \end{cases}. \tag{2.11}$$

According to the variational principle, the spinorbitals are obtained by minimizing the ground state energy (E_0),

$$E_0 = \langle \Psi_0 | \hat{H}_{el} | \Psi_0 \rangle. \tag{2.12}$$

Equation (2.12) is presented in the Dirac notion, which will be used throughout this Chapter. Consider a well behaved functions $\{f_i(x)\}$ and and operator $\hat{A}$, then

$$\langle m | n \rangle \equiv \int f_m^* f_n d\tau, \text{ and} \tag{2.13}$$

$$\langle m | \hat{A} | n \rangle \equiv \int f_m^* \hat{A} f_n d\tau, \tag{2.14}$$

where * indicates the complex conjugate for the arbitrary function $\{f_m(x)\}$ and τ is the integration variable in which the integrating is over. In the case of Equation (2.12), the integration is over both the space and spin component.

After substitution of the determinantal wavefunction Ψ_0 into Equation (2.12) the energy E_0 becomes a functional of the spinorbitals. Obtaining the spinorbitals that minimize E_0 determines the best possible approximation to the ground state of the N-electron system described by the electronic Hamiltonian. This reduces the N-electron problems to a set of one-electron equations, which are known as the Hartree-Fock equations,

$$\hat{F}(\mathbf{x}_1)\psi_i(\mathbf{x}_1) = \epsilon_i\psi_i(\mathbf{x}_1) \qquad i = 1, 2, \ldots, N. \qquad (2.15)$$

The Fock operator ($\hat{F}(\mathbf{x}_1)$) depends on both space (r) and spin (σ); by using the orthonormality of the spinorbitals the HF equations can be written in terms of the spatial orbitals only,

$$\hat{F}(1)\phi_i(1) = \epsilon_i\phi_i(1) \qquad i = 1, 2, \ldots, N, \qquad (2.16)$$

where the label '1' represents the spatial coordinates of a single electron (r_1). $\hat{F}(1)$ represents the one-electron Fock operator, $\phi_i(1)$ is the ith spatial orbital and ϵ_i is the orbital energy. The Fock operator can be written as

$$\hat{F}(1) = \hat{h}^{core}(1) + \hat{v}^{HF}(1), \qquad (2.17)$$

where $\hat{h}^{core}(1)$ is the core Hamiltonian, which describes the motion of a single electron in the field of all nuclei, and $\hat{v}^{HF}(1)$ is the HF potential energy operator, which describes the electron-electron interaction in an averaged electronic distribution. $\hat{h}^{core}(1)$ is given by

$$\hat{h}^{core}(1) = \underbrace{-\frac{1}{2}\nabla_1^2}_{\hat{T}_e} - \underbrace{\sum_{A=1}^{M}\frac{Z_A}{r_{1A}}}_{\hat{V}_{Ne}}, \qquad (2.18)$$

which has the kinetic energy operator ($\hat{T}_e$) for one electron and a potential energy operator ($\hat{V}_{Ne}$) for the attraction between one electron and the nuclei.

For a closed-shell system the interelectron potential operator can be given as

$$\hat{v}^{HF}(1) = \sum_{j=1}^{\frac{N}{2}} [2\hat{J}_j(1) - \hat{K}_j(1)], \tag{2.19}$$

where $\hat{J}_j(1)$ is the Coulomb operator and $\hat{K}_j(1)$ is the exchange operator. The sum over j runs over the occupied spatial orbitals. The Coulomb and exchange operators are defined as

$$\hat{J}_j(1)\phi_i(1) = \left[\int \phi_j(2) \frac{1}{r_{12}} \phi_j(2) dr_2 \right] \phi_i(1) \tag{2.20}$$

and

$$\hat{K}_j(1)\phi_i(1) = \left[\int \phi_j^*(2) \frac{1}{r_{12}} \phi_i(2) dr_2 \right] \phi_i(1). \tag{2.21}$$

The Coulomb operator gives the repulsion experienced by an electron due to the averaged distribution of the others. The exchange operator stems from the Fermion statistics which describes the energy of single particles in a system and is comprised of many identical particles, such as electrons and has no classical analogue.

The Fock operator (given in Equation (2.17)) which is used to calculate the energy using Equation (2.16) depends on its own eigenfunctions through Equations (2.20) and (2.21). Thus, solving the HF equations requires knowledge of the orbitals. For this reason the HF equations are solved iteratively, which is usually done by using an initial guess of the wavefunction and then solving the equation and taking the new wavefunction as the next guess wavefunction. This procedure is repeated until the new wavefunction does not differ significantly from the wavefunction from the previous step of iteration. This method is known as the self-consistent field (SCF) method. Using the HF all-electron (AE) calculations provides useful insight into chemical properties, but the computational cost of AE calculations can sometimes be too great for very large molecules and additional approximations are needed to speed up the calculations.

2.2   Valence electron only methods

Chemical characteristics, atomic properties, and bonding capabilities are directly related to the valence electrons while the core electrons, which are more tightly bound to the nuclei, only provide the core-valence potential for the valence electrons to move in. Hellmann [20, 21] and Gombás [22] in mid 1930's suggested to replace the core-valence electron interaction by a potential. Many methods have since been developed and a brief overview of three of them is given below: effective core potentials (ECP), model core potentials (MCP), and semiempirical (SE) methods.

2.2.1   Effective Core Potentials

In 1976, Kahn et al. [23] introduced a method to generate effective core potentials. The ECP for an atom is obtained by

$$V_l^{eff} = \epsilon_l + \frac{Z_{eff}}{r} - \frac{l(l+1)}{2r^2} + \frac{(\frac{1}{2}\nabla^2 - V'_{val})\chi_l}{\chi_i}, \tag{2.22}$$

where l indicates the angular momentum, V_l^{eff} is the effective potential, Z_{eff} is the effective nuclear charge (nuclear charge minus the number of core electrons), V'_{val} is the valence orbital interaction potential between pseudo-orbitals (pseudo-orbitals are the orbitals that are created by smoothing the valence orbital, see Figure 2.1), r is the radial variable, χ_l is the pseudo-orbital constructed from the Hartree-Fock orbitals, and ϵ_l is the eigenvalue energy of the original HF orbital.

The ECP reduces the number of electrons that are explicitly calculated, which reduces the amount of computation time required and allows for calculations of larger systems that would otherwise be unattainable. When the radial functions of AE and ECP calculations are compared, ECP performs quite well at long distances, but ECPs do not reproduce the nodal structure at short distances (see Figure 2.1). Therefore, ECP offers

a reduction in the computational time and is a good approximation for most chemistry (chemistry that takes place in the outer range of r). The nodal structure becomes increasingly important when the investigated property (such as spin-orbit coupling) is dependent on the orbital shape close to the nucleus and, hence, another method must be employed.

2.2.2 Model Core Potentials

The foundations of MCP method were presented by Bonifacic and Huzinaga in 1974 [24]; a review of the method can be found in reference [25]. There have been several versions of MCP and the details of these variants are given in references [24, 26, 27] and the more recent work done on MCP's with spin-orbit coupling (SOC) by Zeng et al. has been summarized [28–31].

Both MCP and ECP reproduce the shape of valence orbitals at larger radial distances but only the MCP plot overlaps the entire AE radial function, as apparent from Figure 2.1. The core potential in the MCP method is represented by

$$V^{core} = -\frac{Z - N_c}{r_i} \left[1 + \sum_l A_l r_i^{n_l} \exp(\alpha_l r_i^2) \right] \tag{2.23}$$

where Z is the nuclear charge, N_c is the number of core electrons, r_i is the position of electron i, A_l and α_l are adjustable parameters, and n_l is 0 or 1.

The MCP method is an improvement over the ECP method due to the correct nodal structure over the entire AE radial function and, like ECP, MCP provides AE-like results using a fraction of the computational time. The reduction in computational time allows for larger systems to be studied. However, if even larger systems are of interest, another approach to the valence-electron methodology is needed.
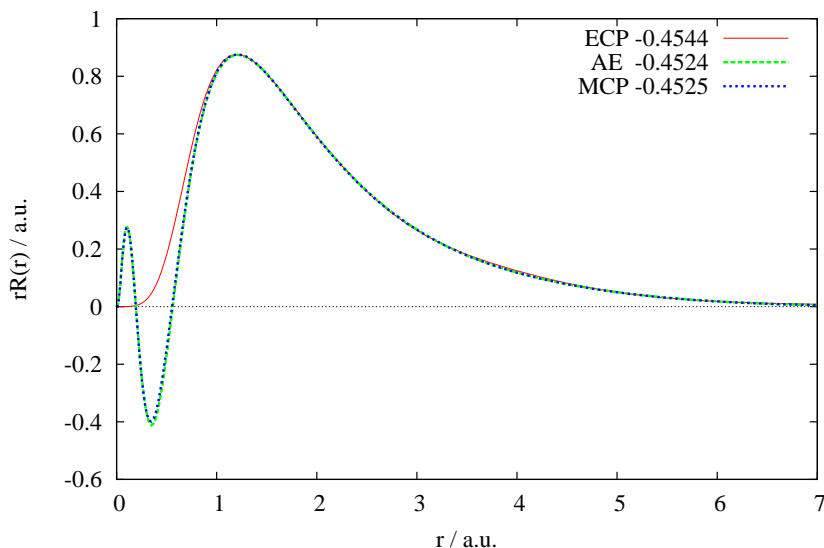
Figure 2.1: The radial functions and orbital energies of the Au 5d orbital from the AE, ECP and MCP calculations. The AE calculation was at the DK3 relativistic level with uncontracted well-tempered basis functions [32]. The ECP calculation was performed using the ECP60MDF [33] potential and basis set. The MCP calculation was performed using the recently developed potentials (at the DK3 relativistic level) and basis set [34]. The orbital energies are listed for comparison. (Image produced and added with permission of Dr. T. Zeng)

2.2.3   Semi-empirical Methods

The semi-empirical approach to quantum chemistry is characterized by the inclusion of empirical parameters from either experimental results or high-level quantum mechanical calculations. Semi-empirical methods are based on the molecular orbital (MO) theory.

The goal of semi-empirical methods was to develop a method that reproduces experimental results while using only a minimal basis set and explicitly treating only valence electrons. Dewar and co-workers worked for many years developing semi-empirical methods that are still in use today. Two of the most popular are the MNDO (Modified Neglect of Diatomic Overlap) and AM1 (Austin Model 1) methods. Both of these methods are

based on the NDDO (Neglect of Diatomic Differential Overlap) formalism.

In order to consider NDDO, which is the basis for MNDO, AM1, and the RM1 methods described later in this Chapter, the self-consistent molecular orbital theory must first be discussed. Self-consistent field (SCF) theory optimizes the orbitals using the variational principle and uses the linear combinations of atomic orbitals (LCAO). The coefficients in the LCAO are chosen to minimize the energy, thus yielding the best linear combination of atomic orbitals for the self-consistent functions (LCAO SCF). In order for this method to be a useful computational method for very large molecules, further approximations must be made. Using LCAO SCF and neglecting the differential overlap, (i.e. neglecting the product of pairs of different atomic orbitals in certain electron-interaction integrals) as well as introducing empirical parameters, creates the basis for NDDO semi-empirical methods.

Using Roothaan's [35] LCAO SCF formalism to calculate MO's, the valence electrons are described by LCAO MOs using Equation (2.24)

$$\psi_i = \sum_{\nu=1}^{k} \phi_\nu c_{\nu i} \tag{2.24}$$

where ϕ_ν are the valence atomic orbitals, and ν denotes atomic orbitals. Substituting Equation (2.24) into the HF equations Equation (2.16) we obtain several equations that describe the LCAO MO formalism in the matrix form. The expanded matrix HF equations are:

$$\sum_{\nu=1}^{k} F_{\mu\nu} c_{\nu i} = \sum_{\nu=1}^{k} S_{\mu\nu} c_{\nu i} \epsilon_i, \quad \nu = 1, 2, ..., k \tag{2.25}$$

where ϵ_i is the orbital energy for the MO ψ_i, and the overlap integral $S_{\mu\nu}$ is defined by

$$S_{\mu\nu} = \int \phi_\mu^* \phi_\nu dr. \tag{2.26}$$

The Fock matrix element is given by,

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu} \tag{2.27}$$

and the $H_{\mu\nu}$ is the matrix element of the one-electron Hamiltonian,

$$H_{\mu\nu} = \int \phi_\mu^* \left[-\frac{1}{2}\nabla^2 - \sum_A V_A(r) \right] \phi_\nu dr, \tag{2.28}$$

where $V_A(r)$ is the sum of potentials for various atoms A in the molecule, and

$$G_{\mu\nu} = \sum_{\lambda=1}^{k} \sum_{\sigma=1}^{k} P_{\lambda\sigma} \left[\langle \mu\nu|\lambda\sigma \rangle - \frac{1}{2} \langle \mu\sigma|\nu\lambda \rangle \right] \tag{2.29}$$

$G_{\mu\nu}$ is the matrix element of the potential due to other valence electrons, which depends on the population matrix $P_{\lambda\sigma}$, and

$$\langle \mu\nu|\lambda\sigma \rangle = \int \int \phi_\mu^*(1)\,\phi_\nu(1)\,(r_{12})^{-1}\,\phi_\lambda^*(2)\,\phi_\sigma(2)\,dr_1 dr_2, \tag{2.30}$$

$$\langle \mu\sigma|\nu\lambda \rangle = \int \int \phi_\mu^*(1)\,\phi_\sigma(1)\,(r_{12})^{-1}\,\phi_\nu^*(2)\,\phi_\lambda(2)\,dr_1 dr_2, \tag{2.31}$$

and

$$P_{\lambda\sigma} = 2 \sum_{i}^{occ} c_{i\lambda}^* c_{i\sigma}. \tag{2.32}$$

The sum in Equation (2.32) is over occupied orbitals only, and formally ϵ_i are the roots of the secular equation,

$$|F_{\mu\nu} - \epsilon S_{\mu\nu}| = 0. \tag{2.33}$$

E_{el} is the total electronic energy of the valence electrons, and the total energy E_{tot} is found by adding in the repulsion energy between the cores, where Z_X is the core charge

on atom X and R_{AB} is the $A - B$ internuclear distance.

$$E_{el} = \frac{1}{2} \sum_{\mu=1}^{k} \sum_{\nu=1}^{k} P_{\mu\nu} \left(H_{\mu\nu} + F_{\mu\nu} \right) \tag{2.34}$$

$$E_{tot} = E_{el} + \sum_{A<B} \sum_{B} \frac{Z_A Z_B}{R_{AB}}, \tag{2.35}$$

Equations (2.25) to (2.35) form the basis of LCAO SCF calculations. Further assumptions are needed for the complete neglect of diatomic overlap (CNDO): ϕ_μ form an orthonormal set; two-electron integrals are neglected; electron interaction integrals depend on the atoms to which the orbitals belong and not the type of orbital; integrals $\langle \mu | V_B | \nu \rangle = 0$ if $\mu \neq \nu$ and ϕ_μ, ϕ_ν are not on atom A. The off-diagonal core matrix elements between orbitals on different atoms are calculated using

$$H_{\mu\nu} = \beta_{\mu\nu} = \beta^0_{AB} S_{\mu\nu}, \tag{2.36}$$

where $S_{\mu\nu}$ is the overlap integral and β^0_{AB} is the parameter depending on the nature of the atoms A and B. These assumptions are made to simplify the calculations and further details can be found in [36].

Neglect of Diatomic Differential Overlap (NDDO)

NDDO as the name indicates, neglects the differential overlap in two-electron integrals for atomic orbitals on different atoms (neglect the two electron integrals which depend on the overlapping of charge densities of different orbitals, unless both orbitals belong to the same atom, shown in Equation (2.30)). All products $\phi_\mu \phi_\nu$ for different orbitals on the same atom are retained. Starting from the LCAO SCF equations shown in the previous section (Equations (2.24) to (2.35)), the following assumptions are made.

Assumption 1:

$\{\phi_\mu\}$ are treated as if they are an orthonormal set, and therefore their overlap integrals ($S_{\mu\nu}$) are equal to zero unless $\mu = \nu$ in which case $S_{\mu\nu} = 1$. The coefficients ($c_{i\mu}$) form an orthogonal matrix and the orthonormality conditions for ψ_i become,

$$\sum_\mu c_{i\mu} c_{j\mu} = \delta_{ij}, \qquad (2.37)$$

where δ_{ij} is the Kronecker delta. Using this approximation, the diagonal matrix elements $P_{\mu\mu}$ are the electron populations of the atomic orbitals ϕ_μ, and

$$\sum_\mu P_{\mu\mu} = 2N \qquad (2.38)$$

where 2N is the number of valence electrons. $P_{\mu\nu}$ is the molecular orbital bond order of neighbouring atoms.

Assumption 2:

All of the two-electron integrals shown in Equation (2.30) depend on the overlapping charge densities of the basis orbitals on different atoms are neglected. This leads to $\langle \mu\nu | \lambda\sigma \rangle$ being equal to zero unless μ, ν belong to the same atom A and λ, σ are centered on atom B (where B could be the same atom as A). All other two-electron integrals are two-centered, with one electron on each atom.

With these two assumptions, the matrix elements are:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_B \sum_{\lambda,\sigma \epsilon B} P_{\lambda\sigma} \left[\langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} \sum_{\lambda,\sigma} \langle \mu\sigma | \nu\lambda \rangle \right] \qquad (2.39)$$

where μ and ν are both on A, and

$$F_{\mu\nu} = H_{\mu\nu} - \frac{1}{2} \sum_{\sigma \epsilon A} \sum_{\lambda \epsilon B} P_{\lambda\sigma} \langle \mu\sigma | \nu\lambda \rangle \qquad (2.40)$$

where μ is on A and ν is on B.

The core Hamiltonian is split into atomic parts, and the matrix elements between orbitals on the same atom are

$$H_{\mu\nu} = U_{\mu\nu} - \sum_{B \neq A} \langle \mu | V_B | \nu \rangle, \qquad (2.41)$$

where V_B is the sum of the kinetic energy and the potential energy in the electrostatic field of the core. The core matrix elements $H_{\mu\nu}$ are calculated using Equation (2.36), where the parameter β_{AB}^0 (which depend on the AB distance) are derived empirically to match either experimental data or calculated results.

Semi-empirical parameters

The semi-empirical methods allow for quick calculations to be done with reasonably accurate results. The semi-empirical calculations scale on the order of O^2 with respect to the size of the system, whereas other quantum computational methods scale with a much larger exponent. Parameters are used to compensate for the approximations used in the integral calculations. The parameters used are derived from experiment and/or high level quantum mechanical calculations; these parameters are described in detail in Appendix H, (see Supplementary Data) of this Thesis. It is the integral approximations made in semi-empirical methods that greatly speed up the calculations, because it is the two-electron integrals calculations that are computationally expensive. If only the previ-

ously described equations and assumptions were used within the semi-empirical method, the semi-empirical calculations would not give accurate or reliable results because of the underlying theory of neglecting or making approximations to some important integrals. It is therefore necessary to compensate for some of the vanishing integrals by introducing parameters. These parameters are optimized by adjusting them until the results of the calculations match the reference data. The accuracy of the semi-empirical methods for a set of conditions is dependent on the quality and quantity of the reference data used in the parameterization process.

The Modified Neglect of Diatomic Overlap (MNDO) Method

The MNDO method started with first row elements (H, C, N, and O) [37] and was later expanded to halogens and second row elements [38]. Finally it was expanded to include transition metals after the inclusion of d orbitals (MNDO-d) [39].

MNDO is similar to the NDDO method in respect to which integral are kept and which are discarded, but differs from NDDO in the manner in which the repulsion integrals are calculated. NDDO calculates both the one-center and two-center repulsion integrals analytically, but these integrals are calculated using spectroscopic experimental data or from semi-empirical expressions in the MNDO approximation. Spectroscopic experimental data, such as ionization potential and electron affinity, was used to calculate the one center repulsion integrals which allows for some correlation effects to be incorporated into the MNDO method, which is not taken into account in the NDDO method. The Fock matrix elements of MNDO method are shown in the following equations:

$$F_{\mu\mu} = \hat{h}_{\mu\mu}^{core} + \sum_{\nu \epsilon A} \left[P_{\nu\nu}\langle\mu\mu|\nu\nu\rangle + \frac{1}{2}P_{\nu\nu}\langle\mu\nu|\mu\nu\rangle \right] \tag{2.42}$$

$$+ \sum_{B \neq A}\sum_{\lambda \in B}\sum_{\sigma \in B} P_{\lambda\sigma}\langle\mu\mu|\lambda\sigma\rangle$$

where

$$\hat{h}_{\mu\mu}^{core} = U_{\mu\mu} - \sum_{B \neq A} V_{\mu\mu,B} \tag{2.43}$$

$$V_{\mu\mu B} = -Z_B\langle\mu\mu|s_B s_B\rangle,$$

where μ is located on atom A and s_B are valence orbitals.

$$F_{\mu\nu} = \hat{h}_{\mu\nu}^{core} + \frac{3}{2}P_{\mu\nu}\langle\mu\nu|\mu\nu\rangle - \frac{1}{2}P_{\mu\nu}\langle\mu\mu|\nu\nu\rangle \tag{2.44}$$

$$+ \sum_{B \neq A}\sum_{\lambda \epsilon B}\sum_{\sigma \epsilon B} P_{\lambda\sigma}\langle\mu\nu|\lambda\sigma\rangle;$$

μ and ν both on A, and

$$\hat{h}_{\mu\nu}^{core} = -\sum_{B \neq A} V_{\mu\nu B} \tag{2.45}$$

$$V_{\mu\mu B} = -Z_B\langle\mu_A\nu_A|s_B s_B\rangle$$

$$F_{\mu\lambda} = \beta_{\mu\lambda} - \frac{1}{2}\sum_{\nu}^{A}\sum_{\sigma}^{B} P_{\mu\nu}\langle\mu\nu|\lambda\sigma\rangle. \tag{2.46}$$

The Fock matrix elements contain one-center one-electron energies $U_{\mu\mu}$, one center two-electron repulsion integrals $\langle\mu\mu|\nu\nu\rangle = G_{\mu\nu}$, one center two-electron exchange integrals $\langle\mu\nu|\mu\nu\rangle = H_{\mu\nu}$, two-center one-electron core resonance integrals $\beta_{\mu\lambda}$ from Equation (2.46), two-center one-electron attraction term between electron distribution at atom A and the core of atom B ($V_{\mu\nu}$), and two-center two-electron repulsion integrals $\langle\mu\nu|\lambda\sigma\rangle$. These elements are determined from experimental data (heats of formation, dipole mo-

ments, ionization energies and molecular geometries) or from semi-empirical expressions [38] that are adjusted to fit experimental data. The core-electron attraction and core-electron repulsion term for MNDO are represented by the following equations, respectively:

$$V_{\mu\nu,B} = -Z_B \langle \mu\nu | s_B s_B \rangle \tag{2.47}$$

$$E_{AB}^{core} = Z_A Z_B \langle s_A s_A | s_B s_B \rangle + f_3 (R_{AB}) \tag{2.48}$$

where $f_3 (R_{AB})$ contains exponential functions in the form $e^{\alpha_A R_{AB}} + e^{\alpha_B R_{AB}}$, but the formulas are varied for different atom pairs. For example for two neutral atoms,

$$f_3 (R_{AB}) = Z_A Z_B (s_A s_A | s_B s_B) [e^{-\alpha_A R_{AB}} + e^{\alpha_B R_{AB}}] \tag{2.49}$$

is used but it was determined that for the pairs *N-H* and *O-H*,

$$f_3 (R_{XH}) = Z_X Z_H (s_X s_X | s_H s_H)(R_{XH}/\text{Å})[e^{-\alpha_X R_{XH}} + e^{\alpha_H R_{XH}}] \tag{2.50}$$

is more accurate, where X = N or O.

The MNDO approximation, like the NDDO approximation, provides a quick way to perform quantum mechanical molecular calculations, but unlike the NDDO approximation, the MNDO approximation provides some electron correlation effects due to the use of one center repulsion integrals. Some of the drawbacks of the MNDO method are that calculations fail to reproduce hydrogen bonds distances, energies are too small for molecules that are sterically crowded, energies are too large for molecules containing four-membered rings, and the activation energies tended to be too large. The next generation of semi-empirical methods, Austin Model 1 (AM1), was developed specifically to

address these issues.

Austin Model 1 (AM1)

In 1985, Dewar et al. [12] developed AM1 using the NDDO approach and the MNDO formalism previously discussed. The shortcoming of MNDO were specifically addressed when this method was developed. It was found that the core repulsion function in the MNDO method, shown in Equation (2.48), could be modified with the addition of Gaussian terms. Therefore Equation (2.48) was modified to:

$$E_{AB}^{core} = Z_A Z_B \gamma_{ss} \left[1 + F\left(A\right) + F\left(B\right) \right] \tag{2.51}$$

where

$$F(A) = e^{-\alpha_A R_{AB}} + \sum_i K_{A_i} e^{[L_{A_i}(R_{AB} - M_{A_i})^2]} \tag{2.52}$$

$$F(B) = e^{-\alpha_B R_{AB}} + \sum_j K_{B_j} e^{[L_{B_j}(R_{AB} - M_{B_j})^2]} \tag{2.53}$$

The L parameters determine the width of the Gaussian and are not included in the optimization of these terms, and the factors K_{A_i}/K_{B_j} and M_{A_i}/M_{B_j} were optimized. K_{A_i}/K_{B_j} are Gaussian multiplier for the nth Gaussian of atom A/B, L_{A_i}/L_{B_j} are Gaussian exponent multiplier for the nth Gaussian of atom A/B and M_{A_i}/M_{B_j} are the radius of center for the nth Gaussian of atom A/B. Two types of Gaussian terms were added to Equation (2.51) to reduce the overestimated interatomic repulsions at large separation. The first type is an attractive Gaussian to compensate the excessive repulsion and the second type is a repulsive Gaussian centred at smaller internuclear separations. The addition of these Gaussians reduced the repulsion at larger internuclear distances. Therefore, some of the drawbacks of the MNDO method have been addressed with the AM1 method and this leads to a better estimate of activation energy for reactions, as

well as the ability to reproduce hydrogen bond distances.

Recife Model 1 (RM1)

Stewart et al. published a series of papers on development of the widely used Parameteri-
zation Model (PMx, where X is 3,4,5,6) semiempirical molecular orbital methods [40–45]
and these methods were utilized along with the previously described AM1 method for
approximately twenty years. In 2006 Rocha et al. reparameterized the AM1 method and
created the Recife Model 1 (RM1) [13] semiempirical molecular orbital method. This
method uses the same parameters as the AM1 method, but unlike the AM1 method
where only a few parameters were optimized, RM1 optimizes all 191 parameters for
ten atoms (hydrogen, carbon, nitrogen, oxygen, phosphorus, sulphur, fluorine, chlorine,
bromine, and iodine), including the one-center two-electron repulsion integrals, which
are taken from experimental ionization potentials and electron affinities [46] in the AM1
method. The optimization procedure for the RM1 method uses a least squares fitting
procedure.

$$F_{resp} = \sum_i (q_i^{calc} - q_i^{exp})^2 w_i^2, \tag{2.54}$$

where q_i^{calc} and q_i^{exp} are calculated and experimental values of property i, and w_i are
weights equal to those used in the MNDO approximation (given in Table 2.1). The
chemical properties that were investigated by Rocha et al. [13] are enthalpies of forma-
tion, dipole moments, ionization potentials, bond lengths, and bond angles, and from
their results it is apparent that RM1 is an improvement over not only AM1, but also
PM3 and in addition it shows to be at least competitive with the PM5 method.

2.3   Beyond Wave Function Theory

Post HF methods such as complete active space (CAS), configuration interaction (CI),
perturbation theory (PT) and coupled cluster (CC), are additional ways of approximating

Table 2.1: Properties and weights used in general semi-empirical fitting for MNDO methods

Property, q_i^{exp}	Weight, w_i
Heat of Formation (ΔH_f)	1kcal^{-1} mol
Ionization potential (IP)	10 eV^{-1}
Dipole moment (μ)	20 D^{-1}
Interatomic distance (R_{ab})	100 Å^{-1}
Bond angle (θ_l)	2/3 degree^{-1}
Dihedral angle (θ_d)	1/3 degree^{-1}

the exact solution of the Schrödinger equation, but unfortunately even for small molecules the calculations are very expensive. The time consuming problem has been addressed by the popular density functional theory (DFT), where the system is described by the electronic density ρ, which only depends on three variables x, y, and z. Therefore, no matter how big the system is the problem is always a 3-dimensional problem in DFT and not a 3N-dimensional (where N is the number of electrons) problem as in the HF-based methods. The DFT functionals have been developed for different problems of interest and a comical yet accurate summary of these developments can be found in [47]. Within this Thesis a general overview of the DFT method will be given.

2.3.1 Density Functional Theory

In 1964, Hohenberg and Kohn [48] proved that for N interacting electrons moving in an external potential $v(\mathbf{r})$ there is a universal functional $F[\rho(\mathbf{r})]$ of the ground-state electron density $\rho(\mathbf{r})$ that minimized the energy functional

$$E[\rho(\mathbf{r})] = F[\rho(\mathbf{r})] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r}. \tag{2.55}$$

The minimum value of E is the ground-state electronic energy. The theory given here only holds for the non-degenerate ground-states, for the case of degenerate ground-states see [49].

The major problem of the above theory is that it provides only an existence proof of a functional, but the form of this functional is still unknown. In 1965, Kohn and Sham [50] developed a method for determining $\rho(\mathbf{r})$. The Kohn-Sham method to determine $\rho(\mathbf{r})$ solves an equation which has a similar form to the HF equation (given in Equation (2.56) below).

$$\hat{F}^{KS}\phi_i^{KS}(\mathbf{r}) = \epsilon_i^{KS}\phi_i^{KS}(\mathbf{r}), \tag{2.56}$$

Equation (2.56) is known as the Kohn-Sham (KS) equation. $\hat{F}^{KS}$ is the effective one-electron KS operator, $\phi_i^{KS}(\mathbf{r})$ are the KS orbitals, and ϵ_i^{KS} are the KS orbital energies. In the KS formalism, the electron density of the system is calculated as

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{\frac{N}{2}} |\phi_i^{KS}(\mathbf{r})|^2. \tag{2.57}$$

The KS operator $\hat{F}^{KS}$ is defined as

$$\hat{F}^{KS}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{1A}} + 2\sum_{j=1}^{\frac{N}{2}} J_j + V_{XC}(\mathbf{r}), \tag{2.58}$$

where $J_j(\mathbf{r_1})$ the Coulomb repulsion term is

$$J_j(\mathbf{r_1}) = \int \frac{\rho(\mathbf{r_2})}{r_{12}} dv_2, \tag{2.59}$$

and $V_{XC}(\mathbf{r})$ the exchange-correlation potential term is the functional derivative

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}. \tag{2.60}$$

The ground-state energy functional in Equation (2.55) is given as

$$E_0[\rho(\mathbf{r})] = 2\sum_{i=1}^{\frac{N}{2}}\langle\phi_i(\mathbf{r})| - \frac{1}{2}\nabla^2|\phi_i(\mathbf{r})\rangle - \sum_{A=1}^{M}\int\frac{Z_A\rho(\mathbf{r_1})}{r_{1A}}d\mathbf{r_1} \qquad (2.61)$$
$$+\frac{1}{2}\int\int\frac{\rho(\mathbf{r_1})\rho(\mathbf{r_2})}{r_{12}}d\mathbf{r_1}d\mathbf{r_2} + E_{XC}[\rho(\mathbf{r})].$$

The form of the exchange-correlation energy $E_{XC}[\rho(\mathbf{r})]$ is not known and is usually approximated by separating $E_{XC}[\rho(\mathbf{r})]$ into two parts, exchange and correlation functionals, given as

$$E_{XC}[\rho(\mathbf{r})] = E_X[\rho(\mathbf{r})] + E_C[\rho(\mathbf{r})]. \qquad (2.62)$$

Two exchange functionals that are used within the calculations in this Thesis: Becke 1993 ($B3$) [51] and Gill 1996 (G) [52] and two correlation functionals are used: Perdew 1986 ($P86$) [53] and Lee, Yang and Parr (LYP) [54].

Chapter 3

Classical Mechanical Methods

3.1 Molecular Mechanics

Molecular mechanics (MM) is the method most commonly used when the system of interest is a large biomolecule. Molecular mechanics is based on Newtonian mechanics,

$$F = \frac{d}{dt}m\mathbf{v}, \tag{3.1}$$

where m is the mass and $\mathbf{v}$ is the velocity of a particle. Molecular mechanics uses two basic assumptions: the first is that the potential energy of the system can be calculated by summing the bonding and nonbonding interactions between the atoms, and the second is that the parameters used in the force field equation can be transferred to other similar systems.

3.1.1 Force Fields

The force field used in molecular mechanics to calculate the potential energy (U) sums contributions from the bond stretches, angle bending, bond rotations, and nonbonding interactions. The basic form of the potential energy equation is

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{electrostatic} + U_{van\ der\ Waals} \tag{3.2}$$

where U_{bond}, U_{angle}, and $U_{dihedral}$ are the bonding terms and $U_{electrostatic}$, and $U_{van\ der\ Waals}$ are the nonbonding terms.

The U_{bond} term describes the stretching and compression of chemical bonds (r_{AB})

$$U_{bond} = \frac{1}{2} \sum_{bonds} k_{AB}(r_{AB} - r_{AB,0})^2, \tag{3.3}$$

where k_{AB} is a force constant for each bond $A - B$, and the equilibrium bond length between atoms A and B is $r_{AB,0}$. k_{AB} and $r_{AB,0}$ are constant for specific bonds between two atoms. Similarly, U_{angle} describes the the bending of the bond angle (ϕ_{ABC})

$$U_{angle} = \frac{1}{2} \sum_{angles} k_{ABC}(\phi_{ABC} - \phi_{ABC,0})^2, \tag{3.4}$$

with a force constant k_{ABC} and a equilibrium bond angle ϕ_{ABC}. The $U_{dihedral}$ term,

$$U_{dihedral} = \frac{1}{2} \sum_{dihedrals} V_{ABCD}[1 + (-1)^{n+1} \cos(n\ \omega_{ABCD} + \phi_{ABCD})], \tag{3.5}$$

describes the contribution to the energy from the dihedral angle of four consecutive atoms, A, B, C, and D. V_{ABCD} is the amplitude of the energy curve for the four atoms. n is the the periodicity of the dihedral angle, ω_{ABCD} is the dihedral angle, and ϕ_{ABCD} is a phase angle. These three terms (U_{bond}, U_{angle}, and $U_{dihedral}$) are the typical bonding terms used within a molecular mechanical description of a molecule. Occasionally additional terms are added, such at the improper dihedral angle term, which describes the dihedral angles between nonconsecutive atoms.

The nonbonding terms (or noncovalent interactions) are the electrostatic and the van der Waals interactions. The electrostatic term describes ionic bonding, hydrogen bonding

and hydrophobic interactions. The electrostatic interaction is calculated by

$$U_{electrostatic} = \sum_{pairs\ A,B} \frac{q_A q_B}{r_{AB}} \tag{3.6}$$

and requires the value a partial charge q_A or q_B on atoms A and B, respectively, which are separated by their distance r_{AB}. The van der Waals interactions are usually calculated by using the Lennard-Jones potential,

$$U_{van\ der\ Waals} = \sum_{pairs\ A,B} \left(\frac{a_{AB}}{r_{AB}^{12}} - \frac{b_{AB}}{r_{AB}^{6}} \right), \tag{3.7}$$

where a_{AB} and b_{AB} are constants that depend on atoms A and B. This potential is sometimes called the 12-6 potential and is the potential often used in molecular mechanical calculations. Sometimes an alternative potential is used, such as the Morse potential.

These five terms (U_{bond}, U_{angle}, $U_{dihedral}$, $U_{electrostatic}$, and $U_{van\ der\ Waals}$) describe the basic composition of the force field. The number and values of each of the parameters in the equations depend on the specific force field used within the molecular mechanical program.

3.2 Molecular Dynamics

Molecular dynamics (MD) is the study of the motion of atoms within the system of interest over a set length of time. Molecular dynamics provides detailed time and space information of molecular behaviour in phase space. Phase space is the coordinate system used in molecular dynamic calculations, which evolves over time. It is a plot of the momentum of particles with respect to their positions. An important concept for molecular dynamics is that the system must explore all microstates in time (the system is ergodic) because the time average must be approximately equal to the phase space average so

that thermodynamic properties can be calculated. The microstate is a specific micro-scopic configuration of a thermodynamic system that the system occupies with a certain probability in the course of its thermal fluctuations. The trajectories of the system are calculated and molecular mechanics uses Newton's equations of motion (EOM),

$$v = u + at \tag{3.8}$$

$$s = \tfrac{1}{2}(u+v)t \tag{3.9}$$

$$s = ut + \tfrac{1}{2}at^2 \tag{3.10}$$

$$s = vt - \tfrac{1}{2}at^2 \tag{3.11}$$

$$v^2 = u^2 + 2as \tag{3.12}$$

$$a = \frac{v-u}{t} \tag{3.13}$$

By substituting (3.8) into (3.9), we can get (3.10), (3.11) and (3.12). (3.13) can be constructed by rearranging (3.8). In these equations s is the distance between initial and final positions (displacement), u is the initial velocity, v is the final velocity, a is the constant acceleration, and t is the time taken to move from the initial state to the final state. One example of an algorithm to solve the equations of motion is the Verlet algorithm:

$$x(t + \Delta t) = \frac{a(t)\Delta t^2 + 2x(t) - x(t - \Delta t)}{\Delta t^2}, \tag{3.14}$$

where t is time Δt is the time step, x is the positions and a is the acceleration. Generally a molecular dynamic simulation progresses in the following manner:

(1) Specify initial position of atoms and select Δt

↓

(2) Calculate force ($\mathbf{F}$) and acceleration ($\mathbf{a}$)

using a solution to the EOM

↓

(3) Update positions of atoms

↓

(4) Add time step ($t = t + \Delta t$)

↓

(5) Return to Step (2) as long as required to explore all microstates.

Upon completion, snapshots of the system can be used to determine various quantities of interest such as structures (radial distribution functions, solvated structures), energetics (energy of a system as a function of time, binding affinities, free energy), and thermodynamics (follow a reaction coordinate, thermodynamic quantities).

3.2.1 Ensembles

The variables used in molecular dynamics calculations are: the number of particles or atoms (N), pressure (P), temperature (T), volume (V), energy (E), and the chemical potential (μ_i). There are three main types of ensembles: micro-canonical, canonical, and isothermal-isobaric. In the micro-canonical ensemble the number of particles, the volume, and the total energy are kept constant during the simulations and the time evolution of position and velocity of the particles is investigated. In the canonical ensemble the number of particles, the volume, and the system's temperature are kept constant during the simulations. This ensemble is used when the studied system needs to be kept at a constant temperature in order to better mimic the conditions found in nature. For example, when the system of interest is a protein found within the body, it is usual

practice to use this ensemble and have the temperature at 310K. The last ensemble is the isothermal-isobaric ensemble which keeps the number of particles, the temperature of the system, and the pressure of the system constant. This ensemble is typically used when the system that is studied needs to mimic experimental conditions that require constant temperature and constant pressure such as biological membranes, lipid bilayers, and surface tension simulations.

3.3   Solvation and Free Energy Calculations

Unlike QM, where if the same method and basis set are used, the energetics of the molecular mechanics system can not be directly compared, because in molecular mechanics each molecule has a distinctive force field. It is necessary to calculate the enthalpies of formation or free energy in order to compare the energetics of different molecules. Also, in order for the calculations to be comparable to experiment, solvation often has to be included in calculations.

3.3.1   TIPnP Water Molecules

Since most biological processes take place in an aqueous environment, it is often necessary for the system of interest to be modeled in the presence of solvent. Therefore, water molecules are frequently explicitly added to the solute. Usually there are significantly more solvent molecules than solute and this drastically increases the amount of computational time required to model the system. Often an approximation is made in the description of the solvent water molecules when these molecules are not participating in the reaction being studied. A frequently used model for such water is the TIPnP model. The n in TIPnP indicates the number of point charges on the water molecule. The water molecule in the TIPnP approximation is rigid and therefore the bond lengths and bond angle do not change. The OH bond length is fixed at 0.9572 Å and a HOH bond angle is fixed at 104.52°. The TIP3P water molecule has 3 point charges of $0.417e$

on the hydrogen atoms and $-0.834e$ on the oxygen atom. The values of the parameters were determined using the pair potential for water-water interaction between the pair of waters labeled m and n is

$$\epsilon_{mn} = \sum_i^{on\ n} \sum_j^{on\ m} \frac{q_i q_j K_c^2}{r_{ij}} + \frac{A}{r_{mn}^{12}} - \frac{C}{r_{mn}^6}, \tag{3.15}$$

and performing Monte Carlo simulations to optimize q, A, and C [55], where r_{mn} is the distance between two oxygen atoms, and K_c is the electrostatic constant (K_c=332.1 Å kcal mol^{-1}). The dipole moment, heats of vaporization, heat capacity, density of liquid water, and the O-O radial distribution function of liquid water were used as experimental references in the fitting [56]. The parameters (charges, bond lengths, and bond angles) vary depending on which TIPnP (where n=3,4,5,6) model is being used. The TIP4P water molecule has an extra point charge around the oxygen atom, TIP5P has two additional point charges around the oxygen atom, and finally TIP6P has three additional point charges around the oxygen atom. TIP4P, TIP5P and TIP6P lead to a better description of liquid water, but they are much more expensive models and usually TIP3P is an adequate approximation for the systems being studied.

3.3.2 Generalized Born equations

The Born equation (3.16) is used to calculate the polarization free energy change of a monatomic ion in atomic units

$$G_p = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\frac{q^2}{a}, \tag{3.16}$$

where q is the charge on the atom (considered in this approximation as a sphere), ϵ is the dielectric constant of the medium in which the sphere is placed, and a is the radius of the sphere. Extending the Born equation to a polyatomic system gives the generalized

Born equation (3.17).

$$G = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\underbrace{\sum_{k,k'}^{atoms} q_k q_{k'} \gamma_{kk'}}_{G_p} + \underbrace{0.0072(SASA)}_{G_{np}}, \tag{3.17}$$

where the sum (k and k') is over all atoms with partial charges q_k and $q_{k'}$, respectively. $\gamma_{kk'}$ has units of inverse length and has the form

$$\gamma_{k,k'} = (r_{kk'}^2 + \alpha_k \alpha_{k'} e^{r_{kk'}^2/d_{kk'}\alpha_k\alpha_{k'}})^{-\frac{1}{2}}, \tag{3.18}$$

where $r_{kk'}$ is the interatomic distance, α_k is is the effective Born radius of atom k, d is a parameter which is typically given a value of 4, and $SASA$ is the solvent accessible surface area. G_p and G_{np} are the polar and the non-polar portion of the equation respectively.

3.3.3 Generalized Born Solvation Area

In order to calculate the change in free energy between two states, the generalized Born solvation area is used. Ideally

$$Drug_{(aq)} + Protein_{(aq)} \overset{\Delta G^\circ_{bind,sol}}{\longleftrightarrow} Complex_{(aq)} \tag{3.19}$$

would be used, but much of the energy contributions would come from solvent-solvent interactions and therefore would introduce a large error. An alternative method to directly calculating the change in free energy of solvation ($\Delta G_{bind,sol}$) is to calculate the change in free energy of gas phase drug to solvated drug ($\Delta G^\circ_{D,sol}$), gas phase ligand to solvated protein ($\Delta G^\circ_{P,sol}$), gas phase complex to solvated complex ($\Delta G^\circ_{C,sol}$), and the change in the free energy of the gas phase binding ($\Delta G^\circ_{bind,gas}$).

$$Drug_{(aq)} \quad + \quad Protein_{(aq)} \quad \overset{\Delta G^{\circ}_{bind,sol}}{\longleftrightarrow} \quad Complex_{(aq)}$$

$$\Big\updownarrow \Delta G^{\circ}_{(D,sol.)} \qquad \Big\updownarrow \Delta G^{\circ}_{(P,sol.)} \qquad\qquad \Big\updownarrow \Delta G^{\circ}_{(C,sol.)}$$

$$Drug_{(g)} \quad + \quad Protein_{(g)} \quad \overset{\Delta G^{\circ}_{bind,gas}}{\longleftrightarrow} \quad Complex_{(g)}$$
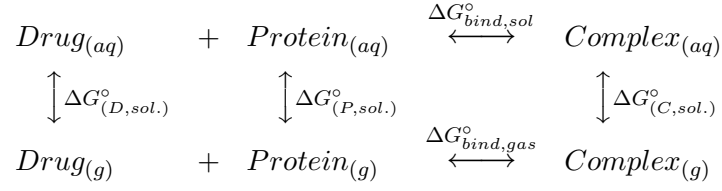
After all required ΔG° values have been calculated, $\Delta G^{\circ}_{bind,sol}$ can be calculated by

$$\Delta G^{\circ}_{bind,sol} = G_{C,sol} - (G_{D,sol} + G_{P,sol}) \tag{3.20}$$

$$\Delta G^{\circ}_{bind,gas} = G_{C,gas} - (G_{D,gas} + G_{P,gas}) \tag{3.21}$$

$$\Delta G^{\circ}_{C,sol} = G_{C,sol} - G_{C_{gas}} \tag{3.22}$$

$$\Delta G^{\circ}_{P,sol} = G_{P,sol} - G_{P_{gas}} \tag{3.23}$$

$$\Delta G^{\circ}_{D,sol} = G_{D,sol} - G_{D_{gas}} \tag{3.24}$$

and substituting Equations (3.22), (3.23), and (3.24) into Equation (3.20), gives

$$\Delta G^{\circ}_{bind,sol} = \Delta G^{\circ}_{C,sol} - \Delta G^{\circ}_{P,sol} - \Delta G^{\circ}_{D,sol} + G_{C,gas} - G_{D,gas} - G_{P,gas}. \tag{3.25}$$

Finally, the last three terms of Equation (3.25) are equivalent to Equation (3.21), and substituting Equation (3.21) into Equation (3.25) one gets

$$\Delta G^{\circ}_{bind,sol} = \Delta G^{\circ}_{bind,gas} + \Delta G^{\circ}_{C,sol} - (\Delta G^{\circ}_{D,sol} + \Delta G^{\circ}_{P,sol}), \tag{3.26}$$

which is the final equation used and allows for the desired change in free energy of the solvated system to be extracted ($\Delta G^{\circ}_{bind,sol}$).

Chapter 4

DFT Studies of Complexes Between Ethylenediamine Tetraacetate and Alkali and Alkaline Earth Cations[*]

4.1   Introduction

Ethylenediamine tetraacetic acid (H_4EDTA) is a polyprotic acid that has been widely used, together with its associated anion ethylenediamine tetraacetate ($EDTA^{4-}$), since the pioneering work on H_4EDTA was done by Schwarzenbach in the 1940s [57]. One of the most useful properties of $EDTA^{4-}$ is its ability to form exceptionally strong complexes with Mn^{2+}, Cu^{2+}, Fe^{3+}, and Co^{3+} [58]; $EDTA^{4-}$ will also form complexes with many other cations. For this reason, $EDTA^{4-}$ can be used in sequestering heavy metal ions, buffering of solutions, and many forms of titrations [59]. Also, some of the properties of $EDTA^{4-}$ allow it to be used in the pulp and paper industry [60], as well as an industrial cleaner [58], a food preservative [61], and as a drug [62]; it is also used as a detergent to complex with Ca^{2+} and Mg^{2+} in order to reduce the hardness of water [63].

H_4EDTA is a polyprotic acid with $pKa_1 = 1.99$, $pKa_2 = 2.67$, $pKa_3 = 6.16$, and $pKa_4 = 10.26$ [64]. This can allow for different ion selectivity based on the pH of the solution: one can adjust the pH of the solution and influence the order in which the ions forms complexes with H_nEDTA^{n-4} (n=0, 1, 2, 3, 4). Crystal structures of EDTA^{4-} complexed with Na$^+$, Mg^{2+}, and Ca^{2+} are available [65–67] and may be used to compare with calculated geometries.

One approach that can be taken to computationally establish the order with which cations bind to EDTA^{4-} is to calculate the energy difference between reactants and products. Based on this energy difference (ΔE_{rxn}), the binding affinities of cations to EDTA^{4-} could be determined by comparing the energy of complexation. Equation (4.1) illustrates the calculation of ΔE_{rxn}, or the stabilization energy (SE), for cation-EDTA^{4-} and the cation-$(H_2O)_6$ complexes:

$$\text{EDTA}^{4-} + \text{M}^{n+} \rightleftharpoons [\text{MEDTA}]^{n-4} \tag{4.1}$$

$$6H_2O + \text{M}^{n+} \rightleftharpoons [\text{M}(H_2O)_6]^{n+}$$
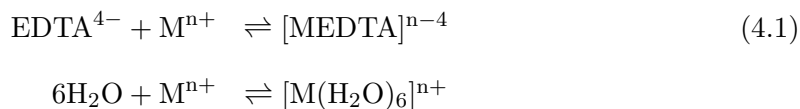
where M $\in$ { Na$^+$, K$^+$, Rb$^+$, Mg^{2+}, Ca^{2+}, and Sr^{2+}}.

DFT calculations that were performed in the present work were carried out for the gas phase; in such a case, Equation (4.1) introduces errors because solvation effects have not been taken into account. Thompson et al. [68] and Glendening et al. [69] suggested that an alternative approach may be used to assess the binding affinity of cations to a ligand in aqueous solution. Their method, which is shown in Equation (4.2) and depicted in Figure 4.1,

$$[\text{M}_1\text{EDTA}]^{n_1-4} + [\text{M}_2(H_2O)_6]^{n_2+} \rightleftharpoons [\text{M}_2\text{EDTA}]^{n_2-4} + [\text{M}_1(H_2O)_6]^{n_1+} \tag{4.2}$$
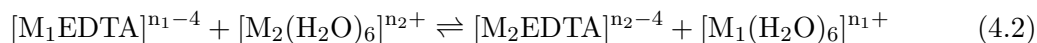
where $M_1 \neq M_2$ and $M_i \in \{$ Na^+, K^+, Rb^+, Mg^{2+}, Ca^{2+}, and $Sr^{2+}\}$, is based on an exchange reaction of two cations which are enclosed in six water molecules and the host ligand, in the present case the $EDTA^{4-}$ anion.
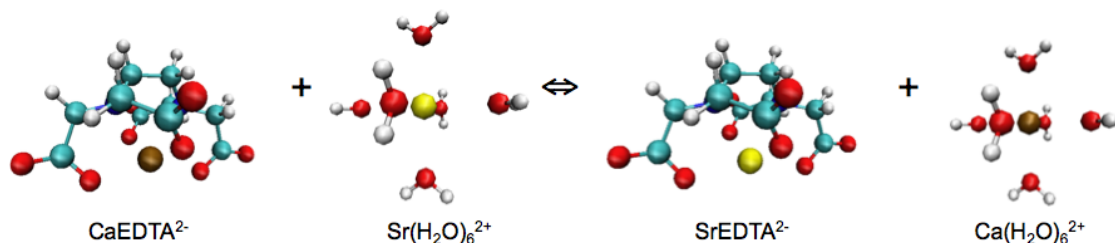
Figure 4.1: Pictorial representation of the exchange reaction used to calculate $\Delta\Delta E_{rxn}$

Glendening et al. [69] demonstrated that this approach will mimic some of the solvation effects that influence the binding affinity of cations to ligands. Their work on crown-ether complexes [69] showed that by using Equation (4.2) changed the ordering of the gas phase calculations binding affinity of cations to 18-crown-6 to match experimental results obtained in aqueous solution. Using Equation (4.2) may also lead to a reduction in errors which are due to inadequate basis set representation because it may be expected that these errors would be approximately equivalent on both sides of Equation (4.2). In the present work, corrections for basis set superposition errors (BSSE) were calculated and it was found that these BSSE corrections were approximately equal for all the cation-$EDTA^{4-}$ complexes. Therefore the BSSE corrections need not be included as they cancel out when using Equation (4.2).

Glendening et al. [70, 71] carried out extensive studies to determine the lowest energy structure for different configurations of water clusters with various cations. The different structures of water complexes are shown in Figure 4.2. Alkali cations have lower energies

for the first two geometries depicted in Figure 4.2, with the first structure possessing S_6 symmetry (Figure 4.2a) and the second C_{2v} symmetry (Figure 4.2b). The results from these calculations show that the energy difference between these two water complex conformations is either negligible or it slightly favors the C_{2v} geometry. Therefore the C_{2v} geometry was chosen for all of the alkali cations. All alkaline earth cations prefer the T_h geometry that is shown in Figure 4.2c. As seen in Figure 4.2, the lowest energy structures for the alkali cations are very different from the lowest energy structures of the alkaline earth cations.

(a)
S_6 symmetry

(b)
C_{2v} symmetry

(c)
T_h symmetry
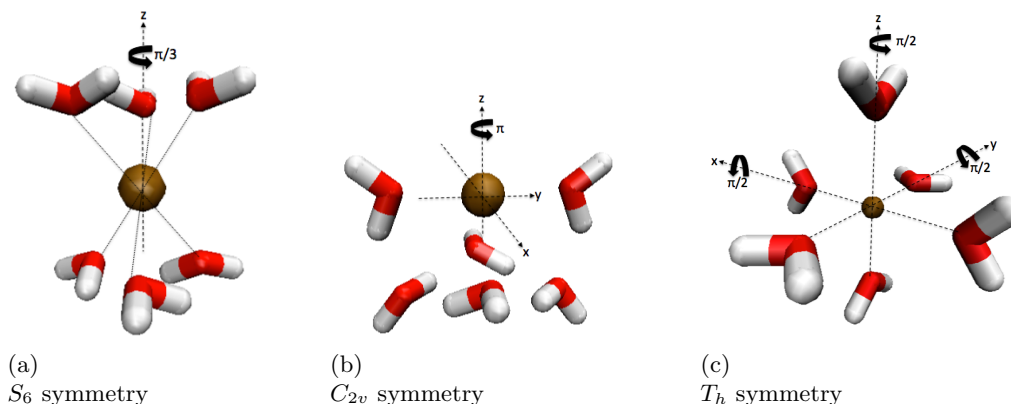
Figure 4.2: Structures of water complexes with alkali and alkaline earth cations

The structures of free EDTA^{4-} and the cation-EDTA^{4-} complexes are quite different and are shown in Figure 4.3. The free EDTA^{4-} has all four carboxylate groups pointing away from the central N-C-C-N dihedral angle, as shown in Figure 4.3a. The negatively charged oxygen atoms on the carboxylate groups cause electrostatic repulsion between the carboxylate side chains. The electrostatic repulsion, along with the flexible geometry of EDTA^{4-}, causes the carboxylate groups to point away from the central N-C-C-N dihedral angle in order to minimize the repulsion. As the cations are sequestered, three or four oxygen atoms envelope the cation (one oxygen atom from each of the carboxylate

groups). The case of three oxygen atoms from three carboxylate groups sequestering a cation is shown in Figure 4.3b and the case of four oxygen atoms from four carboxylate groups sequestering a cation is shown in Figure 4.3c. It is interesting to note that the carboxylate groups do not lose their resonance structure (bond order for each C-O bond is 1.5) upon complexation with a cation: all of the carbon-oxygen bond lengths are approximately 1.26 ± 0.03 Å in all of the calculated geometries and the crystal structure geometries [65–67]. There is little change (± 0.02 Å) in the carbon-oxygen bond lengths between the free $EDTA^{4-}$ and the cation-$EDTA^{4-}$ complex.

(a)
$EDTA^{4-}$

(b)
$[Na\ EDTA]^{3-}$

(c)
$[Ca\ EDTA]^{2-}$
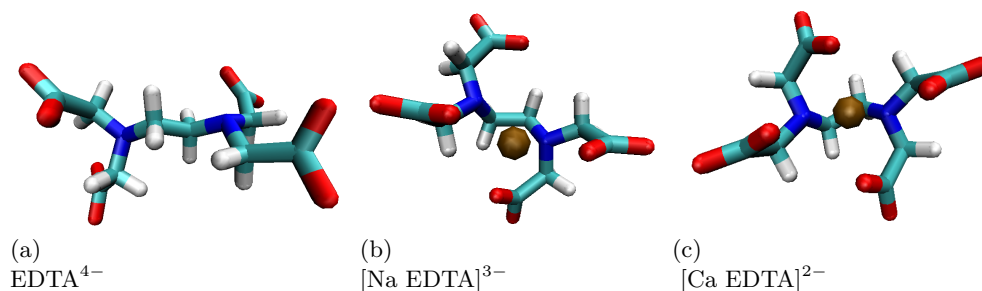
Figure 4.3: Structures of $EDTA^{4-}$ and its complexes

It is the purpose of the present work to evaluate the binding energies of the cation-$EDTA^{4-}$ complexes using two DFT functionals along with the all electron (AE) and the model core potential (MCP) methods.

4.2 Method of Calculations

Density functional theory (DFT) with two different functionals was used to calculate the cation-$EDTA^{4-}$ and the cation-$(H_2O)_6$ complexes. The DFT calculations in the first set were carried out using Gill 1996 exchange [52] and Perdew 1986 correlation functional [53] (GP86) while in the second set the commonly used B3LYP functional [51,54]

was employed. The two functionals, GP86 and B3LYP, were employed in order to ensure that the results were independent of the DFT functional used. Furthermore, two different methods were used to describe electron distribution: the AE and a pseudopotential method, more specifically the MCP. For all the studied systems geometry optimization was performed, followed by hessian calculation and harmonic vibrational analysis. The hessian calculation was done to ensure that the stationary points found in the geometry optimization were minima rather than saddle points.

For the AE calculations, the following composite basis set was used: 6-31G* basis set on carbon, magnesium, calcium, sodium, potassium [72, 73]; 6-31+G* on oxygen, nitrogen, hydrogen [72, 73]; a split-valence all-electron basis set (SVPall) on rubidium and strontium [74]; the SVPall basis is roughly equivalent to 6-31G* [75]. The AE DFT calculations are denoted by B3LYP/AE and GP86/AE.

In order to reduce computational effort as well as to calibrate the pseudopotential method, the MCP method [76] and basis sets were used in the second set of DFT calculations. The MCP method approximates the core electrons by a potential and explicitly treats only the valence electrons and thus offers an effective manner of reducing the basis set size, leading to reduction of CPU time. The recently prepared MCP-DZP parameters and basis sets [77] were used. The MCP DFT calculations are referred to as B3LYP/MCP and GP86/MCP.

All of the computations for the complexes were done for systems in gas phase. Computer programs GAMESS [1] and Gaussian03 [2] were used, running concurrently on computer clusters at the University of Alberta. The GP86 DFT functional is implemented equivalently in both GAMESS and Gaussian03 for the AE calculations, which allowed for both GAMESS and Gaussian03 to be used for the AE/GP86 calculations. Visualiza-

tion of molecular structures, vibrational modes, and molecular orbitals were done using MOLDEN [5] and MacMolPlt [6].

4.3   Results, Discussion, and Conclusions

Glendening et al. [70], who carried out calculations on the 18-crown-6 molecule and calculated the binding affinities to the alkali cations, showed that by using an equation equivalent to Equation (4.1) the ordering of binding affinities matched experimental results that were obtained using a kinetics method in a collision-induced dissociation, liquid secondary ion mass spectrometry study of mixed crown ether/alkali halide complexes which was done by Maleknia and Brodbelt [78]. The calculations that best describe this type of experimental measurement are gas phase calculations. The gas phase calculations done by Glendening et al. and the experimental results given by Maleknia et al. both gave the same order of binding affinities for the crown ether/alkali cation complexes.

Equation (4.2) was used to calculate the binding affinities which incorporates some of the solvation effects, which is necessary since $EDTA^{4-}$ is typically used in aqueous solutions. The equations given below show how the $\Delta\Delta E$ is derived using the equations for the stabilization energies of individual complexes:

$$EDTA^{4-} + M_1^{n+} \rightleftharpoons [M_1EDTA]^{n-4} \qquad (\Delta E_1)$$

$$EDTA^{4-} + M_2^{n+} \rightleftharpoons [M_2EDTA]^{n-4} \qquad (\Delta E_2)$$

$$6H_2O + M_1^{n+} \rightleftharpoons [M_1(H_2O)_6]^{n_2+} \qquad (\Delta E_3)$$

$$6H_2O + M_2^{n+} \rightleftharpoons [M_2(H_2O)_6]^{n_2+} \qquad (\Delta E_4)$$
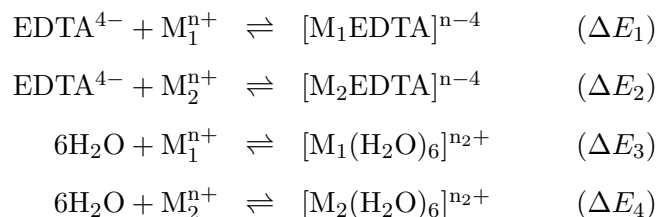
Suitable combination of these equations cancels out $EDTA^{4-}$ and six H_2O molecules, leaving only cation-$EDTA^{4-}$ or cation-$(H_2O)_6$ complexes, with the energy change given

by

$$\Delta\Delta E_{\text{rxn}} = (\Delta E_1 + \Delta E_3) - (\Delta E_2 + \Delta E_4) \tag{4.3}$$

for the reaction

$$[M_1 EDTA]^{n_1-4} + [M_2(H_2O)_6]^{n_2+} \rightleftharpoons [M_2 EDTA]^{n_2-4} + [M_1(H_2O)_6]^{n_1+}$$
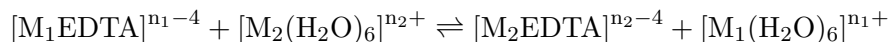
that is equivalent to Equation (4.2). Thus the specific equation that was used to calculate the binding affinities of the cation to $EDTA^{4-}$ is based on Equation (4.2):

$$\Delta\Delta E_{\text{rxn}} = \left(E_{[M_2 EDTA]^{n_2-4}} + E_{[M_1(H_2O)_6]^{n_1+}} \right) - \left(E_{[M_1 EDTA]^{n_1-4}} + E_{[M_2(H_2O)_6]^{n_2+}} \right) \tag{4.4}$$

A negative value of $\Delta\Delta E$ indicates that M_2 binds with greater affinity to $EDTA^{4-}$ than M_1. The results for the calculated order of selectivity for some of the alkali and alkaline earth cations are summarized in Tables 4.1 and 4.2.

The data shown in Tables 4.1 and 4.2 indicate that the relative difference between the $\Delta\Delta E$ for the two different functionals is smaller when using the AE DFT method. More importantly, it is apparent from these tables that the same selectivity of cations to $EDTA^{4-}$ was obtained for all four methods used, B3LYP/AE, GP86/AE, B3LYP/MCP, and GP86/MCP, namely $Mg^{2+} > Ca^{2+} > Sr^{2+} > Na^+ > K^+ > Rb^+$, which matches experimental results except for the reversal of ordering of Ca^{2+} and Mg^{2+} [79]. In order to understand this discrepancy, it must be noted that at the high pH is required for the existence of $EDTA^{4-}$ in aqueous solution, Mg^{2+} would have precipitated out of solution forming $Mg(OH)_2$. The magnesium cation would more likely bind to $HEDTA^{3-}$ according to the reaction,

$$HEDTA^{3-} + Mg^{2+} \rightleftharpoons Mg[HEDTA]^-, \tag{4.5}$$

unlike the calcium cation which reacts with $EDTA^{4-}$ as shown in the generic $EDTA^{4-}$ reaction scheme given in Equation (4.1). In order to address the inconsistency in the binding order, the stabilization energy for the $Mg[HEDTA]^{-}$ complex in Equation (4.5) were calculated and compared it to the stabilization energy from the $Ca[EDTA]^{2-}$ complex using the B3LYP/MCP method. It was found that the stabilization energy of the calcium-EDTA complex was approximately 18 kcal/mol lower than that of the magnesium-HEDTA complex. It may be assumed that if analogous calculations were performed with the replacement of $EDTA^{4-}$ with $HEDTA^{3-}$, the reversal of the Mg^{2+} and Ca^{2+} binding affinities for the complexes would remain consistent across all of the DFT calculations and a correct matching of calculated binding affinities to experimental results would be obtained.

In order to assess whether the same trend in the binding affinities would be seen in the presence of implicit solvent, single point calculations were done using the polarizable continuum model (PCM) [80] in GAMESS [1]. The previously described B3LYP/AE DFT method was employed at the geometry optimized in B3LYP/AE calculations. The uncorrected energies from the PCM calculations were corrected using the ZPE energy corrections from the B3LYP/AE calculations and the results were analyzed using Equation (4.4) to determine the binding affinity for the cation-$EDTA^{4-}$ complexes. The same trend, $Mg^{2+}> Ca^{2+}> Sr^{2+}> Na^{+}> K^{+}> Rb^{+}$, was observed for the PCM results as well as for the gas-phase ones. It can be conjectured that if similar calculations were performed using the other DFT methods described above, the same binding affinity order would be obtained. It should be noted that compared to gas-phase results, all PCM calculations led to more stable, lower energy metal-ligand systems. The alkali cation-$EDTA^{4-}$ complexes are all more stable by approximately 330 kcal/mol where the alkaline earth cation-$EDTA^{4-}$ complexes are all more stable by approximately 190 kcal/mol. A similar trend is seen for the cation-$(H_2O)_6$ complexes with the increase in stability of 60

Table 4.1: Binding affinity (in kcal/mol) for cation-EDTA^{4-} complexes: results from B3LYP/AE calculations (GP86/AE values in parentheses); PCM values in italics.

M_1 / M_2	Na^+	K^+	Rb^+	Mg^{2+}	Ca^{2+}	Sr^{2+}
Na^+	—	6.8	10.6	-333.1	-315.1	-301.8
	—	(7.5)	(11.7)	(-335.3)	(-315.9)	(-299.8)
	—	*2.9*	*6.0*	*-63.3*	*-32.1*	*-15.0*
K^+	-6.8	—	3.7	-339.9	-322.0	-308.6
	(-7.5)	—	(4.2)	(-342.9)	(-323.4)	(-307.3)
	-2.9	—	*3.2*	*-66.1*	*-35.0*	*-17.9*
Rb^+	-10.6	-3.7	—	-343.6	-325.7	-312.3
	(-11.7)	(-4.2)	—	(-347.0)	(-327.6)	(-311.5)
	-6.0	*-3.2*	—	*-69.3*	*-38.2*	*-21.0*
Mg^{2+}	333.1	339.9	343.6	—	17.9	31.3
	(335.3)	(342.9)	(347.0)	—	(19.4)	(35.6)
	63.3	*66.1*	*69.3*	—	*31.1*	*48.2*
Ca^{2+}	315.1	322.0	325.7	-17.9	—	13.4
	(315.9)	(323.4)	(327.6)	(-19.4)	—	(16.1)
	32.1	*35.0*	*38.2*	*-31.1*	—	*17.1*
Sr^{2+}	301.8	308.6	312.3	-31.3	-13.4	—
	(299.8)	(307.3)	(311.5)	(-35.6)	(-16.1)	—
	15.0	*17.9*	*21.0*	*-48.2*	*-17.1*	—

kcal/mol and 190 kcal/mol for the alkali and alkaline earth cation complexes, respectively.

Experimental data for cation-EDTA^{4-} complexes are available for all of the cation-EDTA^{4-} complexes that were studied in this Chapter except for Rb^+ [79]. The experimental results show that on the basis of the logarithms of the stability constants ($\log K$), the binding affinity ladder for the cation-EDTA^{4-} complexes is $Ca^{2+} > Mg^{2+} > Sr^{2+} > Na^+ > K^+$. The $\log K$ values are 1.7, 0.8, 8.8, 10.7, 8.7 for the reaction:

$$EDTA^{4-} + M^{n+} \rightleftharpoons [MEDTA]^{n-4}$$
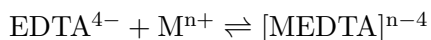
Table 4.2: Binding affinity (in kcal/mol) for cation-EDTA^{4-} complexes: results from B3LYP/MCP calculations (GP86/MCP values in parentheses).

M_1 / M_2	Na^+	K^+	Rb^+	Mg^{2+}	Ca^{2+}	Sr^{2+}
Na^+	—	8.7	9.7	-340.3	-320.3	-310.2
	—	(11.1)	(17.8)	(-342.7)	(-319.9)	(-306.6)
K^+	-8.7	—	1.0	-349.0	-329.0	-318.9
	(-11.1)	—	(6.7)	(-353.8)	(-331.0)	(-317.7)
Rb^+	-9.7	-1.0	—	-350.0	-330.0	-319.9
	(-17.8)	(-6.7)	—	(-360.5)	(-337.7)	(-324.4)
Mg^{2+}	340.3	349.0	350.0	—	20.0	30.1
	(342.7)	(353.8)	(360.5)	—	(22.9)	(36.1)
Ca^{2+}	320.3	329.0	330.0	-20.0	—	10.1
	(319.9)	3(331.0)	(337.7)	(-22.9)	—	(13.2)
Sr^{2+}	310.2	318.9	319.9	-30.1	-10.1	—
	(306.6)	(317.7)	(324.4)	(-36.1)	(-13.2)	—

for Na^+, K^+, Mg^{2+}, Ca^{2+}, and Sr^{2+}, respectively. Using these values of log K, the stability constants for the exchange reaction were calculated using the following equations (in aqueous solution),

$$EDTA^{4-} + M_1^{n+} \rightleftharpoons [M_1EDTA]^{n-4} \qquad (K_1)$$

$$EDTA^{4-} + M_2^{n+} \rightleftharpoons [M_2EDTA]^{n-4} \qquad (K_2)$$

$$[M_2EDTA]^{n-4} + M_1^{n+} \rightleftharpoons [M_1EDTA]^{n-4} + M_2^{n+} \qquad (K_3)$$
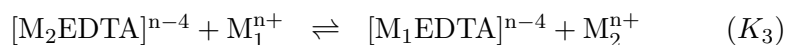
where $K_3 = K_1/K_2$. These values of K_3 inserted into the formula for Gibbs free energy, $\Delta G = -RT \ln K$, allowed for the experimental ΔG_{rxn} values to be calculated for the reaction shown in Equation (4.2), with the exception of Sr^{2+}. These values are summarized in Table 4.3 along with the values of ΔG obtained from the PCM calculations.

Table 4.3: Experimental and B3LYP/AE PCM (in parentheses) values of ΔG values for cation-EDTA^{4-} complexes (in kcal/mol).

M_1 / M_2	Na^+	K^+	Mg^{2+}	Ca^{2+}	Sr^{2+}
Na^+	—	1.17(3.83)	-9.72(-6.36)	-12.3(-25.78)	-9.64(-7.06)
K^+	-1.17(-3.83)	—	-10.9(-63.6)	-13.5(-29.6)	-10.8(-10.9)
Mg^{2+}	9.72(6.36)	10.9(63.6)	—	-2.59(3.40)	0.082(52.7)
Ca^{2+}	12.3(25.8)	13.5(29.6)	2.59(-3.40)	—	2.67(18.7)
Sr^{2+}	9.64(7.06)	10.8(10.9)	-0.082(-52.7)	-2.67(-18.7)	—

The calculated values for ΔG_{rxn} were obtained from the PCM results by using the uncorrected energies ($E_{o(PCM)}$) from the AE B3LYP PCM single-point calculations and adding the thermodynamic Gibbs free energy correction ($G_{corr(GAS)}$) from the gas phase hessian calculations of B3LYP/AE.

$$G_{[M_nEDTA]^{n_2-4}} = E_{o(PCM)} + G_{corr(GAS)}$$
$$G_{[M_n(H_2O)_6]^{n_1+}} = E_{o(PCM)} + G_{corr(GAS)}$$

The values of the ΔG_{rxn} that were obtained for the PCM calculations using Equation (4.6),

$$\Delta G_{rxn} = \left(G_{[M_2EDTA]^{n_2-4}} + G_{[M_1(H_2O)_6]^{n_1+}}\right) - \left(G_{[M_1EDTA]^{n_1-4}} + G_{[M_2(H_2O)_6]^{n_2+}}\right), \quad (4.6)$$

are summarized in Table 4.3. It is apparent from Table 4.3 that the calculated and experimental values of ΔG_{rxn} match except for Mg^{2+}. This is again due to the discrepancy between these calculated binding affinities and the experimental binding affinities, i.e., the reversal of the Mg^{2+} and Ca^{2+} ordering, which was discussed earlier.

The ZPE-corrected energies are shown in Tables 4.4 and 4.5. In these tables ΔE_0 denotes

the stabilization energy, calculated using Equation (4.1). The energies were taken at the optimized geometry (minimum, as confirmed by hessian calculation). ΔZPE is the zero-point-energy (ZPE) correction which when added to ΔE_0 gives the ZPE corrected energy ΔE_{corr}. It is interesting to note that the magnitude of the ΔZPE correction is much larger for the water complexes than it is for the EDTA^{4-} complexes. This difference is due to the way ZPE was evaluated: the ZPE correction is calculated by summing over all normal modes (3N-6 for non-linear molecules, where N is the number of atoms). In the cation/water complexation reaction there is an increase of 33 normal modes when going from reactants to products, while there is an increase of only 3 normal modes in the cation-EDTA^{4-} complexation reaction. This 11-fold increase in the number of normal modes between the two different reactions is the reason that the cation-$(H_2O)_6$ complexation reaction has a much greater ΔZPE correction. The ΔZPE is fairly consistent across the two functionals (B3LYP and GP86) and the basis sets (AE and MCP).

The two different DFT functionals (B3LYP and GP86) provide similar results for the AE and MCP calculations. The corrected stabilization energies (ΔE_{corr}) that are shown in Table 4.4 for B3LYP/AE and GP86/AE and in Table 4.5 for B3LYP/MCP and GP86/MCP, are summarized in Figure 4.4. The results given in Tables 4.4 and 4.5 are in kcal/mol, where mol refers to a single mole of the cation. In addition to revealing that the four methods give the same trend in the corrected stabilization energies, this figure also shows that the MCP calculations for the cation-$(H_2O)_6$ and cation-EDTA^{4-} complexes give a lower stabilization energy.

Table 4.4: All-electron stabilization energies of $EDTA^{4-}$ and hexaaqua complexes (in kcal/mol).

Ion-Complex	B3LYP			GP86		
	ΔE_0	ΔZPE	ΔE_{corr}	ΔE_0	ΔZPE	ΔE_{corr}
$[Na(EDTA)]^{3-}$	-333.2	2.4	-330.8	-322.5	2.3	-320.2
$[K(EDTA)]^{3-}$	-305.5	1.9	-303.6	-297.6	1.6	-296.0
$[Rb(EDTA)]^{3-}$	-296.5	1.3	-295.2	-288.7	1.3	-287.4
$[Mg(EDTA)]^{2-}$	-881.4	5.1	-876.3	-866.2	5.2	-861.0
$[Ca(EDTA)]^{2-}$	-786.0	4.1	-781.9	-775.9	4.2	-771.7
$[Sr(EDTA)]^{2-}$	-733.1	3.4	-729.7	-723.0	3.5	-719.5
$[Na(H_2O)_6]^{+}$	-117.0	15.2	-101.8	-94.3	14.7	-79.6
$[K(H_2O)_6]^{+}$	-96.2	14.7	-81.5	-76.9	13.9	-63.0
$[Rb(H_2O)_6]^{+}$	-91.3	14.6	-76.7	-72.5	14.0	-58.5
$[Mg(H_2O)_6]^{2+}$	-329.1	14.9	-314.2	-299.6	14.5	-285.1
$[Ca(H_2O)_6]^{2+}$	-250.5	12.7	-237.8	-227.9	12.7	-215.2
$[Sr(H_2O)_6]^{2+}$	-210.6	11.7	-198.9	-191.0	11.7	-179.3

Selected bond lengths and the C-N-N-C dihedral angle for the optimized structures of ligands and complexes for the four sets of calculations are shown in Table 4.6. The calculated B3LYP/AE, GP86/AE, B3LYP/MCP, and GP86/MCP geometries exhibit great similarity. The bond lengths and dihedral angles are also given for the crystal structures that were available. However, it is difficult to compare the calculated gas phase structural parameters to the crystal structure data because in all crystal structures there are other cations, water molecules, or other complexes present. The presence of these extra atoms/molecules as well as the crystal packing effects will affect not only the dihedral angle, but also the bond lengths. However, Durand et al. [81] showed in molecular dynamics simulations that the cation-oxygen and cation-nitrogen bond distances are comparable with calculated geometries in the present work. Specifically, Durand et al. found that the cation-oxygen distances are 2.3 Å and 2.5 Å for Ca^{2+} and Sr^{2+} respectively and the cation-nitrogen distances are 2.5 Å and 2.6 Å for Ca^{2+} and Sr^{2+}, respectively, which is

Table 4.5: MCP stabilization energies of EDTA^{4-} and hexaaqua complexes (in kcal/mol).

Ion-Complex	B3LYP			GP86		
	ΔE_0	ΔZPE	ΔE_{corr}	ΔE_0	Δ ZPE	ΔE_{corr}
[Na(EDTA)]$^{3-}$	-351.0	2.3	-348.7	-338.1	2.2	-335.9
[K(EDTA)]$^{3-}$	-319.5	1.7	-317.8	-305.2	1.6	-303.6
[Rb(EDTA)]$^{3-}$	-318.8	1.2	-317.6	-293.3	0.9	-292.4
[Mg(EDTA)]$^{2-}$	-913.3	5.2	-908.1	-896.0	5.3	-890.8
[Ca(EDTA)]$^{2-}$	-799.9	4.4	-795.5	-785.7	4.4	-781.3
[Sr(EDTA)]$^{2-}$	-756.0	3.3	-752.7	-738.2	3.3	-734.9
[Na(H$_2$O)$_6$]$^{+}$	-123.4	14.3	-109.1	-97.6	13.8	-83.8
[K(H$_2$O)$_6$]$^{+}$	-100.8	14.0	-86.8	-75.8	13.3	-62.5
[Rb(H$_2$O)$_6$]$^{+}$	-101.3	13.6	-87.7	-71.3	13.2	-58.1
[Mg(H$_2$O)$_6$]$^{2+}$	-342.8	14.7	-328.1	-310.2	14.2	-296.0
[Ca(H$_2$O)$_6$]$^{2+}$	-248.0	12.4	-235.6	-221.5	12.3	-209.2
[Sr(H$_2$O)$_6$]$^{2+}$	-214.1	11.3	-202.8	-187.2	11.1	-176.1

within 0.1 Å of this works calculated values which are shown in Table 4.6.
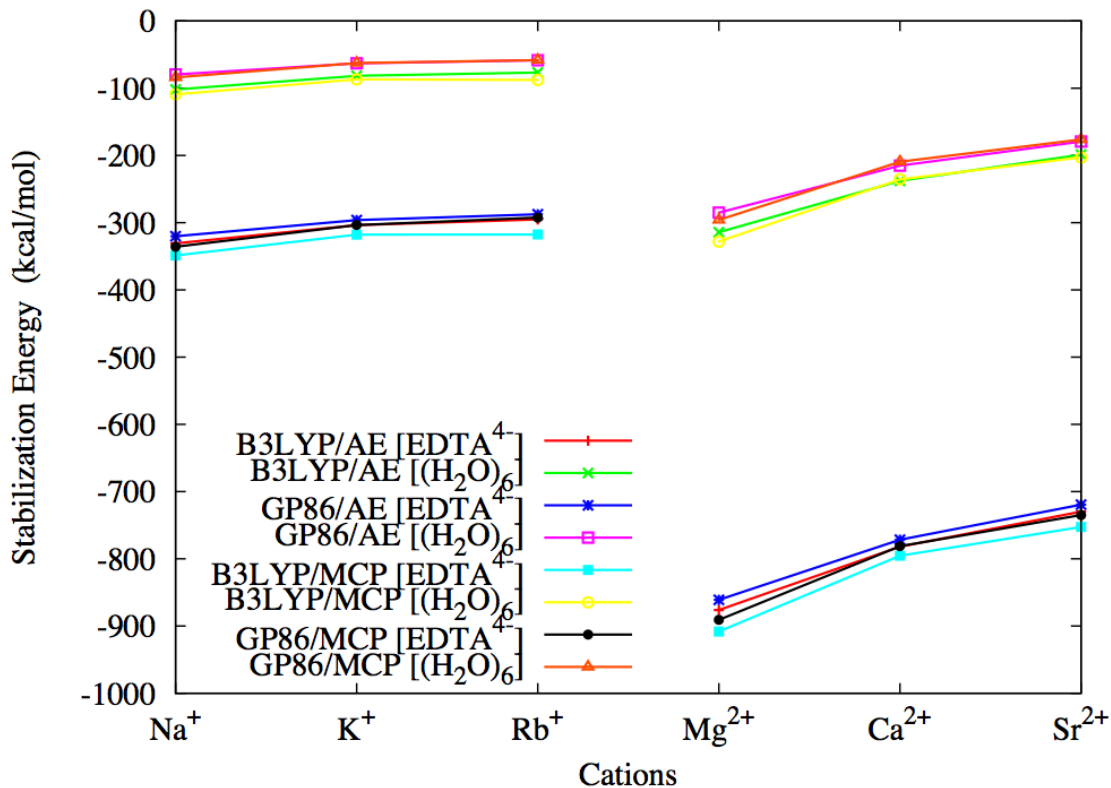
Figure 4.4: Stabilization energies (in kcal/mol)

The Natural Bond Orbital analysis (NBO) [82, 83] was carried out for both cation-$(H_2O)_6$ and cation-EDTA^{4-} complexes, and the obtained atomic charges are summarized in Table 4.7. It may be noticed that cationic charge was reduced by approximately 10% for all cations in both cation-$(H_2O)_6$ and cation-EDTA^{4-} complexes. This indicates that there is little transfer of charge between the ligand and the cation and thus the bonding is mainly electrostatic.

Table 4.6: Selected bond distances (Å) and C-N-N-C dihedral angle (degree) in EDTA^{4-} complexes for AE and MCP basis sets (MCP values in parenthesis); X-ray experimental results shown in italics.

Cation	B3LYP			GP86		
	O-M	N-M	C-N-N-C	O-M	N-M	C-N-N-C
	AE/MCP	AE/MCP	AE/MCP	AE/MCP	AE/MCP	AE/MCP
Na$^+$	2.373	2.578	69.276	2.428	2.559	71.142
	(2.285)	(2.612)	(68.908)	(2.330)	(2.575)	(70.421)
	2.464	*2.534*	*57.50*			
K$^+$	2.894	2.791	72.897	2.951	2.772	75.514
	(2.779)	(2.746)	(71.357)	(2.834)	(2.731)	(73.578)
Rb$^+$	3.017	2.909	76.462	3.057	2.883	78.915
	(2.934)	(2.859)	(74.262)	(3.041)	(2.921)	(79.055)
Mg^{2+}	2.085	2.318	59.027	2.105	2.301	59.584
	(2.031)	(2.280)	(57.736)	(2.050)	(2.255)	(57.977)
	2.182	*2.396*	*57.39*			
Ca^{2+}	2.364	2.559	63.045	2.331	2.536	63.383
	(2.344)	(2.538)	(62.115)	(2.356)	(2.519)	(62.683)
	2.408	*2.67*	*57.83*			
Sr^{2+}	2.554	2.727	66.384	2.562	2.700	67.243
	(2.521)	(2.691)	(65.054)	(2.540)	(2.675)	(66.046)

Table 4.7: NBA charges on metal cations.

Ligand and Method	Cation					
	Na$^+$	K$^+$	Rb$^+$	Mg^{2+}	Ca^{2+}	Sr^{2+}
EDTA^{4-}						
B3LYP/AE	0.911	0.928	0.950	1.792	1.851	1.901
GP86/AE	0.918	0.935	0.934	1.778	1.843	1.900
B3LYP/MCP	0.866	0.903	0.873	1.741	1.809	1.816
GP86/MCP	0.872	0.909	0.935	1.723	1.801	1.833
(H$_2$O)$_6$						
B3LYP/AE	0.958	0.986	0.996	1.826	1.890	1.936
GP86/AE	0.950	0.971	0.989	1.815	1.881	1.933
B3LYP/MCP	0.922	0.964	0.947	1.772	1.867	1.894
GP86/MCP	0.932	0.974	0.988	1.760	1.866	1.916

Chapter 5

DFT Studies of the Electronic Structure and Geometry of 18-Crown-6, Hexaaza[18]annulene, and their Complexes with Cations of the Heavier Alkali and Alkaline Earth Metals*

5.1 Introduction

Ever since their discovery in 1967 [84], crown ethers have been a focus of intense experimental and computational studies. Crown ethers are widely used for their cation-capturing ability [85–87]. This class of host molecules can form complexes with a wide number of cations. The ability of the crown ethers to bind cations is used in industry for the extraction of heavy metal cations from aqueous solutions [85–87]. Crown ethers

---

are also commonly used for the separation of various metal cations from different media [85–87]. Furthermore, crown ethers may stabilize metal cations in organic solvents. In several recent studies the crown ether complexes were employed as a nano-switch [88–90]. The ability to bind metal cations is afforded through the interaction between the metal cation and the lone electrons pairs from the oxygen atoms. Studies of the toxicity of crown ethers indicated the compounds are not genotoxic in mammalian cells despite their cytotoxicity. They have been shown to be highly toxic in prokaryotes and eukaryotes [91]. Calculations of the electronic and geometrical structures of 18-crown-6 (18c6) and their complexes with various cations have been performed [92–95]. Particularly detailed studies were done by Feller and coworkers, who studied the complexes of 18c6 with both light and heavy Group 1 and 2 cations using increasingly more sophisticated levels of theory [69, 70, 96–99].

18c6 was shown to have different symmetry in different solvents [100–104]. Monte Carlo and molecular dynamics simulations demonstrated that 18c6 prefers the prolate C_i conformation in non-polar solvents; this is the same conformation that is observed in crystals [105, 106]. In a polar solvent 18c6 prefers the D_{3d} symmetry [100–104]. The structural differences between these two conformers are shown in Figure 5.1. It is possible to perform molecular dynamics calculations in either polar (H_2O) or non-polar ($CHCl_3$) solvent to view this conversion between the D_{3d} and C_i conformations of 18c6.

Experimental evidence shows different cation selectivity of crown ethers in the gas phase as compared to solvated models [78, 107–109]. The ionic diameters of Rb^+, Cs^+, Sr^{2+}, and Ba^{2+} are 2.96 Å, 3.38 Å, 2.26 Å, and 2.90 Å, respectively [107], while the cavity size of 18c6 measured as the O-O distance is 2.6-3.2 Å [78]. Maleknia and Brodbelt [78] suggested that in the absence of solvent the prediction of the binding affinity for the cations is described best by the "maximum contact point" concept. The maximum contact point
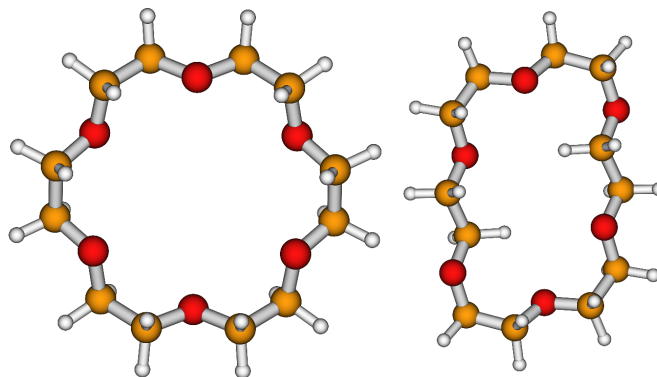
Figure 5.1: Structures of the D_{3d} and C_i configurations of 18-crown-6

concept predicts that 18c6 will have preference for slightly smaller cations because this allows for higher electric field-dipole interaction [78]. This concept explains the trend found in our results, with the stability of the series $Sr^{2+} > Ba^{2+} > Rb^{+} > Cs^{+}$, which was corroborated by experimental evidence [78, 109]. A different ordering of Rb^{+} and Cs^{+}, that was also found [108], may be attributed to the type of experiment used to evaluate the selectivity of the crown ethers [110]. The ordering $Cs^{+} > Rb^{+}$ could result from better size matching of the cation to the ring structure and thus may mimic solution trends. The two papers [78, 109] that give the ordering that agrees with our results, $Sr^{2+} > Ba^{2+} > Rb^{+} > Cs^{+}$, used mass spectrometry and should give results that are closer to the ones obtained in the computational studies of isolated complexes.

The unsaturated macrocyclic polyenes were synthesized in 1959 [111–113]. However, few computational studies of the hexaaza analogs of [18] annulene, $C_{18}H_{18}$, with the simplest of them being $C_{12}H_{12}N_6$, (hexaaza[18]annulene), and corresponding metal complexes were done, even though hexaaza[18]annulene (6aa) has a better configuration for binding cations. The structures of the two molecules, 18c6 and 6aa, are compared in Figure 5.2. In 18c6, the six atoms closest to the centre of the metal-binding cavity are oxygens; in 6aa, the six atoms are nitrogens that replaced the inner six CH groups of

[18] annulene. The cavity diameters of the two structures are remarkably similar: 5.8 Å for 18c6 and 5.5 Å in 6aa (measured as the O-O distance and N-N distance, respectively). In 18c6, out of the six oxygen atoms only every second atom has a donor electron pair that points towards the centre of the ring where the cation is captured. In contrast, 6aa has six nitrogen atoms with all of their donor pairs pointing towards the centre of the ring. The cation thus has a greater access to the donor pairs of the nitrogen atoms which might lead to a greater binding affinity. The X-ray crystal structure of a macrocycle analogue of 6aa, in which four benzene rings are fused at the periphery of 6aa (torands), was determined [114]. Since 1984 Bell studied the complexing capabilities of compounds similar to the one studied crystallographically: he found that these torands have strong binding affinity for several metal ions [115–120]. In particular, he showed that the stability of the potassium complex with an analogue of 6aa is an order of magnitude greater than that of 18c6 [115].
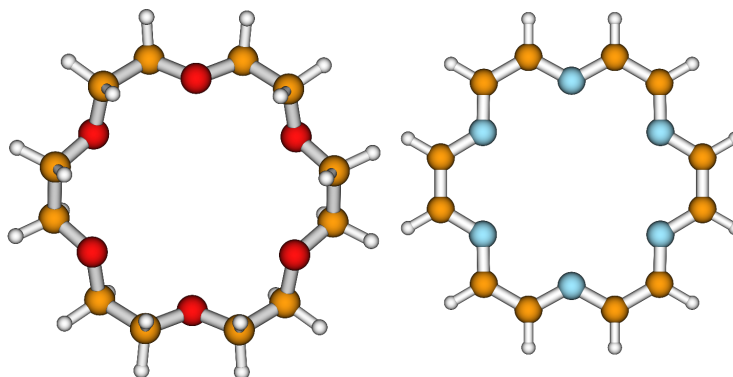
Figure 5.2: Structures of 18-crown-6 (left) and hexaaza[18]annulene (right)

Early computational studies of 6aa include the Hartree-Fock calculations in the π-electron approximation [121]. An aza-derivative of [18] annulene, in which nine of the CH groups were replaced by nitrogen atoms to form an alternating chain of C-N bonds, was studied computationally by Haddon [122] at the Hartree-Fock level with small basis

sets. From Table 5.1 it is apparent that when RHF methods is used there is an alternating chain of C-N bonds, but when DFT is used the C-N bond lengths are all equal. Kao used molecular mechanics based on molecular orbitals method [123] to find that 6aa is planar [124].

It is the purpose of the present work to evaluate the stabilization energies of complexes formed by 6aa with heavier alkali and alkaline earth cations; these values will be compared with those found in 18c6 complexes.

5.2   Method of Calculation

The complexes formed by 18c6 and 6aa with metal cations are flexible structures that require a substantial amount of computational time. A pseudopotential method was needed to approximate the core electrons and explicitly treat only the valence electrons. In the present work we used the model core potential (MCP) method [76]. The MCP method offers an effective way of reducing the size of the basis set, which reduces the amount of CPU time, while at the same time reducing the magnitude of the basis set superposition error caused by inadequate basis sets used for the core electrons. We used the newly developed MCP-DZP and MCP-TZP parameters and basis sets [77]; the quality of these MCP's is comparable to that of all-electron correlation consistent basis set cc-pVDZ [125] or ECP basis sets [126, 127]. The MCP parameters and basis sets for heavier atoms incorporate the scalar relativistic effects [25].

The effects of electron correlation were included via the density functional theory with the B3LYP functional [51, 54]. Geometry optimization at the B3LYP/MCP-DZP level was followed by the evaluation of the hessian in order to ascertain that the stationary points found corresponded to minima at the potential energy surface. Single-point ener-

gies were calculated using the MCP-TZP basis set at the geometries optimized with the smaller MCP-DZP basis set. The stabilization energies, defined as the difference between the energy of the complex and the combined energies of the free ligand and metal ion, included the zero-point energies evaluated at the B3LYP/MCP-DZP level. The basis set superposition errors (BSSE) in the stabilization energies were corrected using the full counterpoise method (CP) [128]. Atomic changes were evaluated using the Natural Population Analysis (NPA) [82, 83], which is less dependent on the basis set than the frequently used Mulliken population analysis [129]. Direct SCF method was used, as the number of basis functions for MCP-DZP basis were 312, 320, 372, and 380 for 6aa, M@6aa, 18c6, and M@18c6, respectively. For the MCP-TZP basis set, the molecular basis set size increased to 792, 810, 972, and 990 functions.

In all calculations the GAMESS [1] program was used, running concurrently on computer clusters at the University of Alberta. Visualization of molecular structures, vibrational modes, and molecular orbitals was accomplished using the MacMolPlt [6] and Molden [5] programs.

5.3   Results and Discussion

In our calculations, the gas-phase C_i conformation was found to have a lower total energy by only 1.2 kcal/mol compared to the D_{3d} conformation. In preliminary geometry optimization on the complexes we found even if the initial structure of the complex was assumed to be of C_i symmetry, the optimized structure had higher symmetry; in subsequent studies we started geometry optimization from a high-symmetry structure, D_{3d} for 18c6 and D_{6h} for 6aa.

For the free ligand 6aa, we investigated the possibility of bond alternation that was

studied for the neutral [18] annulene [130–133]. Using the MCP-DZP basis set, we found that the symmetry of the stable structures of 6aa differs at the RHF and DFT levels of theory, with the DFT-optimized structure showing no bond alternation. Selected bond lengths of 6aa are shown in Table 5.1. Both structures were minima, RHF giving C_{3h} symmetry and DFT – D_{6h}; in both cases the lowest frequencies correspond to the bending modes of the molecular plane.

Table 5.1: Selected bond bond lengths (in Å) of 6aa from RHF and B3LYP calculations

Bond	RHF	DFT
NC	1.253; 1.382	1.323
CC	1.345; 1.491	1.418
CN	1.253; 1.382	1.323

Selected bond lengths in the optimized structures of free ligands and complexes are shown in Tables 5.2 and 5.3 for 18c6 and 6aa, respectively.

Table 5.2: Selected structural parameters (in Å) of 18c6 in the gas phase and in complexes

Cation	18c6				
	symm	CC	CO	OO	d
none	D_{3d}	1.514	1.412	2.937	5.800
Rb^+	D_{3d}	1.515	1.428	2.877	5.709
Cs^+	C_{3v}	1.514	1.426	2.875	5.705
Sr^{2+}	C_{3v}	1.507	1.442	2.713	5.423
Ba^{2+}	C_{3v}	1.514	1.444	2.807	5.599

In the case of 18c6, the relatively large Rb^+ is the the only cation that binds in the centre of the host ring. The other guest cations do not remain at the centre: they are displaced from the plane of the ring and the symmetry is reduced to C_{3v}. Complexation leads to the lengthening of the CO bond, reduction of the cavity's diameter d, and ensuing

Table 5.3: Selected structural parameters (in Å) of 6aa in the gas phase and in complexes

Cation	6aa				
	symm	CC	CN	NN	d
none	D_{6h}	1.418	1.323	2.752	5.504
Rb^+	C_{6v}	1.419	1.329	2.769	5.538
Cs^+	C_{6v}	1.419	1.329	2.767	5.535
Sr^{2+}	D_{6h}	1.416	1.336	2.750	5.500
Ba^{2+}	C_{6v}	1.419	1.337	2.776	5.552

reduction of the distance between adjacent oxygens.

Experimental results indicate that while Rb^+ has a favorable interaction with 18c6, Cs^+ has a less favoured interaction [78]. Maleknia and Brodbelt explain that the less favorable interaction of Cs^+ is probably due to the difficulty of Cs^+ to maximize interaction of the cation and crown, while simultaneously minimizing the van der Waals repulsion between cation and crown [78].

In the case of 6aa, only the smallest cation Sr^{2+} binds in the centre of the host cavity. The other guest cations are displaced from the plane of the ring, and the symmetry of the complex is reduced to C_{6v}. In contrast to 18c6, the rigid framework of the 6aa ring does not allow for significant reorganization of molecular geometry upon complexation and there is only slight variation in the CC and CN bond lengths; furthermore, the diameter of the cavity d increases when complexation leads to reduction of symmetry.

The displacements are summarized in Table 5.4. For 6aa, the displacement (δ) is the distance along the symmetry axis to the plane of the six nitrogen atoms. For the 18c6, the distance was measured from the intersection point of the lines connecting opposite oxygen atoms.

Table 5.4: Displacement of cation δ (in Å)

Cation	18c6		6aa	
	symm	δ	symm	δ
Rb^+	D_{3d}	0.000	C_{6v}	1.023
Cs^+	C_{3v}	2.706	C_{6v}	2.190
Sr^{2+}	C_{3v}	0.103	D_{6h}	0.000
Ba^{2+}	C_{3v}	0.253	C_{6v}	0.842

An argument based on experimental results [134–136] can be used to explain the phenomenon of cations being out of the plane of the crown ether ring. The crystal structures of the metal cations with crown ethers show that the out-of-plane distance of the cation to the crown ether molecule increases with the size of the cation [134–136]. This supports our finding that the metal cations are displaced from the center of the rings. It is hypothesized that Rb^+ is in the center of the ring due to its smaller size and ability to maximize the Rb^+ 18c6 interaction while it minimizes the Rb^+ 18c6 van der Waals repulsion [78].

The stabilization energies and their components for the two families of complexes are shown in Tables 5.5 and 5.6. In these Tables, ΔE_0 denotes the energy difference at the optimized minima, ZPE is the zero-point-energy correction, BSSE stands for the correction from the counterpoise method, and ΔE is the stabilization energy corrected for both zero-point-energy and basis set superposition error. The ZPE correction was evaluated at the B3LYP/MCP-DZP level.

The BSSE correction is rather small and fairly constant for all the studied species. The larger basis set MCP-TZP increases stability for both 18c6 and 6aa complexes. For both 18c6 and 6aa, the stabilization energies decrease with increasing size of the ion, following the pattern found by Feller [69]. As expected, the stabilization energies are larger

Table 5.5: Contributions to stabilization energies from B3LYP/MCP-DZP calculations (in kcal/mol)

Cation	18c6				6aa			
	ΔE_0	ZPE	BSSE	ΔE	ΔE_0	ZPE	BSSE	ΔE
Rb^+	-65.9	2.6	2.5	-60.8	-75.3	2.4	3.1	-69.8
Cs^+	-43.0	2.3	2.6	-38.2	-50.1	1.9	3.8	-44.4
Sr^{2+}	-184.2	4.2	3.6	-176.4	-207.5	4.6	3.7	-199.2
Ba^{2+}	-155.5	3.5	3.4	-148.6	-169.7	4.0	4.0	-161.7

Table 5.6: Contributions to stabilization energies at the B3LYP/MCP-TZP//B3LYP/MCP-DZP level of theory (in kcal/mol)

Cation	18c6				6aa			
	ΔE_0	ZPE	BSSE	ΔE	ΔE_0	ZPE	BSSE	ΔE
Rb^+	-71.2	2.6	3.2	-65.4	-76.6	2.4	2.3	-72.0
Cs^+	-47.1	2.3	3.4	-41.4	-51.9	1.9	3.7	-46.3
Sr^{2+}	-200.2	4.2	3.0	-193.1	-218.7	4.6	2.1	-212.0
Ba^{2+}	-181.4	3.5	3.0	-174.9	-190.2	4.0	2.4	-183.8

for the 6aa complexes. The difference in stabilities is reduced for the TZP basis set as the result of preferential increase of stability of the 18c6 complexes. It is worth noting that the cations are bound to 6aa more strongly than in the case of 18c6, even with the displacement of the guest cation out of the plane of the ring that may indicate some structural strain due to the increased size of the guest ion.

The NPA charge distributions in free ligands and their complexes, with the charges on hydrogens added to the charges of the attached carbon atoms, are shown in Table 5.7. For 18c6 in the lower symmetry (C_{3v}), the charges were averaged for methylene and oxygen. As anticipated, there is but a small variation in the computed atomic and group charges when the basis set is changed from DZP to TZP. As the comparison with the charges in the free ligands shows, presence of a cation induces polarization of electron

distribution in both ligands. Analysis of data in Table 5.7 reveals transfer of electrons from methylene group to oxygen upon complexation, with subsequent donation of electrons to the cation and reduction of cation's formal charge. The bivalent cations polarize charge more than the univalent ones. In the case of the univalent cations, the charge polarization is very similar for both 18c6 and 6aa complexes; for bivalent cations, 6aa exhibits slightly larger redistribution of charge.

Table 5.7: Atomic and group charges from Natural Population Analysis (NPA) using DZP basis set (TZP in parenthesis)

Cation	18c6			6aa		
	M	O	CH_2	M	N	CH
none	—	-0.57(-0.53)	0.29(0 .27)	—	-0.39(-0.35)	0.20(0.18)
Rb^+	0.89(0.87)	-0.62(-0.57)	0.32(0.30)	0.88(0.87)	-0.46(-0.42)	0.24(0.22)
Cs^+	0.95(0.93)	-0.61(-0.56)	0.31(0.29)	0.95(0.94)	-0.46(-0.42)	0.23(0.21)
Sr^{2+}	1.87(1.84)	-0.68(-0.63)	0.35(0.33)	1.81(1.80)	-0.55(-0.51)	0.29(0.27)
Ba^{2+}	1.89(1.84)	-0.68(-0.62)	0.35(0.33)	1.85(1.82)	-0.54(-0.51)	0.28(0.27)

The harmonic vibrational analysis performed at the B3LYP/MCP-DZP level of theory allows for simulating vibrational infrared spectra. The computed spectra for the free ligands and gas-phase complexes are shown on Figure 5.3; these IR spectra were generated using gnuplot [137], assuming a Lorentzian distribution and a half-width of 25 cm^{-1}. In order to determine if the addition of a counterion would change the IR spectra, simulations were done using K^+-18c6, KBr-18c6, and KSCN-18c6 complexes: only minute differences between the three IR spectra were observed, therefore counterions were not utilized to analyze the spectra. The spectra for each family (18c6 and 6aa) show similar behaviour: the intensity of the high-frequency modes (comprised of the carbon-hydrogen stretches) is quenched, with the extent of intensity reduction dependent on the charge of the guest cation; the spectra in that region are slightly blue-shifted. A blue-shift is caused by a strengthening of the bond corresponding to the IR peak. The bond lengths

of the 18c6 and 6aa decrease from approximately 1.110 Å to 1.105 Å to 1.100 Å for 18c6/6aa to 18c6M^{1+}/6aaM^{1+} to 18c6M^{2+}/6aaM^{2+}, respectively, which corresponds to the bond strengthening and therefore the blue shifting of the C-H stretching region of the IR spectra. Also, due to the strengthening of the C-H bond, there is also a reduction in the change in electric dipole during the C-H stretching, which corresponds to the quenching of the C-H bond stretching mode upon the addition of cations. The deformations of the rings (including bends of ring bonds, methylene groups, CH groups and bond stretches) in the $800 - 1500$ cm^{-1} region retain their intensity while being minimally red-shifted.

5.4   Conclusion

Results of B3LYP calculations using MCP-DZP and MCP-TZP basis sets show that Rb$^+$, Cs$^+$, Sr^{2+}, and Ba^{2+} the cations bind more strongly to 6aa than to 18c6. The cations were bonded with affinity greater by about 8-23 kcal/mol, even in the cases where the cations were displaced from the plane of the ring. Complexation of the ligands affects significantly more the geometry of the flexible 18c6 than that of the more rigid 6aa. Presence of the cation induces polarization of electron charge distribution within ligands, with the CH$_2$ and CH groups donating electron density to the lone-pair atoms (oxygens and nitrogens) and these in turn transferring some of the charge to the cation. Computed infrared spectra show that complexation leads to quenching of the carbon-hydrogen stretching modes.
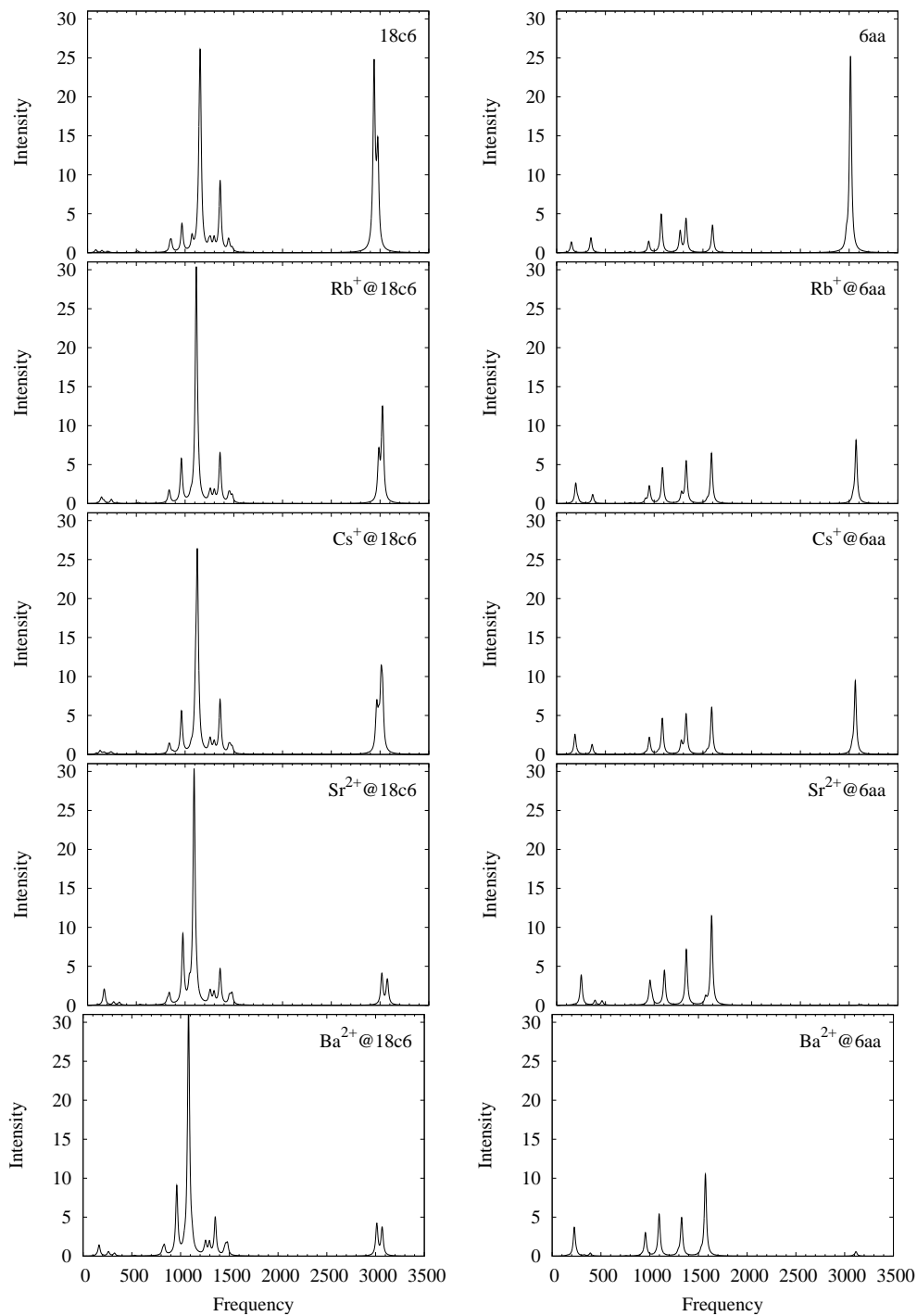
Figure 5.3: Simulated infrared spectra of 18c6 and 6aa systems. Frequencies in cm^{-1}, intensity in Debye^2/(amu Å^2).

Chapter 6

Cancer and the Cell

6.1 Introduction

Cancer has touched the lives of most of the population either directly or indirectly. If cancer is not caught during its early stages it can lead to the reduction of life expectancy. Cancer is a debilitating disease that has taken many lives over the decades.

Cancer is a disease that targets cells, which are the building blocks of the human body. Normal cells grow and divide according to a regulated cycle, know as the cell cycle. The cell cycle is depicted in Figure 6.2 and a brief overview will be given in this Chapter. At the onset of cancer the normal cells mutate due to some unknown factor(s) and the cells begin to grow and divide at an unregulated (very fast) rate. Cancerous cells tend to group together and form clusters, which are known as tumors. Tumors which are growing can kill healthy cells and tissue that is nearby.

Cancer spreads when the cancerous cells detach from the tumor and move to other areas of the body, and start dividing out of control again and form new tumors. The spread of a tumor to another location in the body is known as metastasis. Once cancer

has metastasized it is much more difficult to successfully treat.

Cancer can be treated with surgery, radiotherapy, chemotherapy, as well as combination treatments. Treatment depends on the type and location of the tumor. When surgery is performed it requires an invasive procedure in which both healthy cells and tissues are removed to insure that all of the cancer cells have been resected. Radiotherapy uses high energy electromagnetic waves, such as X-rays, to kill cancer cells. Radiotherapy damages both cancer and healthy cells, but healthy cells can recover from the treatment, whereas cancer cells are damaged in a manner that does not allow them to continue to grow and divide.

Chemotherapy uses drugs to treat cancer. These drugs target cells which are dividing and growing the most rapidly (cancer cells). Although some healthy cells are killed in the process, the cancer cells are killed with greater frequency. Table 6.1 gives a list of the most commonly used antimitotic cancer drugs that are in use today along with their target cancer type. Many chemotherapy drugs work by targeting microtubules (microtubules are essential for cellular division) and once the microtubule has been targeted, the cell is no longer able to divide and eventually dies. Current chemotherapy drugs are highly toxic and have unpleasant side effects. Therefore, finding new drugs which better target cancer cells with a reduced affinity towards healthy cells is an important area of research.

6.2   The Cell

A detailed description of all that is summarized in this Chapter can be found in most biology textbooks, some examples are [138], [139], [140], and [141]. The cell is the basic unit of life. It has a membrane that encloses the cell, separating and protecting it from its surrounding. The membrane is selectively permeable to control what enters and leaves

Table 6.1: Some commercially available chemotherapy drugs

Chemotherapy Drug	Brand Name	Type of Cancer
Docetaxel	Taxotere®	breast, prostate, lung
	Vepesid®	
Paclitaxel	Taxol®	breast, ovarian, lung
Vinblastine	Velbe®	leukemia, lymphoma, breast, lung
Vincristine	Oncovin®	leukemia, lymphoma, breast, lung
Vindesine	Eldisine®	leukemia, lymphoma, melanoma breast, lung
Vinorelbine	Navelbine®	breast, lung

the cell. The cell is filled with cytoplasm, which is a viscous aqueous fluid where most of the cellular activities take place. Within the cytoplasm there are many organelles which perform various functions in the cell; the organelles and their roles within the cell are described in Table 6.2.

6.2.1 The Cell Cycle and Microtubules

The cytoskeleton is composed of proteins which form microfilaments, intermediate filaments, and microtubules. Microtubules are an important factor in many processes such as cytokinesis, vesicular transport, and mitosis. Microtubules (depicted in the left image of Figure 6.1) have an outer diameter of approximately 25 nm, and inner diameter of approximately 15 nm, and they vary in length from 200 nm to 25 μm. Microtubules are formed from α/β-tubulin dimers; an α/β-tubulin dimer is depicted in the right image of Figure 6.1. The right image of Figure 6.1 was generated using VMD and the 1TUB structure obtained from the Protein Data Bank (PDB), which has a resolution of 2.7 Å.

Table 6.2: Components of the cell

Organelle	Function
Nucleolus	ribosomal RNA is transcribed and assembled
Nucleus	contains genetic material
Ribosome	assembles proteins from amino acids
Vesicle	small membrane-enclosed sack
Rough endoplasmic reticulum	synthesize proteins
Golgi apparatus	processes and packages macromolecules
Cytoskeleton	provides the cell with structure and shape
Smooth endoplasmic reticulum	synthesize lipids and steroids
Mitochondria	generate most of the cell's energy
Lysosome	break up waste materials
Vacuoles	vesicle which stores food and waste

Mitosis takes place during cell division, and the mitotic cell cycle is depicted in Figure 6.2. During the cell cycle the cell starts in G1 phase or Gap 1 phase, and is marked by an increase in cell size and a high biosynthesis of ribonucleic acid (RNA). The G1 checkpoint is a control mechanism that ensures the cell is ready for deoxyribonucleic acid (DNA) synthesis. The G1 phase is immediately followed by the S phase or Synthesis phase, when DNA replication occurs. G2 phase or Gap 2 phase occurs next and it is during this phase that the cell continues to grow and α/β-tubulin dimers come together to form microtubules. The G2 checkpoint ensures that the cell is ready to enter mitosis. The cell then enters the mitotic phase.

The mitotic phase consists of four sub-phases: prophase, metaphase, anaphase, and telophase. During prophase, the nuclear membrane disintegrates and the nucleolus disappears. The chromosomes which were in the nucleolus navigate to the center of the enlarged cell and microtubules begin to grow in length towards the center of the cell. It
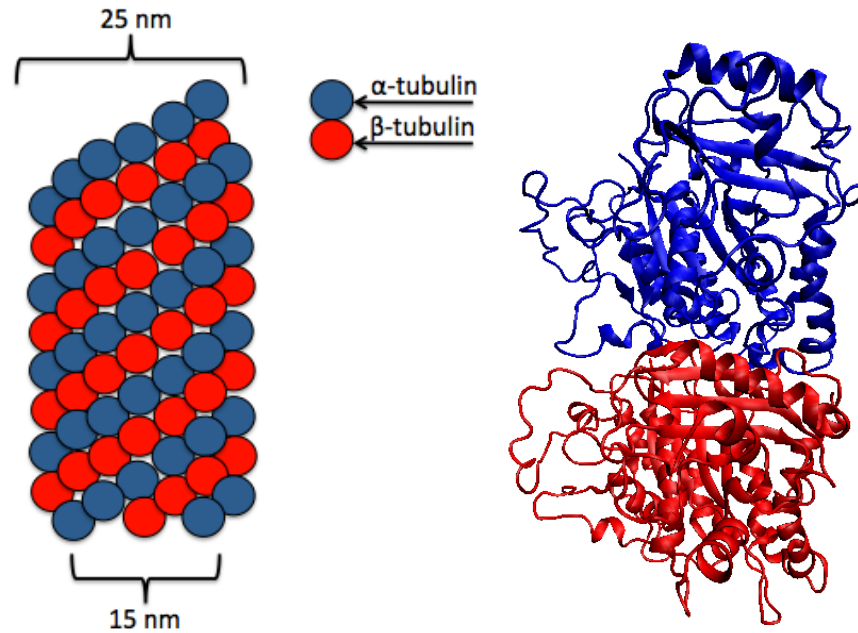
Figure 6.1: Structure of microtubule (left) and α/β-tubulin (right)

is then, during metaphase, that the microtubules attach themselves through kinetochore to the chromosomes. The final checkpoint (metaphase checkpoint) ensures that the cell is ready for complete cell division into two daughter cells. Once the microtubules are attached to the chromosomes, the cell enters anaphase and the chromosomes are divided and pulled to either end of the cell with the aid of the microtubule dynamics. Upon completion, the cell starts telophase and the cell membrane begins to pinch down in the middle of the cell until the cell divides into two (two daughter cells are created). After the cell cycle is complete each new cell either reenters the G1 phase or goes into the resting phase G0. In the G0 phase no part of the cellular division process is taking place and the cell rests until it is ready to reenter the cell cycle in the G1 phase. If the cell is damaged in any way the cell stays in the G0 phase and eventually dies. Cancer cells never enter the G0 phase and this leads to an over production of cancer cells.
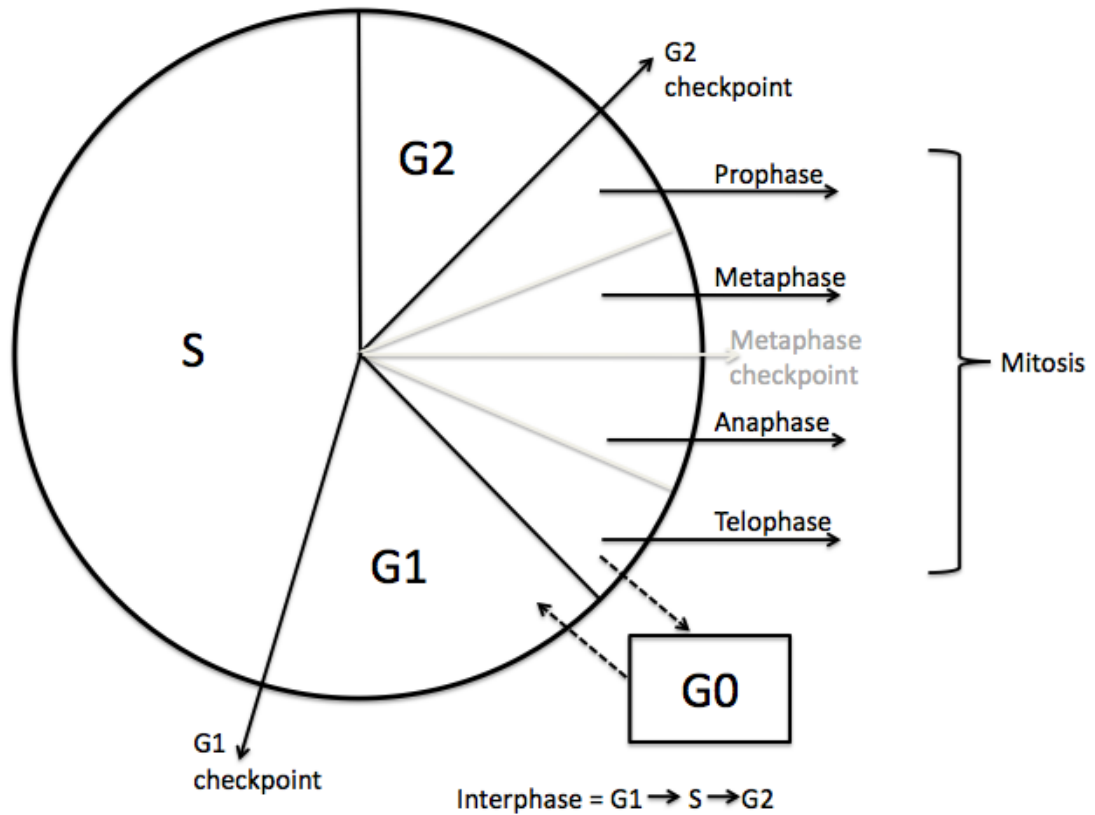
Figure 6.2: Cell cycle

6.2.2   Antimitotic Drugs and α/β-tubulin

Antimitotic drugs interfere with the cell cycle, specifically during mitosis when micro-tubule formation and dynamics are essential for correct cell division. Antimitotic drugs such as paclitaxel, laulimalide, and peloruside are microtubule stabilizing agents. The drugs bind to microtubules and prevent the microtubule from depolymerizing into the α/β-tubulin dimers. The depolymerization/polymerization dynamics of microtubules is essential in cell division, specifically when the chromosomes are being pulled to either side of the cell in anaphase. Another class of antimitotic drugs consists of vinblastine, vincristine, and vindesine; these drugs bind to microtubules and cause the microtubules

to depolymerize into the α/β-tubulin dimers. Without microtubules, the cell is unable to complete the cell cycle and therefore goes into apoptosis (cell death). The last type of antimitotic behaviour is when the drug binds to the α/β-tubulin dimer and prevents tubulin from polymerization to form microtubules; an example of this class of drugs is colchicine. Similarly to the destabilizing drugs, if the microtubules cannot form, the cell cannot undergo division and apoptosis occurs. X-ray and NMR determined structures of several drugs can be found in the PDB and the interested reader is directed there for further information on the structures of the tubulin/drug complexes.

Chapter 7

Stability and Torsional Flexibility of Noscapine and its Halogenated Derivatives

7.1 Introduction

Noscapine ((3S)- 6,7-Dimethoxy-3-[(5R)-5,6,7,8-tetrahydro-4-methoxy-6-methyl-1,3-dioxolo(4,5-g)isoquinolin-5-yl]-1(3H)-isobenzofuranone), is a benzylisoquinoline alkaloid derived from plants of the Papaveraceae (poppy) family. Noscapine has been used as an antitussive (cough suppressant) agent for over half a century; however, its recent usage as an antimitotic agent made it a focus of attention. Noscapine binds to microtubules [142] and exhibits antimitotic properties that could lead to a new class of less toxic anti-cancer drugs. Noscapine was shown to be effective in treating many forms of cancer such as breast [143], small cell lung [144], colon [145], prostate [146], drug resistant cell lines, and drug resistant lymphomas [147]. Experimental studies of halogenated derivatives of noscapine have also been performed revealing that the halogenated complexes have an increased antimitotic activity in several cell lines [142]. The chlorinated and brominated noscapine derivatives show particularly interesting properties. The chlorinated derivative shows promising results against resistant ovarian cancer cell [148], and the

brominated noscapine has shown to be particularly effective as an antimitotic [149] and as an anti-inflammatory agent [150]. Due to the flexible nature of noscapine and the lack of crystal structure data of noscapine bound to the tubulin protein, it is imperative that possible conformations of noscapine be investigated in order to obtain an understanding of the possible binding modes that could occur both in vivo and *in vitro*.

Noscapine (Figure 7.1) has two chiral carbons that connect two ring structures (meconine and hydrocotarnine$^+$; depicted in Figure 7.2). A racemic mixture of RS and SR (α-noscapine) enantiomers [151] forms during the synthesis of noscapine, and therefore both enantiomers are included in this study. Additionally, for the sake of completeness the remaining two enantiomers (SS and RR) are included, leading to four enantiomers (SS, SR, RS, and RR) considered in this study. In addition noscapine is often administered under acidic conditions leading to the protonation of the nitrogen atom. This protonation can occur at either the axial or equatorial position. Using all four possible enantiomers and both protonation sites leads to eight molecules: SS-ax, SS-eq, SR-ax, SR-eq, RS-ax, RS-eq, RR-ax, and RR-eq. These two conformations of the four enantiomers, will be investigated in order to find the lowest energy conformation about the dihedral angle which goes through the two chiral carbons (highlighted in red in Figure 7.1). For each molecule, five derivatives were generated, giving a total of forty molecules whose energies were optimized for several fixed values of the dihedral angle C-C_2^*-C_1^*-N.

7.2 Method of Calculation

All calculations were done using the semiempirical molecular orbital method Recife Model 1 (RM1) [13], which is a reparameterization of the Austin Model 1 (AM1) [12] method. The RM1 method was recently implemented in the GAMESS [1] program and is also implemented in the MOPAC [3] program. All calculations were carried out on
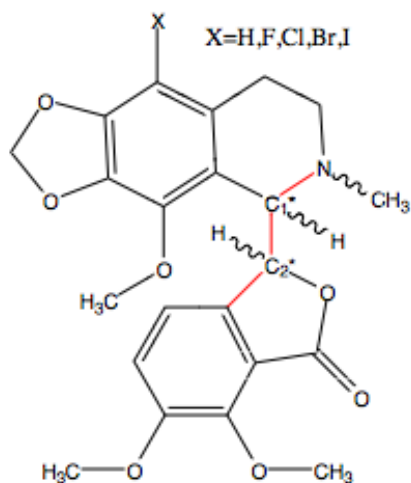
Figure 7.1: Noscapine
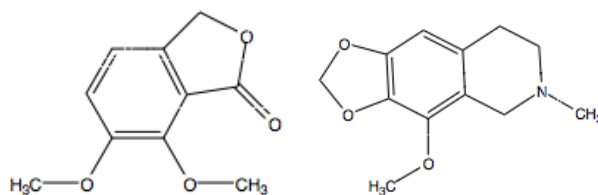
Figure 7.2: Meconine (left) and hydrocotarnine$^+$ (right)

computer clusters at the University of Alberta. Visualization of molecular structures, vibrational modes, and molecular orbitals were done using Molden [5] and MacMolPlt [6].

The first set of calculations was done using the GAMESS program. Geometry optimization at the RM1 level was followed by the evaluation of the hessian in order to ascertain that the stationary points found corresponded to minima or transition states on the slice of the potential energy surface (PES). In the initial geometries the dihedral angle along C-C_2^*-C_1^*-N was between $150°$ and $170°$. The initial value of the dihedral angle was fixed, and partial geometry optimization and hessian calculation were carried out. The dihedral angle was then changed by five degrees, the constraint was placed on

the new value, and geometry optimization followed with the hessian calculations. This process was repeated until the full slice of the PES along the C-C_2^*-C_1^*-N was investigated. The hessian calculations showed that there was only one imaginary frequency, and only around the local maxima corresponding to (local) transition states. The energies, shifted by the lowest energy of the given conformation, are shown in Figures 7.3 and 7.4. The lowest energy value for each conformer and the corresponding dihedral angle at which it occurs are given in Table 7.1.

The second series of calculations was carried out using the MOPAC [3] program where geometry optimization (energy minimization) was performed for each of the constrained dihedral angles. These calculations were again used to create a slice of the PES along the C-C_2^*-C_1^*-N dihedral angle. The first set of MOPAC calculations was done using the RM1 method to ascertain that MOPAC found the same minima as GAMESS. The lowest energy values and the dihedral angles at which it occurs for the MOPAC RM1 calculations are given in Table 7.2 and are shown in Figures 7.5 and 7.6. The change in energy (ΔE) is considered, which is depicted in Figures 7.3, 7.4, 7.5, 7.6 and it is apparent that the same minima were obtained because the energy plots are essentially identical. The second set of MOPAC calculations was performed using the RM1 method together with the conductor-like screening model method (COSMO) [152] solvation model in order to determine how solvation of noscapine affects the geometry minima. The same procedure was used to calculate the slice of the PES by constraining the geometry at the dihedral angle and optimizing with COSMO water (with a dielectric constant of 78.4 at 25°C). The lowest energy values and corresponding dihedral angles from MOPAC RM1/COSMO calculations are given in Table 7.3 and the PES slices are shown in Figures 7.7 and 7.8. The solvation effects on the PES slices are apparent when comparing Figures Figures 7.3 and 7.4, 7.5, and 7.6 to Figures 7.7 and 7.8.

7.3 Results and Discussion

Tables 7.1 and 7.2 summarize the lowest energy of the conformer found in the RM1 calculations about the dihedral angle for GAMESS and MOPAC calculations, respectively. It is interesting to note that all but one of the lowest energy values for a given enantiomer correspond to the same dihedral angle ($\pm$ 10°) regardless which derivative is being investigated or which program is being used. All derivatives of SS-eq have the lowest energy at a dihedral angle of 295°, except the iodinated derivative, which has a lowest energy at 180°. The lowest energy value of SS-eq(I) differs only by 0.04 kcal/mol (7×10^{-5} a.u., see Table 7.1) between the energy at the dihedral angles of 295° and 180°, and thus both could be considered as stable minima.

Table 7.3 gives the lowest energy values and corresponding dihedral angles. For several enantiomers (SS-ax, RS-eq, and SR-ax) the lowest energy geometry does not significantly change ($\pm$ 10°), but for the remaining enantiomers there is a significant change in the lowest energy conformation compared to the gas-phase results. The change in most stable minima is more apparent when comparing Figures 7.3, 7.4, 7.5, and 7.6 to 7.7 and 7.8. These Figures will be discussed in more detail later in this section.

In order to determine how the lowest energy conformations is affected by solvation for the lowest energy conformations of the noscapine derivative in the solvent, a single point calculation of the energy was done in the absence of the solvent. The difference between the two energies defines the solvation energy (given by the difference, $E_{RM1/COSMO}$ - E_{RM1}). The values of the solvation stabilization energies, summarized in Table 7.4, are between 55 and 60 kcal/mol.

Interestingly, for each of the enantiomers the plots for the gas phase and solvated

Table 7.1: Lowest energy conformations of gas-phase noscapine derivatives calculated using GAMESS program and RM1 semi-empirical method. Energy values in a.u., dihedral angles (in parentheses) in degrees

Derivative	H	F	Cl	Br	I
SS-ax	-202.577	-219.955	-216.162	-215.207	-211.192
	(180)	(180)	(180)	(180)	(180)
SS-eq	-202.578	-219.957	-216.164	-215.208	-211.193
	(295)	(295)	(295)	(295)	(180)
SR-ax	-202.574	-219.952	-216.160	-215.204	-211.188
	(175)	(175)	(175)	(175)	(170)
SR-eq	-202.579	-219.958	-216.165	-215.209	-211.194
	(75)	(75)	(75)	(75)	(75)
RS-ax	-202.580	-219.958	-216.165	-215.210	-211.194
	(300)	(305)	(305)	(305)	(305)
RS-eq	-202.575	-219.953	-216.160	-215.205	-211.189
	(170)	(175)	(170)	(170)	(170)
RR-ax	-202.580	-219.958	-216.165	-215.210	-211.195
	(155)	(160)	(160)	(160)	(160)
RR-eq	-202.578	-219.956	-216.163	-215.207	-211.192
	(160)	(165)	(165)	(165)	(175)

molecules (which are shown in Figures 7.3, 7.4, 7.5, 7.6, 7.7, and 7.8), indicate that most of the potential energy energy surfaces along the C-C_2^*-C_1^*-N coordinate are almost identical for each of the derivatives. The enantiomer RR-ax is the sole exception. In this case the hydrogen derivative undergoes a slight shift of the position of the nitrogen at 50°, but all other derivatives (F, Cl, Br, and I) have a corresponding nitrogen shift at 100°. After this nitrogen shift, regardless whether it takes place at 50° or 100°, all derivatives find the lowest energy geometry around 160°.

In some of the slices of the PES there is a sharp increase or decrease in the energy. This rapid change occurs when there is a shift in the position of the nitrogen atom.

Table 7.2: Lowest energy conformations of gas-phase noscapine derivatives calculated using MOPAC program and RM1 semi-empirical method. Energy values in a.u., dihedral angles (in parentheses) in degrees

Derivative	H	F	Cl	Br	I
SS-ax	-202.580	-219.958	-216.166	-215.210	-211.195
	(170)	(170)	(170)	(170)	(170)
SS-eq	-202.581	-219.960	-216.167	-215.211	-211.196
	(295)	(295)	(295)	(295)	(295)
SR-ax	-202.577	-219.955	-216.163	-215.207	-211.192
	(175)	(175)	(175)	(175)	(175)
SR-eq	-202.582	-219.961	-216.168	-215.212	-211.197
	(75)	(75)	(75)	(75)	(75)
RS-ax	-202.583	-219.961	-216.168	-215.213	-211.197
	(300)	(300)	(300)	(300)	(300)
RS-eq	-202.577	-219.956	-216.164	-215.208	-211.193
	(170)	(170)	(170)	(170)	(170)
RR-ax	-202.583	-216.961	-216.168	-215.213	-211.198
	(155)	(155)	(155)	(165)	(155)
RR-eq	-202.580	-219.958	-216.166	-215.210	-211.195
	(160)	(160)	(160)	(160)	(160)

This nitrogen shift is apparent in many of the slices of the PES (e.g. SS-ax, SR-ax, and SR-eq): when there is a significant change in the orientation of the nitrogen, there is also a significant change in the energy (between 10 to 30 kcal/mol). The nitrogen atom is flipping from being in a downward facing positing (toward the meconine ring) to an upward facing position (away from the meconine ring). The flipping of the nitrogen atom is essential at various dihedral angles due to the bulky ring structure that make up noscapine. At various dihedral angles there is steric hinderance due to the methyl group on the nitrogen and considering this steric restraint the nitrogen must shift to accommodate the rotation of these bulky ring structures.

Table 7.3: Lowest energy conformations of noscapine derivatives obtained using MOPAC program and RM1 semi-empirical method with COSMO solvation. Energy values in a.u., dihedral angles (in parentheses) in degrees

Derivative	H	F	Cl	Br	I
SS-ax	-202.668	-220.049	-216.256	-215.300	-211.284
	(160)	(160)	(160)	(165)	(165)
SS-eq	-202.669	-220.051	-216.25	-215.301	-211.281
	(175)	(175)	(175)	(175)	(180)
SR-ax	-202.665	-220.046	-216.252	-215.297	-211.280
	(175)	(175)	(175)	(175)	(175)
SR-eq	-202.669	-220.050	-216.256	-215.300	-211.285
	(190)	(190)	(190)	(195)	(190)
RS-ax	-202.668	-220.050	-216.256	-215.300	-211.284
	(160)	(165)	(165)	(160)	(160)
RS-eq	-202.667	-220.049	-216.255	-215.299	-211.283
	(170)	(165)	(170)	(165)	(165)
RR-ax	-202.669	-220.051	-216.256	-215.300	-211.285
	(185)	(185)	(180)	(185)	(185)
RR-eq	-202.669	-220.050	-216.256	-215.300	-211.284
	(195)	(195)	(195)	(200)	(195)

When comparing the slices of the PES in the gas and solvated phase, it is apparent (Figures 7.3, 7.4, 7.5, 7.6, 7.7, and 7.8) that each of the enantiomers has two distinct minima. These minima are the same for each of of the derivatives within the enantiomeric set. When moving from the gas phase to solvated phase there is a lowering of the potential energy barrier between the two minima of approximately 1 to 3 kcal/mol, which will allow for easier rotation about the dihedral angle when noscapine is in solution. In the case of SR-ax and SR-eq, the global and local minima are swapped and there is again a distinct lowering of the energy barrier. The slices of the PES, by giving possible conformations for noscapine, may give a more detailed understanding of how noscapine in all of its enantiomeric forms and derivatives will behave when administered

Table 7.4: Solvation energies of noscapine derivatives (in kcal/mol)

Derivative	H	F	Cl	Br	I
SS-ax	-56.8	-58.1	-57.5	-57.1	-56.6
SS-eq	-56.3	-57.8	-57.1	-57.0	-55.0
SR-ax	-55.4	-57.3	-56.6	-57.0	-56.2
SR-eq	-58.1	-60.0	-59.2	-59.1	-58.7
RS-ax	-58.1	-59.6	-59.3	-59.4	-58.6
RS-eq	-57.1	-59.3	-58.2	-58.1	-57.8
RR-ax	-56.0	-57.8	-56.6	-56.5	-56.1
RR-eq	-56.6	-58.1	-57.3	-57.6	-56.8

for either antitussive or antimitotic effects.

7.4 Conclusion

Noscapine and its halogenated derivatives have potential to be useful antimitotic drugs and this preliminary study aimed at determining some of the most likely conformations of noscapine which could be used to predict the mode (or modes) of binding to tubulin. We found that both noscapine and its derivatives are flexible molecules which can rotate by 360° about the torsional angle with a barrier that can be as small as 8.5 kcal/mol when in solvent. We predicted that these drug candidates have several stable conformations that could bind to tubulin. Thus, from the perspective of computational drug discovery, it may be a challenging and interesting problem to determine the binding mode (or modes) of noscapine with the tubulin dimer. In contrast to a crystal structure which if known, would probably reveal only one binding mode realized in the solid state; computational simulations may better mimic interactions in the cellular environment.

Figure 7.3: Energy change during 360° rotation of the SS and SR enantiomers of noscapine and the five derivatives using GAMESS program and RM1 semi-empirical method. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.1)

Figure 7.4: Energy change during 360° rotation of the RS and RR enantiomers of noscapine and the five derivatives using GAMESS program and RM1 semi-empirical method. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.1)
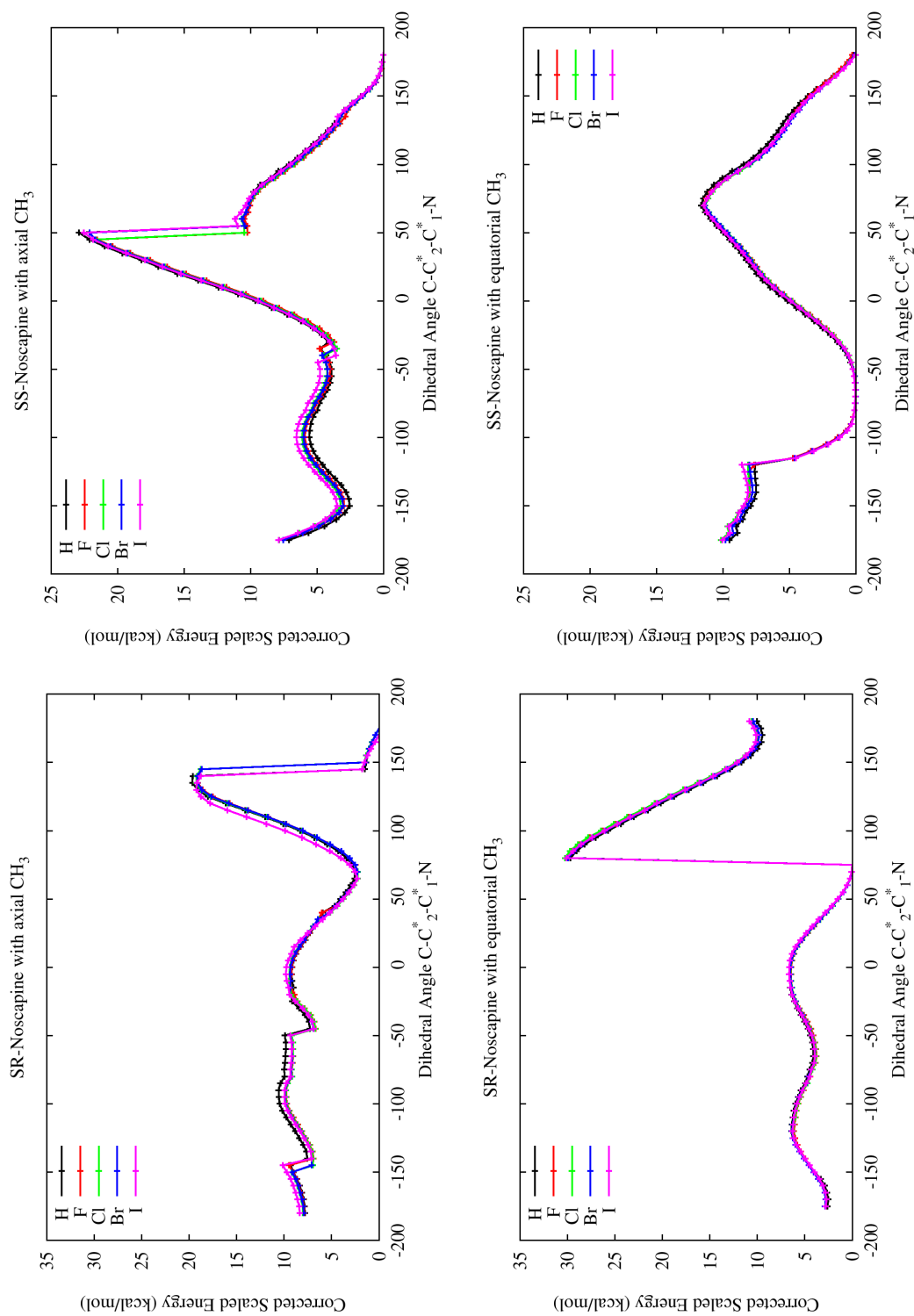
Figure 7.5: Energy change during 360° rotation of the SS and SR enantiomers of noscapine and the five derivatives, using MOPAC program and RM1 semi-empirical method. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.2)

Figure 7.6: Energy change during $360°$ rotation of the RS and RR enantiomers of noscapine and the five derivatives, using MOPAC program and RM1 semi-empirical method. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.2)
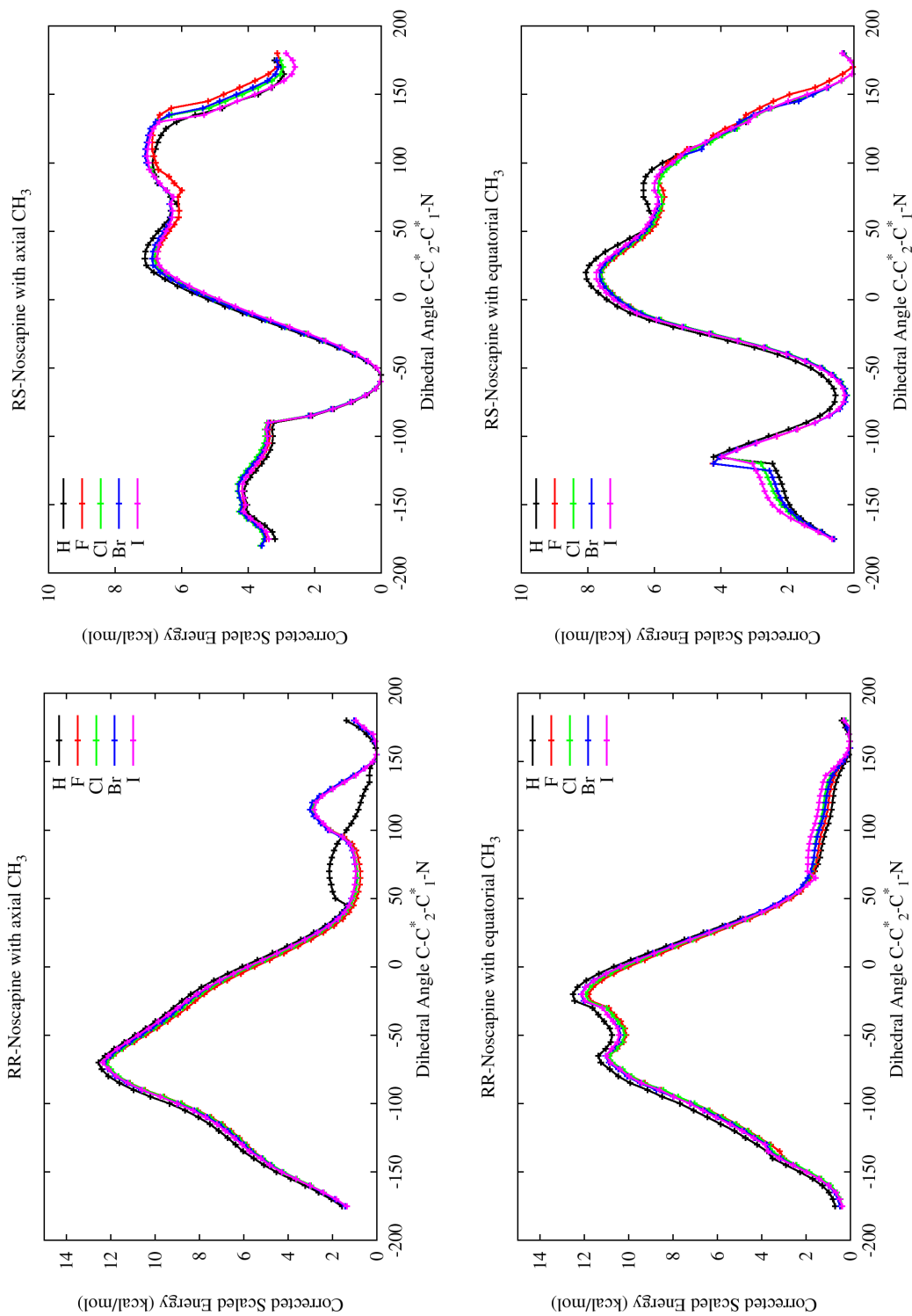
Figure 7.7: Energy change during 360° rotation of the SS and SR enantiomers of noscapine and the five derivatives, using MOPAC program and RM1 semi-empirical method using COSMO solvent. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.3)
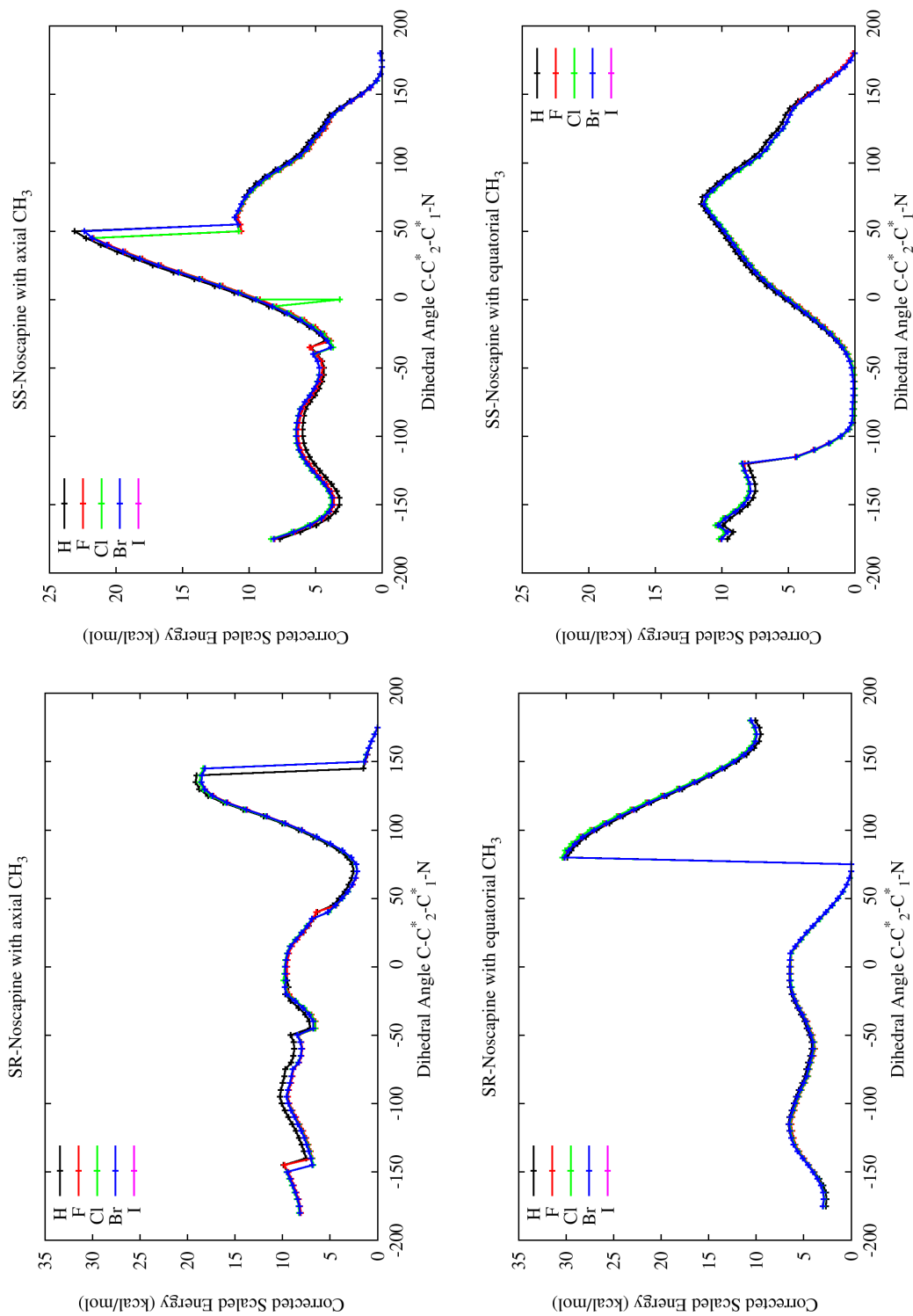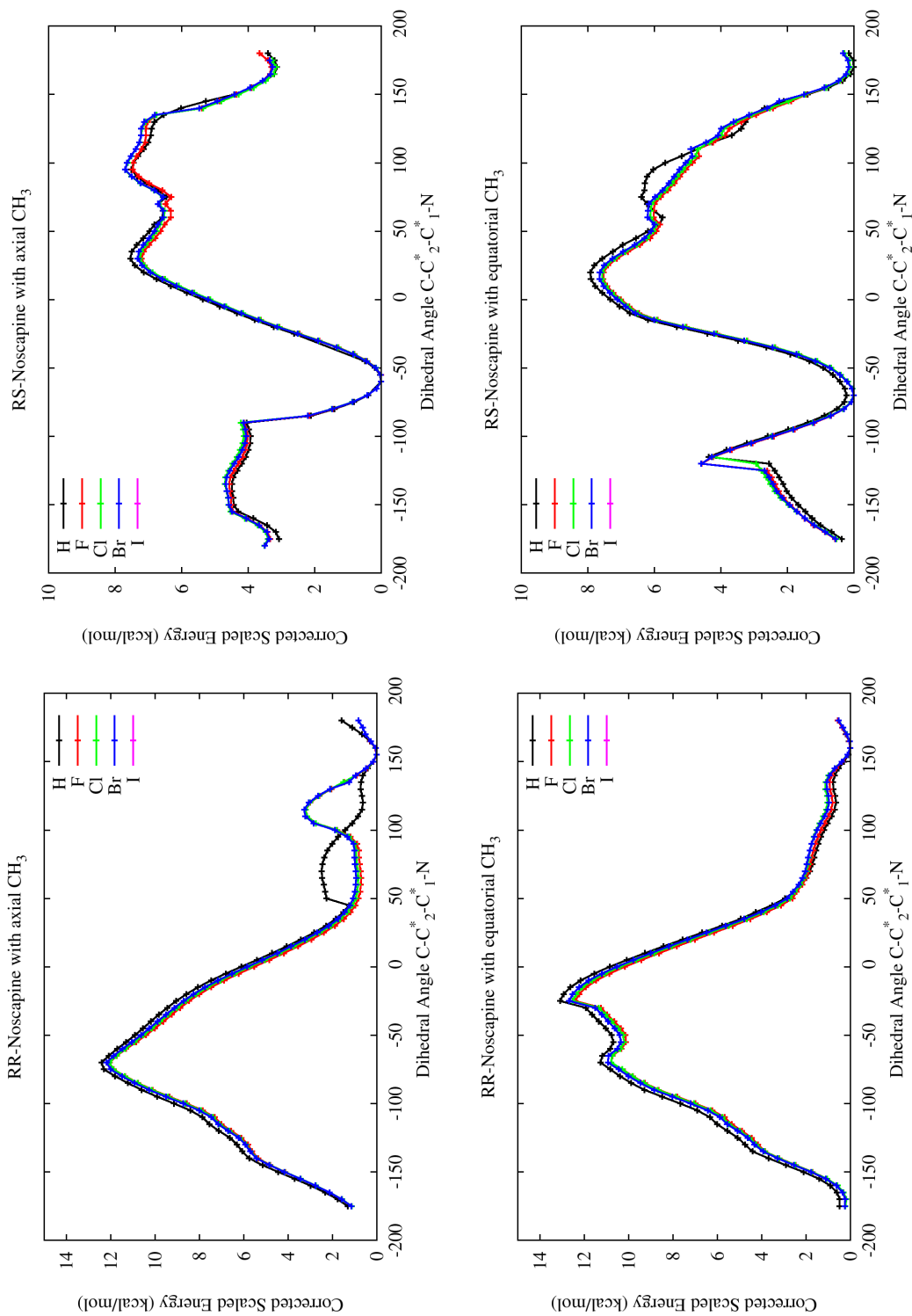
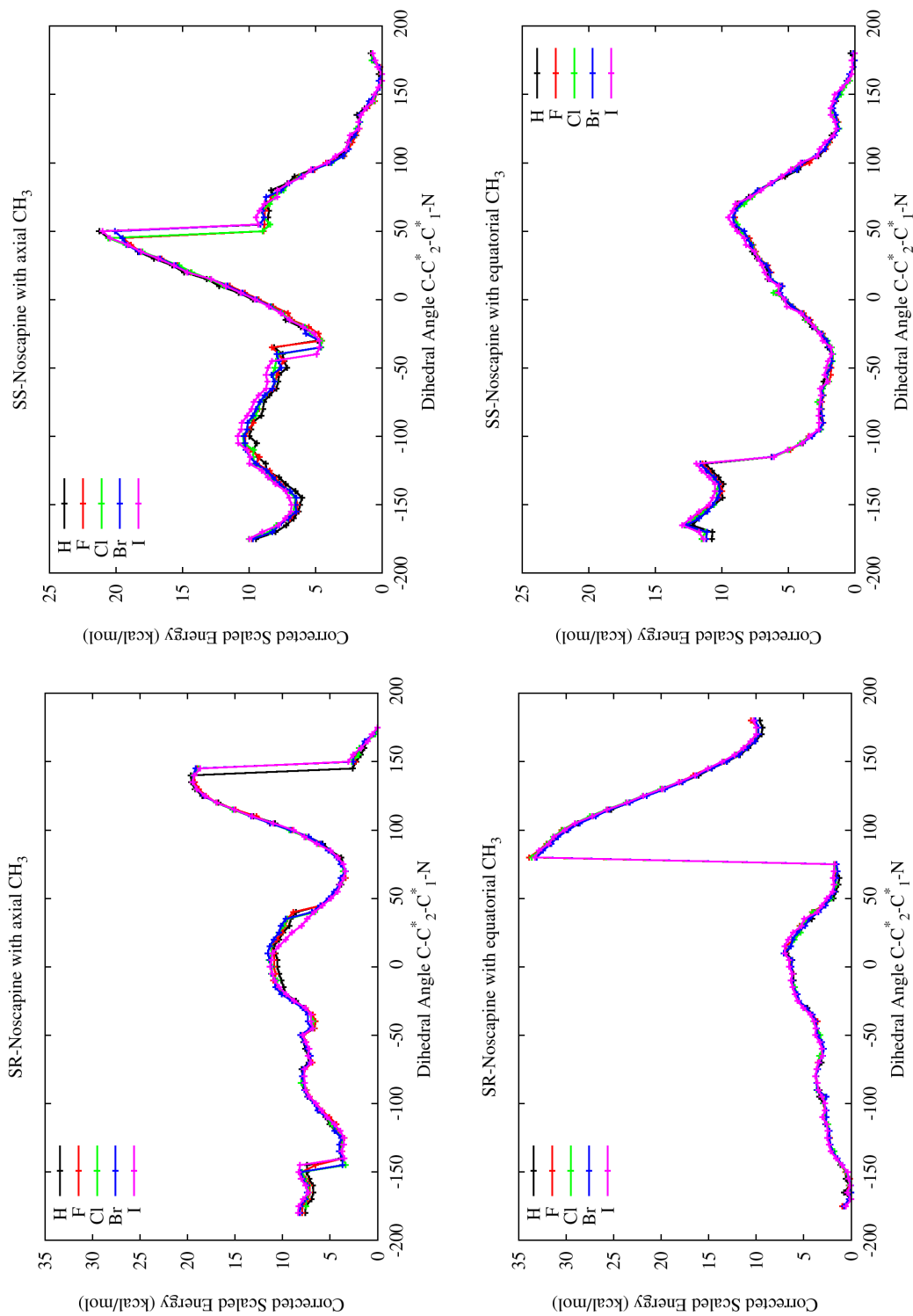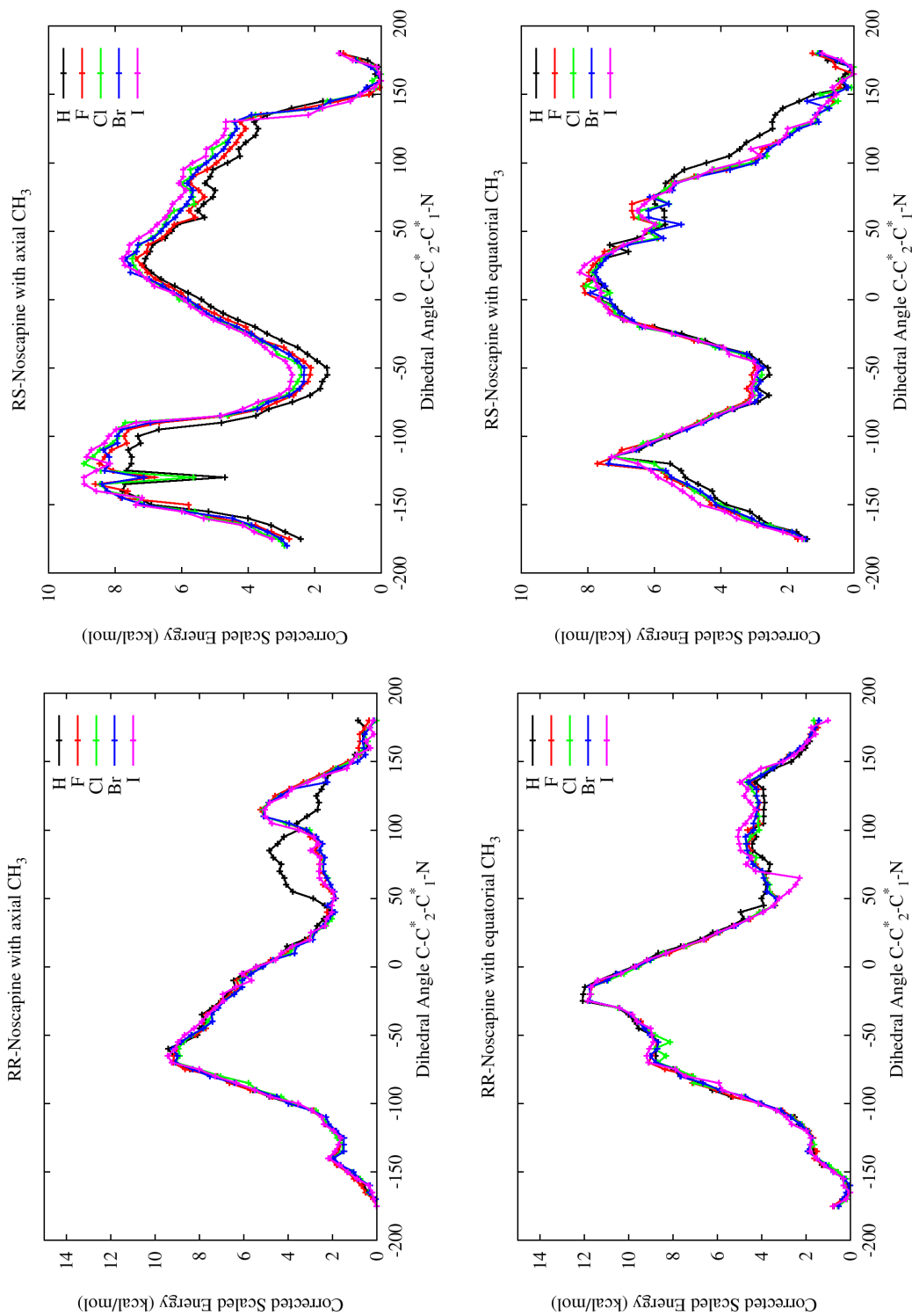Figure 7.8: Energy change during 360° rotation of the RS and RR enantiomers of noscapine and the five derivatives, using MOPAC program and RM1 semi-empirical method using COSMO solvent. Note: all energies are scaled by the lowest energy conformation (values given in Table 7.3)

Chapter 8

Interactions of Laulimalide, Peloruside, and their Derivatives with Ten Isotypes or Mutants and 1TUB from the PDB of β-tubulin

8.1   Introduction

Approximately 76,200 deaths from cancer occurred in Canada in 2010 [153]. An estimated 173,800 new cases of cancer were diagnosed in Canada over the past year, which is an increase by about 4.5% from 2009. Almost 11% of the people in Canada that were newly diagnosed with cancer this year were Albertans (16,000 people). In 2010, 2,900 women and 3,300 men died of cancer in Alberta [153]. To improve the treatment and reduce the number of deaths, there is a critical need for a fundamental understanding of the interaction of cancer drugs with cancer cells.

Most of the cancer drugs used or investigated today have been extracted from natural sources or are modifications of natural products. For example: paclitaxel is found in the bark of the Pacific yew tree, *Taxus brevifolia*, native to Sumatra and Celebes [154];

peloruside is isolated from the marine sponge, *Mycale hentscheli*, growing around New Zealand [155]; and laulimalide is isolated from the marine sponge, *Cacospongia mycofijiensis*, growing in Vanuatu [156]. However, due to the disappearance of our forests and contamination of our oceans, it is possible that a revolutionary antimitotic drug could be destroyed before we have a chance to discover it. We must therefore alleviate some of Mother Nature's burdens by focusing on understanding how cancer drugs interact with cancer cells. With this knowledge, designing a specific cancer drug to attack a specific area of a cancer cell could be accomplished.

Due to the involvement of microtubules in cell division, they are an important focus in anti-cancer research because cancerous cells proliferate by unregulated cell division. Some drugs currently used in chemotherapy work by interfering with the process of cellular division through binding to the α/β-tubulin heterodimer [157]. α/β-tubulin heterodimers provide the building blocks for microtubules. There are eight common isotypes of tubulin that are present in varying degrees in the body, all of which have known mutations. Recent research [158] indicates that different isotypes of tubulin are prevalent in different types of cancer cells. Therefore, augmenting the clinical cancer drugs with molecular-engineered derivatives of peloruside or laulimalide families could increase the efficiency of treatment by a synergistic effect.

Laulimalide and isolaulimalide, shown in Figure 8.1, (isolaulimalide forms from laulimalide under acidic conditions) were investigated for their antimitotic behaviour by Mooberry et al. in 1999 [159] who found that laulimalide had an IC_{50}* only one order of magnitude less potent than paclitaxel, but isolaulimalide is a thousand times less potent. Mooberry and co-workers [160–162] along with Gallagher et al. [163] synthesized several derivatives of laulimalide (ILD, ILG, L01, L02, L03, L04, L05, L13, L14, L15,

---

*half maximal inhibitory concentration; IC_{50} measures the effectiveness of a compound at inhibiting biological or biochemical function; in this case the IC_{50} is the concentration at which 50% of the cells die

L16, L17, L18, LAD, and LAG) and experimentally determined the IC_{50} values (given in Table 8.1) with respect to several cancerous cell lines. Cell lines 1A9, PTX10, PTX22, A8, B10, SK-OV-3, and SKVLB are ovarian cancer cell lines, where PTX10, PTX22, and A8 are paclitaxel resistant cell lines and SK-OV-3 is resistant to diphtheria toxin, cis-platinum and adriamycin while SKVLB is a vinblastine resistant cell line. Cell lines MDA-MD-435, and NCI/SDR are breast cancer cell lines and HCT-116 is a colon cancer cell line. All of the derivatives created did not show improvement over the parent compound laulimalide for the cancer cell lines that they tested. The sixteen laulimalide derivative were used in the present study to determine if the same trend that is found experimentally can be computationally validated. A total of eighteen laulimalide compounds were studied (sixteen derivatives mentioned above along with laulimalide (LAU) and isolaulimalide (ILA)). Two-dimensional structures of all sixteen derivatives can be found in Appendices E and G along with the cartesian coordinates in Appendices D and F, (see Supplementary Data).
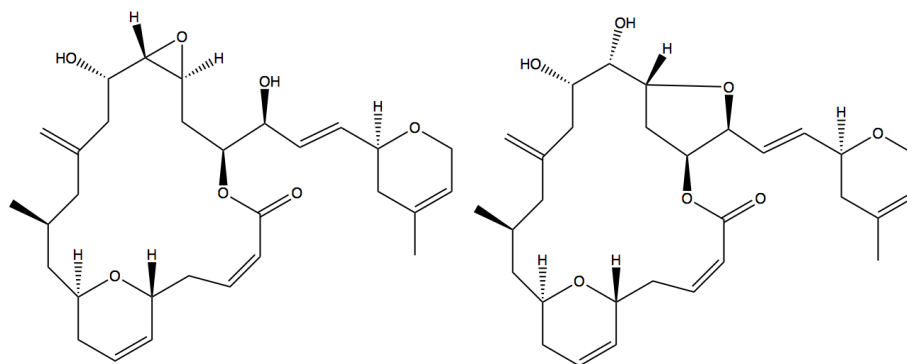

Figure 8.1: Laulimalide (left) and Isolaulimalide (right)

Peloruside A and B (depicted in Figure 8.2) is a new antimitotic agent which has been extensively studied since 2000 [165–169]. De novo drug molecules were designed and a

Table 8.1: Experimental IC$_{50}$ (μm) of laulimalide, peloruside and laulimalide derivatives with several cancer cell lines (where ND indicated that the IC$_{50}$ was not determined)

DER	1A9	PTX10	PTX22	A8	B10	MDA-MB-435	HCT-116	SK-OV-3	SKVLB	NCI/SDR
LAU[a,b,c]	0.0039	0.0060	0.0063	0.0092	0.0150	0.00574	—	0.01153	1.21	ND
PEL-A[d]	0.0150	0.0510	0.0210	0.0170	0.0290	—	—	—	—	—
ILA[c]	—	—	—	—	—	1.970	—	2.57	2.65	—
ILD[c]	—	—	—	—	—	21.4	1.9	—	—	—
ILG[c]	—	—	—	—	—	>50	>50	—	—	—
L01[a]	0.189	0.353	1.22	0.625	—	—	—	—	—	—
L02[a]	0.450	0.862	1.76	0.531	—	—	—	—	—	—
L03[a]	ND	ND	ND	ND	—	—	—	—	—	—
L04[a]	10.9	10.3	9.68	11.6	—	—	—	—	—	—
L05[a]	ND	ND	ND	ND	—	—	—	—	—	—
L13[b]	ND	ND	—	—	—	7.9	—	—	—	22.9
L14[b]	1.0	1.8	—	—	—	1.2	—	—	—	1.2
L15[b]	1.42	3.49	—	—	—	1.35	—	—	—	2.38
L16[b]	7.91	16.4	—	—	—	7.30	—	—	—	12.4
L17[b]	ND	ND	—	—	—	>50	—	—	—	ND
L18[b]	0.424	1.11	—	—	—	0.233	—	—	—	0.433
LAD[c]	—	—	—	—	—	1.3	0.7	—	—	—
LAG[c]	—	—	—	—	—	9.2	14.1	—	—	—

[a] Reference [160]; [b] Reference [161]; [c] Reference [162]; [d] Reference [164]

collection of 25 molecules were studied, (PEL-A, PEL-B and 23 de novo drug designs: P13, P14, P15, P16, P17, P18, P1A, P1B, P1C, P1D, PAD, PAG, PLA, PLX, POX, X13, X14, X15, X16, X17, X18, XAD, and XAG). The two-dimensional structures of the twenty three derivatives are depicted in Appendices E and G, (see Supplementary Data). The IC$_{50}$ for Peloruside A with respect to several cancer cell lines which were previously described are summarized in Table 8.1. Peloruside is a less potent antimitotic agent than laulimalide, but nevertheless it shows promising results against drug resistant cell lines, and unlike laulimalide it is stable in acidic conditions.
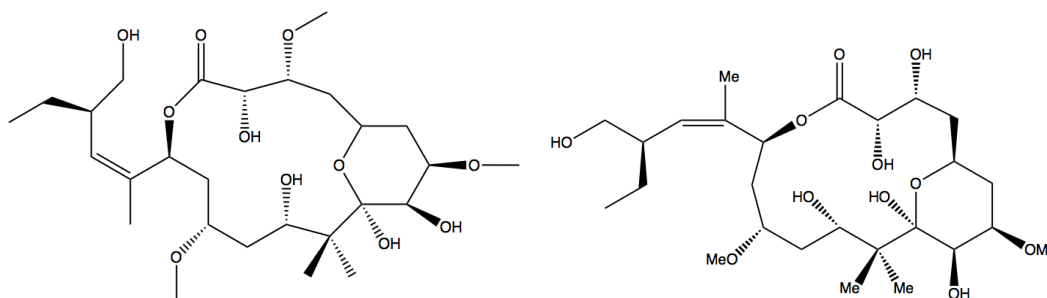

Figure 8.2: Peloruside-A (left) and Peloruside-B (right)

The ADMET program [170] was run to determine the toxicity of the peloruside and laulimalide derivatives, and the computed toxicity risks for all forty-three derivatives is given in Table 8.2. The toxicity risk is calculated based on several properties such as, acute toxicity in rats, carcinogenicity in rodents, and Ames test mutagenicity. A toxicity risk value of less then three corresponds with 90% of the world drug index and several of the derivatives fall within this threshold. ADMET also calculates the Lipinski rule of 5. The Lipinski rules for a compound are that molecular weight must be under 500 Daltons, less than five hydrogen bond donors, less than ten hydrogen bond acceptors and the logarithm of the partition coefficient P must be less than five. Excluding molecular weight (derivatives fall between 460 and 620 Daltons), only four peloruside derivatives fail the remaining Lipinski rules: PELB, PAG, P1C and XAG.

Table 8.2: ADMET toxicity risk

Toxicity Risk	Derivative
1	PELA, PELB, P16, P1A, P1B, P1C, P1D
2	P13, P14, PLA, POX, X13, X14, X15, X16, XAG
	ILD, ILG, L02, L04, L05, L13, L15, L16,LAD, LAG, THL
3	P15, P17, P18, PAD, PAG, PLX, X18, XAD
	LAU, ILA, L01, L03, L14, L17, L18
4	X17

Tubulin isotypes (given in Table 8.3) are found in varying degree throughout healthy cells in the human body; the most predominant source of these isotypes are listed in Table 8.3. The β isotypes consist of 421 amino acid residues and approximately 1% of the residues change between the isotypes [171]. Only one or two residues change within the peloruside/laulimalide binding site, and therefore a single amino acid mutation within the binding site can cause a drug to be ineffective against that tubulin mutation/isotype. It is known [158] that although there is a main form of β-tubulin within a certain organ, other isotypes are also found in lower amounts within the same organ. The percentage of each isotype also changes when tissue turns cancerous; it is possible that this property can be exploited to more effectively eliminate cancer cells. Cancer cells can become resistant to this antimitotic agent, which could be due to the emergence of different isotypes of β-tubulin which are naturally resistant to this drug. If the isotypes distribution of the cancerous tissue is known, along with the effectiveness of many drugs against those isotypes, determining the optimum drug for each patient could be easily accomplished. It is the goal of this work to provide computational predictions to the problem of drug resistance by determining promising new cancer drugs which can kill current drug resistant cancer cells.

Table 8.3: Tissue distribution of β-tubulin isotypes in normal cells

Isoform	Organ Expression	Cellular Expression
βI	Constitutive	Most cells
βIIa/b	Brain, nerve, muscle;	Restricted to particular
	rare elsewhere	cell types
βIII	Brain	Neurons only
	Testes	Sertoli cells
	Colon (very slight amounts)	Epithelial cells only
βIVa	Brain only	Neurons and glia
βIVb	Constitutive	High in ciliated cells,
	(not as widespread as βI)	lower in others
βV	Unknown	Unknown
βVI	Blood, bone marrow, spleen	Erythroid cells, platelets
βVII	Brain	Unknown
βVIII	Unknown	Unknown

A matrix of binding affinities of different isotypes of tubulin (10 isotypes or mutants and the 1TUB PDB structure) to derivatives of peloruside (25 derivatives) and laulimalide (18 derivatives) was created. The binding affinity calculations were accomplished with the AMBER [4] program utilizing molecular mechanical and molecular dynamical calculations. Several peloruside de novo drug designs have shown improvement in tubulin binding affinities.

8.2 Method of Calculations

Docking and mass shift perturbation mapping [172] studies were performed on both laulimalide [173] and peloruside [174], and it was determined that both peloruside and laulimalide share the same binding pocket within β-tubulin. The binding sites/modes for laulimalide and peloruside are depicted in Figure 8.3. The images in Figure 8.3 were

generated using VMD and the 1TUB structure obtained from the Protein Data Bank (PDB) with laulimalide and peloruside docked into their binding sites. The docking calculation were performed by Barakat [173] and Huzil [174] in the Tuszynski lab for laulimalide and peloruside, respectively, and the docking predictions were experimentally validated by Schriemer's lab by mass shift perturbation mapping. The details for this work can be found in the references previously cited. It is these studies that provide the foundation for the present project. The assumption that the docking calculations and the experimental results provide an adequate representation of the binding mode of laulimalide and peloruside was necessary. In addition to the docked structures of laulimalide and peloruside, the ten isotypes of β-tubulin (βI, βIIa, βIIb, βIII, βIVa, βIVb, βV, βVI, βVII, and βVIII) were obtained form work done by Huzil et al. [171] and it was also assumed that these adequately represented the isotypes found within the human body.

The α-tubulin used for all of the α/β-tubulin dimers was from the 1TUB complex. The positions of the guanosine triphosphate (GTP) and guanosine diphosphate (GDP) on the α-tubulin and β-tubulin, respectively, were also conserved from the 1TUB isotype. The alignment of all of the α-tubulin with the β-tubulin isotypes as well as the positioning of the GTP and GDP molecules within the tubulin fragments was done using the VMD [7] program. Molecular mechanics as implemented in the AMBER program was used to perform energy minimization. The systems were solvated with 30,000 TIP3P water molecules. The input parameters to the calculations specified that a maximum of 6000 cycles of steepest descent minimization was used, followed by heating the system to 310K using Langevin temperature scaling was used to keep the system at 310K. An NVT ensemble for the molecular dynamics simulation was used and the simulations was run for 4,250,000 steps with a time step of 2 femtoseconds. Average structures of the α/β-tubulin dimers were obtained from the last 3,000,000 steps of the simulation. The
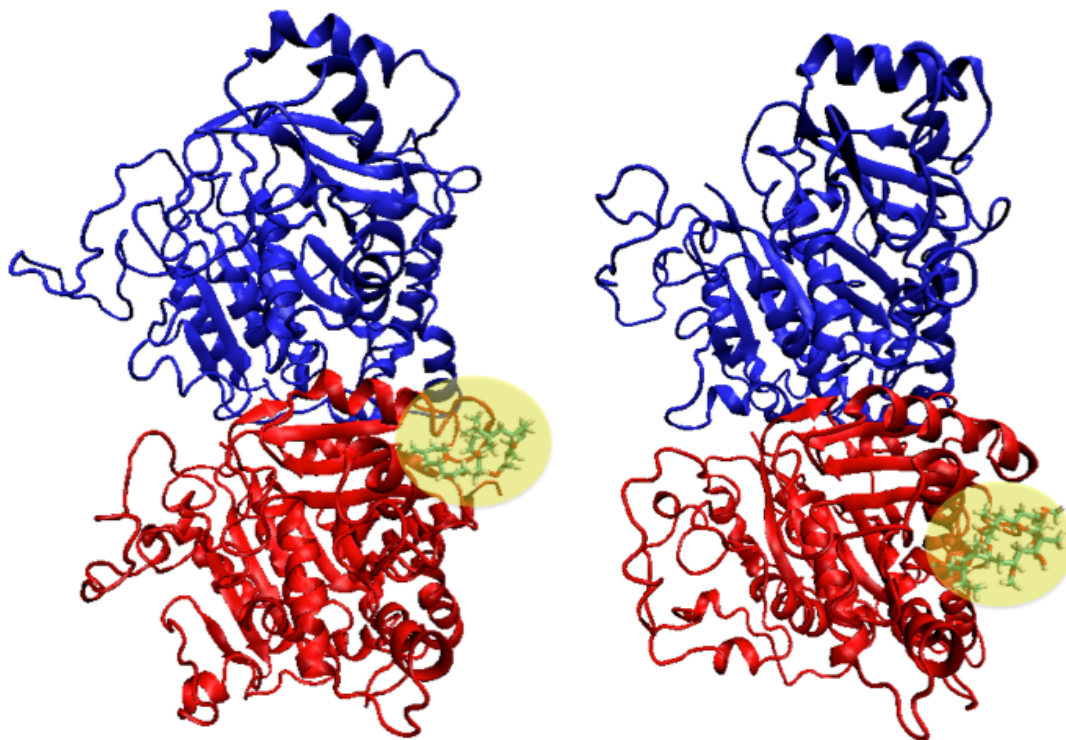
Figure 8.3: Laulimalide (left) and peloruside (right) bound to α-β tubulin dimer

FF99SB force field parameters [175] were used for the α/β-tubulin dimers and the FF03 force field parameters [176] were used for GDP and GTP. The calculations were run on WestGrid and computer clusters at the University of Alberta.

The three-dimensional structures of the laulimalide and peloruside derivatives were created using the MOLDEN [5] program and an energy/geometry minimization was followed by the evaluation of the hessian in order to ascertain that the stationary points found corresponded to minima. The calculations were performed using the semiempirical molecular orbital method Austin Model 1 (AM1) [12] method within the GAMESS [1] program. These calculations were carried out on computer clusters at the University of Alberta. Visualization of molecular structures, vibrational modes, and molecular orbitals were performed with MOLDEN [5] and MacMolPlt [6].

The alignment of the isotypes with drug molecule was performed using the VMD program and a maximum overlap between the different isotypes was obtained. The molecular mechanics and molecular dynamics calculations were run according to the procedure described previously, with the exception that laulimalide, peloruside and the derivatives were run using the generalized AMBER force field (GAFF) parameters [177] because there is no specific set of force field parameters for the new drugs. The calculations were primarily run on the PharmaMatrix cluster which is generously supported by the Allard Foundation, as well as WestGrid clusters, and computer clusters at the University of Alberta.

Using molecular mechanics - generalized Born surface area (MM-GBSA) calculations, the changes in free energies (ΔG) or binding energies were extracted for all of the drug molecules with respect to each of the tubulin isotypes. The results of the calculations are presented in Figures 8.4 and 8.6 and full numerical data can be found in Appendices E and G, (see Supplementary Data). Figures 8.4 and 8.6 provide a simple visual representation of the binding matrix. If the predominant isotype of tubulin is known, determining the optimum drug is simply a matter of finding the isotype on the x-axis and searching along the y-axis (drug derivative) for the lightest square (highest binding affinity).

8.3 Results and Discussion

In Figures 8.4 and 8.6, the tubulin isotype is plotted against the drugs and the colour on the plots ranges from black (no bonding of drug to tubulin isotype) to white (strong bonding of drug to tubulin isotype). Colon, breast, and ovarian cancer have high levels of βI tubulin, furthermore, βIII tubulin is often thought to be the cause of paclitaxel and drug resistant cancer cells [178]. It is apparent from Figure 8.4 that none of the

laulimalide derivatives bond well with any of these tubulin isotypes. It can be conjectured that if a cancer cell line with a higher concentration of βIIb or βIVa isotypes is investigated, the laulimalide derivatives L05 and L04 respectively could be used and these derivatives show greater binding affinity. The lack of bonding of the laulimalide derivatives is corroborated by the experimental data presented in Table 8.1. Table 8.4 demonstrates that there is a distinct change in the ratios of isotypes when a healthy cell turns cancerous. The change in tubulin isotype distribution could possibly be exploited and therefore a specific drug could be chosen to maximize cancer cell death.

Table 8.4: Change in distribution of isotypes of tubulin in several cancer types
($\uparrow$ / $\downarrow$ indicated increase/decrease in tubulin isotype in cancer cells with resect to healthy cell and $\approx$ indicates that there is no change in isotype distribution)

Cancer	βI	βIIa	βIII	βIVb	βV
Colon	$\uparrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Ovarian	$\approx$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$
Kidney	$\uparrow$	$\downarrow$	$\approx$	$\downarrow$	$\uparrow$
Prostate	$\downarrow$	$\uparrow$	$\approx$	$\uparrow$	$\downarrow$
Breast	$\uparrow$	$\downarrow$	$\uparrow$	$\approx$	$\downarrow$
Lung	$\uparrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$

According to Leandro-García et al. [158], colon and prostate cancer cells have βI as the most abundant source of the β-tubulin isotype and βIVb as the second most abundant isotype. Additionally, ovarian, kidney, breast, lung and larynx cancer all have the opposite distribution (βIVb as the most abundant and βI as the second most abundant isotype). Several of the peloruside derivatives show an increased binding affinity to these two tubulin isotypes. P15, P16, P1D, PLX, and X15 derivatives show increased binding over the parent compounds peloruside A/B for the βI isotype and P14, P17, P1B, PLX, X14, and X15 show increased binding over the parent compounds peloruside A/B for the βIVb isotype. Also, increased amount of the βIII isotype in some tissues can

be associated with cancerous cells and thus determining which drug will bind to this isotype is extremely important. An increase in βIII isotype can be observed for lung cancer which is the second most prevalent form of cancer (13.9% of cancer cases) [153] and also accounts for over one-quarter (27%) of all cancer deaths. Finding a drug that binds strongly and specifically to this isotype is of great importance. According to Figure 8.6, the derivative PAD has a strong binding affinity towards βIII and only mediocre or non-bonding to the other tubulin isotypes with the exception of βVII isotype. Peloruside derivative P15, P17, PAD, PAG, and X14 also show improved binding affinity to the βIII isotype. Such data could be extremely useful in targeting a specific cancer with a known tubulin isotype distribution by choosing the drug with the greatest affinity for it. For instance, prostate cancer in particular shows an increase in the βIIa isotype and peloruside derivatives PAG, PLA, PLX, X18 and XAG have the highest binding affinities to this tubulin isotype. Cancer of the larynx, which interestingly has seen a decrease in the number of cases over the past year [153], has an increased amount of βV isotype the and derivatives P13, P16, and P1D bind with greater affinity to that isotype. Once the distribution of β tubulin is known for a specific cancer, an optimum drug that corresponds to that cancer could be easily determined: For example, lung cancer shows an increase in β isotypes I, IIa, III and a decrease in isotypes IVb and V. In Figure 8.5 the derivative that has the most stars (green for increased affinity and red for decreased affinity), is the derivative L14. Choosing laulimalide derivative L14 to treat lung cancer would target the isotypes which have an increase in the cancerous cells, while simultaneously having a lower affinity for the isotypes that have decreased in the cancerous cells. Similarly, the best peloruside derivative to treat breast cancer, which has an increase in β isotypes I and III and a decrease in isotypes IIa and V, would be derivative P15 as shown in Figure 8.7. Once further information about tubulin isotype distribution is known for the wide varieties of cancers with respect to healthy cells, determining the best antimitotic drug could be carried out theoretically.

8.4 Conclusion

Results of the molecular mechanics and molecular dynamics calculations show that within the matrix of binding affinities of different isotypes of tubulin (11 isotypes or mutants) to derivatives of peloruside (25 derivatives) and laulimalide (18 derivatives), there are several peloruside derivatives that could be promising new drugs for many types of cancer. It should also be noted that some laulimalide derivates may be used at attack cancer cells with an increased content of a specific tubulin isotype, regardless of the type of cancer. For example, L04 would theoretically bind very strongly to cancer cells with increased βIVa isotype. Further derivatization of peloruside and laulimalide could be performed in order to obtain a complete set of drugs which bind optimally to one or more of the tubulin isotypes. This research shows that the exploitation of the differentiation of tubulin isotypes in different cell tissues, as well as between healthy and cancerous cells, is possible and exploiting these properties could allow for maximum cancer cell death.
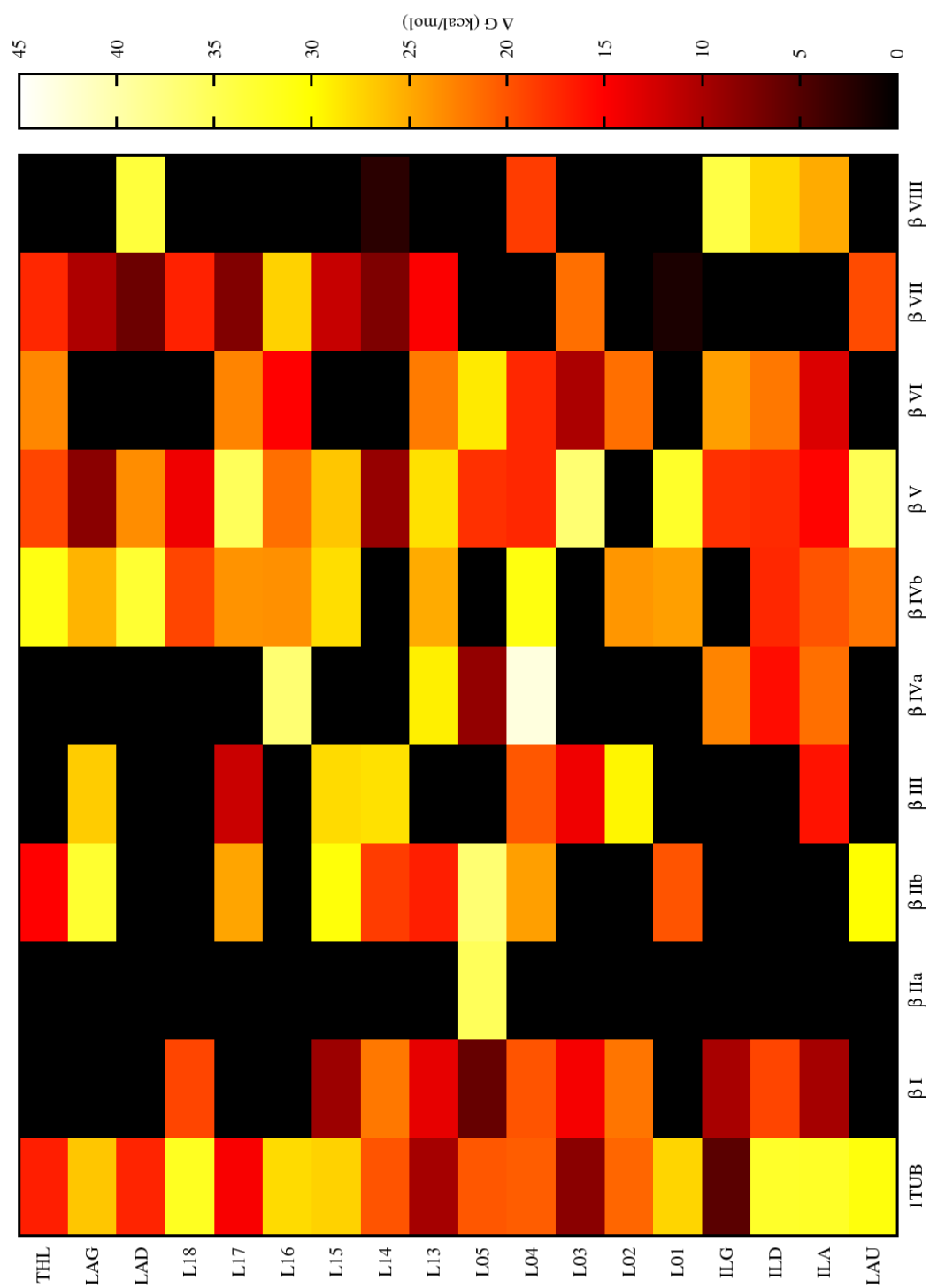
Figure 8.4: ΔG values in kcal/mol of laulimalide and its derivatives

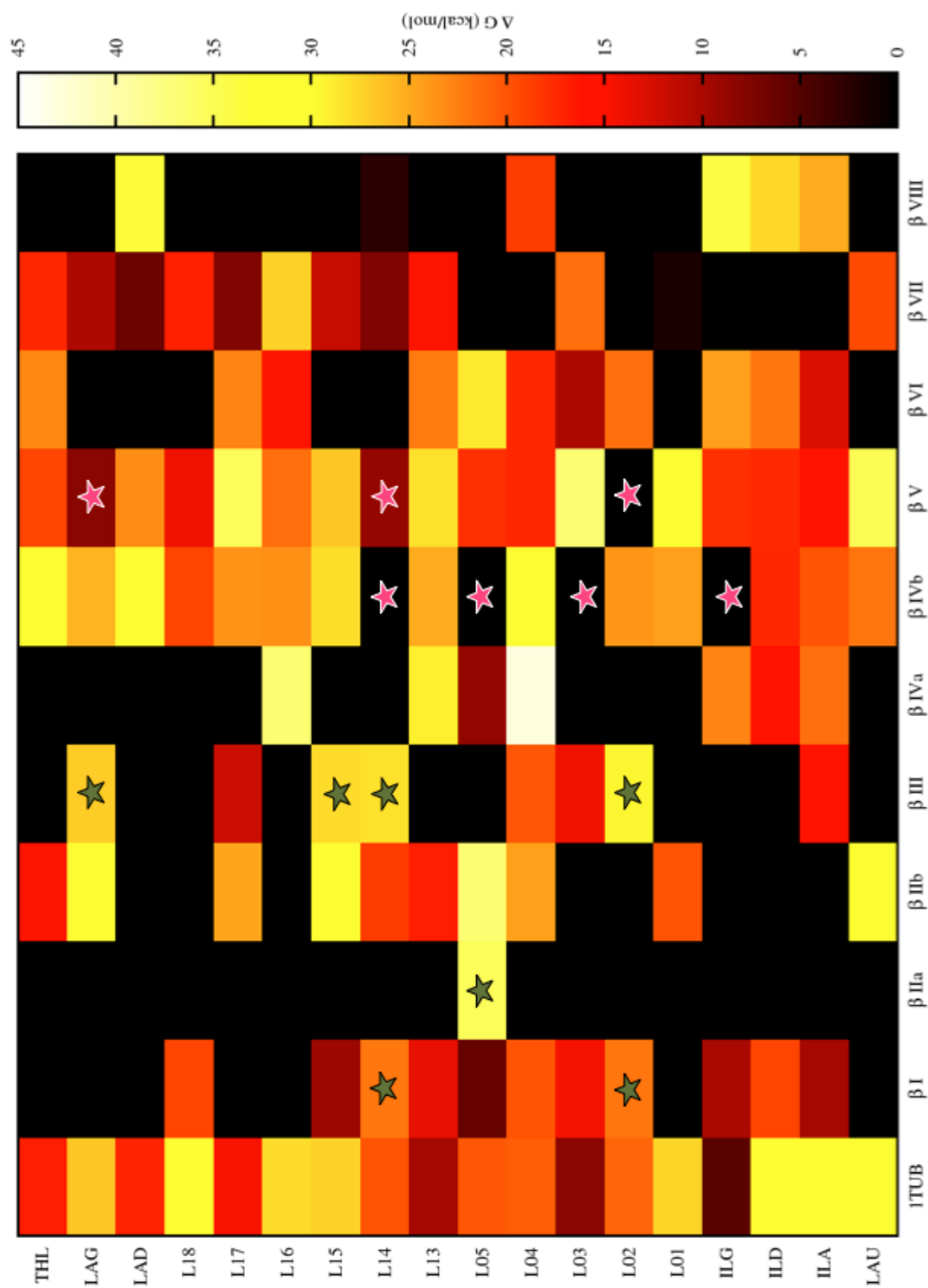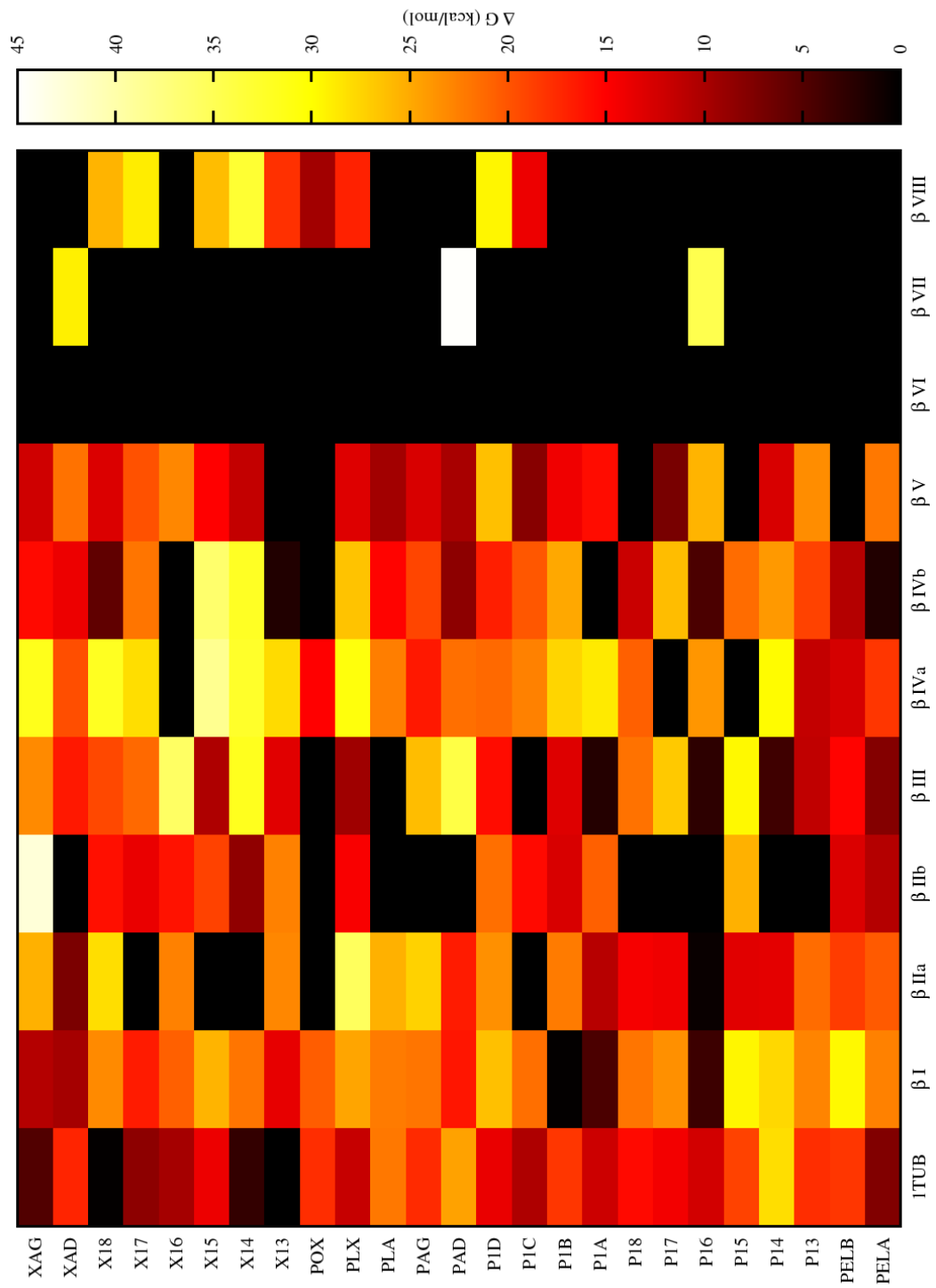Figure 8.5: Utilizing chart to determine optimized laulimalide derivative for lung cancer

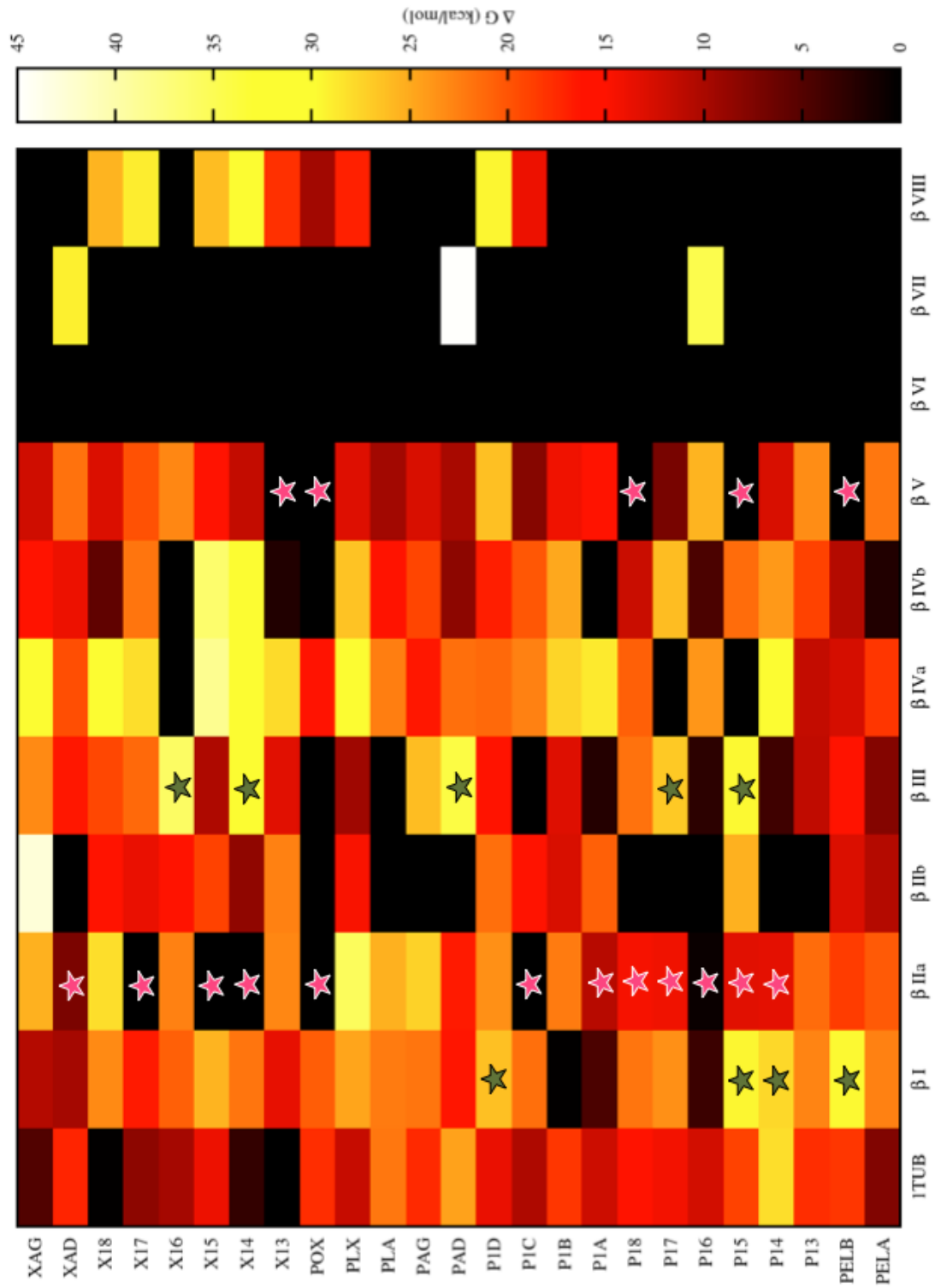Figure 8.6: ΔG values in kcal/mol of peloruside and its derivatives

Figure 8.7: Utilizing chart to determine optimized peloruside derivative for breast cancer

Chapter 9

Future Work and Conclusions

9.1 Cation Capturing Molecules

Using the $\Delta\Delta$E equation shown in Chapter 4 it is possible to mimic the solution trends in experimental data by performing only gas phase calculations. Similar systems could be investigated in the same manner, giving insight into cation-capturing molecules. Additionally, it is evident from this study using either AE or MCP at the DFT level provides no significant differences in the results, so it is possible to use the method which is most computationally inexpensive and/or most convenient.

The ring structures 18c6 and 6aa discussed in Chapter 5 are extremely useful molecules to sequester metal ions, and future computational studies could investigate derivatives of these molecules to optimize the selectivity of cations to these molecules. Furthermore, computing the IR spectra of these molecules could give insight into the complexation reaction that is taking place when a ring structure sequesters a cation.

9.2 Additional Laulimalide Derivatives

Ten new laulimalide derivatives were created based on the outcome of the calculations for the previously described laulimalide derivatives. These ten new derivatives are de-

picted in Figure 9.1. The same procedure was used for these derivatives as for all other drugs that were investigated and this procedure is described in Chapter 8. The data for the calculations is summarized in Table 9.1. Three of the ten derivatives (L21, L24, and L25) show a greater or similar binding affinity towards the 1TUB tubulin as the parent compound laulimalide. It would be a logical next step to perform calculations on these three laulimalide derivative with respect to the ten tubulin isotypes.

Table 9.1: Calculated ΔG values (kcal/mol) of laulimalide derivatives in 1TUB

Isotype	ΔG		error
LAU	-30.67	$\pm$	4.32
L20	-7.19	$\pm$	3.65
L21	-45.13	$\pm$	3.44
L22	-21.07	$\pm$	3.79
L23	>0	$\pm$	—
L24	-30.26	$\pm$	2.87
L25	-33.08	$\pm$	4.74
L26	>0	$\pm$	—
L27	-12.70	$\pm$	4.58
L28	-17.14	$\pm$	2.49
L29	-17.37	$\pm$	3.39

9.3 Fragmented Laulimalide Derivatives

In addition, another approach to de novo drug design will be investigated; it is based on fragmenting the parent drugs (peloruside and laulimalide) into smaller molecules. It is the goal of the fragmentation approach to retain the chemical activity of the parent drug while simultaneously providing a simpler/cheaper chemical synthesis. The ability to chemically synthesize the fragment molecules would permit experimental validation

of the calculations. Some examples of fragmented laulimalide molecules are depicted in Figure 9.2, and preliminary calculations using β-tubulin from 1JFF were performed. The binding mode of the fragments was assumed to be the same as the parent compound laulimalide. The NVT ensemble was run, at 310 K for 1.5 ps of equilibration, followed by 3.0 ps of production, with a time step of 2 fs. MM-GBSA calculation determined the change in free energy (ΔG) or binding affinities for the drug fragment with respect to the β-tubulin. Preliminary results show that some of the fragments bind with similar affinity as laulimalide. Therefore, further calculations are currently being carried out. If the results show promising binding affinities, the fragments will be synthesized by Professor Hall at the University of Alberta, and mass shift perturbation mapping will be performed by Professor Schriemer at the University of Calgary.

It is apparent from the studies described within this Thesis that a merging of medicine and computational studies is imminent. Computational chemistry can have a synergistic relationship by weeding out unpromising compounds, while simultaneously promoting further experimental investigation on compounds that show promising binding affinities to the protein of interest. Although computational validation and prediction is an old concept, it is now within our computer hardware capacity to combine the worlds of experimental and theoretical chemistry. Predicting the binding site of ligands to proteins can be accomplished by using docking protocols, allowing for simpler and easier experimental measures to be implemented to validate the binding mode without waiting until the X-ray crystal structure can be resolved. Once an experimentally validated binding site on a protein is found, derivatization of the ligand could be accomplished using the methods described within this Thesis, and by synthesizing only the promising complexes a drastic reduction in synthesis costs and time could be achieved. In addition, by plotting the calculated data (in the fashion of Figures 8.4 and 8.6), a better understanding is obtained of how mutants of different proteins can drastically change the binding affinities

of the ligands. To my knowledge this is the first time that protein mutants versus ligand interactions have been plotted in such a manner.

9.4 Conclusions

To my knowledge this work presents the first studies to determine the optimum method and basis set for studying a cation-sequestering system of interest. It is apparent from the studies discussed in Chapters 4 and 5 that utilizing the MCP method (where the core electrons are replaced by a potential) does not affect the stabilization energies. Furthermore, adding the first solvation shell can reduce error introduced by using an inadequate basis set and provide results that are comparable to solvated experimental data.

Utilizing the graphs given in 8.5 and 8.7, in the manner described in Chapter 8, is the first attempt at optimizing chemotherapy for an individual patient. In this manner personalized medicine can be accomplished by first determining the tubulin isotype distribution of the patient and then administering the drug that best targets the cancer cells. Additionally, to combat cancer drug resistance, a second tubulin isotype distribution measurement can be performed after the first round of chemotherapy and a lower dose of the new optimized drug can be given in order to target the new isotype distribution; this process can be continued until it is certain that all cancerous cells have been eradicated.

In conclusion, following the procedures outlined within this Thesis can be extended to any similar system, whether the interests lie with rings and metal, in cancer drug conformations, or in protein/drug interactions.
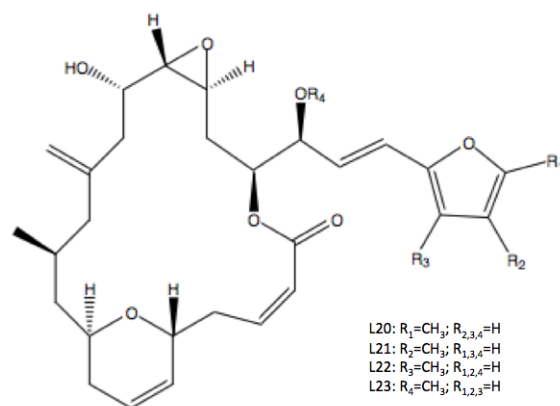
(a)
L20, L21, L22, and L23
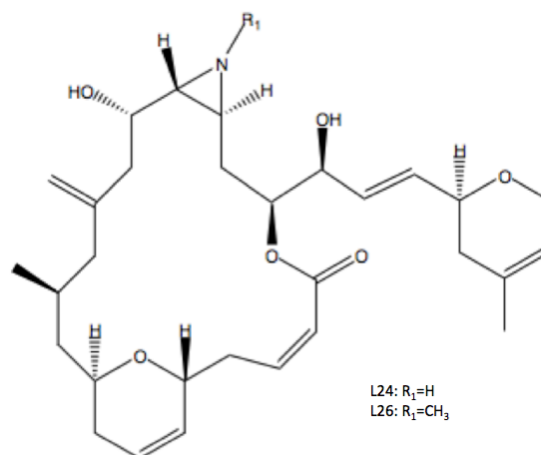
(b)
L24 and L25
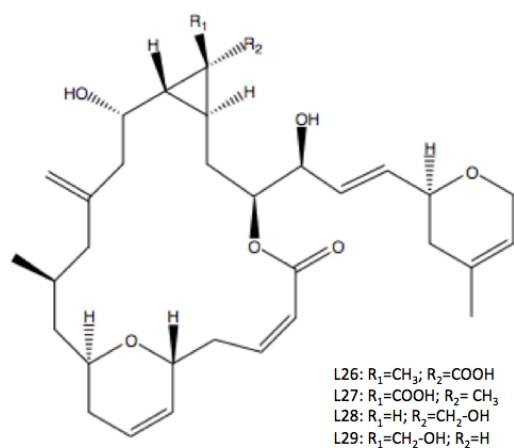
(c)
L26, L27, L28, and L29

Figure 9.1: Laulimalide derivatives: L20, L21, L22, L23 L24, L25, L26, L27, L28, and L29

(a)
A2.0

(b)
A2.1.4

(c)
A3.1

(d)
A3.2

(e)
A3.4

(f)
A4.1

(g)
A4.2

(h)
A4.3

Figure 9.2: Laulimalide fragments: A2.0, A2.1.4, A3.1, A3.2, A3.4, A4.1, A4.2, A4.3

Bibliography

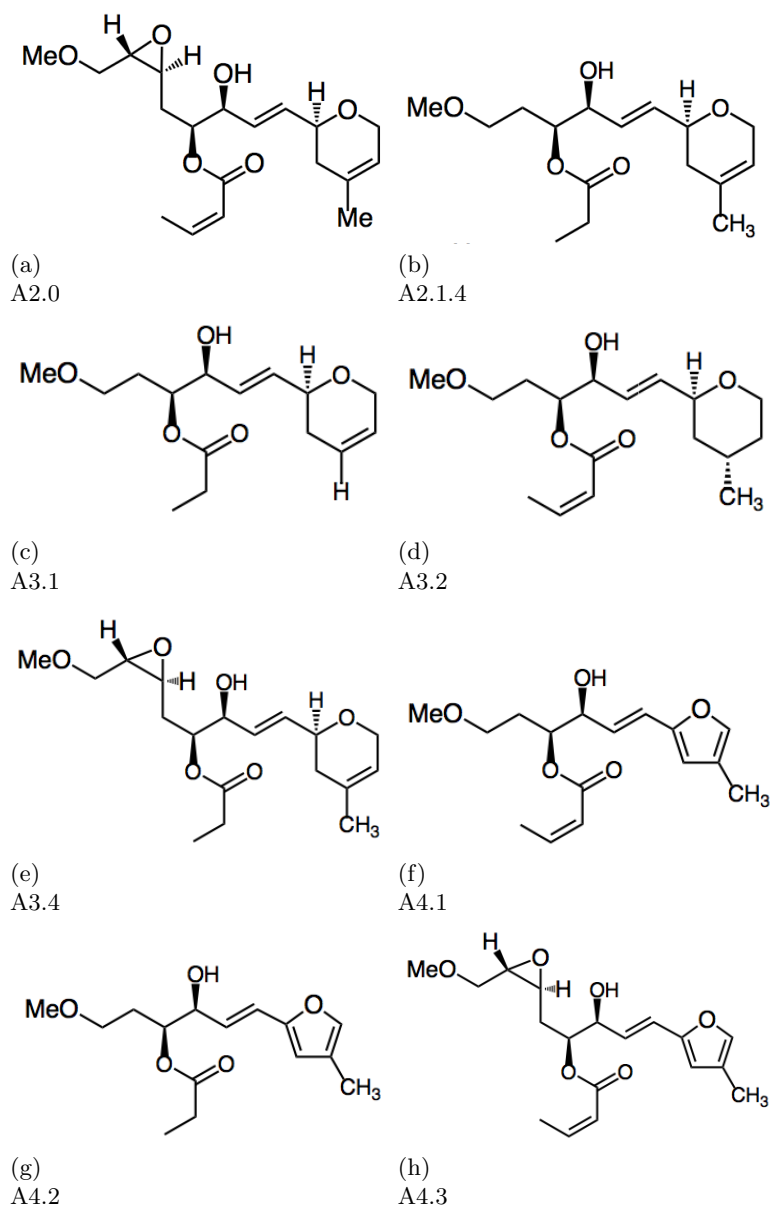[1] Schmidt, M.; Baldridge, K.; Boatz, J.; Elbert, S.; Gordon, M.; Jensen, J.; Koseki, S.; Matsunaga, N.; Nguyen, K.; Su, S.; Windus, T.; Dupuis, M. and Montgomery, J., *J. Comp. Chem.*, 1993, **14**, 1347–1363.

[2] Frisch, M. J. and et al., , Gaussian 03, revision c.02 Wallingford CT, 2004.

[3] Stewart, J. J., *J. Comp. Aid. Mol. Des.*, 1990, **4**, 1–103.

[4] Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S. and Kollman, P. A., Amber 9, University of California, San Francisco.

[5] Schaftenaar, G. and Noordik, J., *J. Comp. Aid. Mol. Des.*, 2000, **14**, 123–134.

[6] Bode, B. M. and Gordon, M. S., *J. Mol. Graph. Model.*, 1998, **16**, 133–138.

[7] Theoretical and Computational Biophysics Group, an NIH Resource for Macromolecular Modeling and Bioinformatics, at the Beckman Institute, University of Illinois at Urbana-Champaign.

[8] Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M. and van de Streek, J., *J. Appl. Cryst.*, 2006, **39**, 453–457.

[9] Vögtle, F., *Host guest complex chemistry : macrocycles : synthesis, structures, applications*, Springer, 1985.

[10] Gajewski, M. and Klobukowski, M., *Can. J. Chem*, 2009, **87**(10), 1492–1498.

[11] Gajewski, M.; Tuszynski, J.; Mori, H.; Miyoshi, E. and Klobukowski, M., *Inorg. Chim. Acta*, 2008, **361**(7), 2166–2171.

[12] Dewar, M. J.; Zoebisch, E. G.; Healy, E. F. and Stewart, J. J., *J. Am. Chem. Soc*, 1985, **107**, 3902–3909.

[13] Rocha, G. B.; Freire, R. O.; Simas, A. M. and Stewart, J. J., *J. Comp. Chem.*, 2006, **27**, 1101–1111.

[14] Nogales, E.; Wolf, S. G. and Downing, K. H., *Nature*, 1998, **391**, 199–203.

[15] Szabo, A. and Ostlund, N. S., *Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc., 1996.

[16] McQuarrie, D. A. and Simon, J. D., *Physical Chemistry: A Molecular Approach*, University Science Books, 1997.

[17] Schatz, G. C. and Ratner, M. A., *Quantum Mechanics in Chemistry*, General Publishing Company Ltd., 2002.

[18] Cramer, C. J., *Essentials of Computational Chemistry Theories and Models*, John Wiley & Sons Ltd. (Canada), 2002.

[19] Thiel, W., *Computational Chemistry: The First Forty Years: Chapter 21: Semiempirical quantum-chemical methods in computational chemistry of Theory and Applications*, Elsevier B.V., 2005.

[20] Hellmann, H., *J. Chem. Phys.*, 1935, **3**, 61.

[21] Hellmann, H. and Kassatotschkin, W., *J. Chem. Phys.*, 1936, **4**, 325–326.

[22] Gombás, P., *Z. Phys.*, 1935, **94**, 473–488.

[23] Kahn, L. R.; Baybutt, P. and Truhlar, D. G., *J. Chem. Phys.*, 1976, **65**, 3826–3853.

[24] Bonifacic, V. and Huzinaga, S., *J. Chem. Phys.*, 1974, **60**, 2779–2786.

[25] Klobukowski, M., *J. Comput. Chem.*, 1983, **4**, 350–361.

[26] Sakai, Y. and Huzinaga, S., *J. Chem. Phys.*, 1982, **76**, 2537–2551.

[27] Katsuki, S. and Huzinaga, S., *Chem. Phys. Lett.*, 1988, **152**, 203–206.

[28] Zeng, T.; Fedorov, G. and Klobukowski, M., *J. Chem. Phys.*, 2010, **in press**.

[29] Zeng, T.; Fedorov, G. and Klobukowski, M., *J. Chem. Phys.*, 2010, **133**, 114107–114118.

[30] Zeng, T.; Fedorov, G. and Klobukowski, M., *J. Chem. Phys.*, 2010, **132**, 74102–74117.

[31] Zeng, T.; Fedorov, G. and Klobukowski, M., *J. Chem. Phys.*, 2009, **131**, 124109–124126.

[32] Klobukowski, M., *Chem. Phys. Lett.*, 1993, **214**, 166–174.

[33] Figgen, D.; Rauhut, G.; Dolg, M. and Stoll, H., *Chem. Phys*, 2005, **311**, 227–244.

[34] Zeng, T. and Klobukowski, M., *J. Chem. Phys.*, 2009, **130**, 204107–204119.

[35] Roothaan, C. C. J., *Rev. Mod. Phys.*, 1951, **23**, 69–89.

[36] Pople, J.; Santry, D. and Segal, G., *J. Chem. Phys.*, 1965, **S43**, 129–135.

[37] Dewar, M. and Thiel, W., *J. Am. Chem. Soc*, 1977, **99**, 4899–4907.

[38] Dewar, M. and Thiel, W., *J. Am. Chem. Soc*, 1977, **99**, 4907–4917.

[39] Thiel, W. and Voityuk, A., *J. Phys. Chem.*, 1996, **100**, 616–626.

[40] Stewart, J. J., *J. Comp. Chem.*, 1989, **10**, 209–220.

[41] Stewart, J. J., *J. Comp. Chem.*, 1989, **10**, 221–264.

[42] Stewart, J. J., *J. Comp. Chem.*, 1991, **12**, 320–341.

[43] Stewart, J. J., *J. Mol. Mod.*, 2004, **10**, 155–164.

[44] Stewart, J. J., *J. Mol. Mod.*, 2007, **13**, 1173–1213.

[45] Stewart, J. J., *J. Mol. Mod.*, 2009, **15**, 765–805.

[46] Oleari, L.; Sipio, L. D. and Michelis, G. D., *Mol. Phys*, 1966, **10**, 97–109.

[47] Gill, P. M., *Aust. J. Chem.*, 2001, **54**, 661–662.

[48] Hohenberg, P. and Kohn, W., *Phys. Rev.*, 1964, **B136**, 864–871.

[49] Parr, R. G. and Yang, W., *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Inc.; New York, 1989.

[50] Kohn, W. and Sham, L. J., *Phys. Rev.*, 1965, **A140**, 1133–1138.

[51] Becke, A., *J. Chem. Phys.*, 1993, **98**, 5648–5653.

[52] Gill, P. M. W., *Mol. Phys.*, 1996, **89**, 433–445.

[53] Perdew, J. P., *Phys. Rev. B*, 1986, **33**, 8822–8824.

[54] Lee, C.; Yang, W. and Parr, R., *Phys. Rev. B*, 1988, **37**, 785–789.

[55] Jorgensen, W. L.; Chandrasekhar, J. and Madura, J. D., *J. Chem. Phys.*, 1983, **79**, 926–936.

[56] Jorgensen, W. L., *J. Am. Chem. Soc.*, 1981, **103**, 335–340.

[57] Wheelwright, E. J.; Spedding, F. H. and Schwarzenbach, G., *J. Am. Chem. Soc.*, 1953, **75**, 4196–4201.

[58] Chalmers, L., *Manuf. Chem. Aero.*, 1978, **49**, 79.

[59] Flaschka, H. A., *EDTA titrations; an introduction to theory and practice*, New York, Pergamon Press, 2 ed., 1964.

[60] Yang, Z.; D'Entremont, M.; Ni, Y. and Heiningen, A. R. P., *Pulp. Pap. Can.*, 1997, **11**, 408–413.

[61] Whalley, G., *Manuf. Chemi.*, 1991, **62**, 22–23.

[62] Cranton, E. M. and Pauling, L., *A Textbook on EDTA Chelation Therapy*, Hampton Roads Publishing Company, Inc., 2 ed., 2001.

[63] Verissimo, M.; Oliveira, J. and Gomes, M., *Sens. Actu.*, 2007, **127**, 102–106.

[64] Cheng, K. L.; Ueno, K. and Imamura, T., *Handbook of Organic Analytical Reagents*, CRC Press Inc., 1982.

[65] Font-Bardia, M.; Solans, X. and Font-Altaba, M., *Acta Cryst. C*, 1993, **49**, 1452–1456.

[66] Passer, E.; White, J. and Cheng, K., *Inorg. Chim. Acta*, 1977, **24**, 13–23.

[67] Barnett, B. and Uchtman, V., *Inorg. Chem.*, 1979, **18**, 2674–2678.

[68] Thompson, M. A.; Glendening, E. D. and Feller, D., *J. Phys. Chem.*, 1994, **98**, 10465–10476.

[69] Glendening, E. D.; Feller, D. and Thompson, M. A., *J. Am. Chem. Soc.*, 1994, **116**, 10657–10669.

[70] Glendening, E. D. and Feller, D., *J. Phys. Chem.*, 1996, **100**, 4790–4797.

[71] Glendening, E. D. and Feller, D., *J. Phys. Chem.*, 1995, **99**, 3060–3067.

[72] Feller, D., *J. Comp. Chem.*, 1996, **17**, 1571–1586.

[73] Schuchardt, K.; Didier, B.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J. and Windus, T., *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.

[74] Ahlrichs, R. and May, K., *Phys. Chem. Chem. Phys.*, 2000, **2**, 943–945.

[75] Turbomole version 5.10. Ahlrichs, R. and et al., ; http://www.cosmologic.de/data/DOK.pdf.

[76] Klobukowski, M.; Huzinaga, S. and Sakai, M. In Leszczynski, J., Ed., *Computational Chemistry: Reviews of Current Trends*, Vol. 3. World Scientific, Singapore, 1999.

[77] Miyoshi, E.; Mori, H.; Hirayama, R.; Osanai, Y.; Noro, T.; Hondo, H. and Klobukowski, M., *J. Chem.Phys.*, 2006, **122**, 74104–74112.

[78] Maleknia, S. and Brodbelt, J., *J. Am. Chem. Soc*, 1992, **114**, 4295–4298.

[79] Martell, A. E. and Smith, R. M., *Critical Stability Constants*, Vol. 1, Plenum Press, 1974.

[80] Miertus, S.; Scrocco, E. and Tomasi, J., *Chem. Phys.*, 1981, **55**, 117–129.

[81] Durand, S.; Dognon, J.-P.; Guilbaud, P.; Rabbe, C. and Wipff, G., *J. Chem. Soc. Perkin Trans. 2*, 2000, **4**, 705–714.

[82] Foster, J. P. and Weinhold, F., *J. Am. Chem. Soc.*, 1980, **102**, 7211–7218.

[83] Reed, A. E.; Curtiss, L. A. and Weinhold, F., *Chem. Rev.*, 1988, **88**, 899–926.

[84] Pedersen, C., *J. Am. Chem. Soc.*, 1967, **89**, 2495–2496.

[85] Humphry-Baker, R.; Gratzel, M.; Tundo, P. and Pelizzetti, E., *Angew. Chem. Int. Ed. Engl.*, 1979, **18**, 630–631.

[86] Gubelmann, M.; Harriman, A.; Lehn, J.-M. and Sessler, J., *J. Chem. Soc., Chem. Comm.*, 1988, **2**, 77–79.

[87] Gubelmann, M.; Harriman, A.; Lehn, J.-M. and Sessler, J., *J. Phys. Chem.*, 1990, **94**, 308–315.

[88] Gromov, S. and Alfimov, M., *Russ. Chem. Bull.*, 1997, **46**, 611–636.

[89] Takeshita, M.; Soong, C. and Irie, M., *Tetra. Lett.*, 1998, **39**, 7717–7720.

[90] Kawai, S., *Tetra. Lett.*, 1998, **39**, 4445–4448.

[91] Arenaz, P.; Bitticks, I.; Pannell, K. and Garcia, S., *Mutat. Res.*, 1992, **280**, 109–115.

[92] Bagatur'yants, A.; Freidzon, A.; Alfimov, M.; Baerends, E.; Howard, J. and Kuz'mina, L., *J. Mol. Struct. (Theochem)*, 2002, **588**, 55–69.

[93] Al-Jallal, N.; Al-Kahtani, A. and El-Azhary, A., *J. Phys. Chem. A*, 2005, **109**, 3694–3703.

[94] Wang, J. and Kollman, P. A., *J. Am. Chem. Soc.*, 1998, **120**, 11106–11114.

[95] Haddon, R. and Raghavachari, K., *J. Am. Chem. Soc.*, 1985, **107**, 289–298.

[96] Glendening, E. D. and Feller, D., *J. Am. Chem. Soc.*, 1996, **118**, 6052–6059.

[97] Feller, D.; Apra, E.; Nichols, J. A. and Bernholdt, D. E., *J. Chem. Phys.*, 1996, **105**, 1940–1951.

[98] Feller, D., *J. Phys. Chem.*, 1997, **A101**, 2723–2731.

[99] Feller, D.; Thompson, M. A. and Kendall, R. A., *J. Phys. Chem.*, 1997, **A101**, 7292–7298.

[100] Baerends, E.; Ellis, D. and Ros, P., *Chem. Phys.*, 1973, **2**, 41–51.

[101] Baerends, E. and Ros, P., *Chem. Phys.*, 1973, **2**, 52–59.

[102] Baerends, E. and Ros, P., *Int. J. Quantum Chem.*, 1978, **12**, 169–177.

[103] Boerrigter, P.; te Velde, G. and Baerends, E., *Int. J. Quantum Chem.*, 1988, **33**, 87–113.

[104] te Velde, G. and Baerends, E., *J. Comput. Phys.*, 1992, **99**, 84–98.

[105] Dunitz, J. D.; Dobler, M.; Seiler, P. and Phizackerley, R. P., *Acta Cryst. B*, 1974, **30**, 2733–2738.

[106] Maverick, E.; Seiler, P.; Schweizer, W. B. and Dunitz, J. D., *Acta Cryst. B*, 1980, **36**, 615–620.

[107] Colton, R.; Mitchell, S. and Traeger, J., *Inorg. Chim. Acta*, 1995, **231**, 87–93.

[108] Brodbelt, J. and Liou, C., *Pure Appl. Chem.*, 1993, **65**, 409–414.

[109] Wong, P.; Antonio, B. and Dearden, D., *J. Am. Soc. Mass Spectrom.*, 1994, **5**, 632–637.

[110] Brodbelt, J., Personal communication.

[111] Sondheimer, F. and Wolovsky, R., *Tetra. Lett.*, 1959, **1**, 3–6.

[112] Sondheimer, F.; Wolowsky, R. and Amiel, Y., *J. Am. Chem. Soc.*, 1962, **84**, 274–284.

[113] Sondheimer, F., *Acc. Chem. Res.*, 1972, **5**, 81–91.

[114] Drew, M.; de O. Cabral, J.; Cabral, M. F.; Esho, F. and Nelson, S. M., *J. Chem. Soc., Chem. Comm.*, 1979, **22**, 1033–1035.

[115] Bell, T. W. and Guzzo, F., *J. Am. Chem. Soc.*, 1984, **106**, 6111–6112.

[116] Bell, T. W. and Firestone, A., *J. Am. Chem. Soc.*, 1986, **108**, 8109–8111.

[117] Bell, T. W. and Guzzo, F., *J. Chem. Soc., Chem. Comm.*, 1986, **10**, 769–771.

[118] Bell, T. W.; Firestone, A.; Guzzo, F. and Hu, L.-Y., *J. Incl. Phenom. Macrocycl. Chem.*, 1987, **5**, 149–152.

[119] Bell, T. W.; Firestone, A. and Ludwig, R., *J. Chem. Soc., Chem. Comm.*, 1989, **24**, 1902–1904.

[120] Bell, T. W.; Guzzo, F. and Drew, M., *J. Am. Chem. Soc.*, 1991, **113**, 3115–3122.

[121] Honeybourne, C., *Tetra.*, 1973, **29**, 1549–1557.

[122] Haddon, R., *Pure Appl. Chem.*, 1982, **54**, 1129–1142.

[123] Kao, J., *J. Am. Chem. Soc.*, 1987, **109**, 3817–3829.

[124] Kao, J., *J. Comput. Chem.*, 1988, **9**, 905–923.

[125] Woon, D. and Dunning Jr., T., *J. Chem. Phys.*, 1993, **98**, 1358–1371.

[126] Peterson, K. A., *J. Chem. Phys.*, 2003, **119**, 11099–11113.

[127] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H. and Dolg, M., *J. Chem. Phys.*, 2003, **119**, 11113–11124.

[128] Boys, S. and Bernardi, F., *Mol. Phys.*, 1970, **19**, 553–566.

[129] Mulliken, R., *J. Chem. Phys.*, 1955, **23**, 1833–1842.

[130] Yoshizawa, K.; Kato, T. and Yamabe, T., *Synth. Met.*, 1997, **86**, 2397–2398.

[131] Baldridge, K. K. and Siegel, J. S., *Angew. Chem. Int. Ed. Engl.*, 1007, **36**, 745–748.

[132] Wannere, C. S.; Sattelmeyer, K. W.; III, H. F. S. and von Rague Schleyer, P., *Angew. Chem. Int. Ed. Engl.*, 2004, **43**, 4200–4206.

[133] Ermer, O., *Helv. Chim. Acta*, 2005, **88**, 2262–2266.

[134] Seiler, P.; Dobler, M. and Dunitz, J., *Acta Cryst. B*, 1974, **30**, 2744–2745.

[135] Dobler, M. and Phizackerley, R., *Acta Cryst. B*, 1974, **30**, 2746–2748.

[136] Dobler, M. and Phizackerley, R., *Acta Cryst. B*, 1974, **30**, 2748–2750.

[137] Gnuplot version 4.4.0; http://www.gnuplot.info/.

[138] Voet, D.; Voet, J. G. and Pratt, C. W., *Fundamentals of Biochemistry*, John Wiley and Sons, Inc., 1999.

[139] Harris, D. A., *Bioenergetics at a Glance*, Blackwell Publishing, 1995.

[140] Pelengaris, S. and Khan, M., *The Molecular Biology of Cancer*, Blackwell Publishing, 2006.

[141] Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K. and Watson, J. D., *Molecular Biology of the Cell*, Garland Publishing, Inc., 1983.

[142] Aneja, R.; Vangapandu, S. N.; Lopus, M.; Viswesarappa, V. G.; Dhiman, N.; Verma, A.; Chandra, R.; Panda, D. and Joshi, H. C., *Biochem. Pharma.*, 2006, **72**, 415–426.

[143] Sebak, S.; Mirzaei, M.; Malhotra, M.; Kulamarva, A. and Prakash, S., *Int. J. Nanomed.*, 2010, **5**, 525–532.

[144] Karna, P.; Sharp, S.; Yates, C.; Prakash, S. and Aneja, R., *Mol. Cancer*, 2009, **8**, 93–104.

[145] Aneja, R.; Ghaleb, A. M.; Zhou, J.; Yang, V. W. and Joshi, H. C., *Cancer Res.*, 2007, **67**, 3862–3870.

[146] Aneja, R.; Miyagi, T.; Karna, P.; Ezell, T.; Shukla, D.; Vij Gupta, M.; Yates, C.; Chinni, S. R.; Zhau, H.; Chung, L. W. K. and Joshi, H. C., *Eur. J. Cancer*, 2010, **46**, 1668–1678.

[147] Aneja, R.; Liu, M.; Yates, C.; Gao, J.; Dong, X.; Zhou, B.; Vangapandu, S. N.; Zhou, J. and Joshi, H. C., *Cancer Res.*, 2008, **68**, 1495–1503.

[148] Aneja, R.; Vangapandu, S. N.; Lopus, M.; Chandra, R.; Dulal, P. and Joshi, H. C., *Mol. Pharm.*, 2007, **72**, 486–486.

[149] Jaiswal, A. S.; Aneja, R.; Connors, S. K.; Joshi, H. C.; Multani, A. S.; Pathak, S. and Narayan, S., *J. Cell. Biochem.*, 2009, **106**, 1146–1156.

[150] Zughaier, S.; Karna, P.; Stephens, D. and Aneja, R., *PLoS One*, 2010, **5**, e9165.

[151] von Langermann, J.; Lorenz, H.; Boehm, O.; Flemming, A.; Bernsdorf, A.; Köckerling, M.; Schinzer, D. and Seidel-Morgenstern, A., *Acta Cryst. E*, 2010, **66**, o570.

[152] Klamt, A. and Schüümann, G., *J. Chem. Soc.*, 1993, **5**, 799–805.

[153] Canadian cancer statistics 2010 (cancer.ca).

[154] Goodman, J. and Walsh, V., *The Story of Taxol*, Cambridge University Press, 2001.

[155] Trojanowski, J. Q.; Smith, A. B.; Huryn, D. and Lee, V. M.-Y., *Cancer Res.*, 2004, **64**, 5063–5067.

[156] Quinoa, E.; Kakou, Y. and Crews, P., *J. Org. Chem.*, 1988, **53**, 3642–3644.

[157] Hari, M.; Yang, H.; Zeng, C.; Canizales, M. and Cabral, F., *Cell Motil. Cytoskeleton*, 2003, **56**, 45–56.

[158] Leandro-García, L. J.; Leskela, S.; Landa, I.; Montero-Conde, C.; Lopez-Jimenez, E.; Leton, R.; Cascon, A.; Robledo, M. and Rodriguez-Antona, C., *Cytoskeleton*, 2010, **67**, 214–223.

[159] Mooberry, S. L.; Tien, G.; Hernandez, A. H.; Plubrukarn, A. and Davidson, B. S., *Cancer Res.*, 1999, **59**, 653–660.

[160] Mooberry, S. L.; Randall-Hlubek, D. A.; Leal, R. M.; Hegde, S. G.; Hubbard, R. D.; Zhang, L. and Wender, P. A., *PNAS*, 2004, **101**, 8803–8808.

[161] Mooberry, S. L.; Hilinski, M. K.; Clark, E. A. and Wender, P. A., *Mol. Pharm.*, 2008, **5**, 829–838.

[162] Johnson, T. A.; Tenney, K.; Cichewicz, R. H.; Morinaka, B. I.; White, K. N.; Amagata, T.; Subramanian, B.; Media, J.; Mooberry, S. L.; Valeriote, F. A. and Crews, P., *J. Med. Chem.*, 2007, **50**, 3795–3803.

[163] Gallagher, B. M., *Curr. Med. Chem.*, 2007, **14**, 2959–2967.

[164] Gaitanos, T. N.; Buey, R. M.; Diaz, J. F.; Northcote, P. T.; Teesdale-Spittle, P.; Andreu, J. M. and Miller, J. H., *Cancer Res.*, 2004, **64**, 5063–5067.

[165] West, L. M. and Northcote, P. T., *J. Org. Chem.*, 2000, **65**, 445–449.

[166] Hood, K. A.; West, L. M.; Rouwe, B.; Northcote, P. T.; Berridge, M. V.; Wakefield, S. J. and Miller, J. H., *Cancer Res.*, 2002, **62**, 3356–3360.

[167] Jimenez-Barbero, J.; Canales, A.; Northcote, P. T.; Buey, R. M.; Andrew, J. M. and Diaz, J. F., *J. Am. Chem. Soc*, 2006, **128**, 8757–8765.

[168] Crume, K. P.; Miller, J. H. and Flamme, A. C. L., *Exp. Biol. and Med.*, 2007, **232**, 607–613.

[169] Miller, J. H.; Singh, A. J. and Northcote, P. T., *Mar. Drugs*, 2010, **8**, 1059–1079.

[170] Portions of these results were generated by ADMET predictor 5.5 software. Simulations Plus, Inc., Lancaster, California, USA.

[171] Huzil, J. T.; Chen, K.; Kurgan, L. and Tuszynski, J., *Cancer Infor.*, 2007, **3**, 159–181.

[172] Percy, A. J.; Slysz, G. W. and Schriemer, D. C., *Anal. Chem.*, 2009, **81**, 7900–7907.

[173] Bennett, M. J.; Barakat, K.; Huzil, J. T.; Tuszynski, J. and Schriemer, D. C., *Chem. Biol.*, 2010, **17**, 725–734.

[174] Huzil, J. T.; Chick, J. K.; Slysz, G. W.; Freedman, H.; Tuszynski, J.; Taylor, R. E.; Sackett, D. L. and Schriemer, D. C., *J. Mol. Biol.*, 2008, **378**, 1016–1030.

[175] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A. and Simmerling, C., *Proteins*, 2006, **65**, 712–725.

[176] Cezard, C.; Vanquelef, E.; Pecher, J.; Sonnet, P.; Cieplak, P.; Derat, E. and Dupradeau, F. 236th ACS National Meeting, Philadelphia, PA, USA, 2008.

[177] Wang, J. M.; Wolf, R.; Caldwell, J. W.; Kollman, P. A. and Case, D. A., *J. Comp. Chem.*, 2004, **25**, 1157–1174.

[178] Banerjee, A., *Biochem. Biophys. Res. Comm.*, 2002, **293**, 598–601.