

# Modeling the Interactions of Anticancer Compounds with DNA and Lipid Membranes

Gustavo Adolfo Cárdenas

Doctoral Thesis

Madrid 2022



Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Química

# Modeling the Interactions of Anticancer Compounds with DNA and Lipid Membranes

A thesis submitted for the degree of Doctor

Candidate: Gustavo Adolfo Cárdenas

Supervisor: Dr. Juan José Nogueira Pérez

Madrid, July 2022

---

# Contents

<b>List of Figures</b>	<b>10</b>
<b>List of Tables</b>	<b>11</b>
<b>Abstract</b>	<b>21</b>
<b>Resumen</b>	<b>25</b>
<b>1 Introduction</b>	<b>26</b>
1.1 Overview of the Research in Anticancer Agents . . . . .	26
1.2 Computational Techniques Applied on Related Anticancer Systems . . . . .	31
1.3 Objectives of the Thesis . . . . .	32
<b>2 Theoretical Methods</b>	<b>35</b>
2.1 Electronic Structure Theory: Ab Initio Methods . . . . .	35
2.1.1 The Born-Oppenheimer Approximation . . . . .	35
2.1.2 The Hartree-Fock Approximation . . . . .	38
2.1.3 Møller-Plesset Perturbation Theory . . . . .	41
2.1.4 Configuration Interaction . . . . .	43
2.1.5 Multiconfigurational Self-Consistent Field Methods . . . . .	45
2.1.6 Complete Active Space Perturbation Theory . . . . .	48
2.2 Electronic Structure Theory: Density Functional Theory . . . . .	53
2.2.1 The Hohenberg-Kohn Theorems . . . . .	54
2.2.2 The Kohn-Sham Equations . . . . .	56
2.2.3 Exchange Correlation Functionals . . . . .	59

---

2.2.4	Time-Dependent Density Functional Theory . . . . .	60
2.3	Electronic Structure Theory: Energy Decomposition Analysis . . . . .	66
2.4	Classical Molecular Dynamics . . . . .	73
2.4.1	Classical Equations of Motion . . . . .	75
2.4.2	Integration Methods . . . . .	77
2.4.3	Classical Potentials: Force Fields and Molecular Mechanics . . . . .	79
2.4.4	Ab Initio Molecular Dynamics . . . . .	80
2.4.5	Enhanced Sampling: the Umbrella Sampling Technique . . . . .	81
2.5	Multiscale Hybrid Methods: Quantum Mechanics/Molecular Mechanics . . . . .	85
<b>3</b>	<b>MoBioTools: A Toolkit to Setup QM/MM Calculations</b>	<b>90</b>
	Gustavo Cárdenas, Jesús Lucia-Tamudo, Henar Mateo-delaFuente, Vito F. Palmisano, Nuria Anguita-Ortiz, Lorena Ruano, Álvaro Pérez-Barcia, Sergio Díaz-Tendero, Marcos Mandado, Juan J. Nogueira	
3.1	Introduction . . . . .	91
3.2	Methodology . . . . .	94
3.2.1	The General Purpose QM/MM Input Generator . . . . .	96
3.2.2	The Active Space Automatic Preservation . . . . .	101
3.3	Applications . . . . .	106
3.3.1	Reduction Potential: Canonical Nucleobases in Acetonitrile . . . . .	106
3.3.2	Energy Decomposition Analysis: Tyrosine in Water . . . . .	110
3.3.3	Absorption Spectrum: p-Diaminoazobenzene Integrated into the Human Nav1.4 Channel . . . . .	112
3.3.4	Emission and Absorption Spectra: Luciferine/Luciferase Complex . . . . .	117
3.4	Conclusions . . . . .	120
<b>4</b>	<b>Active Space Preservation for CASSCF Calculations on Ensembles of Geometries</b>	<b>121</b>
	Gustavo Cárdenas, Juan J. Nogueira	
4.1	Introduction . . . . .	122
4.2	Theory and Implementation . . . . .	125
4.2.1	Calculation of the MO Overlap Matrix . . . . .	125

---

4.2.2	Comparison between the MO Sets . . . . .	130
4.3	Computational Details . . . . .	135
4.4	Results and Discussion . . . . .	137
4.5	Conclusions . . . . .	149
<b>5</b>	<b>The Permeation Mechanism of Cisplatin Through a Dioleoylphosphocholine Bilayer</b>	<b>152</b>
	Lorena Ruano, Gustavo Cárdenas, Juan J. Nogueira	
5.1	Introduction . . . . .	153
5.2	Computational Details . . . . .	156
5.3	Results and Discussion . . . . .	158
5.3.1	Equilibration and Umbrella Sampling Setup . . . . .	158
5.3.2	Characterization of the Intermolecular Interactions . . . . .	163
5.3.3	The Role of Water Dehydration and Membrane Rigidity . . . . .	170
5.4	Conclusions . . . . .	172
<b>6</b>	<b>Characterization of Cisplatin/Membrane Interactions by QM/MM Energy Decomposition Analysis.</b>	<b>176</b>
	Gustavo Cárdenas, Álvaro Pérez-Barcia, Marcos Mandado, Juan J. Nogueira	
6.1	Introduction . . . . .	177
6.2	Methodology . . . . .	180
6.3	Computational Details . . . . .	182
6.4	Results and Discussion . . . . .	183
6.5	Conclusions . . . . .	190
<b>7</b>	<b>Stacking Effects on Anthraquinone/DNA Charge-Transfer Electronically Excited States</b>	<b>192</b>
	Gustavo Cárdenas, Juan J. Nogueira	
7.1	Introduction . . . . .	193
7.2	Computational Details . . . . .	196
7.3	Results and Discussion . . . . .	198
7.3.1	Sampling the Stacking Binding Pocket . . . . .	198

7.3.2	Electronically Excited States: Delocalization and Charge Transfer . . . . .	202
7.4	Conclusions . . . . .	211
<b>8</b>	<b>Permeation Mechanism and Excited States of Anthraquinones inside a Lipid Bilayer</b>	<b>213</b>
	Gustavo Cárdenas, Davide Avagliano, Marco Garavelli, Juan J. Nogueira	
8.1	Introduction . . . . .	214
8.2	Computational Details . . . . .	216
8.3	Results and Discussion . . . . .	219
8.3.1	Umbrella Sampling Simulations . . . . .	219
8.3.2	Electronically Excited States . . . . .	221
8.4	Conclusions and Future Work . . . . .	229
<b>9</b>	<b>Conclusions</b>	<b>231</b>
<b>A</b>	<b>Supplementary Information for Chapter 4</b>	<b>239</b>
A.1	Active Spaces of the Five Canonical Nucleobases . . . . .	239
A.2	Choice of the Reference Active Space of Thymine . . . . .	243
A.3	Analysis of the CI Weights for Two Geometries of Uracil . . . . .	248
<b>B</b>	<b>Supplementary Information for Chapter 5</b>	<b>250</b>
B.1	Umbrella Sampling Setup . . . . .	250
B.2	Permeability Coefficient . . . . .	252
B.3	Force Field Parameters . . . . .	253
<b>C</b>	<b>Supplementary Information for Chapter 8</b>	<b>254</b>
C.1	Electronic Excited States and CI Weights of Rubiadin and Soranjidiol in Explicit Chloroform . . . . .	254

# List of Figures

1.1	Mechanism of action of anticancer molecules . . . . .	28
2.1	Representation of the potential energy profile along the reaction coordinate $\xi$	84
2.2	Representation of the QM/MM partitioning scheme of a molecular system consisting of an anthraquinone molecule (inner region, red) intercalated in a double strand of DNA (outer region, blue). . . . .	86
3.1	Schematic representation of the working principle of the MoBioTools toolkit.	95
3.2	Representation of the contents of the two input files for the main_qminputs.py driver . . . . .	97
3.3	Schematic representation of the working principle of the algorithm to correct the active space on an ensemble of sampled geometries. . . . .	103
3.4	Structure and main dependencies for the pyoverlaps.py script. . . . .	105
3.5	Main and template inputs, and one-electron redox potentials for the nucleobases in acetonitrile. . . . .	109
3.6	Main and template inputs, and probability distributions of the EDA components of tyrosine in water. . . . .	113
3.7	Representation of the molecular orbitals included in the active space. The $n^+$ and $n^-$ orbitals are the lone pairs of the central diazine bond, $\pi_{1-4}$ are four occupied $\pi$ orbitals and $\pi_{1-4}^*$ are four unoccupied $\pi^*$ orbitals. . . . .	115
3.8	Active space of p-diaminoazobenzene. . . . .	117
3.9	Main and template inputs, and emission and absorption spectra of the oxyluciferin/luciferase complex. . . . .	119

---

4.1	Schematic representation of the algorithm to correct the active space of an ensemble of geometries. . . . .	132
4.2	Analysis of the overlap matrix of a sampled geometry of uracil. . . . .	138
4.3	Two examples of geometries of uracil whose active spaces were recovered after a $\mathbf{S}^{MO}$ analysis. . . . .	139
4.4	Comparison of SA-CASSCF excitation energies towards the $S_1$ states between uncorrected and corrected active spaces . . . . .	144
4.5	Comparison of energy differences and density of states between SA-CASSCF calculations with the uncorrected and corrected active spaces. . . . .	147
4.6	Comparison of energy differences and density of states between MS-CASPT2 calculations with the uncorrected and corrected active spaces. . . . .	148
5.1	Analysis of the equilibration process of the DOPC membrane. . . . .	159
5.2	Convergence analysis of the umbrella sampling simulations. . . . .	162
5.3	Representative snapshot of cisplatin interacting with the residue PC382. . .	164
5.4	Time evolution of the number of interatomic contacts between cisplatin and the five residues that present the highest interaction energy with the drug. .	166
5.5	Contribution of phosphocholine (PC) head groups and dioleoyl (OL) tails to the cisplatin/membrane interaction energy along the reaction coordinate. .	168
5.6	Decomposition of the interaction energy between cisplatin and phosphocholine (PC) groups. . . . .	170
5.7	Cisplatin dehydration and membrane ordering relevance. . . . .	173
6.1	The cisplatin molecule embedded in a DOPC lipid membrane, and PMF profile of the permeation process. . . . .	179
6.2	Convergence of the interaction energy and the EDA components as the number of DOPC molecules in the QM region increases. . . . .	185
6.3	Energy distributions of the contributions to the total interaction energy stemming from the QM/MM EDA between cisplatin and a DOPC membrane. .	187
6.4	Linear correlation coefficients of the electrostatic, induction, dispersion and Pauli energies. . . . .	189

---

7.1	Conformational analysis of anthraquinone intercalated in a poly(dG-dC) decamer. . . . .	200
7.2	Analysis of the stacking between anthraquinone and guanine. . . . .	202
7.3	Representation of the excited states formed in the QM system consisting of anthraquinone and guanine . . . . .	204
7.4	Decomposition of the density of states and absorption spectrum of the anthraquinone-guanine system. . . . .	206
7.5	Contribution of monomer, exciton, excimer and charge-transfer states to the density of states for the symmetric 2 configuration. . . . .	209
7.6	Probability distribution of the position of the hole ( $POS_i$ ) and of the position of the excited electron ( $POSt_f$ ). . . . .	210
8.1	Top: Schematic representation of an anthraquinone derivative embedded inside the DOPC lipid membrane. The polar and non-polar regions of the membrane are represented in blue and red, respectively. Bottom: Potential of mean force (PMF) of the permeation of the rubiadin (green line) and the soranjidiol (orange line) molecules, considering the simulation interval time between 4 and 30 ns for each umbrella sampling window. . . . .	220
8.2	Cluster analysis of rubiadin (top) and of soranjidiol (bottom) along a 100 ns classical molecular dynamics simulation in chloroform. The inset of each plot represents the most representative cluster for each of the anthraquinone derivatives. . . . .	223
8.3	Electronic absorption spectra of rubiadin and soranjidiol in $CHCl_3$ . . . . .	225
8.4	(12,9) Active space of rubiadin. . . . .	226
8.5	(12,9) Active space of soranjidiol. . . . .	226
8.6	Top: (black) Experimental absorption spectrum of rubiadin in chloroform; (green) Absorption spectrum of rubiadin at the minimum of the umbrella sampling free energy profile inside the DOPC lipid membrane. Bottom: (black) Experimental absorption spectrum of soranjidiol in chloroform; (orange) Absorption spectrum of soranjidiol at the minimum of the umbrella sampling free energy profile inside the DOPC lipid membrane. . . . .	228

---

A.1	The (14,10) reference active space of uracil. . . . .	240
A.2	The (14,10) reference active space of cytosine. . . . .	240
A.3	The (14,10) reference active space of thymine. . . . .	241
A.4	The (18,13) reference active space of adenine. . . . .	241
A.5	The (20,14) reference active space of guanine. . . . .	242
A.6	Thiel's (16, 11) reference active space for thymine. . . . .	243
A.7	Molecular orbitals 23 and 24 of the sampled geometry 1 of thymine, after the fourth SA-CASSCF optimization cycle. . . . .	244
A.8	Atom labels for thymine. . . . .	245
B.1	Reaction coordinate probability distributions for the windows centred at 10.0, 10.5, 11.0, 11.5 and 12.0 Å. . . . .	251
B.2	Z-dependent diffusion coefficient of cisplatin. . . . .	252

# List of Tables

4.1	$S^{MO}$ analysis on the overall set of sampled geometries for all five canonical nucleobases. . . . .	143
A.1	Atomic orbital contributions to the molecular orbital 28 of the large reference active space of thymine. . . . .	245
A.2	Excitation energies and configuration weights for the excited states of thymine with the large active space . . . . .	246
A.3	Excitation energies and configuration weights for the excited states of thymine with the small active space . . . . .	247
A.4	Difference between the weights of the dominant configuration of the $S_1$ state for the corrected and uncorrected active spaces for two geometries of uracil. . . . .	249
B.1	Force field parameters for cisplatin . . . . .	253
C.1	Absorption spectrum and configuration weights for the excited states of rubiadin. . . . .	255
C.2	Absorption spectrum and configuration weights for the excited states of soranjidiol. . . . .	256



---

# List of Abbreviations

**AIMD** *Ab Initio* Molecular Dynamics.

**AO** Atomic Orbital.

**CASPT2** Complete Active Space Second-order Perturbation Theory.

**CASSCF** Complete Active Space Self Consistent Field.

**CI** Configuration Interaction.

**DFT** Density Functional Theory.

**DOPC** 1,2-Dioleoyl-sn-Glycero-3-Phosphocholine.

**EDA** Energy Decomposition Analysis.

**HF** Hartree-Fock.

**MCSCF** Multiconfigurational Self Consistent Field.

**MD** Molecular Dynamics.

**MO** Molecular Orbital.

**MP2** Møller-Plesset Second-order Perturbation Theory.

**PDT** Photodynamic Therapy.

**PS** Photosensitizer.

---

**QM/MM** Quantum Mechanics/Molecular Mechanics.

**TD-DFT** Time-Dependent Density Functional Theory.



# Acknowledgements

Todos sabemos que las posiciones más importantes (es decir, las primeras que las personas buscan al leer los agradecimientos) son la primera y la última. Así, pues, no puedo no abrir este capítulo agradeciendo al Dr. Juan José Nogueira, el *Jefazo*, porque este trabajo de tesis no habría sido posible sin él. Pero no me refiero solo al manuscrito de la tesis, sino a todas las cosas que he aprendido con él en estos tres años que me han hecho crecer tanto desde el punto de vista científico, como desde el punto de vista humano.

Muchas gracias a mis compañeros de despacho, pasados y presentes, por haber hecho (aún) más afables las jornadas de trabajo: gracias Eva, Darío, Juanito, Ransel, Enrique y Javi por vuestra compañía, por todos los buenos momentos, por los cineforos y por haberme soportado durante estos tres años. Muchas gracias también a todos los miembros del grupo de Juanjo: Vito, Nuria, Jesús, Lorena y Henar, tanto por todas las discusiones científicas como por las (en ocasiones forzadas) pausas café en vuestro despacho. Así mismo, muchas gracias a todos los miembros del Departamento de Química de la UAM, por la buena atmosfera de trabajo que habéis creado.

Me gustaría agradecer también a nuestros colaboradores en Vigo, en particular al Prof. Marcos Mandado y al Dr. Nicolás Ramos, por las discusiones científicas en ausencia, y en presencia, de una botella de licor café. Igualmente, muchas gracias a nuestros colaboradores en Santiago, Emilio y Saulo, con quienes tuve la oportunidad de trabajar al comienzo de la tesis, así como de tomar el caffettino sagrado a las 10 a.m. Y ya que he mencionado a Santiago, aunque no “colaboremos”, no podía dejarte fuera de los agradecimientos Davith (Ferrote), aunque hayas ganado en street fighter con los ojos cerrados.

Grazie al Prof. Marco Garavelli e a tutti i membri del suo gruppo per avermi ospitato a Bologna durante le fasi finali della mia tesi.

---

Muchas gracias a la Comunidad de Madrid, que a través del Proyecto Talento ha financiado este trabajo de tesis.

Thanks a lot to the HPC Europa3 Project, which financed a 3-week collaboration internship in Bologna (application HPC1794KRI), and the CINECA for computational resources.

Agradezco a la Comunidad de Madrid por el soporte cinanciero a través del proyecto de Atracción de Talento Investigador (referencia 2018-T1/BMD-10261) y al centro de computación científica (CCC) de la Universidad Autónoma de Madrid por los recursos computacionales y el tiempo de cálculo ofrecidos.

Muchas gracias a mi familia, a mi papá, a mi mamá y a Giovanni, por el apoyo incondicional que me han dado. E infine, grazie Annina, senza di te non avrei mai raggiunto questo traguardo, sei stata al mio fianco nel bene e nel male.

# Abstract

This thesis presents a systematic theoretical study on some anticancer molecules to gain unprecedented insights on some of the key steps in their mechanism of action. The anti-carcinogenic molecules under study consist of cis-diaminedichloroplatinum(II) (cisplatin), anthraquinone and some of its derivatives, which present different mechanisms of action. Specifically, although both molecules permeate the cell membrane of a cancer cell prior to triggering their anticancer activity, the main difference resides in the fact that cisplatin is active at the electronic ground state, whereas the anthraquinone molecules need to be photoexcited by irradiating them with a suitable wavelength of electromagnetic radiation to trigger their anticancer activity. The mechanism of action also depends on the target tissue of the drug, which could involve the cellular lipid membrane or the DNA present in some organelles such as the cell nucleus or the mitochondria. Cisplatin is perhaps the most representative – as well as the most widely used – compound of a family of Pt(II) and Pt(IV) chemotherapeutic drugs, which present a common mechanism of action consisting of the permeation of the drug inside the carcinogenic cell followed by binding to a DNA strand, in a process that ultimately leads to the cell death. As such, cisplatin is used as scaffold that can be tailored by modifying its functional groups to tune its cytotoxic activity, selectivity, specificity, and other chemical and biological properties. Likewise, anthraquinone can be considered as the “progenitor” of a family of anticancer chromophores which bears its name, that are active following photoexcitation towards a certain electronic excited state. The mechanism of action of anthraquinones differs from that of the platinum drugs, in that they react with the neighbouring tissues (usually the lipid membrane or DNA) by means of either electron transfer or energy transfer reactions that generate reactive radical species, ultimately disrupting the structure of these tissues and inducing cell death. As

---

such, anthraquinone also represents a scaffold prone to chemical modification to obtain more efficient drugs. Thus, it is the fact of cisplatin and anthraquinone of being the “progenitors” of the corresponding families, that poses the need for studying to a great detail the key features of the mechanisms of action of these drugs, so that future tailoring of anticancer derivatives can be performed in an efficient manner.

For this reason, the present thesis focuses on modeling at an atomistic scale the permeation of these drugs into a lipid membrane, as well as the interactions of the drugs with a model of a double strand of DNA. These studies involve state-of-the art computational techniques that include classical molecular dynamics (MD) to study the conformational motion of the systems of interest and quantum mechanics to describe at a high level of accuracy the electronic structure of the drugs under study, which ultimately is the main feature that determines the chemistry of these drugs, and therefore their cytotoxic activity. In this work, a large amount of effort is dedicated to obtaining an accurate description of the interactions between the above mentioned tissues and the drugs, in particular the influence the complex biological environment that surrounds the drug has upon the electronic structure of the latter. Therefore, sophisticated quantum mechanics/molecular mechanics (QM/MM) techniques are employed to accurately compute physical observables, such as interaction energies and electronic absorption spectra, which give an account on the nature of the interactions established between the drug and the target tissue. In this thesis, the approach adopted is to interface the generation of an ensemble of geometries *via* classical MD to account for conformational motion, with the calculation of a given observable on top of each sampled geometry by means of QM/MM so that the observable of interest is obtained as an ensemble average. Although nowadays there are a manifold of software available to either perform the classical MD simulations or the QM/MM calculations, the resources at hand to readily interface both methodologies in the manner described above are limited, and in most cases lack of the generality (in the observables to be computed, the systems to be treated, the difficulty in their application, etc.) required by the present thesis. Therefore, part of this thesis has been dedicated to the development of a set of computational tools that interfaces the MD and the QM/MM methodologies in a straightforward manner, and with the degree of generality required by the systems under study. The present thesis consists of 8 Chapters, plus the conclusions and

---

perspectives, which are organized as follows: Chapter 1 provides a general introduction on the mechanism of action of cisplatin and anthraquinone derivatives, and the state-of-the-art of the computational techniques employed in the study of analogous anticancer drugs. Chapter 2 gives an overview of the theoretical methods used in the present thesis, spanning from the equations of motion in classical MD to the electronic structure theory to describe ground and excited states of molecular systems at different levels of approximation. The main results of this thesis are presented from Chapter 3 to Chapter 8. In particular, Chapters 3 and 4 introduce the computational tools developed to facilitate and automatize the procedure to perform QM/MM computations. A particular emphasis is done in the application of wavefunction methods, such as the Complete Active Space Self Consistent Field (CASSCF), on ensembles of geometries, for which an algorithm has been proposed to preserve the active space along the ensemble of geometries. Chapter 5 focuses on the study of the permeation mechanism of cisplatin through a lipid membrane model, in which classical MD simulations using a force field were performed in conjunction with an enhanced sampling technique called umbrella sampling, to compute the energetics of the permeation process. The interaction energies (and their decomposition in energy components) between cisplatin and the lipid membrane were also computed using a force field. Chapter 6 represents a follow-up of the work in Chapter 5, where the interaction energy and the energy decomposition were computed using a QM/MM version of an energy decomposition analysis based on electron densities. Chapter 7 presents a conformational study of the anthraquinone molecule when intercalated in a double strand of DNA, in which two predominant orientations of the anthraquinone molecule relative to the surrounding nucleobases were identified. The conformational analysis was followed by the computation of the excited states of the anthraquinone molecule on different orientations, whereby the excited states were characterized in terms of their charge transfer and delocalization nature. Chapter 8 provides some results of a work currently in progress, in which the permeation mechanism of some derivatives of the anthraquinone molecule into a lipid membrane model, and the excited states of these molecules in the presence of the lipid environment, have been studied. Finally, Chapter 9 provides some general conclusions and perspectives of the present thesis work.

# Resumen

En esta tesis se presenta un estudio teórico sistemático de algunas moléculas anticancerígenas para obtener información relevante sobre los pasos clave de su mecanismo de acción. Las moléculas anticancerígenas estudiadas son el *cis*-diaminodicloroplatino(II) (cisplatino), la antraquinona y algunos de sus derivados, las cuales presentan diferentes mecanismos de acción. En particular, si bien ambas moléculas permean la membrana lipídica de una célula cancerígena antes de desencadenar su actividad, la diferencia principal reside en que el cisplatino es activo en el estado electrónico fundamental, mientras que las antraquinonas debe ser fotoexcitadas mediante radiación electromagnética a una longitud de onda propicia para desencadenar su actividad anticancerígena. El mecanismo de acción también depende del tejido diana del fármaco, lo que podría involucrar a la membrana lipídica de la célula o el ADN presente en algunos organelos tales como el núcleo celular o las mitocondrias. El cisplatino es probablemente el compuesto más representativo – así como el más ampliamente empleado – de una familia de medicamentos de Pt(II) y Pt(IV) empleados en quimioterapia, los cuales presentan un mecanismo de acción común; éste consiste en la premeación del medicamento al interior de la célula cancerígena seguida de su asociación con una hélice de ADN, en un proceso que conlleva a la muerte celular. Como tal, el cisplatino es utilizado como andamio que puede ser confeccionado modificando sus grupos funcionales para modular su actividad citotóxica, selectividad, especificidad, y otras propiedades químicas y biológicas. Análogamente, la antraquinona puede ser considerada como “progenitora” de una familia de cromóforos anticancerígenos que comparte su nombre, los cuales son activos tras haber sido fotoexcitados a un específico estado electrónico excitado. El mecanismo de acción de las antraquinonas difiere de aquel del cisplatino, en cuanto a que las primeras reaccionan con los tejidos adyacentes (usualmente la membrana lipídica

---

o el ADN) mediante reacciones de transferencia electrónica o transferencia de energía que generan especies de radicales libres activos, alterando la estructura de dichos tejidos y, como consecuencia, induciendo la muerte celular. Como tal, la antraquinona también representa un andamio que puede ser modificado químicamente para obtener medicamentos más eficientes. Por tanto, el hecho de que tanto el cisplatino como la antraquinona sean los progenitores de sus correspondientes familias conduce a la necesidad de estudiar con detalle las características clave de los mecanismos de acción de dichos fármacos, de manera que moléculas anticancerígenas derivadas de ellas puedan ser diseñados de forma eficiente.

Por este motivo, la presente tesis se concentra en la modelización a una escala atómica de la permeación de dichos fármacos al interior de una membrana lipídica, así como sus interacciones con un modelo de doble hélice de ADN. Estos estudios involucran técnicas computacionales de vanguardia que incluyen dinámica molecular clásica (MD) para estudiar el movimiento conformacional de los sistemas estudiados, y mecánica cuántica para describir la estructura electrónica de los fármacos estudiados a un elevado nivel de exactitud; ésta es, a fin de cuentas, la característica principal que determina la química de dichos fármacos y, por consiguiente, su actividad citotóxica. En este trabajo, una gran cantidad de esfuerzo es dedicada a la obtención de una descripción exacta de las interacciones entre los tejidos diana mencionados anteriormente y los fármacos, en particular la influencia del ambiente biológico complejo que rodea al fármaco sobre su estructura electrónica. A tal efecto, técnicas sofisticadas de mecánica cuántica/mecánica molecular (QM/MM) son empleadas para calcular de forma exacta observables físicos tales como energías de interacción y espectros de absorción electrónicos, las cuales proveen información acerca de la naturaleza de las interacciones que se establecen entre el medicamento y el tejido diana. En esta tesis, la estrategia adoptada consiste en interrelacionar la generación de un conjunto de geometrías a través de MD clásica para considerar el movimiento conformacional de las moléculas, con el cálculo de un observable físico sobre cada una de las geometrías del conjunto mediante QM/MM de manera que el observable de interés sea obtenido como una media estadística. Si bien hoy en día hay una variedad de programas disponibles para llevar a cabo las simulaciones MD o los cálculos QM/MM, los recursos a disposición para interrelacionar ambas metodologías con la estrategia mencionada anteriormente son limitados, y en la mayor parte de casos no son suficientemente generales (en los observables

---

que deben ser calculados, en los sistemas tratados, en la dificultad de su aplicación, etc.) con respecto de lo requerido por la presente tesis. Por este motivo, una parte de esta tesis ha sido dedicada al desarrollo de un conjunto de herramientas computacionales que conectan las metodologías MD y QM/MM de manera fácil y suficientemente general con respecto de los sistemas estudiados. La presente tesis consiste en 8 Capítulos, más las conclusiones y las perspectivas, que están organizados de la forma siguiente: el Capítulo 1 proporciona una introducción general sobre el mecanismo de acción de los derivados del cisplatino y de la antraquinona, y de las técnicas computacionales a la vanguardia empleadas en el estudio de fármacos anticancerígenos análogos. El Capítulo 2 da una visión general de los métodos teóricos usados en la presente tesis, abarcando desde las ecuaciones del movimiento de la MD hasta la teoría de la estructura electrónica molecular para describir el estado fundamental y los estados electrónicos excitados de los sistemas moleculares a diferentes niveles de aproximación. Los resultados principales son presentados desde el Capítulo 3 hasta el 8. En concreto, los Capítulos 3 y 4 introducen las herramientas computacionales desarrolladas para facilitar y automatizar el procedimiento para llevar a cabo cálculos QM/MM en un conjunto de geometrías. Se hace particular énfasis en la aplicación de métodos de función de onda, tales como el *complete active space self consistent field* (campo auto-consistente con espacio activo completo, CASSCF), en conjuntos de geometrías, para los cuales un algoritmo ha sido propuesto para preservar el espacio activo a lo largo del conjunto de geometrías. El Capítulo 5 se concentra en el estudio del mecanismo de difusión del cisplatino a través de una membrana lipídica, en el cual simulaciones de dinámica molecular clásica han sido llevadas a cabo utilizando campos de fuerza, junto con una técnica de muestreo eficiente, llamada *umbrella sampling*, para calcular las energías involucradas en el proceso de permeado. Las energías de interacción (así como su descomposición en componentes energéticas) entre el cisplatino y la membrana lipídica también han sido calculadas utilizando un campo de fuerzas. El Capítulo 6 representa una continuación del Capítulo 5, en donde las energías de interacción y la correspondiente descomposición han sido calculadas mediante una versión QM/MM de un análisis de la descomposición de la energía basado en densidades electrónicas. El Capítulo 7 presenta un estudio conformacional de la molécula de antraquinona cuando ésta se encuentra intercalada en una doble hélice de ADN, en el cual dos orientaciones de la antraquinona relativamente a

---

las bases nitrogenadas alrededor de esta, han sido identificadas. El análisis conformacional ha sido seguido por el cálculo de los estados excitados de la molécula de antraquinona, en donde se ha evidenciado que el carácter de transferencia de carga (entre el medicamento y una base nitrogenada) de los estados excitados depende de la orientación relativa entre la el fármaco y las bases nitrogenadas. El Capítulo 8 presenta algunos resultados de un trabajo aún en curso, en el cual se ha estudiado el mecanismo de difusión de algunos derivados de la molécula de antraquinona al interior de un modelo de membrana lipídica, así como los estados electronicos excitados de dichas moléculas en presencia del ambiente lipídico. Finalmente, el Capítulo 9 proporciona algunas conclusiones generales y perspectivas del presente trabajo de tesis.

# Chapter 1

## Introduction

### 1.1 Overview of the Research in Anticancer Agents

The research for more efficient anticarcinogenic agents still represents a topic of intense research since, despite the progress that has been made in the last decades, cancer still represents one of the most significant causes of death worldwide.<sup>1,2</sup> Some of the treatment methodologies consist of targeting the carcinogenic cells and inducing the cell death by means of apoptosis. Although two main mechanisms of cell killing have been devised, namely apoptosis and necrosis,<sup>3,4</sup> the former is preferred as in this case cell death occurs in an ordered manner and without involving adjacent tissues that are non-carcinogenic, unlike in the case of necrosis. The development of anticancer drugs that lead to apoptosis needs to account for several factors, such as the biocompatibility, and the mechanism of transport towards the affected tissue, the mechanism of action, among others. One way of classifying the compounds that have shown interesting anticancer activity, is by identifying them as active in the electronic ground state or in a specific excited state. In this regard, two families have shown promising anticancer activity: platinum-based compounds<sup>5-7</sup> and organic photosensitizers (PSs),<sup>8-10</sup> respectively. The goal of the present thesis work is to explore some fundamental steps of the mechanism of action and the relationship between the structure and the chemical properties of some of the most representative compounds of these two families of anticancer drugs, namely cisplatin and anthraquinone, respectively. For this reason, in what follows some of the main features of these compounds will be

---

introduced.

The family of platinum-based compounds is represented by Pt(II) complexes, which are active at the ground state. These compounds in general display a square planar geometry, and present two firmly bound *cis* amino groups (which could either be chelating groups or different ligands), and two labile ligands with good leaving groups. Among these complexes, perhaps the most widely used is *cis*-diamminedichloroplatinum (II) (*cisplatin*), which was approved by the Food and Drug Administration in 1978,<sup>11</sup> and has been since then used for treating several types of cancer, such as testicular, cervical, head, neck, lung and breast cancer, among others.<sup>12</sup> Other examples of extensively used Pt(II) complexes include oxaliplatin and carboplatin.<sup>8,9</sup> The general mode of action of *cisplatin*-like complexes to induce apoptosis consists of four steps: i) cell uptake, ii) hydrolysis/activation, iii) DNA binding and iv) DNA processing, thus leading to cell death.<sup>13</sup> Despite it is still object of debate, it is widely agreed that the mechanism of *cisplatin* uptake can occur either *via* a passive transport mechanism, in which the drug diffuses across the lipid bilayer of the carcinogenic cell, or *via* active transport, whereby a transmembrane protein plays a key role in the diffusion of *cisplatin* through the lipid membrane (Figure 1.1a).<sup>14</sup> An example of these protein facilitators is represented by the CTR1 copper transport protein.<sup>15</sup> After the *cisplatin* molecule has permeated the lipid membrane, it undergoes a hydrolysis of one (or both) of the two Pt-Cl labile bonds and forms a mono-aqua or a di-aqua Pt(II) derivative. It is at this point that the Pt aqua derivative enters the cell nucleus and eventually reacts by means of nucleophilic substitution with the imidazole N atoms of the DNA purine moieties, especially of the guanine residues.<sup>11,16</sup> The formation of a Pt-DNA adduct induces structural distortions of the DNA double helix which are recognized by some intracellular proteins (Figure 1.1b, right). These proteins trigger cellular processes that ultimately give rise to apoptosis or programmed cell-death.<sup>17</sup> There are several drawbacks stemming from using Pt(II) complexes for anticancer therapy which are mainly associated with their toxicity, low kinetic inertness and the method used for their administration (mainly intravenous infusion).<sup>9,10</sup> For these reasons, there has been interest in developing Pt(IV) complexes as precursors (or prodrugs) of the above mentioned Pt(II) drugs.<sup>18</sup> Their chemistry is significantly different from that of their Pt(II) counterparts, in particular because their saturated coordination sphere is kinetically more inert, and thus they are less

prone to ligand substitutions than the Pt(II) complexes.<sup>19</sup> This makes them more resistant towards undesired biochemical side-reactions that can occur prior to the formation of the Pt-DNA adduct, and therefore, are suitable for oral administration.<sup>20</sup> As a result, there have been efforts in tailoring Pt(IV) complexes that can be activated once in the presence of the target tumor cell, by triggering in a controlled manner their reduction to the active Pt(II) complexes.

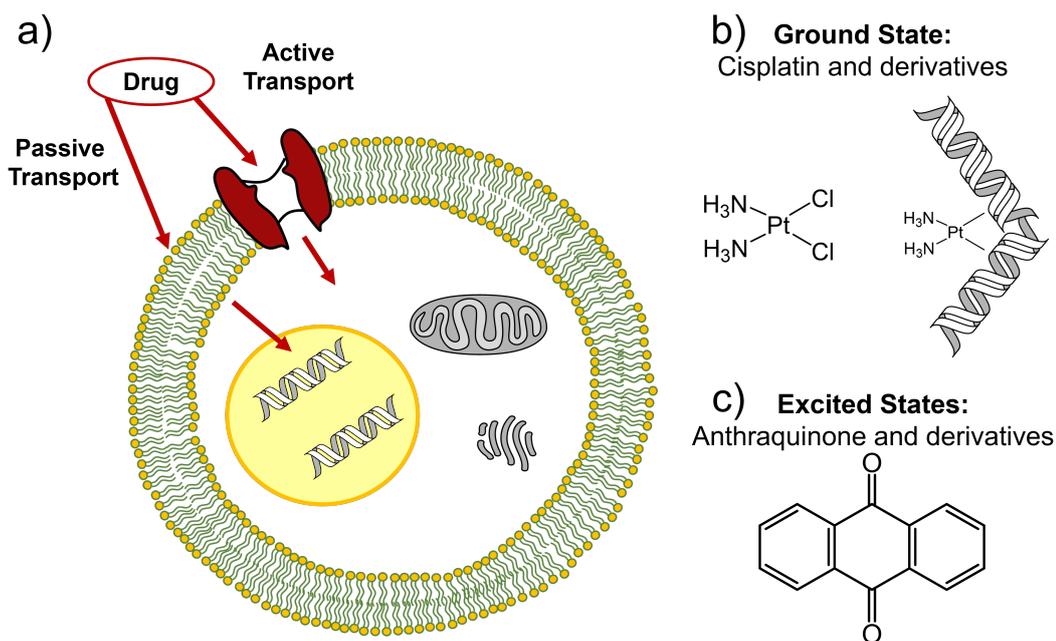


Figure 1.1: a) Schematic representation of the two possible permeation mechanisms of anticancer molecules inside a cell, that are mentioned in the main text: active transport (mediated by transmembrane proteins) and passive transport (through the lipid bilayer). b) Left: Molecular structure of the cisplatin molecule. Right: Representation of the DNA helix distortion induced by the binding of cisplatin. Notice the distortion of the helix induced by the drug. c) Molecular structure of the anthraquinone molecule, which is active following photoexcitation towards an electronic excited state.

The treatment of cancer by using PSs, which trigger apoptosis following photoexcitation towards electronic excited states, falls in the field of photodynamic therapy (PDT). One of the main reasons for which PDT is nowadays widely diffused in the treatment of different types of cancer,<sup>8-10</sup> is the fact of being a non invasive technique in which the activation of the PS occurs locally. This implies that the cell death is triggered exclusively within the region comprising the carcinogenic tissue.<sup>21</sup> Although the mechanism of apoptosis depends

---

on the nature of the tissue where the PS accumulates, and on the PS itself,<sup>22,23</sup> there is consensus in regard with the mechanism of action of the PS following the irradiation with a suitable wavelength to promote its photoexcitation. Specifically, the excited PS undergoes nonradiative decay followed by intersystem crossing towards the triplet state manifold. It is at this point that the PS induces oxidative stress on the molecules present in the surrounding environment (*e.g.*, phospholipids or DNA), either by means of electron transfer to these molecules to produce reactive radical species (type I mechanism) or *via* energy transfer towards nearby oxygen molecules to produce singlet oxygen, which can react with the biomolecules surrounding the PS (type II mechanism). It has been observed that several tumors develop hypoxia,<sup>24</sup> which induces conditions of low flux of oxygen. As a result, there is great interest in studying PSs that induce apoptosis through the type I mechanism. This has been the case for photosensitizers such as transition metal complexes,<sup>25,26</sup> nanoparticles,<sup>27,28</sup> and organic-based compounds.<sup>29,30</sup>

As stated above, the mechanism of cell death depends to a great extent on the tissue the PS accumulates on. In particular, it has been evidenced that it can accumulate on cell components such as the plasma membrane, lysosomes, the Golgi apparatus, the endoplasmic reticulum or the mitochondria.<sup>23</sup> Of particular interest are the latter, since following photoexcitation of the PS and subsequent rupture of the mitochondrial membrane or mutations in the mitochondrial DNA, they release pro-apoptotic components to the cytosol such as cytochrome-c and the pro-apoptosis inducing factor.<sup>31-33</sup> Some examples of families of PSs studied for PDT comprise phenothiazinium dyes, cyanines, porphyrins, phenanthridinium dyes, acridines, and anthraquinones, among others.<sup>34</sup> Among these, anthraquinone derivatives (Figure 1.1c), which are one of the families of anticarcinogenic molecules studied in the present thesis, have displayed promising activity *in vitro* on human carcinogenic tissues, especially on breast cancer cells.<sup>35</sup> The interactions established between some of these PSs and the lipid bilayer, as well as their interactions with DNA have been a relevant object of study, as they provide information regarding the mechanism of apoptosis induced by the PS under study. For example, the mechanism of oxidative stress of DNA induced by the photoexcited PS depends on the binding mode of the latter,<sup>36-38</sup> so that the PSs can be tailored by modifying their substituents<sup>39</sup> in order to induce oxidative stress on DNA in a specific manner.

---

The previous paragraphs highlight the importance of tailoring anticancer drugs in order to enhance their cytotoxic activity or to favor a specific mechanism of action depending on the target tissue. To this aim, it is however of utmost importance to deeply understand the properties of the precursor drugs, as well as the nature of the interactions that establish between these precursors and the target tissue, so as to being able to propose structural modifications for these drugs in the most efficient manner, by exploiting the cytotoxic properties inherent to these precursors. For this reason, the objective of the present thesis is to gain unprecedented insights on the mechanism of action of cisplatin, anthraquinone and some derivatives of the latter, in the presence of different biological targets, by means of state-of-the-art computational modeling techniques. To this end, a secondary objective is to develop a set of computational tools that facilitate and automatize the application of such modeling techniques. As will be evidenced in the following sections, computational modeling has been thoroughly used in the study of anticancer drugs, since it provides information at an atomistic level of the structure and the mechanism of action of these drugs<sup>40</sup> that can improve the interpretation of the experimental results,<sup>41</sup> or even validate the usage of experimental model systems.<sup>40, 42</sup>

---

## 1.2 Computational Techniques Applied on Related Anti-cancer Systems

The previous section has evidenced in a summarized manner the main features of some representatives of two families of anticancer drugs. It has been highlighted the interest for tailoring the drug to enhance its anticarcinogenic activity while reducing potential undesired effects stemming from their mode of action. On equal footing is the interest for understanding the underlying principles of the mechanism of action of these representatives. Computational methodologies have been extensively used to gain insights on the modes of action of anticarcinogenic drugs in the presence of a complex biological environment.<sup>40,43</sup> The methodologies that are employed to describe such systems mainly consist of MD simulations<sup>44</sup> in conjunction with enhanced sampling techniques such as US,<sup>45</sup> as well as multiscale hybrid QM/MM techniques.<sup>46,47</sup> The main goal of classical MD is to explore the configurational space of the system under study. This exploration is required when studying complex biological systems consisting of several thousands of atoms, like those which are the object of study in the present thesis, since the potential energy landscape of these systems could present several minima that can contribute to the properties under study.<sup>48</sup> For example, the permeation mechanism of cisplatin through a lipid membrane has been studied to some extent in the past at the classical MD level of theory.<sup>49,50</sup> In particular, the authors have evidenced that the composition of the lipid membrane and its curvature play a major role in determining the thermodynamics of the permeation itself. As previously discussed, the mechanism of action of the PSs that induces damage in the DNA structure depends on the binding mode of the PS. In this regard, classical MD has also been employed to study the binding modes and the relative binding energies of some of the PSs mentioned in the previous section with DNA. For example, the energetics of the noncovalent binding process of anthraquinone derivatives with DNA, such as the intercalation binding mode, have been performed in several MD studies.<sup>51-54</sup>

Multiscale hybrid QM/MM methodologies have also been employed in conjunction with classical MD to explore the electronic structure properties of anticancer molecules in the presence of complex biological media modeling the target tissues. QM/MM methodologies allow for the subdivision of the system under study into different subsystems, each described

---

at a different level of theory. The simplest case consists of two subsystems (or regions), whereby the region of chemical interest is described at a QM high level of theory, whereas the environment that surrounds the QM region is described at a lower level of theory, usually employing a classical potential.<sup>46,47</sup> The QM/MM methodology can be devised depending on the manner in which the interaction between the QM and the MM subsystems is described. One of the most widely used QM/MM approaches is the electrostatic embedding QM/MM, where the MM subsystem atoms are described as fixed point charges, which polarize the QM subsystem by entering the QM potential as an electrostatic perturbation. This framework has been extensively applied in the study of several complex biological systems, to determine both electronic ground<sup>55-59</sup> and excited state<sup>30,60-70</sup> properties. Within the context of anticancer drugs interacting with target cellular tissue, one example is represented by the work of Nogueira *et. al.*, in which the authors have combined QM/MM MD with X-ray absorption spectroscopy to elucidate the mechanism of hydrolysis and binding with DNA of a cisplatin-alike Pt(II) complex,<sup>41</sup> and have evidenced that selective binding of Pt compounds to adenine or to guanine could be attained by tuning the electron donating/accepting character of the hydrolyzed compounds. The photophysics of several organic PS has also been explored by using QM/MM and MD methodologies. For example, the excited states of methylene blue have been computed in the presence of a DNA strand,<sup>71,72</sup> and in the presence of a lipid membrane,<sup>73</sup> in both cases evidencing the impact of the biological environment on the character of these states. Huix-Rottland *et. al.* have investigated the fast intersystem crossing that the acetophenone photosensitizer undergoes in the presence of a DNA strand following UV irradiation.<sup>74</sup> Gattuso and co-workers have studied the binding modes of nile blue and nile red with a DNA strand, and have simulated the absorption and emission spectra of these chromophores for each of the binding modes.<sup>75</sup> Other examples include chelerythrine,<sup>54</sup> palmatine<sup>76</sup> and porphyrines.<sup>77</sup>

### 1.3 Objectives of the Thesis

The present thesis aims at gaining insights on the mechanism of action of two anticarcinogenic drugs in the presence of different biological targets. Specifically, the cisplatin molecule, which is active at the ground electronic state, and the anthraquinone molecule, which

---

is cytotoxic following photoexcitation to an electronic excited state, will be investigated when interacting with a lipid bilayer and with DNA, respectively. Two derivatives of the anthraquinone molecule will also be studied in the presence of a lipid bilayer. The methodologies adopted to investigate these drugs consist of classical MD to account for the conformational motion of the complex biological systems under study, and electrostatic embedding QM/MM to compute electronic ground and excited state properties of the drug at a high level of theory while accounting for the polarization due to the environment surrounding the drug. In particular, the studies on the cisplatin molecule mainly focus on the simulation of the permeation mechanism of the drug into a dioleoyl phosphatidylcholine model of a lipid membrane bilayer by means of classical MD in conjunction with the umbrella sampling technique, and the subsequent computation of the interaction energy between the drug and the lipid membrane within a novel QM/MM framework. In regard with the anthraquinone molecule, the intercalation binding mode has been studied in a dG-dC (guanine-cytosine) double strand model of DNA, and subsequent QM/MM calculations have been performed to analyze the effect of the DNA environment on the character of the excited states of the anthraquinone molecule. The permeation mechanism of some natural derivatives of the anthraquinone molecule that have shown promising anticancer activity<sup>78</sup> has also been studied, and the excited states of these molecules in the presence of the lipid membrane are currently work in progress.

The practical problem posed by the computation of physical properties within a QM/MM framework on an ensemble of configurations of the systems of interest is also addressed in the present thesis work. This has derived in the development of a software (MoBioTools)<sup>79,80</sup> that allows for the generation of QM/MM inputs for an ensemble of geometries in an automatic manner, irrespective of the physical property under study. In this context, for systems whereby the QM region is described using a wavefunction method such as the CASSCF,<sup>81-83</sup> an algorithm to automatically correct the CAS active space on an ensemble of geometries has also been developed.<sup>84</sup>

In summary, the thesis is organized as follows:

1. Chapter 2 presents an overview of the theoretical methodologies applied in the present thesis.

- 
2. Chapter 3 introduces the MoBioTools software: a program developed to set up in an automated manner QM/MM input files from an ensemble of geometries. This software has been used to some extent throughout the entire thesis.<sup>79,80</sup>
  3. Chapter 4 treats an algorithm developed to correct in an automatic manner the CASSCF active space of an ensemble of geometries to compute electronic excited state properties of complex biological systems.<sup>84</sup>
  4. In Chapter 5 the permeation mechanism of the cisplatin molecule is studied by means of classical MD and US simulations.<sup>85</sup>
  5. Chapter 6 is a follow-up of Chapter 5, whereby the interaction energies between the cisplatin molecule and the lipid membrane are computed at a QM/MM level of theory, and a QM energy decomposition approach of the interaction energy is extended to the QM/MM framework.<sup>86</sup>
  6. In Chapter 7, the intercalation binding mode of the anthraquinone molecule in a dG-dC DNA model is studied *via* classical MD simulations, and the excited states of a QM region including the photosensitizer and a guanine molecule are characterized.<sup>87</sup>
  7. Chapter 8 shows the permeation mechanism of two derivatives of the anthraquinone molecule into a lipid membrane, and the subsequent study of the electronic excited states of these systems in the presence of the lipid bilayer.
  8. Finally, some general conclusions and perspectives are presented in Chapter 9.

## Chapter 2

# Theoretical Methods

### 2.1 Electronic Structure Theory: Ab Initio Methods

#### 2.1.1 The Born-Oppenheimer Approximation

One of the main problems in quantum chemistry consists of finding approximate solutions to the time-independent Schrödinger equation,

$$H\Psi = E\Psi \quad (2.1)$$

where  $\Psi$  is the wavefunction of the system under study,  $E$  is the energy and  $H$  is the non-relativistic time-independent Hamiltonian operator (in atomic units),

$$\begin{aligned} H(r, R) = & - \sum_{A=1}^M \frac{1}{2} \nabla_A^2 + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|R_A - R_B|} + \\ & - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{|R_A - r_i|} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_i - r_j|} \end{aligned} \quad (2.2)$$

where  $R_A$  and  $R_B$  represent nuclear coordinates,  $r_i$  and  $r_j$  represent electron coordinates,  $Z_A$  ( $Z_B$ ) is the nuclear charge of the atom centered at  $R_A$  ( $R_B$ ),  $M$  is the number of nuclei and  $N$  is the number of electrons. In Equation 2.2, the first term represents the nuclear kinetic energy, the second term the nucleus-nucleus coulombic repulsion, the third term the kinetic energy of the electrons, the fourth one represents the nuclear-electron attraction and the last term corresponds to the electron-electron coulombic repulsion. At

---

the beginning of the paragraph it was stated that the main goal is to find approximate solutions for Equation 2.1; the main reason is because no analytic solutions are known even for a system containing three particles, let alone a complex molecular system. In this regard, it represents a formidable task to seek for solutions for a system consisting of  $3M$  nuclear coordinates and  $3N$  electron coordinates, so that it is desirable to apply approximations that have a physical meaning to tackle the problem of solving Equation 2.1. One example consists of evidencing the fact that, since the nuclei are heavier than the electrons, it is expected that the former move slower than the latter. Thus, the motion of the electrons can be considered to occur in the presence of a potential generated by fixed nuclei, so that in practice the positions of the nuclei are considered to be constant. This is the core of the so-called Born-Oppenheimer (BO) approximation,<sup>88</sup> and its implications can be evidenced in Equation 2.2: if the positions of the nuclei are considered to be constant, then the first sum in the equation can be neglected, and the second sum represents an additive constant. With this approximation, the total molecular wavefunction  $\Psi(r, R)$  can be separated into an electronic wavefunction  $\Phi(r; R)$ , which depends parametrically on the nuclear positions, and a nuclear wavefunction  $\Omega(R)$ :

$$\Psi(r, R) = \Phi(r; R)\Omega(R) \quad (2.3)$$

Besides, the last three terms in equation 2.2, which are those displaying a dependence on the electron coordinates, can be gathered into a single operator, which is called the electronic Hamiltonian operator ( $H_{el}$ )

$$H_{el}(r; R) = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_{A=1}^M \sum_{i=1}^N \frac{Z_A}{|R_A - r_i|} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_i - r_j|} \quad (2.4)$$

Having considered the nuclear positions fixed, the electronic part of the total wavefunction is determined by solving the electronic Schrödinger equation:

$$H_{el}(r; R)\Phi(r; R) = E_{el}(R)\Phi(r; R) \quad (2.5)$$

Where  $E_{el}(R)$  is the electronic energy that depends on the nuclear coordinates. It should be emphasized that in order to determine the potential energy, the nuclear-nuclear

---

repulsion term, which in the context of the BO approximation is a constant term, needs to be accounted for. Thus, the potential energy is obtained by adding to the  $E_{el}(R)$  energy this nuclear contribution:

$$U_{el}(r; R) = E_{el}(R) + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{|R_A - R_B|} \quad (2.6)$$

If the electronic Schrödinger equation is solved for different nuclear positions  $R$ , one obtains the potential energy surface (PES) of the electronic state of interest. It should be emphasized that within the BO approximation, it is assumed that the nuclei move on the PES of a specific electronic state. Once the electronic Schrödinger equation has been solved and the potential energy  $U_{el}$  has been determined, it is possible to determine the nuclear part of the molecular wavefunction,  $\Omega(R)$ , by solving the corresponding nuclear Schrödinger equation, whose Hamiltonian consists of the nuclear kinetic energy (first term in Equation 2.2) plus the total electronic energy:

$$\left[ - \sum_{A=1}^M \frac{1}{2} \nabla_A^2 + U_{el}(R) \right] \Omega(R) = E \Omega(R) \quad (2.7)$$

This equation accounts for the vibrational, rotational and translational motion of the nuclei. At this point, the total molecular wavefunction within the BO approximation is given by Equation 2.3, and the total energy is given by  $E(R)$ . It should be emphasized that the BO approximation results to be a formidable approach for several systems of chemical interest, whereby different electronic states are far from one another from an energetic viewpoint. This is the case of the present thesis, in which physical observables such as the interaction energy, electronic energy, absorption spectra and others were computed within this framework. However, there are situations in which the motion of the nuclei and the electrons need to be accounted for simultaneously, for example, in situations of non-radiative decay of a molecule from an electronically excited state to one at a lower energy. These situations formally occur whenever the electronic wavefunction strongly varies as the nuclear positions vary, so that a parametric dependence of the electronic wavefunction on the nuclear coordinates (as in Equation 2.5) is no longer viable. In these situations, the time-dependent molecular Schrödinger equation needs to be solved:

---


$$i\hbar \frac{\partial}{\partial t} \Psi(r, R, t) = H(r, R) \Psi(r, R, t) \quad (2.8)$$

where  $\Psi(r, R, t)$  is the time-dependent molecular wavefunction. A description of the methods to solve Equation 2.8 is beyond the scope of the thesis, and several accounts in that regard can be found elsewhere.<sup>89–95</sup> In what follows, different approximate numerical techniques (within the framework of the BO approximation) to solve the time-independent Schrödinger equation will be discussed.

### 2.1.2 The Hartree-Fock Approximation

Solving the electronic time-independent Schrödinger equation (Equation 2.5) for a molecular system implies considering a system consisting of  $N$  interacting electrons. One of the most important features of a wavefunction describing a system of  $N$  electrons is that it must obey the antisymmetric property with respect to a permutation of two electrons or, what amounts to the same thing, an exchange in the coordinates of two electrons:

$$\Phi(x_1, x_2, \dots, x_i, \dots, x_j, \dots, x_n) = -\Phi(x_1, x_2, \dots, x_j, \dots, x_i, \dots, x_n) \quad (2.9)$$

where  $x_i$  and  $x_j$  represent the coordinates of two electrons which have been exchanged;  $x_i$  not only consists of the cartesian coordinates of the  $i^{th}$  electron, but also its intrinsic spin. Since the only "molecular" systems for which analytic solutions of the corresponding Schrödinger equations consist of hydrogen-like atoms, and the wavefunctions that describe the electronic states are (clearly) one-electron wavefunctions, an approach to obtain a wavefunction for a molecular system consists of using a superposition one-electron wavefunctions (each of which formally describing an electron in the molecule) which obeys the antisymmetry principle. One of the simplest molecular wavefunctions proposed is the Slater determinant,

$$\Phi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_N(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_N(x_2) \\ \dots & \dots & \dots & \dots \\ \phi_1(x_N) & \phi_2(x_N) & \dots & \phi_N(x_N) \end{vmatrix} \quad (2.10)$$

---

where each  $\phi_i(x_j)$  is a one-electron wavefunction called a spin orbital. The Hartree-Fock (HF) method consists of introducing the  $N$ -electron Slater determinant into the electronic Schrödinger equation (Equation 2.5), after which the variation theorem<sup>88</sup> is applied to determine the spin-orbitals that minimize the Rayleigh ratio:

$$E = \frac{\int \Phi_0^* H \Phi_0 dx_1 dx_2 \dots dx_N}{\int \Phi_0^* \Phi_0 dx_1 dx_2 \dots dx_N} \quad (2.11)$$

The application of this procedure provides a set of one-electron Schrödinger-type equations, whose solutions provide the set of spin orbitals that minimize  $E$  in Equation 2.11:

$$f_1(x_1)\phi_i(x_1) = \epsilon_i\phi_i(x_1) \quad (2.12)$$

where  $f_1(x_1)$  is the Fock operator:

$$f_1(x_1) = h_1(x_1) + \sum_i (J_i(x_1) - K_i(x_1)) \quad (2.13)$$

in which  $h_1(x_1)$  is the core hamiltonian

$$h_1(x_1) = -\frac{1}{2}\nabla_1^2 + \sum_{A=1}^N \frac{Z_A}{|R_A - r_1|} \quad (2.14)$$

$J_i(x)$  represents the classical Coulomb repulsion between electrons

$$J_i(x_1)\phi_a(x_1) = \int \frac{\phi_i(x_2)^*\phi_i(x_2)}{|r_1 - r_2|} dx_2 \phi_a(x_1) \quad (2.15)$$

and  $K_i(x)$  represents the exchange interactions between electrons, that arise due to the antisymmetric character of the wavefunction,

$$K_i(x_1)\phi_a(x_1) = \int \frac{\phi_i(x_2)^*\phi_a(x_2)}{|r_1 - r_2|} dx_2 \phi_i(x_1) \quad (2.16)$$

The set of equations 2.12 provide the spin orbitals that minimize the energy. Thus, according according to the variation principle, these spin orbitals give the least upper bound of the (non-relativistic) exact electronic energy. These equations represent a set of

---

nonlinear equations, since the solutions *are* the spin orbitals, but at the same time the Fock operator depends on these spin orbitals. As a result, these equations need to be solved in an iterative manner, starting from a trial set of spin orbitals until a certain convergence criterion is attained. This iterative procedure is known as the self-consistent field (SCF). It should be emphasized that although, in principle, the Fock equations could have infinite solutions, in practice it is possible to solve them for  $M$  spin orbitals, of which  $N$  will correspond to spin orbitals occupied by an electron. A very common strategy adopted to solve these equations consists of expressing each spin orbital as a linear combination of known basis functions  $\chi_j$ , usually called atomic orbitals (LCAO)

$$\phi_i(x_1) = \sum_{j=1}^k C_{ij} \chi_j(x_1) \quad (2.17)$$

so that in practice the problem of determining the optimal set of spin orbitals becomes the problem of determining the optimal coefficients  $C_{ij}$  in Equation 2.17. This method was devised by Roothaan in 1951,<sup>96</sup> and transforms the analytic problem of solving for a set of functions  $\phi_i(x)$  into an algebraic problem: if the spin orbitals in Equation 2.12 are expanded using Equation 2.17, and then one multiplies to the left by a given  $\chi_j$  and integrates over  $x$  (including the spin coordinates), one obtains the Roothaan-Hall matrix equations:

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (2.18)$$

where  $\mathbf{C}$  is the matrix of the atomic orbital coefficients,  $\epsilon$  is a diagonal matrix containing the eigenvalues of the spin orbitals, and

$$\begin{aligned} F_{ij} &= \int \chi_i(r_1)^* f_1(r_1) \chi_j(r_1) dr_1 \\ S_{ij} &= \int \chi_i(r_1)^* \chi_j(r_1) dr_1 \end{aligned} \quad (2.19)$$

where the spin coordinates have been integrated out, so that the matrix elements are represented solely in terms of the spatial coordinate  $r_1$ . The main approximation on the HF method stems from the assumption of the form of the wavefunction (as a Slater determinant, Equation 2.10), which in turn derives in the assumption that each electron

---

interacts with an average potential generated by the other  $N - 1$  electrons (Equations 2.15 and 2.16). This introduces an error in the energy with respect to what would be the exact energy obtained by solving accurately Equation 2.5 - this error is referred to as the correlation energy. This terminology however does not imply that HF lacks of electron correlation; indeed, the Slater determinant obeys the antisymmetry principle, which in turn accounts for a type of correlation between electrons of the same spin, whereby no two electrons having the same spin can occupy the same position in space – the Fermi hole. However, no correlation is accounted for between electrons having opposite spin – the Coulomb hole. Another problem could arise whenever an electronic configuration is close in energy to the ground state configuration, so that the description of the system with a single Slater determinant would no longer be accurate. The correlation energy is defined as the difference between the exact non-relativistic electronic energy of a molecular system ( $E_{exact}$ ) and the HF energy:

$$E_{corr} = E_{exact} - E_{HF} \quad (2.20)$$

Several methodologies have been devised to recover the correlation energy using the HF wave function as a starting point. In most cases, these could be either variational (configuration interaction) or perturbative (perturbation theory). These two cases will be explored in what follows.

### 2.1.3 Møller-Plesset Perturbation Theory

In this section, perturbation theory is applied to the ground state HF energy and wavefunction to obtain an estimate for the correlation energy.<sup>88</sup> In this approach, the unperturbed Hamiltonian,  $H^0$ , is taken as the sum of the Fock operators ( $f_i$  in Equation 2.13) over a set of  $N$  occupied spin orbitals:

$$\begin{aligned} H^0 &= \sum_{j=1}^N f_j \\ &= \sum_{j=1}^N \left[ -\frac{1}{2} \nabla_j^2 + \sum_{A=1}^M \frac{Z_A}{|r_j - R_A|} + \sum_{i=1}^N [J_i(x_j) - K_i(x_j)] \right] \end{aligned} \quad (2.21)$$

The HF wavefunction,  $\Phi_0$ , is an eigenfunction of the unperturbed Hamiltonian, with

eigenvalue  $\sum_{i=1}^N \epsilon_i$ . On the other hand, the perturbed Hamiltonian is given by  $H_{el}$  in Equation 2.4. Thus, the perturbation is given by the difference  $H_{el} - H^0$ , that is

$$\begin{aligned}
H' &= H_{el} - H^0 \\
&= \sum_{j=1}^N \sum_{i>j}^N \frac{1}{|r_i - r_j|} - \sum_{j=1}^N \sum_{i=1}^N [J_i(x_j) - K_i(x_j)] \\
&= \sum_{j=1}^N \sum_{i>j}^N \frac{1}{|r_i - r_j|} - \sum_{j=1}^N v^{HF}(x_j)
\end{aligned} \tag{2.22}$$

where  $v^{HF}(x_j) = \sum_{i=1}^N [J_i(x_j) - K_i(x_j)]$ . In the case of the HF wavefunction (the Slater determinant  $\Phi_0$ ), the first order correction to the energy – considering the unperturbed Hamiltonian  $H^0$ - is given by:

$$\begin{aligned}
E_0^{(1)} &= \langle \Phi_0 | H' | \Phi_0 \rangle \\
&= \langle \Phi_0 | \sum_j \sum_{i>j} \frac{1}{|r_i - r_j|} | \Phi_0 \rangle - \langle \Phi_0 | \sum_{j=1}^N v^{HF}(x_j) | \Phi_0 \rangle \\
&= \sum_j \langle \Phi_0 | \sum_{i>j} r_{12}^{-1} | \Phi_0 \rangle - \langle \Phi_0 | \sum_{j=1}^N v^{HF}(x_j) | \Phi_0 \rangle \\
&= \frac{1}{2} \sum_{j=1}^N \langle \phi_j | v^{HF}(x_j) | \phi_j \rangle - \sum_{j=1}^N \langle \phi_j | v^{HF}(x_j) | \phi_j \rangle \\
&= -\frac{1}{2} \sum_{j=1}^N \langle \phi_j | v^{HF}(x_j) | \phi_j \rangle
\end{aligned} \tag{2.23}$$

where the Slater-Condon rules have been used to compute the first term on the r.h.s. of Equation 2.23, and  $r_{12}^{-1} = \frac{1}{|r_i - r_j|}$ . Thus, the perturbed energy (to the first order) is given by

$$\begin{aligned}
E_0 &= E_0^{(0)} + E_0^{(1)} \\
&= \sum_{j=1}^N \epsilon_j - \frac{1}{2} \sum_{j=1}^N \langle \phi_j | v^{HF}(x_j) | \phi_j \rangle
\end{aligned} \tag{2.24}$$

which is precisely the HF energy. Thus, a second-order correction to the energy is required to improve the HF energy. The second-order correction (using again the HF wavefunction) is given by

---


$$E_0^{(2)} = \sum_{J \neq 0} \frac{|\langle \Phi_J^{(0)} | H' | \Phi_0 \rangle|^2}{E_0^{(0)} - E_J^{(0)}} \quad (2.25)$$

in which  $\Phi_J^{(0)}$  are all possible Slater determinants that can be constructed from the unperturbed wavefunction  $\Phi_0$ . The Slater-Condon rules can be used to compute the integrals at the numerator in Equation 2.25. In particular, integrals involving singly excited, as well as levels of excitation above 3 vanish. Thus, Equation 2.25 becomes:

$$E_0^{(2)} = \sum_a \sum_{b>a} \sum_r \sum_{s>r} \frac{|\langle \phi_a \phi_b | r_{12}^{-1} | \phi_r \phi_s \rangle - \langle \phi_a \phi_b | r_{12}^{-1} | \phi_s \phi_r \rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s} \quad (2.26)$$

where the sum runs over all occupied  $a$  and  $b$ , and all virtual  $r$  and  $s$  spin orbitals. Thus, the total energy corrected up to the second order is given by

$$E_0 = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} \quad (2.27)$$

## 2.1.4 Configuration Interaction

The correlation part of the energy missing on the HF method can also be recovered to some extent using variational methods. An example of such methods is the configuration interaction method (CI),<sup>88</sup> in which the wavefunction is expressed as a linear combination of Slater determinants (configurations):

$$\Psi_0 = \Phi_0 + \frac{1}{1!} \sum_{ar} C_a^r \Phi_a^r + \frac{1}{2!} \sum_{abrs} C_{ab}^{rs} \Phi_{ab}^{rs} + \frac{1}{3!} \sum_{abc}^{rst} C_{abc}^{rst} \Phi_{abc}^{rst} + \dots \quad (2.28)$$

where, for a given set of  $N$  electrons and  $2K$  spin orbitals,  $\Phi_0$  is the ground state HF wavefunction,  $\Phi_a^r$  ( $a = 1, \dots, N; r = (N+1), \dots, 2K$ ) are all possible singly excited Slater determinants that can be formed from any occupied spin orbital  $a$  to any virtual spin orbital  $r$ ;  $\Phi_{ab}^{rs}$  are all possible doubly excited Slater determinants, and so on. The factors  $\frac{1}{n!}$  are included so that each determinant be counted only once. It should be emphasized that in principle, for a complete basis set ( $2K \rightarrow \infty$ ), the sum in Equation 2.28 is infinite. This implies that when the expression for  $\Psi_0$  in Equation 2.28 is used as a trial wavefunction to minimize the Rayleigh ratio (Equation 2.11; variation method), the energy obtained is, in

---

principle, the exact non-relativistic energy of the ground state for the system under study, within the BO approximation. If one represents the CI expansion (as in Equation 2.28) for any given electronic state  $S$ , as follows:

$$\Psi_S = \sum_{J=0} C_{JS} \Phi_J \quad (2.29)$$

where  $J = 0$  corresponds to the HF wavefunction, and  $J > 0$  represents all determinants having at most  $J$  excitations, then the problem of variationally determining the coefficients  $C_{JS}$  becomes that of solving the set of equations:

$$\sum_{J=0} H_{IJ} C_{JS} = E_S \sum_{J=0} S_{IJ} C_{JS} \quad I = 0, 1, \dots \quad (2.30)$$

where

$$\begin{aligned} H_{IJ} &= \int \Phi_I^* H \Phi_J dx_1 \dots dx_N \\ S_{IJ} &= \int \Phi_I^* \Phi_J dx_1 \dots dx_N \end{aligned} \quad (2.31)$$

In practice, however, only a finite basis set can be used, so that the sums in Equations 2.28 (for the ground state) of 2.29 are finite, and the number of determinants present depends on the number of basis functions used (which in turn determines the number of spin orbitals under consideration). If all possible excited determinants are considered in the CI expansion, for a set of  $N$  electrons and  $2K$  spin orbitals, the obtained wavefunction is referred to as the full CI wavefunction. The lowest eigenvalue obtained by solving Equations 2.29 within the full CI framework is an upper bound for the ground state energy of the system, and represents a benchmark for a given finite basis set, as it is the best possible approximation to the exact ground state energy for that basis set. However, the number of determinants scales dramatically fast with the size of the set -the scaling factor is  $\binom{N}{K}$  -of spin orbitals, so that the full CI calculations are practicable only for small systems and basis sets. Therefore, the expansion in Equation 2.29 is usually truncated to consider determinants having at most a certain number  $L$  of excitations. One such example of truncated CI consists of an expansion considering only singly and doubly excited configurations, which is referred to as CISD. A shortcome of the truncated CI approach however, is the lack of size consistency. A method is said to be *size consistent* whenever the

---

computation of the energy of two infinitely separated molecular systems  $A$  and  $B$ ,  $E(AB)$ , is the same as the sum of the energies of the two systems computed separately, that is

$$E(AB)_{R_{AB} \rightarrow \infty} = E(A) + E(B) \quad (2.32)$$

For example, CISD is not size consistent, since the computation of the energy for each of the molecules involves at most doubly excited configurations for each of the monomers; when CISD is applied to the supermolecule  $AB$ , the situation in which both monomers are simultaneously doubly excited is not accounted for, as this would imply to consider a quadruple excitation in the supermolecule  $AB$ . One example of a method that overcomes the size consistency problem when employing a truncated CI expansion is the multiconfigurational self-consistent field method, which will be further explored in the next section.

### 2.1.5 Multiconfigurational Self-Consistent Field Methods

In the multiconfigurational self-consistent field (MCSCF)<sup>82</sup> method, the wavefunction is represented as a linear combination of Slater determinants, or alternatively of configuration state functions (CSFs), which are spin-adapted linear combinations of determinants. Thus, the MCSCF ground state wavefunction is a CI-like expansion of determinants:

$$\Psi_0 = \sum_J C_J^0 \Phi_J \quad (2.33)$$

for the ground state, and

$$\Psi_K = \sum_J C_J^K \Phi_J \quad (2.34)$$

for any given excited state  $K$ . The main difference between the CI and the MCSCF approaches lies on the fact that in the MCSCF case not only the CI coefficients ( $C_J^0$  in Equation 2.33), but also the spin orbital coefficients (the  $C_{ij}$  coefficients of the LCAO in Equation 2.17), are variationally optimized during the minimization of the energy. This implies that the reduction in the number of determinants considered to describe the wavefunction does not occur by setting an upper bound to the number of excitations to consider, as in the case of the truncated CI approach, but instead by limiting the number

---

of spin orbitals involved in the electronic excitations. This approach has the advantage of reducing considerably the number of configurations at hand, while making the method size consistent. For simplicity, in what follows, both the MCSCF ground state wavefunction and the determinants (or CSFs) in Equation 2.33 will be represented as ket vectors:  $\Psi_0 = |0\rangle$  and  $\Phi_J = |J\rangle$ . Thus, Equation 2.33 becomes:

$$|0\rangle = \sum_J C_J^0 |J\rangle \quad (2.35)$$

In order to minimize the energy, both the CI coefficients and the spin orbital coefficients need to be varied. These variations are described by rotations of the MCSCF wavefunction. The orbital rotations are described by the unitary operator  $\exp(T)$  on the wavefunction  $|0\rangle$ , where  $T$  is an anti-Hermitian operator:

$$T = \sum_{p>q} t_{pq} (E_{pq} - E_{qp}) \quad (2.36)$$

where  $E_{pq}$  is the spin-averaged excitation operator, defined in terms of the creation and annihilation operators  $a_{p\sigma}^\dagger$  and  $a_{p\sigma}$ , respectively:

$$E_{pq} = \sum_{pq} a_{p\sigma}^\dagger a_{q\sigma} \quad (2.37)$$

The operator  $T$  can be shown to be anti-Hermitian by using the algebra rules of creation and annihilation operators, a topic widely covered in the literature.<sup>82,88</sup> On the other hand, in order to describe the variations in the CI coefficients of Equation 2.35, the orthogonal complement of  $|0\rangle$  (*i.e.*, the subspace spanned by all the states orthogonal to the ground state  $|0\rangle$ ) is introduced:

$$|K\rangle = \sum_J C_J^{(K)} \Phi_J \quad K \neq 0 \quad (2.38)$$

Notice that each state  $|K\rangle$  is also spanned by the basis of determinants  $\Phi_J$ . In this way, the rotation describing the variations in the CI coefficients of Equation 2.35 is given by another exponential unitary operator,  $\exp(S)$ , in which  $S$  is given by the anti-Hermitian operator

---


$$S = \sum_{K \neq 0} s_{K0} (|K\rangle \langle 0| - |0\rangle \langle K|) \quad (2.39)$$

where the sum runs over all multiconfigurational states but the ground state. Thus, a rotation in the CI and the orbital coefficients is given by the composition of the two exponential operators described above:

$$|0'\rangle = \exp(T)\exp(S)|0\rangle \quad (2.40)$$

Thus, the energy of the MCSCF expansion is given by the following expression:

$$E(s, t) = \langle 0 | \exp(-S)\exp(-T)H\exp(T)\exp(S) | 0 \rangle \quad (2.41)$$

which is a function of the parameters  $t_{pq}$  and  $s_{K0}$  which determine the orbital and state rotations, respectively. By using a power series expansion for each of the exponential operators in Equation 2.41, in conjunction with the definitions of the operators  $S$  and  $T$  (Equations 2.36 and 2.39), it can be shown that the following conditions of stationarity hold:

$$\langle 0 | H | K \rangle = 0 \quad K \neq 0 \quad (2.42)$$

$$\langle 0 | H(E_{pq} - E_{qp}) | 0 \rangle = 0 \quad (2.43)$$

A widely used MCSCF method is the complete active space self-consistent field (CASSCF),<sup>81,83</sup> in which the orbital space is subdivided into three different subspaces: inactive, active and secondary orbitals. Within this framework, a full CI is performed considering exclusively determinants deriving from excitations of orbitals within the active orbital space, whereas the inactive orbitals are kept doubly occupied and the secondary orbitals are kept unoccupied. It can be shown that, for rotations within the inactive and the secondary subspaces, an application of the spin-averaged excitation operator (Equation 2.37) upon the CASSCF wavefunction provides precisely the occupation number if  $p = q$  and zero otherwise. On the other hand, any excitation within the active orbital subspace provides a linear combination of all the state vectors  $|K\rangle$ , since the CSF space is complete

---

in the active orbital subspace.

$$E_{tu} = \sum_K \langle K | E_{tu} | 0 \rangle \quad t, u = \text{active} \quad (2.44)$$

Furthermore, if the orbitals are real, then

$$E_{ut} = \sum_K \langle K | E_{tu} | 0 \rangle \quad t, u = \text{active} \quad (2.45)$$

so that, if one computes  $\langle 0 | H(E_{tu} - E_{ut}) | 0 \rangle$ , the second condition of stationarity (Equation 2.42) is satisfied. This implies that the energy is invariant under rotations within the same orbital subspace, so that in practice only rotations occurring between different orbital subspaces are considered during the optimization process. An advantage from this feature is represented by the fact that no redundant parameters are obtained from the application of two different rotations: redundancies can arise when, for example, two different rotations (either from applying Equation 2.39 or 2.36, or different combinations of both) give rise to the same state  $|0'\rangle$ . This does not happen if the CSF basis is complete within the active orbital space. For this reason, CASSCF is perhaps the most widely used MCSCF method in the literature. The feature that requires the most attention when applying it on molecular systems is the choice of the active orbital subspace, which in general requires chemical intuition, depending on the system under study, although some strategies have been recently devised to allow for an automatic selection of the active space.<sup>97,98</sup>

### 2.1.6 Complete Active Space Perturbation Theory

A caveat of the MCSCF method in general, and the CASSCF method in particular, resides on the fact that only a limited amount of configurations (out of the entire set of excited configurations present on the full CI expansion) is considered in the determinant expansions in Equations 2.33 and 2.35. This implies that there is a remaining part of the correlation energy that is not recovered with the MCSCF wavefunction. The unrecovered correlation energy is associated with the regions in the many-electron coordinate space whereby the distance between two electrons is small.<sup>99</sup> Several approaches have been devised to overcome this caveat, for example using truncated CI expansions of a MCSCF wavefunction, extended basis sets<sup>100</sup> or applying perturbation theory on top of a reference MCSCF wavefunction. In the present thesis, the CASSCF method has been used in conjunction with perturbation theory, since it has provided an accurate description for ground and excited state properties

---

of molecular systems similar to the anticancer molecules treated here.<sup>54,71,74,75</sup> This approach is widely known as the complete active space second order perturbation theory (CASPT2).<sup>99</sup> Within the framework of CASPT2, the Hamiltonian is partitioned into a zeroth-order part  $H^0$  and the perturbation  $H'$  (cf. Equation 2.22):

$$H = H^0 + H' \quad (2.46)$$

Likewise, the Hilbert space is partitioned into a reference space  $P$ , and a secondary space  $Q$ ,

$$P + Q = 1 \quad (2.47)$$

whereby the reference space  $P$  is spanned by the (single) CASSCF reference state  $|\alpha\rangle$ ,

$$P = |\alpha\rangle\langle\alpha| \quad (2.48)$$

which is an eigenfunction of the zeroth-order Hamiltonian,

$$H^0 |\alpha\rangle = E_\alpha^0 |\alpha\rangle \quad (2.49)$$

Let  $\Omega_\alpha$  be the operator that generates the exact wavefunction upon acting on the reference state  $|\alpha\rangle$ . It can be expanded as an order-by-order expansion series,

$$\Omega_\alpha = 1 + \Omega_\alpha^1 + \Omega_\alpha^2 + \dots \quad (2.50)$$

In particular, the first order term,  $\Omega_\alpha^1$  corresponds to the operator that generates the first order correction of the wavefunction when acting on  $|\alpha\rangle$ . The corresponding first order equation is

$$(E_\alpha^0 - H^0)\Omega_\alpha^1 |\alpha\rangle = QH |\alpha\rangle \quad (2.51)$$

The first and the second order corrections of the energy ( $E_\alpha^{(1)}$  and  $E_\alpha^{(2)}$ , respectively) are given by

---


$$E_\alpha^{(1)} = \langle \alpha | H' | \alpha \rangle \quad (2.52)$$

$$E_\alpha^{(2)} = \langle \alpha | H' \Omega_\alpha^1 | \alpha \rangle$$

The zeroth-order Hamiltonian for the reference CASSCF state  $|\alpha\rangle$  is given by

$$H_\alpha^0 = |\alpha\rangle \langle \alpha | F_\alpha | \alpha \rangle \langle \alpha | + Q_K F_\alpha Q_K + Q_{sd} F_\alpha Q_{sd} + Q_{tq\dots} F_\alpha Q_{tq\dots} \quad (2.53)$$

where  $F_\alpha$  is a one-electron operator,  $Q_K$  is a projector into the orthogonal complement of  $|\alpha\rangle$  within the CASSCF space of wavefunctions (cf. Equations 2.38 and 2.39),

$$Q_K = \sum_{\kappa \neq \alpha} |\kappa\rangle \langle \kappa| \quad (2.54)$$

$Q_{sd}$  is a projector into the space of functions  $|pqrs; \alpha\rangle$  generated by applying single and double excitations to  $|\alpha\rangle$ ,

$$Q_{sd} = \sum_{pqrs} |pqrs; \alpha\rangle \langle \alpha; pqrs| \quad (2.55)$$

where the basis functions of this subspace,  $|pqrs; \alpha\rangle$  are given by,

$$|pqrs; \alpha\rangle = E_{pq} E_{rs} |\alpha\rangle \quad (2.56)$$

$Q_{tq\dots}$  is a projector into the space spanned by higher order excitations on  $|\alpha\rangle$ . The choice of the operator  $F_\alpha$  is somewhat arbitrary, however it is required that it be such that the zeroth-order Hamiltonian is equivalent to the Møller-Plesset Hamiltonian in the limit of a single determinant reference state (Equation 2.21). An example of one such operator is given by

$$F_{pq\sigma} = a_{p\sigma} [H, a_{q\sigma}^\dagger] + a_{p\sigma}^\dagger [H, a_{q\sigma}] \quad (2.57)$$

It can be shown<sup>101</sup> that the first order equation 2.51 can be written in terms of solely  $Q_{sd}$  as follows:

$$(E_\alpha^0 - F_\alpha) \Omega_\alpha^1 |\alpha\rangle = Q_{sd} H |\alpha\rangle \quad (2.58)$$

where  $|\alpha\rangle$  is projected solely into  $Q_{sd}$ , and  $F_\alpha$  is the one-electron operator in Equation 2.53. This equation is solved iteratively for  $\Omega_\alpha^1$  by using states obtained from the orthogonalization of  $|pqrs; \alpha\rangle$ , which span the space  $Q_{sd}$ .

There are some situations where a strong mixing between the reference  $|\alpha\rangle$  state

---

and other CASSCF states. For such situations, CASPT2 has been extended to a multi-dimensional perturbative approach where the reference space is spanned by different state-averaged CASSCF (SA-CASSCF) wavefunctions. This approach is referred to as multi-state CASPT2 (MS-CASPT2).<sup>101</sup> In MS-CASPT2, the reference space is spanned by  $d$  SA-CASSCF wavefunctions (cf. Equation 2.48),

$$P = \sum_{\alpha=1}^d |\alpha\rangle \langle\alpha| \quad (2.59)$$

where the SA-CASSCF states  $|\alpha\rangle$  are eigenfunctions of the zeroth-order Hamiltonian  $H_{\alpha}^0$ . Again,  $Q$  spans the rest of the Hilbert space. The approach is similar to the state-specific CASPT2 (SS-CASPT2), however the reference states span a subspace of dimension  $d$ . These reference states are eigenfunctions of the zeroth-order Hamiltonians,

$$H_{\alpha}^0 |\alpha\rangle = E_{\alpha}^{(0)} |\alpha\rangle \quad (2.60)$$

Let  $\Psi_p$  be the exact state of interest, and consider its projection into the P space  $\Psi_p^0$ , given by,

$$P\Psi_p = \Psi_p^0 \quad (2.61)$$

This projecton is clearly spanned by the vectors  $|\alpha\rangle$  ( $\alpha = 1, \dots, d$ ). The exact state of interest  $\Psi_p$  is generated by an operator  $\Omega_P$  analogous to the one defined in Equation 2.50, however it does not act anymore upon a single CASSCF state  $|\alpha\rangle$ , but upon the entire space generated by the SA-CASSCF reference states, as follows,

$$\Omega_P \Psi_p^0 = \Psi_p \quad (2.62)$$

The operator  $\Omega_P$  can be expanded in an order-by-order series expansion, analogously to the case of  $\Omega_{\alpha}$  for the SS-CASPT2 approach (Equation 2.50). In relation with Equation 2.60, whereby there is a zeroth-order Hamiltonian for each reference state, a partitioning needs to be performed for each reference state  $|\alpha\rangle$ ,

$$H = H_{\alpha}^0 + V_{\alpha} \quad (2.63)$$

---

The zeroth-order Hamiltonian is analogous to the SS-CASPT2 case, but now the first term accounts for the projection into the  $d$ -dimensional space  $P$ ,

$$H_\alpha^0 = \sum_{\beta=1}^d |\beta\rangle \langle\beta| F_\alpha |\beta\rangle \langle\beta| + Q_K F_\alpha Q_K + Q_{sd} F_\alpha Q_{sd} + Q_{tq\dots} F_\alpha Q_{tq\dots} \quad (2.64)$$

where  $F_\alpha$ ,  $Q_K$ ,  $Q_{sd}$  and  $Q_{tq\dots}$  have the same meaning as in Equation 2.53. It can be shown that the first-order term  $\Omega_P^1$  can be determined by solving the equations,

$$(E_\alpha^0 - F_\alpha) \Omega_P^1 |\alpha\rangle = Q_{sd} H |\alpha\rangle \quad (2.65)$$

which are analogous to Equation 2.58. This analogy indeed allows for expressing  $\Omega_P$  as a linear combination of all of the  $\Omega_\alpha^1$ , as

$$\Omega_P^1 = \sum_{\alpha=1}^d \Omega_\alpha^1 |\alpha\rangle \langle\alpha| \quad (2.66)$$

An effective Hamiltonian  $H_{eff}$  is defined for the MS-CASPT2 problem as follows,

$$H_{eff} = PH\Omega_P P \quad (2.67)$$

$H_{eff}$  acts exclusively on the  $P$  subspace. The projections of the exact states into  $P$ ,  $\Psi_p^0$ , are the eigenfunctions of the effective Hamiltonian,

$$H_{eff} \Psi_p^0 = E_p \Psi_p^0 \quad (2.68)$$

where  $E_p$  is the exact energy of the state  $\Psi_p$ ,

$$H \Psi_p = E_p \Psi_p \quad (2.69)$$

Considering  $\Omega_p$  up to the second order, the effective Hamiltonian is given by

$$H_{eff}^{2nd} = PHP + PH\Omega_P^1 P \quad (2.70)$$

Since the  $\Omega_P^1$  is a linear combination of the first order state-specific operators  $\Omega_\alpha^1$ , it is easy to construct the effective Hamiltonian matrix by using Equation 2.70 and the SA-

---

CASSCF reference states: the diagonal elements of the matrix are given by the SS-CASPT2 energies,

$$\begin{aligned} \langle \alpha | H_{eff}^{2nd} | \alpha \rangle &= E_{\alpha}^{2nd} \\ (\alpha &= 1, \dots, d) \end{aligned} \tag{2.71}$$

whereas the off-diagonal elements are given by,

$$\begin{aligned} \langle \beta | H_{eff}^{2nd} | \alpha \rangle &= \langle \beta | H\Omega_{\alpha}^1 | \alpha \rangle \\ (\alpha, \beta &= 1, \dots, d) \end{aligned} \tag{2.72}$$

The effective Hamiltonian matrix constructed this way need not be symmetric. Thus, to obtain the MS-CASPT2 energies and eigenstates, it suffices to symmetrize the matrix, after which the symmetrized matrix is diagonalized.

## 2.2 Electronic Structure Theory: Density Functional Theory

An alternative to the *ab initio* methods, explored in the previous sections to recover the correlation energy that the HF approach lacks, relies on two theorems first introduced and proved by Hohenberg and Kohn in 1964,<sup>102</sup> whereby it was determined that the wavefunction and the energy of an  $N$ -electron system can in principle be uniquely determined by the electron density  $\rho(r)$  of that system. Formally, these results reduce the complexity of determining the wavefunction of a system under study – which depends on  $4N$  coordinates including cartesian and spin coordinates – to the problem of determining its electron density, which is a function of three spatial coordinates ( $r = \sqrt{x^2 + y^2 + z^2}$ , in the present thesis represented as  $r_1$ ). These two theorems are of existence and uniqueness of a functional relation between the electron density and the energy of the system, although in practice they provide no hints regarding the form of such a relation. Thus, the real problem of density functional theory (DFT) is that of determining an approximate expression for a *functional*  $E[\rho]$  of the electron density (a functional in the mathematical sense of a function from a space of functions  $V$  to the real numbers,  $E : V \rightarrow \mathbb{R}$ ). Still, the popularity of the DFT method stems from the fact that several functionals of the electron density have been proposed with various degrees of success. Besides, the formalism employed to

---

determine the electron density that provides the best approximation to the exact energy given a specific approximate functional, devised by Kohn and Sham,<sup>103</sup> consists of solving a set of one-electron Schrödinger type equations analogous to the HF approach, so that the computational cost is analogous to the HF method while having the advantage of accounting for part of the correlation energy that is missing within the HF framework.

### 2.2.1 The Hohenberg-Kohn Theorems

The first theorem<sup>102</sup> states the existence of an electron density  $\rho(r)$  which uniquely determines the exact wavefunction and energy of the system under study. In other words, “*Any observable can be written as a functional of the electron density of the [stationary non-degenerate] ground state*”. Given an  $N$ -electron system of interest, characterized by the wavefunction  $\Psi(x_1, \dots, x_N)$ , the corresponding electron density is given by,

$$\rho(r_1) = N \int |\Psi(r_1, r_2, \dots, r_N)|^2 dr_2 \dots dr_N \quad (2.73)$$

where the spin coordinates have been integrated out. In what follows, the electron density will be expressed as  $\rho(r)$  instead of  $\rho(r_1)$ , unless the coordinates of two or more electrons need to be accounted for simultaneously. This electron density, which is also positive definite, must be such that its integral over  $\mathbb{R}^3$  provides the number of electrons  $N$  of the system,

$$\int \rho(r) dr = N \quad (2.74)$$

Thus, the first Hohenberg-Kohn theorem states, in particular, that the electron density uniquely determines the external potential  $v(r)$  that acts upon a system of  $N$  electrons. This implies that, for a molecular system, the electronic energy is a functional of the electron density,

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] \quad (2.75)$$

where  $T[\rho]$  is the kinetic energy,  $V_{ee}[\rho]$  is the electron-electron repulsive potential and  $V_{eN}[\rho]$  is, in the simplest case, the nuclear-electron attractive potential. The first two

---

terms in Equation 2.75 are usually gathered into the so-called Hohenberg-Kohn functional  $F_{HK}[\rho]$ .

If the  $V_{eN}$  term is represented in terms of the external potential  $v(r)$  (for example, in the case of the classical electron-nuclei attractive term) as,

$$V_{eN} = \int v(r_1)\rho(r_1)dr_1 \quad (2.76)$$

then it can be shown that  $v(r)$  is precisely the functional derivative of  $V_{eN}[\rho]$ ,

$$\frac{\delta V_{eN}}{\delta \rho} = v(r_1) \quad (2.77)$$

Thus, Equation 2.75 can be rewritten as follows:

$$E[\rho] = F_{HK}[\rho] + \int \rho(r_1)v(r_1)dr_1 \quad (2.78)$$

The second Hohenberg-Kohn theorem is an immediate consequence of the more general variation principle, and states that “*the electron density of a non-degenerate ground state can be calculated by determining the density that minimizes the energy of the ground state*”. In other words, for a given trial density  $\rho(\tilde{r}_1)$ , the following inequality holds:

$$E_0 \leq E[\rho(\tilde{r}_1)] \quad (2.79)$$

where  $E_0$  is the exact energy of the ground electronic state and  $E[\rho(\tilde{r}_1)]$  is computed using Equation 2.78. Thus, the ground state energy of a molecular system can be determined by finding the electron density for which  $E[\rho]$  is stationary, with the constraint in the electron density given by Equation 2.74. The stationary condition is given by,

$$\frac{\delta}{\delta \rho} \left\{ E[\rho] - \tau \left( \int \rho(r_1)dr_1 - N \right) \right\} = 0 \quad (2.80)$$

$$\frac{\delta F_{HK}[\rho]}{\delta \rho(r_1)} + v(r_1) = \tau$$

where the Lagrange multiplier  $\tau$  has been introduced to account for the constraint in Equation 2.74.

Again, the main problem of the DFT framework is the fact that no analytic expression

---

is known for the  $F_{HK}[\rho]$  functional of the exact density. However, approximate expressions can be proposed, so that the problem in practice becomes that of determining the density that minimizes  $E[\rho]$  for a given molecular system. It is important to emphasize that, strictly speaking, DFT is variational only if the exact functional is known. This implies that when using approximate functionals, DFT formally no longer respects the variational principle.

### 2.2.2 The Kohn-Sham Equations

The problem of determining the electron density  $\rho$  which minimizes the energy in Equation 2.78 was tackled by Kohn and Sham<sup>103</sup> by considering a system of non-interacting electrons. This system was used as reference to develop the working equations to determine  $\rho$ . For a system of non-interacting electrons subject to an external potential  $v_r(r)$ , where the subscript  $r$  is used to denote the reference system, the Hamiltonian is given as a sum of one-electron operators consisting of the kinetic energy plus the external potential,

$$H_r = \sum_i h_i(r_1) \quad (2.81)$$

with

$$h_i(r_1) = -\frac{1}{2}\nabla_i^2 + v_r(r_1) \quad (2.82)$$

(cf. Equation 2.14 for the case of the core Hamiltonian in the HF framework). The exact wavefunction obeying the antisymmetry principle is given by the Slater determinant

$$\Phi_0 = |\phi_1(r_1)\phi_2(r_2)\dots\phi_N(r_N)| \quad (2.83)$$

analogous to that of Equation 2.10 except for the fact that the spin part of the spin orbitals has been integrated out. Thus, the (orthonormal) set of molecular orbitals  $\{\phi_i\}_{i=1,\dots,N}$  that provides the wavefunction  $\Phi_0$  is obtained by solving the one-electron Schrödinger equations:

$$\left[ -\frac{1}{2}\nabla_i^2 + v_r(r_1) \right] \phi_i = \epsilon_i \phi_i \quad (2.84)$$

---

The electron density and the kinetic energy for the non-interacting system are given by

$$\rho(r_1) = \sum_{i=1}^N |\phi_i(r_1)|^2 \quad (2.85)$$

and

$$T_r[\rho] = \sum_{i=1}^N \left[ \langle \phi_i | -\frac{1}{2} \nabla_i^2 | \phi_i \rangle \right] \quad (2.86)$$

respectively. The total energy of the non-interacting system is obtained by using Equation 2.78, excluding the electron-electron repulsive term,

$$E_r[\rho] = T_r[\rho] + \int \rho(r) v_r(r_1) dr_1 \quad (2.87)$$

The electron density of the non-interacting system is precisely the one that minimizes the energy in Equation 2.87, so that it satisfies Equation 2.80, again excluding the electron-electron repulsion term,

$$\begin{aligned} \frac{\delta}{\delta \rho} \left\{ E_r[\rho] - \tau \left( \int \rho(r_1) dr_1 - N \right) \right\} &= 0 \\ \frac{\delta T_r[\rho]}{\delta \rho(r_1)} + v_r(r_1) &= \tau \end{aligned} \quad (2.88)$$

Now, let  $E[\rho]$  be the energy functional of an interacting system of  $N$  electrons. Then, by considering Equations 2.75 and 2.78  $E[\rho]$  can be written as follows:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + \int v(r_1) \rho(r_1) dr_1 \quad (2.89)$$

This expression can be rewritten by adding and subtracting the functional  $T_r[\rho]$ , which corresponds to the exact kinetic energy of the non-interacting system, and the coulomb repulsion term  $J[\rho]$ ,

$$E[\rho] = T_r[\rho] + J[\rho] + \int v(r) \rho(r) dr + [T[\rho] - T_r[\rho]] + [V_{ee}[\rho] - J[\rho]] \quad (2.90)$$

where  $J[\rho]$  is given by

---


$$J[\rho] = \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 \quad (2.91)$$

and the last two terms in squared brackets,

$$\begin{aligned} T_C[\rho] &= T[\rho] - T_r[\rho] \\ W_{XC}[\rho] &= V_{ee}[\rho] - J[\rho] \end{aligned} \quad (2.92)$$

are referred to as the correlation kinetic energy and the exchange correlation potential, respectively. They are embedded in a single term, referred to as the exchange correlation energy,

$$E_{XC}[\rho] = T_C[\rho] + W_{XC}[\rho] \quad (2.93)$$

If  $\rho(r)$  is sufficiently slowly varying in space, it can be shown that the exchange correlation energy can be expressed as follows,

$$E_{XC}[\rho] = \int v_{xc}[\rho(r_1)]\rho(r_1)dr_1 \quad (2.94)$$

where the functional derivative of  $E_{XC}[\rho]$

$$\frac{\delta E_{XC}[\rho]}{\delta \rho} = v_{xc}[\rho(r_1)] \quad (2.95)$$

is termed the exchange correlation potential of the interacting system. Applying the stationarity condition on the energy of the interacting system (as in Equation 2.80) under the constraint 2.74, one obtains,

$$\frac{\delta T_r[\rho]}{\delta \rho} + v_c(r_1) + \mu_{xc}[\rho(r_1)] = \tau \quad (2.96)$$

where

$$v_c(r_1) = v(r_1) + \int \frac{\rho(r_2)}{|r_1 - r_2|} dr_2 \quad (2.97)$$

and

---


$$\mu_{xc}[\rho(r_1)] = \frac{\delta v_{xc}[\rho]}{\delta \rho} \quad (2.98)$$

corresponds to the exchange and correlation contribution of the potential. It should be emphasized that Equation 2.96 is analogous to Equation 2.88, apart from the fact that the term  $v_c(r) + \mu_{xc}[\rho(r)]$  appears instead of the external potential for the non-interacting system  $v_r(r)$ . This implies that the density that satisfies Equation 2.96 can be obtained by solving the one-electron Schrödinger equations:

$$\left\{ -\frac{1}{2}\nabla_i^2 + [v_c(r_1) + \mu_{xc}[\rho(r_1)]] \right\} \phi_i(r_1) = \epsilon_i \phi_i(r_1) \quad (2.99)$$

which are referred to as the Kohn-Sham (KS) equations. The electron density is computed from the molecular orbitals  $\phi_i$  by using Equation 2.85. These equations are analogous to the HF equations (cf. Equations 2.12 and 2.13), but now the exchange correlation potential  $\mu_{xc}[\rho]$  appears instead of the exchange operator  $K$ . Thus, the equations can be solved self-consistently in a similar manner as in the HF approach, usually employing a set of basis functions (Equation 2.17) to expand the molecular orbitals, so that the problem again becomes that of determining the optimal expansion coefficients for these molecular orbitals. Since the exact exchange correlation functional is unknown, approximate functionals need to be employed.

### 2.2.3 Exchange Correlation Functionals

As stated above, a caveat for the application of the KS approach lies on the fact that, as no explicit representation for the exact exchange correlation functional exists, an approximate expression must be employed. Different functional expressions for  $E_{XC}[\rho]$  have been provided, depending on the properties of interest. In most cases, these functionals are provided with a set of parameters that are tuned by fitting with experimental data or *ab initio* calculations. There are four main approaches that have been adopted to propose functional expressions for the exchange correlation energy, depending on the form of the dependence of the potential on the electron density. Perhaps the simplest one is the local density approximation (LDA), where  $E_{XC}[\rho]$  depends exclusively on the electron density of a uniform gas that has the same density as the system under study. A generalized version

---

of this approach is the local spin density approximation (LSDA), in which different electron densities are given to electron with different spins. In general, LDA and LSDA are suitable for systems whose electron densities are not expected to undergo significant variations in space.

The LDA scheme needs to be improved whenever the non-uniformity of the electron density is to be accounted for. A direct extension consists of considering an explicit functional dependence of  $E_{XC}[\rho]$  on the gradient of the electron density,  $\nabla\rho(r)$ . This approach is referred to as the generalized gradient approximation (GGA). Natural extensions to the GGA approach are obtained by considering higher order terms of the Taylor series expansion of the electron density, for example, the meta-GGA approaches, which consider the laplacian  $\nabla^2\rho(r_1)$  of the electron density. The fourth category of widely employed functionals is the so-called det of hybrid functionals. These are the result of a further extension to the GGA (or meta-GGA) hybrid approach, in which a portion of the exact exchange  $K$  from the HF theory is incorporated into the analytic expression of the functional. Some examples of these functionals are the B3LYP,<sup>104–106</sup> the CAM-B3LYP<sup>107</sup> and the M062X,<sup>108</sup> which have been extensively used throughout the present thesis.

## 2.2.4 Time-Dependent Density Functional Theory

### Formalism

The main limitation of the DFT framework explored in the previous sections stems from the fact that it regards exclusively systems in the electronic ground state. Therefore, it was subsequently extended to a time-dependent framework in what is known as time-dependent density functional theory (TD-DFT), which gave rise to the possibility of studying electronic excited states.<sup>109–111</sup> TD-DFT relies on an extension of the first Hohenberg-Kohn theorem to case in which the external potential – hereby represented as  $v(r_1, t)$  – is time-dependent, the Runge-Gross theorem.<sup>112</sup> It states that the initial state  $\Psi_0$ , as well as the initial density  $\rho_0$  are determined at the initial time  $t_0$ . Besides, the time-dependent density  $\rho(t)$  determines the wavefunction up to a time-dependent phase factor,

$$\Psi(t) = e^{-i\xi(t)}\Psi[\rho, \Psi_0](t) \tag{2.100}$$

---

As a consequence, there is a one-to-one correspondence between the time-dependent density  $\rho(t)$  and the external potential  $v(r_1, t)$ . In the time-dependent case, there is a variational theorem which provides a stationarity (but not a minimum) condition, which involves the action  $\mathcal{A}$ ,

$$\mathcal{A} = \int_{t_0}^{t_1} \langle \Psi(t) | i \frac{\partial}{\partial t} - H(t) | \Psi(t) \rangle dt \quad (2.101)$$

The (unconstrained) stationarity condition upon  $\mathcal{A}$  implies that its functional derivative with respect to the density  $\rho(r, t)$  vanishes in the case of the true density of the system. The action can be split into two separate terms,

$$\mathcal{A}[\rho] = B[\rho] - \int_{t_0}^{t_1} \int v(r, t) dr dt \quad (2.102)$$

where  $B[\rho]$  is a universal functional, and  $v(r, t)$  is the time-dependent external potential. As in the time-independent case, it is assumed that there exists an effective (reference) potential  $v_c(r_1, t)$  for a non-interacting system, whose orbitals  $\psi_i(r_1, t)$  give the same electron density as that of the real (interacting) system,

$$\rho(r_1, t) = \sum_{i=1}^N |\psi_i(r_1, t)|^2 \quad (2.103)$$

Under this assumption, the universal functional  $B[\rho]$  can be split as follows:

$$\begin{aligned} B[\rho] = & \sum_i \int_{t_0}^{t_1} \langle \psi_i(t) | i \frac{\partial}{\partial t} - \frac{1}{2} \nabla_i^2 | \psi_i(t) \rangle dt \\ & - \frac{1}{2} \int_{t_0}^{t_1} \int \frac{\rho(r_1, t) \rho(r_2, t)}{|r_1 - r_2|} dr_1 dr_2 dt - \mathcal{A}_{XC}[\rho] \end{aligned} \quad (2.104)$$

where the (unknown) exchange correlation functional  $\mathcal{A}_{XC}[\rho]$  has been defined. If the action is minimized under the constraint 2.103, one obtains a set of one-electron Schrödinger equations analogous to the KS equations in the time-independent case,

$$\left\{ -\frac{1}{2} \nabla_i^2 + v_c(r_1, t) \right\} \psi_i(r_1, t) = i \frac{\partial \psi_i(r_1, t)}{\partial t} \quad (2.105)$$

where

---


$$v_c = v(r_1, t) + \int \frac{\rho(r_2, t)}{|r_1 - r_2|} dr_2 + v_{xc}(\rho, t) \quad (2.106)$$

and  $v_{xc}(\rho, t)$  is the functional derivative of the exchange correlation action functional  $\mathcal{A}_{XC}$ ,

$$v_{xc}(\rho, t) = \frac{\delta \mathcal{A}_{XC}}{\delta \rho(r_1, t)} \quad (2.107)$$

An approximation is made in the case of an external potential which varies slowly with time: it is the adiabatic local density approximation (ALDA), in which the functional  $\mathcal{A}_{XC}[\rho]$  is expressed in terms of  $E_{XC}$ , the exchange correlation functional of the time-independent KS theory,

$$\mathcal{A}_{XC} = \int_{t_0}^{t_1} E_{XC}[\rho_t] dt \quad (2.108)$$

where  $\rho_t$  indicates the electron density evaluated at time  $t$ . This approximation provides the following expression for  $v_{xc}(\rho, t)$ ,

$$v_{xc}[\rho](r_1, t) = \frac{\partial \mathcal{A}_{XC}[\rho]}{\partial \rho(r_1, t)} \approx \frac{\partial E_{XC}[\rho_t]}{\partial \rho(r_1, t)} = v_{xc}[\rho_t](r_1) \quad (2.109)$$

where the term to the right corresponds to the exchange correlation of the time-independent KS theory, and thus does not display an explicit dependence on  $t$ . It should be emphasized that within the ALDA, the exchange correlation potential depends on time only in the sense that it depends on the electron density computed at a specific time  $t$ , but does not depend on the density at previous times. Thus, it is assumed that the potential changes instantaneously following a change in the electron density.

### Linear Response TD-DFT

The working equations in TD-DFT are determined using a perturbative approach. The starting point are the time-dependent KS equations (Equation 2.105), although it is convenient to rewrite the external potential  $v(r_1, t)$  as follows:

$$v(r_1, t) \rightarrow v_N(r_1) + v(r_1, t) \quad (2.110)$$

---

where  $v_N(r_1)$  is the electron-nuclear attraction potential, and  $v(r_1, t)$  now represents the time-dependent external potential for which the response of the system will be computed. At this point, the KS Hamiltonian in Equation 2.105 becomes

$$F = -\frac{1}{2}\nabla_i^2 + v_N(r_1) + v(r_1, t) + v_J(r_1) + v_{xc}(\rho, t) \quad (2.111)$$

where

$$v_J(r_1) = \int \frac{\rho(r_2, t)}{|r_1 - r_2|} dr_2 \quad (2.112)$$

The time-dependent molecular orbitals in 2.105 are expanded in the basis of  $M$  time-independent KS orbitals ( $\phi_i$  in Equation 2.99),

$$\psi_r = \sum_{j=1}^M C_{rj}(t)\phi_j(r_1) \quad (2.113)$$

where the time-dependent coefficients  $C_{rj}(t)$  are to be determined. If Equations 2.111 and 2.113 are inserted in Equation 2.105, one obtains,

$$F \sum_j C_{rj}(t)\phi_j(r) = i \sum_j \dot{C}_{rj}(t)\phi_j(r) \quad (2.114)$$

where  $\dot{C}_{rj}$  is the time derivative of the coefficient  $C_{rj}$ . If then one multiplies to the left by  $\phi_i (i = 1, 2, ..M)$ , where  $M$  is the total number of time-independent KS orbitals, and then integrates, the following matrix equation is obtained,

$$\mathbf{FC} = i\dot{\mathbf{C}} \quad (2.115)$$

If the density matrix is defined as

$$\mathbf{D} = \mathbf{CC}^\dagger \quad (2.116)$$

and if suitable matrix manipulations are employed on Equation 2.115, the latter can be rewritten as follows:

---


$$\begin{aligned}\mathbf{F}\mathbf{D} - \mathbf{D}\mathbf{F} &= i\dot{\mathbf{D}} \\ [\mathbf{F}, \mathbf{D}] &= i\frac{\partial \mathbf{D}}{\partial t}\end{aligned}\tag{2.117}$$

which corresponds to Liouville's equation in the KS case.<sup>110</sup>

A perturbative approach up to the first order is adopted for the operators  $F(t)$  (cf. Equation 2.111) and  $D(t)$ , associated with the matrices  $\mathbf{F}$  and  $\mathbf{D}$ , respectively, in which one considers the parameter  $\lambda \in [0, 1]$ :

$$\mathbf{D}(t) = \mathbf{D}^{(0)} + \lambda \mathbf{D}^{(1)}\tag{2.118}$$

$$\begin{aligned}F(t) &= -\frac{1}{2}\nabla_i^2 + v_N + v_J^{(0)} + \lambda v_J^{(1)} + v_{xc}^{(0)} + \lambda v_{xc}^{(1)} + \lambda v(t) \\ &= F^{(0)} + \lambda(v_J^{(1)} + v_{xc}^{(1)}) + \lambda v(t) \\ &= F^{(0)} + \lambda F^{(1)}\end{aligned}\tag{2.119}$$

where

$$\begin{aligned}F^{(0)} &= -\frac{1}{2}\nabla^2 + v_N + v_J^{(0)} + v_{xc}^{(0)} \\ F^{(1)} &= v_J^{(1)} + v_{xc}^{(1)} + v(t)\end{aligned}\tag{2.120}$$

Now, let  $\rho(r_1, r'_1)$  be the one-particle density matrix, defined in terms of the N-electron wavefunction  $\Psi(x_1, x_2, \dots, x_N)$ ,<sup>88</sup>

$$\rho(r_1, r'_1) = \int \int \Psi^*(r_1, \sigma_1, x_2, \dots, x_N) \Psi(r'_1, \sigma_1, x_2, \dots, x_N) d\sigma_1 dx_2 \dots dx_N\tag{2.121}$$

where  $x_i = (r_i, \sigma_i)$ ,  $\{i = 1, \dots, N\}$  indicate the joint spatial and spin coordinates, and  $r_1$  and  $r'_1$  indicate two different spatial integration variables.  $\rho(r_1, r'_1)$  is a generalization of the ordinary electron density  $\rho(r_1)$  (Equation 2.85). The unperturbed density matrix  $\mathbf{D}^{(0)}$  trivially corresponds to the identity matrix in the basis of the  $M$  time-independent KS spin orbitals. This implies that the (unperturbed) one particle density matrix  $\rho(r_1, r'_1)^{(0)}$  is just the ordinary electron density obtained from the Slater determinant constructed with these orbitals,

$$\rho(r_1, r'_1)^{(0)} = \sum_{r,s=1}^M \phi_r(r_1)^* D_{rs}^{(0)} \phi_r(r'_1) = \sum_{j=1}^N \phi_r(r_1)^* \phi_j(r'_1)\tag{2.122}$$

---

for an  $N$ -electron system. The main goal is to solve Equation 2.117 for  $D$ , which up to a first order perturbation (in the operators  $F(t)$  and  $D(t)$ ) is given by

$$[(\mathbf{F}^{(0)} + \lambda\mathbf{F}^{(1)}), (\mathbf{D}^{(0)} + \lambda\mathbf{D}^{(1)})] = i \left( \dot{\mathbf{D}}^{(0)} + \lambda\dot{\mathbf{D}}^{(1)} \right) \quad (2.123)$$

This equation can be solved by imposing some constraints on the density matrix  $\mathbf{D}$ . In particular, if the matrix  $\mathbf{D}$  is required to be idempotent (in the sense that  $\mathbf{D}\mathbf{D} = \mathbf{D}$ ) to the first order, then it can be easily shown that

$$\mathbf{D}^{(0)}\mathbf{D}^{(1)} + \mathbf{D}^{(1)}\mathbf{D}^{(0)} = \mathbf{D}^{(1)} \quad (2.124)$$

An expression for  $\mathbf{D}^{(1)}$  is obtained by considering the matrix  $\mathbf{D}$  as a 4-block square matrix having blocks OO, OV, VO and VV, where O=occupied and V=virtual orbitals. In this representation, the matrices  $\mathbf{D}^{(0)}$  and  $\mathbf{D}^{(1)}$  are given by

$$\mathbf{D}^{(0)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \mathbf{D}^{(1)} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{C} \end{bmatrix} \quad (2.125)$$

The matrix  $\mathbf{D}^{(1)}$  is determined by using Equation 2.124,

$$\mathbf{D}^{(1)} = \begin{bmatrix} \mathbf{0} & \mathbf{P}^\dagger \\ \mathbf{P} & \mathbf{0} \end{bmatrix} \quad (2.126)$$

so that the linear response (of the one-particle density matrix) to the perturbation  $v(r_1, t)$  is given by,

$$\rho^{(1)}(r_1, r'_1, t) = \sum_{\mu\nu} \phi_\mu(r_1) P_{\mu\nu} \phi_\nu^*(r'_1) + \sum_{\mu\nu} \phi_\nu(r_1) P_{\nu\mu} \phi_\mu^*(r'_1) \quad (2.127)$$

with  $\mathbf{P}$  Hermitian. On the other hand, by computing the functional variations of  $v_j^{(1)}$  and  $v_{xc}^{(1)}$  it can be shown that their matrix elements in the basis of time-independent KS orbitals are given by

$$\langle \phi_r | v_j^{(1)} + v_{xc}^{(1)} | \phi_s \rangle = \sum_{\mu\nu} P_{\nu\mu}(t) \langle r\mu | g + g_{xc} | s\nu \rangle + \sum_{\mu\nu} P_{\nu\mu}^*(t) \langle r\nu | g + g_{xc} | s\mu \rangle \quad (2.128)$$

where

---


$$g = \frac{1}{|r_1 - r_2|}$$

$$g_{xc} = \frac{\delta^2 E_{XC}[\rho]}{\delta\rho(r_1)\delta\rho(r_2)}$$
(2.129)

so that the first order perturbation of the KS Hamiltonian is given by

$$F^{(1)} = \langle\phi_r|v_J^{(1)} + v_{xc}^{(1)}|\phi_s\rangle + V$$
(2.130)

where  $V$  depends on the perturbation  $v(r_1, t)$ . Finally, introducing Equations 2.128 and 2.130 in Equation 2.123, applying the Fourier transform to move to the frequency domain  $\omega$ , and considering a negligible perturbation ( $V \rightarrow 0$ ), one obtains the working equations for TD-DFT,

$$\begin{bmatrix} \mathbf{M} - \mathbf{I}\omega & \mathbf{Q} \\ \mathbf{Q} & \mathbf{M} + \mathbf{I}\omega \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \mathbf{0}$$
(2.131)

with  $X_{\mu\nu} = P_{\nu\mu}(\omega)$ ,  $Y_{\mu\nu} = P_{\nu\mu}^*(-\omega) = P_{\mu\nu}(\omega)$  as functions of the frequency  $\omega$ .  $\mathbf{M}$  and  $\mathbf{Q}$  are given by

$$M_{j\mu,k\nu} = \delta_{jk}\delta_{\mu\nu}(\epsilon_\mu - \epsilon_\nu) + \langle j\nu|g + g_{xc}|\mu k\rangle$$

$$Q_{j\mu,k\nu} = \langle jk|g + g_{xc}|\mu\nu\rangle$$
(2.132)

It can be shown that the eigenvalues of Equation 2.131 correspond to the electronic excitation energies of a molecular system.

## 2.3 Electronic Structure Theory: Energy Decomposition Analysis

Most of the work presented in this thesis is dedicated to the study of the interactions between some anticancer drugs and the biological environment they are embedded in. For example, in Chapter 6 the interaction energies between cisplatin and a lipid membrane – considered as two subsystems, or fragments, of the entire system – have been computed at a quantum mechanical/molecular mechanical (QM/MM) level of theory (an introduction to the QM/MM framework is provided in section 2.5), and an interaction energy partitioning

---

scheme into physically meaningful terms has been employed to study the most important energy contributions to the interaction energy between cisplatin and the lipid membrane. This section aims at introducing the theoretical framework that was applied in Chapter 6 to perform an energy decomposition analysis (EDA) of the interaction energy. There are several different EDA schemes present in the literature based on either variational or perturbative approaches.<sup>113–116</sup> In the present thesis, a perturbative approach based on electronic deformation densities (which describe the first order perturbation of the electron density) devised by Marcos Mandado and collaborators has been employed.<sup>117,118</sup> Consider the case of a quantum mechanical system consisting of two interacting subsystems  $A$  and  $B$ . The interaction energy between  $A$  and  $B$  is defined as the difference between the energy of the complex  $AB$  (*i.e.*, the entire system) minus the energies of the subsystems  $A$  and  $B$ , computed at the exact geometries as they appear in the geometry of the complex, and using the basis functions of the complex. The latter is done in order to correct for the basis set superposition error,<sup>119,120</sup>

$$E_{int} = E_{AB} - (E_A^{AB} + E_B^{AB}) \quad (2.133)$$

where the superscript  $AB$  indicates that the basis set of the complex is being used. Let  $\Psi(x_1, \dots, x_N)$  be the wavefunction describing an  $N$ -electron molecular system. Recall that the one-particle density matrix is given by (cf. Equation 2.121),

$$\rho_1(r_1, r'_1) = \int \int \Psi^*(r_1, \sigma_1, x_2, \dots, x_N) \Psi(r'_1, \sigma_1, x_2, \dots, x_N) d\sigma_1 dx_2 \dots dx_N \quad (2.134)$$

where  $x_i = (r_i, \sigma_i)$ ,  $\{i = 1, \dots, N\}$  indicate the joint spatial and spin coordinates, and  $r_1$  and  $r'_1$  are two different coordinate variables.  $\rho_1(r_1, r'_1)$  is useful to represent non-multiplicative operators in terms of electron densities (*e.g.*, the kinetic energy operator). The two-particle density is defined as follows,

$$\rho_2(r_1, r_2) = N(N-1) \int \int |\Psi(r_1, \sigma_1, r_2, \sigma_2, x_3, \dots, x_N)|^2 d\sigma_1 d\sigma_2 dx_3 \dots dx_N \quad (2.135)$$

$\rho_2(r_1, r_2)$  can be seen as the probability density of finding an electron at  $r_1$  and at the same time another electron at  $r_2$ . As a result,  $\rho_2(r_1, r_2)$  implicitly accounts for the electron correlation between pairs of electrons. Thus, the total energy of a molecular

system consisting of  $N$  electrons and  $M$  nuclei can be expressed in terms of  $\rho_1(r_1, r'_1)$  and  $\rho_2(r_1, r_2)$ , as follows,

$$\begin{aligned}
E &= \langle \Psi | H | \Psi \rangle \\
&= \int h(r_1) \rho_1(r_1, r'_1) |_{r'_1=r_1} dr_1 + \int \int \frac{\rho_2(r_1, r_2)}{|r_1 - r_2|} dr_1 dr_2 + \\
&\quad + \sum_{I=1}^{M-1} \sum_{J>I}^M \frac{Z_I Z_J}{|R_I - R_J|}
\end{aligned} \tag{2.136}$$

where  $h(r_1)$  is the usual one-electron Hamiltonian representing the kinetic energy plus the electrostatic potential of the nuclei  $\nu_N$  (cf. Equation 2.14)

$$h(r_1) = -\frac{1}{2} \nabla^2 + \nu_N \tag{2.137}$$

in which the spin coordinate  $\sigma_1$  has been integrated out, and the last term in Equation 2.136 is the nucleus-nucleus Coulomb repulsion. The two-particle density can be represented as a sum of an uncorrelated term (the product of the ordinary electron densities  $\rho(r_1)$  and  $\rho(r_2)$  at  $r_1$  and  $r_2$ , respectively) plus a term bearing the exchange and correlation effects between pairs of electrons,

$$\rho(r_1, r_2) = \rho(r_1)\rho(r_2) + \rho_{xc}(r_1, r_2) \tag{2.138}$$

The total energy can now be written as follows,

$$\begin{aligned}
E &= - \int \frac{1}{2} \nabla^2 \rho_1(r_1, r'_1) |_{r'_1=r_1} dr_1 + \int \nu_N \rho(r_1) dr_1 + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + \\
&\quad + \frac{1}{2} \int \int \frac{\rho_{xc}(r_1, r_2)}{|r_1 - r_2|} dr_1 dr_2 + \sum_{I=1}^{M-1} \sum_{J>I}^M \frac{Z_I Z_J}{|R_I - R_J|}
\end{aligned} \tag{2.139}$$

The first term on the rhs of Equation 2.139 should be interpreted as follows: the kinetic energy operator acts upon  $\rho_1(r_1, r'_1)$ , after which the coordinate  $r'_1$  is set equal to  $r_1$ . In what follows, a perturbative approach is adopted for some of the operators and for the electron densities in Equation 2.139. Specifically, they will be represented as a sum of contributions from a hypothetical system consisting of two non-interacting fragments  $A$  and  $B$ , plus some perturbative terms that account for the interactions between the two fragments. In this regard, the nucleus-nucleus repulsion can be written as follows,

---


$$\begin{aligned} \sum_{I=1}^{M-1} \sum_{J>I}^M \frac{Z_I Z_J}{|R_I - R_J|} &= \sum_{I=1}^{M_A-1} \sum_{J>I}^{M_A} \frac{Z_I Z_J}{|R_I - R_J|} + \sum_{I=M_A+1}^{M-1} \sum_{J>I}^M \frac{Z_I Z_J}{|R_I - R_J|} + \\ &+ \sum_{I=1}^{M_A} \sum_{J=M_A+1}^M \frac{Z_I Z_J}{|R_I - R_J|} \end{aligned} \quad (2.140)$$

or more succinctly,

$$V_{M,M} = V_{M_A,M_A} + V_{M_B,M_B} + V_{M_A,M_B} \quad (2.141)$$

where  $M_A$  and  $M_B$  is the total number of nuclei in fragments  $A$  and  $B$ , respectively,  $V_{M,M}$  is the total molecular nucleus-nucleus repulsive potential,  $V_{M_A,M_A}$  ( $V_{M_B,M_B}$ ) is the nuclear repulsive potential involving exclusively nuclei of fragment  $A$  ( $B$ ), and  $V_{M_A,M_B}$  is the nuclear repulsive potential involving nuclei of both fragments. The nuclear electrostatic potential operator  $\nu_N$  (present in the second term of the rhs of Equation 2.139) can be split into two terms,

$$\nu_N = \nu_{N_A} + \nu_{N_B} \quad (2.142)$$

The electron density and the (two-particle) exchange correlation density are given by,

$$\rho(r) = \rho_A(r_1) + \rho_B(r_1) + \Delta\rho_{Pauli}(r_1) + \Delta\rho_{pol}(r_1) \quad (2.143)$$

and

$$\rho_{xc}(r_1, r_2) = \rho_{xc,A}(r_1, r_2) + \rho_{xc,B}(r_1, r_2) + \Delta\rho_{x,AB}(r_1, r_2) + \Delta\rho_{xc}(r_1, r_2) \quad (2.144)$$

respectively. The first two terms on the rhs of Equations 2.143 and 2.144 represent unperturbed one and two-particle electron densities for each of the fragments.  $\Delta\rho_{Pauli}(r_1)$  and  $\Delta\rho_{pol}(r_1)$  are referred to as Pauli and polarization deformation densities, respectively, and the last two terms in Equation 2.144 are the interfragment exchange (only) and exchange correlation densities. The integrands in the second and third terms in the rhs of Equation 2.139 (associated with the electrostatic nucleus-electron attraction and the electron-electron repulsion) can be embedded into two fragment-wise unperturbed potentials,

---


$$\begin{aligned}\nu_A(r_1) &= \nu_{N_A}(r_1) + \int \frac{\rho_A(r_2)}{r_2 - r_1} dr_2 \\ \nu_B(r_1) &= \nu_{N_B}(r_1) + \int \frac{\rho_B(r_2)}{r_2 - r_1} dr_2\end{aligned}\tag{2.145}$$

where  $\rho_A(r_2)$  and  $\rho_B(r_2)$  are the unperturbed electron densities of fragments  $A$  and  $B$ . At this point, using the expression of the energy in Equation 2.139 to compute the energy terms in Equation 2.133, and including Equations 2.140-2.145 in the expression of the energy, one obtains the following partitioning of the interaction energy:

$$E_{int} = E_{elec} + E_{exch} + E_{rep} + E_{pol}\tag{2.146}$$

whose explicit expressions are given by Equations 2.147-2.150

$$\begin{aligned}E_{elec} &= \int \nu_{N_A} \rho_B(r_1) dr_1 + \int \nu_{N_B} \rho_A(r) dr + \int \int \frac{\rho_A(r_1) \rho_B(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \\ &+ \sum_{I=1}^{N_A} \sum_{J=N_A+1}^N \frac{Z_I Z_J}{|R_I - R_J|}\end{aligned}\tag{2.147}$$

$$E_{exch} = \frac{1}{2} \int \int \frac{\rho_{x,AB}(r_1, r_2)}{|r_2 - r_1|} dr_1 dr_2\tag{2.148}$$

$$\begin{aligned}E_{rep} &= -\frac{1}{2} \int \nabla^2 \Delta \rho_{Pau}(r_1, r'_1)|_{r'_1=r_1} dr_1 + \int \nu_A \Delta \rho_{Pau}(r_1) dr_1 + \int \nu_B \Delta \rho_{Pau}(r_1) dr_1 + \\ &+ \frac{1}{2} \int \int \frac{\Delta \rho_{Pau}(r_1) \Delta \rho_{Pau}(r_2)}{|r_2 - r_1|} dr_1 dr_2\end{aligned}\tag{2.149}$$

$$\begin{aligned}E_{pol} &= -\frac{1}{2} \int \nabla^2 \Delta \rho_{pol}(r_1, r'_1)|_{r'_1=r_1} dr_1 + \int \nu_A \Delta \rho_{pol}(r_1) dr_1 + \int \nu_B \Delta \rho_{pol}(r_1) dr_1 + \\ &+ \frac{1}{2} \int \int \frac{\Delta \rho_{pol}(r_1) \Delta \rho_{pol}(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \int \int \frac{\Delta \rho_{Pau}(r_1) \Delta \rho_{pol}(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \\ &+ \frac{1}{2} \int \int \frac{\Delta \rho_{xc}(r_1, r_2)}{|r_2 - r_1|} dr_1 dr_2\end{aligned}\tag{2.150}$$

The exchange and repulsion terms of Equations 2.148 and 2.149 both arise as a consequence of the Pauli exclusion principle, so that they are merged into the single term  $E_{Pau}$ . The polarization energy includes the classical induction energy and the inherently quantum mechanical dispersion energy. These two contributions can be separated

exactly within the second-order perturbation theory framework considering the molecular wavefunction  $\Psi(x_1, \dots, x_N)$ .<sup>121</sup> Within that framework, it can be shown that the induction energy is given by,

$$E_{\text{ind}} = \sum_{m \neq 0} \frac{[\int \nu_A \rho_B^{m0}(r_1) dr_1]^2}{E_B^m - E_B^0} + \sum_{n \neq 0} \frac{[\int \nu_B \rho_A^{n0}(r_1) dr_1]^2}{E_A^n - E_A^0} \quad (2.151)$$

where  $\rho_A^{n0}$  and  $\rho_B^{m0}$  are the induced transition one-electron densities within each fragment from the ground state configuration 0 to the excited state  $m$  or  $n$ . The denominators contain the corresponding energy differences.

At the second order perturbation theory, the first order correction to the electron density  $\Delta\rho(r_1)$  (the deformation density) is required to obtain the second order correction for the energy. It can be shown that  $\Delta\rho(r_1)$  is given by<sup>121</sup>

$$\begin{aligned} \Delta\rho(r) &= \Delta\rho_A(r) + \Delta\rho_B(r) \\ &= 2 \sum_{n \neq 0} \frac{\int \nu_B \rho_A^{n0}(r) dr}{E_A^n - E_A^0} \rho_A^{n0} + 2 \sum_{m \neq 0} \frac{\int \nu_A \rho_B^{m0}(r) dr}{E_B^m - E_B^0} \rho_B^{m0} \end{aligned} \quad (2.152)$$

Equation 2.152 represents the polarization densities of fragments  $A$  and  $B$ . Within the framework presented in this section, it is possible to obtain an expression for the induction energy starting from the expression of the deformation density in Equation 2.152 and the polarization contribution to the interaction energy in Equation 2.150. At first, one defines the charge induction energy as the sum of the second and third terms in the rhs of Equation 2.150,

$$E_{\text{ch-ind}} = \int \nu_A \Delta\rho_{\text{pol}}(r) dr + \int \nu_B \Delta\rho_{\text{pol}}(r) dr \quad (2.153)$$

Then, by introducing Equation 2.152 in Equation 2.153, one obtains

$$\begin{aligned} E_{\text{ch-ind}} &= 2 \sum_{m \neq 0} \frac{[\int \nu_A \rho_B^{m0}(r_1) dr_1]^2}{E_B^m - E_B^0} + 2 \sum_{n \neq 0} \frac{[\int \nu_B \rho_A^{n0}(r_1) dr_1]^2}{E_A^n - E_A^0} + \\ &+ 2 \sum_{m \neq 0} \frac{\int \nu_A \rho_B^{m0}(r_1) dr_1 \int \nu_B \rho_B^{m0}(r_1) dr_1}{E_B^m - E_B^0} + 2 \sum_{n \neq 0} \frac{\int \nu_B \rho_A^{n0}(r_1) dr_1 \int \nu_A \rho_A^{n0}(r_1) dr_1}{E_A^n - E_A^0} \end{aligned} \quad (2.154)$$

At this point, the induction energy stemming from second order perturbation theory on

the molecular wavefunction<sup>121</sup> (Equation 2.151) can be identified as one half times the first two terms in the rhs of Equation 2.154. Thus, by introducing Equation 2.151 in Equation 2.153 and solving for  $E_{ind}$ , one obtains,

$$E_{ind} = \frac{1}{2} \left[ E_{ch-ind} - 2 \sum_{m \neq 0} \frac{\int \hat{v}_A \rho_B^{m0}(r_1) dr_1 \int \hat{v}_B \rho_B^{m0}(r_1) dr_1}{E_B^m - E_B^0} - 2 \sum_{n \neq 0} \frac{\int \hat{v}_B \rho_A^{n0}(r_1) dr_1 \int \hat{v}_A \rho_A^{n0}(r_1) dr_1}{E_A^n - E_A^0} \right] \quad (2.155)$$

Now, if the expression for the first-order density (Equation 2.152) is introduced in the last two terms of Equation 2.155, the following expression for  $E_{ind}$  is obtained,

$$E_{ind} = \frac{1}{2} \left[ E_{ch-ind} - \int \hat{v}_A \Delta \rho_A(r_1) dr_1 - \int \hat{v}_B \Delta \rho_B(r_1) dr_1 \right] \quad (2.156)$$

Finally, introducing the definition of the charge induction energy (Equation 2.153) in Equation 2.156, the induction energy becomes,

$$E_{ind} = \frac{1}{2} \left[ \int \hat{v}_A \Delta \rho_B(r_1) dr_1 + \int \hat{v}_B \Delta \rho_A(r_1) dr_1 \right] \quad (2.157)$$

which is in agreement with the classical result for the induction energy.<sup>118</sup> At this point, the energy decomposition of the interaction energy can be expressed in terms of a sum of electrostatic ( $E_{elec}$ ), Pauli ( $E_{Pau}$ ), inductive ( $E_{ind}$ ) and dispersive ( $E_{disp+respol}$ ) terms,

$$E_{ind} = E_{elec} + E_{Pau} + E_{ind} + E_{dis+respol} \quad (2.158)$$

In Equation 2.158, the second order dispersion energy represents the highest contribution to the  $E_{dis+respol}$ . The EDA scheme presented above has been extended to the QM/MM framework (Chapter 6). A specific situation in which the fragments  $A$  and  $B$  consisted of a substrate and the environment, respectively, was considered. In this case, the environment presented both atoms represented quantum mechanically and atoms represented as molecular mechanics fixed point charges. The interactions between the MM region and fragment  $A$  are introduced in the calculation of the different energy terms in Equation 2.149 by adding the potential generated by these point charges,  $V_{FF}$ , to the potential generated by the nuclei of fragment  $B$ ,

---


$$\nu_{N_B}^{\text{QM/MM}}(r) = \nu_{N_B}(r) + \hat{V}_{\text{FF}} \quad (2.159)$$

$V_{FF}$  is also included in the calculation of the nucleus-nucleus interactions in the  $E_{elec}$  term of the EDA scheme (last term in Equation 2.147).

## 2.4 Classical Molecular Dynamics

The study of the time evolution of a molecular system would in principle require to solve the time-dependent version of the Schrödinger equation (Equation 2.8), as both the nuclei and the electrons are treated quantum mechanically. The propagation of Equation 2.8 to obtain a time-dependent wavefunction  $\Psi(r, R, t)$  is embedded in the framework of quantum dynamics, which however is limited to systems consisting of a small number of atoms due to its computational cost. It was pointed out in section 2.1.1 that the fact that the nuclei are much heavier than the electrons could allow for an approximation to the form of the wavefunction  $\Psi(r, R, t)$  in which the motion of the nuclei was decoupled from the motion of the electrons, so that  $\Psi(r, R, t)$  could be represented as a product of an electronic wavefunction (which depends parametrically on the nuclear coordinates) times a nuclear wavefunction (which depends on the nuclear coordinates but is influenced by the potential generated by the electrons)—this was the BO approximation, represented in Equation 2.3. Within the BO approximation, one solves two separate Schrödinger equations, and both electrons and nuclei are treated quantum mechanically. However, even with this approximation, the solution of a time-dependent nuclear Schrödinger equation to study the motion of the nuclei with time is affordable for a limited number of atoms, and thus it is impracticable for model systems of complex environment media, which could involve tens to hundreds of thousands of atoms. To overcome these limitations, an approximation widely adopted consists of assuming that the nuclei behave as structureless particles that interact *via* a potential  $V$  and obey the laws of motion of a classical mechanical system: this approach is referred to as the classical MD methodology.<sup>122,123</sup> The equations of motion that govern the time evolution of a system of classical nuclei have a general character – in the same way as the Schrödinger equation does for a quantum mechanical system –, and derive from fundamental Laws of nature (the next section is dedicated to

---

the derivation of the classical equations of motion from one such principle) although the motion of the particles is defined once the potential  $V$  is known. In the simplest case,  $V$  can be described as an empiric analytic expression consisting of physically meaningful terms bearing parameters that are fit to experimental or quantum mechanical data – a force field (section B.3). The usage of a force field drastically simplifies the description of the system at hand, although it has been widely used with success to determine properties that depend on the motion of a system consisting of several thousands of atoms at an affordable computational cost.<sup>124</sup> The usage of a classical force field to describe the potential  $V$  of a system is suitable whenever the exploration of the configurational space is required. However, there are situations in which the description of the time evolution of the system needs to account for chemical reactivity, that is bond breaking and formation events: these phenomena are not properly described by using an analytic force field. To overcome this limitation, a methodology termed *ab initio* molecular dynamics (AIMD) has been devised by Helgaker<sup>125</sup> *et. al.*, in which the nuclei are treated classically, but the potential  $V$  is obtained by considering the electrons quantum mechanically. AIMD provides a less expensive approach with respect to the solution of the time-dependent Schrödinger Equation as the nuclei are treated classically; however, since the electrons of the system are treated quantum mechanically, this approach is limited to at most a few hundreds of atoms. This approach is explored in section 2.4.4. Finally, there may be situations in which there is the need to describe the configurational space of a system consisting of thousands of atoms, but the most interesting part of the system (for example in the case of a chemical reaction), which may consist of tens or at most a few hundreds of atoms, would require the usage of quantum mechanical methods to describe it. These are the situations considered in most of the systems studied in the present thesis. For these situations, an approach first proposed by Warshel and Levitt,<sup>126</sup> in which the part of chemical interest is described quantum mechanically, whereas the atoms surrounding the region of interest are described with a classical potential, and polarize the quantum mechanical part. It should be emphasized that the BO approximation is assumed, so that only the electrons of the region of interest are described quantum mechanically (Equation 2.5), whereas the motion of the nuclei is described classically. These methodologies will be explored in section 2.5.

---

### 2.4.1 Classical Equations of Motion

The classical equations of motion provide the relation between the interactions (forces, in classical mechanics) that act upon a set of (non)interacting atoms and the positions and momenta of these atoms. These equations can be derived for a general system of  $N$  interacting particles using the Lagrangian formulation of classical mechanics, based on the principle of least action,<sup>122,127,128</sup> which states that "the motion of the system from an initial time  $t_0$  to a final time  $t_1$  is such that the action integral  $\mathcal{A}$  has a stationary value for the actual path of the motion". In this context, action is a functional  $\mathcal{A}$  of the generalized coordinates  $q$  (that describe the degrees of freedom of the above mentioned particles, for example the coordinates in the cartesian space), their time derivatives  $\dot{q}$  and time.

$$\mathcal{A} = \int_{t_0}^{t_1} \mathcal{L}(q, \dot{q}, t) dt \quad (2.160)$$

Applying the stationarity condition upon Equation 2.160 provides the Euler-Lagrange equations,

$$\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{q}_k} \right) = \frac{\partial \mathcal{L}}{\partial q_k} \quad (2.161)$$
$$k = 1, 2, \dots, 3N$$

satisfied by any system of  $N$  particles. The Lagrangian is defined as the difference between the kinetic and the potential energies,

$$\mathcal{L} = T - V \quad (2.162)$$

One then proceeds to define the Hamiltonian  $\mathcal{H}$ , which in the most general case, is given by,

$$\mathcal{H} = \sum_k^{3N} \dot{q}_k p_k - \mathcal{L}(q_k, \dot{q}_k, t) \quad (2.163)$$

where the momenta  $p_k$  have been introduced. Within the Lagrangian formalism, the momenta are defined as

---


$$p_k = \frac{\partial \mathcal{L}}{\partial \dot{q}_k} \quad (2.164)$$

If the potential  $V$  depends exclusively on the positions  $q$  and the kinetic energy is expressed as,

$$T = \frac{p_k^2}{2m_k}$$

where  $m_k$  is the mass of the particle associated with the  $k^{th}$  coordinate, then the Hamiltonian assumes the form,

$$H = \sum_{k=1}^{3N} \frac{p_k^2}{2m_k} + V(q_k) \quad (2.165)$$

and the Euler-Lagrange equations in 2.161 provide Hamilton's equation of motion,

$$\begin{aligned} \dot{q}_k &= \frac{\partial H(q_k, p_k)}{\partial p_k} = \frac{\partial T(q_k)}{\partial p_k} \\ \dot{p}_k &= - \frac{\partial H(q_k, p_k)}{\partial q_k} = - \frac{\partial V(q_k)}{\partial q_k} \end{aligned} \quad (2.166)$$

In cartesian coordinates,  $\dot{q}_k$  can be represented as,

$$\dot{q}_k = p_k/m_k \quad (2.167)$$

Introducing this Equation in the expression involving  $\dot{p}_k$  in Equation 2.166, one can identify the Newton's equations of motion,

$$m_k \ddot{q}_k = -F_k \quad (2.168)$$

with  $\dot{p}_k = m_k \dot{q}_k$  and  $-\frac{\partial V(q_k)}{\partial q_k} = F_k$ , the force along the  $k^{th}$  coordinate. Thus, once the initial conditions are given, one can proceed to solve a either set of  $6N$  first-order differential equations (Equation 2.167) or a set of  $3N$  second-order differential equations (Equation 2.168). In either case, the main problem lies on the fact that for a system of  $N$  particles, with  $N > 2$ , no analytic solutions exist. As a consequence, the equations of motion need to be integrated numerically.

---

## 2.4.2 Integration Methods

In general the numerical method adopted to solve either Equation 2.166 or Equation 2.168 is the so-called finite differences method. Given the  $3N$  coordinates (or  $6N$  coordinates and momenta) of the system of particles at time  $t$ , one proceeds to determine the positions (and momenta) at a time  $t + \Delta t$ , for a time step  $\Delta t$  sufficiently small. The equations of motion are then solved in an iterative fashion one time step at a time. It is important that the time step  $\Delta t$  be sufficiently small (at least smaller than the time scale of the process under study). The following development will focus on the solution of Equation 2.168, for which the positions are represented in cartesian coordinates. Therefore, in what follows the positions and velocities will be represented using the more customary  $r(t)$  and  $v(t)$ , respectively, instead of the generalized coordinates  $q(t)$  and  $\dot{q}(t)$ . The general idea consists of obtaining an estimate for the coordinates and the velocities at the time  $t + \Delta t$  by means of a Taylor expansion of these quantities,

$$\begin{aligned} r_k(t + \Delta t) &= r_k(t) + \dot{r}_k(t)\Delta t + \frac{1}{2}\ddot{r}_k(t)\Delta t^2 + \dots \\ v_k(t + \Delta t) &= v_k(t) + \dot{v}_k(t)\Delta t + \frac{1}{2}\ddot{v}_k(t)\Delta t^2 + \dots \end{aligned} \quad (2.169)$$

The first and the second derivatives of the positions  $r_k$  are the velocities  $v_k$  and the accelerations  $a_k$ , respectively. Perhaps one of the most widely used algorithms to integrate the equations of motion is the Störmer-Verlet algorithm,<sup>122</sup> which provides an estimate for the positions up to the  $\Delta\mathcal{O}(t^4)$  order. The expression for  $r(t + \Delta t)$  is obtained by adding together the Taylor expansions for  $r_k(t + \Delta t)$  and  $r_k(t - \Delta t)$  and truncating the resulting series at the fourth order term, thus obtaining

$$r_k(t + \Delta t) = 2r_k(t) - r_k(t - \Delta t) + \Delta t^2 a_k(t) \quad (2.170)$$

so that the position at times  $t - \Delta t$  and  $t$  needs to be known to determine  $r_k(t + \Delta t)$ . The acceleration  $a_k(t)$  is determined by using Equation 2.168 (with  $r_k(t)$  instead of  $q_k(t)$ ), so that it depends on the specific form of the potential  $V(r)$ . Although the velocity at time  $t$  is not necessary to determine the position  $r_k(t + \Delta t)$ , it is still useful to compute the kinetic energy. It can be determined by subtracting  $r_k(t + \Delta t)$  and  $r_k(t - \Delta t)$ , and truncating at the second order term,

---


$$v_k(t) = \frac{r_k(t + \Delta t) - r_k(t - \Delta t)}{2\Delta t} \quad (2.171)$$

The main problem with the Störmer-Verlet algorithm lies on the fact that it needlessly introduces some numerical imprecision due to the addition of a small term ( $\mathcal{O}(\Delta t^2)$ ) to a difference between two large terms ( $\mathcal{O}(\Delta t^0)$ ).<sup>122</sup> Another problem is associated with the fact that the velocity at time  $t$  can be computed only after  $r_k(t + \Delta t)$  is known. To overcome these issues, a modified version of the Störmer-Verlet has been proposed, in which “mid-step” velocities are computed to determine  $r_k(t + \Delta t)$ ,

$$\begin{aligned} r_k(t + \Delta t) &= r_k(t) + v_k(t + \frac{1}{2}\Delta t)\Delta t \\ v_k(t + \Delta t) &= v_k(t - \Delta t) + a_k(t)\Delta t \end{aligned} \quad (2.172)$$

This algorithm is known as the leap-frog.<sup>129</sup> The velocities are computed as follows:

$$v_k(t) = \frac{1}{2}[v_k(t + \frac{1}{2}\Delta t) + v_k(t - \frac{1}{2}\Delta t)] \quad (2.173)$$

Although it represents an improvement over the Störmer-Verlet algorithm, the velocity is not handled satisfactorily, in the sense that the computation of  $v(t + \Delta t)$  is not involved in the main loop (Equation 2.172), but instead requires an external step to be computed (Equation 2.173). In other words, to compute the position at  $r_k(t + \Delta t)$  only the mid-step velocity  $v_k(t + \frac{1}{2}\Delta t)$  (but not the velocity at the time  $t + \Delta t$ ) is necessary, so that Equation 2.173 represents an extra step in the propagation of the leapfrog algorithm, which is necessary only if the velocities are required. Another modification of the Störmer-Verlet algorithm, which unlike the leapfrog algorithm does include the computation of the velocities in the main loop, is the velocity-Verlet,<sup>130</sup> where the acceleration, the velocity and the position related to the same time step  $t$  are available simultaneously and are used to compute  $r_k(t + \Delta t)$ ,

$$\begin{aligned} r_k(t + \Delta t) &= r_k(t) + v_k(t)\Delta t + \frac{1}{2}a_k(t)\Delta t^2 \\ v_k(t + \Delta t) &= v_k(t) + \frac{1}{2}[a_k(t) + a_k(t + \Delta t)]\Delta t \end{aligned} \quad (2.174)$$

This method is perhaps the most widely used due to its numerical stability, convenience and simplicity.<sup>122</sup> At this point, the last ingredient necessary to integrate the equations of

---

motions is the potential  $V(r_k)$ , whose form depends on the level of accuracy required to describe the system.

### 2.4.3 Classical Potentials: Force Fields and Molecular Mechanics

Considering the fact that in classical MD each atom is described as a structureless particle interacting with the other atoms present in the system under study, it comes natural to employ an analytic expression that depend on empirical parameters that are tuned using experimental data or quantum mechanical calculations as reference. They are constructed based on classical energy terms that describe the interactions that the atoms of the system under study undergo. These analytic expressions are termed force fields although, strictly speaking, a force field would not refer to the expression of the potential  $V(r)$  (a scalar field), but to its gradient  $\nabla V(r)$  (a vector field), which is the actual force term appearing in the equations of motion 2.168. The advantage of these force fields consists of the fact that they allow for the description of systems consisting of several thousands – and even a few millions – of atoms at an affordable computational cost. In general these force fields consist of additive terms describing bonded contributions to the potential that model covalent bonds between atoms, and non-bonded terms which account for the electrostatic interactions, as well as non-electrostatic terms that model interactions between atoms that are intrinsically quantum mechanical (exchange repulsion, dispersion, induction). A very popular functional form of such a potential is provided by the AMBER force field,<sup>131</sup> which is implemented in the AMBER<sup>132</sup> software, and has been thoroughly used throughout the present thesis work to generate ensembles of geometries. The AMBER force field has the following structure:<sup>133</sup>

$$\begin{aligned}
 V^{AMBER} = & \sum_{i \in \text{bonds}} K_{bi}(b_i - b_i^0)^2 + \sum_{i \in \text{angles}} K_{\theta_i}(\theta_i - \theta_i^0)^2 + \\
 & + \sum_{i \in \text{torsions}} \sum_n \frac{1}{2} K_{i,n} [1 + \cos(n\phi_i + \gamma_i)] + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{R_{ij}} \right] \quad (2.175)
 \end{aligned}$$

where  $b_i^0$ ,  $\theta_i^0$  and  $\phi_i$  are the equilibrium values for the bonds, bond angles and dihedral angles respectively, and  $K_{bi}$ ,  $K_{\theta_i}$  and  $K_{i,n}$  are force constants for the same terms.  $n$  describes the periodicity of the torsional angle and  $\gamma_i$  is the phase angle of the  $i^{\text{th}}$  torsional

---

angle (used so that a minimum of the corresponding torsional energy occurs at 180 degrees).  $R_{ij}$  describes interatomic distances;  $A_{ij}$  and  $B_{ij}$  are parameters that account for the exchange repulsion and non-electrostatic attraction between atoms  $i$  and  $j$ , respectively, and  $q_i$  and  $q_j$  represent atomic charges. Although this is a general analytic form that describes the energetics governing the interactions established in a system of  $N$  atoms, the terms mentioned above are parameters that need to be tuned depending on the system under study. In practice in molecular mechanics (MM) it is customary to define different atom types and the tuned parameters are associated with a specific set of atom types. These atom types characterize elements present on a specific atomic environment, so that the same parameters can be used for different systems that present atoms with similar atomic environments. Thus, in practice, whenever one refers to a specific force field, what is actually referred to is not only the analytic expression in Equation 2.175, but also to a set of atom types and parameters characterizing the system under study. In the present thesis different sets of force fields have been used to describe different biological systems. Specifically, the generalized amber force field (GAFF)<sup>131</sup> has been used to describe small organic molecules, whereas the lipid17<sup>134,135</sup> force field has been used to describe a model of a cellular lipid bilayer, the CHARMM36<sup>136</sup> force field has been used to describe proteins, the TIP3P<sup>137</sup> to describe water molecules and the OL15<sup>138</sup> to describe nucleotides.

#### 2.4.4 Ab Initio Molecular Dynamics

The force fields described in the previous section are suitable to explore the conformational space of a system consisting of hundreds of thousands of atoms. However, the form of the potential in Equation 2.175 is not suitable to study phenomena which involve, for example, bond breaking and formation. As the BO approximation is implicit on the classical MD framework, the motion of the nuclei is inherently decoupled from that of the electrons. Therefore, within the law of motion that describes the motion of the nuclei (Equation 2.168), the potential on which the nuclei move can be given by the potential generated by the electrons surrounding the nuclei, whereby this potential is obtained by describing quantum mechanically the electrons. This strategy to solve the equations of motion is referred to as *ab initio* molecular dynamics (AIMD),<sup>125</sup> in which the potential energy  $V(r)$ ,  $r = (r_1, r_2, \dots, r_{3N})$  is computed quantum mechanically by solving the time-

---

independent Schrödinger equation 2.1, after which its gradient  $\nabla V(r)$  is computed and introduced in Equation 2.168. As the classical equations of motion are solved for the coordinates  $r(t)$ , they are propagated numerically by using any of the algorithms presented in section 2.4.2, so that at every time step  $t$ , the coordinates  $r(t)$  are used to compute the energy and the gradient  $-V(r(t))$  and  $\nabla V(r(t))$ , respectively – with a suitable quantum mechanical method, which in turn is used to determine the position of the atoms at time  $t + \Delta t$ . Some examples of applications of AIMD include the exploration of reaction pathways and the research of transition state structures,<sup>139,140</sup> and the computation of thermodynamic properties at a high level of theory.<sup>56,57,141</sup>

### 2.4.5 Enhanced Sampling: the Umbrella Sampling Technique

The classical MD methodology introduced in the previous sections allows for an exploration of the configurational space of the molecular system of interest. This exploration is necessary whenever one is interested in computing thermodynamic properties (*e.g.*, Gibbs or Helmholtz free energies), which are obtainable as ensemble averages over contributions due to different minima on the potential energy landscape which can be populated at the temperature of interest. In order to explore different minima on the potential energy landscape *via* classical MD, the energetic barriers that separate these minima from one another need to be overcome as the equations of motion are integrated. However, there may be situations in which the barrier separating two minima are high in energy (much higher than  $k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is the temperature), so that in practice it is unlikely that two minima are thoroughly sampled in a finite amount of time when such a high barrier separates the two minima. Another example of a problem due to the presence of high energy regions in the potential energy surface may arise when one is interested in studying the energetics of a chemical process along the entire reaction pathway separating two energy minima: for a finite simulation time, the region around either of the minima would be sampled well, whereas the regions of higher energy would be sampled rarely. To overcome these difficulties and to thoroughly sample high energy regions along the potential energy surface, different computational techniques – termed Accelerated Sampling Techniques – have been devised. In the most general case, these techniques can be classified in two categories: collective variable based and

---

collective variable free methods.<sup>142</sup> The collective variable based techniques include umbrella sampling,<sup>45,143</sup> hyperdynamics,<sup>144</sup> metadynamics,<sup>145,146</sup> potential smoothing method,<sup>147</sup> variational enhanced sampling,<sup>148</sup> among others. As the name of the category suggests, these techniques rely on the definition of some pre-defined collective variables (also termed reaction coordinates) along which to guide the simulation. For those situations for which the reaction coordinate of the process under study is not known, the collective variable free category comes into play. Some examples include replica exchange molecular dynamics,<sup>149</sup> temperature-accelerated dynamics,<sup>150</sup> accelerated molecular dynamics simulation,<sup>151,152</sup> integrated tempering sampling,<sup>153,154</sup> among others.

In the present thesis, the umbrella sampling technique was employed to study the permeation mechanism of cisplatin into a lipid membrane. As stated above, in umbrella sampling, a reaction coordinate  $\xi$  is defined *a priori*, so that the objective is to compute the potential of mean force (PMF, which corresponds to the free energy function  $A(\xi)$  defined below) along the reaction coordinate. Let  $Z$  be the canonical partition function, which contains all the necessary information to compute the thermodynamic properties of the system,

$$Z = \int \exp[-\beta V(r)] d^{3N} r \quad (2.176)$$

where  $\beta = (1/(k_B T))$ ,  $V(r)$  is the potential energy and the integration is done over the entire coordinate space. The Helmholtz free energy is, by definition,

$$A = -\frac{1}{\beta} \ln Q \quad (2.177)$$

now, to study the energetics of the chemical process under study along the reaction coordinate  $\xi$ , it is required to compute the probability distribution of the system along  $\xi$ , defined by,

$$P(\xi) = \frac{\int \delta[\xi(r) - \xi] \exp[-\beta V(r)] d^{3N} r}{Z} \quad (2.178)$$

where the integration is done over all degrees of freedom but  $\xi$ . At this point, the PMF is defined as the free energy along the reaction coordinate,

---


$$A(\xi) = -\frac{1}{\beta} \ln P(\xi) \quad (2.179)$$

Therefore, in order to obtain the free energy along the reaction coordinate  $\xi$ , the reaction pathway would need to be thoroughly sampled to obtain  $P(\xi)$ . In practice, the reaction coordinate is subdivided in different windows (Figure 2.1a), and a bias potential  $w_i(\xi)$  is included in the potential term for each of these windows,

$$V^b(r) = V^u(r) + w_i(\xi) \quad (2.180)$$

where the  $b$  superscript refers to biased quantities and the  $u$  superscript refers to unbiased quantities. The main role of the bias potential consists of restraining the system to a neighborhood of  $\xi_i$  for some window  $i$ , so that it samples exhaustively the remaining degrees of freedom within the window of interest.

At this point, a MD simulation is performed on each of the windows using the biased potential  $V^b(r)$ . Thus, one obtains the probability distribution of the reaction coordinate  $\xi$  for each window  $i$ . It should be emphasize that the distributions obtained are the *biased* distributions  $P_i^b(\xi)$  associated with the biased potential  $V^b(r)$ ; however, in order to compute the PMF using Equation 2.179, the *unbiased* potential  $P_i^u(\xi)$  associated with  $E(r)$  (Equation 2.178) is required. It can be shown<sup>45</sup> that  $P_i^u(\xi)$  can be obtained from  $P_i^b(\xi)$  by

$$P_i^u(\xi) = P_i^b(\xi) \exp[\beta w_i(\xi)] \langle \exp[-\beta w_i(\xi)] \rangle \quad (2.181)$$

where  $\langle \exp[-\beta w_i(\xi)] \rangle$  indicates an ensemble average. The PMF associated with the  $i^{th}$  window is then given by,

$$A_i(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - w_i(\xi) + F_i \quad (2.182)$$

where  $F_i = \ln(\langle \exp[-\beta w_i(\xi)] \rangle)$ . If only a single window spanned the entire reaction coordinate, then the PMF would be given by Equation 2.182, so that  $F_i$  would consist of an arbitrary additive constant. However, when different  $A_i(\xi)$ s are used to compute the total PMF, the  $F_i$ s would need to be computed. However, they cannot be directly

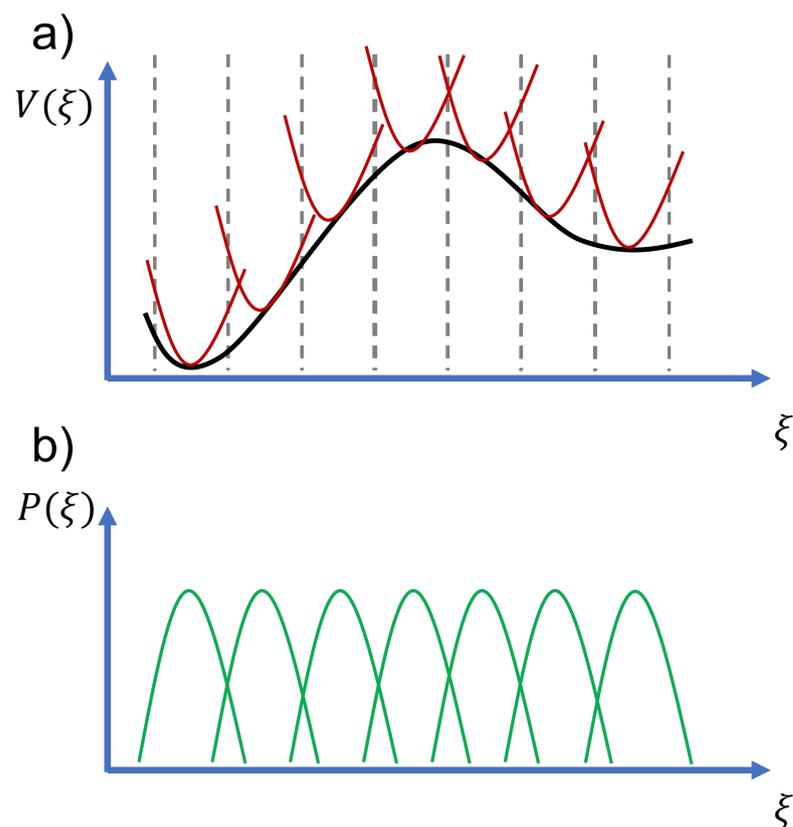


Figure 2.1: a) Schematic representation of a free energy profile along the reaction coordinate  $\xi$ . The dashed vertical bars indicate the subdivision of the reaction coordinate into different windows, and the red parabolas indicate the bias potential applied on each window. b) Probability distributions of the reaction coordinate  $\xi$  obtained from the (biased) sampling performed on each window.

---

computed from the sampled distributions, since their dependence on the ensemble average  $\langle \exp[-\beta w_i(\xi)] \rangle$  implies that the unbiased probability distributions  $P_i^u(\xi)$  would need to be known *a priori* (Equation 2.178). There are several methods in the literature to obtain estimates for  $F_i$ ,<sup>155,156</sup> of which one of the most popular being the weighted histogram analysis approach (WHAM).<sup>157,158</sup>

## 2.5 Multiscale Hybrid Methods: Quantum Mechanics/Molecular Mechanics

The previous sections have evidenced the application of classical mechanics methodologies to study the time evolution of a molecular system (Equation 2.168) and the interactions that occur among the atoms of the system under study (*e.g.*, Equation 2.175). A hybrid quantum/classical framework has also been explored in section 2.4.4, namely AIMD, in which, basen on the BO approximation, the motion of the nuclei is treated classically (again Equation 2.168), whereas the electrons are described quantum mechanically. In this framework, it is possible to describe bond breaking and formation events, which would not be possible to account for with a harmonic analytic potential like that of Equation 2.175, without the need to solve the time-dependent Schrödinger equation. However, the applicability of AIMD is limited to at most a few hundreds of atoms and short timescales (hundreds of picoseconds) due to the computational cost of the quantum mechanical computations, so that it would not be feasible to apply AIMD on systems modeling biological media, consisting of several thousands – to a few millions – of atoms and relatively long times (tens of microseconds). This problem would be alleviated if only a portion of the system were to be treated quantum mechanically: this is precisely the approach adopted by the multiscale hybrid QM/MM methods in which, in the most elementary case, the system is partitioned into two subsystems: an inner region ( $I$ , Figure 2.2), which usually corresponds to the region of chemical interest and is described quantum mechanically, and an outer region ( $O$ ), which is described using a classical potential (*e.g.*, a force field).

The QM/MM methodology can be subdivided into different partition schemes, depending on how the interaction between the inner region and the outer region is accounted

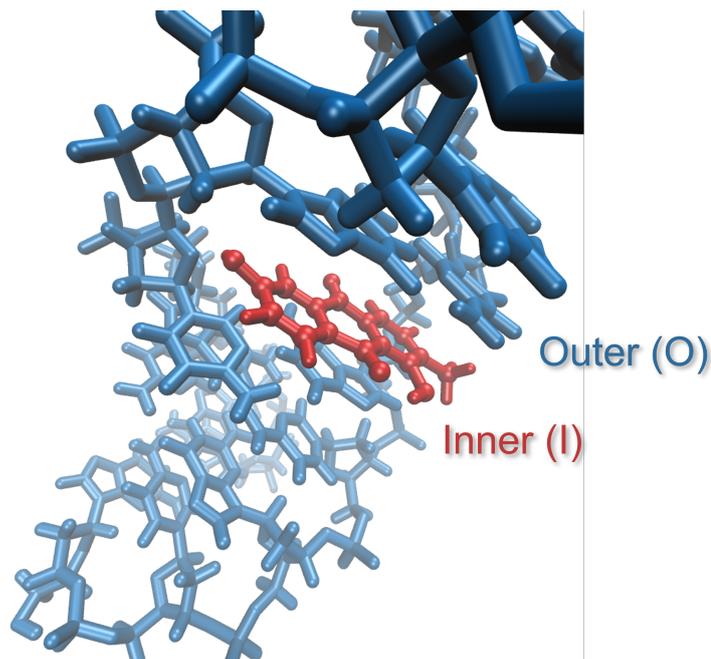


Figure 2.2: Representation of the QM/MM partitioning scheme of a molecular system consisting of an anthraquinone molecule (inner region, red) intercalated in a double strand of DNA (outer region, blue).

for.<sup>47,48,159</sup> In the subtractive scheme, the total energy is determined as the difference between the energy of the inner region computed quantum mechanically ( $E^{QM}(I)$ ), plus the energy of the entire system computed at a molecular mechanical level of theory ( $E^{MM}(I + O)$ ) minus the energy of the inner region, again computed using a MM method ( $E^{MM}(I)$ ),

$$E = E^{QM}(I) + E^{MM}(I + O) - E^{MM}(I) \quad (2.183)$$

where the last term is included to avoid counting twice the energy of the inner subsystem. The main problem of this subtractive scheme consists of the fact that the QM calculation of the inner region is not influenced in any manner by the presence of the atoms belonging to the outer region, so in practice this subtractive scheme consists of including an additive term to the QM energy that accounts for the interactions between the inner and the outer region in a classical manner *via* the  $E^{MM}(I + O)$  term in Equation 2.183. A substantial improvement to the subtractive approach can be made by explicitly including the electrostatic potential generated by the point charges in the outer region inside the

---

Hamiltonian of the quantum mechanical inner region; thus, an effective Hamiltonian is used in the QM computation involving the inner region,

$$H^{eff}\Psi_I = (H_I^0 + H_{I/O})\Psi_I = E\Psi_I \quad (2.184)$$

where  $H_I^0$  is the Hamiltonian of the (unperturbed) inner region, and  $H_{I/O}$  is a perturbation term that describes the electrostatic interactions between the outer region point charges and the electrostatic potential of the inner region,

$$H_{I/O} = \sum_{i=1}^N \sum_{J=1}^M \frac{q_J}{|r_i - R_J|} \quad (2.185)$$

where  $q_J$  is the point charge centered at the  $J^{th}$  outer atom, in  $R_J$ , and  $r_i$  indicates the position of the  $i^{th}$  inner electron. This approach explicitly considers the polarization of the QM inner region due to the presence of the atoms in the outer region. The energy is then computed as follows:

$$E = E_{q(O)}^{QM}(I) + E^{MM}(I + O) - E_{q(O)}^{MM}(I) \quad (2.186)$$

where the term  $E_{q(O)}^{QM}(I)$  is computed by solving Equation 2.185, and the subscript  $q(O)$  indicates that the point charges of the outer region are being included in the computation of the inner region energy term. The term  $E_{q(O)}^{MM}(I)$  is subtracted so that the electrostatic interactions (at a MM level of theory) between the inner and the outer region are accounted for only once.

Another scheme for computing the total energy is the additive scheme, in which the energy is given by the sum of the energies of the inner and the outer regions, computed at the QM and MM level of theory, respectively, plus an additive term that accounts for the interaction of the two regions ( $E^{QM/MM}(I/O)$ ),

$$E = E^{QM}(I) + E^{MM}(O) + E^{QM/MM}(I/O) \quad (2.187)$$

where the coupling term is given by

$$E^{QM/MM}(I/O) = E_{bonded}^{MM}(I/O) + E_{vdW}^{MM}(I/O) + E_{el}(I/O) \quad (2.188)$$

---

where the first two terms on the rhs describe the bonded and the van der Waals interactions between outer and inner region atoms. These terms are described using a classical MM force field. At the same time, additive schemes can be subdivided depending on the methodology used to compute the term  $E^{QM/MM}(I/O)$  in Equation 2.188. The mechanical embedding approach is perhaps the simplest one, in which the electrostatic term is described as a classical Coulomb term. Some of the main drawbacks of mechanical embedding include the fact that the inner region is not polarized by the outer region, and vice versa, the inner region QM charge density does not interact with the outer region atoms. A more sophisticated approach is the electrostatic embedding scheme, in which the polarization of the QM inner region due to the presence of the outer region charges is accounted for by employing an effective Hamiltonian analogous to the one in Equation 2.185. This is the most widely used scheme to study chemical processes in complex media, as effectively no extra computational cost is present, other than the cost of the QM part of the system itself. This is the approach that has been adopted throughout the present thesis work. It should be emphasized that also in this case the polarization of the outer region due to the inner region is not accounted for in the electrostatic embedding scheme.

The mutual polarization between the outer and the inner regions is accounted for by the more sophisticated polarizable embedding scheme,<sup>160–162</sup> in which this mutual polarization is computed in a self-consistent manner. In that case, an effective Hamiltonian is used as in Equation 2.184 to compute the electronic wavefunction of the inner region, but this time not only an electrostatic term is included in  $H_{I/O}$  (as in Equation 2.185), but an extra term is also included to account for the polarization of the outer region charges due to the electric field generated by the charge density of the inner region,

$$\begin{aligned}
 H_{I/O} &= H_{I/O}^{elec} + H_{I/O}^{pol} \\
 &= \sum_{J=1}^M q_J^0(R_J) V^I(R_J) - \sum_{J=1}^M \mu_J^0(R_J) E^I(R_J)
 \end{aligned}
 \tag{2.189}$$

where both sums run over the set of  $M$  outer region point charges,  $V^I(R_J)$  is the electrostatic potential generated by the inner region charge density at  $R_J$ , and  $E^I(R_J)$  is the electric field generated by the same charge density at  $R_J$ .  $\mu_J^0(R_J)$  is the classical electrostatic dipole centered at  $R_J$ . In Equation 2.189, the classical dipoles  $\mu_J^0(R_J)$  in the

---

outer will react to changes in the charge density of the inner region, so that it will be the term accounting for the polarization of the outer region. Some methodologies to describe this polarization effect include using induced point dipoles,<sup>163</sup> fluctuating charges<sup>164</sup> and Drude oscillators.<sup>161</sup> For example, in the case of induced point dipoles, the atomic dipole moments  $\mu_J^0(R_J)$  of the outer region are expressed in terms of the atomic polarizabilities  $\alpha_J$  and the total electric field, which is the sum of the electric field due to the inner region ( $E^I(R_J)$ ) plus the electric field of the outer region,  $E^0(q_J^0, \mu_J^0)$ :

$$\mu_i^0(R_J) = \alpha_J[E^I(R_J) + E^0(q_J^0, \mu_J^0)] \quad (2.190)$$

It should be emphasized that the electric field of the classical region depends on the dipoles  $\mu_i$ , which in turn depend on the electric field: this is the reason why Equations 2.189 and 2.190 need to be solved in a self-consistent manner.

## Chapter 3

# MoBioTools: A Toolkit to Setup QM/MM Calculations

Gustavo Cárdenas<sup>1</sup>, Jesús Lucia-Tamudo<sup>1</sup>, Henar Mateo-delaFuente<sup>1</sup>, Vito F. Palmisano<sup>1</sup>, Nuria Anguita-Ortiz<sup>1</sup>, Lorena Ruano<sup>1</sup>, Álvaro Pérez-Barcia<sup>4</sup>, Sergio Díaz-Tendero<sup>1,2</sup>, Marcos Mandado<sup>4</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

3 – Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, 28049 Madrid, Spain

4 – Department of Physical Chemistry, University of Vigo, Lagoas-Marcosende s/n, ES-36310-Vigo, Galicia, Spain

This Chapter has been submitted for publication. It has been published in pre-print form in *ChemRxiv*, 2022.

*In this Chapter a toolkit that allows for the preparation of QM/MM input files from a conformational ensemble of molecular geometries is presented. This toolkit is extensively used throughout the thesis to set up the QM/MM computations. It can be used in command line, so that no programming experience is required, although it presents some features that*

---

*can also be employed as a python application programming interface. In the present Chapter, the toolkit is applied in four situations in which different electronic-structure properties of organic molecules in the presence of a solvent or a complex biological environment are computed: the reduction potential of the nucleobases in acetonitrile, an energy decomposition analysis of tyrosine interacting with water, the absorption spectrum of an azobenzene derivative integrated into a voltage-gated ion channel, and the absorption and emission spectra of the luciferine/luciferase complex. These examples show that the toolkit can be employed in a manifold of situations for both the electronic ground state and electronically excited states. It also allows for the automatic correction of the active space in the case of CASSCF calculations on an ensemble of geometries, as it is shown for the azobenzene derivative photoswitch case.*

### **3.1 Introduction**

Multiscale hybrid quantum/classical methods have gradually gained popularity since their introduction,<sup>126,165</sup> and are nowadays used routinely for the description of chemical processes that involve complex media.<sup>48,159,166–168</sup> These methods consider the partitioning of the system under study into two (or more) fragments: a quantum mechanical (QM) subsystem, which is the region of chemical interest, and a region described classically, which interacts with the QM subsystem. These multiscale approaches have been classified into two different categories in a recent review<sup>47</sup> depending on the description of the classical subsystem. In the first category, the atoms of the classical region are explicitly represented as point-wise particles in a classical potential, for example, a molecular mechanics (MM) analytic potential, that polarizes the quantum mechanical subsystem. The second category uses an implicit continuum description,<sup>169–172</sup> where the classical region is represented as a mean potential that polarizes and is polarized by the quantum mechanical subsystem.

The classification of the paradigms described above can be further branched into different methodologies, depending on the way in which the interaction between the two subsystems is established.<sup>46</sup> For example, within the quantum QM/MM framework, one of the most widely used schemes is the electrostatic embedding, in which fixed-point charges of the classical atoms polarize the QM subsystem by entering the one electron

---

part of the Hamiltonian as an electrostatic perturbation. This approach accounts for the polarization of the QM region due to the classical point charges, but does not consider the polarization of the classical region by the QM one; however, it has been widely employed in a manifold of situations, such as the study of catalytic reactions,<sup>166,173–179</sup> thermodynamic quantities<sup>55–59</sup> and excited state properties<sup>41,54,60–67,70,73</sup> in biologically and technologically relevant systems. An embedding scheme which accounts for the mutual polarization of the QM and the MM subsystems is the polarizable embedding approach,<sup>168</sup> whereby the classical region is described in terms of fixed point charges plus higher order multipoles. For example, the polarization of the environment can be accounted for by assigning polarizability tensors to the polarizable sites that give rise to an induced charge distribution in the environment. This methodology has been employed in the study of ground state<sup>161,162</sup> and excited state<sup>160</sup> properties of complex biological systems. Its main drawback is the computational cost that arises as self consistency needs to be attained simultaneously for the wavefunction of the QM region and the polarizabilities of the MM region.

An issue that needs to be addressed in the study of complex biological media is that of conformational sampling. Indeed, when the system under study consists of several thousands of atoms, the potential energy landscape often presents several minima that can be populated at the temperature of interest, and conformations corresponding to these minima can contribute to a great extent to the property under study.<sup>48</sup> Thus, this situation is most accurately represented as a statistical distribution arising from the individual conformations, rather than a value due to a specific conformation. In this regard, the application of sampling methods, such as MD,<sup>44</sup> Monte Carlo (MC),<sup>180</sup> and Wigner sampling,<sup>181</sup> is crucial to obtain an accurate description of the system. In this Chapter, the main focus will regard sampling by means of MD. However, whenever generic ensemble of geometries is referred to, and unless otherwise stated, it is implicit that it may be generated by any of the sampling methodologies mentioned above.

The potential energy or the potential energy gradient of the system (or part of it) along the sampling process can be computed by a QM method, for example, in *ab initio* MD or QM/MM MD simulations.<sup>125</sup> However, in most cases, using a high level QM potential to perform the conformational sampling is unfeasible. In these situations, a less sophisticated

---

potential is used to perform the exploration of the potential energy surface, for example, a cheap QM/MM MD approach or MD using analytic force fields.<sup>48</sup> Then, once the sampling is completed, an ensemble of geometries is fetched from the trajectory obtained, and single point QM or QM/MM computations - using a high level QM method of interest - are performed on top of each of these geometries.<sup>62,73</sup> The setup of the hundreds, or even thousands, of single point QM/MM calculations that need to be performed to obtain converged values of the property of interest is an arduous task that has to be automatized.

Nowadays, the computational chemistry community can exploit a manifold of quantum chemistry software that include the electrostatic embedding scheme to perform QM/MM calculations on the one hand, and software aimed at performing sampling at different levels of theory on the other hand, in some cases interfaced with the QM software. There are also packages that act as interfaces between the QM programs and the MD or MC programs, which are either distributed as application programming interfaces (APIs),<sup>182-186</sup> or as executables that require little to no programming experience and solely depend on input files provided by the user. Some examples of the latter case are the Cobramm<sup>187</sup> software, whose main purpose is to interface the Gaussian<sup>188</sup> program and several other QM software with Amber, and the SHARC<sup>189,190</sup> and Newton-X<sup>191</sup> packages which act also as interfaces between QM and MD software but have the main purpose of performing nonadiabatic MD simulations. Each of the paradigms presented above bears several advantages. On the one hand, the usage of APIs in most cases provides a high degree of versatility in the setting up of the QM/MM calculations, but its main drawback is that some degree of programming experience is required. On the other hand, the executables mentioned above are either usable for very specific tasks or require some degree of training before the user gets fully acquainted with them.

Herein, MoBioTools is presented, a simple package that allows for setting up electrostatic embedding QM/MM calculations in an automatic and simple fashion from an ensemble of geometries. Although it is by no means intended to substitute the interfaces cited above, it aims at presenting an interface that requires only familiarity of the user with a specific quantum mechanical software but, at the same time, attempts to be as general as possible in the input. In this way, specialized tasks, for example, the computation of the interaction energy between two monomers, can be performed with all the QM software MoBioTools

---

is interfaced with, considering as less variation of the input format as possible for the different QM interfaces. Currently the package requires a trajectory and a topology file in an Amber<sup>192</sup> compatible format, and has the possibility to generate QM/MM input files for Gaussian (09 and 16), Orca ( $\geq 4.0$ ),<sup>193</sup> NWChem<sup>194</sup> and (Open)Molcas.<sup>195,196</sup> However, the interface is under constant development and further compatibilities with additional QM software are yet to come. The main feature of the package is the generation of the input files for the QM/MM computations, and not the computations themselves. Therefore, in practice, neither of the QM software are required to execute MoBioTools - the only exception is for (Open)Molcas in the case that the active space of a CASSCF calculation<sup>81,83</sup> needs to be corrected on-the-fly using an algorithm proposed by some of us involving the computation of the overlap matrix between orbital spaces (Chapter 4).<sup>84</sup> In the following, the main structure of MoBioTools is explained and the versatility of the algorithm is shown by applying it to four different situations: the computation of the reduction potential of the nucleobases in acetonitrile, the interaction energy between tyrosine and water and its energy decomposition analysis, the absorption spectrum of an azobenzene derivative integrated into a voltage-gated ion channel, and the emission spectrum of the luciferine/luciferase complex.

## 3.2 Methodology

As stated above, the main purpose of the MoBioTools package is to extract an ensemble of geometries that have been previously sampled by means of MD, MC or any other sampling approach, and automatically generate a QM/MM input file for each of these geometries in accordance with the QM software and methodology requested by the user. Figure 3.1 shows a schematic representation of the geometry extraction process and the applicability of the QM/MM single point computations, which are performed on top of the selected snapshots to obtain the desired properties. Some examples of these properties include absorption spectra, interaction energies, free energies, among others (Figure 3.1b-d). In the case in which the property of interest needs to be computed by means of CASSCF (or a related wavefunction method),<sup>81-83</sup> the toolkit performs the automatic correction of the active space for those geometries for which the active space differs from that of a reference

structure (Figure 3.1e).

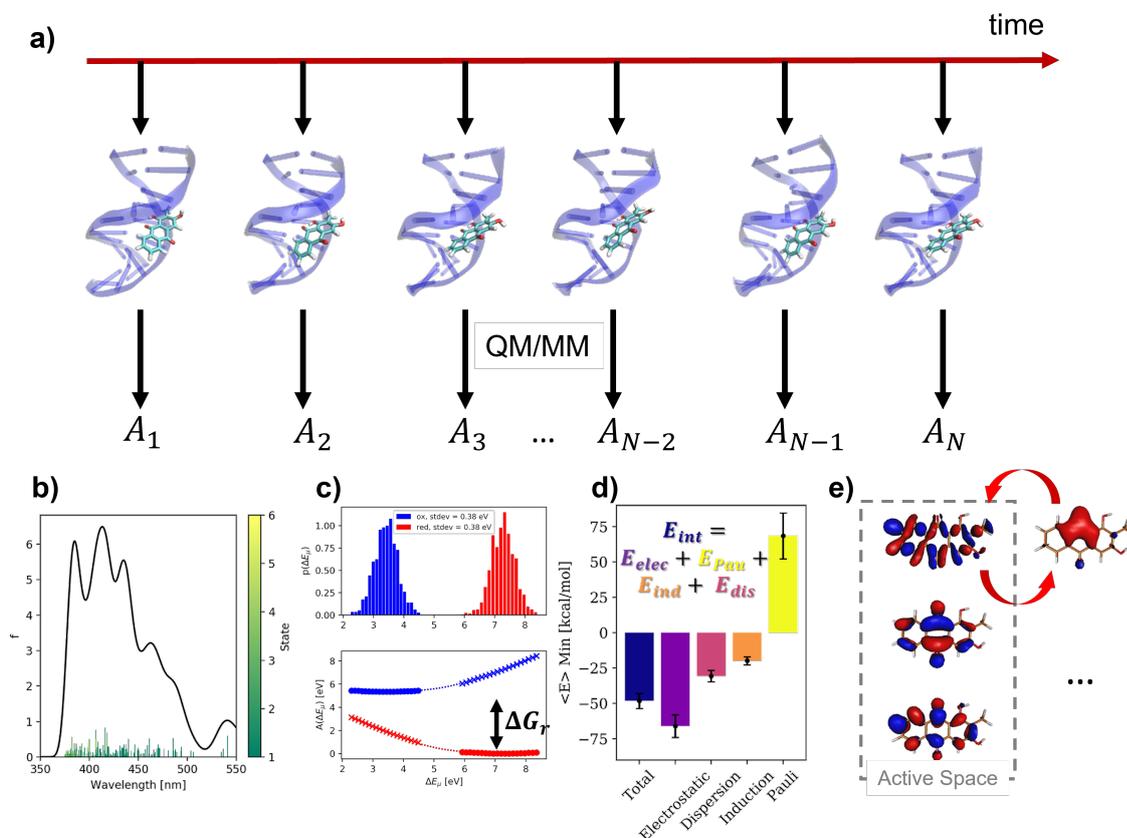


Figure 3.1: Schematic representation of the working principle of the MoBioTools toolkit and its applicability. a) The extraction of an ensemble of geometries (snapshots) from a trajectory obtained by means of MD, MC or other technique. The QM/MM single point calculations performed using the input files generated provide a distribution of the property of interest  $A_N$ , for example, b) the absorption spectrum, c) the reduction free energy, or d) an energy decomposition analysis of the interaction energy between two molecules of a complex. e) Representation of the automatic correction of the CASSCF active space for a specific sampled geometry.

The MoBioTools package is subdivided into two main drivers depending on the task to be performed: a general purpose QM/MM input generator (`main.qminputs.py`), which is executed in command line and requires two input files; and a program to carry out CASSCF computations with an on-the-fly correction of the active space (`pyoverlaps.py`), which works in command line. Although both drivers in principle work similarly, as in both cases a set of QM/MM inputs is generated from an ensemble of geometries, their functionalities and the QM programs they are interfaced with differ considerably, as will

---

be explored in the following sections.

### 3.2.1 The General Purpose QM/MM Input Generator

The script `main.qminputs.py` is the driver that currently provides an interface to generate a sequence of input files for a set of geometries provided as input in the form of an Amber topology/parameters and an Amber trajectory files. The script works in command line and needs to be provided with two input files: a general input file, which does not depend on the QM software to be used and contains information about the set of geometries to be fetched and settings on how to perform the QM/MM partitioning, and a template file for the QM calculations to be performed. Figure 3.2 shows a schematic representation of the options featured by each of these files. Both files are subdivided in sections, introduced by an ampersand symbol (&) and enclosed by an `&end` statement.

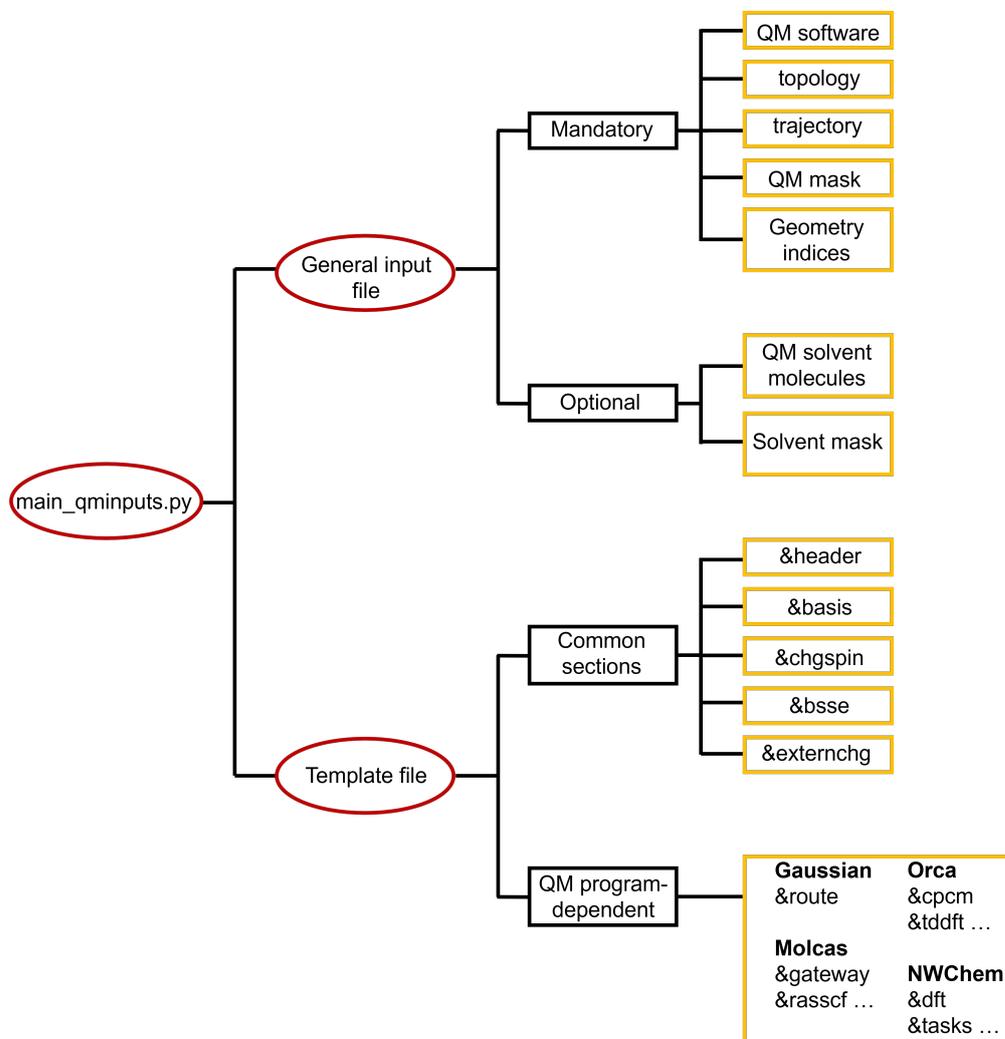


Figure 3.2: Schematic representation of the contents of the two files provided as input to the main\_qminputs.py driver. Red ovals represent files and yellow rectangles represent input sections – introduced by an & symbol – and/or the options these sections enclose. See the text for a detailed description on each of these options.

---

The general input file (hereby and in listing 3.1 referred to as `main.inp`) consists of a single section (`&main...&end`), in which up to seven options can be provided as input by the user on how to treat the ensemble of geometries at hand to generate the input files. In particular, five of these options are mandatory for the software to successfully generate the QM/MM input files: the QM software name (`tpl`), the trajectory (`traj`) and the topology (`top`) file relative to the ensemble of geometries at hand, the definition of the QM region (`qmmask`) and the geometry indices of the geometries for which to generate the input files. The options `solvmask` and `closest` are not mandatory, and refer to the mask of atoms or residues to be treated as solvent molecules, and the number of solvent molecules closest to the QM region, which also need to be treated quantum mechanically, respectively. This is useful whenever a part of the environment that surrounds the quantum mechanical part of the system needs to be included in the QM region, but the indices of the molecules of interest are unknown to the user or they change along the sampled geometries. Here, by mask it is intended the atom or molecule selection, as referred to in the Amber terminology.<sup>192</sup> For the selection mask, the Amber syntax is used so that, for example, `@` refers to atom selection or a colon refers to residue selection. This syntax is used in the general input file as well as in the template file (see listings 3.1 and 2). The discussion above may seem to imply that the system is treated as a solute-solvent system, where the solute is represented by the `qmmask` entry and the solvent is (if requested) introduced by means of the `solvmask` and the `closest` keywords. It should be emphasized that this is purely a notational convention and it by no means implies that only systems in solution can be studied. This is the notation employed by the `cpptraj` and the `pytraj` software,<sup>197</sup> whose machinery is employed by MoBioTools to perform operations on the trajectory objects. However, complex systems containing, for example, proteins and lipid membranes – and not only solvent – can be treated by MoBioTools, as will be evidenced by some of the examples below. If the `solvmask` and the `closest` keywords are not included, a QM/MM calculation (or even a gas phase QM calculation) can be set up for the geometries of interest, where only the `qmmask` entry is treated quantum mechanically.

The template file is also subdivided in sections, some of which are common to all the QM software interfaced with MoBioTools (`header`, `basis`, `chgspin`, `bsse` and `externchg`), whereas some others are exclusive to the QM software under consideration (see Figure

---

3.2 for reference). An example of the template for setting up an ensemble of (QM/MM) calculations using the Gaussian software is provided in listing 3.2. In this example, three common sections are featured: the `head` section, which corresponds to the `Link0` commands in Gaussian (general instructions); the `chgspin` section, which features the charge and the spin multiplicity of the system; and the `externchg` section, which does not take arguments and essentially informs the script that a QM/MM input is requested, so that the point charges that surround the QM region (defined by the `qmmask` keyword in the main input file) also need to be included in the input file. The `route` section is unique to the Gaussian software, and includes the route commands that define the methodology, the basis set and the type of calculation to be performed, among others. The execution of the script `main_qminputs.py` with the files `main.inp` and `template.inp` (listings 3.1 and 3.2, respectively) as arguments will generate a gaussian QM/MM input file for the geometry having index 3 in the `GUA_O2.inpcrd` trajectory file.

The example above describes a simple task of extracting a geometry from a coordinate file and generating a QM/MM input file with the desired methodology and settings. Its usefulness stems from the fact that the same operation can be performed for any of the QM software MoBioTools is interfaced with, and from the ease it is scaled to an ensemble of geometries. The keywords `basis` and `bsse` of the template file provide a further degree of flexibility: within the section `basis` the user can employ customized basis sets (for example, different basis sets for different atoms), and the `bsse` keyword provides a way to subdivide the system into two subsystems to request either a fragment calculation or the computation of a property that involves the two subsystems, such as the interaction energy, for which the basis set superposition error (BSSE)<sup>119,120</sup> needs to be corrected. This procedure has been extensively used for generating input files in the past,<sup>86</sup> and in the following section this approach will be applied to a system of chemical interest.

Listing 3.1: Example of the general input file to generate a set of input files from an ensemble of geometries. The sections in square brackets refer to optional features. In this example, it is requested to generate an input file for the geometry having index 3 in the `GUA_O2.inpcrd` coordinate file. The `qmmask` “:1,2” indicates that molecules 1 and 2 are

---

to be considered as QM molecules.

```
# Name file : main.inp
&main
  tpl      = gaussian
  traj     = GUA_O2.inpcrd
  top      = GUA_O2.prmtop
  qmmask   = :1,2
  geoms    = 3 [start, [stop, step]]
  [solvmask = :WAT]
  [closest = 5]
&end
```

Listing 3.2: Example of a template file for a Gaussian QM/MM calculation. The common sections head, chgspin and externchg, are included, as well as the Gaussian specific route section.

```
# Name file : template.inp
&head
%NprocShared=4
%mem=4GB
&end

&route
#p M062X/cc-pVTZ Charge
&end

&chgspin
0,1
&end

&externchg
&end
```

---

### 3.2.2 The Active Space Automatic Preservation

When CASSCF-based calculations are performed for an ensemble of geometries, it is desirable that the computations of all the geometries include similar molecular orbitals within the active space.<sup>84</sup> If this is the case, the same reference wavefunction is employed along the ensemble of geometries and the different computations can be directly compared or convoluted to obtain, for example, the absorption or emission spectra. The script `pyoverlaps.py` is the driver of the software that allows for the automatic correction of the active space in a CASSCF computation on an ensemble of geometries by comparing, as will be explained below, the orbitals of the active space of the sampled geometries with the orbitals of the active space of a reference geometry. It performs a similar task to the `main_qminputs.py` of the previous section, in that it also generates a set of QM/MM input files from an ensemble of geometries. However, it also presents several differences with respect to the general input generator, perhaps the most important one being the fact that the script `pyoverlaps.py` does not only generate the input files, but it also controls the execution of the QM software, which in this case is the OpenMolcas program. The script works exclusively in command line interface, and requires the presence of four files in the working directory: the trajectory and the topology/parameters files of the ensemble of geometries under study, a molden<sup>198</sup> file bearing the molecular orbitals used as reference to preserve the active space on the single point CASSCF calculations throughout the selected geometries and a template file for the Molcas inputs. An example of the execution of the script is given by the command `pyoverlaps.py -h`, which prints a list of arguments it can be provided with, the most important of which are shown in listing 3.3. Apart from the four files that need to be present in the working directory, the user needs to provide the indices of the molecular orbitals (in the reference geometry) that need to be considered as the reference active space (`-ref` option), the mask of the QM atoms (`-qm`), and the index of the geometry for which to perform the CASSCF calculations and the active space correction, if needed. The execution is related to a single geometry, however, it can (and should) be easily executed in parallel.

---

Listing 3.3: Mandatory arguments for the execution of the `pyoverlaps.py` script.

```
-p TOP           Topology file
-c CRD           Trajectory file
-r REFERENCE     Reference molden file
-tpl TEMPLATE    Molcas template file
-rng RNG [RNG ...] Range of MOs in the active space
                  (e.g: 30 43)
-qm QMMASK       QM mask
-ig IGEOM        Specific frame for which to generate
                  an input file. Default = 0
```

The driver `pyoverlaps.py` presents the implementation of an algorithm<sup>84</sup> proposed by some of the authors that automatically corrects the active space of a specific geometry. This is achieved by computing and analyzing the overlap matrix ( $S_{MO}$ ) between the set of CASSCF molecular orbitals (MOs) of the reference geometry and that of the geometry of interest. The algorithm and its implementation are described in full detail in Chapter 4. In this section only the main features are briefly summarized. The script `pyoverlaps.py` performs a CASSCF calculation on the sampled geometry, after which it computes the overlap matrix  $S_{MO}$  and analyzes its column maxima. The column labels of  $S_{MO}$  represent the MOs of the sampled geometry, whereas the row labels represent the MOs of the reference geometry. In this framework, it is assumed that there exists a bijection between the set of reference MOs and the set of MOs of the sampled geometry (sampled MOs for simplicity). Thus, the maximum value of each column represents the reference orbital that is the most similar to the MO of the sampled geometry for that column. If there is a sampled MO outside of the active space, whose maximum value in  $S_{MO}$  coincides with a MO in the active space of the reference geometry, that MO should be included in the active space of the sampled geometry (see the matrix element  $S_{MO}[4][5']$  on Figure 3.3). If, on the other hand, the maximum value of a sampled MO that belongs to the active space coincides with a reference MO outside the active space, that sampled MO should be removed from the active space. The correction occurs when at least two MOs are each in one of the two situations above, in which case the two MOs are swapped, thus,

creating a new guess wavefunction, and a new CASSCF optimization is performed. This procedure is carried out iteratively by the program until the MOs in the active space of the sampled geometry are in a one-to-one correspondence with the MOs in the active space of the reference geometry. Figure 3.3 shows a schematic representation of the situation in which the positions of two MOs need to be swapped to correct the active space of sampled geometry 3.

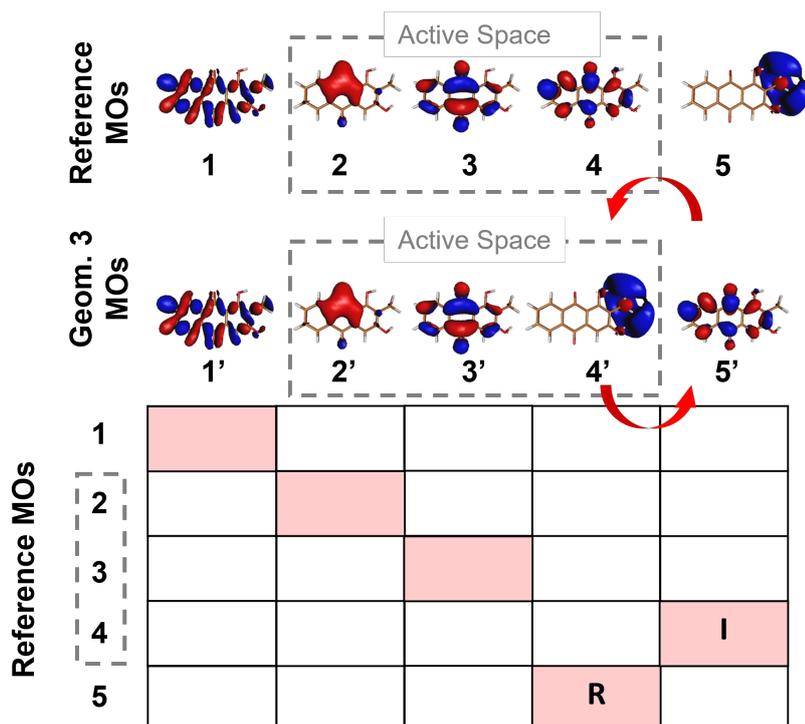


Figure 3.3: Schematic representation of the working principle of the algorithm to correct the active space on an ensemble of sampled geometries.<sup>84</sup> Top: the set of MOs of the reference geometry and the MOs of a sampled geometry (Geom. 3). Bottom: the overlap matrix  $S_{MO}$  between the two MO orbital sets. The column maxima are highlighted in pink. In this example, the maximum of column 4' (that is, MO 4' of geometry 3) coincides with MO 5 from the reference set of MOs, so that it has to be removed from the active space (hence the R label). On the other hand, the maximum of column 5' (MO 5' of geometry 3) coincides with reference MO 4, which is inside the reference active space, and it should be included in the active space (hence the I label). Thus, MOs 4' and 5' need to be swapped to correct the active space of geometry 3.

Although the execution of `pyoverlaps.py` occurs in the command line, with little to no interaction of the user with the code, the script itself relies on a framework that can also be exploited by the user to some extent as an API as well. The general structure

---

and dependencies of the `pyoverlaps.py` script are shown in Figure 3.4. In synthesis, the driver carries out two main tasks: on the one hand, it is responsible for most of the input and output operations, as well as the operations that regard the reading and handling of the trajectory (and topology files), including the definition of the QM and MM regions. On the other hand, and as explained above, it bears the implementation of the active space correction algorithm. However, the machinery for reading and operating with the atomic orbital (AO) and MO information of the two geometries, whose orbital spaces are to be compared, is present in the `parse_molden.py` module. Its most important feature is the class `Mol()`, from which an object is created by providing it with a molden file as an argument. During the execution of `pyoverlaps.py`, two `Mol()` objects are created at each iteration: one for the reference geometry and one for the geometry under study, but no trace of these objects remains at the end of the execution. If the user is interested in analyzing the data of each molden file or computing the  $S_{MO}$  matrix without executing the `pyoverlaps.py`, the `parse_molden.py` module can be executed interactively on a python shell; this will create the two `Mol()` objects and perform the computation of the atomic orbital overlap matrix  $S_{AO}$  and the  $S_{MO}$  matrix. Alternatively, the user can import the `Mol()` class and compute these matrices by their own. An important feature is that the `Mol()` class is similar (although less sophisticated) to the `pyscf.gto.M` class from the PySCF<sup>199</sup> software, in particular it bears the attributes `_bas`, `_env` and `_atm` that can be employed to use the machinery of the PySCF to compute molecular integrals for a given molecular species.<sup>199,200</sup>

Figure 3.4 shows the dependencies of the `parse_molden.py` script/API. These dependencies are modules that provide a lower level framework for the treatment of a `Mol()` object and for the computation of the  $S_{AO}$  and  $S_{MO}$  matrices. The `Align.py` module aligns the sampled geometry with the reference one prior to the computation of the overlap matrices. The `permutations.py` module allows the reordering of the AO basis functions between the PySCF ordering and that from Molcas (and other QM software such as Gaussian or Orca), and the `ovlp_wrapper.py` module is a wrapper for the C++ implementation of the AO overlap integrals, present in the library `intwrap.so`. These AO integrals have been implemented formally using the Obara-Saika recursion relations<sup>201</sup> in an iterative fashion, a procedure which is routinely performed in the literature.<sup>200,202</sup> The reason for

implementing the AO integrals stems from the fact that, to our knowledge, there are no APIs that allow for the computation of AO overlap integrals between two different atom centered basis sets, a crucial feature to compare the MO sets of two different geometries of the same molecular species, as in principle the atom centers will not be equal for both structures.

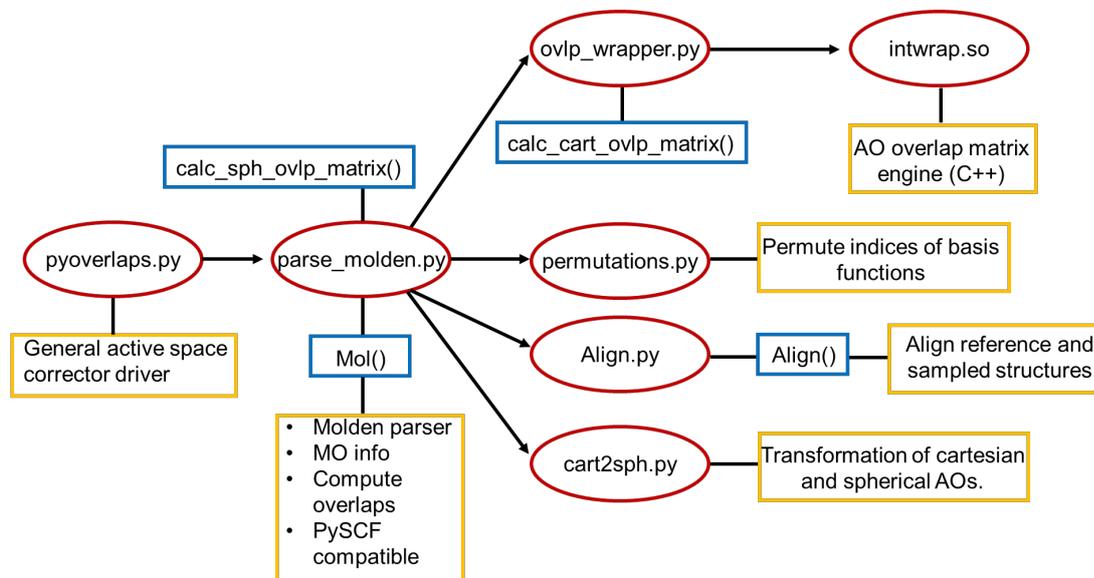


Figure 3.4: Schematic representation of the structure and main dependencies of the `pyoverlaps.py` script for the automatic correction of the active space on an ensemble of geometries. The red ovals represent a script or a program, the blue rectangles represent some functions and classes that can be used as an API, and the yellow boxes provide a brief description of the function of each script.

---

### 3.3 Applications

In this section, the application of the MoBioTools toolkit to set up the QM/MM input files for four different systems is illustrated. Specifically, the toolkit was employed to compute the reduction potential of the five nucleobases in acetonitrile, the interaction energy between tyrosine and water and its energy decomposition analysis, the absorption spectrum of an azobenzene derivative integrated into an voltage-gated ion channel, and the emission spectrum of the luciferine chromophore embedded in the luciferase protein. In the following subsections, for each of the four applications, first, the computational details are introduced, and then the results are briefly discussed.

#### 3.3.1 Reduction Potential: Canonical Nucleobases in Acetonitrile

**Protocol.** In this section the application of the toolkit to compute reduction potentials from an ensemble of geometries is discussed. The reduction potential associated with the half reaction of oxidation of each one of the five canonical nucleobases (adenine, guanine, thymine, cytosine and uracil) has been computed in a solution of acetonitrile. The reduction potential is associated with the free energy of the reduction half reaction  $\Delta G_{red}$  by means of:

$$\Delta E_{red}^0 = \frac{\Delta G_{red}}{nF} - E_{red,SHE}^0 \quad (3.1)$$

where  $F$  is the Faraday constant,  $n$  is the number of exchanged electrons and  $E_{red,SHE}^0$  is the potential of the standard hydrogen electrode (SHE).<sup>203–206</sup> A value of 4.28 V determined in an earlier work is used as  $E_{red,SHE}^0$ . In the present work the Marcus theory<sup>207–210</sup> has been employed in a framework first introduced by Warshel,<sup>211,212</sup> in which it was shown that within a linear response regime, the vertical energy difference between the oxidized and the reduced states could be defined as the reaction coordinate of the redox process to compute the free energy. In this context, two MD simulations were performed: one for the reduced state ( $\mathbf{r}_R$ ) and one for the oxidized state ( $\mathbf{r}_O$ ). In what follows, the energy differences between the two states computed on top of geometries of the phase space of  $\mathbf{r}_R$  ( $\Delta E_{R \rightarrow O}\{\mathbf{r}_R\}$ ) will be referred to as the vertical ionization energy (VIE) and the energy differences computed on top of geometries of the phase space of  $\mathbf{r}_O$  ( $\Delta E_{O \rightarrow R}\{\mathbf{r}_O\}$ ) will be referred to as the vertical attachment energy (VAE). If the linear response conditions are

---

satisfied, it can be shown that the free energy in Equation 3.1 can be determined from the averages of the VIE and the VAE, as follows:

$$\begin{aligned}\Delta G_{red} &= \frac{1}{2}(\langle \Delta E_{R \rightarrow O}\{\mathbf{r}_R\} \rangle - \langle \Delta E_{O \rightarrow R}\{\mathbf{r}_O\} \rangle) - G_{e(g)} \\ &= \frac{1}{2}(\langle VIE \rangle_R - \langle VAE \rangle_O) - G_{e(g)}\end{aligned}\tag{3.2}$$

where the subscripts on the averages refer to the phase space from which the geometries were extracted to compute the vertical energy differences. The term  $G_{e(g)} = -0.867$  kcal/mol<sup>213,214</sup> is the free energy of the electron in the gas phase according to the Fermi-Dirac statistics. It has to be included in the equations because the reference SHE potential value of 4.28 V<sup>215</sup> also includes this contribution. This methodology has been widely used to compute redox potentials of redox half reactions in the past.<sup>56,57,59,141</sup>

For each of the five nucleobases, a classical MD simulation was performed on each of the two oxidation states under study. The system setup for each MD simulation was done using the AmberTools20<sup>192</sup> package and a set of different homemade scripts. For both the oxidized and the reduced forms a geometry optimization was performed at the PBEOP/6-311G(d)<sup>216-218</sup> level of theory using the NWChem<sup>194</sup> software. Implicit solvation effects were introduced by means of the COSMO<sup>219</sup> solvation model, using acetonitrile as solvent. Electrostatic potential (ESP) charges were obtained from the same DFT calculation. Bond and bond angle parameters were obtained using the Hessian matrix of the optimized geometry by means of the Semiario method.<sup>220</sup> Parameters for dihedral angles, improper torsions and van der Waals non-bonded terms were taken from the generalized amber force field (GAFF).<sup>131</sup> Each nucleobase was solvated in a truncated octahedron with a buffer of 25.0 Å with approximately 1200 acetonitrile molecules by using the packmol<sup>221</sup> software. For the oxidized form, a chloride anion was also added to neutralize the system. The force field parameters for acetonitrile were obtained in the same manner as in the case of the five nucleobases.

Each system (for both oxidation states) was minimized for 10000 steps using the steepest descent algorithm<sup>222</sup> for the 5000 first steps and the conjugate gradient algorithm<sup>223</sup> for the last 5000 steps. Afterwards, a progressive heating to 300 K was performed for 1 ns at constant volume (NVT). The first 500 ps were employed to drive the system to the desired temperature and the last 500 ps were destined to equilibrate the structure of the system. The Langevin thermostat<sup>224</sup> was applied to control the temperature taking into account a

---

collision frequency of  $2 \text{ ps}^{-1}$ . After that, the volume of the system and the appropriate density were equilibrated carrying out a 1 ns simulation in the NPT ensemble. Finally, an additional 500 ns production simulation was performed in the NPT ensemble. The pressure was maintained constant at 1 bar employing the Berendsen barostat<sup>225</sup> with isotropic position scaling and a pressure relaxation time of 2 ps. The electrostatic interactions were computed during the full protocol using the particle-mesh Ewald method<sup>226</sup> with a grid spacing of 1.0 Å. In the case of the nonbonded interactions a 10 Å cutoff was chosen. The SHAKE<sup>227</sup> algorithm restrained the bonds involving hydrogen atoms and a time step of 2 fs was used during the heating, equilibration and production stages.

For each of the trajectories in the oxidized and in the reduced states, 200 snapshots were fetched randomly from the last 450 ns of the production trajectories. The script `main.qminputs.py` was employed to select the snapshots and to automatically generate the input files for the sampled geometries in two situations: one to compute the vertical energy differences within an electrostatic embedding QM/MM framework, in which the nucleobases represented the QM part and the acetonitrile molecules were present as fixed-point charges, and one in which the point charges were removed and replaced by COSMO to compute the vertical energy differences by QM/COSMO computations. In what follows the former and the latter situations will be referred to as the dynamic Marcus explicit (DME) and the dynamic Marcus implicit (DMI) approaches, respectively, to compute redox potentials. In both cases the QM subsystem was described at the PBEOP/6-311G(d) level of theory.

**Results.** In this section, the computed reduction potentials of the five canonical nucleobases in a solution of acetonitrile are analyzed. As said above, the computations have been performed within the framework of the Marcus theory using two different solvation models for the acetonitrile solvent: explicit solvation (direct Marcus explicit - DME) and implicit solvation (direct Marcus implicit - DMI). Figures 3.5a,b show the main input and the template files used to generate the QM/MM input files for the DME calculations. It should be emphasized that the same ensemble of geometries was employed for the DMI approach, whereby the explicit point charges were replaced by an implicit solvation cavity. To generate the inputs for the DMI approach, it suffices to replace the

&externchg

---

&end

entry with the following section:

&cosmo

dielec 37.5

&end

a) 

```
&main
tpl = nwchem
top = nucleobase.prmtop
traj = nucleobase.nc
qmmask = :1
geoms = N
&end
```

b) 

```
&header
start geom0
title "geom0.in"
memory total 12 gb
&end

&basis
* library aug-cc-pvdz
&end

&chgspin
0,1
&end

&externchg
&end

&dft
xc pbeop
convergence nolevelshifting
odft
mult 1
odft
&end

&tasks
task dft
&end
```

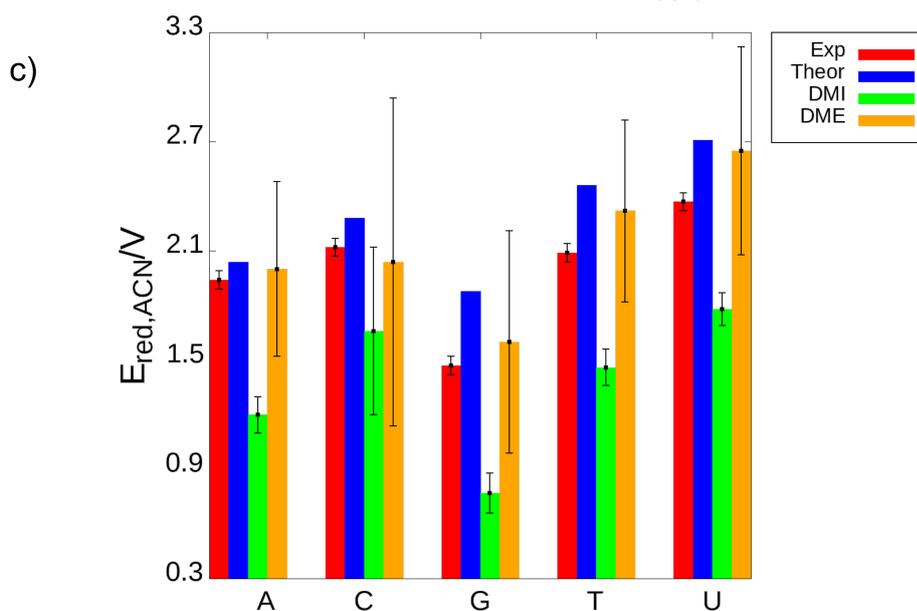


Figure 3.5: a) Main input file for the MoBioTools input generator. b) Template file for the NWChem software used to generate a set of 200 QM/MM input files for each of the nucleobases. The QM/COSMO inputs are generated using the &cosmo section instead of the &externchg entry. c) One electron reduction potentials obtained using the direct Marcus explicit (orange) and the direct Marcus implicit (green) approaches. The experimental results by Seidel<sup>228</sup> and the computational results by Crespo-Hernández<sup>229</sup> (blue) are also reported for comparison. The black lines represent standard deviations.

---

Figure 3.5c shows the reduction potentials of an ensemble of 200 geometries for each one of the nucleobases computed using the DME and the DMI approaches. For comparison, the experimental redox potentials determined by Seidel and co-workers,<sup>228</sup> and the redox potentials computed at the B3LYP/6-31++(d,p) using a thermodynamic cycle by Crespo-Hernández and co-workers,<sup>229</sup> have been reported. It can be evidenced that the DMI approach underestimates the oxidation potentials for all five nucleobases, whereas the DME computations provide the most accurate results, with values that differ at most by 0.1 V from the experimental values. Interestingly, the opposite situation had been previously evidenced by some of the authors<sup>59</sup> in the case of the reduction potentials computed in water, for which the DMI approach had outperformed the DME results. This could indicate that the implicit solvation model performs better for aqueous solvent than for acetonitrile. However, additional calculations and analyses would be necessary to corroborate it.

### 3.3.2 Energy Decomposition Analysis: Tyrosine in Water

**Protocol.** This section aims at illustrating the ability of the MoBioTools toolkit to easily process MD trajectories into input files for the calculation of interaction energies and their different quantum mechanical energy contributions of a system consisting of two fragments, at the QM/MM level of theory. The chosen test system consists of the aminoacid tyrosine in a water solution, which undergoes configurational sampling by means of classical MD, followed by the Energy Decomposition Analysis (EDA) of the resulting geometry ensemble. An electron density based QM/MM-EDA scheme<sup>86,117,118</sup> is used for such purpose, allowing the decomposition of the total interaction energy in its electrostatic, polarization (induction and dispersion), and Pauli or exchange-repulsion components.

The geometry of tyrosine was taken from reference 230. Then, the antechamber and tleap modules from AmberTools20<sup>192</sup> were employed to solvate the aminoacid in a truncated octahedron simulation box with a buffer of 22.0 Å. The tyrosine was described by the GAFF<sup>131</sup> force field, and the water molecules were described by the TIP3P<sup>137</sup> model. Once the system was set up, the next step was to minimize, heat and equilibrate the structure and density of the solvated tyrosine by classical MD using the AMBER20<sup>192</sup> package. At first, a minimization was carried out using the steepest descent method for 5000 steps and the conjugate gradient method for another 5000 steps. Then, the system was heated in

---

the NVT ensemble applying a Langevin thermostat<sup>224</sup> with  $1.0 \text{ ps}^{-1}$  of collision frequency from 0 to 303.15 K for 1 ns. After the heating process, a production simulation of 100 ns was run in the NPT ensemble applying the Monte Carlo barostat to keep the pressure at 1.0 bar. For all these steps, the cutoff and switching distances for nonbonded interactions were limited to 12.0 and 10.0 Å, respectively. During the whole protocol, a time step of 2 fs was employed and the bonds involving hydrogen atoms were constrained by using the SHAKE<sup>227</sup> algorithm.

After the equilibration of the system, single point QM/MM calculations for 100 equally spaced snapshots selected from the last 50 ns of the simulation were performed with Gaussian16<sup>188</sup> using the M062X<sup>108</sup> functional and the cc-pVDZ<sup>231</sup> basis set. The QM region included the tyrosine molecule and the 10 water molecules closest to tyrosine, with the remaining solvent included as point charges assigned by the TIP3P model. To perform the EDA, three single point QM/MM calculations were carried out for each snapshot, corresponding to the full system (QM region and MM point charges) and the fragments on which to compute the interaction energy. The first fragment consisted of the QM solute in the presence of the basis set of the QM water molecules, whilst the second fragment included all QM water molecules as well as the basis set from the solute surrounded by the MM electrostatic embedding. Basis functions of absent atoms are aimed at accounting for the basis set superposition error (BSSE) by means of the Counterpoise correction (CPC).<sup>119,120</sup> The script `main_qminputs.py` was used for the selection of the snapshots from the MD trajectory, the definition of the QM and MM regions, and the automatic generation of the three input files required by the EDA calculation.

**Results.** In what follows the energy decomposition analysis of the interaction energy of tyrosine in water is discussed. Figures 3.6a,b show the main input and the template files used to generate the QM/MM input files from an ensemble of 100 MD geometries with the size of the QM region set to fit 10 water molecules. In this case, the two subsystems for which the interaction energy was computed are the tyrosine molecule on the one hand, and the 10 QM water molecules plus the water molecules defined in the MM region on the other hand. Thus, three input files are generated for each geometry to correct for the BSSE: one for each subsystem (with the basis set of the other subsystem) and one for the complex. This option is enabled by including the `&bsse` section in the template file (Figure

---

3.6b), where the two subsystems can be defined by using Amber masks. In the present case no arguments were provided, so that the toolkit will assume that the two subsystems consist of the `qmmask` argument on the one hand, and the combination of the `solvmask` and the `closest` arguments on the other hand.

Figure 3.6c shows the normal probability distribution of each interaction energy component for the ensemble of 100 MD geometries. The Kolmogórov-Smirnov test<sup>232</sup> (5% significance level) was used to test for normality. In regard with the width of the energy distributions, the relatively large standard deviations observed (electrostatic 11.94 kcal/mol, Pauli 16.08 kcal/mol, polarization 7.92 kcal/mol, dispersion 4.68 kcal/mol, induction 3.60 kcal/mol, total 7.46 kcal/mol) restate the importance of the carried conformational sampling even for simple systems like the one studied here. In addition, as can be seen from the averaged values, the Pauli component is the main energy contribution for tyrosine in water, whilst the electrostatic energy is the most relevant among the attractive components, followed by dispersion and induction. This qualitative distribution of the interaction energy components is in consonance with the structural features of tyrosine and their interaction with a highly polar solvent like water.

### 3.3.3 Absorption Spectrum: p-Diaminoazobenzene Integrated into the Human Na<sub>v</sub>1.4 Channel

**Protocol.** In this section the machinery of the MoBioTools toolkit has been used to generate the input files for the calculation of the absorption spectrum of a photoswitch bound to a protein ion channel at the QM/MM MS-CASPT2<sup>99,101,233,234</sup> level for an ensemble of geometries. In this regard, the script `pyoverlaps.py` was used to perform the generation of the CASSCF input files for the different snapshots selected, the CASSCF calculations themselves *via* its interface with the OpenMolcas<sup>196</sup> software, and the correction of the active space for the ensemble of geometries. The script `main.qminputs.py` was employed to generate the input files for the CASPT2 calculations for those geometries for which the active space was successfully recovered.

The system under study consists of the p-diaminoazobenzene (p-DAZ) molecule inside the human’s brain voltage gated ion channel Na<sub>v</sub>1.4. An ensemble of geometries by means of a 100 ns classical MD simulation has been obtained, for which the initial geometry was

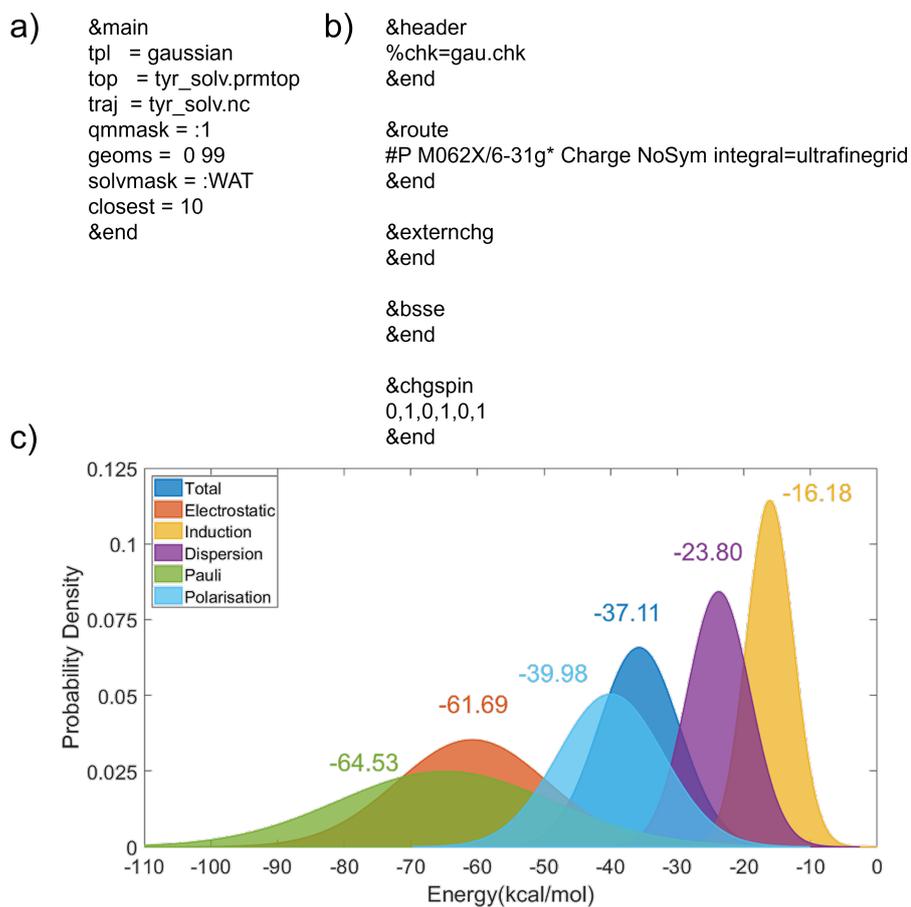


Figure 3.6: a) Main input file for the MoBioTools input generator. b) Template file for the Gaussian software used to generate a set of 100 QM/MM input files for the tyrosine molecule in water. c) Normal probability distributions for each interaction energy component in kcal/mol obtained from a sample of 100 geometries from a MD trajectory. The Pauli exchange-repulsion is represented with opposite sign. The average value of each distribution is reported in kcal/mol

---

retrieved from a previous work.<sup>235</sup> For completeness, in what follows the preparation of the initial geometry and the obtainment of the force field parameters are reported. A truncated pore model of the Nav1.4 was constructed in VMD<sup>236</sup> based on the cryo-electron microscopic structure file of the channel in complex with its  $\beta_1$  subunit (PDB ID: 6AGF).<sup>237</sup> The simulation comprised only the S5-S6 subunits of each domain, that form the central pore. The protein was aligned along the z-axis using the Positioning Proteins in Membrane (PPM) web server,<sup>238</sup> placed inside a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer (xy length: 100 x 100 lipid components) and solvated in a rectangular box with aqueous solvent and NaCl at a concentration of 0.15 mol/L using CHARMM-GUI Bilayer Builder.<sup>239</sup> The potential parameters for the protein and lipids were taken from the CHARMM36m force field,<sup>136</sup> and the TIP3P model was employed for the water molecules.<sup>137</sup> The ligand p-DAZ was constructed with the IQmol molecular viewer,<sup>240</sup> the geometry was optimized at MP2/6-31G\* level of theory and used for the calculation of the electrostatic potential (ESP) charges at the Hartree-Fock/6-31G\* level of theory. The ligand force constants for the dihedral angles  $-C-N=N-C-$  and  $-C-C-N=N-$  were taken from a previous work,<sup>241</sup> while the rest of the parameters were taken from CGenFF.<sup>242</sup> After the system set up, in a previous work Gaussian accelerated MD simulations were run to identify the possible binding pockets of p-DAZ into the Nav1.4 channel.<sup>235</sup> In the present work, one snapshot from the most stable binding pocket was selected as initial snapshot for a 100 ns MD simulation.

The 100 ns MD simulation of this work was run in the isothermal-isobaric ensemble (NPT) with a Monte Carlo barostat and a semiisotropic pressure scaling at 303.15 K and 1.01 bar. A timestep of 2 fs was used and the bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. The nonbonded interactions cutoff radius and switching distance were set to 12.0 Å and 10.0 Å, respectively. A force constant of 20 kcal/(mol Å<sup>2</sup>) was imposed on top of the  $-C-N=N-C-$  and  $-C-C-N=N-$  dihedral angles to avoid unrealistic torsion motions and to keep the photoswitch on the trans configuration. The script `pyoverlaps.py` was employed to fetch 100 equidistant geometries from the MD trajectory to perform the state average (SA) CASSCF computations with an eventual correction of the active space. For each geometry, equal weights of the first 10 states has been used to generate the reference wavefunction for the MS-CASPT2 single point

---

calculations to compute the absorption spectrum. The active space included a total of 12 electrons in 10 orbitals (12,10), which comprises two doubly occupied nitrogen lone pairs, four doubly occupied  $\pi$  orbitals and four unoccupied  $\pi^*$  orbitals, as shown in Figure 3.7.

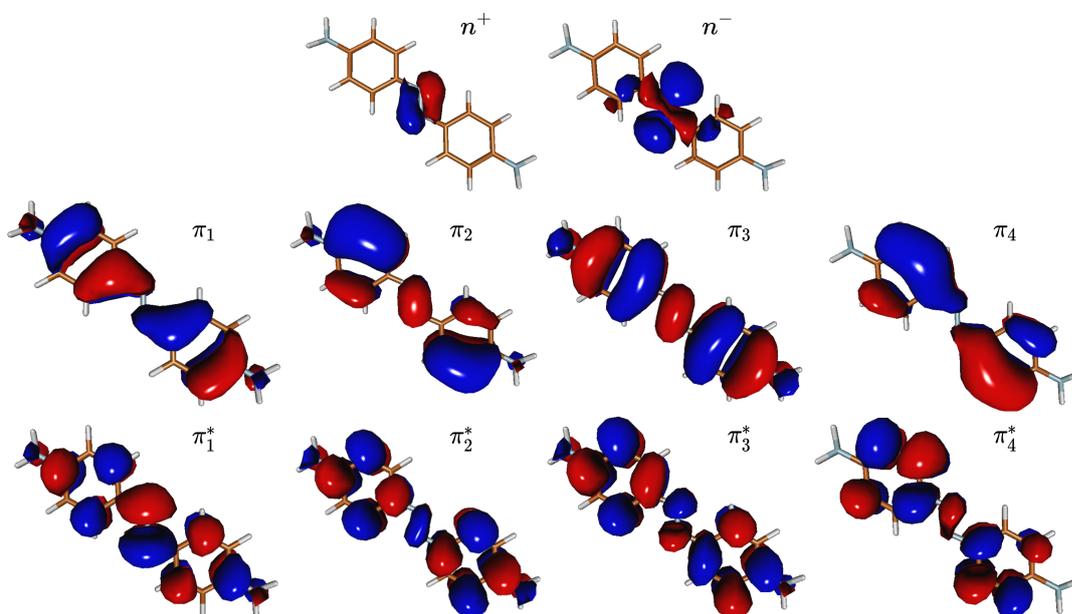


Figure 3.7: Representation of the molecular orbitals included in the active space. The  $n^+$  and  $n^-$  orbitals are the lone pairs of the central diazine bond,  $\pi_{1-4}$  are four occupied  $\pi$  orbitals and  $\pi_{1-4}^*$  are four unoccupied  $\pi^*$  orbitals.

---

The cc-pVDZ<sup>231</sup> basis set was used, and the Resolution of Identity approach was used to treat the two electron integrals.<sup>243</sup> In the MS-CASPT2 calculations, an imaginary shift of 0.2 Ha was added to the zero order Hamiltonian to preclude the inclusion of intruder states,<sup>244</sup> with an IPEA value of 0.0 Ha as previously recommended.<sup>245</sup> The restricted active space interaction (RASSI) protocol was used to compute the transition oscillator strengths to compare transition probabilities among the states studied.

**Results.** As described above, the software was used to calculate the absorption spectrum of an ensemble of geometries obtained from 100 ns of a classical MD trajectory of the photoswitch embedded in the protein. The script `pyoverlaps.py` allowed to correct the active space in correspondence with a reference vacuum calculation. Figure 3.8a,b shows the main input and template files used to generate the QM/MM input files from an ensemble of 100 MD geometries with the size of the QM region set to contain only p-diaminoazobenzene. The absorption spectrum displays a strong UV band ( $\lambda_{\max} = 383$  nm) which arises from a symmetry allowed  $\pi \rightarrow \pi^*$  transition. Intensity borrowing is observed for higher-energy lying  $\pi\pi^*$  states at 220-330 nm as well as for the symmetry forbidden  $n \rightarrow \pi^*$  transition appearing at 550-550 nm. This coupling between different electronic states of different symmetry shows the importance of sampling an ensemble of geometries around the equilibrium geometry, instead of computing the absorption spectrum only at the minimum-energy geometry.

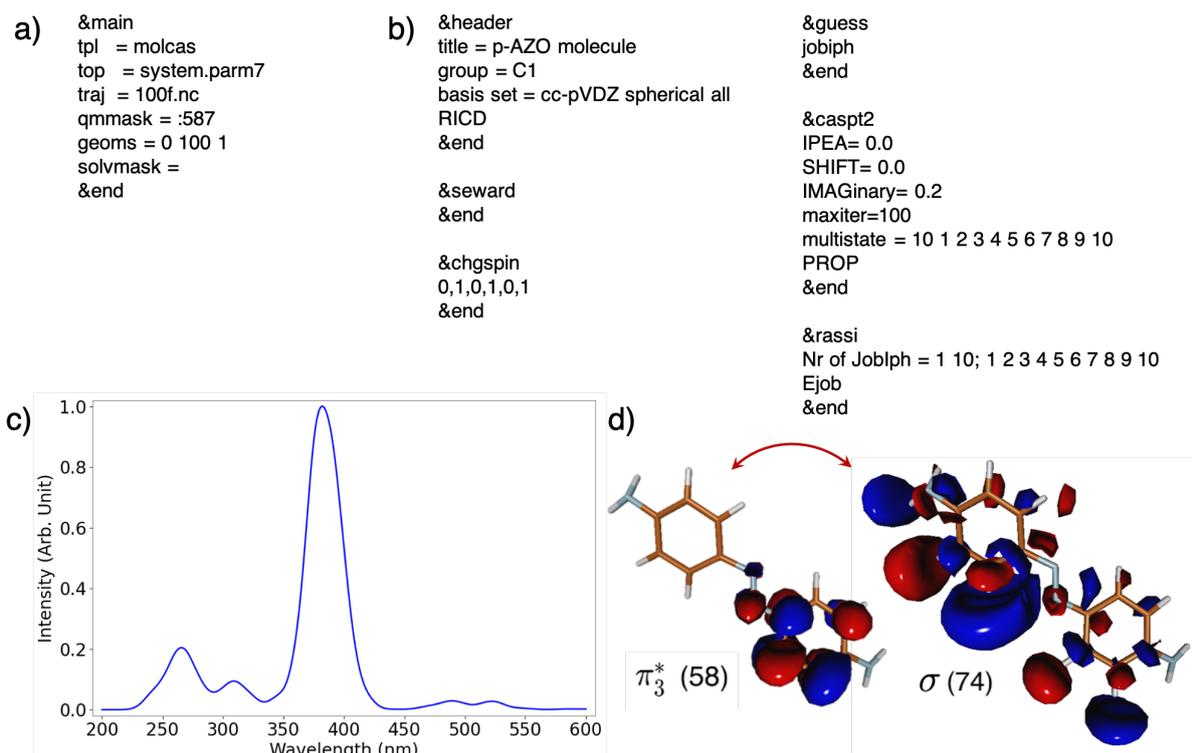


Figure 3.8: a) Main input file for the MoBioTools input generator. b) Template file for the OpenMolcas software used to generate a set of 100 QM/MM input files for the p-diaminoazobenzene molecule embedded in the  $\text{Na}_v1.4$  ion channel. c) Absorption spectrum of p-diaminoazobenzene molecule embedded in the  $\text{Na}_v1.4$  ion channel. d) Example of a  $\sigma$  orbital swapped for a  $\pi$  orbital corresponding to the reference wavefunction.

### 3.3.4 Emission and Absorption Spectra: Luciferine/Luciferase Complex

**Protocol.** As evidenced in the previous sections, the MoBioTools toolkit can be employed to compute electronically excited states for an ensemble of geometries of a molecular system. As a final example, its application is illustrated to compute the absorption and the emission spectra of the oxyluciferin chromophore in the luciferase enzyme system. In this case, two classical MD simulations were evolved using the AMBER20<sup>192</sup> software to generate the ensemble of geometries for the electronic ground state and for the first excited state, as explained below. These trajectories were processed by the `main.qminputs.py` script to generate the QM/MM input files for the subsequent excited state calculations.

The luciferine/luciferase complex was constructed using PyMol<sup>246</sup> based on the X-ray

---

diffraction structure file of firefly luciferase crosslinked in the second catalytic form (PDB ID: 4G37<sup>247</sup>). The complex was solvated with the tleap module of AmberTools20<sup>192</sup> by TIP3P<sup>137</sup> water molecules within a truncated octahedral box, ensuring a solvent shell of at least 12 Å from any solute molecule. The luciferase enzyme was described using the ff19SB force field,<sup>248</sup> whereas the GAFF force field was used to describe the dihedral and improper torsions and Lennard-Jones parameters of the chromophore. Bond and bond angle parameters for both the  $S_0$  and  $S_1$  electronic states of the oxyluciferin molecule were obtained by the Seminario method. The optimized geometries and the Hessian matrices were obtained using the Gaussian16<sup>188</sup> software by means of DFT and TD-DFT calculations, respectively. In particular, for both ground and excited state calculations the B3LYP<sup>104–106</sup>/6-311G(2d,p)<sup>249–251</sup> level of theory was employed. The ESP charges for the two electronic states were computed using the same level of theory.

A minimization was performed for 5000 steps using the steepest descent algorithm, and for another 5000 steps using the conjugate gradient algorithm. Afterwards, a progressive heating was performed from 0 to 300 K for 500 ps in the NVT ensemble using a timestep of 2 fs, followed by a constant temperature simulation for another 500 ps in the same ensemble, using a Langevin thermostat<sup>224</sup> with  $1.0 \text{ ps}^{-1}$  of collision frequency. Then, a production simulation of 100 ns was performed in the NPT ensemble, again with a timestep of 2 fs. A Berendsen barostat<sup>225</sup> and thermostat were used throughout the simulation to keep the pressure at 1.0 bar and the temperature at 300 K, respectively. The electrostatic interactions were computed using the particle-mesh Ewald<sup>226</sup> method, with a grid spacing of 1.0 Å. For the non-bonded interactions cutoff and switching distances of 9.0 and 7.0 Å were chosen, respectively. Bonds involving hydrogen atoms were constrained using the SHAKE<sup>227</sup> algorithm.

For both MD simulations in the  $S_0$  and  $S_1$  states, the first 25 ns of the production run were discarded as the equilibration time of the protein structure. The script `main_qminputs.py` was used to randomly extract 200 geometries from the last 75 ns of the MD trajectory, and to generate the QM/MM input files to compute the absorption spectrum (for the  $S_0$  trajectory) and the emission spectrum (for the  $S_1$  trajectory) of the chromophore. The QM/MM calculations were performed by Gaussian16,<sup>188</sup> where 10 excited states were computed at the B3LYP/6-311G(2d,p) level of theory for each of the

snapshots.

**Results.** As stated above, the toolkit was used to generate 200 Gaussian16 input files from the last 75 ns of the classical MD trajectory in the  $S_0$  state (for the absorption spectrum) and in the  $S_1$  state (for the emission spectrum). As the structure of the input files for the toolkit is the same in both situations, for simplicity only the main and the template files for the computation of the absorption spectrum of the oxyluciferin molecule are reported in Figures 3.9a,b. Figure 3.9c shows the computed absorption and the emission spectra of the oxyluciferin/oxiluciferase complex. In the case of the absorption spectrum, two bands can be observed, which correspond to transitions to the  $S_1$  and  $S_3$  electronic states. The orbitals involved in the first one are shown in Figure 3.9d. The emission spectrum present a band centered at 567 nm, which is in good agreement with the experimental value of 553 nm.<sup>252</sup>

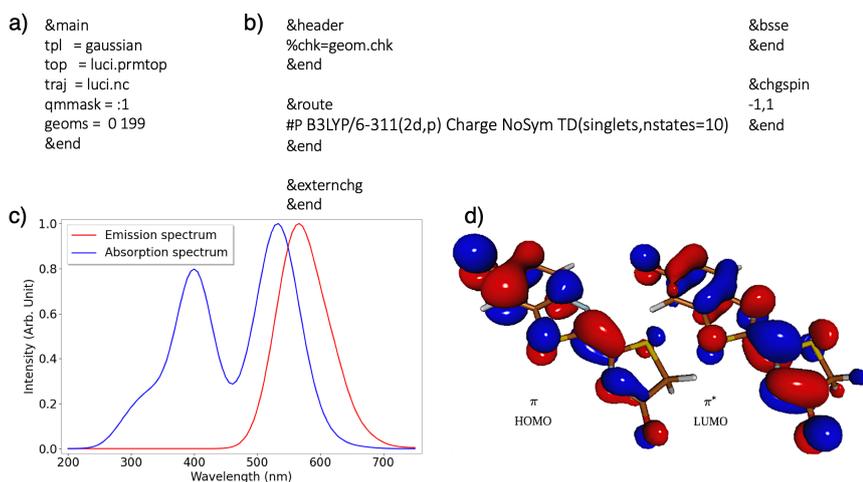


Figure 3.9: a) Main input file for the MoBioTools input generator. b) Template file for the Gaussian16 software used to generate a set of 200 QM/MM input files for the oxyluciferin molecule embedded in the luciferase enzyme system. c) Emission and absorption spectra of the oxyluciferin/luciferase solvated complex. d) Orbitals that participate in the transition from the  $S_1$  and to the  $S_1$  state.

---

## 3.4 Conclusions

In this Chapter the MoBioTools toolkit has been introduced. It allows for a straightforward setting up of a set of QM/MM calculations from an ensemble of geometries, so as to compute various physical properties at a high level of accuracy, while accounting for conformational sampling. The toolkit has been applied to four different situations, showing a large degree of versatility. Specifically, it has been used to set up the calculations of the reduction potentials of the five canonical nucleobases in acetonitrile and the interaction energies (and corresponding energy decomposition analysis) of a system consisting of tyrosine solvated by water. In the latter application, it has been evidenced how the toolkit defines the environment for cases in which part of it needs to be included within the quantum mechanical subsystem. MoBioTools has also been used to compute electronically excited state properties, such as the emission and absorption spectra of some chromophores embedded in complex biological media. In the case that a multiconfigurational method, such as CASSCF, is employed to compute the excited states of the chromophore, it has been shown that the toolkit can also be used to automatically correct the active space for those geometries of the ensemble in consideration for which a first CASSCF wavefunction optimization has provided a different active space from the one used as reference. The toolkit developed and presented here will be useful for the computational chemistry community, as it attempts to carry out in an automatic way tasks that need to be performed within the framework of QM/MM, but that otherwise would be tedious and time consuming.

## Chapter 4

# Active Space Preservation for CASSCF Calculations on Ensembles of Geometries

Gustavo Cárdenas<sup>1</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

This Chapter has been published in:

*International Journal of Quantum Chemistry*, 2021; 121:e26533.

DOI: 10.1002/qua.26533

*When the excitation energies of an ensemble of geometries, computed by the CASSCF or other CAS-based methods, are convoluted to obtain the absorption spectrum, it is advisable that all the considered geometries have the same molecular orbitals within the active space. This Chapter presents an algorithm that evaluates the molecular orbital overlap matrix between a previously selected reference geometry, with the desired active space, and each of the sampled geometries. Based on the value of the overlap matrix elements, the algorithm determines whether one or more pairs of molecular orbitals of the sampled geometry have to be swapped for a subsequent CASSCF calculation. The developed algorithm has been*

---

*applied to CASSCF and CASPT2 calculations for sets of geometries of the five canonical nucleobases in vacuum and in solvent. The algorithm shows a very good efficacy since it recovered the correct active space for 90% of the geometries which presented undesired molecular orbitals in the active space after the first CASSCF wavefunction optimization. In addition, the importance of having the same orbitals within the active space for all the geometries is discussed based on the corrected and uncorrected density of states for the nucleobases.*

## 4.1 Introduction

Multiconfigurational self-consistent field (MCSCF)<sup>82</sup> methods represent nowadays a standard *ab initio* tool used by computational chemists to study systems in situations that cannot be properly described by a single Slater determinant (single reference methods), such as bond dissociations, vertical electronic excitations, and photochemical reaction pathways, among others. Of these methods, the CASSCF<sup>81,83</sup> is perhaps the most popular method due to its conceptual simplicity and the development that it has undergone since its introduction. The CASSCF wavefunction is constructed by first subdividing the molecular orbital (MO) space into three subspaces: inactive, active and secondary orbitals. During the optimization of the wavefunction the inactive and secondary orbitals are assumed to be doubly occupied and completely empty, respectively, whereas the active orbitals are allowed to assume all possible occupations, as long as the overall spin and spatial symmetry of the wavefunction is conserved. Thus, the problem of selecting the most suitable configurations for the chemical problem at hand reduces to that of choosing the set of MOs that compose the active space. This is mainly a problem that involves a strong chemical intuition; however, the size of the set of orbitals is also limited by the size of the system, and in some cases not all suitable orbitals can be incorporated in the active space due to the huge amount of obtainable configurations. Once the CASSCF wavefunction has been obtained, it can be used as the reference wavefunction for methods that recover the so-called electron dynamical correlation due to electrons not present in the active space, such as the CASPT2<sup>233,234</sup> method, the restricted active space second-order perturbation theory (RASPT2),<sup>253</sup> or the n-electron valence second-order perturbation

---

theory (NEVPT2).<sup>254</sup> In the last decades, these quantum mechanical methods that rely on a CASSCF reference wavefunction have been extensively employed to unravel a large list of photochemical and photophysical processes.<sup>255–260</sup>

When the photoinduced behaviour of a chromophore is theoretically investigated, for example, by computing the potential-energy curves that connect different stationary points<sup>71,261,262</sup> or by evolving dynamic trajectories,<sup>263–266</sup> it is usually necessary to compare excitation energies or other electronic properties for different geometries of the chromophore. This is also the case when the absorption spectrum is calculated by considering an ensemble of geometries which are generated by, *e.g.*, MD simulations.<sup>73,267–271</sup> If the electronic-structure calculations are based on the complete active space (CAS) approach, it is advisable to use very similar orbitals from a qualitative perspective – although they present slightly different coefficients – in the active space for all the geometries that are considered. This is especially true when studying non-reactive processes such as nuclear dynamics at the Franck-Condon region, where there is no bond breaking or formation and, thus, new MOs are not formed or destroyed. Although the vibrational motion of the chromophore will modify the coefficients of the MOs, and the active spaces of different geometries will be quantitatively different, the qualitative nature of the MOs should not be altered since a drastic change of the electronic wavefunction is not expected. If a chemical reaction or a non-adiabatic process occurs, where the electronic wavefunction suffers important alterations, the modification of the active space must be done smoothly along the reaction pathway to avoid discontinuities in the potential-energy surfaces and sudden changes of the electronic properties. However, to preserve the same orbitals inside the active space is a very challenging task in many cases, especially when the molecular motion and the solvent or biological environments have to be explicitly considered in the theoretical model. In this situation it might happen that for some of the geometries the optimized active space differs from that of the desirable reference wavefunction, which could be, for example, the active space of the equilibrium geometry in vacuum. In such cases, a second (or even further) CASSCF optimization needs to be performed using the previously optimized active space as the initial guess but swapping the problematic MOs, in the hope that the desired reference active space is obtained along the different optimizations. The usual way to tackle the problem is to visually assess for each of the failed geometries which orbitals are

---

within the active space and which of them need to be swapped by inactive or secondary orbitals.<sup>71,73,272</sup> However, this becomes a formidable task when considering a large amount of sampled geometries which, in many cases, is disregarded.

In this Chapter, an approach to compare the active spaces of a set of geometries with the one of a reference geometry is proposed. This approach, whenever required, swaps the problematic MOs and performs further CASSCF calculations to correct for the wavefunctions of the sampled geometries. The method is based on the calculation of the MO overlap matrix between the CASSCF optimized orbitals of the reference geometry and each of the sampled geometries. The criterion for swapping the MOs will, thus, depend on the value of the overlap integrals of the above mentioned matrix. Although there are several efficient methods for calculating overlap integrals between many-electron wavefunctions in the literature,<sup>273,274</sup> they have not been applied with the aim of selecting or preserving the MOs of the active space along an ensemble of conformers for a given molecular species. In this work, the comparison between the wavefunction of the reference structure and that of a given sampled geometry *via* the MO overlap integrals will prove to give satisfactory results by recovering the desired active space for most of the sampled geometries presented. Density Matrix Renormalization Group methods<sup>275</sup> have been used in conjunction with entropy based MO entanglement selection methods<sup>97,276,277</sup> to choose in an automated and black-box manner the active space for a given molecular structure. The problem of choosing a consistent active space for a set of geometries connected along a specific reaction pathway has been faced by means of the previously mentioned entropy based entanglement approaches,<sup>98</sup> and more recently by analyzing the atomic orbital coefficients and the overlap matrix of the MOs exclusively within the active space of the geometries along the reaction pathway.<sup>278</sup> The approach introduced in the present Chapter aims at preserving the previously selected active space of a given molecular structure - considered to be the reference active space - along several conformers of the same molecular species, which are not necessarily close to each other within the space of coordinates, as it is the case in the computation of the absorption spectrum. Thus, our approach could somehow be regarded as complementary to the methods mentioned above if one aims at a fully automated approach using CAS methods upon an ensemble of geometries.

It should be pointed out that it is not expected that two different geometries of the same

---

molecular species possess exactly the same wavefunction – the configuration-interaction (CI) coefficients and the MO coefficients will inevitably be different. However, it is desirable that they be qualitatively similar in terms of the nature of their active spaces. As evidenced in the present work, for a given geometry different minima of the energy (as a function of the CI and MO coefficients) can be obtained for different initial guess active spaces. Being the CASSCF method a variational one, it is thus of utmost importance to tackle this potential ambiguity by attempting to attain the minimum-energy wavefunction, which is the most suitable CASSCF wavefunction for a given geometry.

In this work the theoretical insights and the implementation of the approach mentioned above are presented, followed by an application on an ensemble of force field MD sampled geometries for each of the five canonical nucleobases in vacuum and in explicit water. Nucleobases represent a suitable test case set as their photophysical and photochemical properties have been thoroughly studied both in vacuum and with implicit solvation models within several different MCSCF approaches.<sup>279–282</sup> Thus, our goal is not to reproduce accurate vertical excitation energies, but rather to show the effectiveness of our approach in recovering the appropriate CASSCF active space of an ensemble of geometries. Moreover, we seek to stress the importance of describing in a proper manner the wavefunction for each of these geometries by comparing the uncorrected CASSCF calculations with their corrected counterparts.

## 4.2 Theory and Implementation

### 4.2.1 Calculation of the MO Overlap Matrix

In the CASSCF (or in general MCSCF) framework, the wavefunction is represented as a superposition of Slater determinants that span the configuration space that is most adapted for the system under study

$$\Psi_0 = \sum_j C_j \Phi_j \quad (4.1)$$

where each Slater determinant  $\Phi_j$  is constructed from a set of  $N$  MOs  $\{\psi_{jk_i}\}_{i \in \{1..N\}}$

$$\Phi_j = |\psi_{jk_1} \dots \psi_{jk_N} \bar{\psi}_{jk_1} \dots \bar{\psi}_{jk_N}| \quad (4.2)$$

---

To construct the  $\Psi_0$  wavefunction, both the expansion coefficients  $C_j$  and the MOs  $\{\psi_{jk_i}\}_{i \in \{1..N\}}$  are variationally optimized.

In what follows a method to recover the active space in an ensemble of sampled geometries of a molecular species is presented; it uses a specific geometry and the corresponding active space as the reference. For simplicity, the Slater determinant index  $j$  will be omitted in the MO representation (Equation 4.2) as we will refer to the entire orbital space of a geometry, for which one index suffices. The orbital indices  $p$  and  $q$  will be used instead of  $k_i$ , and the MO space of the reference geometry will be represented as  $\{\psi_p\}$ , and that of any sampled geometry for which to recover the active space as  $\{\psi'_q\}$ ; thus, throughout the present Chapter, MOs without a prime ( $'$ ) will represent reference MOs, and those with a prime ( $'$ ) will represent MOs of a sampled geometry. The method we propose consists of the computation of the MO overlap matrix  $\langle \psi_p | \psi'_q \rangle$  (for simplicity  $\mathbf{S}^{MO}$ ) between the reference geometry and any of the ensemble geometries. The molecular orbitals are expanded in an atomic orbital (AO) basis

$$\psi_p = \sum_{\mu} C_{\mu p} \phi_{\mu} \quad (4.3)$$

so that the MO overlap matrix is given by

$$\langle \psi_p | \psi'_q \rangle = \sum_{\mu\nu} C_{\mu p} C'_{\nu q} \langle \phi_{\mu} | \phi'_{\nu} \rangle \quad (4.4)$$

where  $C_{\mu p}$  and  $C'_{\nu q}$  are the MO coefficients of the reference and the sampled geometries, respectively, and  $\langle \phi_{\mu} | \phi'_{\nu} \rangle$  is the corresponding AO overlap matrix  $\mathbf{S}^{AO}$ .

The  $\mathbf{S}^{MO}$  matrix could be easily determined from the expansion coefficients of the reference  $\{\psi_p\}$  and the sampled geometry  $\{\psi'_q\}$  MO sets<sup>283</sup> ( $C_{\mu p}$  and  $C'_{\nu q}$  in equation 4.4, respectively), assuming that the AOs are equal for the different geometries. However, the AO basis sets used in the present work comprise atom-centered functions, and since two different geometries of the same molecular system are compared, no such assumption can be made. The only assumption made – for the sake of comparing calculations at the same level of theory – is that the contraction schemes of both AO basis sets are the same. Therefore, the  $\mathbf{S}^{AO}$  needs to be calculated explicitly.

To this end, at first the basis set of atom-centered functions is considered; these functions are linear combinations (contractions) of spherical harmonic Gaussian primitive

---

functions of the type

$$\tilde{\phi}(\zeta, l, m, n, \mathbf{r}) = \tilde{N}(n, \zeta) Y_m^l r^n e^{-\zeta(\mathbf{r}-\mathbf{R})^2} \quad (4.5)$$

where  $\tilde{N}$  denotes the normalization constant,  $\zeta$  is the orbital exponent,  $n$  is the principal quantum number,  $Y_m^l$  is the spherical harmonic having orbital and magnetic angular momentum numbers  $l$  and  $m$ , respectively,  $\mathbf{r}$  denotes the electron coordinates, and  $\mathbf{R}$  denotes the atomic coordinates on which the Gaussian primitive is centered. However, several integral calculation algorithms rely on the usage of Cartesian Gaussian functions instead

$$\phi(\zeta, l_x, l_y, l_z, \mathbf{r}) = N(l_x, l_y, l_z, \zeta, n) (x - R_x)^{l_x} (y - R_y)^{l_y} (z - R_z)^{l_z} e^{-\zeta(\mathbf{r}-\mathbf{R})^2} \quad (4.6)$$

where  $N$  is the normalization constant,  $l_x, l_y, l_z$  are three non negative integers such that, for the orbital angular momentum number,  $l = l_x + l_y + l_z$ , whereas  $\zeta, \mathbf{R}, \mathbf{r}$  and  $n$  have the same meaning as before. In the present implementation, Cartesian Gaussian functions are used, so that instead of calculating the overlap matrix  $\mathbf{S}^{AO}$  of spherical harmonic Gaussians directly, the Cartesian overlap matrix  $\mathbf{S}_{xyz}^{AO}$  is computed.

In what follows, a *shell* will correspond to a set of Cartesian Gaussian functions having the same principal ( $n$ ) and orbital angular momentum ( $l$ ) numbers, and the components of a shell will be the set of basis functions belonging to it. Each Cartesian Gaussian function will be represented as a vector  $\mathbf{a} = (a_x, a_y, a_z)$  having as components the three numbers  $l_x, l_y, l_z$  mentioned above. Having adopted a vector description for the components of a shell, they will be represented using vector notation. Thus, the single component of an s shell having angular momentum  $l = 0$  is given by  $\mathbf{s} = (0, 0, 0)$ . To represent each of the three components of the p shell ( $\mathbf{p}_x, \mathbf{p}_y$  and  $\mathbf{p}_z$ ), and each of the six components of a d shell ( $\mathbf{d}_{x^2}, \mathbf{d}_{xy}, \mathbf{d}_{xz}, \mathbf{d}_{y^2}, \mathbf{d}_{yz}, \mathbf{d}_{z^2}$ ) the unit vector  $\mathbf{1}_i$  ( $i = x, y, z$ ), will be employed. This vector is given by

$$\mathbf{1}_i = (\delta_{ix}, \delta_{iy}, \delta_{iz}) \quad (4.7)$$

in terms of Kronecker deltas. In that case, the  $\mathbf{p}_i$  and the  $\mathbf{d}_{ij}$  components of the p and d shells, respectively, will be given by

---


$$\mathbf{p}_i = \mathbf{1}_i \quad (4.8)$$

$$\mathbf{d}_{ij} = \mathbf{1}_i + \mathbf{1}_j$$

so for example,  $\mathbf{p}_x = (1,0,0)$ ,  $\mathbf{d}_{xy} = (1,1,0)$  and  $\mathbf{d}_{z^2} = (0,0,2)$ . Given two Cartesian Gaussian functions  $\mathbf{a}$  and  $\mathbf{b}$  centered at  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, there are several algorithms that allow for the calculation of the overlap integral  $\langle \mathbf{a} | \mathbf{b} \rangle$  – that is, an arbitrary element of the  $\mathbf{S}_{xyz}^{AO}$  matrix – in a recursive manner,<sup>284–286</sup> starting from integrals of low angular momentum functions to obtain higher angular momentum integrals. In all such cases, the recursion begins with the analytic expression for the overlap integral between two s functions:

$$\langle \mathbf{s}_a | \mathbf{s}_b \rangle = \left(\frac{\pi}{\zeta}\right)^{\frac{3}{2}} e^{-\xi(\mathbf{A}-\mathbf{B})^2} \quad (4.9)$$

where

$$\xi = \frac{\zeta_a \zeta_b}{\zeta_a + \zeta_b} \quad (4.10)$$

In the present implementation, the Obara-Saika recursion relation<sup>201</sup> is used to compute the integral  $\langle \mathbf{a} + \mathbf{1}_i | \mathbf{b} \rangle$  from the lower angular momentum integrals  $\langle \mathbf{a} | \mathbf{b} \rangle$  and  $\langle \mathbf{a} - \mathbf{1}_i | \mathbf{b} \rangle$

$$\langle \mathbf{a} + \mathbf{1}_i | \mathbf{b} \rangle = (P_i - A_i) \langle \mathbf{a} | \mathbf{b} \rangle + \frac{1}{2\zeta} N_i(\mathbf{a}) \langle \mathbf{a} - \mathbf{1}_i | \mathbf{b} \rangle + \frac{1}{2\zeta} N_i(\mathbf{b}) \langle \mathbf{a} | \mathbf{b} - \mathbf{1}_i \rangle \quad (4.11)$$

in which  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{1}_i$  have the same meaning as before and  $N_i$  is the projection operator on the  $i^{\text{th}}$  component of a vector.  $\mathbf{P}$  and  $\zeta$  are defined as follows

$$\zeta = \zeta_a + \zeta_b \quad (4.12)$$

$$\mathbf{P} = \frac{\zeta_a \mathbf{A} + \zeta_b \mathbf{B}}{\zeta_a + \zeta_b} \quad (4.13)$$

so that  $P_i$  and  $A_i$  represent the  $i^{\text{th}}$  components of  $\mathbf{P}$  and  $\mathbf{A}$  in equation 4.11, respectively. The recursion in equation 4.11 stops once a component of either  $\mathbf{a} - \mathbf{1}_i$  or  $\mathbf{b} - \mathbf{1}_i$  is negative (specifically -1), and the value of the corresponding integral is set to zero. Once the  $\mathbf{S}_{xyz}^{AO}$  matrix has been determined, it is transformed into the AO matrix  $\mathbf{S}^{AO}$  of spherical harmonic Gaussians as follows

$$\mathbf{S}^{AO} = \mathbf{T}^T \mathbf{S}_{xyz}^{AO} \mathbf{T} \quad (4.14)$$

The transformation matrix  $\mathbf{T}$  is constructed by using the expansion coefficients of spher-

ical harmonic Gaussians in terms of Cartesian Gaussian functions reported elsewhere.<sup>287</sup> Finally the molecular orbital overlap matrix  $\mathbf{S}^{MO}$  is computed using equation 4.4, which in matrix notation reads

$$\mathbf{S}^{MO} = \mathbf{C}^T \mathbf{S}^{AO} \mathbf{C}' \quad (4.15)$$

where  $\mathbf{C}$  and  $\mathbf{C}'$  are the expansion coefficient matrices of the  $\{\psi_p\}$  and  $\{\psi'_q\}$  spaces, respectively.

It must be pointed out that to have a meaningful comparison between the MO sets of the reference and the sampled geometries, these need to be aligned prior to calculating the  $\mathbf{S}^{MO}$  matrix. Therefore, the center of mass of the sampled geometry is first translated so that it coincides with the center of mass of the reference geometry. Afterwards, the alignment is accomplished by determining the optimal rotation matrix  $\mathbf{W}$  that minimizes the root mean square deviation between the two structures, using the algorithm developed by Kabsch<sup>288</sup> and implemented in the quaternion algebra reformulation due to Coutsiar *et al.*<sup>289</sup> A comprehensive description of the alignment procedure for obtaining the matrix  $\mathbf{W}$  is presented elsewhere;<sup>289–291</sup> here we present only the basic insights. Considering two geometries of the same molecular system of  $N_a$  atoms, let  $\mathbf{X}$  and  $\mathbf{Y}$  be the two  $N_a \times 3$  matrices containing the coordinates of the two geometries, with  $\mathbf{X}$  being the structure to be superimposed (by means of a rotation  $\mathbf{W}$ ) to the structure  $\mathbf{Y}$ . The rotation matrix  $\mathbf{W}$ , which provides the best superposition between the structures  $\mathbf{X}$  and  $\mathbf{Y}$ , is the one that minimizes the following residual function:

$$\begin{aligned} E &= \frac{1}{N_a} \text{tr}((\mathbf{X}\mathbf{W} - \mathbf{Y})^T (\mathbf{X}\mathbf{W} - \mathbf{Y})) \\ &= \frac{1}{N_a} (G_X + G_Y - 2\text{tr}(\mathbf{M}\mathbf{W})) \end{aligned} \quad (4.16)$$

where  $G_X$  is the inner product of structure  $\mathbf{X}$

$$G_X = \text{tr}(\mathbf{X}^T \mathbf{X}) = \sum_i^{N_a} (x_{X,i}^2 + y_{X,i}^2 + z_{X,i}^2) \quad (4.17)$$

with  $x_{X,i}$ ,  $y_{X,i}$  and  $z_{X,i}$  being the x, y and z coordinates of the  $i^{\text{th}}$  atom of structure  $\mathbf{X}$ , and  $\mathbf{M}$  is the matrix product

$$\mathbf{M} = \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \quad (4.18)$$

with

$$S_{xy} = \sum_i^N x_{Y,i} y_{X,i} \quad (4.19)$$

Thus, it can be shown<sup>289</sup> that the optimal rotation matrix  $\mathbf{W}$ , in the quaternion representation, corresponds to the eigenvector having the largest eigenvalue of the following matrix:

$$\mathbf{S}_W = \begin{bmatrix} S_{xx} + S_{yy} + S_{zz} & S_{yz} - S_{zy} & S_{xz} - S_{zx} & S_{xy} - S_{yx} \\ S_{yz} - S_{zy} & S_{xx} - S_{yy} - S_{zz} & S_{xy} + S_{yx} & S_{xz} + S_{zx} \\ S_{zx} - S_{xz} & S_{xy} + S_{yx} & -S_{xx} + S_{yy} - S_{zz} & S_{yz} + S_{zy} \\ S_{xy} - S_{yx} & S_{xz} + S_{zx} & S_{yz} + S_{zy} & -S_{xx} - S_{yy} + S_{zz} \end{bmatrix} \quad (4.20)$$

so that the resulting 4-component eigenvector, which represents the above mentioned quaternion, needs to be transformed into the corresponding 3x3 matrix by a suitable transformation for quaternions.<sup>289</sup> In the case of a QM calculation including point charges, or in general an electrostatic embedding QM/MM calculation, the  $\mathbf{W}$  matrix is determined for the QM part (which coincides in the number of atoms with the reference geometry), and the point charges are rotated using the same  $\mathbf{W}$  matrix.

## 4.2.2 Comparison between the MO Sets

Once the  $\mathbf{S}^{MO}$  matrix has been calculated, the comparison between the reference orbital space  $\{\psi_p\}$  and the orbital space of the sampled geometry  $\{\psi'_q\}$  proceeds as follows: for each column of the  $\mathbf{S}^{MO}$ , the absolute value of the maximum element is fetched and its matrix element indices recorded (the row indices of the  $\mathbf{S}^{MO}$  matrix correspond to the MO labels of the reference structure and the column indices represent the MOs of the sampled geometry). This maximum value is chosen since it qualitatively represents which orbital of the reference space is the most similar to the MO of the sampled geometry under consideration. In what follows we will refer to this analysis as the column-wise analysis.

The orbitals to be swapped for the sampled geometry are determined by generating two lists of orbitals, one for orbitals that need to be removed from the active space (**lrem**) and the other one for orbitals to be added to the active space (**ladd**). An example of the outcome of the column-wise analysis is displayed in Figure 4.1a, in which we consider

---

the hypothetical situation of a 4-dimensional MO space [MOs 1-4] and a 2-dimensional active space [2,3], so an overall 4x4  $\mathbf{S}^{MO}$  matrix is considered. In this situation, each matrix element is represented with a letter that indicates what would be done with the corresponding MO of the sampled geometry (column label) if the maximum value of that column were to happen on that matrix element position (A = add to the active space, K = keep within the active space, I = ignore, R = remove from the active space), and in green are shown the matrix elements that effectively correspond to the column maximum. Thus, it is clear that the lists **ladd** and **lrem** are built depending on the position of the maximum value of each column of  $\mathbf{S}^{MO}$ ; specifically, they are generated as follows (see Figure 4.1a):

- i If the column index is outside of the range of the reference active space, but the row index is inside the range of the reference active space, the orbital corresponding to that column index is added to the **ladd** list.
- ii If both the row and the column indices are within the range of the reference active space, the orbital is kept inside the active space.
- iii If the column index is inside the range of the reference active space, but the row index is outside, the orbital corresponding to that column index is removed from the active space, *i.e.*, it is added to the **lrem** list.
- iv Finally, if both the row and the column indices are outside the range of the reference active space, then the orbital corresponding to that column index is ignored, *i.e.*, it is kept outside the active space.

The algorithm described above - the column-wise analysis - provides by itself an excellent insight on which reference MO is equivalent to each one of the MOs of the sampled geometry (Figure 4.1a). However, some problematic situations could arise because the columns are analyzed independently of one another, and, therefore, the column-wise algorithm needs to be extended to treat these situations, as exemplified in Figure 4.1. Before proceeding, and to facilitate the explanation below, we emphasize the fact that for most of the situations described in Figure 4.1, independently of where the green matrix elements are, the four corner elements have letter I, the four center elements have letter K, the four middle

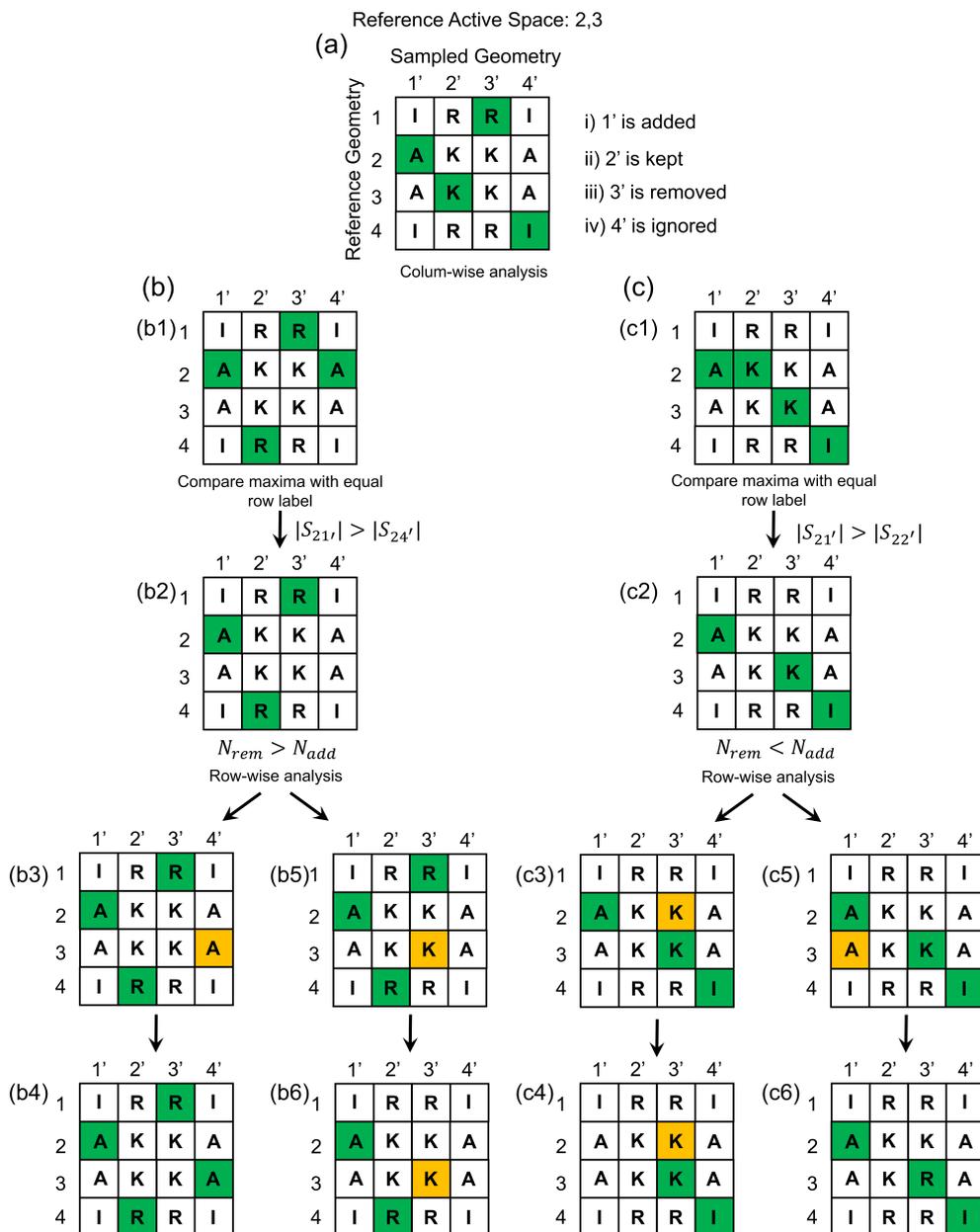


Figure 4.1: Schematic representation of the algorithm for comparing the active space of a reference geometry with the active space of a sampled geometry *via* the  $\mathbf{S}^{MO}$  matrix, in this case represented by a 4x4 matrix. Each matrix element is represented with the letter A, K, R or I indicating that the MO orbital of the sample geometry would be added to the active space, kept within the active space, removed from the active space or ignored, respectively, if the maximum value of that column were on that matrix element position. The matrix elements that effectively correspond to the column maximum are shown in green.

---

Figure 4.1: a) The outcome of the column-wise algorithm, in which a one-to-one correspondence is obtained between the reference MO and the sampled MO sets. b) and c) represent cases for which such one-to-one correspondence is not immediately obtained. The row maximum stemming from the row-wise algorithm is indicated in orange.

elements of rows 1 and 4 have letter R and the four center elements of columns 1' and 4' have letter A. The only exception to this trend is represented by step c6 in Figure 4.1c, on which the element  $S_{33'}$  changes from K to R (see below for details). First of all, a situation could occur in which two columns present a maximum overlap integral with the same reference MO. For example, the step b1 in Figure 4.1b shows that the MOs 1' and 4' of the sample geometry present a large overlap with the reference MO 2. Similarly, the step c1 in Figure 4.1c shows a large overlap of MOs 1' and 2' with the reference MO 2. To avoid such situations and guarantee that the most similar MO (according to the criterion of the overlap integral value) to a given reference MO is being considered, the absolute values of the two column maxima associated with the same reference MO are compared, and only the matrix element with the highest value is considered. Thus, since  $|S_{21'}| > |S_{24'}|$  in Figure 4.1b and  $|S_{21'}| > |S_{22'}|$  in Figure 4.1c, only the MO 1' is added to the **ladd** list. This comparison is performed whenever such situation occurs during the column-wise analysis, and is in practice an integral part of the column-wise algorithm. In most cases, it brings the analysis to a situation analogous to the one shown in Figure 4.1a, that is, one in which there is a one-to-one correspondence between the reference MOs and the MOs of the sampled geometries, and in line with this, the number of elements of the **ladd** list ( $N_{add}$ ) equals the number of elements of the **ldel** list ( $N_{rem}$ ). However, following the column-wise algorithm, there might be situations in which  $N_{add} \neq N_{rem}$  (for example, the situations displayed in the steps b2 and c2 of Figures 4.1b,c). In such cases, we perform a row-wise analysis (*i.e.*, row by row) by determining the maximum value of each row of the  $\mathbf{S}^{MO}$  matrix corresponding to a reference MO, so as to guarantee a one-to-one relation between the reference orbital space  $\{\psi_p\}$  and the orbital space of the sampled geometry  $\{\psi'_q\}$ , and proceed depending on the relation between  $N_{add}$  and  $N_{rem}$ . Specifically, the row-wise analysis proceeds as follows:

- i  $N_{rem} > N_{add}$  (Figure 4.1b, steps b3-b6): If the column label of the row maximum is outside of the sampled active space ( $S_{34'}$  of step b3), add the corresponding MO to

---

the sampled active space (MO 4') provided it had not been added to **ladd** before; in the specific case of steps b3-b4, the sampled MO 4' is added back to **ladd**, as it clearly presents high MO overlaps with MOs of the reference active space. If on the other hand the column label of the row maximum is within the sampled active space ( $S_{33'}$  of step b5), keep the sampled MO whose column label corresponds to the row maximum within the active space (MO 3') (*i.e.*, remove it from the **lrem** list - step b6). The latter situation implies that MO 3' has a high overlap with an MO within the reference active space but also with an MO outside of the reference active space.

- ii  $N_{rem} < N_{add}$  (Figure 4.1c, steps c3-c6): For each row corresponding to a reference MO, compare the absolute value of the row maximum with the absolute value of the element whose column label was determined to be equivalent to that reference MO by the column-wise criterion (steps c3 and c5). If the row maximum is within the active space of the sampled geometry ( $S_{23'}$  of step c3), and the element whose column label was determined to be equivalent to that reference MO ( $S_{21'}$  of step c3) is outside of that active space - so that it is a member of the **ladd** list, remove the element corresponding to that column label from **ladd** (step c4). If, on the other hand, the row maximum is outside of the active space of the sampled geometry ( $S_{31'}$  in step c5), and the element whose column label was equivalent to that reference MO (by the column-wise criterion) is within the active space ( $S_{33'}$  of step c5), include such element in the **lrem** list and ignore the maximum located by the row-wise analysis. Notice that the label of the element  $S_{33'}$  has changed from K to R in the step c6.

The row-wise analysis (analogously to the column-wise analysis) occurs one row at a time until the criterion  $N_{add} = N_{rem}$  is satisfied, so that on Figures 4.1b,c, row 2 is analyzed before row 3 in all cases. This implies that if the criterion is satisfied after the analysis of row 2, the algorithm stops and row 3 is not analyzed. Thus, the column-wise analysis followed by the row-wise analysis guarantees a consistent selection of MOs to be swapped at every SA-CASSCF wavefunction optimization, until the desired active space has been attained for each sampled geometry.

---

### 4.3 Computational Details

The ensembles of geometries for each nucleobase were generated by force field MD simulations performed using the AMBER18 package.<sup>132</sup> An octahedral water solvation box of 20 Å from the center of the box to either of the faces was built around each one of the nucleobases using the tleap program of AmberTools19,<sup>132</sup> and the potential of the water molecules was described by the TIP3P<sup>137</sup> solvation model. The intramolecular and van der Waals parameters for all nucleobases were taken from the general amber force field (GAFF) for organic molecules.<sup>131</sup> The geometries for all nucleobases were retrieved from Thiel’s benchmark set<sup>282</sup> except for guanine, whose geometry was obtained from the work of Wiebeler *et al.*<sup>281</sup> These geometries were reoptimized at the MP2/6-31G\* level of theory using the Gaussian16<sup>188</sup> software, and restrained electrostatic potential (RESP) charges of all the nucleobases were calculated at the B3LYP<sup>104–106</sup>/6-31G\* level of theory using the same software. For each solvated nucleobase, the system was at first minimized for 5000 steps using the steepest descent algorithm, after which the conjugate gradient algorithm was used for another 5000 steps. Afterwards, a constant volume (NVT) heating to 300 K was performed for 300 ps using a time step of 2 fs, in which positional restraints were imposed to the geometry of the nucleobase by applying a force constant of 10 kcal/mol. Three consecutive MD simulations of 1 ns each were run in the NPT ensemble to equilibrate the density and to gradually remove the positional restraints previously applied, so that force constants of 10 kcal/mol, 5 kcal/mol and no force constant were applied correspondingly. Finally, a 100 ns production simulation was run in the NPT ensemble using a Langevin thermostat to keep the temperature constant, and the SHAKE<sup>227</sup> algorithm was used to maintain fixed the bond lengths of bonds involving hydrogen atoms as required when using the TIP3P solvation model; a time step of 2 fs was selected for all the NPT simulations. For each nucleobase, 100 snapshots were fetched from the last 50 ns of the production run to compute the CASSCF/CASPT2 vertical energies and perform the MO analysis described above. For each of the geometries taken from the dynamics the excitation energies were computed in solvent by the electrostatic embedding QM/MM scheme implemented in OpenMolcas,<sup>195,196</sup> where the nucleobases were included in the QM region and the water molecules in the classical region. The

---

excitation energies were also computed for the nucleobases in vacuum by using the same ensemble of geometries as for the QM/MM calculations but removing the solvent molecules. Then, the MOs for each of the geometries were analyzed by the algorithm, as explained above, to evaluate whether the active space corresponds to the active space of the reference geometry or whether one or more pairs of orbitals need to be swapped and a second (or more) CASSCF calculation is required to obtain the desired active space. The reference geometries employed for the MO overlap analysis correspond to the equilibrium geometries of the nucleobases in vacuum; the corresponding CASSCF MOs were optimized by means of the state average (SA) CASSCF with the triple- $\zeta$  basis set developed by Alrichs,<sup>292</sup> using the OpenMolcas software.<sup>195,196</sup> For all nucleobases, no symmetry was used in the CASSCF calculations to have a better comparison with the sampled geometries, as it is expected that the  $C_s$  symmetry breaks down during the MD sampling. For most of the nucleobases, the same active spaces and number of roots as used in a previous study<sup>281</sup> were considered for the SA-CASSCF calculations. Specifically, for uracil we computed the first 10 roots using a (14,10) active space which included all  $\pi$  electrons plus the four nonbonding electrons of the oxygen atoms; for cytosine we computed the first 8 roots with a (14,10) active space. For adenine 11 roots were computed with a (18,13) active space and for guanine we considered a (20,14) active space and computed the first 9 roots. In the case of thymine, an active space smaller than the one reported before<sup>281,282</sup> was used, that is a (14,10) active space, in which we have excluded the molecular orbital localized on the methyl group (See Supporting Information for more details), and the first 10 roots were computed in the SA-CASSCF calculations. The MOs analyzed with the MO overlap algorithm correspond to the state-averaged natural orbitals, as provided by the OpenMolcas SA-CASSCF calculations. To further investigate the effect of the variation of the CASSCF active space on the ensemble of geometries we performed MS-CASPT2 on top of the SA-CASSCF optimized wavefunctions with the undesired active spaces, as well as on top of the SA-CASSCF wavefunctions with the corrected active spaces, using the same number of states as in the SA-CASSCF calculation for each nucleobase, an imaginary level shift of  $0.2 E_h$ <sup>244</sup> and no IPEA correction;<sup>245</sup> otherwise the default parameters of the OpenMolcas software were used. All MO figures were made using the molden software.<sup>198</sup>

---

## 4.4 Results and Discussion

In what follows the algorithm is presented in a visual and schematic way for an example geometry for uracil in water taken from the force field MD simulations. Figure 4.2a shows the orbitals of the active space –orbitals (23) to (32)– for the reference geometry of uracil, which corresponds to the optimized geometry in vacuum. The orbitals after the first CASSCF calculation for the geometry labeled as geometry 29, and those corresponding to the same geometry, but after performing the  $\mathbf{S}^{MO}$  analysis and the second CASSCF wavefunction optimization are also shown. As can be seen, the active space resulting from the first CASSCF calculation for the sampled geometry contains two molecular orbitals, namely (24') and (31'), which should not be part of the active space. Consequently, there are two orbitals – (19') and (37') – which erroneously lie outside the active space. The developed algorithm is able to detect these two pairs of orbitals, which need to be swapped in the subsequent CASSCF calculation, by computing the overlap matrix between the MOs of the sampled and reference geometries, whose relevant rows and columns are shown in Figure 4.2b. The elements of the matrix columns (24') and (31'), which represent the overlap between the orbitals (24') and (31') of the geometry 29 and the orbitals of the active space of the reference geometry, have very small absolute values, indicating that the MOs (24') and (31') have to be removed from the active space. On the contrary, the matrix columns (19') and (37') present large overlap values with the reference MOs (23) and (31) and, therefore, orbitals (19') and (37') have to be included in the active space. For each of the geometries of the ensemble, the algorithm analyzes the MO overlap matrix after the CASSCF computation to determine whether there are orbitals that have to be swapped. If this is the case, a new CASSCF calculation is performed with the modified active space, and this procedure is repeated until the appropriated active space is obtained or until the maximum number of iterations of the  $\mathbf{S}^{MO}$  analysis algorithm is reached. In the example shown in Figure 4.2a the MOs inside the active space for the geometry 29 are the same as the MOs of the active space of the reference geometry after the second iteration of the algorithm and, therefore, the CASSCF optimization is converged.

The algorithm is not only useful to perform the swapping of the MOs in an automatic way along all the geometries of the ensemble, but also to identify the orbitals to be swapped

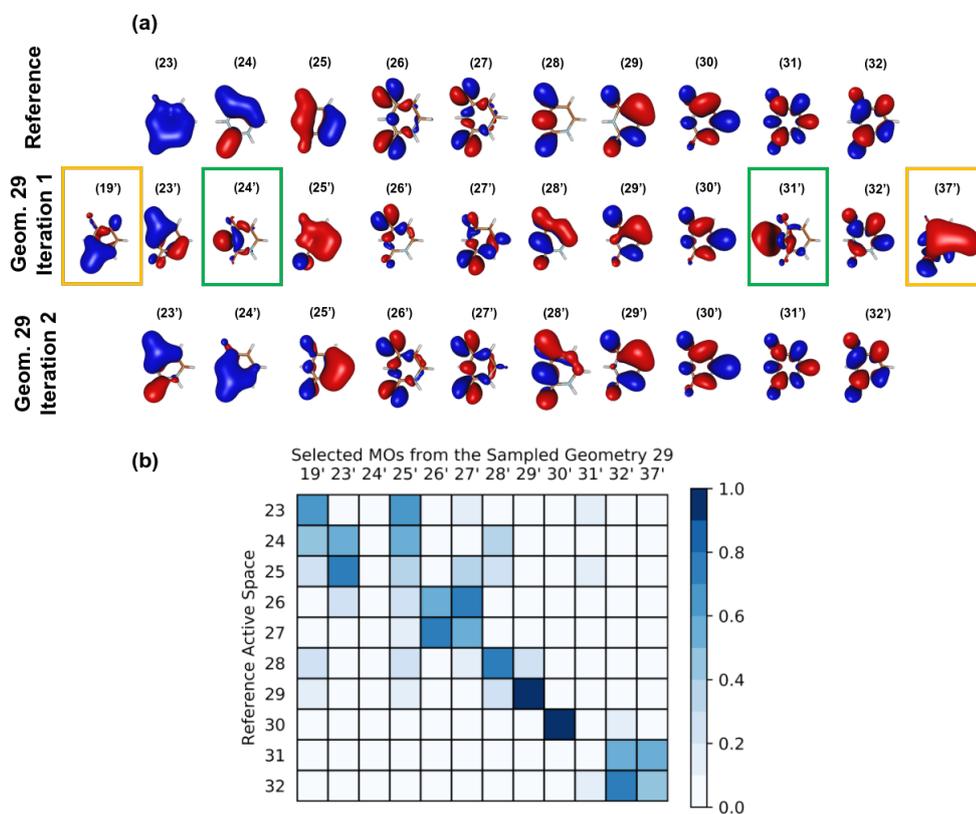


Figure 4.2: An illustration of the working principle of the  $\mathbf{S}^{MO}$  analysis algorithm for sampled geometry 29 of uracil. (a) A comparison among the reference active space, the active space of geometry 29 after the first SA-CASSCF calculation - displaying the MOs that need to be removed (green) and added (orange) to the active space-, and the active space of geometry 29 after the second SA-CASSCF calculation (*i.e.*, following the  $\mathbf{S}^{MO}$  analysis). (b) A portion of the  $\mathbf{S}^{MO}$  matrix showing the absolute values of the overlap integrals. It can be clearly evidenced that MOs 24' and 31' display negligible overlap integrals with the MOs of the reference active space, whereas MOs 19' and 37' show significant overlaps with at least one MO in the reference active space.

in a quantitative and easy way. In some cases, the identification of the undesired and desired orbitals is straightforward and could be done just by visualisation. For example, as seen in Figure 4.3a, it is relatively easy to directly identify the MO (22') of the geometry 55, which lies outside the active space after the first CASSCF calculation, as the equivalent counterpart of the orbital (26) of the reference geometry. Thus, it should be included in the active space. Moreover, the orbital (24'), included in the active space by the first CASSCF wavefunction optimization, is clearly different from all the active-space MOs of the reference geometry of uracil plotted in Figure 4.2a and, therefore, has to be removed from the active space. However, in many other cases, the visual characterization of the MOs is an impossible task if the optimization process did not reach the global minimum and the wavefunction is not converged with respect to the wavefunction of the reference geometry. For example, our algorithm has identified the orbital (37') of the geometry 27, shown in Figure 4.3b, as being equivalent to orbital (31) of the reference geometry. Therefore, orbital (37') needs to be included in the active space in replacement of the

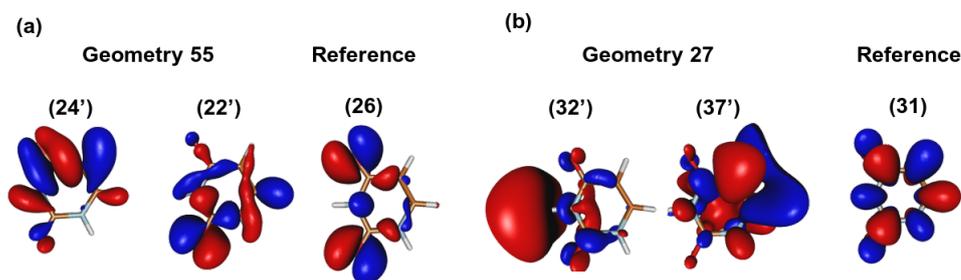


Figure 4.3: Two examples of geometries of uracil whose active spaces were recovered after a  $\mathbf{S}^{MO}$  analysis. For geometry 55 (a), the algorithm correctly identifies MO 22' as being equivalent to the reference MO 26, as can be verified by simple inspection. For geometry 27 (b), the algorithm identifies MO 37' as being equivalent to the reference MO 31, although their similarity is not qualitatively straightforward.

orbital (32'). This assignment would have been impossible to do by simple visualization since both orbitals (32') and (37') significantly differ from the reference orbital (31), and it would be hard to decide which one should be included within the active space.

Although the main goal of the algorithm presented here is for it to be applied in multi-scale QM/MM calculations, it can also be employed in calculations where the chromophore is in vacuum. Thus, 100 geometries have been considered for each of the five nucleobases

---

(deriving from the previously mentioned MD simulations) to perform CASSCF calculations in vacuum (for which the water point charges have been removed) and CASSCF/MM calculations in water (considering the solvent point charges). This was done in order to evaluate whether there are differences in the efficacy of the algorithm associated with the presence of the point charges. In both vacuum and solvent calculations, the initial MOs were taken from the previous CASSCF calculations for the optimized nucleobases in vacuum. Since large geometrical alterations with respect to the vacuum optimized geometries should not occur along the ground-state classical MD simulations, it could be expected that the orbitals of the optimized nucleobases in vacuum is a good initial guess for the calculations of the geometries from the MD simulation. Therefore, it would be reasonable to assume that the vast majority of the CASSCF calculations finish correctly with the right active space. However, as Table 4.1 shows, this is partially true only for adenine, for which 97 out of 100 sampled geometries in the absence of the point charges (vacuum) and 82 out of 100 geometries in the presence of the water point charges present the right active space after the CASSCF calculation. However, for the other canonical nucleobases a large amount of geometries have undesirable MOs in the active space. Specifically, for the sampled geometries in vacuum, 24 geometries for uracil, 20 for cytosine, 36 for thymine and 63 for guanine have an undesired active space after the first CASSCF calculation. In the case of the sampled geometries in the presence of point charges, 39 geometries for uracil, 37 for cytosine, 34 for thymine and 77 for guanine do not present the desired active space after the first CASSCF calculation. Therefore, for all these geometries it was necessary to swap some orbitals based on the analysis of the MO overlap matrix. In the case of the geometries in vacuum, the algorithm introduced here shows an outstanding efficacy since for most of the geometries initially presenting inappropriate MOs within the active spaces, the corresponding active spaces were successfully corrected as can be seen in Table 4.1 (vacuum). In particular, 100%, of the geometries with undesired MOs within the active space have been corrected by the algorithm for uracil, cytosine, thymine, adenine, and 87% of the wrong geometries for guanine. For all the nucleobases, the right active space for the recovered geometries has been achieved, on average, after carrying out one additional CASSCF calculation, in which two pairs of orbitals have been swapped.

To investigate whether there exists a correlation between the degree of geometrical

---

distortion of the sampled geometries with respect to the reference geometries, and the efficacy of the algorithm, we have calculated for each nucleobase the root mean square deviation (RMSD) of the atomic displacements with respect to the corresponding reference geometry. We have evidenced an average RMSD (averaging over the sampled geometries) of 0.10 Å, 0.44 Å, 0.54 Å, 0.37 Å and 0.47 Å for uracil, cytosine, thymine, adenine and guanine, respectively. Interestingly, in spite of the fact that guanine - the only nucleobase which did not present a 100% success rate after the  $\mathbf{S}^{MO}$  analysis - presents a higher average RMSD than uracil, cytosine and adenine, thymine presents an even higher average RMSD with respect to its reference geometry. Therefore, no evident correlation has been observed between the degree of the global geometrical variations of the sampled geometries with respect to the reference geometries and the efficacy of the  $\mathbf{S}^{MO}$  analysis algorithm. However, we do not exclude that specific distortions (*e.g.*, particular normal modes of vibration) play a role in the difficulty with which the desired active space of a sampled geometry is recovered and investigations in this regard need to be done in a systematic manner in the future. When analyzing the results of the calculations regarding the same geometries but in the presence of the point charges, the results are still very satisfactory: 100%, 92%, 91%, 94% and 74% of the wrong geometries have been corrected by the algorithm for uracil, cytosine, thymine, adenine, and guanine, respectively. Noteworthy, the number of geometries that needed to be recovered after the first CASSCF calculation is higher in the presence of point charges, and also the percentages of the recovered geometries were lower than in the case of the geometries in vacuum. This implies that the point charges play a non-negligible role in determining whether it is possible to recover the active space of a given geometry. Since the calculations in the presence of the solvent environment have shown to be more challenging, the QM/MM computations will be analyzed in more detail.

Interestingly, for most of the geometries that were not corrected after five iterations of the  $\mathbf{S}^{MO}$  analysis (and therefore after five consecutive SA-CASSCF wavefunction optimizations), the algorithm identified, outside of the active space, MOs presenting lone electron pairs (associated either with an oxygen atom of a carbonyl group or with a pyridinic nitrogen) that needed to be included in the active space. Specifically, for all the non recovered geometries of cytosine, the MOs to be added to the active space were MOs equivalent to the reference MO 26 /Figure A.2 in Appendix A.1), associated with the lone

---

pairs of the carbonyl oxygen and the pyridinic nitrogen. In the case of thymine, the MOs to be recovered were all equivalent to the reference MO 28 (lone pairs of the carbonyl oxygen atoms, (Figure A.3 in Appendix A.1). For the sole non recovered geometry of adenine such MO was equivalent to the reference MO 31 (lone pairs of pyridinic nitrogen atoms, Figure A.4 in Appendix A.1). Despite guanine represents a special case in terms of the higher number of non recovered geometries, for most of these the algorithm identified MOs that needed to be added to the active space which, again, were associated with reference MOs presenting lone pairs (specifically, MOs equivalent to MO 30, MO 33 and MO 37 on Figure A.5 in Appendix A.1). This might indicate a correlation between the difficulty of recovering the active space in some cases and the MOs that need to be recovered. These orbitals are usually nonbonded MOs centred at the heteroatoms, which are more exposed to the solvent than other orbitals. Therefore, this suggests that the interaction between the nonbonded MOs and the MM point charges is not properly described by an electrostatic-embedding approach. However, this is not the only reason behind the failure in recovering the active space since the algorithm was not able to correct 13% of the failed geometries for guanine in vacuum, where there is not interaction with the environment. Further calculations and analyses, which are beyond the scope of the present work, would be necessary to unveil all the factors that influence the MO analysis.

The preservation of the active space along all the geometries of the ensemble is very important. Different CASSCF – or in general CAS-based calculations – can be combined, for example, to obtain the absorption spectrum or the density of states, only when all the computed geometries have qualitatively similar active spaces. Otherwise, different geometries would have electronic properties, such as excitation energies and oscillator strengths, which rely on qualitatively (in some cases drastically) different CASSCF wavefunctions. The convolution of these energies based on very different wavefunctions to obtain the spectrum would be equivalent to the convolution of energies computed at different levels of theory. The importance of preserving the active space is exemplified in the following analyses. Figure 4.4 displays, for each of the recovered geometries for the five solvated nucleobases, the difference in excitation energies for the  $S_1$  state between the wrong and corrected active-space CASSCF calculations, against the variation of the CI coefficient of the most representative determinant in the corrected active-space calculation.

Phase	Nucleobase	Correct Geometries <sup>a</sup>	Wrong Geometries <sup>b</sup>	Recovered Geometries <sup>c</sup>	Average Number of Iterations	Average Number of Swaps
Vacuum	Uracil	76	24	24 (100%)	1.04	1.46
	Cytosine	80	20	20 (100%)	2.00	1.00
	Thymine	64	36	36 (100%)	1.00	1.69
	Adenine	97	3	3 (100%)	1.00	1.67
	Guanine	37	63	55 (87%)	1.09	2.29
Water	Uracil	61	39	39 (100%)	1.05	2.07
	Cytosine	63	37	34 (92%)	1.00	1.88
	Thymine	66	34	31 (91%)	1.22	2.19
	Adenine	82	18	17 (94%)	1.00	1.76
	Guanine	23	77	57 (74%)	1.12	2.22

Table 4.1: Results of the  $\mathbf{S}^{MO}$  analysis on the overall set of sampled geometries for all five canonical nucleobases, considering the ensemble of geometries in vacuum (*i.e.* removing the point charges of the solvent), and in the presence of the solvent. The wrong geometries represent those for which the active space differed from the reference active space. The recovered geometries are represented with the percentage over the whole set of geometries presenting the wrong active space after the first SA-CASSCF iteration.

<sup>a</sup> Geometries presenting the desired active space after the first SA-CASSCF iteration

<sup>b</sup> Geometries presenting an undesired active space after the first SA-CASSCF iteration

<sup>c</sup> Geometries for which the desired active space was recovered after performing the  $\mathbf{S}^{MO}$  analysis. They are recovered in the sense that they do not need to be discarded for studying the ensemble properties of the system.

It should be clarified that for the uncorrected CASSCF calculation the configuration under consideration will not be identical to that of the corrected calculation, since they will differ in the MOs that have been swapped to correct the calculation, not to mention that the ordering of the MOs need not be the same in both calculations. However, for the MOs that were not swapped - and thus, that were equivalent between the two configurations - eventual differences in ordering have been accounted for by ensuring that the occupations of equivalent MOs were the same in both configurations under comparison. As can be seen in Figure 4.4, many geometries present large energy deviations – in some cases larger than 0.5 eV – after the MO swapping to attain the desired active space, especially for uracil, cytosine, and guanine.

It is also interesting to see in Figure 4.4 that there is no correlation between the difference in excitation energy and the difference in the CI coefficients. This means that large excitation energy variations are not always originated by a significant alteration of the electronic wavefunction. In other words, an important error in the electronic properties can be obtained even when the undesired orbitals in the active space are not orbitals involved in the main electronic transitions of the electronic state under study. Two specific cases

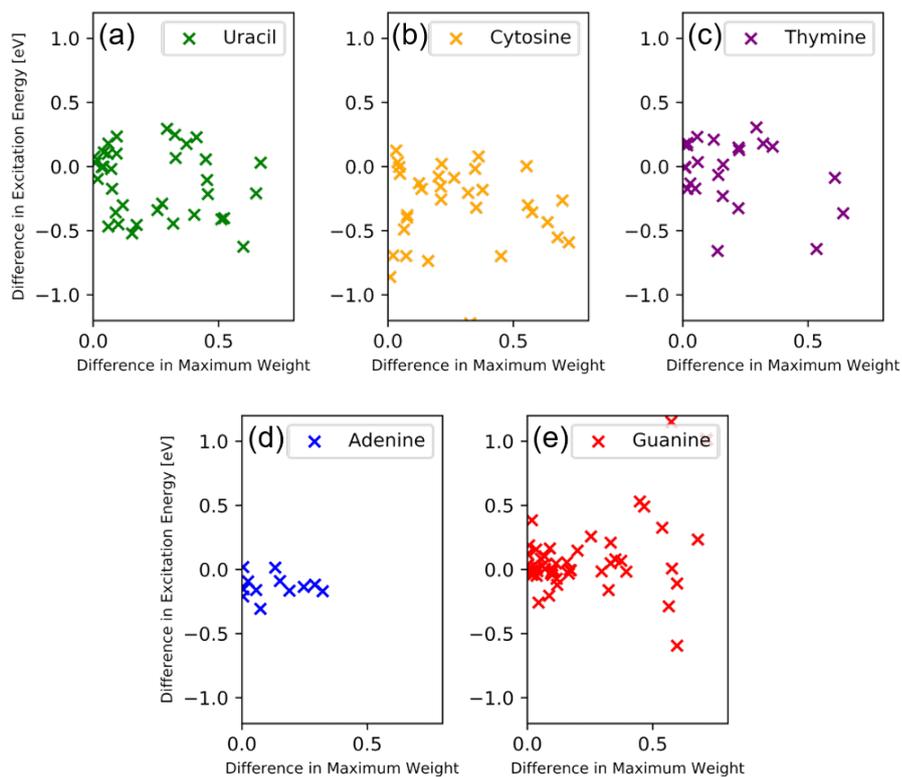


Figure 4.4: Difference in Excitation Energies towards the  $S_1$  state between SA-CASSCF calculations with the corrected and the uncorrected active spaces for uracil (a), cytosine (b), thymine (c), adenine (d) and guanine (e), vs the difference between the coefficient of the most representative determinant in the corrected active space calculation with its corresponding coefficient in the uncorrected calculation

---

are discussed in more detail in Table A.4 in Appendix A.1.

Although it is evident that swapping the problematic MOs to preserve the active space of the reference geometry provides significantly different results than the first SA-CASSCF wavefunction optimization, this does not directly imply that the result with the swapped orbitals is better. However, this can be easily investigated by comparing the SA-CASSCF energies of the electronic states obtained with the corrected and uncorrected active spaces. The active space that provides the lowest energies, and thus that approaches more to the full-CI result, contains the most suitable MOs. Figure 4.5a-e shows the energy difference for the six lowest-lying electronic states between the corrected and uncorrected active-space SA-CASSCF calculations for the geometries whose active space were successfully recovered by the algorithm. As can be seen, the energies of the vast majority of electronic states obtained by employing the corrected active space are lower than the energies of the states obtained after the first SA-CASSCF wavefunction optimization. Specifically, the use of the corrected orbitals gives electronic energies that are, in average, 0.59, 0.72, 0.85, 0.61, and 0.66 eV lower than those of the uncorrected calculations for uracil, cytosine, thymine, adenine, and guanine, respectively. Therefore, the wavefunction corrected by the algorithm, which is composed by the same MOs of the active space of the reference geometry, is more accurate than the wavefunction obtained after the first CASSCF calculation. Since the energies between the corrected and uncorrected calculations are clearly different, the density of states presents also important differences as is shown in Figures 4.5f-j. The band structure of both calculations clearly differs with different number of peaks located at different positions for the five solvated nucleobases.

It is interesting to see that the large energy differences displayed between the corrected and uncorrected SA-CASSCF calculations partially disappear when MS-CASPT2 calculations are performed on top of the SA-CASSCF wavefunctions. Figure 4.6a-e shows that the MS-CASPT2 energies based on the corrected SA-CASSCF wavefunction can be lower or higher than the MS-CASPT2 energies based on the uncorrected SA-CASSCF wavefunction depending on the geometry considered. This is an expected result due to the non-variational nature of the MS-CASPT2 approach, which can provide a better result even when a worse reference wavefunction is employed or *vice versa*. On average, the absolute energy difference between the corrected and uncorrected MS-CASPT2 calculations, taking

---

into account only the solvated geometries for which the active space was recovered, is around 0.1 eV for the five nucleobases. Despite this fortuitous error cancellation observed when a large number of geometries is considered, the band structure of the density of states suffers important modifications after the swapping of the MOs to keep the same orbitals of CASSCF wavefunction of the reference geometry, as can be seen in Figure 4.6f-j. In addition, when individual geometries are considered, the error in the CASPT2 excitation energies is larger than 0.50 eV for several geometries. These errors in the electronic properties introduced by the presence of undesired MOs in the active space are not only relevant when investigating the photoabsorption of the chromophores at the Franck-Condon region by means of static calculations. It is evident that these errors can also be present in excited-state dynamics simulations and, thus, artificially alter the conclusions about the photophysics and photochemistry of the chromophores extracted from the dynamics.

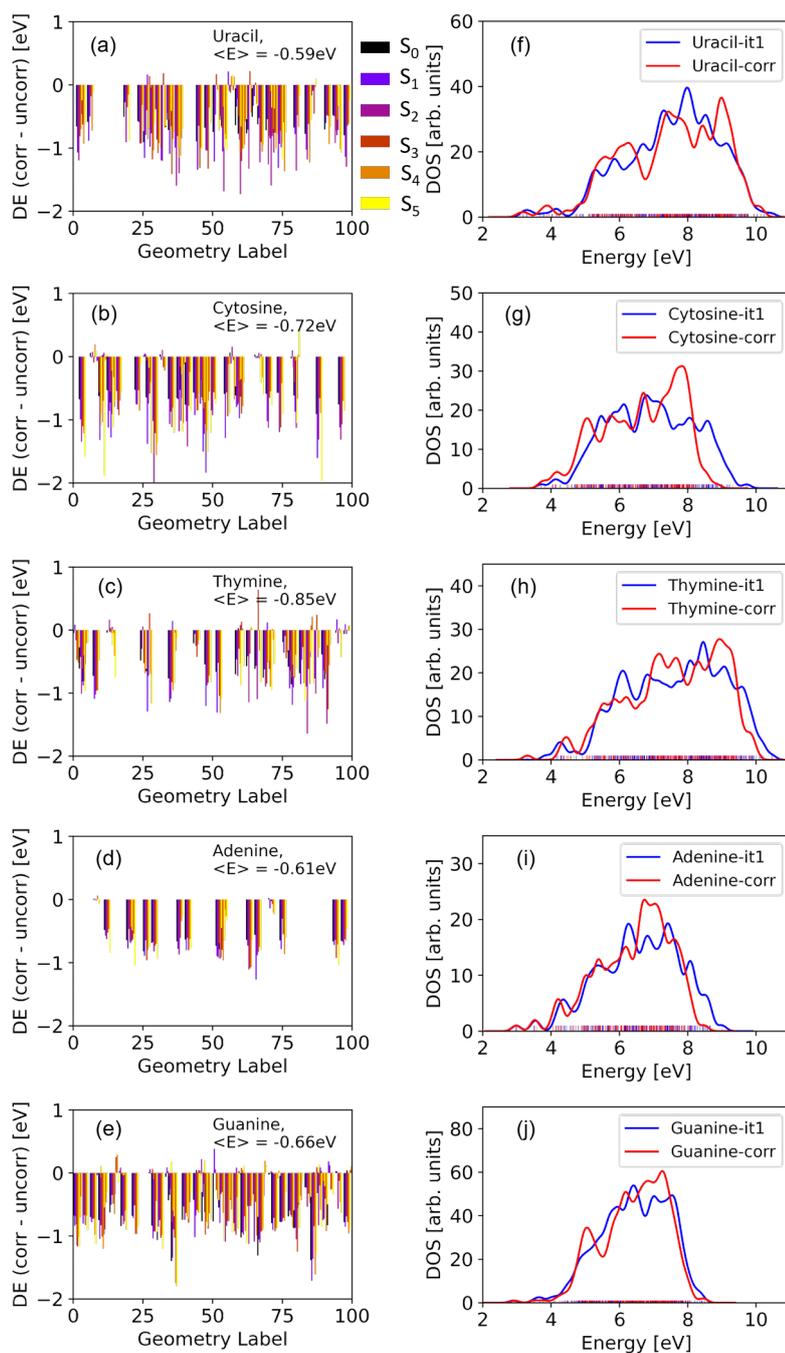


Figure 4.5: (Left). Absolute energy differences DE for the ground state and the first five singlet excited states between the SA-CASSCF calculations for the uncorrected and the corrected active spaces for uracil (a), cytosine (b), thymine (c), adenine (d) and guanine (e). (Right) Comparison of the density of states (SA-CASSCF) obtained with the active space stemming from the first SA-CASSCF wavefunction optimization (blue), and the density of states (SA-CASSCF) obtained using the target active space (red) for uracil (f), cytosine (g), thymine (h), adenine (i) and guanine (j). Only those geometries for which the active space was successfully recovered are considered.

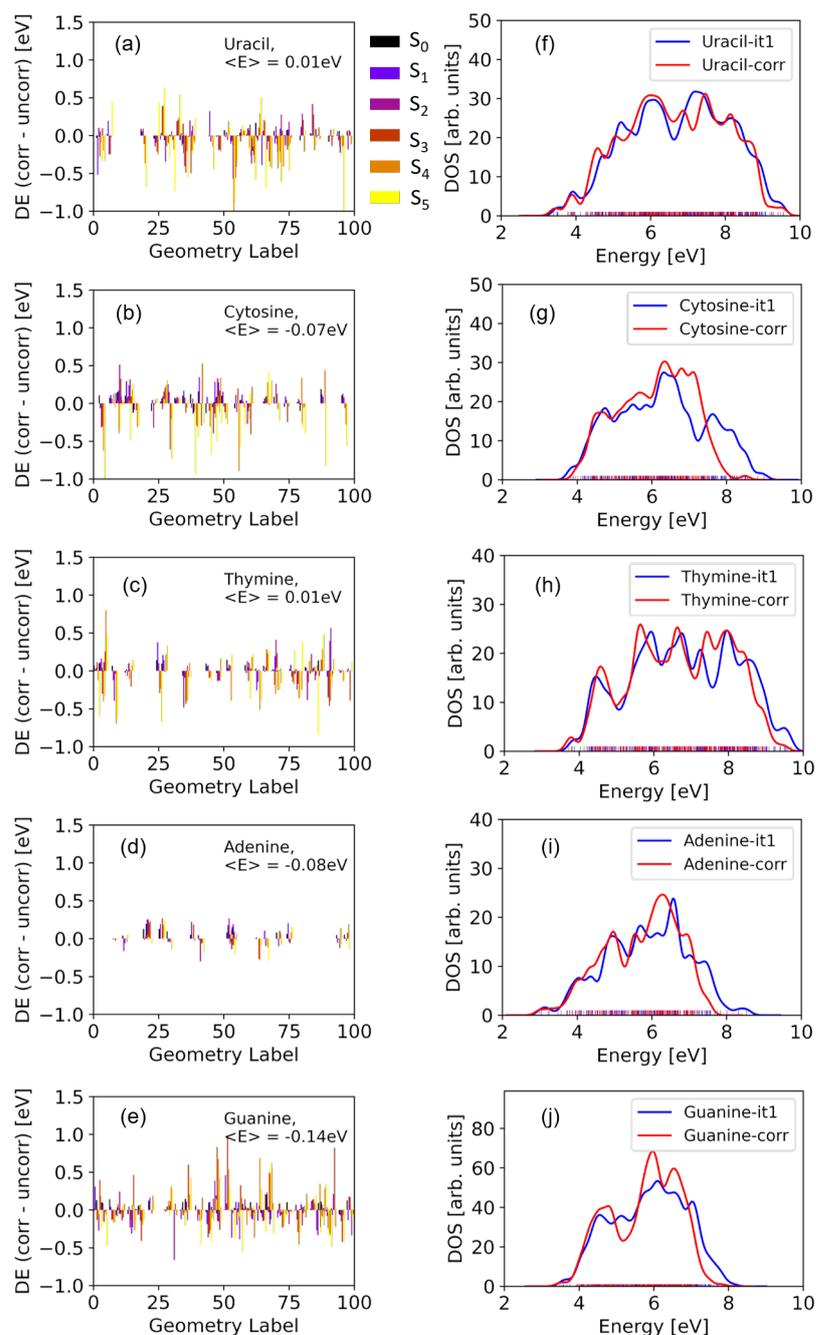


Figure 4.6: (Left). Absolute energy differences for the ground state and the first five singlet excited states between the MS-CASPT2 calculations for the uncorrected and the corrected active spaces for uracil (a), cytosine (b), thymine (c), adenine (d) and guanine (e). (Right) Comparison of the density of states (MS-CASPT2) obtained with the active space stemming from the first SA-CASSCF wavefunction optimization (blue), and the density of states (MS-CASPT2) obtained using the target active space (red) for uracil (f), cytosine (g), thymine (h), adenine (i) and guanine (j). Only those geometries for which the active space was successfully recovered are considered.

---

## 4.5 Conclusions

The analysis and comparison of electronic properties computed by CAS-based approaches for different geometries of the same chromophore requires the use of the same MOs in the active space for all the geometries. This issue is often ignored when, for example, the absorption spectrum or the density of states is computed by selecting an ensemble of geometries which was generated by a sampling technique, such as classical MD. It is generally assumed that the sampled geometries keep the same active space as the reference active space provided as initial guess in the CASSCF calculation. However, in this Chapter it has been demonstrated that this is not necessarily true by performing CASSCF/MM and CASPT2/MM computations for the five canonical nucleobases in vacuum and in water. Specifically, 146 (in vacuum) and 205 (in water) out of the 500 geometries employed in the calculations presented undesired orbitals inside the active space after the CASSCF wavefunction optimization. The correction of the active space by swapping the undesired MOs by the desired ones would be a tremendous task if one aims at performing it manually after visual inspection of the orbitals for all the failed geometries.

In this Chapter it has been presented an algorithm which automatically assesses whether the sampled geometries display the desired active space and, when this is not the case, identifies the MOs that have to be swapped in subsequent CASSCF wavefunction optimizations. The MOs to be swapped are selected based on the computation of the overlap matrix between the MOs of a reference geometry and the MOs of the sampled geometries. The algorithm was able to correct, in average, the orbitals of 95% (in vacuum) and 87% (in water) of the geometries that presented a undesired active space after the first CASSCF calculation. The efficacy of the algorithm was lower for guanine than for the other nucleobases, indicating that the preservation of the active space along the ensemble could be more difficult to achieve for large chromophores. In addition, the visual inspection of the non-recovered geometries showed that nonbonded MOs centred at the heteroatoms, which are directly exposed to the solvent, are usually the problematic orbitals. This could indicate that the interaction of these orbitals with the point charges of the water molecules is not properly described and hinders the insertion of the nonbonded orbitals inside the active space.

---

The importance of having the correct active space was also demonstrated. Specifically, it has been shown that the SA-CASSCF energy of the electronic states is lower, and therefore more accurate, when the orbitals of the active space of the reference geometry are employed than when undesired orbitals appear in the active space. It has also been discussed that the excitation energies can suffer artificial and large shifts when undesired orbitals enter the active space during the wavefunction optimization, even when these orbitals are not involved in the excitation that dominates the electronic state. These energy shifts have induced important modifications on the shape of the density of states bands at both SA-CASSCF and MS-CASPT2 levels. Therefore, the amendment of the active space for all the geometries of the ensemble is highly advisable.

The algorithm has been applied to the calculation of the density of states of the five canonical nucleobases in vacuum and in aqueous solution showing a good performance. However, more challenging situations need to be investigated in the future to assess the robustness of the implemented algorithm. For example, chromophores that undergo larger geometrical alterations within the Franck-Condon region, biological environments that strongly interact with the chromophore, or chemical reactions that modify the relevant orbitals that should be included in the active space along the evolution of a reaction coordinate are more difficult scenarios that could be addressed. In addition, the current algorithm will allow the systematic investigation of the factors that can potentially mess up the active space along a set of sampled geometries, such as the size of the chromophore, the nature of its chemical substituents, the presence and description of environments, and the sampling approach, among others.

Despite in this Chapter the presented algorithm has only been applied to the study of some organic molecules, it is expected that it works with other kinds of molecules, for example transition metal complexes, as the usage of a consistent active space for transition metal complexes along a specific potential-energy surface has already been evidenced in the literature.<sup>278</sup> The reason is that the efficacy of the algorithm mainly depends on a careful selection of the reference active space, as evidenced in the present work in the case of thymine, for which a different active space from that reported in the literature<sup>281,282</sup> needed to be used. Thus, the method can be used in a black-box manner once a suitable reference active space for the system under study has been selected. The current implementation

---

can also be adapted to be used in excited-state MD codes. If the electronic structure is described by a CAS-based method along the excited-state dynamics, it is important to keep the same active space along the trajectories to avoid discontinuities on the potential-energy surfaces and on the electronic properties of the chromophore. This can be achieved by comparing the orbitals of the active space of the current time step with those of the previous time step, and perform orbital swapping based on the analysis of the MO overlap matrix when required.

## Chapter 5

# The Permeation Mechanism of Cisplatin Through a Dioleoylphosphocholine Bilayer

Lorena Ruano<sup>1</sup>, Gustavo Cárdenas<sup>1</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

This chapter has been published in:

*ChemPhysChem* 2021, 22, 1251.

DOI: 10.1002/cphc.202100059

*The investigation of the intermolecular interactions between platinum-based anticancer drugs and lipid bilayers is of special relevance to unveil the mechanisms involved in different steps of the anticancer mode of action of these drugs. This Chapter presents the simulation of the permeation of cisplatin through a model membrane composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine lipids by means of umbrella sampling classical MD simulations. The initial physisorption of cisplatin into the polar region of the lipid membrane is controlled by long-range electrostatic interactions with the choline groups in a first step and, in a second step, by long-range electrostatic and hydrogen bond interactions with the phosphate groups.*

---

*The second half of the permeation pathway, in which cisplatin diffuses through the nonpolar region of the bilayer, is characterized by the drop of the interactions with the polar heads and the rise of attractive interactions with the non-polar tails, which are dominated by van der Waals contributions. The permeation free-energy profile is explained by a complex balance between the drug/lipid interactions and the energy and entropy contributions associated with the dehydration of the drug along the permeation pathway and with the decrease and increase of the membrane ordering along the first and second half of the mechanism, respectively.*

## 5.1 Introduction

The biological activity of cisplatin was accidentally discovered in 1965 while investigating the role of electromagnetic radiation in bacterial cell division.<sup>293</sup> Subsequent tests in mice bearing sarcoma and leukemia showed remarkable tumor regression after administration of cisplatin and other platinum complexes.<sup>294</sup> The first clinical tests in patients were conducted in 1971 and just seven years later cisplatin was approved by the U.S. Food and Drug Administration.<sup>11</sup> Since then, an impressive amount of experimental<sup>16,295–300</sup> and theoretical<sup>8,49,50,301–303</sup> work has been carried out to elucidate the mechanism of action of cisplatin, which is widely employed as chemotherapeutic drug in the treatment of patients with bladder, ovarian, head and neck, lung, testicular, cervical, esophageal, breast and brain cancers.<sup>12</sup> Moreover, the success shown by cisplatin motivates the design and investigation of many other platinum-based compounds with improved biophysiological properties in the last decades.<sup>41,304–309</sup>

The first step of the mode of action of cisplatin and related compounds is the entry of the drug in the cancer cells. This occurs by passive diffusion through the cell membrane and by facilitated transporters, such as the CTR1 copper transport protein.<sup>15,298</sup> Despite its hydrophilic nature, permeability assays have shown that cisplatin is able to diffuse through the lipid bilayer due to its small size and neutral charge.<sup>310</sup> Once in the cytosol, cisplatin is activated by the hydrolysis of the Pt Cl bonds, in which one or two chloride ligands are replaced by water to form the mono-aqua  $[\text{Pt}(\text{NH}_3)_2(\text{OH}_2)\text{Cl}]^+$  or the diaqua  $[\text{Pt}(\text{NH}_3)_2(\text{OH}_2)_2]^{2+}$  platinum complexes.<sup>311</sup> The dissociation of the chloride ligands is favored inside the cell due to the low chloride ion concentration, which is around 13 times

---

lower than in the extracellular fluid.<sup>310</sup>

After hydrolysis, the cationic platinum species are able to enter the cell nucleus and react with nucleophilic molecules, e. g., DNA strands. Specifically, the platinum atom of the hydrolyzed drug undergoes a nucleophilic attack by the N7 atoms of the purine bases, especially guanine, to form different types of DNA lesions, including monoadducts, intrastrand crosslinks and interstrand crosslinks.<sup>11,16</sup> The most common DNA damage is given by the formation of 1,2-d(GpG) intrastrand crosslinks and, to a lesser extent, 1,2-d(ApG) intrastrand crosslinks, where cisplatin binds to two adjacent guanine nucleobases or to an adenine-guanine stacked pair, respectively. The formation of adducts between DNA and cisplatin-or other platinum complexes-induces structural distortions in the DNA helix. These helical alterations are recognized by several cellular proteins, which can initiate different complex cellular processes, including programmed cell death or apoptosis.<sup>11,16,17</sup>

One of the key steps in the mode of action of cisplatin is the transport of the drug across the membrane of the tumor cells. In fact, one possible mechanism by which cancer cells develop resistance to the chemotherapy treatment is a reduced accumulation of cisplatin, in which different cell-membrane processes are involved.<sup>16,310,312</sup> In some cancer cells, the membrane-related resistance pathways are more important than the DNA-related ones and may represent 90% of total resistance.<sup>17</sup> A reduced drug accumulation may be caused either by a decrease in drug uptake or by an increase in drug efflux. The passive permeation of the drug through the bilayer depends on several factors. For example, MD simulations have evidenced that the lipid composition, the cholesterol content and the curvature of the membrane drastically affect the permeation of cisplatin.<sup>49,50</sup> Membrane fluidity also plays a relevant role in the passive diffusion mechanism; it has been suggested that a decrease in membrane fluidity hampers the entry of cisplatin into the cells.<sup>313</sup> The fluidity of the bilayers depends, among other factors, on the insertion of molecules into the bilayer and the interactions between those molecules and the lipids.<sup>298</sup> Atomic force microscopy experiments have revealed that cisplatin-encapsulated liposomes are significantly stiffer and more stable than cisplatin-free liposomes.<sup>314</sup> Therefore, it could be hypothesized that the presence of cisplatin in the membrane might affect its own diffusion and uptake efficacy.

The interaction between cisplatin and membrane rafts, which can induce changes in the biophysical properties of the membrane, is not only involved in resistance mechanisms

---

but also in different apoptotic routes.<sup>298,315</sup> Although it was generally assumed that cisplatin-induced apoptosis occurs mainly through DNA adduct formation, the consideration exclusively of these DNA-mediated pathways is not enough to explain the toxicity exerted by cisplatin,<sup>315</sup> and other mechanisms have to be invoked, including those where the lipid membrane plays a central role. Drug-lipid interactions are also relevant in the liposomal formulations of the drugs aimed to enhance the control over the delivery process and to reduce side effects and resistance of the anticancer compounds.<sup>13</sup> The encapsulation of drugs inside vesicular lipid bilayers facilitates the transport of both hydrophilic and hydrophobic species, which are integrated in the polar and in the lipophilic region of the vesicle, respectively. The liposomal formulation of cisplatin, called lipoplatin, has been successfully developed and administered in clinical trials showing lower side effects than cisplatin.<sup>316</sup> One possible mechanism for the delivery of the encapsulated cisplatin is the fusion of the membrane of the transport vesicle with the membrane of the cancer cell followed by the release of the drug. The efficiency of this process is largely regulated by the interactions between cisplatin and the lipids of both the carrier and the target cell.

Since cisplatin/lipid interactions play a decisive role in the initiation of apoptotic routes, resistance mechanisms and delivery of the drug from liposomal carriers, the rigorous characterization of these interactions is a necessary step towards a comprehensive understanding of the role played by lipid membranes in the mode of action of cisplatin and related species. In the present study, we unveil the intermolecular interactions that regulate the integration of cisplatin inside a model membrane composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipids by means of umbrella sampling classical MD simulations. Moreover, we unveil the role played by the dehydration process of the drug and the change in the ordering of the bilayer in the permeation process. Although it has been recently stated that a single-component DOPC membrane does not reproduce with accuracy the features of cisplatin permeation into more realistic cell membranes,<sup>49</sup> a DOPC membrane was chosen as a model because phosphatidylcholine is one of the major phospholipid components of the plasma membrane and of the membrane of endoplasmic reticulum, Golgi apparatus, mitochondria, endosomes and lysosomes.<sup>317</sup> Due to the biological relevance of phosphatidylcholine lipids, DOPC membranes are widely employed as model in experimental and computational investigations.<sup>298</sup> However, despite the use of DOPC in a large number of publications,

---

including theoretical modelling of the diffusion of cisplatin through the membrane,<sup>13,318</sup> a detailed analysis of cisplatin/membrane interactions has never been performed even for such a simplified membrane model. Therefore, the investigation presented here is an inescapable step before tackling an analysis of the interactions between cisplatin and more complex bilayers. In addition, the use of simplified models facilitates the evaluation of the theoretical methods employed in the simulations, e. g., the accuracy of the force field.

## 5.2 Computational Details

The lipid bilayer formed by 64 molecules of DOPC per layer with a 25 Å of water thickness on each side was built with the help of the CHARMM-GUI website.<sup>239</sup> Furthermore, a concentration of 0.15 M of KCl was added to reproduce the physiological concentration of this salt. Then, cisplatin was manually placed at a distance of 32 Å from the centre of mass of the membrane using the tleap module of AmberTools19,<sup>132</sup> resulting in a system of 35451 atoms. The lipid bilayer was described by the Lipid17 force field, which is an update of the previously developed Lipid11<sup>134</sup> and Lipid14<sup>135</sup> force fields for lipids. Water molecules were described by the TIP3P model<sup>137</sup> and the K<sup>+</sup> and Cl<sup>-</sup> ions by suitable Amber parameters.<sup>319</sup> The bond and angle parameters of cisplatin were obtained from the Cartesian Hessian matrix through the Seminario method<sup>220</sup> employing the MCPB.py module<sup>320</sup> of AmberTools19.<sup>132</sup> The dihedral potentials of cisplatin were neglected because metal-ligand torsion barriers are usually below the thermal energy.<sup>321</sup> Bonded parameters for cisplatin had been previously parameterized based on quantum dynamics trajectories.<sup>318</sup> However, we have chosen to develop new parameters using a simpler methodology that can be easily applied to future research of new metal complexes, allowing a direct comparison with the results presented here. The atomic charges and Lennard-Jones parameters were taken from previous simulations.<sup>322</sup> All the parameters for cisplatin are listed in Table B.1 of Appendix B.3. All the simulations described below have been evolved by the pmemd CUDA implementation<sup>323</sup> of the AMBER18 program.<sup>132</sup>

Once the system was set up, the following step was to equilibrate the structure and the density of the solvated bilayer. At first, the whole system was minimized using the steepest descent method for 5000 steps, and the conjugate gradient method for another

---

5000 steps. Then, it was gradually heated in the canonical (NVT) ensemble, employing a Langevin thermostat with a collision frequency of  $1 \text{ ps}^{-1}$ , to 300 K for 300 ps. Positional restraints were imposed on the lipid bilayer and on the cisplatin molecule during the heating process, by applying force constants of  $10 \text{ kcal}/(\text{mol}\text{\AA}^2)$  and  $5 \text{ kcal}/(\text{mol}\text{\AA}^2)$ , respectively. Once the system was at 300 K, three consecutive MD simulations of 4 ns each were performed at constant pressure (isothermal-isobaric NPT ensemble), in which the positional restraints on the lipid were gradually decreased by applying force constants of  $10 \text{ kcal}/(\text{mol}\text{\AA}^2)$ ,  $5 \text{ kcal}/(\text{mol}\text{\AA}^2)$  and  $0 \text{ kcal}/(\text{mol}\text{\AA}^2)$ , correspondingly, whereas the force constant of  $5 \text{ kcal}/(\text{mol}\text{\AA}^2)$  applied on the cisplatin molecule was kept throughout the three NPT simulations. The Berendsen barostat with a pressure relaxation time of 1 ps was employed to maintain the pressure around 1 bar. Finally, a 100 ns production simulation was performed in the NPT ensemble where the restrain force constant of the cisplatin was maintained at  $5 \text{ kcal}/(\text{mol}\text{\AA}^2)$ . For all the steps of this protocol the Particle Mesh Ewald method<sup>226</sup> was used to calculate the Coulomb interactions, where the direct-space sum was limited to a cutoff of 10 Å. For the computation of van der Waals interactions the same cutoff was applied. In addition, bond distances involving H atoms were restrained by using the SHAKE algorithm<sup>227</sup> and a time step of 2 fs was employed.

After the equilibration of the system, the permeation of cisplatin through the DOPC lipid bilayer was simulated by umbrella sampling MD. The reaction coordinate was defined as the distance between the centre of mass of cisplatin and the centre of mass of the membrane along the z axis (normal to the membrane), as schematically shown in Figure 5.1a. The initial value of the reaction coordinate was 32.0 Å. This distance was divided into 65 windows separated by 0.5 Å. A MD simulation of 15 ns was carried out at constant pressure (NPT) on each of the windows applying a harmonic bias potential with a force constant of  $2.5 \text{ kcal}/(\text{mol}\text{\AA}^2)$  on the reaction coordinate. Therefore, a total time of 975 ns was run to simulate the permeation of cisplatin. These simulations were performed sequentially, i.e., the initial conditions (geometry and velocities) of a window were selected from the last snapshot of the previous window. The same technical parameters employed for the 100 ns production described above were used in the simulation of each window. In order to investigate the convergence of the results the umbrella sampling simulation was extended from the middle of the membrane through the lower leaflet

---

by running additional 65 umbrella sampling windows for 15 ns each. The free-energy profile was obtained by means of the WHAM approach.<sup>157,158</sup> The energy-decomposition, interaction-energy, hydrogen bonding, radial distribution function and SCD analyses were performed using the cpptraj module<sup>197</sup> of the AmberTools19 suite.<sup>132</sup> Visual Molecular Dynamics<sup>236</sup> was used to visualize the trajectories.

## 5.3 Results and Discussion

### 5.3.1 Equilibration and Umbrella Sampling Setup

The diffusion of cisplatin through a lipid DOPC bilayer, schematically represented in Figure 5.1a, has been simulated by umbrella sampling MD simulations. This enhanced sampling approach is widely employed for modelling the slow diffusion process of small molecules across membranes<sup>49,50,73,324,325</sup> because it allows an efficient sampling along the permeation process, which is usually very hard to simulate by conventional dynamics simulations. As first step, a conventional MD simulation has been run for 100 ns to equilibrate the structure and the density of the solvated membrane and to evaluate its stability. The electron density through the lipid bilayer was the first property to be computed to characterize the membrane structure, and it is plotted in Figure 5.1b. The two peaks of the profile that appear at around 18 and 18 Å in the z axis indicate the position of the polar head groups of the lipid chains, which present heteroatoms (O, N and P) with a relatively large number of electrons, while the valley at 0 Å corresponds to the centre of the bilayer, where the electron density is lower. As can be seen in Figure 5.1b, the electron densities computed for the first 10 ns of the simulation and for the entire simulation time (100 ns) are virtually the same. This indicates that the lipid molecules of the bilayer do not suffer important diffusion processes along the simulation and that the bilayer is well equilibrated.

The area per lipid, which is defined as the average area that a single lipid molecule occupies on the interface, is the second property analysed to evaluate the stability of the system. Figure 1C shows that, after a strong drop during the first 5 ns, the area per lipid oscillates around a constant value of 69.6 Å<sup>2</sup> along the 100 ns simulation in good agreement with the experimentally determined value of 67.4 Å<sup>2</sup>.<sup>326</sup> In addition, the area

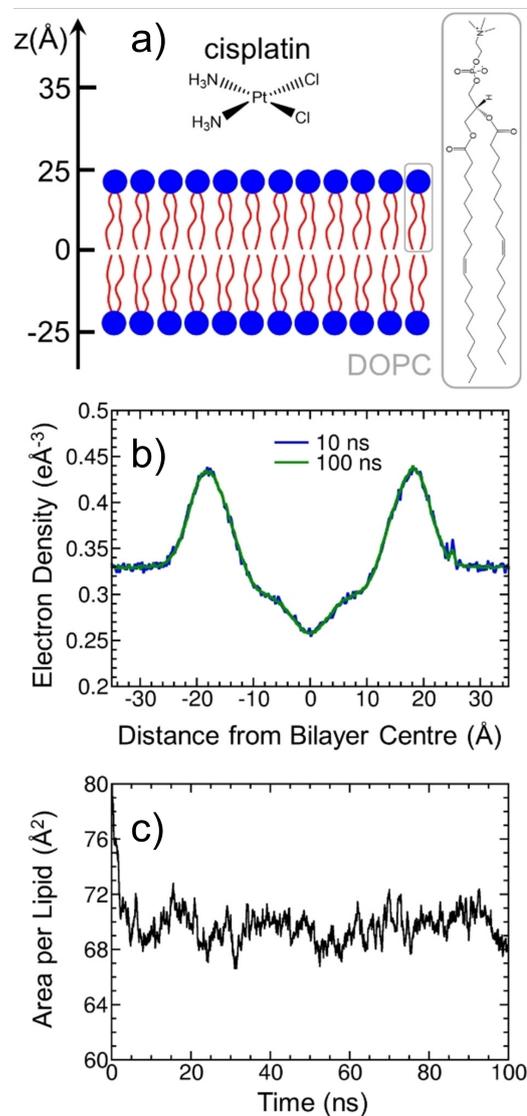


Figure 5.1: Analysis of the equilibration process of the DOPC membrane. a) Schematic representation of cisplatin and the DOPC membrane. Several reference values along the z axis are shown. b) Electron density profile through the membrane computed for the first 10 ns (blue) and for 100 ns (green) during the equilibration MD simulation. c) Time evolution of the area per lipid along the equilibration.

---

compressibility modulus was computed from the average area per lipid and its variance (see reference<sup>327</sup> for more details) providing a value of 156.5 dyn/cm. Different experimental values have been reported in the literature ranging from 188 to 300 dyn/cm,<sup>328-330</sup> showing the difficulty to obtain an accurate value of the area compressibility modulus. However, the value predicted in the present Chapter is of the same order of magnitude as the experimental findings. In summary, the small variation underwent by both the electron density and the area per lipid with time indicates that the initial structure was already close to be equilibrated. In addition, the good agreement of the simulations presented here with experimental properties validates the accuracy of the theoretical model employed. Therefore, the solvated membrane structure after the 100 ns MD simulation is a good initial structure for the subsequent umbrella sampling simulations.

The reaction coordinate was defined as the separation between the center of mass of cisplatin and the center of mass of the DOPC membrane along the z axis (see Figure 5.1). The initial value of the reaction coordinate was set to 32.0 Å, which corresponds to a separation of around 10 Å between cisplatin and the top of the surface of the bilayer. The reaction coordinate was divided into 65 windows separated by 0.5 Å such that the drug diffused from the bulk solvent across half of the membrane and reached the centre of the bilayer in the last window. A value of 2.5 kcal/(molÅ<sup>2</sup>) was chosen as force constant for the bias potential to keep the system inside each window while allowing overlap with the neighbouring windows. As shown in Figure B.1 of Appendix B.1, the reaction coordinate distributions of consecutive windows show a good overlap and, thus, validate the umbrella sampling parameters employed.

Two important factors that determine the accuracy of the calculated free energy from umbrella sampling simulations are the number of snapshots considered per window and the way they are chosen. When the reaction coordinate is modified to drive the system from one window to the next one, it is usually necessary a short equilibration process in the new window such that the system accommodates to the new value of the reaction coordinate. It is advisable to compute the free-energy profile without considering those snapshots that belong to the equilibration stage. Moreover, an accurate free-energy profile is obtained only if the simulation in each window has been run for sufficient time. These two important issues are discussed in the following analysis. Figure 2A displays the potential of mean

---

force computed for the whole simulation time (15 ns per window) and for different time intervals where the initial steps of each window were not considered; specifically, the initial 1, 2, 3, 4 and 5 ns were removed from the analysis. Note that in the profile plotted in Figure 5.2a the reactants (cisplatin in the bulk solvent) and the products (cisplatin integrated inside the bilayer) are located on the right and on the left side of the plot, respectively. Therefore, the uptake of cisplatin in the membrane occurs from right to left. As can be seen in the inset of Figure 5.2a, it is necessary to eliminate the first 3 ns of simulation to obtain a converged profile for the relevant free-energy region located around 12–24 Å, which contains the very small energy barrier (named Top in Figure 2B) at the entrance of the bilayer and the minimum (Min).

In this case, the system adapts to the new reaction coordinate value when it migrates from one window to the next one only after 3 ns are elapsed. Therefore, in the subsequent analyses the first 3 ns of each window will not be considered. Next, we analyse for how long each window must be evolved to get a converged free-energy profile. Figure 5.2b shows that convergence is achieved after running 8 ns (from 3 ns to 11 ns) of dynamics simulation per window. Therefore, it is not necessary to extend the simulations for longer time than the 15 ns per window initially run. In order to further analyse the convergence of the results, the umbrella sampling simulations were extended through the lower leaflet of the membrane and the free-energy profile along the permeation pathway through the whole bilayer was obtained. As Figure 5.2c shows, the free-energy profile is not completely symmetric with respect to the centre of the membrane, as could be expected for a symmetric bilayer as the DOPC one employed here. However, the observed asymmetry is not too pronounced and is within the error of the force field. In addition, the important features of the profile are well reproduced at both sides of the centre of the bilayer: a nearly flat region leading to a minimum located at 15 Å from the center of the membrane. In summary, the results indicate that a reasonably good convergence was achieved. Thus, the energetic analysis discussed below will be performed by considering snapshots from the time interval between 3 ns and 15 ns in each window of the simulation through the upper leaflet of the membrane.

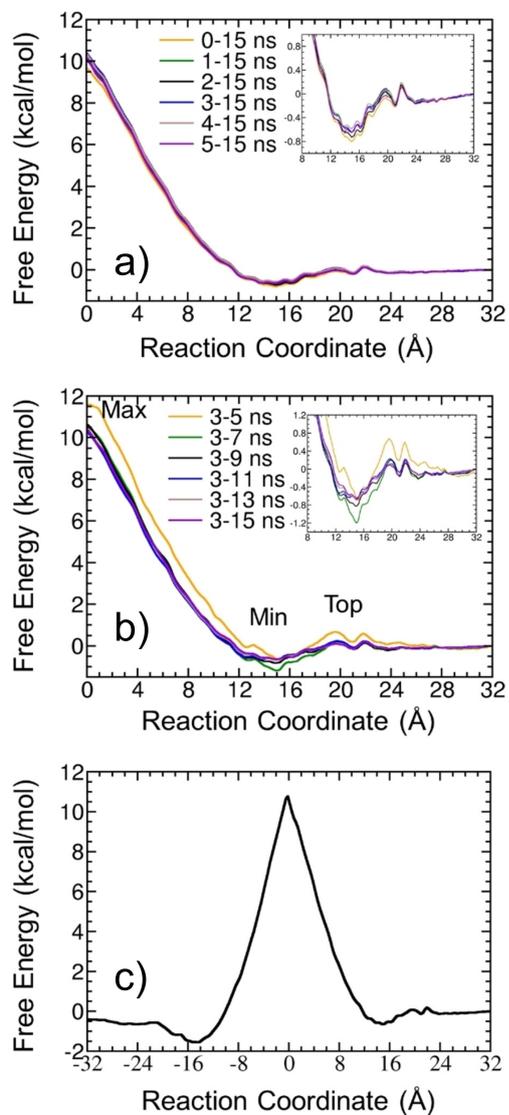


Figure 5.2: Convergence analysis of the umbrella sampling simulations. a) Potential of mean force computed by removing different time intervals from the beginning of each window. b) Potential of mean force computed for different computational times from the initial time of 3 ns. c) Potential of mean force computed along the two leaflets of the bilayer for reaction coordinate values from 32 to 32 Å considering the simulation time from 3 to 15 ns for each umbrella sampling window.

---

### 5.3.2 Characterization of the Intermolecular Interactions

The potential of mean force presents one energy minimum and two energy maxima-called Min, Max and Top in Figure 2B-that will be analysed in detail in the following. When cisplatin approaches the membrane, there is an initial attraction between the drug and the bilayer which results in a free-energy minimum (Min) at 15 Å of 0.65 kcal/mol depth with respect to the energy of cisplatin located in the bulk solvent. This free-energy minimum corresponds to the physisorption of cisplatin on the polar region of the membrane. The minimum is reached after overcoming a very small free-energy barrier of around 0.2 kcal/mol extended along the free-energy curve region located between reaction coordinate values of 19 and 23 Å. However, it is not clear whether this slight energy increase corresponds to an actual energy barrier or whether it is noise whose fluctuations lie within the error of the method, especially considering that it is not present in the left part of the profile in Figure 5.2c. Therefore, this point will be referred to as top surface region (Top) in what follows, due to the position of cisplatin, which is situated at the top of the bilayer in correspondence with this point. In any case, the available thermal energy is sufficient to drive the system through this region of the free-energy surface and accommodate the drug inside the free-energy minimum.

If cisplatin continues the diffusion pathway towards the centre of the bilayer the free energy steadily increases until achieving its absolute maximum (Max) in the middle of the membrane, which is 10.4 kcal/mol higher than the energy of the global minimum Min. This energy barrier agrees very well with previous dynamic simulations, where an energy barrier of around 10 kcal/mol was obtained.<sup>49,318</sup> However, the simulations presented here predict a free-energy minimum of 0.65 kcal/mol at the entrance of the polar region of the bilayer which was not found by previous simulations, where the permeation of cisplatin occurs through an uphill pathway from the solvent to the center of the membrane.<sup>49,318</sup> This difference could be a consequence of the slightly different parameters employed in the simulations, e. g., different simulation times, different criteria for free-energy convergence in the Weighted Histogram Analysis Method (WHAM) calculation, or different intramolecular force field parameters, among others. Nevertheless, it is hard to conclude whether the free-energy profile should present a minimum or not since both types of profile have been

---

previously found in umbrella-sampling simulations for different systems.<sup>331</sup>

The present simulations are also in good agreement with experimental findings. Specifically, the permeability coefficient of cisplatin travelling through DOPC vesicles obtained by kinetic measurements was  $1.1 \times 10^{-8} \text{ ms}^{-1}$  at a Cl concentration of 0.15 M,<sup>310</sup> while the permeability coefficient computed from the Z-position dependent diffusion coefficient of the permeating drug (Figure B.2 in Appendix B.2 and reference 325 for more details) extracted from our simulations is  $3.6 \times 10^{-7} \text{ ms}^{-1}$ . Once the results from the simulations have been validated with previous calculations and experiments, the behavior of cisplatin inside the DOPC bilayer and the intermolecular interactions that lead to the permeation of the drugs will be analyzed below.

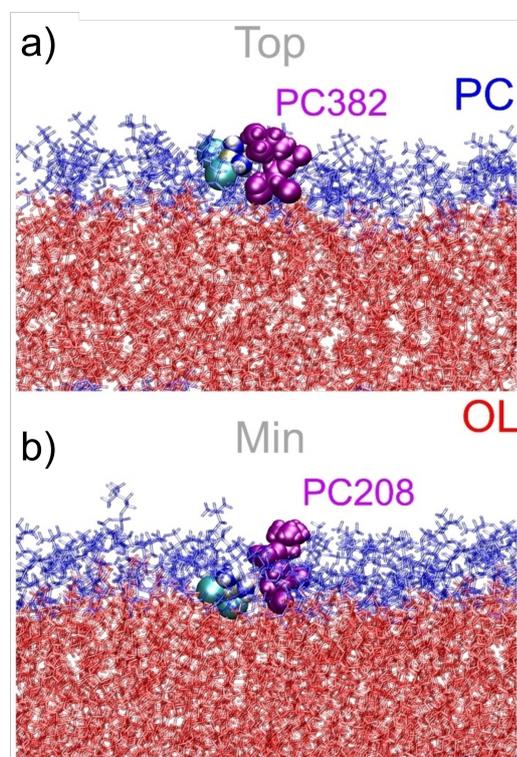


Figure 5.3: Representative snapshot of cisplatin interacting with the residue PC382 at Top (a) and with the residue PC208 at Min (b). Cisplatin is represented with van der Waals spheres with the following atom colour code: ochre for Pt, blue for N, white for H and cyan for Cl. The residues PC382 and PC208 are represented by magenta van der Waals spheres. The polar heads (PC) and the nonpolar tails (OL) of the membrane are represented by blue and red sticks, respectively.

A visual inspection of the umbrella sampling windows located at the Top and Min

---

regions reveals that the environment and the behavior of cisplatin is different in both regions. While cisplatin is located at the water/membrane interface at Top (Figure 5.3a), it is completely inside the polar region of the bilayer formed by the phosphocholine (PC) groups at Min (Figure 5.3b). This situation results in a higher mobility and a shorter interaction time with the lipids when the drug is located at the more external region Top than when it is at Min.

To illustrate the mobility of cisplatin in the different membrane regions, Figure 5.4 shows the time evolution of the number of interatomic contacts between cisplatin and the five lipid residues that present the highest interaction energy with the drug in the umbrella windows located at 20 and 15 Å, which correspond to the free energy regions Top and Min, respectively. These residues are the PC polar heads 358, 382, 307, 385 and 373 for Top, and the PC residues 343, 208, 328 and 196 and the nonpolar dioleoyl (OL) tail 207 for Min. A contact between a cisplatin atom and a lipid atom is arbitrarily considered when the separation between both atoms is lower than 5 Å. As can be seen, interatomic contacts between cisplatin and lipid residues are present for shorter time at Top than at Min. For example, if one considers that the contact between the drug and PC or OL residues is present when the number of interatomic contacts is 10 or higher, the lifetime of the intermolecular contacts is between 6 and 10 ns for Min, but only between 4 and 6 ns for Top. This indicates that the motion of the drug is restrained when it is located at the free-energy minimum Min, while it presents a higher degree of freedom when it is on the region Top.

To get more insight on the nature of the intermolecular interactions that drive the diffusion of cisplatin through the lipid bilayer the interaction potential energy between cisplatin and the membrane was decomposed into different contributions along the permeation pathway. First, we analyze the contribution of the polar PC heads and nonpolar OL tails to the total interaction energy. The different chemical groups that belong to these two regions of the lipid chains are shown in Figure 5.5a. In addition, each of these contributions (PC heads and OL tails) can be further decomposed into van der Waals and electrostatic interactions, which in the force field employed here are described by Lennard-Jones and Coulomb potentials, respectively. As can be seen in Figure 5.5b,c the interaction energy between cisplatin and the PC head groups dominates along the entire reaction coordinate

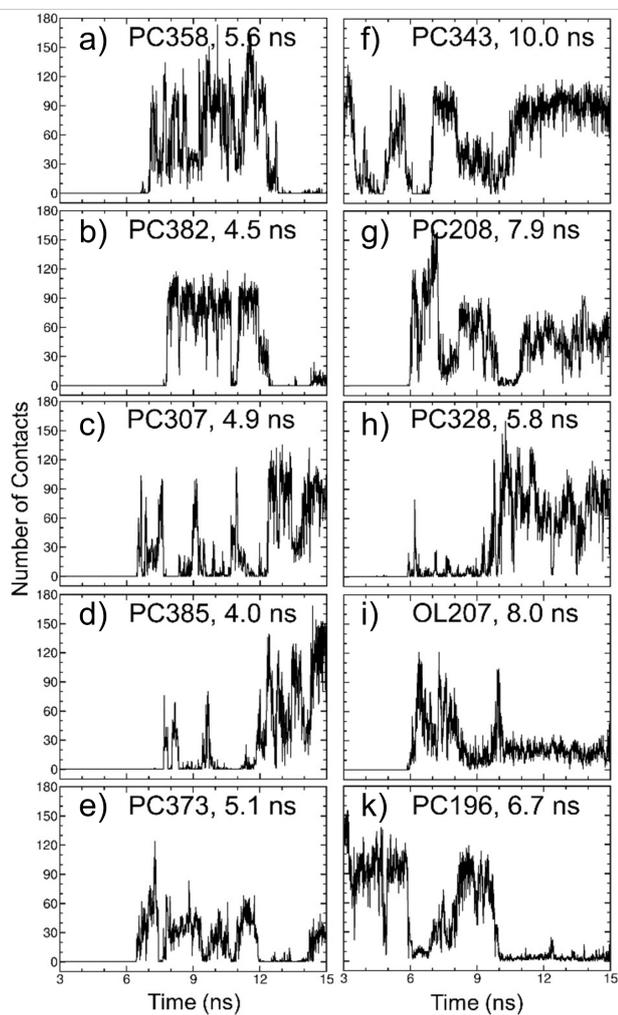


Figure 5.4: Time evolution of the number of interatomic contacts between cisplatin and the five residues that present the highest interaction energy with the drug at Min (a-e) and Top (f-j), at the umbrella-sampling windows located at 15 and 20 Å, respectively. The lifetime of the interaction with each of the residues is also shown. It is assumed that the interaction is present when the number of interatomic contacts is 10 or higher.

---

over the interaction between cisplatin and the OL tails, except in the region of the absolute energy maximum (Max), where the OL contribution is more important than the PC one.

The cisplatin/PC interaction energy steadily increases (in absolute value) along the permeation pathway until the drug reaches the minimum (Min), where the attractive interaction between the drug and the head groups achieves a value around 35 kcal/mol. Then, along the pathway from Min to Max, the cisplatin/PC interaction becomes less attractive with values between 20 and 30 kcal/mol in regions around reaction coordinate values of 10 Å, and between 10 and -5 kcal/mol at the middle of the bilayer (Max). Figure 5.5b also shows that the interaction between the drug and the polar heads is clearly dominated by electrostatic interactions, which represent 85% of the cisplatin/PC interaction energy (as average along the entire permeation pathway), while the van der Waals contribution represents only 15%.

The value of the cisplatin/OL interaction attractive energy steadily increases when the drug diffuses from the bulk solvent to the middle of the membrane, where it reaches a value of around 15 kcal/mol. Contrary to the cisplatin/PC energy, the most important contribution to the interaction between cisplatin and the nonpolar tails is given by the van der Waals interactions, especially on the region between Top and Max where the van der Waals interactions constitute 73% of the total energy. Since the cisplatin/OL attraction steadily increases along the diffusion pathway, but the cisplatin/PC attraction first increases until Min, and then decreases from Min to Max, the first half of the permeation pathway is regulated by electrostatic interactions, while in the second half of the pathway the van der Waals interactions gain relevance. Specifically, the van der Waals interaction represents only 27% of the total interaction energy at the umbrella sampling window located at 15 Å (corresponding to Min), while it is 52% of the total energy at the window located at 0 Å (corresponding to Max).

Since the PC heads are formed by different chemical groups, namely choline, phosphate, and glycerol groups (see Figure 5.5a), it is interesting to evaluate the contribution of each of these moieties to the cisplatin/PC interaction energy along the reaction coordinate and, especially, around the Top-Min free-energy curve region, where the initial physisorption process occurs. Such an energy decomposition is displayed in Figure 5.6. As can be seen, the glycerol groups play an almost irrelevant role along the entire reaction pathway in

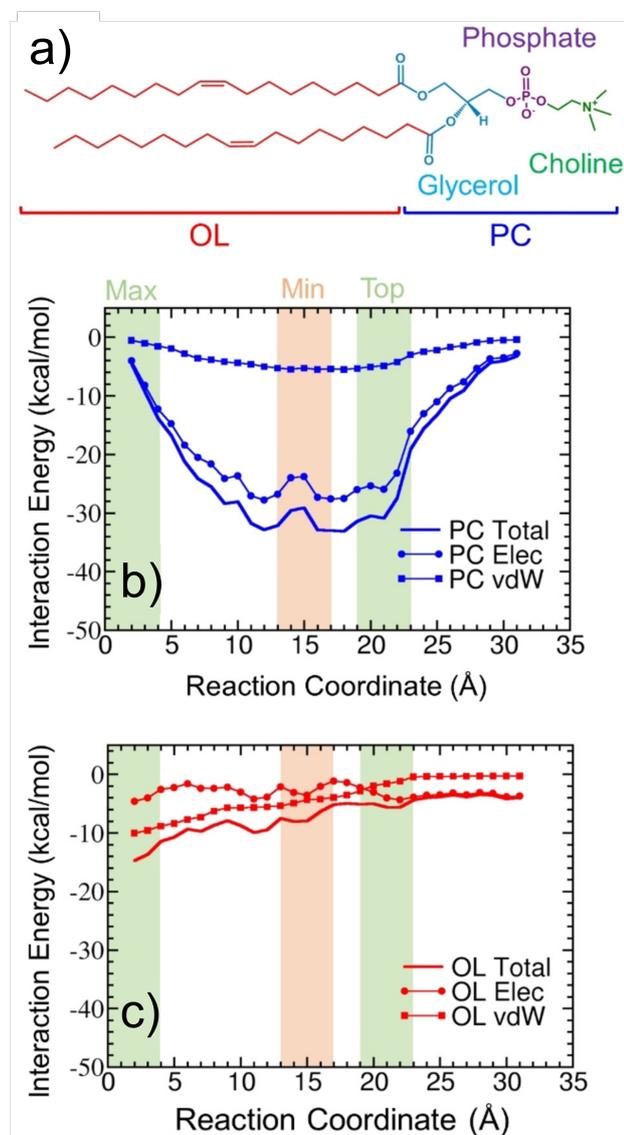


Figure 5.5: Contribution of phosphocholine (PC) head groups and dioleoyl (OL) tails to the cisplatin/membrane interaction energy along the reaction coordinate. a) Different chemical groups of the DOPC lipid chains. b) Interaction energy between cisplatin and the PC groups, and decomposition into electrostatic (Elec) and van der Waals (vdW) contributions. c) Interaction energy between cisplatin and the OL groups, and decomposition into electrostatic (Elec) and van der Waals (vdW) contributions.

---

comparison to the choline and phosphate groups. For these last two groups, the electrostatic interactions with cisplatin largely surpass the van der Waals ones, as was already evidenced when discussing the interaction energy of the entire PC head (Figure 5.5b). The interaction with the phosphate units is slightly repulsive before entering the DOPC membrane (reaction coordinate between 23 and 27 Å), while the attraction with the choline groups counteracts the repulsion with the phosphate units, stabilizing the system. Then, when the drug is physisorbed on the Top region the attraction with the choline residues is maximum and the interaction with the phosphates is also attractive but smaller. From Top the cisplatin/choline attraction steadily decreases until the interaction becomes repulsive at Min, while the cisplatin/phosphate attraction steadily increases until the drug reaches the Min region. At the minimum (Min), the attractive interactions between cisplatin and the phosphate groups clearly dominate the system. Therefore, it can be stated that the electrostatic attractive interactions between cisplatin and choline drive the initial approach of the drug to the bilayer and then, the electrostatic attraction with the phosphate moieties is responsible for the deeper integration of cisplatin in the free-energy minimum. From Min to Max, the attraction exerted by the phosphate groups is partially compensated by the repulsion with the choline units, while the van der Waals interactions with the nonpolar tails gain importance, as seen in Figure 5.5b.

Finally, the large number of heteroatoms present in the PC heads (see Figure 5.5a) could lead to the formation of hydrogen bonds with cisplatin. Specifically, the amino groups of cisplatin (see Figure 5.1a) could participate as hydrogen donors and the oxygen atoms of phosphate and glycerol residues as hydrogen acceptors. The positive charge of the nitrogen atom of choline precludes the participation of this atom in hydrogen bond interactions. The occurrence of hydrogen bonding with the glycerol and phosphate moieties defined as the percentage of the simulation time that a hydrogen bond is present was analyzed along the permeation pathway. It was assumed that a hydrogen bond is formed when the distance between the hydrogen donor and the hydrogen acceptor is shorter than 3.0 Å, and the angle formed by the hydrogen donor, hydrogen, and hydrogen acceptor is larger than 135°, which are arbitrary criteria commonly used in the literature.<sup>332,333</sup> Figure 5.6d shows that the presence of hydrogen bonding with the phosphate groups is relevant during the initial physisorption of the drug around the Top and Min regions, while the

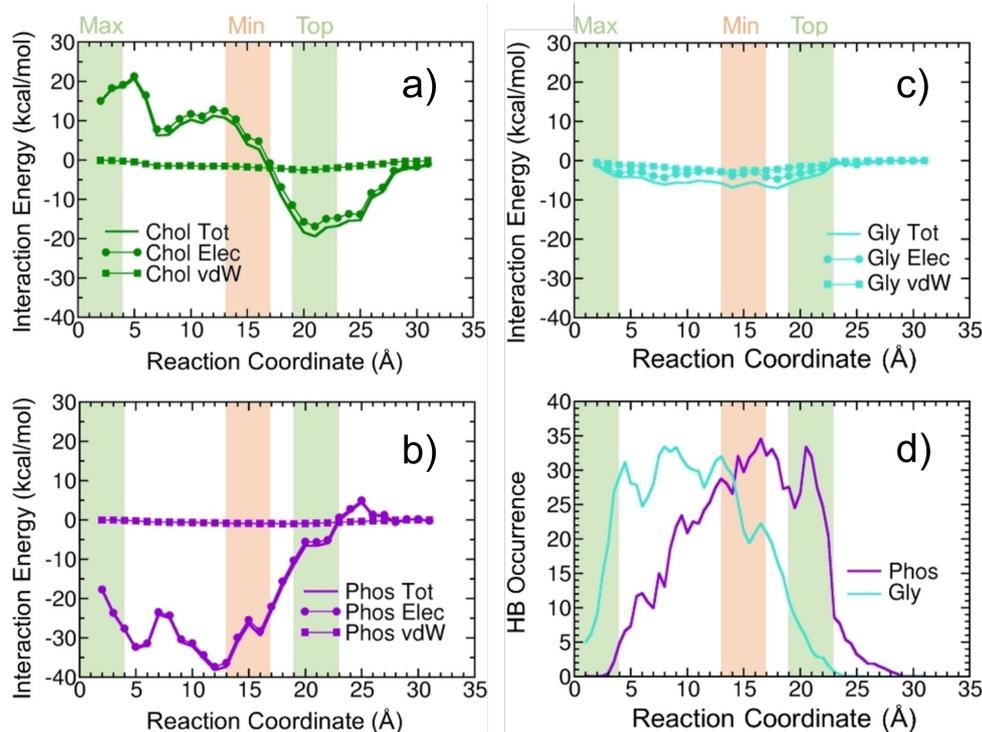


Figure 5.6: Decomposition of the interaction energy between cisplatin and phosphocholine (PC) groups. Interaction energy between cisplatin and the choline (a), phosphate (b), and glycerol (c) groups, together with their van der Waals and electrostatic energies decomposition, along the permeation pathway. (d) Hydrogen bond (HB) occurrence between cisplatin and the phosphate and glycerol groups along the permeation pathway.

occurrence of hydrogen bonding with glycerol is important at a later stage between Min and Max. However, the weak electrostatic interactions observed in Figure 5.6c indicates that, although the presence of hydrogen bonds between cisplatin and glycerol moieties is relatively high, these hydrogen bonds are not strong and, therefore, do not play a relevant role in the transport of cisplatin through the membrane.

### 5.3.3 The Role of Water Dehydration and Membrane Rigidity

The intermolecular interactions of cisplatin with the head groups and the non-polar tails of the DOPC membrane shown in Figure 5B,C are attractive along the entire permeation pathway and, therefore, do not explain the free-energy barrier of 10.4 kcal/mol found in the middle of the bilayer. The attractive cisplatin/membrane interactions are expected to be compensated by unfavorable desolvation energy and/or entropy contributions. It

---

has been previously discussed that the integration of small drugs and metal chlorides from the aqueous solvent to the polar region of the membrane is regulated by the solute dehydration.<sup>334,335</sup> This partial or total loss of the solvation sphere has two opposite effects: an energy penalty due to the breaking of drug/water interactions and an entropy gain due to the disorder increase. In order to investigate this effect, the solvation of cisplatin along the permeation pathway has been analyzed. First, the solvation sphere of the drug was characterized from the Pt(cisplatin)-O(water) radial distribution function, which was computed for the umbrella sampling window whose reaction coordinate value is 32 Å, *i. e.*, when cisplatin is in the bulk solvent. As seen in Figure 5.7a, the Pt O radial distribution function presents the maximum at 5 Å and the minimum at 6 Å. The integration of this first peak (up to 6 Å) gives that the number of water molecules in the first solvation sphere is 17–18. This is in qualitative agreement with previous MD simulations that found from 12 to 28 water molecules in the solvation sphere of cisplatin, depending on the force field employed.<sup>336</sup> Afterwards, the number of water molecules within a sphere of 6 Å along the reaction coordinate has been computed. As can be evidenced from Figure 5.7b shows, cisplatin loses around 10 water molecules from its solvation sphere when it goes from the bulk solvent to the free-energy minimum. This partial desolvation is associated with a large interaction energy penalty which is compensated by the entropy gain related to the loss of water molecules and the favorable cisplatin/lipid interactions analyzed above, especially by the electrostatic interactions with the choline and phosphate moieties, resulting in a free-energy minimum of 0.65 kcal/mol. From the free-energy minimum to the centre of the bilayer the drug loses 4 additional water molecules. Contrary to the first half of the permeation pathway, the unfavorable energy related to this process is neither counteracted by the favorable interactions with the polar heads, which undergo a significant drop (see Figure 5.5b), nor by the favorable interactions with the non-polar tails, which are relatively weak (see Figure 5.5c). In addition, the entropy gain associated with the loss of water molecules is likely less important than the entropy decrease associated with the ordering increase of the bilayer when cisplatin is inside the non-polar region of the membrane. Previous kinetic and thermodynamic measurements for the permeation of chlorpromazine through different membranes have shown that the unfavorable entropy contribution to the free energy can be explained by the less disordered lipid chains found when the drug is

---

inside the bilayer.<sup>337</sup>

The order of the lipid chains was analyzed in terms of the orientation of the C H bonds with respect to the normal to the membrane by means of the so-called deuterium order parameter (SCD) of the acyl chains, a term which comes from NMR measurements where the H atoms are substituted by deuterium (D) atoms resulting in C D bonds. SCD can vary from 0 to 0.5, where large values indicate a high degree of order of the acyl chains and low values indicate low ordering.<sup>134</sup> Figure 5.7c displays the SCD values for 3 different simulation windows with reaction coordinate values of 32, 15 and 0 Å, which correspond to situations where cisplatin is in the bulk solvent, in the free-energy minimum interacting with the polar heads, and in the middle of the bilayer interacting with the acyl chains, respectively. When cisplatin goes from the aqueous solvent to the free-energy minimum the decrease of the SCD parameter shows that the ordering of the membrane is decreased upon the integration of the drug in the polar region of the bilayer. This results in a favorable entropy contribution to the total free energy. When cisplatin is transported from Min to Max, SCD is kept constant for the C2-C10 region of the acyl chains, but it increases for the C11-C16 region. This ordering increase indicates an entropy decrease that contributes unfavorably to the free energy of the process. In summary, the first half of the permeation pathway from the bulk solvent to the free-energy minimum is characterized by favorable drug/lipid interactions and unfavorable dehydration energies. In addition, entropic contributions due to the loss of water molecules from the solvation sphere of cisplatin and the membrane ordering decrease also favor the process. The second part of the permeation process presents a large free-energy barrier which can be explained by the fact that the favorable drug/lipid interactions and favorable dehydration entropy do not compensate the dehydration interaction energy penalty and the entropy decrease due to an increase of the bilayer ordering.

## 5.4 Conclusions

The intermolecular interactions between cisplatin (and other platinum-based drugs) and cell membranes play an important role on different steps of the mode of action of these anticancer drugs, including the uptake process, resistance mechanisms and initiation of

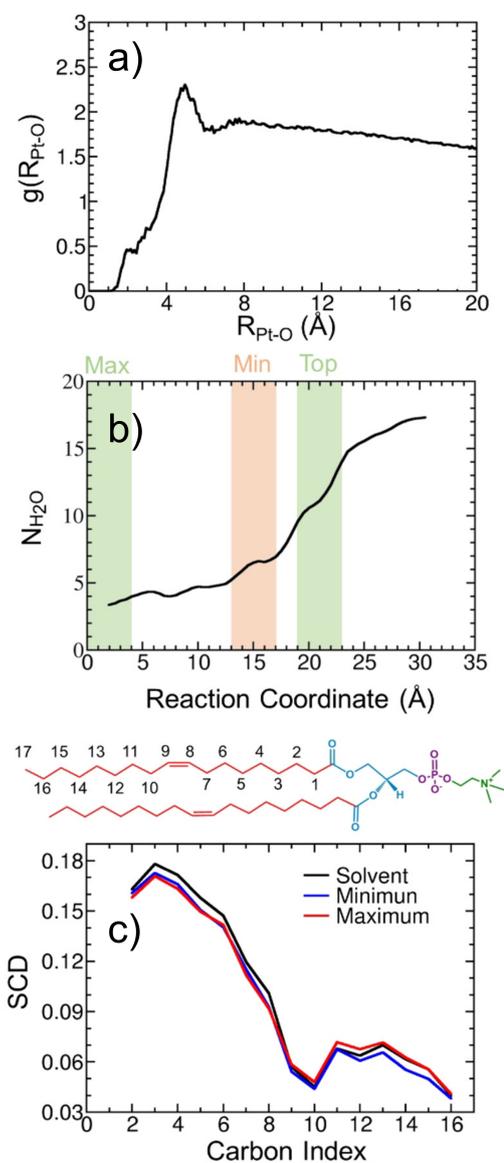


Figure 5.7: Cisplatin dehydration and membrane ordering relevance. a) Pt O radial distribution function  $g(R_{\text{Pt-O}})$  for the umbrella sampling window with reaction coordinate value of 32 Å. b) Variation of the number of water molecules within the first solvation sphere of cisplatin along the reaction coordinate. c) Deuterium order parameter (SCD) of the acyl chains for the umbrella sampling windows with reaction coordinate values of 32 Å (solvent), 15 Å (minimum) and 0 Å (maximum).

---

apoptosis. In addition, these interactions are also relevant on the delivery of the drug from liposomal carriers of encapsulated formulations. Therefore, the investigation of the nature of the intermolecular interactions present in cisplatin/membrane systems is crucial to gain knowledge that can contribute towards the understanding of the molecular mechanism behind the mode of action of platinum drugs and towards the design of new compounds with improved properties. In this work, the diffusion process of cisplatin through a DOPC lipid bilayer has been simulated by means of umbrella sampling MD to unveil the mechanism of permeation of cisplatin through the lipid membrane. Cisplatin is first weakly trapped on a free-energy minimum of 0.65 kcal/mol depth after overcoming a very small energy barrier of 0.2 kcal/mol. At this minimum, cisplatin is completely integrated inside the polar region of the bilayer and presents a lower mobility than when it is around the barrier region, which corresponds to the interface between the solvent and the membrane. This is reflected in the longer lifetime of the interactions between cisplatin and specific polar heads when the drug is trapped in the minimum. During this initial stage, where cisplatin is around the first maximum and the minimum of the free-energy surface (first half of the permeation pathway), the integration of cisplatin into the membrane is driven first by electrostatic interactions with the choline groups and then by electrostatic interactions with the phosphates. The formation of hydrogen bonds with the oxygen atoms of phosphate groups contributes to these electrostatic interactions since hydrogen bonds are present during 30% of the simulation time. Along this first half of the permeation pathway the interaction of cisplatin with the glycerol moieties of the polar heads and with the non-polar tails is insignificant. In addition, the more important interactions with the choline and phosphate units are largely dominated by the electrostatic component (around 85%), while the van der Waals contribution is only minor (around 15%). Moreover, the integration of the drug from the bulk solvent to the free-energy minimum is also regulated by the unfavorable dehydration energy associated to the loss of 10 water molecules from the solvation sphere and the favorable entropic contributions caused by the partial breaking of the solvation sphere and the decrease of the membrane ordering.

Along the second half of the permeation pathway, where the drug moves from the minimum to the global maximum located at the middle of the membrane, the cisplatin/choline repulsion and the cisplatin/phosphate attraction steadily increase and partially cancel each

---

other out. This partial cancellation induces a reduction of the total interaction energy between cisplatin and the head groups. This is accompanied by an increase of the attractive interactions between cisplatin and the non-polar tails, which are dominated by the van der Waals contribution, contrary to the interactions with the polar heads. Therefore, van der Waals interactions gain importance in the non-polar region of the membrane and represent more than 50% of the total cisplatin/membrane interaction energy. These drug/lipid interactions together with the dehydration entropy associated with the loss of 4 water molecules from the cisplatin solvation sphere favorably contribute to the free energy of the transport process. However, these favorable contributions are surpassed by the unfavorable dehydration interaction energy penalty of the drug and the entropy drop related to the membrane ordering increase, causing the occurrence of a free-energy barrier when cisplatin is located in the middle of the bilayer. This energy barrier that the system must overcome to reach the global maximum is 10.4 kcal/mol, a value which agrees very well with previous simulations<sup>49,318</sup> and experimental measurements.<sup>310</sup>

## Chapter 6

# Characterization of Cisplatin/Membrane Interactions by QM/MM Energy Decomposition Analysis.

Gustavo Cárdenas<sup>1</sup>, Álvaro Pérez-Barcia<sup>3</sup>, Marcos Mandado<sup>3</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

3 – Department of Physical Chemistry, University of Vigo, Lagoas-Marcosende s/n, ES-36310-Vigo, Galicia, Spain

This Chapter has been published in:

*Physical Chemistry Chemical Physics*, 2021, 23, 20533-20540.

DOI: 10.1039/D1CP03382D.

*In this Chapter, a quantum mechanical energy decomposition analysis scheme based on deformation electron densities is extended for the first time to a hybrid electrostatic embedding QM/MM framework. The implemented approach is applied to characterize the*

---

*interactions between cisplatin and a dioleoyl-phosphatidylcholine membrane, which play a key role in the permeation mechanism of the drug inside the cells. The interaction energy decomposition into electrostatic, induction, dispersion and Pauli repulsion contributions is performed for ensembles of geometries to account for conformational sampling. It is evidenced that the electrostatic and repulsive components are predominant in both polar and non-polar regions of the bilayer.*

## 6.1 Introduction

The permeation of drugs across cell membranes is a key biological process which largely determines the efficacy of the drugs.<sup>338,339</sup> In general, the transport of small and moderately polar molecules happens by passive diffusion through the lipid bilayer, while larger and more polar compounds enter the cell by active transport mediated by membrane proteins.<sup>340,341</sup> However, many species present intermediate size and polarity and, thus, can be uptaken by the cells through both mechanisms, as it is the case of platinum-based agents.<sup>13,298,342</sup> Platinum complexes are anticancer drugs employed in chemotherapy, whose cytotoxic mechanisms have been intensively investigated in the last decades.<sup>41,304–306,308,309</sup> Among these species, cisplatin is the most frequently used compound to treat different types of human cancers,<sup>299,343–345</sup> as has been evidenced in the previous Chapter. The interaction between the drug and lipid membranes along the passive entrance of the compound inside cancer cells has shown to be fundamental in the mode of action of cisplatin, including the development of resistance mechanisms by some cancer cells<sup>310,312</sup> and the activation of apoptotic routes.<sup>298,315</sup> Therefore, the characterization of the interactions between cisplatin and the lipids that compose the cellular membranes is of great relevance to get fundamental insight into the membrane processes involved in the biological activity of cisplatin and other platinum agents.

The computer simulation of the transport of drugs through lipid membranes is usually carried out by a combination of classical MD and enhanced-sampling approaches.<sup>73,324,346–349</sup> Specifically, the passive diffusion of cisplatin through different lipid bilayers has been investigated by umbrella sampling simulations.<sup>49,50,85,318</sup> Despite the significant mechanistic details obtained from the previous theoretical work, the nature of the drug/lipid interac-

---

tions has been barely characterized in terms of simple force field contributions,<sup>85</sup> as was evidenced in the previous Chapter. However, a more realistic description requires the use of more accurate methodologies, such as quantum mechanical interaction energy partitioning schemes based on variational or perturbational approaches.<sup>113–116</sup> These methods provide different contributions to the total interaction energy, such as Pauli, electrostatic, induction and dispersion terms, some of them not properly described by neither fixed-charge nor polarizable force fields. However, such quantum mechanical approaches have not been widely applied to the investigation of biological systems mainly due to two reasons: (i) the computational cost associated with the calculation of the interaction energy for large systems, and (ii) the need to perform the calculations for ensembles of geometries in order to consider conformational sampling. Thus, energy decomposition analyses (EDA) are usually limited to the investigation of small molecular systems<sup>350–352</sup> or relatively large systems (few tens of atoms) within a static framework.<sup>118, 353–355</sup> In those situations in which EDA had been combined with QM/MM calculations to treat large systems, the interaction energy terms were not directly derived from the QM/MM Hamiltonian but atomic charges and van der Waals parameters were introduced for the QM region.<sup>356, 357</sup> In line with these considerations, it becomes clear that a compromise between the computational cost and the level of theory implemented for the study of the interaction energies needs to be achieved, especially on the description of complex biological systems.

The present Chapter aims at characterizing the interactions that drive the permeation of cisplatin through a dioleoyl-phosphatidylcholine (DOPC) lipid membrane by means of a quantum mechanics/molecular mechanics (QM/MM) EDA approach. The main focus will be the analysis on the minimum (Min) and the maximum (Max) of the free-energy profile (represented in Figure 6.1; cf. Figures 5.2 and 5.3) previously computed by umbrella sampling MD simulations<sup>85</sup> (Chapter 5). The Min region is located at the interface between the polar heads and the non-polar tails of DOPC, while the Max region is at the center of the bilayer. The main goal of this Chapter is threefold: (i) to implement a QM/MM EDA able to treat large systems; (ii) to characterize the nature of each of the components of the cisplatin/membrane interaction energy on the Min and Max regions; and (iii) to evidence the importance of considering conformational sampling in the model.

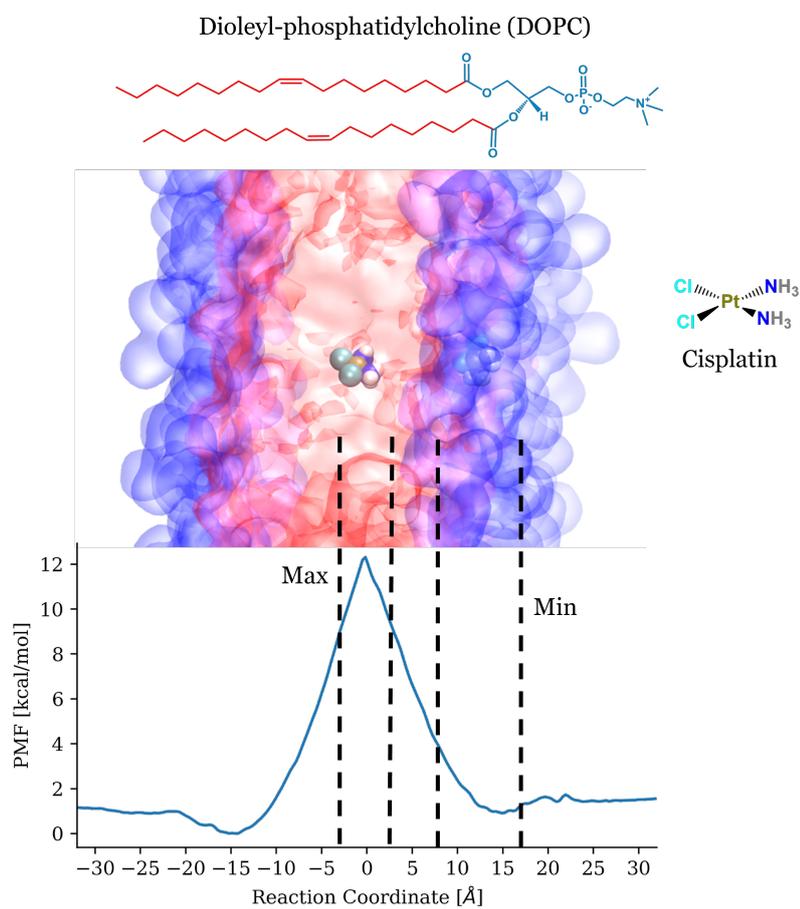


Figure 6.1: Top: Schematic representation of the cisplatin molecule embedded in a dioleoyl-phosphatidylcholine lipid membrane. The polar and non-polar regions of the membrane are represented in blue and red, respectively. Bottom: Potential of mean force (PMF) profile of the permeation of cisplatin inside the lipid membrane<sup>85</sup> displaying the two regions (Min and Max) under study in the present work.

---

## 6.2 Methodology

The interaction energy between two fragments A and B of a molecular system is defined as the difference between the energy of the AB complex and the energies of the isolated A and B fragments.

$$E_{\text{int}} = E_{\text{AB}} - (E_{\text{A}}^{\text{AB}} + E_{\text{B}}^{\text{AB}}) \quad (6.1)$$

In Equation (6.1) the superscript AB indicates that the energies of A and B are calculated considering their geometries in the complex and the basis set of the complex. The latter is introduced to correct the basis set superposition error.<sup>119,120</sup>

In this Chapter the system is divided into two subsystems, namely the cisplatin molecule on the one hand and the DOPC membrane plus the water molecules and the K<sup>+</sup> and Cl<sup>-</sup> ions on the other hand; thus, in accordance with Equation (6.1), these subsystems correspond to the A and B moieties, respectively, and the AB complex is represented by the cisplatin molecule embedded in the solvated lipid membrane. Considering this division, we adopt an EDA scheme based on deformation electron densities, initially developed and implemented for quantum mechanical methods.<sup>117,118</sup> This methodology is based on the definition of the one-electron and the exchange-correlation unperturbed and deformation densities, the latter associated with Pauli and polarization effects. Thus, the one-electron density of the complex may be written as,

$$\rho(r) = \rho_{\text{A}}(r) + \rho_{\text{B}}(r) + \Delta\rho_{\text{Pau}}(r) + \Delta\rho_{\text{pol}}(r) \quad (6.2)$$

where the first two terms in the rhs of Equation (6.2) are the unperturbed densities of the subsystems A and B and the last two terms correspond to the Pauli and polarization deformation densities. In the same way, the exchange-correlation density of the complex may also be written as,

$$\rho_{\text{XC}}(r_1, r_2) = \rho_{\text{XC,A}}(r_1, r_2) + \rho_{\text{XC,B}}(r_1, r_2) + \rho_{\text{X,AB}}(r_1, r_2) + \Delta\rho_{\text{XC}}(r_1, r_2) \quad (6.3)$$

where the first two terms represent contributions from the isolated subsystems and the last two terms from intermolecular electron exchange and polarization interactions, in this

---

order. The definitions given by Equations (6.2) and (6.3) allow for decomposing the total interaction energy into electrostatic, exchange, repulsion and polarization contributions,

$$E_{\text{int}} = E_{\text{elec}} + E_{\text{exc}} + E_{\text{rep}} + E_{\text{pol}} \quad (6.4)$$

$E_{\text{exc}}$  and  $E_{\text{rep}}$  together account for the traditional Pauli repulsion energy,  $E_{\text{Pau}}$ , whereas  $E_{\text{pol}}$  may be split into the second order induction energy,  $E_{\text{ind}}$ , and dispersion energy plus higher-order polarization terms,  $E_{\text{disp+res-pol}}$ . In what follows latter term will be referred to as dispersion energy,  $E_{\text{dis}}$ , because this is by far the dominant contribution. Thus, with the introduction of the perturbation expansion of the one-electron and exchange-correlation densities, one obtains the following partition,

$$E_{\text{int}} = E_{\text{elec}} + E_{\text{Pau}} + E_{\text{ind}} + E_{\text{dis}} \quad (6.5)$$

It should be emphasized here that the lipid membrane is composed by 128 lipid chains formed by 138 atoms each, so that in practice only a part of the lipid membrane can be treated at a quantum mechanical level. In order to account for the presence of the remaining atoms an electrostatic-embedding QM/MM computational scheme has been adopted both in the computation of the interaction energy and in the EDA, where the classical charges of the environment polarize the QM region. Therefore, not only have the polarization effects of the QM region by the MM charges been considered when calculating the interaction energy, but the MM region has also been explicitly included in the membrane subsystem and, thus, in the calculation of the energetic term of Equations 6.4 and 6.5. A QM/MM-EDA has been performed whereby each interaction energy term is still defined quantum mechanically and the electrostatic potential created by the MM region is integrated with the external potential created by the nuclear charges of the part of the membrane included in the QM region. To the knowledge of the authors, this is the first time a QM/MM-EDA is performed where one of the fragments is purely QM (cisplatin) and the other one is a QM/MM hybrid (membrane). An overview on the derivation of this EDA scheme is provided in Chapter 2, section 2.3.

---

## 6.3 Computational Details

The system under study consists of a lipid bilayer formed by two layers of 64 DOPC molecules each, a 25 Å of water thickness on each side and a concentration of 0.15 M of KCl to reproduce the physiological concentration of this salt, and the cisplatin molecule. The 194 geometries analyzed for the Max region and the 190 geometries for the Min region were sampled in an equispaced manner from the last 12 ns of two 15 ns NPT MD simulations performed with the cisplatin molecule placed at the center of a DOPC lipid membrane for the Max region, and at a distance of 15 Å from the center of mass of the lipid membrane for the Min region, respectively; in both cases a harmonic potential with a force constant of 2.5 kcal/(mol Å<sup>2</sup>) was imposed for restraining the position of cisplatin into a particular region of the bilayer. Further details on the classical MD procedure and on the computation of the potential of mean force profile by umbrella sampling simulations can be found in Chapter 5, section 5.2.

The single point QM/MM calculations for the 190 (194) selected snapshots were performed with the Gaussian16 software,<sup>188</sup> using the M062X<sup>108</sup> functional, and the LANL2DZ<sup>358,359</sup> basis set (with the corresponding effective core potential) for the Pt atom, and the 6-31G\* basis set for the C, N, H, O, Cl and P atoms of the QM regions. The point charges for the DOPC molecules in the MM region were obtained from the Lipid17 force field, an updated version of the Lipid11<sup>134</sup> and the Lipid14<sup>135</sup> force fields, those for the water molecules were extracted from the TIP3P water solvation model,<sup>137</sup> and the point charges for the K<sup>+</sup> and the Cl<sup>-</sup> from suitable Amber force field parameters.<sup>319</sup> For each geometry three single point QM/MM calculations were performed: one for the 6 DOPC lipids in the QM region with the basis functions of cisplatin (but not the nuclear charges) and the point charges of the rest of the environment in the MM region; one calculation for the cisplatin molecule in the presence of the basis functions of the 6 DOPC lipids treated by QM; and one considering the entire system that comprises the QM region (cisplatin + 6 DOPC lipids) and the point charges of the MM region. Therefore, almost 1200 QM/MM energy calculations including around 6000 functions have been performed. Therefore, the choice of the smallest QM region after attaining convergence (see below) is crucial to reduce the computational time. The selection of the 6 closest molecules (for the

---

convergence analysis, 1 to 7) of DOPC to the cisplating molecule and the generation of the input files was performed using a development version of the `MoBioTools` toolkit,<sup>79,80</sup> in a similar manner as in the case of Chapter 3, section 3.3.2. In particular, the toolkit employs the `closest` tool of the `cpptraj` module<sup>197</sup> of the `AmberTools19` package<sup>132</sup> to select the desired number of closest residues. The EDA developed and implemented by Mandado *et. al.*<sup>117,118</sup> was extended to a QM/MM framework, as explained above, and used to compute the different components of the interaction energy. The Visual Molecular Dynamics<sup>236</sup> was used to visualize the sampled geometries.

## 6.4 Results and Discussion

As a first step, considering the large size of the system investigated here it is necessary to analyze the convergence of the interaction energy and its different contributions with respect to the size of the QM region and the number of sampled geometries considered in the calculations. For the first purpose, a geometry was selected at random from a 15 ns MD run in the Min region and in the Max region, and in each case the QM/MM energy calculations and the EDA were performed increasing the number of DOPC molecules in the QM region. Moreover, in order to investigate the effect of including the point charges of the classical environment in the EDA, the energy terms were obtained without (Figure 6.2a,b) and with (Figure 6.2c,d) point charges. It can be seen that for the calculations without point charges, the dispersion, induction and Pauli contributions converge after including at least 5 DOPC molecules in the QM region. However, the total interaction energy does not converge even for 7 DOPC molecules in the QM region, as can be evidenced from Figure 6.2a,b; this is due to the fact that the electrostatic contribution assumes an oscillatory behavior between 3 and 7 DOPC molecules. On the other hand, the counterpart calculations that include the point charges that surround the QM region (Figure 6.2c,d) display a good convergence for all the interaction energy terms after including 5 DOPC molecules in the QM region. The faster convergence of the QM/MM interaction energies with respect to the pure QM results, as the QM region increases, can be rationalized by considering that in the QM/MM calculations the polarization effects due to the classical residues are being accounted for to some extent within an electrostatic embedding framework. This

---

means that when the QM region is extended, part of the polarization due to the newly added residues had been (partially) included in the smaller model as point charges. This is not the case for the pure QM computations (the cluster model) as in this case it is being considered exclusively the polarization due to the residues in the QM region, so that extending the QM region results in adding residues that had not been accounted for within the smaller model, hence the variation of the interaction energy is larger than in the QM/MM case. Similar convergence analyses have been performed in the literature in the case of reaction energies<sup>360</sup> and spectroscopic properties,<sup>361</sup> in which a faster convergence has been evidenced for the QM/MM models with respect to the pure cluster models. These results suggest that although considering the entire system quantum mechanically is computationally unfeasible, it would still be necessary to include the polarizing effect of the surrounding environment by an electrostatic-embedding QM/MM approach. Furthermore, in the case of the system under study, it represents an excellent compromise to incorporate the cisplatin molecule plus 6 DOPC molecules to the QM region, and consider the remaining atoms as point charges in the MM region.

Next, the convergence of the energies with the number of geometries has been investigated. This was done by sampling 190 geometries from a previously performed<sup>85</sup> (Chapter 5) MD simulation at the Min region, and 194 geometries from a MD simulation at the Max region. For each of the sampled geometries, the EDA was performed considering the cisplatin molecule plus six DOPC molecules in the QM region, with overall 839 atoms treated quantum mechanically. In order to ensure that 190 (194) geometries are an appropriate statistical ensemble, the convergence of the energy components as the number of averaged geometries increases was analyzed, by selecting these geometries in an equidistant manner from the 190 (194) geometries of the Min (Max) region. As can be evinced from Figure 6.2e,f, the induction and the dispersion energies are mostly converged after considering 50 geometries. However, the convergence is most difficult to achieve for the electrostatic and Pauli components. The former is converged only after considering 100 geometries, and the latter still shows a slight increase when going from 100 to 190 geometries. This highlights the importance of including conformational sampling in the theoretical model. In the following analyses, the 190 (194) geometries sampled at the Min (Max) region of the membrane will be considered.

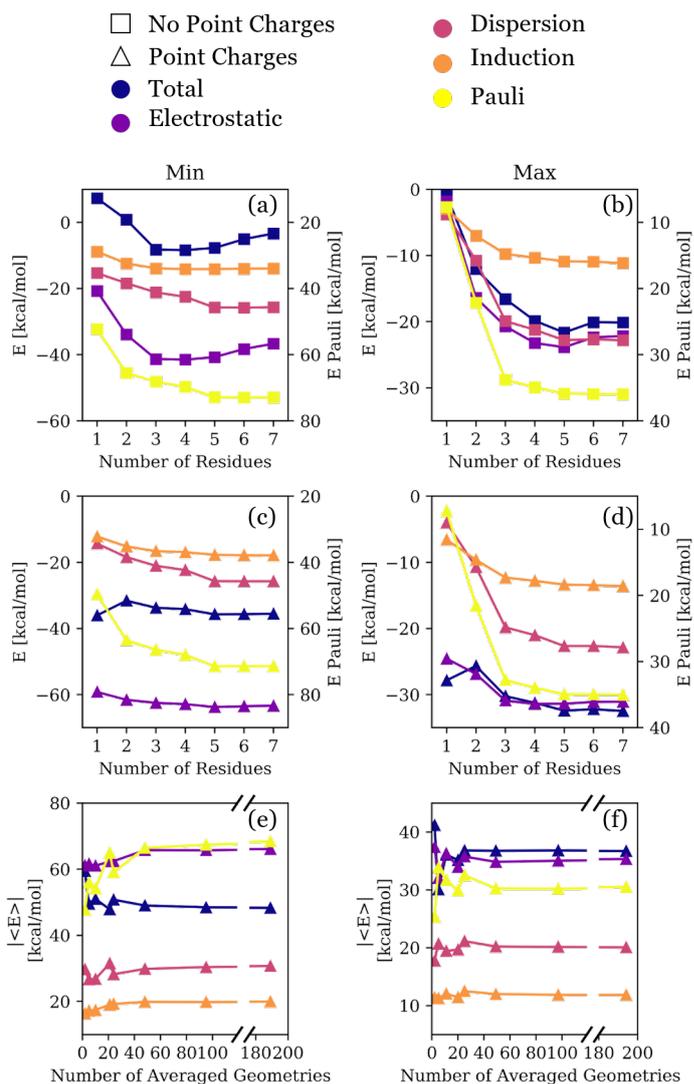


Figure 6.2: Analysis of the convergence of the interaction energy and the EDA components as the number of DOPC molecules in the QM region increases without (a Min, b Max) and with (c Min, d Max) point charges. Convergence of the absolute value of the interaction energy and the EDA components with respect to the number of (equispaced) geometries considered in the analysis (e Min, f Max).

The probability distributions of the total interaction energies and of each of the energy components that stem from the EDA considering the 190 (194) geometries at Min (Max) are displayed in Figure 6.3a,c. It is shown that the different energy components are not distributed similarly, but their distributions present different broadness. Furthermore, for both the situation at Min and at Max the broadness of the distribution is seemingly

---

associated with the magnitude of the average value, as can be seen in Figure 6.3b,d. The total energy and the different energy components present lower average values at Max with respect to the geometries at Min and, therefore, the distributions for the Min region are wider. This is in agreement with the fact that the region of the membrane corresponding to Min is a highly polar and polarizable region, and the interactions between cisplatin and the lipids are stronger than those in the non-polar region (corresponding to Max).

The energy contributions to the interaction energy can be classified into electrostatic and non-electrostatic, the latter corresponding to the sum of dispersion, induction and Pauli repulsion in the present case. In regard with this classification, it is observed that at the minimum of the free-energy profile the overall non-electrostatic interactions amount to 17.8 kcal/mol, thus resulting to be repulsive, while the electrostatic contribution is -66.1 kcal/mol (Figure 6.3b). Therefore, the electrostatic interaction is the main component of the interaction energy; this result is in agreement with the EDA performed considering a classical force field (Chapter 5), although it should be emphasized that in the latter case the non-electrostatic component resulted to be attractive, unlike the quantum mechanical analysis performed in this Chapter. Interestingly, when analyzing the averages of the electrostatic and the non-electrostatic components at the center of the membrane (Figure 6.3d), the electrostatic contribution is still predominant with -35.3 kcal/mol, whereas the non-electrostatic component (albeit being attractive, unlike the minimum) amounts to only -1.4 kcal/mol. These results are in contrast with those previously obtained with a classical force field, for which the non-electrostatic interactions represented the most relevant part of the total interaction energy in the Max region.<sup>85</sup>

All of this can be rationalized by examining the different energy components in terms of attractive (the sum of the electrostatic, dispersion and induction contributions, Figure 6.3b,d) and repulsive (Pauli) interactions. Specifically, when the relative contributions to the overall attractive energy are analyzed, it can be seen that the electrostatic contribution diminishes from 56.6% to 52.5% from the Min to the Max regions, the dispersion component increases from 26.3% to 29.8%, whereas the induction energy barely increases from 17.1% to 17.6%. These percentages place the electrostatic component as the main attractive contribution in both regions of the membrane, and the fact that the non-electrostatic component (*i.e.*, dispersion + induction + Pauli) varies when going from Min to Max is

mainly associated with the decrease in the Pauli repulsive component. These results also suggest that the Pauli repulsion energy is underestimated by the classical force field since the classical non-electrostatic energy resulted to be negative for both the Min and the Max situations, as was evidenced in Chapter 5. These discrepancies between the classical force field and the present QM/MM analysis indicate that a point-charged force field, although capable of reproducing the experimental free-energy barrier to travel across the bilayer, is not adequate to describe the nature of the interactions that govern the permeation of cisplatin through the lipid membrane.

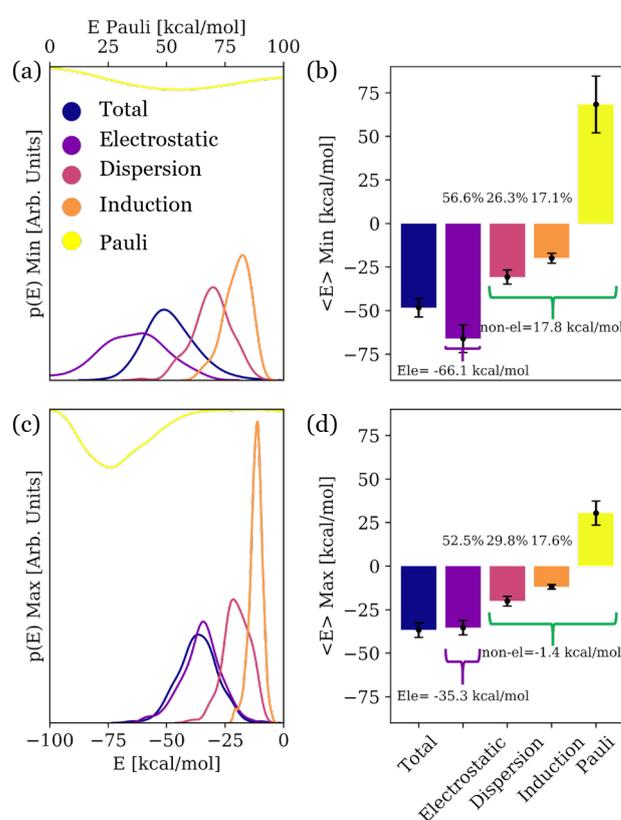


Figure 6.3: Energy distributions of the contributions to the total interaction energy stemming from the EDA on 190 (194) geometries on a) the minimum (Min) and c) the maximum (Max) of the free-energy profile. Averages and standard deviations of the distributions obtained at Min b) and Max d). The percentages are relative to the overall attractive component to the interaction energy. Electrostatic and non-electrostatic energies (sum of Pauli, induction and dispersion components) are also shown.

It is interesting to analyze which atoms and chemical groups of cisplatin and the

---

DOCP membrane are more relevant for each interaction type. Specifically, the correlation between the value of each energy component and the interatomic distances between cisplatin and the bilayer has been studied. It is assumed that these energy terms follow the classical mechanical dependency with the inverse of the interatomic distance ( $1/r^n$ ). Thus, there have been computed the linear correlation coefficient of the electrostatic, induction, dispersion and Pauli energies with respect to  $1/r$ ,  $1/r^4$ ,  $1/r^6$  and  $1/r^{12}$ , respectively. In order to assess the influence of the different functional groups of the DOPC molecule, the latter has been subdivided into four different moieties, namely, choline, phosphate, glycerol and oleyl. Figure 6.4 shows the absolute values of the three most prominent correlation coefficients corresponding to the distance between each cisplatin atom type and each atom of the DOPC moieties. Noticeably, in the region of the minimum the electrostatic and the induction components present relatively strong correlations (between 0.5 and 0.68) that are spread across the different pairwise interactions between cisplatin and the choline group, indicating a similar participation of these interactions to the overall electrostatic and inductive components. On the other hand, the Pauli component presents a strongly localized correlation (whose coefficient amounts to 0.73) upon the Pt - choline pair, in particular in regard with the distance between the Pt atom and the nitrogen of the choline group. The dispersion component of the interaction energy does not show a strong correlation with any pairwise distance, but instead the correlation coefficients are uniformly distributed throughout the four moieties of DOPC, indicating that the dispersion interaction is not dominated by any particular pairwise interaction between cisplatin and DOPC. The situation is different when linear correlations are analyzed at the center of the membrane (Max region), where cisplatin is close the oleyl non-polar tails of the DOPC molecule. It can be concluded from Figure 6.4 (bottom) that neither of the attractive components of the interaction energy present strong correlations with a specific pairwise distance, unlike in the case of the Min region, where the cisplatin molecule was surrounded by the polar heads (choline, phosphate and glycol moieties) of the DOPC molecules. This indicates that these energy components are not strongly influenced by any particular pairwise interaction when cisplatin is at the center of the bilayer.

Finally, two possible limitations of the analysis shown in Figure 6.4 related to the QM/MM setup are discussed. First, the correlation analysis was performed considering

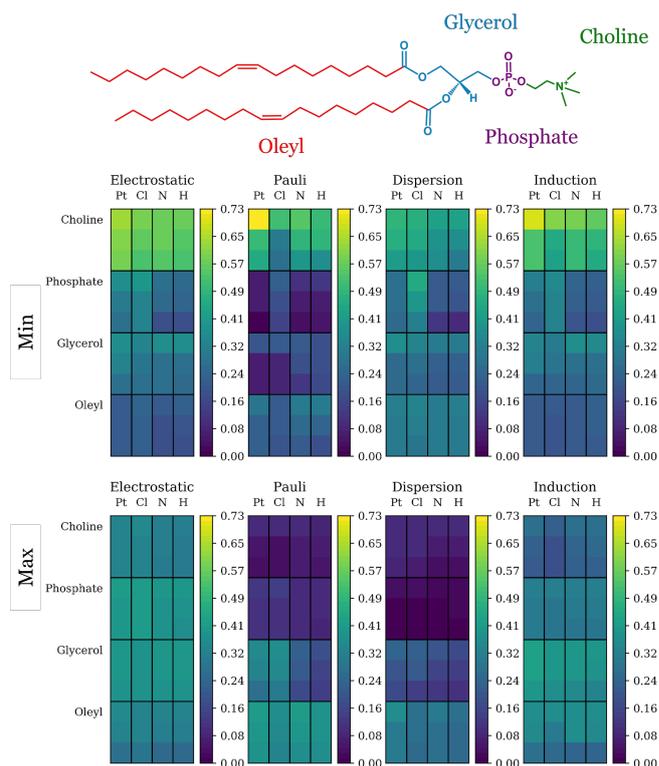


Figure 6.4: Absolute values of the linear correlation coefficients of the electrostatic, induction, dispersion and Pauli energies with  $1/r$ ,  $1/r^4$ ,  $1/r^6$  and  $1/r^{12}$ , respectively, where all the distances between each cisplatin moiety (Pt, Cl, N, H) and each atom in DOPC are considered, for the Min (top) and the Max (bottom) regions. For each moiety of DOPC (choline, phosphate, glycerol and oleyl) only the three most prominent correlation coefficients are shown.

only the interatomic distances between cisplatin and the lipid residues located in the QM region. This means that the lipids of the classical environment are not involved. However, this approximation should not introduce a significant error due to two reasons: (i) the separation between cisplatin and the classical region of the membrane is relatively large, since six lipid residues are included in the QM region, and a strong correlation of the interaction energy with the distances that involve classical atoms is not expected; (ii) the interaction between cisplatin and the classical region of the membrane is not directly computed but, instead, it is considered by an electrostatic embedding approach during both the energy calculations and the EDA analysis. Second, a convergence analysis of the correlation coefficients with the size of the QM region was not carried out because it would be necessary to run hundreds of QM/MM energy calculations and EDA analyses for

---

different QM region sizes. This large number of calculations would be computationally unfeasible. However, the convergence analysis of the interaction energy and its different contributions with respect to the size of the QM region considering two geometries (see Figure 6.2a-d) clearly indicated that the energy is well converged after including six lipid residues in the QM region. Therefore, it is expected that the correlation of the different energy terms with the inverse of the interatomic distance ( $1/r^n$ ) is also converged.

## 6.5 Conclusions

In this work, an extension of an EDA scheme based on deformation densities<sup>117,118</sup> to include a multiscale hybrid electrostatic-embedding QM/MM approach has been implemented. This methodology has been applied to the study of the interaction energy of cisplatin with a DOPC lipid bilayer, whose permeation pathway had been previously simulated by umbrella sampling MD (Chapter 5).<sup>85</sup> It has been evidenced the importance of considering a sufficiently large QM region (6 DOPC molecules plus the cisplatin drug), as well as the need of including the surrounding environment within an electrostatic-embedding QM/MM approach to properly compute the interaction energy components. Conformational sampling needs to be accounted for since wide distributions of energies are obtained even when considering sampled geometries from the same region (Min and Max regions). It has been observed that the interaction energy diminishes when moving from the polar to the non-polar region of the bilayer, a fact that is explained by the larger decrease of all the attractive energy components in comparison with the decrease in the Pauli repulsion. Contrary to the classical result,<sup>85</sup> the electrostatic component is predominant in both regions of the membrane, and the Pauli repulsive component plays a major role in determining the overall non-electrostatic component, suggesting that the repulsive component is underestimated by classical force fields. Finally, it has been observed that when cisplatin is closer to the polar heads, the electrostatic, induction and Pauli components display strong correlations with the distance between cisplatin and the choline moiety of DOPC, whereas the dispersion energy does not show a strong dependence on a particular distance. At the center of the membrane, when cisplatin is closest to the oleyl moieties, neither of the energy components displays correlations with particular groups. The methodology presented here can be

---

applied to further drug/biomolecule complexes where an accurate characterization of the interactions that govern the system is desired. Moreover, comparison with results based on molecular mechanics force fields will allow the assessment of classical methodologies to describe drug/biomolecule interactions.

## Chapter 7

# Stacking Effects on Anthraquinone/DNA Charge-Transfer Electronically Excited States

Gustavo Cárdenas<sup>1</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

This Chapter has been published in:

*Molecules*, 2020, 25, 5927.

DOI: 10.3390/molecules2524592

*The design of more efficient PSs is a matter of great importance in the field of cancer treatment by means of photodynamic therapy. One of the main processes involved in the activation of apoptosis in cancer cells is the oxidative stress on DNA once a PS is excited by light. As a consequence, it is very relevant to investigate in detail the binding modes of the chromophore with DNA, and the nature of the electronically excited states that participate*

---

*in the induction of DNA damage, for example, charge-transfer states. In this Chapter, the electronic structure of the anthraquinone PS intercalated into a double-stranded poly(dG-dC) decamer model of DNA is investigated. First, the different geometric configurations are analyzed by means of classical MD simulations. Then, the excited states for the most relevant poses of anthraquinone inside the binding pocket are computed by an electrostatic-embedding quantum mechanics/molecular mechanics approach, where anthraquinone and one of the nearby guanine residues are described quantum mechanically to take into account intermolecular charge-transfer states. The excited states are characterized as monomer, exciton, excimer, and charge-transfer states based on the analysis of the transition density matrix, and each of these contributions to the total density of states and absorption spectrum is discussed in terms of the stacking interactions. These results are relevant as they represent the footing for future studies on the reactivity of anthraquinone derivatives with DNA and give insights on possible geometrical configurations that potentially favor the oxidative stress of DNA.*

## **7.1 Introduction**

Photodynamic therapy (PDT) is nowadays a widely-employed technique to treat different types of cancer as well as some infectious diseases.<sup>5-7,362</sup> The reason for its widespread usage stems from the fact of being a non-invasive technique which allows for the induction of cell-death through apoptosis on specific target cells, *e.g.*, those present on tumor tissues.<sup>363,364</sup> PDT involves the usage of two main components, namely a PS compound and irradiation of light at a specific wavelength to promote the excitation of the PS.<sup>21</sup> The mechanism of PDT apoptosis induction depends on the nature of the PS and on the tissue where the PS accumulates.<sup>22,23</sup> In regard with the mechanism of action of the PS, it is well recognized that after its photoexcitation and population of the triplet-state manifold, it promotes oxidative stress on the surrounding molecules either via electron transfer directly to these molecules (*e.g.*, DNA bases) to produce free radicals (type I mechanism), or through energy transfer to molecular oxygen, which generates singlet oxygen that can cause damage to nearby biomolecules (type II mechanism). Many tumors develop hypoxia conditions,<sup>24</sup> where the low flux of oxygen strongly limits the use of PSs that operate via type II reactions.

---

As a consequence, there is a growing interest in the development of PSs that are able to induce damage in the absence of oxygen.<sup>25-30</sup>

It has been evidenced that, depending on its nature, the PS can accumulate in specific cell components or organelles.<sup>23</sup> In particular, mitochondria result to be an ideal target for PDT since they release pro-apoptotic factors to the cytosol following mitochondrial dysfunction due to mutations in the mitochondrial genome and rupture of the mitochondrial membrane.<sup>31-33</sup> In addition, PSs can also cause DNA damage and their interaction with the DNA double strand has been of great interest since this allows for understanding the apoptotic way induced by DNA lesion. It has been established that the mechanism of oxidative stress of DNA following the excitation of the PS depends on the binding mode of the latter, as the electronic structure of the PS can be modified in a specific manner, depending on the surrounding environment.<sup>71,75,365,366</sup> There are three different binding modes in which the PS can bind in a noncovalent manner with DNA,<sup>34,367</sup> namely, electrostatic binding, groove binding, and intercalative binding. Although the same PS could bind to DNA through more than one interaction mode,<sup>368-371</sup> it has been evidenced that the preference for a binding mode over the others can be induced by suitably modifying the substituents of the PS<sup>372,373</sup> so that the PS can be tailored to induce oxidative stress on DNA in a specific manner.

Several families of compounds have been studied and tested on PDT.<sup>34,374</sup> In particular, anthraquinone derivatives have shown promising phototoxic activity *in vitro* on human carcinogenic tissues, especially on breast cancer cells.<sup>35,375,376</sup> Several anthraquinone derivatives are known to interact with DNA through intercalation between two consecutive base pairs,<sup>36,38,39</sup> and the cleavage efficiency of DNA is strongly dependent on the substituents present on the anthraquinone scaffold.<sup>39</sup> Therefore, the investigation of the binding modes of these PSs and their influence on the nature of the electronically excited states that lead to photoreactions with DNA is of utmost importance for the design of novel anthraquinone-based phototherapeutic drugs. Molecular modeling has proven itself to be a valuable tool in this regard.<sup>40</sup> For example, MD simulations have been employed to study the binding modes and determine the corresponding binding free energies of some representatives of different PS families.<sup>52,53,377,378</sup> Moreover, MD in conjunction with hybrid QM/MM approaches have been applied to unravel in a comprehensive man-

---

ner the binding modes with DNA and the nature of the excited states that give rise to photochemotherapeutic reactivity of organic PSs, such as acetophenone,<sup>74</sup> palmatine,<sup>76</sup> methylene blue,<sup>71,72</sup> Nile red and Nile blue,<sup>75</sup> and chelerythrine.<sup>54</sup> Although several MD studies have been performed to unveil the energetics of the noncovalent binding process of anthraquinone derivatives with DNA, in particular the intercalation binding mode,<sup>51-54</sup> to the knowledge of the authors a detailed study considering the effect of the DNA surrounding environment on the electronic structure of an anthraquinone derivative has not been performed to this date. Furthermore, in the optics of tailoring more efficient anthraquinone PS derivatives, in particular molecules presenting moieties that favor specific conformations that enhance charge transfer between the PS and the DNA strand, a good place to start would be to consider the pristine anthraquinone (AQ) molecule and to analyze the nature of its electronic structure right after excitation. It is important to note that AQ is not water soluble and, thus, it is unlikely to be employed in PDT mechanisms in biological environments. However, a detailed analysis of its electronically excited states when it is interacting with DNA is important to carry out future comparisons with functionalized water-soluble anthraquinone derivatives with potentially efficient PDT mechanisms.

This Chapter presents the study of the different rotational poses assumed by the AQ molecule when intercalated between two base pairs of a solvated double-stranded poly(dG-dC) polynucleotide model, along with the influence of these different poses on the electronically excited states of AQ at the Franck-Condon region. Poly(dG-dC) has been chosen since guanine presents the lowest oxidation potential of all four DNA nucleobases,<sup>228</sup> and it has been evidenced that when employing different AQ derivatives, DNA oxidative damage occurs by photoinduced electron transfer from a guanine moiety of DNA to the photoexcited PS.<sup>379,380</sup> The exploration of the ground-state potential-energy surface of the solvated AQ-DNA complex is performed by means of classical MD sampling. Four different conformational minima are identified when analyzing the relative orientation between AQ and one of the two guanine-cytosine flanking base pairs. Then, the nature of the excited states of AQ depending on its relative orientation with respect to the guanine-cytosine base pair is investigated by means of a hybrid electrostatic-embedding QM/MM scheme, in which AQ and a guanine molecule are part of the QM region, whereas the surrounding environment is considered by a MM force field. With this approach and by performing

---

a suitable wavefunction analysis,<sup>381</sup> it is possible to characterize the different classes of excited states of the system, including those with a high electron-transfer character from the guanine moiety to the AQ molecule, which are relevant in the PDT mechanism of the PS.

## 7.2 Computational Details

The poly(dG-dC) decamer structure was constructed using the Nucleic Acid Builder (NAB) utility of the AMBER18 software.<sup>132</sup> The poly(dG-dC) polynucleotide consisted of a double strand having a 10-base guanine–cytosine sequence in each strand. AQ was non-covalently bound to the polynucleotide model by manually positioning it between the fifth and sixth guanine–cytosine base pairs (G5-C16 and C6-G15 in Figure 7.1a of the poly(dG-dC) sequence, to emulate the intercalative binding mode of AQ with DNA. The tleap module of AmberTools19<sup>132</sup> was used to solvate the AQ-DNA system with a periodic truncated octahedral water solvation box considering a maximum distance of 10 Å from any solute atom to the faces of the box, and a suitable number of Na<sup>+</sup> ions was introduced to neutralize the phosphate moieties. The polynucleotide was classically described with the OL15 force field,<sup>138</sup> whereas the bonding and the Lennard-Jones nonbonding parameters of AQ were taken from the general AMBER force field for organic molecules.<sup>131</sup> Water molecules were described by the TIP3P<sup>137</sup> solvation model and Na<sup>+</sup> ions by suitable AMBER parameters.<sup>319</sup> The geometry of AQ was optimized at the MP2/6-31G\* level of theory using the Gaussian16<sup>188</sup> software, and restrained electrostatic potential charges (RESP) for AQ were calculated at the HF/6-31G\* level of theory using the same software. The classical MD simulations were performed using the GPU accelerated pmemd software<sup>323</sup> of the AMBER18 package.

The entire system was at first minimized for 5000 steps using the steepest descent algorithm, followed by 5000 steps using the conjugate gradient algorithm. Afterwards, the system was gradually heated for 50 ps at constant volume (NVT ensemble), using a timestep of 2 fs, to the temperature of 300 K. During the heating process, positional restraints were used for both AQ and poly(dG-dC), by applying a force constant of 10 kcal/(molÅ<sup>2</sup>), while harmonic restraints having the same force constant were applied on

---

the base pairs at the top and at the bottom of the polynucleotide structure (that is, on the G1-C20 and the C10-G11 base pairs) to conserve the double helix structure. This harmonic potential used an equilibrium distance of 10.5 Å between the centers of mass of the base-pairing nucleotides. After the heating, the entire system was equilibrated at constant pressure (NPT ensemble) by three consecutive 2 ns MD simulations, on which the positional restraints (but not the harmonic restraints on the G1-C20 and the C10-G11 base pairs) were gradually removed to 10, 5, and 0 kcal/(molÅ<sup>2</sup>). Afterwards, a long 200 ns production simulation was performed in the NPT ensemble using a Langevin thermostat<sup>224</sup> to keep the temperature constant; the SHAKE<sup>227</sup> algorithm was used along the entire protocol to maintain fixed the bonds involving hydrogen atoms. From the resulting 200 ns trajectory, a snapshot was taken from each one of the two symmetric and from each one of the two rotated configurations identified from the distribution of the twist angle (formed between the long axis of AQ and the axis of the G15-C6 base pair; a detailed definition of these configurations is presented in section 7.3.1) along the entire trajectory. These four snapshots were used as starting geometries for a 200 ns MD production each, so that, overall, 1000 ns of classical MD simulation were obtained. From the resulting 1000 ns trajectories we sampled 100 geometries from each of the intervals (0,30), (30,60), (120,150), and (150,180) degrees of the twist angle, using the Metropolis Monte Carlo algorithm so that the sampled geometries reproduced the Boltzmann-distributed twist angles along the 1000 ns MD trajectories.

For each of these 400 selected geometries an electrostatic embedding hybrid QM/MM calculation was performed using the AMBER18 interface with the Gaussian16 software, whose input files were generated using a development version of the MoBioTools software,<sup>79,80</sup> in which the QM region comprised the AQ molecule plus the G15 guanine nucleobase. In addition, for the 100 geometries of the symmetric 2 configuration Figure 7.1c, additional calculations were performed for two different models: The QM region electrostatically embedded in the DNA strand and the QM region in vacuum. For the QM calculation, the first 10 singlet excited states were computed at the TD-DFT level with the CAM-B3LYP<sup>107</sup> long-range corrected functional and Dunning's cc-pVTZ<sup>231</sup> basis set. The CAM-B3LYP functional was used since it has provided an excellent agreement with the experimental absorption spectra of many organic PSs while providing a proper description

---

of the charge-transfer states involved.<sup>382,383</sup> The characterization of the monomer, excimer, Frenkel exciton and charge-transfer states was performed by using the TheoDORÉ program suite.<sup>381</sup>

## 7.3 Results and Discussion

### 7.3.1 Sampling the Stacking Binding Pocket

Molecular PSs formed by fused-ring aromatic moieties are known to non-covalently bind to DNA strands as intercalators between neighboring base pairs, where the PS/DNA complex is stabilized by stacking interactions between the aromatic rings of the drug and the nucleobases.<sup>34,367</sup> This is the case of anthraquinone derivatives, whose DNA intercalative binding have been extensively investigated.<sup>36,39,384,385</sup> For example, the PS can rotate inside the intercalative pocket of DNA, as was found for methylene blue by MD simulations<sup>378</sup> and spectroscopic measurements,<sup>386</sup> or one or two nucleobases can be ejected and replaced by the intercalator, as it was observed for benzophenone by MD simulations.<sup>75,377</sup> Therefore, vibrational sampling must be considered in the theoretical model when investigating the photophysics of the PS, since different spatial configurations of the chromophore and the environment can present different electronic properties.<sup>387</sup>

The AQ molecule was introduced between the fifth and sixth guanine–cytosine base pairs (G5-C16 and C6-G15) in the double-stranded d(GCGCGCGCGC) decamer as shown in Figure 7.1a. Then, a classical MD simulation was evolved for 200 ns. Large rotational motions of AQ inside the pocket are observed; these can be monitored by computing the twist angle formed by the long axis of AQ and the long axis of the G15-C6 base pair. The former is defined as the vector that connects the centers of mass of the two outer benzene rings of anthraquinone (R1 and R2), and the latter is defined as the vector that connects the C10 atoms of the sugars of each nucleoside in the G15-C6 base pair, as is shown in Figure 7.1b. The probability distribution of the twist angle, plotted in Figure 7.1c, presents two intense maxima at the regions of 0–30° and 150–180°, which correspond to spatial configurations with strong stacking interactions between AQ and the flanking base pairs. These two PS orientations will be referred to as symmetric configurations 1 and 2, respectively. These two distribution maxima extend over the regions of 30–60° and

---

120–150° with much less intensity, whereby the chromophore rotated around the axis normal to its aromatic plane and partially broke the stacking interactions with the nucleobases. These two spatial orientations will be named rotated configurations 1 and 2, respectively. Therefore, the twist angle distribution indicates that the PS visits preferentially four regions of the potential-energy surface: two symmetric and two rotated configurations. In order to determine whether these four configurations are stable or whether they are consequence of a bad equilibration of the system along the simulation, four additional MD trajectories of 200 ns each were performed. The initial conditions for these new simulations are taken from four different snapshots selected from the symmetric and rotated configurations of the original simulation. The twist angle probability distributions for the four additional simulations are shown in Figure 7.1d. As can be seen, the same two symmetric and two rotated configurations are clearly identified. Thus, one can conclude that they are stable regions of the potential-energy surface that must be considered in the subsequent electronically excited-state calculations.

The twist angle defined above shows that the stacking interactions between AQ and the flanking bases are stronger for the symmetric configurations than for the rotated ones. However, a more rigorous geometrical analysis can be performed to characterize the stacking interactions for the symmetric and rotated configurations, and the influence of stacking on the electronically excited states of the system. As will be discussed later, the excitation energies of the system were computed by an electrostatic-embedding QM/MM scheme, where AQ and the nucleobase G15 (see Figure 7.1a,b) were included in the QM region. This partitioning of the system allows the investigation of delocalized excitations, where both the chromophore and the nucleobase participate, and of the effect of the stacking interactions on those excitations. Therefore, the stacking interactions present in the symmetric and rotated configurations have been characterized in terms of the relative orientation between AQ and the nucleobase G15. Specifically, two intermolecular coordinates were defined: The shift and slide distances represented in Figure 2a. The shift distances (N1) are computed as the separation between the center of mass of the six-membered ring of guanine and the center of mass of each of the benzene rings of AQ (R1 and R2 in Figure 7.2a) projected on the plane of guanine and along the base-pair direction. The base-pair direction is defined here as the direction given by the vector that connects the center of mass of guanine and

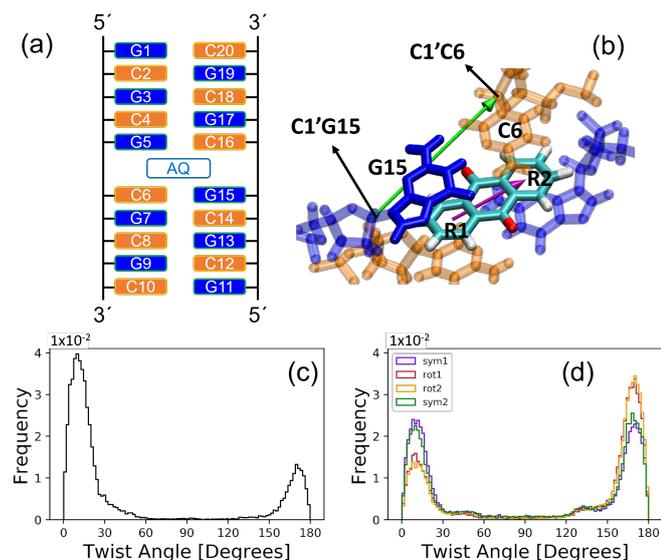


Figure 7.1: a) Schematic representation of anthraquinone (AQ) intercalated into the d(GCGCGCGCGC) decamer sequence between the G5-C16 and C6-G15 base pairs; b) representation of the long axis of AQ (magenta) and long axis of the G15-C6 base pair (green) used to compute the twist angle; c) probability distribution of the twist angle for the initial 200 ns MD simulation; d) probability distribution of the twist angle for the four 200 ns MD simulations which were initialized by four snapshots selected from the four configuration regions (symmetric 1 and 2 and rotated 1 and 2) of the initial simulation. Color code: Guanine nucleotide residues are represented in blue, cytosine nucleotides in orange, and the C, O, and H atoms of AQ in cyan, red and white, respectively.

---

the bisection of the C-C bond opposite to the pyrrole ring of guanine. The slide distance (N2) is defined as the separation between the centers of mass of the six-membered ring of guanine and each of the benzene rings of AQ projected again on the plane of guanine, but along the direction perpendicular to the base-pair direction. This perpendicular direction is given by the vector that connects the center of mass of the six-membered ring of guanine and its carbonyl group, then orthogonalized with respect to N1 via a Gram–Schmidt orthogonalization process.

Two shift distances and two slide distances have been calculated with respect to the rings R1 and R2 of AQ because the visual inspection of the dynamics shows that the AQ ring involved in the stacking interactions with G15 is different depending on the geometric configuration. This can be observed in Figure 7.2b, which displays the probability distributions of the shift and slide distances for the two symmetric and the two rotated configurations. The orientations symmetric 1 and rotated 1 present shorter slide and shift distances for the ring R1 than for the ring R2, while the opposite is true for the symmetric 2 and rotated 2 orientations. This means that the ring R1 is involved in the stacking interactions in the symmetric 1 and rotated 1 orientations, while the ring R2 is the one that interacts with guanine in the symmetric 2 and rotated 2 orientations.

The slide and shift distributions with respect to the ring R1 (R2) for the symmetric 1 orientation are very similar to the distributions for the symmetric 2 orientation with respect to the ring R2 (R1). This indicates that the strength of the stacking interactions is similar for both symmetric configurations. The same holds for the rotated configurations, indicating that the difference in the stacking interactions between the two rotated orientations is not important. However, the rotated configurations exhibit weaker stacking interactions than their symmetric counterparts, as reflected by their widespread slide distributions. The probability distribution for the slide distance (N2) with respect to the ring R1 is extended over larger distances for the rotated 1 orientation than for the symmetric 1 orientation. Similarly, the probability distribution for the slide distance (N2) with respect to the ring R2 is extended over larger distances for the rotated 2 orientation than for the symmetric 2 orientation. The different geometric features observed for the symmetric and rotated configurations, which are related with different stacking scenarios, will be reflected on the electronic properties of the excited states, as will be discussed below.

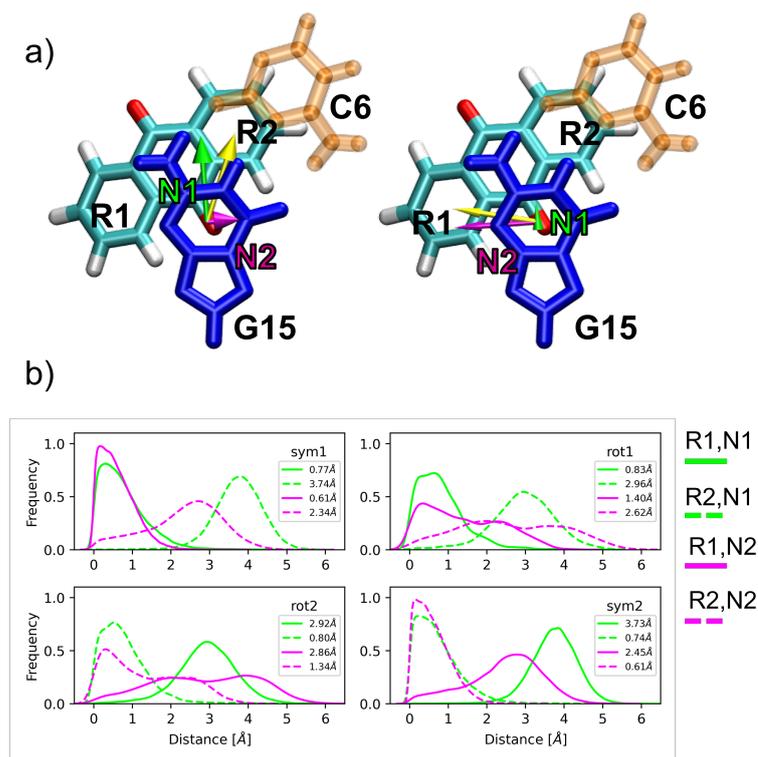


Figure 7.2: a) Shift (N1, green) and slide (N2, magenta) distances between guanine G15 and the rings R1 and R2 of AQ employed in the analysis of the stacking interactions. These distances are computed as the separation between the center of mass of the six-membered ring of guanine and the center of mass of each benzene ring R1 and R2 of AQ (yellow vector) projected on the guanine plane, and along the base-pair direction (N1) and the direction perpendicular to it (N2); b) probability distributions of the shift and slide distances for the two symmetric and the two rotated configurations. Color code: Guanine in blue, cytosine in orange, and the C, O, and H atoms of AQ in cyan, red, and white, respectively.

### 7.3.2 Electronically Excited States: Delocalization and Charge Transfer

The presence of DNA environment can strongly modify the electronic structure of the excited states of chromophores.<sup>71,72,388</sup> In particular, the formation of stacked complexes can result in collective excitations, where the chromophore and the interacting nucleobases actively participate. The electronically excited states of the solvated AQ/DNA complex studied here were computed by an electrostatic-embedding QM/MM scheme, in which the QM region is composed by AQ and G15 (see Figure 7.1a) and described by time-dependent density-functional theory with the CAM-B3LYP functional. Then, the excited states were subsequently characterized by electronic-structure descriptors based on the one-particle

---

transition density.<sup>381</sup> The 10 lowest singlet excited states were computed for 100 snapshots for each of the four geometric configurations discussed above: symmetric 1 and 2 and rotated 1 and 2. Thus, a total of 4000 electronically excited states were computed and characterized. One of the goals of the computations and analyses presented in this Chapter is to characterize the intermolecular charge-transfer states, where electron transfer between the DNA and PS happens, because those states would potentially lead to DNA damage. Since guanine is the most easily oxidizable nucleobase, only one of the neighboring guanines has been included inside the QM region.

The absorption of light by multimeric stacking complexes can lead to the formation of monomer-like excitations, Frenkel excitons, charge-transfer states and excimer states, as represented schematically in Figure 7.3a for the system investigated here formed by two absorbing fragments: AQ and guanine. Frenkel excitons are excitations where both the electron-hole and the excited electron are delocalized over the two fragments with no electron density exchange between fragments. In charge-transfer states the electron-hole and the excited electron are located on different fragments, *i.e.*, the hole is on AQ and the electron on guanine or vice versa. Finally, excimer-like states are a combination of monomer-like and charge-transfer states. It is important to note that the present analysis is aimed to the Franck-Condon region and, thus, excimer species stabilized in an excited-state potential-energy minimum are not formed. However, the term excimer states is used because a strong mix between charge-transfer and exciton states was observed during excimer formation.<sup>389</sup> These four electronic states can be univocally identified by means of the computation of two electronic descriptors, namely average delocalization length ( $DL_{av}$ ) and charge-transfer number (CTN), which have been previously employed to describe the collective excitations in a polyadenine single strand.<sup>390</sup>  $DL_{av}$  is the arithmetic mean between the electron-hole and excited-electron participation ratios and indicates the number of fragments over which the hole and electron are delocalized. For example, pure monomer states and pure exciton states delocalized over AQ and guanine will present a  $DL_{av}$  value of 1 and 2, respectively. CTN provides the fraction of excited electron (or hole) density transferred between different fragments. For example, pure Frenkel exciton states and pure charge-transfer states will have CTN values of 0 and 1, respectively. The different electronic states were classified using the following  $DL_{av}$  and CTN arbitrary thresholds employed in a

previous publication:<sup>390</sup> Electronic states with  $DL_{av} < 1.25$  and  $CTN < 0.2$  are considered monomer-like states; excitons are defined as states with  $DL_{av} > 1.25$  and  $CTN < 0.2$ ; excimers have  $DL_{av} > 1.25$  and  $0.2 \leq CTN \leq 0.8$ ; and finally, a state is classified as a charge-transfer state if  $CTN > 0.8$ . If the 2-dimensional probability distribution function of  $DL_{av}$  and  $CTN$  is computed, the four different types of states can be easily visualized on different regions of the distribution contour plot (see Figure 7.3b).

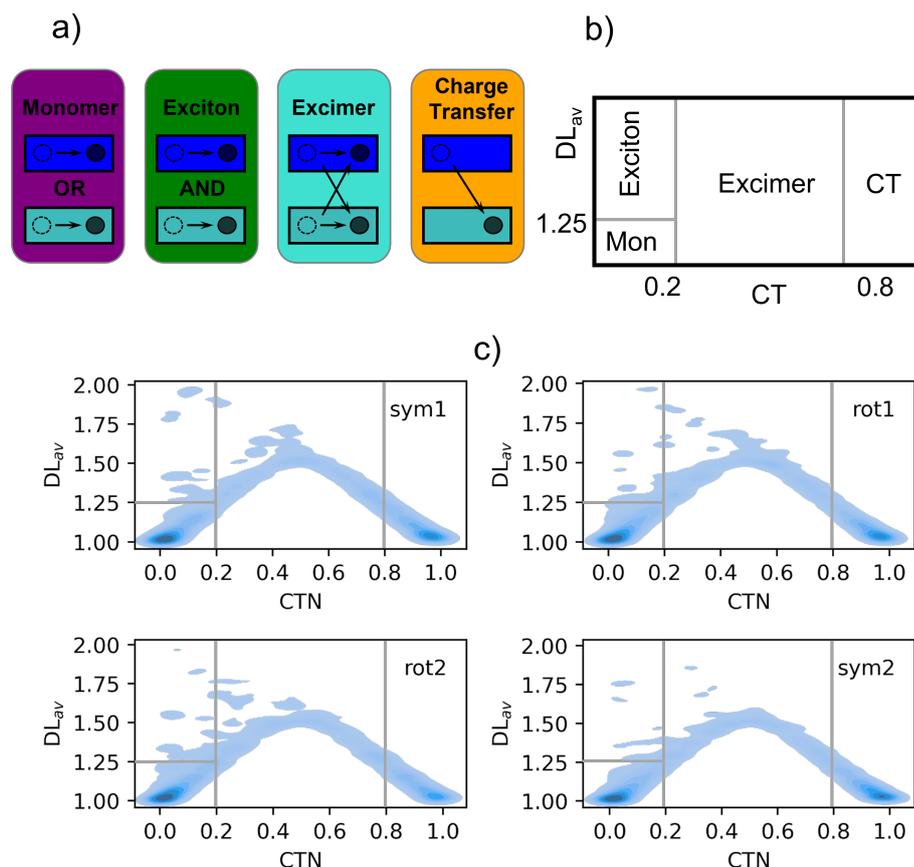


Figure 7.3: a) Representation of the excited states formed in two fragments, *e.g.*, AQ and guanine, represented by rectangles. Empty and filled circles represent the electron-holes and excited electrons, respectively. The thresholds for the average delocalization length ( $DL_{av}$ ) and charge-transfer number (CTN) descriptors to classify the excited states are given; b) Different areas of the 2-dimensional  $CTN/DL_{av}$  distribution where the four types of excited states lie; c) 2-dimensional  $CTN/DL_{av}$  probability distributions for the four geometric configurations identified in the dynamics.

The  $CTN/DL_{av}$  probability distribution for the 100 snapshots of each of the stacking situations (symmetric 1 and 2 and rotated 1 and 2) is plotted in Figure 7.3c. The most

---

important contribution for all the stacking configurations comes from monomer excitations, which are concentrated on a small region of the contour plot at  $DL_{av} = 1$  and  $CTN = 0$  and, thus, correspond to pure monomer states. Charge-transfer states represent the second most important contribution. As for monomer states, the distribution of charge-transfer states is well localized on a small area of the plot, but in this case around the values of  $DL_{av} = 1$  and  $CTN = 1$ . This is especially true for the symmetric orientations, which present a more intense signal than the rotated orientations. This is not surprising since orbital overlap between AQ and guanine is expected to be stronger for the symmetric configurations, where stacking is stronger, leading to more efficient charge-transfer processes. Excimer states are also important but, contrary to monomer-like and charge-transfer states, they are spread over a large area of the distribution map with wide range of CTNs. Finally, exciton states are almost irrelevant and appear mainly on the boundary with excimer and monomer states, that is, the amount of pure Frenkel exciton states are negligible.

The exact contribution of each state class to the total density of states composed by the ten lowest singlet states can be seen in Figure 7.3. Monomer excitations represent around 42% of electronic excited states for the symmetric configurations, while their percentage increases to 45.6% and 50.1% for the orientations rotated 1 and 2, respectively. The percentage of states with exciton character is lower than 5% irrespective of the stacking situation. The contribution of charge-transfer states is larger for the configurations symmetric 1 (30.7%) and symmetric 2 (34.6%) than for the configurations rotated 1 (28.2%) and rotated 2 (23.7%) because, as explained above, orbital overlap between fragments is expected to be stronger for the symmetrically stacked orientations. Since excimer states are a mixture of monomer and charge-transfer states, and those behave in an opposite manner with the stacking interactions, there is no correlation between the contribution of excimer states and the stacking scenario.

If the intensity of each electronic excitation is considered in the analysis, *i.e.*, if the absorption spectrum (and not the density of states) is decomposed into the different contributions, the situation drastically changes. As Figure 7.3 displays, the importance of the charge-transfer states is greatly reduced in the absorption spectrum and, consequently, the percentage of the other electronic-state types increases. In other words, most of the charge-transfer states are dark because the transition dipole moment from the ground

state to the charge-transfer states is small due to the relatively large separation between the PS and guanine, which precludes a strong orbital overlap between the interacting chromophores. The presence of excimer states with mixed exciton and charge-transfer character has also been identified in guanine–cytosine duplexes by means of fluorescence measurements and quantum mechanical calculations.<sup>391</sup> Since these states can evolve to charge-transfer states, which are responsible for DNA photodamage, it would be interesting to investigate in future studies the effect of the PS intercalation on the guanine–cytosine charge-transfer states, especially at low energies, where these states can be easily populated.

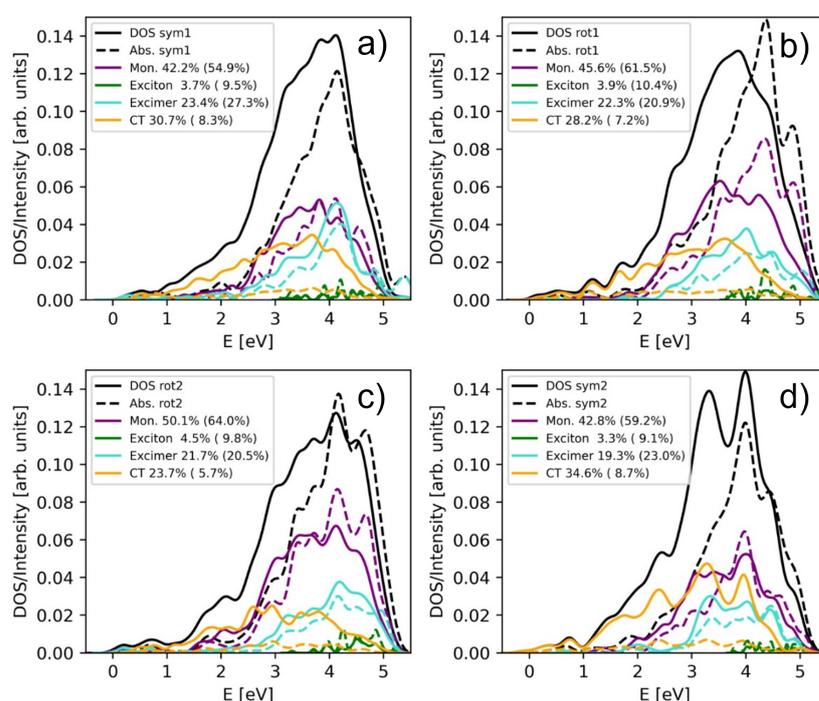


Figure 7.4: Decomposition of the density of states (solid lines) and absorption spectrum (dashed lines) into monomer, Frenkel exciton, excimer, and charge-transfer states for the a) symmetric 1, b) rotated 1, c) rotated 2 and d) symmetric 2 configurations. The percentage of the different contributions to the density of states (absorption spectrum) are given in the legend.

The role of the DNA and solvent environment on the nature of the excited states has also been investigated. Specifically, additional calculations were carried out for the 100 snapshots previously selected from the symmetric 2 configuration. In particular, the solvent molecules and Na<sup>+</sup> ions have been removed from the model and the excited-state QM/MM

---

calculations have been performed for the AQ-G15 QM region electrostatically embedded in the DNA strand. Then, the classical strand has also been removed and the excitations have been computed only for the QM region in vacuum. In this way, the effect of the classical DNA strand and of the solvent on the nature of the excited states can be easily disentangled. Figure 7.5a shows the contribution of the different electronic states to the total density of states for the three different models: Full system, QM region embedded in the classical DNA strand, and QM system in vacuum. As can be seen, when the solvent is removed from the model the contribution of charge-transfer states drastically decreases and, consequently, the monomer-like states become relevant. In addition, the percentage of excimer states suffers a slight drop due to the smaller amount of charge-transfer states that are available to be mixed with exciton states to form excimers. When the strand is also removed from the model, the most significant alteration is seen in the excimer states, whose contribution decreases. Thus, the presence of aqueous solvent favors the formation of charge-transfer states—a fact that is not surprising—and the presence of the DNA strand—described as a point-charge electrostatic embedding—favors the formation of excimer states. An additional factor that deserves attention is the role of the  $\text{Na}^+$  ions. In particular, the presence of a positive ion close to the absorbing region of the system could drastically affect the charge-transfer states by stabilizing the excited electron or destabilizing the electron-hole by Coulomb interactions. Figure 7.5b displays the probability distribution of the separation between the center of mass of the AQ-G15 region and the nearest  $\text{Na}^+$  ion. As can be seen, the position of the nearest ion can adopt a wide range of distances from the QM region from around 1 to 20 Å. However, the CTN of the electronic states does not show a clear trend with the ion position and its value is always 0.4.

It is interesting to characterize in more detail the monomer and charge-transfer states. As explained above, monomer states are formed by excitations where both the electron-hole and excited-electron densities are located on the same fragment. To unravel the fragment that is involved in the monomer excitations of Figures 7.3 and 7.4, the probability distribution of the position of the electron-hole ( $\text{POS}_i$ ) and of the position of the excited electron ( $\text{POS}_f$ ) have been obtained from the transition-density analysis,<sup>61,381</sup> and are plotted in Figure 7.6a,b. Both distributions peak at fragment 1, which in the present case is AQ, and only a small fraction of the distribution appears at fragment 2 (guanine) for

---

the four orientation configurations. This means that most of the monomer-like electronic states that are involved in the density of states are located on the PS. The same analysis performed for the charge-transfer states, plotted in Figure 7.6c,d, reveals that the electron-hole and excited electron are located at fragments 2 and 1, respectively, independently of the orientation of the chromophore inside the binding pocket. This means that the electron flow in charge-transfer states occurs mainly from guanine to AQ, and only a small percentage of electronically excited states present electron transfer in the opposite direction.

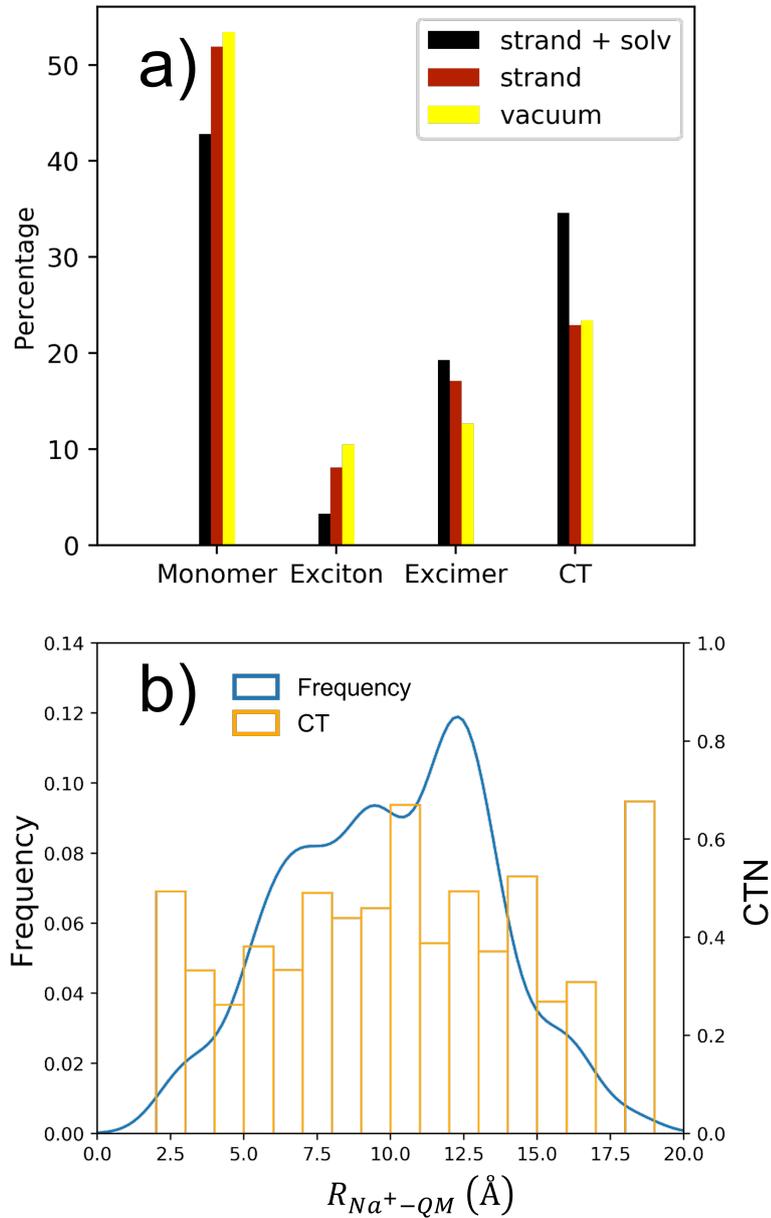


Figure 7.5: a) Contribution of monomer, exciton, excimer and charge-transfer states to the density of states for the symmetric 2 configuration. The percentages of the full system, the quantum mechanics (QM) region (AQ and G15) embedded in the DNA strand and the QM region in vacuum are shown in black, brown and yellow, respectively. b) Probability distribution of the separation  $R_{Na^+-QM}$  between the center of mass of the QM region and the nearest  $Na^+$  ion to the QM region (blue) and variation of the CTN with that separation (yellow).

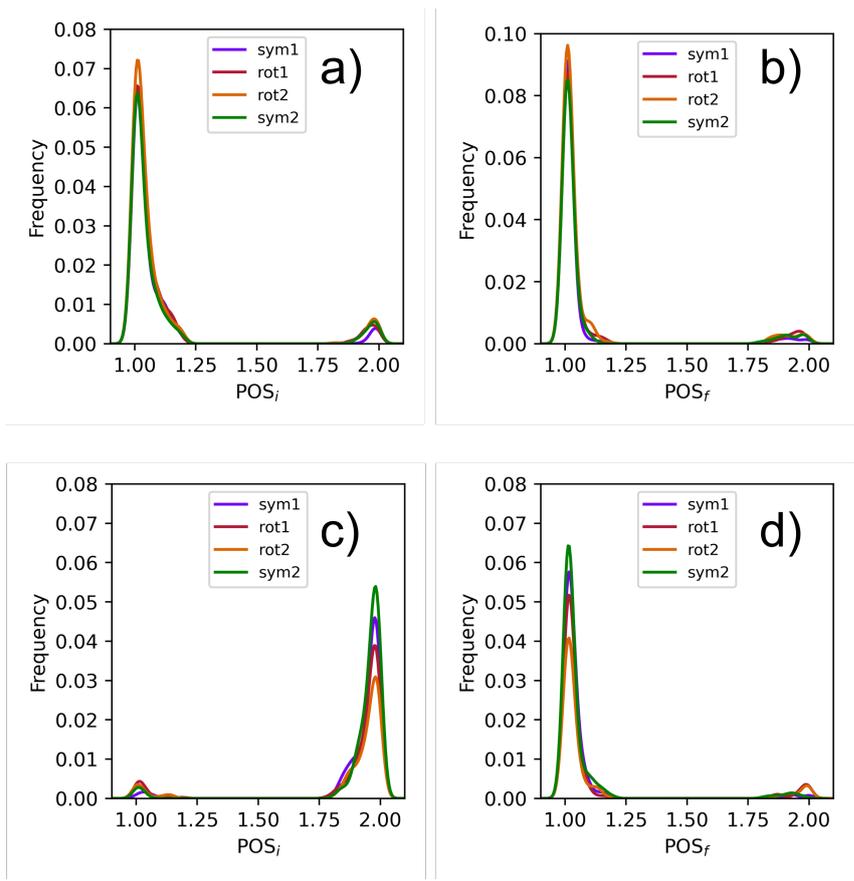


Figure 7.6: Probability distribution of the position of the hole ( $POS_i$ ) and of the position of the excited electron ( $POS_f$ ) for (a,b) the monomer-like states and (c,d) the charge-transfer states. Positions 1 and 2 correspond to AQ and guanine, respectively.

---

## 7.4 Conclusions

Anthraquinone derivatives are known to participate in PDT mechanisms through the interaction with DNA strands by, mainly, an intercalative binding mode. After absorption of UV light, the PS is involved in DNA oxidative damage, where electron transfer from guanine nucleobases to the excited PS occurs. Therefore, an efficient photoinduced DNA damage pathway requires the existence of charge-transfer states energetically accessible. In this Chapter, the electronically excited states at the Franck-Condon region of AQ intercalated into a solvated double-stranded d(GCGCGCGCGC) decamer was investigated by means of a combination of classical MD simulations, QM/MM excited-state calculations and one-electron transition-density matrix analysis.

Classical MD simulations evolved for 1  $\mu$ s showed that the chromophore can adopt four different stable poses inside the intercalative pocket of DNA: Two symmetric configurations, where the stacking interactions between AQ and the flanking nucleobases are strong, and two rotated configurations, where the stacking interactions are partially broken. The density of states and the absorption spectrum of the AQ-DNA solvated complex for the four different geometric configurations were computed by means of an electrostatic-embedding QM/MM scheme, where the chromophore and one of the flanking guanine residues were described by TD-DFT in the QM region. For each of the geometric configurations, 100 geometries were considered in the computation of the 10 lowest singlet excited states to take into the account the vibrational sampling of the system. Moreover, the analysis of one-electron transition densities allowed the characterization of the excited states as monomer, Frenkel exciton, excimer and charge-transfer states. The density of states is mainly dominated by monomer-like states located on AQ and charge-transfer states, where an electron is transferred from guanine to the chromophore. Moreover, the charge-transfer states are more relevant for the symmetric configurations, where orbital overlap between AQ and guanine is expected to be more important, than for the rotated configurations. Excimer states are also relatively important in the density of states band, and Frenkel excitons are virtually negligible since they represent less than 5% of the signal. The contribution of charge-transfer states is drastically reduced, while the contribution of the other electronic-state classes increases, when the absorption spectrum is computed, a fact that indicates

---

that the AQ/guanine charge-transfer states are mainly dark.

In conclusion, the binding of AQ to DNA induces the formation of charge-transfer excited states between the PS and guanine. It has been shown that these states are energetically accessible at the Franck-Condon region in the singlet manifold. It is very likely that those states are also available in the triplet manifold after intersystem crossing, from where the DNA oxidative damage is initiated. However, further simulations are needed to investigate the efficiency of intersystem crossing when the chromophore is interacting with DNA, and of the charge-transfer process once the system is in the triplet manifold. In addition, theoretical modeling would also be beneficial to rationally functionalize the AQ scaffold and obtain new PSs with improved photophysical properties. The excited-state electronic structure of these functionalized anthraquinone derivatives could be compared with the present calculations for AQ to evaluate whether the functionalization of AQ has led to a modification of the electronic-state features, *e.g.*, the charge-transfer character, that could enhance DNA damage.

## Chapter 8

# Permeation Mechanism and Excited States of Anthraquinones inside a Lipid Bilayer

Gustavo Cárdenas<sup>1</sup>, Davide Avagliano<sup>3</sup>, Marco Garavelli<sup>3</sup>, Juan J. Nogueira<sup>1,2</sup>

1 – Department of Chemistry, Universidad Autónoma de Madrid, 28049, Madrid, Spain

2 – Institute for Advanced Research in Chemistry (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

3 – Department of Industrial Chemistry “Toso Montanari”, Università degli Studi di Bologna, 40136, Bologna, Italy

This Chapter is work in progress, and as such, it has not been published yet.

*This brief Chapter presents an overview of a project under progress. The object of study of this work consists of two anthraquinone derivatives – rubiadin and soranjidiol – which have been shown to present promising anticancer activity on colorectal tumor cancer cells. In this Chapter, the permeation mechanism of these derivatives inside a 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) lipid bilayer model is studied by means of umbrella sampling classical dynamics simulations. It is observed that for both molecules the permeation is*

---

*thermodynamically favorable, and each of them attains an energy minimum at 10.5 Å from the center of the lipid membrane. One of the main goals of the present Chapter is to study the excited state properties of these anthraquinone derivatives inside the lipid membrane at the QM/MM MS-CASPT2 level of theory. In this regard, a crucial step is the determination of a suitable active space to perform these calculations. A classical MD simulation of each of the anthraquinone derivatives is performed in chloroform to obtain a representative geometry for which to compute the absorption spectrum of each molecule, to perform a benchmark of different active spaces by comparing the absorption spectra with the experiment. A (12,9) active space is chosen to compute the absorption spectrum of rubiadin and soranjidiol at the free energy minimum of the permeation profile inside the lipid membrane, which displays an overall red shift with respect to the experimental spectrum in chloroform. Future investigation will involve the study of the intersystem crossing process of these molecules inside the lipid membrane, and the investigation of yet another anthraquinone derivative, the parietin molecule.*

## **8.1 Introduction**

Photodynamic therapy (PDT) is a non-invasive technique that is nowadays widely used due to its high selectivity and specificity,<sup>364,392,393</sup> as well as its efficacy in treating infectious diseases and different types of cancer.<sup>5-7,362</sup> In PDT, a photosensitizing agent (PS) is administered in a selective manner to the target tissue, after which it is irradiated with radiation of a suitable wavelength in the presence of molecular oxygen, so that the PS undergoes reactions of electron transfer or energy transfer with the surrounding tissues. In the case of a carcinogenic tissue, these reactions provoke the rupture of the nearby tissue, triggering processes that ultimately lead to cell death of the carcinogenic cells in an ordered manner (apoptosis). It has been evidenced that the mechanism of cell death strongly depends on the tissue where the PS localizes as well as on the structure of the PS.<sup>22,23</sup> It is commonly recognized that after photoexcitation of the anticancer chromophore, it undergoes nonradiative relaxation, in which it populates the triplet excited state manifold *via* intersystem crossing. It is at this point that the PS either undergoes electron transfer reactions that bring about the formation of reactive oxygen species (*e.g.*, the superoxide

---

radical anion  $O_2^-$ ), which in turn provoke the oxidation of nearby biological molecules (type I mechanism), or the PS undergoes an energy transfer reaction towards molecular oxygen (type II mechanism) generating the highly reactive singlet oxygen species  $^1O_2$ .<sup>394,395</sup> The PS can accumulate on different cellular tissues depending on their nature.<sup>22</sup> In particular, mitochondria result to be an ideal target tissue for PDT, as they release pro-apoptotic components such as cytochrome c to the cytosol following the photoinduced damage of the lipid membrane by the PS.<sup>31-33</sup> Several families of organic PSs have been object of study for PDT, including cyanines, phenanthridinium dyes, porphyrins, phenothiazinium dyes, acridines, and anthraquinones.<sup>34</sup> In particular, the family of anthraquinones has shown promising cytotoxic activity both *in vitro* and *in vivo* against carcinogenic tissues, especially breast cancer and colorectal cancer cells.<sup>35,78,375,376</sup> Specifically, after PDT treatment with two anthraquinone derivatives – rubiadin and soranjidiol – the rupture of the lipid membrane was observed at the early stages of apoptosis.<sup>78</sup> Thus, the investigation of the permeation mechanism of these PSs inside the lipid membrane and the influence of the lipid environment in the excited states of the PSs could provide valuable insights into the mechanism of action of the PSs.

Molecular modeling techniques at an atomistic scale are suitable to fulfill this task.<sup>40</sup> Indeed, MD in conjunction with multiscale hybrid quantum mechanics/molecular mechanics (QM/MM) methodologies have been widely employed to study the interactions of several representatives of the PS families mentioned above with biologically relevant target tissues. For example, MD simulations have been performed to study the binding modes of anthraquinones<sup>51-53</sup> and other PS<sup>76,377,378</sup> with DNA. The absorption and emission spectra of Nile red and Nile blue have been computed considering all possible binding modes of these chromophores with DNA,<sup>75</sup> and the mechanism of fast intersystem crossing of acetophenone in the presence of a double strand of DNA following photoexcitation with UV radiation has also been investigated using QM/MM techniques.<sup>74</sup> MD techniques have been employed in the investigation of the different binding modes of methylene blue in DNA,<sup>378</sup> and a combination of MD and QM/MM techniques has also been employed to investigate the influence of DNA on the electronically excited states of this PS.<sup>72</sup> The permeation mechanism and the excited states of methylene blue inside a lipid bilayer have also been investigated using enhanced sampling classical MD techniques and QM/MM methodolo-

---

gies.<sup>73</sup> Other examples of PSs investigated *via* these computational methodologies include chelerythrine,<sup>54</sup> porphyrines,<sup>77</sup> and palmatine.<sup>76</sup>

To this date, however, no similar work has been performed in regard with the study of anthraquinones in the presence of a phospholipid bilayer. For this reason, in the present work the permeation mechanism of the two anthraquinone derivatives mentioned above (rubiadin and soranjidiol) is modeled by means of umbrella sampling<sup>45,143</sup> classical MD simulations. The present work also aims at computing the electronically excited states of these PSs in the presence of a 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid membrane. To this end, multiconfigurational methods such as the complete active space self consistent field (CASSCF)<sup>81–83</sup> in conjunction with multistate complete active space perturbation theory (MS-CASPT2)<sup>99,101</sup> are employed, as they provide a more accurate description for the wavefunction of the system in situations of near-degeneracy between electronically excited states, such as in the case of the intersystem crossing process than single configuration methods. The present Chapter provides preliminary work in the choice of the active space by testing different active spaces in the simulation of the absorption spectrum of rubiadin and of soranjidiol in chloroform, by employing a QM/MM optimized structure of each chromophore in explicit chloroform. A (12,9) active space is chosen to compute the absorption spectrum of rubiadin and soranjidiol at the minimum of the potential of mean force (PMF) obtained from the umbrella sampling MD simulations. These spectra are computed using an ensemble of geometries of the chromophores inside the lipid membrane, so that the (12,9) active space optimized in chloroform is used as an initial guess to optimize the wavefunction for each sampled geometry. In this regard, the `pyoverlaps.py` script of the MoBioTools<sup>79,80</sup> toolkit is used to conserve the active space. Future work will include the investigation of the intersystem crossing process for rubiadin and soranjidiol inside the lipid membrane. The parietin molecule – another anthraquinone derivative investigated along with rubiadin and soranjidiol<sup>78</sup> – will also be studied.

## 8.2 Computational Details

The systems studied in the present work consist of a DOPC lipid bilayer containing two leaflets of 64 lipid molecules each, a 25 Å of water thickness on each side of the bilayer,

---

a concentration of 0.15 M of KCl to emulate physiological conditions, and the rubiadin and soranjidiol molecules. The lipid membrane has been prepared and equilibrated in a previous work,<sup>85</sup> (Chapter 5) in which the CHARMM-GUI<sup>239</sup> software was used to build the initial structure of the lipid membrane. The lipid molecules were described using the lipid17 force field – an updated version of the lipid11<sup>134</sup> and the lipid14<sup>135</sup> force fields, the water molecules were described with the TIP3P water model,<sup>137</sup> and the K<sup>+</sup> and Cl<sup>-</sup> ions with suitable amber parameters.<sup>319</sup> In this work, the structures of rubiadin and soranjidiol were optimized at the MP2/6-31G\* level of theory. The bonded and Lennard-Jones force field parameters of the anthraquinones were retrieved from the GAFF force field,<sup>131</sup> whereas the atomic charges were obtained by means of an electrostatic potential fitting with a HF/6-31G\* level of theory. Each of the anthraquinones was manually placed at a distance of 32Å from the center of mass of the lipid membrane using the tleap module of AmberTools20,<sup>192</sup> giving a total of 31911 atoms. Each system consisting of the solvated lipid membrane plus an anthraquinone derivative was minimized using the steepest descent algorithm for 5000 steps, and the conjugate gradient algorithm for another 5000 steps. Afterwards, it was heated in the (NVT) ensemble to 300 K employing a Langevin<sup>224</sup> thermostat with a collision frequency of 1 ps<sup>-1</sup>. A 10 kcal/(mol Å<sup>2</sup>) force constant was imposed upon the lipid molecules during the heating process. Following the heating process, three consecutive molecular dynamics simulations of 4 ns each were performed at constant pressure (NPT ensemble), where the force constants applied on the lipid molecules were gradually decreased from 10 kcal/(molÅ<sup>2</sup>) to 5 kcal/(molÅ<sup>2</sup>) and finally 0 kcal/(molÅ<sup>2</sup>). A Berendsen barostat<sup>225</sup> was used with a pressure relaxation time of 1 ps to maintain the pressure fixed at around 1 bar. Finally a 100 ns production run was performed to equilibrate the density of the system. For the heating and the production runs, a timestep of 2 fs was used. Throughout these steps, the Particle Mesh Ewald method<sup>226</sup> was used to compute the coulomb interactions, in which a cutoff of 10 Å was used to limit the direct-space sum; the same cutoff was employed to compute the van der Waals interactions. All bond distances involving H atoms were restrained by means of the SHAKE algorithm.<sup>227</sup> For the entire protocol, positional restraints were imposed upon each of the anthraquinone derivatives by applying a force constant of 5 kcal/(molÅ<sup>2</sup>).

Following the equilibration of the systems, umbrella sampling MD was employed to

---

simulate the permeation process of each anthraquinone derivative inside the lipid membrane. In this context, the reaction coordinate was defined as the distance between the center of mass of the anthraquinone derivative and the center of the lipid membrane, so that the permeation occurs perpendicular to the plane of the membrane (Figure 8.1). The pathway along the reaction coordinate was subdivided into 65 windows separated each by 0.5 Å, starting from an initial distance of 32 Å, in which each anthraquinone derivative is outside the lipid membrane, up to a distance of 0 Å, whereby each anthraquinone derivative is at the center of the membrane. For each window a 30 ns of molecular dynamics simulation was performed, using the same settings as the initial production run. A 2.5 kcal/(mol Å<sup>2</sup>) harmonic bias potential was applied on each of the windows. The free-energy profiles for rubiadin and for soranjidiol were obtained by means of the WHAM approach.<sup>157,158</sup>

To determine a suitable methodology to compute the absorption spectrum of each anthraquinone inside the lipid membrane, the absorption spectra were first computed on a system consisting of each anthraquinone inside a solvation box of chloroform, so that the resulting spectra could be compared with the experimental results.<sup>396</sup> To do so, each anthraquinone derivative was placed inside a cubic solvation box of chloroform using the `tleap` module of `AmberTools20`, in which the distance between any solute molecule and the boundaries of the box was at least 30 Å. The chloroform molecule was described by means of the GAFF force field. Each system was minimized for 5000 steps using the steepest descent algorithm, after which 2000 steps of the conjugate gradient were performed. The systems were then heated for 100 ps in the NVT ensemble, after which an equilibration of 110 ns was run in the NPT ensemble. For these calculations the same barostat, thermostat Ewald summation and van der Waals cutoffs were used as in the case of the equilibration of the lipid membrane. A cluster analysis using the `k means` algorithm<sup>397,398</sup> was performed upon the last 100 ns of the MD simulation, with a sieve of 20 geometries applied randomly along the trajectory. From this analysis 100 clusters were identified, of which the most populated one was selected as the initial structure for an electrostatic embedding QM/MM geometry optimization considering the anthraquinone derivative in the QM region, an a radius of 4 Å around the solute molecule was applied to include chloroform molecules in the mobile MM region. The QM/MM optimization was performed using the `COBRAMM`<sup>187</sup> software, which interfaces the `Gaussian16`<sup>188</sup>

---

and the AMBER20<sup>192</sup> programs. The absorption spectra of rubiadin and soranjidiol in chloroform were computed on the QM/MM optimized geometries using an electrostatic embedding QM/MM scheme, where the QM region was described at the MS-CASPT2 level of theory with a cc-pVTZ<sup>231</sup> basis set. A benchmark of different active spaces was performed, including (12,9), (12,10), (14,10) and (14,12). Finally, a (12,9) active space was selected to compute the absorption spectrum of rubiadin and soranjidiol inside the lipid bilayer on top of an ensemble of 100 geometries for each molecule; these geometries stem from the last 20 ns of molecular dynamics simulation on the umbrella sampling window corresponding to the minimum of the free energy profile. The module `pyoverlaps.py` of the MoBioTools software was employed to generate the QM/MM inputs from the 20 ns trajectory, including the analysis (and the eventual correction) of the active space, whereby the active spaces of the QM/MM optimized geometries were used as reference (Figures 8.4 and 8.5). The Visual Molecular Dynamics<sup>236</sup> software was used to visualize the trajectories, and the molden<sup>198</sup> program was employed to generate the images of the molecular orbitals.

## 8.3 Results and Discussion

### 8.3.1 Umbrella Sampling Simulations

The PMF profiles relative to the permeation of rubiadin and soranjidiol inside the DOPC lipid membrane are represented in Figure 8.1. As pointed out elsewhere<sup>85</sup> (cf. Chapter 5), whenever the reaction coordinate of a slow process is known *a priori*, it is advisable to employ enhanced sampling techniques such as umbrella sampling, as it allows for the exploration of the potential energy landscape at a reasonable computation time, a task that would otherwise be unachievable using conventional classical MD. The permeation mechanism of different molecules inside a model of a lipid bilayer has been studied extensively in the literature by means of umbrella sampling simulations,<sup>49,50,73,85,318</sup> and this is the approach adopted in the present work as well. Figure 8.1 shows the PMF profiles for rubiadin (green line) and for soranjidiol (orange line), in which the first 4 ns of simulation time for each umbrella sampling window have been considered as the equilibration time, and thus have been discarded.

It can be evidenced from Figure 8.1 shows that the permeation process for both

## Dioleoyl-phosphatidylcholine (DOPC)

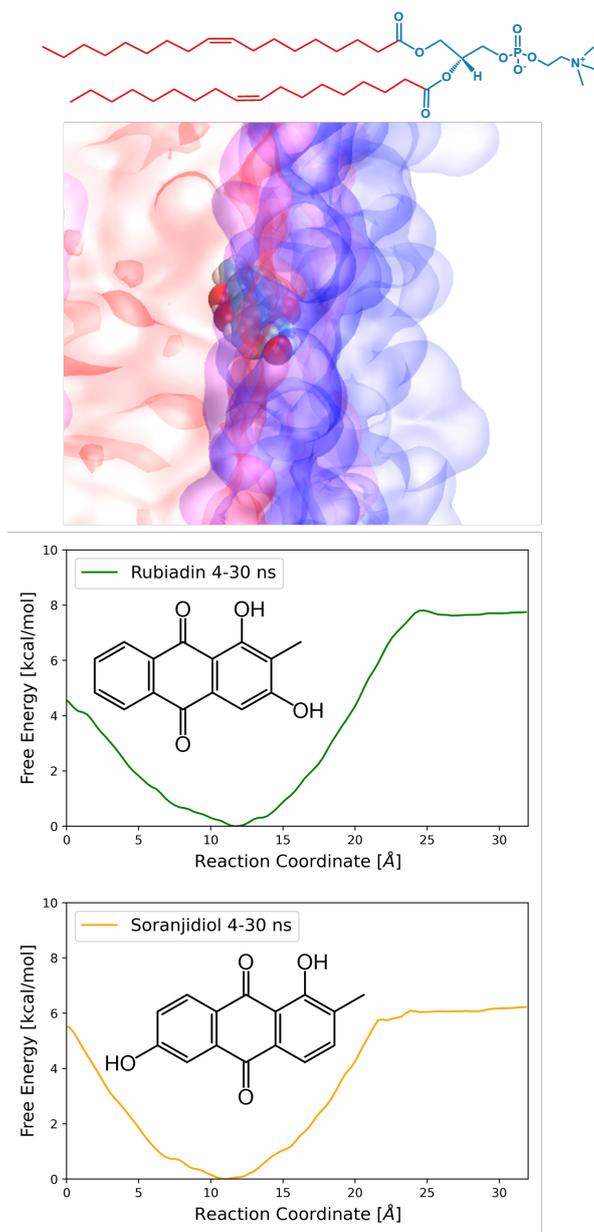


Figure 8.1: Top: Schematic representation of an anthraquinone derivative embedded inside the DOPC lipid membrane. The polar and non-polar regions of the membrane are represented in blue and red, respectively. Bottom: Potential of mean force (PMF) of the permeation of the rubiadin (green line) and the soranjidiol (orange line) molecules, considering the simulation interval time between 4 and 30 ns for each umbrella sampling window.

---

rubiadin and soranjidiol is energetically favorable, and both molecules attain the PMF global minimum at a distance of 10.5 Å from the center of mass of the lipid membrane, at the interface between the polar heads and the non-polar tails (see Figure 8.1, Top). Noticeably, despite the structural similarity between the two molecules, they display different lineshapes and energetics for the permeation process. Specifically, rubiadin presents a maximum of 0.5 kcal/mol at a distance of 25 Å from the center of the membrane, before attaining the global minimum of the PMF; on the other hand, soranjidiol presents a flat PMF until it arrives at a distance of 25 Å from the center of the membrane, whereby the profile presents a small step followed by a steady decrease towards the global minimum. The presence of the global minimum inside the lipid membrane is in agreement with the experimental finding of anthraquinone derivatives accumulating inside the lipid membrane, where they promote the lipid membrane rupture following photoexcitation.<sup>78</sup> Both molecules evidence similar energetics with respect to the permeation process, with rubiadin presenting a reaction free energy of 8 kcal/mol and soranjidiol presenting a reaction free energy of 6 kcal/mol. This similarity in the reaction free energies of the permeation process for both molecules is in agreement with the experimental observation that rubiadin and soranjidiol present a similar cytotoxicity.<sup>78,376</sup> However, it should be emphasized that the systems under study are model systems of the much more complex environment present in a carcinogenic cell. It should also be pointed out that rubiadin and soranjidiol need to be photoactivated in order to trigger their anticancer activity, so that a key property that needs to be explored to gain further insights into the activity of these drugs are their electronically excited states. This is done in the following section.

### 8.3.2 Electronically Excited States

The goal of the present project is to gain insights into the anticancer activity of rubiadin and soranjidiol. It is well known that the anticancer activity of the anthraquinone derivatives is triggered by means of photoexcitation with light having a suitable wavelength. Indeed, the PS is excited towards the singlet excited state manifold, after which it undergoes intersystem crossing towards the triplet manifold.<sup>399</sup> It is at this point that the PS is able to promote the formation of singlet oxygen or of reactive oxygen species that react with the lipid membrane. To study these properties at an accurate level of theory, it

---

is advisable to employ multiconfigurational methods to better describe the situations in which different excited states are almost degenerate (this is particularly important nearby crossing regions between two potential energy surfaces). For this reason, in the present work the SA-CASSCF methodology, in conjunction with MS-CASPT2, has been adopted to calculate the excited states of rubiadin and soranjidiol in the presence of the lipid membrane. A caveat of this method (in particular SA-CASSCF) consists of the choice of a suitable active space. This problem is tackled here by considering a set of different active spaces and comparing the simulated absorption spectra with the experimental spectra, which have been recorded in chloroform.<sup>396</sup> To this end, and to account for the fact that the molecules are in a solution of chloroform, a 110 ns classical MD simulation of rubiadin and of soranjidiol has been performed considering a solvation box of chloroform, of which the first 10 ns have been discarded. Afterwards, a cluster analysis along the 100 ns trajectory was performed, using the k means method and considering a maximum number of 100 clusters, to determine the most representative configuration of the simulation.

Figure 8.2 shows the most representative configuration for rubiadin and for soranjidiol stemming from the above mentioned cluster analyses. These structures were used as initial geometries to perform an MP2/6-31G\* QM/MM geometry optimization including each PS in the QM region and all the chloroform molecules within a radius of 4.0 Å from any atom of the solute in the mobile MM region, using the COBRAMM program. The geometries were optimized so as to obtain representative structures from the classical MD simulations, while at the same time relaxing the structure with the QM potential. Figure 8.3 shows a comparison of the absorption spectra of rubiadin and of soranjidiol in explicit chloroform computed at the MS-CASPT2 level of theory with different active spaces, and the corresponding experimental spectra. These spectra were computed on top of the previously QM/MM optimized geometries. It can be evidenced from Figure 8.3 that none of the active spaces reproduces accurately the entire lineshape of the experimental spectra. A source of error could be evidenced in the fact that a larger active space might be needed to account for all the underlying electronic transitions in the range between 450 and 200 nm, as only a portion of the valence orbitals for rubiadin and for soranjidiol is considered within the active spaces employed. However, it should be emphasized that for both rubiadin and for soranjidiol, the computations with the largest active space considered

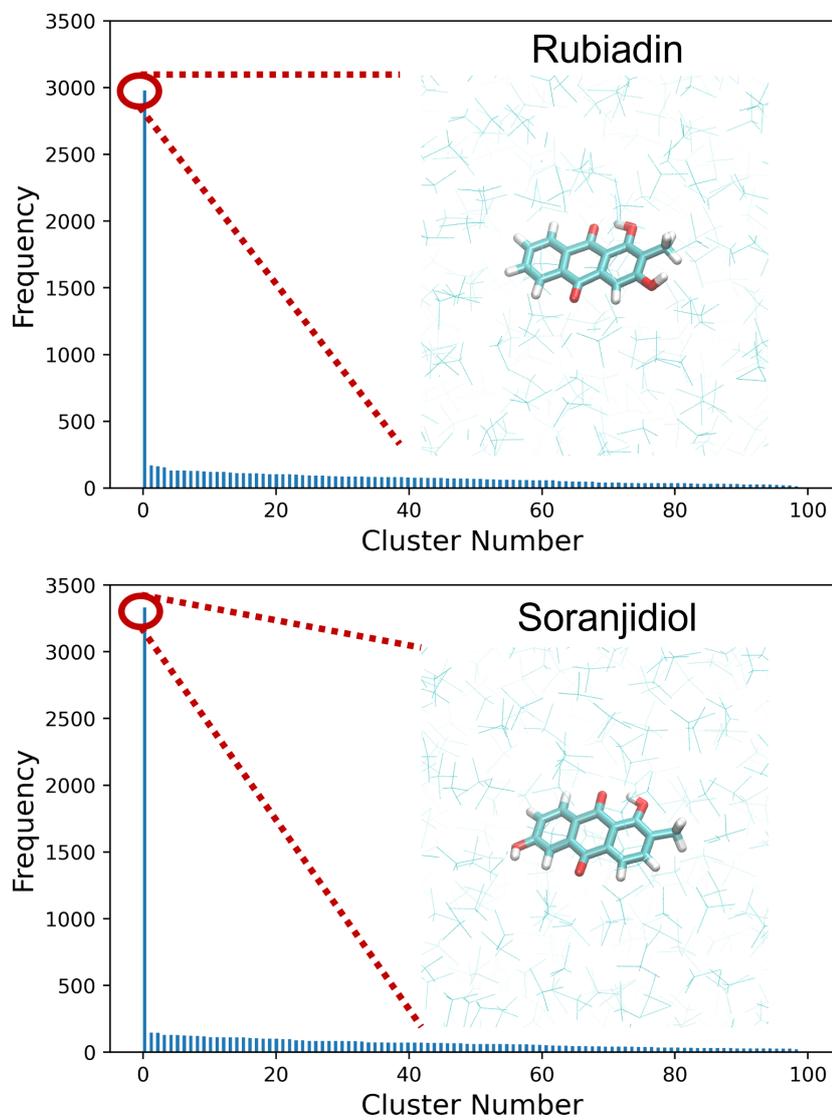


Figure 8.2: Cluster analysis of rubiadin (top) and of soranjidiol (bottom) along a 100 ns classical molecular dynamics simulation in chloroform. The inset of each plot represents the most representative cluster for each of the anthraquinone derivatives.

---

here – (14,12) – do not properly reproduce the first band of the corresponding experimental spectra. On the other hand, this band is reproduced by the calculations employing the smaller active spaces; in particular, in the case of rubiadin, the first bright electronic excitation is centered at 401 nm, that is exactly at the center of the first experimental band. Another source of error could be the fact that the QM/MM optimized geometry need not be the global minimum of the system, but instead one of the several possible local minima present in a complex system consisting of more than 1000 atoms. Indeed, although the cluster analysis mentioned above did provide the most representative geometry along the classical MD trajectory, it need not be the case that the optimized geometry corresponds to a representative geometry of the QM/MM potential. Thus, the effect of conformational sampling would need to be accounted for to better reproduce the lineshape of the spectrum. This is part of the ongoing work for the present project.

As the (12,9) active space (Figure 8.4 and 8.5 for rubiadin and soranjidiol, respectively) provided a decent description of the first band of the absorption spectrum of the two anthraquinone derivatives under study, it is worth analyzing the character of the excited states that compose this band. Tables C.1 and C.1 show the absorption spectra of rubiadin and soranjidiol, respectively, along with the most prominent electronic configurations. The first band for both molecules is characterized by the electronic transitions towards the  $S_1$  and  $S_2$  excited states. For both rubiadin and soranjidiol, the most prominent electronic excitation present in the first band corresponds to the excitation towards the  $S_1$  excited state (Tables 8.4 and 8.5). In the case of rubiadin, the  $S_1$  excited state presents a mixed contribution from the configurations referred to the  $\pi_4 \rightarrow \pi_1^*$  (HOMO  $\rightarrow$  LUMO),  $\pi_1 \rightarrow \pi_1^*$  and  $n_1 \rightarrow \pi_1^*$  excitations (see Figures 8.4 and 8.5 for reference). In the case of soranjidiol, the  $S_1$  state presents contributions from the  $\pi_4 \rightarrow \pi_1^*$  (HOMO  $\rightarrow$  LUMO) and the  $\pi_2 \rightarrow \pi_1^*$  excitations.

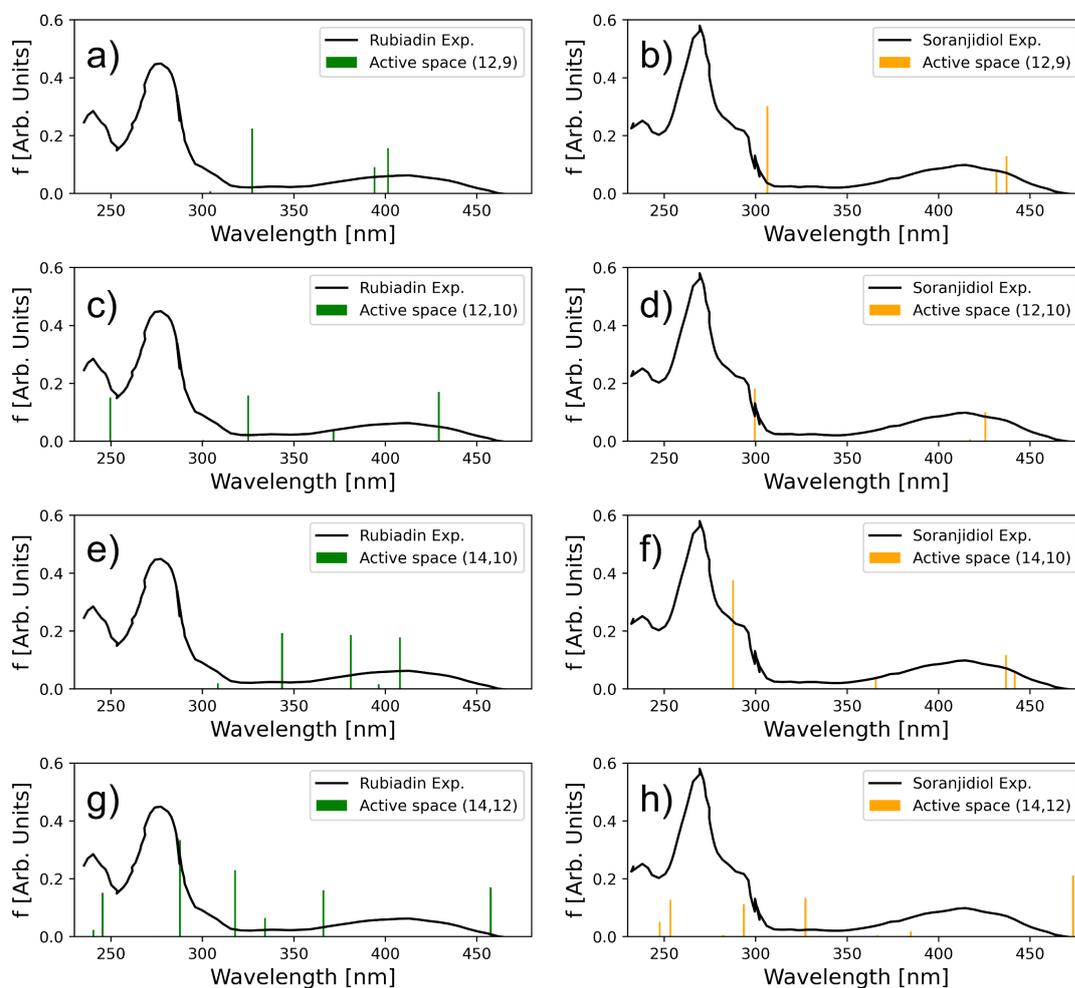


Figure 8.3: Electronic absorption spectrum of rubiadin (green) computed at the QM/MM optimized geometry in chloroform at a MS-CASPT2 level of theory, with an active space of a) (12,9), c) (12,10), e) (14,10), g) (14,12). Electronic absorption spectrum of soranjidiol (orange) computed at the QM/MM optimized geometry in chloroform at a MS-CASPT2 level of theory, with an active space of b) (12,9), d) (12,10), f) (14,10), h) (14,12). The experimental spectrum of each molecule is shown as a black line.

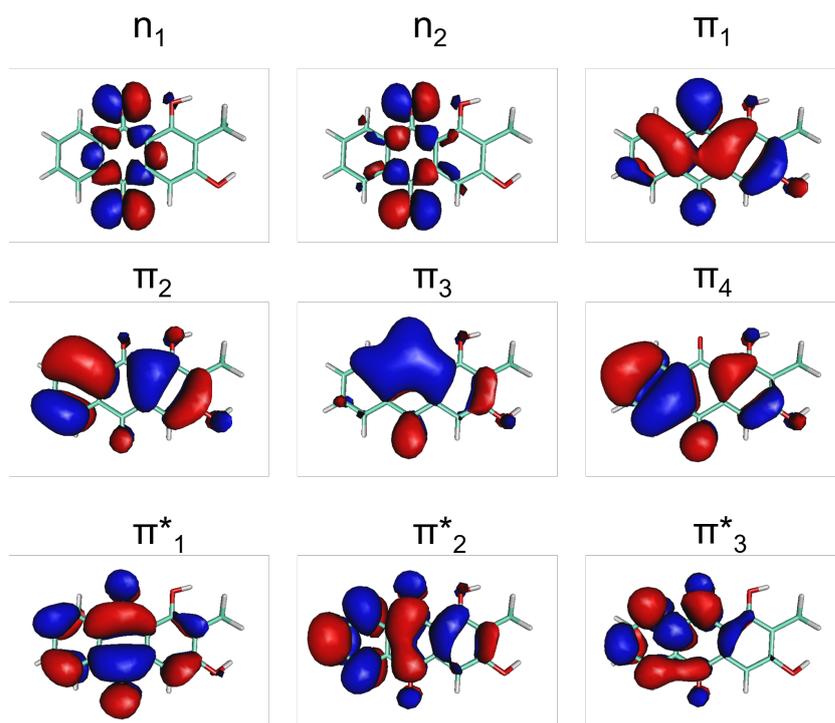


Figure 8.4: (12,9) Active space of rubiadin.

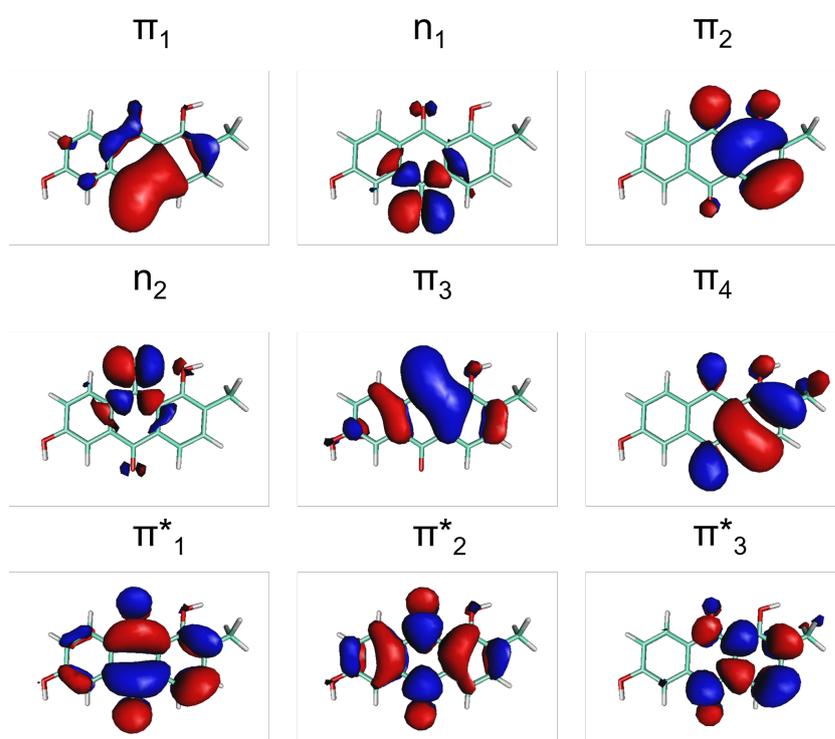


Figure 8.5: (12,9) Active space of soranjidiol.

---

The active space (12,9) was chosen as the reference active space to compute the absorption spectrum of rubiadin and soranjidiol at the minimum of the free energy profile (Figure 8.1). Thus, 100 geometries were sampled from the classical MD simulation of each molecule in the window corresponding to the minimum of the PMF (at a distance of 10.5 Å from the center of the membrane in both cases). Afterwards, the `pyoverlaps.py` script of the MoBioTools toolkit was used to generate the inputs for the single point QM/MM state average (SA)-CASSCF and subsequent MS-CASPT2 calculations on top of the sampled geometries, and to eventually correct the active space in cases in which it differed from the reference one (Chapters 3 and 4). In this work, the active spaces of the optimized geometries of rubiadin and soranjidiol in chloroform (Figures 8.4 and 8.5, respectively) were used as the reference active space. Using these active spaces as the initial guess, 88 geometries of rubiadin presented the correct active space after the first `pyoverlaps.py` iteration (and CASSCF wavefunction optimization), 2 geometries were recovered after the second iteration, and 10 were not recovered after 5 iterations with the `pyoverlaps.py` script. On the other hand, for soranjidiol 94 geometries presented the correct active space after the first CASSCF wavefunction optimization, and it was not possible to recover the reference active space for 6 geometries. Figure 8.6 shows the absorption spectrum of rubiadin (top) and soranjidiol (bottom) computed at the MS-CASPT2 level of theory with the (12,9) active space, in comparison with the corresponding experimental ones in a solution of chloroform. The intensities of the computed spectra have been normalized to the number of sampled geometries for each sampled anthraquinone derivative. It can be evidenced that for both molecules the spectrum inside the lipid membrane is red-shifted with respect to the spectrum in chloroform. Interestingly, the *in vitro* experiments carried out with these two PSs<sup>78</sup> have been performed using an excitation window of  $420 \pm 17$  nm, that is an intermediate region between the lineshapes of the experimental and the computed spectra in figure 8.6. However, to obtain a consistent comparison, the computation of the absorption spectrum of rubiadin and of soranjidiol in chloroform, considering an ensemble of geometries, would be auspicious. This is currently work in progress, and will provide a complete picture of the behavior of the excited states of rubiadin and soranjidiol in different media.

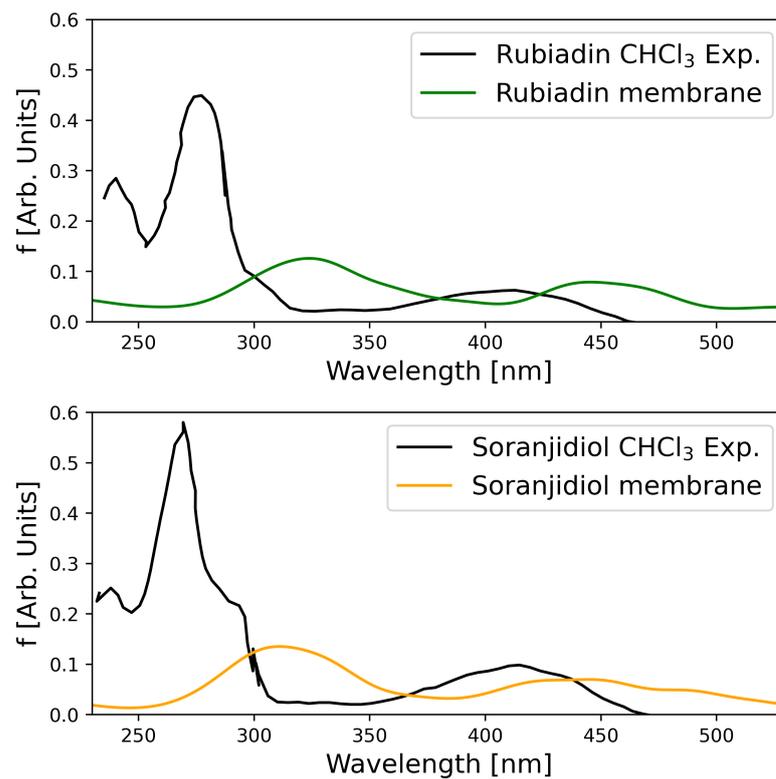


Figure 8.6: Top: (black) Experimental absorption spectrum of rubiadin in chloroform; (green) Absorption spectrum of rubiadin at the minimum of the umbrella sampling free energy profile inside the DOPC lipid membrane. Bottom: (black) Experimental absorption spectrum of soranjidiol in chloroform; (orange) Absorption spectrum of soranjidiol at the minimum of the umbrella sampling free energy profile inside the DOPC lipid membrane.

---

## 8.4 Conclusions and Future Work

In the present work, the permeation mechanism of rubiadin and of soranjidiol into a DOPC lipid membrane has been studied by means of umbrella sampling classical MD. It has been evidenced that the permeation of both molecules is thermodynamically favorable, with rubiadin presenting a reaction free energy of permeation of 8 kcal/mol and soranjidiol a reaction free energy of 6 kcal/mol. This is in agreement with the capability of anthraquinone derivatives to accumulate inside the cellular lipid membrane, whereby following photoexcitation, they are capable of inducing cellular apoptosis by triggering the disruption of the lipid membranes. The similarity in the PMF for rubiadin and soranjidiol is also in agreement with the experimental observation that both molecules evidence similar cytotoxicities.<sup>78</sup> However, the study of the permeation process is associated with the cell uptake of the anticancer drug, and there are several other factors that contribute to the cytotoxicity of the PS. In particular, it is of utmost importance to investigate the electronic excited states of rubiadin and soranjidiol in the presence of the lipid membrane. In general the mechanism of action of anthraquinone derivatives involves an intersystem crossing towards the triplet manifold; therefore, the MS-CASPT2 methodology has been chosen to study the electronic excited states of rubiadin and soranjidiol, since only multiconfigurational methods can properly describe the wavefunction of a system in the presence of two quasi-degenerate electronic states (*e.g.* at crossing regions).

In order to determine an appropriate active space, the absorption spectra of rubiadin and soranjidiol were computed in the presence of explicit chloroform with different active spaces. The spectra were computed on top of a QM/MM MP2/6-31G\* optimized geometry of each anthraquinone in a solvation box of chloroform. It was evidenced that for both molecules the first experimental absorption band was reproduced by a MS-CASPT2 computation using a (12,9) active space. The same active space was employed as the reference active space to compute the absorption spectrum of rubiadin and soranjidiol at the minimum of the free energy profile of permeation inside the lipid membrane, whereby the automatic correction of the active space provided by the `pyoverlaps.py` script has provided the correct active space for 90 geometries of rubiadin and 94 geometries of soranjidiol (out of 100 sampled geometries in both cases). The absorption spectra of both anthraquinones

---

inside the lipid membrane presents a red shift with respect to the experimental spectra in chloroform, although a consistent comparison would require the computation of the absorption spectrum on an ensemble of geometries in chloroform, which is a current work in progress. A next step consists of the QM/MM computation of the intersystem crossing pathway to unveil the mechanism, including the effect of the lipid environment on the triplet state population.

## Chapter 9

# Conclusions

In this thesis, some of the key steps of the mechanism of action of cisplatin and of anthraquinone and some derivatives have been studied by means of state-of-the-art computational modeling techniques. Specifically, the permeation mechanism of cisplatin into a lipid membrane was studied by means of the umbrella sampling MD technique. The energetics of the permeation pathway itself, as well as the energetics of the interactions that are established between the cisplatin molecule and the lipid membrane have also been computed. In particular, the latter were computed using a molecular mechanics analytic potential, as well as a more accurate QM/MM procedure, in which case it was also possible to decompose the interaction energy in physically meaningful energy terms that provide further insights into the nature of the interactions established. In regard with the anthraquinone molecule, its conformational freedom has been studied in the intercalation binding mode in a poly(dG-dC) double strand model of DNA by means of classical MD, whereby there were identified two prominent configurations. The electronically excited states of the photosensitizer were computed considering these two predominant configurations. The permeation mechanism of two derivatives of anthraquinone – rubiadin and soranjidiol – has also been computed by means of the US MD technique. The absorption spectrum of both anthraquinone derivatives has been computed in configurations corresponding to the US window of the minimum of the free energy profile obtained from the study of the permeation mechanism. Of great importance for the present thesis work has also been the development and application of the MoBioTools toolkit, which allows

---

for a straightforward setting up of the QM/MM calculations considering an ensemble of geometries and, thus, has been extensively used in most of the work presented here. Thus, the following conclusions of the present thesis work can be drawn:

- The simulation of the permeation mechanism of cisplatin into a lipid membrane with a force field has provided a detailed account on the energetics of the overall process. In particular, it was observed that, as cisplatin permeates the membrane, it is weakly trapped inside a free energy minimum of 0.65 kcal/mol; afterwards, it overcomes a small barrier of 0.2 kcal/mol and gets trapped inside the global energetic minimum of the permeation pathway, which coincides with cisplatin being incorporated in the polar region of the lipid membrane. At this stage, the decomposition of the interaction energy between cisplatin and the lipid molecules suggested that energetics of the process were dominated by the electrostatic interactions between cisplatin and the polar heads of the lipids. However, as cisplatin approached the center of the membrane – corresponding to the non-polar tails of the lipids –, it was observed that the free-energy of the permeation process increased, so that ultimately the penetration of cisplatin across the membrane would imply that an overall energetic barrier of 10.4 kcal/mol barrier would need to be overcome. In this region, the attractive interactions between cisplatin and the membrane are dominated by non-polar van der Waals interactions, however, these attractive interactions were ultimately overruled by the unfavorable energy associated with the dehydration of cisplatin and the loss of entropy due to the increase of ordering in the lipid bilayer induced by the presence of cisplatin. Thus, certain types of functionalization such as the anchorage of long alkyl chains to the cisplatin scaffold might mitigate the unfavorable entropic contribution associated with the membrane ordering increase. This strategy has been applied in the case of platinum(IV) complexes with variable success.<sup>400</sup>
- The computation of the interaction energy between cisplatin and the lipid molecules within a QM/MM framework, and the subsequent energy decomposition using a deformation density-based scheme, has provided further insights into the nature of the interactions established between the two subsystems involved. In particular, it was observed that the interactions at the minimum and at the maximum of the

---

free-energy profile are mainly dominated by the electrostatic interactions. It was also evidenced that the Pauli repulsive interactions are in fact more significant than when computed using an analytic force field, thus indicating that this contribution is underestimated by the latter. Finally, a broad energy distribution was obtained for each energy component of the interaction energy when considering an ensemble of geometries. This evidences that conformational sampling is fundamental when treating complex biological systems to obtain a complete characterization of the interaction energy and its EDA.

- In the case of the anthraquinone molecule, it was observed that when intercalated in a DNA double strand, two conformations were predominant along the entire MD simulation time. These conformations were considered to study the excited states of the anthraquinone molecule by means of QM/MM at the TD-DFT level of theory, including a guanine molecule as well as anthraquinone in the QM region. Of particular importance are the CT states between guanine and anthraquinone, as these are the relevant states involved in the photoinduced electron transfer that leads to DNA lesion. It was observed that the CT contribution strongly depends on the conformation assumed by the anthraquinone molecule in the intercalated binding pocket, as the CT contribution was of 30-34% when the anthraquinone molecular axis was parallel to the hydrogen bond axis of a nearby guanine-cytosine nucleobase pair (in this situation, anthraquinone displays the highest overlap with one of the neighbouring guanine molecules), whereas the CT character reduced to 23-28% when the axis of anthraquinone was not parallel to that of the guanine-cytosine pair. With these results at hand, a strategy to obtain anthraquinone-based PSs with enhanced cytotoxic activity could be to functionalize the anthraquinone scaffold so that the conformational freedom of the anthraquinone derivative in the intercalated binding pocket be limited to configurations in which stacking is favoured.
- In regard with the anthraquinone derivatives rubiadin and soranjidiol, the study of the permeation of each of these molecules inside a DOPC lipid membrane has provided two free energy profiles showing that the diffusion process is favorable for both derivatives, whereby the reaction energy for rubiadin is 8 kcal/mol and for soranjidiol

---

6 kcal/mol. These results show the similarity of the energetics of the permeation process for rubiadin and soranjidiol, a finding in agreement with the experimental observation that rubiadin and soranjidiol present a similar cytotoxicity.<sup>78</sup> However, to obtain a more complete picture in regard with the anticancer activity of these molecules, their electronic excited states need to be studied inside the lipid bilayer. In this regard, the absorption spectra of the two molecules inside the lipid membrane have been computed whereby a red shift with respect to the experimental spectrum recorded in chloroform was observed. Further studies are aimed at investigating the intersystem crossing mechanism of the two anthraquinone derivatives inside the lipid membrane.

- Finally, the MoBioTools software<sup>79,80</sup> developed and applied throughout the thesis work has shown to be a fundamental tool to interface previously performed MD simulations with QM/MM computations on ensembles of geometries extracted from such MD simulations. Its versatility has been evidenced not only through the worked examples presented during its introduction (Chapter 3), but also in its application on the wide variety of situations posed by the different systems treated in the present thesis. It is available as an open-source code.<sup>80</sup>

# Conclusiones

En el presente trabajo de tesis, algunas de las etapas fundamentales del mecanismo de acción del cisplatino, de la antraquinona y de dos de sus derivados, han sido estudiadas mediante métodos a la vanguardia de modelización computacional. En particular, el mecanismo de difusión del cisplatino al interior de una membrana lipídica ha sido estudiado mediante MD clásica en conjunto con la técnica de *Umbrella Sampling*. La energética del camino de difusión, junto con las energías de interacción entre el cisplatino y las moléculas de la membrana, también han sido calculadas. En particular, las energías de interacción han sido obtenidas utilizando un potencial analítico de mecánica molecular, y también empleando un método más exacto de QM/MM, en cuyo caso ha sido posible descomponer la energía de interacción en diferentes contribuciones energéticas con un significado físico, lo que provee información relevante acerca de la naturaleza de las interacciones que se establecen. Con respecto de la molécula de antraquinona, se ha estudiado la libertad conformacional que ésta posee cuando se encuentra intercalada entre pares de nucleobases en un modelo de poly(dG-dC) de doble hélice del ADN, mediante MD clásica, en cuyo caso han sido identificadas dos principales conformaciones. Los estados excitados de la molécula de antraquinona han sido calculados considerando estas dos conformaciones. El mecanismo de difusión de dos derivados de la antraquinona – rubiadina y soranjidiol – también ha sido calculado mediante la técnica de *umbrella sampling* MD. El espectro de absorción de cada una de dichas moléculas ha sido calculado para configuraciones correspondientes al mínimo del perfil de energía libre obtenido del estudio del mecanismo de difusión. De gran importancia para el presente trabajo de tesis ha sido también el desarrollo y la aplicación del conjunto de herramientas computacionales MoBioTools, el cual permite una preparación rápida del sistema estudiado para realizar cálculos QM/MM a partir de un conjunto de

---

geometrías, por lo que ha sido utilizado ampliamente durante la realización de la presente tesis. Con lo anterior, se pueden identificar las siguientes conclusiones:

- La simulación del mecanismo de difusión del cisplatino dentro de una membrana lipídica ha proporcionado información detallada sobre la energética de dicho proceso. En particular, se ha observado que, a medida que el cisplatino penetra la membrana lipídica, éste es atrapado débilmente al interior de un mínimo de energía libre de 0.65 kcal/mol; sucesivamente, el cisplatino supera una pequeña barrera energética de 0.2 kcal/mol para después ser atrapado en el interior del mínimo global de energía libre, que corresponde a la situación en la cual el cisplatino se encuentra en la región polar de la membrana lipídica. En este punto, la descomposición de la energía de interacción entre el cisplatino y la membrana lipídica sugiere que la energética del proceso es dominada por las interacciones electrostáticas entre el cisplatino y las cabezas polares de los lípidos. Sin embargo, a medida que el cisplatino se aproxima al centro de la membrana – correspondiente a la región con las colas apolares de los lípidos – se observa que la energía libre del proceso de difusión aumenta, por lo que al final, la difusión del cisplatino a través de la membrana lipídica implicaría una barrera energética de 10.4 kcal/mol tendría que ser superada. En dicha región, las interacciones atractivas entre el cisplatino y la membrana son dominadas por interacciones no electrostáticas de van der Waals; sin embargo, dichas interacciones atractivas son anuladas por la energía desfavorable de deshidratación del cisplatino y la pérdida de entropía debida al aumento del orden en la membrana debido a la presencia del cisplatino. Con base en estos resultados, ciertos tipos de funcionalización tales como el anclaje de una cadena alquílica al esqueleto del cisplatino podría mitigar la contribución entrópica desfavorable a la energía asociada con el aumento del orden de la membrana lipídica. Esta estrategia ha sido aplicada en el caso de compuestos de platino(IV), con éxito variable.<sup>400</sup>
- El cálculo de las energías de interacción entre el cisplatino y los lípidos de la membrana en un ámbito de QM/MM, y la sucesiva descomposición de la energía de interacción utilizando un método basado en densidades electrónicas de deformación, ha revelado más información acerca de la naturaleza de las interacciones que se establecen entre

---

los dos fragmentos involucrados. En particular, se ha observado que las interacciones en el mínimo y en el máximo del perfil de energía libre son dominadas principalmente por interacciones electrostáticas. Además, se ha evidenciado que los términos de interacción repulsivos de Pauli son, de hecho, más significativos que en el caso en el que la energía de interacción es calculada utilizando un campo de fuerzas analítico, lo que indica que la contribución de Pauli es subestimada por este último. Finalmente, el hecho de que se haya obtenido una distribución de energías para cada componente de la energía de interacción, al considerar un conjunto de geometrías, indica que el muestreo conformacional es fundamental para estudiar la energía de interacción y su descomposición en sistemas biológicos complejos.

- En el caso de la molécula de antraquinona, se ha observado que cuando se encuentra intercalada en una doble hélice de ADN, dos conformaciones han sido predominantes a lo largo de la trayectoria de MD. Dichas conformaciones han sido consideradas para estudiar los estados excitados de la molécula de antraquinona mediante métodos de QM/MM utilizando TD-DFT, incluyendo una molécula de guanina y la molécula de antraquinona en la región QM. Los estados excitados de CT entre estas dos moléculas son de particular importancia, ya que estos son los estados relevantes involucrados en los procesos de transferencia electrónica fotoinducida que conlleva a la lesión del DNA. Se ha observado que la contribución de los estados CT depende fuertemente de la conformación que asume la antraquinona cuando se encuentra intercalada, puesto que la contribución de CT es del 30-34% cuando el eje de la antraquinona es paralelo al eje de los puentes de hidrógeno de uno de los pares de guanina-citosina que rodean la antraquinona (en esta situación, la antraquinona se encuentra con el máximo solapamiento posible con una de las guaninas adyacentes), mientras que el carácter de CT se reduce al 23-28% cuando el eje de la antraquinona no es paralelo al del eje del par guanina-citosina. Como consecuencia, una estrategia para obtener PS derivados de la antraquinona con mejor actividad citotóxica podría consistir en funcionalizar la molécula de antraquinona de manera que la libertad conformacional del derivado de antraquinona, cuando se encuentra intercalado en una doble hélice de DNA, sea limitada solo a configuraciones para las cuales el solapamiento sea favorecido.

- 
- Con respecto a los derivados de la antraquinona rubiadina y soranjidiol, el estudio de la difusión de cada una de dichas moléculas al interior de una membrana lipídica de DOPC ha producido dos perfiles de energía libre mostrando que la difusión de cada una de las moléculas es un proceso termodinámicamente favorable, mostrando una energía libre de reacción de 8 kcal/mol para la rubiadina y 6 kcal/mol para el soranjidiol. Estos resultados muestran que desde el punto de vista energético los procesos de difusión de la rubiadina y del soranjidiol, un hecho en línea con la similar citotoxicidad de las dos moléculas observada experimentalmente.<sup>78</sup> Sin embargo, para obtener una descripción más completa de la actividad anticancerígena de estas moléculas, es necesario estudiar sus estados electrónicos excitados en el interior de la bicapa lipídica. En este contexto, los espectros de absorción de los dos derivados de la antraquinona han sido calculados en el interior de la membrana lipídica, y muestran un desplazamiento hacia el rojo con respecto de los correspondientes espectros experimentales en cloroformo. En el futuro se planea investigar el mecanismo de cruce entre sistemas de los dos derivados de la antraquinona al interior de la membrana lipídica.
  - Finalmente, el programa MoBioTools,<sup>79,80</sup> desarrollado y aplicado a lo largo de la tesis, ha demostrado ser una herramienta fundamental para interrelacionar simulaciones de MD realizadas anteriormente con cálculos QM/MM en conjuntos de geometrías extraídas de dichas simulaciones. Su versatilidad ha sido evidenciada no solo a través de los ejemplos presentados durante su introducción (Capítulo 3), sino también a través de su aplicación en la variedad de situaciones propuestas por los diferentes sistemas tratados en la tesis. Actualmente se encuentra disponible en la modalidad *open-source*.<sup>80</sup>

## Appendix A

# Supplementary Information for Chapter 4

### A.1 Active Spaces of the Five Canonical Nucleobases

All active spaces were optimized at the SA-CASSCF level of theory. The choice of the number of roots and the active space electrons and orbitals for each nucleobase was made following Thiel *et. al.*,<sup>282</sup> except for thymine, for which a smaller active space was employed, as explained below. The  $C_s$  symmetry of the nucleobases was not considered in the SA-CASSCF calculations, so the state-average procedure was applied to all states considered, for each nucleobase.

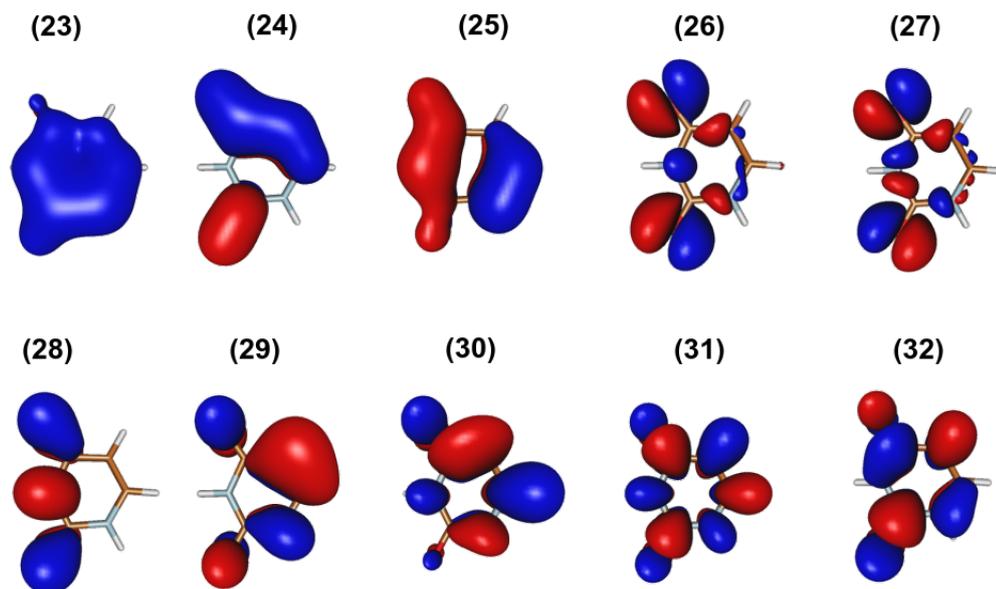


Figure A.1: The (14,10) reference active space of uracil, calculated in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 10 roots.

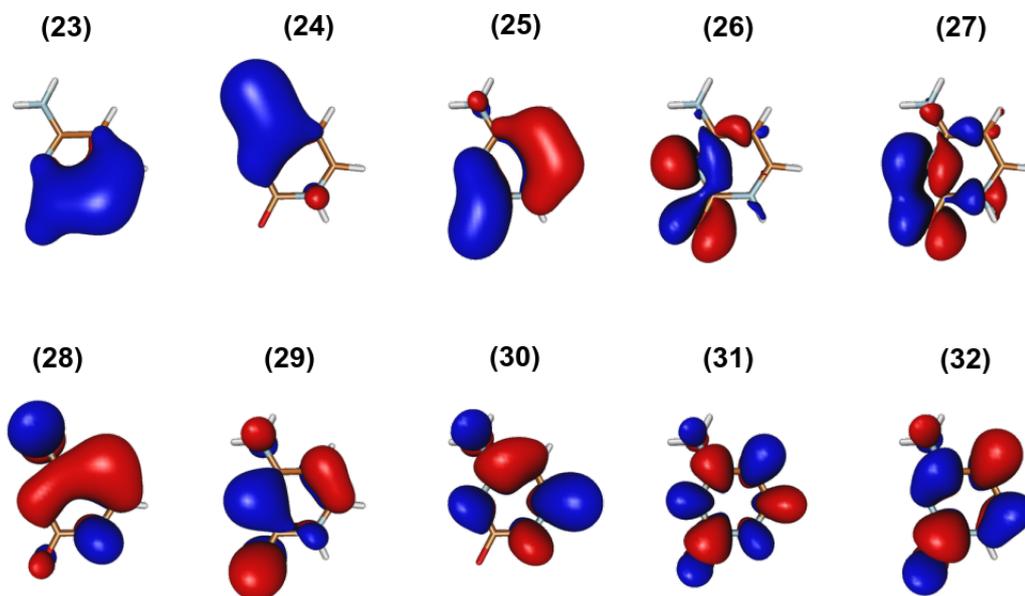


Figure A.2: The (14,10) reference active space of cytosine, calculated in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 8 roots.

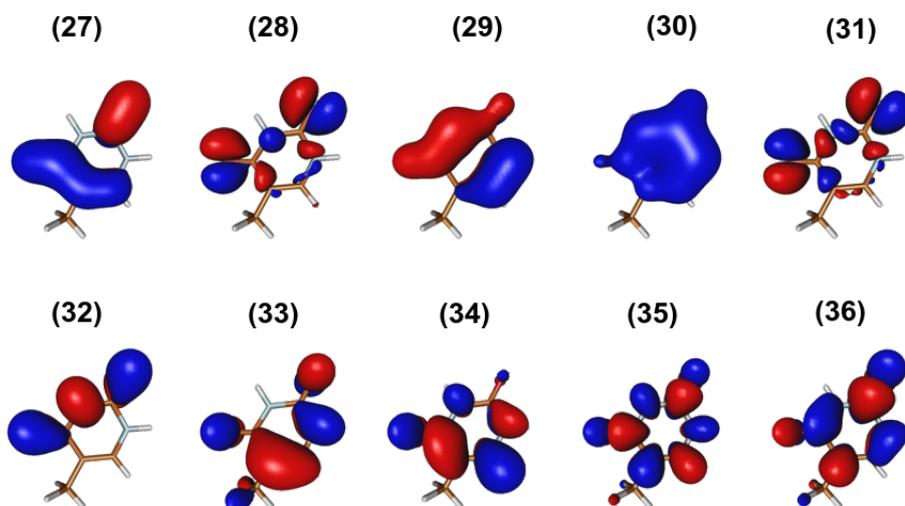


Figure A.3: The (14,10) reference active space of thymine, calculated in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 10 roots.

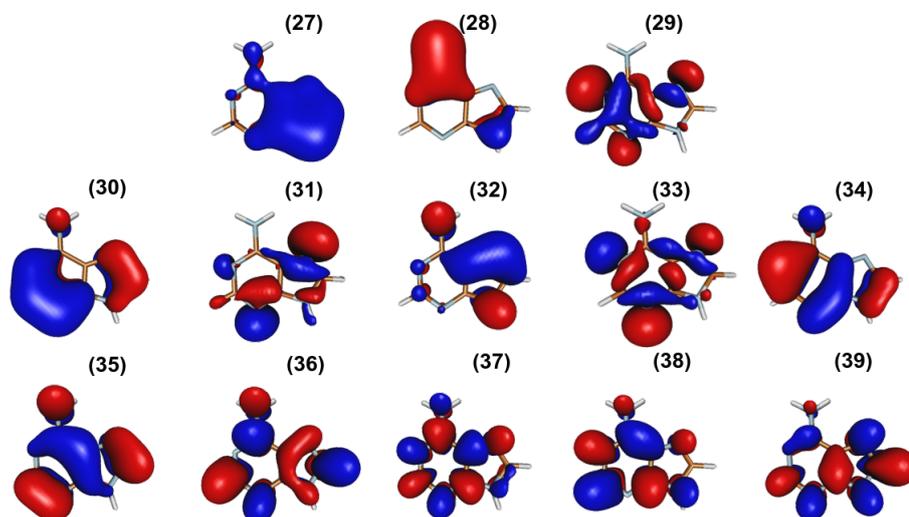


Figure A.4: The (18,13) reference active space of adenine, calculated in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 11 roots.

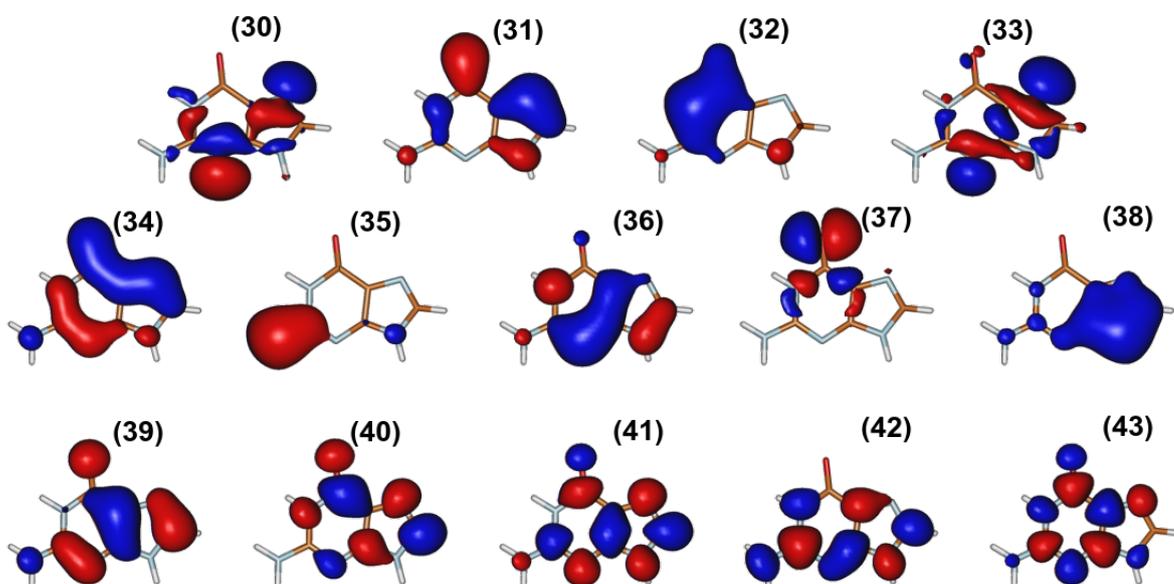


Figure A.5: The (20,14) reference active space of guanine, calculated in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 9 roots.

---

## A.2 Choice of the Reference Active Space of Thymine

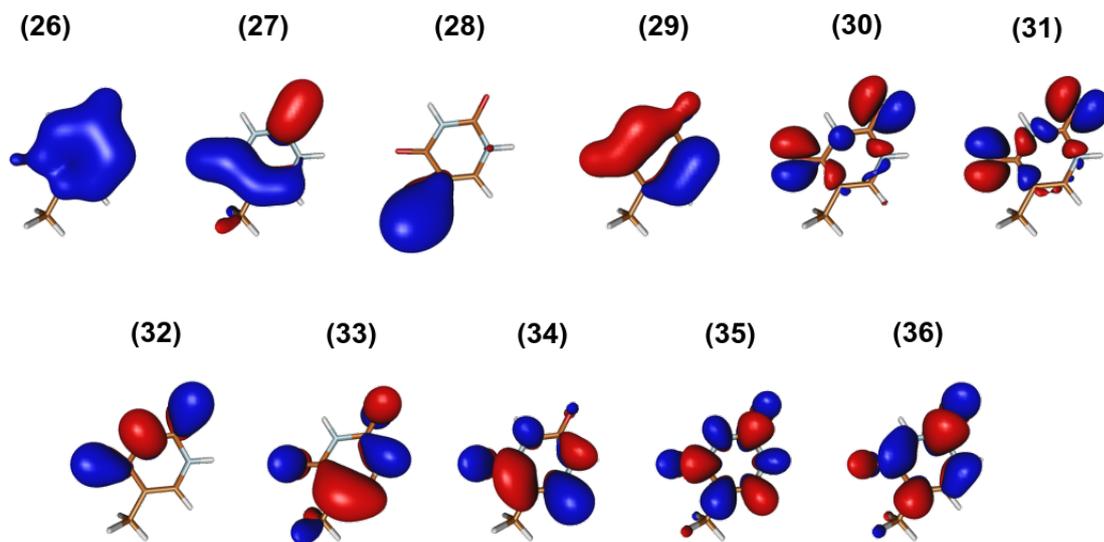


Figure A.6: The (16, 11) reference active space of thymine, according to Thiel’s reference calculations (large active space), computed in gas phase at the MP2/6-31G\* optimized geometry with the SA-CASSCF scheme considering the first 10 roots.

Thymine poses a particular case when considering as the reference active space the (16,11) space - which will be referred to as the large reference active space - employed by Thiel<sup>282</sup> and by Wiebler<sup>281</sup> (Figure A.6). This is due to the fact that the SA-CASSCF calculation on 100 MD sampled geometries of thymine on water, and the subsequent  $\mathbf{S}^{MO}$  analysis, did not provide the correct active space after 5 iterations of the `pyoverlaps.py` script for 99 geometries. This indicates that, while the active space could be appropriated for the ground-state optimized geometry in vacuum, it is not a suitable active space for the sampled geometries in water. It was evidenced that in all cases the molecular orbital (MO) presenting mainly a contribution from a p orbital localized on the methyl group (labeled as 28 in the reference active space, figure A.6) was placed outside of the active space during the CASSCF orbital optimization. Therefore, it could be hypothesized that at least for the sampled geometries, this orbital should not be placed within the active space. This hypothesis stems from two observations:

1. The rotation of the methyl group about the C-C bond axis must be considered when studying the molecular motion of thymine. This could pose a problem when

---

determining the MO analogous to the reference MO 28, as for some conformations this MO could ambiguously correspond to two MOs whose largest contributions stem from the two p orbitals of the methyl C atom perpendicular to the C-C bond (e.g: MOs 23 and 24 of geometry 1, figure A.7).

2. The choice of including MO 28 in the active space of thymine results counterintuitive once the main AO contributions to that MO are considered: indeed, Table A.1 shows that the MO 28 mainly has contributions from the s orbitals of two H atoms of the methyl group (H5 and H6, Figure A.8) and from p orbitals of the methyl C atom (C11), and only to a negligible extent has it contributions from the p orbitals of C9. Thus, this MO has a predominant  $\sigma$  character, and is not expected to play a major role in any of the transition in which the first 9 excited states of thymine are involved, which are either of  $\pi \rightarrow \pi^*$  or of  $n \rightarrow \pi^*$  character.

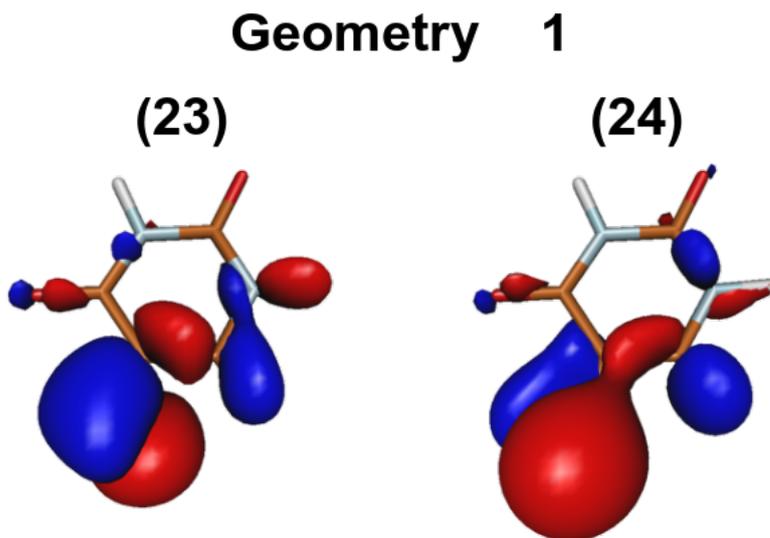


Figure A.7: MOs 23 and 24 of the sampled geometry 1 of thymine, after the fourth SA-CASSCF optimization cycle. Qualitatively, both orbitals are analogous to the MO 28 of the (large) reference active space.

Observation 2 prompts to consider that MO 28 should not be included even within the reference active space of the equilibrium geometry, as its contributions to the 9 excited states considered should be negligible. Therefore, a MS-CASPT2 calculation with an imaginary level shift<sup>244</sup> of 0.2 Hartree and no IPEA correction<sup>245</sup> was performed on top of

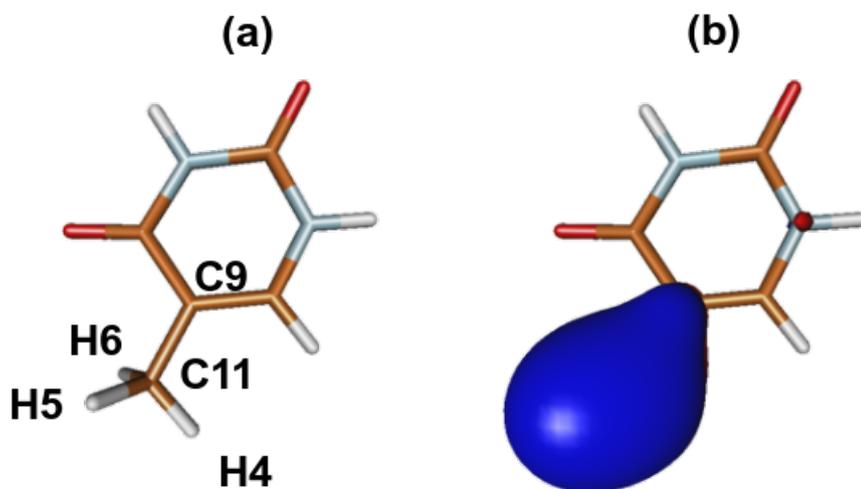


Figure A.8: (a) Equilibrium geometry of thymine, with the atom labels of the atoms having the highest AO contribution to the reference MO 28 of thymine. (b) MO 28 of the (large) reference active space of thymine.

Basis function	Atom	AO character	Coefficient
25	H5	1s	0.14
26	H5	*s	0.21
27	H5	*s	0.09
31	H6	1s	-0.14
32	H6	*s	-0.21
33	H6	*s	-0.09
82	C9	*px	0.06
85	C9	*py	0.06
120	C11	*px	0.31
123	C11	*py	0.27
126	C11	*pz	0.10

Table A.1: Atomic orbital contributions ( $|Coeff.| > 0.05$ ) to the MO 28 of the large reference active space of thymine.

the SA-CASSCF optimized wavefunction, with the large reference active space. Table A.2 shows the results for the first nine excited states of thymine. It should be emphasized that the excitation energies slightly differ from those reported by Wiebler *et. al.*,<sup>281</sup> mainly because in the present case the calculations have not been performed considering the  $C_s$  symmetry of the molecule, so all the 10 states under study have been averaged; this in

turn makes the excitation energies obtained consistent with one another, as the reference ground state is the same for all the excited states. The configurations displayed in table A.2 show the MOs in the large reference active space, following the MO ordering in figure A.6; thus, the first position in a configuration corresponds to MO 26, the second position is MO 27 and so on, until MO 36. It can be evidenced that none of the reported excited states presents a configuration associated with an excitation from MO 28 whose weight is larger than 0.1, so that the contributions from any such configuration are seemingly negligible.

State	Energy (eV)	Configuration	Weight
S <sub>0</sub>	0.000	22222222000	0.90
S <sub>1</sub>	4.667	22222u22d00	0.90
S <sub>2</sub>	4.844	2222222ud00	0.70
S <sub>3</sub>	5.714	22222u2200d	0.89
S <sub>4</sub>	6.004	222222u2d00	0.62
S <sub>5</sub>	6.016	22222u2d200	0.12
		22222u2200d	0.38
		2222u222d00	0.16
S <sub>6</sub>	6.519	2222u222d00	0.60
S <sub>7</sub>	6.868	2222222u00d	0.61
S <sub>8</sub>	6.933	222222ud200	0.11
		22222220u0d	0.12
		2u222222d00	0.22
S <sub>9</sub>	8.147	2222222ud00	0.12
		2222222u00d	0.25
		222222u200d	0.40

Table A.2: Vertical excitation energies and configurations with weight  $> 0.1$  for the first nine excited states of thymine, calculated at the MS-CASPT2 level of theory with the large reference active space. Each configuration is represented by the orbitals within the active space - following the MO ordering in figure A.6 - with their corresponding occupations: 2 represents a doubly occupied MO, u a singly occupied MO with spin up, d a spin down occupied MO, and 0 a unoccupied MO.

To study the effect of removing MO 28 from the reference active space a SA-CASSCF/MS-CASPT2 calculation was performed with the (14,10) active space reported in figure A.3 - which will be referred to as the small reference active space - considering the same settings

as for the large active space calculations. It can be evinced from table A.3 that the variation in the excitation energies with respect to the large active space calculations is at most of 0.1 eV ( $S_9$ ) although in most cases it is no larger than 0.02 eV. Moreover, the configurations whose weights are greater than 0.1 are exactly the same for each excited state in both the large and the small active space calculations (it should be emphasized that the ordering of the MOs is different in the large and the small active spaces, so that this variation needs to be considered when comparing the configurations in Tables A.2 and A.3). This allows for concluding that the small active space (14,10) should indeed be used as the reference space for thymine.

State	Energy (eV)	Configuration	Weight
$S_0$	0.000	222222000	0.90
$S_1$	4.694	2222u22d00	0.90
$S_2$	4.838	22222ud00	0.70
$S_3$	5.696	2222u2200d	0.89
$S_4$	5.998	22222u2d00	0.62
$S_5$	6.021	2222u2d200	0.13
		2222u2200d	0.41
		2u22222d00	0.14
$S_6$	6.538	2u22222d00	0.63
$S_7$	6.695	22222u00d	0.60
$S_8$	6.940	2222220u0d	0.14
		22u2222d00	0.10
		u222222d00	0.18
$S_9$	8.253	22222u2d00	0.14
		22222u00d	0.34
		22222u200d	0.26

Table A.3: Vertical excitation energies and configurations with weight  $> 0.1$  for the first nine excited states of thymine, calculated at the MS-CASPT2 level of theory with the small reference active space. Each configuration is represented by the orbitals within the active space - following the MO ordering in figure A.3 - with their corresponding occupations: 2 represents a doubly occupied MO, u a singly occupied MO with spin up, d a spin down occupied MO, and 0 a unoccupied MO.

---

### A.3 Analysis of the CI Weights for Two Geometries of Uracil

This section aims at showing in detail the analysis of the CI coefficients for geometries 73 and 74 of uracil to complement the discussion following Figure 4.4 of chapter 4 (see Table A.4). Specifically, it will be computed the difference between the weight of the most prominent configuration of the  $S_1$  state following the SA-CASSCF calculation with the corrected active space, with the weight that the same configuration (in the sense that MOs that are equivalent in terms of the MO overlap analysis present the same occupations) has in the SA-CASSCF calculation with the uncorrected active space (deriving from the last but one iteration of the `pyoverlaps.py` program). In table A.4 (and in Figure 4.4 of chapter 4) these differences are referred to as difference in maximum weight. Geometries 73 and 74 have been chosen since in both cases a high difference in the excitation energies of the  $S_1$  state is displayed, independently of whether the difference in maximum weight is high (geometry 73, 0.599) or low (geometry 74, 0.155). This fact evidences that large energy differences can be obtained even when the character of the wavefunction is qualitatively similar between the uncorrected and the corrected SA-CASSCF calculations, that is, even when the problematic MOs are not involved in the electronic excitations that dominate the description of the wavefunction. Thus, the two wavefunctions are inherently different, and this difference is evidenced in the excitation energies, making it desirable to use an appropriate active state for the calculation of such observables.

---

Geometry	DDE [eV] <sup>a</sup>	Maximum weight last iteration	Maximum weight l.b.o. <sup>b</sup> iteration	Difference in maximum weight	Corresponding configuration <sup>c</sup>
73	-0.62	0.627	0.028	0.599	222222ud00
74	-0.52	0.393	0.238	0.155	22222u2d00

Table A.4: Detailed analysis of the difference between the weight of the dominant configuration of the corrected wavefunction for the  $S_1$  state of the corrected SA-CASSCF optimization and the weight of the same configuration for the  $S_1$  state on the uncorrected SA-CASSCF wavefunction for two recovered geometries of uracil. The configurations are the same in the sense that the MOs that participate to the excitations are equivalent in the sense of the MO overlap comparison. The difference in excitatio energy between the two calculations is also displayed.

<sup>a</sup> DDE = Difference betwee the  $S_1$  excitation energy of the last SA-CASSCF calculation and the last but one SA-CASSCF calculation.

<sup>b</sup> l.b.o. = Last but one.

<sup>c</sup> **OpenMolcas** notation: 2 = doubly occupied; u = singly occupied, spin +1/2; d = singly occupied, spin -1/2; 0 = unoccupied. Only the MOs of the active space are shown.

## Appendix B

# Supplementary Information for Chapter 5

### B.1 Umbrella Sampling Setup

It is well known that the number of windows and the force constant of the bias harmonic potential must be properly chosen to have a good overlap among the reaction-coordinate probability distributions of consecutive windows. Such a good overlap is required to obtain an accurate free-energy profile along the reaction coordinate, also called potential of mean force, from the biased probability distributions, especially when the Weighted Histogram Analysis Method (WHAM) approach is employed.<sup>45</sup> Figure B.1 shows the computed probability distributions for the umbrella sampling windows centred at 10.0, 10.5, 11.0, 11.5 and 12.0 Å as example windows. It can be seen that the distributions of neighbouring windows present a strong overlap, which indicates a good sampling along the reaction coordinate and validates the choice of the simulation parameters employed.

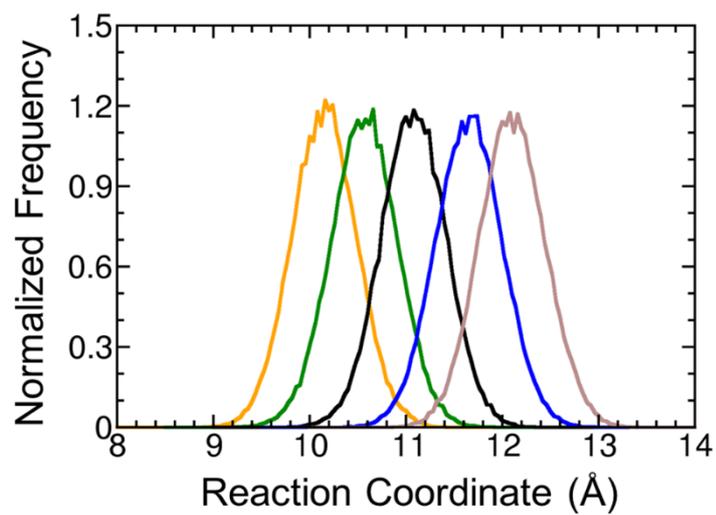


Figure B.1: Reaction coordinate probability distributions for the windows centred at 10.0, 10.5, 11.0, 11.5 and 12.0 Å.

---

## B.2 Permeability Coefficient

The permeability coefficient of  $3.6 \times 10^{-7} \text{ ms}^{-1}$  was computed upon integration of the resistance coefficient along the permeation pathway. In turn, the resistance coefficient can be computed from the Z-dependent diffusion coefficient (plotted in Figure B.2) and from the potential of mean force (plotted in Figure 5.2b). The details about the full protocol are discussed elsewhere.<sup>325</sup>

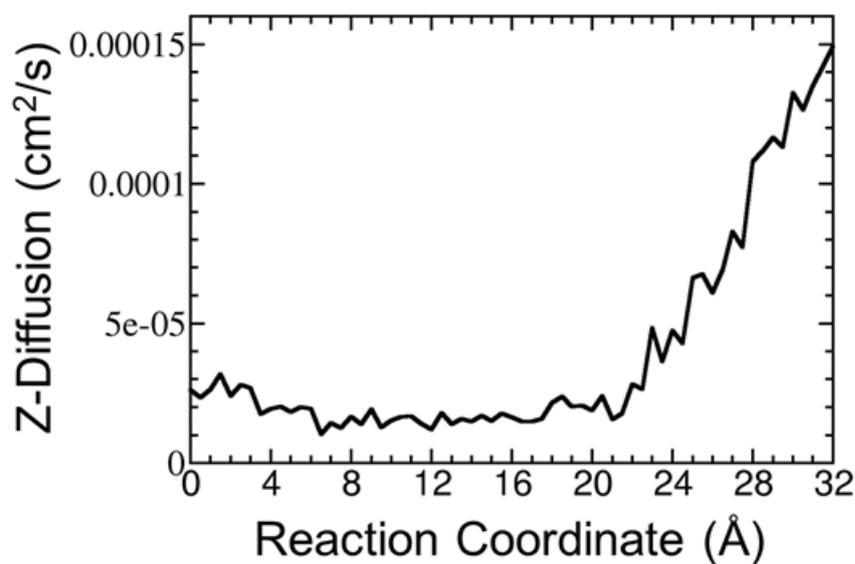


Figure B.2: Z-dependent diffusion coefficient of cisplatin.

### B.3 Force Field Parameters

Bond Parameters			
Bond	$k_b$ (kcal/[molÅ <sup>2</sup> ])	$R_0$ (Å)	
Pt-Cl	123.100	2.305	
Pt-N	88.900	2.111	
N-H	392.400	1.019	
Angle Parameters			
Angle	$k_a$ (kcal/[mol rad <sup>2</sup> ])	$\theta_0$ (degrees)	
N-Pt-H	132.790	98.410	
Cl-Pt-N <sub>cis</sub>	105.950	83.010	
Cl-Pt-N <sub>trans</sub>	106.350	178.580	
Cl-Pt-Cl	88.200	95.570	
Pt-N-H	52.880	110.140	
H-N-H	41.400	106.40	
Intermolecular Parameters			
Atom	q (a.u.)	$R_{min}/2$ (Å)	$\epsilon$ (kcal/mol)
Pt	0.069	2.053	1.055
Cl	-0.375	2.597	0.038
N	-0.410	1.896	0.045
H1 <sup>a</sup>	0.239	0.052	0.018
H2 <sup>b</sup>	0.256	0.052	0.018

Table B.1: Force field parameters for cisplatin: bond force constants  $k_b$  , equilibrium distances  $R_0$  , angle force constants  $k_a$  , equilibrium angles  $\theta_0$  , atomic charges q and Lennard-Jones parameters  $R_{min}/2$  and  $\epsilon$ .

<sup>a</sup> Hydrogen atoms that do not belong to the plane formed by N(Cl)-Pt-N(Cl).

<sup>b</sup> Hydrogen atoms that belong to the plane formed by N(Cl)-Pt-N(Cl).

## Appendix C

# Supplementary Information for Chapter 8

### C.1 Electronic Excited States and CI Weights of Rubiadin and Soranjidiol in Explicit Chloroform

In this section the absorption spectra and the CI weights of the most prominent configurations of rubiadin and soranjidiol in the presence of explicit chloroform are listed. The absorption spectra reported in Tables C.1 and C.1 have been computed at the MS-CASPT2 level of theory with a (12,9) active space (Figures 8.4 and 8.5) and a cc-pVTZ<sup>231</sup> basis set. The geometries on top of which the absorption spectra have been computed have been optimized with an electrostatic embedding QM/MM approach, whereby the QM region has been described at the MP2/6-31G\* level of theory and a cutoff of 4.0 Å has been employed to include MM chloroform molecules inside the mobile region, so that these molecules were included in the gradient calculation and the geometry optimization.

---

State	Wavelength [nm]	Osc. Strength	Configuration	Weight
S <sub>0</sub>	0.0		222222000	0.85
S <sub>1</sub>	401.43	1.57E-01	22222ud00	0.26
			22u222d00	0.14
			u22222d00	0.27
S <sub>2</sub>	394.06	9.14E-02	2u2222d00	0.52
S <sub>3</sub>	327.22	2.24E-01	2u222d200	0.42
S <sub>4</sub>	304.32	8.20E-03	22222ud00	0.21
			u22222d00	0.22
			u222220d0	0.15
S <sub>5</sub>	226.35	4.80E-02	2222u2d00	0.15
S <sub>6</sub>	213.65	1.13E-02	2ud222200	0.29
			22u222d00	0.40
S <sub>7</sub>	208.43	3.51E-03	222220200	0.20
S <sub>8</sub>	203.51	1.05E-02	ud2222200	0.46
S <sub>9</sub>	185.90	1.21E-02	2u22d2200	0.28

Table C.1: Absorption spectrum and configurations with weight  $> 0.1$  for the first nine excited states of rubiadin, calculated at the MS-CASPT2 level of theory with the (12,9) active space. Each configuration is represented by the orbitals within the active space - following the MO ordering in figure 8.4 - with their corresponding occupations: 2 represents a doubly occupied MO, u a singly occupied MO with spin up, d a spin down occupied MO, and 0 a unoccupied MO.

---

State	Wavelength [nm]	Osc. Strength	Configuration	Weight
S <sub>0</sub>	0.0		222222000	0.84
S <sub>1</sub>	437.16	1.29E-01	22222ud00	0.22
			22u222d00	0.52
S <sub>2</sub>	431.58	7.69E-02	222u22d00	0.65
			222u220d0	0.19
S <sub>3</sub>	375.50	2.52E-04	2u2222d00	0.56
			2u22220d0	0.16
S <sub>4</sub>	306.32	3.03E-01	22222ud00	0.40
			2222u2d00	0.11
			22u222d00	0.18
S <sub>5</sub>	254.19	4.19E-05	222u2d200	0.27
			22ud22200	0.13
			2ud222200	0.18
S <sub>6</sub>	241.21	5.47E-05	222u2d200	0.20
			222u220d0	0.34
S <sub>7</sub>	229.56	1.09E-04	2u2d22200	0.74
S <sub>8</sub>	227.98	2.39E-04	222u2d200	0.12
			2ud222200	0.42
S <sub>9</sub>	212.00	3.28E-02	2222u2d00	0.23
			22u2220d0	0.23
			220222200	0.14

---

Table C.2: Absorption spectrum and configurations with weight  $> 0.1$  for the first nine excited states of soranjidiol, calculated at the MS-CASPT2 level of theory with the (12,9) active space. Each configuration is represented by the orbitals within the active space - following the MO ordering in figure 8.5 - with their corresponding occupations: 2 represents a doubly occupied MO, u a singly occupied MO with spin up, d a spin down occupied MO, and 0 a unoccupied MO.

# Bibliography

- [1] Ma Xiaomei and Herbert Yu. Global burden of cancer. *The Yale Journal of Biology and Medicine*, 79(3-4):85–94, 2006.
- [2] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1):7–34, 2019.
- [3] Vassiliki Nikolettou, Maria Markaki, Konstantinos Palikaras, and Nektarios Tavernarakis. Crosstalk between apoptosis, necrosis and autophagy. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(12):3448–3459, 2013.
- [4] D. Kanduc, A. Mittelman, R. Serpico, E. Sinigaglia, A. A. Sinha, C. Natale, R. Santacroce, M. G. Di Corcia, A. Lucchese, L. Dini, P. Pani, S. Santacroce, S. Simone, R. Bucci, and E. Farber. Cell death: Apoptosis versus necrosis (review). *International Journal of Oncology*, 21(1):165–170, 2002.
- [5] Wesley M. Sharman, Cynthia M. Allen, and Johan E. van Lier. Photodynamic therapeutics: basic principles and clinical applications. *Drug Discovery Today*, 4(11):507–517, 1999.
- [6] Aisling E O'Connor, William M Gallagher, and Annette T Byrne. Porphyrin and nonporphyrin photosensitizers in oncology: preclinical and clinical advances in photodynamic therapy. *Photochem Photobiol*, 85(5):1053–1074, August 2009.
- [7] David Phillips. Toward targeted photodynamic therapy. *Pure and Applied Chemistry*, 83(4):733–748, 2011.

- 
- [8] Stephen G. Chaney, Sharon L. Campbell, Ekaterina Bassett, and Yibing Wu. Recognition and processing of cisplatin-and oxaliplatin-dna adducts. *Critical Reviews in Oncology/Hematology*, 53(1):3–11, 2005.
- [9] Markus Galanski, A. Michael Jakupec, and K. Bernhard Keppler. Update of the preclinical situation of anticancer platinum complexes: Novel design strategies and innovative analytical approaches. *Current Medicinal Chemistry*, 12(18):2075–2094, 2005.
- [10] Verena Pichler, Simone Göschl, Samuel M. Meier, Alexander Roller, Michael A. Jakupec, Markus Galanski, and Bernhard K. Keppler. Bulky n(n)-(di)alkylethane-1,2-diamineplatinum(ii) compounds as precursors for generating unsymmetrically substituted platinum(iv) complexes. *Inorganic Chemistry*, 52(14):8151–8162, 2013.
- [11] Lloyd Kelland. The resurgence of platinum-based cancer chemotherapy. *Nature Reviews Cancer*, 7(8):573–584, Aug 2007.
- [12] S. A. Aldossary. Review on pharmacology of cisplatin: Clinical use, toxicity and mechanism of resistance of cisplatin. *Biomedical and Pharmacology Journal*, 12:7–15, 2019.
- [13] Timothy C. Johnstone, Kogularamanan Suntharalingam, and Stephen J. Lippard. The next generation of platinum drugs: Targeted pt(ii) agents, nanoparticle delivery, and pt(iv) prodrugs. *Chemical Reviews*, 116(5):3436–3486, 2016.
- [14] Jorgelindo da Veiga Moreira, Laurent Schwartz, and Mario Jolicoeur. Targeting mitochondrial singlet oxygen dynamics offers new perspectives for effective metabolic therapies of cancer. *Frontiers in Oncology*, 10, 2020.
- [15] Saliha Harrach and Giuliano Ciarimboli. Role of transporters in the distribution of platinum-based drugs. *Frontiers in Pharmacology*, 6, 2015.
- [16] Cara A. Rabik and M. Eileen Dolan. Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treatment Reviews*, 33(1):9–23, 2007.
- [17] Zahid H. Siddik. Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene*, 22(47):7265–7279, Oct 2003.

- 
- [18] Hristo P. Varbanov, Michael A. Jakupec, Alexander Roller, Frank Jensen, Markus Galanski, and Bernhard K. Keppler. Theoretical investigations and density functional theory based quantitative structure–activity relationships model for novel cytotoxic platinum(iv) complexes. *Journal of Medicinal Chemistry*, 56(1):330–344, 2013.
- [19] Valentina Gandin, Cristina Marzano, Giorgio Pelosi, Mauro Ravera, Elisabetta Gabano, and Domenico Osella. trans,cis,cis-bis(benzoato)dichlorido(cyclohexane-1r,2r-diamine)platinum(iv): a prodrug candidate for the treatment of oxaliplatin-refractory colorectal cancer. *ChemMedChem*, 9(6):1299–1305, 2014.
- [20] Mauro Ravera, Elisabetta Gabano, Ilaria Zanellato, Ilaria Bonarrigo, Manuela Alessio, Fabio Arnesano, Angela Galliani, Giovanni Natile, and Domenico Osella. Cellular trafficking, accumulation and dna platination of a series of cisplatin-based dicarboxylato pt(iv) prodrugs. *Journal of Inorganic Biochemistry*, 150:1 – 8, 2015.
- [21] Dennis E.J.G.J. Dolmans, Dai Fukumura, and Rakesh K. Jain. Photodynamic therapy for cancer. *Nature Reviews Cancer*, 3(5):380–387, May 2003.
- [22] L. Wyld, M. W. R. Reed, and N. J. Brown. Differential cell death response to photodynamic therapy is dependent on dose and cell type. *British Journal of Cancer*, 84(10):1384–1386, May 2001.
- [23] Ana P. Castano, Tatiana N. Demidova, and Michael R. Hamblin. Mechanisms in photodynamic therapy: part one—photosensitizers, photochemistry and cellular localization. *Photodiagnosis and Photodynamic Therapy*, 1(4):279–293, 2004.
- [24] B. Muz, P. de la Puente, F. Azab, and A. K. Azab. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia*, 3:83–92, 2015.
- [25] Susan Monro, Katsuya L. Colón, Huimin Yin, John Roque, Prathyusha Konda, Shashi Gujar, Randolph P. Thummel, Lothar Lilge, Colin G. Cameron, and Sherri A. McFarland. Transition metal complexes and photodynamic therapy from a tumor-centered approach: Challenges, opportunities, and highlights from the development of tld1433. *Chemical Reviews*, 119(2):797–828, 2019.

- 
- [26] John A. Roque, Patrick C. Barrett, Houston D. Cole, L. M. Lifshits, Ge Shi, Susan Monro, David von Dohlen, Susy Kim, Nino Russo, Gagan Deep, Colin G. Cameron, Marta E. Alberto, and Sherri A. McFarland. Breaking the barrier: an osmium photosensitizer with unprecedented hypoxic phototoxicity for real world photodynamic therapy. *Chemical Science*, 11:9784–9806, 2020.
- [27] Yuhao Cheng, Hao Cheng, Chenxiao Jiang, Xuefeng Qiu, Kaikai Wang, Wei Huan, Ahu Yuan, Jinhui Wu, and Yiqiao Hu. Perfluorocarbon nanoparticles enhance reactive oxygen levels and tumour growth inhibition in photodynamic therapy. *Nature Communications*, 6(1):8785, Nov 2015.
- [28] Jonghoon Kim, Hye Rim Cho, Hyejin Jeon, Dokyoon Kim, Changyeong Song, Nohyun Lee, Seung Hong Choi, and Taeghwan Hyeon. Continuous o<sub>2</sub>-evolving mnfe<sub>2</sub>o<sub>4</sub> nanoparticle-anchored mesoporous silica nanoparticles for efficient photodynamic therapy in hypoxic cancer. *Journal of the American Chemical Society*, 139(32):10992–10995, 2017.
- [29] Van-Nghia Nguyen, Sujie Qi, Sangin Kim, Nahyun Kwon, Gyoungmi Kim, Yubin Yim, Sungnam Park, and Juyoung Yoon. An emerging molecular design approach to heavy-atom-free photosensitizers for enhanced photodynamic therapy under hypoxia. *Journal of the American Chemical Society*, 141(41):16243–16248, 2019.
- [30] Antonio Francés-Monerris, Iñaki Tuñón, and Antonio Monari. Hypoxia-selective dissociation mechanism of a nitroimidazole nucleoside in a dna environment. *The Journal of Physical Chemistry Letters*, 10(21):6750–6754, 2019.
- [31] J. M. Runnels, N. Chen, B. Ortel, D. Kato, and T. Hasan. Bpd-ma-mediated photosensitization in vitro and in vivo: cellular adhesion and  $\beta$ 1 integrin expression in ovarian cancer cells. *British Journal of Cancer*, 80(7):946–953, Jun 1999.
- [32] Janet Morgan and Allan R Oseroff. Mitochondria-based photodynamic anti-cancer therapy. *Advanced Drug Delivery Reviews*, 49(1):71–86, 2001.

- 
- [33] Keshav K. Singh, James Russell, Barbara Sigala, Yonggong Zhang, Jerry Williams, and Kylie F. Keshav. Mitochondrial dna determines the cellular response to cancer therapeutic agents. *Oncogene*, 18(48):6641–6646, Nov 1999.
- [34] Heiko Ihmels and Daniela Otto. *Intercalation of Organic Dye Molecules into Double-Stranded DNA – General Principles and Recent Developments*, pages 161–204. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [35] Yanyan Yan, Xiaodong Su, Yongju Liang, Jianye Zhang, Chengjun Shi, Yu Lu, Lianquan Gu, and Liwu Fu. Emodin azide methyl anthraquinone derivative triggers mitochondrial-dependent cell apoptosis involving in caspase-8-mediated Bid cleavage. *Molecular Cancer Therapeutics*, 7(6):1688–1697, 06 2008.
- [36] Jonathan B. Chaires, S. Satyanarayana, Dongchul Suh, Izabela Fokt, Teresa Przewloka, and Waldemar Priebe. Parsing the free energy of anthracycline antibiotic binding to dna. *Biochemistry*, 35(7):2047–2053, 1996.
- [37] Claude Monneret. Recent developments in the field of antitumour anthracyclines. *European Journal of Medicinal Chemistry*, 36(6):483–493, 2001.
- [38] Giorgio Minotti, Pierantonio Menna, Emanuela Salvatorelli, Gaetano Cairo, and Luca Gianni. Anthracyclines: Molecular advances and pharmacologic developments in antitumor activity and cardiotoxicity. *Pharmacological Reviews*, 56(2):185–229, 2004.
- [39] David T. Breslin, Changjun Yu, Danith Ly, and Gary B. Schuster. Structural modification changes the dna binding mode of cation-substituted anthraquinone photonucleases: Association by intercalation or minor groove binding determines the dna cleavage efficiency. *Biochemistry*, 36(34):10463–10473, 1997.
- [40] Antonio Monari and Elise Dumont. Understanding dna under oxidative stress and sensitization: the role of molecular modeling. *Frontiers in Chemistry*, 3, 2015.
- [41] Joanna Czapla-Masztafiak, Juan J. Nogueira, Ewelina Lipiec, Wojciech M. Kwiatek, Bayden R. Wood, Glen B. Deacon, Yves Kayser, Daniel L. A. Fernandes, Mariia V.

- 
- Pavliuk, Jakub Szlachetko, Leticia González, and Jacinto Sá. Direct determination of metal complexes' interaction with dna by atomic telemetry and multiscale molecular dynamics. *The Journal of Physical Chemistry Letters*, 8(4):805–811, 2017.
- [42] Shuang Ding, Alexander Kolbanovskiy, Alexander Durandin, Conor Crean, Vladimir Shafirovich, Suse Broyde, and Nicholas E. Geacintov. Absolute configurations of dna lesions determined by comparisons of experimental ecd and ord spectra with dft calculations. *Chirality*, 21(1E):E231–E241, 2009.
- [43] Mariangela Garofalo, Giovanni Grazioso, Andrea Cavalli, and Jacopo Sgrignani. How computational chemistry and drug delivery techniques can support the development of new anticancer drugs. *Molecules*, 25(7), 2020.
- [44] Stewart A. Adcock and J. Andrew McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615, 2006.
- [45] J. Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.
- [46] Dirk Bakowies and Walter Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *The Journal of Physical Chemistry*, 100(25):10580–10594, 1996.
- [47] Filippo Lipparini and Benedetta Mennucci. Hybrid qm/classical models: Methodological advances and new applications. *Chemical Physics Reviews*, 2(4):041303, 2021.
- [48] Sérgio Filipe Sousa, António J. M. Ribeiro, Rui P. P. Neves, Natércia F. Brás, Nuno M. F. S. A. Cerqueira, Pedro A. Fernandes, and Maria João Ramos. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(2):e1281, 2017.

- 
- [49] Semen Yesylevskyy, Timothée Rivel, and Christophe Ramseyer. Curvature increases permeability of the plasma membrane for ions, water and the anti-cancer drugs cisplatin and gemcitabine. *Scientific Reports*, 9(1):17214, Nov 2019.
- [50] Timothée Rivel, Christophe Ramseyer, and Semen Yesylevskyy. The asymmetry of plasma membranes and their cholesterol content influence the uptake of cisplatin. *Scientific reports*, 9(1):5627–5627, Apr 2019.
- [51] Arnab Mukherjee, Richard Lavery, Biman Bagchi, and James T. Hynes. On the molecular mechanism of drug intercalation into dna: A simulation study of the intercalation pathway, free energy, and dna structural changes. *Journal of the American Chemical Society*, 130(30):9747–9755, 2008.
- [52] Shirlene Jackson Beckford and Dabney W. Dixon. Molecular dynamics of anthraquinone dna intercalators with polyethylene glycol side chains. *Journal of Biomolecular Structure and Dynamics*, 29(5):1065–1080, 2012.
- [53] Matthieu Wilhelm, Arnab Mukherjee, Benjamin Bouvier, Krystyna Zakrzewska, James T. Hynes, and Richard Lavery. Multistep drug intercalation: Molecular dynamics and free energy studies of the binding of daunomycin to dna. *Journal of the American Chemical Society*, 134(20):8588–8596, 2012.
- [54] Alessio Terenzi, Hugo Gattuso, Angelo Spinello, Bernhard K. Keppler, Christophe Chipot, François Dehez, Giampaolo Barone, and Antonio Monari. Targeting g-quadruplexes with organic dyes: Chelerythrine–dna binding elucidated by combining molecular modeling and optical spectroscopy. *Antioxidants*, 8(10), 2019.
- [55] Xiya Lu, Dong Fang, Shingo Ito, Yuko Okamoto, Victor Ovchinnikov, and Qiang Cui. Qm/mm free energy simulations: recent progress and challenges. *Molecular Simulation*, 42(13):1056–1078, 2016.
- [56] Polydefkis Diamantis, Jérôme Florian Gonthier, Ivano Tavernelli, and Ursula Rothlisberger. Study of the redox properties of singlet and triplet tris(2,2'-bipyridine)ruthenium(ii) ([ru(bpy)<sub>3</sub>]<sup>2+</sup>) in aqueous solution by full quantum and

- 
- mixed quantum/classical molecular dynamics simulations. *The Journal of Physical Chemistry B*, 118(14):3950–3959, 2014.
- [57] Polydefkis Diamantis, Ivano Tavernelli, and Ursula Rothlisberger. Redox properties of native and damaged dna from mixed quantum mechanical/molecular mechanics molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 16(10):6690–6701, 2020.
- [58] Gustavo Cárdenas, Philipp Marquetand, Sebastian Mai, and Leticia González. A force field for a manganese-vanadium water oxidation catalyst: Redox potentials in solution as showcase. *Catalysts*, 11(4), 2021.
- [59] Jesús Lucia-Tamudo, Gustavo Cárdenas, Nuria Anguita-Ortiz, Sergio Díaz-Tendero, and Juan J. Nogueira. Computation of oxidation potentials of solvated nucleobases by static and dynamic multilayer approaches. *ChemRxiv*, 2022.
- [60] Mirjana Eckert-Maksić, Mario Vazdar, Matthias Ruckebauer, Mario Barbatti, Thomas Müller, and Hans Lischka. Matrix-controlled photofragmentation of formamide: dynamics simulation in argon by nonadiabatic qm/mm method. *Physical Chemistry Chemical Physics*, 12:12719–12726, 2010.
- [61] Felix Plasser, Adelia J. A. Aquino, William L. Hase, and Hans Lischka. Uv absorption spectrum of alternating dna duplexes. analysis of excitonic and charge transfer interactions. *The Journal of Physical Chemistry A*, 116(46):11151–11160, 2012.
- [62] Juan J. Nogueira, Felix Plasser, and Leticia González. Electronic delocalization, charge transfer and hypochromism in the uv absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis. *Chemical Science*, 8:5682–5691, 2017.
- [63] Philipp Marquetand, Juan J. Nogueira, Sebastian Mai, Felix Plasser, and Leticia González. Challenges in simulating light-induced processes in dna. *Molecules*, 22(1), 2017.

- 
- [64] Martina De Vetta, Maximilian F. S. J. Menger, Juan J. Nogueira, and Leticia González. Solvent effects on electronically excited states: Qm/continuum versus qm/explicit models. *The Journal of Physical Chemistry B*, 122(11):2975–2984, 2018.
- [65] Eliot Boulanger and Jeremy N Harvey. Qm/mm methods for free energies and photochemistry. *Current Opinion in Structural Biology*, 49:72–76, 2018.
- [66] Marco Marazzi, Hugo Gattuso, Maria Fumanal, Chantal Daniel, and Antonio Monari. Charge-transfer versus charge-separated triplet excited states of [rei(dmp)(co)3(his124)(trp122)]+ in water and in modified pseudomonas aeruginosa azurin protein. *Chemistry A European Journal*, 25(10):2519–2526, 2019.
- [67] Irina Dokukina, Artur Nenov, Marco Garavelli, Christel M. Marian, and Oliver Weingart. Qm/mm photodynamics of retinal in the channelrhodopsin chimera c1c2 with om3/mrci. *ChemPhotoChem*, 3(2):107–116, 2019.
- [68] Nuno Manuel de Almeida Barbosa, Madjid Zemmouche, Pauline Gosset, Cristina García-Iriepa, Vincent Ledentu, Isabelle Navizet, Pascal Didier, and Nicolas Ferré. pH-dependent absorption spectrum of oxyluciferin analogues in the presence of adenosine monophosphate. *ChemPhotoChem*, 3(12):1219–1230, 2019.
- [69] Cristina Garcia-Iriepa and Isabelle Navizet. Effect of protein conformation and amp protonation state on fireflies’ bioluminescent emission. *Molecules*, 24(8), 2019.
- [70] Davide Avagliano, Matteo Bonfanti, Marco Garavelli, and Leticia González. Qm/mm nonadiabatic dynamics: the sharc/cobramm approach. *Journal of Chemical Theory and Computation*, 17(8):4639–4647, 2021.
- [71] Juan J. Nogueira, Markus Oppel, and Leticia González. Enhancing intersystem crossing in phenotiazinium dyes by intercalation into dna. *Angewandte Chemie International Edition*, 54(14):4375–4378, 2015.
- [72] Juan J. Nogueira, Sven Roßbach, Christian Ochsenfeld, and Leticia González. Effect of dna environment on electronically excited states of methylene blue evaluated by a three-layered qm/qm/mm oniom scheme. *Journal of Chemical Theory and Computation*, 14(8):4298–4308, 2018.

- 
- [73] Juan J. Nogueira, Maximilian Meixner, Marius Bittermann, and Leticia González. Impact of lipid environment on photodamage activation of methylene blue. *ChemPhotoChem*, 1(5):178–182, 2017.
- [74] Miquel Huix-Rotllant, Elise Dumont, Nicolas Ferré, and Antonio Monari. Photo-physics of acetophenone interacting with dna: Why the road to photosensitization is open. *Photochemistry and Photobiology*, 91(2):323–330, 2015.
- [75] Hugo Gattuso, Vanessa Besancenot, Stéphanie Grandemange, Marco Marazzi, and Antonio Monari. From non-covalent binding to irreversible dna lesions: Nile blue and Nile red as photosensitizing agents. *Scientific Reports*, 6(1):28480, Jun 2016.
- [76] Élise Dumont and Antonio Monari. Interaction of palmatine with dna: An environmentally controlled phototherapy drug. *The Journal of Physical Chemistry B*, 119(2):410–419, 2015.
- [77] Jan C. Schöneboom, Shimrit Cohen, Hai Lin, Sason Shaik, and Walter Thiel. Quantum mechanical/molecular mechanical investigation of the mechanism of c-h hydroxylation of camphor by cytochrome p450cam: Theory supports a two-state rebound mechanism. *Journal of the American Chemical Society*, 126(12):4017–4034, 2004.
- [78] Ingrid Sol Cogno, Pamela Gilardi, Laura Comini, Susana C. Núñez-Montoya, Jose Luis Cabrera, and Viviana Alicia Rivarola. Natural photosensitizers in photodynamic therapy: In vitro activity against monolayers and spheroids of human colorectal adenocarcinoma sw480 cells. *Photodiagnosis and Photodynamic Therapy*, 31:101852, 2020.
- [79] Gustavo Cárdenas, Jesús Lucia-Tamudo, Henar Mateo-delaFuente, Vito F. Palmisano, Nuria Anguita-Ortiz, Lorena Ruano, Álvaro Pérez-Barcia, Sergio Díaz-Tendero, Marcos Mandado, and Juan Jose Nogueira. Mbiotools: A toolkit to setup qm/mm calculations. *ChemRxiv*, 2022.
- [80] Gustavo Cárdenas and Juan J. Nogueira. Mbiotools: A simple yet versatile toolkit to setup quantum mechanical/molecular mechanical (qm/mm) calculations from

- 
- molecular dynamics trajectories. <https://github.com/mobiochem/MoBioTools>, 2022. Accessed: 2022-05-13.
- [81] Björn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chemical Physics*, 48(2):157 – 173, 1980.
- [82] Ron Shepard. The multiconfiguration self-consistent field method. In Lawley K. P., editor, *Advances in Chemical Physics*, chapter 2, pages 63–200. John Wiley & Sons, Ltd, 1987.
- [83] Björn O. Roos. The complete active space self-consistent field method and its applications in electronic structure calculations. In Lawley K. P., editor, *Advances in Chemical Physics*, chapter 7, pages 399–445. John Wiley & Sons, Ltd, 1987.
- [84] Gustavo Cárdenas and Juan J. Nogueira. An algorithm to correct for the casscf active space in multiscale qm/mm calculations based on geometry ensembles. *International Journal of Quantum Chemistry*, 121(6):e26533, 2021.
- [85] Lorena Ruano, Gustavo Cárdenas, and Juan J. Nogueira. The permeation mechanism of cisplatin through a dioleoylphosphocholine bilayer. *ChemPhysChem*, 22:1251–1261, 2021.
- [86] G. Cárdenas, A. Pérez-Barcia, M. Mandado, and J. J. Nogueira. Characterization of cisplatin/membrane interactions by QM/MM energy decomposition analysis. *Physical Chemistry Chemical Physics*, 23:20533, 2021.
- [87] Gustavo Cárdenas and Juan J. Nogueira. Stacking effects on anthraquinone/dna charge-transfer electronically excited states. *Molecules*, 25(24), 2020.
- [88] Attila Szabó and Neil S. Ostlund. *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory*. Mineola (N.Y.) : Dover publications, 1996.
- [89] D.J. Tannor. *Introduction to Quantum Mechanics: A Time-Dependent Perspective*. University Science Books, 2007.

- 
- [90] Dmitrii V. Shalashilin. Quantum mechanics with the basis set guided by ehrenfest trajectories: Theory and application to spin-boson model. *The Journal of Chemical Physics*, 130(24):244101, 2009.
- [91] Lea M. Ibele and Basile F. E. Curchod. Dynamics near a conical intersection—a diabolical compromise for the approximations of ab initio multiple spawning. *The Journal of Chemical Physics*, 155(17):174119, 2021.
- [92] Yorick Lassmann and Basile F. E. Curchod. Aimswiss—ab initio multiple spawning with informed stochastic selections. *The Journal of Chemical Physics*, 154(21):211106, 2021.
- [93] A Picón, L Plaja, and J Biegert. Attosecond x-ray transient absorption in condensed-matter: a core-state-resolved bloch model. *New Journal of Physics*, 21(4):043029, apr 2019.
- [94] Giovanni Cistaro, Luis Plaja, Fernando Martín, and Antonio Picón. Attosecond x-ray transient absorption spectroscopy in graphene. *Physical Review Research*, 3:013144, Feb 2021.
- [95] Piero Decleva, Mauro Stener, and Daniele Toffoli. Continuum electronic states: The tiresia code. *Molecules*, 27(6), 2022.
- [96] C. C. J. Roothaan. New developments in molecular orbital theory. *Reviews of Modern Physics*, 23:69–89, Apr 1951.
- [97] Katharina Boguslawski, Pawel Tecmer, Örs Legeza, and Markus Reiher. Entanglement measures for single- and multireference correlation effects. *The Journal of Physical Chemistry Letters*, 3(21):3129–3135, 2012.
- [98] Christopher J. Stein and Markus Reiher. Automated identification of relevant frontier orbitals for chemical compounds and processes. *CHIMIA International Journal for Chemistry*, 71(4):170–176, 2017.
- [99] Kerstin Andersson and Björn O. Roos. Multiconfigurational second-order perturbation theory. In Yarkony R. D., editor, *Modern Electronic Structure Theory*, pages 55–109. 1995.

- 
- [100] Charles W. Bauschlicher Jr., Stephen R. Langhoff, and Peter R. Taylor. Accurate quantum chemical calculations. In I. Prigogine and Stuart A. Rice, editors, *Advances in Chemical Physics*, chapter 3, pages 103–161. John Wiley & Sons, Ltd, 1990.
- [101] James Finley, Per Åke Malmqvist, Björn O. Roos, and Luis Serrano-Andrés. The multi-state caspt2 method. *Chemical Physics Letters*, 288(2):299 – 306, 1998.
- [102] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864–B871, Nov 1964.
- [103] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133–A1138, Nov 1965.
- [104] Axel D Becke. Becke’s three parameter hybrid method using the LYP correlation functional. *The Journal of Chemical Physics*, 98(492):5648—5652, 1993.
- [105] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.
- [106] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, 1988.
- [107] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1):51–57, 2004.
- [108] Y. Zhao and D. G. Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(120):215–241, 2008.
- [109] Mark E. Casida. *Time-Dependent Density Functional Response Theory for Molecules*, pages 155–192.

- 
- [110] Ivo Cacelli. Corso di laurea magistrale in chimica-chimica teorica. <https://people.unipi.it/static/Cacelli/didattica/chimicateorica.htm>, 2009. Accessed: 2022-05-25.
- [111] Miquel Huix-Rotllant, Nicolas Ferré, and Mario Barbatti. *Time-Dependent Density Functional Theory*, chapter 2, pages 13–46. John Wiley & Sons, Ltd, 2020.
- [112] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52:997–1000, Mar 1984.
- [113] Keiji Morokuma. Molecular orbital studies of hydrogen bonds. iii. c=O...h-o hydrogen bond in h<sub>2</sub>co...h<sub>2</sub>o and h<sub>2</sub>co...2h<sub>2</sub>o. *The Journal of Chemical Physics*, 55(3):1236–1244, 1971.
- [114] M. J. S. Phipps, T. Fox, C. S. Tautermann, and C. K. Skylaris. Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns. *Chemical Society Reviews*, 44:3177–3211, 2015.
- [115] Lili Zhao, Moritz von Hopffgarten, Diego M. Andrada, and Gernot Frenking. Energy decomposition analysis. *WIREs Comput. Mol. Sci.*, 8:e1345, 2018.
- [116] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chemical Reviews*, 94:1887–1930, 1994.
- [117] M. Mandado and J. M. Hermida-Ramón. Electron Density Based Partitioning Scheme of Interaction Energies. *Journal of Chemical Theory and Computation*, 7(3):633–641, 2011.
- [118] N. Ramos-Berdullas, I. Pérez-Juste, C. Van Alsenoy and M. Mandado. Theoretical study of the adsorption of aromatic units on carbon allotropes including explicit (empirical) DFT dispersion corrections and implicitly dispersion-corrected functionals: the pyridine case. *Physical Chemistry Chemical Physics*, 17(1):575–587, 2015.
- [119] S.F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970.

- 
- [120] Sílvia Simon, Miquel Duran, and J. J. Dannenberg. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics*, 105(24):11024–11031, 1996.
- [121] I. G. Kaplan. *Intermolecular Interactions: Physical Picture, Computational Methods, Model Potentials*. John Wiley & Sons, Chichester, England, 2006.
- [122] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, USA, 1989.
- [123] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*. Academic Press, Inc., USA, 2nd edition, 2001.
- [124] Charles L. Brooks, David A. Case, Steve Plimpton, Benoît Roux, David van der Spoel, and Emad Tajkhorshid. Classical molecular dynamics. *The Journal of Chemical Physics*, 154(10):100401, 2021.
- [125] Trygve Helgaker, Einar Uggerud, and Hans Jørgen Aa. Jensen. Integration of the classical equations of motion on ab initio molecular potential energy surfaces using gradients and Hessians: application to translational energy release upon fragmentation. *Chemical Physics Letters*, 173(2):145–150, 1990.
- [126] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976.
- [127] R.P. Feynman, R.B. Leighton, M. Sands, and EM Hafner. *The Feynman Lectures on Physics; Vol. I*, volume 33. AAPT, 1965.
- [128] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley, 1980.
- [129] M. Amini, J.W. Eastwood, and R.W. Hockney. Time integration in particle models. *Computer Physics Communications*, 44(1):83–93, 1987.
- [130] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the

- 
- formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.
- [131] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.
- [132] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, and P.A. Kollman. Amber 2018, 2018.
- [133] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.
- [134] Åge A. Skjevik, Benjamin D. Madej, Ross C. Walker, and Knut Teigen. Lipid11: A modular framework for lipid simulations using amber. *The Journal of Physical Chemistry B*, 116(36):11124–11136, 2012.
- [135] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. Lipid14: The amber lipid force field. *Journal of Chemical Theory and Computation*, 10(2):865–879, 2014.
- [136] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L De Groot, Helmut Grubmüller, and Alexander D MacKerell. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, 2017.

- 
- [137] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [138] Marie Zgarbová, Jiří Šponer, Michal Otyepka, Thomas E. Cheatham, Rodrigo Galindo-Murillo, and Petr Jurečka. Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z- and b-dna. *Journal of Chemical Theory and Computation*, 11(12):5723–5736, 2015.
- [139] Saulo A. Vázquez, Xose L. Otero, and Emilio Martínez-Nunez. A trajectory-based method to explore reaction mechanisms. *Molecules*, 23(12), 2018.
- [140] Emilio Martínez-Núñez, George L. Barnes, David R. Glowacki, Sabine Kopec, Daniel Peláez, Aurelio Rodríguez, Roberto Rodríguez-Fernández, Robin J. Shannon, James J. P. Stewart, Pablo G. Tahoces, and Saulo A. Vazquez. Automekin2021: An open-source program for automated reaction discovery. *Journal of Computational Chemistry*, 42(28):2036–2048, 2021.
- [141] Jochen Blumberger, Ivano Tavernelli, Michael L. Klein, and Michiel Sprik. Diabatic free energy curves and coordination fluctuations for the aqueous  $ag^+/ag^{2+}$  redox couple: A biased born-oppenheimer molecular dynamics investigation. *The Journal of Chemical Physics*, 124(6):064507, 2006.
- [142] Yi Isaac Yang, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics*, 151(7):070902, 2019.
- [143] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.
- [144] Arthur F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.*, 78:3908–3911, May 1997.
- [145] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

- 
- [146] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annual Review of Physical Chemistry*, 67(1):159–184, 2016. PMID: 26980304.
- [147] Lucjan Piela, Jaroslaw Kostrowicki, and Harold A. Scheraga. On the multiple-minima problem in the conformational analysis of molecules: deformation of the potential energy hypersurface by the diffusion equation method. *The Journal of Physical Chemistry*, 93(8):3339–3346, 1989.
- [148] Omar Valsson and Michele Parrinello. Variational approach to enhanced sampling and free energy calculations. *Phys. Rev. Lett.*, 113:090601, Aug 2014.
- [149] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1):141–151, 1999.
- [150] Mads R. So/rensen and Arthur F. Voter. Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21):9599–9606, 2000.
- [151] Donald Hamelberg, John Mongan, and J. Andrew McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*, 120(24):11919–11929, 2004.
- [152] Yinglong Miao, William Sinko, Levi Pierce, Denis Bucher, Ross C. Walker, and J. Andrew McCammon. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*, 10(7):2677–2689, 2014.
- [153] Yi Qin Gao. An integrate-over-temperature approach for enhanced sampling. *The Journal of Chemical Physics*, 128(6):064105, 2008.
- [154] Lijiang Yang, Cheng-Wen Liu, Qiang Shao, Jun Zhang, and Yi Qin Gao. From thermodynamics to kinetics: Enhanced sampling of rare events. *Accounts of Chemical Research*, 48(4):947–955, 2015. PMID: 25781363.
- [155] Alan M. Ferrenberg and Robert H. Swendsen. Optimized monte carlo data analysis. *Physical Review Letters*, 63:1195–1198, Sep 1989.

- 
- [156] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1):275–282, 1995.
- [157] Shankar Kumar, John M. Rosenberg, Djamel Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.
- [158] Alan Grossfield. Wham: the weighted histogram analysis method. [http://membrane.urmc.rochester.edu/wordpress/?page\\_id=126](http://membrane.urmc.rochester.edu/wordpress/?page_id=126). version 2.09.
- [159] Hans Martin Senn and Walter Thiel. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.
- [160] Carles Curutchet, Aurora Muñoz-Losa, Susanna Monti, Jacob Kongsted, Gregory D. Scholes, and Benedetta Mennucci. Electronic energy transfer in condensed phase studied by a polarizable qm/mm model. *Journal of Chemical Theory and Computation*, 5(7):1838–1848, 2009.
- [161] Eliot Boulanger and Walter Thiel. Solvent boundary potentials for hybrid qm/mm computations using classical drude oscillators: A fully polarizable model. *Journal of Chemical Theory and Computation*, 8(11):4527–4538, 2012.
- [162] Alexey Savelyev and Alexander D. MacKerell Jr. All-atom polarizable force field for dna based on the classical drude oscillator model. *Journal of Computational Chemistry*, 35(16):1219–1239, 2014.
- [163] Arieh Warshel, Mitsunori Kato, and Andrei V. Pisliakov. Polarizable force fields: History, test cases, and prospects. *Journal of Chemical Theory and Computation*, 3(6):2034–2045, 2007.
- [164] Pier Paolo Poier and Frank Jensen. Describing molecular polarizability by a bond capacity model. *Journal of Chemical Theory and Computation*, 15(5):3093–3107, 2019.
- [165] Jean-Louis Rivail and Daniel Rinaldi. A quantum chemical approach to dielectric solvent effects in molecular liquids. *Chemical Physics*, 18(1):233–242, 1976.

- 
- [166] Marc W. van der Kamp and Adrian J. Mulholland. Combined quantum mechanics/molecular mechanics (qm/mm) methods in computational enzymology. *Biochemistry*, 52(16):2708–2728, 2013.
- [167] Arieh Warshel. Multiscale modeling of biological functions: From enzymes to molecular machines (nobel lecture). *Angewandte Chemie International Edition*, 53(38):10020–10031, 2014.
- [168] Mattia Bondanza, Michele Nottoli, Lorenzo Cupellini, Filippo Lipparini, and Benedetta Mennucci. Polarizable embedding qm/mm: the future gold standard for complex (bio)systems? *Physical Chemistry Chemical Physics*, 22:14433–14448, 2020.
- [169] Christopher J. Cramer and Donald G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chemical Reviews*, 99(8):2161–2200, 1999.
- [170] Modesto Orozco and F. Javier Luque. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chemical Reviews*, 100(11):4187–4226, 2000.
- [171] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3094, 2005.
- [172] Andreas Klamt. The cosmo and cosmo-rs solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):699–709, 2011.
- [173] Maite Roca, Juan Andrés, Vicent Moliner, Iñaki Tuñón, and Juan Bertrán. On the nature of the transition state in catechol o-methyltransferase. a complementary study based on molecular dynamics and potential energy surface explorations. *Journal of the American Chemical Society*, 127(30):10648–10655, 2005.
- [174] Hans Martin Senn and Walter Thiel. Qm/mm studies of enzymes. *Current Opinion in Chemical Biology*, 11(2):182–187, 2007.
- [175] Natércia F. Brás, Sara A. Moura-Tamames, Pedro A. Fernandes, and Maria J. Ramos. Mechanistic studies on the formation of glycosidase-substrate and glycosidase-inhibitor covalent intermediates. *Journal of Computational Chemistry*, 29(15):2565–2574, 2008.

- 
- [176] Anna L. Bowman, Ian M. Grant, and Adrian J. Mulholland. Qm/mm simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate. *Chemical Communications*, pages 4425–4427, 2008.
- [177] Andrew J. Adamczyk, Jie Cao, Shina C. L. Kamerlin, and Arieh Warshel. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proceedings of the National Academy of Sciences*, 108(34):14115–14120, 2011.
- [178] Marc W. van der Kamp, Jolanta Żurek, Frederick R. Manby, Jeremy N. Harvey, and Adrian J. Mulholland. Testing high-level qm/mm methods for modeling enzyme reactions: Acetyl-coa deprotonation in citrate synthase. *The Journal of Physical Chemistry B*, 114(34):11303–11314, 2010.
- [179] Kara E. Ranaghan, William G. Morris, Laura Masgrau, Kittusamy Senthilkumar, Linus O. Johannissen, Nigel S. Scrutton, Jeremy N. Harvey, Frederick R. Manby, and Adrian J. Mulholland. Ab initio qm/mm modeling of the rate-limiting proton transfer step in the deamination of tryptamine by aromatic amine dehydrogenase. *The Journal of Physical Chemistry B*, 121(42):9785–9798, 2017.
- [180] Kenneth W. Borrelli, Andreas Vitalis, Raul Alcantara, and Victor Guallar. Pele: Protein energy landscape exploration. a novel monte carlo based technique. *Journal of Chemical Theory and Computation*, 1(6):1304–1311, 2005.
- [181] Jens Peder Dahl and Michael Springborg. The morse oscillator in position space, momentum space, and phase space. *The Journal of Chemical Physics*, 88(7):4535–4547, 1988.
- [182] Martin J. Field. The pdynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *Journal of Chemical Theory and Computation*, 4(7):1151–1161, 2008.
- [183] Jan Řezáč. Cuby: An integrative framework for computational chemistry. *Journal of Computational Chemistry*, 37(13):1230–1237, 2016.

- 
- [184] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, jun 2017.
- [185] You Lu, Matthew R. Farrow, Pierre Fayon, Andrew J. Logsdail, Alexey A. Sokol, C. Richard A. Catlow, Paul Sherwood, and Thomas W. Keal. Open-source, python-based redevelopment of the chemshell multiscale qm/mm environment. *Journal of Chemical Theory and Computation*, 15(2):1317–1328, 2019.
- [186] Sergio Martí. Qmcube (qm3): An all-purpose suite for multiscale qm/mm calculations. *Journal of Computational Chemistry*, 42(6):447–457, 2021.
- [187] Oliver Weingart, Artur Nenov, Piero Altoè, Ivan Rivalta, Javier Segarra-Martí, Irina Dokukina, and Marco Garavelli. Cobramm 2.0 — a software interface for tailoring molecular electronic structure calculations and running nanoscale (qm/mm) simulations. *Journal of Molecular Modeling*, 24(9):271, Sep 2018.
- [188] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith,

- 
- R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian 16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.
- [189] Martin Richter, Philipp Marquetand, Jesús González-Vázquez, Ignacio Sola, and Leticia González. Sharc: ab initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings. *Journal of Chemical Theory and Computation*, 7(5):1253–1258, 2011.
- [190] Sebastian Mai, Philipp Marquetand, and Leticia González. Nonadiabatic dynamics: The sharc approach. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(6):e1370, 2018.
- [191] Mario Barbatti, Matthias Ruckebauer, Felix Plasser, Jiri Pittner, Giovanni Granucci, Maurizio Persico, and Hans Lischka. Newton-x: a surface-hopping program for nonadiabatic molecular dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(1):26–33, 2014.
- [192] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O’Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber 2020, 2020.
- [193] Frank Neese. Software update: the orca program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1327, 2018.
- [194] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong. NWChem: A

- 
- comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.*, 181(9):1477–1489, September 2010.
- [195] Ignacio Fdez. Galván, Morgane Vacher, Ali Alavi, Celestino Angeli, Francesco Aquilante, Jochen Autschbach, Jie J. Bao, Sergey I. Bokarev, Nikolay A. Bogdanov, Rebecca K. Carlson, Liviu F. Chibotaru, Joel Creutzberg, Nike Dattani, Mickaël G. Delcey, Sijia S. Dong, Andreas Dreuw, Leon Freitag, Luis Manuel Frutos, Laura Gagliardi, Frédéric Gendron, Angelo Giussani, Leticia González, Gilbert Grell, Meiyuan Guo, Chad E. Hoyer, Marcus Johansson, Sebastian Keller, Stefan Knecht, Goran Kovačević, Erik Källman, Giovanni Li Manni, Marcus Lundberg, Yingjin Ma, Sebastian Mai, João Pedro Malhado, Per Åke Malmqvist, Philipp Marquetand, Stefanie A. Mewes, Jesper Norell, Massimo Olivucci, Markus Oppel, Quan Manh Phung, Kristine Pierloot, Felix Plasser, Markus Reiher, Andrew M. Sand, Igor Schapiro, Prachi Sharma, Christopher J. Stein, Lasse Kragh Sørensen, Donald G. Truhlar, Mihkel Ugandi, Liviu Ungur, Alessio Valentini, Steven Vancoillie, Valera Veryazov, Oskar Weser, Tomasz A. Wesolowski, Per-Olof Widmark, Sebastian Wouters, Alexander Zech, J. Patrick Zobel, and Roland Lindh. Openmolcas: From source code to insight. *Journal of Chemical Theory and Computation*, 15(11):5925–5964, 2019.
- [196] Francesco Aquilante, Jochen Autschbach, Alberto Baiardi, Stefano Battaglia, Veniamin A. Borin, Liviu F. Chibotaru, Irene Conti, Luca De Vico, Mickaël Delcey, Ignacio Fdez. Galván, Nicolas Ferré, Leon Freitag, Marco Garavelli, Xuejun Gong, Stefan Knecht, Ernst D. Larsson, Roland Lindh, Marcus Lundberg, Per Åke Malmqvist, Artur Nenov, Jesper Norell, Michael Odellius, Massimo Olivucci, Thomas B. Pedersen, Laura Pedraza-González, Quan M. Phung, Kristine Pierloot, Markus Reiher, Igor Schapiro, Javier Segarra-Martí, Francesco Segatta, Luis Seijo, Saumik Sen, Dumitru-Claudiu Sergentu, Christopher J. Stein, Liviu Ungur, Morgane Vacher, Alessio Valentini, and Valera Veryazov. Modern quantum chemistry with [open]molcas. *The Journal of Chemical Physics*, 152(21):214117, 2020.
- [197] Daniel R. Roe and Thomas E. Cheatham. Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, 2013.

- 
- [198] G. Schaftenaar and J. H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, 14(2):123–134, 2000.
- [199] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018.
- [200] Qiming Sun. Libcint: An efficient general integral library for gaussian basis functions. *Journal of Computational Chemistry*, 36:1664–1671, 2015.
- [201] S. Obara and A. Saika. Efficient recursive computation of molecular integrals over cartesian gaussian functions. *The Journal of Chemical Physics*, 84(7):3963–3974, 1986.
- [202] E. F. Valeev. Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions. <http://libint.valeyev.net/>, 2021. version 2.7.1.
- [203] Donald G. Truhlar, Christopher J. Cramer, Anne Lewis, and John A. Bumpus. Molecular modeling of environmentally important processes: Reduction potentials. *Journal of Chemical Education*, 81(4):596, 2004.
- [204] Donald G Truhlar, Christopher J Cramer, Anne Lewis, and John A Bumpus. Journal of chemical education 2004, 81, 596–604. *Journal of Chemical Education*, 84(6):934, 2007.
- [205] Abdirisak A. Isse and Armando Gennaro. Absolute potential of the standard hydrogen electrode and the problem of interconversion of potentials in different solvents. *The Journal of Physical Chemistry B*, 114(23):7894–7899, 2010.
- [206] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *The Journal of Physical Chemistry B*, 110(32):16066–16081, 2006.

- 
- [207] R. A. Marcus. On the theory of oxidation-reduction reactions involving electron transfer. i. *The Journal of Chemical Physics*, 24(5):966–978, 1956.
- [208] Rudolph A. Marcus. On the theory of oxidation-reduction reactions involving electron transfer. iii. applications to data on the rates of organic redox reactions. *The Journal of Chemical Physics*, 26:872–877, 1957.
- [209] R. A. Marcus. On the theory of oxidation-reduction reactions involving electron transfer. v. comparison and properties of electrochemical and chemical rate constants. *The Journal of Physical Chemistry*, 67(4):853–857, 1963.
- [210] Rudolph A Marcus. On the theory of electron-transfer reactions. vi. unified treatment for homogeneous and electrode reactions. *The Journal of Chemical Physics*, 43(2):679–701, 1965.
- [211] Arieh Warshel. Dynamics of reactions in polar solvents. semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *The Journal of Physical Chemistry*, 86(12):2218–2224, 1982.
- [212] Gregory King and Arieh Warshel. Investigation of the free energy functions for electron transfer reactions. *The Journal of Chemical Physics*, 93(12):8682–8692, 1990.
- [213] John E. Bartmess. Thermodynamics of the electron and the proton. *The Journal of Physical Chemistry*, 98(25):6420–6424, 1994.
- [214] John E. Bartmess. Thermodynamics of the electron and the proton [erratum to document cited in ca121:19495]. *The Journal of Physical Chemistry*, 99(17):6755–6755, 1995.
- [215] Aleksandr V. Marenich, Junming Ho, Michelle L. Coote, Christopher J. Cramer, and Donald G. Truhlar. Computational electrochemistry: Prediction of liquid-phase reduction potentials. *Physical Chemistry Chemical Physics*, 16:15068–15106, 2014.
- [216] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77:3865–3868, Oct 1996.

- 
- [217] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple [physical review letters 77, 3865 (1996)]. *Physical Review Letters*, 78:1396–1396, Feb 1997.
- [218] Takao Tsuneda, Toshihisa Suzumura, and Kimihiko Hirao. A new one-parameter progressive colle–salvetti-type correlation functional. *The Journal of Chemical Physics*, 110(22):10664–10678, 1999.
- [219] A. Klamt and G. Schüürmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, pages 799–805, 1993.
- [220] Jorge M. Seminario. Calculation of intramolecular force fields from second-derivative tensors. *International Journal of Quantum Chemistry*, 60(7):1271–1277, 1996.
- [221] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009.
- [222] Juan C. Meza. Steepest descent. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):719–722, 2010.
- [223] A. Galántai. The theory of newton’s method. *Journal of Computational and Applied Mathematics*, 124(1):25–44, 2000. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.
- [224] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B*, 17(3):1302, 1978.
- [225] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNicola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81:3684, 1984.
- [226] Tom Darden, Darrin York, and Lee Pedersen. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics*, 98:10089–10092, 1993.

- 
- [227] S. Miyamoto and P.A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992. cited By 4794.
- [228] Claus A. M. Seidel, Andreas Schulz, and Markus H. M. Sauer. Nucleobase-specific quenching of fluorescent dyes. 1. nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *The Journal of Physical Chemistry*, 100(13):5541–5553, 1996.
- [229] Carlos E. Crespo-Hernández, David M. Close, Leonid Gorb, and Jerzy Leszczynski. Determination of redox potentials for the watson-crick base pairs, dna nucleosides, and relevant nucleoside analogues. *The Journal of Physical Chemistry B*, 111(19):5386–5395, 2007.
- [230] M. Ropo, M. Schneider, C. Baldauf, and V. Blum. First-principles data set of 45,892 isolated and cation-coordinated conformers of 20 proteinogenic amino acids. *Scientific Data*, 3, 2016.
- [231] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.
- [232] F.J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.
- [233] Kerstin Andersson, Per Aake. Malmqvist, Bjoern O. Roos, Andrzej J. Sadlej, and Krzysztof. Wolinski. Second-order perturbation theory with a casscf reference function. *The Journal of Physical Chemistry*, 94(14):5483–5488, 1990.
- [234] Kerstin Andersson, Per-Åke Malmqvist, and Björn O. Roos. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics*, 96(2):1218–1226, 1992.
- [235] Vito F. Palmisano, Carlos Gómez-Rodellar, Hannah Pollak, Gustavo Cárdenas, Ben Corry, Shirin Faraji, and Juan J. Nogueira. Binding of azobenzene and p-

- 
- diaminoazobenzene to the human voltage-gated sodium channel nav1.4. *Physical Chemistry Chemical Physics*, 23(5):3552–3564, 2021.
- [236] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.
- [237] Xiaojing Pan, Zhangqiang Li, Qiang Zhou, Huaizong Shen, Kun Wu, Xiaoshuang Huang, Jiaofeng Chen, Juanrong Zhang, Xuechen Zhu, and Jianlin Lei. Structure of the human voltage-gated sodium channel nav1. 4 in complex with  $\beta 1$ . *Science*, 362(6412):eaau2486, 2018.
- [238] Mikhail A Lomize, Irina D Pogozheva, Hyeon Joo, Henry I Mosberg, and Andrei L Lomize. Opm database and ppm web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*, 40(D1):D370–D376, 2012.
- [239] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. Charmm-gui: a web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [240] Anna I Krylov and Peter MW Gill. Q-chem: an engine for innovation. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(3):317–326, 2013.
- [241] Martin McCullagh, Ignacio Franco, Mark A Ratner, and George C Schatz. Dna-based optomechanical molecular motor. *Journal of the American Chemical Society*, 133(10):3452–3459, 2011.
- [242] Kenno Vanommeslaeghe, Elizabeth Hatcher, Chayan Acharya, Sibsankar Kundu, Shijun Zhong, Jihyun Shim, Eva Darian, Olgun Guvench, P Lopes, Igor Vorobyov, et al. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2010.
- [243] Francesco Aquilante, Roland Lindh, and Thomas Bondo Pedersen. Unbiased auxiliary basis sets for accurate two-electron integral approximations. *The Journal of Chemical Physics*, 127(11):114107, 2007.

- 
- [244] Niclas Forsberg and Per-Åke Malmqvist. Multiconfiguration perturbation theory with imaginary level shift. *Chemical Physics Letters*, 274(1):196–204, 1997.
- [245] J. Patrick Zobel, Juan J. Nogueira, and Leticia González. The ipea dilemma in caspt2. *Chemical Science*, 8:1482–1499, 2017.
- [246] LLC Schrödinger and Warren DeLano. Pymol.
- [247] Jesse A. Sundlov, Danielle M. Fontaine, Tara L. Southworth, Bruce R. Branchini, and Andrew M. Gulick. Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. *Biochemistry*, 51(33):6493–6495, aug 2012.
- [248] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Miguez, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 2020.
- [249] Michelle M. Francl, William J. Pietro, Warren J. Hehre, J. Stephen Binkley, Mark S. Gordon, Douglas J. DeFrees, and John A. Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics*, 77(7):3654–3665, 1982.
- [250] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics*, 72(1):650–654, 1980.
- [251] A. D. McLean and G. S. Chandler. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms,  $Z=11-18$ . *The Journal of Chemical Physics*, 72(10):5639–5648, 1980.
- [252] David M. Mofford, Gadarla Randheer Reddy, and Stephen C. Miller. Aminoluciferins extend firefly luciferase bioluminescence into the near-infrared and can be preferred substrates over D-luciferin. *Journal of the American Chemical Society*, 136(38):13277–13282, 2014.

- 
- [253] Per Åke Malmqvist, Kristine Pierloot, Abdul Rehaman Moughal Shahi, Christopher J. Cramer, and Laura Gagliardi. The restricted active space followed by second-order perturbation theory method: Theory and application to the study of  $\text{CuO}_2$  and  $\text{Cu}_2\text{O}_2$  systems. *The Journal of Chemical Physics*, 128(20):204109, 2008.
- [254] Yang Guo, Kantharuban Sivalingham, Edward F. Valeev, and Frank Neese. Sparsemaps—a systematic infrastructure for reduced-scaling electronic structure methods. iii. linear-scaling multireference domain-based pair natural orbital  $n$ -electron valence perturbation theory. *The Journal of Chemical Physics*, 144(9):094111, 2016.
- [255] Raúl Montero, Asier Longarte, Álvaro Peralta Conde, Carolina Redondo, Fernando Castaño, Israel González–Ramírez, Angelo Giussani, Luis Serrano-Andrés, and Manuela Merchán. Photophysics of 1-aminonaphthalene: A theoretical and time-resolved experimental study. *The Journal of Physical Chemistry A*, 113(48):13509–13518, 2009.
- [256] Ida Josefsson, Kristjan Kunnus, Simon Schreck, Alexander Föhlisch, Frank de Groot, Philippe Wernet, and Michael Odellius. Ab initio calculations of x-ray spectra: Atomic multiplet and molecular orbital effects in a multiconfigurational scf approach to the l-edge spectra of transition metal complexes. *The Journal of Physical Chemistry Letters*, 3(23):3565–3570, 2012.
- [257] Sergey I. Bokarev, Marcus Dantz, Edlira Suljoti, Oliver Kühn, and Emad F. Aziz. State-dependent electron delocalization dynamics at the solute-solvent interface: Soft-x-ray absorption spectroscopy and ab initio calculations. *Physical Review Letters*, 111:083002, Aug 2013.
- [258] Igor Schapiro, Kantharuban Sivalingham, and Frank Neese. Assessment of  $n$ -electron valence state perturbation theory for vertical excitation energies. *Journal of Chemical Theory and Computation*, 9(8):3567–3580, 2013.
- [259] Frédéric Gendron, Valerie E. Fleischauer, Thomas J. Duignan, Brian L. Scott, Matthias W. Löble, Samantha K. Cary, Stosh A. Kozimor, Hélène Bolvin, Michael L. Neidig, and Jochen Autschbach. Magnetic circular dichroism of  $\text{UCl}_6^-$  in the ligand-

- 
- to-metal charge-transfer spectral region. *Physical Chemistry Chemical Physics*, 19:17300–17313, 2017.
- [260] Neus Aguilera-Porta, Giovanni Granucci, Jordi Munoz-Muriedas, and Inés Corral. Unveiling the photophysics of thiourea from caspt2/casscf potential energy surfaces and singlet/triplet excited state molecular dynamics simulations. *Computational and Theoretical Chemistry*, 1151:36 – 42, 2019.
- [261] Lara Martínez-Fernández, Serra Arslançan, Dmytro Ivashchenko, Carlos E. Crespo-Hernández, and Inés Corral. Tracking the origin of photostability in purine nucleobases: the photophysics of 2-oxopurine. *Physical Chemistry Chemical Physics*, 21:13467–13473, 2019.
- [262] Flavia Aleotti, Lorenzo Soprani, Artur Nenov, Roberto Berardi, Alberto Arcioni, Claudio Zannoni, and Marco Garavelli. Multidimensional potential energy surfaces resolved at the raspt2 level for accurate photoinduced isomerization dynamics of azobenzene. *Journal of Chemical Theory and Computation*, 15(12):6813–6823, 2019.
- [263] Jaroslaw J. Szymczak, Mario Barbatti, Jason T. Soo Hoo, Jaclyn A. Adkins, Theresa L. Windus, Dana Nachtigallová, and Hans Lischka. Photodynamics simulations of thymine: Relaxation into the first excited singlet state. *The Journal of Physical Chemistry A*, 113(45):12686–12693, 2009.
- [264] Mario Barbatti, Adélia J. A. Aquino, Jaroslaw J. Szymczak, Dana Nachtigallová, Pavel Hobza, and Hans Lischka. Relaxation mechanisms of uv-photoexcited dna and rna nucleobases. *Proceedings of the National Academy of Sciences*, 107(50):21453–21458, 2010.
- [265] Lara Martínez-Fernández, Jesús González-Vázquez, Leticia González, and Inés Corral. Time-resolved insight into the photosensitized generation of singlet oxygen in endoperoxides. *Journal of Chemical Theory and Computation*, 11(2):406–414, Feb 2015.

- 
- [266] Sebastian Mai, Anna-Patricia Wolf, and Leticia González. Curious case of 2-selenouracil: Efficient population of triplet states and yet photostable. *Journal of Chemical Theory and Computation*, 15(6):3730–3742, 2019.
- [267] Clemens Rauer, Juan J. Nogueira, Philipp Marquetand, and Leticia González. Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics. *Journal of the American Chemical Society*, 138(49):15911–15916, 2016.
- [268] Juan J. Nogueira and Leticia González. Computational photophysics in the presence of an environment. *Annual Review of Physical Chemistry*, 69(1):473–497, 2018.
- [269] ZongRong Xu and Spiridoula Matsika. Combined multireference configuration interaction/ molecular dynamics approach for calculating solvatochromic shifts: Application to the  $n \rightarrow \pi^*$  electronic transition of formaldehyde. *The Journal of Physical Chemistry A*, 110(43):12035–12043, 2006.
- [270] Irene Conti, Artur Nenov, Siegfried Höfner, Salvatore Flavio Altavilla, Ivan Rivalta, Elise Dumont, Giorgio Orlandi, and Marco Garavelli. Excited state evolution of dna stacked adenines resolved at the caspt2//casscf/amber level: from the bright to the excimer state and back. *Physical Chemistry Chemical Physics*, 17:7291–7302, 2015.
- [271] Angelo Giussani, Irene Conti, Artur Nenov, and Marco Garavelli. Photoinduced formation mechanism of the thymine–thymine (6–4) adduct in dna; a qm(caspt2//casscf):mm(amber) study. *Faraday Discussions*, 207:375–387, 2018.
- [272] J. Patrick Zobel, Juan J. Nogueira, and Leticia González. Quenching of charge transfer in nitrobenzene induced by vibrational motion. *The Journal of Physical Chemistry Letters*, 6(15):3006–3011, 2015.
- [273] Per Åke Malmqvist and Björn O. Roos. The casscf state interaction method. *Chemical Physics Letters*, 155(2):189 – 194, 1989.
- [274] Felix Plasser, Matthias Ruckebauer, Sebastian Mai, Markus Oppel, Philipp Marquetand, and Leticia González. Efficient and flexible computation of many-electron wave function overlaps. *Journal of Chemical Theory and Computation*, 12(3):1207–1219, 2016.

- 
- [275] Dominika Zgid and Marcel Nooijen. The density matrix renormalization group self-consistent field method: Orbital optimization with the density matrix renormalization group method in the active space. *The Journal of Chemical Physics*, 128(14):144116, 2008.
- [276] Katharina Boguslawski and Paweł Tecmer. Orbital entanglement in quantum chemistry. *International Journal of Quantum Chemistry*, 115(19):1289–1295, 2015.
- [277] Christopher J. Stein and Markus Reiher. Automated selection of active orbital spaces. *Journal of Chemical Theory and Computation*, 12(4):1760–1771, 2016.
- [278] Ruocheng Han and Sandra Luber. Complete active space analysis of a reaction pathway: Investigation of the oxygen–oxygen bond formation. *Journal of Computational Chemistry*, 41(17):1586–1597, 2020.
- [279] Johan Lorentzon, Markus P. Fuelscher, and Bjoern O. Roos. Theoretical study of the electronic spectra of uracil and thymine. *Journal of the American Chemical Society*, 117(36):9265–9273, 1995.
- [280] Markus P. Fülcher, Luis Serrano-Andrés, and Björn O. Roos. A theoretical study of the electronic spectra of adenine and guanine. *Journal of the American Chemical Society*, 119(26):6168–6176, 1997.
- [281] Christian Wiebeler, Veniamin Borin, Adalberto Vasconcelos Sanchez de Araújo, Igor Schapiro, and Antonio Carlos Borin. Excitation energies of canonical nucleobases computed by multiconfigurational perturbation theories. *Photochemistry and Photobiology*, 93(3):888–902, 2017.
- [282] Marko Schreiber, Mario R. Silva-Junior, Stephan P. A. Sauer, and Walter Thiel. Benchmarks for electronically excited states: Caspt2, cc2, ccsd, and cc3. *The Journal of Chemical Physics*, 128(13):134110, 2008.
- [283] Ron Shepard. The analytic gradient method for configuration interaction wave functions. In Yarkony R. D., editor, *Modern Electronic Structure Theory*, chapter 7, pages 345–458. World Scientific Publishing Co Pte Ltd, 1995.

- 
- [284] S. F. Boys and Alfred Charles Egerton. Electronic wave functions - i. a general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 200(1063):542–554, 1950.
- [285] Richard C. Raffenetti. General contraction of gaussian atomic orbitals: Core, valence, polarization, and diffuse basis sets; molecular integral evaluation. *The Journal of Chemical Physics*, 58(10):4452–4458, 1973.
- [286] Larry E McMurchie and Ernest R Davidson. One- and two-electron integrals over cartesian gaussian functions. *Journal of Computational Physics*, 26(2):218 – 231, 1978.
- [287] H. Bernhard Schlegel and Michael J. Frisch. Transformation between cartesian and pure spherical harmonic gaussians. *International Journal of Quantum Chemistry*, 54(2):83–87, 1995.
- [288] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 9 1976.
- [289] Evangelos A. Coutsiaris, Chaok Seok, and Ken A. Dill. Using quaternions to calculate rmsd. *Journal of Computational Chemistry*, 25(15):1849–1857, 2004.
- [290] Douglas L. Theobald. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallographica Section A*, 61(4):478–480, Jul 2005.
- [291] Pu Liu, Dimitris K. Agrafiotis, and Douglas L. Theobald. Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*, 31(7):1561–1563, 2010.
- [292] Ansgar Schäfer, Christian Huber, and Reinhart Ahlrichs. Fully optimized contracted gaussian basis sets of triple zeta valence quality for atoms li to kr. *The Journal of Chemical Physics*, 100(8):5829–5835, 1994.
- [293] Barnett Rosenberg, Loretta van Camp, and Thomas Krigas. Inhibition of cell division in escherichia coli by electrolysis products from a platinum electrode. *Nature*, 205(4972):698–699, Feb 1965.

- 
- [294] Barnett Rosenberg, Loretta van Camp, James E. Trosko, and Virginia H. Mansour. Platinum compounds: a new class of potent antitumour agents. *Nature*, 222(5191):385–386, Apr 1969.
- [295] Elizabeth R. Jamieson and Stephen J. Lippard. Structure, recognition, and processing of cisplatin-dna adducts. *Chemical Reviews*, 99(9):2467–2498, 1999.
- [296] Suzanne E. Sherman, Dan Gibson, Andrew H. J. Wang, and Stephen J. Lippard. X-ray structure of the major adduct of the anticancer drug cisplatin with dna: cis-[pt(nh<sub>3</sub>)<sub>2</sub>d(pGpG)]. *Science*, 230(4724):412–417, 1985.
- [297] Franck Legendre, Véronique Bas, Jiří Kozelka, and Jean-Claude Chottard. A complete kinetic study of gg versus ag platination suggests that the doubly aquated derivatives of cisplatin are the actual dna binding species. *Chemistry – A European Journal*, 6(11):2002–2010, 2000.
- [298] Nuno Martinho, Tânia C. B. Santos, Helena F. Florindo, and Liana C. Silva. Cisplatin-membrane interactions and their influence on platinum complexes activity and toxicity. *Frontiers in Physiology*, 9, 2019.
- [299] Shaloam Dasari and Paul Bernard Tchounwou. Cisplatin in cancer therapy: Molecular mechanisms of action. *European Journal of Pharmacology*, 740:364–378, 2014.
- [300] Tomaz Makovec. Cisplatin and beyond: molecular mechanisms of action and drug resistance development in cancer chemotherapy. *Radiology and Oncology*, 53(2):148–158, 2019.
- [301] Yogita Mantri, Stephen J. Lippard, and Mu-Hyun Baik. Bifunctional binding of cisplatin to dna: Why does cisplatin form 1,2-intrastrand cross-links with ag but not with ga? *Journal of the American Chemical Society*, 129(16):5023–5030, 2007.
- [302] Mu-Hyun Baik, Richard A. Friesner, and Stephen J. Lippard. Theoretical study of cisplatin binding to purine bases: Why does cisplatin prefer guanine over adenine? *Journal of the American Chemical Society*, 125(46):14082–14092, 2003.

- 
- [303] Johan Raber, Chuanbao Zhu, and Leif A. Eriksson. Theoretical study of cisplatin binding to dna: The importance of initial complex stabilization. *The Journal of Physical Chemistry B*, 109(21):11006–11015, 2005.
- [304] Shahana Dilruba and Ganna V. Kalayda. Platinum-based drugs: past, present and future. *Cancer Chemotherapy and Pharmacology*, 77(6):1103–1124, Jun 2016.
- [305] Timothy C. Johnstone, Ga Young Park, and Stephen J. Lippard. Understanding and improving platinum anticancer drugs – phenanthriplatin. *Anticancer Research*, 34(1):471–476, 2014.
- [306] Qiling Chen, Yuanyuan Yang, Xun Lin, Wen Ma, Gui Chen, Wenliang Li, Xuefeng Wang, and Zhiqiang Yu. Platinum(iv) prodrugs with long lipid chains for drug delivery and overcoming cisplatin resistance. *Chemical Communications*, 54:5369–5372, 2018.
- [307] Reece G. Kenny, Su Wen Chuah, Alanna Crawford, and Celine J. Marmion. Platinum(iv) prodrugs – a step closer to ehrlich’s vision? *European Journal of Inorganic Chemistry*, 2017(12):1596–1612, 2017.
- [308] Srinivas Ramachandran, Brenda Temple, Anastassia N. Alexandrova, Stephen G. Chaney, and Nikolay V. Dokholyan. Recognition of platinum–dna adducts by hmgb1a. *Biochemistry*, 51(38):7608–7617, 2012.
- [309] Daniele Veclani, Andrea Melchior, Marilena Tolazzi, and José P. Cerón-Carrasco. Using theory to reinterpret the kinetics of monofunctional platinum anticancer drugs: Stacking matters. *Journal of the American Chemical Society*, 140(43):14024–14027, 2018.
- [310] Nasma D. Eljack, Ho-Yu M. Ma, Janine Drucker, Clara Shen, Trevor W. Hambley, Elizabeth J. New, Thomas Friedrich, and Ronald J. Clarke. Mechanisms of cell uptake and toxicity of the anticancer drug cisplatin. *Metallomics*, 6:2126–2133, 2014.
- [311] Jiří Kozelka. Molecular origin of the sequence-dependent kinetics of reactions between cisplatin derivatives and dna. *Inorganica Chimica Acta*, 362(3):651–668, 2009. Protagonists in Chemistry: Bernhard Lippert.

- 
- [312] Shang-Hung Chen and Jang-Yang Chang. New insights into mechanisms of cisplatin resistance: From tumor cell to microenvironment. *International Journal of Molecular Sciences*, 20(17), 2019.
- [313] Yanyan Zhang, Wenjuan Zeng, Feifei Jia, Juan Ye, Yao Zhao, Qun Luo, Zihua Zhu, and Fuyi Wang. Cisplatin-induced alteration on membrane composition of a549 cells revealed by tof-sims. *Surface and Interface Analysis*, 52(5):256–263, 2020.
- [314] Srinivasan Ramachandran, Arjan P. Quist, Sashi Kumar, and Ratnesh Lal. Cisplatin nanoliposomes for cancer therapy: Afm and fluorescence imaging of cisplatin encapsulation, stability, cellular uptake, and toxicity. *Langmuir*, 22(19):8156–8162, 2006.
- [315] Amélie Rebillard, Xavier Tekpli, Olivier Meurette, Odile Sergent, Gwenaëlle LeMoigne-Muller, Laurent Vernhet, Morgane Gorria, Martine Chevanne, Markus Christmann, Bernd Kaina, Laurent Counillon, Erich Gulbins, Dominique Lagadic-Gossmann, and Marie-Thérèse Dimanche-Boitrel. Cisplatin-Induced Apoptosis Involves Membrane Fluidification via Inhibition of NHE1 in Human Colon Cancer Cells. *Cancer Research*, 67(16):7865–7874, 08 2007.
- [316] G P Stathopoulos and T Boulikas. Lipoplatin formulation review article. *Journal of Drug Delivery*, 2012:581363, August 2011.
- [317] Gerrit van Meer and Anton I. P. M. de Kroon. Lipid map of the mammalian cell. *Journal of Cell Science*, 124(1):5–8, 01 2011.
- [318] S. Yesylevskyy, Bruno Cardey, S. Kraszewski, Sarah Foley, Mironel Enescu, Antônio M. da Silva, Hélio F. Dos Santos, and Christophe Ramseyer. Empirical force field for cisplatin based on quantumdynamics data: case study of new parameterization scheme for coordinationcompounds. *Journal of Molecular Modeling*, 21(10):268, Sep 2015.
- [319] Pengfei Li, Lin Frank Song, and Kenneth M. Merz. Systematic parameterization of monovalent ions employing the nonbonded model. *Journal of Chemical Theory and Computation*, 11(4):1645–1657, 2015.

- 
- [320] Pengfei Li and Kenneth M. Merz. Mcpb.py: A python based metal center parameter builder. *Journal of Chemical Information and Modeling*, 56(4):599–604, 2016.
- [321] Pengfei Li and Kenneth M. Merz. Metal ion modeling using classical mechanics. *Chemical Reviews*, 117(3):1564–1686, 2017.
- [322] Juliana Fedoce Lopes, Victor Ströele de A. Menezes, Hélio A. Duarte, Willian R. Rocha, Wagner B. De Almeida, and Hélio F. Dos Santos. Monte carlo simulation of cisplatin molecule in aqueous solution. *The Journal of Physical Chemistry B*, 110(24):12047–12054, 2006.
- [323] Romelia Salomon-Ferrer, Andreas W. Götz, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 2. explicit solvent particle mesh ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, 2013.
- [324] Pedro A. Sánchez-Murcia, Juan J. Nogueira, and Leticia González. Exciton localization on ru-based photosensitizers induced by binding to lipid membranes. *The Journal of Physical Chemistry Letters*, 9(4):683–688, 2018.
- [325] Christopher T. Lee, Jeffrey Comer, Conner Herndon, Nelson Leung, Anna Pavlova, Robert V. Swift, Chris Tung, Christopher N. Rowley, Rommie E. Amaro, Christophe Chipot, Yi Wang, and James C. Gumbart. Simulation-based approaches for determining membrane permeability of small compounds. *Journal of Chemical Information and Modeling*, 56(4):721–733, 2016.
- [326] Norbert Kučerka, John F. Nagle, Jonathan N. Sachs, Scott E. Feller, Jeremy Pencer, Andrew Jackson, and John Katsaras. Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. *Biophysical Journal*, 95(5):2356–2367, 2008.
- [327] Callum J. Dickson, Lula Rosso, Robin M. Betz, Ross C. Walker, and Ian R. Gould. Gafflipid: a general amber force field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*, 8:9617–9627, 2012.

- 
- [328] Stephanie Tristram-Nagle, Horia I. Petrache, and John F. Nagle. Structure and interactions of fully hydrated dioleoylphosphatidylcholine bilayers. *Biophysical Journal*, 75(2):917–925, 1998.
- [329] W. Rawicz, K.C. Olbrich, T. McIntosh, D. Needham, and E. Evans. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophysical Journal*, 79(1):328–339, 2000.
- [330] E. Evans, W. Rawicz, and B. A. Smith. Concluding remarks back to the future: mechanics and thermodynamics of lipid biomembranes. *Faraday Discussions*, 161:591–611, 2013.
- [331] Chris Neale and Régis Pomès. Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(10):2539–2548, 2016. Biosimulations of lipid membranes coupled to experiments.
- [332] Martina De Vetta, Leticia González, and Juan J. Nogueira. Hydrogen bonding regulates the rigidity of liposome-encapsulated chlorin photosensitizers. *ChemistryOpen*, 7(6):475–483, 2018.
- [333] Qiankun Gong, Haomiao Zhang, Haozhe Zhang, and Changjun Chen. Calculating the absolute binding free energy of the insulin dimer in an explicit solvent. *RSC Adv.*, 10:790–800, 2020.
- [334] Benjamin Klasczyk, Volker Knecht, Reinhard Lipowsky, and Rumiana Dimova. Interactions of alkali metal chlorides with phosphatidylcholine vesicles. *Langmuir*, 26(24):18951–18958, 2010.
- [335] Virginia Miguel, Maria E. Defonsi Lestard, María E. Tuttolomondo, Sonia B. Díaz, Aida Ben Altabef, Marcelo Puiatti, and Adriana B. Pierini. Molecular view of the interaction of s-methyl methanethiosulfonate with dppc bilayer. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(1):38–46, 2016.
- [336] Andrea Melchior, José Manuel Martínez, Rafael R. Pappalardo, and Enrique Sánchez Marcos. Hydration of cisplatin studied by an effective ab initio pair potential

- 
- including solute–solvent polarization. *Journal of Chemical Theory and Computation*, 9(10):4562–4573, 2013.
- [337] Patrícia T. Martins, Adrian Velazquez-Campoy, Winchil L. C. Vaz, Renato M. S. Cardoso, Joana Valério, and Maria João Moreno. Kinetics and thermodynamics of chlorpromazine interaction with lipid bilayers: Effect of charge and cholesterol. *Journal of the American Chemical Society*, 134(9):4184–4195, 2012.
- [338] C. Peetla, A. Stine, and V. Labhasetwar. Biophysical interactions with model lipid membranes: Applications in drug discovery and drug delivery. *Molecular Pharmaceutics*, 6:1264–1276, 2009.
- [339] R. Zhang, X. Qin, F. Kong, P. Chen, and G. Pan. Improving cellular uptake of therapeutic entities through interaction with components of cell membrane. *Drug Delivery*, 26:328–342, 2019.
- [340] Nicole J. Yang and Marlon J. Hinner. Getting across the cell membrane: An overview for small molecules, peptides, and proteins. *Methods in Molecular Biology*, pages 29–53, 2015.
- [341] Annela M. Seddon, Duncan Casey, Robert V. Law, Antony Gee, Richard H. Templer, and Oscar Ces. Drug interactions with lipid membranes. *Chemical Society Reviews*, 38:2509–2519, 2009.
- [342] S. Zhang, K. S. Lovejoy, J. E. Shima, L. L. Lagpacan, Y. Shu, A. Lapuk, Y. Chen, T. Komori, J. W. Gray, X. Chen, S. J. Lippard, and K. M. Giacomini. Organic cation transporters are determinants of oxaliplatin cytotoxicity. *Cancer Research*, 66:8847–8857, 2006.
- [343] I. E. Smith and D. C. Talbot. Cisplatin and its analogues in the treatment of advanced breast cancer: a review. *British Journal of Cancer*, 65(6):787–793, Jun 1992.
- [344] K. Barabas, R. Milner, D. Lurie, and C. Adin. Cisplatin: a review of toxicities and therapeutic applications. *Veterinary and Comparative Oncology*, 6(1):1–18, 2008.

- 
- [345] Gwo Yaw Ho, Natasha Woodward, and Jermaine I.G. Coward. Cisplatin versus carboplatin: Comparative review of therapeutic management in solid malignancies. *Critical Reviews in Oncology/Hematology*, 102:37–46, 2016.
- [346] Chris Neale, Jenny C.Y. Hsu, Christopher M. Yip, and Régis Pomès. Indolicidin binding induces thinning of a lipid bilayer. *Biophysical Journal*, 106(8):L29–L31, 2014.
- [347] Hugo A. L. Filipe, Maria João Moreno, Tomasz Róg, Ilpo Vattulainen, and Luís M. S. Loura. How to tackle the issues in free energy simulations of long amphiphiles interacting with lipid membranes: Convergence and local membrane deformations. *The Journal of Physical Chemistry B*, 118(13):3572–3581, 2014.
- [348] D. Bochicchio, E. Panizon, R. Ferrando, L. Monticelli, and G. Rossi. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *The Journal of Chemical Physics*, 143:144108, 2015.
- [349] C. Neale and R. Pomès. Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858:2539–2548, 2016.
- [350] Ka Un Lao and John M. Herbert. Accurate intermolecular interactions at dramatically reduced cost: Xpol+sapt with empirical dispersion. *The Journal of Physical Chemistry Letters*, 3:3241–3248, 2012.
- [351] Yuezhi Mao, Paul R. Horn, and Martin Head-Gordon. Energy decomposition analysis in an adiabatic picture. *Physical Chemistry Chemical Physics*, 19:5944–5958, 2017.
- [352] Yuezhi Mao and Martin Head-Gordon. Probing blue-shifting hydrogen bonds with adiabatic energy decomposition analysis. *The Journal of Physical Chemistry Letters*, 10:3899–3905, 2019.
- [353] Federico Calle-Vallejo, Philippe Sautet, and David Loffreda. Understanding adsorption-induced effects on platinum nanoparticles: An energy-decomposition analysis. *The Journal of Physical Chemistry Letters*, 5(18):3120–3124, 2014.

- 
- [354] M. Bamdad, H. Farrokhpour, B. Najafi, and M. Ashrafizaadeh. Energy decomposition analysis of the intermolecular interaction energy between different gas molecules (h<sub>2</sub>, o<sub>2</sub>, h<sub>2</sub>o, n<sub>2</sub>, co<sub>2</sub>, h<sub>2</sub>s, and co) and selected li<sup>+</sup>-doped graphitic molecules: Df-sapt (dft) calculations. *Theoretical Chemistry Accounts*, 137:1–12, 2018.
- [355] Kevin Carter-Fenk, Ka Un Lao, Kuan-Yu Liu, and John M. Herbert. Accurate and efficient ab initio calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion. *The Journal of Physical Chemistry Letters*, 10:2706–2714, 2019.
- [356] Hajime Hirao. The effects of protein environment and dispersion on the formation of ferric-superoxide species in myo-inositol oxygenase (miox): A combined oniom(dft:mm) and energy decomposition analysis. *The Journal of Physical Chemistry B*, 115(38):11278–11285, 2011.
- [357] Hajime Hirao. Energy decomposition analysis of the protein environmental effect: The case of cytochrome p450cam compound i. *Chemistry Letters*, 40(10):1179–1181, 2011.
- [358] P. Jeffrey Hay and Willard R. Wadt. Ab initio effective core potentials for molecular calculations. potentials for the transition metal atoms sc to hg. *The Journal of Chemical Physics*, 82(1):270–283, 1985.
- [359] Willard R. Wadt and P. Jeffrey Hay. Ab initio effective core potentials for molecular calculations. potentials for main group elements na to bi. *The Journal of Chemical Physics*, 82(1):284–298, 1985.
- [360] Chris Vanessa Sumowski and Christian Ochsenfeld. A convergence study of qm/mm isomerization energies with the selected size of the qm region for peptidic systems. *The Journal of Physical Chemistry A*, 113(43):11734–11741, 2009.
- [361] Marius Retegan, Frank Neese, and Dimitrios A. Pantazis. Convergence of qm/mm and cluster models for the spectroscopic properties of the oxygen-evolving complex in photosystem ii. *Journal of Chemical Theory and Computation*, 9(8):3832–3842, 2013.

- 
- [362] Brian C Wilson and Michael S Patterson. The physics, biophysics and technology of photodynamic therapy. *Physics in Medicine & Biology*, 53(9):R61–109, April 2008.
- [363] Irene E. Kochevar, Mary C. Lynch, Shougang Zhuang, and Christopher R. Lambert. Singlet oxygen, but not oxidizing radicals, induces apoptosis in hl-60 cells. *Photochemistry and Photobiology*, 72(4):548–553, 2000.
- [364] Ron R. Allison and Keyvan Moghissi. Oncologic photodynamic therapy: Clinical strategies that modulate mechanisms of action. *Photodiagnosis and Photodynamic Therapy*, 10(4):331–341, 2013.
- [365] Qiu-ying Chen, Dong-hui Li, Huang-hao Yang, Qing-zhi Zhu, Jin-gou Xu, and Yang Zhao. Interaction of a novel red-region fluorescent probe, Nile blue, with dna and its application to nucleic acids assay. *Analyst*, 124:901–906, 1999.
- [366] Kazutaka Hirakawa, Kazuhiro Ota, Junya Hirayama, Shinji Oikawa, and Shosuke Kawanishi. Nile blue can photosensitize dna damage through electron transfer. *Chemical Research in Toxicology*, 27(4):649–655, 2014.
- [367] Lucjan Strekowski and Beth Wilson. Noncovalent interactions with dna: An overview. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 623(1):3–13, 2007. Non-covalent chemical/DNA interactions: Mechanisms and Consequences.
- [368] W. David Wilson, Fariat A. Tanious, Henryk J. Barton, Robert L. Jones, Lucjan Strekowski, and David W. Boykin. Binding of 4',6-diamidino-2-phenylindole (dapi) to gc and mixed sequences in dna: intercalation of a classical groove-binding molecule. *Journal of the American Chemical Society*, 111(13):5008–5010, 1989.
- [369] Teresa A. Larsen, David S. Goodsell, Duilio Cascio, Kazimierz Grzeskowiak, and Richard E. Dickerson. The structure of dapi bound to dna. *Journal of Biomolecular Structure and Dynamics*, 7(3):477–491, 1989.
- [370] Seog K. Kim, Svante Eriksson, Mikael Kubista, and Bengt Norden. Interaction of 4',6-diamidino-2-phenylindole (dapi) with poly[d(g-c)2] and poly[d(g-m5c)2]: evidence for major groove binding of a dna probe. *Journal of the American Chemical Society*, 115(9):3441–3447, 1993.

- 
- [371] Debapriya Banerjee and Samir Kumar Pal. Dynamics in the dna recognition by dapi: Exploration of the various binding modes. *The Journal of Physical Chemistry B*, 112(3):1016–1021, 2008.
- [372] Anette Larsson, Christina Carlsson, Mats Jonsson, and Bo Albinsson. Characterization of the binding of the fluorescent dyes yo and yoyo to dna by polarized light spectroscopy. *Journal of the American Chemical Society*, 116(19):8459–8465, 1994.
- [373] Barbara Gatto, Giuseppe Zagotto, Claudia Sissi, Cinzia Cera, Eugenio Uriarte, Giorgio Palù, Giovanni Capranico, and Manlio Palumbo. Peptidyl anthraquinones as potential antineoplastic drugs: Synthesis, dna binding, redox cycling, and biological activity. *Journal of Medicinal Chemistry*, 39(16):3114–3122, 1996.
- [374] Alexandra B. Ormond and Harold S. Freeman. Dye sensitizers for photodynamic therapy. *Materials*, 6(3):817–840, 2013.
- [375] L.R. Comini, I.M. Fernandez, N.B. Rumie Vittar, S.C. Núñez Montoya, J.L. Cabrera, and V.A. Rivarola. Short communication. *Phytomedicine*, 18(12):1093–1095, 2011.
- [376] Natalia Belen Rumie Vittar, Laura Comini, Ivana Maria Fernadez, Elizabeth Agostini, Susana Nuñez-Montoya, Jose Luis Cabrera, and Viviana Alicia Rivarola. Phototherapy using natural anthraquinones: Rubiadin and soranjidiol sensitize human cancer cell to die by apoptosis. *Photodiagnosis and Photodynamic Therapy*, 11(2):182–192, 2014.
- [377] Elise Dumont and Antonio Monari. Benzophenone and dna: Evidence for a double insertion mode and its spectral signature. *The Journal of Physical Chemistry Letters*, 4(23):4119–4124, 2013.
- [378] Juan J. Nogueira and Leticia González. Molecular dynamics simulations of binding modes between methylene blue and dna with alternating gc and at sequences. *Biochemistry*, 53(14):2391–2412, 2014.
- [379] Danith Ly, Yongzhi Kan, Bruce Armitage, and Gary B. Schuster. Cleavage of dna by irradiation of substituted anthraquinones: Intercalation promotes electron

- 
- transfer and efficient reaction at gg steps. *Journal of the American Chemical Society*, 118(36):8747–8748, 1996.
- [380] Gary B. Schuster. Long-range charge transfer in dna: Transient structural distortions control the distance dependence. *Accounts of Chemical Research*, 33(4):253–260, 2000.
- [381] F. Plasser. Theodore: A toolbox for a detailed and automated analysis of electronic excited state computations. *The Journal of Chemical Physics*, 152(8):084108, 2020.
- [382] Adèle D. Laurent and Denis Jacquemin. Td-dft benchmarks: A review. *International Journal of Quantum Chemistry*, 113(17):2019–2039, 2013.
- [383] Stephan Kupfer, Julien Guthmuller, and Leticia González. An assessment of rasscf and tddft energies and gradients on an organic donor–acceptor dye assisted by resonance raman spectroscopy. *Journal of Chemical Theory and Computation*, 9(1):543–554, 2013.
- [384] Farial A. Tanious, Terence C. Jenkins, Stephen Neidle, and W. David Wilson. Substituent position dictates the intercalative dna-binding mode for anthracene-9,10-dione antitumor drugs. *Biochemistry*, 31(46):11632–11640, 1992.
- [385] Dennis Kreft, Ying Wang, Michael Rattay, Katja Toensing, and Dario Anselmetti. Binding mechanism of anti-cancer chemotherapeutic drug mitoxantrone to dna characterized by magnetic tweezers. *Journal of Nanobiotechnology*, 16(1):56, Jul 2018.
- [386] Bengt Nordén and Folke Tjerneld. Structure of methylene blue–dna complexes studied by linear and circular dichroism spectroscopy. *Biopolymers*, 21(9):1713–1734, 1982.
- [387] Juan J. Nogueira and Leticia González. Computational photophysics in the presence of an environment. *Annual Review of Physical Chemistry*, 69(1):473–497, 2018.
- [388] Elise Dumont, Meilani Wibowo, Daniel Roca-Sanjuán, Marco Garavelli, Xavier Assfeld, and Antonio Monari. Resolving the benzophenone dna-photosensitization

- 
- mechanism at qm/mm level. *The Journal of Physical Chemistry Letters*, 6(4):576–580, 2015.
- [389] Felix Plasser, Adelia J. A. Aquino, William L. Hase, and Hans Lischka. Uv absorption spectrum of alternating dna duplexes. analysis of excitonic and charge transfer interactions. *The Journal of Physical Chemistry A*, 116(46):11151–11160, 2012.
- [390] Lea M. Ibele, Pedro A. Sánchez-Murcia, Sebastian Mai, Juan J. Nogueira, and Leticia González. Excimer intermediates en route to long-lived charge-transfer states in single-stranded adenine dna as revealed by nonadiabatic dynamics. *The Journal of Physical Chemistry Letters*, 11(18):7483–7488, 2020.
- [391] Miquel Huix-Rotllant, Johanna Brazard, Roberto Improta, Irene Burghardt, and Dimitra Markovitsi. Stabilization of mixed frenkel-charge transfer excitons extended across both strands of guanine–cytosine dna duplexes. *The Journal of Physical Chemistry Letters*, 6(12):2247–2251, 2015.
- [392] Patrizia Agostinis, Kristian Berg, Keith A. Cengel, Thomas H. Foster, Albert W. Girotti, Sandra O. Gollnick, Stephen M. Hahn, Michael R. Hamblin, Asta Juzeniene, David Kessel, Mladen Korbelik, Johan Moan, Pawel Mroz, Dominika Nowis, Jacques Piette, Brian C. Wilson, and Jakub Golab. Photodynamic therapy of cancer: An update. *CA: A Cancer Journal for Clinicians*, 61(4):250–281, 2011.
- [393] Aleksandra Kawczyk-Krupka, Andrzej M. Bugaj, Wojciech Latos, Katarzyna Zaremba, Katarzyna Wawrzyniec, Marek Kucharzewski, and Aleksander Sieroń. Photodynamic therapy in colorectal cancer treatment—the state of the art in pre-clinical research. *Photodiagnosis and Photodynamic Therapy*, 13:158–174, 2016.
- [394] Charles J. Gomer, Marian Luna, Angela Ferrario, and Natalie Rucker. Increased transcription and translation of heme oxygenase in chinese hamster fibroblasts following photodynamic stress or photofrin ii incubation. *Photochemistry and Photobiology*, 53(2):275–279, 1991.

- 
- [395] Thomas J. Dougherty, Charles J. Gomer, Barbara W. Henderson, Giulio Jori, David Kessel, Mladen Korbelik, Johan Moan, and Qian Peng. Photodynamic Therapy. *JNCI: Journal of the National Cancer Institute*, 90(12):889–905, 06 1998.
- [396] Susana C. Núñez Montoya, Laura R. Comini, Martín Sarmiento, Cecilia Becerra, Inés Albesa, Gustavo A. Argüello, and José L. Cabrera. Natural anthraquinones probed as type i and type ii photosensitizers: singlet oxygen and superoxide anion production. *Journal of Photochemistry and Photobiology B: Biology*, 78(1):77–83, 2005.
- [397] SP Lloyd. Least square quantization in pcm. bell telephone laboratories paper. published in journal much later: Lloyd, sp: Least squares quantization in pcm. *IEEE Transactions of Information Theory(1957/1982)*, 18:11, 1957.
- [398] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [399] Klaus Gollnick, Stephan Held, Daniel O. Mártire, and Silvia E. Braslavsky. Hydroxanthraquinones as sensitizers of singlet oxygen reactions: quantum yields of triplet formation and singlet oxygen generation in acetonitrile. *Journal of Photochemistry and Photobiology A: Chemistry*, 69(2):155–165, 1992.
- [400] Ilaria Zanellato, Ilaria Bonarrigo, Donato Colangelo, Elisabetta Gabano, Mauro Ravera, Manuela Alessio, and Domenico Osella. Biological activity of a series of cisplatin-based aliphatic bis(carboxylato) pt(iv) prodrugs: How long the organic chain should be? *Journal of Inorganic Biochemistry*, 140:219–227, 2014.

