INAUGURAL – DISSERTATION

Zur Erlangung der Doktorwürde der

Naturwissenschaftlich-Mathematischen Gesamtfakultät der

Ruprecht-Karls-Universität

Heidelberg

The Mode of Action of Embelin as an Alternative Photosensitizer for Photodynamic Therapy

Vorlegt von

Michael Rogo Opata, Master of Science, Chemistry

aus Bondo, Kenia

The Mode of Action of Embelin as an Alternative Photosensitizer for Photodynamic Therapy

Michael Rogo Opata

Gutacher:

Prof. Dr. Andreas Dreuw

Prof. Dr. Petra Tegeder

Tag der Mündlichen Prüfung: 09.12.2021

To my Family who sailed with me through every storm

Hawi loyo Rieko

~ Luo Proverb

Abstract

Effective treatment of cancer using existing drugs remains a challenge due to side effects experienced by cancer patients during therapy. Research towards discovering new anticancer agents with less side effects is an ongoing effort which unfortunately requires huge investments from lead compound discovery to clinical trials. Conventional methods used in treating cancer include chemotherapy, radiation, as well as surgery. These methods have seen tremendous improvements, but a lot remains to be done to slay the cancer dragon. In the recent past, attention has turned to exploring the use of photo active molecules to treat cancer and other ailments such as microbials in a process known as photodynamic therapy (PDT). These molecules are commonly known as photosensitizers (PS) and should possess adequate structural characteristics that promote synergy effects when exposed to certain wavelengths of light.

Ideally, the absorption wavelength of PS should range between ~ 500-800 nm which is critical for tissue penetration. Popular PS include Porphyrin whose derivatives have been successfully applied in PDT treatment. Alternatives to the widely studied porphyrin scaffold also exist, and more scaffolds need to be added onto the wide therapeutics required to treat cancer. In this work, a theoretical investigation of embelin which has a benzoquinone scaffold has been carried out. Interesting properties are explored including electronic resonance, proton transfer pathways and the ability to generate singlet oxygen. Furthermore, its ability to bind the X-linked inhibitor of apoptosis is investigated and the binding mechanisms as well as binding energies are reported. Advanced hybrid QM/MM studies have also been carried out to investigate its spectral signatures within the protein environment.

This work begins by addressing embelin's ground- and -excited state properties in **chapter 3**. The aim of this chapter is to highlight embelin's individual molecular properties. The calculations are carried out with and without solvation models. Results obtained are the excitation energies of embelin, the nature of its geometrical parameters in the ground and excited state potential energy surfaces (PES), and its general ability to generate singlet oxygen. The Algebraic Diagrammatic Construction

(ADC) and Time-dependent Density Functional Theory (TD-DFT) methods are extensively used to approximate the excitation energies and molecular properties of embelin. Particularly, ADC(3) accurately reproduces the excitation energy at 4.32 eV compared to the experimental absorption energy of 4.31 eV. In the ground state, embelin is found to exhibit a weak intramolecular hydrogen bond confirming previous reported results on the existence of electronic resonance within its structure. This property is enhanced on photon absorption by hyperbonds within the structure, meaning that photon absorption stabilizes embelin in the exited state and promotes resonance. In fact, a detailed investigation using potential energy surface (PES) scans resulted in intermediates which are in complete agreement with the excited state stabilizations which favor excited state intramolecular proton transfer (ESIPT). The intermediates formed in the excited state further promote efficient intersystem crossing to the triplet state with enough excess energy to generate reactive oxygen species (ROS) such as singlet oxygen.

In **chapter 4**, the binding pattern of embelin against X-linked Inhibitor of apoptosis (XIAP) is presented. Here, focus is turned towards identifying the "*active site*" where embelin prefers to bind the baculoviral inhibitor of the apoptosis (BIR3) domain of XIAP. XIAPs are known to bind caspases hence interfere with normal cell death which results in proliferation of abnormal cells. Embelin is found to bind mostly to Glu and Tyr residues (which have been experimentally identified as one of the critical amino acids in embelin's binding pattern). The binding is measured as a function of the hydrogen bonding percentage with values as high as 60%. The dominating binding pattern is a hydrogen donating mechanism with the protein residue pairs acting as proton acceptors. The binding energy is calculated to be -24.7 kJ/mol which agrees nicely with -25.1 kJ/mol (the experimental value of Human serum albumin).

Since the overall aim of the thesis is to address the applicability of embelin in PDT which occurs within a protein environment, advanced hybrid QM/MM calculations are presented in **chapter 5**. The major results in this chapter are the spectral signature of embelin embedded within the BIR3 environment. The reported vertical excitation energies are comparable to the values obtained in **chapter 3**. More importantly, embelin is stable within the protein environment and absorbs light with similar wavelengths reported by experiment. The absorption wavelengths are however

slightly shifted when only electrostatics are taken into consideration. Inclusion of polarization effects simulated using PE-ADC resulted in slightly red-shifted excitation energies. In general, the results obtained within this chapter are consistent and predict the ability of embelin to generate singlet oxygen.

Zusammenfassung

Die wirksame Behandlung von Krebs mit bestehenden Medikamenten bleibt eine medizinische Herausforderung, da Krebspatienten während der Behandlung Nebenwirkungen verspüren. Die Forschung zur Entdeckung neuer Krebsmedikamente mit weniger Nebenwirkungen ist ein ständiges Unterfangen, das leider enorme Investitionen von der Entdeckung der Leitmoleküle bis hin zu klinischen Versuchen erfordert. Zu den konventionellen Methoden der Krebsbehandlung gehören Chemotherapie, Bestrahlung und Operation. Diese Methoden haben zu enormen Verbesserungen in der Behandlung geführt. In jüngster Zeit hat sich die Aufmerksamkeit auf die Erforschung photoaktiver Moleküle zur Behandlung von Krebs und anderen Krankheiten, wie z. B. Mikroben, in einem als photodynamische Therapie (PDT) bekannten Verfahren gerichtet. Diese Moleküle werden gemeinhin als Photosensibilisatoren (PS) bezeichnet und sollten geeignete strukturelle Merkmale aufweisen, die Synergieeffekte fördern, wenn sie bestimmten Wellenlängen von Licht ausgesetzt werden.

Im Idealfall sollte die Absorptionswellenlänge der PS zwischen 500 und 800 nm liegen, was für die Gewebepenetration günstig ist. Zu den beliebten PS gehören Porphyrine, deren Derivate bei PDT-Behandlungsansätzen erfolgreich eingesetzt wurden. Es gibt jedoch auch Alternativen zu dem weithin untersuchten Porphyrin-Gerüst, und es werden weitere gesucht, um die breite Palette der für die Krebsbehandlung erforderlichen Therapeutika zu ergänzen. In dieser Arbeit wurde eine theoretische Untersuchung von Embelin durchgeführt, das ein Benzochinon-Gerüst hat. Es werden seine interessanten Eigenschaften erforscht, darunter electronische Resonanz, Protonentransfer und die Fähigkeit, Singulett-Sauerstoff zu erzeugen. Darüber hinaus wird seine Fähigkeit untersucht, den X-chromosomalen Apoptoseinhibitor zu binden, und es wird über die Bindungsmechanismen und Bindungsenergien in diesem Zusammenhang berichtet. Es wurden fortgeschrittene hybride QM/MM-Studien durchgeführt, um die spektralen Signaturen von Embelin innerhalb der Proteinumgebung zu untersuchen.

In dieser Arbeit wird zunächst in **Kapitel 3** auf die Eigenschaften von Embelin im Grundzustand und im angeregten Zustand eingegangen. Das Ziel dieses Kapitels ist es, die molekularen Eigenschaften im Grundzustand und im angeregten Zustand zu betrachten. Die Berechnungen werden sowohl mit als auch ohne Lösungsmodelle durchgeführt. Die Ergebnisse umfassen die Anregungsenergien von Embelin, den Charakter seiner geometrischen Parameter auf der Potentialhyperfläche (PES) im Grundzustand und im angeregten Zustand sowie seine Fähigkeit, Singulett-Sauerstoff zu erzeugen. Die ADC- und TD-DFT-Methoden werden genutzt, um die Anregungsenergien und molekularen Eigenschaften von Embelin zu approximieren. Insbesondere ADC(3) reproduziert die Anregungsenergie mit 4,32 eV sehr genau, verglichen mit der experimentellen Absorptionsenergie von 4,31 eV. Im Grundzustand weist Embelin eine schwache intramolekulare Wasserstoffbrückenbindung auf, was frühere Ergebnisse über das Vorhandensein von elektronischer Resonanz in seiner Struktur bestätigt. Diese Eigenschaft wird bei Photonenabsorption durch das Vorhandensein von Hyperbindungen innerhalb der Struktur verstärkt, was bedeutet, dass die Photonenabsorption die Resonanz fördert. Die Untersuchung seiner dynamischen Eigenschaften mit Hilfe von PES scans ergab Zwischenstufen, die mit der Stabilisierung im angeregten Zustand völlig übereinstimmen, welche den intramolekularen Protonentransfer im angeregten Zustand (ESIPT) begunstigen. Die im angeregten Zustand gebildeten Zwischenprodukte fördern außerdem ein effizientes Intersystem Crossing zum Triplett-Zustand. Im Triplett-Zustand verfügt Embelin über einen ausreichenden Energieüberschuss, um reaktive Sauerstoffspezies (ROS) wie Singulett-Sauerstoff zu erzeugen.

In **Kapitel 4** wird das Bindungsmuster von Embelin an den X-linked Inhibitor of Apoptosis (XIAP) vorgestellt. Hier liegt der Schwerpunkt auf der Identifizierung der "aktiven Stelle", an der Embelin bevorzugt an die BIR3-Domäne von XIAP bindet, dem baculoviralen Inhibitor der Apoptose. XIAPs sind dafür bekannt, dass sie Caspasen binden und damit den normalen Zelltod stören, was zu einer Vermehrung abnormaler Zellen führt. Embelin bindet hauptsächlich an Glu- und Tyr-Reste (die experimentell als eine der kritischen Aminosäuren im Bindungsmuster von Embelin identifiziert wurden). Die Bindung wird als Funktion des Wasserstoffbrückenbindungsanteils gemessen, wobei die Werte bis zu 60 % betragen.

Das vorherrschende Bindungsmuster ist ein Wasserstoff spendender Mechanismus, bei dem die Proteinrestpaare als Protonenakzeptoren fungieren. Die Bindungsenergie wird mit -24,7 kJ/mol berechnet, was gut mit -25,1 kJ/mol (dem experimentellen Wert von Humanserumalbumin) übereinstimmt.

Da das übergeordnete Ziel der Arbeit darin besteht, den Nutzen von Embelin als PDT zu untersuchen, dessen Anwendung in einer Proteinumgebung stattfindet, werden in **Kapitel 5** fortgeschrittene hybrid QM/MM-Berechnungen vorgestellt. Das wichtigste Ergebnis in diesem Kapitel ist das Absorptions- Spektrum von Embelin, eingebettet in die BIR3-Umgebung. Die vertikalen Anregungsenergien sind mit den in **Kapitel 3** ermittelten Werten vergleichbar. Noch wichtiger ist, dass Embelin in der Proteinumgebung stabil ist und Licht in einem Wellenlängenbereich absorbiert, der auch im Experiment ermittelt wurde. Unter alleiniger Berücksichtigung der Elektrostatik sind die Absorptionswellenlängen jedoch leicht blau verschoben. Die mit PE-ADC simulierten Polarisationseffekte führen zu einem leicht rotverschobenen Spektrum. Im Allgemeinen sind die in diesem Kapitel erzielten Ergebnisse konsistent und belegen die Fähigkeit von Embelin, Singulett-Sauerstoff zu erzeugen.

Table of Contents

ABBREVIATIONS

ADC – Algebraic Diagrammatic Construction

BIR – Baculoviral Inhibitor of Apoptosis Repeat

cc-PVTZ – Correlation Consistent - Polarized *Valence* Triple Zeta

CI – Configuration Interaction

CT – Charge Transfer

DFT – Density Functional Theory

DNA – Deoxyribonucleic Acid

ER - Endoplasmic Reticulum

ESIPT – Excited State Intramolecular Proton Transfer

FDA– Food and Drug Administration

GGA – Generalized Gradient Approximation

HF – Hartree Fock

ISC -Intersystem Crossing

T – Kinetic Energy

KS-DFT – Kohn Sham DFT

libwfa – Wave Function Library

MBPT – Many Body Perturbation Theory

MPPT – Møller Plesset Perturbation Theory

NBO – Natural Bond Orbitals

NRT – Natural Resonance Theory

NTO – Natural Transition Orbital

PCM – Polarizable Continuum Model

PDB – Protein Data Bank

PDT – Photodynamic Therapy

PES – Potential Energy Surface

PS – Photosensitizers

QM/MM – Hybrid Quantum Mechanics / Molecular Mechanics

ROS – Reactive Oxygen Species

SD – Slater Determinant

TD-DFT – Time-Dependent Density Functional Theory

TTET - Triplet-Triplet Energy Transfer

XIAP – X-Linked Inhibitor of Apoptosis

Chapter - 1

Introduction

The therapeutic properties of light have played a significant role in disease treatment since ancient times and continue to be employed in modern medicinal practice. This approach of treatment utilizes effective chemical molecules known as photosensitizers (PS) which on interaction with light sources trigger synergy effects that alleviate disease symptoms or result in cure.[1] Treatment approaches using a combination of light sources, a PS, and singlet oxygen is known as Photodynamic Therapy (PDT) and has been applied in tackling various diseases such as cancer and microbial infections[1–3] among others.

Increased PDT applications have surged particularly in attempts to treat cancer due to its selective nature and ability to cause site specific subcellular photodamage.[1,2,4] The mechanisms through which it acts on tumor cells thereby causing cell death (also referred to as apoptosis) are known to cause minimal side effects and provides for a safer alternative to cancer treatment. This major advantage makes PDT preferrable over traditional chemotherapy approaches and remains an active promising area of research.

To achieve such a mean feat, efficient PS must be identified. These molecules should possess qualities that are responsible for the underlying photophysical processes which in turn produce desired pharmacological effect. Typically, the known therapeutic window for an effective PS fall within the wavelength range of ~ 500 to 850 nm to achieve tissue penetration.[5] Therefore, molecules with ability to accumulate in the cell as well as absorb light at these wavelengths qualify as promising candidates. Porphyrin scaffold and it's derivatives remain the most studied molecular class with hematoporphyrin being the earliest candidate.[6,7] It's chlorin and bacteriochlorin derivatives absorb light at red shifted wavelengths of around 630 nm, resulting in enhanced photochemical properties suitable to act as anti-tumor agents.[8] Some of the

current food drug and Administration (FDA) approved drugs for PDT use on various cancer ailments include Photfrin®, Levulan ®, and Metvixia®.

The success of PDT mechanism is dictated by the photochemical signatures of the chosen PS. On absorption of light, the PS is subsequently converted to an electronically excited state. These states can be of different spins. However, in the first excited state the spin is usually conserved and is hence a singlet. Due to the fact that singlets have short lifetimes ($10^{-6} - 10^{-9}$ s), the excited species can either decay back to the ground state in a radiationless fashion or via fluorescence (emitting a photon). Alternatively, it can undergo spin conversion to a long-lived triplet excited state through intersystem crossing. It is at this state that the PS can undergo two interesting types of reactions which are biologically relevant to PDT. Reaction type I involves the interaction of the triplet excited PS and molecular oxygen (which is a triplet in the ground state) through energy transfer reaction to produce singlet reactive oxygen species (ROS) which lead to photodamage.[9,10] Type I photoreaction is usually the most desirable mechanism since it produces the oxidized form of biomolecules and has been widely researched upon.[5,11,12] Reaction type II involves interaction of the PS and biomolecule directly through hydrogen atom or electron transfer producing radical species of either the PS and or the substrate.[6] The species can then react directly with oxygen thereby forming superoxide ions, hydroxyl radicals and or peroxides. Figure 1 summarizes the reaction mechanism that a suitable PS undergoes to produce a singlet oxygen and other products.

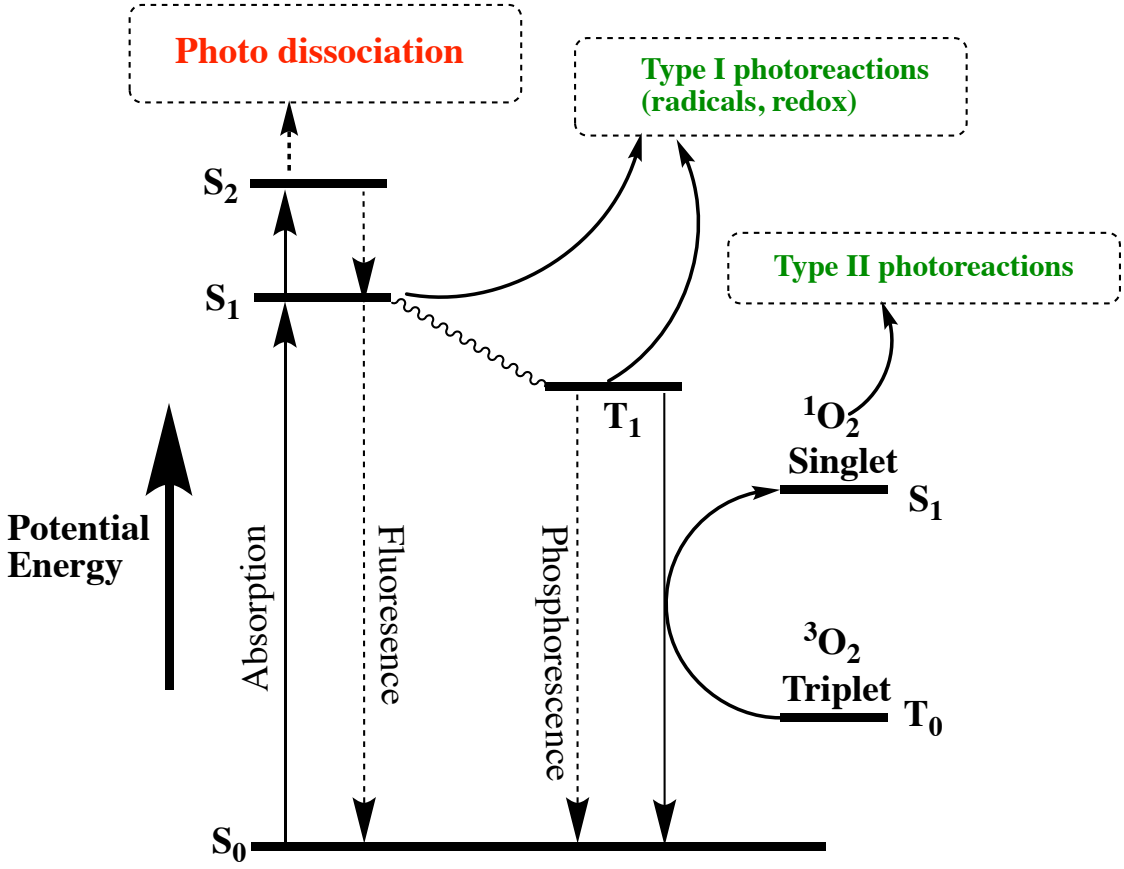


Figure 1: Photo reaction mechanism known to facilitate PDT process via Type I and Type II reactions.

The singlet oxygen generated by the PS can selectively interact with tumor cells thereby curing cancer. In essence, programmed cell death (apoptosis) which is a key biological process and occurs through triplet-triplet energy transfer (TTET) is promoted. ROS driven apoptosis has been previously reported to occur in lysosome, endoplasmic reticulum (ER), and mitochondria.[2] On the other hand, failure of such a programmed cell death would result in chain reactions leading to abnormal cell growth, which may eventually lead to cancer and related ailments.

There are specific biomolecules mentioned above known to promote abnormal cell growth which include presence of caspase binding proteins that interfere with key cell death pathways. The X-linked inhibitor of apoptosis proteins (XIAP) are one such group of biomolecules whose activity hinder cell death in both intrinsic and extrinsic pathways.[13] XIAP achieve its function through zinc-binding baculovirus IAP repeat (BIR) motifs that have a substrate preference to caspases. From a cellular point of view, caspases play an important role of triggering important cell death pathways.[14] Once blocked, then cell death processes are also hindered. The BIR2 and BIR3

domains of the XIAP have been found to be major apoptosis antagonists and can therefore be used to develop drug-like apoptosis agonists. As such, studies can be conducted to probe known, as well as new binding pockets within these domains to develop new lead molecules that can be applicable to PDT. Once the BIR domains of XIAP are inhibited, caspases would then be *"free"* to perform their crucial roles of initiating cell death reactions. Part of the goals in this work was to investigate BIR3 binding mechanism using a molecule known to experimentally bind on its surface groove thereby promoting apoptosis.

Apart from porphyrins, there exist other natural chromophores that are known to act as PS and their suitability have been reported in literature.[15] Benzoquinones and its polycyclic derivatives such as anthraquinones, perylenequinones and phenalenones have particularly seen the light of day as promising alternatives to porphyrins. In fact, this group of new PS are known to exist naturally and absorb light within the wavelength range applicable for PDT.[16] It remains a subject of intense research to discover new scaffolds that can be used to treat cancer and other ailments using PDT approach.

In search of alternative PS, I have carried out a theoretical investigation on the potential that embelin holds as a PDT candidate. Embelin is a naturally occurring quinone scaffold which has been previously isolated from various plants including *Embelia ribes*[17] and *Embelia schimperi* [18] among other species. It is one of the many isolated molecules from these plant species and have proven to have a broad biological activity spectrum against cancer, inflammation, diabetes and microbials.[19–21] The molecule has two *para-hydroxyl* and quinonic moieties which is believed to modulate its ground state biological activity. Most relevant to my work is the fact that embelin's use in photoreaction applications have been recently reported and therefore pose as a promising PS.[3,22] This work is motivated by the latter findings. Specifically, an in depth understanding of the underlying photochemical phenomena of embelin resulting from its interaction with light given its unique structure is sought. Insights of such understanding has been further transferred to a protein environment where embelin-BIR3 interactions have been explored to investigate its binding mechanism within the BIR3 protein environment using molecular dynamics as well as hybrid quantum mechanics/molecular mechanics (QM/MM) methods. I believe that this

work lays a fundamental background to understanding the basic photo mechanism of how embelin as well as other hydroquinone moieties, modulate broad biological activity spectrum both in ground and excited states. It could also serve as a stepping-stone to advanced structural modifications to enhance the absorption bands of the base benzoquinone scaffold in a quest to develop alternative PDT compounds.

In most cases, traditional lead compound discovery programs involve several approaches but relies majorly on molecular mechanics principles such as docking and charge fitting. These approaches typically involve investigation of static poses and orientations within a known pocket that closely mimic co-crystalized ligands bound to the proteins obtained from the PDB databank. In this work, a multistep approach has been used and is divided in three major chapters. I begin with results obtained from pure quantum chemical studies of embelin presented in **chapter 3** to explore the static as well as dynamic spectral signatures of embelin computed in vacuum using the ADC and TD-DFT theories. Implicit solvent effects are then included using PCM as well as explicit solvent models. Since embelin is known to have resonance within its structure, relaxed PES results are also presented in both ground and excited states in a bid to investigate the character of its dynamic signatures. This is followed by molecular mechanics studies in **chapter 4** using docking, molecular dynamics and Markov State Model calculations between embelin and BIR3 domain of the XIAP. A further hybrid quantum mechanics/molecular mechanics study is carried out in **chapter 5** to investigate the spectral signatures of embelin in a protein environment.

Chapter - 2

Theoretical Methods

This chapter briefly introduces the underlying theoretical methods behind the models used to carry out all of the simulations in this work. For the sake of compactness, mathematical rigor as well as in-depth expansion of the given formulae is sacrificed. Readers are therefore encouraged to refer to the given references which contain, in most cases, the original derivations and extensive reviews behind each theory.

In quantum chemistry, theoretical approximations to the energies as well as prediction of other molecular properties are achieved by solving the electronic time-independent Schrödinger equation which is always an eigenvalue problem.

$$\hat{H}_n|\Psi_n\rangle = E_n|\Psi_n\rangle \tag{1}$$

Where $\hat{H}_n$ is the Hamiltonian, $|\Psi_n\rangle$ is the wavefunction, and E_n is the energy of the n^{th} electron. The total Hamiltonian for *N* electrons and *K* nuclei can be written in atomic units as

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{K=1}^{M}\frac{Z_K}{r_{iK}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{2}$$

The first term is the electron kinetic energy $\hat{T}_e$, the second term is the potential energy of the electron in the field of the nuclei $\hat{V}_{ne}$, and the third term represents the electron-electron repulsion $\hat{V}_{ee}$, respectively. The total energy of the molecule in the Born-Oppenheimer approximation for a fixed nucleus can be obtained by adding the constant nuclear repulsion $\hat{V}_{nn}$. A brief introduction to the formalism of the methods used in this work will be reviewed.

.

2.1 Hartree-Fock Theory (HF)

HF theory formulates the starting point of all quantum chemistry methods and can be equated to a single slater determinant (SD) theory.[23] In SD theory, a set of spin orbitals $\{\chi_a\}$ that

forms a single slater determinant $|\Psi_o\rangle$ is systematically varied to minimize the electronic energy through the variational principle which states the energy of a system is an upper bound of its exact energy.

$$\left|\Psi_0\right\rangle = \left|\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\right\rangle \tag{3}$$

The spin orbitals used in the SD are often constrained to remain orthonormal and give rise to a minimized ground state electronic energy. In physicist notation, the energy is given as

$$E_0 = \left\langle\Psi_0\left|H\right|\Psi_0\right\rangle = \sum_a \left\langle a\left|h\right|a\right\rangle + \tfrac{1}{2}\sum_{ab}\left\langle ab\left|\right|bb\right\rangle \tag{4}$$

With the first and second parts of the equation being the one- and two-electron summations respectively. Writing the above equation using HF operators, integrating out spin, and simplifying steps, we arrive at a more visually appealing equation which contains the coulombic and exchange parts of equation (4).

$$E_0 = \sum_i h_{ii} + \tfrac{1}{2}\sum_{ij} J_{ij} - K_{ij} \tag{5}$$

Where h_{ii} is the single-electron kinetic and potential energy for attraction to the nuclei, J_{ij} is the one-electron coulomb potential which represents the average local potential, and K_{ij} the exchange term arising from antisymmetric nature of a SD. Rigorous derivation of HF theory is well documented in Modern quantum chemistry: introduction to advanced electronic structure theory.[23]

2.2 Electron Correlation Methods

2.2.1 Configuration Interaction

HF theory is a mean-field theory and accounts for electron exchange but lacks electron correlation which can be defined as the energy difference between the exact energy of a system and the Hartree-Fock energy.

$$E_0^{corr} = E_0^{exact} - E_0^{HF} \tag{6}$$

Beyond the HF approximation, it is desirable to find electron correlation with methods such as the configuration interaction (CI), many-body perturbation theory (MBPT), coupled cluster (CC) and density functional theory (DFT). A brief overview of these correlated methods is given below.

CI as a method is determined variationally as superposition of determinants, the so-called linear combination of slater determinants, as a way of including the correlation energy to obtain an improved ground state wavefunction and the energy of an *n*-electron system.[24] All the possible electron configurations i.e. (singles , doubles ...) are obtained from a reference HF determinant.

$$\left|C\right\rangle = \sum_i C_i \left|i\right\rangle \tag{7}$$

Where $\left|i\right\rangle$ is the ground state configuration

Minimization of the ground state variationally leads to the electronic energy

$$E_{CI} = \min_C \frac{\langle C|\hat{H}|C\rangle}{\langle C|C\rangle} \tag{8}$$

The expansion of the CI ground state wavefunction $\left|\Psi_0^{CI}\right\rangle$ consists of a chosen HF determinant $\left|\Phi_0\right\rangle$ and a set of singly and doubly excited determinants resulting from replacing one or more occupied orbitals *i,j*…. with unoccupied (virtual) orbitals *a,b* ….up to *N*-tuply excited determiants

$$\left|\Phi_0\right\rangle = c_0\left|\Psi_0\right\rangle + \sum_{ia} c_i^a\left|\Psi_i^a\right\rangle + \sum_{\substack{i<j \\ a<b}} c_{ij}^{ab}\left|\Psi_{ij}^{ab}\right\rangle \ldots \tag{9}$$

The resulting wave function in equation (9) is multiconfigurational and can be rewritten using a factor $(1/n!)^2$ in front of the summations to ensure excitations counting occurs only once to get

$$\left|\Phi_0\right\rangle = c_0\left|\Psi_0\right\rangle + (\tfrac{1}{1!})^2\sum_{ia} c_i^a\left|\Psi_i^a\right\rangle + (\tfrac{1}{2!})^2\sum_{\substack{i<j \\ a<b}} c_{ij}^{ab}\left|\Psi_{ij}^{ab}\right\rangle \ldots \tag{10}$$

The equation above can be expanded to the full CI matrix to obtain the full CI method which in essence would include all the necessary excited determinants for *n*-electron system. This matrix becomes large even for small systems and the expansion must therefore be truncated, usually done at the double excitations. More importantly, the structure of the full CI matrix provides for the couplings between different excitations as shown below where only doubles directly interact with the HF ground state.

$$\begin{array}{c} \\ \\ \langle\Psi_0| \\ \langle S| \\ \langle D| \\ \langle T| \\ \langle Q| \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \left|\Psi_i^a\right\rangle & \left|\Psi_{ij}^{ab}\right\rangle & \left|\Psi_{ijk}^{abc}\right\rangle & \left|\Psi_{abc}^{rst}\right\rangle & \left|\Psi_{abcd}^{rstu}\right\rangle & \cdots \\ \left|\Psi_0\right\rangle & \left|S\right\rangle & \left|D\right\rangle & \left|T\right\rangle & \left|Q\right\rangle & \cdots \end{array} \\
\begin{bmatrix}
\langle\Psi_0|H|\Psi_0\rangle & 0 & \langle\Psi_0|H|D\rangle & 0 & 0 & \cdots \\
 & \langle S|H|S\rangle & \langle S|H|D\rangle & \langle S|H|T\rangle & 0 & \cdots \\
 & & \langle D|H|D\rangle & \langle D|H|T\rangle & \langle D|H|D\rangle & \cdots \\
 & & & \langle T|H|T\rangle & \langle T|H|D\rangle & \cdots \\
 & & & & \langle Q|H|Q\rangle & \vdots
\end{bmatrix}
\end{array}$$

2.2.2 Møller-Plesset Perturbation Theory (MPPT)

Attempts to improve the HF wavefunction can also be achieved by applying perturbation theory which has been quite successful especially using the Raleigh-Schrödinger perturbation theory, commonly known as MPPT theory. MPPT belongs to the family of the perturbation theories which involve partitioning the total

Hamiltonian into the zeroth order (H_0) with known eigenfunctions and eigenvalues and a perturbation (V_0), commonly known as a fluctuating potential according to equation (11) for N-electron systems.

$$H\left|\Phi_i\right\rangle = (H_0 + V_0)\left|\Phi_i\right\rangle = E_i\left|\Phi_i\right\rangle \tag{11}$$

The eigenfunctions and eigenvalues of H_0 are known using the relation.

$$H_0\left|\Psi_i^{(0)}\right\rangle = E_i^{(0)}\left|\Psi_i^{(0)}\right\rangle \tag{12}$$

The trial eigenfunction $\left|\Psi_i^{(0)}\right\rangle$ and $E_i^{(0)}$ are reasonably close to the exact $\left|\Phi_i\right\rangle$ and E_i if the fluctuating potential V is small. The aim is to systematically improve the trial functions and eigenfunctions of the known Hamiltonian so that they approximate the total Hamiltonian H by introducing an ordering parameter λ and then expand the exact eigenfunctions and eigenvalues using a Taylor series in λ (eqs. (14) and (15)). The parameter is initially set to 1 to obtain

$$H = H_0 + \lambda V \tag{13}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots \tag{14}$$

$$\left|\Phi_i\right\rangle = \left|i\right\rangle + \lambda\left|\Psi_i^{(1)}\right\rangle + \lambda^2\left|\Psi_i^{(2)}\right\rangle + \ldots \tag{15}$$

Where $E_i^{(n)}$ refers to the n$^{\text{th}}$-order energy. After applying standard MP2 derivation, one obtains the second order equations in the perturbation.

$$E_i^2 = \sum_n \frac{\langle i|V|n\rangle\langle n|V|i\rangle}{E_i^{(0)} - E_n^{(0)}} = \sum_n \frac{\left|\langle i|V|n\rangle\right|^2}{E_i^{(0)} - E_n^{(0)}} \tag{16}$$

Thus, the HF energy is equivalent to the sum of both the zeroth- and the first-order corrections

$$E_{HF} = E_{MP}^{(0)} + E_{MP}^{(1)} \ . \tag{17}$$

The total Møller-Plesset energy is obtained by adding the second correction,

$$E_{MP2} = E_{MP}^{(0)} + E_{MP}^{(1)} + E_{MP}^{(2)} \tag{18}$$

The major advantage of MP2 is that it is a size consistent correction, accurate, as well as a low-cost method.[25] MP2 has therefore proven to be a reliable method to approximate ground state energies and geometries for advanced excited state methods.

2.2.3 Density Functional Theory

DFT as a theory rests upon two fundamental theorems formulated and proved by Hohenberg and Kohn.[26] Theorem one states that: *The ground-state energy from Schrödinger's equation is a unique functional of the electron density.* Meaning there is a one-to-one mapping between the ground-state electron density and the ground-state wave function. The second theorem states that: *The electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger equation.* If this theorem holds, the "true" functional form would be known, and the electron density would be varied until the energy from the functional is minimized[27] A brief mathematical background of DFT is provided.

In DFT, the electronic energy, $E_e[\rho(\boldsymbol{r})]$ within the Born-Oppenheimer approximation can be written as a functional of the electron density [28],

$$E_e[\rho(r)] = T[\rho(r)] + V_{en}[\rho(r)] + J[\rho(r)] + Q[\rho(r)] \tag{19}$$

Here; $T[\rho(r)]$ is the Kinetic energy (K.E) of the electrons, $V_{en}[\rho(r)]$ the nuclear-electron attraction energy, $J[\rho(r)]$ the classical electron-electron repulsion energy, and $Q[\rho(r)]$ is the quantum (non-classical) electron-electron interaction energy. The aim of DFT is to develop accurate functionals that approximate $T[\rho(r)]$ and $Q[\rho(r)]$. Terms two and three of equation (19) can be calculated using equations (20) and (21) .

$$V_{en}[\rho(r)] = -\sum_{A=1}^{M} \int \frac{Z_A}{|r-R_A|} \rho(r) dr \tag{20}$$

$$J[\rho(r)] = \tfrac{1}{2}\int\int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1\, dr_2 \tag{21}$$

An accurate approximation of the kinetic energy functional is therefore important since the K.E contribution is the largest unknown term. The Thomas-Fermi model which is known to be exact for a finite uniform gas is the simplest approximation to the K.E but suffers a disadvantage of being inadequate for chemical bond descriptions.

Kohn and Sham proposed an accurate approximation of the K.E to circumnavigate the Thomas-Fermi model using Slater determinants (orbitals $\{\phi_i\}$) which describe fictitious system of non-interacting electrons with the same density as the exact electronic wave function. This is commonly known as Kohn Sham Density functional theory (KS-DFT) with the original flavor known as orbital free DFT. KS DFT is given by the equation

$$T_s[\{\phi_i\}] = -\tfrac{1}{2}\sum_{i=1}^{n}\int \phi_i^*(r)\nabla^2\phi_i(r)dr \tag{22}$$

The difference between the non-interacting K.E and the exact Thomas-Fermi model plus $Q[\rho(r)]$ defines what is commonly referred to as exchange-correlation energy, $E_{xc}[\rho(r)]$,

$$E_{xc}[\rho(r)] = T[\rho(r)] - T_s[\{\phi_i\}] + Q[\rho(r)] \tag{23}$$

Therefore, the unknown term in DFT is the E_{xc} commonly written as individual sums of E_x and E_c. Functionals which approximate the XC have been developed over the years and take the form of the so-called "Jacob's" Ladder hierarchy of John Perdew[29] which begins with Hartree Fock all the way up to the double hybrid functionals with the aim of achieving chemical accuracy. In this work, effort has been taken to estimate chemical properties of embelin using a number of hybrid GGA exchange functionals (B88[30] , PBE [31]) and GGA correlation functionals (LYP[32] , PBE[31]) as well as the range-separate functional CAM-B3LYP.[33] CAM-B3LYP has been extensively used because it provides comparable results with the state-of-the-art ADC method as will be seen in the next chapter. The reader is referred to an excellent publication by

Mardirossian and Head-Gordon[28] and Sholl & Steckel[27] for the mathematical rigor of the various functionals and their chemical accuracy.

2.2.4 Time-Dependent Density Functional Theory (TD-DFT)

Just like in the ground state DFT, one-to-one correspondence of the real and fictitious systems is obtained in the time-dependent version. The equations of motion were proved by Runge and Gross.[34] The underlying theorem treats two densities $n(r,t)$ and $n'(r,t)$ evolving from an initial state $\psi(t=0)$. Both densities are under the influence of two potentials which are Taylor expandable at $t=0$, $v_{ext}(r,t)$ and $v'_{ext}(r,t)$ should differ if their potential differs by more than a purely time-dependent function (which is r-independent)

$$\Delta v_{ext}(r,t) = v_{ext}(r,t) - v'_{ext}(r,t) \neq c(t) \tag{24}$$

One to one mapping between densities and potentials can be obtained under these conditions implying that the potential is a *functional* of the density. One only needs to prove that the current densities must differ over time. An important factor to consider is that the density determines the potential up to a time-dependent constant. Therefore, a time-dependent phase is used to determine the wavefunction. This ensures that the expectation value of any operator (which is canceled by the phase) is a functional of the time-dependent density and the initial state.

Most computations are carried out using KS equations which boils down to a time-dependent Kohn-Sham equation,

$$i\frac{\partial \varphi_j(r,t)}{\partial t} = \left[-\frac{\nabla^2}{2} + v_{KS}[n](r,t)\right]\varphi_j(r,t) \tag{25}$$

Where, $n(r,t) = \sum_{j=1}^{N}\left|\varphi_j(r,t)\right|^2$ is the density of the real system. The potential $v_{KS}(r,t)$ yields a unique density by virtue of one-to-one mapping, hence an exchange-correlation potential can be written as

$$v_{KS}(r,t) = v_{ext}(r,t) + v_H(r,t) + v_{xc}(r,t)\ . \tag{26}$$

This function is used to perform all time-dependent related calculations as it contains the history of the density, $n(r,t)$, the initial KS wavefunction, $\phi(0)$ and the wavefunction $\psi(0)$.

2.2.5 Algebraic Diagrammatic Construction Scheme (ADC)

The ADC method offers efficient and size-consistent means to calculate excited state properties of molecules. It falls within methods derived in the language of Green's function and propagator formulation.[35] The advantage of propagators is that they allow one to directly access basic excitation processes of a system such as:

i. Polarization propagator $\Pi(\omega)$ gives access to electronic excitation
ii. The one-particle Green's function $G(\omega)$ enables single electron removal or attachment.
iii. The two-particle (hole) propagator $P(\omega)$ enables double ionization or attachment.

A great effort has been taken by early researchers to derive ADC equations and the reader is referred to the works of Schirmer, Dreuw, and co-workers.[35–40] A major advantage of the ADC as a method is its ability to correctly scale with system size, the so-called size consistency. This has seen immense applications in several areas of research including spectroscopy [41], PCM models [42], as well as in biomolecular systems.[43]

Within the Green's function theory framework, the polarization propagator, whose spectral representation describes the electronic excitations consists of two parts written as:

$$\Pi_{pq,rs}(\omega) = \Pi^{+}_{pq,rs}(\omega) + \Pi^{-}_{pq,rs}(\omega) \tag{27}$$

The relationship of the representation is such that, $\Pi^{-}_{pq,rs}(\omega) = \Pi^{+}_{pq,rs}(-\omega)$. We can expand the Π^{+} such that

$$\Pi^{+}_{pq,rs}(\omega) = f^{+}_{pq}(\omega 1 - M + \imath\eta)^{-1} f_{rs} \tag{28}$$

With **M** being the energy shifted Hamiltonian matrix containing the matrix elements

$$M_{IJ} = \left\langle \phi_I \middle| \hat{H} - E_0 \middle| \phi_J \right\rangle \tag{29}$$

Where ϕ_I and ϕ_J are the intermediate states. The spectral amplitude vectors are obtained from f_{pq} and are given by

$$f_{pq,I} = \left\langle \phi_I \middle| \hat{c}_p^\dagger c_q \middle| \Psi_0 \right\rangle \tag{30}$$

Where E_0 and $\left|\Psi_0\right\rangle$ are the ground state energy and wave function, respectively.

The more commonly used intermediate state representation (ISR)[35] entails application of excitation operators for particle-hole (p-h) configurations, 2p-2h configurations and so on to the MP ground state.

$$\hat{C}_I \in \{\hat{c}_a^\dagger \hat{c}_i, \hat{c}_a^\dagger \hat{c}_b^\dagger \hat{c}_i \hat{c}_j, a < b, i < j; \ldots\}$$

A set of the resulting intermediate states (IS) wave functions are subjected to the Gram-Schmidt orthogonalization and expanded in the basis of the intermediate states, $\{\left|\tilde{\Psi}_J\right\rangle\}$.

$$\left|\Psi_n\right\rangle = \sum_J X_{nj} \left|\tilde{\Psi}_j\right\rangle \tag{31}$$

The resulting equations can be inserted in equations (29) and (30) to yield

$$M^{[n]} = \sum_{i=0}^{n} M^{(i)} \text{ and } f_{pq}^{[n]} = \sum_{i=0}^{n} f_{pq}^{(i)} \tag{32}$$

Where (i) refers to the *ith* order contributions. This leads to a perturbation expansion of the ADC matrix where MP2 and MP3 yields ADC(2) and ADC(3), respectively. Clearly, intermediate states lead to CI-like blocks (see Figure 2) of singly (p-h), doubly (2p-2h) and higher excited configurations. Aside from the fact that ADC is size consistent, the fact that it includes doubles makes it a suitable choice as a correlated method for different chemical phenomena such as two photon absorption and has been used extensively in this work. The matrix $M^{[n]}$ gives the $n-th$ order

approximations to the excited states and excitation energies via solving the Hermitian eigenvalue problem

$$M^{[n]}X = X\Omega, \quad X^{\dagger}X = 1 \ . \tag{33}$$

Where $\Omega_I = E_I^{[n]} - E_0^{[n]}$ yields the excitation energies (eigenvalues) and X_I refers to the excited states (eigenvectors).

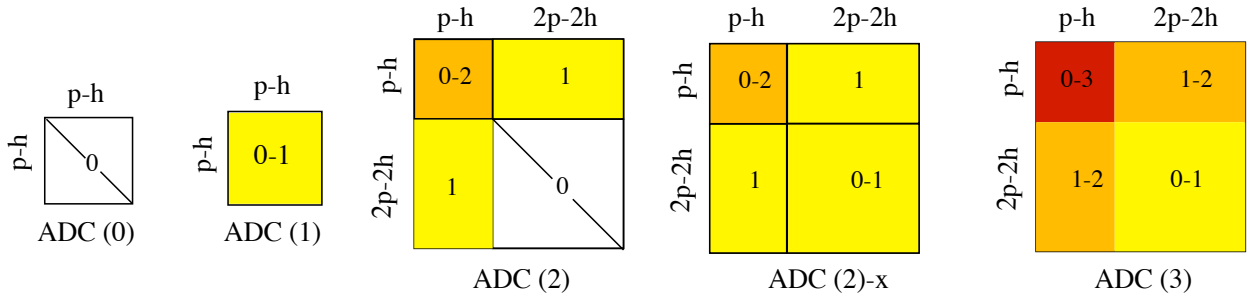


Figure 2: Block structure of the ADC secular matrix. The zeroth- and first-order contain the p-h block while the second- and third-order has [p-h, 2p-2h],[2p-2h,p-h] and [2p-2h,2p-2h] blocks. The numbers within the box indicate the orders of terms to be considered in the perturbation expansions of the matrix elements.

2.3 Molecular Dynamics (MD)

In the following, a brief derivation of the MD equation will be provided using the Hamiltonian formulation. This formulation describes a system in terms of generalized coordinates (q_k) and generalized momenta (p_k). The changing of basis from the set (q_k, q_k, t) to (q_k, p_k, t) is obtained via a Legendre transformation, defined by the function:

$$H(q,p,t) = \sum_k p_k \dot{q}_k - L(q,\dot{q},t). \tag{34}$$

Taking the differential on the LHS,

$$dH = \frac{\partial H}{\partial t} dt + \sum_k \frac{\partial H}{\partial q_k} dq_k + \sum_k \frac{\partial H}{\partial p_k} dp_k \tag{35}$$

Or from the RHS,

$$dH = \sum_k p_k d\dot{q}_k + \sum_k q_k dp_k - \sum_k \frac{\partial L}{\partial q_k} dq_k - \sum_k \frac{\partial L}{\partial \dot{q}_k} d\dot{q}_k - \frac{\partial L}{\partial t} dt \tag{36}$$

And designating the canonical momenta to be

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \tag{37}$$

The first and the fourth summation cancel, and the remaining terms can be identified with the corresponding terms, i.e.

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \tag{38}$$

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \tag{39}$$

$$\dot{p}_k = -\frac{\partial H}{\partial q_k} \tag{40}$$

Therefore, the function $H(q,p,t)$ is the Hamiltonian and equations (39) and (40) are the Hamiltonian equation of motion (EOM).

2.3.1 Mathematical Form of a Trajectory

EOM describes the unique evolution of the coordinates and momenta subject to a set of initial coordinates. Equations (39) and (40) specify a trajectory

$$x(t) \equiv (q_1(t),...,q_F(t), p_1(t),..., p_F(t)) . \tag{41}$$

In phase-space, starting from an initial point $x(0)$. The EOM can be expressed using a matrix or *symplectic* notation. The time derivative of the phase space vector $x(t)$ can be written as

$$\dot{x}(t) = \left(\frac{\partial H}{\partial p_1},...,\frac{\partial H}{\partial p_F},-\frac{\partial H}{\partial q_1},...,-\frac{\partial H}{\partial q_F} \right) \tag{42}$$

Hence the Hamiltonian's EOM can be recast as

$$\dot{x} = M \frac{\partial H}{\partial x} \tag{43}$$

Where **M** is a matrix expressible in block form as

$$M = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{44}$$

Where 0 and I are the $F \times F$ zero and identity matrices, respectively. Here, F stands for *3N* dimensions. The time-dependence of a property $A(t)$, that is a function of phase space $A(t)=A(x(t))$, is formally given by

$$\frac{d}{dt} A(x(t)) = \sum_{\alpha=1}^{F} \left[\frac{\partial A}{\partial q_\alpha} \dot{q}_\alpha + \frac{\partial A}{\partial p_\alpha} \dot{p}_\alpha \right] \tag{45}$$

$$= \sum_{\alpha=1}^{F} \left[\frac{\partial A}{\partial q_\alpha} \frac{\partial H}{\partial p_\alpha} - \frac{\partial A}{\partial p_\alpha} \frac{\partial H}{\partial q_\alpha} \right] = \{A, H\} \tag{46}$$

Where the *Poisson bracket* notation has been introduced

$$\{A, B\} \equiv \sum_{\alpha}^{E} \left[\frac{\partial A}{\partial q_\alpha} \frac{\partial B}{\partial p_\alpha} - \frac{\partial B}{\partial q_\alpha} \frac{\partial A}{\partial p_\alpha} \right] \tag{47}$$

Which defines the *Liouville operator iℒ*

$$i\mathcal{L}\ldots\equiv\{\ldots,H\} \tag{48}$$

where $i=\sqrt{-1}$. This equation can be written explicitly as a differential operator

$$i\mathcal{L}=\sum_{\alpha=1}^{F}\left[\dot{q}_\alpha\frac{\partial}{\partial q_\alpha}+\dot{p}_\alpha\frac{\partial}{\partial p_\alpha}\right] \tag{49}$$

The time-dependence of an arbitrary phase-space function *A(x(t))* is given by

$$\frac{d}{dt}A=i\mathcal{L}A \tag{50}$$

With a formal solution of

$$A(t)=e^{i\mathcal{L}t}A(0) \tag{51}$$

The operator $e^{i\mathcal{L}t}$ is known as the classical propagator and by the introduction of the imaginary unit i into the operator makes it resemble the quantum propagator $exp(-i\hat{H}/\hbar)$. The time evolution of the phase-space vector x(t) can be written as

$$x(t)=e^{iLt}x(0) \tag{52}$$

This equation describes the central numerical problem in MD simulations to obtain the time-dependent trajectory $x(t)$ in phase-space given an initial condition $x(0)$.

2.4 Quantum Mechanics/Molecular Mechanics (QM/MM)

Hybrid methods generally involve treating one part of a large complex using high level *ab initio* method while the rest of the molecule is treated using a cheaper molecular mechanics, commonly, parametrized force fields. There exists a plethora of methods that have been implemented to address the technicalities and evolution of the hybrid QM/MM partition. The key difference lies in the type of physicochemical effects used to describe the interactions between the QM and MM environments. Known effects include polarization, electrostatics, and repulsion. Among the widely used QM implementation that includes polarization is the so-called continuum solvation model (CSM) which consist of various flavors such as the conductor-like

screening model (COSMO)[44] and the polarizable continuum model (PCM).[45–47] The disadvantage of this approach is the lack of sampling which is solved by many advanced hybrid QM/MM implementations. In these approaches, effort is taken to include at least one or all of the effects in a sampled ensemble. A brief overview of two commonly used QM/MM implementations namely, electrostatic and the polarizable embedding schemes is provided.

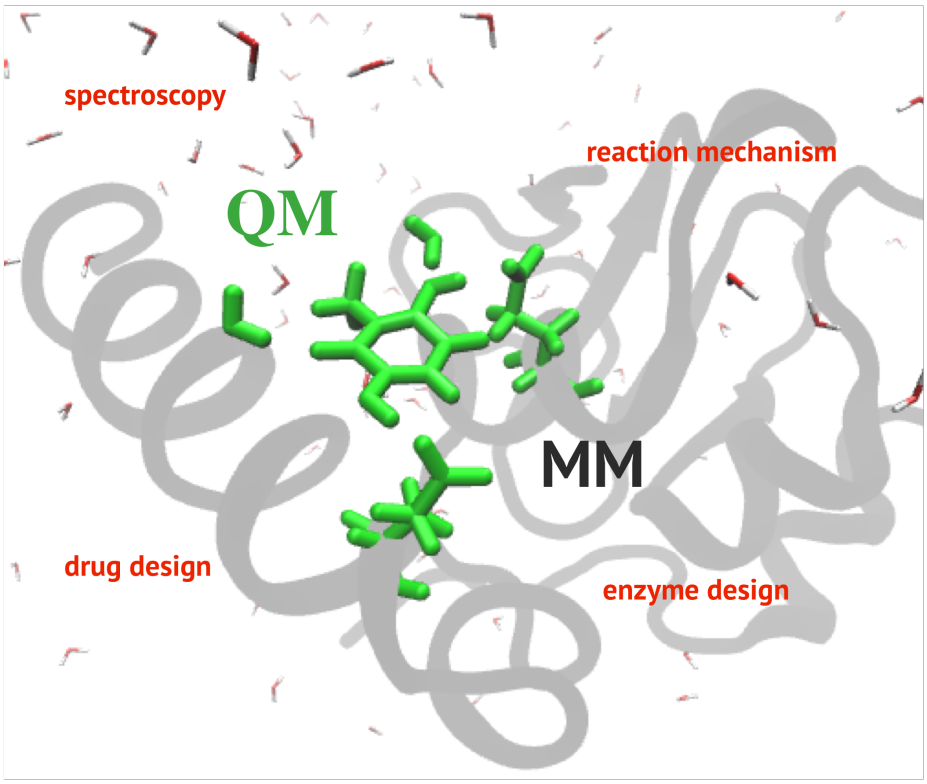


Figure 3: Pictorial representation of a typical hybrid QM/MM complex system. The green part of the system is the QM region while the grey and water molecules constitute of the MM region parametrized by force fields.

2.4.1 Electrostatic Embedding

Electrostatic embedding is among the most common approaches used in treating large systems. The non-polarizable general equation form corresponds to parametrized Coulombic interactions of the electron charge density with point charges from the MM region and the nuclei of the QM nuclei. Additional non-electrostatic contribution analogous to Lennard-Jones potential effective between atoms in the MM region , $E_{\mathrm{LJ}}^{\mathrm{QM-MM}}$ are included in the equation:

$$E_{QM-MM} = \sum_{A\in QM}\sum_{I\in MM}\frac{Z_I Z_A}{|R_A - R_I|} + \sum_{I\in MM} Z_I \int \frac{\rho(r)}{|R_A - R_I|} + E_{LJ}^{QM-MM}(|R_A - R_I|) \quad (53)$$

A and *I* indices run over the atoms in the QM and MM regions, respectively. The equation above can be implemented using Gaussian basis functions. In this case, the partial MM charges are explicitly included within the QM Hamiltonian enabling polarization of the electron density arising from the electrostatic field generated by the environment. In this work, this type of embedding approach has been utilized since it is commonly applied in spectroscopy studies.[48]

2.4.2 Polarizable Embedding (PE)

PE is a QM/MM method that aims to solve the problems faced by electrostatic embedding highlighted above, namely, the mutual polarization of the core region and the environment. Quantum mechanical calculations are usually used to derive advanced force field representations of the environment. Further, the electrostatic potential is represented by assigning multipole expansion to each molecule in the system. Dipole-dipole polarizability tensors are further placed on all expanded centers thus allowing polarization of the electrostatic embedding potential.[49] The main advantage of this approach is that it ensures optimization of the core region's ground state electron density and at the same time taking into consideration of the explicit environment electrostatic interactions. A brief theory behind PE is highlighted. For a complete derivation, see references.[49,50]

Within PE, an effective Hamiltonian is constructed at the HF/DFT methods using the Fock/Kohn-sham operator

$$\hat{f}^{eff} = \hat{f}^{vac} + \hat{v}^{PE} \quad (54)$$

Where $\hat{f}^{vac}$ is the vacuum HF/DFT operator and $\hat{v}^{PE}$ the operator describing the environmental potential experienced by the core region. Within this operator, we have

$$\hat{\mathrm{v}}^{PE} = \hat{v}^{es} + \hat{\mathrm{v}}^{ind} \quad (55)$$

$\hat{V}^{es}$ is the electrostatic operator describing the environmental fragments' permanent charge distributions (nuclei and electron densities) on the QM region and can be expanded as

$$\hat{v}^{es} = \sum_{s=1}^{N} \sum_{|k|=0}^{K} \frac{(-1)^{|k|}}{K!} M_s^{(k)} \hat{V}_{s,el}^{(k)} \tag{56}$$

Where, $M_s^{(k)}$ are the multipole expansion sites s , $\hat{V}_{s,el}^{(k)}$ operator gives derivative of the electric potential at site s. This equation can be expanded up to the second order

$$\hat{v}^{es} = \sum_{s=1}^{N} \left(q_s \hat{V}_{s,el} - \sum_{\alpha} \mu_s^{\alpha} \hat{V}_{s,el}^{\alpha} + \sum_{\alpha,\beta} \Theta_s^{\alpha\beta} \hat{V}_{s,el}^{\alpha\beta} \right) \tag{57}$$

Where cartesian coordinates are represented by α and β , charge as q , dipoles as μ, and quadrupoles as Θ. The induction due to the polarization of the environment on the QM region from the polarized charge distributions is given by

$$\hat{v}^{ind} = -\sum_{s=1}^{N} \mu_s^{ind}(F_{tot}) \hat{F}_{s,el} \tag{58}$$

μ_s^{ind} is the induced dipole set up by the total electric field at each polarizable site s, whereas F_{tot} consist of the permanent multipole moments and the induced dipoles of the environment and the electrons and nuclei in the core region. $\hat{F}_{s,el}$ yields the electronic electric field at sites a. The induced dipoles are obtained first to derive $\hat{v}^{ind}$:

$$\mu_s^{ind}(F_{tot}) = \alpha_s F_{tot}(R_s) = \alpha_s \left(F(R_s) + \sum_{s' \neq s} \mathrm{T}_{ss}^{(2)} \mu_{s'}^{ind} \right) \tag{59}$$

Where $F(R_s)$ represents the electric field at site s from electrons, permanent multipole moment and nuclei. $T_{ss}^{(2)}$ is the dipole-dipole interaction tensor. We can introduce a column vector containing the induced dipoles $\mu^{ind} = (\mu_1^{ind}, \mu_2^{ind}, ..., \mu_1^{ind})^T$

and $F = (F(R_1), F(R_2), ..., F(R_N))^T$. The induced moments can then be solved as a matrix-vector equation

$$\mu^{ind} = BF \tag{60}$$

Where B is a $(3S \times 3S)$ - dimensional response matrix which connects a set of induced dipoles and the electric fields.

$$B = \begin{pmatrix} \alpha_1^{-1} & \mathrm{T}_{12}^{(2)} & \cdots & \mathrm{T}_{1S}^{(2)} \\ \mathrm{T}_{21}^{(2)} & \alpha_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathrm{T}_{(S-1)S}^{(2)} \\ \mathrm{T}_{S1}^{(2)} & \cdots & \mathrm{T}_{S(S-1)}^{(2)} & \alpha_S^{-1} \end{pmatrix}^{-1} \tag{61}$$

Chapter - 3

Embelin as a Potent Photosensitizer

The results presented in this section have been published in a peer-reviewed journal titled:

Michael Rogo Opata and ***Andreas Dreuw***: Embelin's Versatile Photochemistry Makes it a Potent Photosensitizer for Photodynamic Therapy. *J. Phys. Chem. B 2021, 125, 3527-3537*

3.1 Abstract

The search for desirable and effective photosensitizers (PS) is a subject of continued research. Apart from the widely used porphyrin scaffold, the quinone scaffolds pose as alternative photosensitizers due to their simple but unique structure and their presence in various naturally occurring plants. Embelin is one such quinone commonly isolated from various *Embelia* genus and has shown promising biological activity in preliminary *in-vitro* studies. Most of the reported isolation has been from the *Embelia ribes* species. Numerous studies have been conducted on its prospective use in pharmacology with recent studies suggesting its potential application in photodynamic therapy. Despite existence of several literature on embelin, there lacks a mechanistic understanding on its photochemistry and subsequent exploration on its potential use as a PS. This section seeks to address this knowledge gap in understanding embelin's photo mechanism. The approach taken is a theoretical one with the results compared to existing experimental data. Using state-of-the-art quantum methods, interesting photochemistry that comprise of embelin's ability to transfer proton intramolecularly, generate singlet oxygen and undergo oxidation reactions is hereby presented. Its photochemistry has been studied using ground and excited state methodologies which encompass investigation of both its static and

dynamic behavior. Particularly, the mechanisms through which embelin can transfer a proton and generate ROS is dealt with in greater depth.

The implication of such an exploration is to have an understanding of the molecular behavior of embelin upon photon absorption. This is particularly important as a building block to decipher possible mechanisms through which embelin inhibit XIAP proteins. Such efforts are pursued given the fact that not only can an excited state proton transfer impair a protein's function but can also mediate critical processes such as singlet oxygen generation which is a key requirement of PDT as a treatment method. Therefore, pure quantum mechanical study of embelin act as a starting point and a building block for advanced explorations such as hybrid QM/MM study. Results obtained in this chapter will be heavily referred to in terms of comparison and consequent discussions of the proceeding chapters.

3.2 Introduction

Natural products also known as phytomedicines play a critical role in today's medicinal folklore. Due to their rich naturally occurring scaffolds which can be synthetically modified to improve potency, several classes have been used over years. However, one class of the natural products that remains overlooked and unexplored are the so-called photoactive phytomedicines.[16] These are molecules with promising scaffolds that are isolated from plants and can absorb photons within PDT wavelengths hence qualify as efficient alternative PS, a key ingredient for practical medical application. Some of the phytomedicine scaffolds include furanocoumarins, curcumins, alkaloids (such as benzylisoquinolines, chinolin-alkaloids, pterins, indigo and beta-carbolines), polyacetylenes, xanthenoids, thiophenes, porphyrins, phenalenones, anthraquinones and perylenequinones[16] shown in Figure 4 together with their absorption wavelengths.

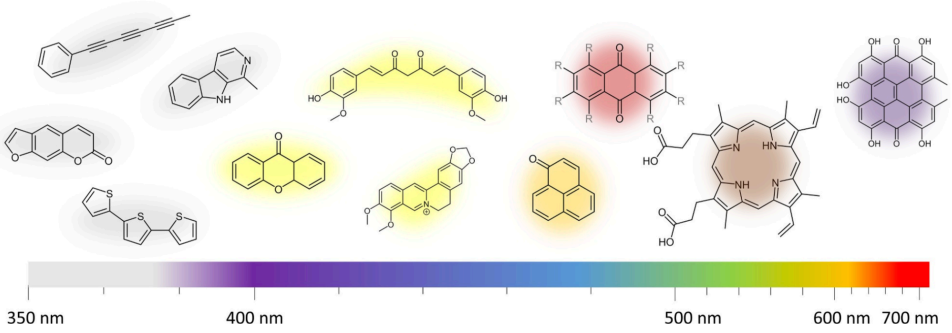


Figure 4: Natural PS scaffolds with their respective absorption max wavelengths. Adopted from Siewer and Stuppner 2019.

The photo pharmaceutical application of these scaffolds varies from treatment of skin-diseases, viruses, filamentous fungi, and cancer[51–53] among others. Therefore, utility of various scaffolds in PDT is viable even though there exist a number of controversial issues regarding some scaffolds' mechanism of action (see references[54,55] as an example for the case of Mangostin). In this section, extensive explanation on the photochemistry of embelin which belong to the quinone scaffold is provided.

Embelin (2,5-dihydroxy-3-undecylcyclohexa-2,5-diene-1,4-dione) (Figure 5) is a benzoquinone isolated from the fruits of *Embelia ribes* and *schimperi.*[18,56] It has been previously shown to have a broad biological activity spectrum comprising anti-inflammatory, analgesic as well as anticancer effects. A number of studies have

specifically focused on its ability to act as anti-tumor through inhibition of XIAP.[57–59] These studies have been conducted using traditional ligand-protein binding experimental approaches which in most cases does not utilize light sources. Recently, research has shifted focus into using PDT approach for cancer treatment [1,60] necessitating the search for more efficient PS other than porphyrins. PDT approach involves the use of PS that can selectively accumulate in malignant tissues and cause photodamage on absorbing a photon of specific wavelength, normally around 600 nm. This method has clear advantages over traditional cancer treatment approaches in that it is spatially localized hence reduces the extent of tissue damage and side effects normally experienced with chemotherapy for example.

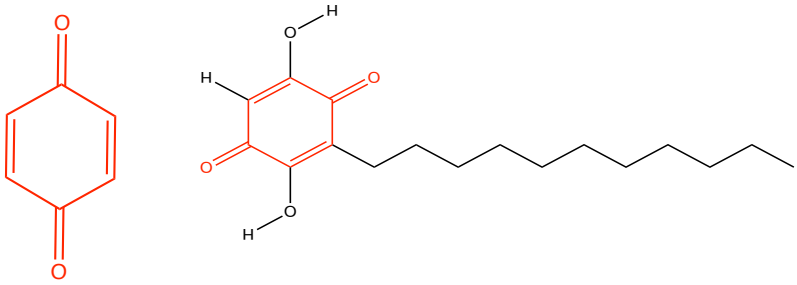


Figure 5 :Benzoquinone scaffold to which embelin belong and the molecular structure of embelin. The hydrogen bonds between the hydroxyl and quinone moieties are thought to play a critical biological activity role via resonance.

In PDT, there are two types of reactions between the PS and the respective biomolecules. Reaction type I involves the generation of reactive oxygen species (essentially singlet oxygen) which are known to be aggressive towards tumor cells. Biological responses towards the PS in this type of reaction are localized within a particular area hence are deemed safer [1]. Type II reactions involves the PS reacting with the biomolecule directly to produce peroxides, superoxide ions, and hydroxyl radicals. In addition, some PS are known to take part in different reactions (this however occurs to a lesser extent) such as in the photoaddition to DNA.[61] Nevertheless, a potent PS should possess several modes of action to produce a harming effect to the tumor cell via a combination of different mechanisms. This ensures that either of the aforementioned type of reactions can be achieved on absorption of a photon hence a broad photo biological activity spectrum.

The most common PDT reactions (Type I) relies on energy transfer between the PS and the quenching molecule (usually molecular oxygen) via a process known as

triplet-triplet energy transfer (TTET) to generate a singlet. TTET can be considered as two simultaneous electron transfer processes with different spin, i.e. ($\alpha \rightarrow \alpha, \beta \rightarrow \beta$) [62] and is always considered as the reverse mechanism reaction of Singlet Fission (SF).[63] This is because in TTET, the dominating couplings always arise from the exchange integral as opposed to Dexter exchange in SF.[62] The equations governing TTET are a two-step process as shown below.

$$Mol \xrightarrow{hv} {}^{1}Mol^{*} \xrightarrow{isc} {}^{3}Mol^{*} \tag{62}$$

$${}^{3}Mol^{*} + {}^{3}O_{2} \xrightarrow{SE} Mol + {}^{1}O_{2}{}^{*} \rightarrow oxidativedamage \tag{63}$$

Where hv, isc, and SE refers to the energy of the absorbed photon, intersystem crossing and spin exchange respectively.

The core challenge with embelin despite its promising potential is its poor solubility in water that reduces its oral bioavailability thereby hindering its application in clinical trials. This limitation has seen a number of synthetic structural modifications to enhance its bioavailability[64–66] especially in cancer treatment. This work seeks to investigate ways to circumnavigate this problem by using embelin as a PDT agent. In this way, we solve two problems. First, the oral bioavailability limitation is avoided since intravenous (IV) administration methods can be used within a localized region. Secondly, embelin's potency is enhanced using a light source to generate ROS which has been shown to improve its cytotoxicity *in vitro*.[22] An understanding of the basic molecular mechanism underlying embelin's photochemistry on absorption of a photon in gas phase or within an environment is therefore important. The data could be used to design more potent PS and further optimize photon absorption at certain wavelengths via structural modifications of its existing scaffold, hence performing specific desired reactions.

Structurally, embelin has an interesting scaffold consisting of two unique intramolecular hydrogen bonds between its quinone and hydroquinone functional groups. This structural arrangement has been previously documented to trigger its biological activity action due to the possibility of donating or accepting hydrogen which enhances resonance.[67] Within a protein environment where there exist acidic and basic functional groups, such structural arrangement could be desirable since

most reactions rely on donor/acceptor pairs. An additional advantage to embelin's motif is the fact that other carbonyl containing compounds with the ability to act as PS and generate singlet oxygen have been reported in literature.[68–70] The various possible mechanisms indicate that embelin can react via different modes of action as a PS. Its ability as a PS was first reported in *in vitro* studies targeting malignant tissues by Joy.[22] In the study, cytotoxicity of embelin was enhanced upon irradiation suggesting that embelin has synergy effects with light sources. The increased reactivity has been suggested to originate from singlet oxygen generation resulting from photo "activation" of molecular oxygen. In another study, embelin was found to play a critical role in photochemical generation of silver and gold nanostructures for photothermal therapy.[3] The researchers found embelin to act as a stabilizing and reducing agent for silver nanoparticles. These experimental evidences point towards embelin's important photochemical role in certain reactions.

Although a number of studies have outlined embelin's biological potential, details of its photo mechanism remains poorly understood. The knowledge gap could be tackled on different fronts; experimental and theoretical. A plethora of experimental data exist in literature as explained above and a theoretically attempt to characterize embelin's reactions in the ground state level has been reported in literature.[71] However, a theoretical investigation on its photochemistry is lacking. This work seeks to tackle embelin's photochemistry using advanced methods implemented with the correct physics to capture different chemical and photo chemical phenomena of embelin. The ground- and -excited state properties as well as dynamically photoinduced processes are reported in this work. For the dynamic component, the focus lies on the investigation of excited state intramolecular proton transfer (ESIPT) pathways. Proton transfer reactions are in line with the fact that resonance is reported to play a critical role in embelin's biological mechanism. More relevant to PDT applications is embelin's ability to generate singlet oxygen, a commonly known type I reaction mechanism.

3.3 Computational Methods

Functional groups play a critical role in chemistry particularly in photochemistry. Embelin contain a long undecyl chain that play an insignificant photochemical role. Thus, a model of the molecule was created (still referred to as embelin herein) by replacing the undecyl group with a methyl group (Figure 6 **A**). This part of the molecule may however play an important role within a proteins' or enzymes' pocket. Additionally, it could dictate the lipophilicity of embelin which is important for pharmacological applications such as drug delivery to intended target. The system numbering used in this chapter is as depicted in Figure 6 **A**.

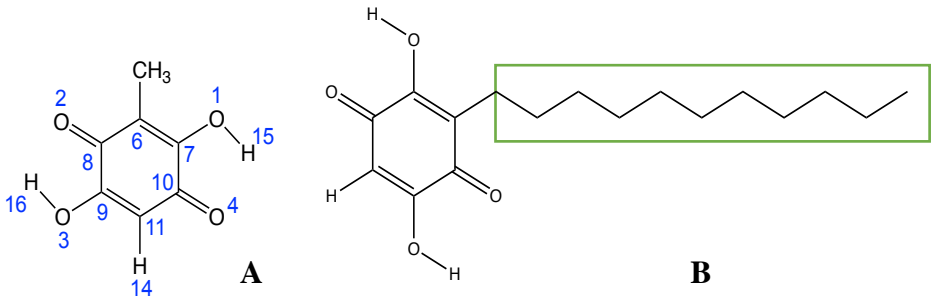


*Figure 6: Molecular structure of embelin used as a model for theoretical calculations (**A**), Full embelin structure depicting truncated undecyl chain (**B**).*

All calculations in this chapter unless stated otherwise were carried out using Dunning's cc-PVTZ basis set.[72] The Algebraic diagrammatic construction (ADC) scheme for the polarization propagator at the second (ADC(2))[35,40] and third order of perturbation theory (ADC(3))[38,39] together with linear-response time-dependent density functional theory (TD-DFT)[28,34] was used for calculating the relevant energetically low-lying excited electronic states. The exchange-correlation (xc) functionals, Becke88[30] and Lee, Yang, and Parr (BLYP)[32] functional, the static hybrid DFT functionals B3LYP[48], the half-and-half functional BHHLYP[73], and the long-range-corrected functional CAM-B3LYP [33]. The choice of these functionals was motivated by order of the nonlocal orbital exchange for the DFT methods i.e., BLYP < B3LYP < BHHLYP < CAM-B3LYP which allows for internal consistency checks within TD-DFT values obtained. The corresponding ground state geometries for the excited states were calculated using ground state DFT[74] and Møller-Plesset perturbation theory of the second order (MP2)[75] for ADC and TD-DFT respectively

as implemented in Q-Chem 5.1. Advanced exciton analysis was carried out using the *libwfa* package[76] as implemented in Q-Chem. The natural bond orbital (NBO) package version 6[77] was used for the natural bond orbital analysis as a stand-alone software.

Geometry optimizations and excitation energies were also calculated including implicit solvent effects using polarizable continuum model (PCM). The integral equation formalism (IEF)[78] model implemented in Gaussian 16[79] was used. The solvents used are acetonitrile (ACN $\epsilon = 35.688$), DiMethylSulfoxide (DMSO $\epsilon = 46.826$), and methanol (MeOH $\epsilon = 32.613$). Up to four explicit water solvents were also added around embelin to take into account the influence of intermolecular bonds on its excitation energies. The structures were reoptimized while including the bulk PCM effects. The scientific question sought to be answered with these models is the effect of both intermolecular and intramolecular hydrogen bonding on the optical properties of embelin since it can form up to four bonds with water.

A dynamic study was further carried out to gain insights into possible photoinitiated and thermal proton-transfer processes of embelin. This was achieved by carrying out potential energy surfaces (PES) scans. Both ground- and excited-state relaxed scans along embelin's two major intramolecular hydrogen bonds as reaction coordinates moving hydrogens H16 towards O2 and H15 towards O4. The motivation behind the dynamic study is the fact that this proton resonance has been previously reported to be a contributing feature to embelin's biological activity.[67] The fundamental question significant to PDT is whether embelin prefers to transfer one or both protons on photo excitation and ability to generate oxygen. TD-DFT//CAM-B3LYP method was used for this purpose as implemented in Gaussian 16 and Q-chem 5.1. For all relaxed scans, Pople's 6-31G*[80] basis set was used. Subsequent frequency calculations were carried out to confirm for the nonexistence of transition states and hence geometrical minima. The obtained PESs suggested that both ground and the first excited state PES facilitated interesting pathway back to the ground state. A conical intersections (CIs) search between the first excited state and the ground state was therefore carried out and the identified CIs optimized using the spin-flip TD-DFT/CAM-B3LYP//6-31G* model. The branching-plane updating method[81] as implemented in Q-chem 5.1[82] was

used for this purpose. A mapping of the energies on these surfaces was also done to ascertain that embelin can possibly generate singlet oxygen.

3.4 Results and Discussion

3.4.1 Ground- and Excited-State Properties of Embelin.

3.4.1.1 Electronic Resonance in the Ground State

As highlighted above, the unique structure of embelin enables existence of electronic resonance as a result of its *para-hydroxyl* and *para-quinonic* moieties. Early reports have indicated that this feature is responsible for its antioxidant activity due to the ease of electron donation as a quinone.[67] Bond properties and bond orders would enable this resonance feature characterization on the structure of embelin. A detailed analysis was therefore carried out to gain further insights into the relevance of electronic resonance using NBO theory on the ground-state density. The optimized geometry applied for this purpose was calculated at the DFT/CAM-B3LYP level of theory. Within the NBO analysis, the best single "natural" Lewis structure (NLS) representation of the wavefunction or density can be identified hence simplifying complicated values to an interpretable chemical picture. Mathematical correction to the resulting NLS is then associated with weakly occupied non-Lewis-type "acceptor" NBOs, such as valence antibond of $\sigma^* - or\ \pi^* -$ type, which are formally vacant in the elementary Lewis structure picture. The theory can thus be easily connected to the usual "resonance theory" which reveal electronic delocalization effects within the structure. A second order perturbation theory is hereby used to quantify the resonance stabilization which is provided by each donor-acceptor delocalization in the orbitals.

The analysis reveals a strong π-type intramolecular stabilization between the nonbonding electrons of oxygen one (O1) and the π^* antibond between C6 and C7(π^*_{C6-C7}) with an energy of 52.48 kcal/mol. A similar stabilization is observed between O3 and the π^* antibond between C9 and C11 π^*_{c9-c11} with an energy of 56.42kcal/mol. There are also other weak stabilizations between the nonbonding orbitals of O4 and O2 and the σ^* antibond of H15 and O1 (an energy of 7.12 kcal/mol) as well as σ^* antibond between H16 and O3 (energy of 6.86 kcal/mol). The energies signify a

nearly equivalent intramolecular hydrogen bond pattern in the molecule. The bond orders of 1.5 for bonds O1-C7 and O3-C9 also allude to the fact that there is resonance in the ground state for this molecule. Natural resonance theory (NRT) is a component of perturbation theory of NBO analysis that was further used to generate resonance weights of the top contributing structures. Pictorial depictions of the possible resonance structures are generated from the density and their average weights presented. As can be seen from Figure 7 the top dominating figures generated by NRT corresponds to the values above. Contributions of the remaining structures were insignificant hence are not shown. It is evident from the structures that embelin is stable in the ground state as it is shown to possess intramolecular hydrogen bonds on both sides. Naturally, this has however a low weighting percentage as the molecule prefers to exist in a "non-resonance sate" (40%). Combining the two nearly equal hydrogen bonds (at ~8.7 %) lead to two intramolecular bonds within the molecule signifying a pronounced electronic resonance between the hydroxyl and carbonyl functional groups. As will be discussed later, these features play a significant role in embelin's dynamic photochemistry both in the ground- and -excited states.

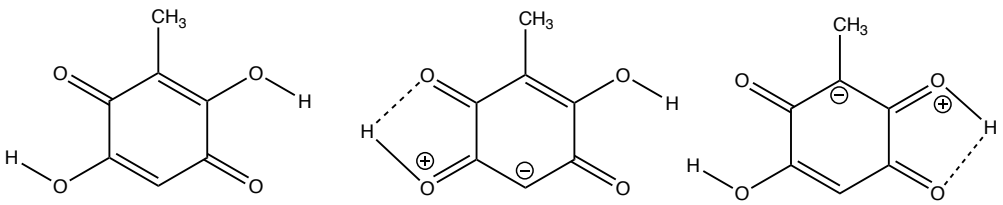


Figure 7: Top contributing resonance weighted structures obtained from the DFT ground state density.

Since chemical reactions take place in solutions, especially in aqueous media in the case of biological processes, solvent effects implemented in the PCM model were also included in the electronic ground state calculations. Water, DMSO, methanol and acetonitrile were used as solvents in this work. The ground state PCM results obtained showed that the influence of solvent was insignificant to the energies obtained above. A small increase in the ground state dipole moment from 0.62 D in gas-phase to 0.72-0.73 D in water and DMSO, methanol, or acetonitrile was however observed. This is due to the mutual polarization of embelin and the dielectric field of the solvent used.

3.4.1.2 Static Optical Properties and Vertical Excited State Results

In this subsection, the benchmarked vertical excitation energies for the first four low-lying gas phase singlet and triplet states are presented. The ADC(2) and ADC(3) together with TD-DFT (BLYP, BHHLYP, B3LYP and CAM-B3LYP xc-functionals) models used for computing these energies at the corresponding ground state geometries. There is a consistent agreement within the methods used say for the small excitation energy deviations. What's really remarkable in the results presented is the good agreement between the more computationally efficient TD-DFT/CAM-B3LYP//cc-PVTZ model and the state-of-the-art ADC (3) method at the same level of theory. The energies are presented in Table 1.

Table 1: complete Gas phase vertical excitation energies (eV) for the first four low-lying singlet and triplet excited states of embelin at the level of ADC(2), ADC(3) and TD-DFT (BLYP, B3LYP, CAM-B3LYP and BHHLYP) xc-functionals. Oscillator strengths.

State	ADC(2)	ADC(3)	BLYP	B3LYP	CAM-B3LYP	BHHLYP
S_1	3.03 (0.00)	3.09 (0.00)	2.32(0.00)	2.86 (0.00)	3.08 (0.00)	3.3567(0.00)
S_2	3.22 (0.00)	3.63 (0.00)	2.70(0.00)	3.24 (0.00)	3.66 (0.00)	3.8784(0.00)
S_3	3.46 (0.00)	3.90 (0.00)	3.00(0.00)	3.55 (0.00)	3.97(0.00)	4.209(0.00)
S_4	4.57 (0.39)	4.90 (0.46)	3.74(0.13)	4.57 (0.39)	4.81(0.33)	5.13(0.388)
T_1	1.77	2.29	1.68	1.87	1.99	2.01
T_2	2.26	2.77	2.33	2.49	2.50	2.34
T_3	3.36	3.35	2.41	2.72	3.12	3.31
T_4	3.50	3.64	2.61	2.77	3.47	3.64

The internal consistency checks within the TD-DFT excitation energies agrees with known non-local exchange behavior (see section 3.3). From the table, a comparison can be made between the computed CAM-B3LYP and ADC(3) singlet states energies. In the case of the singlet states, the difference in energy is less than 0.1 eV while for the triplet state case, ADC exceeds that of CAM-B3LYP with 0.3 eV. An assumption can therefore be made regarding the use of TD-DFT/CAM-B3LYP to accurately describe the photochemistry of embelin since it closely resembles ADC excitation energies. It is also important to note that there is no state ordering or large energy deviations between the other models (ADC(2) and TD-DFT) used.

The states can be characterized according to the nature of the excitation. Here, a number of approaches can be used namely direct analysis of the excited state wavefunction, difference in chemical bonding between ground and excited state, nature of the electronic transition, one-electron transition density between two states, transition dipole moments, and attachment/ detachment densities. The detachment/detachment density is the choice of analysis chosen in this case. The "detachment density" is the part of one-particle density corresponding to removed electron density from the ground state density distribution. The "attachment" density corresponds to the replacement of the detachment density in the distribution of the specific excited state. In other words, it is the returned detachment electron density.[83] The S_1 and S_4 are $\pi\pi^*$ transitions judging from their detachment and attachment densities. At the ADC(3) level of theory, S_1 has an energy of 3.09 eV and a very weak oscillator strength. The lowest bright state with the highest oscillatory strength is the S_4 which is consistent across all the methods and functionals used. States S_1 to S_3 are dark states gauging from their oscillation strengths. The vertical excitation energy at this state is 4.9 eV for ADC(3) compared with 4.881 eV. The remaining states S_2 and S_3 are $n\pi^*$ transitions as revealed by the attachment/ detachment densities. Figure 8 shows the respective attachment/detachment densities of embelin for S_1 to S_4 states.

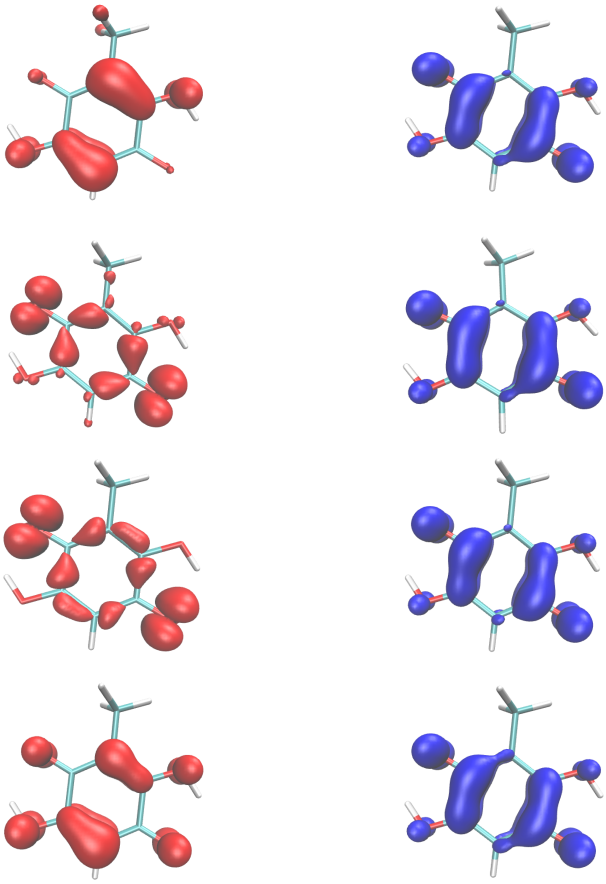


Figure 8: Attachment (blue)/detachment(red) densities for the first four low-lying states of embelin calculated at the TD-DFT/CAM-B3LYP model.

With respect to the detachment and attachment densities, the triplet states are analogous to the singlet states with T_1 and T_4 being $\pi\pi^*$ while T_2 and T_3 are $n\pi^*$ (see Figure 9). As with the singlets the state ordering is also maintained. The energy ordering presents a scenario where efficient intersystem crossing can be realized since the triplet energy T_1 and T_2 is lower than the singlet S_1. Therefore, reaction type I could also be possible in cases where energy criterions are met and will be explored in the proceeding sections.

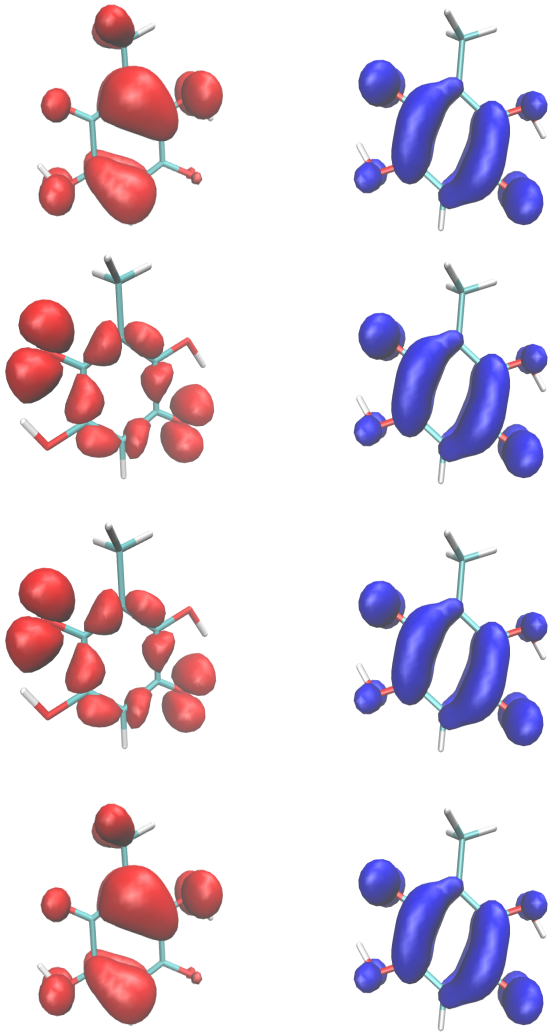


Figure 9: Triplet Attachment/detachment densities for embein calculated at the TDDFT/CAM-B3LYP model.

3.4.1.3 Exciton Properties and Bonding Patterns of the S_1 State Geometry

Kasha's rule[84] state that "*The emitting level of a given multiplicity is the lowest excited level of that multiplicity*". According to this rule, the internal conversion from higher excited states is faster than photochemistry or radiation decay. Photo reactions are therefore assumed to happen on the S_1 state thus analysis of the electronic structure for the S_1 excited state is provided in detail. One of the tools that provide for an easy interpretation of the one-electron density matrix is the exciton analysis tool.[85] It gives access to the *electron-hole(e-h)* picture and a direct way to classify the excited states

based on excited electron (exciton) descriptors. The descriptors include, participation ratios of the natural transition orbitals (PR_{NTO}), correlation coefficients, *e-h* distances which reveal key photochemistry features of excitation such as static or dynamic charge-transfer characters. The summarized S_1 to S_4 exciton descriptors for embelin using ADC and TD-DFT density matrices are summarized in Table 2.

Table 2: S1 to S4 Exciton descriptors of embelin computed at ADC (2) and ADC (3) and TD-DFT xc-functionals. They are represented by energy (E), Participation Ratio(PR_{NTO}), electro-hole distance (e-h), correlation coefficient (R).

S1	Method	E	PR_{NTO}	r_{e-h}	R
	ADC(2)	3.03	1.0252	0.29	0.033
	ADC(3)	3.09	1.0296	0.22	0.037
	B3LYP	2.72	1.0363	0.35	-0.027
	CAM-B3LYP	3.08	1.0355	0.35	-0.027
S2					
	ADC(2)	3.22	1.08	0.41	0.189
	ADC(3)	3.63	1.18	0.08	0.28
	B3LYP	3.22	1.023	0.30	0.10
	CAM-B3LYP	3.26	1.03	0.29	-0.03
S3					
	ADC(2)	3.50	1.1	0.34	0.22
	ADC(3)	3.89	1.23	0.05	0.32
	B3LYP	3.54	1.04	0.23	0.10
	CAM-B3LYP	3.60	1.04	0.21	0.12
S4					
	ADC(2)	4.57	1.09	0.19	-0.03
	ADC(3)	4.90	1.15	0.06	-0.02
	B3LYP	4.33	1.28	0.21	-0.16
	CAM-B3LYP	4.40	1.27	0.21	-0.16

Generally, there is consistency in the descriptor values for the states across all methods used. Multireference character of the excited state is measured by the participation ratio of natural transition orbitals (PR_{NTO}). The number of the PR_{NTO} indicates how many NTO pairs are involved in the excitation of a particular state. In all the methods, there is only one ratio of participating NTO pair meaning only one pair is needed to characterize each state. On the other hand, the electro-hole *(e-h)* distances measure the extent to which the hole and electron are separated and can be used to judge whether an excitation is of charge transfer (CT) nature or not. Large separations signify CT excitations while small distances depict absence of CT. The

distances range between ~ 0.2 – 0.4 Å and can be characterized based on the attachment/detachment analysis as a typical $\pi\pi^*$ and $n\pi^*$ excitation with no CT character. The correlation coefficient, R, shows whether the *(e-h)* pair moves together in a correlated manner or otherwise. It is clear that at TD-DFT level of theory, the electron and hole are anticorrelated (low R values) while ADC (higher R values) is weakly correlated. This trend is widely known for the TD-DFT method which heavily underestimate *e-h* correlation as well as CT characters.[86]

NBO analysis is also applicable to the excited state electronic structure density. As seen earlier, NBO reduces complex electronic structure to a simplified Lewis-like picture which is important for the excited state geometry analysis. However, the complex open-shell character and pronounced electron delocalization of excited state wavefunctions usually go beyond the localized Lewis theory. These electronic states are thus challenging for NBO and subsequent NRT analysis.[87] For embelin case however, NBO analysis of the excited state wavefunction provides useful chemically intuitive picture of its excited state structural analysis.

The analysis is first done at the Frank Condon geometry followed by optimized excited state geometry. The NBO 6 program was used to analyze the optimized geometry at the DFT/CAM-B3LYP//cc-PVTZ level of theory. Using density analysis, embelin exhibits a 3c,4e hyperbond between O3, C9, and C11 with corresponding occupancies of 3.71 and a lone pair at C11. Interestingly, upon vertical excitation, only one hyperbond is observed on one side of the molecule (see Figure 10 **A**). Due to the symmetry of the molecule, two hyperbonds are expected on both sides of the molecule because of the presence of two hydroxyl groups which are known to enhance resonance within the molecule at the ground state geometry. The scenario is however

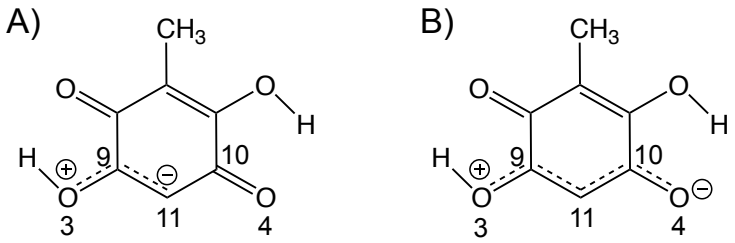


Figure 10: Main Lewis structures characterizing the S_1 state in the ground state equilibrium geometry (A) and at the energetically lowest S_1 equilibrium geometry (B) as obtained by NBO analyses.

different at the S_1 equilibrium geometry optimized at TD-DFT/CAM-B3LYP//cc-PVTZ model. Here, two hyperbonds are identified at O3, C9 and C11 as well as C11, C10 and O4 consisting of electron occupancies of 3.78 and 3.50 (Figure 10 **B**). The pattern can be clearly seen in the contributing natural orbitals (see Figure 11) and the bond lengths C9-C11 (1.347 Å) and C11-C10 (1.438 Å) for the ground state geometry and 1.402 Å and 1.417 Å at the S_1 equilibrium geometry. The hyperbonds show a pronounced *"resonance"* from O3 to O4 which in essence results into an *"ionic"* structure. There also exists a lone pair at C6 with no hyperbond on the CH3 side of the molecule. The Hyperbond values are summarized in Table 3.

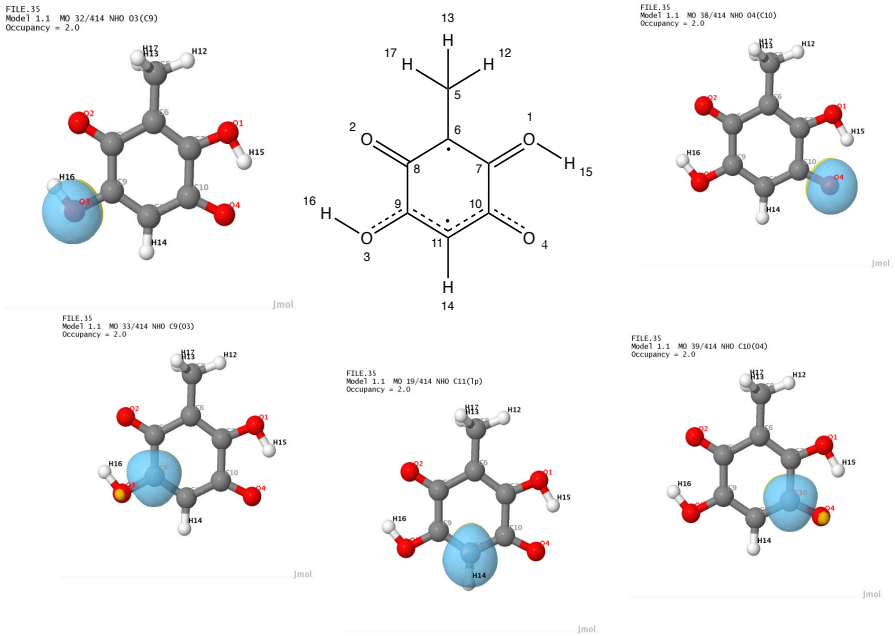


Figure 11: Contributing NHOs for Excited S1 state of Embelin depicting electron resonating from O3 to O4.

The resonating electron between O3 and O4 enhances embelin's photoactivity since at the S_1 equilibrium geometry a negative charge is exhibited at O4. The implication is that the transfer of proton from O1 to O4 should be energetically favorable. This could particularly be important in aqueous solutions or within a protein environment where hydrogen donor/acceptor reactions play critical roles. Hence one can speculate that O4 is able to abstract a proton from a hydrogen-bonded molecule, while O1 is likely to donate one.

Table 3: Hyperbonds formed by embelin on vertical excitation and excited state geometry.

	A: –B– :C	**%A-B/%B-C**	**Occupancy**
Frank Condon geometry	O3:– C 9–: C11	34.1/65.9	3.7051
Excited state geometry	O3:– C 9–: C11	34.1/65.9	3.7834
	O4:– C 10–: C11	42.4/57.6	3.4968

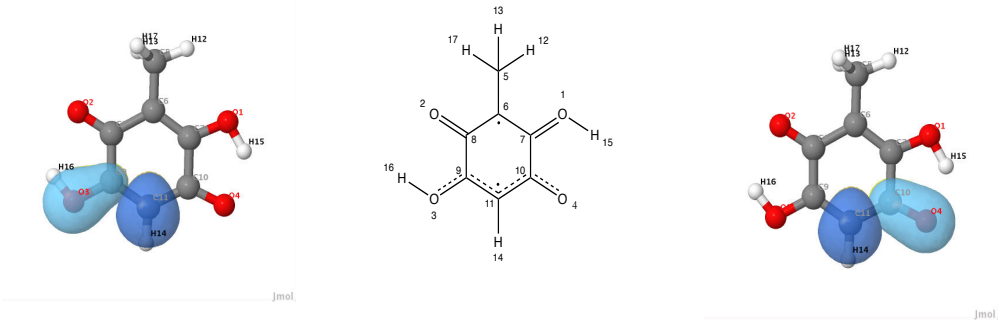


Figure 12: Excited states NBO hyperbond orbitals depicting electron from O3 to O4. This electron delocalization effect is responsible for the stability of embelin in the excited state yielding I_H. A similar behavior is not observed for I_{CH3}.

3.4.1.4 Influence of Solvation on Excitation Energies

In this subsection, the experimental spectra of embelin in solution is compared with calculated theoretical values. Embelin is reported to have poor water solubility but dissolves better in methanol and ethanol for which experimental data exist.[88] However, it is also important to interrogate the influence of water on its spectral signatures since biological reactions occur in aqueous solution. The maximum absorption wavelength of embelin is known to be λ_{max} = 288 nm (4.31 eV) in methanol or ethanol solution. Embelin has a vertical excitation energy of 4.32 eV for the S_4 state at the TD-DFT/CAM-B3LYP level of theory. This value is obtained by including two explicit methanol molecules in an implicit polarizable continuum model (PCM). The first excited state is a bright state with a significant oscillation strength of 0.71 in the lower region of the spectrum and is in perfect agreement with experimental values. S_1 which is responsible for the subsequent photochemistry has an energy of 2.79 eV. Table 4 shows all the other solvent values which are nearly the same given their similar dielectric constant.

Table 4: PCM excitation energies of embelin for Acetonitrile, DMSO, and Methanol solvents. The experimental value provided is in methanol.

	PCM-TD-DFT (CAM-B3LYP)			Expt.
	Acetonitrile	DMSO	Methanol	
S_1	3.00	2.98	2.97	2.7552
S_2	3.67	3.66	3.65	
S_3	4.00	4.00	4.00	
S_4	4.29	4.76	4.48	

Apart from the above-mentioned solvents, the influence of water on the excitation energy of embelin was investigated. This is because the focus lies on its photo biological role which in turn influence biomolecular mechanism. Therefore, water molecules were placed around it to investigate the effect of explicit solvents coupled with PCM model on excitation energies. The calculations were carried out at the TD-DFT/CAM-B3LYP level of theory. Figure 13 shows the respective water clusters used to obtain energies shown in Table 5.

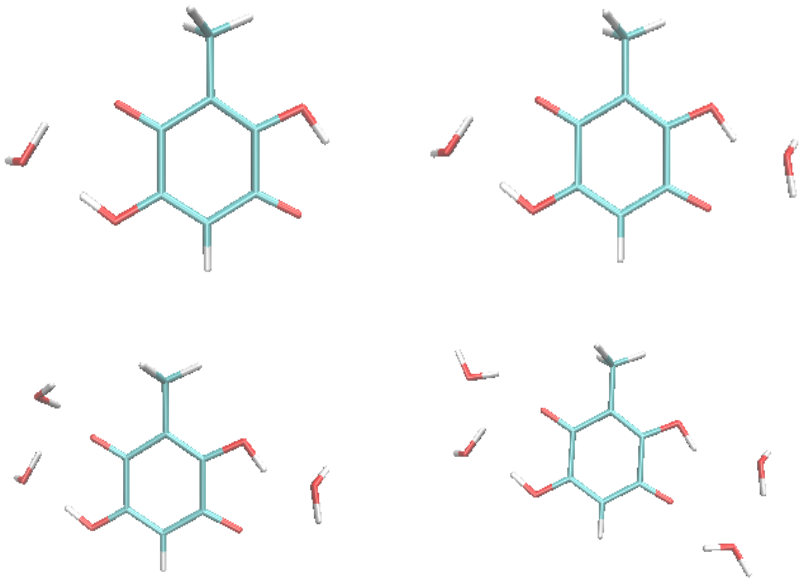


Figure 13: Equilibrium structures of water clusters (1-4) used to obtain excitation energies of embelin in PCM model.

Table 5: Excitation energies of embelin in explicit water clusters at the level of TD-DFT/CAM-B3LYP with n=0-4 representing the number of explicit water molecules. Osc. Strengths are given in parentheses.

State	n=0	n=1	n=2	n=3	n=4
S_1	2.99 (0.00)	2.92 (0.00)	2.81 (0.00)	2.81 (0.00)	2.81 (0.00)
S_2	3.66 (0.00)	3.57 (0.00)	3.56 (0.00)	3.57(0.00)	3.58 (0.00)
S_3	4.00 (0.00)	3.89 (0.00)	3.72 (0.00)	3.74 (0.00)	3.76 (0.00)
S_4	4.76 (0.13)	4.68 (0.35)	4.60 (0.38)	4.59 (0.40)	4.59 (0.42)

From the table, addition of one water molecule results in a red shift for all excited states by about 0.07 – 0.11 eV. A similar effect is observed with addition of two water molecules. Minor solvent effects are however observed with addition of three and four water molecules. It is thus clear that addition of two water molecules is sufficient to

capture the explicit interactions present in aqueous solution. This is important due to the fact that adequate intermolecular hydrogen bonds can be formed around embelin and the water molecules. Notably, the states S_1 to S_4 are negligibly influenced by the presence of water molecules thus one can safely neglect explicit solvents to get qualitative insights into processes taking place in the S_1 excited state.

3.4.2 Photobiological Activity of Embelin as a Photosensitizer

The utility of a PS in PDT against tumor cells encompasses a broad spectrum of mechanisms of action. Embelin can possibly act as a proton donor or acceptor, a reducing agent, and a singlet oxygen generator due to its chemical structure and hence provides a broad and robust mode of action upon irradiation. Furthermore, the redox chemistry of quinones is also known in literature and their roles in biochemistry have been widely documented.[89,90] In this work however, the mode of action explored is the excited-state proton transfer as well as embelin's capability to undergo intersystem crossing and subsequent oxygen generation. ROS generation is the mechanism which is more interesting for PDT applications especially in the case of cancer treatment approaches as well as other diseases.

3.4.2.1 Embelin as a Proton Donor or Acceptor

Experimental data shows that embelin possesses biological activity both in ground state as well as increased cytotoxic activity within PDT. Consideration of its proton donor and -accepting capabilities in the ground state as well as in the first excited electronic state is therefore informative in an effort to improve its photo biological applications. The hypothesis applied for this case is that there is a direct correlation between its ability to donate a proton in an aqueous solution or a protein environment to the possibility of undergoing intramolecular proton transfer. Meaning, the easier it is for embelin to perform intramolecular proton transfer, the more likely it is to undergo intermolecular transfer within a protein or when bound to another molecule via hydrogen bond networks.

As previously seen, two intramolecular hydrogen bonds exist in the ground state and its electronic structure has been shown to be characterized by pronounced electronic resonance. Effort has been taken to investigate whether it is possible for embelin to

undergo proton transfer which could occur via a single or double transfer mechanism owing to the nature of its structure. These reactions are driven by the fact that OH protons are adjacent to the carbonyl groups.

The results from the ground state PES reveal that a single proton transfer produces energetically high-lying, short-lived intermediates, which are named depending on which side of the molecule the proton is transferred (i.e., intermediate H (**I_H**) and intermediate CH_3 (**I_{CH3}**). This reaction occurs in a stepwise manner in both directions resulting in the same starting structure (isomer). Likewise, simultaneous double proton transfer does not change embelin's chemical structure. In other words, single and double intramolecular proton transfer within the molecule leads to identical reaction or a degenerate arrangement. Figure 14 shows the possible transfer paths and corresponding energetics calculated using DFT/CAM-B3LYP//6-31G* model. The corresponding ESIPT reaction processes is also carried out using a comparable TD-DFT/CAM-B3LYP/6-31G* model.

From the energies of the different intramolecular proton transfers, it is apparent that the stepwise proton transfer is favored to simultaneous double proton transfer in the electronic ground state. This reaction proceeds along almost two identical sequences with energy barriers of 72.8 and 71 kJ/mol for I_H and I_{CH3} respectively. On the other hand, the energy barrier for simultaneous double proton transfer is as large as 103.2 kJ/mol at DFT/CAM-B3LYP level of theory. These energies therefore exclude the latter to be thermally relevant for embelin's proton transfer mechanism. Additionally, the intermediates highlighted lie 63.8-66 kJ/mol above the most stable ground-state with small separating energy barriers of about 7.2 and 6.8 kJ/mol respectively and are hence short-lived. The ground state described here is presented as a 2D picture in Figure 15 (red colored).

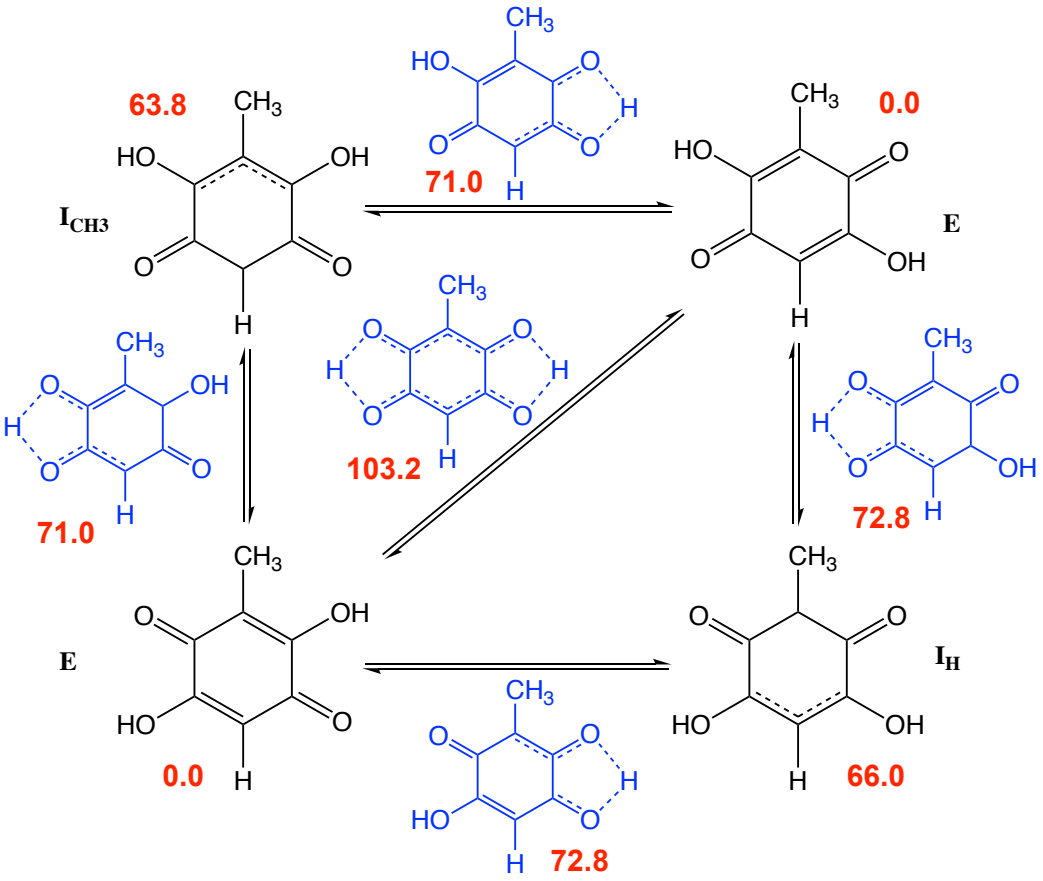


*Figure 14:Possible proton transfer pathways of embelin (**E**) in the electronic ground state. The reaction occurs via intermediates I_{CH3} and I_H calculated using DFT/CAM-B3LYP model. The black structures correspond to local minima on the potential energy surface, while blue ones are transition states. The corresponding energies are highlighted in red (kJ/mol) relative to the most stable ground state equilibrium structure.*

Having defined the ground state PES, focus is now turned to the excited PES and excited state proton transfer reactions will be explained in detail. Recall that the bright state of embelin is the S_4 which has the highest oscillator strength. This is the state that is most likely excited when embelin is irradiated. However, one can assume that consequent photoinitiated processes occur within the first excited state S_1 as defined earlier by Kasha's rule. Hence internal conversion is usually much faster than regular chemical processes. As a result, stable structures **E**, $\mathbf{I}_{CH3}$, and $\mathbf{I}_H$ identified in the ground-state were reoptimized in the S_1 state using TD-DFT/CAM-B3LYP//6-31G* model. These structures correspond to stable isomers in the S_1 state albeit with different geometrical parameters as summarized in Table 6.

Table 6: Geometry differences between intermediate structures formed in the ground and excited state PES of embelin. It is evident that at the excited state geometry, the dihedral angles increase in correspondence with the coupling vectors.

Parameter	**I_H(gs)**	**I_{CH3}(gs)**	**I_H(es)**	**I_{CH3}(es)**
O3-C9	1.32	1.28	1.33	1.23
C9-C11	1.37	1.38	1.38	1.45
C10-C11	1.38	1.40	1.39	1.45
C10-O4	1.31	1.27	1.34	1.23
O2-C8	1.26	1.31	1.23	1.34
C8-C6	1.4	1.39	1.46	1.40
C6-C7	1.39	1.38	1.5	1.40
C7-O1	1.27	1.39	1.23	1.34
C8-C6-C7 (degrees)	114.32	112.94	120.06	118.34
C9-C11-C10 (degrees)	117.95	119.10	120.14	121.46

Notably, the energies of the isomers in the S_1 are inverted (see Figure 15). The implication is that, at the first excited state, the intermediates $\mathbf{I}_H$ and $\mathbf{I}_{CH3}$ discussed above are now stable compared to their ground state counterparts. Structure $\mathbf{I}_H$ has an energy of 58.6 kJ/mol (0.61 eV) relative to the most stable ground state structure **E** and is more stable than $\mathbf{I}_{CH3}$ by a difference of 17.2 kJ/mol (0.18 eV) in energy. This stability can be rationalized by the fact that at the excited state, embelin exhibit delocalization of electrons between O3 and O4 forming hyperbonds as previously explained using NBO analysis. Inspecting the 2-D relaxed scan of the S_1 PES along the two proton-transfer coordinates Figure 15, shape of the S_1 PES in comparison to the ground state depicts a readily apparent inversion which strongly supports the fact that excited state intramolecular proton transfer (ESIPT) of embelin readily occurs on photon absorption. Additionally, the transfer of proton towards the intermediate $\mathbf{I}_H$ is slightly more favored to $\mathbf{I}_{CH3}$ due to higher stability and is in agreement with the hyperbonds from NBO theory analysis previously explained.

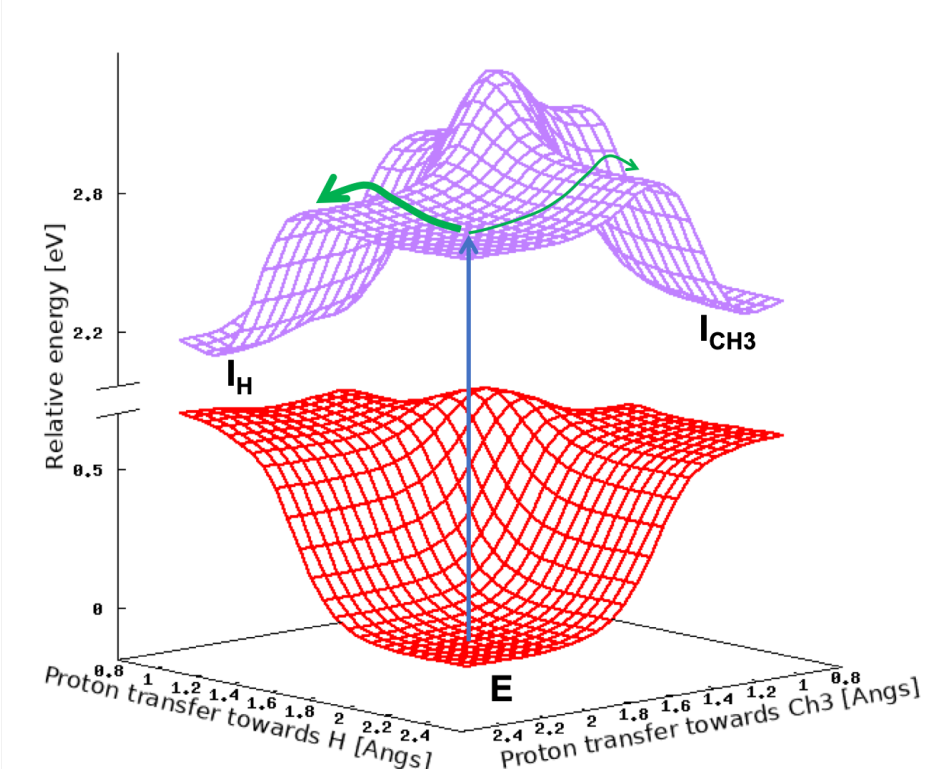


Figure 15: Two-dimensional relaxed scan of the S_1 potential energy surface at TDDFT/CAM-B3LYP level along the two proton transfer coordinates. Upon optical excitation of embelin (sketched as vertical blue arrow), the molecules arrive at the E structure on the S_1 surface and can decay via single proton transfer into the more stable isomers I_{CH3} and I_H (green arrows), while the latter is more likely due to the smaller energy barrier and lower final energy.

The PES of the ground and excited S_1 states approach each other at the $\mathbf{I}_H$ and $\mathbf{I}_{CH3}$ structures. The energy gap also becomes as small as 0.86 and 1.05 eV respectively leading to a hypothesis that nonradiative transitions from S_1 back to the ground state can be readily facilitated. Very often, nonradiative decay is mediated by a conical intersection (CI), a so-called photochemical funnel and local minima in excited states are indeed often thought of as "missed" CIs. A search for nearby CIs between the ground state and the S_1 starting at the local minima $\mathbf{I}_H$ and $\mathbf{I}_{CH3}$. Since it is well known that CIs are multireference in nature, spin-flip TD-DFT was employed for the minimum energy crossing point (MECP) calculations in connection with the branching-plane updating method. Researcher have previously shown that this methodology can reproduce correct topologies for statically correlated geometries in different diradical and triradical systems especially at the CIs.[91–93] The identified CIs

in the spatial vicinity to the local minima $\mathbf{I}_H$ and $\mathbf{I}_{CH3}$ are herein referred to as $\mathbf{CI}_H$ and $\mathbf{CI}_{CH3}$, respectively (Figure 16). These conical intersections are characterized by their branching spaces defined by the gradient different vector **g** and the derivative coupling vector **h**, which reflect the nuclear motions that lift the degeneracy of S_0 and S_1. The g-vectors are dominated by an asymmetrically coupled C-C stretch vibration within the six-membered ring at the side of the hydroxy group, i.e., the part of the ring which contains the newly formed hyperbonds in S_1. On the other hand, the *h*-vectors correspond to the C-C stretch vibrations on the opposite side of the ring (see Figure 16).

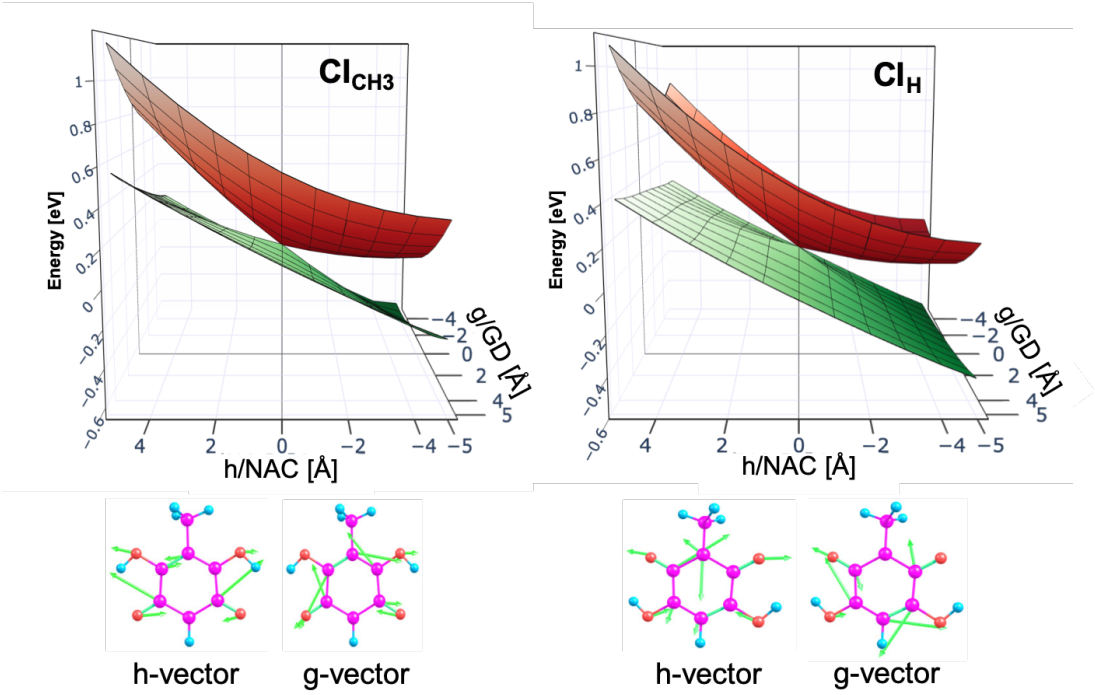


Figure 16: Optimized minimum energy crossing points of the two S_1/S_0 conical intersections CICH3 (left) and CIH (right) present in the vicinity of the ICH3 and IH intermediates, respectively. The h- and g-vectors correspond to the derivative coupling and the gradient difference vectors characterizing the conical intersections.

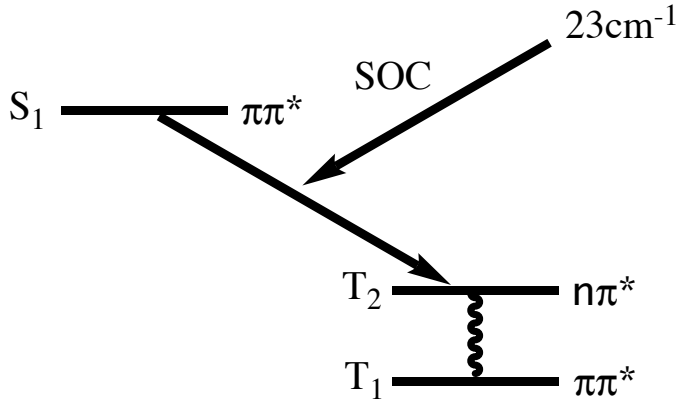
The pathway embelin follows from the S_1 back to the S_0 state is quite complex. The obtained CIs lie well above their corresponding minimum structures. This is to mean that $\mathbf{CI}_H$ lie 0.43 eV above $\mathbf{I}_H$ and $\mathbf{CI}_{CH3}$ lie 0.74 eV above $\mathbf{I}_{CH3}$. But since the ESIPT process is fast and embelin possess a lot of excess energy in the excited state when relaxing to the intermediate structures. This energy is likely sufficient to reach the CIs identified. What this means is that on absorption on photon, there are possible decay channels back to the ground state which include hitting a CI and relaxing back to

previously explained ground state intermediate structures (refer back to Figure 14 and Figure 15). This efficiency of this path is corroborated by the nature of the CIs which are very similar in topology and can be categorized as the "sloped" CIs.[94] Sloped CIs are known to facilitate only one ground-state relaxation pathway and in this case, the relaxation leads to the ground-state intermediate structures ($\mathbf{I}_H$ and $\mathbf{I}_{CH3}$) which as explained above are high in energy, unstable and are short-lived. Eventually, the back transfer of proton to form embelin's isomer which is the original structure takes place on the ground state PES.

So far, clear dynamics of what takes place when embelin absorbs a photon has been highlighted. The ESIPT dynamics on the PES can be captured well using nonadiabatic dynamics simulations that can aid in quantifying reaction yields. This is however not feasible due to the large system size. It is nevertheless conclusive from the analysis that ESIPT is favored when embelin absorbs a photon in the gas phase. Since the aim is to use embelin in a protein environment, a different mechanism or modes of action when it is bound to hydrogen donor/acceptor pairs could be favored. In this case, embelin might prefer to undergo intermolecular proton transfer thereby resulting in either structural change or electron transfer leading to protein damage.

3.4.2.2 Singlet Oxygen Sensitization Capability of Embelin

TTET facilitate the generation of singlet oxygen. This subsection looks at the ability of embelin to efficiently cross to the triplet state and whether it does so with sufficient energy to generate singlet oxygen. The triplet state is efficiently reached *via* spin-orbit coupling and consequent intersystem crossing. Generally, there are rules with which PS can generate singlet oxygen. One of the fundamental requirements is that the



excitation energy of the triplet state $S_0 \rightarrow T_1$ must be higher than the excitation energy of the singlet oxygen $^3\sum_g^- \rightarrow {}^1\Delta_g$. Here, the assumption made is that the triplet excitation energies are the sole generator of singlet oxygen. This is because of the short-lived excited singlet state and efficient CIs pathways explained above.

The energies represented in Table 1 at the ground state equilibrium structure **E (**Figure 15**)** of embelin shows that T_1 and T_2 have lower energy values than S_1. Moreover, they are $n\pi^*$ in nature while the S_1 is of $\pi\pi^*$ character. Due to El Sayed's rules, it is possible for embelin to efficiently cross to the triplet manifold favored by the nature of the transition characters. The spin-orbit calculation (SOC) values of embelin have been calculated using TD-DFT/CAM-B3LYP//cc-PVTZ model to support this picture. The values between S_1 and T_1 is 0.001 cm^{-1} while that between S_1 and T_2 is 23 cm^{-1}. This is a relatively weak number compared to known molecules such as porphyrins which exhibit large SOC values. However, there exist molecules with small SOC values, but faster rates are known in literature especially those containing aromatic carbonyls.[68] Examples of such molecules include benzaldehyde,[70] benzophenone,[69] and substituted quinone derivatives e.g., chloranil, and molecules that contain same scaffold as embelin.[95]

Another factor that could facilitate large intersystem crossing rate according to Fermi's Golden rule is the singlet triplet gap i.e., the smaller the difference between S_1 and T_2, the larger the *isc* rate. From Table 1, the singlet triplet gap between S_1 and T_2 is 0.58 eV at TD-DFT/CAM-B3LYP method and 0.32 eV at ADC(3) level of theory. Since T_2 is lower than S_1, state crossings occur along the proton transfer pathway on the singlet surface. This is because at the identified conical intersections, the T_2 state lie energetically above the CIs by 1.62 eV and 1.09 eV respectively. Intersystem crossing is thus facilitated by this crossing. It could be put in simple words that, inter system crossing in embelin is mediated by proton transfer on the first excited state surface (ESIPT). Internal conversion then plays an important role in the decay of the molecule from T_2 to T_1. Normally, the T_1 state is a long-lived state and, in this case, the triplet should be able to mediate singlet oxygen generation *via* TTET

process. It is known experimentally that the excitation energy of triplet oxygen $^3O_2$ ($^3\sum_g^-$) to singlet oxygen $^1O_2$ ($^1\Delta_g$) is 0.98 eV.[5,9] Ideally, this occurs first through vibrational relaxation from ($^1\sum_g^+$) which is less stable to a highly reactive and metastable $^1\Delta_g$. The T_1 excitation energy of embelin at the T_1 -optimized equilibrium geometry is 1.33 eV using TD-DFT/CAM-B3LYP//cc-PVTZ model. This energy is enough to be quenched by the triplet oxygen and consequent conversion into singlet oxygen.

3.4.3 Additional Factors for PDT Application of Embelin

Singlet oxygen generation is commonly observed in virtually every chemical system involving absorption of light in the presence of oxygen. As mentioned above, there are three major reasons why photosensitization of singlet oxygen is common in nature.[9]

i. For both the S1 and T1, energy transfer quenching is spin-allowed due to the special configuration of molecular oxygen. This is in contrast to the many competing deactivation processes.
ii. The energy required to activate singlet oxygen, particularly $O_2(^1\Delta_g)$ are low.
iii. Singlet oxygen rapidly diffuses in many media due to its very small size. Hence intermolecular interactions are likely to be quenched by triplet oxygen for excited states that are sufficiently long-lived.

From a medical perspective however, there are a number of characteristics that a molecule must possess to qualify as a PS. Without these qualities, then the generated singlet oxygen capability cannot be medically useful for PDT applications. Some of the known characteristics highlighted in refence[96] are summarized in Table 7.

Table 7: Some of the known requirements for a molecule to qualify as a drug.

Purity	The PS should be pure and consist of a well-known composition which is stable at room temperature
ADMET	A good PS should possess medically allowed descriptors for Absorption, Distribution, Metabolism, Excretion and Toxicity.
Quantum yield	Should have high singlet oxygen quantum yield (Φ_{Δ})
Carcinogenicity/Mutagenicity	A good PS should not result in other carcinogenic effects. This is commonly observed in the form of long-term skin side effects.
Cost/ Availability	Since cancer is a killer disease, an inexpensive PS would promote extensive utilization of treatment.

3.5 Conclusion

In this section, the general ground state chemistry as well as the photochemistry of embelin in both gas phase and in solvent models have been highlighted. Both static and dynamic behavior of embelin have been studied in detail using previously reported experimental features. The DFT, MP2 methods for ground state and the corresponding TD-DFT and ADC methods for excited states have been exclusively used in this chapter. It is important to note that the ADC (3) method used for the excited states is able to accurately reproduce experimental absorption spectra as well as electronic descriptors of embelin. On the other hand, DFT has been used to predict the experimentally observed resonance within the structure. Additional information is revealed by the ground state PES which serves as a pivotal pathway for embelin on photon absorption. Key intermediates on the ground state PES have also been shown to be useful for characterization of embelin's excited state decay channels especially on its back to the ground state.

Due to its unique molecular structure with hydrogen bonds on both sides, embelin has been shown to possess a versatile photochemistry. This characteristic of embelin is useful for PDT application in the sense that it has capabilities to modulate various photo mechanisms and pathways. First, it has the ability to undergo resonance which at the excited state mediate important mechanisms. A possible mechanism is the possibility to donate or accept a proton on photoexcitation which occurs *via* ESIPT on the first excited state potential energy surface. This reaction proceeds along embelin's two intramolecular hydrogen bonds that leads to two different low-lying excited state intermediate structures, $\mathbf{I}_{H}$ and $\mathbf{I}_{CH3}$. The formation of these intermediates depends on the direction of the proton transfer. Reactions that can be influenced by proton donation include redox coupling as well as electron transfer that play critical role in biological systems. On the other hand, direct reactions with a biomolecule are possible when embelin is bound to proteins for example.

The second possible pathway is generation of singlet oxygen. Along the ESIPT photo mechanism pathway, efficient routes for *isc* to a long-lived T_1 are facilitated by the nature of the S_1 ($\pi\pi^*$) and T_2 ($n\pi^*$) transition pairs. At T_1, embelin has been shown to possess enough energy that can be transferred to a nearby molecule such as

molecular oxygen. This process has been proven experimentally to be aggressive towards tumor cells and thus is a useful process in PDT for cancer treatment.

The section does not however address what happens when embelin is bound to a protein or another molecule. Intermolecular hydrogen bonds might be dominant compared to the intramolecular counterparts discussed in this section. For this purpose, advanced hybrid QM/MM studies have been carried out to investigate how embelin behaves when it is "embedded" within an environment. This forms the basis of chapter 6. The proceeding chapter addresses embelin's binding patterns when bound to the BIR3 domain of the XIAP protein.

Chapter - 4

Binding Patterns of Embelin and X-linked Inhibitor of Apoptosis

This section builds up on the previous chapter using the known facts about embelin especially its ability to act as a donor/acceptor. The main aim is to explore the binding mechanism of embelin when bound to a stable pocket of the XIAP (BIR3) protein. In terms of research design, docking simulations are first carried out, followed by molecular dynamics simulation coupled with Markov state models. Finally, the binding energy between embelin and BIR3 obtained using the MM/PBSA method is presented. The results discussed herein have been submitted for publication under the title:

Michael Rogo Opata *and* ***Andreas Dreuw****:* Characterization of Embelin-BIR3 binding patterns using molecular dynamics and Markov state models. *J. Phys. J. Phys. Chem. B 2021, under review.*

4.1 Abstract

Developing drug-like molecules requires an understanding of the binding modes between the ligand(s) and the active site of the protein under study. Key factors that play critical roles include hydrogen bonds, pocket fitting, as well as isomerization. These factors serve as guiding footprints when choosing a reliable ligand which can be used as is or modified to act as a lead molecule or drug. Typical drug development approaches from a theoretical point of view involve docking, molecular dynamics, optimization, and advanced simulations such as free energy perturbation simulations. In this chapter, embelin which has been explored in the previous chapter is found to bind the BIR3 domain of the X-linked inhibitor of apoptosis (XIAP) in a stable pocket that is shifted by ~ 3 Å from a previously reported experimental site. Molecular dynamics and Markov state models reveal the metastable states and binding modes that take place between the ligand-protein complex. The dominating binding mechanism occurs via its hydroxyl groups which play an important hydrogen donation role with two Glu 239 (E 239) and Glu 220 (E 220). A second binding mechanism comprising of a hydrogen accepting ability is identified, though this binding is comparatively weak to the first one. The two mechanisms have been widely referred to as the "*binding modes"* in the following sections.

Binding energies within these modes are equally important and have also been reported using the MM/PBSA method. To obtain the energies, the MD trajectories have been categorized based on their binding modes. Mode one has an average binding energy of -18.67 kJ/mol while mode two has a binding energy of -30.80 kJ/mol. An average of the two is comparable with experimentally obtained energies of embelin in a human serum albumin (HSA). The nearly equivalent average theoretical binding energy to experiment is satisfactory leading to a conclusion that the embelin-BIR3 binding perform useful biological functions which is capable of restoring apoptosis.

4.2 Introduction

There are several reasons why abnormal cell growth occurs within the human body. One of the contributing factors is the interference of programmed cell death process which is also known as apoptosis. The term is always used to mean that cell death occurs in certain locations and at specific body sites in a predictable manner,[97] a process which is important for generation of new cells. Failure of cell death processes may result in biological malfunctions and diseases such as cancer within the body. Activation of cell death processes are well known and normally occur in either caspase-dependent or caspase-independent modes.[98] In both cases the mitochondria play a critical role. For the caspase dependent pathway, mediation of cell death occurs through the deactivation of cell integrity by caspases which are a family of synthesized endoproteases capable of hydrolyzing peptide bonds.[14] Caspases are classified according to their mechanism of action: (a) The initiator (caspase-8 and -9) caspases and (b) the executioner/ effector (caspase-3, -6 and -7). Initiator caspases are involved in reactions that trigger production of executioner/effectors which in turn induce apoptosis. These reactions depend on catalytic cysteine residues sites and cleave only after specific aspartic acid residues. Hence, they are known to bind to specific substrate sites. Other proteins containing aspartic residue sites can as well bind to them as antagonists. This can be very detrimental in the case of cell death because blocking caspase means cell death processes becomes totally uncontrollable.

Inhibitor of apoptosis (IAP), especially the X-linked type (XIAP) are known to interfere with programmed cell death. They execute their functions mostly via baculoviral repeat (BIR) domains which prefer to bind to caspases. Such a mechanism therefore promotes proliferation of aged cells leading to dangerous diseases such as cancer. There are two commonly known BIR domains of XIAP namely, the BIR2 and BIR3 which play different roles in apoptosis processes. This work is focused on the BIR3, a known apoptosis antagonist and is mainly an inhibitor of the initiator caspase, caspase-9. Functionally, caspase-9 play critical roles especially in the processing of the executioner caspase-3 which trigger intracellular disintegration.[98] BIR3 is a zinc-binding domain inhibited by natural mitochondrial polypeptides such as SMAC/DIABLO.[13] The N-terminal of SMAC enables its pro-apoptotic functions in

an intrinsic pathway and therefore its binding patterns can be used to design new peptide or non-peptide molecules to displace it and bind BIR3. In fact, this concept has previously been used in deriving small peptides just by mimicking specific functionalities of SMAC.[99] Details presented in this chapter follow a similar idea of mimicking the N-terminal binding of SMAC using embelin in a bid to identify a non-peptide BIR3 inhibitor. Identification of the binding site, mechanism involved in the binding and the energy of binding is of prime interest.

An alternative means with which cells can die is the caspase-independent pathway, a process in which other molecules mediate apoptosis. In this scenario, it has been proven that inhibiting caspases does not block cell death.[100–103] In other words, studies have shown that in certain cases, even if caspase are inhibited, apoptosis still occurs. The loss of mitochondrial transmembrane potential is one key feature observed with this mode of apoptosis.[98] More importantly, caspase-independent apoptosis mode is mostly mediated by ROS. In the case of ROS, several model systems have been observed to undergo cell death in caspase-independent processes for example in HIV-induced T cells death.[104] The mechanism within a biological context relies on the strength of the initiator molecule to act as powerful antioxidant which mediate superoxide and other radical species. Candidates with this capacity include, metallic porphyrins, and its derivatives of benzoic acids such as manganese (III) tetrakis (5, 10, 15, 20-benzoic acid) porphyrin. The dominating pathway or reactions involved here is the PDT. This path has been explored in the preceding chapter and will be expounded in the next chapter when the ligand is embedded within a protein environment.

Given the above knowledge, it is therefore important to find small non peptide molecules that can inhibit BIR3 which then "*frees*" caspase to induce apoptosis. Alternatively, the molecule should be able to activate cell death *via* caspase-independent route acting as a powerful antioxidant to interfere with the cell membrane. In this respect, embelin serves as a suitable candidate for exploration as a Porphyrin alternative. It is a small natural product belonging to the quinone scaffold that has been extensively studied to inhibit XIAPs.[57,58,105] Structurally, embelin is similar to the known bioactive drug coenzyme $Q_{10,}$ a known antioxidant and contains

unique *benzoquinone* and *hydroquinone* groups (see Figure 6). Nikolovska-Coleska and co-workers' experimental data found embelin to bind to the groove surface of BIR3. The defined pocket on this surface are the four residues bound to SMAC which are (alanine-valine-proline-isoleucine (AVPI)). Furthermore, this binding mechanism has been found to be competitive with caspase-9 and makes the pocket suitable for designing new molecules that can perform similar functions. It is a common approach in drug discovery to use the binding patterns of natural inhibitors or X-ray co-crystallized ligands as guiding factors to develop new or similar molecules with the identical biological functions.

As highlighted in Chapter - 3 above, previous studies have explored the use of embelin as a PDT and the results from the pure quantum mechanical studies confirmed this possibility. In a protein environment, several dynamics play important roles and therefore investigating how different physiochemical effects acting on the ligand of interest under parametrized forces is critical to understanding its binding. A stable dynamics profile would indicate proper binding while an unstable one signifies weak or no binding.

This chapter deals with the dynamics of embelin in BIR3 domain of XIAP. The approach used is a classical molecular dynamics study that has been carried out to investigate whether embelin can bind in a similar pattern to SMAC. The most important results being sought after is the pattern of binding that replicates how BIR3 bind to caspase since SMAC and caspase binding mechanisms are known to be naturally competitive.[58] In classical molecular dynamics advanced techniques can be used to include free-energy calculations, integration algorithms, polarizable force fields, and enhanced-sampling. The main shortcoming of MD is that long timescales are needed to describe biological processes which is most times unfeasible. This problem can be solved by combining short MD trajectories with Markov state models (MSM) to bridge the timescale hence reveal processes that occur over a long period of time.

Results obtained from classical molecular dynamic approach has been analyzed using Markov state models and MM/PBSA to characterized how embelin binds BIR3. In MSM, a solution to the rate equation also known as the master equation enables

resolution of the transition states constructed from discretized trajectories.[106,107] Memoryless time transition matrices between two states that describe longtime physical events are created and used to trace the pathway of the ligand into the binding site. This technique has been successfully applied to study protein folding as well as ligand-binding studies which include locating hidden allosteric sites[108–110] and has been considered in this section. The general workflow applied in studying embelin in this chapter in shown in Figure 17 .

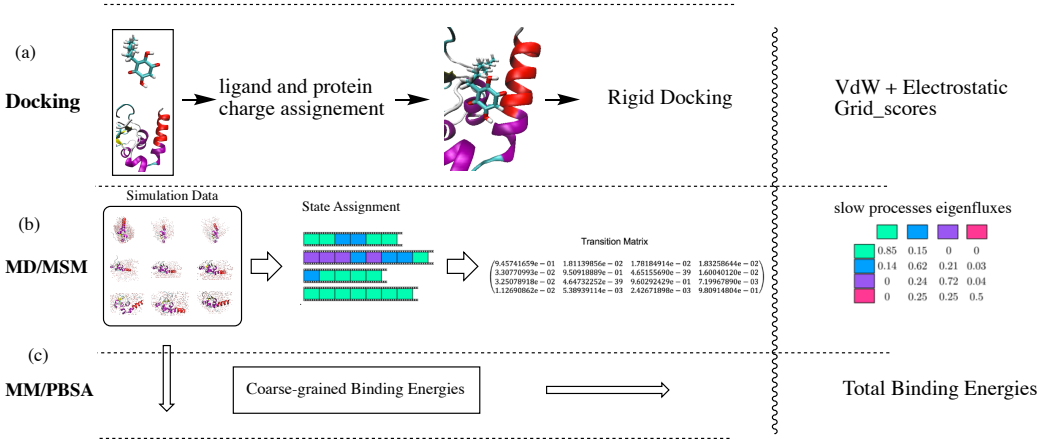


Figure 17: Overall design workflow of the studies carried out in this chapter. (a) Depiction of rigid molecular docking where an initial ligand-complex was first created. The poses obtained were then used to run MD simulations. (b) MD/MSM simulation where data is first obtained using classical MD trajectories. The data is then subjected to clustering and slow processes identified. (c) The obtained trajectories are subjected to MM/PBSA binding energy studies.

4.3 Markov State Model Theory (MSM)

A brief theory on MSM is highlighted. For a comprehensive derivation, see the work in references.[107,111–113] MSM theory applies Markov chains as its underlying mathematical framework which uses the idea of random variables.[114] In essence these are functions with certain values having probabilities associated with them. The MSM framework involves decomposing phase/configurational space occupied by a dynamical system into a set of disjoint, discrete states and a transition matrix denoting the conditional probability of finding a system in state *j* at time $t+\tau$ given that it was in state i at time t.

Recall *macrostates* are collections that have some attribute in common. We can define a set of indicator functions, $\Omega^{(i)}(x),$ classifying microstates to which macrostates they belong.

$$\Omega^{(i)}(x)=\begin{cases} 1 & \textit{if microstates x is in macrostate i} \\ 0 & \textit{if not} \end{cases} \tag{64}$$

In MSM, we compute transition matrices that describe temporal evolution of the system.[106] Then transition matrices are computed from transition functions and time correlation functions of the indicator functions. Formerly, both of these functions are averaged over canonical ensembles of information derived from energy conserving (microcanonical) trajectories . Mathematically, this can be expressed as.

Suppose we have M Boltzmann weighted starting states from which trajectories $x_m(t), m=1,...,M,$ have been computed from $t=0$ to $t=T_m$. We can then estimate the time correlation function between indicator functions i and j, $C_{ij}(\tau)$:

$$C_{ij}(\tau)\equiv\left\langle \Omega^{(i)}(x(\tau))\Omega^{(j)}(x(0))\right\rangle \tag{65}$$

$$=\frac{\int dx(0)e^{-\beta H(x(0))}\Omega^{(i)}(x(\tau))\Omega^{(i)}(x(0))}{\int dx e^{-\beta H(x)}} \tag{66}$$

$$\simeq \frac{1}{M}\sum_{m=1}^{M}\frac{1}{T_m-\tau}\int_0^{T_m-\tau}dt\left[\sum_i \Omega^{(i)}(x_m(t+\tau))\right]\Omega^{(j)}(x_m(t)) \qquad (67)$$

Where, $H(x)$ is the Hamiltonian, $\beta = 1/kT$ is the Boltzmann constant and T is the temperature. We can also compute the probability of finding the system in a microstate that is consistent with some particular microstate *i*.

$$P^{(i)} = \frac{\int dx e^{-\beta H(x)}\Omega^{(i)}(x)}{\int dx e^{-\beta H(x)}} \qquad (68)$$

$$\simeq \frac{1}{M}\sum_{m=1}^{M}\frac{1}{T_m-\tau}\int_0^{T_m-\tau}dt\Omega^{(j)}(x_m(t)) \qquad (69)$$

The transition functions, $T_{ij}(\tau)$ are defined and computed as

$$T_{ij}(\tau) \equiv \frac{\int dx(0)e^{-\beta H(x(0))}\Omega^{(i)}(x(\tau))\Omega^{(j)}(x(0))}{\int dx e^{-\beta H(x)}\Omega^{(j)}{}_{(X)}} = C_{ij}(\tau)/P^{(j)} \ . \qquad (70)$$

The term τ is the *lag* time, which refers to the period one "waits" before characterizing the system, after having seen to be in some condition at time zero.

4.4 Methods

4.4.1 Ligand-Protein Complex Preparation

The crystal structure of BIR3 domain of XIAP was obtained from the protein data bank (PDB ID 1G73).[58] Chains B and C were deleted since the protein has two similar repeating units. It is important to note that the protein did not contain any water molecules and thus standard protein and ligand preparation procedures were applied to obtain a complex of embelin and BIR3. Chain A was truncated to use only the first fifteen residues. Ideally, only the first five would suffice but for stability purposes, more residues were included. Standard charges were added to the receptor using AMBER-BCC charges utility in chimera program.[115] A similar step was carried out for the ligand. The resulting ligand-protein complex was then prepared using a rigid docking procedure as implemented in DOCK 6 program.[116] As explained in the

introduction section, the binding pattern of SMAC can be used to design new molecules and, in this case, the first few N-terminus residues of SMAC were included together with the BIR3 domain to define an initial binding pocket for embelin. Residues within a radius of 12 Å were included to create a binding sphere using Sphgen utility program. The sphere represents the "active site" for intended calculations. After defining a sphere, a grid was generated using the grid generation program. The distance chosen to serve as the boundary of the pocket of the new receptor was 18 Å padded as a square. This ensures equal chances of the ligand binding in any side of the receptor and in areas where the electrostatics and van der Waals forces are dominant.

4.4.2 Rigid Docking

The docking method that was chosen for obtaining the complex was a rigid-conformer-search rather than a flexible one. This is because the goal was to obtain static binding conformers which could then be used for further MD calculations. A total of ten conformer search was requested so that different ligand clusters would be sampled for binding certainty. Consequently, a ranking was obtained using a total Grid_score constituting of individual electrostatics and van der Waals interactions between the receptor and the ligand. An initial consideration in the formal docking results was the hydrogen bonding patterns between embelin and the protein residues. Therefore, the conformers which had favorable binding scores were further pursued. Hydrogen bond analysis which complemented the docked poses was therefore carried out using the FindHbond utility in Chimera. From the resulting top-ranked conformers, promising hits were chosen for further MD calculations.

4.4.3 Molecular Dynamics Simulations

The chosen conformers from the docking procedure were chosen for MD simulations. Initial preparations were carried out using VMD visualization software and its utility programs. Constant pressure and temperature (NPT) classical MD simulations were carried out using the NAMD program. Force Fields were parameterized using the CHARMM27 FF for all the simulations carried out. Standard minimization using explicit solvents in a TIP3 water model was performed in an NPT ensemble. Short-

range, non-bonded interactions distances were cut off at 12 Å. A further annealing step was performed raising the temperature from 60 to 310 K in a 0.06 ns of simulation. As for the minimization step, 12 Å was used as a cut of distance to short-range non-bonded interactions. The system was then equilibrated for 2.5 ns simulation using similar parameters.

After equilibration, ten 150 ns MD simulations were carried out totaling more than one microsecond long timescale. The temperature was maintained at 310K using Langevin dynamics at a pressure of 1 bar. In this case, both short- and long-range electrostatics were taken into account. The short-range non-bonded interactions were taken care of by setting the cut-off distance to 12 Å. While the particle-mesh Ewald (PME) was used to treat long-range electrostatic interactions. Integration of the equation of motion was achieved using the r-RESPA multi time step scheme and was used to update short-range interactions every 1 step and long-range interactions every 2 steps. A time step of 2 fs was used for all the MD simulations where all the atoms (i.e., the backbone) were unconstrained to allow for free-energy simulation.

4.4.4 Calculations of Ligand-Binding Affinities

In this subsection, the molecular mechanics/Poisson-Boltzmann theory approach (MM/PBSA) was used to calculate the strength of binding between embelin and BIR3 during simulation. The *g_mmpbsa* package[117] which utilizes Gromacs was used for this purpose. Energies obtained using this method are approximate, but the implementation is well suited for relative energies especially when dealing with instances where different conformations or clusters of a ligand are involved. As such, snapshots from conformers obtained from MD simulations with different poses can be energetically compared. Additionally, the obtained energies can be compared with experimentally observed results.

The total energy can be broken down into individual contributing energies which are the Vacuum energies E_{MM}, the polar energy G_{polar} and the nonpolar energy $G_{nonpolar}$. For this purpose, snapshots from the trajectory were taken each 1 ns interval giving rise to 150 snapshots per simulation. The E_{MM} energies were calculated using the Lennard Jones and Coulomb potential. In the case of $G_{polar,}$ a box containing an initial

grid was created using the extreme coordinates of the complex in each dimension. A 1.5-fold factor was used to expand the box to obtain a coarse grid. The grid was then expanded by adding 5 Å to obtain a further finer grid-box. To neutralize the system, a concentration of 0.15 M NaCl with a radius of 1.81 and 0.95 Å for chloride and sodium respectively was employed. The energies of the $G_{nonpolar}$ were calculated using a solvent accessible surface area (SASA) model with a probe radius of 1.4 Å, gamma value of 0.02267 kJ/molÅ^{-2} and an offset of 3.8428 kJ/mol. These values are the default in the package and the $G_{nonpolar}$ equation is given by:

$$G_{nonpolar} = \gamma A + b \tag{71}$$

Where, γ is the surface tension related coefficient and b a fitting parameter.

These three terms were calculated for each residue and summed up to give an average binding contribution. Water as a solvent was used with a dielectric constant of 80 and the solute had a dielectric constant of 1. The average binding energy for each of the snapshots were calculated using equation:

$$\Delta G_{bind} = G_{BIR3-embelin} - (G_{BIR3} + G_{embelin}) \tag{72}$$

4.5 Results and Discussion

4.5.1 Binding BIR3: The Role of Embelin and Its Binding Pattern

The binding patterns between embelin and BIR3, ligand-protein binding energies and binding mechanisms from the MD trajectories are presented in this section. From the docking procedure, the top 3 docked poses were used to run initial 50 ns simulation from which the final coordinates were picked for the next ten 150 ns replicate for each pose. The obtained trajectories were subjected to energetic and structural analyses to answer three major questions.

(i) Where is the preferred binding pocket/site of BIR3 by embelin?
(ii) What are the modes of binding manifested by embelin within this pocket?
(iii) What is the binding energy within this site?

4.5.2 Rigid Docking

The resulting complex yielded nine of the ten poses requested and the results were ranked based on the electrostatics, van der Waals, and the binding poses as shown in Figure 18. Cluster and hydrogen bond analysis was performed in a post docking method using Chimera. From the table, the top three molecules had promising van der Waals energies and were chosen. A further hydrogen bond analysis reveals that one of the embelin conformers binds to Glutamine 221 (Q 221) of BIR3 and Alanine 1 (A 1) of SMAC at distances 2.324 Å and 2.583 Å respectively. The binding forms a bridge-like orientation suggesting that embelin could displace SMAC in a competitive binding mechanism to BIR3 as is the case of caspase (see Figure 18). A previous experimental nuclear magnetic study (NMR) study by Nikolovska *et al* found embelin to bind BIR3 at Tryptophan 323 (W 323) and Tyrosine 324 (Y 324) and serve as the crucial binding pocket. Additionally, it was noted that other residues also play important roles in the binding. Furthermore, embelin was found to bind BIR3 on the surface groove which is similar to how SMAC binds BIR3. Therefore, a further step was taken to use the top docked poses in MD simulations to identify a stable pocket where embelin would prefer to bind. The total binding energy as depicted by the Grid_score was used to choose the top three poses for further calculations.

Table 8: Grid_score binding energies as obtained from docking procedure obtained using DOCK 6.

Pose	vdw	es	Grid_score
1	-29.265	-0.537	-29.802
2	-24.477	-1.387	-25.865
3	-15.779	-0.264	-16.044
4	-11.237	-1.810	-13.0476
5	-7.297	-0.823	-8.119
6	-7.385	0.079	-7.306
7	-8.437	1.636	-6.801
8	284.687	0.511	285.198
9	31083.949	-0.466	31083.4824

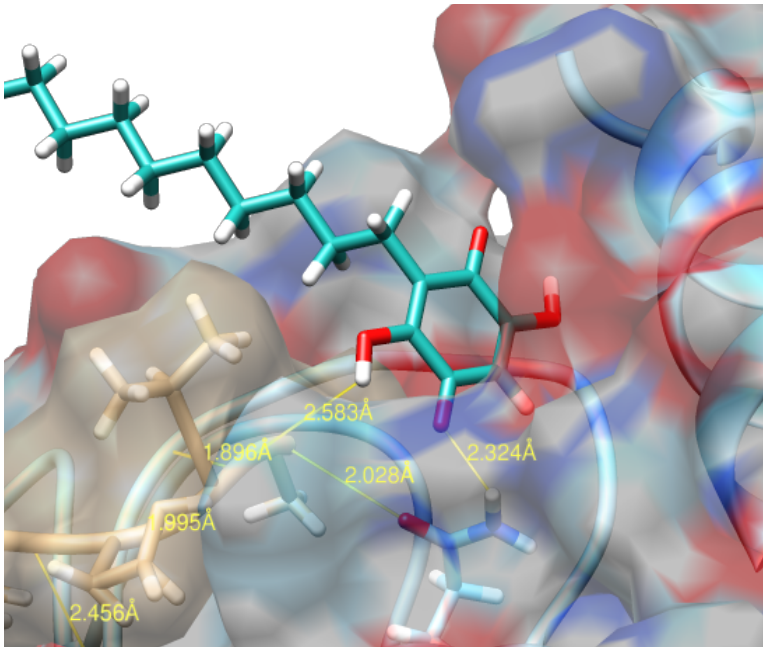


Figure 18: Embelin's bridge-like binding which served as an initial pose for further MD calculations.

4.5.3 Molecular Dynamics

4.5.3.1 Characterization of The Binding Site

Preliminary root mean square (rsmd) and fluctuations (rmsf) analysis of the obtained trajectories was carried out. Various software and packages were used including VMD[118] and Bio3d library[119] to cluster the ligand conformers based on the orientation within the binding pocket. Snapshots of each system were taken at an interval of 1 ns and the stability of the simulation ascertained using rmsd. The first step was to align the trajectories to the initial frame of reference and then subject them to periodic boundary corrections using the *pbc* utility in VMD. Every residue was wrapped around the protein backbone within a unit cell and the obtained trajectories used for further analysis.

Using the data above, the binding site has been characterized. This site consists of the most stable pocket where a large number of clusters prefer to bind. The residues found within the site are 37 in total ranging from K 224 (LYS 224) to E 259 (Glu 259), namely: LYS 224, TRP 225, TYR 226, PRO 227, GLY 228, CYS 229, LYS 230, TYR 231, LEU 232, LEU 33, Glu 234, GLN 235, LYS 236, GLY 237, GLN 238, Glu 239, TYR 240, ILE 241, ASN 242, ASN 243, ILE 224, HSE 245, LEU 246, THR 247, HSE 248, SER, 249, LEU 250, Glu 251, Glu 252, CYS 253, LEU 254, VAL 255, ARG 256, THR 257, THR 258, Glu 259. The site lies at the α-helix section of the protein. Compared to the experimental residues suggested by Nikolovska-Coleska and co-workers, it slightly deviates as will be explained in later subsections. The experimental residues correspond to Tryptophan 225 (W 225) and Tyrosine (Y 226) in the protein used in this work.

The clustering also reveals that embelin prefers to bind BIR3 in two major orientations as can be seen in **Figure 19** labelled **A** and **B**. Pose A and B lie at an angle difference of approximately 180° and clusters falling within these poses were categorized into two groups. To arrive at these definitions, a distance of 3.5 Å and orientation criterion of 120° was used to define the stability of each ligand within the site. Stability was defined when the ligand remains in one orientation for at least 20ns of simulation time in each trajectory. In case one pose is switched before this time, it was dropped. Orientations A and B were stable for 55 ns and 43 ns of simulation time respectively. This pattern was observed in replicated simulations.

The other important factor considered was the deviations of individual trajectories from the starting structure where a distribution density was extracted are shown in Table 9.

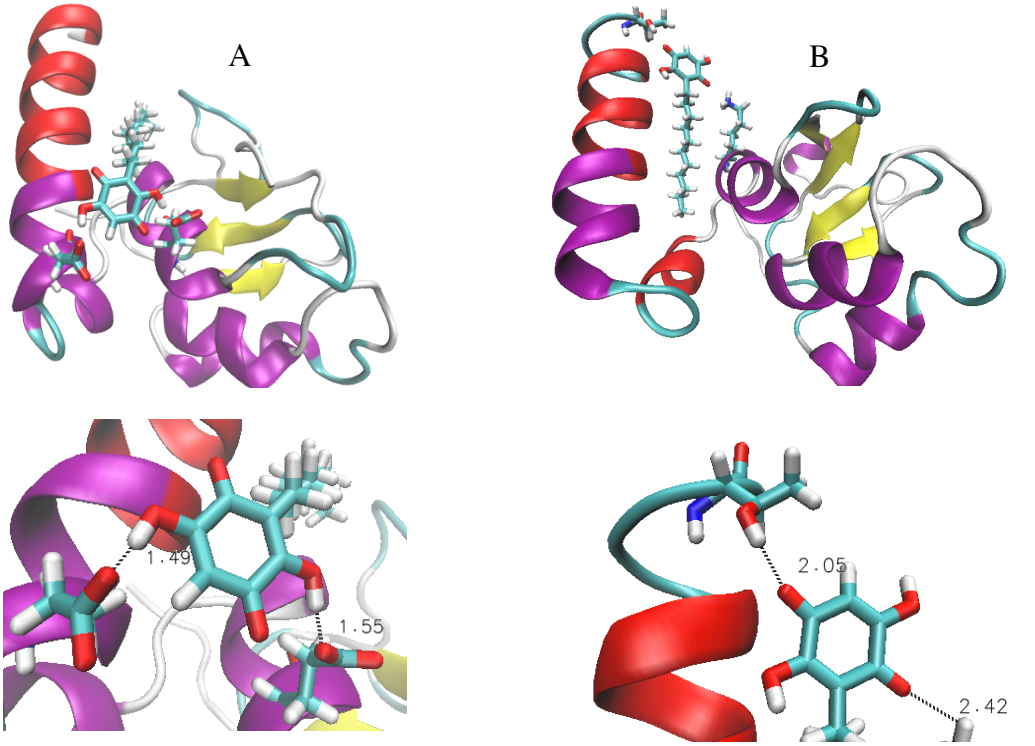


Figure 19: (Top) Structural orientations of embelin in the alpha-helix BIR3 binding site. (Bottom) Binding pairs distances of embelin in mode one and two. The criteria that were used in defining the modes followed a convention limiting of each distance to 3.5 angstroms.

Table 9: Maximum density distribution of ligand-protein binding for each simulation.

Simulation	1	2	3	4	5	6	7	8	9	10
Distribution density	4	4	5	5.5	4	4.5	4.5	4.5	3	3

The distribution varies approximately between distances 2 and 4.5 Å from the lowest to the highest deviation for all the trajectories. However, there are distances that are less than 2 Å which form the basis of the stable binding within the site as will be seen later. These values confirm that indeed embelin bind BIR3 within a stable site. Advanced calculations such as proton transfer, electron transfer, and or charge transfer reactions can therefore be performed using snapshots obtained from this analyzed data and forms the basis of the next chapter.

Particularly, proton transfer reactions as well as singlet oxygen generation ability are explored in detail.

The measure of the deviation of embelin from the initial bridge-like pose obtained in docking to the stable site is approximately 3 Å. Since the N-termini binding of the SMAC/DIABLO (the antagonist) was mimicked in the docking studies, the deviation suggests that embelin could prefer a different binding mechanism. In other words, embelin and SMAC do not bind with the same pattern according to the analysis conducted so far. Since it is known that BIR3 can bind both caspase and SMAC in a competitive process, it is important to investigate the nature of embelin-BIR3 binding as well as the mechanisms involved. Ideally, such mechanisms should resemble biological functions performed by SMAC and caspase to reproduce the desired effect of deactivating BIR3 anti apoptotic functions. These mechanistic binding modes forms the basis of the next analysis section.

4.5.4 Embelin-BIR3 Binding Modes

Preliminary results discussed above has shown that there exist at least two orientations that embelin can bind to BIR3. In this subsection, effort is made to investigate the binding modes of embelin in BIR3 using the Markov state model technique. Once the modes are identified, the underlying mechanism can be mapped out to check whether indeed embelin can replicate SMAC and or caspase bindings to BIR3. MSM models are known to provide useful insight into certain slow processes which may not be possible to capture using classical MD. These properties include rates in kinetics and allostery in protein sites which have been successfully reported using MSM in ligand-protein simulations[106,109] and in hidden sites.[120] The main framework involves application of variational techniques that reduce high dimensional data to low dimensions hence enabling identification of collective variables which represent slow motion processes. Captured slow processes within MSM data tend to display dominant phenomena of interest and therefore contain important mechanistic information obtained from short MD simulations. More importantly, the variational technique employed in this section utilizes a time-structure-based independent component analysis (TICA) that maximizes autocorrelation of transformed coordinates from high to low dimension.[111,121,122] Using high dimensional data from MD simulations as input, TICA solves for the eigenvalues and eigenfunctions of the

Markov operator and the solutions obtained used to construct Markov state models. From these models, adequate information can be obtained by identifying metastable states which when decomposed further reveal the modes involved in the binding. The PyEmma package[123] together with Pycontact was used to achieve the MSM analysis presented in this subsection.

MSM analysis of the phase space was majorly based on conformational signature characterization of both the ligand and the protein. The technique can also reveal information on the kinetics of the binding, however in this work, kinetics analysis was not pursued further due to lack of experimental data to use as a reference. As such, it was considered to be beyond the scope of the study. The obtained kinetic information was quantitatively ignored and was only used as a guidance for the transition path of the metastable states of interest. Specifically, the kinetic flux was used to aid in analysis of the dominating transition paths of the binding. Three parameters were used for the conformational analysis namely, hydrogen bonds (HB), the angle of orientation *w.r.t* the active binding site, and RMSD of the ligand. The choice to include HB analysis was informed by the fact that from previous chapter, embelin was seen to undergo intramolecular proton transfer reaction. When embedded within the protein, it is also possible that embelin could undergo intermolecular proton transfer analogous to the intramolecular process, which modulates its biological activity.

The stated facts combined with the knowledge of the binding site characterized above were applied to define parameters, also known as *features,* for MSM creation. Hydrogen bonds were labelled using an initial distance metric of 3.5 Å for normal hydrogen bonding pattern as well as 4.5 Å for "long-range bonding" and were included because most distribution density fell within this distance (refer to section 4.5.3). The second parameter criterion was the orientation of the ligand within the site. An angle difference of at least 120° was used to cluster the binding orientation of the ligand when switching from one orientation to the other (see Figure 19). Consequently, two MSMs were created based on these *features* and were labelled MSM one and two respectively. The naming convention corresponds to the orientations labelled **A** and **B** respectively in section 4.5.3. More importantly, the

ligand orientation parameter was considered as the leading criterion for choosing the modes followed by the hydrogen bonding distances. This is to mean that a pose, regardless of its hydrogen bond pattern or proximity to the stable binding site, was used to group modes as long as it deviated 120° from the defined binding site. An additional filter of the HBs was applied to cluster ligands into meaningful ligand-residue pair distances for the purpose of HB analysis. After obtaining the MSMs, four coarse grain macrostates were obtained to group each microstate into respective metastable states. Consequently, four representative structures were generated and stored in a file format which was then used for further ligand-residue analysis. A coarse-grained model of the MSM was used to obtain Hidden Markov State Models (HMSM) for visualization of the transition path theory.

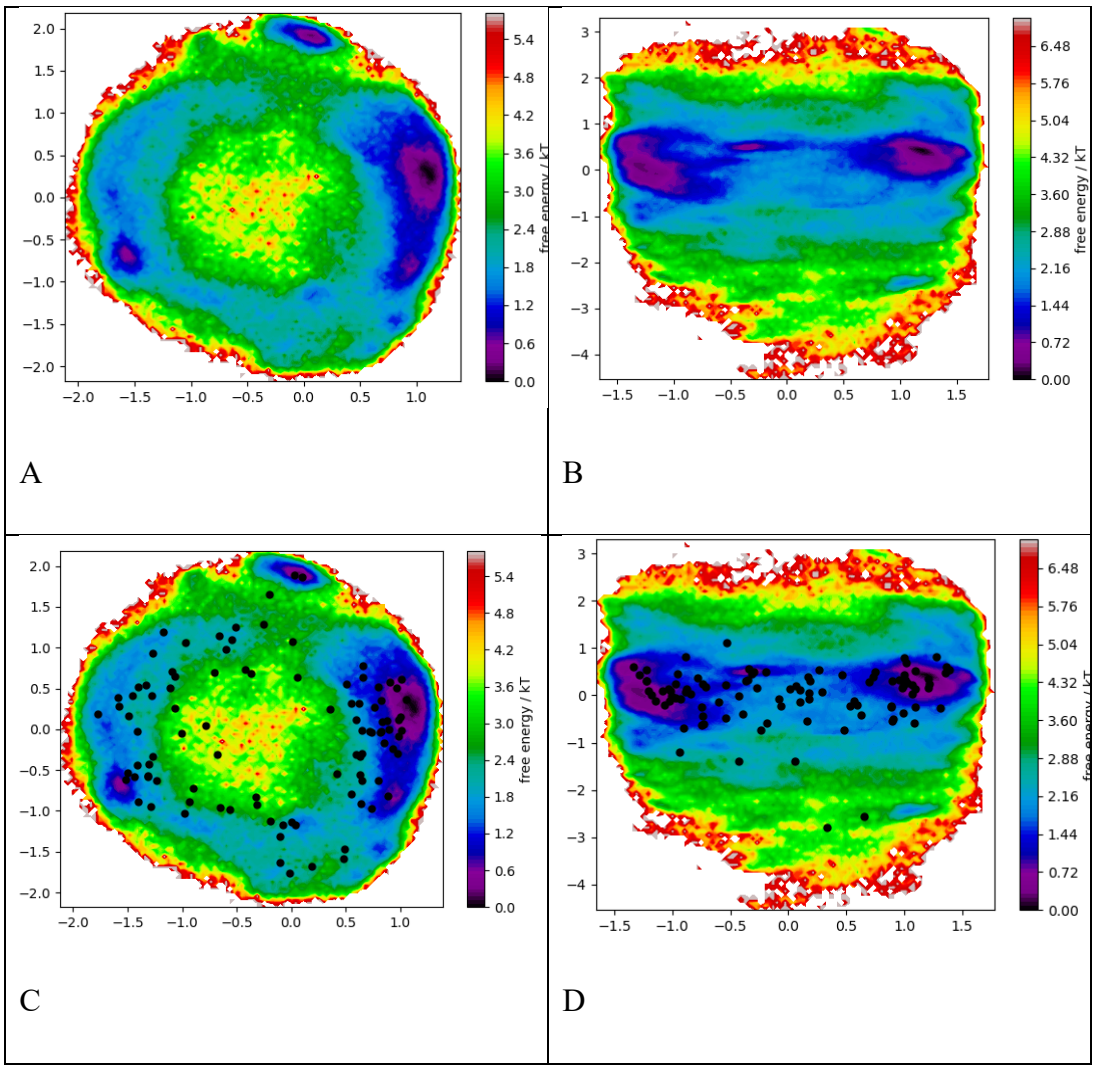
MSM Discussion

The aim of MSM construction is to obtain the collective variables that best characterize the rare-event transitions within the analysis.[123] There are a number of steps involved in achieving a meaningful MSM model namely, featurization, dimensionality reduction, clustering, model construction, and model validation. The MSM results will be discussed in this order for clarity purposes.

Featurization: The chosen distance *features* resulted into 8505 dimensions for each MSM category. The closest-heavy atom groups were used to select the ligand-residue pairs satisfying the distance of interest. The first 10 features are:

['RES_DIST (closest-heavy) LIG261 - ALA1',
'RES_DIST (closest-heavy) LIG261 - VAL2',
'RES_DIST (closest-heavy) LIG261 - PRO3',
'RES_DIST (closest-heavy) LIG261 - ILE4',
'RES_DIST (closest-heavy) LIG261 - ALA5',
'RES_DIST (closest-heavy) LIG261 - GLN6',
'RES_DIST (closest-heavy) LIG261 - LYS7',
'RES_DIST (closest-heavy) LIG261 - SER8',
'RES_DIST (closest-heavy) LIG261 - Glu9',
'RES_DIST (closest-heavy) LIG261 - PRO10'].

Dimensionality reduction & clustering: Having obtained the features, a transformation from high to low dimension was performed. The transformed features using TICA resulted in 4310 dimensionally transformed space and were used to obtain a kinetic map at 95% kinetic variance for both MSMs. Consequently, using *k-means* initialized by *k_means*++ initializer, 547 cluster centers were used to generate TICA coordinates. These TICA coordinates gives the so-called "*energy-sinks*" and can be plotted for the first two independent coordinates. Figure 20 shows the minima of the "sinks" in the ensemble as captured by free energy plot mapped onto the first two independent components obtained using TICA. Four populated metastable states with three distinct resolved processes can be interpreted as minima within the plot A and two minima in B which correspond to trajectories for MSM one and two respectively. Consequently, the clusters can be visualized as shown in Figure 20 **C** and **D**.



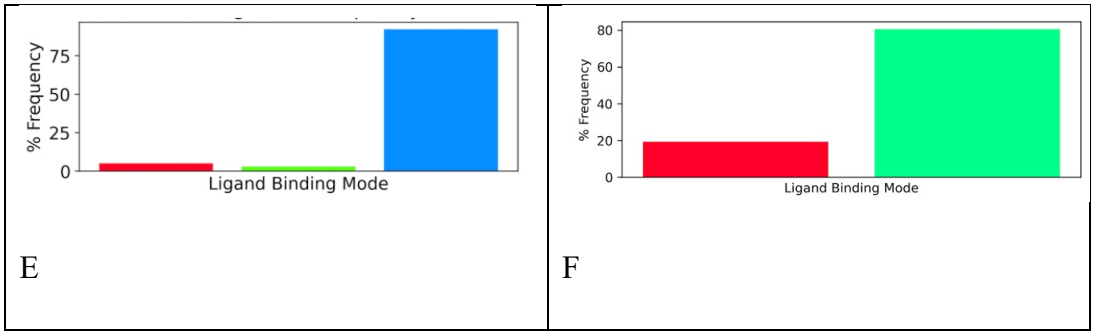


Figure 20: Maximum-likelihood and free energy MSMs of embelin-BIR3 projected on the first 2 TICA components. (A) represents mode one while (B) represents mode two. The microstates were generated using a set of parameters (orientation and hbond parameters).

Model creation: Before creating the model, I will discuss the idea of implied time scale (*its*) which is central to MSM. In simple terms, *its* refers to the relaxation timescales of a molecule implied by a Markov model transition matrix estimated at a lag time τ. This parameter is a physical property of the simulated system while τ is a model parameter and are hence independent of each other. Spectroscopic techniques can be used to obtain *its*[124,125] but the parameter can be also obtained using the eigenvalues of the Markov model via the relationship,

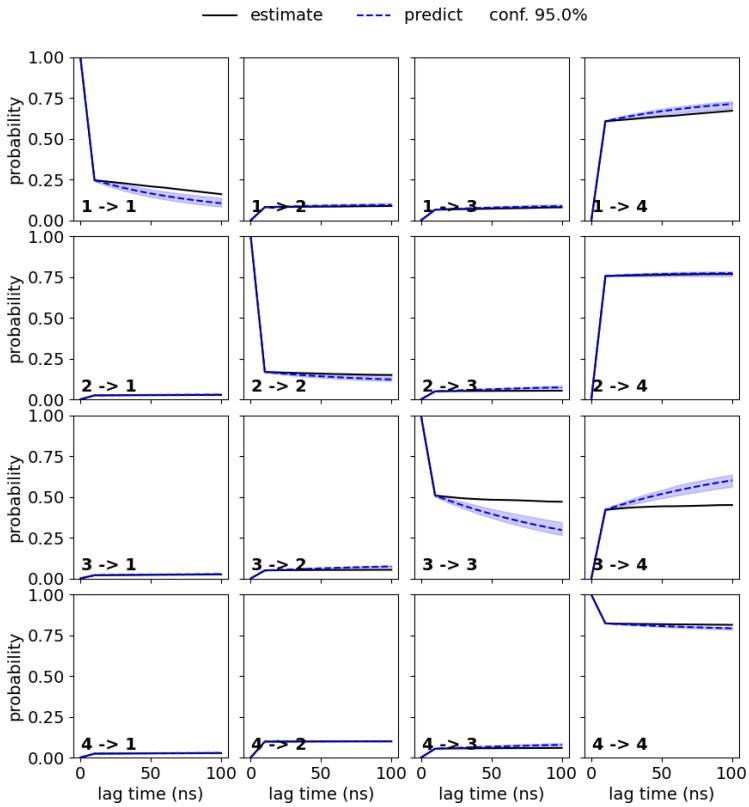
$$t_i = \frac{-\tau}{\ln|\lambda_i(\tau)|} \quad .$$

Where, λ_i is the eigenvalues of the i^{th} state, τ is the lag time given by $\tau = n_{lag}.t_{stride}$, here n_{lag} is an integer (always given as a list) range and t_{stride} is the trajectory time-stride.

In other words, to create our MSM model, we need to choose an appropriate timescale where our data is properly discretized and use the chosen time scale (*lag time* in this case) to extract slow processes information. The so-called *lag-time* is our model parameter equivalent to the explained *its* above for the discretization purpose. The lag times used for estimating the Bayesian hidden Markov models ranged between 1-200. In an ideal scenario, a good indication of a well discretized phase space is a constant graph of implied relaxation time scales as a function of lag time as shown in Figure 22. This independence arises from the fact that MSM is memoryless and the transition matrix count of the current state does not depend on the previous states.[106] The nature of the relaxation timescale shows that the features and consequent discretization meets

the quality threshold of a model that can describe the underlying slow processes. Convergence in the curve (Figure 22A) is achieved at around 20 timesteps for both cases and a $\tau = 20$ was therefore chosen for further analysis. The model that was used to create the MSM is the Bayesian model which incorporates the transition matrices. Similar to the MSM *its*, the Bayesian *its* can be obtained within a confidence interval containing 95% Bayesian samples.

Model Validation: The model created using the lag time has to be validated to ensure consistency of the obtained or expected results. All the MSM models used in this case were validated using the chapman-Kolmogorov (CK) test[107], which is a standard practice. This test shows the transition predicted probabilities from the Markov model and the corresponding estimated probabilities using the data. Figure 21 shows the CK data for the two MSM created. The expectation is that both lines would perfectly fit in case of a good Markovian model proving that the model used can predict the slow processes of interest since the two lines are in most probabilities as close as possible. There is near perfect agreement between the estimate and the prediction. Of course, in this case, the model is not strictly Markovian. Such a shortcoming can be addressed using Hidden Markov state models (HMSM) which does not require that dynamics between microstates to be Markovian.[126]



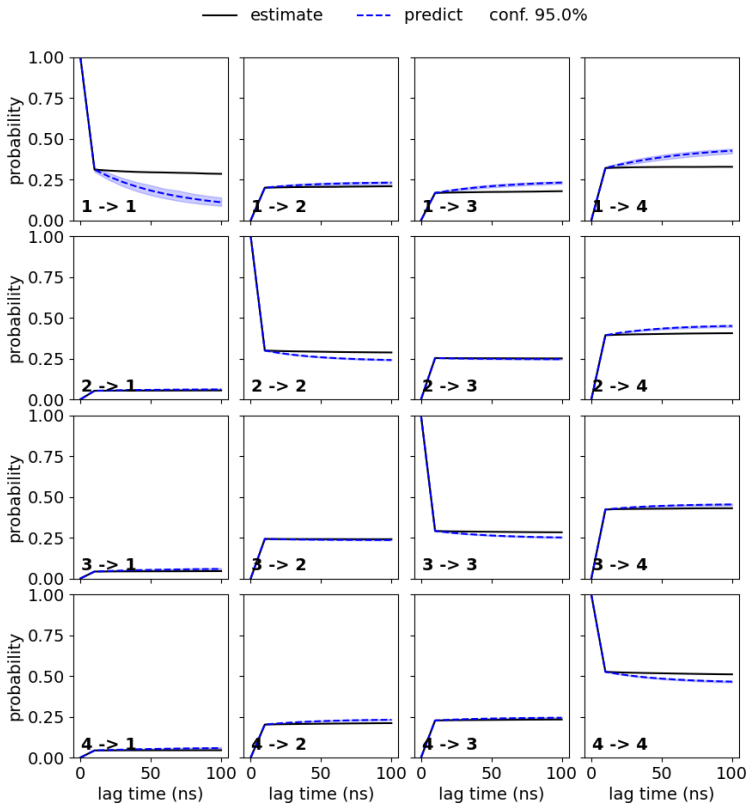


Figure 21: Chapman-Kolmogorov test used to test the MSMs created using a time lag of 20. In both cases, there are three near perfect agreement between the estimate and the prediction.

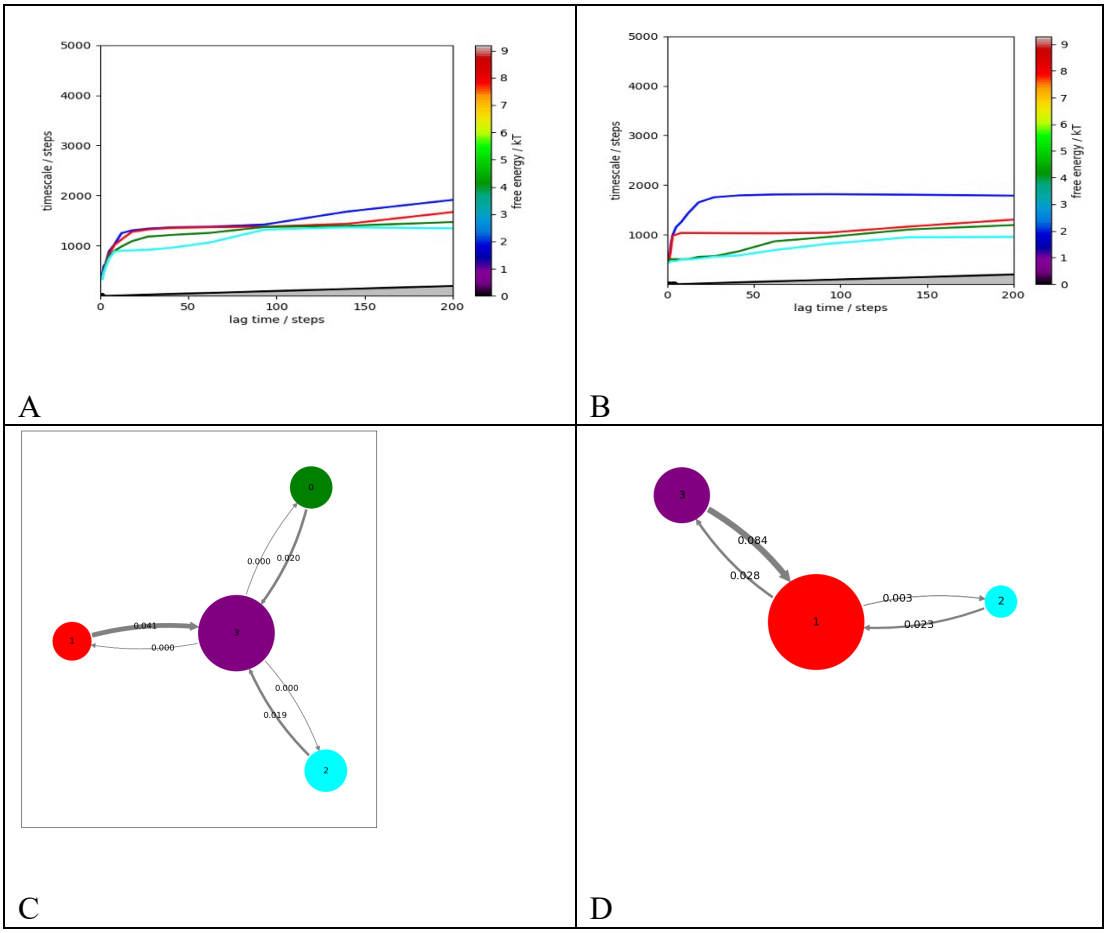
The minima observed in Figure 20, in principle represent the metastable states which are resolved by the model. Such metastable states can be equated to "mechanisms or modes" with which the ligand bind the protein and can be used to map interesting binding pathways, i.e., whether the ligand is either in bound, unbound, or in intermediate states. To map out the pathway, the constructed MSM was coarse grained to construct HMSM. Within HMSM it is easier to use the Transition path theory (TPT) to trace the binding pathway of embelin. [127] TPT provides distinctive map out of the system flux with kinetic details of the systems' pathway representing each "binding or unbinding" state. In simple terms, TPT tells us which route is followed by the ligand to the binding site.

The advantage of HMSM is that they are robust and would overcome poor discretization problems in cases where the process is not strictly Markovian . Additionally, HMSM offer a coarse grain of the so-called "*metastable hidden states".* Using the HMSM, representative structures which revealed critical slow processes were obtained and will be analyzed in the next sub section.

Figure 22 C shows the first four-state HMSM flux of the coarse grained MSM. Three unbound states (red, green and cyan) communicate with the center state (purple) which represents the bulk of bound state. These unbound states do not, however, communicate with each other indicating distinct intermediate states where the ligand "*hit*" on its way to the binding site. Furthermore, the binding is not reversible i.e., there is no transition from the bound state to the unbound state given the zero committor values. This can be clearly seen in the representative structures shown in Figure 22E where each pathway has been mapped to the representative structure using the same colors. Our main interest is the purple structures which lie within our previously categorized binding site. The green and cyan structures are however close to the bound purple and could represent a near intermediate states before the ligand finally moves to the binding site.

In the case of the second HMSM, two clear processes are resolved. In this case, we have 3 metastable states with the purple (unbound) intermediate state moving into the red (bound) state. Another state (cyan) can be seen although its commitor values are low. In fact, the representative structures for this state are a huge distant from the

"*active site*". As discussed earlier and evident from Figure 22 F, the overall binding for this mechanism is weak. There are few clusters within the representative structures. Furthermore, the unbound intermediate lies at a distant from the binding site. Surprisingly, HMSM reveal the correct orientation of embelin within this second HMSM which differ from the first one thereby confirming that even though weak, embelin also bind BIR3 *via* a hydrogen acceptor mechanism. These results further confirm that embelin prefers to bind BIR3 within the initially characterized site. From this analysis alone, one can conclude that both hydrogen donating and accepting mechanisms play a critical role when embelin bind BIR3. This is not surprising and is expected owing to its structural arrangement. The residue-pairs involved in the binding is explored in the next subsection.



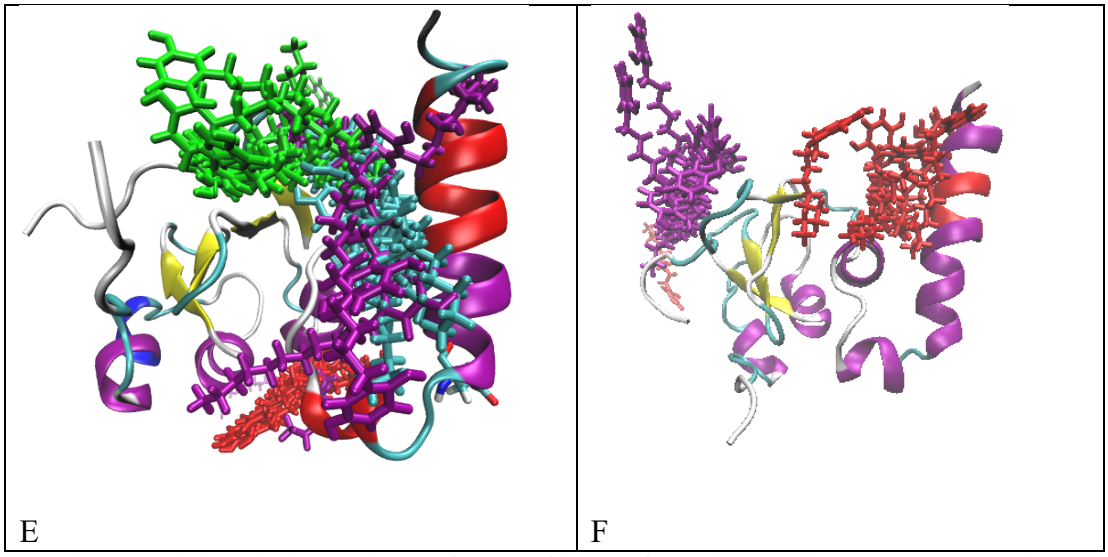


*Figure 22: (Top **A** and **B**) Implied timescales as a function of lag time for mode one and two MSMs respectively. Transition path flux of mode one (C) and mode two (D).*

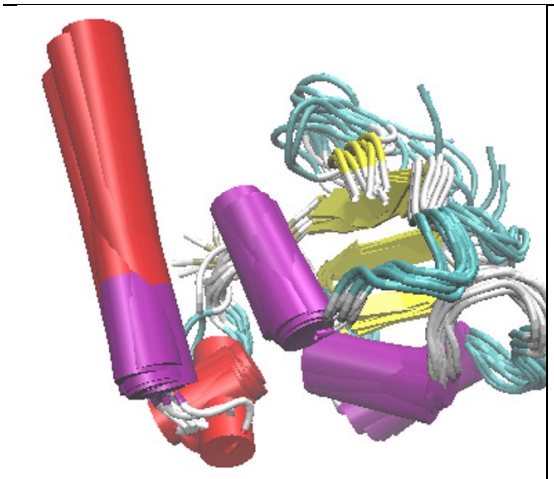
4.5.5 Residue-Ligand Pairs and Poses from Representative Structures

Conformational signatures evaluated from the metastable states highlight the structural differences of ligand-protein dynamics and consequently the underlying binding mechanism. The dominant factor considered here is the representative structures obtained from the most populated metastable states in the two MSMs representing stable binding modes. The first analysis consists of the structural differences in the BIR3 backbone. Secondly, the interactions between embelin and the residue chain pairs are analyzed. The assumption made in the case of backbone analysis is that stable binding is indicated by a relatively immobile backbone while a mobile backbone is an indication of unstable binding given the fact that BIR3 is not a folding protein.

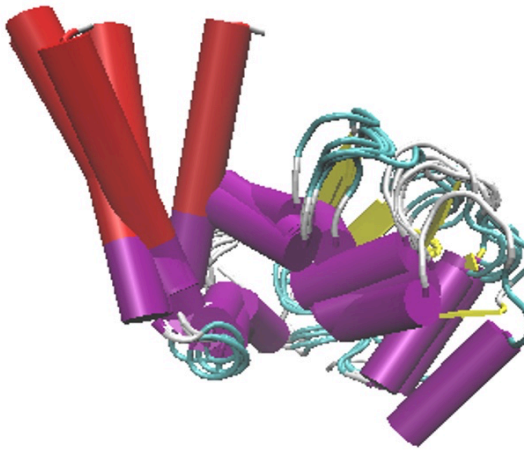
Protein Conformation: The associated TICA coordinates (trajectory frames) data was assigned to the nearest *k_means* cluster centers obtained using Perron-Cluster Cluster Analysis (PCCA)[128] resulting in classification explained above. The rigid and compact nature of the ∝-helix in BIR3 (Figure 23 A) shows a conformationally stable backbone for the longest living metastable state in MSM one. Within this metastable state, the ratio of ligand clusters is higher in comparison to those unbound. In fact, it is within this metastable state that favorable poses are observed. Distances between embelin and residue pairs range between 1.55 to 4.7 Å (see Figure 23 C). The poses within these distances were used in the next chapter for advanced hybrid QM/MM

calculations as will be later explained. This therefore qualifies MSM one to likely represent the dominant binding pattern through which embelin binds BIR3. The other two metastable states do not contain adequate representative structures with definitive bindings. Instead, they have mixed poses and orientations with very weak binding i.e., the distances between the ligand and the side chains are quite large. HMSM also show this trend as can be seen in the green and red ligand clusters in Figure 22C. A quantification of the binding pairs will be clearly highlighted in the proceeding section when the analysis of the ligand structures is presented.

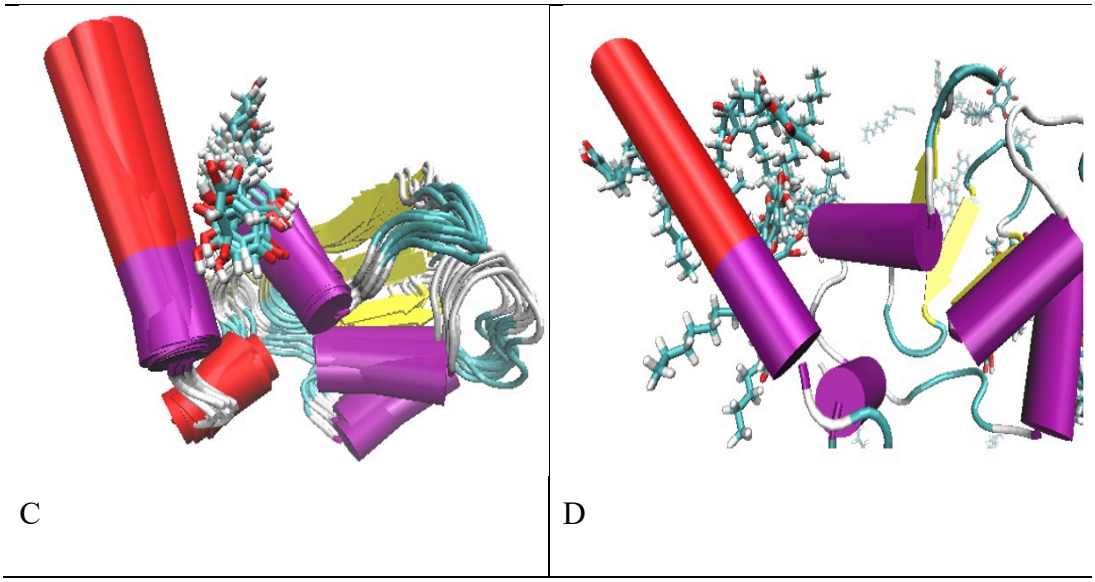
On the other hand, an inward tilting of the $\propto -$helix (Figure 23B) is observed for the case of MSM two with several ligand outliers. The ligands in this mode do not lie within the binding site with a correct pose but do fall within the orientation parameter hence satisfy mode two classification (Figure 23 C). Aside from the fact that there are fewer clusters close to the site, the nature of the binding is also different. Dominant interactions are of the hbond acceptor type where embelin prefers to bind *via* its *para-quinonic* functional groups. However, as will be seen in the ligand-residue pair analysis, it is indeed a weak binding. The second metastable state within MSM two does not also offer any meaningful information as there are mixed orientations. Further binding information in these states are revealed using ligand-residue pair analysis.



A



B



C D

Figure 23: Mode one and two macrostate obtained from embelin-BIR3 MD trajectories. These are the dominating binding metastable states which represent the two leading binding mechanisms. i.e., hydrogen donating (left) and hydrogen accepting capabilities of embelin.

Ligand-residue pairs: This subsection will first begin with a discussion of the MSM ligand pairs analysis followed by the HMSM counterparts. The binding mechanism is based on the pair interactions between the ligand and the protein side chains within the first metastable state in MSM one. PyContact was used to analyze the coarse-grained representative structures of the metastable states saved as individual output files. Similar distance parameters employed in the MSM creation are used here as well. The binding percentages for each respective residue is shown in **Figure 24**. In metastable state one, Glu 239 and Glu 220 pairs make the bulk of interactions with the ligand hitting a threshold of above 50% binding. In comparison to the other two metastable states which are less than 10%, the pairs form the basis of the first binding mechanism between embelin and BIR3. The significance of the binding is explored below, however, it is clear that within these interactions, embelin acts as a hydrogen donor while the Glu residues are the hydrogen acceptors. This is in agreement with known quantum mechanical facts about embelin where it prefers to donate its *para-hydroxyl* protons intramolecularly. The remaining two metastable states contain mixed pairs which satisfies both mode one and two classifications. In a nutshell, a conclusion can be drawn that embelin binds through hydrogen donating mechanism.

In the case of mode two, the binding strength is significantly reduced with LYS 224 providing the highest Hbond count at 7%. This mechanism has a hydrogen donating pattern and can be classified as the second mechanism. It is however very weak since

the other residue pairs consist of only 2 % Hbond counts making the mechanism less significant.

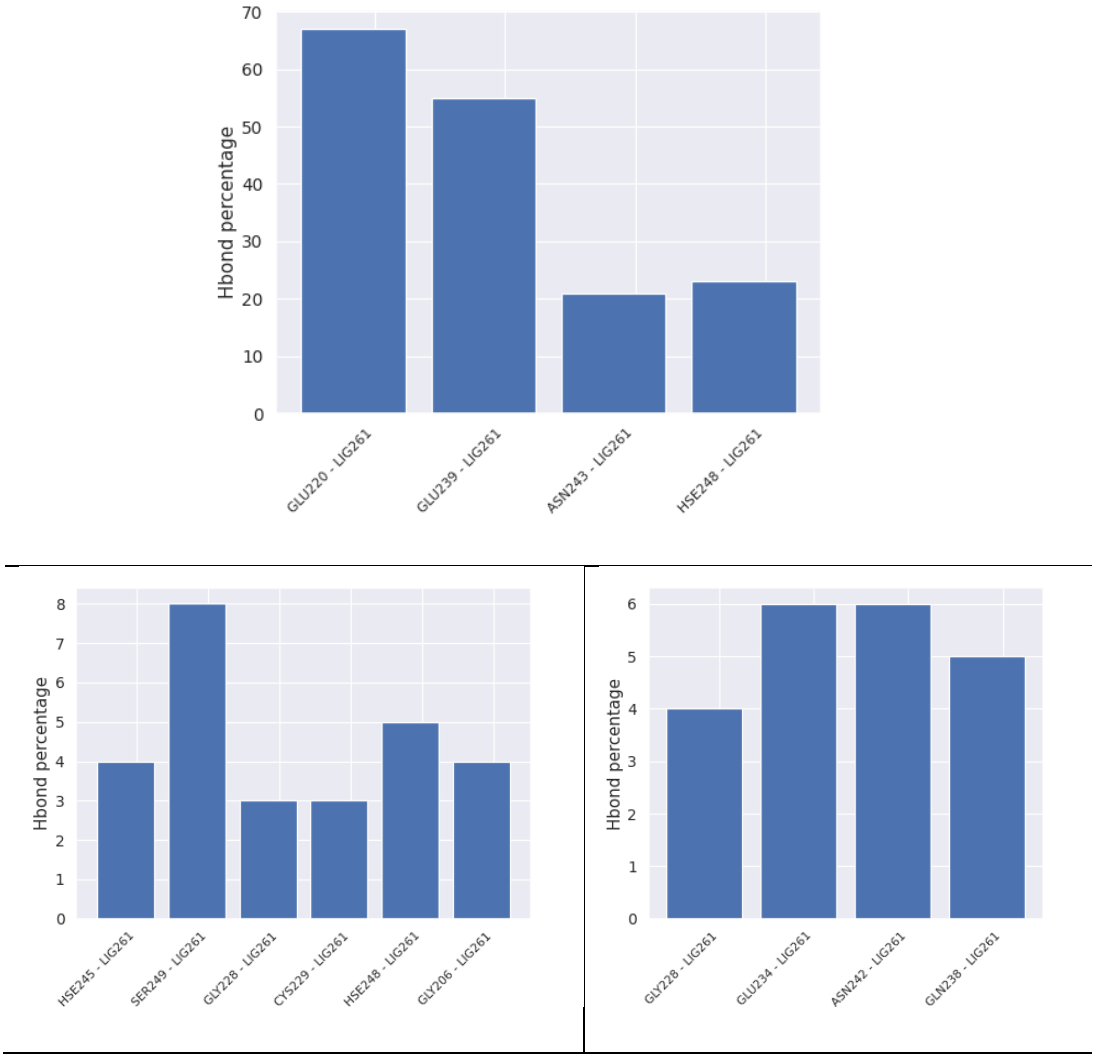


Figure 24: Ligand-residue pairing obtained from MSM one. The top metastable state one corresponds to the highest binding with hydrogen percentage above 50%.

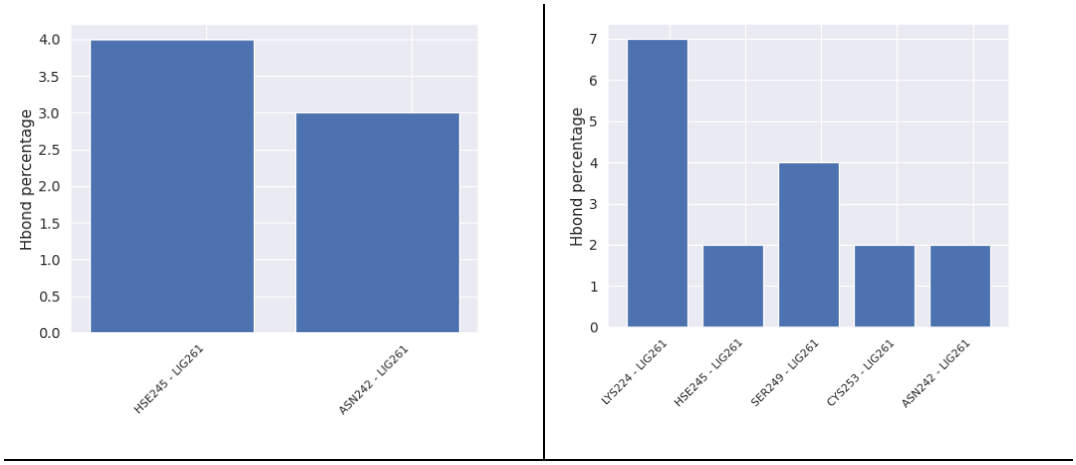


Figure 25: Ligand-residue pairing obtained from MSM two.

Analysis of the distances within the dominating hydrogen donating mechanism shows that embelin strongly bonds at short distances which could result in either a proton or hydrogen transfer. These types of reactions are known to trigger electron transfer or chain reactions within the protein. Furthermore, short hydrogen bonds in proteins may facilitate molecules to undergo *trans-cis* isomerization leading to proton transfer on exposure to light.[129] Possibilities where proton transfer reactions can be realized as a result of favorable hydrogen bond which could be useful in PDT application areas since embelin's ability to undergo *isc* is facilitated by a similar reaction as revealed by quantum mechanical calculations in the previous chapter. These possibilities can be achieved using advanced hybrid QM/MM calculations and will be explored in the next chapter.

HMSM pairs revealed a similar binding pattern albeit with different percentages (Figure 26). The highest scored residue in terms of hydrogen percentage is Hse 245 (H 245) at just above 50%. The Glu residues are quantitively lower as they are identified by 25 % and 18% for Glu 239 (E 239) and Glu220 (E 220) respectively. Interestingly, one of the experimentally highlighted residue Tyr 226 (Y 226) termed by Nikolovska[58] to influence embelin's binding on the surface groove of BIR3 using NMR studies is identified in metastable three. This finding is encouraging since the HB percentage score is quite high hence the results obtained using MSM in this chapter are reliable means and reproduce experimental results. Consistency has also been demonstrated since Glu residue pairs dominating in the MSM previously

discussed are replicated here. Using these findings, a conclusion can be drawn that embelin prefers to bind BIR3 *via* hydrogen donating mechanism. Furthermore, Hse variant as used in NAMD is in a neutral state with a proton at the epsilon position thereby leaving the nitrogen at the delta position to act as an acceptor. These two protonation states are interchangeable in normal biological systems enabling donor/acceptor pair mechanisms supporting our earlier hypotheses.

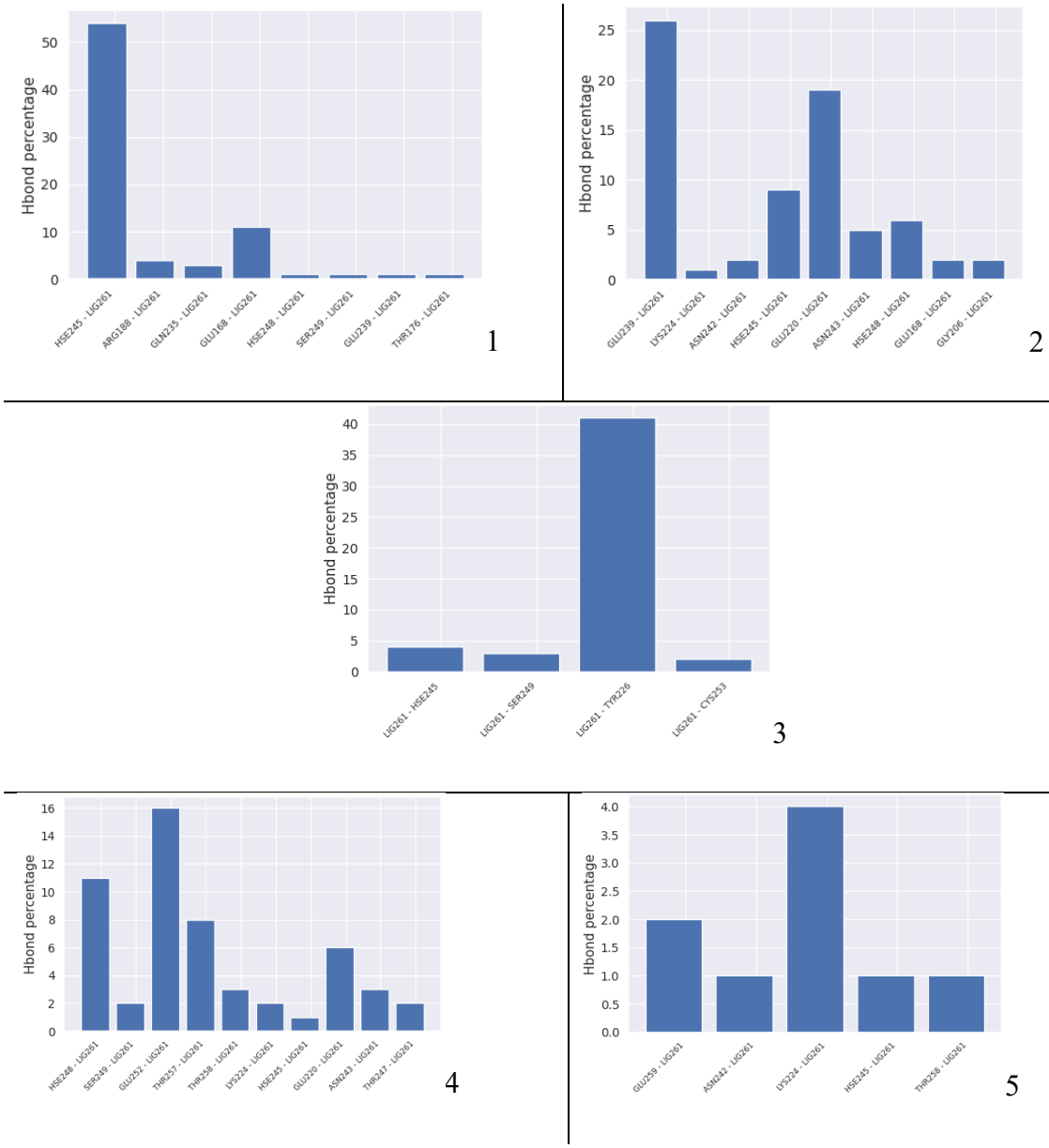
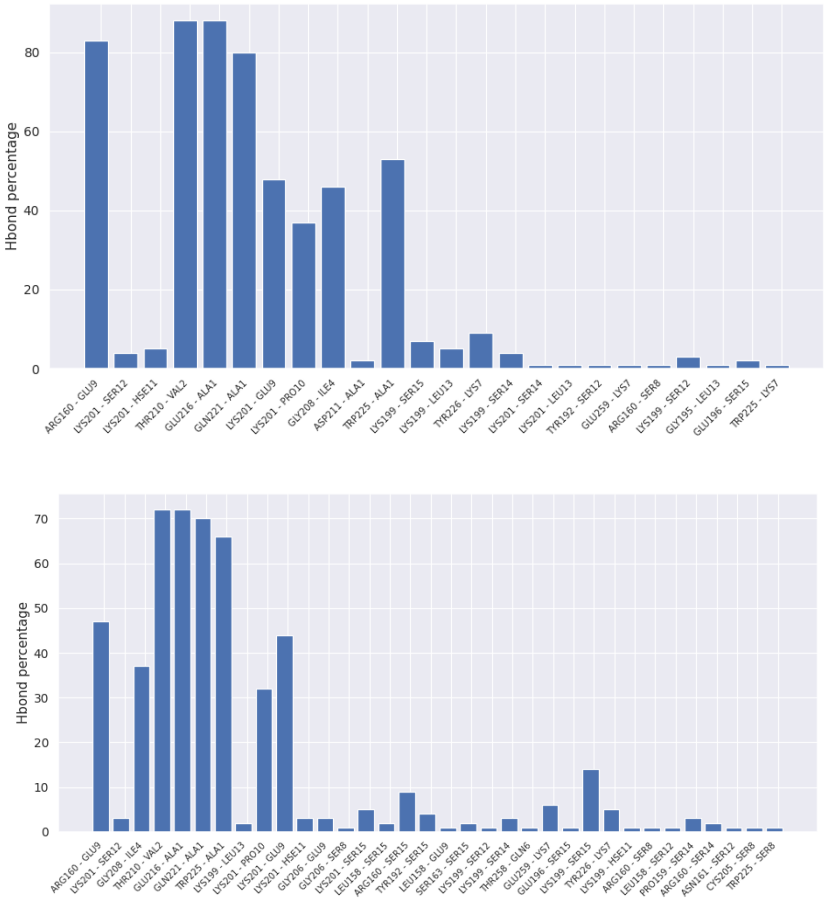


Figure 26: HMSM ligand-residue pairs generated from the coarse-grained MSM model. Graphs 1-3 represents HMSM one and Graphs 4-5 represents HMSM two.

4.5.6 How does SMAC Binding Compare?

A similar analysis applied to SMAC is necessary to compare its binding patterns with that of embelin within these metastable states. As earlier discussed, BIR3 binds SMAC and caspases in a competitive biological binding process and therefore may also prefer to bind the same "*active site"* where embelin binds if at all the pathways are competitive. But this is not the case as SMAC binds at a different site depicted by graphs in Figure 27. The dominating pairs are Arg 160, Thr 210 (T 210), Glu 216 (E 216), Gln 221 (Q 221), and Trp 225 (W 225) HB with the highest bond network. At approximately 80% HB network, the binding can be said to be relatively stronger than that of embelin which is expected since it is the natural antagonist. Notably, SMAC also prefers to bind nearly the same BIR3 residue pairs within all metastable states. This is evident in the strength of the binding replicated in the dominating pairs.



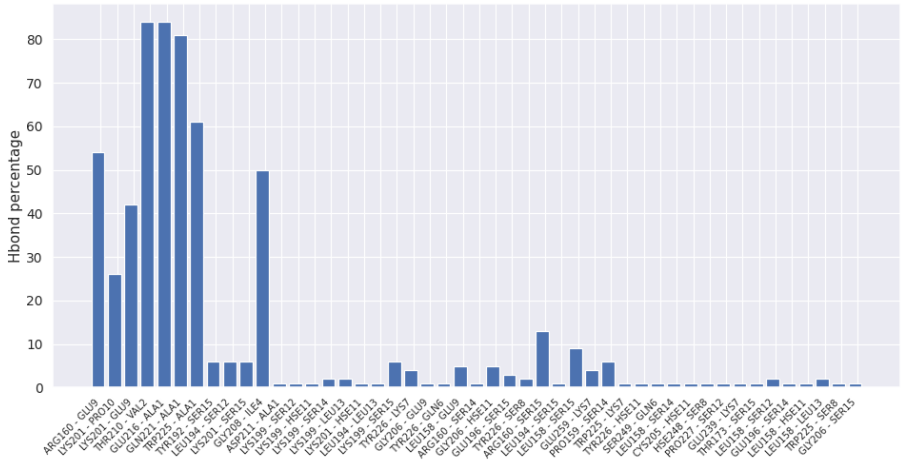


Figure 27: Representative structures of SMAC-BIR3 binding pairs. The percentages are above 80% and are consistent across the metastable states.

4.5.7 Significance of LIG261-BIR3 Binding Mechanisms

As explained in Chapter - 3, the unique structure of embelin enables it to act both as hydrogen donor and acceptor. The structural hydroxyl and carbonyl arrangement have previously been reported in literature to enhance its radical scavenging ability[67] as well as electron transport which is common for quinone molecules.[130] Going by the previous analysis, embelin binds BIR3 at sites where it primarily acts a hydrogen bond donor. In some cluster poses, the distances are as close as 1.49 and 1.55 for both hydroxyls respectively which are within the classification of covalent bonds.

The significance of this binding should however mirror the biological consequence produced by BIR3-caspase or BIR3-SMAC bindings. BIR3 mainly binds caspase-9 which is known to have a substrate preference to aspartate.[131] It is therefore crucial to bind Asp residues within BIR3 since caspase bind them too. In the work presented here however, embelin mainly binds to Glu, Tyr, and Hse. Structurally, Glu and Asp residues are similar which permits their discussion side by side since they are both dicarboxylic and differ only by a methylene. Both acids are also known to be “hard” bases and bind hard metals *via* their carboxylate groups.[132] Biologically, there exist a cleavage relationship where just like in the case of Asp site cleavage, apoptosis is induced through cleavage of non-Asp, Glu sites meaning that in some instances, apoptosis functions can be achieved by cleaving Glu. Furthermore, apparent factors

such as Gene Ontology (GO) enrichments and degrees of conservation that are known to exist in Asp cleavage are observed in the Glu cleavage sites. However, the rate at which these cleavages occur differs in certain scenarios. For example, cleaving Asp is 200 times faster than Glu cleavage when using Asp-N proteases but only two-fold for caspase-3 and -7 substrates. Other researchers have also reported that Glu/Asp pairs play significant roles as binding motifs especially in Grave's disease.[133] These related facts affirm the interchangeability of Asp/Glu binding roles where both would, in exceptional cases, prefer more or less similar substrates and or perform identical biological functions which of course is the case in apoptosis related processes. embelin therefore prefers to bind Glu residues analogous to how caspase-Asp binding. The process occurs on the surface groove of BIR3 which is in agreement with experimental findings promoting cell death in a caspase dependent pathway. The strength of the binding is also important and has been calculated.

4.5.8 Binding Energies between Embelin and BIR3

The energies corresponding to the ligand poses highlighted above are presented in this subsection and serve the purpose of quantifying the strength between the ligand and the proteins' side chain. A plethora of methods are available to achieve this objective including free energy perturbation (FEP)[134,135]and Thermodynamic integration (TI).[136,137] The main advantage of these methods is that they approximate molecular translation over a reversible path from an initial to final state and are therefore considered to be alchemical in nature.[138] Other methods which strike a balance between computation effort and accuracy include molecular Poisson-Boltzmann surface area (MM/PBSA), linear interaction energies (LIE)[139,140], and molecular mechanics Generalized Born surface area (MM/GBSA). MM/PBSA is chosen in this work since it is path independent and give energies of the initial and final states of ensemble while ignoring entropy contributions in a cheap but efficient manner. As a method, MM/PBSA provides intermediate results although the same method has been successfully used to predict drug bindings in dendrimer complexes in another work.[141]

The average relative free binding energy (ΔG_{bind}) of embelin-BIR3 bindings were extracted as a total summation of van der Waals, Polar solvation, Electrostatics, and SASA energies. A high correlation between the binding energies and the stabilities is reflected with mode one having the most stable energy of -30.78 kJ/mol. See Table 11 for individual energies. Clearly, trajectories where embelin binds BIR3 *via* hydrogen-donating pattern has low energies reiterating the fact that this is the main mechanism through which embelin mimic SMAC or caspase deactivation of BIR3.

An effort has further been made to compare the obtained values with experimentally existing data. Here, experimental value is obtained from inhibition constant (K_i) which is equivalent to dissociation constant (K_d) and can be related to the energies using the formula.

$$\triangle G_{expt} = -RT \ln(1/K_i) \tag{73}$$

The experimental data was obtained from a study of human serum albumin (HSA)[142] which reported embelin to bind with an energy of $-5.0 kcalmol^{-1}(-20.9 kJmol^{-1})$. Smac peptide has also been reported to bind BIR3 with an energy of approximately $-2.0 kcalmol^{-1}(-8.37 kJmol^{-1})$.[58] The reported average predicted binding energy of embelin to BIR3 is $-6 kcalmol^{-1}(-25.1 kJmol^{-1})$ as summarized in Table 10. The implication of this result is that embelin can competitively bind BIR3 in a similar fashion as does caspase thereby promoting apoptosis in a caspase dependent pathway. Possible caspase-independent pathway is also possible as embelin is known to generate ROS. So far, it is clear embelin meets both criterion for undergoing reactions type I and II reaction mechanisms. The exact energies within the protein environment is a subject of the next chapter.

Table 10: Relative binding affinities of embelin-BIR3 complex for mode one and two binding mechanism.

Binding mode	ΔG_{bind}**(kJ/mol)**
1	-18.67
2	-30.80
Average	- 24.73
Experimental	-20.9

Table 11: Relative binding affinities of embelin with Smac/Diablo-BIR3 complex.

Simulation	ΔG_{bind}(kJ/mol)	Dominating binding mode
1	-25.255±-27.429	Mode one
2	-23.036±37.240	Mode two
3	-39.122±32.951	Mode one
4	6.282±524.211	Mode two
5	-36.738±34.348	Mode one
6	-31.456±27.719	Switched
7	-22.044±28.595	Mode one
8	-10.138±35.866	Mode two
9	-24.119±33.471	Mode two
10	-17.379±30.378	Mode two

4.6 Conclusion

BIR3 serves as a potent druggable pocket that can be used to develop new promising drug-like molecules which can induce apoptosis. This can be achieved by mimicking the binding patterns of SMAC/DIABLO binding patterns which is known to inhibit BIR3 using its N-terminal residues. Additional knowledge of caspase binding substrates can supplement such a study since they also bind BIR3 competitively with SMAC.

In this chapter, stable binding patterns of embelin have been presented and critical mechanisms through which it binds BIR3 proposed. The binding occurs on the surface groove of BIR3 majorly via Glu and Hse residues which is in agreement with previously reported experimental data. The active site deviates by about 3 Å from the originally proposed binding site where Tryptophan and Tyrosine residues were identified to play a critical binding role. Glu 220 and 239 are reported to be the dominating residues interacting with embelin mainly via a hydrogen donating mechanism. Additionally, Tyrosine and Histidine bindings consist of high hydrogen donating percentages. In a nutshell, the ligand-residue pairs confirm the strong hydrogen donor interaction between embelin and BIR3.

Further mechanism data has been generated using the MSM model. Here, the dominating hydrogen donating ability of embelin which is the preferred mechanism, is clearly manifested. Three resolved binding modes have been identified and include, hydrogen donation, accepting and mixed donor-acceptor mechanisms. Although the donor mechanisms play a critical role which is biologically relevant. An additional weak mechanism is can be seen to occur between the carbonyl moeity and Lysine but cannot be overemphasized since the binding pairs percentages are quite low. Overall, the binding is then compared to available experimental data which closely match. In this work, embelin is reported to bind BIR3 with an energy of $-6 kcalmol^{-1} (-25.1 kJmol^{-1})$ which is close to the reported experimental value of $-5.0 kcalmol^{-1} (-20.9 kJmol^{-1})$.

4.6.1 Summary, Outlook, and Recommendation

Clear binding mechanisms have been reported in this chapter where it is seen that embelin bind in a stable binding site *via* hydrogen donating ability which could inhibit BIR3 in a caspase dependent pathway. In the previous chapter embelin was reported to donate protons intramolecularly using quantum mechanical calculations. The hydrogen donor mechanism reported in this chapter is in agreement with this donation trend. Furthermore, it is also a known fact that some of the distances involved in the binding are close to covalent distances which could facilitate proton transfer reactions. Advanced QM/MM studies can be carried out to explore such possibilities.

Another possibility of embelin inducing apoptosis is *via* caspase-independent pathway where it could act as an antioxidant. In fact, recent literature has studied its synergetic effects with light sources which reported increased cytotoxicity towards tumor cells. These studies can be carried out in a hybrid QM/MM where the potential of embelin to act as a photosensitizer within a protein environment thereby generating ROS could be explored and forms the basis of the next chapter.

Chapter - 5

Can Embelin generate ROS within the Protein Environment?

In this chapter, advanced hybrid quantum mechanical/molecular mechanics (QM/MM) methods have been used to investigate the photo chemical signatures of embelin embedded within the BIR3 site. It is a build up from the previous chapters and combines the quantum chemical and the molecular dynamics knowledge already discussed. This chapter's main results are the vertical excitation energies which are important in mapping out the photo mechanism of embelin within BIR3. The results are obtained using both electrostatics and polarizable embedding approaches and are explained in detail. The choice of these embedding methods was influenced by the need to produce reliable results comparable to experimental values. Furthermore, the excitation energies, more so the triplets can be used to improve the viability of embelin's application in PDT cancer treatment approaches. TD-DFT was employed for the electrostatic studies while Algebraic diagrammatic construction-Polarizable embedding (ADC-PE) was used for the polarization part.

5.1 Abstract

The application of photoactive molecules to treat diseases such as cancer and microbials continue to play a critical role in the medicinal field. A great amount of research has been devoted to studying photosensitizers (PS) which are central to the effectiveness of the medical approach. Several approaches are available for use towards the realization of effective PS which can be used in photodynamic therapy. Ultimately, these reactions occur within the body and any theoretical venture should include most if not all physicochemical effects in the study model.

Correct treatment of photosensitizer reactions especially in complex environments is therefore an important step towards obtaining reliable results that can be used to understand underlying photo mechanisms. In this section, accomplished QM/MM tools have been used to explore the photochemistry of embelin within a previously identified stable site i.e., within the BIR3 domain of the X-linked Inhibitor of apoptosis protein (see **Chapter - 4**). Embelin is seen to undergo a stable ground state proton transfer reaction with a pair of Glu side chains remaining bound at the active site. On photon absorption, the triplet energies, character of the excited state transition, as well as state ordering are quite similar to the previously discussed gas phase results implying that embelin is able to generate reactive singlet oxygen species. This consistency is useful in a number of medical applications including targeting tumor cells *via* PDT method.

Investigation of photo reactions within the protein environment provide a more realistic picture of the processes involved when embelin absorbs photon *in vivo*. However, such processes are not easy to capture in a strict sense, hence it is necessary to compare the values obtained to experimental results. Taking both electrostatics and polarization effects into account, the results are consistent when compared to experimental values indicating that the treatment of embelin-BIR3 complex with advanced hybrid QM/MM approach is successful. The PE-ADC excitation energy at the ADC(2) level of theory when polarization effects are included is 3.07 eV and is comparable to the experimental value of 2.76 eV.

5.2 Introduction

In recent years, research has shifted towards using light together with chemical molecules in treatment of diseases. This approach which is generally known as photodynamic therapy involves systematic administration of photosensitizers (PS) within a localized area of interest and a specific wavelength of light.[6] Success in this area depends on the effectiveness of the photosensitizer which provide necessary photochemical and photo pharmacological characteristics needed to trigger desired reactions, which is mainly apoptosis. The basic principle of PDT is the excitation of a PS to the first excited state (S_1) and since S_1 is short-lived, the PS could undergo fluorescence (emitting the absorbed photon) on its way back to the ground state. Alternatively, if it has a high spin-orbit coupling (SOC) value it could, *via* intersystem crossing, convert to a long-lived triplet state (T_1). An effective PS with excess energy can then transfer its energy to a nearby biomolecule or singlet oxygen consequently generating antagonistic ROS reactions which lead to tumor cell damage.

Porphyrins and its derivatives are among the first major PS which proved to have efficient photochemical characteristics to be used in PDT application. Other PS, which are non-porphyrin, based have also been developed depending on their modes of action and include hypericin (naturally occurs in *Hypericum* species) and dyes (Cyanines and Phenothiazinium).[96] The common characteristics of these PS is the ability to efficiently generate singlet oxygen known to be very reactive and selective towards abnormal cells. In addition, effective PS should possess desired medicinal properties which includes ability to obey acceptable absorption, distribution, metabolism, excretion, and toxicity (ADMET) limits, be chemically pure, be selective within tumor cells, and should not be carcinogenetic or cause mutagenicity.

The reactions which these PS undergo determines the extent to which they are used in PDT applications. Proton transfer followed by electron-coupled reaction is one of the common reactions which indicate the usefulness of light induced processes within biological systems, with photosynthesis being a perfect example.[96] Such reactions have been also found to be useful in the DNA repair of photolyase which has been extensively reviewed in reference. [143] A major promoting factor of the hydrogen

transfer is the favorable distance of the hydrogen bonds formed either intramolecularly or intermolecularly within the systems of interest. Short distances are particularly known to act as triggers to important biological reactions with desired biological consequences, commonly occurring along isolated waters, and polar or charged residues.[129] Within these distances, chain reactions can be consequently induced via light absorption processes that lead to specific mechanisms such as energy or even charge transfer. The main known reaction in PDT is the energy transfer which occurs through a triplet-triplet energy transfer (TTET) process. In TTET, the process can be considered as two simultaneous electron transfer process involving a donor-acceptor pair as shown in Figure 28 and is of the Dexter type of exchange. These reactions can be theoretically modeled within a complex environment using advanced tools.

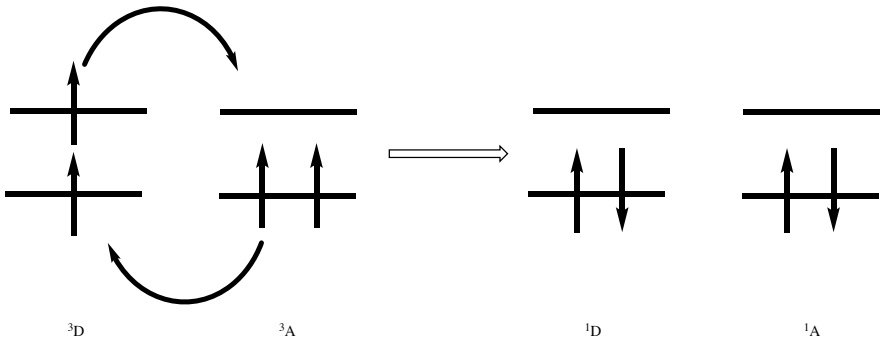


Figure 28: $\alpha,\alpha \rightarrow \beta,\beta$ electron exchange involving a donor-acceptor pair in TTET.

Theoretical treatment of PDT reactions and other biological processes within a complex environment remains a nontrivial task. Various methods have been developed over years to study processes within the protein environment including spectroscopy,[48] response properties[43,50], as well as proton transfer reactions.[129] Generally, these methods belong to the class of hybrid QM/MM type where a small region also known as core is treated quantum mechanically while the rest of the system known as the environment is treated using molecular mechanics. Just like in the case of normal quantum calculations, the core region can be calculated using MP2, DFT, HF or any other method of choice for ground state and TD-DFT, EOMCC, and the ADC family of methods employed for the excited state. Differences between the

various QM/MM methods lies in the physiochemical effects taken care of within the underlying theory representing the environment. These include electrostatics, polarization, and repulsion effects with the cost of calculation increasing in the same order, i.e., inclusion of one, two or all effects increases the cost as well as accuracy of the calculation.

From a drug design perspective, all of these effects play critical roles in describing biological processes within a complex environment and are therefore necessary to be included in a simulation. Electrostatic and polarizable embedding methods have recently become a popular choice to treat complex environments reproducing experimental results with the former employing classical force fields for the environment while the latter involves generation of a potential field to represent both electrostatics and polarization effects respectively. Polarization Embedding works through mutual induction of fields between the core and the environment. In this chapter, both two effects, the electrostatic- and polarization embedding hybrid QM/MM methods have been employed to simulate the spectroscopy of embelin and BIR3 environment. The key interest is to account for the electrostatics as well as higher order multipole polarizabilities offered by polarizable embedding potentials to capture relevant effects which influence embelin-BIR3 spectral signatures. Since both methods have an advantage of sampling geometries, previously generated MD trajectories have been used, with their last coordinates serving as starting coordinates for the calculations.

The aforementioned effects have been recently found to lead to more accurate results in proteins especially in cases where charge transfer processes occur.[43] As seen in section 3.4.2.1 embelin undergoes proton transfer reactions with ESIPT being favored as a result of stable intermediates formation. Proton transfer reactions could as well occur within the protein environment but in a different context. In the previous chapter, the dominating binding mechanism was hydrogen donation and therefore embelin could possibly undergo intermolecular proton transfer. To this end, effort has also been made to investigate the possibility of ground state hydrogen transfer reactions within the BIR3 environment.

5.3 Theoretical Methods

Methods in this chapter follow similar approach to the steps followed in obtaining chapter 3 results. Snapshots extracted from cluster one representing a stable site in the previous chapter (refer to subsection 4.5.4) were used to perform all the simulations in this chapter. The cluster consisted of 100 samples out of which 10 random snapshots were extracted and used to perform scans to investigate whether embelin undergoes intermolecular proton transfer. There is a possibility that the proton transfer to the nearby amino acid groups occur owing to the fact that embelin binds Glu 220 & 239 residues via its *para-hydroxyl* moieties. Consequent calculations using solvation model as well as QM/MM were carried out using the same snapshots.

Ground state QM scans: A scan between H25 towards O1 of Glu 220 and H26 towards O2 of Glu 239 (Figure 29) is therefore deemed to be comparable to the ground state scans in section **3.4.2.1**. The terminals of the carboxylic and the amine moieties were capped with acetyl groups to mimic unreactive protecting groups. Graph 1, shows the distances between embelin and both Glu protein residues within the aforementioned cluster representing a stable site from which the snapshots were taken. Only 57 out of the 100 samples (57%) from the extracted cluster met a criterion of ~ 5.5 Å and below which is a favorable distance to perform scans. The performed scans ranged from a maximum distance of ~3.5 Å. The resulting PES were then compared with the intramolecular results obtained in Chapter - 3. All scans were produced using the DFT/CAM-B3LYP//6-31G* model.

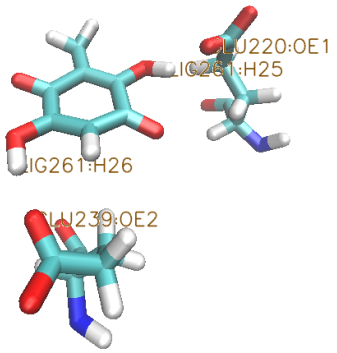
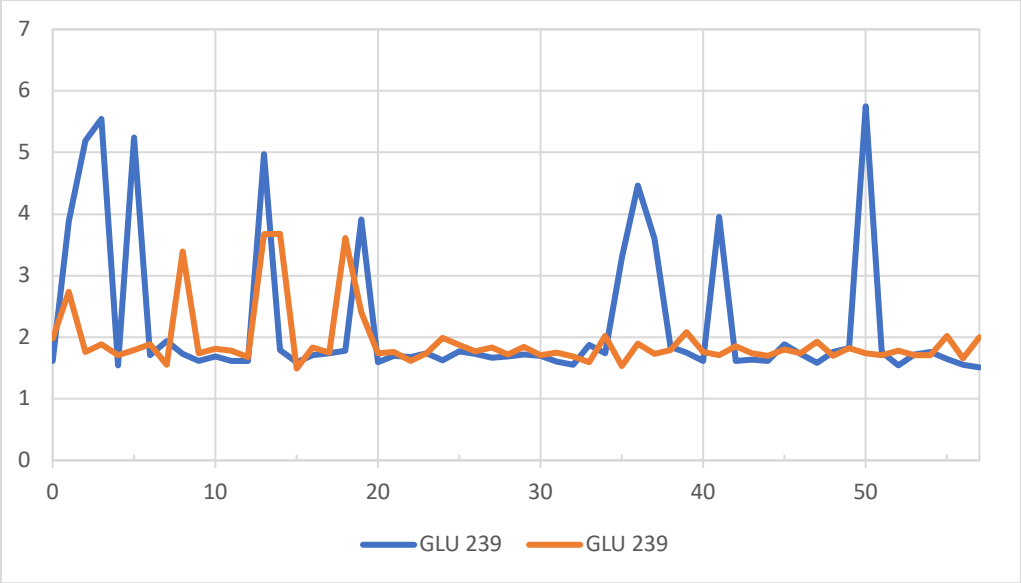


Figure 29: Representative geometry of embelin and Glu 220 and Glu 239 used to perform ground state proton transfer reactions.



Graph 1: A graph showing the H-bond distances between embelin and Glu residues within the stable site.

Pure QM calculations: The same geometries extracted from the cluster as explained above were subjected to PCM calculations by selecting water molecules ~ 3.5 Å from embelin to investigate the influence of water molecules on its spectroscopy. The main aim was to compare PCM energies from sampled random geometries with PE-ADC (includes polarization effects) and electrostatic embedding energies (includes electrostatic effects) when the environment is taken into consideration. Low lying vertical energies were calculated for the 10 snapshots using TD-DFT/CAM-B3LYP/6-31G* model. A model of embelin previously used was employed in all the calculations. Water was used as the dielectric medium for all the implicit PCM calculations.

QM/MM MD: embelin's spectrum within BIR3 environment is simulated in this section using hybrid QM/MM studies. Only the poses within the stable binding site previously clustered using Markov state model (see subsection 4.5.5) were considered since they are likely to bind long enough to absorb photons. In other words, only the geometries that showed significant binding were further pursued. Initial ground state QM/MM calculations were carried out for each snapshot followed by respective excited state calculations. Q-chem and Gaussian programs were used to model the QM region while NAMD was used for the MM region. The DFT/CAM-B3LYP/6-

31G* model and DFT-D3/PBEh-3c/def2-mSVP model were used to calculate electrostatic effects for the studied systems. The final coordinates of the ground state trajectories were extracted and used as inputs for exited states calculations.

A set of four systems were constructed to be included in the QM region. The choice was guided by previous pure QM calculations which consisted of vacuum and explicit solvents around embelin (see subsections 3.4.1 and subsections 3.4.1.4). As noted from the earlier results, a system of more than two water molecules have insignificant influence on the excitation energy of embelin. This was coincidentally an advantage owing to the prohibitive cost of QM/MM simulations.

The first system consisted of embelin alone in the QM region, the second and the third were obtained by selecting one nearest water molecules on either side of the ligand (Figure 30). The fourth system had both water molecules on each side of the molecule. Table 12 summarizes the set of systems used in this subsection. A QM/MM bond was placed between C7 and C9 within embelin (Figure 30) to ensure that the results are comparable to previously obtained gas phase results which were simulated on a truncated embelin model. Truhlar's redistributed charge and dipole (RCD) method[144] was used to prevent over polarization at the QM/MM boundary. An integration timestep of 0.5 fs was used for all the calculations. Short minimizations of 100 steps were initially done for each simulation followed by a 12 ps QM/MM run. The simulation conditions were maintained at 310K and 1 bar pressure. Short- and long-range effects were treated in a similar manner as in the MD setup explained in section **4.4.3**. Point charges passed onto the QM region were embedded using electrostatics embedding at a cutoff distance of 12 Å.

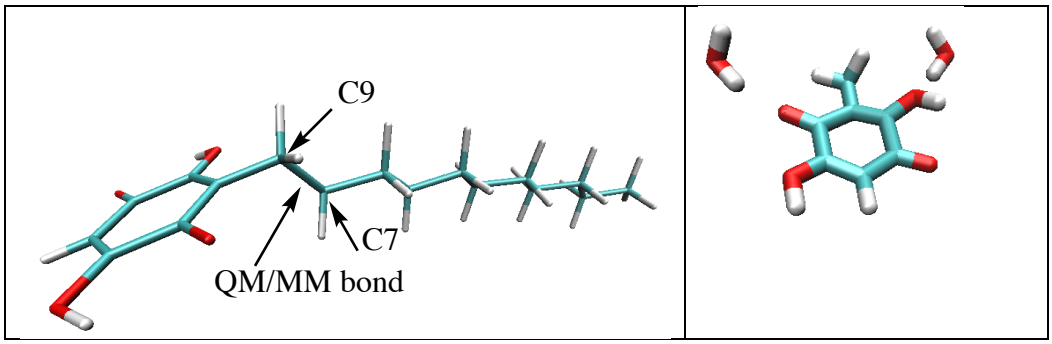


Figure 30: QM/MM bond placed between C7 and C9 as a model to investigate embelin with BIR3 protein. Embelin water systems used for calculating the hybrid QM/MM and PE spectral signatures of embelin.

Table 12: Systems used for hybrid QM/MM simulations of embelin within the BIR3 environment.

System	System Group	Number of Atoms	Snapshots
I	Embelin	17	10
II	Embelin + H_20a	20	2
III	Embelin + H_20b	20	2
IV	Embelin + $2H_2O$	23	2

QM/MM PE: The final geometries of the QM/MM MD simulations were used to generate embedding potentials for systems in Table 12. A fragment-based molecular fractionation with conjugate caps (MFCC)[145] scheme was used to cap the fragments of the environment. For each system, DFT/CAM-B3LYP/6-31G* model was used to parametrize the protein for multipoles up to the second order and dipole-dipole polarizabilities were obtained using the localized properties (LoProp) model.[145] All Ions (sodium and chloride) were modeled using the standard "SEP" potentials model. Water up to 15 Å and 20 Å were modeled using second order multipoles and polarizabilities using LoProp as implemented in the Dalton program. The PyFraME[50] package which provides automated Dalton scripts out of the box for the potential file generation was used for this purpose. Using the generated potential, PE-ADC calculation was thereafter carried out at the ADC(2)/6-31G* level of theory as implemented in the Q-Chem 5.2 program. A redistribution scheme was used to remove site polarizabilities between the quantum and the molecular region by turning the *border* keyword to *true* and *border_type* to *redist*. This technique prevents over polarization at the sites and ensures that the core/environment region is properly represented. The general scheme of the studies carried out in this chapter is shown in Figure 31 below.

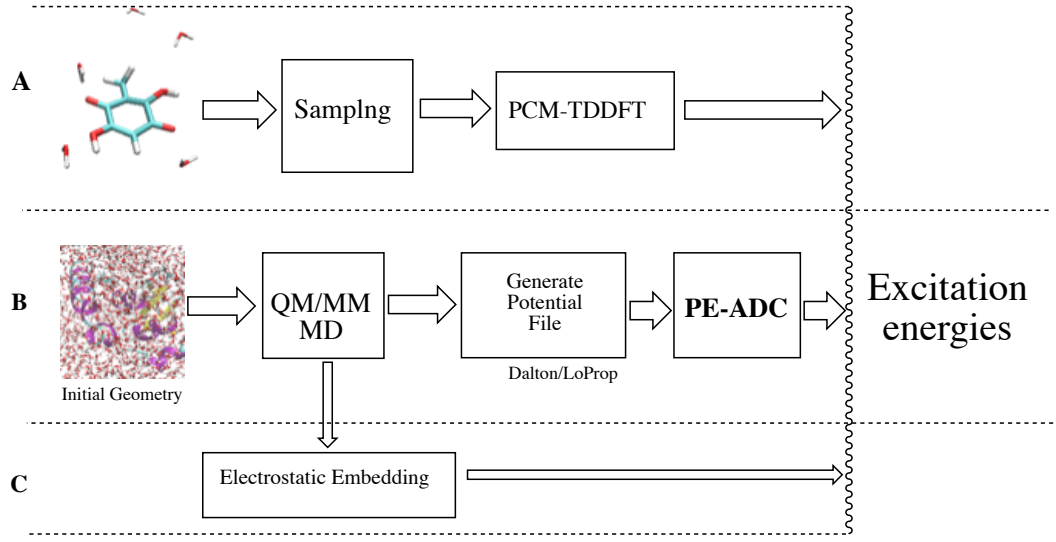


Figure 31: Explicit solvent simulation from sampled MD snapshots(A). Hybrid QM/MM of embelin-BIR3 workflow simulated using PE-ADC (B) and Electrostatic Embedding (TD-DFT) (C).

5.4 Results and Discussion

Influence of sampled explicit water snapshots on excitation energy: Influence of explicit water molecules around ligands in a protein environment contribute to important hydrogen bond network and may influence excitation energies. Table 13 shows the PCM excitation energies obtained from sampled trajectories consisting of water molecules within ~3.5 Å distance of embelin. In case of embelin, gas phase simulations did not realize any significant influence on its excitation energy as seen in chapter 3. However, the advantage of sampling is in this case utilized to generate PCM energies that can be compared with values obtained from hybrid QM/MM.

Table 13: Low lying vertical excited state of sampled embelin snapshots with explicit water at a radius of within 3.5 Å. The implicit PCM solvent used was water, n=x represents the number of explicit water molecules used.

snapshot / system State	n=5	n=5	n=0	n=8	n=6	n=2	n=7	n=7	n=8	n=3
	1	2	3	4	5	6	7	8	9	10
S_1	3.0449	2.9988	3.0758	2.8695	3.0805	3.0118	2.9881	3.0699	3.02	3.1043
S_2	3.3224	3.286	3.3448	3.2653	3.3402	3.3086	3.229	3.2347	3.2753	3.2201
S_3	3.5055	3.5416	3.4581	3.5524	3.4926	3.4529	3.3632	3.5075	3.47	3.5062
S_4	4.268	4.2201	4.2877	4.1221	4.2903	4.3656	4.2524	4.2861	4.242	4.2958
T_1	1.9753	1.9722	2.045	1.856	2.0083	1.9251	1.7604	1.9921	1.9737	2.0205
T_2	2.1784	2.1671	2.1677	2.1474	2.1734	2.2567	2.1325	2.1662	2.1453	2.1953
T_3	2.8155	2.7212	2.7949	2.7536	2.8183	2.7753	2.6985	2.7103	2.7707	2.6532
T_4	2.9845	3.053	2.9469	3.1079	2.9803	2.9334	2.8508	2.9895	2.9446	3.0332

From the data, there is consistency in the energies regardless of the number of water system in the snapshots illustrating the little influence of explicitly adding water on embelin's excitation energies. The S_1 energies are approximately 3 eV and as earlier seen the S_4 state is the bright state. Similarly, the transitions of the states are identical to the previously identified HOMO→ LUMO transitions etc. It is however necessary to compare these values when embelin is embedded within the protein environment to investigate the significance of the protein and additional effects (electrostatic and polarization) that are added.

A look at the triplet energies point to general consistent trend similar to the singlets. More importantly, T_1, T_2, and T_3 are all lower that S_1. In case ESPIT exists within the protein environment, then efficient *isc* is favored to facilitate the much-desired TET processes consequently promoting PDT. It is nevertheless worth noting that proton transfer in the presence of bulk solution could result in different mechanisms hence different outcomes.[129]

QM Scans: Ligands bound within active sites may experience several physiochemical effects which may lead to reactions such as proton transfer, electron transfer, and even charge transfer. In the presence of bulk solvents and protein environment, existing complex hydrogen bond networks may trigger proton transfer pathways *via* polar residues, water molecules, and ionizable residues.[129] An understanding of the possible proton transfer reactions is important in identifying specific physicochemical effects affecting the ligands of interest when bound to protein residues.

The possibility of embelin undergoing proton transfer reactions is enhanced by existence of resonance in the ground state, a feature that is previously cited in literature.[67] As seen from Figure 15, the ground state intramolecular proton transfer leads to unstable intermediates depending on which side of the molecule the proton transfer takes place. A second step of proton transfer must take place to form an isomer of the starting geometry. Within the protein environment, intermolecular proton transfer is more likely to take place compared to the previously explored intramolecular transfer pathway considering the pose of embelin within the stable site is different (see the difference between Figure 32 A and B). Graph 2 shows the ground state proton transfer scans between embelin and Glu 220 and Glu 239 residues. There is an immediate rise in energy followed by a "plateau" and thereafter an energy drop. The rise in energy corresponds to the instability of the embelin system on losing a proton. Before the energy drop, a proton is then captured from the positively charged amide group by the unstable embelin leading to its re-stabilization and consequent decrease in energy. This means that the proton exchange between embelin and the

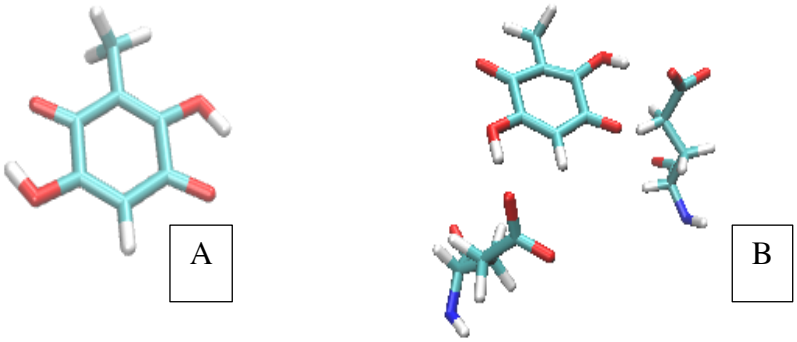
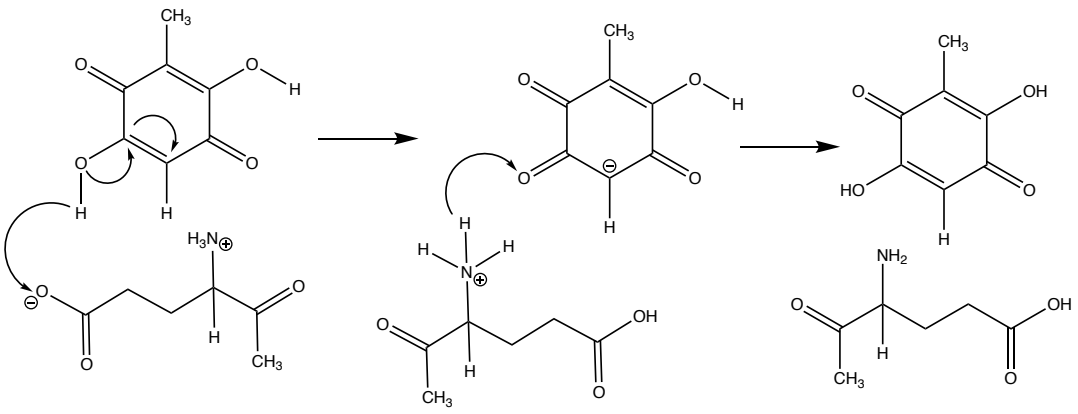


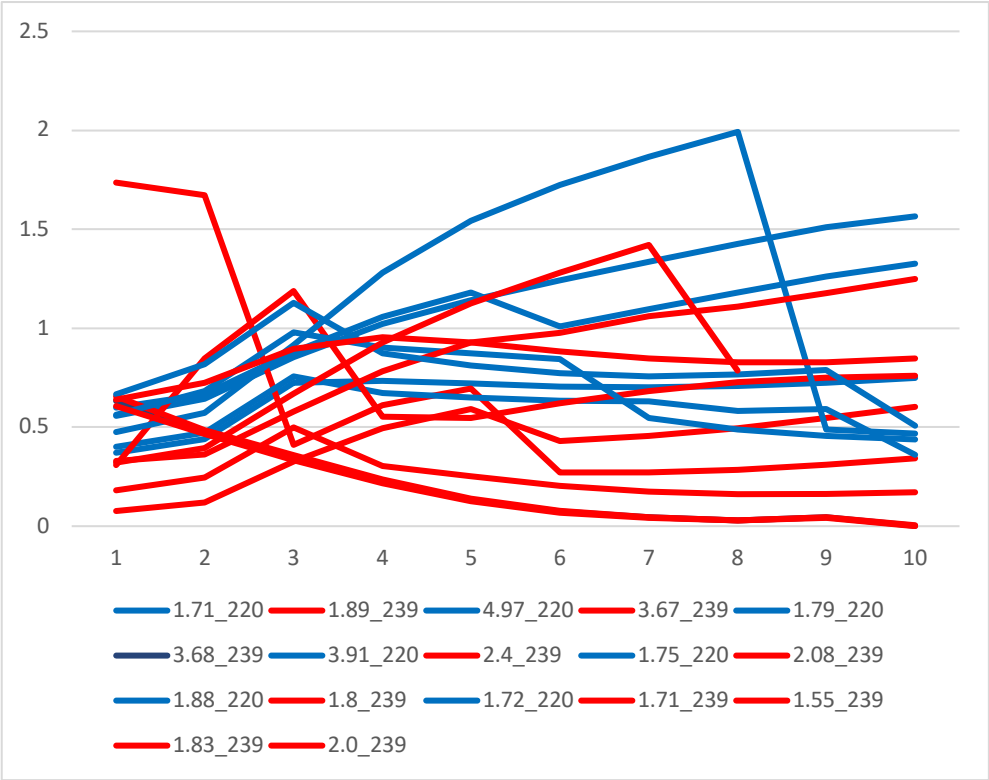
Figure 32: Poses of embelin's optimized geometry depicting both hydroxyl groups oriented towards the carbonyls (A) and within a protein environment where both of the hydrogens are oriented towards Glu oxygen atoms (B).

Glu residues generally leads to an overall stable system for exchanges taking place within a distance of ~ 3.5 Å and below.

Such stability is indeed dependent on the proximity between embelin and the two amino acid residues. A different picture could in principle be observed at long distances since nearby water molecules may also donate or accept their protons. More importantly, stability of the reaction implies that embelin remains bound within the site and can therefore absorb a photon regardless of any ground state proton transfer reactions. Once stability in the ground state is ascertained, advanced spectroscopy hybrid QM/MM simulations of embelin within the BIR3 environment can be carried out. A proposed general proton transfer reaction mechanism between embelin and both Glu residues is shown in Scheme 1.



Scheme 1: Proposed proton transfer pathway between embelin and Glu residues. A similar reaction would take place on the other side of the molecule due to the favorable pose shown in figure 5.



Graph 2: Intermolecular proton transfer between embelin and Glu 220 (blue) and 239 (red) residues, x_220 and x_239 in the legend represents the distance between embelin and the residues.

Hybrid QM/MM simulations:

Singlet states: In the electrostatic embedding QM/MM layout, the MM region does not modify its charge distribution with respect to the QM region. The opposite is however true. In a nutshell, this layout utilizes the MM charge model together with the QM charge density which are periodically updated.[146] Polarization thus occur between the QM and the MM regions hence the provision of a more accurate and advanced description of the physicochemical effects the environmental (MM) has on the core region (QM). Inclusion of electrostatic effects is thus important in characterizing the spectroscopic signatures of embelin within BIR3 environment. One of the known challenges when setting up QM/MM simulations is how charges at the QM/MM interface are handled.

Caution needs to be taken especially when treating the charge density at the boundary to avoid problems arising from over polarization. Over polarization may occur when

the MM charges are in close proximity to the QM electron density[146] and the problem can be solved by either removing or redistributing the excess charge at the boundary. For this reason, a redistribution charge and dipole (RCD) scheme[144] has been employed to take care of this problem. RCD is specifically designed to handle link atoms, especially when covalent bonds are involved between the QM and the MM regions.

The low-lying vertical excitation energies obtained from the various systems of interest are shown in Table 14. A comparison is first made between the PCM snapshots (Table 13) and the values obtained by adding effects from the environment. It is evident that the QM/MM energies are slightly red shifted compared to the obtained PCM energies reported in Table 13. This trend holds across the systems studied indicating that within the protein environment, embelin absorbs photons at relatively higher wavelengths when electrostatic effects (TD-DFT/CAM-B3LYP/6-31G* model) are included in the simulation.

A similar effect is observed for the vertical energies obtained at the TD-DFT/D3-BJ level of theory. The D3BJ method includes dispersion corrections, specifically the Becke and Johnson (BJ) dumping for small interatomic distances and is best suited in cases where hydrogen bond network is present.[147] In fact, BJ-damping takes care of noncovalently bonded structures within short distances better than other schemes thereby making it a suitable functional to use within the protein environment. Energies obtained in Table 14 using this method are also consistent with the data in subsection 3.4.1.2. There is however a clear distinction when D3BJ and CAM-B3LYP energies are compared. A general trend indicate that D3BJ values are slightly blue shifted for all the systems although the difference in the energies is less than 1 eV. Once again, the character of the S_1, T_1 and T_2 transitions are similar to the results previously discussed. Therefore, similar rationalizations regarding embelin's photo reaction mechanism still hold up to this point.

Table 14: QM/MM vertical excitation energies for various embelin systems (oscillator strengths for the bright state in bracket).

		QM/MM		PE-ADC (2)	Expt (meOH)
		D3-BJ	CAM-B3LYP		
Embelin					4.31 eV
Emb + 0 H_2O	S_1	2.4	1.84	3.13	
	S_2	3.05	2.82	3.37	
	S_3	3.34	3.27	3.35	
	S_4	4.70	4.09	4.91	
Embelin + 1_a (02) H_2O	S_1	1.95	2.03	2.76	
	S_2	2.90	2.99	3.16	
	S_3	3.34	3.07	3.40	
	S_4	4.45 (0.32)	4.09 (0.32)	4.57 (0.48)	
Embelin + 1_b (O4) H_2O	S_1	2.27	1.96	2.89	
	S_2	3.02	2.95	3.16	
	S_3	3.56	3.37	3.48	
	S_4	4.55 (0.36)	4.08 (0.28)	4.64 (0.45)	
Embelin + 2 H_2O	S_1	2.45	2.03	2.93	
	S_2	3.12	2.99	3.00	
	S_3	3.46	3.04	3.50	
	S_4	4.69 (0.39)	3.98 (0.31)	4.30 (0.13)	

Simulated spectrum of embelin within BIR3: It is a general interest to report the overall UV-vis spectrum of embelin within the protein environment which was obtained using system I (Table 12) using a convoluted Gaussian function represented in (74). Each of these systems were run for 2.5 ps and the excitation energies used to compute the spectrum. QM/MM MD simulations on these snapshots ensured a well sampled configuration. The general formular used to convolute the spectra is,

$$\varepsilon_i(\tilde{\nu}) = \varepsilon_i^{\max} \exp\left[-\left(\frac{\tilde{\nu}-\tilde{\nu}_i}{\sigma}\right)\right] . \qquad (74)$$

Since there are more than one electronic excitation, the overall spectrum was obtained as a sum of individual bands as in

$$\varepsilon(\tilde{\nu}) = \sum_{i=1}^{n} \varepsilon_i(\tilde{\nu}) = \left(\sum_{i=1}^{n} 1.3062974 \times 10^8 . \frac{f_i}{\sigma} \exp\left[-\left(\frac{\tilde{\nu}-\tilde{\nu}_i}{\sigma}\right)^2\right]\right) . \qquad (75)$$

The spectrum is shown in Figure 33. In this case, state transitions for the excitations are also similar to the ones discussed in (3.4.1.2). i.e., $\pi \rightarrow \pi^*$ for S1 and S4 and $n \rightarrow \pi^*$ for S2 and S3 as can be seen from the attachment-detachment Figure 8.

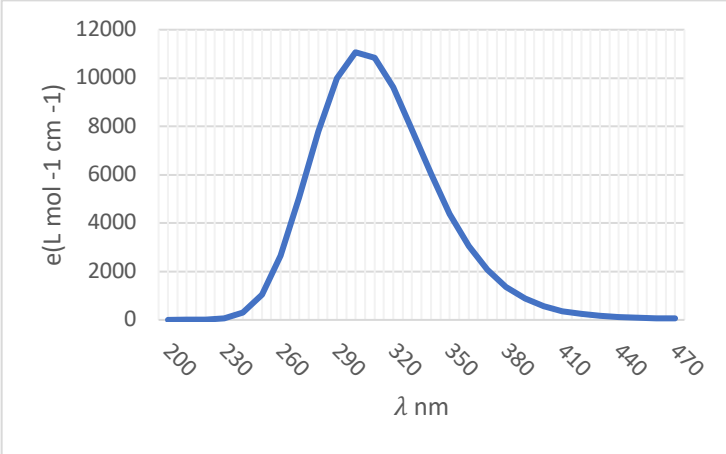


Figure 33: Convoluted UV-vis spectrum of embelin within BIR3 protein.

Table 15: Excitation energies used to obtain a convoluted spectrum of embelin within the BIR3 protein environment.

system / State	1	2	3	4	5	6	7	8	9	10
S_1	1.84	1.7	1.98	2.42	1.62	2.11	2.01	1.94	1.47	1.81
S_2	2.82	3.03	2.97	3.05	2.74	2.97	3.23	3.10	2.77	2.92
S_3	3.27	3.17	3.31	3.48	3.39	3.21	3.27	3.28	3.26	3.05
S_4	4.09	4.00	4.12	4.33	4.05	4.09	4.15	4.07	4.01	4.01

PE-ADC simulations: The PE method was developed with an aim of describing general excited states and molecular properties of embedded systems.[148] Ideally, the method can be combined with configurational sampling to include polarization and electrostatic effects described above. An advantage of PE is that it works with both heterogenous and homogenous molecular systems and is implemented for various methods, including HF and DFT[49] and ADC family of methods.[43]

In summary, the PE model strives to describe the first order and permanent induced charge distribution of the environment using distributed multipole expansions by

taking the molecules of the environment as expansion points. Anisotropic dipole-dipole polarizability as well as permanent multipole is assigned to each site in the environment. In the end mutual polarization between the core region and the environment is achieved in an efficient SCF procedure.[49] More relevant to this study is the PE-ADC variant implemented in Q-chem as of version 5.2. The method allows for studying complex biomolecular systems in an efficient manner to produce reliable ADC excitation energies.

Table 14 shows the summarized excitation energies obtained with PE-ADC. Compared to CAM-B3LYP and D3-BJ methods, they are further blue shifted especially in the case of $s_0 \rightarrow s_1$ transition. The other states are relatively less blue shifted in comparison to the DFT methods. A similar conclusion can be observed when ADC values are compared to PCM-DFT values in (Table 13) but are in this case almost identical.

Systems treated with the method in the original implementation also reported identical observation where blue shifted PE-ADC values were observed.[43] Even though experimental values for embelin in protein does not exist to the best of my knowledge, the consistency of the PE-ADC method indicates that the results obtained in this subsection are reliable and represent what would be obtained experimentally. Additionally, as expected, perturbative corrections (both state specific $ptSS$ and linear response $ptLR$) which indicate the direction to which the overestimated energies are corrected show zero $ptSS$ and $ptLR$ for system I and increase as the number of water molecules around embelin is increased. This has been formerly reported to be normal in the original implementation. PE-ADC state characterization are also similar with the previously obtained attachment/detachment orbital analysis and hence reliable for triplet energy discussion.

Triplet states:

The most relevant photochemical process in PDT applications is the *isc* pathway to the triplet state which is facilitated by an electron flip from the singlet manifold to triplet manifold. If the molecule crossing to the triplet state possess enough energy, it is able to undergo a triplet energy transfer reaction in a donor/acceptor mechanism

with molecular oxygen to generate singlet oxygen (commonly known as ROS). Within a biological context process, damage of tumor cells consequently follows after generation of ROS.

The triplet energies of embelin in BIR3 environment are shown in Table 16. Similar to the singlet states, the CAM-B3LYP values are blue shifted compared to the D3-BJ and PE-ADC(2) values further confirming consistency in the results obtained using these methods. This trend is valid across all systems investigated demonstrating that even in the protein environment, embelin does not undergo significant state reordering or drastic photochemical changes, at least when it is bound to the protein. One can therefore discuss the triplets and their fate with respect to PDT as it was previously in subsection 3.4.2.2. Both T_1 and T_2 lie energetically lower than S_1 as previously observed. Thus, the possibility of embelin undergoing efficient *isc* within this arrangement is high since the respective triplet states lie below the S_1 state. Additionally, given that the gap between the S_1 and T_1 is not large, ~ 1 eV for D3-BJ and CAM-B3LYP values, embelin can efficiently cross to the triplet state. ADC(2) values are surprisingly low with 0.63 eV for the highest value and 0.34 eV for the lowest and hence the likelihood of crossing to the triplet state is even higher.

The character of the excitation transition is also maintained. I.e., S_1 has a $\pi\pi^*$ while T_1 and T_2 are $n\pi^*$ and $\pi\pi^*$ respectively. This arrangement efficiently facilitates *isc* in the gas phase photochemistry of embelin and a similar argument can be made when embelin is bound to BIR3. In other words, embelin can undergo stable reactions, particularly photoactivated reactions which is useful for PDT application. Stability resulting from the scans obtained above further support this hypothesis. It is hence sufficient to state that embelin can undergo two types of reactions:

(i) Stable proton transfer reactions akin to type I reaction mechanism in which the ligand of interest reacts directly with the biomolecule and
(ii) ROS generation reactions *via* energy transfer to a possible nearby oxygen molecule. These finding supports the *in vivo* experimental results initially proposed by Joy.[22]

Table 16: QM/MM triplet excitation energies for various embelin systems. The singlet-triplet gap is also provided for the first excited state.

		QM/MM				PE-ADC (2)	
		D3-BJ	ΔE_{ST}	CAM-B3LYP	ΔE_{ST}		ΔE_{ST}
Embelin							
Emb + 0 H_2O	T_1	1.31	1.09	1.02	0.82	2.68	0.45
	T_2	1.90		1.14		2.88	
	T_3	2.55		1.86		3.08	
	T_4	2.73		2.38		3.16	
Embelin + 1_a (02) H_2O	T_1	1.31	0.64	0.96	1.07	2.13	0.63
	T_2	2.05		1.37		2.79	
	T_3	2.60		1.99		3.00	
	T_4	2.97		2.43		3.23	
Embelin + 1_b (O4) H_2O	T_1	1.25	1.02	0.84	1.12	2.55	0.34
	T_2	2.12		0.74		2.78	
	T_3	2.71		2.06		2.97	
	T_4	3.22		2.15		3.11	
Embelin + 2 H_2O	T_1	1.52	0.93	0.86	1.17	2.50	0.43
	T_2	2.45		1.48		2.65	
	T_3	2.52		2.25		2.75	
	T_4	3.13		2.64		2.99	

5.5 Conclusion

In this chapter, the ground state proton transfer reactions and the spectroscopy signatures of embelin have been explored. Embelin is found to remain stable and bound within the identified site even when undergoing hydrogen transfer reactions within a binding distance ~ 3.5 Å. Two intermolecular proton transfer pathways analogous to the previously explained intramolecular pathway has been discussed. The general reaction mechanism for loosing either of the protons is identical. In this scenario, both reactions proceed via an unstable "*intermediate*" before restabilizing in a two-step proton exchange mechanism. The binding distance chosen ensures that embelin remains bound to the site since competitive proton transfer reactions such as towards water could take place at larger distances.

The PCM excitation energies from the QM/MM snapshots highlight the consistency in the photochemistry of embelin even with solvents around the protein environment. Energies obtained are comparable to the previously mentioned gas phase and experiment. A clear conclusion regarding the influence of water system surrounding

embelin can be made. Therefore, it is evident that there is no significant change in embelin's excitation energy beyond a two-water system.

A striking finding in the QM/MM excitation energies is the consistency across the methods. It is however clear that CAM-B3LYP provide slightly red shifted values compared to D3-BJ and ADC (2) values which are blue shifted. The ordering of the excited states and the character of the transitions has also been found to favor *isc.* Since embelin possess excess energy as has previously been explained, it can generate singlet oxygen which is useful in PDT applications. These facts lead to positive confirmation of the proposed hypothesis that embelin's cytotoxicity is improved in the presence of light. Furthermore, it remains stable thermally to undergo photo reactions.

Global Summary and Conclusions

The photochemistry of embelin and its potential as photosensitizer have been presented in this thesis. The main aim was to provide theoretical evidence that embelin possesses unique photochemical properties and hence can be utilized in photodynamic therapy. Effort has been dedicated to study both its static and dynamic properties using the TD-DFT and ADC family of methods. Analyses of the properties of interest have been carried out using attachment/detachment densities, natural bond orbital theory and exciton properties.

In Chapter - 3 pure quantum chemical calculations have been used to explore the properties of embelin. **Static properties**: The experimental absorption energy of embelin is 4.31 eV and has been well reproduced using ADC (3) yielding a gas phase value of 4.32. The interest however lies in the photochemical application of embelin within biological systems and hence the influence of water was investigated. Only addition of up to two water molecules influenced the excitation energies. Adding more explicit water molecules did not have any further influence. An explanation to this observation boils down to the fact that embelin forms intermolecular "*ring"* with two water molecules and more water molecules do not affect its bonding properties. In the ground state equilibrium geometry, a unique hyperbond is observed on one side of the molecule from O3 to C9. At the S_1 optimized geometry two hyperbonds from O3 to O4 are formed. This hyperbond is one of the features that activates most if not all of embelin's photochemistry as it leads to stable intermediates at the excited state. Bond parameters of this hyperbond confirms that photon absorption enhances resonance of embelin.

Dynamic properties: The explained phenomena above are clearly manifested using dynamic ground and excited state PES scans, by moving the two hydroxyl protons towards the carbonyl groups. Two intermediates (I_H and I_{CH3}) are formed which are critical in explaining the photo chemistry of embelin. Particularly, in the ground state PES, these intermediates are unstable while at the S_1 PES, they are seen to be stable, a feature attributed to the existing hyperbonds. It is important to note that the stability

is however more pronounced at the I_H than I_{CH3} in the excited state while the opposite is true in the ground state. In the vicinity of these intermediates, conical intersections were found and since embelin possess excess energy, it can hit these CIs on its way back to the ground state. More relevant to PDT applications are the triplet energies which were found to lie below the S_1 state. Due to Kasha's rule and the character of the excitation coupled with a SOC value of 23 cm^{-1}, embelin can efficiently cross to the triplet state and generate ROS. This finding is important and corroborate the previously reported experimental observations.

In Chapter - 4, the binding patterns of embelin within the BIR3 protein environment have been reported. Here, the aim was to identify and investigate the "*active site"* where embelin prefers to bind, its mechanism and the binding energies. A preference to Glu 220 and Glu 230 residues of BIR3 led to the conclusion that embelin prefers to bind BIR3 *via* a hydrogen donating mechanism. The binding is found to not compete with that of SMAC since it is known that SMAC competes caspase within its own binding mechanism. Markov state model theory was used to reveal these binding mechanisms. Even though embelin binds mostly to Glu residue pairs, it was found to bind to Tyr 226 as previously reported in experiments. The computed binding energy of -24.7 kJ/mol is also comparable to experimental results which has been reported at -25.1 kJ/mol.

In chapter 5, the photochemical properties of embelin enabling it to act as a PS within a protein have also been further investigated using hybrid QM/MM approaches. When electrostatic effects are included using a TD-DFT/MM combination, embelin's absorption spectrum is seen to be slightly blue shifted especially when using the PBE/D3-BJ xc-functional while inclusion of polarization red shifts the spectrum. These energies are comparable to the gas phase and PCM values from snapshot geometries extracted from MD trajectories. Using the reliable PE-ADC method, the energy values obtained shifts of $\sim$ 1 eV is observed. More importantly, the triplet values are within the known singlet oxygen generation capability range. This confirms the utility of embelin as photosensitizer when bound to a protein environment. Furthermore, proton transfer reactions between embelin and Glu residue pairs result in stability. The conclusion can therefore be drawn that embelin remains photochemically stable within the BIR3 environment regardless of proton transfer

reactions taking place and that no photobleaching will occur. This stability promotes PDT applications since the probabilities of photon absorption is increased.

This work has therefore laid a solid background on the utility of embelin especially its potential application in the area of photodynamic therapy (PDT). The data presented herein can be used to improve the absorption wavelength of the benzoquinone scaffold in general and embelin in particular. Known functional groups can be used to substitute the undecyl group and or the hydrogen at positions 6 and 11 respectively of Figure 6 to improve absorption wavelength range.

References

(1) Dolmans, D. E. J. G. J.; Fukumura, D.; Jain, R. K. Photodynamic Therapy for Cancer. *Nat. Rev. Cancer* **2003**, *3* (5), 380–387. https://doi.org/10.1038/nrc1071.

(2) Kessel, D. Apoptosis, Paraptosis and Autophagy: Death and Survival Pathways Associated with Photodynamic Therapy. *Photochem. Photobiol.* **2019**, *95* (1), 119–125. https://doi.org/10.1111/php.12952.

(3) Sasidharan, S.; Poojari, R.; Bahadur, D.; Srivastava, R. Embelin-Mediated Green Synthesis of Quasi-Spherical and Star-Shaped Plasmonic Nanostructures for Antibacterial Activity, Photothermal Therapy, and Computed Tomographic Imaging. *ACS Sustain. Chem. Eng.* **2018**, *6* (8), 10562–10577. https://doi.org/10.1021/acssuschemeng.8b01894.

(4) Berkman. Targeted Photodynamic Therapy for Prostate Cancer: Inducing Apoptosis via Activation of the Caspase-8/-3 Cascade Pathway. *Int. J. Oncol.* **2010**, *36* (4). https://doi.org/10.3892/ijo_00000553.

(5) Simone, B. C. D.; Mazzone, G.; Sang-aroon, W.; Marino, T.; Russo, N.; Sicilia, E. Theoretical Insight into Joint Photodynamic Action of a Gold(i) Complex and a BODIPY Chromophore for Singlet Oxygen Generation. *Phys. Chem. Chem. Phys.* **2019**, *21* (7), 3446–3452. https://doi.org/10.1039/C8CP04848G.

(6) Gomer, C. J.; Rucker, N.; Ferrario, A.; Wong, S. Properties and Applications of Photodynamic Therapy. *Radiat. Res.* **1989**, *120* (1), 1–18. https://doi.org/10.2307/3577632.

(7) Spikes, J. D.; Jori, G. Photodynamic Therapy of Tumours and Other Diseases Using Porphyrins. *Lasers Med. Sci.* **1987**, *2* (1), 3–15. https://doi.org/10.1007/BF02594124.

(8) Gomer, C. J. Preclinical Examination of First- and Second-Generation Photosensitizers Used in Photodynamic Therapy. *Photochem. Photobiol.* **1991**, *54* (6), 1093–1107. https://doi.org/10.1111/j.1751-1097.1991.tb02133.x.

(9) Schmidt, R. Photosensitized Generation of Singlet Oxygen. *Photochem. Photobiol.* **2007**, *82* (5), 1161–1177. https://doi.org/10.1562/2006-03-03-lR-833.

(10) Jockusch, S.; Turro, N. J.; Thompson, E. K.; Gouterman, M.; Callis, J. B.; Khalil, G. E. Singlet Molecular Oxygen by Direct Excitation. *Photochem. Photobiol. Sci.* **2008**, *7* (2), 235–239. https://doi.org/10.1039/B714286B.

(11) Karotki, A.; Kruk, M.; Drobizhev, M.; Rebane, A.; Nickel, E.; Spangler, C. W. Efficient Singlet Oxygen Generation upon Two-Photon Excitation of New Porphyrin with Enhanced Nonlinear Absorption. *IEEE J. Sel. Top. Quantum Electron.* **2001**, *7* (6), 971–975. https://doi.org/10.1109/2944.983301.

(12) Bai, S.; Barbatti, M. Divide-to-Conquer: A Kinetic Model for Singlet Oxygen Photosensitization. *J. Chem. Theory Comput.* **2017**, *13* (11), 5528–5538. https://doi.org/10.1021/acs.jctc.7b00619.

(13) Wu, G.; Chai, J.; Suber, T. L.; Wu, J.-W.; Du, C.; Wang, X.; Shi, Y. Structural Basis of IAP Recognition by Smac/DIABLO. *Nature* **2000**, *408* (6815), 1008–1012. https://doi.org/10.1038/35050012.

(14) McIlwain, D. R.; Berger, T.; Mak, T. W. Caspase Functions in Cell Death and Disease. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (4), a008656. https://doi.org/10.1101/cshperspect.a008656.

(15) Rajendran, M. Quinones as Photosensitizer for Photodynamic Therapy: ROS Generation, Mechanism and Detection Methods. *Photodiagnosis Photodyn. Ther.* **2016**, *13*, 175–187. https://doi.org/10.1016/j.pdpdt.2015.07.177.

(16) Siewert, B.; Stuppner, H. The Photoactivity of Natural Products – An Overlooked Potential of Phytomedicines? *Phytomedicine* **2019**, *60*, 152985. https://doi.org/10.1016/j.phymed.2019.152985.

(17) Chitra, M.; Sukumar, E.; Suja, V.; Devi, S. Antitumor, Anti-Inflammatory and Analgesic Property of Embelin, a Plant Product. *Chemotherapy* **1994**, *40* (2), 109–113. https://doi.org/10.1159/000239181.

(18) Atlabachew, M.; Mehari, B.; Combrinck, S.; McCrindle, R. Single-Step Isolation of Embelin Using High-Performance Countercurrent Chromatography and Determination of the Fatty Acid Composition of Seeds of Embelia Schimperi. *Biomed. Chromatogr.* **2017**, *31* (12), e4018. https://doi.org/10.1002/bmc.4018.

(19) Mahendran, S.; Badami, S.; Ravi, S.; Thippeswamy, B. S.; Veerapur, V. P. Antioxidant, Analgesic and Anti-Inflammatory Properties of New Ninhydrin Adduct of Embelin. *Pharm. Chem. J.* **2011**, *45* (9), 547–551. https://doi.org/10.1007/s11094-011-0676-x.

(20) Ko, J.-H.; Lee, S.-G.; Yang, W. M.; Um, J.-Y.; Sethi, G.; Mishra, S.; Shanmugam, M. K.; Ahn, K. S. The Application of Embelin for Cancer Prevention and Therapy. *Molecules* **2018**, *23* (3), 621. https://doi.org/10.3390/molecules23030621.

(21) Naik, S. R.; Niture, N. T.; Ansari, A. A.; Shah, P. D. Anti-Diabetic Activity of Embelin: Involvement of Cellular Inflammatory Mediators, Oxidative Stress and Other Biomarkers. *Phytomedicine* **2013**, *20* (10), 797–804. https://doi.org/10.1016/j.phymed.2013.03.003.

(22) Joy, B.; Kumar, S. N.; Radhika, A. R.; Abraham, A. Embelin (2,5-Dihydroxy-3-Undecyl-p-Benzoquinone) for Photodynamic Therapy: Study of Their Cytotoxicity in Cancer Cells. *Appl. Biochem. Biotechnol.* **2015**, *175* (2), 1069–1079. https://doi.org/10.1007/s12010-014-1266-7.

(23) Szabo, A.; Ostlund, N. S. *Sch*; Dover Publications: Mineola, N.Y, 1996.

(24) Pople, J. A.; Seeger, R.; Krishnan, R. Variational Configuration Interaction Methods and Comparison with Perturbation Theory. *Int. J. Quantum Chem.* **1977**, *12* (S11), 149–163. https://doi.org/10.1002/qua.560120820.

(25) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: Chichester ; New York, 2000.

(26) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. https://doi.org/10.1103/PhysRev.136.B864.

(27) Sholl, D. S.; Steckel, J. A. *Density Functional Theory: A Practical Introduction*; Wiley: Hoboken, N.J, 2009.

(28) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372. https://doi.org/10.1080/00268976.2017.1333644.

(29) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the Design and Selection of Density Functional Approximations: More Constraint Satisfaction with Fewer Fits. *J. Chem. Phys.* **2005**, *123* (6), 062201. https://doi.org/10.1063/1.1904565.

(30) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098.

(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(32) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. https://doi.org/10.1103/PhysRevB.37.785.

(33) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1), 51–57. https://doi.org/10.1016/j.cplett.2004.06.011.

(34) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52* (12), 997–1000. https://doi.org/10.1103/PhysRevLett.52.997.

(35) Schirmer, J. Closed-Form Intermediate Representations of Many-Body Propagators and Resolvent Matrices. *Phys. Rev. A* **1991**, *43* (9), 4647–4659. https://doi.org/10.1103/PhysRevA.43.4647.

(36) Schirmer, J.; Trofimov, A. B. Intermediate State Representation Approach to Physical Properties of Electronically Excited Molecules. *J. Chem. Phys.* **2004**, *120* (24), 11449–11464. https://doi.org/10.1063/1.1752875.

(37) Mertins, F.; Schirmer, J. Algebraic Propagator Approaches and Intermediate-State Representations. I. The Biorthogonal and Unitary Coupled-Cluster

Methods. *Phys. Rev. A* **1996**, *53* (4), 2140–2152. https://doi.org/10.1103/PhysRevA.53.2140.

(38) Trofimov, A. B.; Stelter, G.; Schirmer, J. Electron Excitation Energies Using a Consistent Third-Order Propagator Approach: Comparison with Full Configuration Interaction and Coupled Cluster Results. *J. Chem. Phys.* **2002**, *117* (14), 6402–6410. https://doi.org/10.1063/1.1504708.

(39) Harbach, P. H. P.; Wormit, M.; Dreuw, A. The Third-Order Algebraic Diagrammatic Construction Method (ADC(3)) for the Polarization Propagator for Closed-Shell Molecules: Efficient Implementation and Benchmarking. *J. Chem. Phys.* **2014**, *141* (6), 064113. https://doi.org/10.1063/1.4892418.

(40) Dreuw, A.; Wormit, M. The Algebraic Diagrammatic Construction Scheme for the Polarization Propagator for the Calculation of Excited States. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5* (1), 82–95. https://doi.org/10.1002/wcms.1206.

(41) Knippenberg, S.; Eisenbrandt, P.; Šištík, L.; Slavíček, P.; Dreuw, A. Simulation of Photoelectron Spectra Using the Reflection Principle in Combination with Unrestricted Excitation ADC(2) to Assess the Accuracy of Excited-State Calculations. *ChemPhysChem* **2011**, *12* (17), 3180–3191. https://doi.org/10.1002/cphc.201100485.

(42) Mewes, J.-M.; You, Z.-Q.; Wormit, M.; Kriesche, T.; Herbert, J. M.; Dreuw, A. Experimental Benchmark Data and Systematic Evaluation of Two a Posteriori, Polarizable-Continuum Corrections for Vertical Excitation Energies in Solution. *J. Phys. Chem. A* **2015**, *119* (21), 5446–5464. https://doi.org/10.1021/jp511163y.

(43) Scheurer, M.; Herbst, M. F.; Reinholdt, P.; Olsen, J. M. H.; Dreuw, A.; Kongsted, J. Polarizable Embedding Combined with the Algebraic Diagrammatic Construction: Tackling Excited States in Biomolecular Systems. *J. Chem. Theory Comput.* **2018**, *14* (9), 4870–4883. https://doi.org/10.1021/acs.jctc.8b00576.

(44) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc. Perkin Trans. 2* **1993**, *0* (5), 799–805. https://doi.org/10.1039/P29930000799.

(45) Cossi, M.; Barone, V. Solvent Effect on Vertical Electronic Transitions by the Polarizable Continuum Model. *J. Chem. Phys.* **2000**, *112* (5), 2427–2435. https://doi.org/10.1063/1.480808.

(46) Giovannini, T.; Macchiagodena, M.; Ambrosetti, M.; Puglisi, A.; Lafiosca, P.; Gerfo, G. L.; Egidi, F.; Cappelli, C. Simulating Vertical Excitation Energies of Solvated Dyes: From Continuum to Polarizable Discrete Modeling. *Int. J. Quantum Chem.* **2019**, *119* (1), e25684. https://doi.org/10.1002/qua.25684.

(47) Improta, R.; Barone, V.; Scalmani, G.; Frisch, M. J. A State-Specific Polarizable Continuum Model Time Dependent Density Functional Theory Method for Excited State Calculations in Solution. *J. Chem. Phys.* **2006**, *125* (5), 054103. https://doi.org/10.1063/1.2222364.

(48) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramírez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118* (7), 4071–4113. https://doi.org/10.1021/acs.chemrev.8b00026.

(49) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theory Comput.* **2010**, *6* (12), 3721–3734. https://doi.org/10.1021/ct1003803.

(50) Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. Response Properties of Embedded Molecules through the Polarizable Embedding Model. *Int. J. Quantum Chem.* **2019**, *119* (1), e25717. https://doi.org/10.1002/qua.25717.

(51) Conforti, F.; Menichini, G.; Zanfini, L.; Tundis, R.; Statti, G. A.; Provenzano, E.; Menichini, F.; Somma, F.; Alfano, C. Evaluation of Phototoxic Potential of Aerial Components of the Fig Tree against Human Melanoma. *Cell Prolif.* **2012**, *45* (3), 279–285. https://doi.org/10.1111/j.1365-2184.2012.00816.x.

(52) Kitamura, N.; Kohtani, S.; Nakagaki, R. Molecular Aspects of Furocoumarin Reactions: Photophysics, Photochemistry, Photobiology, and Structural Analysis. *J. Photochem. Photobiol. C Photochem. Rev.* **2005**, *6* (2), 168–185. https://doi.org/10.1016/j.jphotochemrev.2005.08.002.

(53) Postigo, A.; Funes, M.; Petenatti, E.; Bottai, H.; Pacciaroni, A.; Sortino, M. Antifungal Photosensitive Activity of Porophyllum Obscurum (Spreng.) DC.: Correlation of the Chemical Composition of the Hexane Extract with the

Bioactivity. *Photodiagnosis Photodyn. Ther.* **2017**, *20*, 263–272. https://doi.org/10.1016/j.pdpdt.2017.10.023.

(54) Hafeez, B. B.; Mustafa, A.; Fischer, J. W.; Singh, A.; Zhong, W.; Shekhani, M. O.; Meske, L.; Havighurst, T.; Kim, K.; Verma, A. K. α-Mangostin: A Dietary Antioxidant Derived from the Pericarp of Garcinia Mangostana L. Inhibits Pancreatic Tumor Growth in Xenograft Mouse Model. *Antioxid. Redox Signal.* **2013**, *21* (5), 682–699. https://doi.org/10.1089/ars.2013.5212.

(55) Sun, D.; Zhang, S.; Wei, Y.; Yin, L. Antioxidant Activity of Mangostin in Cell-Free System and Its Effect on K562 Leukemia Cell Line in Photodynamic Therapy. *Acta Biochim. Biophys. Sin.* **2009**, *41* (12), 1033–1043. https://doi.org/10.1093/abbs/gmp099.

(56) Srinivas, K. Anticancer and Antimicrobial Activity of Embelin Derivatives. **2010**, *1* (1), 4.

(57) Lee, Y.-J.; Park, B.-S.; Park, H.-R.; Yu, S.-B.; Kang, H.-M.; Kim, I.-R. XIAP Inhibitor Embelin Induces Autophagic and Apoptotic Cell Death in Human Oral Squamous Cell Carcinoma Cells. *Environ. Toxicol.* **2017**, *32* (11), 2371–2378. https://doi.org/10.1002/tox.22450.

(58) Nikolovska-Coleska, Z.; Xu, L.; Hu, Z.; Tomita, Y.; Li, P.; Roller, P. P.; Wang, R.; Fang, X.; Guo, R.; Zhang, M.; Lippman, M. E.; Yang, D.; Wang, S. Discovery of Embelin as a Cell-Permeable, Small-Molecular Weight Inhibitor of XIAP through Structure-Based Computational Screening of a Traditional Herbal Medicine Three-Dimensional Structure Database. *J. Med. Chem.* **2004**, *47* (10), 2430–2440. https://doi.org/10.1021/jm030420+.

(59) Wehrkamp, C. J.; Gutwein, A. R.; Natarajan, S. K.; Phillippi, M. A.; Mott, J. L. XIAP Antagonist Embelin Inhibited Proliferation of Cholangiocarcinoma Cells. *PLoS ONE* **2014**, *9* (3), e90238. https://doi.org/10.1371/journal.pone.0090238.

(60) Juarranz, Á.; Jaén, P.; Sanz-Rodríguez, F.; Cuevas, J.; González, S. Photodynamic Therapy of Cancer. Basic Principles and Applications. *Clin. Transl. Oncol.* **2008**, *10* (3), 148–154. https://doi.org/10.1007/s12094-008-0172-2.

(61) Epstein, J. H. Phototoxicity and Photoallergy. *Semin. Cutan. Med. Surg.* **1999**, *18* (4), 274–284. https://doi.org/10.1016/S1085-5629(99)80026-1.

(62) You, Z.-Q.; Hsu, C.-P.; Fleming, G. R. Triplet-Triplet Energy-Transfer Coupling: Theory and Calculation. *J. Chem. Phys.* **2006**, *124* (4), 044506. https://doi.org/10.1063/1.2155433.

(63) Miyata, K.; Conrad-Burton, F. S.; Geyer, F. L.; Zhu, X.-Y. Triplet Pair States in Singlet Fission. *Chem. Rev.* **2019**, *119* (6), 4261–4292. https://doi.org/10.1021/acs.chemrev.8b00572.

(64) Aravindhan, R.; Sreelatha, T.; Perumal, P. T.; Gnanamani, A. Synthesis, Characterization and Biological Profile of Metal and Azo-Metal Complexes of Embelin. *Complex Met.* **2014**, *1* (1), 69–79. https://doi.org/10.1080/2164232X.2014.886963.

(65) Viault, G.; Grée, D.; Das, S.; Yadav, J. S.; Grée, R. Synthesis of a Focused Chemical Library Based on Derivatives of Embelin, a Natural Product with Proapoptotic and Anticancer Properties. *Eur. J. Org. Chem.* **2011**, *2011* (7), 1233–1241. https://doi.org/10.1002/ejoc.201001627.

(66) Singh, B.; Guru, S. K.; Sharma, R.; Bharate, S. S.; Khan, I. A.; Bhushan, S.; Bharate, S. B.; Vishwakarma, R. A. Synthesis and Anti-Proliferative Activities of New Derivatives of Embelin. *Bioorg. Med. Chem. Lett.* **2014**, *24* (20), 4865–4870. https://doi.org/10.1016/j.bmcl.2014.08.052.

(67) Singh, D.; Singh, R.; Singh, P.; Gupta, R. S. Effects of Embelin on Lipid Peroxidation and Free Radical Scavenging Activity against Liver Damage in Rats. *Basic Clin. Pharmacol. Toxicol.* **2009**, *105* (4), 243–248. https://doi.org/10.1111/j.1742-7843.2009.00429.x.

(68) Turro, N. J.; Ramamurthy, V.; Scaiano, J. C. *Principles of Molecular Photochemistry : An Introduction*; Sausalito, Calif. : University Science Books, 2009.

(69) Yamauchi, S.; W. Pratt, D. Oxygen-17 Hyperfine Structure in the Zero-Field Odmr Spectrum of Triplet State Benzophenone. *Chem. Phys. Lett.* **1978**, *60* (1), 145–149. https://doi.org/10.1016/0009-2614(78)85730-3.

(70) Ou, Q.; Subotnik, J. E. Electronic Relaxation in Benzaldehyde Evaluated via TD-DFT and Localized Diabatization: Intersystem Crossings, Conical Intersections, and Phosphorescence. *J. Phys. Chem. C* **2013**, *117* (39), 19839–19849. https://doi.org/10.1021/jp405574q.

(71) Joshi, R.; Ghanty, T. K.; Mukherjee, T. Formation of Semiquinone Radical in the Reaction of Embelin (2,5-Dihydroxy-3-Undecyl-1,4-Benzoquinone) with Reductants as Well as Oxidants. Characterization by Pulse Radiolysis and Structure Investigation by Quantum Chemical Study. *J. Mol. Struct.* **2009**, *928* (1–3), 46–53. https://doi.org/10.1016/j.molstruc.2009.03.010.

(72) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796–6806. https://doi.org/10.1063/1.462569.

(73) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(74) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133.

(75) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A Direct MP2 Gradient Method. *Chem. Phys. Lett.* **1990**, *166* (3), 275–280. https://doi.org/10.1016/0009-2614(90)80029-D.

(76) Plasser, F.; Thomitzni, B.; Bäppler, S. A.; Wenzel, J.; Rehn, D. R.; Wormit, M.; Dreuw, A. Statistical Analysis of Electronic Excitation Processes: Spatial Location, Compactness, Charge Transfer, and Electron-Hole Correlation. *J. Comput. Chem.* **2015**, *36* (21), 1609–1620. https://doi.org/10.1002/jcc.23975.

(77) Glendening, E. D.; Landis, C. R.; Weinhold, F. NBO 6.0 : Natural Bond Orbital Analysis Program. *J. Comput. Chem.* **2013**, *34* (16), 1429–1437. https://doi.org/10.1002/jcc.23266.

(78) Cossi, M.; Barone, V. Time-Dependent Density Functional Theory for Molecules in Liquid Solutions. *J. Chem. Phys.* **2001**, *115* (10), 4708–4717. https://doi.org/10.1063/1.1394921.

(79) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian~16 Revision B.01*; 2017.

(80) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54* (2), 724–728. https://doi.org/10.1063/1.1674902.

(81) Maeda, S.; Ohno, K.; Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. *J. Chem. Theory Comput.* **2010**, *6* (5), 1538–1545. https://doi.org/10.1021/ct1000268.

(82) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.;

Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215. https://doi.org/10.1080/00268976.2014.952696.

(83) Head-Gordon, M.; Grana, A. M.; Maurice, D.; White, C. A. Analysis of Electronic Transitions as the Difference of Electron Attachment and Detachment Densities. *J. Phys. Chem.* **1995**, *99* (39), 14261–14270. https://doi.org/10.1021/j100039a012.

(84) Kasha, M. Characterization of Electronic Transitions in Complex Molecules. *Discuss. Faraday Soc.* **1950**, *9* (0), 14–19. https://doi.org/10.1039/DF9500900014.

(85) Plasser, F.; Wormit, M.; Dreuw, A. New Tools for the Systematic Analysis and Visualization of Electronic Excitations. I. Formalism. *J. Chem. Phys.* **2014**, *141* (2), 024106. https://doi.org/10.1063/1.4885819.

(86) Mewes, S. A.; Plasser, F.; Dreuw, A. Communication: Exciton Analysis in Time-Dependent Density Functional Theory: How Functionals Shape Excited-State Characters. *J. Chem. Phys.* **2015**, *143* (17), 171101. https://doi.org/10.1063/1.4935178.

(87) Weinhold, F.; Landis, C. R. *Discovering Chemistry with Natural Bond Orbitals*; Wiley: Hoboken, NJ, 2012.

(88) Parmar, K.; Patel, J.; Sheth, N. Formulation and Development of Embelin Liquisolid Systems Using Quality by Design Approach. *J. Pharm. Investig.* **2016**, *46* (6), 547–556. https://doi.org/10.1007/s40005-016-0239-y.

(89) Morton, R. A. *Biochemistry of Quinones*; Academic Press, 1965.

(90) El-Najjar, N.; Gali-Muhtasib, H.; Ketola, R. A.; Vuorela, P.; Urtti, A.; Vuorela, H. The Chemical and Biological Activities of Quinones: Overview and Implications in Analytical Detection. *Phytochem. Rev.* **2011**, *10* (3), 353. https://doi.org/10.1007/s11101-011-9209-1.

(91) Lefrancois, D.; Tuna, D.; Martínez, T. J.; Dreuw, A. The Spin-Flip Variant of the Algebraic-Diagrammatic Construction Yields the Correct Topology of S_1/S_0 Conical Intersections. *J. Chem. Theory Comput.* **2017**, *13* (9), 4436–4441. https://doi.org/10.1021/acs.jctc.7b00634.

(92) Krylov, A. I. Spin-Flip Equation-of-Motion Coupled-Cluster Electronic Structure Method for a Description of Excited States, Bond Breaking, Diradicals, and Triradicals. *Acc. Chem. Res.* **2006**, *39* (2), 83–91. https://doi.org/10.1021/ar0402006.

(93) Zimmerman, P. M.; Bell, F.; Goldey, M.; Bell, A. T.; Head-Gordon, M. Restricted Active Space Spin-Flip Configuration Interaction: Theory and Examples for Multiple Spin Flips with Odd Numbers of Electrons. *J. Chem. Phys.* **2012**, *137* (16), 164110. https://doi.org/10.1063/1.4759076.

(94) Atchity, G. J.; Xantheas, S. S.; Ruedenberg, K. Potential Energy Surfaces near Intersections. *J. Chem. Phys.* **1991**, *95* (3), 1862–1876. https://doi.org/10.1063/1.461036.

(95) Hubig, S. M.; Bockman, T. M.; Kochi, J. K. Identification of Photoexcited Singlet Quinones and Their Ultrafast Electron-Transfer vs Intersystem-Crossing Rates. *J. Am. Chem. Soc.* **1997**, *119* (12), 2926–2935. https://doi.org/10.1021/ja963907z.

(96) O'Connor, A. E.; Gallagher, W. M.; Byrne, A. T. Porphyrin and Nonporphyrin Photosensitizers in Oncology: Preclinical and Clinical Advances in Photodynamic Therapy. *Photochem. Photobiol.* **2009**, *85* (5), 1053–1074. https://doi.org/10.1111/j.1751-1097.2009.00585.x.

(97) Chapter 10 - Programmed Cell Death. In *Medical Cell Biology (Third Edition)*; Goodman, S. R., Ed.; Academic Press: San Diego, 2008; pp 291–307. https://doi.org/10.1016/B978-0-12-370458-0.50015-3.

(98) Hildeman, D. A.; Mitchell, T.; Teague, T. K.; Henson, P.; Day, B. J.; Kappler, J.; Marrack, P. C. Reactive Oxygen Species Regulate Activation-Induced T Cell Apoptosis. *Immunity* **1999**, *10* (6), 735–744. https://doi.org/10.1016/S1074-7613(00)80072-2.

(99) Bai, L.; Smith, D. C.; Wang, S. Small-Molecule SMAC Mimetics as New Cancer Therapeutics. *Pharmacol. Ther.* **2014**, *144* (1), 82–95. https://doi.org/10.1016/j.pharmthera.2014.05.007.

(100) Déas, O.; Dumont, C.; MacFarlane, M.; Rouleau, M.; Hebib, C.; Harper, F.; Hirsch, F.; Charpentier, B.; Cohen, G. M.; Senik, A. Caspase-Independent Cell Death Induced by Anti-CD2 or Staurosporine in Activated Human Peripheral T Lymphocytes. *J. Immunol.* **1998**, *161* (7), 3375–3383.

(101) Susin, S. A.; Lorenzo, H. K.; Zamzami, N.; Marzo, I.; Snow, B. E.; Brothers, G. M.; Mangion, J.; Jacotot, E.; Costantini, P.; Loeffler, M.; Larochette, N.; Goodlett, D. R.; Aebersold, R.; Siderovski, D. P.; Penninger, J. M.; Kroemer, G. Molecular Characterization of Mitochondrial Apoptosis-Inducing Factor. *Nature* **1999**, *397* (6718), 441–446. https://doi.org/10.1038/17135.

(102) Ink, B.; Zörnig, M.; Baum, B.; Hajibagheri, N.; James, C.; Chittenden, T.; Evan, G. Human Bak Induces Cell Death in Schizosaccharomyces Pombe with Morphological Changes Similar to Those with Apoptosis in Mammalian Cells. *Mol. Cell. Biol.* **1997**, *17* (5), 2468–2474.

(103) Xiang, J.; Chao, D. T.; Korsmeyer, S. J. BAX-Induced Cell Death May Not Require Interleukin 1 Beta-Converting Enzyme-like Proteases. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (25), 14559–14563. https://doi.org/10.1073/pnas.93.25.14559.

(104) Buttke, T. M.; Sandstrom, P. A. Redox Regulation of Programmed Cell Death in Lymphocytes. *Free Radic. Res.* **1995**, *22* (5), 389–397. https://doi.org/10.3109/10715769509147548.

(105) Schimmer, A. D.; Welsh, K.; Pinilla, C.; Wang, Z.; Krajewska, M.; Bonneau, M.-J.; Pedersen, I. M.; Kitada, S.; Scott, F. L.; Bailly-Maitre, B.; Glinsky, G.; Scudiero, D.; Sausville, E.; Salvesen, G.; Nefzi, A.; Ostresh, J. M.; Houghten, R. A.; Reed, J. C. Small-Molecule Antagonists of Apoptosis Suppressor XIAP Exhibit Broad Antitumor Activity. *Cancer Cell* **2004**, *5* (1), 25–35. https://doi.org/10.1016/S1535-6108(03)00332-5.

(106) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52* (1), 99–105. https://doi.org/10.1016/j.ymeth.2010.06.002.

(107) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and

Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105. https://doi.org/10.1063/1.3565032.

(108) Taylor, B. C.; Lee, C. T.; Amaro, R. E. Structural Basis for Ligand Modulation of the CCR2 Conformational Landscape. *Proc. Natl. Acad. Sci.* **2019**, *116* (17), 8131–8136. https://doi.org/10.1073/pnas.1814131116.

(109) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. *J. Chem. Phys.* **2009**, *131* (12), 124101. https://doi.org/10.1063/1.3216567.

(110) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a β-Hairpin Peptide. *J. Phys. Chem. B* **2004**, *108* (21), 6582–6594. https://doi.org/10.1021/jp037422q.

(111) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 015102. https://doi.org/10.1063/1.4811489.

(112) Swope, W. C.; Pitera, J. W.; Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108* (21), 6571–6581. https://doi.org/10.1021/jp037421y.

(113) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to Elucidate the Functional Conformational Changes of Complex Biomolecules. *WIREs Comput. Mol. Sci.* **2018**, *8* (1), e1343. https://doi.org/10.1002/wcms.1343.

(114) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140* (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.

(115) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(116) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132–1156. https://doi.org/10.1002/jcc.23905.

(117) Kumari, R.; Kumar, R.; Open Source Drug Discovery Consortium; Lynn, A. G_mmpbsa —A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54* (7), 1951–1962. https://doi.org/10.1021/ci500020m.

(118) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38, 27–28.

(119) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461.

(120) Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S. Discovery of Multiple Hidden Allosteric Sites by Combining Markov State Models and Experiments. *Proc. Natl. Acad. Sci.* **2015**, *112* (9), 2734–2739. https://doi.org/10.1073/pnas.1417811112.

(121) Molgedey, L.; Schuster, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **1994**, *72* (23), 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634.

(122) Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Wiley: New York, 2002.

(123) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

(124) Chan, C.-K.; Hu, Y.; Takahashi, S.; Rousseau, D. L.; Eaton, W. A.; Hofrichter, J. Submillisecond Protein Folding Kinetics Studied by Ultrarapid Mixing. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (5), 1779–1784.

(125) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. The Speed Limit for Protein Folding Measured by Triplet–Triplet Energy Transfer. *Proc. Natl. Acad. Sci.* **1999**, *96* (17), 9597–9601. https://doi.org/10.1073/pnas.96.17.9597.

(126) Wehmeyer, C.; Scherer, M. K.; Hempel, T.; Husic, B. E.; Olsson, S.; Noé, F. Introduction to Markov State Modeling with the PyEMMA Software [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1* (1). https://doi.org/10.33011/livecoms.1.1.5965.

(127) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and Hidden Markov Models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* **2013**, *139* (18), 184114. https://doi.org/10.1063/1.4828816.

(128) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151* (1), 146–168. https://doi.org/10.1006/jcph.1999.6231.

(129) Ishikita, H.; Saito, K. Proton Transfer Reactions and Hydrogen-Bond Networks in Protein Environments. *J. R. Soc. Interface* **2014**, *11* (91). https://doi.org/10.1098/rsif.2013.0518.

(130) Tellone, E.; Galtieri, A.; Barreca, D.; Ficarra, S. Chapter 2.8 - Coenzyme Q10 and Embelin. In *Nonvitamin and Nonmineral Nutritional Supplements*; Nabavi, S. M., Silva, A. S., Eds.; Academic Press, 2019; pp 69–73. https://doi.org/10.1016/B978-0-12-812491-8.00009-6.

(131) Knight, T.; Luedtke, D.; Edwards, H.; Taub, J. W.; Ge, Y. A Delicate Balance – The BCL-2 Family and Its Role in Apoptosis, Oncogenesis, and Cancer Therapeutics. *Biochem. Pharmacol.* **2019**, *162*, 250–261. https://doi.org/10.1016/j.bcp.2019.01.015.

(132) *Amino Acids, Peptides and Proteins in Organic Chemistry*; Hughes, A. B., Ed.; Wiley-VCH: Weinheim, 2009.

(133) Inaba, H.; Martin, W.; Ardito, M.; De Groot, A. S.; De Groot, L. J. The Role of Glutamic or Aspartic Acid in Position Four of the Epitope Binding Motif and Thyrotropin Receptor-Extracellular Domain Epitope Selection in Graves' Disease. *J. Clin. Endocrinol. Metab.* **2010**, *95* (6), 2909–2916. https://doi.org/10.1210/jc.2009-2393.

(134) Dixit, S. B.; Chipot, C. Can Absolute Free Energies of Association Be Estimated from Molecular Mechanical Simulations? The Biotin−Streptavidin System Revisited. *J. Phys. Chem. A* **2001**, *105* (42), 9795–9799. https://doi.org/10.1021/jp011878v.

(135) Kollman, Peter. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417. https://doi.org/10.1021/cr00023a004.

(136) Genheden, S.; Nilsson, I.; Ryde, U. Binding Affinities of Factor Xa Inhibitors Estimated by Thermodynamic Integration and MM/GBSA. *J. Chem. Inf. Model.* **2011**, *51* (4), 947–958. https://doi.org/10.1021/ci100458f.

(137) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free Energy Calculations for Theophylline Binding to an RNA Aptamer: Comparison of MM-PBSA and Thermodynamic Integration Methods. *Biopolymers* **2003**, *68* (1), 16–34. https://doi.org/10.1002/bip.10270.

(138) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10* (5), 449–461. https://doi.org/10.1517/17460441.2015.1032936.

(139) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. In *Advances in Protein Chemistry*; Protein Simulations; Academic Press, 2003; Vol. 66, pp 123–158. https://doi.org/10.1016/S0065-3233(03)66004-3.

(140) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35* (6), 358–365. https://doi.org/10.1021/ar010014p.

(141) Vergara-Jaque, A.; Comer, J.; Monsalve, L.; González-Nilo, F. D.; Sandoval, C. Computationally Efficient Methodology for Atomic-Level Characterization of Dendrimer–Drug Complexes: A Comparison of Amine- and Acetyl-Terminated PAMAM. *J. Phys. Chem. B* **2013**, *117* (22), 6801–6813. https://doi.org/10.1021/jp4000363.

(142) Yeggoni, D. P.; Rachamallu, A.; Subramanyam, R. Protein Stability, Conformational Change and Binding Mechanism of Human Serum Albumin upon Binding of Embelin and Its Role in Disease Control. *J. Photochem. Photobiol. B* **2016**, *160*, 248–259. https://doi.org/10.1016/j.jphotobiol.2016.04.012.

(143) Faraji, S.; Dreuw, A. Insights into Light-Driven DNA Repair by Photolyases: Challenges and Opportunities for Electronic Structure Theory. *Photochem. Photobiol.* **2017**, *93* (1), 37–50. https://doi.org/10.1111/php.12679.

(144) Lin, H.; Truhlar, D. G. Redistributed Charge and Dipole Schemes for Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Phys. Chem. A* **2005**, *109* (17), 3991–4004. https://doi.org/10.1021/jp0446332.

(145) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–Molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119* (7), 3599–3605. https://doi.org/10.1063/1.1591727.

(146) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48* (7), 1198–1229. https://doi.org/10.1002/anie.200802019.

(147) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.

(148) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited States in Large Molecular Systems through Polarizable Embedding. *Phys. Chem. Chem. Phys.* **2016**, *18* (30), 20234–20250. https://doi.org/10.1039/C6CP03834D.

List of Tables

List of Figures

Publications

1. **Michael Rogo Opata** and **Andreas Dreuw**: Embelin's Versatile Photochemistry Makes it a Potent Photosensitizer for Photodynamic Therapy. *J. Phys. Chem. B 2021, 125, 3527-3537.*
2. **Michael Rogo Opata** and **Andreas Dreuw**: Characterization of Embelin-BIR3 binding patterns using molecular dynamics and Markov state models. J. Phys. J. Phys. Chem. B 2021, (Submitted)
3. **Michael Rogo Opata** and **Andreas Dreuw**: The Photochemistry of embelin in BIIR3 protein environment: A QM/MM study (*in preparation*).

Acknowledgements

My sincere thank you message goes to Prof. Dr. Andreas Dreuw who not only accepted me as a Ph.D. student in his research group but offered me a home for the last four years. Thank you for the useful discussions that we had and most importantly being patient with me during my studies here in Heidelberg as a father would be to his own child. I have had tremendous academic freedom to explore many ideas under your guidance which helped me appreciate the beauty of science in general and computational chemistry in particular. It was a rare gesture and I hope we will continue engaging intellectually after my graduation.

The AK Dreuw group has been a source of inspiration and help. Thank you, Dr. Maximillian Karämer, Dr. Adrian Dempwolff, Dr. Marvin Hoffman, Dr. Jie Han, Dr. Manuel Hodecker, Dr. Dirk Rhen, Reena Sen, Maximillian Scheurer, Sebastian Thielen, Benjamin Thomitzni, and Mikael Scott for the useful discussions and memories we shared as a group over the years. Not to forget Miss. Ellen Vogel, our ever-keen secretary who handled all my paperwork and reminded me of any upcoming deadline and most of all, great advice on how to overcome looming challenges during my studies. Thanks to Herr Manfred (a.k.a root), our computer administrator for all the assistance you offered me.

It would be impossible to survive without financial aid during my Ph.D. Special thanks goes to my main sponsors Deutscher Akademischer Austauschdienst (DAAD) who awarded me a scholarship to pursue a lifetime dream. I am grateful for the opportunity that you gave me not only to explore science but also to learn a new language, people, and culture. You made my life easier while staying in Germany.
Lastly, I would like to thank my parents for waiting so long to see their first-born son make it this far. You have always been around my success and failures this far. Special thanks go to my wife and son who were a source of inspiration at my lowest moments.

Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die Naturwissenschaftlich-Mathematische Gesamtfakultät der Universität Heidelberg

1. Bei der eingereichten Dissertation zu dem Thema

   "The Mode of Action of Embelin as an Alternative Photosensitizer for Photodynamic Therapy" handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Fol- gen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe

Ort/Datum

Unterschrift