U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PhD

3.º
CICLO

FCUP
2016

U.PORTO

Metals' Data for Biomolecular Force Fields

Rui Pedro Pimenta das Neves

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

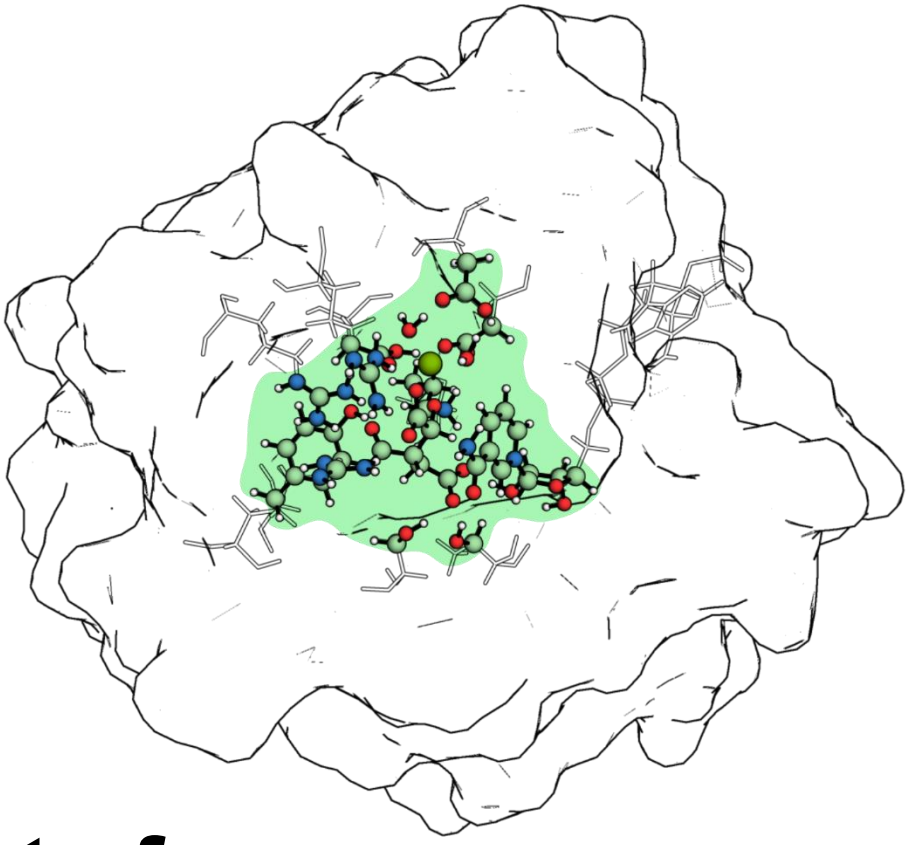
Metals' Data for Biomolecular Force Fields

Rui Pedro Pimenta das Neves

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Química

2016

PhD



Metals' Data for Biomolecular Force Fields

Rui Pedro Pimenta das Neves

Programa Doutoral em Química

Departamento de Química e Bioquímica

2016

Supervisor

Maria João Ribeiro Nunes Ramos, Full Professor, FCUP

Acknowledgements

First of all, I thank my supervisor, Professor Maria João Ramos: by seeing potential in me, when I had no prior research experience; by welcoming me in her group; by allowing me to strive and create unrestrainedly (within few limits); and by providing guidance (which is not easy with her flooded schedule). Professor Maria is not only a mentor, but also a full-hand leader, and I have definitely learned a lot from her.

I would also like to thank Professor Pedro Fernandes, which, together with Professor Maria, has provided guidance, creativity, and discussion over the course of this cycle.

I am also thankful to Sérgio Sousa. He was a pillar during my first steps as a “researcher”; he has shown patience, guidance, and has done far more than he was addressed to during my first year in the group.

To Fabiola Medina and Ana Barbosa, I am to apologize for not being able to keep up with you throughout these last six months. I want to praise you for your fast learning skills, and your interest. I hope that I have been able to aid, provoke and guide you; and that you will become bright investigators in the future to come.

To every member of this group (either here or gone): I want to thank for all the support, patience and company that you have offered me. To Alexandra Carvalho, Ana Teixeira, Daniel Dourado, and Juliana Garcia; you are not to be forgotten. To Diana Gesto, Eduardo Oliveira and Marco Lerario; my fellow companions in this four-year journey. To Cátia Moreira and Diogo Martins; good friends I have made here. To António Ribeiro and Óscar Passos; two persons that I greatly cherish and look up to (in several ways). To Natércia Brás (or Braz), Rita Calixto and Sílvia Martins; three invaluable friends with who I have shared much more than work.

A special thank you to Pablo Venegas for the patience, teachings and discussion on the mathematical formalism in some sections of this thesis.

Finally, my family (João Coimbra included), and friends Miguel Beça, Ricardo Oliveira, and Maria João Moura. I do not need to extend myself to thank my family, for obvious reasons: I choose my path, because they are my family. Even though I standalone individually, these people love me in a way that largely surpasses my research skills, and that I will never be able to fully understand.

I would like to thank Fundação para a Ciência e Tecnologia for the grant SFRH/BD/78397/2011.

Minha imaginação doente. Meu pensar de loucura. Entender. Porquê a obsessão de entender o que não tem entendimento possível? Porquê a obsessão de ter de haver uma resposta, apenas porque houve uma pergunta? Todo o entender é no impossível que tem o seu limite. Mas o impossível é a medida do homem e a sua vocação. Aí sou. Aí estou.

Vergílio Ferreira, in Pensar (1992)

One cannot escape the feeling that these mathematical formulae have an independent existence and intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them.

Heinrich Rudolf Hertz

Abstract

The driving force that motivated this Doctor of Philosophy thesis resided in the will to better understand the interactions between metals, particularly transition metals, and proteins. To produce insight in this field we have employed a combination of theoretical models and computational techniques to describe and emphasize the contribution of metals to protein dynamics and enzyme catalysis. Initially, we had planned to provide a classical description of the metal interactions in proteins to enhance the applicability of current classical force fields (FF) methodologies to any metalloprotein; however the insight that we can draw from molecular dynamics simulations is limited by current FF equations, which are chemically non-dynamic. Hence, to expand the study of the role of metals in proteins, we have also combined quantum mechanics/molecular mechanics (QM/MM) methodologies to study the effects of metal:protein interactions in enzyme catalysis.

Following this line of reasoning, this thesis comprehends four main lines of discussion: (1) the definition of the field of metalloproteins, detailing the main metals binding proteins and some of the most relevant processes in which they are involved; (2) the discussion of the chemical character of metal-ligand binding, followed by a depiction of the main limitations of computational methodologies in describing metal-ligand interactions; (3) the depiction of strategies currently employed when studying metalloproteins, either by MD simulations or electronic structure calculations; (4) and the presentation of the work developed throughout the extent of this PhD cycle, which resulted in three (four) more relevant publications.

The first work that we present here (Chapter 3) concerns the search for a linker between the fields of classical and quantum mechanics, and focuses on the parameterization of a set of 12 single-cluster manganese metalloproteins employing a bonded approach. The main reasons that motivated this work are: (1) classical Molecular Dynamics (cMD) are, currently, the sole methodology we have available to provide atomistic biophysical and biochemical insight at a timescale that can range from nanoseconds (ns) to milliseconds (ms); (2) metals are important elements of protein structure that confer either stability or function to most of the proteins that have been crystallographed to date; and (3) a database including parameters to describe the covalent character for several metals, particularly the most abundant transition metals (Zn, Fe, Mn and Cu), is still lacking. The results we present are important to broaden the applicability of several computational techniques, such as QM/MM, free-energy perturbation (FEP), thermodynamic integration (TI) or enhanced sampling MD, to manganese containing systems.

In a second work (Chapter 4), we have studied the thiol-disulphide exchange reaction with density functional theory (DFT) and accurate *ab initio* methods (MP2, CCSD(T)). This is a key reaction in enzymes holding iron-sulphur clusters, as observed in the Photosystem I (PSI) for instance. Even though there is extensive literature in DFT benchmarking studies of metal interactions with several biological units, these alone are insufficient to provide fully accurate insight in a chemical reaction where the metal is a part of the chemical reacting system. Therefore, work that evaluates the performance of DFT methods for critical reactions in biochemistry, in particular the redox reactions, is of utmost importance. We have crossed a plethora of density functionals (DFs) against reference values derived from CCSD(T) calculations so as to characterize and rank the chemical description of thiol-disulphide exchange by a set of 92 DFs, representative of the different "families" of exchange-correlation functionals. The results of this benchmarking are now being applied to the study of the catalytic power of protein disulphide isomerase (PDI) (Chapter 6). In particular, we are studying the reduction of glutathione disulphide, a small substrate that is fundamental to prevent oxidative damage to cells. We have determined the catalytic mechanism of PDI through ONIOM calculations, proposing the first thiol-disulphide exchange reaction, in which the PDI:glutathione intermediate is formed, as the rate-limiting step of the reaction.

The third work that we present here (Chapter 5) concerns the characterization of the catalytic mechanism by the Mg/Mn-dependent isocitrate dehydrogenase (ICDH), an enzyme essential in the Krebs cycle and with relevant clinical applications known to date. We have employed QM/MM calculations, with the ONIOM methodology, to characterize the energy profile and atomistic/electronic pathway by which isocitrate and $NADP^+$ are converted to α-ketoglutarate and NADPH. Our work has also focused on the role that Mg^{2+} exhibits in ICDH, particularly from analysis from charge transfer and coordination sphere geometry considerations. From this study, future studies will target the comparison of the performance of ICDH in the presence of Mg^{2+} and Mn^{2+}, the latter often observed in several bacteria and plant forms, and in the absence of the metal cofactor.

Besides these studies, several other contributions were developed in- and out- of the scope of the main theme of the thesis. Overall, we have provided relevant chemical insight in metal behaviour in metalloproteins.

Keywords: Metalloproteins, molecular modelling, biomolecular force fields, metallocentre parameterization, density functional theory, hybrid QM/MM methodologies, enzymatic catalysis.

Resumo

A força motriz desta dissertação de doutoramento derivou da vontade de compreender as interações entre metais, em particular metais de transição, e proteínas. Os resultados que produzimos resultam da combinação de modelos teóricos e metodologias computacionais, e visam enfatizar a contribuição dos metais na dinâmica das proteínas e na catálise pelas enzimas. Numa fase inicial planeamos potenciar a aplicabilidade das metodologias que fazem uso dos campos de forças biomoleculares (de natureza empírica) a qualquer metaloproteína. No entanto, o conhecimento que podemos construir com simulações de dinâmica molecular está limitado pelo formalismo empírico inerente a esses campos de força, quimicamente inertes. No seguimento desta limitação, e para melhor compreender a função dos metais em metaloproteínas, combinamos as simulações de dinâmica molecular com metodologias híbridas, que combinam cálculos de estrutura eletrónica e mecânica molecular, para estudar as consequências da interação metal:proteína na catálise enzimática.

Consequentemente, propomos a divisão desta dissertação de doutoramento em quatro frentes: (1) o estado atual do conhecimento no campo das metaloproteínas, onde nos centramos no tipo de metais que ligam frequentemente a proteínas e discutimos alguns dos processos mais relevantes que envolvem a participação das mesmas; (2) a discussão do caráter químico da interação metal-ligando, seguindo-se uma reflexão sobre os principais obstáculos que se colocam do ponto de vista das metodologias computacionais que são comumente empregues para estudar esta interação; (3) a exploração das estratégias atualmente utilizadas para estudar as metaloproteínas, quer por cálculos de simulação de dinâmica molecular (MD), quer por cálculos de estrutura eletrónica (QM); e (4) a apresentação dos trabalho especificamente desenvolvido no âmbito do período em que decorreu este projeto de doutoramento, do qual resultaram três publicações (com uma quarta em preparação) que temos como mais relevantes para este projeto.

O primeiro trabalho que apresentamos (Capítulo 3), debruça-se sobre o processo através do qual a mecânica clássica e a quântica se encontram, e versa, especificamente, a determinação de parâmetros de mecânica molecular para 12 enzimas que empregam o catião manganês como cofator, usando um modelo no qual o catião metálico interage com o átomo dador dos ligandos por intermédio de potenciais harmónicos. As motivações para este trabalho prendem-se com: (1) o facto da dinâmica molecular ser, ainda, a única metodologia que permite inferir sobre a biofísica e a bioquímica das metaloproteínas, na escala atómica, quando os processos em estudo se

estendem na escala do nanossegundo (ns) ao milissegundo (ms); (2) o facto de os metais serem elementos importantes na estabilidade, estrutura e função da larga maioria de proteínas que foram já cristalografadas; e (3) o facto de não estar ainda implementado um repositório de parâmetros de mecânica clássica para os metais de transição mais abundantes em biologia (Zn, Fe, Mn e Cu). Os resultados deste trabalho são importantes para potenciar a aplicabilidade de várias técnicas computacionais que façam uso de potenciais de mecânica molecular, em sistema proteicos que usem o manganês como cofator. Tais podem compreender: técnicas que empreguem cálculos híbridos de mecânica quântica/mecânica molecular (QM/MM) ou que façam uso do formalismo da perturbação de energia livre (FEP), integração termodinâmica (TI), ou técnicas que de amostragem mais avançadas.

Num segundo trabalho (Capítulo 4), procedemos ao estudo a permuta tiol-dissulfureto, usando a teoria do funcional de densidade (DFT) e o métodos *ab initio* com inclusão da correlação eletrónica (MP2, CCSD(T)). Esta reação é muito importante em enzimas com núcleos de ferro-enxofre, como se pode observar no fotossistema I, por exemplo. Apesar de poderem ser encontrados na literatura estudos que avaliam a performance do DFT em centros metálicos, estes podem ser insuficientes em situações em que o metal é apenas uma das partes integrantes na reação química. Assim sendo, há necessidade de desenvolver trabalho que avalie a performance do DFT em reações que são ubíquas em bioquímica, e na classe das oxidorredutases em particular. Neste trabalho, comparamos o desempenho de vários funcionais de densidade com valores de referência que determinamos com CCSD(T). Desta forma, procuramos classificar a performance de um conjunto de 92 funcionais de densidade, atendendo a diferentes aproximações para o cálculo da energia de permuta-correlação eletrónicas, face a vários aspetos da reação (posição reativa, coordenada reacional, e termodinâmica da reação). Os resultados deste trabalho estão agora a ser utilizados para estudar o mecanismo catalítico da isomerase de dissulfuretos proteicos (PDI), nomeadamente na redução da glutationa dissulfureto, um dos seus substratos mais pequenos e que intervém na prevenção de dano oxidativo na célula. Fomos já capazes de determinar o mecanismo catalítico da PDI (Capítulo 6) usando a metodologia subtrativa ONIOM, e verificado que a primeira permuta tiol-dissulfureto, que origina o intermediário PDI:glutationa, constitui o passo limitante desta reação.

O terceiro trabalho que aqui apresentamos (Capítulo 5) versa a determinação do mecanismo catalítico da enzima isocitrato desidrogenase (ICDH), que requere magnésio ou manganês como cofator. Esta enzima é essencial no ciclo de Krebs, e tem já aplicações clínicas que têm sido amplamente discutidas). Para estudar este mecanismo, realizamos cálculos de QM/MM, com a

metodologia ONIOM, para determinar o perfil termodinâmico e o caminho atomístico pelo qual a ICDH converte o isocitrato e o $NADP^+$ em α-cetoglutarato e NADPH. O nosso trabalho também presta especial enfoque ao papel que o Mg^{2+} desempenha no ciclo catalítico da ICDH, recorrendo à análise de transferências de carga e considerações sobre coordenação ao metal, ao longo da reação. Partindo deste trabalho, pretendemos, futuramente, fazer um estudo comparativo sobre os efeitos do Mn^{2+} e a ausência de cofator metálico no mecanismo catalítico da ICDH.

Para além destes estudos, foram desenvolvidas outras contribuições dentro e fora do âmbito desta dissertação. No fim deste ciclo, consideramos que apresentamos resultados relevantes, embora ainda germinais, sobre o comportamento químico de metais em metaloproteínas.

Palavras-chave: Metaloproteínas, modelação molecular, campos de força biomoleculares, parametrização de centros metálicos, teoria do funcional de densidade, metodologias híbridas QM/MM, catálise enzimática.

List of abbreviations

ABS	Absorption spectroscopy
AIM	Atoms in molecules
αKG	α-ketoglutarate
AM1-BCC	Austin model 1 with bond charge corrections
AMBER	Assisted Model Building with Energy Refinement
CASSCF	Complete active-space self-consistent-field
CBS	Complete basis set
CC	Coupled-cluster
CCSD	Coupled-cluster with single and double excitations
CCSD(T)	Coupled-cluster with single and double excitations, and third order perturbation theory corrections
CD	Circular dichroism
CHARMM	Chemistry at HARvard Molecular Mechanics
CHELP	Charges from Electrostatic Potentials
CHELPG	Charges from Electrostatic Potentials using a Grid-based method
CI	Configuration interaction
CISD	Configuration interaction with single and double excitations
CPMD	Car-Parrinello molecular dynamics
CSF	Configuration state function
CVFF	Consistent valence force field
DF	Density functional
DFT	Density functional theory
ECP	Effective core potentials
ENDOR	Electron-Nuclear double resonance
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
ESEEM	Electron spin echo envelope modulation
ESP	Electrostatic potential
EVB	Empirical Valence Bond
EXAFS	Extended X-ray absorption fine structure
FAD	Flavin adenine dinucleotide

FCI	Full configuration interaction
FEP	Free-energy perturbation
FES	Free-energy surface
FF	Force field
GGA	Generalized-gradient approximation
GROMOS	GROningen MOlecular Simulation
GSH	Glutathione
GSSG	Glutathione disulphide
GTO	Gaussian-type orbitals
HF	Hartree-Fock
h-GGA	Hybrid-generalized-gradient approximation
hh-GGA	Double-hybrid generalized-gradient approximation
hm-GGA	Hybrid-meta-generalized-gradient approximation
HOMO	Highest occupied molecular orbital
ICDH	Isocitrate dehydrogenase
ICT	Isocitrate
IRC	Intrinsic reaction coordinate
KS	Kohn-Sham
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LUMO	Lowest unnocupied molecular orbital
MC	Monte Carlo
MCSCF	Multiconfiguration self-consistent-field
MD	Molecular dynamics
m-GGA	Meta-generalized-gradient approximation
MK	Merz-Kollman
MM	Molecular mechanics
MP*n*	n^{th}-order perturbation Møller-Plesset theory
MPPT	Møller-Plesset Perturbation theory
MRCI	Multi-reference configuration interaction
MSE	Mean signed error
MUE	Mean unsigned error
NADP	Nicotinamide adenine dinucleotide phosphate

NMR	Nuclear magnetic resonance
NPT	isobaric-isothermal ensemble
NVT	Canonical ensemble
ONIOM	Our own *N*-layered Integrated molecular Orbital and molecular Mechanics
OPLS	Optimized Potentials for Liquid Simulations
PAPS	Ribonucleotide, 3'-phosphoadenosine-5'-phosphosulfate
PBC	Periodic boundary conditions
PDI	Protein disulphide isomerase
PES	Potential energy surface
PME	particle-mesh-Ewald
QCISD(T)	Quadratic configuration interaction with single and double excitations, and third order perturbation theory corrections
QM	Quantum mechanics
QM/MM	Quantum mechanics/molecular mechanics
RESP	Restrained electrostatic potential
RMSd	Root-mean-square deviation
rs-h-GGA	Range-separated hybrid-generalized-gradient approximation
SCF	Self-consistent-field
STO	Slater-type orbitals
TCA	Tricarboxylic acid
TI	Thermodynamic integration
TIP3P	Transferable intermolecular potential 3P
TS	Transition state
WFT	Wavefunction theory
XANES	X-ray absorption near edge structure
XAS	X-ray absorption spectroscopy
ZPE	Zero-point energy

Index

Acknowledgements

Abstract ..... **VII**

Resumo ..... **IX**

List of abbreviations ..... **XIII**

Index

Index of figures **7**

Index of schemes **13**

Index of tables **15**

Chapter 1: The Universe of Metalloproteins **17**

1.1. Overview of the Field 18

1.1.1. Molecular Mechanics calculations on metalloproteins *18*

1.1.2. Quantum Mechanics calculations in metalloproteins *20*

1.2. Metals in protein structure and function 23

1.2.1. Orbital overlapping and stereochemical effects *26*

1.2.2. Bioavailability and kinetics *27*

1.3. Metalloenzymes 28

1.3.1. Considerations on enzyme catalysis *29*

1.3.2. Metals in enzymatic catalysis *30*

1.4. Overview of Experimental Approaches to Metalloproteins 33

1.5. Modelling of metalloproteins 36

1.5.1. Describing a metal complex with classical mechanics *36*

1.5.2. Parameterization of Intramolecular Mechanical terms *40*

1.5.2.1. First-principles method 40

1.5.2.2. The Seminario Method 41

1.5.3. Parameterization of Electrostatic Atomic Charges *42*

1.5.4. Parameterizing the Lennard-Jones term *44*

1.5.5. Quantum Mechanics study of the Catalysis by Metalloenzymes *45*

1.5.5.1. Building an enzyme model 46

1.5.5.2. Establishing the catalytic mechanism 47

Chapter 2: Strategies to tackle metalloproteins 49

2.1. Molecular Dynamics 50

2.1.1. Expose the problem 50

2.1.2. Empirical Classical Force Fields 50

2.1.2.1. 1–2 Bond stretch 51

2.1.2.2. 1–3 Bond angle bend 52

2.1.2.3. 1–4 Dihedral and Improper torsions 52

2.1.2.4. Charge and dispersion interactions 54

2.1.3. Molecular minimization in Empirical Force Fields 56

2.1.4. Molecular Dynamics simulations by Empirical Force Fields 58

2.1.4.1. Trajectory propagator for MD simulations 58

2.1.4.2. Integration time step in MD simulations 59

2.1.4.3. Conditions of the phase space ensemble in MD simulations 60

2.2. Quantum Mechanics methods 65

2.2.1. Expose the problem 65

2.2.2. Building the molecular wave function 65

2.2.2.1. Linear Combination of Atomic Orbitals 67

2.2.2.2. Determining the molecular wave function: the secular equation 70

2.2.2.3. Many-electron molecular wave function: electron spin and Pauli exclusion principle 71

2.2.3. Hartree-Fock method 72

2.2.3.1. Hartree-Fock Hamiltonian 72

2.2.3.2. Roothaan-Hall approximation 74

2.2.3.3. Hartree-Fock limit: electron correlation 74

2.2.4. Post-Hartree-Fock methods 76

2.2.4.1. Multiconfiguration Self-Consistent-Field 76

2.2.4.2. Configuration Interaction 78

2.2.4.3. Møller-Plesset Perturbation Theory 79

2.2.4.4. Coupled-cluster Theory 82

2.2.5. Density Functional Theory: Electron Exchange and Correlation 83

2.2.5.1. Kohn-Sham formalism 85

2.2.5.2. Exchange-correlation density functionals 86

2.2.5.3. Density Functional Theory or Wave Function Theory? 88

2.3. Hybrid Methodologies 90

2.3.1. Expose the problem 90

2.3.2. Quantum Mechanics/Molecular Mechanics 90

2.3.2.1. The ONIOM subtractive method 91

2.3.3. The interaction between different layers 93
2.3.3.1. Mechanical embedding 94
2.3.3.2. Electrostatic embedding 95
2.3.4. The boundary description 97
2.3.5. Kinetics and phase space 98
Chapter 3: Parameters for Molecular Dynamics Simulations of Manganese-containing Metalloproteins 101
3.1. Introduction 103
3.2. Computational Details 105
3.2.1. Building parameterizable models 105
3.2.2. Bond and angle parameters 110
3.2.3. Atomic single charges 111
3.2.4. van der Waals parameters 111
3.2.5. Dihedral parameters 111
3.2.6. Validation procedure 111
3.3. Results and Discussion 113
3.3.1. Bond stretching parameters 113
3.3.2. Angle bending parameters 114
3.3.3. RESP charges calculation 116
3.3.4. Parameter Validation 118
3.3.4.1. RMSd values 118
3.3.4.2. Bond stretching and angle bending coordinates 120
3.3.4.3. Frequency and Normal Mode Analysis 122
3.3.4.4. Correlations between the Parameters 124
3.4. Conclusions 129
3.4.1. Supporting Information 130
Chapter 4: Benchmarking of Density Functionals for the Accurate Description of Thiol-disulphide Exchange 131
4.1. Introduction 133
4.2. Computational Methods 136
4.2.1. Model system 136
4.2.2. Reference structures for thiol-disulphide exchange 136
4.2.3. Reference energies for thiol-disulphide exchange 137
4.2.4. Geometry Benchmarking. 138

4.2.5. Electronic Energy and PES gradient benchmarking 139
4.3. Results and Discussion 140
4.3.1. Characterization of thiol-disulphide exchange. 140
4.3.2. Benchmarking geometry of the thiol-disulphide exchange model 142
4.3.3. Benchmarking the PES along the reaction coordinate 145
4.3.4. Benchmarking of the activation energy for thiol-disulphide exchange 149
4.3.5. DFT performance in the thiol-disulphide exchange reaction 154
4.4. Conclusions 155
4.4.1. Supporting Information 157
Chapter 5: Unveiling the Catalytic Mechanism of NADP$^+$-dependent Isocitrate Dehydrogenase with QM/MM calculations 159
5.1. Introduction 161
5.1.1. Relevance of the work. 161
5.1.2. General features 161
5.1.3. Catalytic/Kinetic insights 163
5.2. Computational Methods 165
5.2.1. Molecular model for hICDH 165
5.2.2. Defining the layers for the model and real system 166
5.2.3. QM/MM calculations 167
5.3. Results and Discussion 169
5.3.1. Modelling of the enzyme's active site 169
5.3.2. Which base facilitates NADP$^+$ reduction? 170
5.3.3. Reduction of NADP+. 171
5.3.4. Oxidative β-decarboxylation of OXS 173
5.3.5. Enolate protonation 175
5.3.6. Gibbs free-energy profile for the whole cycle 178
5.4. Conclusions 181
5.4.1. Supporting Information 183
Chapter 6: On the reduction of Glutathione disulphide by Protein Disulphide Isomerase: mechanistic insights provided by QM/MM methods 185
6.1. Introduction 187
6.1.1. Motivation 187
6.1.2. Structure and function of PDI 187

6.1.3. Glutathione/glutathione disulphide buffer in catalysis by hPDI 189
6.2. Computational Methods 193
6.2.1. Modelling the hPDI:GSSG complex 193
6.2.2. Building the ONIOM model 195
6.2.3. Establishing the catalytic mechanism.....197
6.3. Results and Discussion.....199
6.3.1. From validating the human PDI model, to optimizing the ONIOM model.....199
6.3.2. Considerations on the S_N2 nucleophilic attack of Cys53 to GSSG 201
6.3.3. S_N2 nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide 203
6.3.4. Oxidation of the a-domain: cleavage of the mixed-disulphide intermediate 207
6.4. Conclusions 211
References

Index of figures

Figure 1.1. Relative distribution of PDB structures containing at least one metal site. (accessed from the Protein Data Base on the 4th April 2015) 23

Figure 1.2. Relative distribution of the metalloenzymes with the prevalent eight metals, among each of the six enzyme classes. (adapted from Andreini *et al*.[77]).....28

Figure 1.3. Model representation of the non-bonded (on the left) and cation dummy (on the right) approaches to parameterize metal coordination spheres in metalloproteins. The *n*+ stands for the charge of the metal ion, and the *m* specifies the number of dummies that will be placed around the metal cation. The *r*-, *s*- and *t*- stand for the charge of the donor atoms in the ligands of the metal. 37

Figure 1.4. Schematic representation of the regions let loose or constrained during: (A) the bond, and (B) angle and torsion linear transit scans. The white-coloured ball and stick representation defines the set of atoms that are kept fixed during the procedure. 41

Figure 1.5. Representation of two types of modelling for large biomolecules: a hybrid QM/MM or QM/QM approach (on the left); a QM cluster approach (on the right). The atoms with contours are treated with QM methods. In the QM cluster model approach, the grey coloured spherical atoms concern hydrogen link atoms added to fill the valence of the cleaved model, also described at a QM level.....46

Figure 2.1. Normal modes employed in common implementations of empirical biomolecular force fields. 53

Figure 2.2. Comparison between the Coulomb and 12-6 Lennard-Jones potentials for a pair of atoms i and j, in kcal·mol^{-1}. The reference parameters adopted were 0.10 au for q_i and q_j, 0.10 kcal·mol^{-1} for ε_i and ε_j, and 2 Å for R_i and R_j.....55

Figure 2.3. Representation of the periodic boundary conditions for a protein system (in lime green) and its substrates (in cyan), solvated with water molecules (in magenta). The parallelepipedic unit cell is highlighted in solid black lines. 61

Figure 2.4. On the right, comparison of each pure primitive Gaussian function (pGTO) of the STO-3G *basis set* against the 1*s* Slater orbital of the hydrogen atom; on the left, comparison of the cGTO from the STO-3G *basis set* (combination of the three pGTOs at the right) against the 1*s* Slater orbital of the hydrogen atom. 69

Figure 3.1. Scheme on the semi-flexible approach took towards the study of manganese metallocentres. The grayish area delimits the atoms that were frozen during the geometry optimization (α-carbon, terminal nitrogen atom, and carbonyl group). 108

Figure 3.2. Representation of the 12 models parameterized in the presented study. The main residues are coloured in blue, red, orange, purple and green for histidines, aspartates, glutamates, asparagines and the manganese centre, respectively. Hydrogen, carbon and oxygen atoms are coloured in white, pale green and red, respectively. 109

Figure 3.3. (A) Bond scan procedure and (B) angle scan procedure. In both figures, the darker areas delimit the set of atoms frozen during the optimization prior to the PES determination and in the PES determination. 110

Figure 3.4. Potential energy surfaces for bond stretching for all models. 113

Figure 3.5. Angle force constants and equilibrium angles from the scan of the studied models, by donor atom type and coordination number. 115

Figure 3.6. Representation of the single point Merz-Kollman charges fitted with the RESP methodology for three models concerning manganese superoxide dismutase, and global charge of the residues for the three models derived from manganese superoxide dismutase. Charges are written in atom units (au). 116

Figure 3.7. Global charge distribution for the main residues of the studied models, by residue and coordination number of the model. 117

Figure 3.8. RMSd graphics (in thin lines) and the accumulated averages (thicker lines) for the models (4) to (7), during the 10 ns of MD simulation. 118

Figure 3.9. Distribution for the bonds of the model (6) with quartile representation and Gaussian curve fitting, for the last 4 ns of the MD run. All the representations and further validation on the gaussian fitting are in SI. 121

Figure 3.10. Distribution of the data for the angles in model (6) with quartile distribution and approaching a Gaussian curve, for the last 4 ns of the MD run, shown as an example. Other distribution data can be found in the SI. 122

Figure 3.11. Comparison of vibrational frequencies for the model HHDH[HO] (III) optimized with the AMBER force field and at B3LYP/SDD:6-31G(*d*,*p*) level of theory (on the left) and normal modes involving the donor and acceptor atoms (on the right). 123

Figure 3.12. Linear regression for the main ligands in manganese coordination spheres: aspartate, glutamate, histidine and water, in red, orange, blue and green, respectively. 125

Figure 3.13. Second order polynomial regression for the main donor atoms in manganese coordination spheres, oxygen and nitrogen, in red and blue colours, respectively.....125

Figure 3.14. Average equilibrium angles and force constants for main donor atoms concerning the equilibrium angles for the main geometries in manganese. 127

Figure 4.1. The enzyme thioredoxin glutathione reductase (2X8H) is shown in cartoon representation, with the Cys-X-X-Cys motif and the glutathione ligand in stick representation. Our model is highlighted by the black contour in the ball and stick representation. S_{lg}, S_{ctr} and S_{nuc} stand for the leaving group, central and nucleophilic sulphides, respectively. 136

Figure 4.2. Reaction states for thiol-disulphide exchange, reagent (R), transition state (TS) and product (P), obtained at the MP2/aug-cc-pVDZ level of theory.....137

Figure 4.3. PES profile for the CBS extrapolated CCSD(T), HF and $CCSD(T)_{correlation}$ energies. The left axis corresponds to HF/CBS and $CCSD(T)_{corr}$/CBS energies, and the right axis corresponds to the CCSD(T)/CBS energies obtained from Varandas' extrapolation scheme. .140

Figure 4.4. Energy profile for the thiol-disulphide reaction obtained with the CBS extrapolation scheme of Varandas, for the vacuum and implicit solvation models, using the geometries optimized in solvent. 141

Figure 4.5. Comparison of the PES scans for the selected functionals with the 6-31G(*d*) basis set, and the reference PES from CCSD(T)/CBS calculations. We analyse DF performance according to E_{xc} approximation for each set: (i) LDA, (ii) GGA and NGA, (iii) m-GGA and m-NGA, (iv) h-GGA and h-NGA, (v) hm-GGA and hm-NGA and (vi) hh-GGA. The CCSD(T)/CBS calculated PES is shown with grey dashes.....147

Figure 4.6. Effect of splitting, polarization and diffuse functions in the DF energies. The bars represent the energy determined with the smaller basis set and the arrows show the mean basis set truncation error obtained with larger basis sets in relation to the smaller basis set. The green region marks the limiting error of 1.00 $kcal \cdot mol^{-1}$. 150

Figure 4.7. Signed Error for the range separated version for a set of pure DFs from the study of the activation energy. The results are presented for the 6-311++G(2*d*,2*p*) basis set.....152

Figure 5.1. (A) hICDH, colored by domain (blue for the 'large domain', red for the 'small domain' and yellow for the 'clasp domain'). ICT, the Mg^{2+} ion and the $NADP^{+}$ cofactors are shown as green spheres; the monomer on the left-hand side is represented in cartoon and transparent surface, and the monomer on the right-hand side is represented with an opaque surface. (B) Residues within a 4 Å radius of the substrate (in ball and stick representation) plus the Mg^{2+} and $NADP^{+}$ cofactors; the protein carbon atoms are colored according to their domain (large, small or clasp), and the proposed catalytic residues are outlined with a black contour. 162

Figure 5.2. The QM/MM geometry–optimized active site of hICDH. The names of the catalytic residues are highlighted in bold. Mg^{2+}, ICT and $NADP^{+}$ are shown in black-stick and sphere representation. The residues shown in cyan-sticks establish important hydrogen bonds and ionic bridges with ICT, the nicotinamide moiety of $NADP^{+}$ and the coordination sphere of Mg^{2+}. 169

Figure 5.3. Stationary points of the $NADP^{+}$ reduction step – ICT, TS_{dehyd}, and OXS. The atomic charge variation, relative to the ICT state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations. 172

Figure 5.4. Stationary points of the OXS β-decarboxylation step – OXS, TS_{decarb}, and ENO. The atomic charge variation, relative to the OXS state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations. 174

Figure 5.5. Stationary points of the ENO protonation step-ENO, TS_{prot} and αKG. The atomic charge variation, relative to the ENO_{C1} state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations. 176

Figure 5.6. Thermodynamic profile for the $Lys212^{B}$-assisted catalysis of hICDH. The relative Gibbs free-energies (blue) and electronic energies (green) are read in the upper graphic, and the relative entropic contributions at 310 K and 1 bar (red) are read in the lower graphic. In the graphic depicting the $-T\Delta S$ contribution, all entropy-driven contributions are presented relative to the ICT state. The TS_{chel} and TS_{unchel} states do not present Gibbs free-energy corrections. 178

Figure 6.1. Depiction of the structure and function of human PDI (hPDI). The domains of hPDI are represented by the different colours of the surface representation. Cartoon representations in

magenta represent regions of the enzyme that have not been obtained from X-ray crystallography, and are thus estimated from modelling. 188

Figure 6.2. Result of the alignment of the backbone of the Cys12–Pro13–Tyr14–Cys15 from glutaredoxin with that of the Cys53–Gly54–His55–Cys56 of the *a*-domain of hPDI. 195

Figure 6.3. Optimized model of the domains *a*- and *b*- from hPDI complexed with GSSG. On the right, representation of the full ONIOM model; on the left, detailed representation of the DFT layer of the ONIOM model, with some of the more relevant interactions in it. 196

Figure 6.4. 2-dimensional rmsd for the four thioredoxin-folds (on the left), and the *a*- and *b*-domains (on the right) of hPDI throughout the 50 ns *NPT* cMD performed. 199

Figure 6.5. Secondary structure of the domain *a* and *b* of the hPDI, throughout the 50 ns cMD simulation in the *NPT* ensemble. 200

Figure 6.6. ONIOM energy of the nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide (in kcal·mol^{-1}), for two test models with a DFT layer of 99 atoms (grey) and 141 atoms (black). 201

Figure 6.7. Representation of the DFT layers for the two ONIOM models in which the His55-imidazole is either neutral (on the left, in palegreen) or in the cationic form (on the right, in palecyan). The GSSG substrate is represented in limegreen (on the left) and in greencyan (on the right). The figure depicts the conformation in the DFT layer when the nucleophilic Cys53-thiolate is at *circa* 2.40 Å from the GSSG-disulphide (which is usually the distance for which the trisulphide anion transition state is formed). The grey shades indicate the regions where residues occupied substantially different poses in both models. 202

Figure 6.8. Stationary points for the formation of the stable mixed-disulphide intermediate (GSSG, TS_{redox1}, GSM···GSX-Cyx53, Cys56···WAT, TS_{deprot}, and GSX-Cys53), and main distances throughout the transformation, in Å. Distances in bold correspond to distances directly related to the reaction coordinate explored through linear transit scans. Blue-to-red shades in atoms represent the variation in atomic charge relative to the GSSG state (from 0.07 to 0.30 a.u.); blue stands for decrease in atomic charge, and red stands for increase in atomic charge. 204

Figure 6.9. Thermodynamic profile for the formation of the mixed-disulphide intermediate between the domain *a* of hPDI and the GSSG substrate. All contributions are represented relatively to the initial hPDI:GSSG complex (GSSG). 206

Figure 6.10. Stationary points for the cleavage of the mixed-disulphide intermediate by the Cys56-thiolate (GSX-Cys53···Cys56, TS_{redox2}, GSM···Cys53-Cys56), and main distances throughout the transformation, in Å. Distances in bold correspond to distances directly related to the reaction coordinate explored through linear transit scans. Blue-to-red shades in atoms represent the variation in atomic charge relative to the GSX-Cys53···Cys56 state (from 0.07 to 0.30 a.u.); blue stands for decrease in atomic charge, and red stands for increase in atomic charge. 208

Index of schemes

Scheme 2.1. Flowchart for an MD simulation with an *NPT* ensemble.....64

Scheme 2.2. Summary of the iterative operations in an ONIOM calculation that occur during the mechanical, electrostatic and polarizable embedding schemes. 94

Scheme 5.1. Catalytic cycle proposed for the metal-dependent $NADP^+$-linked β-hydroxyacid oxidative decarboxylases (hICDH included).[440,457,458].....163

Scheme 6.1. Mechanism for the reduction of glutathione disulphide (GSSG) by the reduced *a*-domain of hPDI. 191

Index of tables

Table 1.1. Description of the parameters required by the molecular mechanics Hamiltonian for most currently implemented force fields. 19

Table 1.2. Physical description of the terms of the one-electron Kohn-Sham operator, in DFT. 21

Table 1.3. Patterns in metal centres from metalloproteins, for metal cations with biological significance.[67,70,73-75] 25

Table 1.4. Summary of advantages and disadvantages of non-bonded and bonded models for the parameterization of metal complexes 39

Table 2.1. Perturbation equations for the zeroth to fourth power in Møller-Plesset theory 81

Table 2.2. Scaling of the Hartree-Fock and Post-Hartree-Fock methods with the number of basis functions, N, to form the molecular wave function. 84

Table 3.1. List of parameterized manganese enzymes.[†] 106

Table 3.2. RMSd averages and standard deviations for the proteins (right column) and the small models (left column), in Å 119

Table 3.3. RMSd for the optimized models, in Å, from the comparison of the QM and MM optimized structures. 124

Table 3.4. Estimate values for the equilibrium angles in frequent manganese coordination geometries 128

Table 4.1. Electronic energies from single-point calculations for the R, TS and P states with MP2 and CCSD(T) methods, and CBS extrapolation. DZ stands for the aug-cc-pVDZ basis set, TZ stands for the aug-cc-pVTZ basis set and QZ stands for the aug-cc-pVQZ basis set. 141

Table 4.2. Best performing density functionals for the 6-31G(*d*) (grey coloured) and the 6-31+G(*d*) (white coloured) basis sets. The first four columns show the unsigned error with respect to the reference values for the tested DFs. The MUE refers to the mean unsigned average error of the S_{ctr}-S_{lg}, S_{ctr}-S_{nuc} and (S_{nuc}-S_{ctr}) – (S_{lg}-S_{ctr}) lengths. 143

Table 4.3. MUEs, in kcal·mol^{-1}, for a set of 29 DFs used to determine the PES from CCSD(T)/CBS calculations (see Figure 4.4). All DFT calculations were performed with the 6-31G(*d*) basis set. 146

Table 4.4. Set of DFs, within the 1 kcal·mol^{-1} error from the reference value for the 6-311++G(2*d*,2*p*) basis set. The DFs are displayed by exchange correlation energy (E_{xc}) approximation and increasing error. 149

Table 4.5. MSE in the activation energy for dispersion corrected DFs using the 6-311++G(2*d*,2*p*) basis set.....153

Table 5.1. Gibbs free-energies for Asp279^{A}- and Lys212^{B}-based reaction mechanisms. The double-split line separating ENO and ENO_{C2} refers to the exit of CO_2 from the active site; a process that we have not approached in our study. ICT, AKO, OXS, ENO, ENO_{C2} and αKG, refer to the stationary minima in the reaction profile; TS_{deprot}, TS_{dehyd}, TS_{decarb} and TS_{prot} refer to the saddle points in the reaction profile for the deprotonation of ICT, the dehydrogenation of AKO, the decarboxylation of OXS and the protonation of ENO, respectively. 171

Table 6.1. Structural and function data for hPDI.....189

Table 6.2. Gibbs free-energies for the catalytic cycle of the *a*-domain of our model of hPDI. All energies are presented in kcal·mol^{-1}; energies represented in curve brackets are given in Ha. GSSG, GSM···GSX-Cys53, Cys56···WAT, GSX-Cys53, GSX-Cys53···Cys56 (pre), GSX-Cys53···Cys56 and GSM···Cys53-Cys56 correspond to minimum energy stationary points at the mPW1N/6-31G(d):PARM99SB level of theory. TS_{redox1} and TS_{deprot} and TS_{redox2} correspond to transition state stationary points at the mPW1N/6-31G(d):PARM99SB level of theory. 198

Chapter 1: The Universe of Metalloproteins

As scientists move towards a more comprehensive understanding of the role of metals in biology, bioinorganic chemistry will be an increasingly important component of chemical biology.

*in Nature Chemical Biology **4**, 143 (2008)*

Bioinorganic Chemistry has emerged as a branch of Chemistry in the early 1950s, and has largely devoted its research interests to study the role of metals in biology, at both cellular and biochemical levels. The developments in the field were highly enhanced as several spectroscopic techniques were implemented, such as nuclear magnetic resonance (NMR), X-ray diffraction, electron paramagnetic resonance (EPR), among others. So far, these techniques have successfully made use of metals to provide for insight in protein structure, active-site mapping, drug-transport, among others.[1] Currently, metalloproteins constitute a large fraction of the protein structures that have been crystallographed so far, and intensive work keeps on being published to provide more and new structural data on other metalloproteins, as well as to unveil several protein mechanisms in which metals are involved.

1.1. Overview of the Field

Theoretical and computational chemistry has accompanied the trends of modern science, providing research of high scientific impact, either on its own or alongside experimental investigation. The field that concerns the study of metalloproteins' structures and functions is one such case. However, there are still major obstacles that are posed to the theoretical/computational chemist, while studying these systems, such as:

(1) Which requirements should a given Hamiltonian satisfy to adequately describe a metalloprotein?
(2) Which are the relevant properties that we intend to accurately determine when studying metal-containing systems?

The answers to these questions depend on the system, timescale, nature of the process and desired accuracy. Current methodologies employed in the study of metalloproteins are essentially built on classical mechanics or quantum mechanics (QM) considerations or both.

1.1.1. Molecular Mechanics calculations on metalloproteins

In a classical mechanics approach the atoms are bulky spheres that constitute the basic building blocks of the system. To construct the Hamiltonian (also known as force field) that describes the interactions between these spherical atoms, the motions of atoms are decoupled in intramolecular/mechanical (bond stretching, angle bending and dihedral torsion) and intermolecular/electrostatic (coulombic and van der Waals) terms, which are then described by simple potentials.

$$U(\mathbf{r}) = \sum_{l} K_l (l - l_0)^2 + \sum_{\theta} K_\theta (\theta - \theta_0)^2 + \sum_{\rho} K_\rho [1 + cos(n\rho - \gamma)] + \sum_{j>i} \sum_{i} \left(\varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}} \right)$$

Equation 1.1

Equation 1.1 illustrates a simple Hamiltonian that is the template of most force field implementations (*e.g.* AMBER,[2] OPLS,[3-5] CHARMM,[6,7] GROMOS[8] and CVFF)[9]. This equation

allows atomic forces to be derived quite quickly, for systems with thousands of atoms; hence one can easily obtain trajectories by applying simple equations of motion from classical mechanics to metalloprotein systems. Nowadays, these molecular dynamics (MD) simulations can already range from the femtosecond (fs) to millisecond (ms) timescales, with moderate computational cost. Hence, MD simulations have been extensively used to sample the conformational space of different biological systems with atomistic detail, assisting experimental techniques such as X-ray crystallography and NMR.[10] Additionally, MD-based free-energy integration methods, *e.g.* thermodynamic integration (TI) or free-energy perturbation (FEP), have also been employed to determine accurate protein-ligand binding or solvation free-energies;[11-14] and molecular mechanics (MM) parameters have been derived to combine with hybrid quantum mechanics/molecular mechanics (QM/MM)[15] methods.[16-24]

Table 1.1. Description of the parameters required by the molecular mechanics Hamiltonian for most currently implemented force fields.

Parameters	Description
K_l, l_0	Bond force constant, bond reference value
K_θ, θ_0	Angle force constant, angle reference value
K_ρ, γ, n	Torsion force constant, phase dihedral, periodicity
ε_{ij}, R_{ij}	Lennard-Jones potential energy depth, van der Waals radii of non-bonded atoms
q_i, q_j	Punctual single charges in non-bonded atoms

However, current force fields present a significant empirical nature, which means that extensive parameterization and validation are required, so that one can have confidence in the MD simulation's results. Table 1.1 presents a description of the parameters required for the simplest of the force field equations. The particular case of metalloproteins has been widely discussed in the literature throughout the past years. The main topics of discussion in the current literature are related to: (1) the adequacy of current classical force fields to describe the metal-ligand interaction, (2) the lack of a generalized force field to systematically describe the metal-ligand interactions, and (3) the definition of the parameterization scheme that best represents the metal-ligand interaction in current established force fields.[16,25-28] These issues remain unanswered to a

certain extent, nevertheless current research has addressed more closely the description of metallocentres by employing MM potentials similar to those in Equation 1.1.

1.1.2. Quantum Mechanics calculations in metalloproteins

Despite the fact that MD simulations have provided valuable insight on the way metalloproteins function in solution and in the cell, the chemical processes underlying their functions are hardly understood from a classical perspective. Chemical reactions in which metallocentres participate are often involved in hydrogen- or charge-transfer reactions, comprehending a scale where electron-electron interactions should be explicitly accounted for in the molecular Hamiltonian; hence QM calculations are required to provide detailed electronic insight in metal-guided transformations such as: electron excitations, coupled hydrogen-electron transfer, thiol-disulphide exchange or phosphodiester bond cleavage, among others. In this type of calculations two major aspects should be carefully considered: (1) the quality of the Hamiltonian and (2) the quality of the wave function for a metallosystem.

One major obstacle is presented immediately: calculations at the QM level are significantly more expensive than those performed at the MM level. Moreover, there is the question of "*Which Hamiltonian accurately describes an N-electron wave function from a metallosystem?*". The Hartree-Fock (HF) method is hardly adequate to describe such a system, since it neglects the substantial role that orbital degeneracy and electron-correlated motions play in these models. On the other hand, post-HF methods have been developed to account for these correlation effects; however, they are hardly applicable to systems of considerable size (with N_{atoms} ≥ 50 atoms), since the number of electron-integrals to be solved is circa N_{atoms}^2 times larger than those resulting from solving the N-electron Kohn-Sham (KS) Hamiltonian.

Currently, Density Functional Theory (DFT)[29-31] is the preferred methodology to perform such calculations in metalloprotein models, since it can account for electron-correlations effects at a moderate computational cost, for a few hundreds of atoms. Particularly in the last decades, such calculations have provided substantial contributions to the field of catalysis by metalloenzymes.[32,33]

$$\mathcal{H}^{KS}(\rho[\mathbf{r}']) = \sum_{e_i}^{e_N}\left[-\frac{1}{2}{\nabla_{e_i}}^2 - \sum_{n_i}\frac{Z_{n_i}}{r_{n_i,e_i}} + \frac{1}{2}\iiint\frac{\rho(\mathbf{r}')}{r'_{e_i}}d\mathbf{r}' + \mathcal{V}_{e_i}^{XC}\right] \qquad \textbf{Equation 1.2}$$

In DFT, the ground state Hamiltonian is completely described by a finite sum of one-electron Kohn-Sham operators, $\hat{h}^{KS}$. Contrarily to the HF theory, for which the wave function $|\Psi\rangle$ is built for a system of $3N$ coordinates, in DFT the N-electron system is described by an electron density that depends solely on the three-dimensional Euclidean space $\rho[\mathbf{r}']$. Such approximation speeds up substantially the calculation of the electronic repulsive potential terms in the Hamiltonian of Equation 1.2. In Table 1.2, there is a description of the main terms of the one-electron Kohn-Sham Hamiltonian.

Table 1.2. Physical description of the terms of the one-electron Kohn-Sham operator, in DFT.

$\hat{h}^{KS}$ terms	Description
$-\frac{1}{2}\nabla_{e_i}{}^2$	Operator for the i^{th}-electron kinetic energy, $\hat{t}_{e_i}$
$-\sum_{n_i}\frac{Z_{n_i}}{r_{n_i,e_i}}$	Operator for the attractive potential on the i^{th}-electron, from all nuclei of the system, $\hat{v}_{e_i}^{n}$
$\iiint \frac{\rho(\mathbf{r}')}{r'_{e_i}} d\mathbf{r}'$	Operator for the repulsive potential on the i^{th}-electron, from the average field of the N-electron system, $\hat{v}_{e_i}^{e}$
$\hat{v}_{e_i}^{XC}$	Exchange-correlation operator of the i^{th}-electron; no exact form is yet defined

Nevertheless, the practical applicability of DFT is questioned, and still dubiously applicable when experimental data is lacking. Despite being formally accurate, there is no analytical mathematical description for the exchange-correlation operator $\hat{v}^{XC}$, which is particularly problematic for metal systems. Hence, the choice of an adequate form for this operator to tackle a metal system has been widely tested and discussed in literature.[34] Currently, sets of exchange-correlation density functionals have been suggested to describe specific properties of these systems: geometrical features, spectroscopic data, thermochemistry or kinetics, among others.

On the other hand, the current description of the molecular wave function of a metal-containing system is based on the fact that the N-electron molecular wave function is a product of one-electron molecular wave functions, with the latter expressed in terms of one-electron atomic orbitals (which are user-defined). Several authors have parameterized families of basis sets that can be used to generate the initial guess for the N-electron molecular wave function. However, transition metals are mostly found in open *d*-shell configurations and exhibit a large density of

electronic states in their outer valence shell. Furthermore, with the increase in nuclear charge, relativistic effects may become relevant in the electrons of the inner core of the metal atoms and ions.[35,36] Hence, standard basis sets developed for main group atoms (*e.g.* Pople, Ahlrichs or Dunning basis sets)[37-49] may not describe adequately the one-electron atomic orbitals of transition metals. To overcome this, specific basis sets or pseudopotentials (which are less expensive) have been developed to accurately describe atomic properties of metals, such as electronic spectra, excitation energies or ionization energies. For most metallosystems, the 6-31G(*d*,*p*)/SDD hybrid basis set[19] is employed. The Stuttgart-Dresden pseudopotential (SDD)[35,50] treats the core electrons in the nucleus in an implicit way, and describes the valence electrons of the metal with the (8s7p6d)/[6s5p3d]-GTO *basis set*. Despite that it works well in most geometry optimizations and linear transit scan operations, the accurate derivation of atomic charge populations and mass-weighted frequencies generally requires *basis sets* with expanded polarization and diffuse properties.[19,20,28,51,52]

Despite the fact that QM methods have provided valuable data to unveil the importance of metal ions in areas such as enzyme catalysis,[33,53] several challenges remain to be overcome. The first, and more obvious, is the number of atoms to be included in the model to describe a metalloprotein. Secondly, there is the non-dynamic character of the Schrödinger's equation in most theoretical methodologies, which do not cover most of the system's phase space and lack the inclusion of accurate thermal and entropic contributions. To overcome such limitations, hybrid quantum mechanics/molecular mechanics (QM/MM) methodologies have been employed in protein systems;[32,54-58] and both enhanced sampling techniques and time-dependent QM-potentials have already been developed and implemented.[59-64] Nevertheless, such improvements come at a cost as long as technical feasibility is a limiting factor in simulation; and either numeric or chemical accuracy must be balanced when a computational experiment is designed.

Throughout the next sections of this chapter we will discuss specific characteristics of metalloproteins and some of the theoretical background on metal-coordinated centres. Then, we will proceed to discuss more specifics on methodologies that aim to improve the description of metal centres in biology and study biological processes in metalloproteins at an atomistic scale.

1.2. Metals in protein structure and function

Proteins are important biomolecules in the array of processes that operate the biological machinery of any organism. Within this category, proteins can be classified according to their function (as enzymes, chaperones, transporters, sensors…), and can depend on several cofactors that activate or potentiate their function. In particular, the role of metals in protein biology has been discussed since bioinorganic chemistry has emerged.

Metals are key elements in biology, with numerous biophysical and biochemical functions. Their relation to proteins is quite unique, since several metal ions exhibit low aqueous solubility (*e.g.* iron or copper) or an ionic character that makes it harder for passive diffusion to occur across the membrane.[65,66] Hence their transport is highly dependent on protein transporters. On the other hand, they can act homeostatically as charge carriers or regulators of the osmotic pressure in cells; and catalytically either in electron-transfer reactions or as structural cofactors for the active site of several enzymes. Thus, they are essential to the well-functioning of cell machinery.[67]

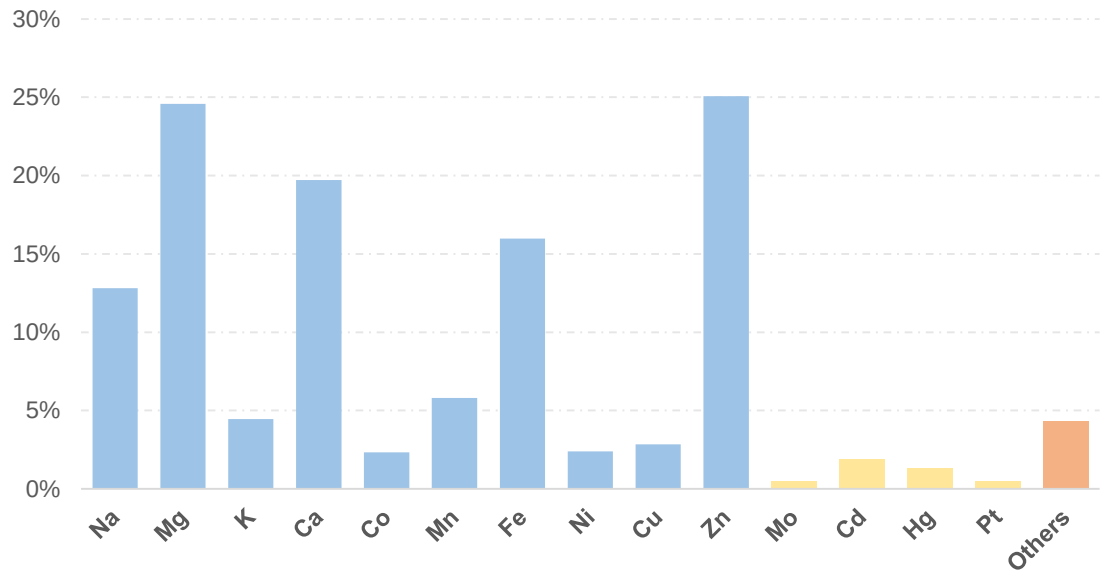


Figure 1.1. Relative distribution of PDB structures containing at least one metal site. (accessed from the Protein Data Base on the 4th April 2015)

It is not a surprise, though, that *circa* 40% of the current available crystallographic structures in the Protein Data Bank[68] have been described as having metals with either structural or catalytic function.[69] Among all the metals involved in the pathways that rule the cell's biology, those from the first three rows of the Periodic Table of Elements (alkali, alkali earth and first-row transition metals), along with molybdenum, are prevalent in the cells of most living beings. From these statistics, Na, Mg, K, Ca, Co, Mn, Fe, Ni, Cu and Zn are the metals that are more often found as natural cofactors of protein structure and function.[70] In Figure 1.1, this prevalence is emphasized for these ten metals (in light blue), and for more residual elements such as Mo, Cd, Hg and Pt (in light yellow).

While for alkali and alkali earth metals the oxidation states are completely known, most transition metals may exhibit more than one oxidation state in proteins. However, in the cell few oxidation states can be found. Most metals exhibit the divalent oxidation state (except for Mo and V, which rapidly form oxoanions), and trivalent metal cations are seldom found in biology (they are mostly involved in redox reactions, *e.g.* Fe^{3+} present in iron-sulphur cluster, or Mn^{3+} in the oxygen-evolving complex of Photosystem II).[71,72] This is a result of the abundant water solvent in the cell, which allows only for a low range of redox potentials in cells (from -0.4 V to 0.8 V). Therefore, we can make use of fundamental chemistry to understand the trends that metals exhibit in biology, particularly for the case of divalent metals.

These trends are mostly bound to the metal's charge/radius ratio, and the ionization energy required to transform the metal in its native state (M) to the ionized state (M^+). In this way, we define three main categories: hard, borderline and soft metals.[65,67]

- Hard metals exhibit higher ionization energies and electronic affinities and are prone to exhibit strong ionic interactions; hence they bind more effectively with small charged ligands with oxygen-donor atoms (mostly alkoxyde- and carboxylate-derived);
- Soft metal cations exhibit a stronger dative covalent character for the metal-ligand interaction, since the polarizability and directionality of paired electrons contribute to the bond in a larger extent. Nitrogen- and sulphur-derived ligands are most often found bound to these metal cations.

Table 1.3. Patterns in metal centres from metalloproteins, for metal cations with biological significance.[67,70,73-75]

Metal	Ligands	No. ligands (Symmetry)	Spin multiplicity
Na^+	Asp, Glu, O-carbonyl, H_2O	6 (O_h)	Singlet
K^+	Asp, Glu, O-carbonyl, H_2O	5 – 7	Singlet
Mg^{2+}	Asp, Glu, phosphate, H_2O	6 (O_h)	Singlet
Ca^{2+}	Asp, Glu, O-carbonyl, O-phosphate, H_2O	6 – 8	Singlet
Mn^{2+}	Asp, Glu, His, H_2O, O-phosphate	6 (O_h)	Sextet
Mn^{3+}	Asp, Glu, His, HO^-, O-phosphate	6 (O_h)	Quintet
Fe^{2+}	Cys	4 (T_d)	Quintet
	Asp, Glu, His, N-porphyrin	6 (O_h)	
Fe^{3+}	Cys	4 (T_d)	Sextet
	Asp, Glu, His	6 (O_h)	
Co^{2+}	Cys, Met, His	4 (T_d)	Quartet
	His, Asp, H_2O	5 (C_{4v}, D_{3h}), 6 (O_h)	Quartet
Co^{3+}	His, Asp, H_2O	6 (O_h)	Singlet
Ni^{2+}	Cys, His, N-cofactor F430	4 (D_{4h})	Singlet
	His, Asp, H_2O	6 (O_h)	Triplet
Cu^+	Cys, Met, His	4 (T_d)	Singlet
Cu^{2+}	Cys, Met, His	4 (T_d)	Doublet
	His, Asp, Glu	4 (D_{4h}), 5 (C_{4v})	Doublet
Zn^{2+}	Cys, His, Asp, Glu	4 (T_d)	Singlet
	His, Asp, Glu, H_2O	5 (C_{4v}, D_{3h})	Singlet
Mo^{4+}	Asp, Tyr, Cys, O-oxide, S-sulphide	6 (O_h)	Triplet
Mo^{5+}	Asp, Tyr, Cys, O-oxide, S-sulphide	6 (O_h)	Doublet
Mo^{6+}	Asp, Tyr, Cys, O-oxide, S-sulphide	6 (O_h)	Singlet

Therefore, it comes with no surprise that, in proteins, metal cations often coordinate to the amino acid's backbone, and polar (Asn, Gln, His) or negatively charged sidechains (Asp, Glu, Cys, Tyr).[65,67,70] However, the binding of transition metals with biological ligands does not depend solely on ligand charge and in the metal cation characteristics (charge, ionic radius, polarizability). Other properties, such as donor atom, steric environment, spin-pairing stabilization or bioavailability, are also key features to understand the metal-ligand distribution in cells.

1.2.1. Orbital overlapping and stereochemical effects

As referred, metal cations bind mostly to oxygen (in Asp, Glu, Asn, Gln), nitrogen (His) and sulphur (Cys, Met) donors.[70] Primarily, we may be tempted to associate their electron affinity and ionization potential to explain this statistic fact. In alkali and alkali earth metals this should be in fact true. However, in transition-metal cations orbital overlapping often plays an even larger role stabilizing the metallocentre in proteins.[67] All of the biological ligands found in metallocentres exhibit valence molecular orbitals with high directionality (*p*-type), which are prone to overlap significantly with the degenerate *d*-orbitals of transition-metal cations. This effect is more relevant when the binding valence orbitals have a more diffuse character, as it occurs for the sulphydril moiety of Cys or the π-system present in imidazole moiety of His. In this way, a larger energy gap occurs between the resulting *ligand*- and *antiligand*-molecular orbitals of the metallocentre, resulting in a very stable metal-protein complex. The latter π-delocalization is also the basis for the stabilization of several metal complexes, including metals such as iron, cobalt or nickel, by large aromatic cofactors, *e.g.* porphyrin, corrin, which present a high mechanical stability that results from the large π-system formed on the plane of the ring.[76] Metals such as copper, zinc, nickel or cobalt are often found organized in such manner. These systems are frequently discovered as tetra-coordinated centres: zinc and copper usually present a distorted tetrahedral geometry, while nickel or cobalt, which mostly bind to tetrapyrroles, are often found in a square planar geometry. However, these metals can also result in distorted octahedral geometries, when they bind to small ligands available in the cell.[76]

The relative metal/ligand size is also relevant for the binding and the geometry of the metal-protein centre. In particular, the stereochemistry of ligands can constrain the affinity of ligands for certain metals, contributing to ligand selectivity by the latter. In particular, large polydentate biological ligands are constantly being synthesized in the body to trap several metals, taking into account the size of the metal cation. However, it is not easy to establish quantitative relations considering

the size of metal cations, since several of them exhibit selective coordination geometry properties. As a result, these properties will also compete in ligand binding, aside from its size and stereo specificity.[65]

1.2.2. Bioavailability and kinetics

Aside from the purely chemical affinities of metals ligands, the bioavailability and kinetics of ligand uptake by proteins and the membrane are also quite important in biology. In the first row of divalent transition-metal cations, copper presents the softest character and also the higher binding constants for ligands with nitrogen-, oxygen- and sulphur-donor atoms. As a consequence, ligand concentrations have to be regulated, in and out of the cell, so that metal-ligand competition with the remaining metals in the cell is not compromised. This is traduced in a synthesis of these ligands in higher concentrations than that of the metal cation availability to the cell. Therefore, Cu^{2+} binds prevalently to nitrogen- and sulphur-donors, to which it exhibits higher binding constants; then Zn^{2+}, which is present in higher concentrations in the cell, captures the remaining nitrogen- and sulphur-donors and may bind to some carboxylate-derived ligands. On the other hand, Ni^{2+} and Co^{2+}, which exhibit similar binding constants to Zn^{2+}, are found in scarce concentrations in the cell; therefore, binding to Zn^{2+} is favoured by kinetics, and these metals are generally chelated by specific cofactors produced/uptaken in the cell (vitamin B12 and factor F-430). Finally, Fe^{2+} and Mn^{2+}, with the lowest binding constants, bind more often to nitrogen- and oxygen-donor atoms. The Fe^{2+} and Mn^{2+} centres exhibit the highest coordination (generally six-coordinated octahedral geometry), and weak ligand fields. This is traduced in lower binding constants, with rapid ligand exchange rates and lower concentration of ligands competing in metal-ligand binding. For trivalent metal cations, these 'rules' do not follow that easily.[65]

A downregulation of any of these mechanisms results in severe health conditions. In fact, metal trafficking has been identified to result in several diseases that affect human health at a global scale. The abnormal trafficking of metals, such as copper, zinc or iron, in the body has been related to several gene mutations that result in diseases related to the accumulation of metal ions in several tissues leading to liver failure, neurological disorder conditions, arthritis or diabetes. These biological malfunctions are of great interest for current fundamental and applied research in proteins that are involved in these processes, so as to develop new mechanisms to target these conditions. [69]

1.3. Metalloenzymes

A large fraction of the known metalloproteins are enzymes. In all six enzyme classes, metal cofactors have been identified in a significant extent: 44% in oxidoreductases, 40% in transferases, 39% in hydrolases, 36% in lyases, 36% in isomerases and 59% in ligases.[77] Figure 1.2 depicts the distribution of the known metalloenzymes among these six enzyme classes, for each of the eight prevalent metals with catalytic function (Mg, Zn, Fe, Mn, Ca, Co, Mo/W and Cu).

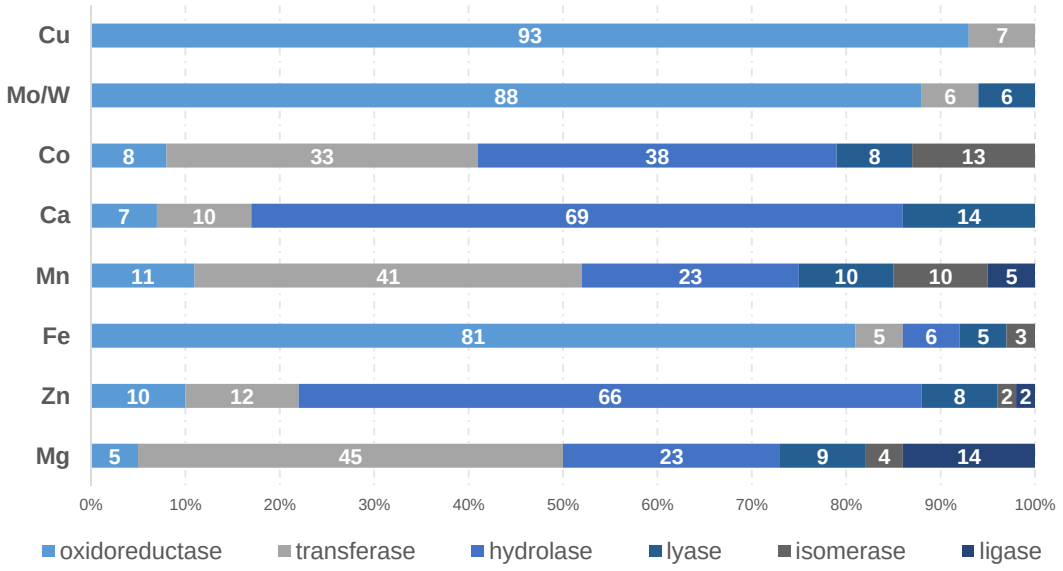


Figure 1.2. Relative distribution of the metalloenzymes with the prevalent eight metals, among each of the six enzyme classes. (adapted from Andreini *et al.*[77])

In metalloenzymes, metals have specific roles in the catalysis by the enzyme: they can function as anchors to place exogenous substrates in an adequate pose for catalysis to occur; or they can participate in charge transfer processes, in which they function as donor/acceptor redox agents in the reaction by the enzyme. In the latter, the respective acceptor/donor pairs for the redox reaction are normally other metal ions binding in a multinuclear metal cluster,[78] reactive oxygen species or electron carriers (such as NAD(P)H,[79] $FADH_2$ or blue-copper proteins[80]).

1.3.1. Considerations on enzyme catalysis

To further discuss the role of metals in enzymes, we first need to approach the models that are currently accepted for enzymatic catalysis. Enzymes are essential biocatalysts that regulate the metabolic and catabolic pathways in cells. One could think of the function of enzymes as two-fold:

(1) Provide an alternative kinetic pathway that can be thermodynamically feasible in regards to the conditions required for a particular individual;
(2) Model a mechanical and an electrostatic environment that can favour the transformation of the substrates required for the functioning of the biological machinery.

Hence, enzymes are both biochemical and biophysical catalysts. Amino acids and cofactors are their chemical catalytic basic units; they are those that participate in acid/base and redox reactions, which provide for unstable intermediates that can be easily reconverted to hold the products required by the remaining pathways of the organism's biology. In the same enzyme environment a network of non-covalent interactions is responsible for properly orienting the catalytic residues and the substrate to provide lower activation barriers. By determining these activation free-energies we are able to predict the rate at which enzymes work. In particular, from the activation free-energy of the rate-limiting step of the catalysed reaction, $\Delta G^{\ddagger}$, we can derive catalytic rates of enzymes acting on several substrates. Such assumptions are provided by the transition state theory developed by Eyring and coworkers, and summarized in Equation 1.3 by the Eyring-Polanyi equation.[81] This equation is derived from the study of a Boltzmann distribution of a system of nuclei (whose forces are derived from electronic quantum mechanical effects) for both the reacting and the activated complex states, found at a temperature T in a quasi-equilibrium state. It presents the catalytic rate, k_{cat}, as a function of a constant κ, the transmission coefficient (often close to unity), and the $\Delta G^{\ddagger}$ for the transformation.

$$k_{cat} = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{k_B T}} \qquad \textbf{Equation 1.3}$$

There are certain premises that should be verified: the transformation of the reactant into the activated complex should occur adiabatically so that the distribution of states for the activated complex can be described by a Boltzmann distribution, the activated complex state should be in quasi-equilibrium so that the rate of transformation of activated complex depends only on the vibration of the reaction coordinate, and the reaction coordinate should be described by a simple translational reaction coordinate.[82] Despite that the transition state theory is still valid for most

enzymatic reactions, in which activation free-energies are often higher than 5 kcal·mol^{-1}, phenomena that occur extremely quick, such as charge transfer or hydrogen tunnelling, are not adequately described, since these are non-equilibrium processes in which quantum effects are prevalent.[83,84]

1.3.2. Metals in enzymatic catalysis

It is now pertinent to question what is the role that the metal cation plays in enzyme:substrate models. Metalloenzymes function as ternary complexes in which the metal cofactor and the substrate follow a sequential binding mechanism. In these complexes the metal ion is generally the first one to bind to amino acids with significant nucleophilic character (for instance Asp, Glu, His, Cys or Tyr) that can be found in more flexible peptide sequences of an enzyme. Additionally, it may also bind weakly to water molecules in order to stabilize the ligand field. This coordination is proposed to allow for the enzyme to adopt a scaffold in which a selective tunnel is formed to regulate the substrate transport to the active site. The tunnel lowers the access of the solvent to the active site and exhibits high specificity towards natural substrates of the enzyme. At this point, the metal-ligand field would lead to a decrease in both translational and rotational entropy in the enzyme's active site, thus enhancing the allocation of the substrate in a proper catalytic pose. This conception of the enzyme:substrate complex is called the entatic state model.[85] In the entatic state, it is assumed that in the enzyme:substrate complex the energy required to drive the kinetics of the substrate-to-product conversion is stored in the several modes that can only be observed in a catalytic conformation of the enzyme. Therefore, during the enzymatic mechanism, major domain motions in the enzyme should not be observed, which would mean that the catalytic site is pre-organized for the transition state to occur. Taking into account these premises, metals generally exhibit one of two functions during the chemical reaction by the enzyme:

(1) They can electronically activate the enzyme through an initial redox reaction that will change some characteristics at the metallocentre (generally its ligand field) inducing changes in the active site that are prior to the proceeding steps of catalysis. This is the case for most copper (involved mostly in oxidation reactions for oxygen activation, oxygen reduction to water and denitrification processes), iron (which can be found in mononuclear and Rieske dioxygenases) and molybdenum enzymes.[33,69]

(2) Alternatively, they can function as electrophiles. In this case, they bind to catalytic residues in the active site, or to the substrate, providing for an environment that favours the chemical reaction by the enzyme. This category includes most of the metalloenzymes that have been characterized to date, in particular those that include metals such as magnesium, calcium, manganese, cobalt and zinc.[69]

Metal cations usually bind to enzymes in a sequential way, preceding the binding of the substrate. Hence, as for most enzymes, catalysis by metalloenzymes usually follows a Michaelis-Menten kinetics. Equation 1.4 depicts the overall catalytic cycle by metalloenzymes.

$$\mathrm{M}^{n+}\mathrm{:Enz} + \mathrm{Sub} \underset{k_{bind}^{inv}}{\overset{k_{bind}}{\rightleftharpoons}} \mathrm{M}^{n+}\mathrm{:Enz:Sub} \xrightarrow{k_{cat}} \mathrm{M}^{n+}\mathrm{:Enz} + \mathrm{Prod} \qquad \textbf{Equation 1.4}$$

The rate at which the product, Prod, is formed depends on the rate of conversion of the enzyme:substrate complex, $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$, and the binding of the substrate, Sub, to the enzyme, $\mathrm{M}^{n+}\mathrm{:Enz}$.

One of the main assumptions in a Michaelis-Menten kinetics is that the formation of the $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$ complex is a fast equilibrium process, which means that the rate at which Sub is converted to Prod (k_{cat}) is much slower than the rate at which Sub dissociates from $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$ (k_{bind}^{inv}). In this case, we can establish that as soon as Sub is converted to Prod, another $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$ is generated to replace the converted substrate, in such a way that one could say that the concentration of $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$, $[\mathrm{M}^{n+}\mathrm{:Enz:Sub}]$, does not vary over time. Moreover, a Michaelis-Menten kinetics relies on the fact that the catalysis by the enzyme is an irreversible phenomenon, which should be true if the media is saturated with substrate or if the free-energy of the process is very negative towards the product. If this latter condition is verified, then the rate of product formation follows a first-order kinetics, where it is only dependent of the formation of the $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$ complex and the catalytic rate of the reaction k_{cat} (see Equation 1.4). Equation 1.5 shows the conservation of enzyme:substrate complex in a steady-state kinetics, as is the case for the Michaelis-Menten model.

$$k_{bind}[\mathrm{M}^{n+}\mathrm{:Enz}][\mathrm{Sub}] - \left(k_{bind}^{inv} + k_{cat}\right)[\mathrm{M}^{n+}\mathrm{:Enz:Sub}] = 0 \qquad \textbf{Equation 1.5}$$

From Equation 1.5, we can derive the dissociation constant for this type of kinetics, K_{diss}, which can be expressed as a function of the velocity constants (k_{bind}, k_{bind}^{inv} and k_{cat}) for the process.

For the particular case in which k_{cat} is significantly lower than k_{bind}^{inv}, this dissociation constant can also be called the Michaelis-Menten dissociation constant, K_M. Despite that both K_{diss} and K_M can be used to quantify the affinity of the enzyme towards the substrate, they are intrinsically different. While K_{diss} is an equilibrium constant with a thermodynamic interpretation, K_M does not require for the system to be in equilibrium. Instead, K_M requires that the substrate is in large excess relatively to the concentration of enzyme, and it can be interpreted as the available concentration of substrate when half of the enzyme's population is occupied by it. Experimentally, it is more frequent to impose the condition in which the substrate is in large excess over the enzyme, and thus, it is more frequent to determine K_M instead of K_{diss}.

Recalling that the amount of enzyme available in solution, $[\mathrm{M}^{n+}\mathrm{:Enz}]$, depends on the amount of the enzyme that is in the $\mathrm{M}^{n+}\mathrm{:Enz:Sub}$ form, $[\mathrm{M}^{n+}\mathrm{:Enz}]$ can be easily determined over time in terms of the initial concentration of enzyme, $[\mathrm{M}^{n+}\mathrm{:Enz}]_0$, and $[\mathrm{M}^{n+}\mathrm{:Enz:Sub}]$. Then, the first-order kinetics for the rate of product formation can be easily re-written in terms of the substrate concentration (see Equation 1.6).

$$v_{steady-state} = k_{cat}\frac{[\mathrm{M}^{n+}\mathrm{:Enz}]_0[\mathrm{Sub}]}{K_M + [\mathrm{Sub}]} \quad , \quad K_M = \frac{k_{bind}^{inv} + k_{cat}}{k_{bind}} \qquad \textbf{Equation 1.6}$$

Equation 1.6 depicts a steady-state kinetics of enzymes, as a function of substrate concentration, where the substrate is in large excess over the enzyme. Subsequently, we can derive the k_{cat} and the K_M for the enzymatic reaction, determining quantitatively the enzyme's affinity, and kinetics. The calculation of the k_{cat} provides insight on the turnover rate of the enzyme (which usually varies from 1 – 10^5 s^{-1}), and it estimates the activation free-energy for the rate-limiting step of the catalysed reaction. The K_M indicates the concentration of substrate at which the enzyme kinetics is at half of its maximum reaction rate (usually lies within 10^{-1} – 10^{-7} mol·dm^{-3}); and together with the k_{cat}, it is a measure of the specificity of the enzyme for a given substrate; it is closely related to the binding equilibrium constant for the enzyme:substrate complex; and it varies with temperature, pH or ionic strength, providing information on the optimal conditions for an enzyme to function.

We have come to realize that metalloenzymes are highly dynamic entities, as a result of their metal coordination sphere, and this observation has provided several important publications over the past few years, by both computational and experimental means.[51,72,86-88]

1.4. Overview of Experimental Approaches to Metalloproteins

We have been discussing metalloproteins in the last two sections: enumerating metal-ligand patterns, speculating the reactivity of metallocentres, addressing thermochemical and kinetic matters of metal-protein binding. However, we have not yet addressed how it became clear that metalloproteins are an essential part of life's biology. Let us then return to some keywords we have not highlighted above.

First, we will address the Protein Data Bank, which is currently the largest repository of structural data for biological units. Metalloprotein structure is an invaluable tool for computational chemists, nowadays. We have already referred how the confidence from the DFT results is dependent of our geometry guess and how several changes in proteins are resultant from metal-guided binding or redox reactions. X-ray crystallography is the most used technique to determine protein structure; however, the X-ray interference patterns of metals result in diffraction maps that are more difficult to solve by current energy minimization algorithms. Hence, the errors in bond and angle prediction in metalloprotein crystals are larger than for proteins with no metal cofactors. Neutron diffraction can also be used to refine the model and map its hydrogen contacts; however the quality of the determination of the structure of the metalloprotein will still be very dependent on the pre-orientation of the crystallographer throughout the process.[70] Alternatively, Nuclear Magnetic Resonance (NMR) can also be employed to provide dynamic insight into the three-dimensional structure of metalloproteins in solution.[89] The main advantage of NMR resides in its dynamic character, which increases the probability to characterize a correct structure for the protein. However, the resolution of the technique is inferior to that of X-ray crystallography. Moreover, the larger the size of the system of interest, the larger the overlapping of chemical shifts that result from the interaction of the nuclear spins with the source magnetic field.[90,91]

We have realized by now that diffraction and NMR methods are insufficient to identify and characterize metal complexes in proteins. Hence, nowadays, these structural determinations are routinely combined with spectroscopic methods that provide access to the geometric and electronic properties of metal complexes in the realm of proteins.[70,92-94] The core of bioinorganic chemistry is made of a large plethora of spectroscopic techniques that have been developed throughout the years, ranging from the study of the ground state to core-electron excitations.[95,96] Some of the most popular are, in increasing order of energy magnitude:

- Electron resonance (EPR, ENDOR and ESEEM) and magnetic Mössbauer, which provide the spectrum for the hyperfine coupling between the unpaired electrons of the transition-metal and the electron delocalization effects when an external magnetic field is applied to raise the anisotropy of the magnetic momentum of the system. From such studies, we can infer the *spin* multiplicity and the covalent character of the metal complex;[96]
- Circular dichroism, which allows the study of ligand field excitations in the sample, from the analysis of the spectrum of absorption from circularly polarized light. Alternatively, electron absorption spectroscopy (ABS) can also allow this study. However, due to the high degree of parity of the orbitals of the ligand field, these transitions are rarely observed when induced solely by an electric field, resulting in very weak signals. Hence, ABS is usually combined with circular dichroism and resonance Raman spectroscopy to study charge transfers between metal and ligand orbitals that do not present parity.[95,96]
- X-ray absorption spectroscopy (XAS, XANES and EXAFS), which can provide the spectrum of X-ray absorption from core electrons of either the metal or the ligand. The main advantages of the technique are related to the fact that the absorption patterns are very specific for a given element in a given coordinating environment, and can, thus, provide an accurate description of the nature and number of donor ligands, as well as metal-ligand bond distances (EXAFS), or the redox state and symmetry of the metal in the complex (XANES). Furthermore, excitations of 1*s*-electrons of the ligand to metal *d*-orbitals (K-edge) or from the (*n*-1)*p*-electrons of the metal to *d*-orbitals (L-edge) are frequently combined to provide insight on the *d*-character of the HOMO orbital and, as a result, the covalent character of metal-ligand binding.[94,96,97] However, two problems can arise: the multiple scattering provided by XAS can difficult the analysis of the output data, particularly when the sample presents elements with similar atomic number; and the employment of X-ray radiation can compromise the integrity of the sample.[98]

At this point, we have already understood that the main advantage of spectroscopic methods is their high selectivity towards the metal complex, in detriment of the protein matrix and the solvent. From the combination of the results from this diversity of techniques, one can derive detailed insight on several metal complex properties, such as: oxidation state, spin multiplicity, ligand field, metal-ligand binding distances and geometry.[70,92,95,96] However, we want to enforce that an accurate study of the complexity of a metalloprotein should favour a combination of atomic (X-ray crystallography, neutron diffraction and NMR methods) and electronic structure methods (EPR, CD, ABS, resonance Raman and XAS, to name a few).

The recent years have witnessed the advent of computational tools in chemical prediction of properties of metalloproteins. If computers were once mere accessories to store, process and fit the data from experimental methods, they are nowadays fundamental tools to improve chemistry through simulation and prediction. In particular, DFT methods, to which we will refer in the next chapter, have been widely used to perform electronic structure calculations, walking alongside experiment in this field.[70,94,96,98]

1.5. Modelling of metalloproteins

We have closed the last section referring to the advent of computational methodologies, walking hand-in-hand with experiment to provide new insight for biological systems, in particular for metalloproteins. Since the past decades, computational biochemistry and bioinformatics have been actively employed in structure and reactivity predictions.[33,53,56-64,99-102] Prior to any computational simulation there is homework to do: characterize the target system to study, and set up a model system. The latter is a key point in computational biochemistry.

In particular for metal-containing systems, studying transition metals complexes introduces several difficulties, either in quantum or classical reality, since their chemistry is far more complex than the chemistry of CNOH-based only macromolecules.[25,27,33] We have already discussed how *d*-shell orbital polarization, *spin* coupling and energy spectra can change from metal to metal and with changing coordinating environments. Despite that several metalloproteins' behaviour has been unravelled through the combination of structural and spectroscopic experimental methods, computational methods have contributed significantly to extend the understanding of several phenomena in this field.[33,51,53,103-105]

Currently, it is only possible to model full metalloproteins at the MM level of theory. DFT calculations can only be feasibly applied to a few hundreds of atoms in extreme situations, and post-HF are not applicable in practical terms. QM/MM methodologies have been employed to circumvent this issue and describe only a small region of the protein; however a compromise between sampling and accuracy has to be achieved, and available timescales are generally bellow the picosecond (ps) region.

1.5.1. Describing a metal complex with classical mechanics

MD simulations at the MM level provide for extensive sampling of the phase space of the protein. However, it is required that the metallocentre is characterized differently from the remaining chemical interactions of the protein. In particular, relevant empirical biomolecular force fields for metallocentres must reproduce the geometrical properties of their coordination sphere, typical bond lengths and atomic charge densities [25,106].

Most common biomolecular force fields employ an all-atom approach where bond and angle interactions are described by harmonic potentials, torsions are described from simple sinusoidal

potentials derived from Fourier expansions, and 1-4 and above interactions are described by Coulomb and Lennard-Jones potentials.[2,107] Applying this line of reasoning to metal complexes, bond stretching and angle bending modes, together with electrostatics, are the terms that need the most accurate description; dihedrals are commonly set to zero and have been shown not to be determinant in the energetics of metal-centred coordination spheres.[19,20,108] Van der Waals interactions, described by the Lennard-Jones potential, are often considered as nearly invariable in similar solvent media, and the parameters required are mostly derived from experimental data of solvation free energies or *ab initio* calculations.[109-111] Other approaches to force field developments have also been introduced[27,104] to implement electronic effects in dispersion and polarization or *spin* derived effects; however, to date, current parameterization schemes are still based in the potential functions from Equation 1.1.[16,18,20,22,24] We will briefly discuss the non-bonded model approach[18] and the cationic dummy atom approach,[17] and we will describe more thoroughly the bonded model approach,[16] which we have employed in our work.

Non-bonded and cationic dummy atom schemes provide a description of metal-ligand interactions based on Coulomb and Lennard-Jones interactions. The main difference resides in the way the charge of the metal is described: in the former the charge is centred in the metal and interacts spherically with the charge from the donor atoms, while in the latter the charge of the metal is described by a set of metal-centred dummy atoms that are geometrically positioned to interact spherically with each donor atom (see Figure 1.3).

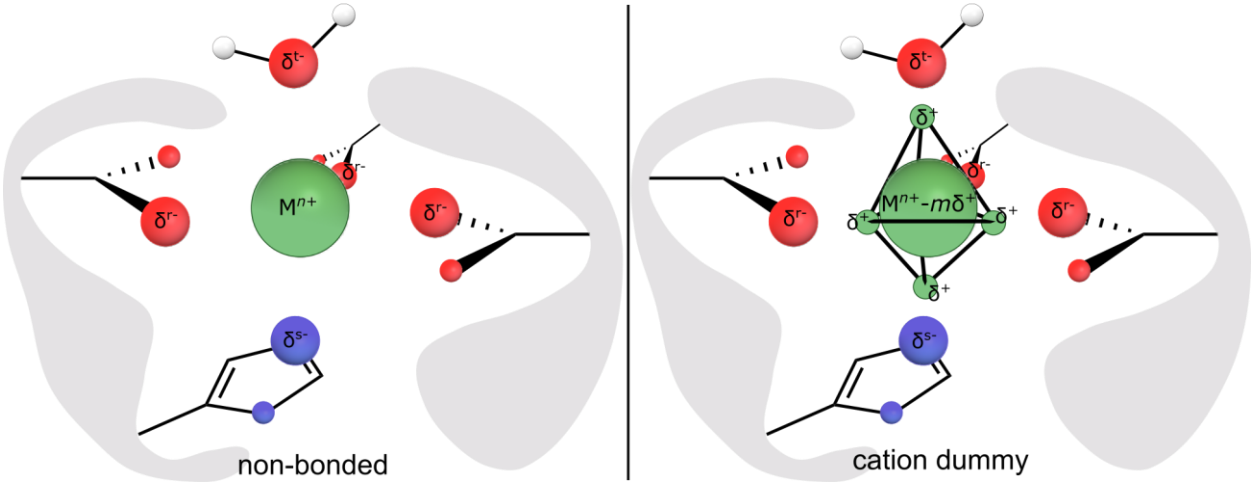


Figure 1.3. Model representation of the non-bonded (on the left) and cation dummy (on the right) approaches to parameterize metal coordination spheres in metalloproteins. The *n*+ stands for the charge of the metal ion, and the *m* specifies the number of dummies that will be placed around the metal cation. The *r*-, *s*- and *t*- stand for the charge of the donor atoms in the ligands of the metal.

These approaches provide the fastest way to parameterize metal complexes and have been remarkably successful in studies with alkali and alkali earth metals or highly negatively charged coordinating environments.[17,112,113] However they are mostly at fault in their capability to describe correct topologies over large timescales, and reproduce *spin*-derived phenomena (for instance pronounced Jahn-Teller effects).[25,106] Cationic dummy atom parameterization schemes keep being developed, and validated from experimental data of coordinating environments, hydration solvation energies and solvent radial distributions of metal complexes;[24,114] however there is no consistent force field for metal complexes derived from this approach.[28]

Bonded model approaches are the more computationally and time demanding parameterizations of metal complexes. In opposition to non-bonded approaches, the philosophy of bonded approaches assumes that, together with coulombic interactions, the classical mechanical terms of the force field potential are important to describe the covalent character of metal-ligand binding.[16] Hence, bond stretching and angle bending modes, and their respective force constants, have to be determined for every metal-ligand combination. These values are mostly derived from QM calculations in simple metal complex models.[16,19,20,22,23] DFT is the preferred level of theory for these systems, since it can be as efficient as MP2, and has been widely validated through benchmarking studies for several transition metals.[34,115,116]. The main advantage of the bonded model resides in its ability to resolve bonds and angles of metal complexes in enzymes with considerable accuracy (with relative errors mostly inferior to 10%), throughout long MD simulations. On the other hand, there is no possibility of ligand-exchange, a phenomena that is commonly observed in coordination spheres exhibiting weak ligand fields.[65,117] Even so, bonded model approaches have shown to sample adequately the conformational space of several metal complexes, allowing the study of interactions between ligands and metals, drug-membrane interactions, while ensuring a reasonable description of the structure of several different biomolecules.[51,118,119] In Table 1.4, we sum up the main strengths and limitation of non-bonded and bonded models.

Table 1.4. Summary of advantages and disadvantages of non-bonded and bonded models for the parameterization of metal complexes

NON-BONDED MODEL	BONDED MODEL
Strengths	***Strengths***
– It only requires Lennard-Jones parameters to be derived; – Allows for fast ligand exchange to occur; – Cationic dummy atom approaches have provided results for several divalent metals, in particular Mg^{2+} and Zn^{2+}.	– It describes quite reasonably the structure of almost any metal complex; – It can reproduce Jahn-Teller effects, and other *spin*-derived phenomena.
Limitations	***Limitations***
– It neglects the covalent character of metal-ligand binding; – It often fails to reproduce metal-ligand effects that are a consequence of ligand field theory, such as Jahn-Teller distortions; – Despite that ligand-exchange is allowed, undesired ligand-exchange may occur frequently; – It presents weaker results when the ligands have no net charge.	– It demands a tedious determination of parameters for bond stretching, angle bending, Coulomb and Lennard-Jones interactions; – Parameters are less transferable among coordination complexes; – The description of coordinating solvent molecules is difficult due to the neglecting of 1-3 non-bonded interactions; – It cannot dynamically reproduce ligand-exchange or coordination shifts.

Despite that it is almost consensual that non-bonded models do not describe accurately the potential of most metal complexes, there is no definite approach to parameterize these systems. In the end, it all depends on the type of MD study conducted and the validation scheme developed. New classical potentials and polarisable force fields could improve the performance of empirical force fields in the study of metalloproteins; however, current developments are still scarcely implemented, and do not allow for simulations of large timescales in such large systems.[120-123] Hence, considerable efforts are constantly being directed to develop parameters that can describe the behaviour of metal-containing biological systems, with current empirical classical force fields.[19,20,22,24,110,124,125]

1.5.2. Parameterization of Intramolecular Mechanical terms

We have already referred that the mechanical parameters to describe bond stretching and angle bending are the most computationally demanding to calculate. These mechanical parameters refer to: bond stretch, angle bend and dihedral torsion modes, as referred in Equation 1.1. We emphasize the two computational methodologies that are employed, to most extent, in the determination of these parameters for metal complexes.

1.5.2.1. First-principles method

A first designed procedure to determine bond, angle and dihedral force constants concerns the direct fitting of the potential postulated for the force field to a PES obtained from the linear transit scan of the bond, angle and torsional modes. The PES study is performed by imposing some constraints in the system, pondered by the user. In a first-principles line of thinking, it is important that each mode should be as independent as possible from the remaining ones. This means that the surroundings of the metal-ligand interaction must be highly constrained so that the potential obtained from the linear transit scan is completely dependent on the atoms directly involved in the mode that is being studied.[16,19,22] In metalloproteins, this procedure can be divided in two stages: the definition of the metallocentre in the protein, and the selection of the modes to be parameterized. After this is achieved, the model system is optimized with a QM potential to characterize a minimum energy configuration that reproduces the same properties that it exhibits in the protein, and a linear transit scan for each of the modes previously defined follows. To assert that the PES obtained for each linear transit scan is exclusively dependent on the atoms directly involved in the scanned coordinate, all the atoms of the ligands that are not involved in this coordinate are constrained; the backbone of the ligands directly involved in the interaction may also be fixed. The PES is then calculated by increasing and decreasing the scanned coordinates to obtain a potential that may *a posteriori* be fitted to the potential of the empirical force field that will be employed after. For most empirical force fields, a harmonic fitting is required to determine force constants and equilibrium positions for internal bonds and angles, while for torsions sinusoidal terms are generally used.

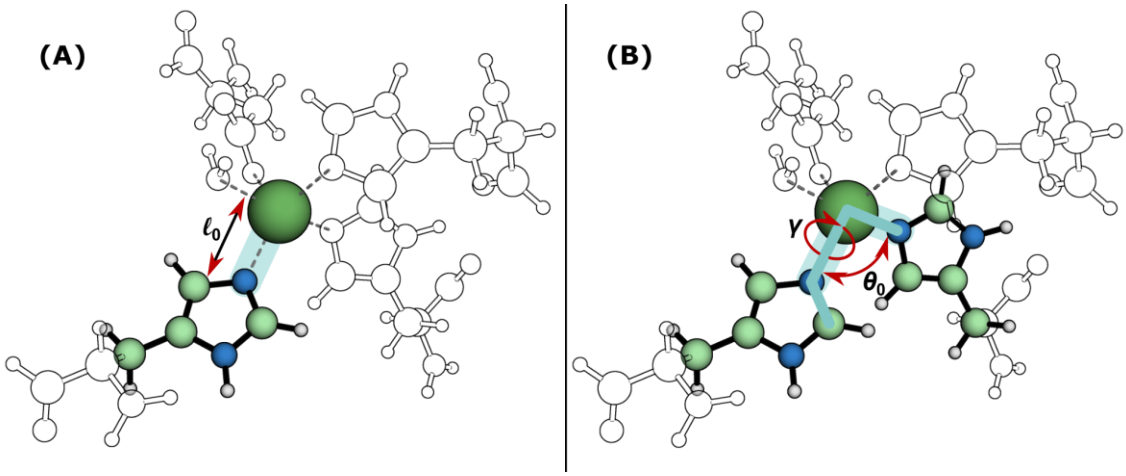


Figure 1.4. Schematic representation of the regions let loose or constrained during: (A) the bond, and (B) angle and torsion linear transit scans. The white-coloured ball and stick representation defines the set of atoms that are kept fixed during the procedure.

Typical restraints refer to the freezing of the atoms non-directly involved in the scanned interaction, or shortening of the models replacing the metal atoms' ligands by smaller units which present the same groups in similar positions[19,51]. Figure 1.4 exemplifies some of the restraints referred.

The main drawbacks of this approach are its highly computational cost and the dependence on the users' considerations on the constraints, which should be imposed to the system. Moreover, it is often required that 1-4 Coulomb and Lennard-Jones interactions are scaled in MD simulations; and this is particularly important when solvent molecules are coordinating to the metal.[22] As a result, a careful validation of the parameters is required.

1.5.2.2. The Seminario Method

The Seminario Method has been proposed in 1996,[126] and is based on the determination of the intramolecular mechanical parameters from the analysis of the Hessian matrix of the metal complex. As for other methods based on the analysis of the vibrational frequencies of the metal complex system, the method assumes that the second derivative of the energy of the system with respect to the equilibrium positions of the N-nuclei in the system returns the harmonic spring constant that describes vibrational and bending modes in near-equilibrium regions. In Seminario's method, the Hessian matrix is determined at the DFT level in a previously optimized geometry of the metal complex and the vibrational frequencies are outputted in Cartesian coordinates, reducing the dependence of the modes on the reference internal coordinates.[28] Bond and angle

force constants can be extracted from the Hessian matrix by solving 3x3 square matrices with the force constants describing the interaction of each pair of atoms involved in stretching and bending modes. As an example, for the bond between i^{th} and j^{th} atoms, with coordinates r_i and r_j, the 3x3 square matrix would be as shown in Equation 1.7.

$$\hat{k}_{ij} = \begin{pmatrix} k_{xx}^{ij} & k_{xy}^{ij} & k_{xz}^{ij} \\ k_{yx}^{ij} & k_{yy}^{ij} & k_{yz}^{ij} \\ k_{zx}^{ij} & k_{zy}^{ij} & k_{zz}^{ij} \end{pmatrix} \qquad \textbf{Equation 1.7}$$

Calculating the *eigenvalues*, λ_n, from $\left|\hat{k}_{ij} - \lambda_n \hat{1}_3\right| = 0$, the *eigenvectors* determined, $|e_n\rangle$, establish an orthonormal space from which the force constant can be determined from the projection of the binding vector $\left|r_{ij}\right\rangle$, as described in Equation 1.8.

$$k_l^{ij} = \frac{1}{r_{ij}} \sum_{n=1}^{3} \lambda_n \left\langle r_{ij}{}^{\dagger} \middle| e_n \right\rangle \qquad \textbf{Equation 1.8}$$

With this formalism one is able to determine all bond constants that concern intramolecular interactions. Additionally, these bond force constants can also be geometrically combined to determine angle, dihedral and crossed-term interaction force constants transferable for force fields in which these terms are described by harmonic potentials.[126]

The Seminario Method has been extensively used in standard parameterization of metallocentres.[20,23,52] However, despite that its derivation is physically reasonable and that it provides for a quicker way to derive force constants for bending and stretching modes, there is a possibility that, due to the few constraints that are imposed to the system, the modes evaluated might account for some non-mechanical contributions and cross-term interactions.[106] In theory that could lead to overestimation of the systems' energy by overcompensation of the bond and angle energy contributions.

1.5.3. Parameterization of Electrostatic Atomic Charges

The energy resulting from the electrostatic interaction of atomic charges constitutes the greater fraction in the non-bonding interaction energy. However, it is difficult to establish an accurate methodology to determine these charges, since atomic charge is not an observable of a system and cannot be experimentally derived. Furthermore, atomic charges are usually very

conformational dependent.[127] As a result, the usual procedure to parameterize atomic charges is based on the analysis of electron density populations from QM calculations on a single or several representative structures of the model. This procedure can also be supported by data from several related observable quantities, such as dipole moments or free energies, to further enhance the validity of the method.

According to Maciel *et al*[128] a general atomic charge parameterization scheme is expected to be independent of the level of theory, topology or orientation of the system, reproduce several observable quantities concerning electrostatics, and be transferable system-wise. A current issue is that there is no atomic charge scheme that fulfils all these requisites, and current theoretical approaches are not consensual about what quantities should be reproduced by any parameterization scheme. The known schemes so far approach molecular orbital, topological or surface potential considerations. The latter is the most addressed scheme by the computational chemistry community, since they accurately represent, to a considerable extent, the surface potential and most dipole moments of the system[19,20,129,130]; however often they show considerable dependency upon structure considerations. Among the most commonly known surface potential fitting schemes are: Merz-Kollman (MK)[131], CHELP[132], CHELPG[133] and restrained electrostatic surface potential (RESP)[134] population analysis. The first three differ mainly in the criteria to select the points in the molecular space which will reproduce the Potential Surface of the molecule[135]. The RESP scheme, on the other hand, starts from the result of the fitting of any ESP scheme and applies a hyperbolic penalty function that introduces the criteria of charge equivalence for atoms that present similar chemistry or symmetry. To ensure that the original ESP is still reproduced the penalty function is iteratively solved through minimization of the least squares matrix[135] A drawback of these methods is the difficulty to reproduce the charge of buried atoms, mainly saturated carbons.[127,128,135]. Moreover, two questions that immediately come to mind are *'how to weight each point of the potential surface'* and *'how to define which electron densities should be considered in the calculation of the atomic charge'*. For the methods referred, the common approach is to consider the density that is farther from the nuclei, preferably at a distance larger than the van der Waals radii of the atom, in order to significantly account for the atoms interacting with it; however distances that are too far from the nuclei might not be significant and overestimate the atomic charge. In literature, a more detailed discussion of the considerations of the MK, CHELP and CHELPG methods can be found.[128,135]

Alternatively, to avoid the computational cost of *ab initio* calculations on single structure surface potentials, semi-empirical schemes have been developed to reproduce atomic charges derived

from *ab initio* methods. The most employed is the AM1-BCC method, proposed by Jakalian *et al*.[127,136] In this method, the atomic charges are primarily derived from the AM1 semi-empirical wave function;[137] then a bond charge correction (BCC) is established for a set of user-defined atom and bond types, in order to reproduce ESP-charges derived from *ab initio* methods. The validation of this method has shown a good agreement with ESP and RESP methods; however it has only been conducted in organic molecules. That said, the RESP methodology is, to date, the one that is customarily taken as the reference method to derive atomic charges in the monopole treatment. For metal complexes, the RESP methodology is usually applied to atomic charges previously derived from an MK population, calculated at the DFT level of theory with large Gaussian-basis sets.[19,20,52,138] Other methods, such as the Mulliken population analysis[139] or the atoms in molecules theory (AIM),[140] provide a more analytical procedure to determine atomic charges. However, these methods are highly dependent on the level of theory, and inaccurately describe the surface potential and dipole moments of several system, unless higher electric pole expansion are performed.[135] Hence, they are not used as common practice in the parameterization of atomic charges for empirical force fields.

Nevertheless, we have already emphasized that this is not a consensual field. The choice of method is pending on the observables under study and the employed force fields' parameterization scheme. However, there is still a major drawback that refers to force fields themselves, and that is the static character of atomic charges throughout any MD simulation.

1.5.4. Parameterizing the Lennard-Jones term

Parameterization schemes on van der Waals parameters are remarked as the most difficult to obtain, either due to the computational cost of their calculation or to the lack of parameters and experimental data.[109] In general, van der Waals parameters are derived from combinations of theoretical calculations and thermodynamic quantities, such as sublimation/vaporization enthalpies and densities, or from crystallographic data and self-diffusion coefficients. In particular, the properties in the gas phase are not very useful since these interactions are largely reduced by the average long distances between atoms.[109] However, the parameterization scheme is mostly dependent on the empirical force field employed. The parameters for the AMBER and OPLS force fields are derived from fittings to quantities from liquid simulations, CHARMM employs the fitting to sublimation or vaporization enthalpies or densities, while CVFF exclusively uses crystallographic data and sublimation enthalpies as the fitting quantities.[109]

Even though van der Waals interactions quickly vanish to zero, short range dispersion interactions can be very important, particularly if there is significant interaction between pairs of atoms. In fact, in current biomolecular force-fields, a large part of the potential energy of the system is stored in these interactions. Hence, the parameterization of these interactions is very relevant for biomolecular force-fields. Frequently, QM calculations are employed to accurately determine the attractive and repulsive parameters of the Lennard-Jones potential. However, high levels of theory and large *basis sets* are required to accurately describe all the electronic interactions between each pair of atoms. Then, the parameterization can be carried out by calculating interaction energies for an optimized model and average them over several rotational positions,[109] or by fitting the classical model with the optimized structure and dimerization energies from QM calculations.[110,111] Monte Carlo (MC) and MD fittings to experimental data can also be used; however they require that some intramolecular interactions should be described accurately, as is the case of the torsion modes.[110] Hence, MC and MD methods are mainly used as auxiliary techniques in free-energy determinations by computational means. In particular for metals, thermodynamic integration (TI) has been largely used to fit hydration free-energies of several main group and transition metals.[110,111,124,125]

1.5.5. Quantum Mechanics study of the Catalysis by Metalloenzymes

When we refer to the study of the catalytic effects of metalloenzymes, these can only be approached through QM calculations. MM methods rely only on nuclear motions, which are the most important aspect in most biophysical processes, but do not include electronic transformations that are a crucial aspect of chemical transformations. The problem is, and we have addressed that before, that most enzymes cannot be fully modelled at a QM level. QM calculations in metalloenzymes are still limited to a few hundreds of atoms, generally no more than 300 atoms.[141,142]

Figure 1.5 sums up the two most common approaches to tackle the catalysis of metalloenzymes: (1) the QM cluster approach, and (2) the hybrid QM/MM approach. While in the former the enzyme will be fully described by a cluster of atoms that may be surrounded by vacuum or an implicit solvent; the latter approach considers an important QM region that encloses the chemistry of the reaction, and an MM environment that embeds the QM region providing for either mechanical or electrostatic constraints that also take part in the optimization of the QM region.[100,143]

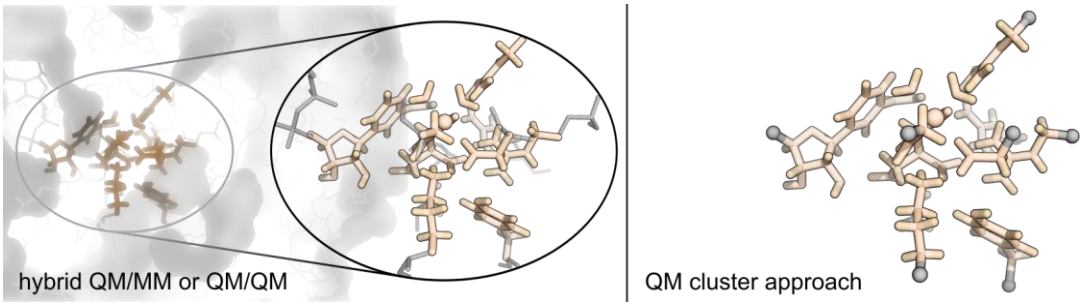


Figure 1.5. Representation of two types of modelling for large biomolecules: a hybrid QM/MM or QM/QM approach (on the left); a QM cluster approach (on the right). The atoms with contours are treated with QM methods. In the QM cluster model approach, the grey coloured spherical atoms concern hydrogen link atoms added to fill the valence of the cleaved model, also described at a QM level.

The starting point of a QM or QM/MM model is usually an X-ray model of the metalloenzyme we wish to study. However, since most of the catalytic conformations of enzymes are obtained from mutagenesis studies or complexation with inhibitors or substrate analogues, standard modelling is often required to design an initial catalytic state for the metalloenzyme. Moreover, we should also account for different protonation states in the amino acid chain, especially when regarding possible catalytic ones. The modelled structure is generally refined by employing simple MM minimizations or running constrained MD simulations. In extreme situations, the enzyme's structure can lack its substrate or cofactors; and as result, docking methodologies can complement the modelling process for the reactive enzyme. Additionally, homology modelling may be employed to refine incomplete enzyme structures. These extreme situations must be carefully validated and not used blindly.

1.5.5.1. Building an enzyme model

The proceeding step is the choice of an adequate set of atoms that can adequately represent the reaction that will occur in the metalloenzyme. The rule of thumb is that any model must ensure that not only the reactive substrate and catalytic residues are represented, but also that other intermolecular interactions are described. Such interactions account for: hydrogen bonds, π-interactions, and the most important ionic interactions. In metalloenzymes, it is important to include, at least, a representative region of the metal's first coordination shell; and often the most acidic residues of the second coordination shell.[32,33] An additional problem, particularly in metalloenzymes complexed with transition metals, is the *spin* multiplicity of the model. To decide on the *spin* multiplicity of the system, a common procedure is to combine QM energy and structure

calculations with experimental data (such as the discussed in Section 1.4). Particularly in QM/MM calculations, the evaluation of the energy of different optimized models may be misleading, since at every step of the process we must evaluate the energy changes in the different layers of the model. Hence, experimental data will further support the observed geometry optimizations and energy minimizations from QM and QM/MM calculations, thus providing a more robust modelling of the system.

Optimizing the QM model is the step that follows. An initial optimization of the system is quite straightforward. However, to optimize a model that can be representative of the chemistry by the metalloenzyme can often require more extensive QM calculations, since these are very sensitive to the set of internal coordinates provided to optimize the model system. Consequently, it is sometimes advisable that different guess structures are generated from MM minimizations or MD simulations. These structures are usually similar, if our system is adequately parameterized, and can provide slightly different conformations for the model system. Furthermore, we can start by performing optimization calculations with softer convergence criteria (either for the *self-consistent-field* equations and the structure convergence parameters), different levels theory, and then proceed to more systematic approaches.

1.5.5.2. Establishing the catalytic mechanism

We have already highlighted that metalloenzymes can contribute to catalysis in different ways: providing structural support for the active site of the enzyme, or intervening in redox reactions that assist enzyme catalysis. Here, we will address to the metal site throughout bond formation and cleavage.

As for any enzyme, tackling the catalysis by metalloenzymes requires an initial set of possible mechanistic guesses. This set is generally derived from experimental and computational studies, and from chemical intuition.[58,100] Specifically, our attention is generally drawn towards strong electrophiles, nucleophiles and hydrogen bonding residues. Additionally, the presence of conserved waters should also be carefully evaluated. This subject is quite sensitive since not all crystallized waters are either catalytic or conserved waters, and Michaelis-Menten complexes are not generally obtained from the natural intervenient of the reaction.

After achieving a stable initial guess model, we are in a position to start exploring different hypothesis for our reaction to take place. At this stage, linear transit scans are usually performed to explore the PES of the coordinates directly involved in our mechanistic guess. Since structural

changes are less sensitive to the choice of basis set, the 6-31G(*d*) *basis set* is usually preferred.[115,116,144] A proper linear transit scan for a reaction coordinate will exhibit a maximum peak of energy, among all optimized internal coordinates. This maximum is expected to be close to the transition state (TS) for the transformation. The optimization of the determined TS is conducted by a search of the saddle point in the PES, which corresponds to a minimum energy configuration in every direction of the PES except that of the reaction coordinate. In practice, this is achieved by determining the nuclear normal modes from the Hessian matrix of the nuclear displacements, which should exhibit only one imaginary frequency.[102]

At each step of the reaction, all states involved should be freely optimized. Alternatively, intrinsic reaction coordinates scans (IRC) can be conducted to obtain the reagents and products from a previously optimized TS. The final step of the study of an enzyme catalysis study comprises the thermodynamic characterization of the stationary states of the chemical space. In particular, we are interested in generating free-energy profiles, which are the ones more directly comparable with experimental data. A standard procedure is to calculate the electronic energy for the reaction and the free-energy corrections to the energy for a given pressure and temperature. The latter contributions are generally determined from single-point frequency calculations for all stages of the reaction, at the same level of theory as that of the optimization calculations. On the other hand, the electronic energy is usually determined with a larger *basis set* to better account for the electron-electron interactions at each stage.[51,88,145,146]

Until this point, we have provided a general, yet summarized, overview of the role of metals in bioinorganic chemistry, and the challenges that computational biochemistry faces to further extend the knowledge in this category of systems. Nevertheless, we have not yet discussed the pillars that support the investigation in current computational biochemistry. We will provide further discussion in the proceeding chapter.

Chapter 2: Strategies to tackle metalloproteins

Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would occasion a rapid and widespread degeneration of that science.

Auguste Compte, in Cours de Philosophie Positive (1830)

Theoretical chemistry is an emerging field from the 20th century whose main interest is the study of chemical systems at an atomic scale. The field is built on the premise that every phenomenon in Nature is ruled by universal laws that can be mathematically formalized. As a result, we should be able to draw structural, thermodynamic and kinetic insight from simplified models of chemical systems under the action of physically motivated potentials. However, the complexity of the nature and the potentials that define chemical systems at an atomic level is still an obstacle for current computational methods to overcome. Despite of the ever-increasing technological development of computers (their processing power, memory allocation or disk storage), and of the increasing efficiency of current computational algorithms, the computational chemist is always faced with a compromise between the level of detail, the size-scale of the system and the timescale that is required to study a given phenomenon. Computational methods are currently assembled in two main categories: the molecular mechanics (MM) and the quantum mechanics (QM) levels of theory. In the past decades a third branch has been increasingly being employed to combine both the benefits of the former two: the hybrid quantum mechanics/molecular mechanics (QM/MM) level of theory. The main challenge for theoretical and computational chemists remains: it is necessary that the level of chemical detail of the system is improved, and that new and more efficient algorithms and potentials are built, enhancing the possibilities of computer simulations without compromise of the quality of the description of the system.

2.1. Molecular Dynamics

2.1.1. Expose the problem

One of the greatest challenges to current computational simulations is to study biochemical systems on a timescale that is within that of the processes at the atomic scale (fs to μs). However, time-dependent simulations with quantum mechanics calculations are still extremely costly and, in practice, they are still only possible for small systems and short timescales (a few hundreds of picoseconds). As a response to this, semi-empirical potentials have been developed from profound approximations and extensive parameterization, which have shortened the number of one-electron molecular orbitals and two-electron integrals to solve computationally.[147] However, despite that these methods have allowed larger systems to be studied with QM calculations, not only is their performance irregular, due to the extensive parameterization of overlap coefficients and unaccounted core-electron and two-electron repulsion integrals, as time-dependent simulations are still far from the ns-timescale.[57,59,61] Hence, empirical force fields are the alternative that is currently employed, when referring to the study of chemical systems throughout the ns- to ms- timescale. Despite that intrinsic electronic effects are neglected with this methodology, a careful combination of extensive sampling and statistical data treatment can provide invaluable thermodynamic and structural insight on biophysical and biochemical phenomena.[148-152]

2.1.2. Empirical Classical Force Fields

Empirical force fields are defined over simple equations based on a classical mechanical description of molecules as an aggregate of spherical particles (with a mass m_i, electrostatic charge q_i, and a van der Waals radii R_i and interaction energy ε_i) that are coupled by harmonic springs. The potential that drives the interactions between particles is parcelled into simple terms that describe the internal modes of the system in equilibrium from simple harmonic/sinusoidal potentials, making use of classic electrostatics to describe charge density interactions (refer to Equation 1.1 to exemplify). Since this potential energy function is highly parameterized, it is extremely fast to solve for systems that can range up to millions of atoms. Moreover, this function is differentiable, and can therefore be solved to derive forces that allow for time-dependent trajectories to be determined.

The most generic biomolecular empirical force fields describe the potential energy function as a sum of six parcels, as stated in Equation 2.1.

$$U(\mathbf{r}) = E_{bonds} + E_{angles} + E_{dihedrals} + E_{impropers} + E_{electrostatic} + E_{van\ der\ Waals} \quad \textbf{Equation 2.1}$$

Particularly, for the AMBER Force Field, which we have employed in the work we present, the potential energy function employed is that of Equation 1.1. We further discuss each of the terms of the potential energy function.

2.1.2.1. 1–2 Bond stretch

The bond stretching mode is a mechanical mode that describe the type of bond between the atoms in the molecule (single, double, triple or aromatic). For simplicity, this interaction is generally approximated by a harmonic potential described by two parameters: the equilibrium bond length l_0 and the force constant K_l that describes the harmonic regime near the equilibrium position l_0, as can be seen in Equation 2.2.

$$E_{1-2} = \sum_{bonds} K_l(l - l_0)^2 \quad \textbf{Equation 2.2}$$

This potential is quite easy to parameterize and is fastly computed to allow for many covalent bonds in the system. However, its validity is obviously limited to small stretching movements, which are circumscribed to systems in equilibrium. In this regime, the atoms vibrate around an equilibrium distance that corresponds to the minimum of potential energy of interaction. This would not be equally accurate if we wanted to study a reactive force field, where a chemical bond would be better described by a Morse potential (see Equation 2.3). Comparatively to Equation 2.2, the computation of this potential would require one additional parameter: the bond dissociation energy E_{diss}.

$$E_{bond} = E_{l_0}\left[1 - e^{-\sqrt{\frac{K_l}{E_{diss}}}(l-l_0)}\right]^2 \quad \textbf{Equation 2.3}$$

However, in this way the interaction between the bonded atoms would vanish as the atoms went farther away from the equilibrium position. In practical terms, this potential compromises computational efficiency since the mathematical form of this term is more complex. Moreover, accurate determinations of E_{diss} should be much more difficult, since the fitting of the potential is

much more demanding. One possible approximation would be to include additional terms from the Taylor's expansion (up to the 4th power) to the harmonic potential, which would mimick reasonably well the Morse potential;[25] however the harmonic description of bond stretching is quite satisfying for a relative stretching of up to 10% of the equilibrium bond length and is computationally cheaper than any of the above proposals.[153]

2.1.2.2. 1–3 Bond angle bend

As for the bond stretching mode, angle bending is also usually described with the harmonic approximation (Equation 2.4), and it has been employed with relative success so far.[153] In some cases, a sinusoidal form is preferred to better describe the change in the energy for very large amplitudes, since it describes the potential energy in a smoother way for bends in the lower and upper limits of angle bending.[25] However, despite that it works quite well for non-equilibrium linear interactions, where the restoring force is more smooth, it might also result in more exotic results for other types of angles. Other force fields have also introduced harmonic potentials to describe 1–3 bond stretching,[154] to provide an additional restraint to angle bending.

$$E_{1-3} = \sum_{angles} K_\theta (\theta - \theta_0)^2 \qquad \textbf{Equation 2.4}$$

Once more, the introduction of higher order potentials from the Taylor's expansion on the equilibrium bond angle can enhance the description of the bond angle bending; however, taking as a reference the compromise between calculation efficiency and accuracy, the employment of the harmonic potential is still regarded as the best answer.

2.1.2.3. 1–4 Dihedral and Improper torsions

Dihedral interactions are among those that are the most demanding to parameterize, since even for a simple system there is a large number of possible dihedral interactions. These modes are described for groups of four atoms linked by harmonic springs, and are classically treated as rigid rotors in which the 1st and 4th atoms of the dihedral rotate around the axis of the bond formed by the 2nd and 3rd atoms. These transitions present the lower energy barriers and can, therefore, exhibit periodic behaviours every time a complete rotation occurs. Hence, it is a requisite that the potential chosen to describe dihedral modes presents a periodic behaviour that can characterize the different minima during a complete torsion. The most accurate way to do this is to consider a Fourier expansion of a sinusoidal potential, where each minima and maxima are accurately

defined.[153] However, due to matters of computational efficiency and to the low energy values from these transitions, only the first term of the Fourier's expansion is considered (see Equation 2.5). This sinusoidal term requires the determination of three parameters: the maximum of potential energy that describes the torsion K_ρ, the number of minima provided by a complete torsion n, and the phase of lowest potential energy γ. These parameters are usually determined by a least squares fitting of QM profiles with MM data.[106]

$$E_{1-4} = \sum_{dihedrals} K_\rho[1 + cos(n\rho - \gamma)] + \sum_{impropers} K_\varphi(\varphi - \varphi_0)^2 \qquad \textbf{Equation 2.5}$$

Still addressing 1–4 interactions, improper potentials were also derived to represent the strain that results from the π-molecular orbitals that are formed from the binding of sp^2-hybridized atoms. These atoms are spatially arranged in a planar conformation that cannot be described by any of the potentials we have described so far. As for the bond angle bending potential, it is generally described by a harmonic or sinusoidal potential. However, the coordinate that is approached is the angle that is formed by two non-concurrent vectors in the rigid plane.[107,153]

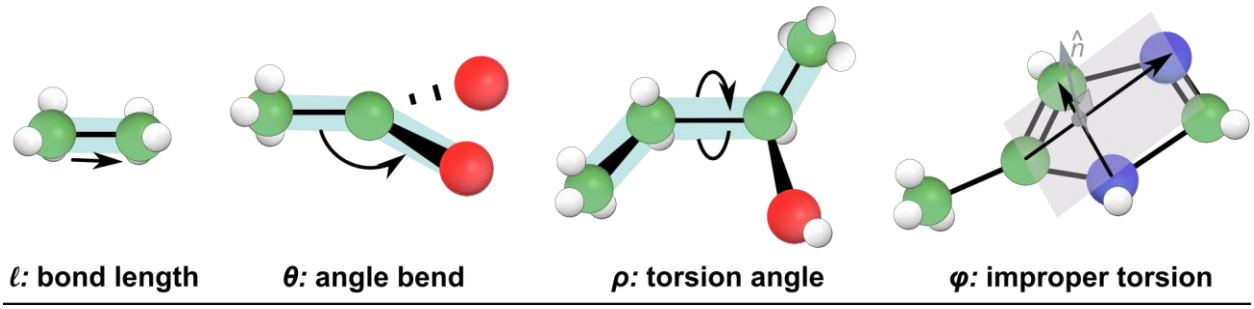


Figure 2.1. Normal modes employed in common implementations of empirical biomolecular force fields.

In Figure 2.1, we systematize the normal mode coordinates that are employed in current empirical biomolecular force fields. More refined force fields introduce potentials concerning crossed interactions or higher order expansions; however most of these corrections stand for the refinement of particular quantities, in particular the simulation of spectra.[107,153] Other than that, such refinements are often accompanied by an increase in computational cost which does not often suit the purpose of the MD simulations, which we will discuss below.

2.1.2.4. Charge and dispersion interactions

We now discuss the non-bonded interactions that are employed in biomolecular force fields. In particular we will address the role of electrostatics and long range interactions. These interactions are only accounted for 1–4 interactions and beyond, to prevent numerical instabilities that could result from the short distance interactions between 1–2 and 1–3 bonded atoms.[107,153] Electrostatic interactions are commonly calculated by the Coulomb's potential, while long range interactions are preferentially calculated from the 12-6 Lennard-Jones potential. A close inspection of Equation 2.6, shows that, provided the relative positions of the atoms i and j, we only need the charge q_i and q_j to describe the Coulomb's interaction between a given pair of atoms, and the ε_{ij} and R_{ij} interaction parameters to describe the Lennard-Jones potential.

$$E_{non-bonded} = \sum_{j>i}\sum_{i}\left(\frac{q_i q_j}{\epsilon r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right]\right) \qquad \textbf{Equation 2.6}$$

The parameters of the Coulomb's potential are a direct property of each spherical atom, but the interaction terms ε_{ij} and R_{ij} have to be calculated from the ε_i, ε_j, R_i, R_j atomic parameters, from the Lorentz-Berthelodt combination rules.[107] From all these considerations, it follows easily that non-bonded interactions involve the most pairs of atoms, and that they are the key limiting-step to solve the potential energy function at the MM level of theory.

An additional difficulty that derives from the choice of these potentials is that both the Coulomb's and the Lennard-Jones' potential are slowly convergent sums. Such requires that a truncation criterion should be imposed so that a finite contribution from the explicit counting of these interactions can be calculated. In practical terms, periodic boundary conditions (PBC) are combined with a *cutoff* that is generally inferior to half of the largest length of the system. From this procedure, several replicas of a model unit cell are created, and translated in every direction from its centre of mass, providing a list of atoms that will interact explicitly with each other. This results in solving the problem of how to study the boundaries of the cell as well as how to select the atoms that should interact explicitly with each other. However, if for van der Waals interactions this approximation has shown that atoms beyond a *cutoff* of about 8 Å could be treated as an isotropic distribution of particles without major loss in accuracy, this type of truncation has shown to result in several artefacts when applied to the treatment of electrostatic interactions.[107,153]

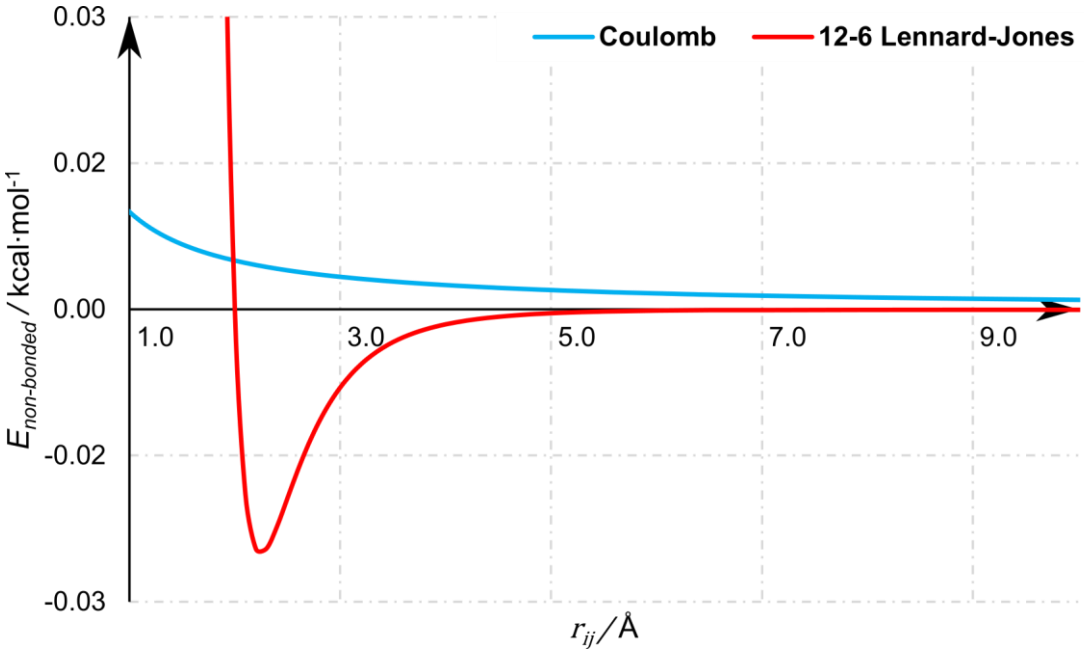


Figure 2.2. Comparison between the Coulomb and 12-6 Lennard-Jones potentials for a pair of atoms i and j, in kcal·mol^{-1}. The reference parameters adopted were 0.10 au for q_i and q_j, 0.025 kcal·mol^{-1} for ε_i and ε_j, and 2 Å for R_i and R_j.

This is mostly due to the lower rate at which the intensity of the coulombic interactions vanishes comparatively to Lennard-Jones interactions, as seen in Figure 2.2. In current biomolecular force fields, these issues are overcome by the inclusion of Ewald sums for the coulombic contributions.[155] In the method of Ewald summation, atom-centred Gaussian distributions of opposite sign are added to the atoms within the *cutoff* distance so that the coulombic interaction becomes negligible at that distance (see $E_{Coulomb,real}$ at Equation 2.7).

$$E_{Coulomb,real} = \sum_{j>i}\sum_{i}\frac{q_i q_j}{r_{ij}}\left(1-\frac{2}{\sqrt{\pi}}\int_{r_{ij}}^{\infty} e^{-\alpha t^2}dt\right)$$

$$E_{Coulomb,Fourier} = \frac{1}{2V_{unit\ cell}}\sum_{\mathrm{k}\neq 0}\frac{4\pi}{\mathrm{k}^2}\left(\sum_{i} q_i e^{i\mathrm{k}r_i}\right)e^{-\frac{\mathrm{k}^2}{4\alpha}}$$

Equation 2.7

On a second step, since the former Gaussian distributions were artefacts that were used to force the electrostatic potential to converge at the truncation distance, the same number of Gaussian distributions with the same sign as q_i are added in the reciprocal space of $\{\vec{r}\}$ so that the overall charge of the system is balanced (see $E_{Coulomb,Fourier}$ at Equation 2.7). With these operations,

and through a combination of the right parameter α and wave number k, the accuracy of the calculation of coulombic interactions can be calibrated to exhibit very small errors. The main problem with this approach is that the Ewald summation at the reciprocal space considers the system as a periodic one; however through the employment of PBC, we are able to simulate such systems.[153]

2.1.3. Molecular minimization in Empirical Force Fields

The first stage of any MD simulation is the refinement of an initial model. To do this, the common procedure is to minimize the potential energy of the system using the empirical potential function that we have systematized above. However, there are two aspects that we must address to perform this energy minimization: to start with, our system is a result of complex modelling, from X-ray refinement to algorithmic addition of hydrogens, solvent and counterions, or even from human-inspection; on second hand our problem is a multidimensional problem, since at the end of the energy minimization we need to have found a full set of Cartesian coordinates that describes the minimum energy configuration of the model, at this level of theory.

As a response to the first issue we have raised, the common procedure is to perform constrained energy minimizations to gradually relax the model system. In particular for protein-solvent systems, the most common procedure comprehends a minimization of the explicit solvent, followed by a minimization of the hydrogens placed by a computational algorithm, and finally a full minimization of the full model. To address our second issue, two minimization algorithms are more commonly employed: the steepest descent[156] and the conjugate gradients.[157]

To further discuss the strategy to minimize the energy of our model, we have to clearly define our problem, which is: we need to find a set of coordinates $\mathbf{r_{min}}$ that minimizes the energy of our system U_{min}. For any of these algorithms, the goal is to minimize a multidimensional linear equation, which in our case is Equation 2.8.

$$U(\mathbf{r_{min}}) = U_{min} \qquad \textbf{Equation 2.8}$$

Neither $\mathbf{r_{min}}$ or U_{min} are known; hence our solution will have to be provided by iteratively generating configurations $\mathbf{r_i}$, until a minimum energy configuration is achieved.

To start off, we provide a configuration $\mathbf{r_0}$ that has a potential U_0. Then, to evaluate the nature of $\mathbf{r_0}$, a new configuration $\mathbf{r_1}$, with potential U_1 is extrapolated. From the difference between U_1 and

U_0, it is possible for us to determine a search vector $\mathbf{\nabla} U_0$, which is no other than the gradient of the potential energy along the $\mathbf{r_1}$ to $\mathbf{r_0}$ direction. At this point, it is clear that we need to iteratively generate new search vectors $\mathbf{\nabla} U_i$, until the difference between U_i and U_{i+1} is very close to zero. For both the steepest descent and the conjugate gradients method the $(i+1)^{\text{th}}$ iteration is generated from the i^{th} iteration by a linear addition of a weighted contribution from $\mathbf{\nabla} U_i$, as stated in Equation 2.9.[158]

$$\mathbf{r_{i+1}} = \mathbf{r_i} + g_i \mathbf{\nabla} U_i \qquad \textbf{Equation 2.9}$$

The generated $(i+1)^{\text{th}}$ search vector is orthogonal to that of the i^{th} iteration, which is somehow intuitive since in the current direction of the search vector we are already at a minimum energy, which is not that of the minimum energy configuration that we are looking for, $\mathbf{r_{min}}$. When the search vector is sufficiently close to zero in all directions, we assume that our final set of coordinates is sufficiently close to $\mathbf{r_{min}}$, and the system is then optimized.

The main difference between the steepest descent and conjugate gradients methods is the way the search vector is built. While for the steepest descent the only requirement of $\mathbf{\nabla} U_{i+1}$ is the orthogonality to $\mathbf{\nabla} U_i$, in conjugate gradients $\mathbf{\nabla} U_{i+1}$ has a memory of the directions that have been searched and it is required for it to be orthogonal to all of the previous $\mathbf{\nabla} U_i$. This occurs because the operation that we perform to generate $\mathbf{\nabla} U_{i+1}$ from $\mathbf{\nabla} U_i$ is repetitive, which means that starting from $\mathbf{\nabla} U_0$, $\mathbf{\nabla} U_1$ is generated by $g\mathbf{\nabla} U_0$, $\mathbf{\nabla} U_2$ will be $g\mathbf{\nabla} U_1$, and so on, being g the weight of $\mathbf{\nabla} U_i$ for the configuration $\mathbf{r_{i+1}}$, as presented in Equation 2.9. The set of solutions of the search vector is of the form $G_i = \{\mathbf{\nabla} U_0, g^1 \mathbf{\nabla} U_0, g^2 \mathbf{\nabla} U_0, \dots, g^i \mathbf{\nabla} U_0\}$, and if we require $\mathbf{\nabla} U_{i+1}$ to be a linear combination of the search vectors of G_i, as well as orthogonal to $\mathbf{\nabla} U_i$, then we can assure that no direction is repeated for this vector.[158] The main advantage of this procedure it that it prevents that the search vector starts to point back and forth, which may occur when we get nearer the minimum energy configuration. However, the convergence is slower than for steepest descent when the generated configurations are farther from the minimum energy configuration. Moreover, these algorithms take the second derivative to be constant for each iteration, which, despite being a good approximation, implies that numerical instabilities and *roundoff* errors may propagate in the calculation of the energy gradient as the number of iterations grows.[159]

2.1.4. Molecular Dynamics simulations by Empirical Force Fields

Despite that we were able to find our minimum energy configuration from MM considerations, this configuration is only one among a large number of configurations that the system can adopt. The minimum energy of the system comprehends an ensemble of configurations for a given set of thermodynamic variables, such as temperature (*T*), pressure (*P*) or composition (μ). Moreover, any system presents a remnant energy that is accumulated in the vibronic states that characterize the ground state of the system. As a result, to compute macroscopic observables we are required to statistically treat this ensemble of configurations. MD is a way to explore this phase space by solving the classical motion equations that can be derived from the potential energy function of the system and the Newton's law of motion.

$$-\frac{\boldsymbol{\nabla} U(\mathbf{r})}{m} = \frac{d^2}{dt^2}\mathbf{r}(t) \qquad \textbf{Equation 2.10}$$

Equation 2.10 is a second order differential equation which describes how the acceleration of an atom, with mass m_i and in a configuration $\mathbf{r}$, can be determined from the force $-\boldsymbol{\nabla} U(\mathbf{r})$ that results of the potential U_i that acts on the atom. If we were able to solve this equation analytically, we could calculate the trajectory of the atom for a given timescale, Δt. However, this differential equation is a system of coupled polyatomic equations that can only solved numerically.[153,160]

2.1.4.1. Trajectory propagator for MD simulations

To solve our problem in a much simpler way, we can consider that starting from the position $\mathbf{r_{i,t_i}}$, the position of the atom at $t_i + \Delta t$ can be approximated by a Taylor's expansion, as in Equation 2.11.

$$\mathbf{r}(t_i + \Delta t) = \mathbf{r}(t_i) + \frac{1}{1!}\frac{d}{dt}\mathbf{r}(t)\Big|_{t=t_i} \Delta t^1 + \frac{1}{2!}\frac{d^2}{dt^2}\mathbf{r}(t)\Big|_{t=t_i} \Delta t^2 + \mathcal{O}(\Delta t^3) \qquad \textbf{Equation 2.11}$$

This is our guess function to describe the trajectory defined by the atom during the interval Δt, and it resembles the solution for the linear trajectory of a particle acted by a constant force, where $\frac{d}{dt}\mathbf{r}(t)\Big|_{t=t_i}$ is the velocity of the atom at the instant t_i, $\mathbf{v}(t_i)$, and $\frac{d^2}{dt^2}\mathbf{r}(t)\Big|_{t=t_i}$ is the acceleration of the atom for that same instant, $\mathbf{a}(t_i)$. The initial velocities for the atoms in the system are generally derived from a Maxwell-Boltzmann distribution at a given temperature, $T \geq 0$ K, in such a way that the velocity of the centre of mass of the system is null.[153,161]

The fact that Equation 2.11 describes a conservative trajectory and it is relatively fast to solve, makes it a good candidate as a propagator for the trajectory of our biomolecular system. However, this algorithm shows instabilities and inaccuracies, mainly due to neglecting of higher order terms in the expansion. To overcome this issue, the Verlet algorithm[162] is more commonly employed. Additionally to the fact that the position $\mathbf{r}(t_i + \Delta t)$ depends of the previous $\mathbf{r}(t_i - \Delta t)$ and the current $\mathbf{r}(t_i)$ positions, it is also independent of the velocities of the atoms and it does cancel the error that would be introduced from the contributions of the even powers from the Taylor's expansion of $\mathbf{r}(t_i + \Delta t)$ (see Equation 2.12). Similarly to Equation 2.11, it also reproduces a conservative trajectory and does not require significantly larger storage memory to solve the propagator's equation.

$$\mathbf{r}(t_i + \Delta t) = 2\mathbf{r}(t_i) - \mathbf{r}(t_i - \Delta t) - \Delta t^2[\boldsymbol{\nabla} U \cdot \mathbf{r}(t_i)] \qquad \textbf{Equation 2.12}$$

Despite that it works well to study properties in which only the position of the system is of interest, the lack of explicit particle-velocities presents an obstacle to control the temperature throughout the MD simulation. Therefore, the velocity Verlet integrator has also been developed to keep track of velocities during the simulation.[153] Despite that these type of algorithms exhibit increasing numerical instabilities and *roundoff* errors as simulations are extended, they are still preferred when increasing simulation times are desired.[163]

Aside from the issues with the type of propagator to employ in MD simulations, there are several additional problems that we have to address so that the results provided by these are statistically significant, namely the time step employed to maximize the simulation time within the limits of computational efficiency, and the control over the conditions of the phase space ensemble (microcanonical, canonical or isobaric-isothermal) that we wish to study.[153]

2.1.4.2. Integration time step in MD simulations

Firstly, we need to remind ourselves that the goal of MD simulations is to explore the phase space of a given biomolecular system by generating successive configurations over time. We cannot guarantee that this space is effectively explored unless the simulation is sufficiently long to allow that the computation of averages over time, $\langle A \rangle_{time}$, is approximate to the averages over the phase space of the system in the simulated ensemble, $\langle A \rangle_{phase\ space}$.[164] On the other hand, for that to happen we now need to find the best compromise between the integration time step, Δt, and the limitations of our trajectory propagator. The assumption that we make to obtain Equation 2.11 and Equation 2.12 is that for a given Δt each atom is acted by a constant force vector, which

is truly not the case if we attend that most internal modes are fitted to periodic potentials, and thus, are acted by varying restorative forces that drive these modes to their equilibrium positions. Moreover, the atoms are deflected also by neighbouring atoms due to non-bonded interactions that deviate their assumed linear trajectory. Taking all this into consideration, the integration time step should be smaller than the fastest molecular vibration in the biomolecular system (which rounds 10 fs for C–H bonds), and smaller than the average collision time in an atomic fluid (which is of about 5 fs for an argon fluid).[153] On the other hand, if we are to simulate large time intervals a time step of 1 fs demands a very high number of operations, and it is often not significant to provide for statistical significance. A scheme that is commonly employed is to constrain the fastest vibration mode, thus allowing that a larger integration time step is employed. The SHAKE algorithm is such an example.[165] In this algorithm a contribution from each constrained distance is added to the solution of Equation 2.12, so that these constrained distances, r_{ij}, are kept at a fixed value, d_{ij}. Then, a set of coefficients similar to force constants, λ_{ij}, is optimized to provide the restoring force required to keep r_{ij} at the value d_{ij}.[166]

2.1.4.3. Conditions of the phase space ensemble in MD simulations

Secondly, to study biomolecular systems we also need to provide the conditions that define its phase space. These can range from constant temperature or energy to constant pressure or volume, but for most biomolecular systems we are generally interested in exploring the phase space of a system with constant number of particles, pressure and temperature (*NPT*), the so called isobaric-isothermal ensemble.

To start off, we need to discuss the dimensionality of our system. The basic units of a biomolecular system are the solute (*e.g.* proteins, lipids, carbohydrates, cofactors and other organic units) and the solvent (for most biological processes we have water). Most of the solute units in biochemistry are large macromolecules in both mass and volume, and are solvated to a larger extent by the water solvent. This means that the explicit inclusion of solvent demands that a considerable radii within the solute should be accounted for. Such considerations result in an increasing system, and in more calculations of forces for each iteration. Another issue is that, since each solvent molecule will also interact with the neighbouring solvent molecules, our system should be infinitely expanded or truncated by a surface tension to define a finite volume for our system.

Implicit solvation models could constitute an alternative to explicit solvation models. In these models, the effects that derive from the solvent could be included from introducing a friction

coefficient to simulate the viscosity of the solvent, or a dielectric constant that would screen the effect of electrostatic interactions to a shorter range; they could also be based in empirical potential energy functions derived from considerations on the solvent accessible area of the atoms of the solute, or non-polar contributions to the solvation energy.[167] However, the results of implicit solvent simulations have been shown unable to provide sampling that can be used to either estimate thermodynamic quantities or even realistic trajectories for most biomolecular systems.[167,168] Thus, we do often have to return to explicit solvation models. To solve the problem of the dimensionality, we should remember again that truncation schemes for long-range interactions mostly rely on PBC and *cutoff* interaction distances.

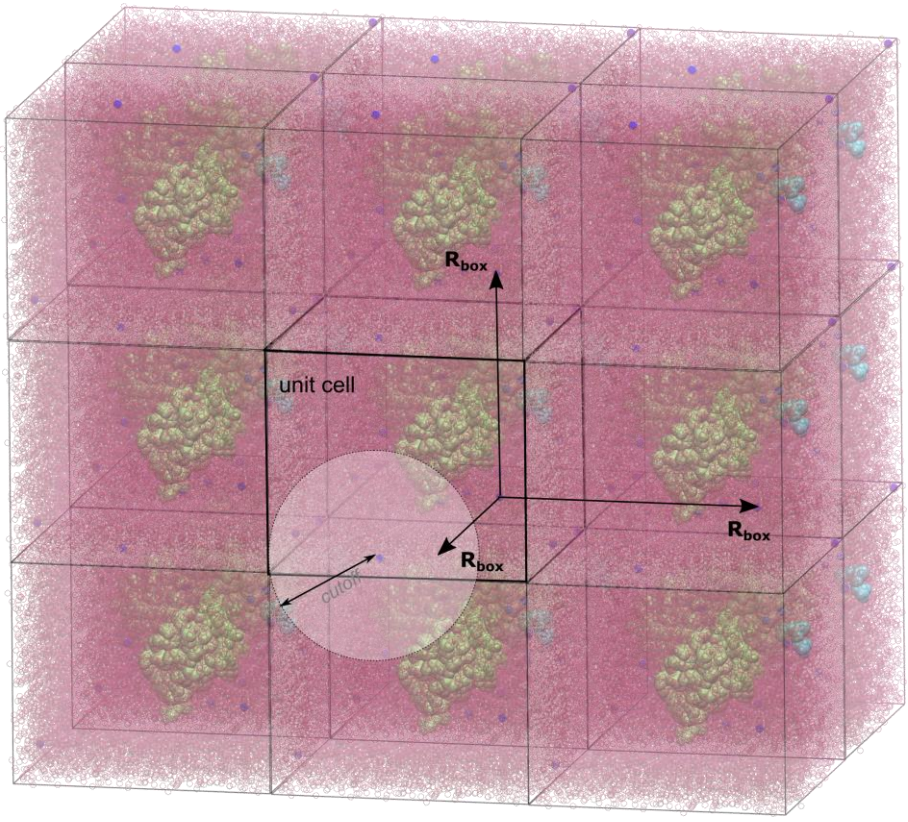


Figure 2.3. Representation of the periodic boundary conditions for a protein system (in lime green) and its substrates (in cyan), solvated with water molecules (in magenta). The parallelepipedic unit cell is highlighted in solid black lines.

In Figure 2.3 we summarize the PBC approach that is commonly employed in MD simulations. Our system is a unit cell (usually either parallelepipedic or octahedral) that is infinitely repeated in every direction of the space. What this means is that when an atom moves across the boundaries of the unit cell, it will occupy a new position that is not $\mathbf{r_{i+1}}$, but $\mathbf{r_{i+1}} - \mathbf{R_{box}}$, being $\mathbf{R_{box}}$ the

translation vector required for the unit cell to generate the neighbouring replica in which $\mathbf{r_{i+1}}$ would be located. As a result, the number of atoms inside the unit cell is constant throughout the entire MD simulation.

We have already discussed that long range interactions are accounted differently after a given *cutoff* is reached, since Coulomb and Lennard-Jones potentials are divergent. Now that our system is a periodic system, the Ewald summation method can be applied to calculate the long range coulomb interactions in the reciprocal space of the phase space.[169] Additionally, we can easily fix the volume of our system, by univocally defining the dimensions of the unit cell. However, several self-interaction artefacts may result from this approach if our box is not large enough. Such self-interaction artefacts can be solved by expanding the box, but the number of pairwise interactions will also increase significantly; hence, this expansion of the box should be performed carefully. On a final note, a common procedure when performing explicit solvent MD simulations is to build solvent boxes that include more than 8 Å of solvent around the solute molecule.

Regarding the simulations at constant pressure and temperature, this regulation is more complex to achieve, since we have already discussed above that our propagator does not explicitly account for these properties. Hence, we need to evaluate them at every iteration throughout the simulation and provide for adequate potentials that can counterbalance their gradient. We will discuss, in particular, the Berendsen barostat[170] and Langevin dynamics,[171] that are commonly employed to regulate pressure and temperature, respectively.

The starting point of the Berendsen barostat is that the pressure of the system enclosed in the unit cell has a contribution from the average velocities of the particles of the system, which move freely as in an ideal gas, and an additional contribution that accounts for the potential raised from the interactions of these particles with each other.

$$P(t) = \underbrace{\rho_{unit\ cell} k_B T(t)}_{\textbf{ideal gas}} + \underbrace{\frac{1}{3V_{unit\ cell}} \sum_{ij,j>i} \mathbf{F_{ij}}(t) \cdot \mathbf{r_{ij}}(t)}_{\textbf{virial correction}} \qquad \textbf{Equation 2.13}$$

Equation 2.13 decomposes the pressure of an isotropic system for each direction of a cubic unit cell. The ideal gas contribution is simply the result of the equipartition theorem to all atoms in the unit cell, while the virial correction accounts for the potential energy of interaction, $\mathbf{F_{ij}}(t) \cdot \mathbf{r_{ij}}(t)$, for every pair of atoms in the unit cell with volume $V_{unit\ cell}$. We cannot guarantee that the new

pressure of the system is our constant pressure P_0, thus we have to add a coupling system (the barostat) that can restore the pressure $P(t)$ back to P_0, within a relaxation time that we will name τ_P. This coupling will generate a friction coefficient $\frac{k_T \Delta P}{3\tau_P}$ that will scale the velocity of the atoms, adapting the temperature and volume of the system to the constant pressure P_0.[170] From these considerations, it proceeds that we have to provide: our solvent with an isothermal compressibility constant k_T, which is commonly that of water at the temperature of 298 K; and our barostat with a relaxation time τ_P that should be sufficiently large comparatively to our integration step, so that the pressure in the coupled system changes in an infinitely slow fashion.

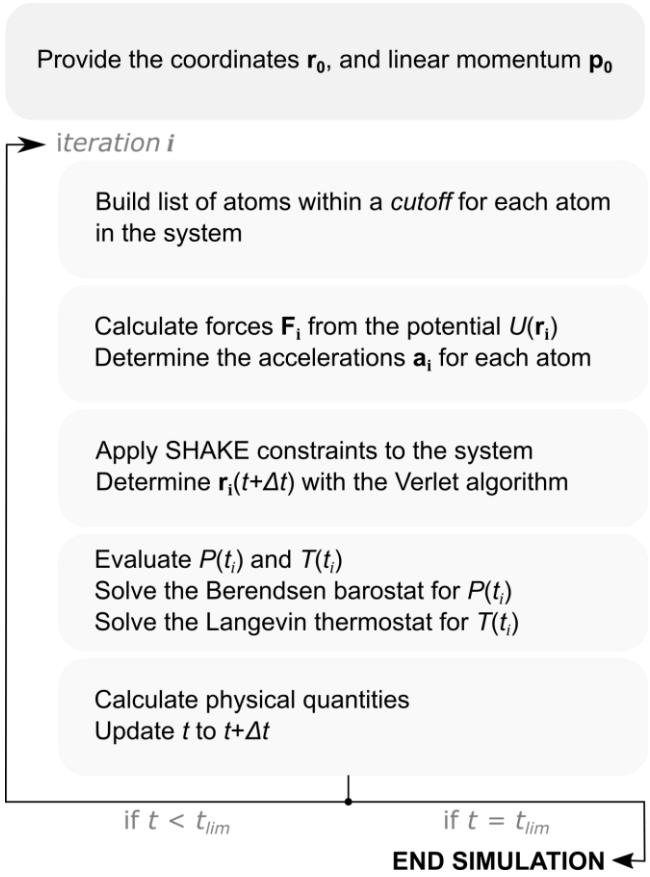
Finally, the control of temperature from Langevin dynamics resides in its stochastic character. Contrarily to the fundamental equation of Newtonian dynamics (Equation 2.10), the Langevin's equation (Equation 2.14) has two additional terms (in square brackets) that concern the collisions between the solute and solvent molecules: one that accounts for the friction of the solvent that cannot be described by the force field potential, and another that accounts for the statistical error that results from our deterministic potential energy function.[166,172]

$$\frac{d^2}{dt^2}\mathbf{r}(t) = -\frac{\boldsymbol{\nabla} U(\mathbf{r})}{m_i} + \left[\underbrace{-\gamma \frac{d}{dt}\mathbf{r}(t)}_{\textbf{friction term}} + \underbrace{\left(\frac{2\gamma k_B T_0}{m_i}\right)^{\frac{1}{2}} \delta(t)}_{\textbf{stochastic term}} \right] \qquad \textbf{Equation 2.14}$$

In the friction term of Equation 2.14, the velocity of each particle is scaled by a friction coefficient that resembles that of the mechanics of a shear viscous fluid, which in our case is the equivalent to an implicit solvent. It results from the fact that in our system there are non-elastic collisions between the solvent and the solute, and that these cannot be described by the conservative potential energy function. The coefficient γ is an average of the number of collisions that is required to damp the highest vibration amplitude in about e^{-1}, and for water it typically rounds 50 ps^{-1}.[172]

The stochastic term arises from the fact that in real systems random collisions between the solute and the solvent occur, aside from the shear friction by the solvent.[166,172] These collisions occur in every direction and their frequency should depend of the temperature, T_0, of the heat bath that surrounds the N-particle system. In Langevin dynamics, it is assumed that there is a change in linear momentum resulting from these random collisions, which are also called Langevin forces. These Langevin forces follow a Gaussian distribution with zero average and variance $2\gamma k_B T_0 m_i$, and rely on information from the past events of the simulation to generate these random

forces.[166,173] In this way, while the friction term dissipates energy from the accelerated particle to the surrounding thermal bath, Langevin forces provide random momentum increments that result from the interaction of the thermal bath of temperature T_0 with the particles of the system. Despite that the use of Equation 2.14 is limited to systems of spherical particles and in which correlation times between successive collisions are much shorter that other relaxation times of the system, it efficiently regulates the temperature of the system and it provides for a better search over the phase space of our system in study.[173]



Scheme 2.1. Flowchart for an MD simulation with an *NPT* ensemble.

At this stage, we can perform MD simulations with *NVT* and *NPT* ensembles to study the biophysics of biomolecular systems. Moreover, we can also post-treat the data from MD simulations to determine thermodynamic quantities, from statistical physics considerations. We enclose the discussion on MD simulations with a summary of the operational tree for common MD simulations performed with the isobaric-isothermal ensemble (*NPT*), in Scheme 2.1.

2.2. Quantum Mechanics methods

2.2.1. Expose the problem

MD simulations are now routinely employed in ever increasing systems (up to millions of atoms) and to larger time-scales (from a few ns to hundreds of μs).[148,149,174] Its large applicability results from the combination of the constant upgrading and refinement of current empirical force fields, the improvement and widespread availability of computational resources, and the development of new improved methodologies to enhance phase space exploration and statistical treatment of MD simulation data.[175-178]

Biomolecular systems are, nevertheless, large and complex aggregates with enigmatic behaviours that are often enclosed in particles much smaller that the bulk spherical atom. The truth is that most of the biochemistry occurs in a reality far different from the classical regime we have been discussing until now. To describe the electronic phenomena that occur at a sub-atomistic scale, we have to resort to QM calculations. We have already discussed that the trade-off that we have to make to obtain such detailed insight is that we have to confine ourselves to systems that often cannot reach up to 200-300 atoms.[141,142]

DFT methods have been the preferred choice to tackle the chemistry of biosystems, since they have provided for accurate results in the study of systems with a considerable dimensionality (*circa* a hundred atoms).[179] Currently, the main drawback in DFT is that up to now we still do not know what formulation the exchange-correlation density functional (DF) should assume, so that this parcel of the energy can be accounted for in an exact fashion. As a result, several DFs have been developed over the years.[179,180] These models are generally optimized by obeying to specific physical principles and by reproducing particular properties, such as activation barriers, reaction energies, electron affinities, among others, for several sets of chemical reactions.[180] Post-HF methodologies could be an alternative to DFT methods, since they rely solely on physical principles to estimate the real energy of the system; however they are not employed on a regular basis since once more the dimensionality problem prevails.[179,180]

2.2.2. Building the molecular wave function

Quantum mechanics marks a change in our understanding of the way in which the events of the microscopic world occur. The realization of the dual character of matter has provided a new paradigm for the quantization of physical observables. At an atomic scale, the state of a physical

system is fully described by a wave function, $|\Psi\rangle$, which allows the determination of any physical quantity that one wishes to measure. In chemical systems, the wave function of the molecule represents the space of the interactions between nuclei or electrons from the atoms of the system, and thus, the space vector that describes a given molecular state should be given by a function of the coordinates of the nuclei, $\mathbf{r_N}$, and electrons, $\mathbf{r_e}$, of the system. The major obstacle in any QM calculation is to provide for an accurate representation of this molecular wave function, $|\Psi_{r_N,r_e}\rangle$, since there is no way to analytically determine the state wave function for polielectronic systems.

We must resort, alternatively, to numerical methods, as we have done previously in molecular mechanics calculations. The goal is to find a molecular wave function, $|\Psi_{r_N,r_e}\rangle$, that can minimize the energy of the system that results from the action of the Hamiltonian operator, $\hat{\mathcal{H}}$, in the molecular wave function. These mathematical objects are related through the fundamental equation of Quantum Mechanics, the Schrödinger equation (Equation 2.15).

$$\hat{\mathcal{H}}|\Psi_{r_N,r_e}\rangle = E|\Psi_{r_N,r_e}\rangle \qquad \textbf{Equation 2.15}$$

Equation 2.15 has not made things any easier, since we have now two big problems to solve:

(1) We do not know the exact QM Hamiltonian to determine the energy of our system;
(2) We have no clue of what our wave function will look like.

The first issue will be addressed in the subsequent sections. We will now focus on how to provide a guess wave function that can be optimized to accurately estimate the real wave function of our chemical system.

We will start by simplifying the dimensionality of our problem, since our current wave function is a function of both the nuclei and the electrons of the system. We start by taking into account that the velocity at which the electrons rearrange themselves while the nuclei translate in space should be close to instantaneous; thus the movement of electrons and nuclei should be practically uncorrelated. This assumption (also called Born-Oppenheimer approximation) greatly simplifies our problem because now our molecular wave function $|\Psi_{r_N,r_e}\rangle$ can be solved independently for the nuclei $|\Phi_{r_N}\rangle$ and the electrons $|\psi_{r_e}\rangle$.[181] Moreover, we can assume that the nuclei interact as classical particles, since chemical transformations result mostly from the electronic rearrangements that occur between the atoms. At the end of this process, our problem now sums up to provide a guess electronic wave function $|\psi_{r_e}\rangle$ and minimize its energy.

2.2.2.1. Linear Combination of Atomic Orbitals

A possible way to generate the guess electronic wave function is to think of the molecular wave function as a linear combination of atomic orbitals. Following this line of reasoning we would need to parameterize an atomic orbital for each electron in each atom of the molecule, forming a space of atomic orbitals $\{|\phi_i\rangle\}$, which is commonly named as the *basis set*. Then, by solving Equation 2.15 we are now able to determine the energy *eigenvalues* for each electron of the molecule, and calculate orthogonal one-electron molecular orbitals through the linear combinations of the starting atomic orbitals (see Equation 2.16).

$$|\psi_i\rangle = \sum_i c_i|\phi_i\rangle \quad ,\langle\phi_j|\phi_i\rangle = \delta_{ij}(r) \qquad \textbf{Equation 2.16}$$

Now, our problem lies in determining the atomic orbitals space, which is equally challenging since, as we have referred above, they are only exactly known for the hydrogen atom. As a result, atomic orbitals are built from differentiable families of functions that can reproduce properties that atomic orbitals must possess, such as angular momentum, $\vec{\mathcal{L}}$ and electron probability distributions. This choice of family of functions will affect the computational efficiency of the optimization of the molecular wave function. Ideally, we should employ the least number of functions as possible, to reduce the number of integrals to calculate, and choose functions with well-known mathematical properties, to lower the computational requisites for the calculation. The Gaussian type orbitals (GTOs) and Slater type orbitals (STOs) exemplify two families of functions that can approximate the behaviour of one-electron atomic orbitals. Their core shape is presented in Equation 2.17, for GTOs and STOs respectively.[160,179]

$$\phi_{GTO}(\vec{r}) = N{r_x}^{m_x}{r_y}^{m_y}{r_z}^{m_z}e^{-\alpha r^2}$$
$$\phi_{STO}(\vec{r}) = Nr^{n-1}Y_{l,m_l}(\theta,\varphi)e^{-\alpha r} \qquad \textbf{Equation 2.17}$$

N is the normalization constant that refers to the probabilistic nature of the squared wave function, and r is the radial distance from the electron to the atom nucleus. The GTO family of functions also exhibits the set (r_x, r_y, r_z), which stand for the displacement vectors, $\vec{r}$, of the electron relatively to the nucleus, and the *eigenvectors* of the angular momentum operator $\vec{\mathcal{L}}\,(m_x, m_y, m_z)$, both projected in the Cartesian space. In an analogous way, in the STO family of functions we have the set (r, θ, φ) and the spherical harmonics $Y_{l,m_l}(\theta,\varphi)$[182] that account for the projection of the position and angular momentum vectors of the electron in the Spherical coordinate space.

While the coefficient α can be parameterized to describe how fast does the electron states distribution vanish for regions farther from the nucleus, the displacement and angular momentum vectors can be combined to reproduce the degenerate orbitals that can be occupied as the dimension of the *eigenvector* space for a given angular momentum increases.

Despite that the STO family of functions represents more accurately the cusps of the electron wave function near the nucleus, most *basis sets* employed nowadays combine GTOs to build the atomic orbitals space. The main advantage of these functions is that the mathematical properties of these functions are well-known, particular in what recalls the computing of integrals of GTO products for N-electron interactions. The drawbacks are that GTOs do not capture the exponential decay that occurs in atomic orbitals, and they are unable to reproduce the cusps that occur near the nucleus unless several GTOs are combined.[160] This results in an increase in the number of primitive functions that will require optimization in posterior energy minimization calculations. On the other hand, fewer STOs are required to describe atomic orbitals, but the calculation of the N-electron interactions can be extremely expensive.

One thing that can be done to overcome the physical problems derived from the use of GTOs is to use subfamilies of GTOs in which primitive Gaussians are linearly combined to reproduce the overall behaviour of the atomic orbitals. These subfamilies, also called contracted Gaussian type orbitals (cGTOs), are still GTOs, and, as result, possess the same mathematical properties.

$$\phi_{cGTO}(\vec{r}) = \sum_i \beta_i \phi_{GTO}(\vec{r}, \alpha_i) \qquad \textbf{Equation 2.18}$$

The subsequent problem is a coefficient optimization problem in which the coefficients α_i and β_i (in Equation 2.18) are optimized through QM calculations to reproduce accurate atomic orbital energies determined from experimental values.

Several basis sets have been developed in the past decades, such as the Pople's,[37-44] Ahlrich's[45,46] and Dunning's[47-49] *basis sets*. However, the most common *basis sets* found in literature are STO-3G[179] or 6-31G. In the first *basis set* STOs are reproduced by using a cGTO subfamily that combines three Gaussian primitives to reproduce all atomic orbitals, while the second uses a cGTO of six Gaussian primitives to build core electron orbitals, a cGTO of three Gaussian primitives to describe the orbitals of the inner valence electrons and one Gaussian primitive to describe the orbitals of the outer valence electrons.[160] Figure 2.4 depicts the

comparison of the STO-3G *basis set* and each of its three Gaussian primitives against the 1*s* Slater-type orbital for the hydrogen atom.

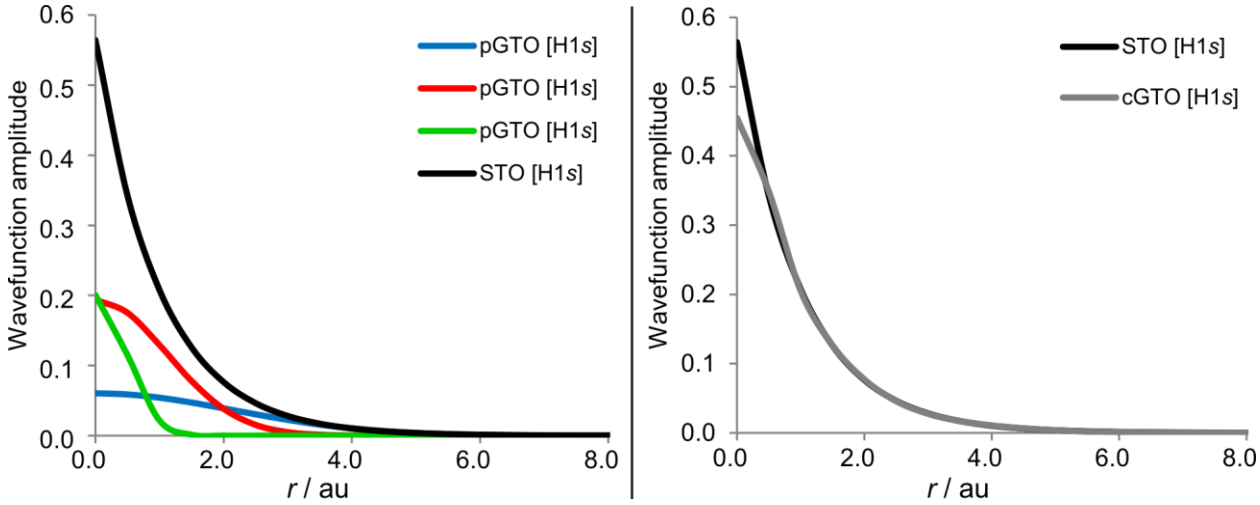


Figure 2.4. On the right, comparison of each pure primitive Gaussian function (pGTO) of the STO-3G *basis set* against the 1*s* Slater orbital of the hydrogen atom; on the left, comparison of the cGTO from the STO-3G *basis set* (combination of the three pGTOs at the right) against the 1*s* Slater orbital of the hydrogen atom.

Despite that these types of *basis sets* have been extensively used for first- and second-row elements,[115,116] the representation of the atomic orbitals for heavier atoms from GTOs is much more demanding, since their core electrons exhibit significant relativistic effects that are not accounted for in most Hamiltonians employed in computational chemistry. As an alternative, we can reckon that the inner electrons of heavy atoms must not exhibit a significant chemical behaviour. Hence, effective core potentials (ECPs) have been developed to describe the core electrons of these heavy atoms from one-electron operators, $\hat{h}_{core}$, that interact with the valence electrons of the heavy atom, with the latter described with GTOs. The $\hat{h}_{core}$ operator accounts for a potential function that describes the Coulomb and exchange-correlation interactions of the inner electrons (as well as any relativistic effects that are observed in these electrons), and a projection operator to represent the wave functions of valence electrons in the space generated by the frozen-core orbitals of the heavy atom.[36] As for optimization of GTO_cs, ECPs may be optimized from the accurate calculation of several atomic properties with relativistic Hamiltonians and large *basis sets*, such as ionization potentials or electronic affinities.

2.2.2.2. Determining the molecular wave function: the secular equation

We have discussed the tools to build guess one-electron molecular orbitals as linear combinations of atomic orbitals. Recalling Equation 2.16, we have to optimize the coefficients c_i that indicate the relative contribution of each atomic orbital to the final one-electron molecular orbital. We start off by using Equation 2.15 to determine the energy of the one-electron wave function. To do this, we apply the reciprocal $\langle\psi_i|$ in both sides of it, and solve it to obtain the energy of the one-electron molecular orbital ε_i. Equation 2.19 systematizes the operation, and shows its result when we expand the one-electron wave function as a sum of GTOs from our *basis set*.

$$\varepsilon_k = \frac{\langle\psi_k|\hat{\mathcal{H}}|\psi_k\rangle}{\langle\psi_k|\psi_k\rangle} \xrightleftharpoons{|\psi_k\rangle=\sum_i c_i|\phi_i\rangle} \varepsilon_k = \frac{\sum_{j\geq i}\sum_i c_j c_i \langle\phi_j|\hat{\mathcal{H}}|\phi_i\rangle}{\sum_{j\geq i}\sum_i c_j c_i \langle\phi_j|\phi_i\rangle} \qquad \textbf{Equation 2.19}$$

We emphasize that the functions the space of GTOs, $\{\phi_i\rangle\}$, are invariant, which means that, for a given configuration of atoms, the matrix elements $\langle\phi_j|\hat{\mathcal{H}}|\phi_i\rangle$ and $\langle\phi_j|\phi_i\rangle$ are known in every iteration, and so the problem becomes simply dependent of the choice of coefficients c_i. These matrix elements are called the resonance integrals, $\mathcal{H}_{ij}$, and the overlapping integrals, $\mathcal{S}_{ij}$, respectively.[160] The former integrals account for the energy that is stored in the overlapping of each set of GTOs, while the overlapping integrals are Dirac integrals that derive from the fact that our Gaussian-type atomic orbitals do not form an orthogonal space.

In the step that follows we apply the variational principle in such a way that the energy ε_k is minimized for every coefficient c_i, which is equivalent to find every c_i so that the first derivative of ε_i, $\frac{\partial\varepsilon_k}{\partial c_i}$, is zero. Differentiating Equation 2.19 in this way, we obtain a set of differential equations that should be simultaneously satisfied for every coefficient c_i (Equation 2.20).

$$\sum_{j\geq i} c_j\left(\mathcal{H}_{ij} - \varepsilon_k\mathcal{S}_{ij}\right) = 0 \quad , \begin{cases}\mathcal{H}_{ij} = \langle\phi_j|\hat{\mathcal{H}}|\phi_i\rangle \\ \mathcal{S}_{ij} = \langle\phi_j|\phi_i\rangle\end{cases} \qquad \textbf{Equation 2.20}$$

This is equivalent to calculate the roots of the determinant of a square matrix that crosses all of the $\frac{\partial\varepsilon_k}{\partial c_i}$ solutions. Equation 2.21 depicts such a determinant, which is also known as the secular equation, with N the number of basis functions that are used to optimize $|\psi_i\rangle$.[160]

$$\begin{vmatrix} \mathcal{H}_{11} - \varepsilon_k \mathcal{S}_{11} & \dots & \mathcal{H}_{1N} - \varepsilon_k \mathcal{S}_{1N} \\ \vdots & \ddots & \vdots \\ \mathcal{H}_{N1} - \varepsilon_k \mathcal{S}_{N1} & \dots & \mathcal{H}_{NN} - \varepsilon_k \mathcal{S}_{NN} \end{vmatrix} = 0 \qquad \textbf{Equation 2.21}$$

The solutions of the secular equation will be the several ε_k (that may or may not be degenerate) that will allow for a new system of linear equations to be built from Equation 2.19, thus retrieving the optimized coefficients that will form the optimized one-electron wave functions.

2.2.2.3. Many-electron molecular wave function: electron spin and Pauli's exclusion principle

We have already seen that the resolution of the secular equation provides optimized one-electron molecular wave functions that, contrarily to the GTO *basis set*, form an orthogonal space of functions $\{|\psi_{e_i}\rangle\}$. Due to this orthogonality condition, the polielectronic wave function should be a product of one-electron wave functions.

However, we have not yet discussed the fact that electrons have an intrinsic magnetic momentum (the spin) that interacts with the orbital angular momentum. This is particularly relevant in polielectronic systems, where each orbital is doubly occupied. The Stern-Gerlach experiment has shown that electrons are particles of half-integer spin, forming a space composed of two *eigenvectors*, $\{|\sigma_\uparrow\rangle, |\sigma_\downarrow\rangle\}$.[183,184] Two consequences arise from these results:

(1) Each occupied orbital, $|\psi_{e_i}\rangle$, may be doubly occupied, with electrons described by a spin-orbital wave function that is the product of the orbital and spin wave functions, $|\psi_{r,\sigma_\uparrow}\rangle$ or $|\psi_{r,\sigma_\downarrow}\rangle$;

(2) The polielectronic wave function must be antisymmetric, which is a result of the Pauli's exclusion principle. This is a result of the fact that electrons are indistinguishable particles that can permute between orbitals without changing the form of the polielectronic wave function, and that no electron can be characterized by more than one state (orbital).

It is clear that the product of one-electron wave functions we have referred above cannot be our solution, since upon permuting the coordinates of any two electrons we would end up with the same wave function. Moreover, the antisymmetry condition is not met. Our solution can be derived by knowing that the permutation operator $\hat{\mathcal{P}}_{ij}$ is a unitary operator, which means that a double permutation of a set of coordinates in a wave function, results in that same wave function. Taking

this into consideration, our N-electron wave function can be written as in Equation 2.22 (the Slater determinant).[185]

$$|\psi_e(e_1,\dots,e_N)\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} |\psi_{1,e_1}\rangle & \cdots & |\psi_{N,e_1}\rangle \\ \vdots & \ddots & \vdots \\ |\psi_{1,e_N}\rangle & \cdots & |\psi_{N,e_N}\rangle \end{vmatrix} \quad , \quad |\psi_{i,e_j}\rangle = \begin{cases} |\psi_{i,r_j,\sigma_\uparrow}\rangle \\ |\psi_{i,r_j,\sigma_\downarrow}\rangle \end{cases} \qquad \textbf{Equation 2.22}$$

The Slater determinant represents every combination that may result from our system of N-electron. The electronic wave function is still described by products of orthogonal one-electron spin-orbital wave functions; however it is now a linear combination of every permutation of electrons labelled with the same spin, within this product. Moreover, it is an antisymmetric wave function, due to the fact that the Slater determinant is solved through a Laplace's expansion, which changes the signal of the determinant for every permutation of columns.[186]

2.2.3. Hartree-Fock method

We are now capable of defining the Hamiltonian for our molecular system. This Hamiltonian should account for the kinetic and potential energy of both the nuclei and the electrons of the system of interest. However, the Born-Oppenheimer approximation states that we can treat them separately, in a space where the nucleus is a particle with no kinetic energy (the average temperature of the system is $\langle T\rangle = 0$) that interacts classically with the neighbouring nuclei. Hence our QM Hamiltonian will be the sum of the kinetic energy of the electrons, $\hat{\mathcal{T}}_{e_i}$, the nucleus-electron attractive potential, $\hat{\mathcal{V}}_{e_i}^{n}$, and the electron-electron interacting potential, $\hat{\mathcal{V}}_{e_i}^{e}$. While the former two operators result in one-electron wave function integrals, the operator that accounts for electron-electron interactions requires for more complex integrals to be solved.

2.2.3.1. Hartree-Fock Hamiltonian

The Hartree-Fock method is the simplest method to determine the energy and orbital configuration of a polielectronic system. It belongs to a class of computational methods that provides the wave function of the N-electron system from one single Slater determinant, $|\psi_e\rangle$. The energy of this configuration is determined from the Hartree-Fock Hamiltonian, $\hat{\mathcal{F}}$, which is depicted in Equation 2.23 in atomic units.

$$\hat{\mathcal{F}} = \underbrace{-\frac{1}{2}\sum_{e_i} {\nabla_{e_i}}^2}_{\hat{\mathcal{T}}_{e_i}} \underbrace{-\sum_{n_i}\sum_{e_i}\frac{Z_{n_i}}{r_{n_i,e_i}}}_{\hat{\mathcal{V}}^n_{e_i}} + \underbrace{\sum_{e_i,e_j>e_i}\sum_{e_i}\frac{1}{r_{e_i,e_j}}}_{\hat{\mathcal{V}}^e_{e_i}} \qquad \textbf{Equation 2.23}$$

A form of the $\hat{\mathcal{T}}_{e_i}$ and $\hat{\mathcal{V}}^n_{e_i}$ operators is quite usual. The $\hat{\mathcal{T}}_{e_i}$ describes the energy stored in the momentum of the electron wandering in the volume enclosed by the wave function; and $\hat{\mathcal{V}}^n_{e_i}$ describes the interaction between the wave function of the electron e_i and the single positive charge of the nucleus n_i through a Coulomb potential centred at the nucleus. The electron-electron operator also follows the form of the Coulomb potential; however once we apply it to a two-electron Slater determinant $\left|\psi_{e_i,e_j}\right\rangle$, we are confronted with an unexpected contribution (see Equation 2.24).

$$\mathcal{V}^e_{e_i} = \int + \underbrace{\psi^2_{\mu,e_i}\frac{1}{r_{e_i,e_j}}\psi^2_{\nu,e_j}}_{\mathcal{J}_{ij}} - \underbrace{\psi_{\mu,e_i}\psi_{\nu,e_j}\frac{1}{r_{e_i,e_j}}\psi_{\mu,e_j}\psi_{\nu,e_i}}_{\mathcal{K}_{ij}}\, dr_i dr_j d\sigma_i d\sigma_j \qquad \textbf{Equation 2.24}$$

While the first parcel of Equation 2.24, $\mathcal{J}_{ij}$, is the interaction of the probability density of the two electrons, i and j, in their one-electron orbitals, ψ_μ and ψ_ν, which is also called the Coulomb repulsion energy; the second energy, $\mathcal{K}_{ij}$, does not stand for the same interpretation. Additionally, this latter term is a result of the antisymmetric character of the two-electron wave function, which implies that this stabilization energy results from the permutation operator in the Laplacian expansion of the Slater determinant wave function.

The energy $\mathcal{K}_{ij}$ is commonly called the exchange energy or the Fermi energy, since it results from the fact that electrons are fermions (half-integer spin particles) that can be permuted interchangeably between one-electron orbitals. This energy of interaction stands for the probability of finding two electrons with the same spin closer together, and it is more pronounced when significant overlapping of ψ_μ and ψ_ν occurs at shorter distances. The result for $\mathcal{K}_{ij}$ in Equation 2.24 also enforces that this energy is only different of zero if the interacting electrons possess the same spin (otherwise the product of the orthogonal spin *eigenvectors* $\langle\sigma_\uparrow|\sigma_\downarrow\rangle$ nullifies this contribution). This result is quite representative of how different are the quantum phenomena from the classical ones, even when our Hamiltonian is mostly built from classical physics considerations.

2.2.3.2. Roothaan-Hall approximation

To solve the Hartree-Fock Hamiltonian in a sequential way, we can take profit from the fact that Equation 2.23 can be re-written as a linear combination of one-electron operators (called one-electron Fock-operators, $\hat{\mathbb{f}}_i$). The one-electron Fock-operator is represented in Equation 2.25, with the one-electron operators for the kinetic, nucleus-electron attractive, electron-electron repulsion and electron exchange operators represented by the equivalent small letters of the operators in Equation 2.23 and Equation 2.24.

$$\hat{\mathbb{f}}_i = \overbrace{\hat{t}_{e_i} + \sum_{n_i} \hat{v}_{e_i}^{n}}^{\hat{h}_{e_i}} + \overbrace{\sum_{e_j > e_i} \left[\hat{J}_{e_i}\{e_j\} - \hat{K}_{e_i}\{e_j\}\right]}^{\hat{v}_{e_i}^{e}\{e_j\}} \qquad \textbf{Equation 2.25}$$

As a result, we can make use of the LCAO approximation (Equation 2.16) and recover the secular equation (Equation 2.21) for the electron-occupied orbitals. Solving the secular equation will now hold for a problem of coefficient optimization, c_i, instead of $|\psi_{e_i}\rangle$ optimization. The LCAO approximation applied to the Hartree-Fock Hamiltonian is known as the Roothaan-Hall approximation or LCAO self-consistent-field,[187] and each matrix element of the secular equation is called a Roothaan equation. By solving the roots for the determinant of the secular equation, we determine the energies, ε_i, for each one-electron molecular spin-orbital, which will allow for the calculation of the optimized coefficient matrix, C. The one-electron molecular spin-orbitals are finally calculated from the linear combination of the $c_i|\phi_i\rangle$ that form the corresponding *eigenvector* of the *eigenvalue* ε_i. In the end of these algebraic operations we obtain a set of orthogonal antisymmetric one-electron molecular spin orbitals $\{\psi_{e_i}(r,\sigma)\}$.

2.2.3.3. Hartree-Fock limit: electron correlation

The main advantage in the Hartree-Fock method is that it relies on a variational scheme. This means that, despite that the energy calculated from the *eigenvalues* of the Hartree-Fock Hamiltonian should be always larger than the actual energy of the minimum energy state, the molecular orbitals derived from a known finite *basis set* will always describe a configuration of minimum energy for the Hartree-Fock Hamiltonian. If we were to extend the energy resulting from the Hartree-Fock Hamiltonian to an infinite *basis set*, the complete *basis set* (CBS), we would end up with the exact energy of the system at the Hartree-Fock level of theory.[160] However, the Hartree-Fock method was shown to provide only reasonable results for system of closed-shell

electronic configurations (with all electrons paired, in a singlet spin state), and for systems with low degeneracy of one-electron molecular orbitals. The main issue comes from the fact that we assume that there is only one electronic configuration for an energy minimum (the electron configuration is univocally defined by one Slater determinant), and that only occupied orbitals contribute to the energy of the system. Why should this be not true?

The main assumption for the Hartree-Fock theory is that electrons can be treated as non-interacting particles that occupy the volume of its wave function under the influence of the nuclei of atoms and the average field generated by the remaining electrons of the molecule. From this perspective, the molecular wave function is a single determinant where every electron can occupy every possible one-electron wave function.

In reality, electrons are interacting particles whose energy and wave function depend on explicit electron-electron spin-orbital interactions and the N-electron configuration itself. This additional stabilization, which is the difference between the real energy of the molecular system and that of the Hartree-Fock limit, is called the correlation energy, and it corrects for the overestimation of the electron repulsion from the Coulomb integrals. Generally, this energy holds a small contribution to the total energy of the system; however, for open-shell systems or electron transfer reactions, the correlation effects can be of outmost importance to describe the chemical reactivity of several systems.[160,188] In order to appropriately account for this energy, there is always a compromise between the computational feasibility of the calculation and the quality of the Hamiltonian, and the molecular wave function.

Before exploring in which ways we can tackle the problem electron correlation, we shall discuss, in a more detailed manner, the physical interpretation of electron correlation. First, we have assumed that our polielectronic wave function may be represented by one Slater determinant, which accounts for every permutation of every electron between two spin-orbitals with the same spin. After that, we have also assumed that the N-electron Hamiltonian can be described by a sum of one-electron linear operators, which result in a set of N energies attributed solely to each of the N electrons. What consequences can be derived from these approximations?

(1) Our molecular wave function is confined to a space of occupied one-electron spin-orbitals $\{\psi_{e_i}(r,\sigma)\}$. This implies that there are no optimized virtual orbitals that can be occupied by electrons so that a configuration with lower energy can be generated. The approximation can be quite acceptable in cases for which there is low degeneracy of the orbitals of the valence electrons. However, particularly for systems in which valence orbitals may exhibit

a considerable degree of degeneracy (as it may occur for orbitals with high angular momentum, high spin multiplicity or high electron delocalization), often the occupied orbitals are not suited to fully describe the molecular wave function. In these cases, the electrons can be better described by mixed states of occupied and virtual orbitals. This correlation energy is called the static correlation energy, or Fermi correlation energy, since it is a result of the way the electronic configuration is described, and not of the way the electrons interact with one another.

(2) On the other hand, the more accurate way to study the electron-electron interaction would be to evaluate the effect of each electron and group of electrons on the remaining electrons of a given configuration, if other orbitals of similar energy could be occupied by these. In this way, the configuration would respond in a dynamical way as each electron or group of electrons was excited to these virtual orbitals. The correlation energy resulting from the dynamic response of the electronic configuration to these electron excitations is called the dynamic correlation energy.

2.2.4. Post-Hartree-Fock methods

Post-Hartree-Fock methods are an *ab initio* attempt to refine the Hartree-Fock method. Every method based on first-principles considerations that has been proposed so far, points out to the fact that the problem of electron correlation must be a result of the plasticity that we have attributed to our wave function and the electronic Hamiltonian. Specifically, we will address two types of post-Hartree-Fock methods: the multiconfigurational methods (that try to tackle the problem of electron static correlation), and electron excitation methods (that try to tackle the problem of dynamic electron correlation).

2.2.4.1. Multiconfiguration Self-Consistent-Field

We will start this discussion by recalling the constraints that we have imposed to our molecular wave function throughout the last sections: firstly, we have assumed that the molecular wave function should be accurately determined by taking into account the guess atomic orbitals of each electron in the system (according to the electronic configuration of each atom); secondly, each one-electron orbital is composed from a finite *basis set*.

The general idea behind multiconfiguration methods arises from the fact that molecular orbitals are quite different from atomic orbitals. As a result, it is possible that molecular orbitals resulting

of the combination of occupied and/or unoccupied orbitals may present similar energies, increasing the degeneracy of the molecular system. This is particularly important for open-shell systems or systems with heavy atoms, in which valence orbitals are increasingly closer in energy.

We will define our wave function after a larger set of molecular orbitals (either occupied or unoccupied). Hence, our correct wave function should be a linear combination of every electronic configuration for our new space of orbitals $\{\psi_{occupied}, \psi_{virtual}\}$.[160]

$$|\psi_{MCSCF}\rangle = c_0|\psi_{HF}^0\rangle + \sum_{i=1}^{N} c_i|\psi_{HF}^i\rangle \qquad \textbf{Equation 2.26}$$

Equation 2.26 refers to our multiconfiguration Hartree-Fock equation, $|\psi_{MCSCF}\rangle$, which is a sum over N Slater determinants, $|\psi_{HF}^i\rangle$, with contributions c_i to the overall molecular wave function. We start-off the calculation by determining the molecular wave function from the Hartree-Fock method, $|\psi_{HF}^0\rangle$, in the space of orbitals $\{\psi_{occupied}, \psi_{virtual}\}$.[189] The remaining configuration state functions (CSF), $|\psi_{HF}^i\rangle$, will be generated by permuting electrons between occupied and virtual orbitals in such a way that the angular momentum of the system is conserved. After that, two matrices of coefficients will be solved: one concerning the Roothaan equations for each Slater determinant, and one concerning the linear combination of CSFs in Equation 2.26. In this way, we can determine each optimal CSF and its relative contribution to the overall $|\psi_{MCSCF}\rangle$. As for the Hartree-Fock method, these coefficients are solved in a self-consistent fashion, and this process is called the Multiconfiguration Self-Consistent-Field (MCSCF).

The number of CSFs, N, to include in this calculation depends on the number of virtual orbitals that were added to the space of orbitals, m, and the number of electrons that will be exchanged between occupied and virtual orbitals, n. In the case where all permutations of electrons between occupied and virtual orbitals is allowed, our space of orbitals is a complete active space (CAS) in which all orbitals possess the minimum energy and are required to be orthogonal. The $|\psi_{MCSCF}\rangle$ that results of this calculation is then called $|\psi_{CASSCF}\rangle$, where CASSCF stands for Complete Active Space Self-Consistent-Field.

While multiconfigurational methods can account for most of the static correlation energy in a molecular system, they are unable to rationally introduce the dynamic correlation of electrons since the coefficients within each CSF correspond to a different ground state electronic configurations. To determine the dynamic correlation, we have to study in which way electronic

configurations interact with one another as electrons are excited from the ground to excited states. At this point, it is not hard to realize that our problem is much more complicated, since now there is the need to rationalize what type of excitations should be accounted for to accurately account for dynamic correlation. Then, the problem resides in the optimization of the coefficients that enclose the contribution of each configuration to the overall minimum energy state.

2.2.4.2. Configuration Interaction

The most direct application of the concept of interaction between every coefficient in the Hartree-Fock wave function and every coefficient c_i from the MCSCF wave function is the Configuration Interaction (CI).[190-192] It works as linear response method in which the electrons of $|\psi_{HF}^0\rangle$ are excited from its occupied orbitals to every virtual orbital of the molecular wave function, as summarized in Equation 2.27.

$$|\psi_{CISDT}\rangle = c_0|\psi_{HF}^0\rangle + \underbrace{\sum_{i}^{occ.}\sum_{r}^{virt.} c_i^r|\psi_i^r\rangle}_{\textbf{single excitation}} + \underbrace{\sum_{i<j}^{occ.}\sum_{r<s}^{virt.} c_{ij}^{rs}|\psi_{ij}^{rs}\rangle}_{\textbf{double excitation}} + \underbrace{\sum_{i<j<k}^{occ.}\sum_{r<s<t}^{virt.} c_{ijk}^{rst}|\psi_{ijk}^{rst}\rangle}_{\textbf{triple excitation}} \qquad \textbf{Equation 2.27}$$

As for the MCSCF method, the calculation starts with the general optimization of the one Slater-determinant wave function from the Hartree-Fock method, employing the occupied and virtual orbitals space $\{\psi_{occupied}, \psi_{virtual}\}$. In the step that follows, excitations of one-, two-, three-, and N- electrons are performed between occupied orbitals (i, j, k, …) and virtual orbitals (r, s, t, …), thus generating single, double, N-tuple excited-state wave functions ($|\psi_i^r\rangle$, $|\psi_{ij}^{rs}\rangle$, $|\psi_{ijk}^{rst}\rangle$, …). The final wave function is a multi-determinant square matrix (as exemplified in Equation 2.28),[160] with contributions from every generated configuration.

Fortunately, we can reduce some of the elements in the multi-determinant, in particular the energy contribution from the interaction of $|\psi_{HF}^0\rangle$ and the $|\psi_i^r\rangle$ configurations, as a result of the Brillouin theorem. The latter states that, since every occupied and virtual orbitals are *eigenvectors* of the Fock operator, the resonance integrals of these configuration should be null. However, Equation 2.28 also shows that accounting further multiple excitations still provides for several non-null resonance integral matrices that, despite being increasingly sparser, still enclose significate contributions if we are attempting very accurate determinations. Moreover, in closed-shell systems the corrections introduced by single excitations are mostly negligible.

$\hat{\mathcal{F}}$	$\lvert\psi_{HF}^{0}\rangle$	$\lvert\psi_{i}^{r}\rangle$	$\lvert\psi_{ij}^{rs}\rangle$	$\lvert\psi_{ijk}^{rst}\rangle$	...
$\langle\psi_{HF}^{0}\rvert$	E_{HF}	0	dense	very sparce	negligible
$\langle\psi_{i}^{r}\rvert$	0	dense	sparse	extremely sparce	negligible
$\langle\psi_{ij}^{rs}\rvert$	dense	sparse	sparse	extremely sparce	negligible
$\langle\psi_{ijk}^{rst}\rvert$	very sparce	extremely sparce	extremely sparce	extremely sparce	negligible
⋮	negligible	negligible	negligible	negligible	negligible

Equation 2.28

Ideally, we would like to perform a Full Configuration Interaction (FCI), in which every possible excitations are performed. The major drawback is that, if we decide to include double and triple excitations in the multi-determinant wave function, than the number of coefficients to be optimized continues to increase exponentially with the number of basis functions that describes the molecular system (up to sixth power), and the complexity of the molecular system ends up severely limited.[160,179] Consequently, most CI calculation are limited to the explicit calculation of single (S) and double (D) excitations (CISD).

The main advantage of the CI method is that, as for the Hartree-Fock method, it is a variational method, which assures that for larger *basis sets* we will determine more accurate energies and molecular wave functions. However, it is not a size-consistent method unless the spin of each spin-orbital is allowed to change during excitation, which requires additional computational resources. As a consequence, unless this last requirement is satisfied, the sum of the energy of the isolated species in the molecular system will not be the same as for when they are infinitely apart.

2.2.4.3. Møller-Plesset Perturbation Theory

Alternatively to the CI method, which is a first-principles-based method, we can make use of perturbation theory to develop a scheme to determine the electron correlation energy. The Møller-Plesset Perturbation Theory (MPPT)[193] is an example of such application.

We already know that the Hartree-Fock Hamiltonian does not accurately describe the electron correlation energy. However, we can determine its *eigenvalues* (energies) and *eigenfunctions* (orbitals) in an accurate way, provided a limited *basis set*. Perturbation theory states that we can determined the true Hamiltonian, from the application of a series of perturbations, $\lambda\hat{\mathcal{V}}$, to the Fock operator, $\hat{\mathcal{F}}$. To determine the *eigenvalues* and *eigenfunctions* of the perturbation operator, $\hat{\mathcal{V}}$, we can think that the exact energy, E_{MPn}, and the molecular wave function, $|\psi_{MPn}\rangle$, should be accurately described by a Taylor's expansion of the Hartree-Fock minimum energy solution, as presented in Equation 2.29.

$$|\psi_{MPn}\rangle = \left|\psi_{HF}^{0}\right\rangle + \sum_{i\leq n} \lambda^{i} \left|\psi_{HF}^{(i)}\right\rangle \quad , \quad \left|\psi_{HF}^{(i)}\right\rangle = \frac{1}{i!}\frac{\partial^{i}}{\partial\lambda^{i}}|\psi_{HF}^{0}\rangle\Big|_{\lambda=0}$$

$$E_{MPn} = E_{HF}^{0} + \sum_{i\leq n} \lambda^{i} E_{HF}^{(i)} \quad , \quad E_{HF}^{(i)} = \frac{1}{i!}\frac{\partial^{i}}{\partial\lambda^{i}} E_{HF}^{0}\Big|_{\lambda=0}$$

Equation 2.29

The terms $\left|\psi_{HF}^{(i)}\right\rangle$ and $E_{HF}^{(i)}$ are new Slater determinants built from excitations of i-electrons from occupied to virtual orbitals, and their respective energy contribution to the molecular wave function $|\psi_{MPn}\rangle$.

In Møller-Plesset theory, the perturbation operator (which is the energy correlation operator) is the difference between the real electron-electron repulsion and that from the Hartree-Fock theory. Given that the powers λ^{i} are constants, we can solve each perturbation, MPi, independently for each i, determining the different contributions to the Hartree-Fock solution.

The perturbation equations in Table 2.1 show that the zeroth order term of MPPT, $MP0$ is equivalent to the result obtained from the Hartree-Fock theory. Furthermore, the first order perturbation equation (corresponding to $MP1$) shows that there is no correction to the Hartree-Fock energy when only one orbital is perturbed. This results from the fact that the perturbation operator is being applied on the Hartree-Fock wave function, thus providing for the exact same electron-electron repulsion energy that would be calculated from solving the Fock operator. On the other hand, there is non-null energy contribution, $E_{HF}^{(2)}$, from the second order perturbation ($MP2$). Here, all single- and double-excited Slater determinants, $\left|\psi_{HF}^{(1)}\right\rangle$ and $\left|\psi_{HF}^{(2)}\right\rangle$ interact with the Hartree-Fock wave function, $\left|\psi_{HF}^{0}\right\rangle$. We now recall, from Equation 2.28, that single-excitations do not interact with the $\left|\psi_{HF}^{0}\right\rangle$, and therefore do not contribute for $E_{HF}^{(2)}$; however double-excitations

exhibit a dense block of non-null resonance integrals with $\left|\psi_{HF}^{0}\right\rangle$, and they will result in a significant correction for the Hartree-Fock energy, E_{HF}^{0}. Furthermore, the $MP2$ correction does not account for triple excitations of the ground state $\left|\psi_{HF}^{0}\right\rangle$, which can only be introduced from third order perturbations ($MP3$). Despite that these interactions should not account for a large electron correlation correction, due to the sparse nature of their interaction, they can be relevant in open-shell systems.

Table 2.1. Perturbation equations for the zeroth to fourth power in Møller-Plesset theory

$\left(\hat{\mathcal{F}}+\lambda\hat{\mathcal{V}}\right)\sum_{i\leq n}\lambda^{i}\ \left\|\psi_{HF}^{(i)}\right\rangle=\sum_{i\leq n}\lambda^{i}E_{HF}^{(i)}\sum_{i\leq n}\lambda^{i}\ \left\|\psi_{HF}^{(i)}\right\rangle \quad , \quad \hat{\mathcal{V}}=\left(\hat{\mathcal{J}}-\hat{\mathcal{K}}\right)-\left(\mathcal{J}-\mathcal{K}\right)_{HF}$	
$MP0$	$\hat{\mathcal{F}}\left\|\psi_{HF}^{0}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{0}\right\rangle$
$MP1$	$\hat{\mathcal{F}}\left\|\psi_{HF}^{(1)}\right\rangle+\hat{\mathcal{V}}\left\|\psi_{HF}^{0}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{(1)}\right\rangle+E_{HF}^{(1)}\left\|\psi_{HF}^{0}\right\rangle$
$MP2$	$\hat{\mathcal{F}}\left\|\psi_{HF}^{(2)}\right\rangle+\hat{\mathcal{V}}\left\|\psi_{HF}^{(1)}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{(2)}\right\rangle+E_{HF}^{(1)}\left\|\psi_{HF}^{(1)}\right\rangle+E_{HF}^{(2)}\left\|\psi_{HF}^{0}\right\rangle$
$MP3$	$\hat{\mathcal{F}}\left\|\psi_{HF}^{(3)}\right\rangle+\hat{\mathcal{V}}\left\|\psi_{HF}^{(2)}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{(3)}\right\rangle+E_{HF}^{(1)}\left\|\psi_{HF}^{(2)}\right\rangle+E_{HF}^{(2)}\left\|\psi_{HF}^{(1)}\right\rangle+E_{HF}^{(3)}\left\|\psi_{HF}^{0}\right\rangle$
$MP4$	$\hat{\mathcal{F}}\left\|\psi_{HF}^{(4)}\right\rangle+\hat{\mathcal{V}}\left\|\psi_{HF}^{(3)}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{(4)}\right\rangle+E_{HF}^{(1)}\left\|\psi_{HF}^{(3)}\right\rangle+E_{HF}^{(2)}\left\|\psi_{HF}^{(2)}\right\rangle+E_{HF}^{(3)}\left\|\psi_{HF}^{(1)}\right\rangle+E_{HF}^{(4)}\left\|\psi_{HF}^{0}\right\rangle$
⋮	⋮
MPn	$\hat{\mathcal{F}}\left\|\psi_{HF}^{(n)}\right\rangle+\hat{\mathcal{V}}\left\|\psi_{HF}^{(n-1)}\right\rangle=E_{HF}^{0}\left\|\psi_{HF}^{(n)}\right\rangle+\left(\sum_{1\leq i<n}E_{HF}^{(i)}\left\|\psi_{HF}^{(n-i)}\right\rangle\right)+E_{HF}^{(n)}\left\|\psi_{HF}^{0}\right\rangle$

The computation of higher order perturbations results in a number of calculations that increases geometrically with the number of basis functions that describes the N-electron system (from the 5th power for the $MP2$ contribution to the 7th power for $MP4$, and so on). At the $MP4$ level, it is estimated that, for several systems, *circa* 95% of the electron correlation energy is accounted for,[160] which is quite a considerable fraction of the electron correlation energy. Moreover, MPPT is a size-consistent method,[194,195] which is an advantage in comparison to the CI method, and it is quite useful to determine dissociation and reaction energies.

However, the MPPT is a non-variational method, which means that a more complete *basis set* does not hold necessarily for a more exact solution for the problem in study. Moreover, accounting for higher order perturbation terms does not necessarily gets us closer to the exact energy of the system; for instance, accounting for third order perturbations often holds an energy, E_{MP3}, that is closer to E_{HF}^{0}. This feature resides in the conditions to which a Taylor's expansion is valid, which for electron correlation stands on the fact that this energy contribution should be very small.[160] However, electron correlation energy can be quite significant for open-shell systems and systems with small gaps between occupied and virtual orbitals, which questions to what extent does MPPT accurately describes electron correlation.

2.2.4.4. Coupled-cluster Theory

As an alternative to both CI and MPPT, we have Coupled-cluster Theory (CC).[196,197] The approach of the method is similar to that of the CI method, in which there is an excitation operator, $\hat{T}_i$, that promotes i-electrons from their occupied spin-orbitals to virtual spin-orbitals (Equation 2.27). In the CC method, the exact wave function, $|\psi_{CC}\rangle$, is expanded from the $|\psi_{HF}^{0}\rangle$, through the action of the non-linear operator $e^{\hat{T}_i}$, according to Equation 2.30.

$$|\psi_{CC}\rangle = e^{\hat{T}_i}|\psi_{HF}^{0}\rangle \quad , \quad \hat{T}_i = \frac{1}{(i!)^2}\sum_{g_1<g_i}^{occ.}\sum_{e_1<e_g\equiv}^{virt.} t_{g_{1,\cdots,g_i}}^{e_{1,\cdots,e_i}} \left|\psi_{g_{1,\cdots,g_i}}^{e_{1,\cdots,e_i}}\right\rangle \qquad \textbf{Equation 2.30}$$

The main advantage of the CC theory, is that when the $e^{\hat{T}_i}$ operator is written as a truncated Taylor's expansion, we can account for multiple excitations from one single operator, assuring size consistency (as it is also observed for MPPT). In this way, if we want to perform single and double excitations on $|\psi_{HF}^{0}\rangle$, we can expand the $e^{\hat{T}_1}$ and $e^{\hat{T}_2}$ operators, obtaining Equation 2.31, and determine a wave function, $|\psi_{CCSD}\rangle$, that also introduces contributions of higher order excitations in electron correlation energy.

$$|\psi_{CCSD}\rangle = \left(1 + \hat{T}_1 + \sum_{i>1}\frac{\hat{T}_1^{\,i}}{i!}\right)\left(1 + \hat{T}_2 + \frac{\hat{T}_2^{\,2}}{2!} + \frac{\hat{T}_2^{\,3}}{3!} + \sum_{i>3}\frac{\hat{T}_2^{\,i}}{i!}\right)|\psi_{HF}^{0}\rangle \qquad \textbf{Equation 2.31}$$

As a result, CC theory is able to provide more accurate results than the CI method with the same calculation effort than CI, for the same type of excitations.[160]

However, as for both CI and MPPT, the accurate determination of the contribution from triple excitations is also computationally unfeasible, since the number of calculations to perform scales with the 8th-power of the number of basis functions of the N-electron molecular system. Nevertheless, the inclusion of these triple excitations is often significant for the electron correlation energy in systems with pronounced non-covalent interactions.[198] In order to account for these, they are estimated from perturbation theory considerations, in what is called the CCSD(T) method.[199]

Currently, CCSD(T) provides results that are slightly more accurate than the ones determined from methods, such as MP4 or QCISD(T), thus standing as the "gold standard" in what concerns the accurate computation of the energy and wave function for a given molecular system.[200,201] In any case, CCSD(T) is a non-variational method, due to the truncation of the Taylor's expansion of the exponential excitation operator, and must be used cautiously with increasing *basis sets*. This is particularly important when energies are extrapolated to the CBS. Additionally, this "rule of thumb" holds valid only in the scope of single-reference calculations, since we have neglected the effects of Fermi correlation throughout the discussion of the CI, MPn and CC methods. In systems with significant degeneracy, a multi-reference wave function should provide for a more accurate description of the overall electron correlation energy of the molecular system, and generally Multi-Reference Configuration interaction methods (MRCI) are favoured relatively to CCSD(T).

2.2.5. Density Functional Theory: Electron Exchange and Correlation

Throughout our discussion of the post-Hartree-Fock methods as a response to the insufficiencies of the Hartree-Fock theory, we have come to the conclusion that the calculation effort to achieve such accurate results is very high. The main difficulty arises from the geometrical growth of the number of linear equations to be solved as the molecular wave function is described by an increasing number of Slater determinants. Current computational algorithms do not allow for these methods to be applied to molecular systems with more than *circa* 20 to 30 atoms. The set of systems that can be modelled under these circumstances is negligible. Most non-covalent phenomena do not fit in it, and hardly any realistic environment can be simulated with such limitation. In Table 2.2, we present the scaling of the Hartree-Fock and post-Hartree-Fock methods with the number of basis functions, N, where we can infer that calculations get more and more expensive with more complete *basis sets* and a larger number of atoms.

An alternative approach to fasten these calculations is to reduce the number of coordinates in our molecular system, thus reducing the number of two-electron integrals to solve. To address this issue, we will discuss the foundations of Density Functional Theory (DFT),[29-31] which is currently almost indispensable for any computational chemist.

Table 2.2. Scaling of the Hartree-Fock and Post-Hartree-Fock methods with the number of basis functions, N, to form the molecular wave function.

Scaling with N basis functions	Wave function methods
N^4	HF
N^5	MP2
N^6	MP3, CISD, CCSD, QCISD
N^7	MP4, CCSD(T), QCISD(T)
N^8	MP5, CISDT, CCSDT
N^9	MP6
N^{10}	MP7, CISDTQ, CCSDTQ

The basic concept underlying DFT is that: a system of N_e-electrons enclosed in a molecular volume can be approximated by a single continuous electron distribution under the influence of an external potential (the nuclear potential) with known charge and position, Z_{n_i} and r_{n_i}. However, this is only true in a quantum mechanical picture due to two fundamental theorems by Hohenberg and Kohn, the existence theorem and the variational theorem:[31]

(1) The existence theorem states that, for a given ground-state molecular wave function, there is only one density functional that presents the same energy as that of the molecular wave function, $E(\mathbf{r_1}, \cdots, \mathbf{r_{N_e}}) = E[\rho(\mathbf{r})]$. This means that this functional describes univocally the corresponding wave function, and does not depend on the form of the potential that operates on it;

(2) The variational theorem implies that, due to the fact that DFT is an exact method, the energy for a given minimum energy configuration for a finite number of basis functions should be always higher or equal to the exact minimum energy.

In this way, we have gained the ability to turn our system of N_e interacting electrons, with $4N_e$-coordinates, into a homogeneous electron gas that is only perturbed by the external potential of the nuclei, and that can be described by 3-coordinates.

2.2.5.1. Kohn-Sham formalism

We now recall the Equation 1.2 that stands for the Kohn-Sham Hamiltonian. It easily comes to mind that, as for the Hartree-Fock method, the Kohn-Sham Hamiltonian, $\widehat{\mathcal{H}}^{KS}$, can also be written as a sum of N_e one-electron Kohn-Sham operators, $\hat{\mathcal{h}}^{KS}$.

$$\hat{\mathcal{h}}^{KS} = -\frac{1}{2}\nabla_{e_i}{}^2 - \sum_{n_i}\frac{Z_{n_i}}{r_{n_i,e_i}} + \frac{1}{2}\iiint\frac{\rho(\mathbf{r}')}{r'_{e_i}}d\mathbf{r}' + \mathcal{V}_{e_i}^{XC} \quad , \quad \rho(\mathbf{r}') = \sum_{i\leq N}\left|\psi_{e_i}(\mathbf{r})\right|^2 \qquad \textbf{Equation 2.32}$$

However, when we compare the one-electron Fock operator (Equation 2.25) and the one-electron Kohn-Sham operator (Equation 2.32), we state a different approach to determine the electron-electron interactions. While in the Fock operator the electron-electron interactions are calculated from two-body operators (the Coulomb operator, $\hat{\jmath}_{e_i}\{e_j\}$, and the Fermi operator, $\hat{\mathcal{k}}_{e_i}\{e_j\}$), in the Kohn-Sham Hamiltonian this interaction is calculated from one-body operators (the Coulomb repulsion and exchange-correlation energy are calculated from the interaction of each electron with the electron density, $\rho(\mathbf{r}')$, that is generated by every-electron of the system). In this way, we have N_e-integrals less to solve for each electron, which makes DFT calculations much faster that HF's.

The calculation of the Kohn-Sham orbitals occurs similarly to what happens in the Roothaan-Hall approximation in HF theory. First, guess one-electron orbitals are generated from the LCAO approximation; then, the minimum energy for each Kohn-Sham orbital, ε_k, is determined from solving the secular equation (Equation 2.21), and the space of Kohn-Sham orbitals is formed. Then, this scheme is solved repeatedly until the coefficients matrix of the i^{th} iteration is the same of the $(i-1)^{\text{th}}$ iteration.

The main problem of DFT resides in the nature of the exchange-correlation operator. Despite that its existence is exact, its true form is not exactly known. Consequently, if the HF formalism provides exact results from an approximated Hamiltonian, DFT provides approximate results from

an exact formalism. The problem now is centred on which functional to introduce to describe the electron exchange-correlation energy as exact as possible.[180]

2.2.5.2. Exchange-correlation density functionals

The problem with exchange-correlation effects is derived from the fact that electrons are not non-interacting particles. The electron's wave function is dependent of the position and *spin* momentum of each of the other electrons in the system, and this is particularly more pronounced when we refer to electrons with similar energy. As a result from this fact, electrons will tend to diminish the probability of being close to each other, thus leading to a decrease in the Coulomb repulsion. Hence, what is required for the exchange-correlation density functional is for the local inhomogeneity of the electron gas to be introduced as a function of the interelectronic distance $\mathbf{r}'$, and the electron density $\rho(\mathbf{r}')$. In particular, we need to develop an interaction potential that can solve the self-interaction problem that arises from the Coulomb repulsion, thus preventing two electrons with the same spin from being infinitely close; a phenomenon that is known as the exchange-correlation hole. The form of the exchange-correlation potential should resemble that of Equation 2.33, being $\rho(\mathbf{r}')$ our electron density, and $g_{XC}^{\sigma,\sigma'}(\mathbf{r},\mathbf{r}')$ the pair-electron correlation function.[202] For this latter function, no analytical form is known; it is generally interpolated from quantum mechanics Monte Carlo calculations of the homogeneous electron gas model in the regions of lower and higher density limits.[203] It is evident that $g_{XC}^{\sigma,\sigma'}(\mathbf{r},\mathbf{r}')$ must possess specific properties in order to be physically meaningful: first of all, it cannot assume negative values because electrons either do ($g_{XC}^{\sigma\uparrow,\sigma\downarrow} > 0$) or do not ($g_{XC}^{\sigma,\sigma} = 0$) interact; and secondly, it must vanish when the electron density of same-*spin* Kohn-Sham orbitals is being evaluated for shorter interelectronic distances, which is a requisite of the Pauli's exclusion principle.[204]

$$\mathcal{V}_{e_i}^{XC} = \frac{1}{2}\iiint \underbrace{\rho(\mathbf{r}')g_{XC}^{\sigma,\sigma'}(\mathbf{r},\mathbf{r}')}_{\mathbf{exchange-correlation\ hole}} - \underbrace{\rho(\mathbf{r}')}_{\mathbf{self-interaction}} d\mathbf{r}' \qquad \textbf{Equation 2.33}$$

An analysis of Equation 2.33 shows that in regions with lower electron correlation the self-interaction is dominant, and the electron-electron repulsion tends to vanish. If electron correlation is high, then the exchange-correlation contribution dominates over the self-interaction term and the classical Coulomb's repulsion is more pronounced. This result has intuitive physical meaning: if electrons do not interact strongly with one another, then they should wander more freely in the electron cloud, which would explain that a lower Coulomb repulsion should be verified; otherwise,

the motion of one electron in the electron cloud should be very dependent of the instantaneous motions of the remaining electrons of the system, thus resulting in a higher local Coulomb repulsion.

In practical terms, the development of density functionals assumes that contributions from electron-exchange and electron-correlation can be accounted separately. Such approximation is not true, since we know the exact form of the exchange operator (from the HF theory), and that form is not the same as the exchange operator for a homogeneous electron gas. In reality, both the exchange and correlation density functionals are contaminated with exchange-correlation electron effects.[180] Moreover, since the homogeneous electron gas approximation is not exact to describe a molecular system, current exchange and correlation density functionals resort to a posterior parameterization of these model to reproduce different observables determined from experiment and from highly accurate post-HF calculations, such as reaction energies, vibrational spectra, atomization energies, among others.[34,180] Another practical issue is that the exchange and correlation functionals that have been developed so far, cannot be solved analytically. Instead, the volume that encloses the N_e electrons is described by a regular grid where the electron density is evaluated, so that they can be computed numerically. Once more, the fact that the molecular electron density does not follow that of the uniform electron gas approximation, particularly when we refer to valence electrons, leads to a faulted evaluation of the electron density along the grid. The fact is that a local density approximation (LDA) of the exchange, $\mathcal{V}_{e_i}^{X}$, and correlation, $\mathcal{V}_{e_i}^{C}$, operators based on Equation 2.33 is not robust enough to describe the local exchange-correlation energy of the system. The alternative is to include further knowledge of the local interaction of the Kohn-Sham orbitals with the electron density by including the gradient and kinetic energy density of the electron density in the exchange-correlation density functional.

$$\mathcal{V}_{e_i}^{XC} = \underbrace{\underbrace{\underbrace{\varepsilon_{e_i}^{XC}[\rho(\mathbf{r}')]}_{\mathbf{LDA}} + \Delta\varepsilon_{e_i}^{XC}[\rho(\mathbf{r}'),\nabla\rho(\mathbf{r}')]}_{\mathbf{GGA}} + \Delta\Delta\varepsilon_{e_i}^{XC}[\rho(\mathbf{r}'),\nabla\rho(\mathbf{r}'),\nabla^2\rho(\mathbf{r}')]}_{\mathbf{m-GGA}} \qquad \textbf{Equation 2.34}$$

In Equation 2.34, we present the different categories of density functionals according to their dependencies in $\rho(\mathbf{r}')$, $\nabla\rho(\mathbf{r}')$ and $\nabla^2\rho(\mathbf{r}')$: the local density approximation (LDA), the generalized-gradient approximation (GGA) and the meta-generalized-gradient approximation (m-GGA), respectively. Despite that these improvements in exchange-correlation density functionals were able to overcome some of the problems rose by the inhomogeneous character of the electron density in molecules, the resulting functionals were unable to satisfy all the requirements of an

exact exchange-correlation density functional. More importantly, they were not able to account properly for the self-interaction energy that arises from the Coulomb's repulsion between the Kohn-Sham orbitals and the electron density. On a final note, most of the density functionals performed very differently for different classes of systems.[160,179,180]

Part of the problem results from the fact that the electron-exchange is not entirely a local phenomenon and that the exchange density functional for the homogeneous electron gas does not entirely correct the self-interaction that arises from the interaction of the Kohn-Sham orbitals with the electron density. To tackle this issue we recall that the form of the non-local exchange operator is exactly known from HF theory. Hence, a strategy has been to combine a constant fraction of the HF exchange and the remaining fraction of the exchange density functional. The amount of HF exchange to include is determined by crossing DFT calculations of these hybrid density functionals against data sets with accurate experimental and/or high-level *ab initio* determinations.[179,180] The fitting of the amount of HF exchange to include in GGA and m-GGA density functionals has allowed for computational chemists to perform calculations with an accuracy resembling that of post-HF methods with significant reduced computation efforts. Currently, several categories of hybrid density functionals are known: the hybrid generalized-gradient approximation (h-GGA), the hybrid meta-generalized-gradient approximation (hm-GGA), range-separated hybrid generalized-gradient approximation (rs-h-GGA), and double-hybrid generalized-gradient approximation (hh-GGA). The former h-GGA and hm-GGA follow the procedure we have just described. As for the rs-h-GGA, it introduces a different fraction of HF exchange depending on the interelectronic distance function: for short-range exchange the local exchange is the dominant term in the interaction, while for long-range exchange the non-local HF exchange dominates the interaction.[205,206] The hh-GGA density functionals introduce a contribution from two-electron excitations from occupied to virtual Kohn-Sham orbitals by employing second order perturbation theory (similarly to $MP2$). In this way, there is an attempt to account for some of the non-local character of electron correlation.[207,208]

2.2.5.3. Density Functional Theory or Wave Function Theory?

Our considerations on DFT and wave function theory (WFT) have provided the main elements to discuss its applicability in the frame of theoretical and computational chemistry.

One of the main fundamental problems in DFT, in comparison to WFT, lies on the physical interpretation of the Kohn-Sham orbitals. While the solutions of WFT are one-electron spin orbitals which describe the state of each electron in the molecular system, in DFT the Kohn-Sham orbitals are non-interacting electron densities, which possess no quantum mechanical equivalent and, thus, lack physical insight other than quantitative.[160] However, pure DFT, in particular, can describe some quantum mechanics observables more accurately that HF theory, and at a reduced computational effort. This is specifically important in open-shell configurations, where HF often returns significant spin contamination in the molecular wave function.[209,210]

Additionally, despite that DFT can provide useful data in a more efficient way than WFT methods, the Kohn-Sham formalism cannot be applied to every chemical process. DFT is only applicable if the interacting N_e-particle wave function can be described as an electron density of non-interacting particles, and this is only possible for single Slater determinant wave functions. Hence, DFT has not been properly developed to describe non-interacting electron densities for excited-states chemistry or for systems with degenerate wave functions (in which Fermi correlation plays a significant role). In these cases, the results from DFT calculations should be approached cautiously, and should be closely followed with experiment of post-HF calculations.

One of the major practical problem with DFT calculations is the inexact form of current exchange-correlation functionals. Since they are approximated by ideal non-interacting systems and suffer from extensive parameterization, the variational principle that would hold true for exact DFT cannot be assured for its current implementations. Nevertheless, even though DFT may appear hardly trustworthy, the fact is that these current implementations have provided invaluable advances in chemical reactivity that would not be allowed by WFT methods.

DFT is a very robust theory for which QM calculations provide results similarly accurate to those calculated from post-HF methods. However, these are achieved at a glance of the computational effort that the latter would demand. Extensive research has been developed to benchmark the performance of several exchange-correlation density functionals against post-HF methods. These studies have allowed the comparison of the performance of DFT on thermochemistry, kinetics, spectroscopy and optimization calculations, against the performance of post-HF methods,[116,188,211-213] enforcing their application to more and more diverse chemistry. Moreover, DFT calculations have been shown to converge quickly with expanded Gaussian-type *basis sets*, in opposition to WFT methods.[160] As a result, DFT calculations can be extended to systems with hundreds of atoms, and permit that reliable and accurate computational studies can be performed.

2.3. Hybrid Methodologies

2.3.1. Expose the problem

Despite that current DFT implementations allow for QM calculations to be performed for systems ranging hundreds of atoms, extending these types of calculations to systems with thousands of atoms still requires substantial approximations. One way to look at the problem is to think that our system can be partitioned in several regions of interest, depending on their contribution to the overall chemical phenomenon. Since these regions will pose different contributions to the process, they can be described with different levels of theory regarding their relevance. Hybrid methodologies enclose this group of modelling strategies, and we will address more closely those that employ quantum mechanics/molecular mechanics calculations (QM/MM) to tackle large molecular ensembles.[15]

The main issues with QM/MM methods are concerned with the way the several layers of the system interact, the description of the boundary between each layer, and the additional computational effort that accompanies this type of calculations. Moreover, there is a large debate about *'how well can a single reference structure describe the thermodynamics of a chemical reaction, when the dimensionality of the system is that large'*.[56-58] All these problems have resulted in different approximations and several different implementations. Nevertheless, QM/MM studies have provided invaluable advances in computational chemistry and biochemistry.

2.3.2. Quantum Mechanics/Molecular Mechanics

In QM/MM approaches, the system is generally divided in a region of chemical interest, and an environment region. The region of chemical interest may comprehend a few hundred atoms and is usually studied with WFT and DFT methods, while the environment region can account for up to thousands of atoms and is often treated within a molecular mechanics formalism (similar to that in Equation 2.1). The new system can now be devised as: a model system, an environment region, and a boundary region.

Developing a way to approach the model system and the environment may seem quite straightforward, since any of the methods that we have described can be used independently for each of these regions. However, an accurate study of the system should account for the fact that the model system and the environment region do interact with each other, and, thus, their individual Hamiltonians should be different from those in Equation 1.1 and Equation 1.2.

Moreover, the treatment of the boundary must also be devised carefully. In conclusion, the Hamiltonian for a QM/MM system would resemble that of Equation 2.35: the Hamiltonian of an N-layered system, $\widehat{\mathcal{H}}_{N-\text{layer}}$, is the sum of the individual Hamiltonians of each layer, $\widehat{\mathcal{H}}_i$, that may or may not interact with each other, and the boundary region between the i^{th} and the $(i+1)^{\text{th}}$ layers, $\widehat{\mathcal{B}}_j^{i\to i+1}$. This intuitive interpretation of the problem is also known as a QM/MM additive scheme, and is widely employed with empirical biomolecular force fields.[57,58]

$$\widehat{\mathcal{H}}_{N-\text{layer}}\left|\Psi\left(r^1,\cdots,r^N\right)\right\rangle = \sum_{i=1}^{N}\underbrace{\widehat{\mathcal{H}}_i\left|\Psi\left(r^i\right)\right\rangle}_{\boldsymbol{i^{\text{th}}\ \text{layer}}} + \sum_{j=1}^{N-1}\underbrace{\widehat{\mathcal{B}}_j^{i\to i+1}\left|\Psi\left(r^i,r^{i+1}\right)\right\rangle}_{\textbf{from } \boldsymbol{i^{\text{th}}} \textbf{ to } \boldsymbol{(i+1)^{\text{th}}}} \qquad \textbf{Equation 2.35}$$

From a close inspection of Equation 2.35, we can also observe where the issues addressed in the upper section reside. We must address the shape of the Hamiltonians for each partition, $\widehat{\mathcal{H}}_i$, and each boundary, $\widehat{\mathcal{B}}_j^{i\to i+1}$. Additionally, we still need to account for the calculation effort that will result from more complex mathematical objects, as the dimensionality problem increases, which is particularly important when carrying QM calculations. These considerations on how to describe the way different layers interact and their boundaries will be subjected to a more detailed discussion in the sections below. However, we will start off by discussing in which Furthermore, despite that QM/MM methods still cannot surpass effectively the problems that concern the phase space exploration of the large chemical aggregates.

2.3.2.1. The ONIOM subtractive method

While this additive method constitutes a more intuitive way to look at the partition of the molecular aggregate, it immediately poses the problem of how to formalize the way the boundary is described, $\widehat{\mathcal{B}}_j^{i\to i+1}$. From a molecular mechanics point of view, intramolecular interactions and van der Waals can easily be accounted for through standard molecular mechanics calculations due to the fact that their effects are more pronounced at shorter distances.[57] On the other hand, and as referred in the previous discussion, Coulomb interactions vanish in a much slower way, and can thus interact unrealistically with QM electron densities. One way to circumvent this problem is to think that the QM and MM regions are fractions of a larger system in which they are still included. In particular, we will address the ONIOM methodology,[214] which is a commonly employed subtractive scheme that is based in this assumption.

We start off by postulating that the QM layer can be thought as an individual system, or instead as all that is not the MM layer, ($\mathrm{QM}_{\mathrm{layer}} = \overline{\mathrm{system} \cap \mathrm{MM}_{\mathrm{layer}}}$). The same can be applied to the MM layer, which can also be understood as all that is not the QM layer, ($\mathrm{MM}_{\mathrm{layer}} = \overline{\mathrm{system} \cap \mathrm{QM}_{\mathrm{layer}}}$). This can be generalized, and every smaller system, i, can be seen as an integrant part of a larger system, $i+1$. Such statement would hold that Equation 2.35 can be rewritten in a different way: the N-layer Hamiltonian, $\widehat{\mathcal{H}}_{N-\mathrm{layer}}$, is the sum of every i^{th} layer (that now includes the layers that are comprised within the new boundary, $\widehat{\mathcal{B}}_j^{i\to i+1}$), but now the boundary $\widehat{\mathcal{B}}_j^{i\to i+1}$ would be a term required to correct for the double-counting resultant from the interaction of the smaller i^{th} system with $\widehat{\mathcal{H}}_i$ and the $\widehat{\mathcal{H}}_{i+1}$.

$$\widehat{\mathcal{H}}_{ONIOM}\left|\Psi\left(r^{sm}, r^{lm}\right)\right\rangle = \widehat{\mathcal{H}}_{QM}\left|\Psi(r^{sm})\right\rangle + \widehat{\mathcal{H}}_{MM}\left|\Psi\left(r^{fm}\right)\right\rangle - \widehat{\mathcal{H}}_{MM}\left|\Psi(r^{sm})\right\rangle \qquad \textbf{Equation 2.36}$$

Equation 2.36 exemplifies the ONIOM Hamiltonian, $\widehat{\mathcal{H}}_{ONIOM}$, for a two-layer QM/MM model. First, the $\widehat{\mathcal{H}}_{QM}$ operator determines the energy of the smaller model $\left|\Psi(r^{sm})\right\rangle$ at the QM level, then the $\widehat{\mathcal{H}}_{MM}$ operator acts of the full model $\left|\Psi\left(r^{fm}\right)\right\rangle$ to determine the corresponding MM energy, and finally $\widehat{\mathcal{H}}_{MM}$ operates on the smaller model $\left|\Psi(r^{sm})\right\rangle$ so that the system is not described twice with different levels of theory.

The good news is that this procedure can be extended to a system of any number of layers, without loss of generality, and it only requires for a complete set of parameters to describe the full model. Furthermore, since the method is relatively straightforward to implement, the determination of analytic gradients and second derivatives follows easily, and the computation of several observable quantities can be performed similarly to the way it happens for pure QM calculations.[102] Lastly, there is no need for a specific parameterization to account for the description of the boundary between the two layers. Instead, it is automatically accounted for by allowing the $\widehat{\mathcal{H}}_{MM}$ to determine the energy of the full model, since it includes the interaction between the QM and MM regions at the MM level of theory.

The flipside of the ONIOM methodology does reside on this last aspect. Since the interaction between the QM and MM layers is solely described at the MM level, the full model must be parameterized with a robust set of MM parameters both for the QM and MM regions of the model. Moreover, rely on a MM description between the QM and MM regions should be trusted in a cautious way.[57,160] This is significantly important since the charge of the QM layer is expected to

vary significantly, if the chemistry of the QM region gets significantly different from that described by the MM parameters that were provided *a priori* to the system.

2.3.3. The interaction between different layers

We cannot expect that the QM region and MM region interact simply at the MM level, particularly when we are aware that Coulomb interactions are long range interactions. However, it is not possible to explicitly account for every electron-electron interaction in the system, since that was the fact that confronted us with the need to resort to hybrid calculations. Hence, we need to develop simplified ways that can reasonably account for the polarization effects that are expected to occur between the QM and MM regions.

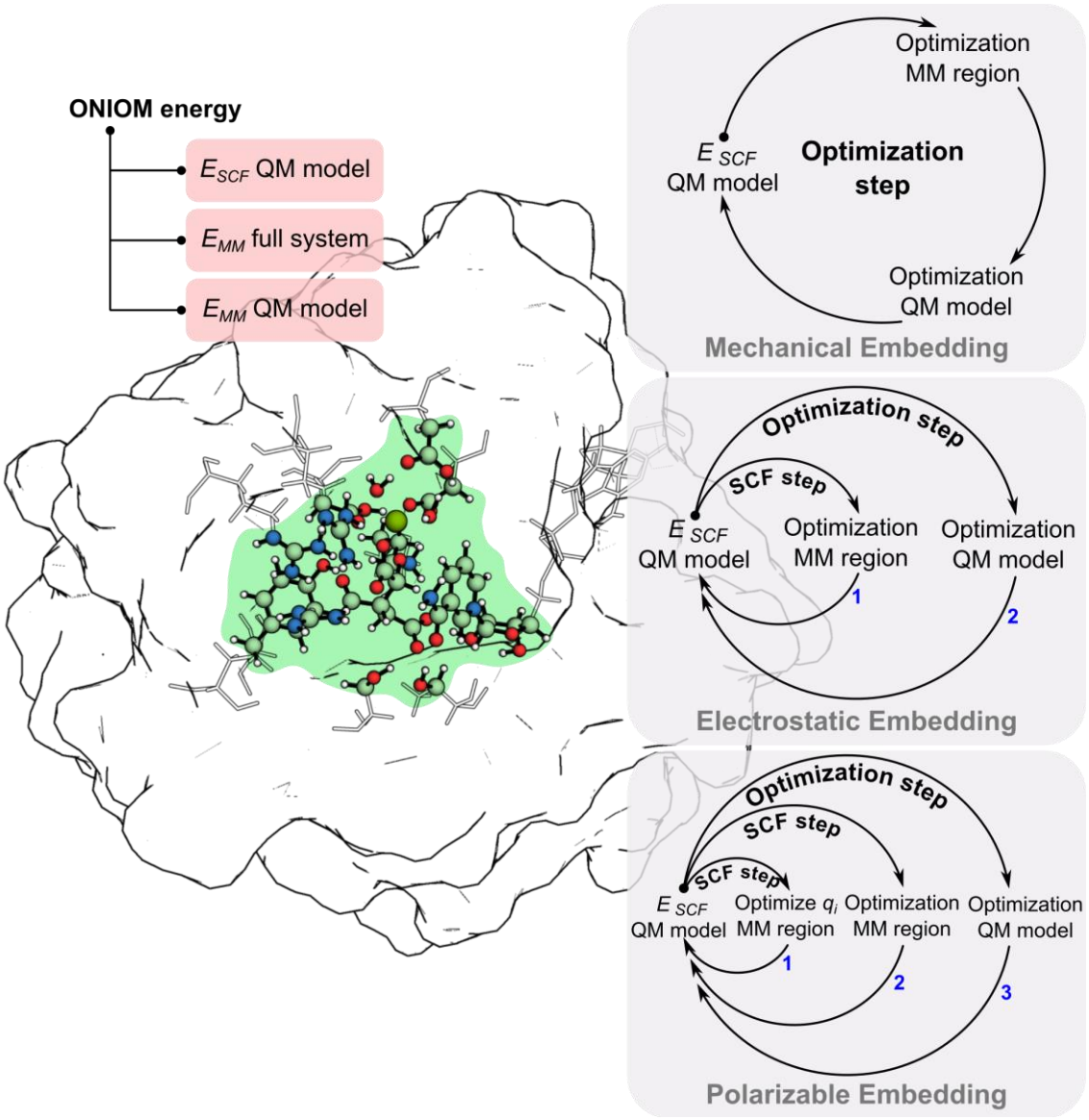
In order to search for approximate solutions, we can start by putting the problem into perspective: we can think that the electrons in the QM region do not 'see' the electron charge in the MM layer in the same way they 'see' the remaining QM electrons. There should be essentially three ways in which the MM electron charge can be 'seen' by the QM electrons:

(1) It cannot be seen at all, and the QM region is embedded in vacuum;
(2) It can be seen as an electron density that is so concentrated that it resembles a punctual charge, in a way similar to that in which QM electrons interact with nuclei, $q_{n_i^{MM}}$;
(3) It can be seen as a spherical smeared electron density, $q_{n_i^{MM}}\delta\left(r_{n_i^{MM},e_i^{QM}}\right)$.

Interchangeably, in the MM Hamiltonian, the MM electron charge interacts with atomic electron charges, not with each electron of the QM region. Hence, we ought to consider the way in which the MM electron charge 'senses' the average electron charge of each nucleus:

(1) It does not respond and the nucleus is simply attracted/repulsed to the QM nucleus;
(2) It can be polarized by the average electron charge of the QM nucleus, while being attracted/repulsed to it.

Accordingly, three main schemes are currently employed, in increasing order of computational effort: the mechanical embedding scheme,[214] the electrostatic embedding scheme,[215,216] and the polarizable embedding scheme.[217] These schemes are summed up in Scheme 2.2.



Scheme 2.2. Summary of the iterative operations in an ONIOM calculation that occur during the mechanical, electrostatic and polarizable embedding schemes.

2.3.3.1. Mechanical embedding

In the mechanical embedding QM/MM calculation, the optimization of each layer proceeds independently. One example would be the implementation of the ONIOM method that we described in Equation 2.36. The optimization of the QM layer proceeds through a SCF energy minimization employing the Hamiltonian in Equation 1.2, and the optimization of the MM layer proceeds through an energy minimization with the Hamiltonian in Equation 1.1. We are in the

situation in which the QM layer is 'blind' to the effect of the remaining environment, and the environment reacts statically to the atomic charges of QM layer (that are kept as parameterized throughout the all process). The fact that these calculations are not coupled provides that the sole difference in the computational cost, comparatively to standard QM calculations, is the MM minimization (which should be close to instant). However, this approximation is very rough, unless carefully monitored. Static atomic charges in the QM layer can be a good approximation when a minimum energy state is being characterized, since atomic charges of biomolecular force fields were parameterized to describe such states. On the other hand, when the study of chemical events is being conducted, this is hardly a good approximation, and such is even more dangerous as we approach transition states and pronounced changes in the nature chemical bonds of atoms. One way to smooth these charge effects is to update atomic charges in the QM layer through a ESP fitting, as we are exploring a given PES.[218] Another significant problem is that assuming that the remaining layers are merely a sterile bulk mass to the QM layer is often wrong, since there are several reactions in which a polarizing environment is fundamental to stabilize stationary points across a PES. To account for such effects we would require expanding the QM layer, but with compromise of the computational efficiency.

2.3.3.2. Electrostatic embedding

As a response to the inefficiencies of the mechanical embedding scheme, we can state that the QM layer does interact with the MM environment (Equation 2.37), and that, conversely, the MM layer can experience the atomic charge fluctuations that are occurring in the QM layer as atoms rearrange. In this electrostatic embedding scheme we can consider three stages: first, the QM layer is minimized through an SCF procedure in which the spin orbitals of the Slater determinant interact with the atomic charges of the MM layer (which can be either punctual or smeared charge distributions); secondly, the MM layer is optimized through minimization of the MM Hamiltonian, interacting with newly derived ESP charges in the QM layer; finally, the energy of the new configuration of the system is compared with that of the previous configuration in a self-consistent fashion. Hence we have two distinct self-consistent cycles: one corresponds to the optimization step of the QM layer, and the other corresponds to the optimization of the MM layer along with the coefficients of the spin orbitals of the QM layer.

$$\widehat{\mathcal{H}}_{EE}^{QM} = \sum_{e_i}^{e_N} \left[\hat{h}_{e_i}^{QM} + \sum_{n_i^{MM}} \frac{q_{n_i^{MM}}}{r_{n_i^{MM}, e_i^{QM}}} \right] \quad , \quad q_{n_i^{MM}} = \begin{cases} q_{n_i^{MM}} \\ q_{n_i^{MM}} \delta \left(r_{n_i^{MM}, e_i^{QM}} \right) \end{cases} \qquad \textbf{Equation 2.37}$$

The electrostatic embedding scheme present obvious advantages in comparison with the mechanical embedding scheme, since it allows for the QM layer to be polarized by the MM layer, and the other way around, self-consistently. However, the atomic charge derived in current biomolecular force fields are parameterized to describe stationary points at an MM level and, thus, are not necessarily the best approach to describe the interaction between a reactive QM region and the MM environment. In particular, the treatment of the boundary between the layers is generally the major source or error in the calculation, due to the likely overpolarization effects that can occur near this interface.[56,57,160] To tackle these effects several strategies have been developed to damp the MM atom charges near the boundary. The schemes where the MM atom charges interact with the QM electrons as smeared Gaussian distributions[216] can significantly reduce these overpolarization effects. However, the simplest implementation is to turn off the charges that are very close to the boundary, let us say *circa* one to three covalent bonds apart from the QM atoms where the system is cut.[215] Alternatively, one can also sparse the charge among the closest MM atoms in the cut region. Finally, the more radical situation would be to expand the QM layer, so that the boundary region is sufficiently far from the reaction center, thus damping the overpolarization that can be introduced by the MM charges near the boundary. Overall, the electrostatic embedding scheme is still the most employed in QM/MM calculation, assuming that its limitations are properly taken into account: (1) the use of a *basis set* with a moderate size is advised, since *basis sets* with diffuse functions interact more strongly with near MM charges; (2) the QM region should be modelled in such a way that the reaction center is not very close to regions where MM charges are high, unless these are properly scaled to account for overpolarization effects.

A natural improvement of the electrostatic embedding scheme would be to allow that not only the QM region could interact with the charges in the MM layer, as it could also polarize these MM charges. This polarized embedding scheme would probably correct for some of the overpolarization effects that can derive from the electrostatic embedding scheme. Comparatively to the MM Hamiltonian of the electrostatic embedding scheme (which is analogous to Equation 1.1), the MM charge would now be allowed to float accordingly to a set of new atomic parameters, such as atom dipole tensor or atom polarizability, aside from the simple first-order Coulomb potential.[57,219] Although this alternative should hold for an improved description of large molecular

ensembles, and this would hold true for both MD simulations and QM/MM calculations, polarized force constructs fields are still scarce in literature, and those that have been established are mostly not implemented in most computational chemistry software packages.[220-223] Additionally, the employment of polarizable force fields should require additional computation, particularly in what concerns the self-consistent charge optimization between the QM region and the MM region.

2.3.4. The boundary description

We have addressed in what way the Hamiltonian of each layer can be described at the QM and MM levels of theory. However, we now recall Equation 2.35 to pay closer attention to the Hamiltonian of the boundary between two layers described with different levels of theory, $\widehat{\mathcal{B}}_j^{i\to i+1}$. This discussion is particularly important when the truncation of a layer comprehends the cut of covalent bonds in a QM region. The problem does not reside fundamentally on the way electron-electron and electron-nucleus interactions occur, but in the chemical character of the system itself.

For a covalent bond to be broken it must occur homolytically or heterolytically. Either way, the valence of both atoms involved is altered in the process. However, truncating a layer is not a chemical process, it is simply a way to define two regions that will be treated differently. Hence, and particularly for QM calculations, we need to keep the nature of the truncated atoms by describing their valence adequately. Currently, there are three main approaches that are mainly used to tackle this issue: (1) the frozen localized orbitals approach, (2) the boundary atom approach, and (3) the link atom approach.[57]

In the frozen localized orbitals approach, we resort to the parameterization of the orbitals that describe the truncated covalent bonds to fill the valence of the atoms involved in the bond. These orbitals can be build up to resemble the truncated covalent bond between the QM and MM atoms: if they are centred in the QM atom they are parameterized to resemble a doubly-occupied orbital lying in the axis of the truncated covalent bond, and are kept fixed throughout the SCF calculations;[224] otherwise, if they are centred in the MM atom, all of the orbitals that are not involved in the covalent bond are kept fixed in the SCF calculations, while the remaining are allowed to be optimized in the SCF cycle.[225] Despite that these methods approach more rigorously the problems derived from the boundary, since they are essentially problematic at a QM level, they are harder to implement: they demand for a previous parameterization from smaller QM models, and they introduce more constraints in the SCF cycle (mainly orthogonality constraints). The result is that, in practice, they can be as effective as the link atom approach.

The link atom approach is a more common implementation to tackle the issues that arise from the QM/MM boundary.[226,227] In this scheme an atom is placed along the axis of the truncated covalent bond, between the QM atom and the MM atom. In that sense, it seems easier to implement than the frozen localized orbitals approach described above. Ideally, the hydrogen atom is the preferred choice for a link atom: it is a monovalent atom, it is small and, thus, easy to compute, and it is extensively parameterized in MM force fields. However, despite that this is highly undesirable, other link atoms can be added depending on the character of the truncated covalent bond.[228] The optimization of the link atom does not proceed in the same way as the remaining atoms. Even though its spin orbitals are treated like every other spin orbitals in the electronic Hamiltonian, which means that they are optimized in the SCF calculation, their position is kept fixed relatively to the new positions on the QM and MM atoms in the truncated covalent bond. Thus, the position of the link atom is completely determined by these new positions, excluding the three additional degrees of freedom that would normally result from introducing a new atom into the system. The major drawback of this approach comprehends the interaction of the MM layer with the QM layer, particularly when the latter is polarized by the MM layer. Since the link atom, which is present in the Slater determinant of the QM layer, is close to the MM atom in the truncated covalent bond, there is a large probability that overpolarization effects can occur near the boundary.[56,57,102] To tackle such problems, charge truncation and smearing protocols can be particularly important to minorate the effects.

2.3.5. Kinetics and phase space

Hybrid methodologies allow for the model system to be embedded in a large MM atomic environment, providing mechanical and/or electrostatic constraints to mimic the anisotropy of the chemical environment. In this way, contrary to pure QM cluster model approaches, it is expected that the model system's conformation should be closer to the expected in the biological target. Furthermore, the inclusion of the enzymatic environment, even at a lower level of theory, may provide insight to understand how and which short and long range interactions lead to higher stabilization of the transition state in a chemical reaction step [144,229-232].

One of the major reservations that has been described for the QM/MM approach comprehends the fact that the studied chemical phenomena are generally described by single structure calculations. However, the phase space of a large molecular aggregate in a given thermodynamic ensemble presents an increasingly more complex structural landscape.[233-238] To what extent is an

averaged single conformation an adequate representation of the system, is still a subject of debate.[239] The Hamiltonian employed in QM/MM methods is not affected by external potentials such as temperature or pressure, it relies simply on fundamental conservative potentials. As a result, the phase space of the system is not properly explored. This means that the points in the PES for a given set of reaction coordinates are related to the previous configurations of the system, but do not account for small thermal transitions that can shift the overall minimum energy configuration of the system as the transformation occurs in a time-dependent fashion.

Another relevant problem arises from this time-independent character of the QM/MM Hamiltonians. During the study of a PES through these methods, the QM and MM environments are allowed to be optimized through an energy minimization procedure. Since this procedure does not introduce time as a variable, we can only assert that each point in the PES is an equilibrium configuration over an ensemble of states; however, we cannot assert that the time that it takes for a given configuration to reach a new one is within that of the chemical step that is being studied. This is particularly more important when reactions take place in an atto to picosecond timescale, or when the study of non-adiabatic processes is being conducted, *e.g.* hydrogen tunnelling, proton-electron coupled transfer, or photoexcitations.[62,84]

To tackle the problem of single conformation QM/MM methods there are mainly two approaches: (1) combine QM/MM methods with MD simulations to improve the statistical representability of phase space of the system, (2) employ time-dependent potentials to account for the dynamics of the QM and MM environments during the exploration of the PES of a given set of reaction coordinates. Examples of the former are multi-conformation QM/MM,[240-246] free energy perturbation (FEP) methods, such as QM/MM-FEP,[55,236] metadynamics,[247-250] or the empirical valence bond method (EVB).[64,251,252]. Except for the EVB, which employs a semi-empirical potential to explore the chemical phase space of the system, other methods usually resort to extended MD simulations that provide the phase space exploration, and combine these results with accurate QM/MM calculations that are required to determine accurate thermodynamic and kinetic quantities. Regarding time-dependent methods, we refer to QM/MM Car-Parrinello MD,[62,253] and QM/MM MD.[59,227,254]

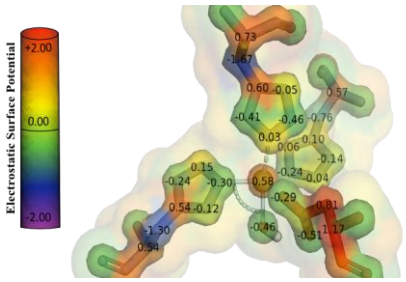
These methods have shown that activation barriers for chemical reactions may vary considerably, even when the reactive site seems to be suited for the reaction to occur. More importantly, they have allowed for further studies to be conducted in non-adiabatic regimes (where standard QM/MM approaches are not adequate). However, this gain in providing a more realistic picture of

the chemical events comes at a price. To direct computational resources to further explore the phase space of a given system, one must sacrifice either the QM region of the system or the quality of the Hamiltonian to treat it. This requires for more previous studies that can provide higher confidence levels in the approximations we are implementing. In conclusion, despite that other advanced hybrid approaches, such as multi-conformation QM/MM[242,245,246] or QM/MM MD,[240] have provided a more realistic picture of how chemical events may proceed in large molecular aggregates, single-conformation QM/MM can often provide similar qualitative information, but providing more accurate results with the same computing time.

In the near future we can expect much from hybrid QM/MM methodologies. In the age of computation and with the continuous development of methodological approaches, the potentialities are immense. Throughout this overview of QM/MM methods in for large macromolecules, we highlighted that employing polarisable QM/MM schemes and increasing sampling from QM/MM MD simulations might be the way to go. Additionally, graphical processor unit (GPU) implementation of QM calculations may increase the size of the QM models built, providing a more accurate electronic description of regions of chemical interest.

Chapter 3: Parameters for Molecular Dynamics Simulations of Manganese-containing Metalloproteins

The study of the metalloproteins with molecular dynamics (MD) simulations depends on the availability of parameters to describe the coordinating environment of the metal cofactor. In studies regarding first- and second-row metals (such as Na, K, Mg or Ca), non-bonded models have shown to provide good sampling of the phase on several metalloproteins. However, the description of the coordinating environment of transition metals requires for more sophisticated parameterization techniques. In this way, MD simulations have been shown to better reproduce several metallocentre features, such as geometry, bond covalence, Jahn-Teller effects, or number or coordination. Despite that literature offers a wide discussion on how to tackle the parameterization of metallocentres, current methodologies are still built over the approximations provided by empirical force field potentials.



$$E_{bonds} = K_\ell (\ell - \ell_0)^2$$

$$E_{angles} = K_\theta (\theta - \theta_0)^2$$

$$E_{coulombic} = \sum_{j>i} \sum_{i} \frac{q_i q_j}{\epsilon r_{ij}}$$

In this chapter, we address the parameterization of manganese metal complexes with a bonded model approach. Later we will discuss the case of the isocitrate dehydrogenase (ICDH), which is a metalloenzyme which requires divalent magnesium or manganese as a metallic cofactor. We have studied its catalytic mechanism through QM/MM calculations, and expect to employ the parameters we have derived to explore the differences in the catalytic mechanism and the thermodynamics of ICDH.

Specifically for this work, we have developed a set of parameters for twelve single manganese metal complexes, from nine metalloproteins. This set of parameters includes: equilibrium modes and force constants for Mn-ligand bonds and Mn-ligand angles, and the determination of atomic point charges with the restrained electrostatic potential charges for each of the 74 residues in the first metal coordination sphere of these metal complexes. Despite that these parameters were

initially developed to be employed in the AMBER force field, we expect that these are, ultimately, transferable to other force fields with a similar philosophy.

We have also validated these parameters by performing several MD simulations, and through a thorough statistical treatment of the former. Additionally, to validate the parameterized models, we performed frequency and normal mode calculations for each metal complex, and we compared optimized metal complexes at the QM and MM levels of theory.

We also performed linear and polynomial fittings to estimate Mn-ligand bond force constants for generic manganese centres. Furthermore, we proposed averages for the main Mn-ligand angle interactions of typical manganese coordination centres: axial, square and triangular equatorial planes, and tetrahedral positions, for the different combinations of donor atoms from waters and hard ligands.

All the calculations were performed by Rui Pedro Pimenta das Neves, as for the writing of the manuscript, which was revised through contributions of all authors. This work has been published in the Journal of Chemical Theory and Computation, and the content that follows is a mostly integral transcription of its published version.

Rui P. P. Neves, Sérgio F. Sousa, Pedro A. Fernandes and Maria J. Ramos, **Parameters for Molecular Dynamics Simulations of Manganese-Containing Metalloproteins**, *J. Chem. Theory Comput.*, 2013, 9 (6), 2718-2732. (DOI: 10.1021/ct400055v)

3.1. Introduction

Biomolecular force fields are very important tools in current computational chemistry and biochemistry, because they allow for a large sampling of the phase space of biological systems and their respective studies at an atomistic level.

Among the fundamental units of biochemistry, biological systems have ions and ligands that participate in the function or organization of these systems. In current biomolecular force fields all units are described by a set of constant parameters. In particular, metalloproteins represent a challenge in the field due to the complex behaviour of inorganic cofactors and transition-metal ions.[26,28,104,108] This translates in large difficulties to introduce general criteria for these systems and in the lack of available parameters.

Following this premise, a search for a unified force field for metal biological systems is a contemporary matter, given its current importance not only in biochemical systems but also for the nanotechnology and pharmaceutical industries.[70,103,255]

Manganese is one of the more frequent metals that figure in the available crystallographic files in the Protein Data Bank.[70,256] A search, both in the Protein Data Bank[68] and in the literature,[65,117,257] on manganese metalloproteins, shows that the main donor atoms to this metal are oxygen (from carboxylate or main carbonyl groups) and nitrogen (from imidazole rings).[70]

Manganese is found in all six classes of enzymes.[70,117] Computational studies have been conducted on the redox properties of manganese systems, such as the superoxide dismutase and the photosystem II or manganeses catalases and peroxidases.[70,257,258] These systems are found to be relevant in water oxidation or oxygen synthesis.[65,117]

Manganese is known to be generally 5- or 6- coordinated, although a typical coordination number is not defined in the literature.[65,117] Its geometries are usually derived from an octahedron. Manganese is usually in the Mn(II) oxidation state. However, in manganese clustered systems, it can be found in combinations of Mn(III)/Mn(IV) or Mn(II)/Mn(IV) and is generally involved in electron transport chains.[65] Manganese-coordinated systems are described as high-spin complexes. This occurs because manganese complexes often exhibit weak ligand fields, which lead to high-spin multiplicity configurations.[257] The high-spin multiplicity also allows one to collect atomistic level, higher resolution data since Mn ions ensure high EPR signals.[117,259] As a result, Mn can be soaked in protein crystals or co-crystallized in place of magnesium, calcium, or zinc,

which have no EPR signal response.[259] Moreover, Mn can exchange easily with Zn or Mg even in crystallized proteins.[117]

Several approaches have been described for the parameterization of metalloproteins,[28,106] all of them having their own limitations. Among the most commonly cited, there are the non-bonded,[18] the cationic dummy atom,[17] and the bonded model approaches.[16] The latter is the most complete in what concerns the description of the internal coordinates associated to the metal coordination centre.

The use of the non-bonded or cationic dummy atom approaches has been shown to be ineffective when dealing with several metalloproteins,[21,26,28,260,261] leading often to significant geometrical distortions and ligand exchange, unrealistic from a biological point of view. Bonded model approaches, although more computationally and time demanding, introduce classical mechanical parameters. In this model, metal-ligand stretching and bending modes are described by molecular mechanical potentials.[16] The flipside is that, in several biological systems coordination spheres change, especially when considering weak ligand fields, and this process is limited by this approach. Even so, bonded model approaches have allowed studies of the interactions between ligands and the metal ion centres to be made,[28,51,262-265] as well as ensuring a good structural description of the protein.

Parameters in a force field for biologically relevant metallocentres must reproduce the geometrical properties of their coordination sphere, typical metal-ligand modes, and charge density.[25,106] An all-atom approach is used the most in common force fields,[2-7] describing bond and angle interactions with harmonic potentials and intermolecular interactions (for atoms separated by more than 3 bonds) using electrostatic and dispersion terms. In addition, torsion sinusoidal terms can be added using Fourier expansions and improper constraints can be taken into account, described as harmonic potentials between coplanar non-concurrent bonds.[25,106,107] The terms that highly influence the specificity of the coordination sphere and nature of the metal, thus needing accurate description, are bond stretching and angle bending modes, and electrostatics.[16,20]

In the proposed parameterization scheme, we chose to parameterize single Mn coordination spheres with a bonded model approach.

Ultimately, correlations between the calculated parameters and their coordination spheres can be established. We will focus on properties such as sphere geometry or ligand position/combination. Such, we expect, will allow some level of transferability of the bonded-model parameters determined, for other Mn metalloproteins or force fields, not directly evaluated in this study.

3.2. Computational Details

The parameterization scheme proposed follows from the AMBER[2] force field potential formalism (Equation 3.1). The potential energy determined from the force field equation is a function of the cartesian coordinates of each atom, from which are calculated bond stretching, angle bending, torsional variables and atomic distances. The remaining terms in the equation are parameters that have to be determined either from theoretical or experimental means.

$$U(\vec{r}) = \sum_{\ell} K_{\ell}(\ell - \ell_0)^2 + \sum_{\theta} K_{\theta}(\theta - \theta_0)^2 + \sum_{\rho} K_{\rho}[1 + cos(n\rho - \gamma)] + \sum_{j>i}\sum_{i}\left(\varepsilon_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon r_{ij}}\right)$$

Equation 3.1

In the proposed scheme bond and angle parameters, K_{ℓ} and K_{θ}, and equilibrium values, ℓ_0 and θ_0, are determined from linear transit scans along the internal coordinates associated with metal-ligand interactions. Electrostatic charges, q_i, are determined from a RESP fitting of Merz-Kollman charges.

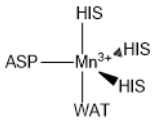
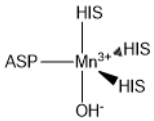
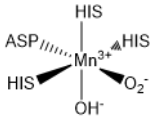
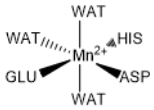
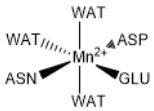
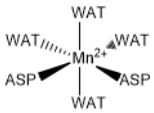
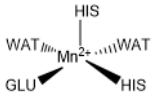
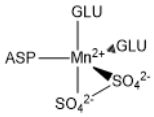
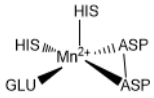
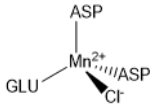
Dihedral force constants, K_{ρ}, involving the metal ion are set to zero, while transferable van der Waals atomic parameters are taken from literature.[19,20,52,266]

3.2.1. Building parameterizable models

We listed all relevant manganese metalloproteins,[65,117] defined their coordination spheres and selected the adequate crystallographic structures of those macromolecules.

From the benched manganese proteins, a set of single manganese proteins was chosen to build models smaller than a hundred atoms. Table 3.1 lists the twelve parameterized models from the crystallographic structures.

Table 3.1. List of parameterized manganese enzymes.[†]

Name	PDB code	Model code	Resolution	Geometry	Spin multiplicity	Global charge
Superoxide Dismutase (resting state)	1N0J[267]	**HHDHW (III)**	2.20		5	+2
Superoxide Dismutase[(a)] (resting state)	1N0J[267]	**HHDH[HO] (III)**	2.20		5	+1
Superoxide Dismutase[(b)] (enzyme-substrate complex)	1MNG[268]	**HHDH[HO]O (III)**	1.80		4	0
Cytochrome C Oxidase	3HB3[269]	**HDEWWW (II)**	2.25		6	0
D-Glutarate Dehydratase[(c)]	3NXL[270]	**DENWWW (II)**	1.89		6	0
Integrase	1A5V[271]	**DDWWWW (II)**	1.90		6	0
Muconate Cyclosisomerase[(c)]	3CT2[272]	**DEDWWW (II)**	1.80		6	-1
Pyruvate Kinase	2G50[273]	**ED[PY2]WW (II)**	1.65		6	-1
Homoprotocatechuate 2,3-dioxygenase	1F1U[274]	**HHEWW (II)**	1.50		6	+1
Mandelate racemase[(d)]	2MNR[275]	**DEE[SO2] (II)**	1.90		6	-3
Pneumococcal Surface Antigen Adhesin A[(c)]	1PSZ[276]	**HHE[D2] (II)**	2.00		6	0
Chloromuconate cycloisomerase	2CHR[277]	**DED[CL] (II)**	3.00		6	-2

† The first three models concern the superoxide dismutase system, then the systems are organized by decreasing coordination number and in alphabetical order. Specific codes were attributed to each model, according to the identity of its ligands. H, D, E, N stands for the one-letter amino acid code of histidine, aspartate, glutamate and glutamine, respectively; W, HO, O, PY, SO, CL stands for waters, hydroxides, superoxides, pyruvates, sulfates and chloride; bidentate ligands are identified by the number 2 associated with the codes specified above. The oxidation state is identified by the roman numbering at the end.

(a) The water ligand suggested from the crystallographic structure was treated as a hydroxide.

(b) The azide ligand, considered as a superoxide analogue,[278] was rebuilt as a superoxide ion.

(c) 3NXL and 3CT2 were obtained from crystallized structures complexed with Mg^{2+} as metal centre and 1PSZ was complexed with Zn^{2+}. For the calculations performed here these ions were replaced by Mn^{2+}, attending to Mn^{2+} exchangeability with these ions.

(d) Coordinated waters suggested by the crystallographic structure were removed, attending to the reviewed literature.[117]

The initial models were obtained from the initial crystallographic structures by selecting only the Mn ion and the residues bonded to it. Note that this process involved a set of assumptions, as the coordination shell of Mn in some of the structures could not be unequivocally assigned.[86,117,279] Coordinated water molecules are many times unresolved in the X-ray structures and sometimes the exceed the maximum number of coordination that Mn allows for (six). Literature points towards 1–3 waters coordinated in citochrome *c* oxidase[280]. In pyruvate kinase, studies refer to the existence of one water molecules,[281] but the structures that we have analysed always have two. In the structure from mandelate racemase, we observed two waters within a 2 Å distance from the Mn ion, even though they were not mentioned in the literature.[117]

To determine the force field parameters defined above, the 12 models built from the crystallographic structures were optimized at the B3LYP/SDD:6-31G(*d*,*p*) level of theory using the Gaussian09 package[282]. B3LYP[202,283,284] is currently the most used functional, yielding good results in the estimate of several thermodynamic quantities[179,285] and the use of the effective core potentials (ECPs) is a common feature in the optimization of systems with transition-metals.[51,286,287] On the latter, the SDD,[35,288] implemented in the Gaussian software, has been widely used in several studies.[19,21,51,115,255,289-291]

The initial model was relaxed using a semi-flexible model approach. In this approach, the backbone atoms (α-carbon, terminal nitrogen atom and carbonyl group) of the bonded amino acids were frozen during the optimization. In Figure 3.1, we present a scheme on the restraints imposed to the metallocentre built from the crystallographic structures approached.

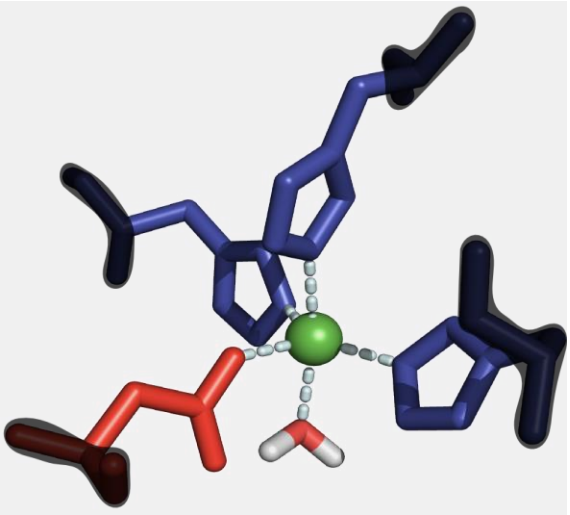


Figure 3.1. Scheme on the semi-flexible approach took towards the study of manganese metallocentres. The grayish area delimits the atoms that were frozen during the geometry optimization (α-carbon, terminal nitrogen atom, and carbonyl group).

In the treatment of ligands such as sulphates we have protonated distant basic groups. With this procedure, we intended to avoid artifacts such as deprotonations or the establishment of unexpected hydrogen bonds, because of the lack of environment. Some atoms further from the metal centre were frozen in order to avoid large changes in geometry relatively to the initial crystallographic structure, ensuring a correct description of the chemistry of the enzyme by the model.

From the geometry optimizations performed at the quantum level, for the 12 initial models considered, a good agreement between the literature predictions,[70,256,257] the X-ray structures, and the computational optimization was observed.

Figure 3.2 shows the initial optimized structures: six of the models are 6-coordinated and present distorted octahedral geometries; five models are 5-coordinated, from which three present a distorted bi-pyramidal trigonal geometry and two exhibit a distorted quadrangular pyramidal geometry; and the remaining model is 4-coordinated exhibiting a distorted tetrahedral geometry.

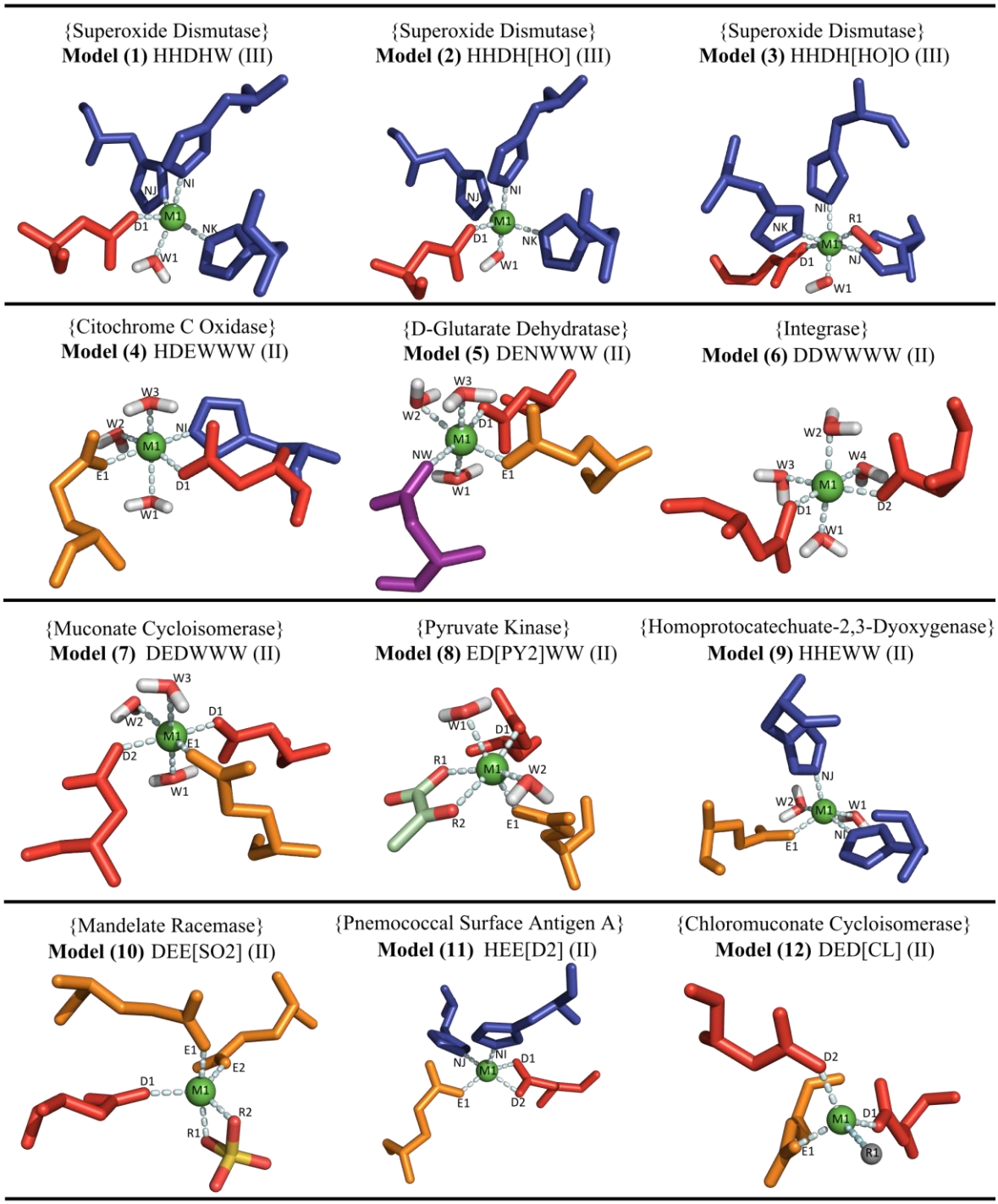


Figure 3.2. Representation of the 12 models parameterized in the presented study. The main residues are coloured in blue, red, orange, purple and green for histidines, aspartates, glutamates, asparagines and the manganese centre, respectively. Hydrogen, carbon and oxygen atoms are coloured in white, pale green and red, respectively.

In the model HHE[D2] we observed an asymmetrical bonding of the aspartate residue in a bidentate way. This model was built substituting the Zn ion for a Mn^{2+} ion, as suggested from the literature.[117,276]

3.2.2. Bond and angle parameters

In order to determine bond and angle equilibrium values to all bonds with the metal ion, a different semi-flexible approach was used, as described in Figure 3.3. Optimizations prior to the determination of the potential energy surface (PES) were performed by freezing the non-scanned ligands and the backbone of scanned ligands, at the B3LYP/SDD:6-31G(*d*,*p*) level of theory.

Figure 3.3 also illustrates the semi-flexible approach scheme used to determine the PES for each of the parameterized Mn-ligand bond and angle coordinates. In this approach we freeze all degrees of freedom except the one being parameterized, to be consistent with the independence of all terms in the force field equation.

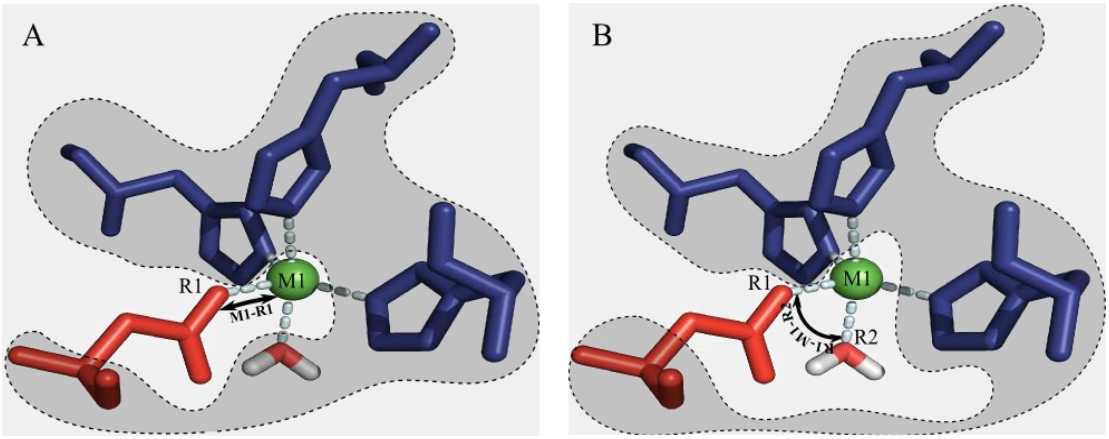


Figure 3.3. (A) Bond scan procedure and (B) angle scan procedure. In both figures, the darker areas delimit the set of atoms frozen during the optimization prior to the PES determination and in the PES determination.

The PES scans were performed with a range of 30 increments for both sideways of the PES minima, resulting in a total metric range of 0.12 Å and 30º around the PES minima. Bond and angle force constants were determined from the best fitting to the harmonic potential approximation.

All angles involving the metal ion as a terminal atom were determined using the same methodology with 20 increment steps, approaching 20º around the PES minima.

3.2.3. Atomic single charges

Atomic point charges were calculated using the Restrained Electrostatic Potential (RESP) methodology,[134] considering the Coulomb potential for the calculation of electrostatic interactions.

The standard AMBER force fields are parameterized at a HF/6-31G(*d*) level of theory.[2] The standard procedure often underestimates the effects of charge transfer with the metal atom[292,293], therefore density functional theory (DFT) is considered to be a more successful approach to calculate RESP charges.[19,20,52,130,138] Atomic charges were calculated on the optimized structures, at the B3LYP/6-311++G(3*df*,3*pd*) level of theory. A default Merz-Kollman radii of 2.00 Å was assigned to the metal centre.[135] Charges for the carbonyl and nitrogen of the amine, from the backbone of the ligands, were fixed according to the parameters established by Cornell *et al*.[2]

3.2.4. van der Waals parameters

van der Waals interactions are considered transferable. Therefore, the van der Waals (vdW) parameters were transferred from the PARM99[294] force field of the AMBER package (parameters for the water hydrogen were considered equivalent to those in hydrogen sulphide).

The van der Waals parameters of the Mn ions were transferred from the works of Babu *et al*,[110] on the determination of vdW parameters from the hydration free energies of hydrated divalent-metal cation shells.

3.2.5. Dihedral parameters

The force constants concerning torsional movements were set to zero, because it is a standard procedure in the parameterization of metal centres.[16,19,292,295] In fact, unless there is π-delocalization this assumption has been verified both experimentally and theoretically.[159] This consideration simplifies the systems' parameterization approach, because the number of dihedrals in a coordination sphere is higher than the number of bonds or metal-centred angles.

3.2.6. Validation procedure

To check how well the Molecular Mechanics (MM) parameters reproduced the geometric properties of the Quantum Mechanics (QM) systems, we reoptimized the latter at the MM level and compared the MM and QM geometries. We have also compared the MM and QM vibrational frequencies. In addition, we checked if the MM parameters could reproduce the geometries of the

Mn center within an enzymatic scaffold. For that purpose, we ran 12 MD simulations of nine different metalloenzymes containing manganese. MD simulations were performed using the AMBER10 package.[296]

Letter and number codes were established to label the residues parameterized (three-letter code) and the bonding atoms (two-letter code). The labelling criteria of parameterized residues and coordinating atoms is presented in Supporting Information (SI).

For each of the 12 biological systems studied, we built models from the crystallographic structure of the protein, constituted by at least one monomer of the unit, to validate the parameters determined. The mechanical parameters of the organic ligands coordinated to the metal centre were parameterized using the ANTECHAMBER module[297] of AMBER10, with the GAFF force field.[298] All remaining ligands and cofactors not included in the coordination sphere under study were removed from the initial crystallographic structures. We added counter ions to the system in order to neutralize its global charge and created a rectangular solvent box with TIP3P waters,[299] so that the distance between the protein and the edges of the box was set to 12 Å. This value fits the average conditions commonly used in several other studies with a same prospect,[19,20,138] since the *cutoff* for the discrete account of vdW interactions is generally within the range of 9-15 Å.[300]

For the built systems, 10 ns of MD were performed, preceded by a four step minimization, to minimize waters, added hydrogens, residue side chains, and all the atoms of the system. An equilibration step of 40 ps was applied, in order to gradually heat the temperature of the system to the standard temperature of 310 K, at a constant pressure of 1 bar. MD proceeded on an NPT ensemble. A constant temperature of 310 K was maintained by the use of Langevin Dynamics implemented in AMBER10 package and a fixed pressure of 1 bar was set. Periodic boundary conditions were imposed to the models built to account for long range interactions. Short and long range interactions were calculated with Ewald summation methods.[155] The SHAKE algorithm[165] was applied in all bonds involving hydrogen atoms, for a mechanical relaxation time of 2 ps. To calculate long range vdW interactions, a *cutoff* distance of 10 Å was established.

Different scaling factors for the 1–4 electrostatic interactions were tested, resulting in a final scaling factor of 60% in all MM calculations applying the parameters determined.

MD results were obtained with the PTRAJ module of the AMBER package. Values analysed included the root-mean-square deviations (RMSds) for the α-carbons of all the models, and for all the atoms in the metal coordination sphere. Average and standard deviation values for all the bonds and angle involving the metal atoms were taken from the final 4 ns of MD simulations.

3.3. Results and Discussion

3.3.1. Bond stretching parameters

From a total of 64 bond force constants determined, 14 of them characterize Mn–N bonds and the other 50 characterize Mn–O bonds. Figure 3.4 illustrates the PES obtained for all the Mn-ligand bond scans.

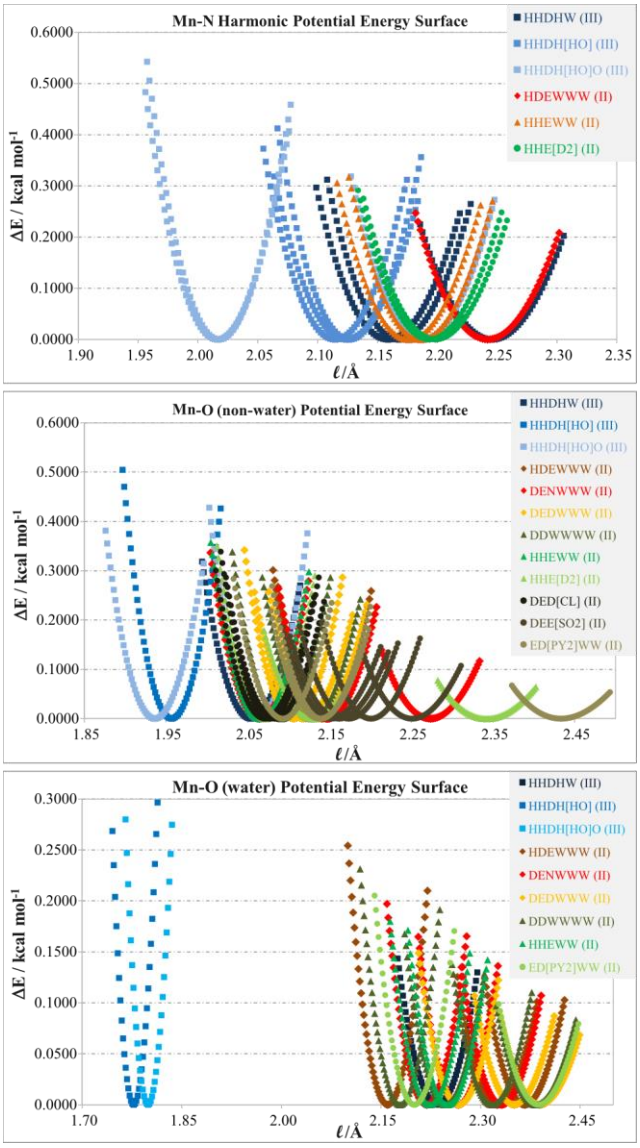


Figure 3.4. Potential energy surfaces for bond stretching for all models.

All the equilibrium bond lengths are found in the interval]2.00; 2.45[Å in all systems. However, for smaller charged ligands this value can be smaller, as seen in the models with the hydroxide ion. For the studied amplitude in bond stretching energetics, increments below 0.5000 kcal·mol^{-1} are frequent. Most of the force constants are in the 60–80 kcal·mol^{-1}·Å^{-2} range.

We can observe similarities in the PES when the equilibrium bond lengths are 2.15 Å–2.20 Å, for nitrogen donor atoms, and ~2.10 Å and ~2 .30 Å for non-water and water oxygen donor atoms, respectively. These observations are in qualitative agreement with the general knowledge regarding transition-metal chemistry and the expected relative magnitude of the corresponding force constants.[70,257]

For similar bond lengths, histidines exhibit higher force constants than glutamates or aspartates. Waters are the weakest ligands, with larger typical bond lengths and shallower PESs.

It is important to highlight, because of the large amount of information in literature concerning the manganese superoxide dismutase,[86,257,278,301] some relevant discussion from the determined parameters. Conformational changes between the two models of manganese superoxide dismutase occur mainly in the axial ligands, when the hydroxide ligand is present instead of the water ligand. As the hydroxide ligand deepens into the coordination sphere, we observe that the axial histidine approaches the metal centre and its force constant increases. By bonding with the superoxide substratem the model becomes 6-coordinated with distorted octahedral symmetry and exhibits a Jahn-Teller effect along the NI–W1 axis, as described in the literature.[257]

The analysis of the internal bonding terms also suggests that similar coordination spheres do not differ substantially either in bond length or bond force constant.

3.3.2. Angle bending parameters

From the total parameterized 146 metal-centred bond angles (the detailed full set of parameterized metal-centred angles is provided in SI) their distribution is as follows: 11 for N–Mn–N, 22 for N–Mn–O, 16 for N–Mn–water, 31 for O–Mn–O, 46 for O–Mn–water and 17 for water–Mn–water (in this distribution hydroxides have been counted together with waters).

Analysis of Figure 3.5 shows that angle force constants are often <75 kcal·mol^{-1}·rad$^{-2.}$ The cases in which this is not verified mainly concern interactions with histidines.

The bidentate ligands studied exhibit force constants >100 kcal·mol^{-1}·rad^{-2}, and >200 kcal·mol^{-1}·rad^{-2} for the bidentate ligands whose donor atoms are bonded to the same atom.

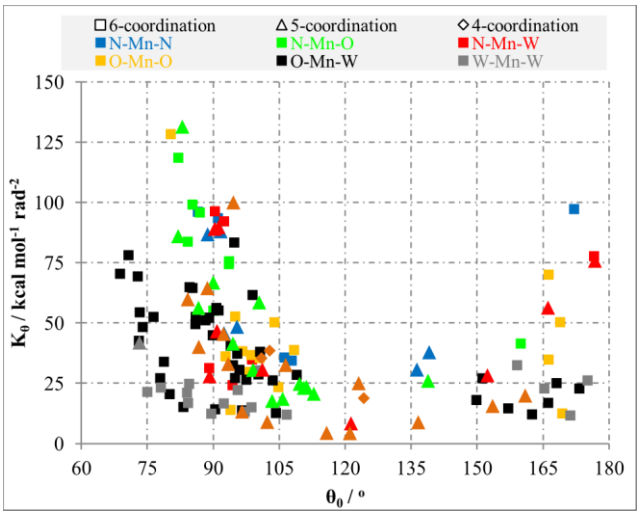


Figure 3.5. Angle force constants and equilibrium angles from the scan of the studied models, by donor atom type and coordination number.

For 6-coordinated models, interaction between non-water donor atoms and water oxygen atoms is frequent, establishing mainly orthogonal interactions with force constants abundantly between 25 and 75 kcal·mol^{-1}·rad^{-2}. In the axial positions force constants are 25 kcal·mol^{-1}·rad^{-2}.

Regarding the 5-coordinated complexes, two possible geometries are observed: bipyramidal trigonal and square pyramidal. In the triangular equatorial plane, interactions with nitrogen donor atoms are more frequent and with low force constants (<25 kcal·mol^{-1}·rad^{-2}), while for the axial positions they are generally near 25 kcal·mol^{-1}·rad^{-2}. In square pyramidal geometries, most interactions are observed from histidines with force constants slightly below 25 kcal·mol^{-1}·rad^{-2}.

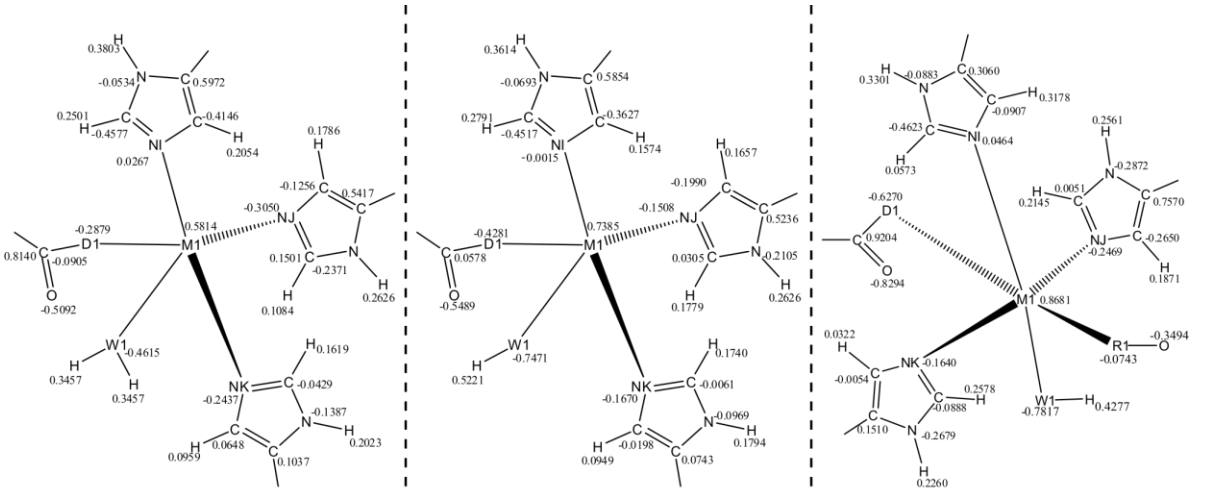
The 4-coordinated model has a distorted tetrahedral geometry with force constants slightly above the 25 kcal·mol^{-1}·rad^{-2} for angles of 109.5º.

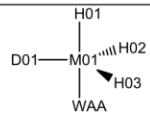
Analysing the angle force constants we can conclude that their amplitude is not strongly concerned with the geometrical position of the ligands for the equilibrium bending amplitudes frequently observed, despite some geometrical similarities observed in models (4) to (7) (recall Figure 3.2). Looking further into the donor atom type, it seems that oxygen donor atoms from aspartates and glutamates establish stronger angle bending interactions with waters than N atoms from histidines, when in equatorial positions; however for orthogonal interactions with axial waters, higher angle force constants are generally observed with histidine ligands.

3.3.3. RESP charges calculation

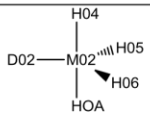
Single-point electrostatic potentials were calculated using the large 6-311++G(3*df*,3*dp*) basis set.

Figure 3.6 shows the charge schemes for the donating groups of the ligands in the models from manganese superoxide dismutase (models (1) to (3)) and the global charge of each residue from their coordination sphere (a detailed description of the RESP charge maps for the 12 metallocentres parameterized is presented in SI). Overall, RESP atomic charges were determined for 74 residues, in the 12 models used for parameterization.

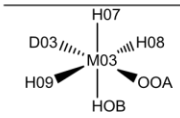




Charge of residue	
M01	0.581
D01	0.005
H01	0.404
H02	0.416
H03	0.363
WAA	0.230



Charge of residue	
M02	0.739
D02	-0.517
H04	0.261
H05	0.381
H06	0.361
HOA	-0.225



Charge of residue	
M03	0.868
D03	-0.724
H07	0.227
H08	0.192
H09	0.215
HOB	-0.354
OOA	-0.424

Figure 3.6. Representation of the single point Merz-Kollman charges fitted with the RESP methodology for three models concerning manganese superoxide dismutase, and global charge of the residues for the three models derived from manganese superoxide dismutase. Charges are written in atom units (au).

As we can see in a dispersion pattern for manganese and its main biological ligands, in Figure 3.7, there is a large dispersion in the global charges of the residues, mainly on the Mn ion. That might concern the fact that the Mn ion is buried within the coordination sphere by its ligands, which is consequence from the ESP methodologies extensively described in literature.[128,135] An average estimate seems to agree with the classical expected zero-charge only for waters.

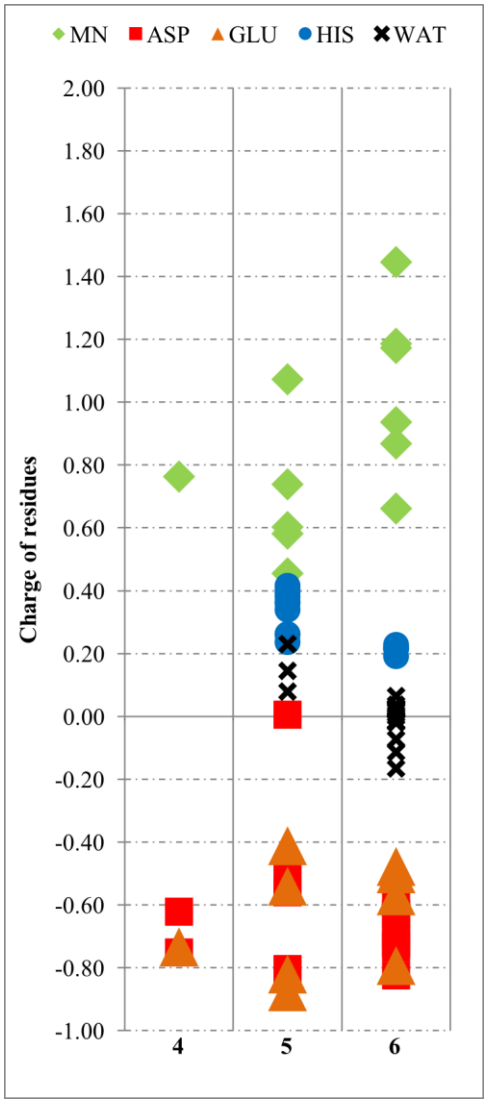


Figure 3.7. Global charge distribution for the main residues of the studied models, by residue and coordination number of the model.

3.3.4. Parameter Validation

A validation protocol was designed to verify the stability of the protein backbone and the coordination sphere. We employed MD and MM calculations to generate data populations and normal mode information for each of the twelve models.

The scaling factor for the 1–4 electrostatic interactions was defined for classical mechanics calculations in order to reproduce better the equilibrium geometries, determined by quantum calculations in vacuum (validation of the electrostatic scaling is supplied in the SI).

3.3.4.1. RMSd values

The results for the RMSd of the models parameterized are shown below (RMSd for the protein backbone chain and the Mn centres are presented in SI). Figure 3.8 shows the RMSd plots for 4 of the 12 parameterized metallocentres (models (4) to (7)) during the 10 ns MD simulation.

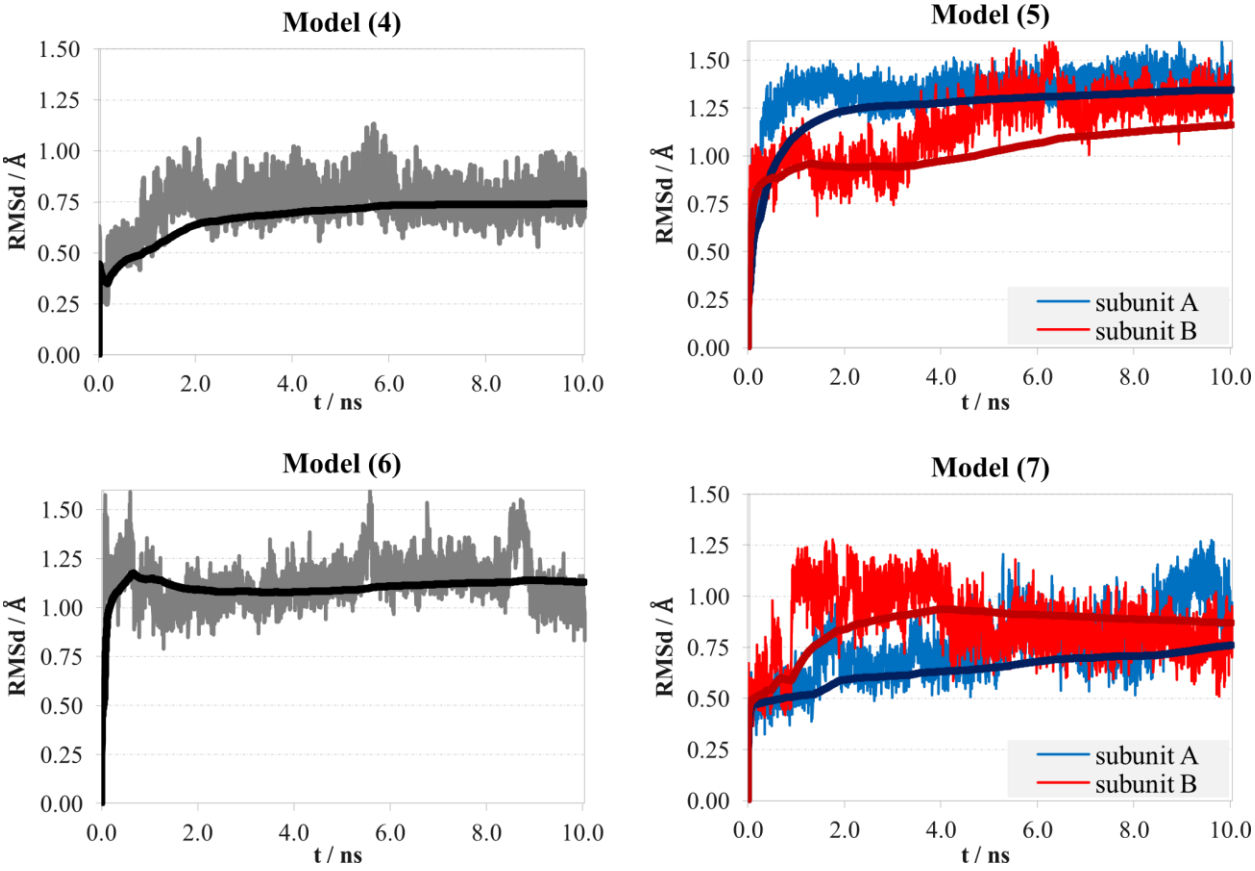


Figure 3.8. RMSd graphics (in thin lines) and the accumulated averages (thicker lines) for the models (4) to (7), during the 10 ns of MD simulation.

An initial overview shows that, in the majority of the models, the backbone stabilized within the first 6 ns of the MD simulation.

The RMSd values of the Mn centres parameterized are generally smaller than 1.25 Å and, as for the accumulated averages, are generally stabilized within the last 4 ns of MD. Therefore we defined this time range as the scope of our statistical analysis.

In Table 3.2 we present the RMSd averages and standard deviations from the last 4 ns of the MD simulations. RMSd values are higher than expected. However, we also observed significant differences between the crystallographic and the respective quantum optimized coordinates. The differences between the QM and MM model are generally smaller. Moreover, in a few cases, we exchanged the metal in the X-ray structure for Mn and, in those cases, differences between the bond length of the experimental and calculated structures are expected.

Table 3.2. RMSd averages and standard deviations for the proteins (right column) and the small models (left column), in Å

Code of the model	$\langle \mathbf{RMSd} \rangle \pm \sigma_{\mathbf{RMSd}}$ metallocentre	$\langle \mathbf{RMSd} \rangle \pm \sigma_{\mathbf{RMSd}}$ protein
HHDHW (III)	1.00 ± 0.09	3.3 ± 0.2
HHDH[HO] (III)	1.21 ± 0.09	2.7 ± 0.2
HHDH[HO]O (III)	0.41 ± 0.08	2.3 ± 0.2
HDEWWW (II)	0.76 ± 0.07	2.9 ± 0.1
DENWWW (II)	1.35 ± 0.09	3.5 ± 0.2
DDWWWW (II)	1.2 ± 0.1	2.9 ± 0.3
DEDWWW (II)	0.8 ± 0.1	2.2 ± 0.1
ED[PY2]WW (II)	0.6 ± 0.1	3.4 ± 0.3
HHEWW (II)	0.7 ± 0.2	1.9 ± 0.1
DEE[SO2] (II)	0.07 ± 0.02	1.6 ± 0.1
HHE[D2] (II)	0.41 ± 0.06	2.5 ± 0.1
DED[CL] (II)	0.47 ± 0.06	2.9 ± 0.1

We emphasize, nonetheless, that crystallographic structures are estimations over electronic density determined maps and differ among them.[302,303] Therefore, deviations from the simulations may be analysed taking into account additional considerations, such as mechanistic information on the coordination sphere of the metal ion, which allows further detail than the crystallographic structures alone.

3.3.4.2. Bond stretching and angle bending coordinates

The results of the bond lengths between the metal and its ligands, taken from the crystallographic structure $\ell_{0\ \mathrm{crystal}}$, the quantum mechanical optimized model $\ell_{0\ \mathrm{opt}}$ and the average from the 4 ns MD population $\langle \ell \rangle_{\mathrm{MD}}$, are presented in the SI.

The standard deviations of the 4 ns bond length population are proximate to the maximum stretching scanned in the PES calculation. However, this has not been verified for most of the water ligands. Considering the range $\langle \ell \rangle_{\mathrm{MD}} \pm \sigma_{\ell}$ for each of the bonds established, 82% of the optimized Mn-ligand bonds are included in the given range. For twice this range, 98% of the optimized bond equilibrium values parameterized are included in the given interval.

For the same given interval $\langle \theta \rangle_{\mathrm{MD}} \pm \sigma_{\theta}$ only 66% of the parameterized angles fit in the range defined. However, standard deviations determined are often <5º, which is a low value if attending to the flatness of the PESs for these small force constants. For an error estimate of 1 kcal·mol^{-1}, an angle bending often higher than twice the standard deviation is required. Considering a new range $\langle \theta \rangle_{\mathrm{MD}} \pm 2\sigma_{\theta}$, 89% of the optimized equilibrium angle bending values are included in the defined interval, for the determined force constants.

To analyse the information on the population studied, quartile and histogram distributions were produced. Quartile distributions allow a glimpse of the symmetry of the population and the range of the full sampling, while histogram/normal distributions allow further considerations on the flatness of the population and the comparison of the optimized and averaged values.

Figure 3.9 and Figure 3.10 present the quartile distributions and normal curves for the bonds and angles involving the Mn atom for model (6), as an example. Quartile and Gaussian distributions are presented in detail in the SI, as well as further validation of the Gaussian curve fitting.

A quartile distribution represents a distribution of the population in blocks of 25% of the data from the last 4 ns of MD simulation, in an increasing order. The first and last 25% of the population values (the first and fourth quartiles) are represented as the tails of the distribution, while the second and third quartiles are represented by boxes, divided by the central value of the population (the median value). From the quartile distribution, we can see the general range of the distribution and the concentration of the sampling around the median value.

The normal curve fitting, validated by the histogram representation, estimates the predominance of each single coordinate in the population.

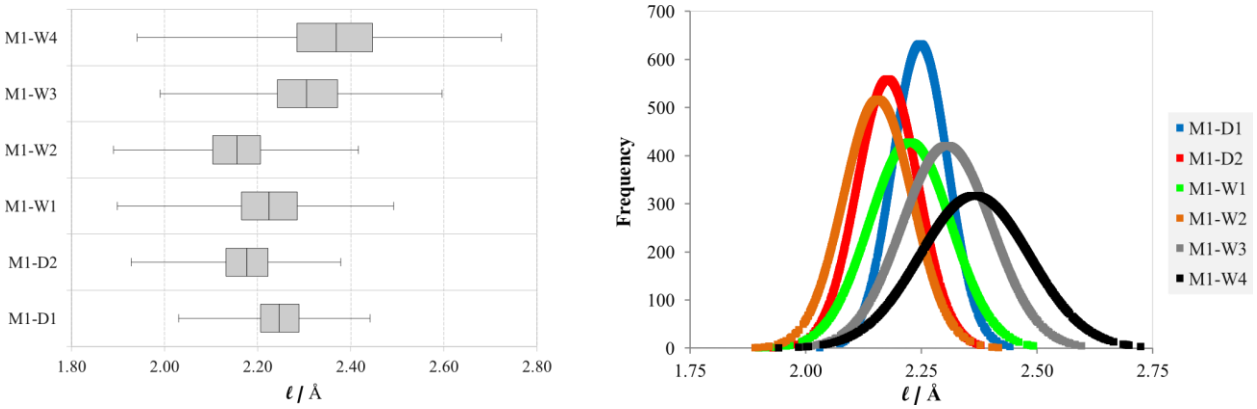


Figure 3.9. Distribution for the bonds of the model (6) with quartile representation and Gaussian curve fitting, for the last 4 ns of the MD run. All the representations and further validation on the gaussian fitting are in SI.

To establish the histogram representation, data on the final 4 ns population was organized in intervals with a range close to the standard deviation values obtained (0.05 Å for bonds and 5º for angles).

Generally, quartile and histogram plots for both bond and angle coordinate show that the median and average values are close, thus supporting the average values determined. The ranges of the boxes around the median, from the quartile representation, point towards an apparent symmetry in the considered populations. A more spread distribution of the bond length values can be observed for farther Mn-ligand distances. Even though only 69% of the QM optimized bond lengths are counted in the two quartiles around the median, these quartiles, however, are on average spread around 0.08 Å from the median.

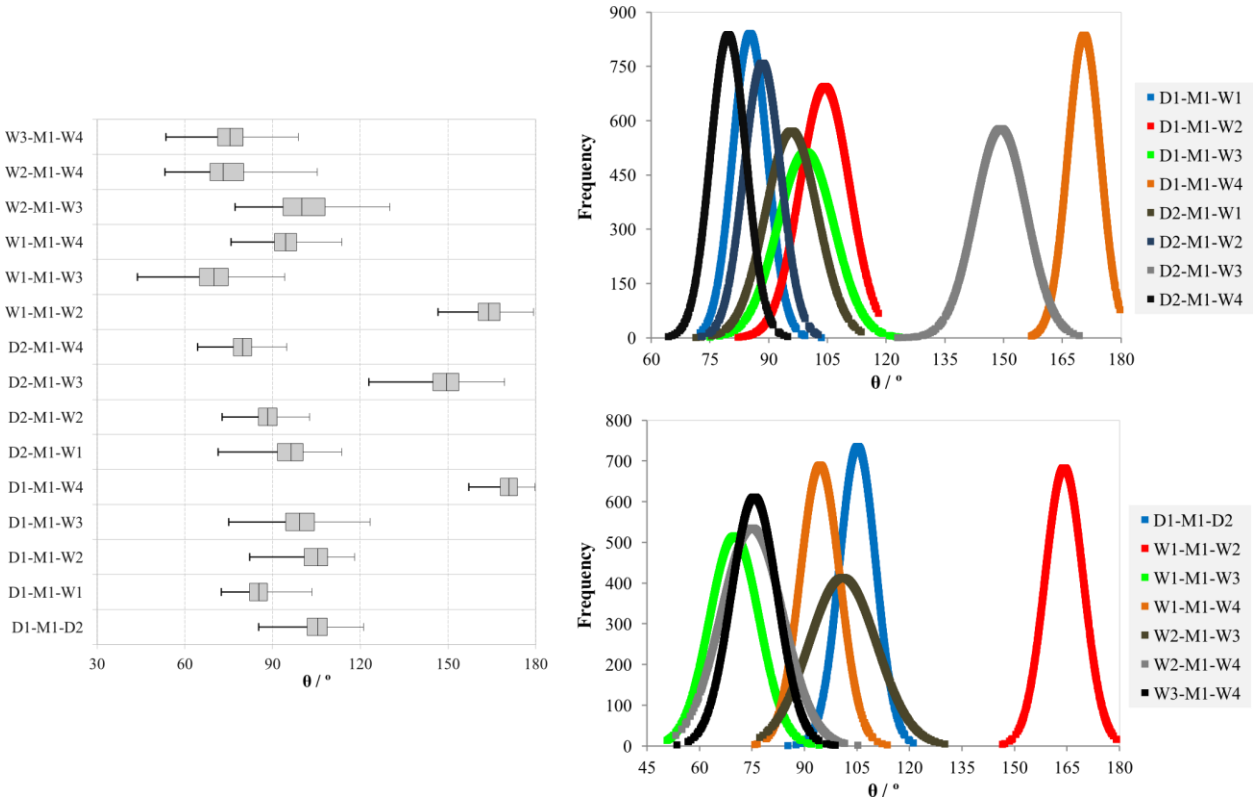


Figure 3.10. Distribution of the data for the angles in model (6) with quartile distribution and approaching a Gaussian curve, for the last 4 ns of the MD run, shown as an example. Other distribution data can be found in the SI.

From Figure 3.10 we observe that, for similar angle bending amplitudes, populations involving O–Mn–water and N–Mn–water interactions have larger standard deviations than the remaining interactions. The length of both boxes accounting for 50% of the population, around the median is, on average, 6º. However, the normal fitting curve shows that angle populations have flatter distributions than quartile plots might resemble. In Gaussian curves, it can be seen that, in most cases, for a range of 8º–10º, only 67% of the angle population would be found around the average value. Since the average and median values are close, it means that, for an increase in range up to 4º, only an additional 17% of the population is included.

All the statistical validation seems to point towards good reliability in the parameters determined.

3.3.4.3. Frequency and Normal Mode Analysis

Comparison of the frequencies of the parameterized models and the original QM models shows that the parameters calculated reproduce the frequency profile expected from quantum

mechanics calculations. Figure 3.11 shows a graphical comparison of the QM and MM vibrational frequencies for one of the parameterized models.

This, however, does not guarantee that normal modes for both models match. Particularly, it would be helpful to find equivalent normal modes in both models and compare their mass-weighted frequencies. Therefore, we show, for four of the twelve models (models HHDH[HO] (III), HDEWWW (II), HHE[D2] (II) and DED[CL] (II)), the correlation for MM vs. QM frequencies, describing similar normal modes. We searched normal modes for which displacements of metal-ligand bonds were similar in QM and MM models.

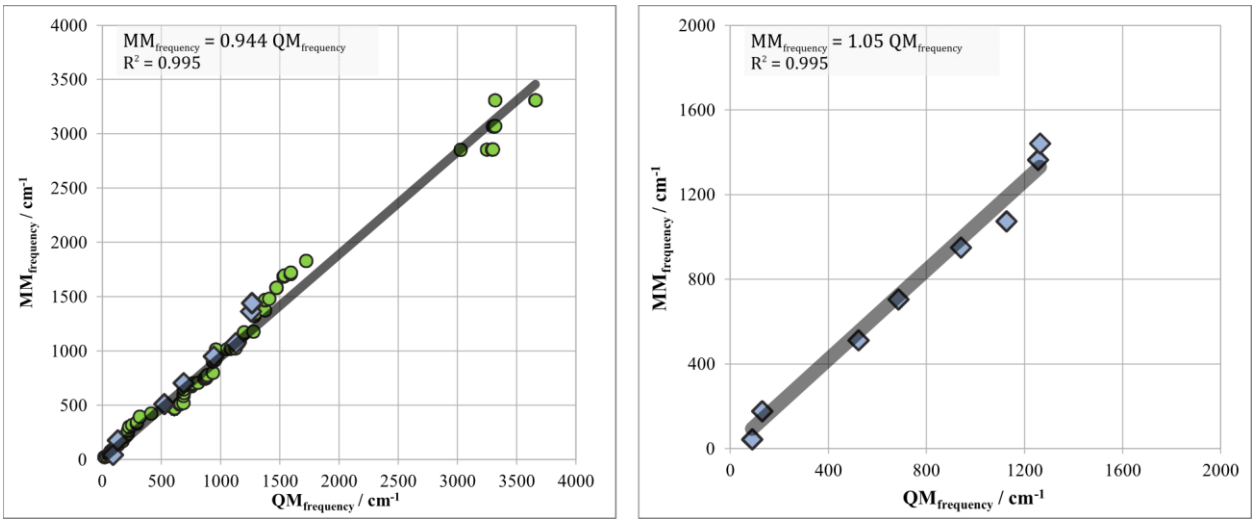


Figure 3.11. Comparison of vibrational frequencies for the model HHDH[HO] (III) optimized with the AMBER force field and at B3LYP/SDD:6-31G(*d*,*p*) level of theory (on the left) and normal modes involving the donor and acceptor atoms (on the right).

We calculated first the RMSd between the MM optimized models and the QM optimized models (detailed plots follow in the SI). Large differences are registered in the backbone atoms and in the carboxylate oxygens not coordinated to the manganese centre. Since the backbone atoms were not targeted by our parameterization scheme, we performed a RMSd calculation based on the heavy atoms of the organic ligands and side chains of bonded amino acids, as can be seen in Table 3.3 RMSd values for our parameterized models are smaller than 1.4 Å.

The differences between the bond lengths calculated in the QM and MM geometry optimizations ranged from 0.002 Å to 0.20 Å (average 0.04 Å).

Table 3.3. RMSd for the optimized models, in Å, from the comparison of the QM and MM optimized structures.

RMSd / Å	Heavy Atoms	Heavy Atoms not from backbone
HHDHW (III)	1.22	0.75
HHDH[HO] (III)	1.83	0.92
HHDH[HO]O (III)	1.92	1.07
HDEWWW (II)	0.90	0.51
DENWWW (II)	1.73	1.10
DDWWWW (II)	1.55	0.50
DEDWWW (II)	1.82	0.98
ED[PY2]WW (II)	1.88	0.87
HHEWW (II)	1.61	1.18
DEE[SO2] (II)	1.92	1.06
HHE[D2] (II)	2.81	1.38
DED[CL] (II)	1.80	1.04

3.3.4.4. Correlations between the Parameters

Beyond the scope of the parameter determination, we expect our results to lead to some estimates on parameter behaviour in manganese-containing proteins. It is expectable and desirable that the parameters would be transferable for systems with similar coordination spheres. Moreover, we also expect to provide estimates in force constants concerning typical donor atoms or geometrical dispositions in Mn-coordinated centres and, ultimately, save computational resources in future projects involving this class of proteins.

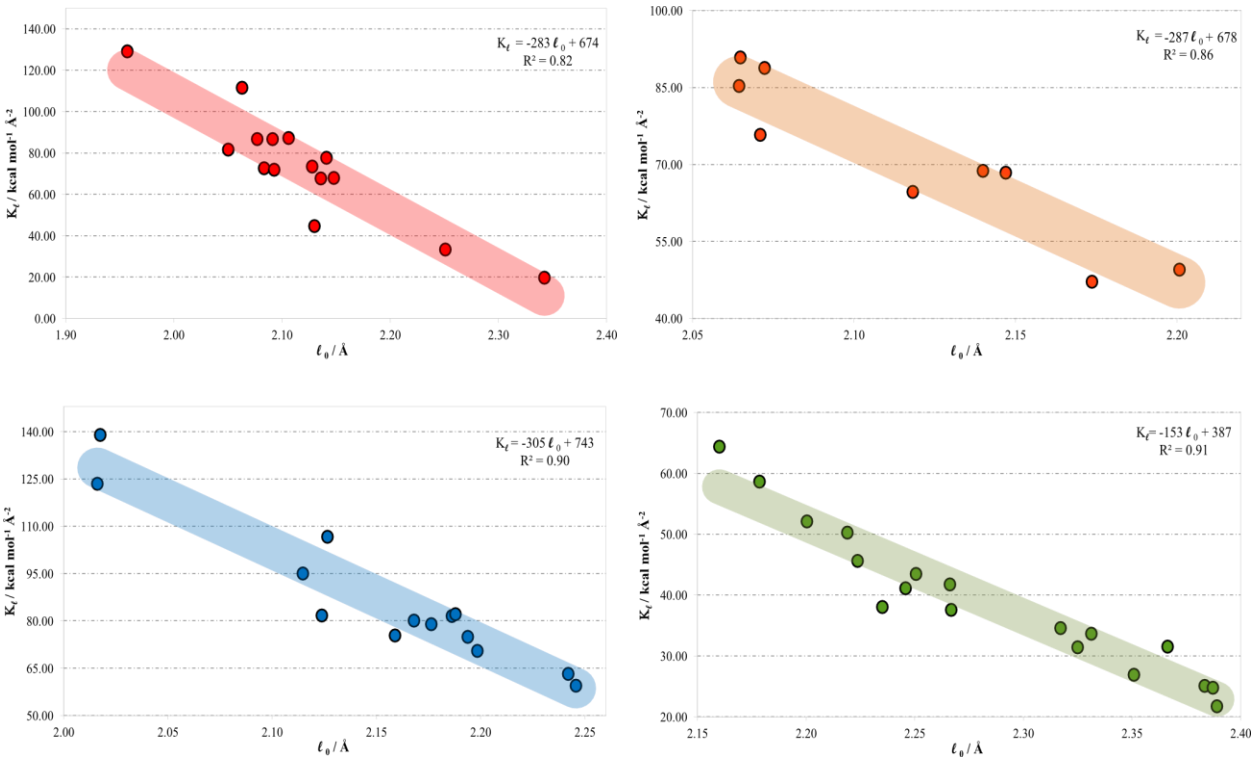


Figure 3.12. Linear regression for the main ligands in manganese coordination spheres: aspartate, glutamate, histidine and water, in red, orange, blue and green, respectively.

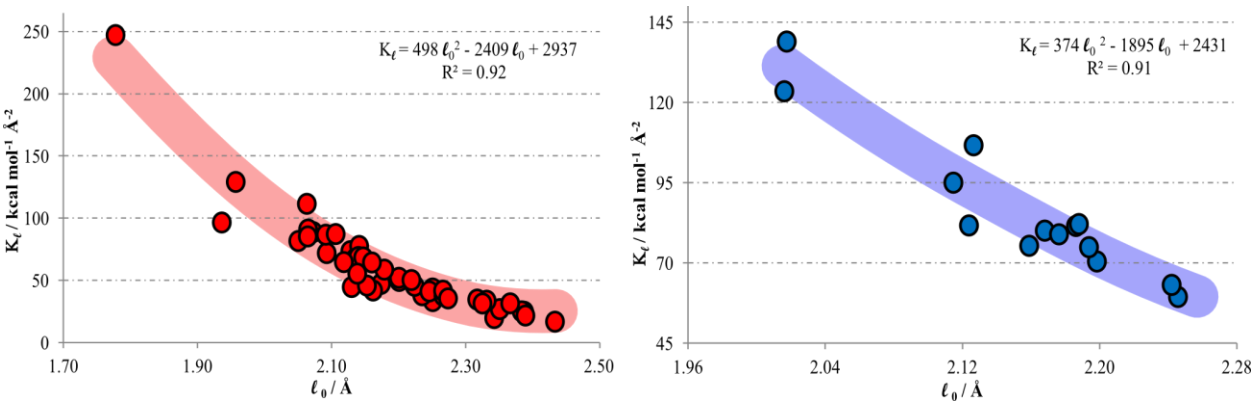


Figure 3.13. Second order polynomial regression for the main donor atoms in manganese coordination spheres, oxygen and nitrogen, in red and blue colours, respectively.

Figure 3.12 and Figure 3.13 illustrate all the 65 bond force constants, taking into account the equilibrium bond lengths of the corresponding interactions. These are grouped by specific Mn-ligand type interaction (by residue) and donor atom (by element). Regarding bond force constants,

linear and second-order polynomial regressions were performed to establish correlations for force constants and equilibrium bond lengths on ligands and donor atoms.

Waters are small, flexible, and distant ligands in their coordination with the metal centre; therefore, they are difficult to predict. For the 18 bonds between the Mn ion and waters, a good sampling was obtained and an average correlation between bond force constants and bond lengths was established. Therefore, rough force constants can be estimated for any water in a given coordination sphere. Such correlation seems to be unrelated to the charge of the manganese centre or the coordination.

We calculated correlations for the remaining frequent ligand amino acids; however, sampling is lower and generally has a wider distribution, particularly for aspartates and glutamates.

The linear regressions established seem to support similar bonding behaviours for aspartates and glutamates. Extending the fitting to large bond lengths, we should expect that the force constants converge to zero; therefore, force constants must be obtained by *interpolation* rather than *extrapolation*. A range of 2.00–2.30 Å and 2.00–2.20 Å is defined for the aspartate and glutamate bonds, respectively, whereas for histidine, an interval from 2.00 Å to 2.25 Å is suggested, and for water, an interval from 2.15 Å to 2.40 Å is suggested.

To estimate force constants for less common ligands, the second order polynomial regression plotted in Figure 3.13 can be used. Although good correlations for rough estimates have been attained, the estimated force constants can be affected by errors in the range of 17 kcal·mol^{-1}·Å^{-2} and 8 kcal·mol^{-1}·Å^{-2}, respectively. Also, the graphical representations in Figure 3.12 and Figure 3.13 seem to support the premise that coordination number does not affect the force constants significantly.

From the representation of the angle force constants versus angle equilibrium value, only dispersion patterns are observed. Angle force constants depend on bond lengths between donor atoms and the metal centre, and angle amplitude, meaning that a three-variable dependence representation, $\mathbf{K_\theta(\theta_0, \ell_{Mn\text{-}L1}, \ell_{Mn\text{-}L2})}$, could allow a more regular understanding of their behaviour.

Although a clear pattern was not achieved by analysis of Mn-centered angle results, rough estimates are proposed in the Table 3.4 and are shown in Figure 3.14 with the standard deviation assigned to it. The estimate that followed originated from the equilibrium coordinate for a given interaction in a range of 8º from the equilibrium position from the non-distorted geometry, except for the axial positions, in which a 15º amplitude was considered (anticlockwise only).

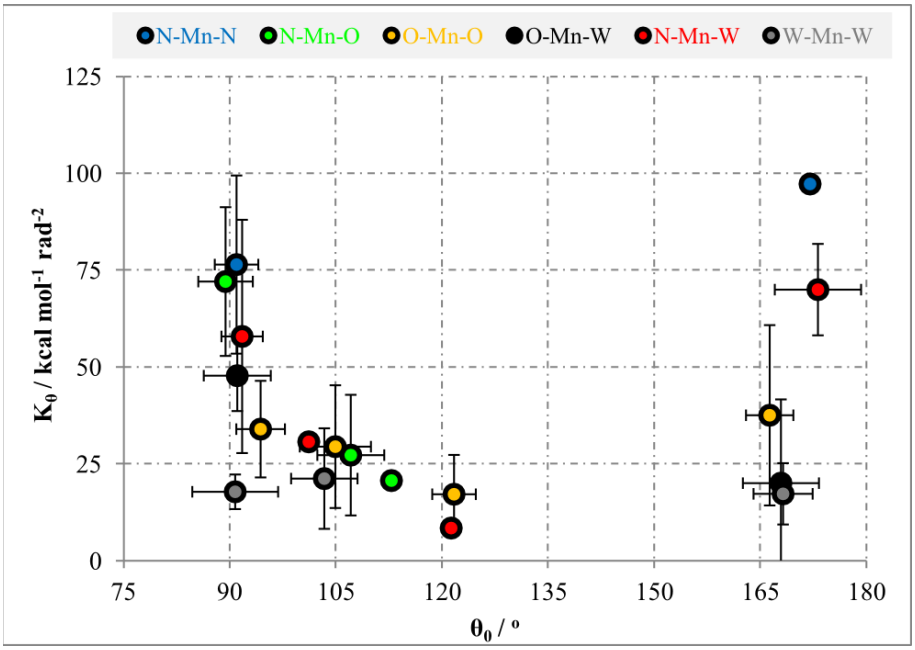


Figure 3.14. Average equilibrium angles and force constants for main donor atoms concerning the equilibrium angles for the main geometries in manganese.

We can observe that angle force constants generally diminish as the average equilibrium amplitude increases; however, for the axial interaction of equal non-water ligands, a slight increase in the force constant is verified. Also, large standard deviations are observed, independently of the sampling dimension considered, for the average force constants determined. Considering the maximum standard deviations observed in the MD simulations, maximum energy errors can reach 8 kcal·mol^{-1} for angles involving water ligands or 3 kcal·mol^{-1} for the remaining ligands.

Table 3.4. Estimate values for the equilibrium angles in frequent manganese coordination geometries

Angles	$\langle \theta_0 \rangle \pm \sigma_{\theta_0}$	$\langle K_\theta \rangle \pm \sigma_{K_\theta}$	Number of cases
N–Mn–N_{axial}	172.06	97	1
N–Mn–$N_{equatorial\ 90^o}$	91 ± 3	(8 ± 2) x10	6
N–Mn–$N_{tetrahedral}$	107 ± 1	35.0 ± 0.9	2
N–Mn–$O_{equatorial\ 90^o}$	89 ± 4	(7 ± 2) x10	9
N–Mn–$O_{equatorial\ 120^o}$	113	21	1
N–Mn–$O_{tetrahedral}$	107 ± 5	(3 ± 2) x10	6
O–Mn–O_{axial}	166 ± 3	(4 ± 2) x10	5
O–Mn–$O_{equatorial\ 90^o}$	94 ± 3	(3 ± 1) x10	10
O–Mn–$O_{equatorial\ 120^o}$	122 ± 3	(2 ± 1) x10	6
O–Mn–$O_{tetrahedral}$	105 ± 5	(3 ± 2) x10	8
O–Mn–W_{axial}	168 ± 5	20 ± 7	3
O–Mn–$W_{equatorial\ 90^o}$	91 ± 5	(5 ± 2) x10	23
O–Mn–$W_{tetrahedral}$	104 ± 4	27 ± 9	5
N–Mn–W_{axial}	173 ± 6	(7 ± 1) x10	3
N–Mn–$W_{equatorial\ 90^o}$	92 ± 3	(6 ± 3) x10	10
N–Mn–$W_{equatorial\ 120^o}$	121	8	1
N–Mn–$W_{tetrahedral}$	101	31	1
W–Mn–W_{axial}	168 ± 4	17 ± 8	2
W–Mn–$W_{equatorial\ 90^o}$	91 ± 6	18 ± 5	8
W–Mn–$W_{tetrahedral}$	103 ± 5	(2 ± 1) x10	2

3.4. Conclusions

We have determined bond, angle and electrostatic parameters for a set of 12 semi-flexible Mn coordination spheres from 9 metalloproteins, following the philosophy of the AMBER force field.

The determined parameters include bond stretching and angle bending force constants, and their equilibrium values, as well as the RESP charges for all the residues in each of the coordination spheres. We have calculated 65 bond force constants, 146 angle force constants and 12 RESP map charges, from a total of 74 residues, for the Mn centres chosen from the literature.

Validation of the determined parameters was carried out by analysis of MD simulations on the metalloproteins chosen in the literature and an analysis of Mn-involving bond and angle parameters from stable metalloproteins was obtained from averaged values and statistical distributions (quartile, histogram, and Gaussian distributions), highlighting some important aspects in the behaviour of the coordination sphere. Typical bond stretching and angle bending values were observed and confidence intervals were established to compare average values with those obtained from the quantum optimizations of the parameterized models. Approximately symmetrical populations were observed for both evaluated coordinates.

In addition, we propose correlations taking into consideration the atom type and typical ligands in Mn centres; therefore other Mn centres in biology can be targeted using the bonded model approach. Linear correlations were established to estimate bond force constants for typical residues in manganese centres. Higher order correlations were proposed in order to estimate rough bond force constants, given typical donor atoms. Angle bending interactions were evaluated by analysis of an obtained dispersion pattern. Some estimates were proposed for the main positions exhibited by Mn centres: axial, quadrangular planar, triangular planar and tetrahedral.

The important results obtained so far strongly support that recurrence can be found in other metalloproteic systems. The established parameters will allow for future relevant mechanistic and catalytic insight on the manganese proteins.

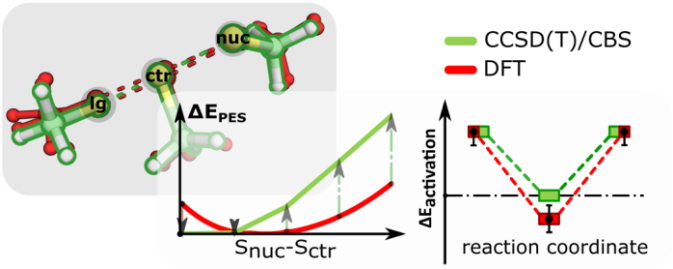
3.4.1. Supporting Information

The Supporting Information for the manuscript can be consulted at http://pubs.acs.org/doi/suppl/10.1021/ct400055v. It contains:

- Description of the nomenclature defined to identify coordinating residues and the donor/acceptor atoms.
- Detailed description of the considerations and literature support for the optimization procedure for each model.
- Tables with detailed information on the Mn-ligand bond and angle force constants and equilibrium coordinates.
- Charge mapping for the 12 models used for parameterization.
- Validation procedure of the 1-4 electrostatic factor applied for MD simulations.
- RMSd plots of the backbone and metallocentre for the 10 ns MD simulations applied in each structure containing the metallocentre.
- Detailed tables comparing crystallographic, optimized and average equilibrium Mn-ligand bond and angle values.
- Quartile and Gaussian distributions for the 12 metalloenzymes simulated.
- Histogram distribution and Gaussian curve approximation validation for the 12 simulated metalloenzymes.
- Frequency calculations at the MM and QM levels of theory for the 12 parameterized models and normal modes analysis on 4 of the parameterized models.
- Internal coordinated analysis for the donating and acceptor atoms and atomic fluctuations on heavy atoms from the optimizations at the MM and QM levels of theory.

Chapter 4: Benchmarking of Density Functionals for the Accurate Description of Thiol-disulphide Exchange

The accuracy of QM calculations is a key aspect in the determination of the thermodynamics of metalloenzyme catalysis. Current DFT calculations have been able to provide highly accurate results for several enzyme mechanisms. However, it is still a requisite that the performance of exchange-correlation density functionals available nowadays should be crossed against that of post-HF methods. The *ab initio* character of these latter methods implies that several of its limitations and mathematical properties are already known and, thus, provide the level of confidence that QM calculations should provide for any given system. However, we have already discussed that the application of post-HF methods can only be employed for systems with a few dozens of atoms, which is an insufficient dimension to describe enzymatic catalytic environments. Several benchmarking studies have been developed to study general chemistry reactions. However, several metalloenzymes catalyse reactions that do not follow such a general chemistry, as is the example of most of the iron-sulphur cluster enzymes. These enzymes are generally involved in redox reactions involving sulphur pairs of atoms, and play a substantial role in cell signalling.



In this study, we provide a benchmarking of a set of 92 density functionals that were employed to accurately characterize the thiol-disulphide exchange reaction. In further work that we will present, we will employ the insight obtained from this benchmarking in the study of the catalytic mechanism of protein disulphide isomerase (PDI) with QM/MM methods. Despite that PDI is not a metalloenzyme, it is a promiscuous enzyme that assists the folding of several proteins through bond and cleavage of disulphide bonds employing the thiol-disulphide exchange reaction. That said, we expect that this study will provide more

insight regarding the core of the reaction in an enzymatic environment. Further work is expected to concern the study of such reaction in metalloenzymes.

The properties that we have benchmarked throughout this study include: the geometry of a 15 atoms model system, the potential energy surface (PES), the activation barrier and the energy of reaction for thiol-disulphide exchange. Reference energies were determined at the extrapolated CCSD(T)/CBS//MP2/aug-cc-pVDZ level of theory, and reference geometries were calculated at the MP2/aug-cc-pVTZ level. M11-L, M06-2X, M06-HF, N12-SX, PBE1PBE, PBEh1PBE and OHSE2PBE described better the geometry of the model system, with average deviations of 0.06 Å in bond lengths (0.06 Å in bond-breaking lengths), and 1.9º in bond angles. On the other hand, the potential energy surface and its gradient were more accurately described by the hybrid density functional BHandH, closely followed by mPW1N, mPW1K and mPWB1K. The barrier height and energy of reaction were better reproduced by the BMK and M06-2X functionals (deviations of 0.17 and 0.07 kcal·mol^{-1}, respectively) for a set of 10 Pople's basis sets. MN12-SX and M11-L showed very good results for the widely used 6-311++G(2*d*,2*p*) basis set, with deviations of 0.02 and 0.05 kcal·mol^{-1}, respectively. We studied the effect of the split-valence, diffuse and polarized functions in the activation barrier of thiol-disulphide exchange, for a set of 10 Pople's basis sets. While increasing the splitting and polarization may increase the activation barrier in approximately 1 kcal·mol^{-1}, diffuse functions generally contribute to decreasing it in no more than 0.10 kcal·mol^{-1}. In general, 13 functionals provided energies within 1 kcal·mol^{-1} of the reference value. The BB1K density functional is one of the best density functionals to characterize thiol-disulphide exchange reactions; however several density functionals with modified Perdew-Wang exchange and about 40% Hartree-Fock exchange, such as mPW1K, mPW1N and mPWB1K, show a good performance too.

All the calculations were performed by Rui Pedro Pimenta das Neves, as for the writing of the manuscript, which was revised through contributions of all authors. This work has been published in the Journal of Chemical Theory and Computation, and the content that follows is a mostly integral transcription of its published version.

Rui P. P. Neves, Pedro A. Fernandes, António J. C. Varandas and Maria J. Ramos,
Benchmarking of Density Functionals for the Accurate Description of Thiol−Disulfide Exchange, *J. Chem. Theory Comput.*, 2014, 10 (11), 4842-4856. (DOI: 10.1021/ct500840f)

4.1. Introduction

Disulphide bonds are abundant in proteins.[304-306] They are known to participate in some folding pathways and are part of the catalytic cycle of some enzymes. In cells, the thiol/disulphide ratio acts as a regulator of the cellular redox potential, it is involved in electron transfer processes across membranes and in the secreted proteins pathway.[304,305,307,308] Atomistic insight of the involvement of these structures in cell regulation may, thus, result in rational pharmacological proceedings toward abnormalities in these processes.

The thiol/disulphide ratio in cells is mainly regulated by thiol-disulphide exchange, and is generally assisted by the reduced glutathione/oxidized glutathione pair. Enzymes that perform thiol-disulphide exchange require an additional step to restart their catalytic cycles, *e.g.* ribonucleotide, 3'-phosphoadenosine-5'-phosphosulfate (PAPS) or methionine sulfoxide reductases.[304] This step may be performed by a small molecule, *e.g.* glutathione, or by specialized enzymes. Disulphide oxidoreductases are the class of enzymes responsible for the formation, reduction and isomerisation of disulphide bonds, through thiol-disulphide exchange. This class of enzymes often possesses a characteristic Cys-X-X-Cys motif (being X any of the natural aminoacids) and a considerable structural similarity to thioredoxin.[307] Equation 4.1 shows the overall reaction for thiol-disulfide exchange.

$$\mathrm{RSH} + \mathrm{R'SSR''} \leftrightarrow \mathrm{RSSR'} + \mathrm{R''SH} \qquad \textbf{Equation 4.1}$$

This is usually a S_N2 reaction in which a cysteine or glutathione is being deprotonated and subsequently acting as a nucleophile, attacking the disulphide bond established by two cysteine (Cys) residues in a Cys-X-X-Cys motif. The attacking sulphur is usually named S_{nuc} (nucleophilic sulphur), the attacked sulphur is named S_{ctr} (central sulphur) and the sulphur that leaves the disulphide bond is named S_{lg} (leaving sulphur). The attack is driven along the disulphide bond axis and the reaction is expected to be intrinsically thermoneutral,[309,310] even though the environment can provide different stabilities to S_{nuc} and S_{lg}, promoting the reaction towards one or another direction, depending on the physiological role of the specific enzyme. While in gaseous phase the trisulphide anion transition state is in fact a minimum energy state,[309,311] in aqueous media the experimentally determined barrier is around 14 kcal·mol^{-1}.[312,313] The transition state structure is approximately symmetrical relative to the attacked sulphur atom.[309,311,314] The charge density delocalization through the three sulphur atoms in the transition state indicates that

hydrophobic environments are better catalysts for the reaction.[309] This evidence is also supported by pK_a studies in cysteines from the active site of disulphide oxidoreductases.[315]

Computational studies have been performed, regarding thiol-disulphide exchange using quantum mechanics calculations, in the recent past.[311,316-318] Density Functional Theory (DFT)[29-31] has been the preferred theoretical level when dealing with large chemical systems, even though there is no exchange-correlation approximation that consistently describes the energy of this interaction. The latter has been the target of several approaches over the past decades, from density functional (DF) development to correlated Hamiltonian methods.[199,319-323] So far, in the DF field we have up to five main approximations for the description of the exchange-correlation energy: local density approximation (LDA), generalized gradient approximation (GGA), meta-generalized gradient approximation (m-GGA), hybrid-generalized gradient approximation (h-GGA) and hybrid-meta generalized gradient approximation (hm-GGA). Other approximations to calculate this term have appeared in the recent years including the range-separated approach (rs),[324-329] nonseparable gradient approximation (NGA)[330-332] and double-hybrid generalized gradient approximation (hh-GGA).[208,333-338]

Given the large number of DFs to calculate a given property of a chemical system, we must choose wisely. In the literature, we can find a plethora of benchmark studies to rank DFs for properties such as ionization and atomization energies, reaction energies, intermolecular and covalent interactions, proton and electron affinities or structural parameters.[116,211-213] Furthermore, several databases that compile specific reactions for chemical properties can be developed to test or rank any DF.[180,339]

We provide here a study on DFT performance for thiol-disulphide exchange, in particular the linkage and dissociation of disulphides by thiolate attack. To our knowledge there are no recent benchmark studies for this reaction, even though it is prevalent in biochemistry.[307,308,340,341] Our benchmarking is designed to rank the performance of DFs to reproduce structural and energy properties of this reaction, from the optimization of model systems to reaction coordinate linear transit scans, and ultimately providing accurate energies for the thiol-disulphide exchange reaction. Therefore, the main goals of our work are defined as: i) to provide a benchmarking on geometry accuracy on a representative model of the thiol-disulphide exchange reaction, ii) to propose DFs to perform linear transit scans along the S_{nuc}-S_{ctr}-S_{lg} reaction coordinate, iii) to determine a set of DFs that best describes the activation barrier and the energy of the thiol-disulphide exchange reaction. Our results are crossed with recent benchmark studies on several

important chemical properties,[180,211-213,342,343] to account for the wide spectra of chemical interactions that a complex medium possesses. The tests are performed with a group of basis sets ranging the double and triple split, polarized and diffuse valence shells. We also test a number of range-separated DFs and several dispersion corrections.[344-347]

To perform the benchmarking we need reference values derived from accurate methods, such as coupled-cluster (CC)[348,349] or configuration interaction (CI).[320,321] These methods require a high computational power and are currently restricted to small systems (less than 30 atoms), if using large basis sets. In this study, we performed single-point energy calculations, at the CCSD(T)[197,322,348-352] level of theory, to determine CCSD(T)/CBS//MP2/aug-cc-pVDZ reference energies. We employ complete basis set (CBS) extrapolation methods[353,354,355,356,357] to extrapolate both the Hartree-Fock (HF) and correlation energies.

4.2. Computational Methods

4.2.1. Model system

A 15 atom model system was built from thioredoxin glutathione reductase (PDB code: 2X8H).[358] Thiol-disulphide exchange occurs through the nucleophilic attack of an external thiolate (mostly from gluthatione) to the disulphide bond of the Cys-X-X-Cys motif. We have kept the terminal methylthiolate from glutathione (GSH1595) and the dimethyldisulphide from the Cys28-Pro29-Tyr30-Cys31 motif of thioredoxin reductase.

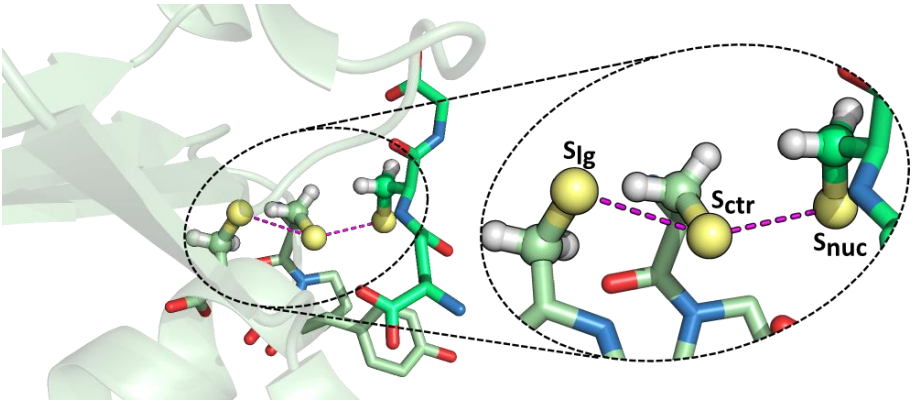


Figure 4.1. The enzyme thioredoxin glutathione reductase (2X8H) is shown in cartoon representation, with the Cys-X-X-Cys motif and the glutathione ligand in stick representation. Our model is highlighted by the black contour in the ball and stick representation. S_{lg}, S_{ctr} and S_{nuc} stand for the leaving group, central and nucleophilic sulphides, respectively.

The computational procedure comprehends two stages, the *determination of reference structures and energies* and the *benchmarking of density functionals*. All the calculations were run with the Gaussian 09 software.[282]

4.2.2. Reference structures for thiol-disulphide exchange

We performed a linear transit scan in vacuum along the S_{nuc}-S_{ctr} distance at the MP2/aug-cc-pVDZ level of theory. The S_{nuc}-S_{ctr}-S_{lg} angle was constrained with three ghost atoms (Gh), by fixing Gh-S_{ctr}-S_{lg} and Gh-S_{nuc}-S_{ctr} angles (see Figure S1 in Supporting Information (SI)). These three ghost atoms establish a geometrical plane that constrains the S_{nuc}-S_{ctr}-S_{lg} angle to 180º throughout the potential energy surface (PES) calculation. This configuration is assumed as

representative for a general nucleophilic attack. Ghost atoms have no charge or basis function information; therefore, they do not interfere in the QM calculations. The PES has a parabolic shape, with only one stationary state (a minimum, the trisulphide anion). The minimum was reoptimized with MP2/aug-cc-pVTZ and served as reference geometry for the DF geometry benchmarking.

4.2.3. Reference energies for thiol-disulphide exchange

We performed geometry optimization calculations at the MP2 level of theory with the aug-cc-pVDZ basis set,[47-49] in water, using the implicit conductor-like polarized continuum model (C-PCM)[359,360] with a dielectric constant of *78.4*. The reaction profile has shown three stationary points (R, TS and P*,* Figure 4.2), whose electronic energy was recalculated, without the solvent, at the CCSD(T)/CBS//MP2/aug-cc-pVDZ level. The continuum solvent was taken out at the end to measure the errors from the functional only.

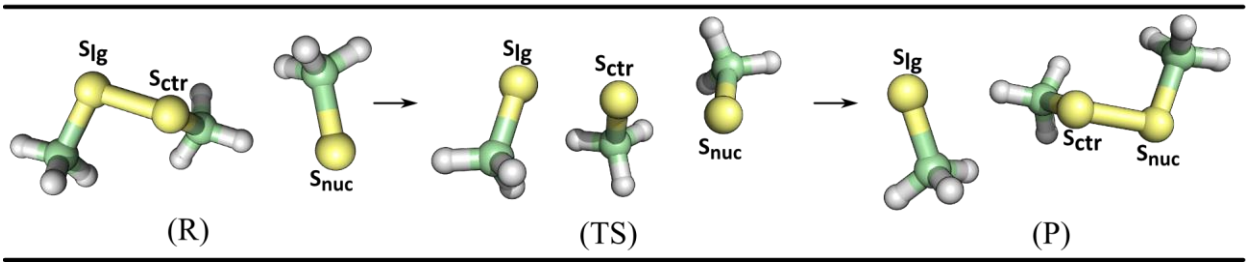


Figure 4.2. Reaction states for thiol-disulphide exchange, reagent (R), transition state (TS) and product (P), obtained at the MP2/aug-cc-pVDZ level of theory.

To investigate the PES shape in vacuum, we calculated the geometry and energy of a series of 20 increments of 0.010 Å, on both sides of the S_{nuc}-S_{ctr} equilibrium bond length, at the MP2/aug-cc-pVDZ level of theory. We then used five equidistant steps, around the minimum of the PES, to calculate our reference PES. Such number of points was required to perform calculations at the CCSD(T)/CBS level.

We have employed the CBS extrapolation schemes for both the reaction energy profile and the S_{nuc}-S_{ctr} relaxed PES (for further discussion, see Table S4 to Table S6 and Figure S4 in SI). We used energies from the MP2/aug-cc-pVXZ (X = 2, 3, 4) levels of theory to extrapolate for the MP2/CBS level, as recommended in the original schemes.[353,354,355,356,357] CBS extrapolation schemes that use only the aug-cc-pVDZ and aug-cc-pVTZ basis sets may lead to inaccurate

results,[353,354,357] and calculations at the CCSD(T)/aug-cc-pVQZ level were too computationally demanding to be feasible. Therefore, we used the MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ single point energies to perform the CBS extrapolation for the correlation energy. We then determined the CCSD(T)/CBS extrapolated energies, assuming that the differences in correlation energy between CCSD(T) and MP2 were the same when calculated using the aug-cc-pVTZ basis set or CBS (see the "CBS extrapolation schemes employed in the study" section, in SI).[211,213,361-363] We did this for all stationary points of the reaction in water and for all points of the PES in vacuum. Despite the limitations that this approach might have, the final quality of the CCSD(T)/CBS energy is much superior to the DFT energies, and is adequate to be taken as a reference to benchmark the density functionals.

4.2.4. Geometry Benchmarking

Geometry optimizations were performed in vacuum, with the 92 density functionals employed in this study, for the 6-31G(*d*)[37,38,41,43,364,365] and the 6-31+G(*d*)[37,38,41-43,364,365] basis sets. These basis sets are frequently employed to optimize large biological systems, where the number of QM atoms usually ranges 100-300 atoms. To rank the performance of the set of DFs in geometry accuracy, we used the MP2/aug-cc-pVTZ optimized structure, in vacuum, as reference. We initially ranked our density functionals by root mean square deviation (RMSd) relatively to the reference structure. However, this criterion is insufficient, since it is an average measure of structure similarity and a more significant deviation in a given coordinate may pass unnoticed in the average. To avoid such situations, we checked for the most relevant internal coordinates and ranked separately (by RMSd) the functionals that did not violate any of the two premises that follow: i) S_{nuc}-S_{ctr}-S_{lg} angle deviations smaller than 3.0º; ii) errors in any of the S_{nuc}-S_{ctr}, S_{ctr}-S_{lg} and S_{nuc}-S_{lg} distances no greater than 0.1 Å. Among these, those with lower RMSd were considered to be more accurate. Even though the thresholds of 0.1 Å and 3.0º might seem arbitrary, in fact, results remain the same even if we change their value. A final note will be given on the threshold for bond lengths. A tenth of an Ångstrom seems to be too large an error for DFT, where common functionals have a bond length accuracy of about 0.01 – 0.03 Å. However, the potential energy surface along the sulphur-sulphur distance in the trisulphide anion in vacuum is extremely flat, allowing for bond elongations at a very modest energy cost.

4.2.5. Electronic Energy and PES gradient benchmarking

A set of 29 DFs, representative of the different exchange-correlation energy approximations and also based on the previous geometry benchmarking, were used to calculate the PES using the S_{nuc}-S_{ctr} bond as the reaction coordinate, since it represents the nucleophilic attack of the cysteine to the glutathione ligand (see Figure 4.1). The PES for the 29 DFs were established with the 6-31G(*d*) basis set, which is the most common choice to perform PES calculations.[87,229,311,317,366-369] We took into consideration also the popularity of some DFs, *e.g* B3LYP.[202,283,284]

Single point energy calculations were carried out for the 92 DFs with 11 basis sets, namely 6-31+G(*d*,*p*),[37,38,41-43,364,365] 6-31+G(2*d*,2*p*), [37,38,41-43,364,365] 6-311G(*d*,*p*),[37-41,43,44] 6-311G(2*d*,2*p*),[37-41,43,44] 6-311G(2*df*,2*p*),[37-41,43,44] 6-311G(2*df*,2*pd*),[37-41,43,44] 6-311+G(*df*,*p*),[37-44] 6-311++G(*df*,*p*),[37-44] 6-311+G(2*d*,2*p*),[37-44] 6-311++G(2*d*,2*p*)[37-44] and TZVP,[45] to investigate the effect of valence splitting, polarization and diffusion functions independently.

We checked the impact of the use of different integration grids, for three basis sets – 6-311G(2*df*,2*p*), 6-311++G(*df*,*p*) and 6-311++G(2*d*,2*p*) – due to the sensitivity stated for some DFs, particularly the Minnesota family of functionals.[373] We used three pruned grids: the 75,302 default grid, the 99,590 grid and the 150,974/225,974 grid. All deviations to the default pruned 75,302 grid are near 0.00 kcal·mol^{-1}. Exceptions stand for the M06-L (0.06 kcal·mol^{-1}), M06-2X, M06 (0.03 kcal·mol^{-1}), BMK (0.09 kcal·mol^{-1}) and ωB97X (0.04 kcal·mol^{-1}) functionals, for both the 99,590 and the 150,974/225,974 grids. The default grid seldom led to inaccuracies, and when they occur they are very small (below 0.1 kcal.mol^{-1}). Additionally, we carried out single point energy calculations for the range separated version[206] of 34 pure DFs and 26 dispersion corrected DFs available in Gaussian 09. The parameters to employ in Grimme's dispersion and with Becke-Johnson damping DFs were retrieved from Grimme's work.[344-346] We employed a large number of the DFs to observe the differences of several DF approximations towards thiol-disulphide exchange, by introducing screened exchange or dispersion corrections.

4.3. Results and Discussion

4.3.1. Characterization of thiol-disulphide exchange

Figure 4.3 shows the PES profile for the HF, CCSD(T) correlation and CCSD(T) electronic energies, extrapolated to the CBS limit, calculated for five relevant S_{nuc}-S_{ctr} distances.

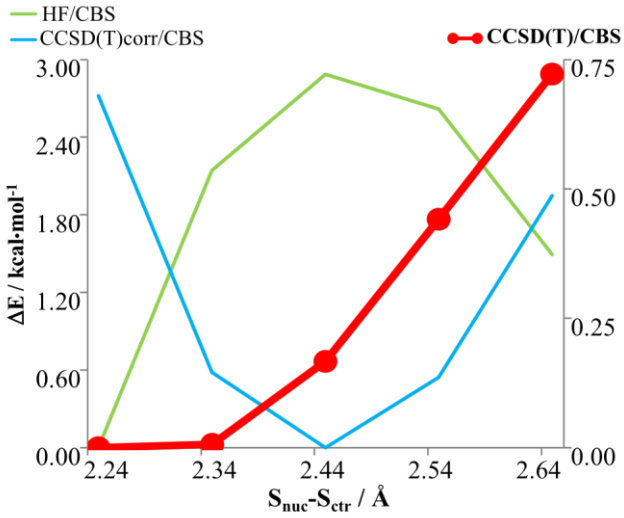


Figure 4.3. PES profile for the CBS extrapolated CCSD(T), HF and CCSD(T)$_{correlation}$ energies. The left axis corresponds to HF/CBS and CCSD(T)$_{corr}$/CBS energies, and the right axis corresponds to the CCSD(T)/CBS energies obtained from Varandas' extrapolation scheme.

A minimum in the correlation energy can be observed when the three sulphur atoms are nearly equidistant (S_{lg}-S_{ctr} and S_{ctr}-S_{nuc} differ in 0.02 Å, see Figure 4.1) and the orbital overlap between the three sulphur atoms is maximum. For this same configuration, the HF energy is at a maximum, since the electron-electron repulsion due to orbital overlapping is high. We observe also that a S_{nuc}-S_{ctr} stretch of 0.20 Å leads to energy differences up to 3 kcal·mol^{-1} and completely different energy profiles at the HF and CCSD(T) levels of theory. Figure 4.3 exemplifies quite well how strongly the correlation energy influences the geometry and energy for our thiol-disulphide exchange model. Therefore the careful choice of an adequate density functional is necessary and fully justified.

Our calculations performed in vacuum failed to provide a local minima representative of the configuration expected in an enzymatic environment, since in X-ray structures of enzymes the thiolate anion is either in an oxidized or reduced form, hence we performed calculations with

implicit aqueous solvent (C-PCM), which provided closer results to the observed in X-ray structures. Figure 4.4 shows the energy profile obtained from the CBS extrapolation scheme of Varandas, for both vacuum and aqueous implicit solvent.

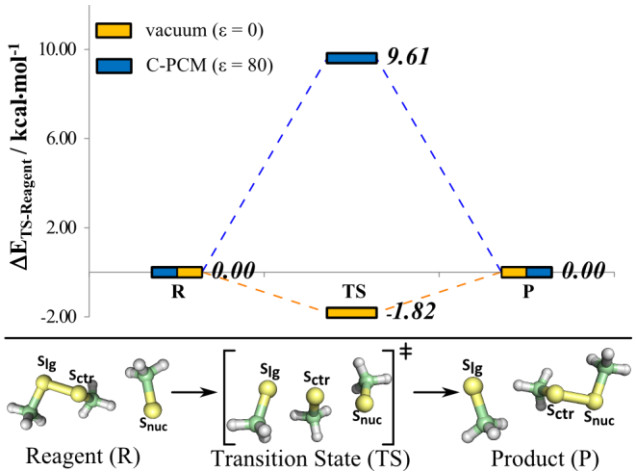


Figure 4.4. Energy profile for the thiol-disulphide reaction obtained with the CBS extrapolation scheme of Varandas, for the vacuum and implicit solvation models, using the geometries optimized in solvent.

Table 4.1 presents the energy of the TS (relative to the R), in vacuum ($\Delta E_{PES\ depth}$) and implicit water C-PCM ($\Delta E_{activation}$), obtained at the MP2/aug-cc-pVXZ (X=2-4), CCSD(T)/aug-cc-pVXZ (X=2, 3) and CBS extrapolated levels of theory.

Table 4.1. Electronic energies from single-point calculations for the R, TS and P states with MP2 and CCSD(T) methods, and CBS extrapolation. DZ stands for the aug-cc-pVDZ basis set, TZ stands for the aug-cc-pVTZ basis set and QZ stands for the aug-cc-pVQZ basis set.

Method / Basis Set	$\Delta E_{PES\ depth}$ / kcal·mol^{-1} (vacuum)	$\Delta E_{activation}$ / kcal·mol^{-1} (water)
$E_{MP2/DZ}$	-5.05	6.24
$E_{MP2/TZ}$	-4.38	7.05
$E_{MP2/QZ}$	-4.09	7.36
$E_{MP2/CBS}$	-4.06	7.38
$E_{CCSD(T)/DZ}$	-3.31	7.90
$E_{CCSD(T)/TZ}$	-2.13	9.28
$E_{CCSD(T)/CBS}$	-1.82	9.61

Comparing the MP2/aug-cc-pVXZ and the MP2/CBS energies to the CCSD(T)/aug-cc-pVXZ and CCSD(T)/CBS energies, one can observe that MP2 energies converge faster towards the CBS limit with the increase of the basis set size. The CCSD(T)/CBS results show an activation energy of 9.61 kcal·mol^{-1}, different from the 14 kcal·mol^{-1} experimental estimate in the literature.[312,313] Our calculations have used an implicit solvent model (with well known limitations) and do not include entropic or zero point energy (ZPE) corrections. We estimated the ZPE and thermal free energy corrections in 8.05 kcal·mol^{-1}, from MP2/aug-cc-pVDZ calculations.

4.3.2. Benchmarking geometry of the thiol-disulphide exchange model

Since disulphide crossed-links have been described as strongly relying in the method employed,[374] we have benchmarked the quality of the molecular geometry, in vacuum, with the 6-31G(*d*) and 6-31+G(*d*) basis sets, against the MP2/aug-cc-pVTZ optimized structure.

These basis sets are representative of this type of calculations in large systems and are extensively used in computational chemistry. Larger basis sets could be employed here but would not be of practical use in studies of applications to realistic biological systems. Of course, the criticism may be that any good agreement with a small basis set may just be accidental, an argument that we cannot unfortunately counterargue. However, we are interested on interaction energies and hence such considerations may not have drastic implications. Table 4.2 shows the DFs that have fulfilled the criteria that we established in the methods section (further results can be found in SI, Tables S24 and S25). Overall, the average error in bond lengths is slightly above of what is typically expected from DFT methods (generally bellow 0.05 Å),[116,179,375] which is due to the unusual flatness of the PES of this system for the sulphur-sulphur distance. As verified in a paper by Goerigk and Reimers[374] geometries obtained with double-split valence basis sets predominantly overestimate sulphur-sulphur stretching, relatively to the MP2/aug-cc-pVTZ structure. Among the best performing DFs, m-GGAs show the highest deviations to the reference values and increasing inclusion of HF exchange shortens these differences.

The Minnesota functionals, M11-L, M06-2X, M06-HF and N12-SX, along with PBE1PBE, PBEh1PBE and OHSE2PBE show the best performance for the set of two basis sets tested. We remark the less good performance of the range separated DFs. B3LYP was not among the best

functionals, nevertheless it showed a similar performance for both 6-31G(*d*) and 6-31+G(*d*) basis sets (ranked 60th and 48th, respectively), with a RMSd of 0.51 Å and 0.52 Å for this system.

Table 4.2. Best performing density functionals for the 6-31G(*d*) (above grey line) and the 6-31+G(*d*) (below grey line) basis sets. The first four columns show the unsigned error with respect to the reference values for the tested DFs. The MUE refers to the mean unsigned average error of the S_{ctr}-S_{lg}, S_{ctr}-S_{nuc} and (S_{nuc}-S_{ctr}) – (S_{lg}-S_{ctr}) lengths.

	S_{ctr}-S_{lg} / Å	S_{ctr}-S_{nuc} / Å	S_{lg}-S_{ctr}-S_{nuc} / º	(S_{ctr}-S_{lg}) – (S_{ctr}-S_{nuc}) / Å	MUE / Å	RMS / Å
MP2/aug-cc-pVTZ	2.42	2.40	179.8	0.02	–	–
M11-L	0.08	0.06	-2.0	0.04	0.06	0.43
PW91B95	0.10	0.09	-0.5	0.03	0.07	0.44
PBEB95	0.10	0.10	-0.7	0.03	0.08	0.44
SVWN	0.03	0.02	-0.5	0.02	0.02	0.45
SVWN5	0.03	0.03	-0.3	0.02	0.03	0.45
MN12-SX	0.08	0.06	-0.8	0.04	0.06	0.45
M06-2X	0.07	0.03	-0.4	0.05	0.05	0.45
M06	0.08	0.04	-1.3	0.06	0.06	0.45
M06-HF	0.04	0.03	-0.2	0.03	0.03	0.46
DSD-BLYP	0.08	0.06	-2.1	0.03	0.06	0.47
N12-SX	0.05	0.02	-2.1	0.04	0.04	0.47
B2GPPLYP	0.08	0.06	-2.3	0.04	0.06	0.47
mPW2PLYP	0.09	0.07	-2.4	0.04	0.07	0.47
mPW1B95	0.10	0.10	-2.2	0.02	0.07	0.48
mPWB95	0.10	0.10	-2.2	0.02	0.07	0.48
PBE1PBE	0.06	0.04	-2.5	0.04	0.05	0.48
B2PLYP	0.09	0.08	-2.7	0.04	0.07	0.48
PBEh1PBE	0.07	0.04	-2.5	0.05	0.05	0.48
OHSE2PBE	0.07	0.04	-2.6	0.05	0.05	0.48
mPW1PBE	0.07	0.04	-2.8	0.04	0.05	0.49
mPW1PW91	0.07	0.04	-2.8	0.04	0.05	0.49
APF	0.07	0.05	-2.8	0.04	0.05	0.49
TPSSh	0.10	0.08	-2.3	0.03	0.07	0.50
B3P86	0.07	0.05	-2.9	0.04	0.05	0.50
TPSS1KCIS	0.10	0.08	-2.8	0.04	0.07	0.50
...	*...*	*...*	*...*	*...*	*...*	*...*
B3LYP	0.07	0.04	3.4	0.05	0.05	0.51

M11-L	0.02	0.09	-2.6	-0.05	0.05	0.44
M062-X	-0.01	0.10	-0.9	-0.09	0.07	0.45
SVWN	0.00	0.04	-0.3	-0.02	0.02	0.46
SVWN5	0.01	0.04	-0.5	-0.02	0.02	0.46
M06-HF	0.01	0.06	-0.6	-0.03	0.03	0.47
N12-SX	0.00	0.06	-2.6	-0.04	0.03	0.48
PBEh1PBE	0.01	0.08	-2.8	-0.05	0.05	0.48
PBE1PBE	0.01	0.08	-2.9	-0.05	0.05	0.48
OHSE2PBE	0.02	0.08	-2.9	-0.05	0.05	0.48
...	*...*	*...*	*...*	*...*	*...*	*...*
B3LYP	0.02	0.09	3.9	0.05	0.05	0.52

The five hh-GGA DFs show good performance with the 6-31G(*d*) basis set. However with the 6-31+G(*d*) basis set they were unable to reproduce the similarity in the S_{ctr}-S_{lg} and S_{ctr}-S_{nuc} bonds exhibited by the reference structure. Other works have already shown that hh-GGAs do not present good results when double-split valence basis sets are employed,[212,333,376] therefore employing hh-GGAs to obtain molecular geometries is limited to smaller systems and may not be adequate to carry out QM calculations in biological systems, where the size of the system must be considerable to account for the most significant interactions in the chemical environment.

A final remark on the set of DFs employed: some current works highlight the role of dispersion as well as basis set error corrections to accurately describe both the system's energy and structure.[212,374,377-379] In particular, when small basis sets, such as 6-31G(*d*), empirical long range dispersion and basis set error considerations have led to significant improvement in molecular geometry, particularly when conformers of the molecular PES show very low barriers.[374,379] Nevertheless, as larger environments are treated, these empirical corrections may be required to obtain either more accurate geometries or energies. Our study did not account for such corrections since we used a small model, based on the fact that we carried out extensive work on the thioredoxin family of enzymes,[309,316,380-383] which employs the reactive system of thiol-disulphide exchange for catalysis. Most of our past studies on these enzymes have been performed on such small systems, but the results[309] have been validated by single-molecule force spectroscopy.[384-386]

4.3.3. Benchmarking the PES along the reaction coordinate

Table 4.3 shows the average MUEs for the five equidistant CCSD(T)/CBS energies in the S_{ctr}-S_{nuc} linear transit scan, for a set of 29 functionals. We have used the 6-31G(*d*) basis set since it is with this basis set (or with basis sets of similar size) that the PES is frequently explored for systems of considerable size. Therefore this study does not present the accuracy of functionals in terms of electronic energy, but instead the accuracy with which they generate the PESs. These DFs were chosen to cover the several existing families and were based on the geometry benchmarking presented beforehand.

Our results show that density functionals, with little or no HF exchange do not reproduce accurately the CCSD(T)/CBS reference energies for the several S_{ctr}-S_{nuc} distances in the PES or the overall interaction. Hybrid functionals with 40-50% HF exchange and the modified Perdew and Wang exchange[387-392] have shown good accuracy in describing the thiol-disulphide exchange energies (see also single-point calculations for the $\Delta E_{PES\ depth}$ in SI). This percentage of HF exchange has been described in previous studies to be needed to accurately describe barrier heights.[207,393,394] In general, pure DFs show the highest MUE among the set of DFs tested. Among the several DF approximations, the hm-GGA set of DFs consistently describes the energies for the five points of the PES with the lowest errors, all rounding 0.20-0.30 kcal·mol^{-1}. For the hh-GGA set of DFs, we observe increasing MUE values as the percentage of HF exchange increases. B3LYP gives relative energies higher than the reference energies when the S_{ctr}-S_{nuc} distance is lower than 2.45 Å. However, the same is observed for almost all other DFs tested. Overall, the BHandH, mPW1N, mPW1K, mPWB1K and BB1K functionals show a good and similar performance for energy calculations. We notice that again the functional BB1K stands out as one of the best.

An energy analysis from our set of DFs is not enough to choose those that best represent the PES energy for thiol-disulphide exchange. The adequate functional must be accurate in both energy and energy gradient towards the CCSD(T)/CBS PES. Indeed, the fact that the CCSD(T)/CBS and the DFT/6-31G(*d*) curves must be parallel is key to certify similar dynamical attributes for the density functional performance. Our analysis turns now to the energy gradient along the S_{ctr}-S_{nuc} reaction coordinate, where we compare four equidistant intervals from the extrapolated CCSD(T)/CBS PES of the reaction coordinate.

Table 4.3. MUEs, in kcal·mol^{-1}, for a set of 29 DFs used to determine the PES from CCSD(T)/CBS calculations (see Figure 4.4). All DFT calculations were performed with the 6-31G(*d*) basis set.

S_{ctr}-S_{nuc} / Å			**2.25**	**2.35**	**2.45**	**2.55**	**2.65**	
$\Delta E(S_{ctr}$-$S_{nuc})$ / kcal·mol^{-1}			0.00	0.01	0.17	0.44	0.72	
DF	**E_{xc}**	**% E_x^{HF}**	**Error / kcal·mol-1**					**MUE**
SVWN	LDA	—	1.21	0.19	-0.15	-0.02	0.54	0.42
HCTH407	GGA	—	0.97	0.32	-0.13	-0.42	-0.51	0.47
PBE	GGA	—	2.04	0.67	-0.09	-0.40	-0.30	0.70
BP86	GGA	—	2.08	0.70	-0.08	-0.41	-0.33	0.72
PW91	GGA	—	2.11	0.75	-0.05	-0.42	-0.39	0.75
BLYP	GGA	—	2.84	1.21	0.17	-0.43	-0.61	1.05
M11-L	m-GGA	—	0.79	0.20	-0.16	-0.35	-0.34	0.37
VSXC	m-GGA	—	1.43	0.52	-0.07	-0.43	-0.50	0.59
PW91TPSS	m-GGA	—	1.87	0.59	-0.11	-0.38	-0.26	0.64
TPSS	m-GGA	—	2.05	0.71	-0.07	-0.41	-0.35	0.72
BHandH	h-GGA	50.00	0.09	0.00	-0.11	-0.26	-0.29	0.15
mPW1N	h-GGA	40.60	0.24	0.02	-0.15	-0.30	-0.34	0.21
mPW1K	h-GGA	42.80	0.20	0.01	-0.15	-0.33	-0.40	0.22
PBE1PBE	h-GGA	25.00	0.81	0.17	-0.17	-0.28	-0.15	0.31
mPW1PBE	h-GGA	25.00	0.83	0.18	-0.17	-0.29	-0.17	0.33
mPW1PW91	h-GGA	25.00	0.84	0.18	-0.17	-0.30	-0.19	0.34
B3LYP	h-GGA	20.00	1.34	0.44	-0.12	-0.41	-0.43	0.54
mPWB1K	hm-GGA	44.00	0.11	-0.01	-0.13	-0.30	-0.39	0.19
BB1K	hm-GGA	42.00	0.18	0.01	-0.15	-0.34	-0.42	0.22
M06-2X	hm-GGA	54.00	0.64	0.12	-0.16	-0.26	-0.16	0.27
M06-HF	hm-GGA	100.00	0.77	0.13	-0.16	-0.23	-0.06	0.27
BMK	hm-GGA	42.00	0.31	0.00	-0.14	-0.34	-0.57	0.27
N12-SX	h-NGA	25.00/ – *	0.77	0.13	-0.16	-0.23	-0.06	0.27
MN12-L	m-NGA	—	0.30	0.04	-0.16	-0.39	-0.54	0.29
MN12-SX	hm-NGA	25.00/ – *	0.64	0.15	-0.17	-0.35	-0.34	0.33
N12	NGA	—	0.80	0.14	-0.16	-0.25	0.73	0.42
B2GPPLYP	hh-GGA	53.00	0.67	0.17	0.00	-0.08	-0.36	0.35
DSD-BLYP	hh-GGA	65.00	0.72	0.20	-0.01	-0.08	-0.39	0.36
B2PLYP	hh-GGA	70.00	1.02	0.30	-0.02	-0.06	-0.36	0.44

* The starred DFs show screened HF exchange at short and long ranges, respectively

Figure 4.5 shows our results for the several approaches to the exchange-correlation (E_{xc}) term. We performed a comparison of the energy gradient as we go farther from the equilibrium distance of the PES. In this way we evaluate both the energy and energy gradient accuracies, towards the CCSD(T)/CBS PES, taking into account that new geometries are being obtained from the DF itself.

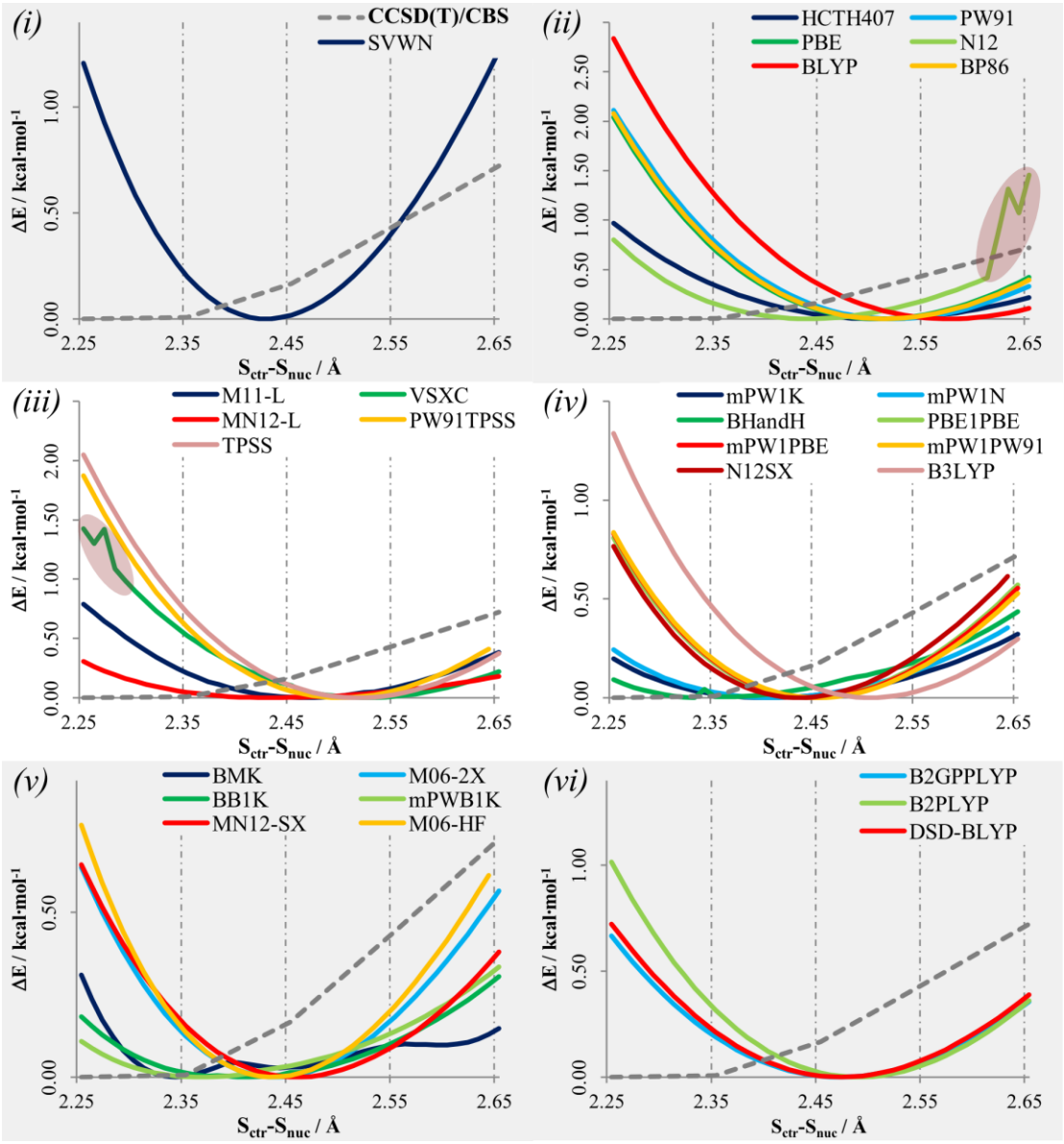


Figure 4.5. Comparison of the PES scans for the selected functionals with the 6-31G(*d*) basis set, and the reference PES from CCSD(T)/CBS calculations. We analyse DF performance according to E_{xc} approximation for each set: (i) LDA, (ii) GGA and NGA, (iii) m-GGA and m-NGA, (iv) h-GGA and h-NGA, (v) hm-GGA and hm-NGA and (vi) hh-GGA. The CCSD(T)/CBS calculated PES is shown with grey dashes.

An overall look at Figure 4.5 shows that no DF reproduces accurately the energy gradient of the reference PES, which we observe to be very flat. We also notice that the energy minimum from the PES of each density functional does not seem to be moving close to the 2.25-2.35 Å distance.

Regarding the GGA functionals, we observe that the energy minimum is shifted towards larger S_{ctr}-S_{nuc} values, relatively to the reference PES. Comparatively to the GGA approximation, the introduction of kinetic spin interaction in the Hamiltonian improves the PES obtained from m-GGAs. The M11-L functional, followed by MN12-L, shows the best performance in reproducing the CCSD(T)/CBS PES for the thiol-disulphide exchange reaction in this class of DFs. The N12 density functional shows a discontinuous behaviour farther from the 2.60 Å distance and VSXC presents a slight discontinuity for values of S_{ctr}-S_{nuc} close to 2.30 Å, hence these regions were not considered in this part of our discussion.

Analysing the h-GGA class of functionals, we can observe a clear dependence on energy profile and HF exchange. DFs with 25% HF exchange (PBE1PBE, mPW1PBE and mPW1PW91) show an identical behaviour, closely followed by the N12-SX functional. The BHandH, mPW1N and mPW1K functionals, with 40-50% HF exchange, show the lowest energy deviations. BHandH shows the closest results to the CCSD(T)/CBS extrapolated PES in both energy accuracy and energy gradient. We emphasize, particularly, the performance of B3LYP due to its high popularity. It overestimates the energy for short S_{ctr}-S_{nuc} distances and the energy gradient profile along the reaction coordinate becomes more consistent with the CCSD(T)/CBS profile for values larger than the equilibrium distance (see Figure 4.5). The results from the hm-GGA class of DFs do not differ substantially from h-GGA DFs, with the mPWB1K and BB1K functionals showing the best overall performance. While the BMK functional performs accurate energy calculations for the reaction coordinate, comparatively to the CCSD(T)/CBS PES, the kinetics of the S_{ctr}-S_{nuc} bond throughout the linear transit scan failed to be reproduced.

B2PLYP, B2GPPLYP and DSD-LYP functionals, all hh-GGAs, behaved similarly in this part of the study. Despite none of them being able to accurately describe the PES gradient at the CCSD(T)/CBS level of theory, MUEs were relatively small comparatively to the CCSD(T)/CBS PES (in the 0.30-0.40 kcal·mol^{-1} range). This gradient is more similar between these DFs, as we move for larger S_{ctr}-S_{nuc} distances.

4.3.4. Benchmarking of the activation energy for thiol-disulphide exchange

A set of 92 DFs[31,202,206-208,283,284,324-338,387-393,395-437] was benchmarked against CCSD(T)/CBS extrapolated activation energies. All calculations show the overall reaction to be thermoneutral, with errors often lower than 0.1 kcal·mol^{-1}.

Table 4.4. Set of DFs, within the 1 kcal·mol^{-1} error from the reference value for the 6-311++G(2*d*,2*p*) basis set. The DFs are displayed by exchange correlation energy (E_{xc}) approximation and increasing error.

Method / Basis Set		$\Delta E_{PES\ depth}$ **/ kcal·mol^{-1}**	
CCSD(T)/CBS		-1.82	
DF	E_{xc} [ρ]	% E_x^{HF}	Error
OVWN5	GGA	—	-0.21
OPL	GGA	—	-0.27
OVWN	GGA	—	-0.48
M11-L	m-GGA	—	-0.05
mPW1K	h-GGA	42.80	0.14
mPW1N	h-GGA	40.60	-0.23
BMK	hm-GGA	42.00	0.11
M06-2X	hm-GGA	54.00	-0.29
BB1K	hm-GGA	42.00	0.46
mPWB1K	hm-GGA	44.00	0.57
mPWKCIS1K	hm-GGA	41.00	0.95
MN12-L	m-NGA	—	0.43
MN12-SX	hm-NGA	25.00/ – *	0.02

* The starred DFs show screened HF exchange at short- and long-range, respectively

LDA DFs present the less exact activation energy ($\Delta E_{PES\ depth}$) for thiol-disulphide exchange, with errors around 10 kcal·mol^{-1} for any of the 11 basis sets tested, while B3LYP shows errors within the 3 kcal·mol^{-1} underestimation error attributed to this density functional. This has been observed also for several other properties in other benchmark studies.[211-213] In Table 4.4, we show the set of density functionals that performed better (within 1.00 kcal·mol^{-1}) for the 6-311++G(2*d*,2*p*) basis set, which is commonly used in single-point calculations in mechanistic studies.[87,229,366,367] The Minnesota family of DFs shows good performance towards the reaction we have studied, as there are 4 of these density functionals that show an error smaller than 1 kcal·mol^{-1} from the

CCSD(T)/CBS extrapolated values, in 10 of the tested DFs. The results for the whole 92 density functionals are given in SI (see Table S7). An error of 1 kcal·mol^{-1} might seem large in sight of the magnitude of the interaction and reaction energies of thiol-disulphide exchange. However, this is illusory, since the activation and reaction energies are differences between large numbers, and if they are not accurate the differences may take large positive/negative values. Hence, to calculate a relative error does not make sense here because the limits of the scale are not determined.

We tested the systematic behaviour of the 92 DFs with basis sets having different valence splitting, polarization and diffuse gaussian functions, and show in Figure 4.6 the results for a set of 17 DFs that performed within 1.00 kcal·mol^{-1} for at least half of the 10 Pople's basis sets tested, and B3LYP.[202,283,284] The objective of the next part of our discussion is not, to choose the best basis set to approach with a given functional but, instead we try to give insight on the confidence the user can get from selecting a given DF to study the thiol-disulphide exchange reaction.

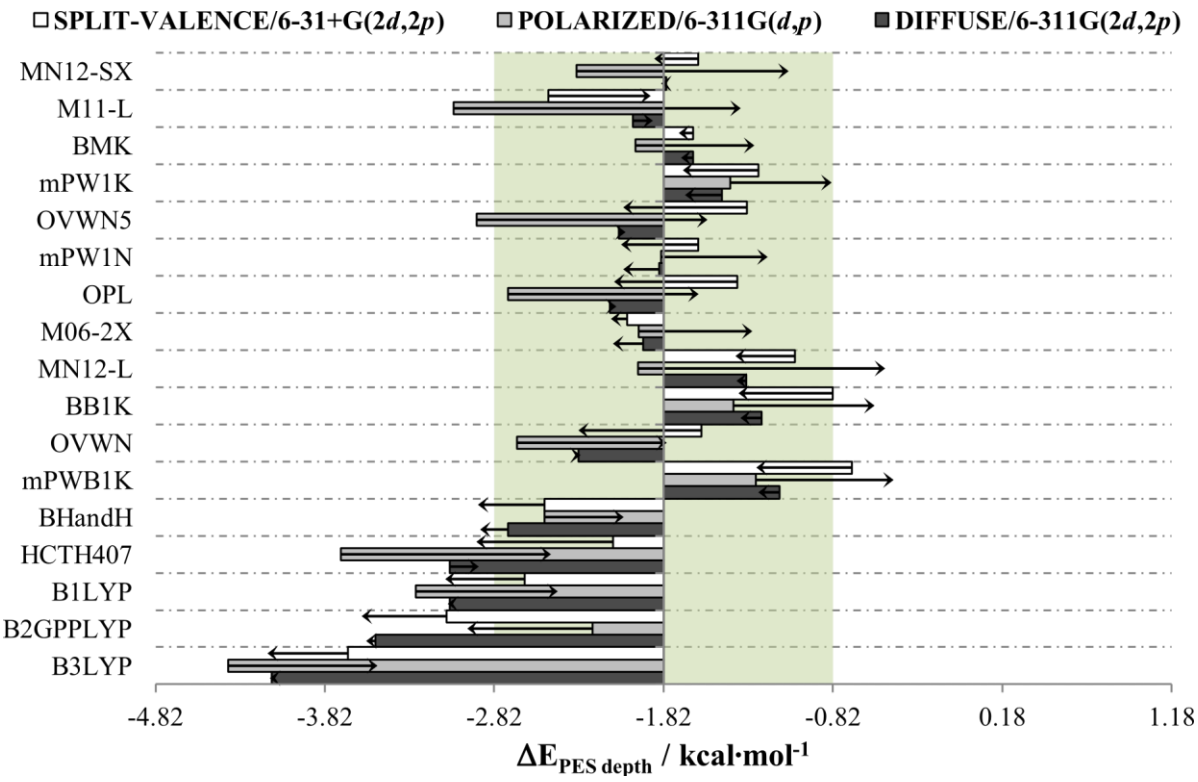


Figure 4.6. Effect of splitting, polarization and diffuse functions in the DF energies. The bars represent the energy determined with the smaller basis set and the arrows show the mean basis set truncation error obtained with larger basis sets in relation to the smaller basis set. The green region marks the limiting error of 1.00 kcal·mol^{-1}.

We observe that inclusion of diffuse basis sets in heavy atoms decreases the $\Delta E_{PES\ depth}$ for most density functionals, in about 0.10 kcal·mol^{-1}, relative to the 6-311G(2*d*,2*p*) energy. As we introduce diffuse functions in hydrogen atoms, we notice that errors from the reference energy are nearly the same (see dark grey bars in Figure 4.6). The cases in which this deviation is larger are observed for h-GGA DFs with high HF exchange (higher than 40%).

Comparing the energy calculations from the double and triple-split valence basis sets we observe that the $\Delta E_{PES\ depth}$ decreases for almost all density functionals (see white bars in Figure 4.6). These decrements are always smaller than 1 kcal·mol^{-1}. Only for M11-L[437] and M06-HF,[431,432] larger split-valence leads to a higher $\Delta E_{PES\ depth}$. The m-GGA and hh-GGA sets of DFs show the largest standard deviations.

The effect of increasing polarization can be observed in light grey bars in Figure 4.6. This property shows the highest variation in the thiol-disulphide exchange reaction. Figure 4.6 shows that increasing polarization leads to higher $\Delta E_{PES\ depth}$, as we approach the 6-311G(*d*,*p*), 6-311G(2*d*,2*p*) and 6-311G(2*df*,2*pd*) basis sets. Differences in $\Delta E_{PES\ depth}$, among these basis sets, can be larger than 1.00 kcal·mol^{-1}, although that was only observed for 6 DFs. In general, as polarization increases the $\Delta E_{PES\ depth}$ obtained for each DF increases in a non-linear manner. The exception stands for the hh-GGA set of DFs. In this latter set, we observe a decrease in the PES depth from 6-311G(*d*,*p*) to 6-311G(2*d*,2*p*), and an increase from the 6-311G(2*d*,2*p*) to 6-311G(2*df*,2*pd*) basis sets. The $\Delta E_{PES\ depth}$ for the 6-311G(2*d*,2*p*) basis set are the farthest from the reference value. Density functionals from the m-GGA set show more equally spaced $\Delta E_{PES\ depth}$ deviations (see Table S21 in SI). The m-GGA and hm-GGA set of DFs provide larger $\Delta E_{PES\ depth}$ for highly polarized basis sets; the latter show non-linear variations of $\Delta E_{PES\ depth}$ with increasing polarization – introduction of *f* and *d* orbitals in heavy and hydrogen atoms, respectively, shows a higher increase in $\Delta E_{PES\ depth}$ for the thiol-disulphide exchange.

The choice of basis set is of outmost importance in any DFT energy calculation. As a general conclusion, we tentatively propose that the combination of triple-valence basis sets with highly polarized functions and diffuse orbitals in heavy atoms should be employed in the thiol-disulphide exchange.

Additionally, we have tested density functionals with screened HF exchange terms or dispersion corrections for three basis sets out of the eleven used. Density functionals with screened HF exchange show no improvement relatively to pure DFs, as can be seen in Figure 4.7. Contrary to

pure DFs, which underestimate the reference activation energy, we observe that the activation energy is constantly overestimated over the CCSD(T)/CBS extrapolated energy, except for M11-L, in which deviations are the same from the pure DF. Out of all the screened HF exchange functionals, CAM-B3LYP performs best with a MUE of 1.42 kcal·mol^{-1}.

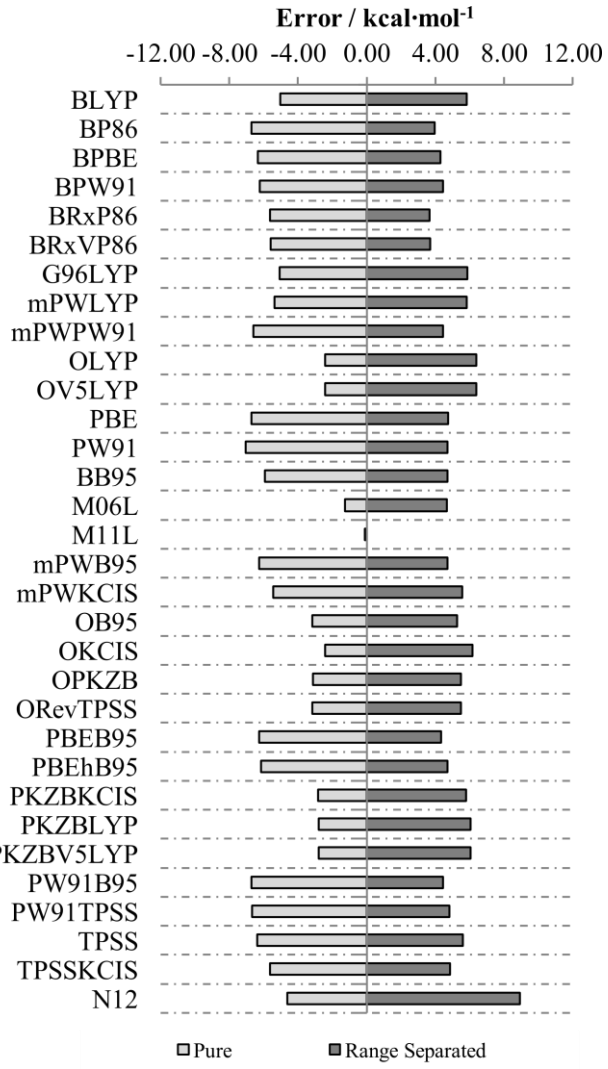


Figure 4.7. Signed Error for the range separated version for a set of pure DFs from the study of the activation energy. The results are presented for the 6-311++G(2*d*,2*p*) basis set.

We used dispersion corrections for 23 DFs available in Gaussian 09, with published dispersion parameters. For most cases, the dispersion correction does not improve the DF performance in our 15 atom model system, as can be seen in Table 4.5. The CAM-B3LYP and BHandHLYP

functionals are relevantly improved by the addition of Grimme's corrections with Becke-Johnson damping, with MSE of 0.79 kcal·mol^{-1} and 1.47 kcal·mol^{-1}.

Table 4.5. MSE in the activation energy for dispersion corrected DFs using the 6-311++G(2*d*,2*p*) basis set.

DFs	no dispersion corrections	dispersion corrections	
		D3	D3-BJ
PBE	-6.74	-6.81	-7.50
TPSS	-6.38	-6.44	-7.36
BLYP	-5.04	-5.04	-6.88
B97	–	-4.25	-5.76
B3LYP	-2.31	-2.40	-3.85
B2PLYP	-2.58	-2.61	-3.36
BP86	-6.71	-6.75	-8.25
BPBE	-6.36	-6.37	-8.09
mPWLYP	-5.39	-5.47	-6.16
TPSSh	-4.87	-4.95	-5.76
OPBE	–	-4.67	-7.52
OLYP	-2.43	-3.64	-6.59
B3PW91	-3.27	-3.34	-4.73
PBE1PBE	-2.98	-3.05	-3.64
M06	-2.12	-2.16	-0.28
B1B95	-1.81	-1.83	-3.26
M06-L	-1.26	-1.25	-1.69
BMK	0.11	0.08	-1.22
mPWB1K	0.57	0.56	-0.01
CAM-B3LYP	1.42	1.38	0.79
BHandHLYP	2.84	2.79	1.47
LC-ωPBE	4.50	4.45	3.75
B2GPPLYP	-1.73	-1.74	-2.63
DSD-BLYP	-2.01	-2.02	-3.02

Despite no significant improvement is observed by employing both screened-exchange or dispersion corrections, we emphasize that our system under study is a small representative model and that for larger systems these corrections might be significant, even though past evidence is on our side.[309,384-386] Overall, the BMK[393] and M06-2X[433] DFs show a good performance towards our CCSD(T)/CBS extrapolated energies and are widely validated as good candidates for very good energy calculations, in several other recent benchmark studies.[180,211-213,342,343] The MN12-SX functional also shows overall good performance, and should be considered as a proper candidate in this type of calculations.

4.3.5. DFT performance in the thiol-disulphide exchange reaction

The final section of the paper presents an overall discussion of the DFs most suited to perform calculations for the thiol-disulphide exchange reaction. A quest for a density functional that may describe both thermodynamics and structure of a system is, ultimately, desired. Our benchmarking indicates that there is a clear prevalence of meta- or hybrid- density functionals with 40-50% HF exchange that perform best for our designed thiol-disulphide exchange model. The mPWB1K, mPW1N, mPW1K and BB1K density functionals can be considered as good candidates to conduct calculations for this reaction; even though they were not the best candidates to reproduce the geometry of our 15 atom model, they are the most accurate to provide thermodynamics for thiol-disulphide exchange. Despite that M11-L, MN12-SX and M06-2X show good results in all properties benchmarked in this study, they were not the best candidates to determine the PES for this model reaction.

4.4. Conclusions

Our study provides insight on the performance of a set of 92 density functionals characterizing thiol-disulphide exchange. This class of reactions is very important in biochemistry, therefore the benchmarking of important thermodynamic and kinetic properties will allow for more accurate computational studies in large systems.

As found in other benchmarking studies, we have opted to design a small model (15 atoms) and use computationally demanding post-HF methods – MP2 and CCSD(T) – with correlation consistent basis sets (aug-cc-pVXZ, X = 2, 3, 4). These calculations provided us with accurate reference values, from the CBS extrapolation method of Varandas, which was employed to benchmark our set of density functionals.

Regarding the molecular geometry, several functionals reproduced well the geometry of the 15 atoms model. In particular M11-L, M06-2X, M06-HF, N12-SX, PBE1PBE, PBEh1PBE and OHSE2PBE functionals produced geometries similar to MP2/aug-cc-pVTZ, with both the 6-31G(*d*) and the 6-31+G(*d*) basis sets. The hh-GGAs show very different results for both basis sets, and were unable to reproduce the reference MP2/aug-cc-pVTZ geometry for the 6-31+G(*d*) basis set. Additionally, we highlight the less accurate performance of range-separated density functionals, in particular for LC-ωPBE.

We selected 29 density functionals to reproduce the CCSD(T)/CBS PES along the reaction coordinate; our observations led us to conclude that no DF accurately reproduces both the energy and energy gradient of the S_{nuc}-S_{ctr} attack. Nevertheless h-GGA functionals showed the best performance, in particular for BHandH, mPW1N, mPW1K and mPWB1K. DFs using 40–50% HF exchange result in an error for the reaction energy and PES in vacuum, often lower than 1.00 kcal·mol^{-1}.

We tested eleven different basis sets – 6-31+G(*d*,*p*), 6-31+G(2*d*,2*p*), 6-311G(*d*,*p*), 6-311G(2*d*,2*p*), 6-311G(2*df*,2*p*), 6-311G(2*df*,2*pd*) 6-311+G(*df*,*p*), 6-311+G(2*d*,2*p*), 6-311++G(*df*,*p*), 6-311++G(2*d*,2*p*) and TZVP – to evaluate the effect of the basis set in density functional performance towards the activation energy of thiol-disulphide exchange. Diffuse functions in heavy atoms may lead to a decrease of 0.10 kcal·mol^{-1} in relative activation energies for the reaction. The split-valence has a higher influence in the relative energy predicted, since the comparison between double and triple-split valences may lead to differences in relative energies

of about 1.00 kcal·mol^{-1}. As polarization functions are added to make the basis set more complete, relative activation energies from DFT calculations increase. While in the m-GGA set of density functionals these increments seem to provide more predictable differences, in most remaining cases the addition of *f* and *p* sets of functions to heavy atoms and hydrogens is more pronounced than the increasing of the *d* and *p* sets of functions. Overall, the M06-2X and BMK functionals provide the most accurate results against our CCSD(T)/CBS activation energies. The M11-L and MN12-SX functionals showed activation barriers with errors lower than 0.10 kcal·mol^{-1} from the CCSD(T)/CBS extrapolated energies, for the typical 6-311++G(2*d*,2*p*) basis set (with MUEs of 0.05 and 0.02 kcal·mol^{-1}, respectively). The calculations we performed for range separated and dispersion corrected functionals show no improvement, for all the functionals, except for the CAM-B3LYP functional, in which the MUE lowers from 1.42 kcal·mol^{-1} to 0.79 kcal·mol^{-1}, with the Becke-Johnson damping. Dispersion in larger systems may be important but it will not be due to the thiol-disulfide exchange reaction itself. Instead, it will emanate from the many different molecular skeletons to which the thiols may be connected. This would be a problem of benchmarking intermolecular interactions, and not thiol-disulphide exchange. Therefore, we believe it is appropriate to say that dispersion does not have a meaningful role in the thiol-disulphide exchange reaction. Such corrections may turn out to be significant to accurately describe thermodynamic properties, among others, in larger systems. Despite the fact that the B3LYP functional does not show a good performance for any of the benchmarked properties, it shows a consistent average performance throughout the spectra of properties we have tested in our study.

Considering all aspects of the thiol-disulphide exchange reaction (geometry, PES, and activation energy) the mPWB1K, mPW1N, mPW1K and BB1K density functionals are better suited to perform calculations for this reaction.

We believe that our current study shows a good systematic approach towards the benchmarking of specific chemical interactions, and will hopefully prove to be useful in future computational studies concerning thiol-disulphide exchange reactions in complex systems.

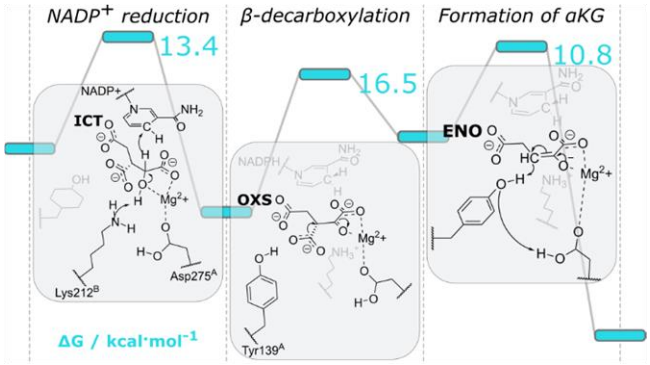
4.4.1. Supporting Information

The Supporting Information for the manuscript can be consulted at http://pubs.acs.org/doi/suppl/10.1021/ct500840f. It contains:

- Linear Transit Scan for the S_{ctr}-S_{nuc} Reaction Coordinate with the MP2 Level of Theory.
- Single-point energy calculations for MP2 and CCSD(T) levels of theory.
- CBS extrapolation schemes employed in the study.
- Rank of basis set from Single-point MP2 calculations relative to CCSD(T)/CBS values.
- Benchmark of the 92 DFs for the 6-311G++g(2*d*,2*p*) basis set. Single-point calculations for the basis set 6-31+G(*d*,*p*), 6-31+G(2*d*,2*p*), 6-311G(*d*,*p*), 6-311G(2*d*,2*p*), 6-311G(2*df*,*p*), 6-311G(2*df*,2*pd*), 6-311+G(*df*,*p*), 6-311++G(*df*,*p*), 6-311+G(2*d*,2*p*), 6-311++G(2*d*,2*p*), TZVP.
- Comparison of the relative energies for the 11 basis sets studies. MSE for the different integration grids for the basis set 6-311G(2*df*,2*p*), 6-311++G(*df*,*p*) and 6-311++G(2*d*,2*p*).
- MSE for the range separated version of a set of GGAs, for the basis set 6-311G(2*df*,2*p*), 6-311++G(*df*,*p*) and 6-311++G(2*d*,2*p*).
- MSE for the Dispersion corrected DFs, for the basis set 6-311G(2*df*,2*p*), 6-311++G(*df*,*p*) and 6-311++G(2*d*,2*p*).
- Benchmark of optimization calculations for the 6-31G(*d*,*p*) and 6-31+G(*d*,*p*) basis set relative to the MP2/aug-cc-pVDZ optimized structure.
- MUE for the five equidistant steps in the S-S PES for the set of 30 DFs.

Chapter 5: Unveiling the Catalytic Mechanism of NADP$^+$-dependent Isocitrate Dehydrogenase with QM/MM calculations

In the field of catalysis by enzymes, there are two major problems: molecular mechanics potentials are unable to describe the formation and cleavage of covalent bonds, and quantum mechanics potentials cannot be applied to a system with a dimension of even the smallest of the enzymes. Hybrid methodologies are the preferred option to tackle these systems, since they enable the study of the chemistry of the enzyme's reaction (even if for a small part of the enzyme model), and also allow that the bulk enzymatic environment can play a role in the process (being described with more simple potentials).



In this chapter, we resorted to hybrid quantum mechanics/molecular mechanics calculations to tackle isocitrate dehydrogenase, a metalloenzyme that requires either divalent magnesium or manganese to catalyse the transformation of isocitrate and NAD(P)$^+$ in α-ketoglutarate and NAD(P)H. In this study we will approach the role of magnesium in the overall catalysis by the enzyme. However, in future studies we will make use of the parameters we have developed in Chapter 3 to study the differences provided in its catalysis, as the magnesium ion is replaced by the divalent manganese.

We have determined the catalytic mechanism of the human cytosolic homodimeric isocitrate dehydrogenase (hICDH), an enzyme involved in the regulation of tumorogenesis. Our study constitutes the first theoretical attempt to describe the entire catalytic cycle of hICDH. In agreement with earlier experimental proposals, the catalysis was shown to proceed in three steps: (1) NADP$^+$ reduction by the isocitrate substrate with the help of the Lys212^B base, (2) β-decarboxylation of the resulting oxalosuccinate, generating an enolate, and (3) protonation of this

intermediate by Tyr139^{A}, giving rise to the α-ketoglutarate product. Our study supports that the β-decarboxylation of oxalosuccinate is the most likely rate-limiting step, with an activation Gibbs free-energy of 16.5 kcal·mol^{-1}. The calculated values are in close agreement with the 16−17 kcal·mol^{-1} range, derived by the application of transition state theory to the reaction rates determined experimentally (11 s^{-1} to 38 s^{-1}). We emphasize the role of the Mg^{2+} and Asp275^{A}, whose acid/base properties throughout the catalytic cycle were found to lower the barrier to physiologic competent values. Aside from its chemical dual role (as a base, deprotonating Lys212^{B}, and as an acid, protonating the basic Tyr139^{A} deprotonated by the enolate intermediate), it also establishes hydrogen bonds with Arg132^{A} and Tyr 139^{A} that become shorter at critical transition states. These residues were shown to influence both the rate and the efficiency of hICDH. The knowledge drawn in this study provides new insights to future clinical and bioengineering applications of hICDH, namely in the development of techniques to regulate the growth of glioblastomas and to capture and storage carbon dioxide. Moreover, it further extends the comprehension: (1) the hydrogen/charge transfer mechanism that regulates the hydrogenation of $NADP^{+}$ to NADPH, an ubiquitous biochemical reaction, and (2) the role of divalent metals as key structure elements in the family of NAD(P)$^{+}$-dependent β-decarboxylases.

All the calculations were performed by Rui Pedro Pimenta das Neves, as for the writing of the manuscript, which was revised through contributions of all authors. This work has been published in the ACS Catalysis, and the content that follows is a mostly integral transcription of its published version.

Rui P. P. Neves, Pedro A. Fernandes, Maria J. Ramos, **Unveiling the Catalytic Mechanism of NADP^{+}-Dependent Isocitrate Dehydrogenase with QM/MM Calculations**. *ACS Catalysis*, 2015, 357-368. (DOI: 10.1021/acscatal.5b01928)

5.1. Introduction

5.1.1. Relevance of the work

Throughout the last decades, several experimental studies have proposed possible catalytic pathways for isocitrate dehydrogenase (ICDH) forms of *Escherichia coli*, *Micobacterium tuberculosis*, *Sus scrofa* and *Homo sapiens*.[438-443] These studies have clarified the main intermediates along the cycle but a set of questions still remain unanswered, partly due to the total absence of atomic-level theoretical studies detailing/validating/proposing any mechanism for ICDH. During the last few years clinical attention was driven towards the human isocitrate dehydrogenase (hICDH), as several mutations of its cytosolic and mitochondrial homodimeric isoforms have been linked to early diagnosis of tumorogenesis, emphasizing their potential as important biomarkers for several types of cancer. Among the mutations studied, those of the active site Arg132 in hICDH led not only to an underproduction of the α-ketoglutarate (αKG) product, but also to the release of a side product of the reaction, 2-hydroxyglutarate.[441,444,445] The latter product competes with αKG for the regulation of the activity of several αKG-dependent enzymes involved in histone and DNA methylation, processes that have been related to eventual tumorogenesis.[446-448] Additionally, recent findings indicate that during the tricarboxylic acid cycle (TCA), ICDH can also catalyze reverse reductive carboxylation cycles under hypoxia conditions producing acetyl-CoA, which is fundamental for the synthesis of many macromolecules by the cell.[449,450]

Considering the importance of the many diverse cellular pathways in which hICDH is involved (such as DNA methylation, oxidative stress response, fatty acid synthesis, among others),[451-453] together with its clinical relevance, and the current interest in the chemistry of hydride transfer reactions,[84,454] high-level theoretical calculations of its reaction mechanism are of the outmost importance to finally understand the catalytic process, and provide rigorous computational support to the experimentally proposed mechanisms.

5.1.2. General features

ICDH is responsible for the catalytic conversion of isocitrate (ICT) and $NADP^+$ to α-ketoglutarate (αKG), NADPH and CO_2. The hICDH that we have studied is an asymmetric dimer with two similar active sites, each with contributions from residues of both subunits. Each hICDH monomer

comprises three main domains: a large domain (Met1 – Leu103 and Gly286 – Leu414), a small domain (Gly104 – Gly136 and Asp186 – Tyr285) and a clasp domain (Asp137 – Gln185). Both the small and clasp domains are responsible for the dimerization of hICDH. Additionally, the clasp domain is of significant importance to hold the enzyme's catalytic site together in the dimer.[455] Several experimental studies have shown that these structural features, as well as key catalytic residues, are conserved in different organisms, and different organelles within an organism, for metal-dependent NADP-linked β-hydroxyacid oxidative decarboxylase enzymes.[442,443,455,456]

The catalysis takes place in a closed-conformation quaternary complex,[455,456] and involves significant conformational changes as the divalent metal (Mg^{2+} or Mn^{2+}), the $NADP^{+}$ cofactor and the trianionic form of the isocitrate substrate (ICT) sequentially bind. Figure 5.1 shows a representation of the fully-closed conformation (A), and of one of the two catalytic sites (B).

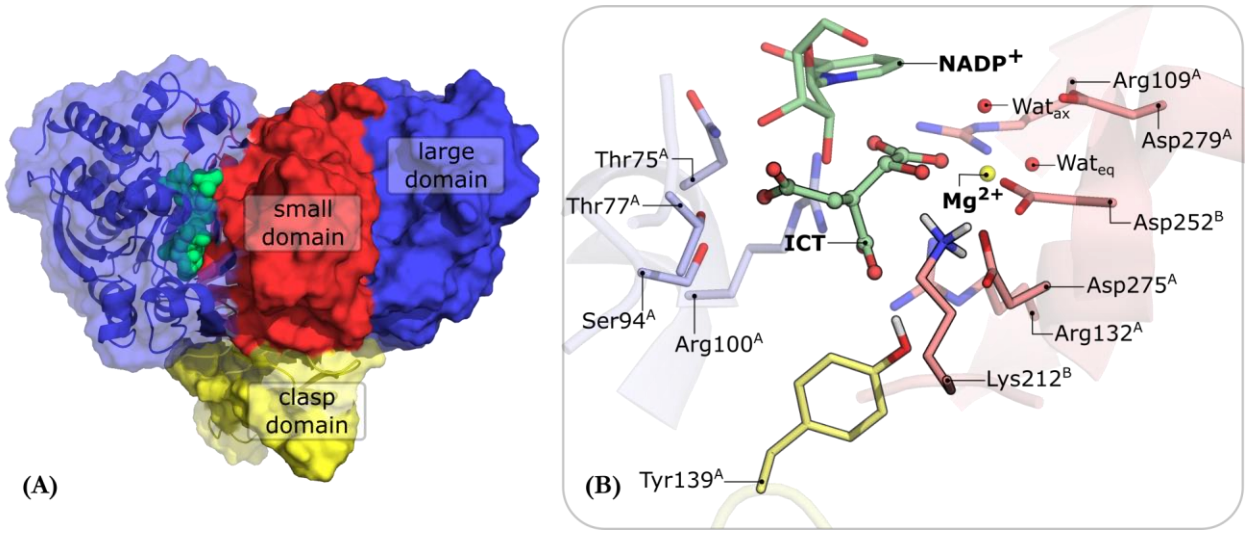
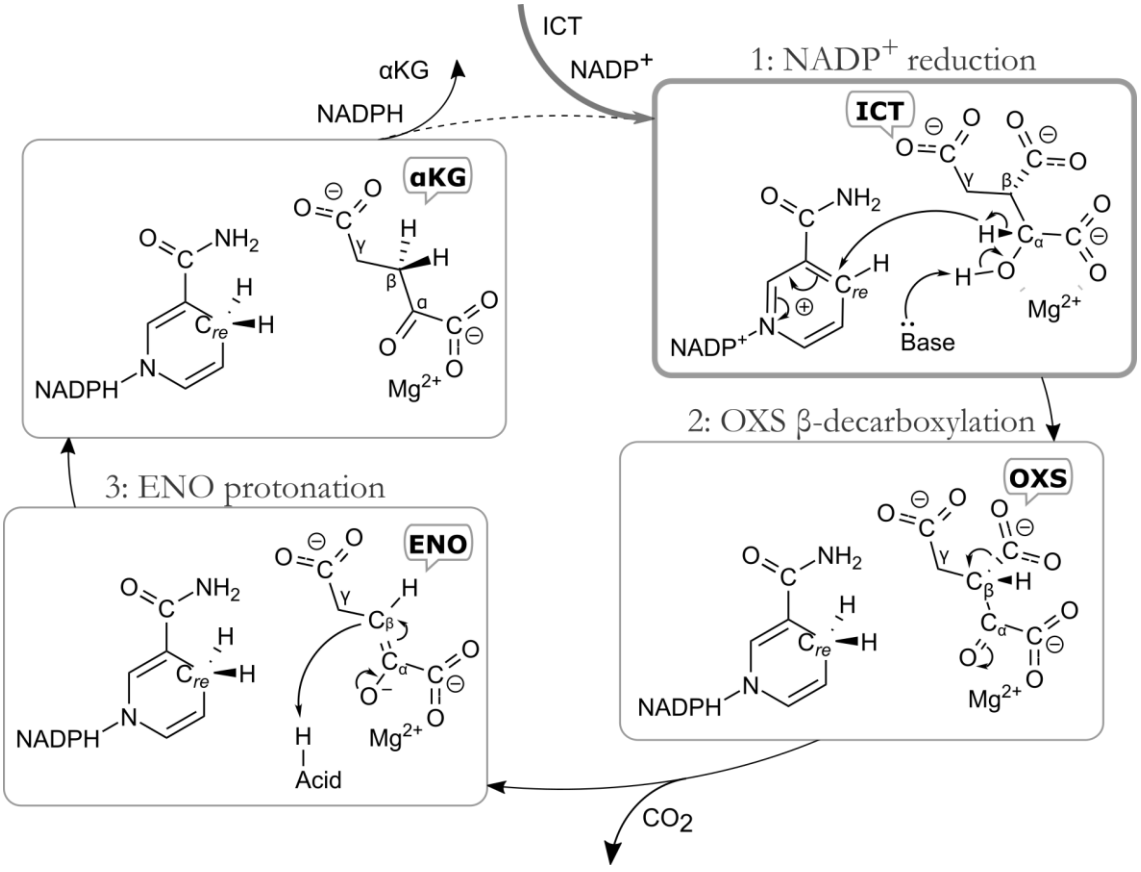


Figure 5.1. (A) hICDH, colored by domain (blue for the 'large domain', red for the 'small domain' and yellow for the 'clasp domain'). ICT, the Mg^{2+} ion and the $NADP^{+}$ cofactors are shown as green spheres; the monomer on the left-hand side is represented in cartoon and transparent surface, and the monomer on the right-hand side is represented with an opaque surface. (B) Residues within a 4 Å radius of the substrate (in ball and stick representation) plus the Mg^{2+} and $NADP^{+}$ cofactors; the protein carbon atoms are colored according to their domain (large, small or clasp), and the proposed catalytic residues are outlined with a black contour.

In the catalytically productive conformation, ICT is anchored by a network of hydrogen bonds as well as by a positively charged environment generated by four Arg/Lys residues close by, shown in Figure 5.1B. The mutation of the Arg around ICT ($Arg100^{A}$, $Arg109^{A}$, $Arg132^{A}$) to Ala has been shown to compromise the catalytic efficiency. Together with $Thr77^{A}$, $Ser94^{A}$ and $Asp275^{A}$, these three Arg are also responsible for substrate recognition.[441]

5.1.3. Catalytic/Kinetic insights

The chemical mechanism is proposed to proceed in three stages and to follow a steady state random mechanism (Scheme 5.1).[440,457,458] First, reduction of $NADP^+$ to NADPH occurs by dehydrogenation of the C_α-carbon of ICT, generating oxalosuccinate (OXS). This is followed by β-oxidative decarboxylation of OXS to form enolate (ENO) and CO_2, the latter leaving the catalytic site; ENO is then protonated by a general base, resulting in αKG.[458] The k_{cat} of ICDH in several bacteria was experimentally determined to be in the range between 11 s^{-1} and 38 s^{-1},[441,443] and the rate-limiting step of the catalytic cycle is the release of the product and of NADPH, which is about 16 times slower than the chemical reaction.[459] Still, there is no consensus on which is the rate-determining step in what concerns the chemical catalysis by the ICDH.[440,443,458]



Scheme 5.1. Catalytic cycle proposed for the metal-dependent $NADP^+$-linked β-hydroxyacid oxidative decarboxylases (hICDH included).[440,457,458]

Currently, an universal catalytic mechanism was proposed for the NAD(P)$^+$/metal-dependent enzymes responsible for the decarboxylation of β-hydroxyacid substrates.[458] The proposal considers a residue triad (Lys212^B-Asp275^A-Tyr139^A in ICDH) to be responsible for catalysis. Specifically in ICDH, several structural and kinetic studies have shown that Lys212^B and Tyr139^A have indeed a significant effect in the affinity for ICT. Moreover, the optimal pK_a (~ 5.2) of fully-closed ICDH suggests that Lys212^B might be deprotonated by a nearby aspartate (Asp275^A, Asp279^A or Asp252^B), due to the very positive electrostatic environment generated by the conserved arginines in the active site (Arg100^A, Arg109^A, Arg132^A).[460,461] According to the universal proposal, in the first step of the mechanism, the basic Lys212^B would deprotonate the ICT-hydroxyl, facilitating the hydride transfer from ICT to NADP$^+$.[439,440] The resulting OXS should be unstable, and the acidic metal ion should quickly promote its decarboxylation, generating ENO, which would be protonated by Lys-212^B to give origin to an enol intermediate (ENL). In the last step, Tyr-139^A is proposed to protonate the C_β-carbon of ENO and generate αKG.[440]

The most stable intermediates have been detected and characterized experimentally.[459,462,463] However, several questions remain unanswered: (1) during NADP$^+$ reduction, Asp279^A or Asp252^B were proposed to deprotonate the hydroxyl of ICT, facilitating the dehydrogenation of C_α-carbon from ICT,[438,463] but there is no confirmation of this; (2) it is not known if the two reactions cited in point (1) are concomitant, as this alkoxyde intermediate (AKO) has not been detected so far; (3) even though the exit of the αKG product is the rate-limiting step of the catalysis, it has not been determined which is the rate-limiting step of the chemical reaction; (4) after the protonation of ENO that produces αKG, there is still no consensus if the protonated state of Tyr139^A is restored by a bulk solvent molecule or by Asp275^A. At the end of our study, we were able to provide insight into further clarifying these unanswered questions.

5.2. Computational Methods

5.2.1. Molecular model for hICDH

Since there was no structure of the catalytically competent form of hICDH, and few fully-closed conformations of the enzyme have been obtained so far,[442] we built our model from two X-ray structures: a fully-closed structure of the human enzyme complexed with Ca^{2+}, αKG and NADPH (PDB code: 3INM),[444] and the fully-closed quaternary complex of Escherichia coli *ICDH*, complexed with Ca^{2+}, ICT and $NADP^{+}$ (PDB code: 4AJ3).[442] In fact, several studies have shown that both the structure and the catalytic sites of cytosolic ICDH are extremely conserved across organisms,[442,458,464,465] and this is also verified for the *H. sapiens* and *E. coli* forms of the enzyme.[442] We completed the N- and C-terminus of the monomer (Met1 to Lys4 and Ala410 to Leu414) of the cytosolic *Ca^{2+}-ICDH* published by Dang and co-workers[444] with a cytosolic hICDH in an open conformation (PDB code: 1T0L).[455] We used the PISA server[466] to obtain the homodimeric *Ec. Ca^{2+}-ICDH*, and then modelled each of the active sites of our hICDH with the relevant conserved residues from the active site of the *E. coli* enzyme in the closed conformation ($Lys72^{A}_{100}$, $Thr77^{A}_{105}$, $Ser94^{A}_{113}$, $Asn96^{A}_{115}$, $Arg100^{A}_{119}$, $Arg109\ ^{A}_{129}$, $Arg132^{A}_{153}$, $Tyr139^{A}_{160}$, $Asp275^{A}_{307}$, $Glu306^{A}_{336}$, $Lys212^{B}_{230}$, $Asp252^{B}_{283}$), and the ICT, $NADP^{+}$ and Ca^{2+} counterparts of the quaternary complex (see Figure S1 in Supporting Information (SI)). Furthermore, we have replaced the Ca^{2+} cofactor by Mg^{2+}, which has been shown to be the natural metal cofactor of hICDH, together with Mn^{2+}.[444,467] All crystallographic waters in a range of 8 Å from the catalytic site of the template hICDH were conserved in the final model of the enzyme. A PDB file with the model was deposited in SI.

The enzyme was geometry optimized using Molecular Mechanics calculations, with the AMBER 12 software,[468] considering two possible protonation states for the $Asp275^{A}$/$Lys212^{B}$ side chains: both charged (as in water at neutral pH), and both neutral, in agreement with the experimental suggestions.[458] The FF99SBildn force field[469-471] was employed to describe the enzyme, while ICT was parameterized with GAFF[298] and Merz-Kollman charges,[131] derived from a RESP fitting[134] of the electrostatic potential calculated in vacuum at the HF/6-31G(d) level, using the Antechamber tool.[472] We employed a single conformation for the ICT, retrieved from the *Ec. ICDH* PDB file, since we have taken its active site conformation as the most representative of a catalytically competent ICDH. Parameters for $NADP^{+}$ were taken from the literature.[472] After stripping all waters farther than 3 Å from ICT, we have obtained a model for the enzyme with 13269 atoms.

To perform classical Molecular Dynamics (cMD) simulations for both protonation states, the heating of the systems was conducted linearly for 50 ps from 0 K to 310 K, in an NVT ensemble, and followed by another 50 ps of NVT MD at 310 K. With this approach, the pressure in the system increases with temperature in a controlled manner, while the density of the system is kept fixed. Subsequently, we ran a 20 ns cMD with an NPT ensemble for the two different protonation states, with the system solvated by TIP3P water molecules,[299] in a rectangular box whose faces were positioned at least at 12 Å from any protein atom. The NPT ensemble will reproduce the conditions that are expected to be verified in the cytoplasm of mammal cells, and will regulate the extent of the system's expansion. We employed the Langevin thermostat and the Berendsen barostat to fix the temperature and pressure of the system (310 K and 1 bar), and treated non-bonded coulombic interactions with the particle-mesh-Ewald method.[155] We defined a radius of 10 Å as *cutoff* for short-range electrostatic and Lennard-Jones explicit interactions. The SHAKE algorithm[165] was used to constrain all bonds involving hydrogen atoms to use an integration step of 2 fs. We have monitored the root-mean square deviation (RMSd) for the residues closest to ICT, Mg^{2+} and $NADP^{+}$ and for the whole protein. The RMSd shows that the simulation with protonated $Lys212^{B}$/deprotonated $Asp275^{A}$ shows the highest deviation to the crystallographic model for both the protein and the active site residues (refer to Figure S2 to detailed analysis). The set of residues highlighted in Figure 5.1B was employed to determine a representative structure for each simulation. To do so, we have performed a cluster analysis with the average linkage algorithm,[473] following the RMSd for the referred residues, from which 10 representative structures resulted. We have retrieved the representative structure of the most populated cluster in each simulation to compare its active site with that of our starting model of hICDH (from crystallography). Further discussion is provided in SI (Figure S3).

5.2.2. Defining the layers for the model and real system

Recent years witnessed the use of a wealth of different computational approaches to determine enzyme reaction mechanisms.[33,53,56-64,99-102] Here we resorted to QM/MM PES calculations using DFT in the QM region. Despite that these type of calculations present limitations, namely: the lack of sampling to accurately describe the environment during the chemical reaction, the smaller size of the QM system in opposition with cluster QM model approaches, and the treatment of the interactions between the QM and MM layers; overall, they still present one of the best compromises between the accuracy of DFT calculations and the average representation of the

anisotropy of the enzyme's steric and electrostatic environment during a chemical reaction, with moderate computational time required.[57,62]

We built a two-layered ONIOM model using Density Functional Theory (DFT) and Molecular Mechanics (MM). The DFT region included 134 atoms. We included the residues of the active site identified in the literature as being the most relevant for catalysis and substrate specificity, together with ICT, Mg^{2+} and $NADP^+$. The magnesium ion was included with its two water molecules, the sidechains of $Asp275^A$ and $Asp252^B$, and the sidechain of $Asp279^A$ in its second coordination sphere. The sidechains of $Tyr139^A$ and $Lys212^B$ (starting from the C_γ-carbon), which are known to be catalytic residues, were also included. We also included the guanidinium moieties of $Arg100^A$, $Arg109^A$ and $Arg132^A$, which establish electrostatic interactions with the α- and β-carboxylates of ICT. The hydroxyl-methyl region of the sidechain of $Thr77^A$ and $Ser94^A$, that were hydrogen bonded to the γ-carboxylate of ICT, were also included. $NADP^+$ was represented by its nicotinamide moiety. The remaining atoms were described at the MM level. A PDB file with the complete list of atoms included in each layer was deposited in SI.

5.2.3. QM/MM calculations

The QM/MM calculations were performed using the electrostatic embedding scheme as implemented in Gaussian 09.[282] With a careful selection of reaction coordinates, we ran the ONIOM geometry optimizations at the B3LYP/6-31G(*d*):FF99SB[37,38,41,202,283,284,364,365,474] level to determine the potential energy surfaces (PESs) for $NADP^+$ reduction, oxidative β-decarboxylation and formation of αKG. As B3LYP has been extensively benchmarked (its strengths and drawbacks are well-known),[188,212] and no DFT benchmark has been published for the specific chemistry of this enzyme, we considered it a good choice to study the PES for this kind of reactions; B3LYP has been successfully used in several other QM/MM studies of the catalytic mechanism of enzymes.[87,475,476]

We characterized all stationary states of the catalytic cycle (minima and saddle points, each one confirmed by the right number of imaginary frequencies). For the complete ONIOM model, we calculated the harmonic zero-point energy (ZPE), thermal energy and entropy for each stationary state with the harmonic approximation as implemented in the Gaussian09 software package.[228] The ONIOM electronic energy was calculated with a larger triple-zeta basis set (6-311+G(2*d*,2*p*)),[37-42,474,477] and with B3LYP (with and without the D3 correction to dispersion),[345]

and the M06-2X density functionals.[433] The latter has been pointed out in several benchmark studies as accurate for general main group thermochemistry and kinetics.[188,212] Our calculations indicated that the D3 correction increases the energy of reaction for OXS, ENO and αKG formation in about 3 kcal·mol^{-1}, and does not significantly alters the reaction barriers (only the barrier for the $NADP^+$ reduction decreases in 1.8 kcal·mol^{-1}). Table S1 in SI compares the PESs obtained with B3LYP, B3LYP-D3 and M06-2X.

The Gibbs free-energy profile for the reaction was obtained at the M06-2X/6-311+G(2*d*,2*p*):FF99SB//B3LYP/6-31G(*d*):FF99SB level of theory, with geometry, ZPE and thermal corrections determined at the B3LYP/6-31G(*d*):FF99SB level of theory. Since the active site of ICDH is significantly buried in the enzyme, and the Gibbs free-energy solvation methods for large QM/MM still lack the accuracy of electronic calculations, we did not account for the very small contribution from the solvent during catalysis. We performed an analysis of the Hirshfeld charge population,[478-480] at the M06-2X/6-311+G(2*d*,2*p*):FF99SB//B3LYP/6-31G(*d*):FF99SB level of theory in all the stationary states.

5.3. Results and Discussion

5.3.1. Modelling of the enzyme's active site

The QM/MM geometry optimization of the minimized structures of the charged protonation state of ICDH indicated that Lys212^{B} and Asp275^{A} should be in their neutral form, since Asp275^{A} spontaneously deprotonates the positively charged Lys212^{B}. Our optimized structure presents an active site, which is very similar (RMSd=0.95 Å) to the X-ray structure of *E. Coli* ICDH. This result agrees with the proposal by Aktas *et al*,[440,458] in which the authors indicate that Lys212^{B} should be deprotonated by Asp275^{A} in order to subsequently deprotonate the C_{α}-hydroxyl from ICT. Further detail on the optimization of our QM/MM model are provided in SI.

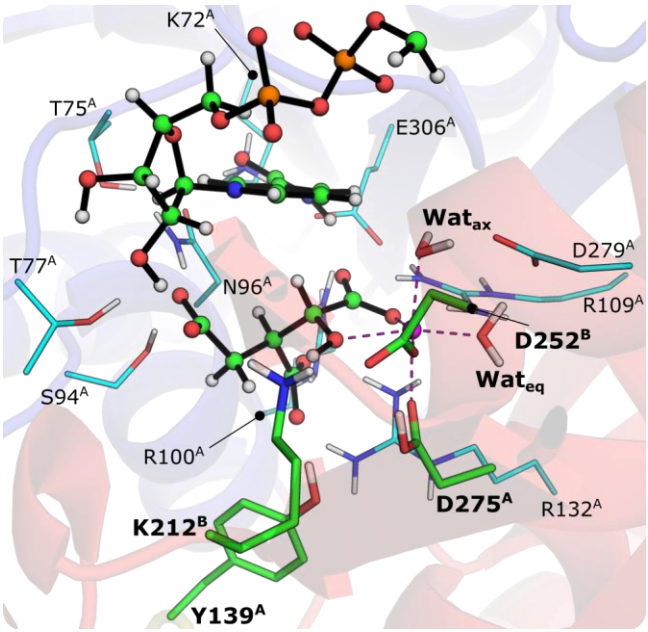


Figure 5.2. The QM/MM geometry–optimized active site of hICDH. The names of the catalytic residues are highlighted in bold. Mg^{2+}, ICT and $NADP^{+}$ are shown in black-stick and sphere representation. The residues shown in cyan-sticks establish important hydrogen bonds and ionic bridges with ICT, the nicotinamide moiety of $NADP^{+}$ and the coordination sphere of Mg^{2+}.

In Figure 5.2, we outline the large network of ionic and hydrogen bond interactions that anchor ICT and the nicotinamide moiety of $NADP^{+}$. The Mg^{2+} cofactor is hexa-coordinated to the C_{α}-hydroxyl and α-carboxylate groups of ICT, to the sidechains of Asp275^{A} and Asp252^{B}, and to two orthogonal water molecules, as referred in the literature.[456] The hydroxyl group of ICT is oriented also to Lys212^{B} while Arg109^{A} interacts closely with the α-carboxylate of ICT with a coplanar ionic

interaction. Additionally, two other Arg are found in the surroundings of the α- and β-carboxylates of ICT; the Arg100^{A} establishes ionic interactions with both the α- and β-carboxylates, and the Arg132^{A} exhibits a short distance to the β-carboxylate and Asp275^{A}. Tyr139^{A} is oriented towards the β-carboxylate of ICT establishing a short hydrogen bond, which has been referred in the literature to be essential for the reduction of $NADP^+$.[457] In the γ-carboxylate side of ICT, we can observe a large network of hydrogen bonds from Thr77^{A}, Ser94^{A}, Asn96^{A} and the $NADP^+$ ribose.[442,465] Finally, $NADP^+$ is anchored in the active site by a tetrad composed of Lys72^{A}-Thr75^{A}-Asn96^{A}-Glu306^{A}, similar to the one found in *E. coli*.[442]

Additionally, we have performed also a 20 ns cMD simulation of hICDH with standard protonation states, and with the neutral Lys212^{B}/acid Asp275^{A} pair, as proposed in the literature.[440,443,458] Overall, our observations indicate that the state in which Lys212^{B} and Asp275^{A} are neutral shows the highest similarity to the active site of the fully-close ICDH of *E. coli*. The analysis of the RMSd of the protein and the residues in our model system, along with a detailed atomistic representation of the active site of ICDH for both protonation states, is given in SI (Figure S2 and Figure S3).

In the following sections, we will discuss the kinetics and thermodynamics of the reaction mechanism. In SI (Figure S6 to S8) we provide the comparison between our characterized intermediates and those experimentally determined.[442,444,457]

5.3.2. Which base facilitates $NADP^+$ reduction?

In this section we discuss which base will be more likely to deprotonate the hydroxyl from ICT to lower the barrier for ICT dehydrogenation. Despite the fact that both experimental studies and our ONIOM optimized model favor the Lys212^{B}/Tyr139^{A} pair,[440,442,443] literature has suggested that an Asp279^{A} may also function as base, assisted by an Mg-coordinated water.

We attempted to study this hypothesis starting from the model of ICDH with charged Lys212^{B} and Asp275^{A} (see Figure S5 in SI; a PDB file with the structure is also appended). The resulting alkoxyde (AKO) is not a stationary point of the Gibbs free-energy profile of $NADP^+$ reduction ($\Delta G^{\ddagger}$ = 12.2 kcal·mol^{-1} and ΔG_{deprot} = 13.8 kcal·mol^{-1}), and the C_α-hydride is transferred to the C_{re}-carbon of $NADP^+$ with an additional cost of 7.6 kcal·mol^{-1}, originating OXS (ΔG_{dehyd} = -4.4 kcal·mol^{-1}). The overall activation Gibbs free-energy for the whole process is 21.4 kcal·mol^{-1}, which is much beyond the 16 kcal·mol^{-1} Gibbs free-energy limit estimated by the experimental k_{cat}. When OXS is formed, the bond between the C_β-carbon and the β-carboxylate in OXS is already elongated

and β-decarboxylation of OXS follows quickly (with a $\Delta G^{\ddagger}$ of 11.5 kcal·mol^{-1}), forming ENO. After CO_2 exits the active site, the protonation of the C_β-carbon of ENO by Tyr139^A occurs with an activation Gibbs free-energy of 18.9 kcal·mol^{-1}, once again quite larger than the experimental limit; moreover, the structure with the αKG product is more unstable than ENO (ΔG_{prot} = 0.4 kcal·mol^{-1}), contrary to what would be expected as the exit of αKG and NADPH are the slowest steps of the full catalytic process.[443] Summing up all evidences, the catalysis should not proceed through the Asp279^A/Tyr139^A pair, since two of the most important reactions in the catalytic cycle seem to be highly compromised in such conditions. Hence, we conclude that the mechanism in which Lys212^B acts as a base for the substrate is the only alternative. In Table 5.1 we provide an overview of the electronic energy profiles for both the Asp279^A and Lys212^B driven mechanisms. In the next sections, we will discuss only the mechanism based on the Lys212^B/Tyr139^A neutral pair.

Table 5.1. Gibbs free-energies for Asp279^A- and Lys212^B-based reaction mechanisms. The double-split line separating ENO and ENO$_{C2}$ refers to the exit of CO_2 from the active site; a process that we have not approached in our study. ICT, AKO, OXS, ENO, ENO$_{C2}$ and αKG, refer to the stationary minima in the reaction profile; TS$_{deprot}$, TS$_{dehyd}$, TS$_{decarb}$ and TS$_{prot}$ refer to the saddle points in the reaction profile for the deprotonation of ICT, the dehydrogenation of AKO, the decarboxylation of OXS and the protonation of ENO, respectively.

	ΔG / kcal·mol^{-1}	
	Asp279^A/Tyr139^A	*Lys212^B/Tyr139^A*
ICT	0.0 (-3522.0916)	0.0 (-3522.5262)
TS$_{deprot}$	12.2	1.5
AKO	13.8	4.7
TS$_{dehyd}$	21.4	13.4
OXS	-4.4	-7.6
TS$_{decarb}$	7.1	8.9
ENO	-5.8	1.42
ENO$_{C2}$	0.0 (-3333.3820)	0.0 (-3333.8310)
TS$_{prot}$	18.9	10.8
αKG	0.4	-23.8

5.3.3. Reduction of NADP+

Starting from the optimized model depicted in Figure 5.2, the Lys212^B nitrogen is adequately positioned to deprotonate the C_α-hydroxyl of ICT (N_{Lys212^B}−HO_{ICT}: 1.72 Å) and the C_{re}-carbon of NADP$^+$ is properly positioned to receive the C_α-hydride ($C_\alpha H$−$C_{re}H$: 2.60 Å and angle $C_\alpha H\cdots C_{re}$−H

of 106.59°). Our calculations have shown that the hydroxyl deprotonation corresponds to a stationary point in the zero-Kelvin electronic energy PES but not to a minimum in the thermal Gibbs free-energy profile ($\Delta G^{\ddagger}$ is 1.5 kcal·mol^{-1} and ΔG_{deprot} is 4.7 kcal·mol^{-1}); hence, the resulting AKO is not a stable intermediate of the cycle. No significant nuclear motions are observed in this reaction. However, there is an increase in electron density in the resulting C_{α}-alkoxide that is symmetric to the change of the Lys212^{B}-nitrogen (about 0.15 au) that should strongly favour the charge transfer for the nicotinamide moiety (see Figure 5.3).

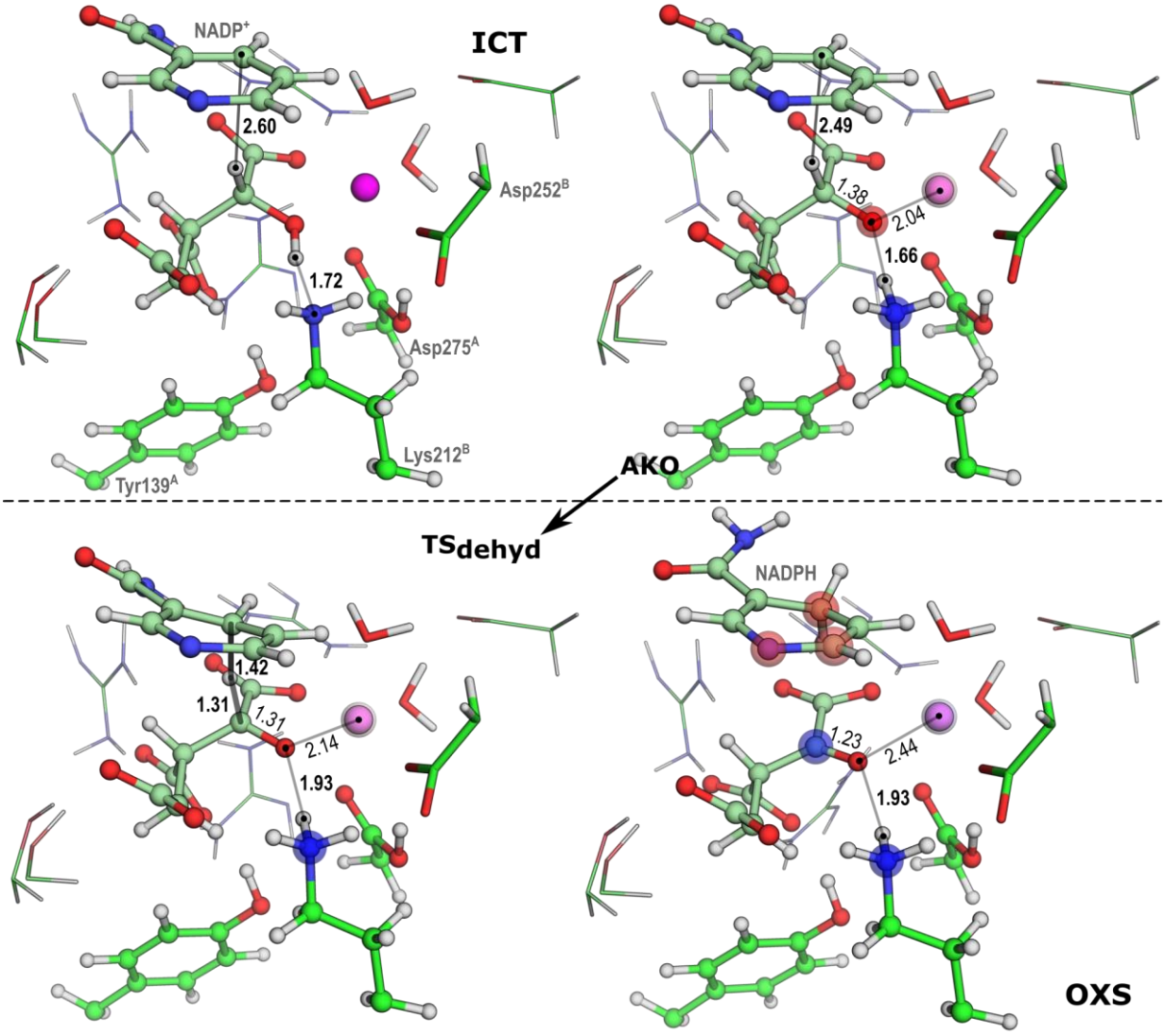


Figure 5.3. Stationary points of the NADP^{+} reduction step – ICT, TS$_{dehyd}$, and OXS. The atomic charge variation, relative to the ICT state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations.

$NADP^+$ reduction occurs rapidly, with the C_α-hydride approaching the C_{re}-carbon to produce OXS. The transition state (TS_{dehyd}) has the C_α-hydrogen at a distance of 1.41 Å from the C_{re}-carbon and 1.30 Å of C_α-carbon. The activation Gibbs free-energy of the step is of 13.4 kcal·mol^{-1}, a value that is well within the range of the values of the literature.[441,443,457] We analysed correlations between the energy profile for the reaction and several coordinates that change considerably from AKO to OXS, and our attention was drawn to the stretching of the bond between Mg^{2+} and the C_α-alkoxide of AKO. Before the TS_{dehyd} its increase is monotonic but not pronounced; however, after the TS_{dehyd} both the Gibbs free-energy and this distance vary more quickly. This observation indicates that this interaction should be the main source of strain during the hydride transfer from AKO to $NADP^+$. In the TS_{dehyd}, there has been a significant loss of electron density in the C_α-alkoxide group, and the bending of the C_{re}-carbon has placed the carbamide group of $NADP^+$ outwards the C_α-carboxylate plane. These changes provide for small shifts in the C_α- and C_β-carboxylate, resulting in a slight stretch of the bond between the C_β-carbon and the β-carboxylate 1.54 Å to 1.57 Å. This small variation anticipates the oxidative β-decarboxylation of OXS. The reaction Gibbs free-energy is of -4.7 kcal·mol^{-1} and it occurs with small structural changes in the active site, as previously described in literature.[457,481] The re-hydrogenation of OXS back to the ICT substrate is unlikely to occur ($\Delta G_{OXS \rightarrow ICT}$=21.0 kcal·mol^{-1}). In SI we provide the coordinates for all the stages of the $NADP^+$ reduction.

5.3.4. Oxidative β-decarboxylation of OXS

No major structural changes take place at the active site during the stretching of the bond between the C_β-carbon and the β-carboxylate of OXS that leads to the transition state (TS_{decarb}). We note, in particular, the shortening of the distance between the C_α-carbonyl and Mg^{2+} (from 2.44 Å to 2.09 Å) at a linear rate with the decarboxylation of OXS. In Figure 5.4, an analysis of the atomic charges indicates that this synchronous behaviour might be related to the increase in electron density in the C_α-carbonyl of OXS (δ^- ~ 0.20 au) that is most likely coming from the leaving CO_2, where the largest loss of electron density occurs (δ^+ ~ 0.30 au).

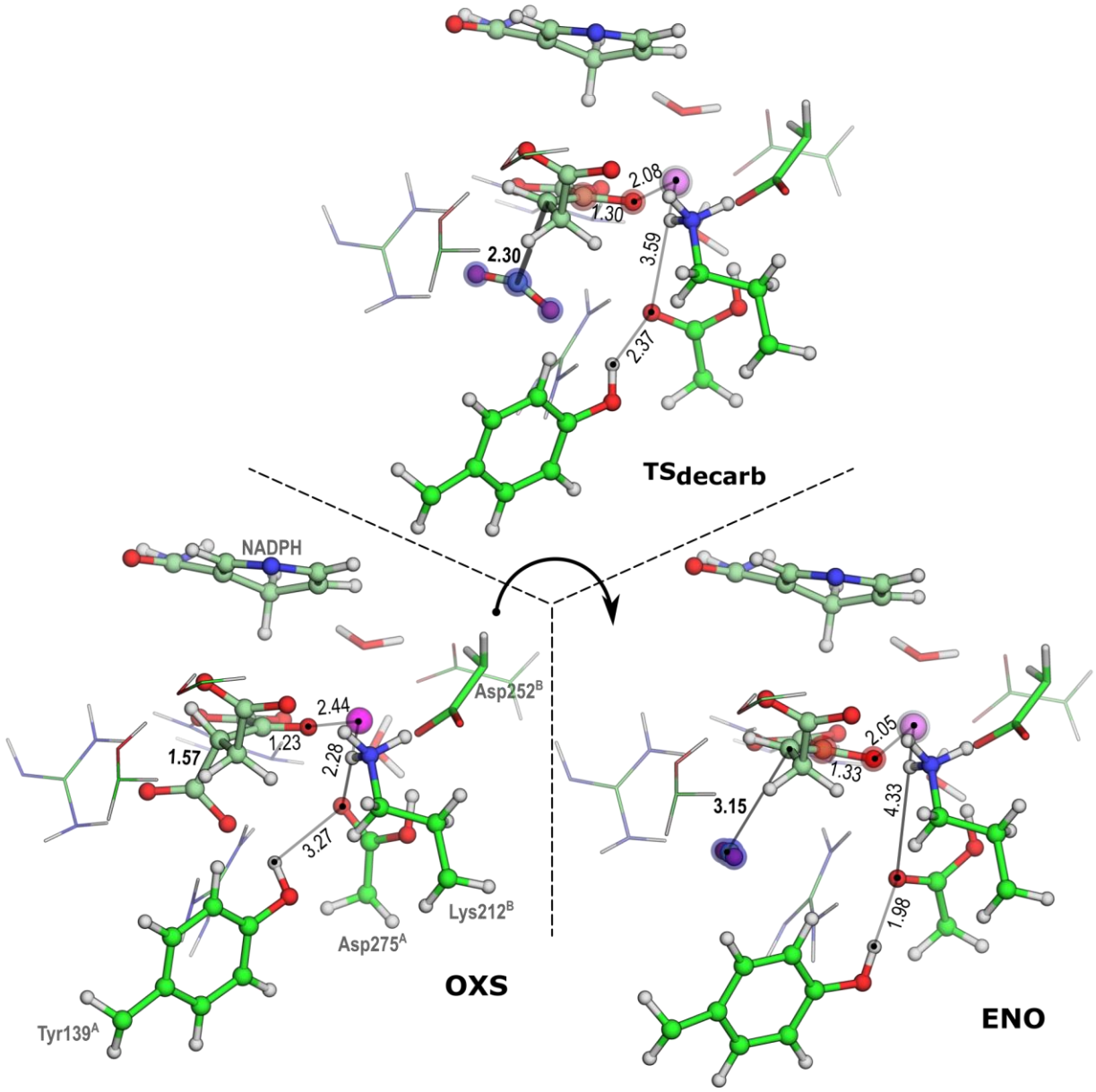


Figure 5.4. Stationary points of the OXS β-decarboxylation step – OXS, TS_{decarb}, and ENO. The atomic charge variation, relative to the OXS state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations.

At the TS_{decarb}, the leaving CO_2 is 2.30 Å away from the C_β-carbon, and the stretching of $Mg{-}O_{Asp275}$ to 3.59 Å breaks the distorted octahedral geometry of Mg^{2+}. The $Asp275^A$-oxygen is now facing the $Arg132^A$-guanidinium and the hydroxyl from $Tyr139^A$ (2.92 Å and 2.37 Å, respectively). We suggest that this conformational change should be a key feature to stabilize the active site in TS_{decarb}, since it flexibilizes the active site to shield the repulsion in the positive Arg

environment that results from the decarboxylation. In fact, as TS_{decarb} is converted to ENO, $Asp275^{A}$ establishes shorter hydrogen bonds with $Arg132^{A}$ (1.86 Å) and $Tyr139^{A}$ (1.89 Å), that were previously anchoring the leaving CO_2. The activation Gibbs free-energy for this step is 16.5 kcal·mol^{-1}, which is near the upper activation Gibbs free-energy expected for the limiting-step of the catalysis (~ 16.3 kcal·mol^{-1}). We will discuss below how the gain in entropy is critical for this step to be feasible, and in what extent does the metal contributes to the process.

Figure 5.4 provides for further geometric detail on the chemical step that we discussed. Additionally, we provide the coordinates for all the stages of β-decarboxylation in the SI. The β-decarboxylation of OXS is an endergonic step, with a reaction Gibbs free-energy of 9.0 kcal·mol^{-1}. After CO_2 leaves the active site, significant conformational changes take place. $Asp275^{A}$ exhibits a large bond to Mg^{2+} ($Mg{-}O_{Asp275}$: 2.74 Å), overcoming an activation barrier lower than 3 kcal·mol^{-1} and the hydroxyl moiety of $Tyr139^{A}$ moves towards the C_β-carbon of ENO (to a final distance of 2.29 Å). At this point, $Tyr139^{A}$ is properly oriented to be deprotonated by the C_β-carbon (the angle of the attack $O_{Tyr139}{-}H{\cdots}C_\beta$ is 174°).

5.3.5. Enolate protonation

It was suggested that the ENO alkoxide moiety should be quickly protonated by $Lys212^{B}$ generating an enol intermediate (ENL), which would tautomerize to produce αKG.[440,458] We studied the protonation of the alkoxide moiety of ENO using the two closest acids, $Lys212^{B}$ and $Asp275^{A}$, (1.83 Å and 2.77 Å). Protonation by $Asp275^{A}$ did not lead to a stationary point. Protonation by $Lys212^{B}$ could generate an unstable ENL through a barrier of ~ 6 kcal·mol^{-1} and an almost identical reaction energy, just a few tenths of kcal.mol^{-1} below. Anyway, ENL reverses to ENO in the following reaction step, as $Tyr139^{A}$ approaches the C_β-carbon to protonate it, which suggests that ENL should not be a relevant species in the last step of the catalysis. Therefore we followed the mechanism starting from ENO.

The protonation of ENO by $Tyr139^{A}$ exhibits an electronic activation energy of about 14 kcal·mol^{-1}, and the reverse electronic activation energy is of *c.a.* 1 kcal·mol^{-1}, which means that this reaction, to be thermodynamically favourable, has to include an acid that restores the protonation state of $Tyr139^{A}$, lowering the energy of the products. The neighbouring residues $Lys212^{B}$, $Asp275^{A}$ or $Arg132^{A}$ are the most likely candidates. Furthermore, experimental data has consistently suggested that the protonation of the C_β-carbon of ENO should be aided by a catalytic $Asp275^{442}$ or by a proton relay system from the bulk solvent.[443,482]

Despite the fact that Asp275^{A} was far from Tyr139 (4.24 Å) and not in a proper orientation to be deprotonated (O_{Asp275}−H···O_{Tyr139} angle of 61.74°), the calculated PES for the approximation of Tyr139^{A} and Asp275^{A} shows that it is Asp275^{A} that exhibits the most significant changes (see ENO_{C1} and ENO_{C2} in Figure 5.5).

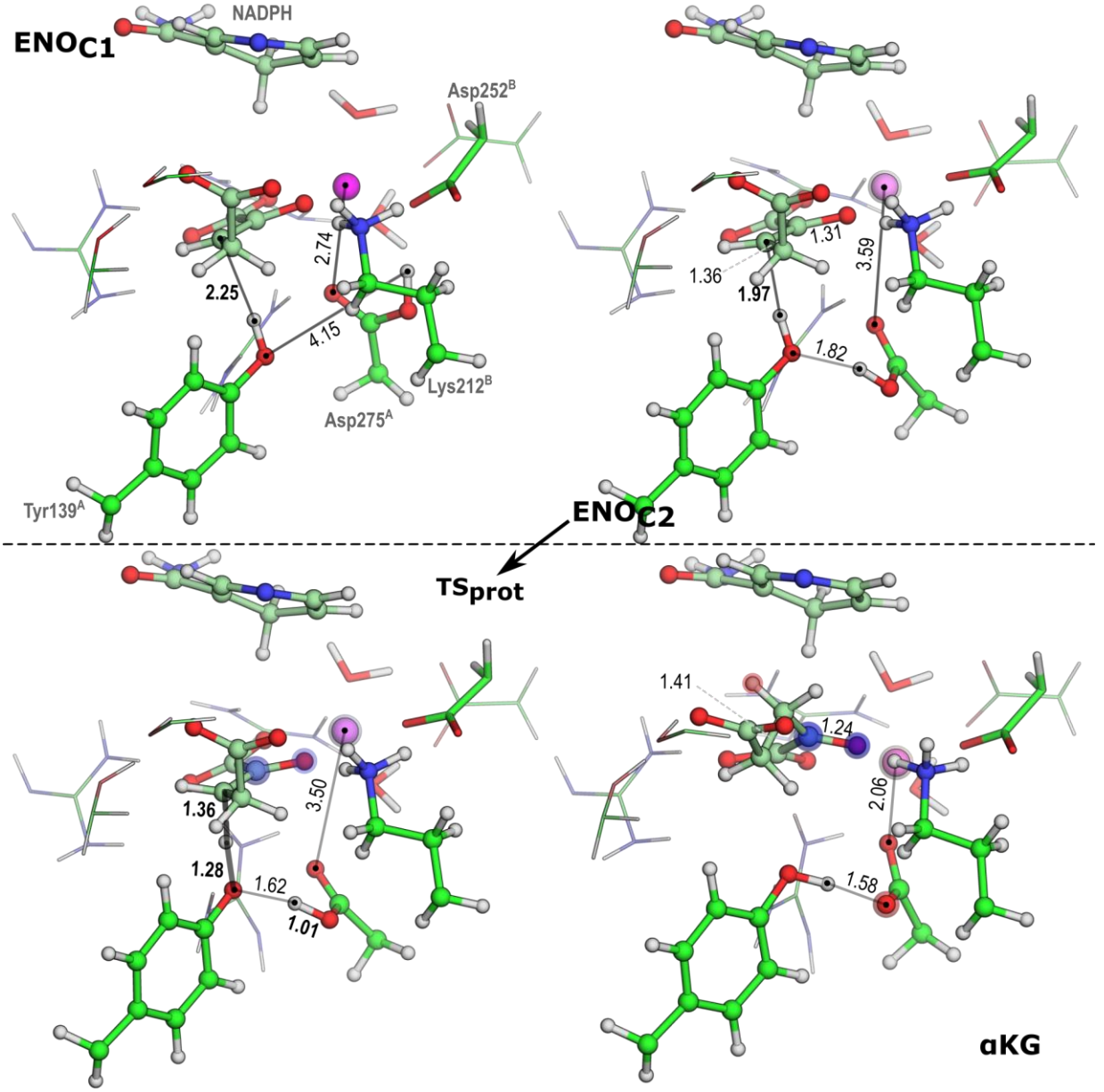


Figure 5.5. Stationary points of the ENO protonation step-ENO, TS_{prot} and αKG. The atomic charge variation, relative to the ENO_{C1} state, is represented by blue-shadowed (decrease of electronic density) and red-shadowed (increase of electronic density) spheres; and it is only highlighted for the atoms exhibiting the largest electron density variations.

The activation barrier for this conformational change is of ~ 6 kcal·mol^{-1}, and the Tyr139^{A} ends up significantly closer to the C_β-carbon of ENO. From this stage on, the deprotonation of the Tyr139-hydroxyl by the C_β-carbon is facile ($\Delta G^{\ddagger}_{prot}$=10.8 kcal·mol^{-1}). At TS_{prot} the transferred proton is at a distance of 1.36 Å from the C_β-carbon and 1.28 Å from the Tyr139-oxygen. Figure 5.5 shows the loss of electron density on the C_α-carbonyl group (δ^+ ~ 0.18 au) as the proton is captured from Tyr139^{A}. In the same step, the Asp275-hydroxyl is deprotonated by the Tyr139-oxygen in a highly asynchronous manner, and Asp275^{A} again chelates to Mg^{2+} (2.06 Å). An important observation is the increase in electron density in the carboxylate moiety of Asp275^{A} (δ^- ~ 0.22 au), depicted in Figure 5.5, which lead us to propose that ENO is in fact protonated by Asp275^{A} via Tyr139^{A}.

During the chelation of Asp275^{A} to Mg^{2+}, there is also a $HC_\beta-C_\gamma H$ conformational change in αKG that places the C_β- and C_γ-carbons in an *anti-staggered* conformation. This rearrangement makes the protonation of ENO a highly exergonic step (ΔG_{prot} = -23.8 kcal·mol^{-1}), with the C_β-carbon of αKG 4.43 Å away from the Tyr139-hydroxyl, and no species nearby that can react with the C_β-hydrogens. Hence, the network of interactions of αKG in the active site is reduced, and compromises the transformation of αKG back to ENO ($\Delta E_{\alpha KG \rightarrow ENO}$ ~ 34.6 kcal·mol^{-1}). Further structural insight can be consulted in the PDB files of the fully characterized states for ENO protonation provided in the SI.

We have checked also if a solvent water molecule could play the role of Asp275^{A} during the deprotonation of Tyr139^{A} by ENO. To do so, we have modelled a water molecule inside the active site in the appropriate position to perform the reaction. The electronic activation energy of the transfer of a proton to ENO via Tyr139^{A} is only about 3 kcal·mol^{-1} higher than the one in the mechanism with the acidic Asp275^{A}; however the reaction is almost thermoneutral. The negatively charged Tyr139^{A} is hydrogen bonded to Arg132^{A} and to the solvent water molecule that we have placed nearby, and it does not deprotonate the water spontaneously. Hence, despite being a step likely to occur, since the activation energy for both reactions is similar, we propose that the deprotonation of Tyr139^{A} should be prevalently aided by Asp275^{A}, which is already in the active site. Moreover, there will be a Gibbs free-energy barrier to drive the water into the active site (otherwise it would be there from the beginning), which would make the reaction with water even more unfavourable.

5.3.6. Gibbs free-energy profile for the whole cycle

In Figure 5.6 we show the Gibbs free-energy profile for the catalysis by the Lys212^{B}-Asp275^{A}-Tyr139^{A} triad. Our Gibbs free-energy profile supports the mechanism proposed by Aktas and Cook,[458] the determinations of Bolduc *et al* for the rate of the dehydrogenation and decarboxylation steps of ICDH,[457] and the k_{cat} determined experimentally.[441,443]

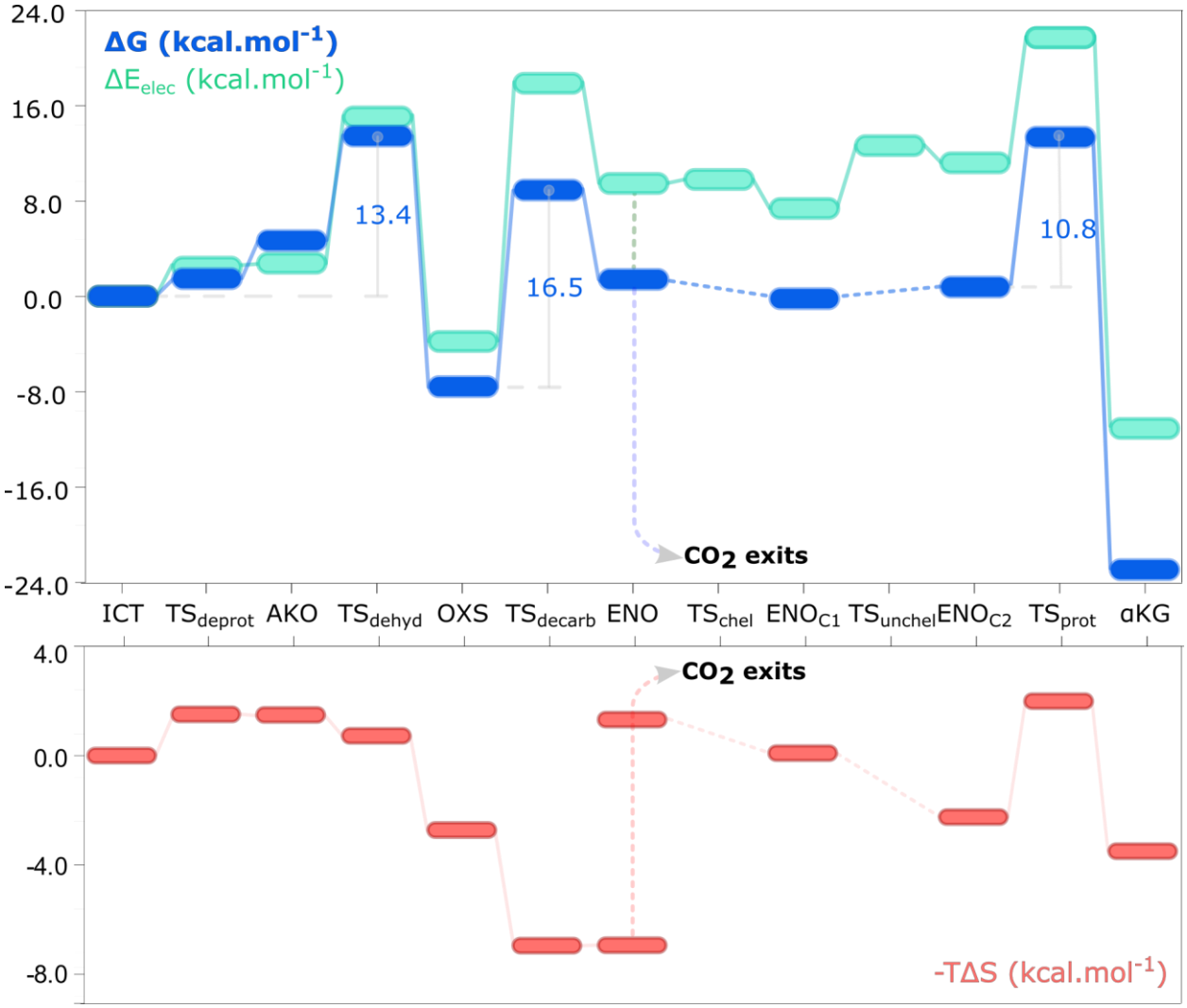


Figure 5.6. Thermodynamic profile for the Lys212^{B}-assisted catalysis of hICDH. The relative Gibbs free-energies (blue) and electronic energies (green) are read in the upper graphic, and the relative entropic contributions at 310 K and 1 bar (red) are read in the lower graphic. In the graphic depicting the $-T\Delta S$ contribution, all entropy-driven contributions are presented relative to the ICT state. The TS_{chel} and TS_{unchel} states do not present Gibbs free-energy corrections.

A close inspection of the Gibbs free-energy profile shows that the highest Gibbs-energy barrier for the reaction is 16.5 kcal·mol^{-1}, corresponding to the β-decarboxylation of OXS; however since the activation Gibbs free-energy for the β-decarboxylation differs in *circa* 3 kcal·mol^{-1} from that of the $NADP^+$ reduction (13.4 kcal·mol^{-1}), the rate-limiting step cannot be assertively defined. Furthermore, the β-decarboxylation of OXS has a significant entropic contribution, which lowers both the barrier and the reaction Gibbs free-energy by 7.0 kcal·mol^{-1}, and it is the most entropy-driven transformation. In opposition, $NADP^+$ reduction by ICT exhibits minor entropy contributions, being mostly lower than 2 kcal·mol^{-1}. The formation of AKO is not a stationary point in the Gibbs free-energy profile of hICDH. An analysis of the TΔS profile shows that the term is similar to both TS_{deprot} and AKO, which means that these states are similar (as can be inferred from the comparison of ICT and AKO in Figure 5.3). Hence, the value in Figure 5.6 is a result of the ZPE correction to the ground state electronic energy.

We emphasize that at this point we have not considered the contribution of the CO_2 exit for the proceeding reaction. Thus, we have assumed that after the β-decarboxylation the CO_2 leaves the active site with no significant Gibbs free-energy barrier, and that its exit from the active site should be very quick. Additionally, the entropy gain with the CO_2 exit (-6 to -10 kcal·mol^{-1} for CO_2 concentrations in the cytoplasm between μM–mM)[87,483,484] should make the reaction thermodynamically favourable As a result we have proceeded with the ICDH:ENO complex, from the ICDH:ENO:CO_2 complex. This does not mean that its energy contribution is not significant; however it does not affect the overall kinetics of the chemical cycle.

Figure 5.6 also shows that minor energies are involved in the rearrangement of the active site after CO_2 leaves the active site. In fact, the Gibbs-energies between the ENO and ENO_{C2} states are mostly lower than 3 kcal·mol^{-1}. Furthermore, similar *TΔS* profiles are observed for these transformations and the β-decarboxylation step, suggesting that these should be mostly driven by changes in the metal-ligand bonds and in the network of hydrogen bonds. The catalysis is concluded with the protonation of ENO by $Asp275^A$, via $Tyr139^A$, with a Gibbs activation energy of 10.8 kcal·mol^{-1}. The Gibbs-energy of the overall cycle is -24.4 kcal·mol^{-1} (taking the MgICDH:ENO complex as the reference state). We highlight that, in Figure 5.6, we have not characterized the barriers for the ENO → ENO_{C1} and ENO_{C1} → ENO_{C2} transformations. We present, nevertheless, the electronic energy barrier that we have obtained from the PESs that describe these transformations. Since these processes are favoured by entropy (see *–TΔS*

contribution in Figure 5.6), the corresponding Gibbs free-energy barriers should not influence at any level the transformation of ENO → αKG.

Furthermore, the analysis of the $-T\Delta S$ profile, together with the geometric detail we have provided in previous sections, sheds a light on the role of the metal in the catalysis: (1) the interaction of Mg^{2+} with the C_α-alkoxide/carbonyl of the substrate, and (2) the bond between Mg^{2+} and the $Asp275^A$-carboxylate. Throughout the entire catalysis the C_α-alkoxide/carbonyl group functions as an electron acceptor, a phenomenon that seems to be a requirement to lower the activation barrier of the $NADP^+$ reduction and the β-decarboxylation of OXS. The divalent metal plays a key role in stabilizing the reactive C_α-alkoxide that is formed as a result, by shielding it from the Asp moiety that is nearby. Moreover, the $T\Delta S$ parcel that lowers the Gibbs free-energy throughout the β-decarboxylation and preceding the protonation of the ENO definitely seems to be related to the stretching of the metal-ligand binding. This is observed in the formation of the OXS intermediate, where $T\Delta S$ lowers the Gibbs-energy of OXS in more than 3 kcal·mol^{-1} relatively to the TS_{dehyd}; and more pronouncedly in the Mg^{2+}–$Asp275^A$ interaction, where there is a large increase in the entropy of the active site ($T\Delta S \geq 5$ kcal·mol^{-1}), thus leading to a feasible direct OXS decarboxylation, and providing for a lower barrier for the ENO protonation. These observations enforce that the metal plays an active role in lowering the entropy of the active site for catalysis.

On a final note, we discuss the approach we took to estimate the ΔS contribution for the catalytic cycle of ICDH. We estimated the $-T\Delta S$ contributions from the change in the vibrational entropy of the system. The vibrational entropy was calculated for all the atom nuclei of the model from the partition function of a set of harmonic oscillators, as is commonly employed in statistical thermodynamics calculations. We are aware that other contributions should be accounted in the ΔS of the reaction, namely: translational, rotational, electronic and solvation terms. However, vibrational modes should be the dominant term in the Gibbs free-energy profile of a single-conformation mechanism, when there are no electron excitations involved. Solvation contributions require sampling for accurate calculation and should be scarce in such a buried active site. The remaining contributions should change negligibly throughout the course of the reaction.

5.4. Conclusions

We have determined computationally the most plausible catalytic mechanism of hICDH, through a combination of MD simulations and ONIOM calculations. The catalytic cycle proceeds in three steps, as pointed out in recent literature. The results we provide are consistent with the experimentally-derived k_{cat}, suggesting that the rate-limiting step of the reaction is *c.a.* 16 kcal·mol^{-1}. Furthermore, we are now able to clarify the questions that lack assertive answers in the scientific community, and that we have outlined in the Introduction section.

(1) During NADP+ reduction (first step), the dehydrogenation of isocitrate (ICT) is facilitated by Lys212^{B}, which deprotonates ICT to an alkoxide intermediate (AKO). We have tested also the mechanism in which an Asp279^{A} is hypothesized to generate the AKO intermediate via an Mg^{2+}-coordinated water; however the activation Gibbs free-energy of *c.a.* 21.4 kcal·mol^{-1} for the rate-limiting step is significantly larger than the expected 16 kcal·mol^{-1} barrier.

(2) The deprotonation of the ICT-hydroxyl by Lys212^{B} is concerted (but earlier and asynchronous) with the dehydrogenation of the C_{α}-hydrogen of ICT by $NADP^{+}$. AKO was shown not to be a stationary point of the Gibbs free-energy profile; our results show that there is a direct charge transfer from Lys212^{B} to ICT, with no significant nuclear motions aside from the proton transfer to Lys212^{B}. At the end of the first step, ICT and $NADP^{+}$ form OXS and NADPH. The Gibbs free-energy barrier is 13.4 kcal·mol^{-1}, and OXS becomes -4.7 kcal·mol^{-1} more stable than the reactant.

(3) Our calculations refer to the β-decarboxylation step as the most likely rate-limiting step of the reaction, with an activation Gibbs free-energy of 16.5 kcal·mol^{-1}. However, the Gibbs free-energy for $NADP^{+}$ reduction is nearly close to this upper barrier (13.4 kcal·mol-1); hence, in certain conformations, it might also contribute to the rate-limitation. During oxidative β-decarboxylation of OXS the β-carboxylate leaves the C_{β}-carbon as CO_2 and the bond of Asp275^{A} to Mg^{2+} stretches to secure both Arg132^{A} and Tyr139^{A}, previously anchoring the β-carboxylate of OXS. This conformational change should be fundamental to facilitate the dissociation and release of CO_2 from the enzyme's active site. The resulting ENO is higher in Gibbs free-energy than OXS ($\Delta G_{OXS \rightarrow ENO}$=9.0 kcal·mol^{-1}). The oxidative β-decarboxylation is the most entropy-favoured reaction step (lowering the activation and reaction energy by 7.0 kcal·mol^{-1}).

(4) Lastly, the protonation of ENO is preceded by conformational changes in the coordination of Asp275^{A} to Mg^{2+} and reorientation of the Asp275-hydroxyl to the acidic Tyr139^{A}, which involve

small Gibbs free-energy changes, and are accomplished by a concerted step in which the deprotonation of $Tyr139^{A}$ by the C_{β}-carbon of ENO is accompanied by the deprotonation of the neighbouring $Asp275^{A}$ by $Tyr139^{A}$. The latter observation leads us to propose that it is in fact $Asp275^{A}$, and not $Tyr139^{A}$, that is responsible for the protonation of ENO. The reaction occurs with an activation Gibbs free-energy of 10.8 kcal·mol^{-1}, and the resulting αKG product exhibits a very high reverse activation Gibbs free-energy (34.6 kcal·mol^{-1}).

The role of $Asp275^{A}$ throughout catalysis was found to be remarkable, significantly complementing the mechanistic knowledge that has been drawn so far: (1) it acts as a base to deprotonate $Lys212^{B}$ and facilitates the deprotonation of the ICT's hydroxyl by $Lys212^{B}$, lowering the activation barrier of $NADP^{+}$ reduction, (2) it establishes hydrogen bonds with $Arg132^{A}$ and $Tyr139^{A}$ facilitating the exit of the CO_2 from the β-decarboxylation of OXS, and (3) it acts as an acid to protonate ENO, via $Tyr139^{A}$, producing αKG. We have also observed the crucial relevance of the divalent metal cation in the structure and electrostatic environment of the enzyme. Mg^{2+} enhances the reactivity of the enzyme:substrate complex by increasing the electron acceptor character of the C_{α}-oxygen, lowering the activation Gibbs free-energies for both the $NADP^{+}$ reduction and OXS β-decarboxylation. Furthermore, it is a key piece to lower the entropy of the enzyme's active site. In particular for ICDH, the lability of the coordination sphere of Mg^{2+} is responsible for the stabilization of several stages throughout the reaction in up to 7 kcal·mol^{-1}.

This work characterizes the catalytic cycle of hICDH with atomistic detail. Such insight has not been obtained so far by any computational or experimental means. Hence, the geometric and thermodynamic insight we present here is important to complement and validate the wide knowledge already available for ICDH and the remaining metal-dependent $NAD(P)^{+}$-linked β-decarboxylases. From a clinical point of view, we provide detailed atomistic data on all stationary points of hICDH catalysis, which should be useful in the future design of chemical biomarkers, with similar affinity for the natural and R132H mutated forms of hICDH as ICT. The drug design can also provide inhibitors (transition state analogs) to regulate the production of the 2-hydroxyglutarate from hICDH mutants, and the catalytic cycle of α-ketoglutarate dependent enzymes. Moreover, since there is a significant similarity between the active site and protein structure of the hICDH and some of its mitochondrial (ICDH2) and periplasmic homodimeric forms, we expect a broad transferability of these results either to other organelles or species. Therefore, the catalytic potentialities of hICDH in other organisms can be of use in the development of biocatalysts or biosensors in the fields of health and environment.

5.4.1. Supporting Information

The Supporting Information for the manuscript can be consulted at http://pubs.acs.org/doi/suppl/10.1021/acscatal.5b01928. It contains:

- Figure with the RMSd of the active-site from the human and *E. coli* ICDH.
- Computational protocol and additional results for the Molecular Dynamics simulations of the $Lys212^{B}$ and $Asp275^{A}$ protonation states.
- Figure of the ONIOM-optimized starting model of hICDH, with the electronic embedding scheme (for the $Lys212^{B}$- and $Asp279^{A}$-based mechanism).
- Table depicting the electronic energies for the hICDH catalysis for B3LYP, B3LYP-D3 and M06-2X.
- Comparison of the main reaction intermediates with experimental X-ray crystallography results.
- Video presenting the mechanistic pathway for hICDH.
- PBD coordinates for all the fully characterized stages throughout the hICDH catalysis (for both the Asp279/Tyr139 and Lys212/Tyr139 mechanisms).
- PDB coordinates for all the stationary point of the reaction.

Chapter 6: On the reduction of Glutathione disulphide by Protein Disulphide Isomerase: mechanistic insights provided by QM/MM methods

The setup of benchmarking studies on the performance of DFT against post-HF methodologies is fundamental to tackle specific chemical reactions in the framework of enzyme catalysis. In Chapter 4:, we have performed such a study for the thiol-disulphide exchange reaction, a key reaction in the oxidative folding of proteins that present disulphide bonds. However, to perform this study we had to resort to a small 15-atom model system. Such dimensionality is not adequate to infer on the catalytic power of enzymes, and the results that we can draw from benchmarking studies must be, ultimately, transferred for systems in which the number of atoms to study with QM calculations may reach up to a few hundred atoms.

In this work, we present a pioneer exploration of the catalytic mechanism of the reduction of glutathione disulphide (GSSG) by the reduced *a*-domain of a human form of protein disulphide isomerase (hPDI), with an atomistic resolution. To do that, we recur to molecular dynamics (cMD) and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. The reaction proceeds in two stages: (1) a first thiol-disulphide exchange between a Cys53-thiolate and a GSSG-disulphide that releases one glutathione molecule (GSH), and (2) after deprotonation of a Cys56-thiol by a Glu47-carboxylate, via a water molecule, a second thiol-disulphide exchange between the Cys56-thiolate and the mixed-disulphide formed in the first step, to release the second GSH molecule. The thiol-disulphide exchange between the Cys53-thiolate and the GSSG-disulphide exhibited a Gibbs activation free-energy of 16.3 kcal·mol^{-1}, while the second reaction, between the Cys56-thiolate and the mixed-disulphide intermediate presented a Gibbs activation free-energy of 7.4 kcal·mol^{-1}. These values agree with the overall pseudo-first-order rates predicted in literature, which indicate a Gibbs activation free-energy of 17.6 kcal·mol^{-1}. However, the second reaction has been proposed to be rate-limiting step of the cycle. We also provide structural and thermodynamic discussion on the main stationary points of the reaction. In

particular, we refer to the role of the hydrogen bonds from the His55-backbone and Cys56 in the nucleophilic attack of the Cys53-thiolate, and the role of the solvent in stabilizing the intermediate stages of the catalysis of GSSG by the *a*-domain of hPDI. Furthermore, we discuss the increase in entropy in the active site of the *a*-domain upon formation of the mixed-disulphide intermediate, and the way in which it contributes to stabilize this intermediate. Finally, we discuss the nature of the rate-limiting step of the catalysis of GSSG by the *a*-domain.

6.1. Introduction

6.1.1. Motivation

Protein disulphide isomerase (PDI) is a multifunctional enzyme able to catalyse disulphide bond formation, cleavage and isomerisation, in the endoplasmic reticulum (ER) of eukaryotic cells. Despite that PDI is not the most effective disulphide redox catalyst, its importance in the secretory protein pathway, where the formation of native disulphides is a rate-limiting step, has been widely recognized.[307,485,486] In fact, in recent years, PDI deletion has been related with diseases involving unfolded protein response, such as Alzheimer's, Parkison's or type-II diabetes.[487-489] This unfolded response is mostly a result of the unbalance in the prevalent redox buffers in the ER – the glutathione/glutathione disulphide (GSH/GSSG), and the hydrogen peroxide/molecular oxygen (H_2O_2/O_2) pairs –[490,491] which are of critical importance to catalyse the oxidative folding of the proteins in the organelle. As a result from this misfolding process, these proteins form aggregates from the packing of hydrophobic exposed regions, resulting in increasing stress in the ER.[492,493]

Understanding the way in which PDI interacts with the ER redox buffer should be of critical importance to: (1) further investigate the role of PDI in the redox balance of the ER; (2) provide for kinetic and thermodynamic data on the catalysis by PDI, complementing the rather scarce kinetic studies that can be found in the literature;[486,494,495] (3) draw mechanistic insight on the catalysis by PDI, since there are no known studies with atomistic detail for this enzyme; (4) provide configurations that may allow for the development of inhibitors to tackle pathophysiological conditions related to the oxidoreductase activity of PDI.

6.1.2. Structure and function of PDI

Human PDI (hPDI) is a *U*-shaped enzyme with *circa* 508 residues. Among the PDI-like family of proteins, PDI (or PDI-1) is the most abundant in cells, constituting about 0.8% of the cell's protein machinery.[496] Its tertiary structure is composed of four thioredoxin-like domains (*a*, *b*, *b'*, and *a'*), and a fifth tail-shaped *c*-domain.[497,498] In Figure 6.1, a depiction of the structure and function of hPDI can be found.

Similarly to thioredoxin, the *a*- and *a'*-domains show a catalytic Cys–X–X–Cys motif (which for hPDI is Cys53–Gly54–His55–Cys56 and Cys397–Gly398–His399–Cys400), near the N-terminus of the α_1-helix of the thioredoxin-domain,[315] which is thought to be related to the high redox potential of PDI (–180mV). However, these motifs are responsible for most, but not for all, of the performance of hPDI as a thiol-disulphide catalyst. Although both of them can function independently, mutations on the N-terminal catalytic cysteines led to loss of k_{cat} and mutations at the C-terminal caused an increase in K_m, suggesting that only at saturating concentrations of substrate do they contribute equally to catalysis.[499] Furthermore, a pioneer study by Darby *et al* has shown that full activity of PDI is enhanced when all domains of PDI contribute synergistically to its function.[500]

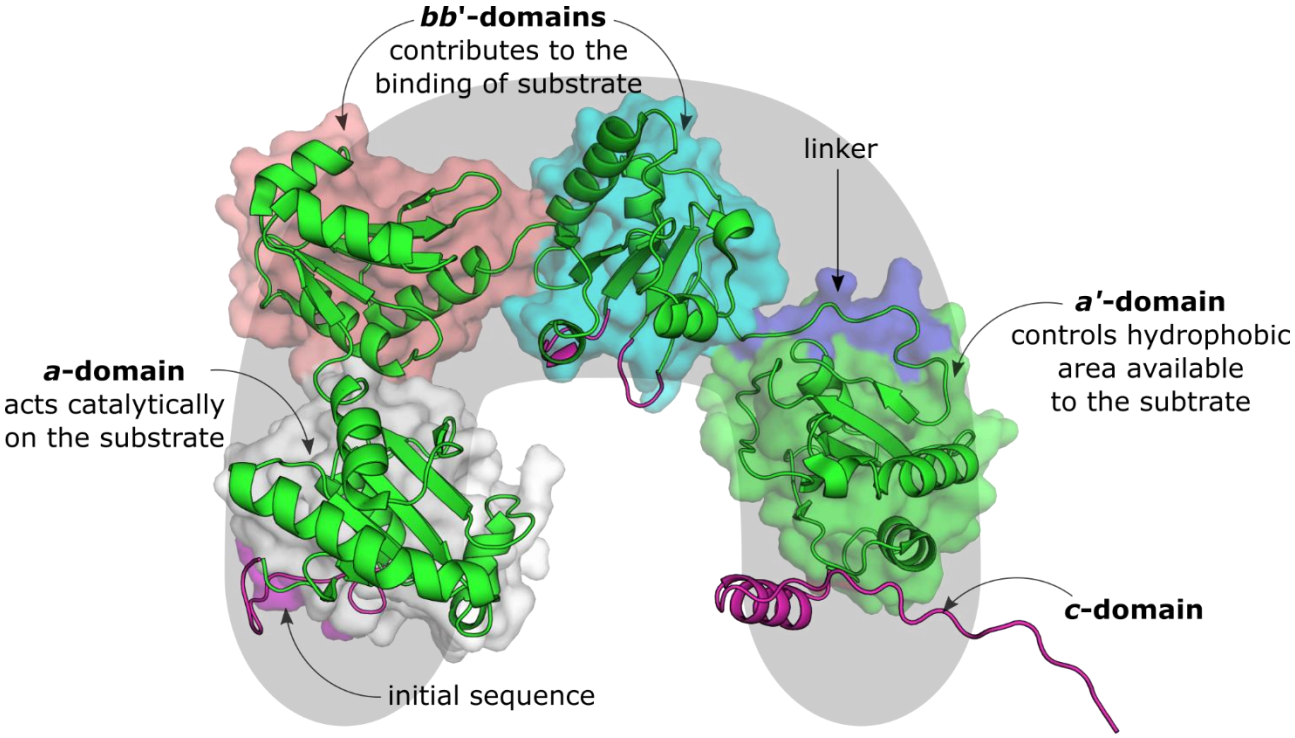


Figure 6.1. Depiction of the structure and function of human PDI (hPDI). The domains of hPDI are represented by the different colours of the surface representation. Cartoon representations in magenta represent regions of the enzyme that have not been obtained from X-ray crystallography, and are thus estimated from modelling.

Despite the structural similarity to thioredoxin, the *b*- and *b'*-domains do not exhibit the usual Cys–X–X–Cys thioredoxin-motif, and share a reduced sequence identity of *circa* 16.5%.[501] Instead, the *bb'*-region provides a large hydrophobic pocket that has been shown to bind reversibly a wide variety of peptides and misfolded proteins.[502-506] In particular, the *b'*-domain has been widely

referred to significantly improve the activity of PDI towards misfolded protein substrates.[498,500,507] Attached to the *b'*-domain, there is a 19-amino acid segment (linker *x*) that connects the *b'*- and *a'*-domains. This segment is involved in the response of these domains when the oxidation/reduction of the Cys397/Cys400 pair occurs, by expanding or shortening the hydrophobic cleft to accommodate the substrates of PDI.[505] In Table 6.1, we summarize the main characteristics and function of the different domains of hPDI.

Table 6.1. Structural and function data for hPDI

Domain	Sequence	Length	Main function
initial sequence	Met1–Glu22	22 residues	*signalling sequence cleaved when PDI enters the ER*
domain *a*	Glu23–Gly134	112 residues	*catalytic Cys53–Gly54–His55–Cys56 motif able to catalyse thiol-disulphide exchange*
interregion *a*/*b*	Pro135	1 residue	–
domain *b*	Ala136–Val237	102 residues	*contributes to substrate binding*
domain *b'*	Ile238–Gly349	112 residues	*largest hydrophobic cleft to bind misfolded proteins*
linker *x*	Lys350–Trp364	15 residues	*responds to changes in the Cys397/Cys400 pair by controlling the hydrophobic area available to the substrate*
domain *a'*	Asp365–Asp476	112 residues	*catalytic Cys397–Gly398–His399–Cys400 motif able to catalyse thiol-disulphide exchange*
domain *c*	Gly477–Leu508	32 residues	*acidic tail, which may be involved in substrate binding*

6.1.3. Glutathione/glutathione disulphide buffer in catalysis by hPDI

The main catalytic cycle of PDI can be summed up in three main stages: (1) oxidation of the Cys397/Cys400 pair to form the Cys397–Cys400 disulphide; (2) formation of the mixed-disulphide PDI:substrate intermediate, by the nucleophilic attack of the Cys53-thiolate on a non-native

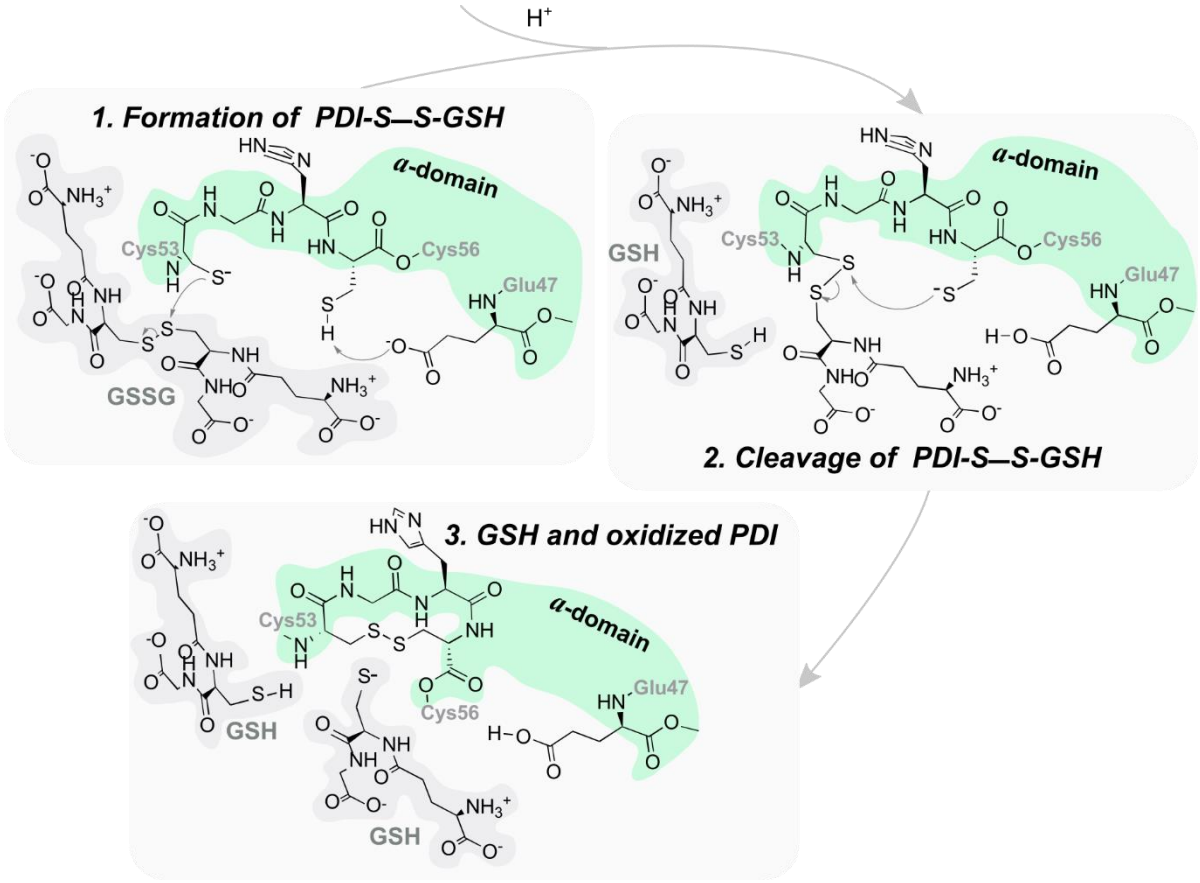
disulphide of the misfolded protein; and (3) cleavage of the mixed-disulphide intermediate by a second nucleophilic attack from the buried Cys56-thiol or a neighbouring thiolate species.[495,507,508]

The oxidation of the Cys397/Cys400 pair in the *a'*-domain has been recently observed to induce domain motions in the *b'–x–a'*-region, and increase the available hydrophobic surface in the cavity of hPDI; consequently, this feature is described to enhance the binding of misfolded proteins to hPDI and increase its isomerase and chaperone activities.[505] As a result, since hPDI can be found in both the oxidized and reduced forms,[509] the oxidation of the Cys397/Cys400 pair should be a critical stage in the PDI redox cycle. Initially, GSSG was thought to play a role in this process; however, currently it is believed that this oxidation occurs through endoplasmic reticulum oxidoreductins (ERO),[510-512] since these enzymes have shown significant interaction with the *b'*- and *a'*-domains of hPDI.[513]

After the substrate binds to hPDI, the thiol-disulphide exchange reaction will proceed in the *a*-domain of the enzyme. Here, the Cys53 and Cys56 exhibit very different properties. The Cys53 is the solvent-exposed residue in the Cys53–Gly54–His55–Cys56 motif and presents a lower pK_a (~4.5) than would be expected for equivalent cysteines in thioredoxin-like folds (~7.1),[514-516] while the Cys56 is more buried within the *a*-domain and presents a much higher pK_a (~12.8).[517] As a result, the Cys53 is a very nucleophilic species, and thus favours the isomerisation reaction by hPDI. The reaction occurs through an S_N2 nucleophilic substitution in which the Cys53-thiolate attacks the disulphide bond of the substrate with concomitant dissociation of the latter, and it has been explored in several theoretical studies of cluster models in the past years.[188,309,316] The turnover of the reaction can vary with the substrate that binds to hPDI; however, for the particular case of the thiol-disulphide exchange between the Cys53-thiolate and the disulphide of GSSG, the turnover rate was determined to be 1.4 s^{-1} (which is about 18 kcal·mol^{-1}, from transition state theory considerations).[495]

At the end of the second step of the catalysis, there are two pathways that will compete to complete the catalytic cycle: either the mixed-disulphide becomes trapped and the buried Cys56 is deprotonated to perform a nucleophilic attack to the mixed disulphide, or another thiolate species performs a nucleophilic attack to the mixed-disulphide intermediate completing the catalytic cycle and restoring the reduced state of hPDI. The former step is fundamental to prevent the formation of kinetically trapped intermediates.[490,518,519] For it to occur, the pK_a of the Cys56 has to lower in such a way that the terminal thiol becomes a thiolate and performs the second thiol-disulphide exchange, releasing the kinetically trapped intermediate. Up to now, Arg120 has

been indicated to contribute to this lowering of the pK_a.[517] Furthermore, X-ray crystallography has shown that a buried Lys81/Glu47 pair is located nearby the buried Cys56, and that it could responsible for the deprotonation of the latter.[497,514,520-522] Since mixed-disulphide intermediates were observed to be long-lived species,[495] the cleavage of the disulphide from the mixed-disulphide intermediate has been referred to be the rate-limiting step of the reaction (for the PDI–S–S–GSH intermediate, a turnover of 0.23 s^{-1} was determined).[495]



Scheme 6.1. Mechanism for the reduction of glutathione disulphide (GSSG) by the reduced *a*-domain of hPDI.

The thermodynamic and kinetic contours of the isomerization process by PDI are still very scarce in literature.[508,523] It is known that the actual isomerisation rate of PDI is lower than the rate of thiol-disulphide exchange reactions, and that the relative potency of each catalytic site in PDI is dependent of the substrate and the synergistic effects in its environment.[499,500,524] Nevertheless, regarding the substrates for which PDI presents a better affinity, the k_{cat} of the isomerization by

PDI may be up to 10^3 times larger than that of the uncatalized reaction (which exhibits a turnover rate of about 1 min^{-1}).[525]

In our study, we will tackle the oxidation of the *a*-domain of PDI by GSSG, which is one of the smallest substrates of PDI (see Scheme 6.1). GSSG is not among the best substrates of PDI.[526] However, the contributions of the GSH/GSSG buffer for the catalysis by PDI have been subject of ongoing investigation in the near past;[495,509,510] in particular, regarding the role of GSSG and endoplasmic reticulum oxidoreductins (ERO) in the oxidation of the Cys^{53}/Cys^{56} and Cys^{397}/Cys^{400} pairs, in its catalytic domains. Moreover, it is expected that this buffer can also assist the formation and cleavage of disulphide bonds in the ER, and thus, it can form mixed-disulphide intermediates that can also require PDI in order to catalyse the disulphide isomerization.[527-529] On a final note, one study by Lappi *et al* details experimental kinetic data from the catalysis of GSSG by the *a*-domain of PDI,[495] thus providing important quantitative data that can be used to validate theoretical calculations in PDI enzyme models.

6.2. Computational Methods

6.2.1. Modelling the hPDI:GSSG complex

Our template molecular model was built from the X-ray crystallographic structure of a human oxidized PDI ($hPDI_{ox}$) – with PDB code: 4EL1.[498] It included the four thioredoxin-like domains and the linker sequence of the enzyme. A peptide of 18 residues from the initial sequence was lacking, as well as the *c*-domain of hPDI. To model the initial sequence we modelled the first two residues (Met1–Leu2) with the LEaP module of the AmberTools13 package,[530] and the remaining (Arg3–Asp18) were transferred from the reduced form of human PDI ($hPDI_{red}$) – with PDB code: 4EKZ.[498] Regarding the *c*-domain, the Ala476···Val504 gap was modelled from the sequence Val488–Glu504 of yeast PDI (yPDI) – with PDB code:2B5E[497] – which includes the α-helix that can be found in this domain;[497] the remaining four residues (Lys505–Leu508) were also modelled with the LeaP module. We emphasize that the segments modelled with the LEaP module exhibit no apparent secondary structure; we expect that no significant modelling error should derive from this protocol. Regarding the use of the X-ray structure from yPDI, we note that despite the low sequence identity between PDI from different organisms, the activity and folding of hPDI and yPDI have been shown to be similar.[524] Finally, we have also modelled two gaps at the *b'*-domain: the Phe249···Thr255 gap was modelled from the same domain in the reduced form of human PDI ($hPDI_{red}$) – with PDB code: 4EKZ;[498] and the Thr319···Met324 gap was modelled from the structural equivalent Gln329···Ala334 in the yPDI.[497] All residues were protonated at the standard physiological pH values, except for the Cys53, Cys397 and Cys, which states are extensively described in literature:[498,514-516] Cys53 was modelled as a thiolate, and the Cys397 and Cys400 pair were covalently bonded to hold a disulphide bridge. All MM parameters to describe the model enzyme were drawn from the FF99SB*ildn* force field;[469-471] the mechanical parameters for GSSG were determined using the Antechamber tool and the PARM99 force field,[2,294] while atomic charges were derived from a restrained electrostatic potential (RESP) approach of a population of Merz-Kollman (MK) charges[134] calculated from a single-point energy calculation at the HF/6-31G(d) level of theory.

To minimize the energy of our model at the molecular mechanics (MM) level of theory, we build a rectangular box solvated with TIP3P waters[299] within a radius of 12 Å from the surface of the enzyme. This protocol comprehended four stages: relaxation of the solvent in the rectangular box, relaxation of all the light atoms in the model, relaxation of the side chains of the amino acid residues of the modelled $hPDI_{ox}$, and free energy minimization of the full system. The system was then heated during 100 ps in an *NVT* ensemble with classical molecular dynamics (cMD) simulations: in the first 50 ps the temperature was linearly increased up to 310 K, and in the remaining 50 ps it was kept fixed. Afterwards, we have performed a 1 ns cMD simulation with the backbone of $hPDI_{ox}$ fixed with 20 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic force constants (to equilibrate de density of the model), and a 10 ns cMD simulation with the backbone of the four thioredoxin-like domains kept fixed with the same harmonic force constant. We have used the SHAKE algorithm[165] to constrain the motion in H-including bonds in order to use a 2 fs integration step, and have employed the particle-mesh-Ewald summation method (PME)[155] to account for the electrostatic interactions beyond the 10 Å *cutoff* for non-bonded interactions. We validated the robustness of our model through a classical molecular dynamics (cMD) simulation of 50 ns in the isobaric-isothermal ensemble (*NPT*), employing the Berendsen barostat[170] and the Langevin thermostat[171]. From this simulation, we have followed the root-mean-square deviation (rmsd), and the secondary structure for each domain of the enzyme model.

The next step for the modelling of our complex is to place the GSSG substrate at the active site of the *a*-domain of our model $hPDI_{ox}$. To do this, we used the X-ray of a glutaredoxin complexed with GSSG – PDB code: 4TR0.[531] Glutaredoxin is a thioredoxin-like enzyme that catalyses the reduction of GSSG to GSH; moreover it also possesses the catalytic Cys–X–X–Cys motif (Cys12–Pro13–Tyr14–Cys15 in glutaredoxin),[382] similarly to the *a*- and *a'*-domains of the PDI family of enzymes. We have transposed the GSSG substrate and all waters within 6 Å to the model $hPDI_{ox}$. To do so, we have aligned the heavy atoms of the backbone of the Cys–X–X–Cys in the glutaredoxin and the *a*-domain of $hPDI_{ox}$, registering an rmsd of 0.27 Å. In Figure 6.2, we show the result of the alignment of the Cys–X–X–Cys in both enzymes, highlighting the position of GSSG relatively to catalytic N-terminal Cys.

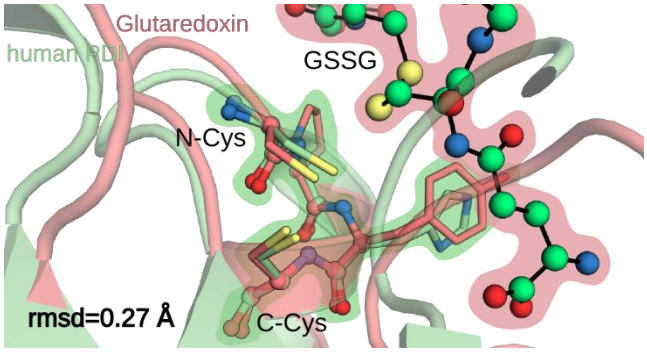


Figure 6.2. Result of the alignment of the backbone of the Cys12–Pro13–Tyr14–Cys15 from glutaredoxin with that of the Cys53–Gly54–His55–Cys56 of the *a*-domain of hPDI.

We have minimized the energy of the $hPDI_{ox}$:GSSG complex with the same procedure that we followed for the $hPDI_{ox}$ model; the same applied to the heating of the system and the density equilibration. We then performed an 10 ns cMD simulation with the heavy atoms of the $hPDI_{ox}$:GSSG complex kept fixed with harmonic force constants of 20 $kcal \cdot mol^{-1} \cdot Å^{-2}$. With this approach, we expected to keep the structure of the complex as a rigid block, allowing for solvent waters to interact with it. We took the last structure of the cMD run as the representative catalytic conformation of the $hPDI_{ox}$:GSSG complex.

6.2.2. Building the ONIOM model

To establish the catalytic mechanism of hPDI, we employed QM/MM calculations with the ONIOM methodology.[214] In this part of the work, we have only approached the complex between the *a*- and *b*-domains of $hPDI_{ox}$ and GSSG; we have also included all water within 3 Å of the domain *a* and *b* of hPDI, and all waters within 6 Å of the active site of the *a*-domain (comprising Cys53–Gly54–His55–Cys56, Glu47, Arg120, and GSSG). Overall, the model presented a charge of -12. Despite that we could have used the full enzyme model, we believe that these two domains are the most representative in the catalysis of GSSG, since the latter is a small substrate and it is very distant from the remaining domains of hPDI. Hence, the insight drawn for this study is not intended to represent, in any level, the chaperone activity of hPDI, but solely the thiol-disulphide exchange reaction that is catalysed by either the *a*- or *a'*-domains of hPDI towards GSSG.

Recent years have witnessed the use of a wealth of different computational approaches to determine enzyme reaction mechanisms, ranging from the use of fully QM cluster enzyme models[33,53,58,99-101] to QM/MM PES calculations using DFT in large QM regions,[56-58,100,102] QM/MM

molecular dynamics,[59,60] Car-Parrinello Molecular Dynamics,[61,62] and Empirical Valence Bond theory.[63,64] Nevertheless, we believe that QM/MM PES calculations offer one of the best compromises between the accurate insight that can be drawn from quantum mechanics (QM) calculations, and the description of the enzymatic environment to the catalysis by the enzyme.[57,100,102] On the other way around, we will draw thermodynamic and kinetic insight from single-conformation calculations, taking our X-ray-based model as representative of the average ensemble of catalytic poses for the hPDI.

We have modelled our system in two layers: a DFT layer comprising mostly the Cys53–Gly54–His55–Cys56, the Glu47 and Arg120 next to Cys56, and a substantial part of the GSSG substrate; and an MM layer comprising the *a*- and *b*-domains of our $hPDI_{ox}$ model. In our final model, we have also included all waters within 6 Å of the sulphur atoms in the active site, and the oxygen atoms from Glu47 in the DFT region, ending up with a DFT layer of 167 atoms. The DFT layer, presented an overall charge of -1, and a singlet spin multiplicity. In Figure 6.3, we present our starting QM/MM model, optimized at the mPW1N/6-31G(d)/FF99SB level of theory, with the electrostatic embedding scheme.

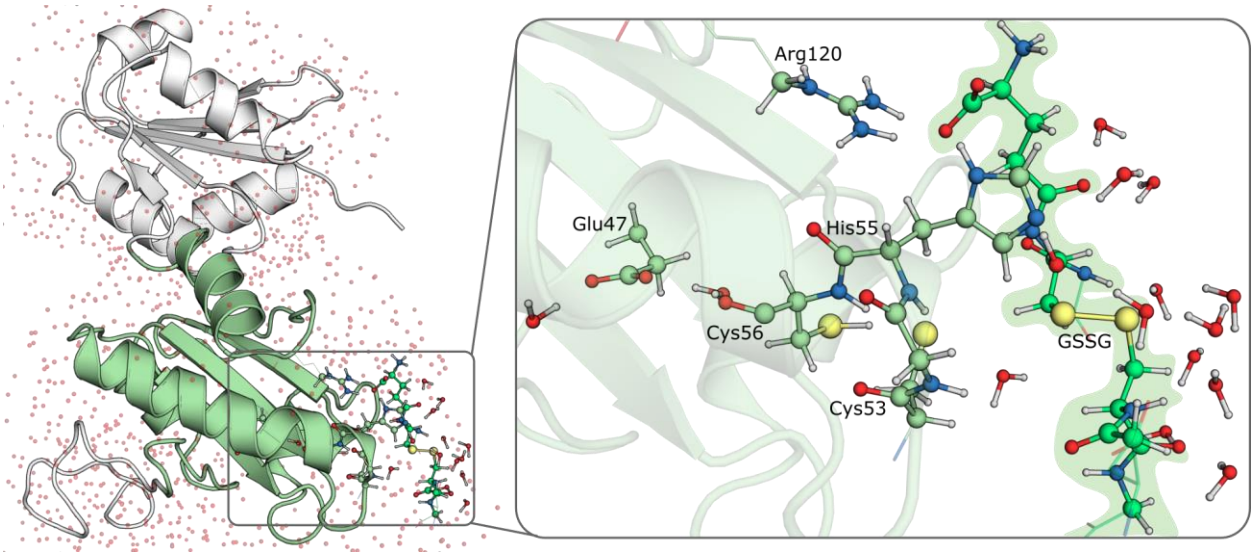


Figure 6.3. Optimized model of the domains *a*- and *b*- from hPDI complexed with GSSG. On the right, representation of the full ONIOM model; on the left, detailed representation of the DFT layer of the ONIOM model, with some of the more relevant interactions in it.

A previous study of ours has shown that mPW1N, [387-392,420] along with mPW1K,[387-392,416,423] BB1K[202,207,283] and mPWB1K, [283,387-392,422,426] is one of the best density functionals to adequately describe the thermodynamic and kinetics of thiol-disulphide exchange.[188] Despite that we could have adopted any of these, mPW1N slightly outperformed other density functionals following the hybrid-generalized-gradient-approximation (h-GGA); and thus, it was adopted to conduct QM calculations in this work. Throughout all ONIOM calculations, all waters included in the MM layer were kept fixed as a rigid block.

6.2.3. Establishing the catalytic mechanism

All calculations to study the reaction coordinate space were performed with the ONIOM methodology and electrostatic embedding scheme,[215] as implemented in the Gaussian09 package.[282] The mPW1N/6-31G(d):FF99SB level of theory was the preferred to establish the potential energy surface (PES) for the several guess coordinates throughout the mechanism. Several works in our group have enforced that the 6-31G(d)[37,38,41-43,364,365] basis is adequate to explore the reaction coordinate space in enzyme catalysis with hybrid methods.[58] In addition, the benchmarking that we have conducted has also pointed out that the combination of the mPW1N density functional with the 6-31G(d) basis set holds small errors (mostly below 0.5 kcal·mol^{-1}).[188]

For all stationary points along the reaction coordinate space, we performed relaxed geometry optimization calculations, and evaluated the character of these points through evaluation of nuclear vibrational frequencies (one imaginary frequency for transition states, and null imaginary frequencies for minimum energy states). To build the thermodynamic and kinetic profile of the reaction, we determined zero-point energy (ZPE), and Gibbs free energy corrections (G_{corr}) at 310 K and 1 bar (physiologic cell conditions), from statistical mechanics considerations (using the harmonic oscillator approximation to account for vibrational modes, the rigid rotor approximation to account for the rotation modes, and the particle-in-a-box for the translational components of the model). These calculations were also performed at the mPW1N/6-31G(d):FF99SB level of theory. Concerning the 0 K-electronic energy contributions, they were determined from single-point energy calculations at the mPW1N/6-311+G(2d,2p):FF99SB//mPW1N/6-31G(d):FF99SB level of theory. We have also performed a comparison the performance of mPW1N against BB1K. The results state that the thermodynamics of the reaction is similar (refer to Table 6.2).

Table 6.2. Gibbs free-energies for the catalytic cycle of the *a*-domain of our model of hPDI. All energies are presented in kcal·mol^{-1}; energies represented in curve brackets are given in Ha. GSSG, GSM···GSX-Cys53, Cys56···WAT, GSX-Cys53, GSX-Cys53···Cys56 (pre), GSX-Cys53···Cys56 and GSM···Cys53-Cys56 correspond to minimum energy stationary points at the mPW1N/6-31G(d):PARM99SB level of theory. TS_{redox1} and TS_{deprot} and TS_{redox2} correspond to transition state stationary points at the mPW1N/6-31G(d):PARM99SB level of theory.

	Optimization and ΔG corrections	**Single-point energy calculation**		
	mPW1N/6-31G(d):PARM99SB	*mPW1N/ 6-31G(d)*	*mPW1N/ 6-311+G(2d,2p)*	*BB1K/ 6-311+G(2d,2p)*
Formation of the mixed-disulphide	GSSG	0.0 (-5671.3928)	0.0 (-5673.1237)	0.0 (-5672.5605)
	TS_{redox1}	14.4	16.3	16.7
	GSM···GSX-Cys53	2.6	2.7	2.4
	Cys56···WAT	4.1	4.5	3.2
	TS_{deprot}	2.8	3.7	4.3
	GSX-Cys53	-10.1	-9.5	-11.6
Cleavage of the mixed-disulphide	GSX-Cys53···Cys56 (pre)	0.0 (-5671.9909)	0.0 (-5673.7124)	0.0 (-5673.1490)
	GSX-Cys53···Cys56	0.8	-1.1	-1.9
	TS_{redox2}	6.9	6.3	6.3
	GSM···Cys53-Cys56	-5.5	-7.5	-6.4

We have not determined the solvation contributions to the mechanism studied. However, since we have included a cap of 3 Å of solvent around the enzyme model and a cap of 6 Å around the active site of the *a*-domain, some of these effects should be accounted for in electronic structure calculations. Moreover, we emphasize that a considerable number of waters was included nearby the catalytic site, explicitly weighting their effect in the reaction.

6.3. Results and Discussion

6.3.1. From validating the human PDI model, to optimizing the ONIOM model

The 50 ns MD simulation that we performed in our model $hPDI_{ox}$ (represented in Figure 6.1), shows considerable changes in the model, relative to X-ray structure, with the rmsd of the backbone atoms in the enzyme ranging to up to 12 Å. However, if we analyse solely the more bulk regions of the enzyme (excluding the signal sequence and the *c*-domain of hPDI), then the rmsd of most conformations differs in less than 4 Å from the X-ray structure. Despite that these results exhibit substantial differences between the X-ray structures and the results from cMD simulations, there are three points that we address to support our modelling: (1) our starting X-ray structure provided a dimerized form of hPDI; (2) hPDI presents a *U*-shaped form which exhibits a large solvent accessible area, particularly in the oxidized state of the *a*'-domain that we have studied; and (3) there is a linker sequence of 19 amino acids that allows for rotation of the *a*'-domain, and that exhibits no secondary structure. Figure 6.4 compares the rmsd for each frame of the cMD simulation against every other frame, for the four thioredoxin-like domains (*a*-, *b*-, *b'*-,*a'*-) and the *a*- and *b*-domains, which were employed in the ONIOM study. These coloured maps show that the conformations from domains *a* and *b* exhibit significantly lower rmsd than for the complete enzyme (mostly lower than 2 Å).

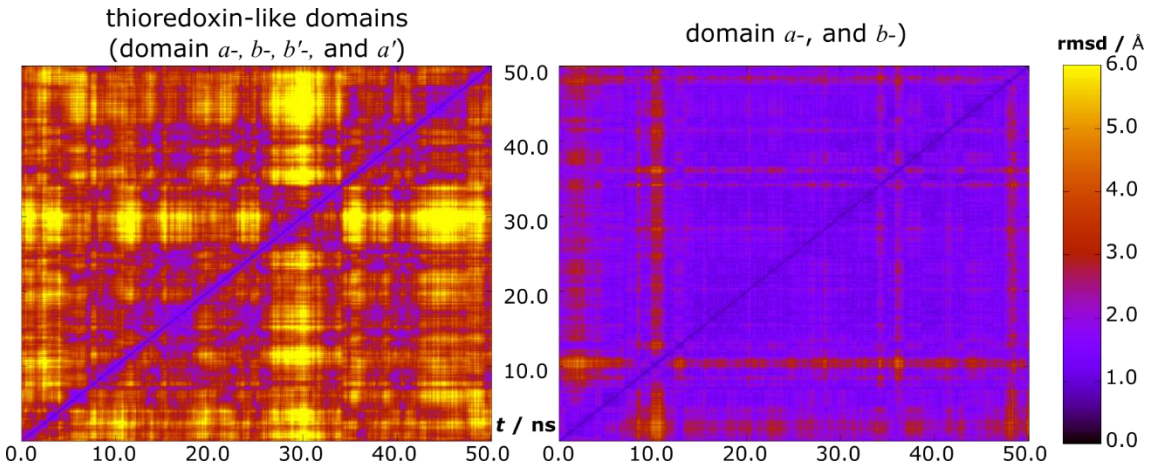


Figure 6.4. 2-dimensional rmsd for the four thioredoxin-folds (on the left), and the *a*- and *b*-domains (on the right) of hPDI throughout the 50 ns *NPT* cMD performed.

Furthermore, secondary structure analysis suggested that the secondary structure of each of the four domains is conserved across the complete cMD simulation. In Figure 6.5, we present the secondary structure of the *a* and *b* domains throughout the 50 ns cMD simulation. We emphasize that the thioredoxin-like fold (β–α–β–α–β–α–β–β–α)[382,532] is conserved for both domains during the entire simulation.

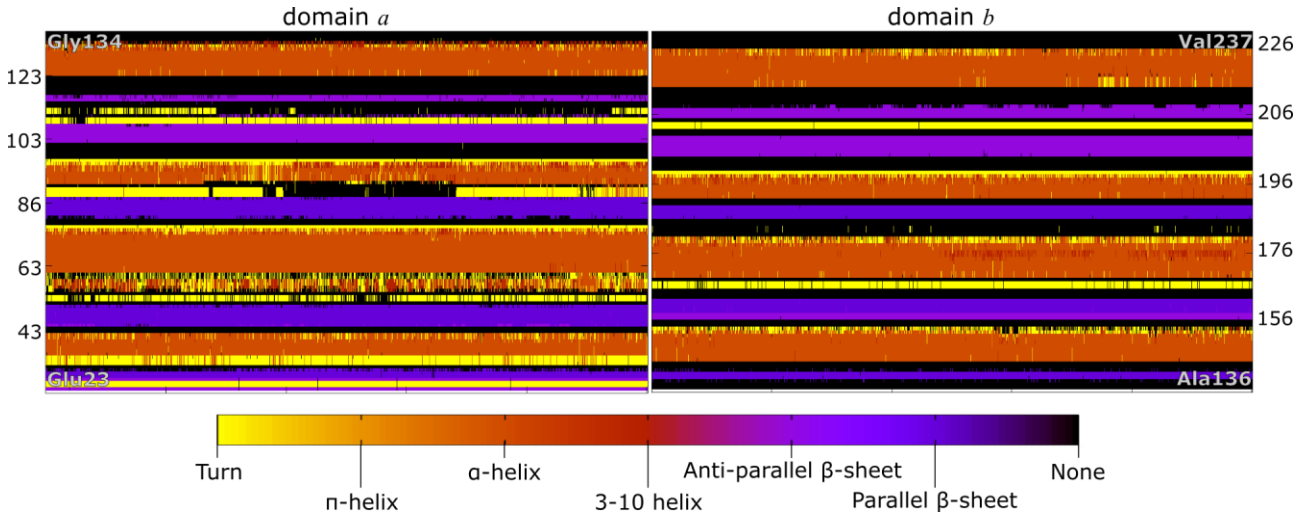


Figure 6.5. Secondary structure of the domain *a* and *b* of the hPDI, throughout the 50 ns cMD simulation in the *NPT* ensemble.

Afterwards, we have modelled the GSSG substrate in our model hPDI (as described in *Modelling the hPDI:GSSG complex*). The optimized ONIOM model maintains the main characteristics that are observed in the active site of glutaredoxin: the disulphide of GSSG is linearly aligned with the Cys53-thiolate (forming an angle of 164.5°), the NH-terminus of His55 and Cys56 is oriented to stabilize the Cys53-thiolate. In addition, the Glu47-carboxylate, which should be able to deprotonate the Cys56-thiol is stabilized by two hydrogen bonds with water molecules (these bonds are 1.64 Å and 1.80 Å long); the Arg120-guanidinium, which has been referred to lower the pK_a of Cys56 is anchored to the carboxylate in the glutamyl group of GSSG; the His55-imidazole is establishing a hydrogen bonded to GSSG-glutamyl, and accepting another with a nearby water molecule (2.09 Å and 2.02 Å, respectively). Relatively to the GSSG substrate, it is mostly solvated by water molecules.

6.3.2. Considerations on the S_N2 nucleophilic attack of Cys53 to GSSG

Relatively to the first thiol-disulphide exchange occurring between Cys53 and the GSSG substrate, we will discuss three main aspects that we have observed: (1) the environment required to stabilize the mixed-disulphide intermediate; (2) the protonation of the His55-imidazole; and (3) the most favourable kinetics for the nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide.

We will start by addressing the subject of the environment around Cys53 and GSSG. We built two test QM/MM models to perform such a study: in the first, we included in the DFT layer the conserved waters from crystallography near the active site – DFT layer of 99 atoms; in the second, we included all waters and GSSG within 6 Å of the sulphur atoms of the Cys53-thiolate and the GSSG-disulphide – DFT layer of 141 atoms. In Figure 6.6, we present a representative part of the DFT layer around Cys53 and GSSG, for both test models, and the ONIOM energies from the linear transit scan performed along the reaction coordinate.

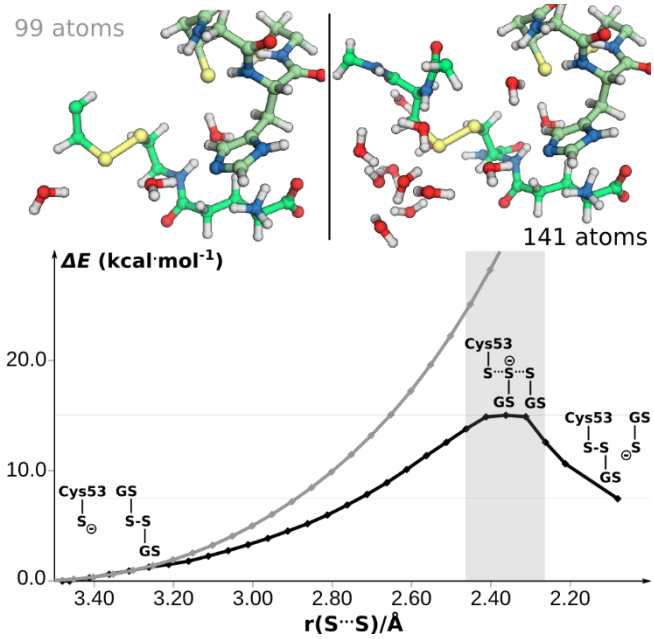


Figure 6.6. ONIOM energy of the nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide (in kcal·mol^{-1}), for two test models with a DFT layer of 99 atoms (grey) and 141 atoms (black).

Only for the model in which we include a considerable amount of solvent and substrate, we do observe the formation of the mixed-disulphide intermediate. In the smaller model, we cannot observe a transition state, and the potential energy surface (PES) for the reaction goes far beyond

the 25 kcal·mol^{-1} limit for an efficient enzymatic catalysis. These observations support that the solvent should play an important role in the reaction, stabilizing the glutathionate (GSM) that is formed upon cleavage of the GSSG-disulphide.

The second aspect of this step concerns the protonation of the His55-imidazole in the Cys53–Gly54–His55–Cys56 motif. To evaluate this aspect, we built a model complex hPDI$_{ox}$:GSSG in which the His55-imidazole was protonated holding the imidazolium cation; this model accounted for 171 atoms in the DFT layer. Despite that the most accepted hypothesis is that His55 is in its neutral form,[382] it has also been referred that the His55-imidazolium cation could stabilize the Cys53-thiolate. Our calculations show that, if the latter hypothesis should be true, the nucleophilic attack of the Cys53-thiolate would not hold any transition state (refer to the right side in Figure 6.7).

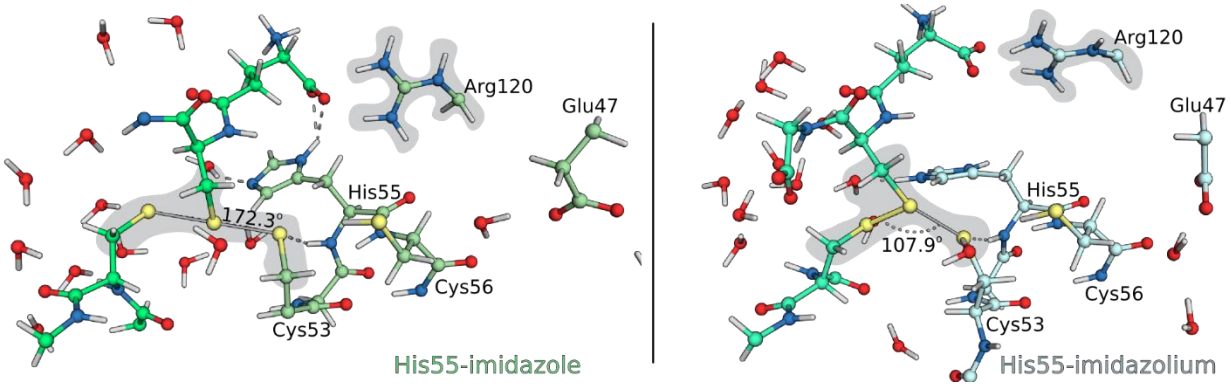


Figure 6.7. Representation of the DFT layers for the two ONIOM models in which the His55-imidazole is either neutral (on the left, in palegreen) or in the cationic form (on the right, in palecyan). The GSSG substrate is represented in limegreen (on the left) and in greencyan (on the right). The figure depicts the conformation in the DFT layer when the nucleophilic Cys53-thiolate is at *circa* 2.40 Å from the GSSG-disulphide (which is usually the distance for which the trisulphide anion transition state is formed). The grey shades indicate the regions where residues occupied substantially different poses in both models.

The PESs for the nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide resemble those presented in Figure 6.6: the nucleophilic attack in which we find the His55-imidazolium exhibits no possible transition state, and the ONIOM energy of the attack increases exponentially throughout the entire linear transit scan and there is no formation of any mixed-disulphide intermediate; and the nucleophilic attack in which we have the His55-imidazole exhibits a clear transition state (refer to the left side of Figure 6.7), forming a mixed-disulphide intermediate. An analysis of the conformations represented in Figure 6.7, indicates that the imidazolium form is only accepting hydrogen bonds from solvent molecules, and thus it should be more flexible than

the imidazole form, which is a hydrogen bond donor to the carboxylate of the glutamyl moiety of GSSG and a hydrogen bond acceptor from the solvent.

At this point, we have asserted that for the thiol-disulphide exchange reaction to occur, we require that there should be solvent in the DFT layer to stabilize the GSM that will be released upon formation of the mixed-disulphide intermediate, and that the His55-imidazole in the Cys53–Gly54–His55–Cys56 motif should be in the neutral form for the attack of the Cys53-thiolate to the GSSG-disulphide to occur linearly.

6.3.3. S_N2 nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide

The formation of the mixed-disulphide requires two different types of reactions: a redox reaction, to reduce GSSG and form the mixed-disulphide intermediate; and an acid-base reaction to deprotonate the Cys56-thiol via the Glu47-carboxylate, holding a very stable mixed-disulphide. In Figure 6.8 and Figure 6.9, we exhibit the main stages of these chemical steps, and the variation of the thermodynamic quantities (Gibbs free-energy, ONIOM electronic energy, and entropy) along them.

Our reactant state shows that the Cys53-thiolate is initially very distant from the GSSG-disulphide (4.49 Å). In this state the Cys53-thiolate is well anchored to the backbone-NH of His55 and Cys56, and to the Cys56-thiol though hydrogen bonds. These interactions should be critical to hold the reactant state together. Nearby the Cys56-thiol, there is a Glu47-carboxylate which is hydrogen bonded to the backbone-CO of Cys56 via a water molecule. In addition, the His55-imidazole is also stabilized by the solvent and the carboxylate terminus of the GSSG-glutamyl through hydrogen bonds. These bonds will hold throughout the entire catalytic cycle, and thus should not be involved in the chemical step itself; instead they should mostly secure the catalytic conformation of the active site of the domain *a*.

In Figure 6.8, we observe that during the nucleophilic attack of the Cys53-thiolate, its hydrogen bonds increase considerably, particularly for those from Cys56. In fact, at the transition state (TS_{redox1}) the main vibrational modes are those concerned to the linear trisulphide anion, and the hydrogen bonds from the backbone-NH of His55 and Cys56. A closer look at these hydrogen bonds shows that the hydrogen bond with the backbone-NH of His55 is the one that increases the less until the TS_{redox1} is observed; afterwards, the hydrogen bonds that Cys53 establishes with Cys56 shorten, contrary to that with His55, as the covalent character of the mixed-disulphide

increases. There is also an evident charge transfer occurring from the Cys53-thiolate to the GSM resulting from the cleavage of the GSSG-disulphide bond. The trisulphide anion exhibits a $S_{GSSG}\cdots S_{GSSG}\cdots S_{Cys53}$ angle of 172°, the S_{Cys53} is at a distance of 2.43 Å from the central sulphur atom, and the GSSG-sulphurs are 2.57 Å farther apart.

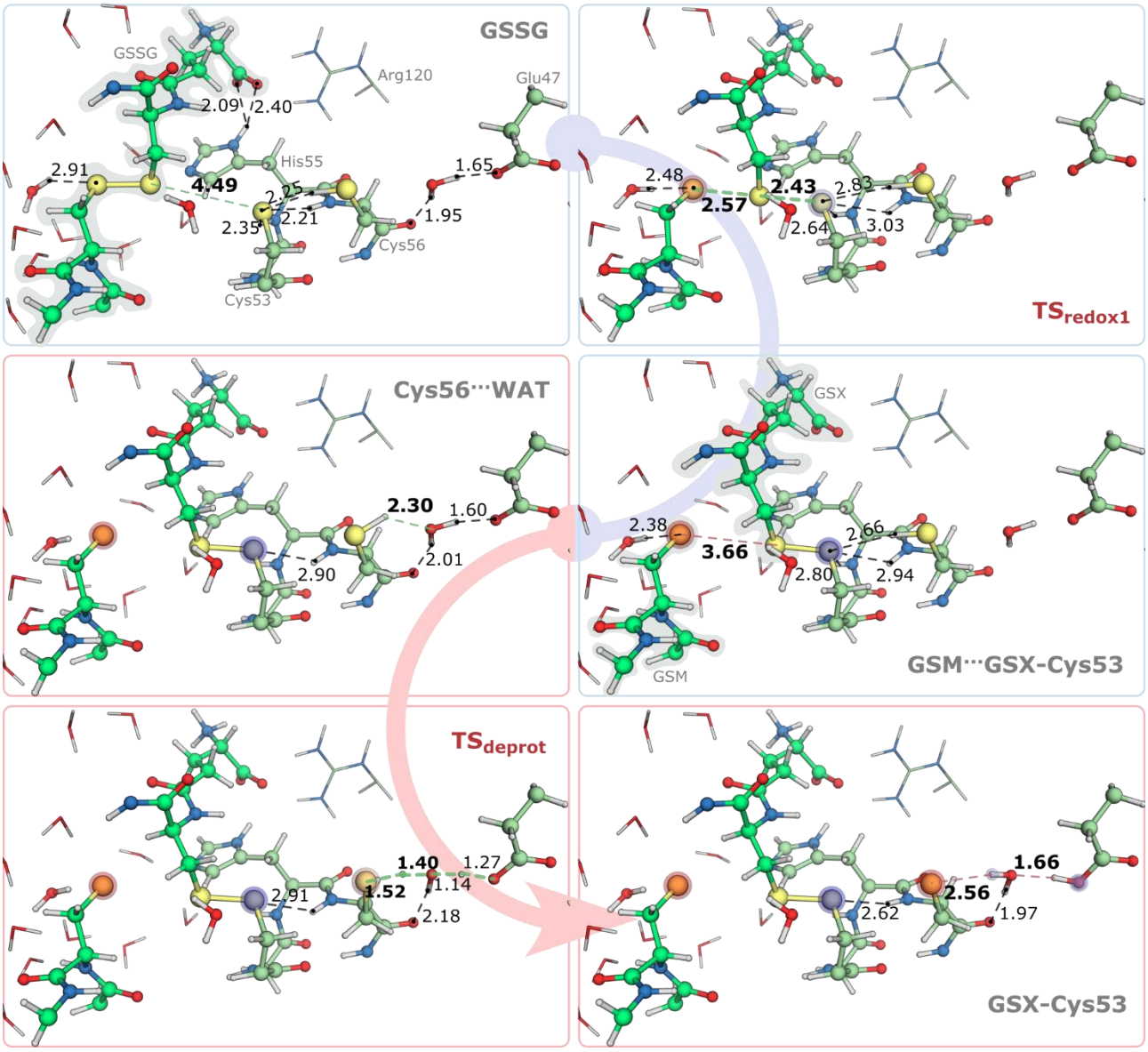


Figure 6.8. Stationary points for the formation of the stable mixed-disulphide intermediate (GSSG, TS_{redox1}, GSM···GSX-Cyx53, Cys56···WAT, TS_{deprot}, and GSX-Cys53), and main distances throughout the transformation, in Å. Distances in bold correspond to distances directly related to the reaction coordinate explored through linear transit scans. Blue-to-red shades in atoms represent the variation in atomic charge relative to the GSSG state (from 0.07 to 0.30 a.u.); blue stands for decrease in atomic charge, and red stands for increase in atomic charge.

In addition, we observe a water in the close vicinity of the exiting GSM that approximates substantially to a position in which it exhibits a shorter hydrogen bond to this species. The Gibbs activation free-energy for the process is 16.3 kcal·mol^{-1}, which corresponds to a turnover of about 21 s^{-1} from transition state theory. This turnover rate is higher than that predicted experimentally, which is of about 2.6 s^{-1} for a pseudo-first-order kinetics in a medium of 10 mM in GSSG and 50 μM of hPDI.[495] However, if we compare our Gibbs activation free-energy with the derived from the experimental turnover (17.6 kcal·mol^{-1}), they differ in about 1.3 kcal·mol^{-1}, a margin that fits well within the error usually associated to the method employed. Moreover, it is the cleavage of the mixed-disulphide intermediate has been referred as the rate-limiting step for catalysis.[495,533]

On a side note, we want to emphasize that the comparison of kinetic rates should attend to the fact that the formation of the mixed-disulphide follows a second order kinetics (it depends on the concentration of both reduced hPDI and GSSG). Hence, our first-order turnover rate is only valid in the limit where GSSG is in large excess, relatively to PDI. The concentration of hPDI in the endoplasmic reticulum (ER) is in the range 200–500 μM,[534] and the redox GSH/GSSG buffer is *circa* 6 mM in GSH and 2 mM in GSSG;[495] from the hPDI available in the ER, only about 30 μM should resemble our modelled $hPDI_{ox}$ (*a*-domain reduced and ***a'***-domain oxidized).[535] Since, there is approximately a hundred times more GSSG than $hPDI_{ox}$, a first-order kinetics for the formation of the mixed-disulphide intermediate should be valid.

The reaction proceeds to hold a mixed-disulphide species between the *a*-domain of hPDI and the GSSG-glutathione (the state GSM···GSX–Cys53). At this stage, the exiting GSM is fully reduced, and the Cys53-sulphur is now oxidized. The GSM is establishing hydrogen bonds to neighbouring water molecules, and Gibbs reaction free-energy is of 2.8 kcal·mol^{-1}. The mixed-disulphide is not yet a stable intermediate, but the approximation of the Cys56-thiol to the nearby water occurs with minimal Gibbs free-energy changes: the Cys56-thiol forms a hydrogen bond with the water close by exhibiting a Gibbs free-energy increase of 1.7 kcal·mol^{-1}, and it then protonates the Glu47-carboxylate via that same water molecule with no Gibbs activation free-energy. The final mixed-disulphide intermediate (GSX–Cys53) is very stable, since the Gibbs activation free-energy to return the intermediate to the reactant state is of about 25.7 kcal·mol^{-1}. The complete reaction step holds a Gibbs reaction free-energy is of -9.46 kcal·mol^{-1}. We believe that the reason for which this intermediate is so stable, comparatively to the one formed upon the thiol-disulphide exchange reaction, lies in the hydrogen interactions that occur between Cys56 and Glu47, via the water molecule found nearby.

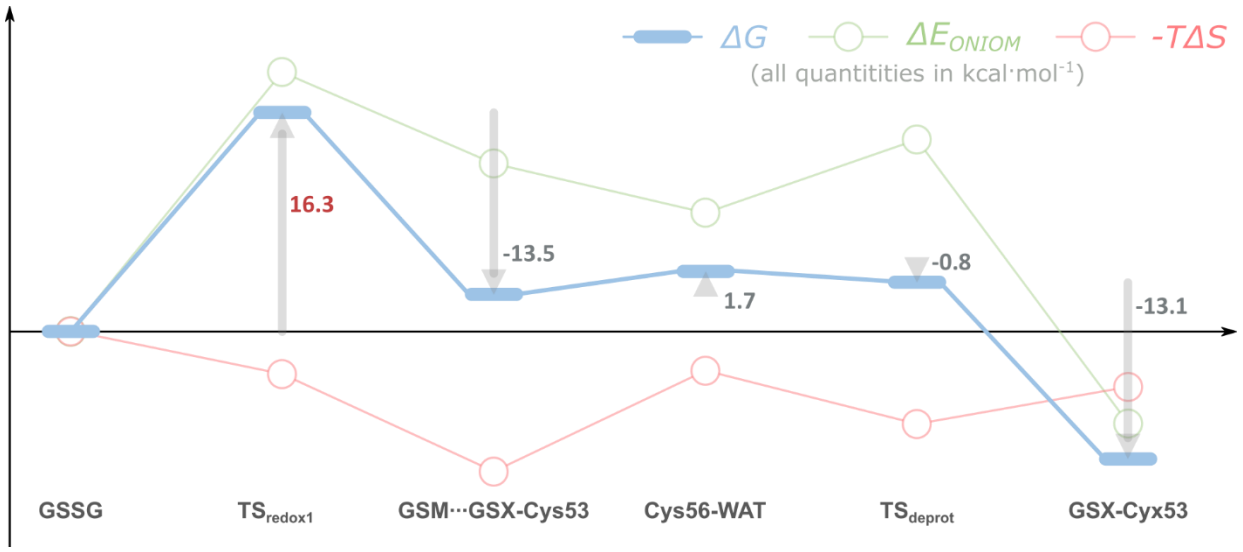


Figure 6.9. Thermodynamic profile for the formation of the mixed-disulphide intermediate between the domain *a* of hPDI and the GSSG substrate. All contributions are represented relatively to the initial hPDI:GSSG complex (GSSG).

A more detailed analysis of the thermodynamic profile of reaction emphasizes that entropy plays a significant role in the stabilization of the mixed-disulphide intermediate that is formed from the nucleophilic attack of the Cys53-thiolate to the GSSG-disulphide. In fact, a comparison of the potential energy surface (PES) and free-energy surface (FES) for the reaction shows that the energy gradient for the protonation of the Glu47-carboxylate by the Cys56-thiol is opposite. This process occurs almost with no activation costs, aside from the rotation of the Cys56-thiol group. The transition state in which there is a concomitant proton transfer from the Cys56-thiol to water's oxygen and from the water molecule to the accepting Glu47-carboxylate group (TS_{deprot}) is only observed for the 0K-PES, and does not correspond to a transition state in physiological conditions. We speculate that the large increase in entropy that leads to the stabilization of the GSM···GSX-Cys53 ($T\Delta S$ differs in about 10.5 kcal·mol^{-1} from the reactant state, GSSG) concerns the formation of the hydrophobic patch in the disulphide moiety of the mixed-disulphide intermediate. Since disulphide bonds are weak hydrogen bond acceptors, the Cys56-thiol is loose after the TS_{redox}, and can rotate freely nearby the disulphide moiety. To further support this behaviour, we highlight that when the hydrogen bond is formed with the water close by (Cys56···WAT) the $T\Delta S$ of the enzyme model lowers in about 7.5 kcal·mol^{-1}, but few other changes occur in the active site (see Figure 6.8). This change in the environment of Cys56 can also support the decreasing in the pK_a of the Cys56-thiol.

On a final note, we enforce that all the discussion that we provide rely on single conformation calculations. In such approach, only the change in vibrational entropy contributes for the thermodynamics of the reaction. These vibrational modes are derived from electronic structure calculations of the forces that act on the nuclei in the model; the remaining contributions (translational and rotational) are derived from statistical mechanics considerations, and do not depend on electronic structure calculations. Hence, the results that we draw from these calculations are limited by the scope of single conformation approaches. We consider that our model resembles that of a catalytically competent *a*-domain of hPDI, and thus, it provides significant thermodynamic and kinetic insight on the events that occur in physiological conditions.

6.3.4. Oxidation of the *a*-domain: cleavage of the mixed-disulphide intermediate

After the formation of the mixed-disulphide intermediate, there are two pathways that can occur to solve the mixed-disulphide state: either the mixed-disulphide goes through a second thiol-disulphide exchange by a nearby thiol group from a protein substrate, or it undergoes an intramolecular oxidation through the deprotonated Cys56-thiol. Here, we have primarily studied the nucleophilic attack of the Cys56-thiolate to the mixed-disulphide intermediate in the presence of the GSM product. However, the PES of the reaction coordinate exhibited no transition state. This result clearly shows that this intramolecular rearrangement will seldom occur in the presence of thiolate species that can compete with that of the interior Cys56, and it supports the fact that the intramolecular rearrangements will mainly occur to rescue trapped mixed-disulphides.

To attempt the study of the intramolecular oxidation of the *a*-domain of hPDI, we protonated the GSM product to obtain glutathione (GSH). This transformation should be legitimate, since GSM is mostly surround by water, and it should exhibit a pK_a similar to that determined in water (about 8.75)[507]. The former optimized state showed that the GSH-thiol is preferably directed to the solvent, despite that, as it occurred previously for Cys56, the conformation in which the GSH-thiol is directed towards the mixed-disulphide state is entropically favoured. However, when we performed electronic energy calculations with the larger 6-311+G(2d,2p) basis set,[37-44] the latter state was found to be slightly more favoured (a ΔG of -1.1 kcal·mol^{-1} is observed).

In Figure 6.10, we present the main stationary points in the new PES for the nucleophilic attack of the Cys56-thiolate to the mixed-disulphide. The PES exhibits a transition state (TS_{redox2}) with a low Gibbs activation free-energy (7.4 kcal·mol^{-1}), and the oxidized disulphide-state of the *a*-domain is formed with a Gibbs reaction free-energy of -6.4 kcal·mol^{-1}.

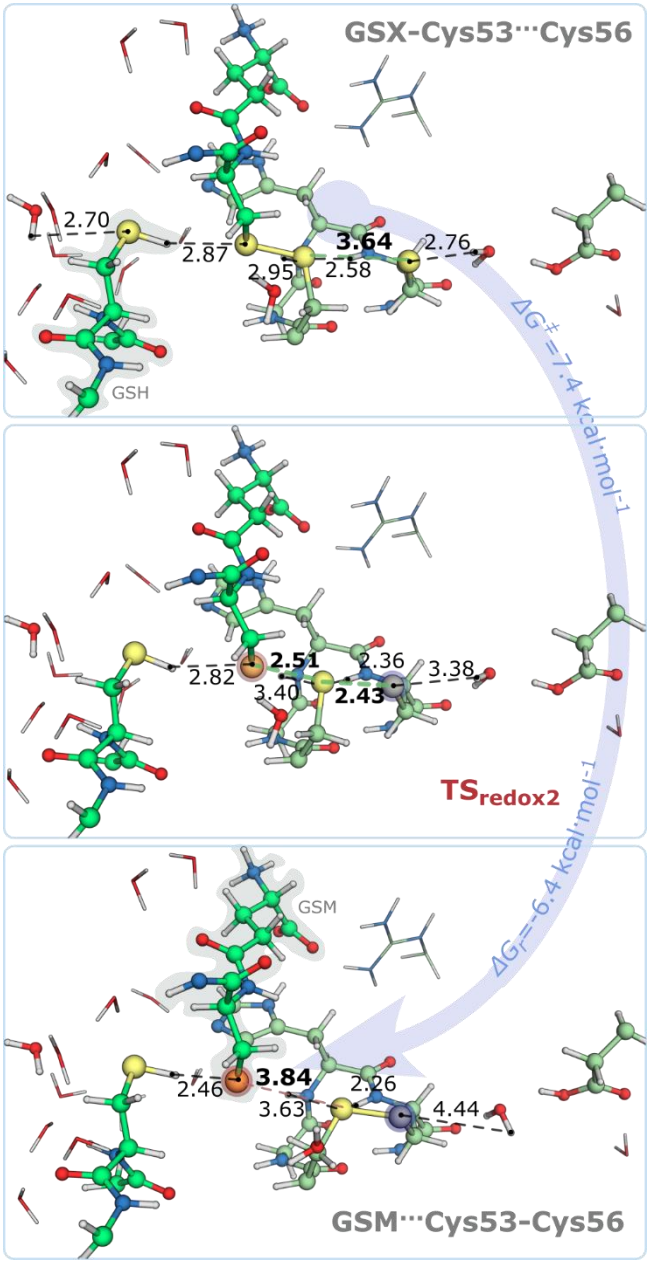


Figure 6.10. Stationary points for the cleavage of the mixed-disulphide intermediate by the Cys56-thiolate (GSX-Cys53···Cys56, TS_{redox2}, GSM···Cys53-Cys56), and main distances throughout the transformation, in Å. Distances in bold correspond to distances directly related to the reaction coordinate explored through linear transit scans. Blue-to-red shades in atoms represent the variation in atomic charge relative to the GSX-Cys53···Cys56 state (from 0.07 to 0.30 a.u.); blue stands for decrease in atomic charge, and red stands for increase in atomic charge.

The reaction is very similar to that in Figure 6.8: the Cys56-thiolate initially forms a hydrogen bond to a nearby water molecule, and it is in a nearly linear position with the mixed-disulphide (*circa* 160°). However, contrarily to the reactant state (GSSG), the Cys56-thiolate is only accepting one more hydrogen bond from the backbone-NH of Ala50 (in the MM layer). In the transition state, there is substantial stretching of the hydrogen bond established between the Cys56-thiolate and the water molecule, while the hydrogen bond between the Cys53-sulphur and the backbone-NH of Cys56 shortens. This latter interaction is not observed in TS_{redox1}, in which there are no hydrogen bonds with the central GSSG-sulphur of the linear trisulphide anion (with an $S_{GSSG}\cdots S_{GSSG}\cdots S_{Cys53}$ angle of 167°), and, together with the lower number of hydrogen bond donors for the Cys56-thiolate in the reactant state (GSX–Cys53···Cys56), it may be responsible for the lower Gibbs activation free-energy that is registered for the step. When the new GSM product is formed (GSM···Cys53–Cys56), it only forms a hydrogen bond with the neighbouring GSH from the first step of the reaction.

At this stage, we have to refer that, contrarily to what is stated in experimental studies,[495,533] the oxidation of the active site of the domain *a* of hPDI does not seem to be the rate-limiting step of the reduction of GSSG to GSH. However, we recall our initial observations regarding the oxidation of the *a*-domain in the presence of thiolates, and the apparently contradictory results of Lappi *et al*, in which the oxidation of the *a*-domain after being initially reduced by GSH has shown a turnover rate of 430 s^{-1}.[495] It seems that other mechanisms can be involved in this chemical step. One of them could be related to the competitivity between the oxidation of PDI by GSSG and its reduction by GSH. Since GSH is available in higher concentrations than GSSG, it is more likely that a mixed-disulphide between $hPDI_{ox}$ and GSH is formed. However, since GSH is shielded mostly by solvent, the attack of the Cys56-thiolate should be effective to restore the oxidized state of domain *a*. We note that, in fact, the proposed activation barrier for this reaction (~14 kcal·mol^{-1})[495] is very similar to the reverse barrier for the cleavage of the mixed-disulphide intermediate by Cys56 that we have determined (13.8 kcal·mol^{-1}). In addition, for a GSH concentration of 50 mM (the highest concentration employed in the study by Lappi *et al*) the pseudo-first-order rate constant for the first reduction of the oxidized *a*-domain by GSH is of about 560 s^{-1}, which corresponds to a Gibbs activation free-energy of about the same as that obtained for the reverse reaction. To conclude, we believe that this same reaction during the oxidation of the reduced *a*-domain will comprehend additional steps, such as the protonation of the exiting GSM to GSH, or its unbinding from the catalytic site to the solvent media. Regarding this last phenomenon, we have performed rigid-backbone cMD simulations for two models: $hPDI_{ox}$–

GSX:GSM and $hPDI_{ox}$–GSX:GSH. These simulations have shown that GSM and GSH exit the active site of the *a*-domain in less than 10 ns and 1 ns, respectively, for a concentration of approximately 1 mM. However, we expect that at larger concentrations, and in a more complex environment, this unbinding period may differ considerably. On a final note, we want to point out that we believe that the rate-limiting nature of the cleavage of the mixed-disulphide intermediate is mainly concerned with the nature of the environment of the *a*-domain, and not with the thiol-disulphide exchange by the Cys56-thiolate itself.

6.4. Conclusions

The catalytic mechanism of the reduction of glutathione disulphide (GSSG) by the *a*-domain of human protein disulphide isomerase (hPDI) has been established through a combination of molecular dynamics simulations (cMD), the ONIOM methodology and several insights from experimental studies. The *a*-domain has been referred to be intrinsically related to the activity of hPDI; hence, this study also provides mechanistic insight that should be transferable to the catalysis of hPDI, as a whole, in the oxidative protein folding pathway.

Our study has shown that the reaction is conducted in two stages: (1) nucleophilic attack of a Cys53-thiolate in the GSSG-disulophide, followed by instant deprotonation of a Cys56-thiol by Glu47, via a bridging water molecule, and formation of a glutathionate species (GSM); (2) following the protonation or the exit of GSM from the active site, a second nucleophilic attack is conducted by the Cys56-thiolate on the mixed-disulphide intermediate (between the Cys53-sulphur and a glutathione-sulphur), forming a second GSM species and oxidizing the *a*-domain through a Cys53-S–S-Cys56 disulphide bond. We provide atomistic and thermodynamic detail for every step along these reactions.

The first thiol-disulphide exchange reaction exhibits the higher Gibbs activation free-energy, comparatively to the second one (16.3 kcal·mol^{-1} vs 7.4 kcal·mol^{-1}), and it is in good agreement with experimentally derived pseudo-first-order turnover rates that predict a Gibbs activation free-energy of 17.6 kcal·mol^{-1} for the complete oxidation of the *a*-domain. Nevertheless, we cannot affirm that it is the rate-limiting step of the full catalytic cycle, since most of the kinetics of the cycle relies on the ratio of thiol/disulphides in the environment of the enzyme. In agreement with experimental data, the mixed-disulphide intermediate is very stable, relative to the reduced form of the *a*-domain (exhibiting an inverse Gibbs activation free-energy of 25.7 kcal·mol^{-1}). Furthermore, we propose that the decreasing in the pK_a of the Cys56-thiol can be related to the hydrophobic patch that occurs upon the formation of the mixed-disulphide with the Cys53-sulphur. To support this hypothesis, we refer that the deprotonation of Cys56-thiol is very favourable thermodynamically, and it is not accompanied with any major changes in the active site of the *a*-domain. In addition, we also observe that there is a large increase in the entropy of the active site upon formation of the mixed disulphide ($T\Delta S$ of about 10 kcal·mol^{-1} relatively to reactant state), which should also be related to the formation of the disulphide bond between the Cys53-thiolate and the GSSG-disulphide. The Gibbs reaction free-energy determined for the step was -9.5 kcal·mol^{-1}.

Relatively to the oxidation of the *a*-domain, it results from the cleavage of the mixed-disulphide intermediate by the Cys56-thiolate. We propose that this second nucleophilic attack should only occur in this way, if there is a polar environment that can stabilize the exiting GSM. Should that occur, the reaction occurs quickly. Despite that our results seem to be contrary to those proposed so far, we do not see this as the case. In fact, our results support, as it was observed for the reduction of the *a*-domain, that the reaction indeed occurs in a rate similar to that registered for the reverse step of the formation of the mixed-disulphide intermediate by the nucleophilic attack of glutathione (GSH). Since the mixed-disulphide intermediate was not obtained experimentally so far, we believe that the rate measured for the intramolecular oxidation of the *a*-domain (which was indicated to be 460 s^{-1}, or ~14 $kcal \cdot mol^{-1}$ from transition state theory considerations), may in fact be, in these latter conditions, that of nucleophilic attack of GSH to the oxidized *a*-domain (which our calculations predicted to have a Gibbs activation free-energy of 13.8 $kcal \cdot mol^{-1}$).

A closer analysis of the behaviour of the Cys53–Gly54–His55–Cys56 motif supports that the backbone-NH groups of His55 and Cys56 are very important to stabilize the nucleophilic Cys53-thiolate. Our study as pointed out that the His55-imidazole should be in its neutral state throughout the catalysis. The His55-imidazolium form seems to prevent the linear arrangement of the trisulphide anion (in the transition state), hampering the formation of the mixed-disulphide intermediate.

Overall, our study has provided clarification in the catalysis by hPDI that was not yet displayed by other experimental works. We realize that the kinetics of this mechanism may differ in the endoplasmic reticulum; however, we believe that this structural and thermodynamic detail that we have provided, will be of great use for future assays concerning the regulation/inhibition of the activity of hPDI. Furthermore, we are aware that hPDI is promiscuous enzyme that functions as an oxidoreductase, an isomerase and a chaperone. Hence, there should be diverse factors that influence its activity, aside from its catalytic sites. However, improving the knowledge of the way in which its domains catalyse reactions should be of outmost importance to control most of its function in cells.

References

(1) Metals in chemical biology. *Nat. Chem. Biol.* **2008**, *4*, 143-143.

(2) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(3) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. OPLS Potential Functions for Nucleotide Bases - Relative Association Constants of Hydrogen-Bonded Base-Pairs in Chloroform. *J. Am. Chem. Soc.* **1991**, *113*, 2810-2819.

(4) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955-1970.

(5) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.

(6) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(7) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

(8) Hermans, J.; Berendsen, H. J. C.; Vangunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **1984**, *23*, 1513-1518.

(9) Dauberosguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand-Binding to Proteins - Escherichia-Coli Dihydrofolate Reductase Trimethoprim, a Drug-Receptor System. *Proteins* **1988**, *4*, 31-47.

(10) Becker, O. M.; Alexander D. MacKerell, J.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics*. Marcel Dekker, Inc.: 2001.

(11) Woo, H. J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.

(12) Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Curr. Med. Chem.* **2010**, *17*, 767-785.

(13) de Ruiter, A.; Boresch, S.; Oostenbrink, C. Comparison of thermodynamic integration and Bennett's acceptance ratio for calculating relative protein-ligand binding free energies. *J. Comput. Chem.* **2013**, *34*, 1024-1034.

(14) Goldfeld, D. A.; Murphy, R.; Kim, B.; Wang, L. L.; Beurning, T.; Abel, R.; Friesner, R. A. Docking and Free Energy Perturbation Studies of Ligand Binding in the Kappa Opioid Receptor. *J. Phys. Chem. B* **2015**, *119*, 824-835.

(15) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.

(16) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force-Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262-8270.

(17) Pang, Y. P.; Xu, K.; El Yazal, J.; Prendergast, F. G. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9*, 1857-1865.

(18) Stote, R. H.; Karplus, M. Zinc-Binding in Proteins and Solution - a Simple but Accurate Nonbonded Representation. *Proteins* **1995**, *23*, 12-31.

(19) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Effective tailor-made force field parameterization of the several Zn coordination environments in the puzzling FTase enzyme: opening the door to the full understanding of its elusive catalytic mechanism. *Theor. Chem. Acc.* **2007**, *117*, 171-181.

(20) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theor. Comp.* **2010**, *6*, 2935-2947.

(21) Wu, R. B.; Lu, Z. Y.; Cao, Z. X.; Zhang, Y. K. A Transferable Nonbonded Pairwise Force Field to Model Zinc Interactions in Metalloproteins. *J. Chem. Theor. Comp.* **2011**, *7*, 433-443.

(22) Neves, R. P. P.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Parameters for Molecular Dynamics Simulations of Manganese-Containing Metalloproteins. *J. Chem. Theor. Comp.* **2013**, *9*, 2718-2732.

(23) Carvalho, A. T.; Teixeira, A. F.; Ramos, M. J. Parameters for molecular dynamics simulations of iron-sulfur proteins. *J. Comput. Chem.* **2013**, *34*, 1540-8.

(24) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C. L. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Phys. Chem. B* **2014**, *118*, 4351-4362.

(25) Norrby, P. O.; Brandt, P. Deriving force field parameters for coordination complexes. *Coordin. Chem. Rev.* **2001**, *212*, 79-109.

(26) Banci, L. Molecular dynamics simulations of metalloproteins. *Curr. Opin. Chem. Biol.* **2003**, *7*, 143-149.

(27) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. Molecular modelling for transition metal complexes: Dealing with d-electron effects. *Coordin. Chem. Rev.* **2009**, *253*, 795-816.

(28) Hu, L. H.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J. Chem. Theor. Comp.* **2011**, *7*, 2452-2463.

(29) Thomas, L. H. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542-548.

(30) Fermi, E. Un metodo statistico per la determinazione di alcune proprietà dell' atomo. *Rend. Accad. Naz. Lincei* **1927**, 602-607.

(31) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864-B871.

(32) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *WIREs Comput. Mol. Sci.* **2012**, *2*, 327-350.

(33) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R. Z.; Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601-3658.

(34) Cramer, C. J.; Truhlar, D. G. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757-10816.

(35) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. Energy-Adjusted Ab initio Pseudopotentials for the 1st-Row Transition-Elements. *J. Chem. Phys.* **1987**, *86*, 866-872.

(36) Cao, X.; Dolg, M. Relativistic Pseudopotentials. In *Relativistic Methods for Chemists*, Barysz, M.; Ishikawa, Y., Eds. Springer Netherlands: 2010; Vol. 10, Chapter 6, pp 215-277.

(37) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213-222.

(38) Hariharan, P. C.; Pople, J. A. Accuracy of AH n equilibrium geometries by single determinant molecular orbital theory. *Mol. Phys.* **1974**, *27*, 209-214.

(39) Mclean, A. D.; Chandler, G. S. Contracted Gaussian-Basis Sets for Molecular Calculations .1. 2nd Row Atoms, Z=11-18. *J. Chem. Phys.* **1980**, *72*, 5639-5648.

(40) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *J. Chem. Phys.* **1980**, *72*, 650-654.

(41) Gordon, M. S. The isomers of silacyclopropane. *Chem. Phys. Lett.* **1980**, *76*, 163-168.

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

(43) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods .23. A Polarization-Type Basis Set for 2nd-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654-3665.

(44) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-Consistent Molecular-Orbital Methods .25. Supplementary Functions for Gaussian-Basis Sets. *J. Chem. Phys.* **1984**, *80*, 3265-3269.

(45) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829-5835.

(46) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian-Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571-2577.

(47) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.

(48) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.

(49) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358-1371.

(50) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. *Theor. Chim. Acta* **1990**, *77*, 123-141.

(51) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. The carboxylate shift in zinc enzymes: A computational study. *J. Am. Chem. Soc.* **2007**, *129*, 1378-1385.

(52) Lin, F.; Wang, R. X. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theor. Comp.* **2010**, *6*, 1852-1870.

(53) Siegbahn, P. E. M.; Himo, F. The quantum chemical cluster approach for modeling enzyme reactions. *WIREs Comput. Mol. Sci.* **2011**, *1*, 323-336.

(54) Cui, Q.; Riccardi, D.; Schaefer, P. S. Pushing the limit of QM/MM methods for studying complex biomolecular systems. *Abstr. Pap. Am. Chem. Soc.* **2004**, *228*, U229-U229.

(55) Rosta, E.; Klahn, M.; Warshel, A. Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *J. Phys. Chem. B* **2006**, *110*, 2934-2941.

(56) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185-199.

(57) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Edit.* **2009**, *48*, 1198-1229.

(58) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis - clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12431-12441.

(59) Ferrer, S.; Ruiz-Pernia, J.; Marti, S.; Moliner, V.; Tunon, I.; Bertran, J.; Andres, J. Hybrid Schemes Based on Quantum Mechanics/Molecular Mechanics Simulations: Goals to Success, Problems, and Perspectives. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 81-142.

(60) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708-2728.

(61) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. Investigating biological systems using first principles Car-Parrinello molecular dynamics simulations. *Curr. Opin. Struc. Biol.* **2007**, *17*, 149-156.

(62) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.

(63) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *106*, 3210-3235.

(64) Kamerlin, S. C. L.; Warshel, A. The empirical valence bond model: theory and applications. *WIREs Comput. Mol. Sci.* **2011**, *1*, 30-45.

(65) Fraústo da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*. 2nd ed.; Oxford University Press: New York, 2001; p 600.

(66) Lyons, T. J. Transport and Storage of Metal Ions in Biology. In *Biological Inorganic Chemistry: Structure and Reactivity*, 1st ed.; Bertini, I.; Gray, H.; Stiefel, E.; Valentine, J. S., Eds. University Science Books: 2007; Chapter 5, pp 57-77.

(67) Holm, R. H.; Kennepohl, P.; Solomon, E. I. Structural and functional aspects of metal sites in biology. *Chem. Rev.* **1996**, *96*, 2239-2314.

(68) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.

(69) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J. Metalloproteins and metal sensing. *Nature* **2009**, *460*, 823-830.

(70) Harding, M. M.; Nowicki, M. W.; Walkinshaw, M. D. Metals in protein structures: a review of their principal features. *Crystallogr Rev.* **2010**, *16*, 247-302.

(71) Bill, E. Iron-sulfur clusters-new features in enzymes and synthetic models. *Hyperfine Interact.* **2012**, *205*, 139-147.

(72) Siegbahn, P. E. M. Substrate Water Exchange for the Oxygen Evolving Complex in PSII in the S-1, S-2, and S-3 States. *J. Am. Chem. Soc.* **2013**, *135*, 9442-9449.

(73) Roat-Malone, R. M. *Bioinorganic Chemistry: A Short Course*. 2nd ed.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2007; p 544.

(74) Sousa, S. F.; Lopes, A. B.; Fernandes, P. A.; Ramos, M. J. The Zinc proteome: a tale of stability and functionality. *Dalton Trans.* **2009**, 7946-7956.

(75) Bras, N. F.; Ribeiro, A. J. M.; Oliveira, M.; Paixao, N. M.; Tamames, J. A.; Fernandes, P. A.; Ramos, M. J. Analyses of cobalt-ligand and potassium-ligand bond lengths in metalloproteins: trends and patterns. *J. Mol. Model.* **2014**, *20*.

(76) Metzler, D. *Biochemistry: The Chemical Reactions of Living Cells*. 2nd ed.; Elsevier Academic Press: 2001; Vol. 1, p 968.

(77) Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol. Inorg. Chem.* **2008**, *13*, 1205-1218.

(78) Dismukes, G. C. Manganese enzymes with binuclear active sites. *Chem. Rev.* **1996**, *96*, 2909-2926.

(79) Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J. Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* **2014**, *10*, 50-55.

(80) Solomon, E. I.; Heppner, D. E.; Johnston, E. M.; Ginsbach, J. W.; Cirera, J.; Qayyum, M.; Kieber-Emmons, M. T.; Kjaergaard, C. H.; Hadt, R. G.; Tian, L. Copper Active Sites in Biology. *Chem. Rev.* **2014**, *114*, 3659-3853.

(81) Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **1935**, *3*, 107-115.

(82) Laidler, K. J.; King, M. C. The Development of Transition-State Theory. *J. Phys. Chem.* **1983**, *87*, 2657-2664.

(83) Pineda, J. R. E. T.; Schwartz, S. D. Protein dynamics and catalysis: the problems of transition state theory and the subtlety of dynamic control. *Philos. T. R. Soc. B* **2006**, *361*, 1433-1438.

(84) Layfield, J. P.; Hammes-Schiffer, S. Hydrogen Tunneling in Enzymes and Biomimetic Models. *Chem. Rev.* **2014**, *114*, 3466-3494.

(85) Vallee, B. L.; Williams, R. J. Metalloenzymes: the entatic nature of their active sites. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *59*, 498-505.

(86) Srnec, M.; Aquilante, F.; Ryde, U.; Rulisek, L. Reaction Mechanism of Manganese Superoxide Dismutase Studied by Combined Quantum and Molecular Mechanical Calculations and Multiconfigurational Methods. *J. Phys. Chem. B* **2009**, *113*, 6074-6086.

(87) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. The Catalytic Mechanism of HIV-1 Integrase for DNA 3′-End Processing Established by QM/MM Calculations. *J. Am. Chem. Soc.* **2012**, *134*, 13436-13447.

(88) Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Gonzalez, P. J.; Moura, J. J. G.; Ramos, M. J. The Sulfur Shift: An Activation Mechanism for Periplasmic Nitrate Reductase and Formate Dehydrogenase. *Inorg. Chem.* **2013**, *52*, 10766-10772.

(89) Li, H.; Sun, H. NMR Studies of Metalloproteins. In *NMR of Proteins and Small Biomolecules*, Zhu, G., Ed. Springer Berlin Heidelberg: 2012; Vol. 326, Chapter 214, pp 69-98.

(90) Chatham, J. C.; Blackband, S. J. Nuclear Magnetic Resonance Spectroscopy and Imaging in Animal Research. *ILAR Journal* **2001**, *42*, 189-208.

(91) Rule, G. S.; Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*. 1 ed.; Springer Netherlands: Dordrecht, 2006; p XXVI, 532.

(92) Sommerhalter, M.; Lieberman, R. L.; Rosenzweig, A. C. X-ray crystallography and biological metal centers: Is seeing believing? *Inorg. Chem.* **2005**, *44*, 770-778.

(93) Arcovito, A.; Benfatto, M.; Cianci, M.; Hasnain, S. S.; Nienhaus, K.; Nienhaus, G. U.; Savino, C.; Strange, R. W.; Vallone, B.; Della Longa, S. X-ray structure analysis of a metalloprotein with enhanced active-site resolution using in situ x-ray absorption near edge structure spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6211-6216.

(94) Cotelesage, J. J. H.; Pushie, M. J.; Grochulski, P.; Pickering, I. J.; George, G. N. Metalloprotein active site structure determination: Synergy between X-ray absorption spectroscopy and X-ray crystallography. *J. Inorg. Biochem.* **2012**, *115*, 127-137.

(95) Solomon, E. I. Preface forum: "Functional insight from physical methods on metalloenzymes". *Inorg. Chem.* **2005**, *44*, 723-726.

(96) Lehnert, N.; George, S. D.; Solomon, E. I. Recent advances in bioinorganic spectroscopy. *Curr. Opin. Chem. Biol.* **2001**, *5*, 176-187.

(97) Koningsberger, D. C.; Prins, R. M. *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES*. John Wiley & Sons: New York, 1988; Vol. 92, p 688.

(98) Yano, J.; Yachandra, V. K. X-ray absorption spectroscopy. *Photosynth Res.* **2009**, *102*, 241-254.

(99) Leopoldini, M.; Russo, N.; Toscano, M. Which one among Zn(II), Co(II), Mn(II), and Fe(II) is the most efficient ion for the methionine aminopeptidase catalyzed reaction? *J. Am. Chem. Soc.* **2007**, *129*, 7776-7784.

(100) Ramos, M. J.; Fernandes, P. A. Computational enzymatic catalysis. *Acc. Chem. Res.* **2008**, *41*, 689-698.

(101) Amata, O.; Marino, T.; Russo, N.; Toscano, M. Human Insulin-Degrading Enzyme Working Mechanism. *J. Am. Chem. Soc.* **2009**, *131*, 14804-14811.

(102) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.

(103) Seebeck, B.; Reulecke, I.; Kamper, A.; Rarey, M. Modeling of metal interaction geometries for protein-ligand docking. *Proteins* **2008**, *71*, 1237-1254.

(104) Zhang, S. T.; Yan, H.; Wei, M.; Evans, D. G.; Duan, X. Valence Force Field for Layered Double Hydroxide Materials Based on the Parameterization of Octahedrally Coordinated Metal Cations. *J. Phys. Chem. C* **2012**, *116*, 3421-3431.

(105) Baxter, E. L.; Zuris, J. A.; Wang, C.; Vo, P. L. T.; Axelrod, H. L.; Cohen, A. E.; Paddock, M. L.; Nechushtai, R.; Onuchic, J. N.; Jennings, P. A. Allosteric control in a metalloprotein dramatically alters function. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 948-953.

(106) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Molecular Dynamics Simulations: Difficulties, Solutions and Strategies for Treating Metalloenzymes. In *Kinetics and Dynamics: From Nano- to Bio-*

Scale, Paneth, P.; Dybala-Defratyka, A., Eds. Springer: Dordrecht, The Netherlands, 2010; Chapter Chapter 11, p 299−330.

(107) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584-1604.

(108) Comba, P.; Ströhle, M.; Hambley, T. W. The Directionality of d-Orbitals and Molecular-Mechanics Calculations of Octahedral Transition-Metal Compounds. *Helv. Chim. Acta* **1995**, *78*, 2042-2047.

(109) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Direct derivation of van der Waals force field parameters from quantum mechanical interaction energies. *J. Phys. Chem. B* **2003**, *107*, 9601-9609.

(110) Babu, C. S.; Lim, C. Empirical force fields for biologically active divalent metal cations in water. *J. Phys. Chem. A* **2006**, *110*, 691-699.

(111) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing Solute-Water van der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116*, 4524-4534.

(112) Aqvist, J.; Warshel, A. Computer-Simulation of the Initial Proton-Transfer Step in Human Carbonic Anhydrase-I. *J. Mol. Biol.* **1992**, *224*, 7-14.

(113) Oelschlaeger, P.; Klahn, M.; Beard, W. A.; Wilson, S. H.; Warshel, A. Magnesium-cationic dummy atom molecules enhance representation of DNA polymerase beta in molecular dynamics simulations: Improved accuracy in studies of structural features and mutational effects. *J. Mol. Biol.* **2007**, *366*, 687-701.

(114) Liao, Q. H.; Kamerlin, S. C. L.; Strodel, B. Development and Application of a Nonbonded Cu^{2+} Model That Includes the Jahn-Teller Effect. *J. Phys. Chem. Lett.* **2015**, *6*, 2657-2662.

(115) Sousa, S. F.; Carvalho, E. S.; Ferreira, D. M.; Tavares, I. S.; Fernandes, P. A.; Ramos, M. J.; Gomes, J. A. N. F. Comparative analysis of the performance of commonly available density functionals in the determination of geometrical parameters for zinc complexes. *J. Comput. Chem.* **2009**, *30*, 2752-2763.

(116) Sousa, S. F.; Pinto, G. R. P.; Ribeiro, A. J. M.; Coimbra, J. T. S.; Fernandes, P. A.; Ramos, M. J. Comparative analysis of the performance of commonly available density functionals in the determination of geometrical parameters for copper complexes. *J. Comput. Chem.* **2013**, *34*, 2079-2090.

(117) Weatherburn, D. C. Manganese-containing Enzymes and Proteins. In *Handbook on Metalloproteins*, Bertini, I.; Siegel, A.; Sigel, H.; B, B., Eds. Marcel Dekker: New York, 2001; Chapter 8, p 193−268.

(118) Coimbra, J. T. S.; Moniz, T.; Bras, N. F.; Iyanova, G.; Fernandes, P. A.; Ramos, M. J.; Rangel, M. Relevant Interactions of Antimicrobial Iron Chelators and Membrane Models Revealed by Nuclear

Magnetic Resonance and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2014**, *118*, 14590-14601.

(119) Cerqueira, N. M. F. S. A.; Coelho, C.; Bras, N. F.; Fernandes, P. A.; Garattini, E.; Terao, M.; Romao, M. J.; Ramos, M. J. Insights into the structural determinants of substrate specificity and activity in mouse aldehyde oxidases. *J Biol. Inorg. Chem.* **2015**, *20*, 209-217.

(120) Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Opin. Struc. Biol.* **2001**, *11*, 236-242.

(121) Semrouni, D.; Isley, W. C.; Clavaguera, C.; Dognon, J. P.; Cramer, C. J.; Gagliardi, L. Ab Initio Extension of the AMOEBA Polarizable Force Field to Fe^{2+}. *J. Chem. Theor. Comp.* **2013**, *9*, 3062-3071.

(122) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theor. Comp.* **2013**, *9*, 5430-5449.

(123) Yang, Z. Z.; Wang, J. J.; Zhao, D. X. Valence State Parameters of All Transition Metal Atoms in Metalloproteins-Development of ABEEM sigma pi Fluctuating Charge Force Field. *J. Comput. Chem.* **2014**, *35*, 1690-1706.

(124) Li, P. F.; Song, L. F.; Merz, K. M. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theor. Comp.* **2015**, *11*, 1645-1657.

(125) Li, P. F.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883-895.

(126) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271-1277.

(127) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132-146.

(128) Maciel, G. S.; Garcia, E. Study of charges transferability for use in force fields. *Chem. Phys. Lett.* **2006**, *420*, 497-502.

(129) Martin, F.; Zipse, H. Charge distribution in the water molecule—A comparison of methods. *J. Comput. Chem.* **2005**, *26*, 97-105.

(130) Johansson, M. P.; Kaila, V. R. I.; Laakkonen, L. Charge parameterization of the metal centers in cytochrome c oxidase. *J. Comput. Chem.* **2008**, *29*, 753-767.

(131) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129-145.

(132) Chirlian, L. E.; Francl, M. M. Atomic Charges Derived from Electrostatic Potentials - a Detailed Study. *J. Comput. Chem.* **1987**, *8*, 894-905.

(133) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990**, *11*, 361-373.

(134) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

(135) Sigfridsson, E.; Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* **1998**, *19*, 377-395.

(136) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641.

(137) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The Development and Use of Quantum-Mechanical Molecular-Models .76. Am1 - a New General-Purpose Quantum-Mechanical Molecular-Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.

(138) Smith, D. M. A.; Xiong, Y.; Straatsma, T. P.; Rosso, K. M.; Squier, T. C. Force-Field Development and Molecular Dynamics of [NiFe] Hydrogenase. *J. Chem. Theor. Comp.* **2012**, *8*, 2103-2114.

(139) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428-3439.

(140) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*. Oxford University Press: New York, 1994; p 456.

(141) Hu, L.; Söderhjelm, P.; Ryde, U. Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations. *J. Chem. Theor. Comp.* **2013**, *9*, 640-649.

(142) Friesner, R. A.; Dunietz, B. D. Large-Scale ab Initio Quantum Chemical Calculations on Biological Systems. *Acc. Chem. Res.* **2001**, *34*, 351-358.

(143) Brás, N. F.; Coimbra, J. T. S.; Neves, R. P. P.; Cerqueira, N. M. F. S. A.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational Biochemistry. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier: 2015; Chapter.

(144) Bras, N. F.; Ramos, M. J.; Fernandes, P. A. DFT studies on the beta-glycosidase catalytic mechanism: The deglycosylation step. *J. Mol. Struc.-Theochem* **2010**, *946*, 125-133.

(145) Ribeiro, A. J. M.; Alberto, M. E.; Ramos, M. J.; Fernandes, P. A.; Russo, N. The Catalytic Mechanism of Protein Phosphatase 5 Established by DFT Calculations. *Chem. - Eur. J.* **2013**, *19*, 14081-14089.

(146) Neves, R. P. P.; Fernandes, P. A.; Ramos, M. J. Unveiling the Catalytic Mechanism of $NADP^+$-Dependent Isocitrate Dehydrogenase with QM/MM Calculations. *ACS Catalysis* **2015**, 357-368.

(147) Lewars, E. G. Semiempirical Calculations. In *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 2nd ed.; Springer Science: 2011; Chapter 6, pp 391-444.

(148) Xue, Y.; Ward, J. M.; Yuwen, T. R.; Podkorytov, I. S.; Skrynnikov, N. R. Microsecond Time-Scale Conformational Exchange in Proteins: Using Long Molecular Dynamics Trajectory To Simulate NMR Relaxation Dispersion Data. *J. Am. Chem. Soc.* **2012**, *134*, 2555-2562.

(149) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341-346.

(150) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theor. Comp.* **2012**, *8*, 2997-3002.

(151) Deng, N. J.; Zheng, W. H.; Gallicchio, E.; Levy, R. M. Insights into the Dynamics of HIV-1 Protease: A Kinetic Network Model Constructed from Atomistic Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 9387-9394.

(152) Palazzesi, F.; Barducci, A.; Tollinger, M.; Parrinello, M. The allosteric communication pathways in KIX domain of CBP. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14237-14242.

(153) González, M. A. Force fields and molecular dynamics simulations. *JDN* **2011**, *12*, 169-200.

(154) Urey, H. C.; Bradley, C. A. The Vibrations of Pentatonic Tetrahedral Molecules. *Phys. Rev.* **1931**, *38*, 1969-1978.

(155) Ewald, P. P. The calculation of optical and electrostatic grid potential. *Ann. Phys. (Berlin)* **1921**, *64*, 253-287.

(156) Debye, P. Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Math. Ann.* **1909**, *67*, 535-558.

(157) Hestenes, M. R.; Stiefel, E. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Nat. Bur. Stand.* **1952**, *49*, 409-436.

(158) Shewchuck, J. R., An Introduction to the Conjugate Gradient Method without the Agonizing Pain. School of Computer Science Carnegie Mellon University: Pittsburgh, 1994.

(159) Comba, P.; Hambley, T. W.; Martin, B. *Molecular Modeling of Inorganic Compounds*. Wiley: Weinheim, Germany, 2008.

(160) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*. 2nd ed.; John Wiley & Sons, Ltd: New York, 2004.

(161) Leach, A. *Molecular Modelling: Principles and Applications*. 2nd ed.; Prentice Hall: Harlow, England, 2001; p 784.

(162) Verlet, L. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98-+.

(163) Omelyan, I. P.; Mryglod, I. M.; Folk, R. New optimized algorithms for molecular dynamics simulations. *Cond. Matt. Phys.* **2002**, *5*, 369–390.

(164) Skeel, R. D. What Makes Molecular Dynamics Work? *SIAM J. Sci. Stat. Comput.* **2009**, *31*, 1363-1378.

(165) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

(166) Paquet, E.; Viktor, H. L. Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review. *Biomed Res. Int.* **2015**.

(167) Feig, M.; Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struc. Biol.* **2004**, *14*, 217-224.

(168) Kleinjung, J.; Fraternali, F. Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struc. Biol.* **2014**, *25*, 126-134.

(169) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, Rna, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193-4194.

(170) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(171) Langevin, P. Sur la théorie du mouvement brownien. *Comptes-rendus de l'Académie des Sciences* **1908**, *146*, 530-532.

(172) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090-2098.

(173) Pottier, N. Brownian motion: the Langevin equation. In *Nonequilibrium Statistical Physics: Linear Irreversible Processes*, Oxford University Press: New York, 2009; Chapter 10, pp 235-375.

(174) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struc. Biol.* **2013**, *23*, 58-65.

(175) Borhani, D. W.; Shaw, D. E. The future of molecular dynamics simulations in drug discovery. *J. Comput. Aid. Mol. Des.* **2012**, *26*, 15-26.

(176) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429-452.

(177) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106*, 1589-1615.

(178) van der Kamp, M. W.; Mulholland, A. J. Computational enzymology: insight into biological catalysts from modelling. *Natural Product Reports* **2008**, *25*, 1001-1014.

(179) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. *J. Phys. Chem. A* **2007**, *111*, 10439-10452.

(180) Peverati, R.; Truhlar, D. G. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos. Trans. R. Soc., A* **2014**, *372*.

(181) Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys. (Berlin)* **1927**, *389*, 457-484.

(182) Cohen-Tannouodji, C.; Diu, B.; Laloë, F. *Quantum Mechanics*. John Wiley & Sons: Paris, 1977; Vol. 1.

(183) Gerlach, W.; Stern, O. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik* **1922**, *9*, 349-352.

(184) Gerlach, W.; Stern, O. Das magnetische Moment des Silberatoms. *Zeitschrift für Physik* **1922**, *9*, 353-355.

(185) Slater, J. C. The theory of complex spectra. *Phys. Rev.* **1929**, *34*, 1293-1322.

(186) Pavarini, E.; Koch, E.; Schollwock, U. Emergent Phenomena in Correlated Matter. In *Lecture Notes of the Autumn School on Correlated Electrons 2013*, Pavarini, E.; Koch, E.; Schollwock, U., Eds. Jülich Forschungsztrum: 2013; Vol. 3, Chapter 2, p 520.

(187) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69-89.

(188) Neves, R. P. P.; Fernandes, P. A.; Varandas, A. J. C.; Ramos, M. J. Benchmarking of Density Functionals for the Accurate Description of Thiol-Disulfide Exchange. *J. Chem. Theor. Comp.* **2014**, *10*, 4842-4856.

(189) Ruttink, P. A.; van Lenthe, J. The optimization of MCSCF functions by application of the generalized brillouin theorem: The LiH2 potential energy surface. *Theor. Chim. Acta* **1977**, *44*, 97-107.

(190) Raghavachari, K.; Pople, J. A. Calculation of One-Electron Properties Using Limited Configuration-Interaction Techniques. *Int. J. Quantum Chem.* **1981**, *20*, 1067-1071.

(191) Krishnan, R.; Schlegel, H. B.; Pople, J. A. Derivative Studies in Configuration-Interaction Theory. *J. Chem. Phys.* **1980**, *72*, 4654-4655.

(192) Pople, J. A.; Seeger, R.; Krishnan, R. Variational Configuration Interaction Methods and Comparison with Perturbation-Theory. *Int. J. Quantum Chem.* **1977**, 149-163.

(193) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 0618-0622.

(194) Brueckner, K. A. Many-Body Problem for Strongly Interacting Particles .2. Linked Cluster Expansion. *Phys. Rev.* **1955**, *100*, 36-45.

(195) Brandow, B. H. Linked-cluster expansions for the nuclear many-body problem. *Rev. Mod. Phys.* **1967**, *39*, 771-828.

(196) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256-4266.

(197) Cížek, J. *Correlation Effects in Atoms and Molecules*. Wiley Interscience: New York, 1969; Vol. 14

(198) Rezac, J.; Simova, L.; Hobza, P. CCSD[T] Describes Noncovalent Interactions Better than the CCSD(T), CCSD(TQ), and CCSDT Methods. *J. Chem. Theor. Comp.* **2013**, *9*, 364-369.

(199) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479-483.

(200) Rezac, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theor. Comp.* **2013**, *9*, 2151-2155.

(201) Ramabhadran, R. O.; Raghavachari, K. Extrapolation to the Gold-Standard in Quantum Chemistry: Computationally Efficient and Accurate CCSD(T) Energies for Large Molecules Using an Automated Thermochemical Hierarchy. *J. Chem. Theor. Comp.* **2013**, *9*, 3986-3994.

(202) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(203) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press: New York, 1989; Vol. 16.

(204) Rajagopal, A. K.; Kimball, J. C.; Banerjee, M. Short-Ranged Correlations and Ferromagnetic Electron-Gas. *Phys. Rev. B* **1978**, *18*, 2339-2345.

(205) Leininger, T.; Stoll, H.; Werner, H. J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **1997**, *275*, 151-160.

(206) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540-3544.

(207) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and Assessment of a New Hybrid Density Functional Model for Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108*, 2715-2719.

(208) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.

(209) Andrews, J. S.; Jayatilaka, D.; Bone, R. G. A.; Handy, N. C.; Amos, R. D. Spin Contamination in Single-Determinant Wave-Functions. *Chem. Phys. Lett.* **1991**, *183*, 423-431.

(210) Cohen, A. J.; Tozer, D. J.; Handy, N. C. Evaluation of < S-2 > in density functional theory. *J. Chem. Phys.* **2007**, *126*.

(211) Ribeiro, A. n. J. M.; Ramos, M. J.; Fernandes, P. A. Benchmarking of DFT Functionals for the Hydrolysis of Phosphodiester Bonds. *J. Chem. Theor. Comp.* **2010**, *6*, 2281-2292.

(212) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670-6688.

(213) Brás, N. F.; Perez, M. A. S.; Fernandes, P. A.; Silva, P. J.; Ramos, M. J. Accuracy of Density Functionals in the Prediction of Electronic Proton Affinities of Amino Acid Side Chains. *J. Chem. Theor. Comp.* **2011**, *7*, 3898-3908.

(214) Maseras, F.; Morokuma, K. Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* **1995**, *16*, 1170-1179.

(215) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theor. Comp.* **2006**, *2*, 815-826.

(216) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. Quantum mechanics/molecular mechanics electrostatic embedding with continuous and discrete functions. *J. Phys. Chem. B* **2006**, *110*, 13682-13684.

(217) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *J. Chem. Theor. Comp.* **2010**, *6*, 3721-3734.

(218) Tao, P.; Fisher, J. F.; Shi, Q. C.; Vreven, T.; Mobashery, S.; Schlegel, H. B. Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue. *Biochemistry* **2009**, *48*, 9839-9847.

(219) Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *J. Chem. Theor. Comp.* **2014**, *10*, 1588-1598.

(220) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. Combined QM/MM Molecular Dynamics Study on a Condensed-Phase SN2 Reaction at Nitrogen: The Effect of Explicitly Including Solvent Polarization. *J. Chem. Theor. Comp.* **2007**, *3*, 1499-1509.

(221) Illingworth, C. J. R.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. Classical Polarization in Hybrid QM/MM Methods. *J. Phys. Chem. A* **2006**, *110*, 6487-6497.

(222) Zhang, Y.; Lin, H.; Truhlar, D. G. Self-Consistent Polarization of the Boundary in the Redistributed Charge and Dipole Scheme for Combined Quantum-Mechanical and Molecular-Mechanical Calculations. *J. Chem. Theor. Comp.* **2007**, *3*, 1378-1398.

(223) Zhang, Y.; Lin, H. Flexible-Boundary Quantum-Mechanical/Molecular-Mechanical Calculations: Partial Charge Transfer between the Quantum-Mechanical and Molecular-Mechanical Subsystems. *J. Chem. Theor. Comp.* **2008**, *4*, 414-425.

(224) Xavier, A.; Nicolas, F.; Jean-Louis, R. The Local Self-Consistent Field Principles and Applications to Combined Quantum Mechanical-Molecular Mechanical Computations on Biomacromolecular Systems. In *Combined Quantum Mechanical and Molecular Mechanical Methods*, American Chemical Society: 1998; Vol. 712, Chapter 15, pp 234-249.

(225) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **1998**, *102*, 4714-4721.

(226) Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7*, 718-730.

(227) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700-733.

(228) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struc.-Theochem* **1999**, *461*, 1-21.

(229) Bras, N. F.; Moura-Tamames, S. A.; Fernandes, P. A.; Ramos, M. J. Mechanistic Studies on the Formation of Glycosidase-Substrate and Glycosidase-Inhibitor Covalent Intermediates. *J. Comput. Chem.* **2008**, *29*, 2565-2574.

(230) Liao, R. Z.; Yu, J. G.; Himo, F. Mechanism of tungsten-dependent acetylene hydratase from quantum chemical calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 22523-22527.

(231) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theor. Comp.* **2012**, *8*, 3793-3803.

(232) Pinto, G. P.; Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A.; Toscano, M.; Russo, N. New insights in the catalytic mechanism of tyrosine ammonia-lyase given by QM/MM and QM cluster models. *Arch. Biochem. Biophys.* **2015**, *582*, 107-115.

(233) Cannon, W. R.; Singleton, S. F.; Benkovic, S. J. A perspective on biological catalysis. *Nat. Struct. Biol.* **1996**, *3*, 821-833.

(234) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. Frozen density functional free energy simulations of redox proteins: Computational studies of the reduction potential of plastocyanin and rusticyanin. *J. Am. Chem. Soc.* **2003**, *125*, 5025-5039.

(235) Crespo, A.; Marti, M. A.; Estrin, D. A.; Roitberg, A. E. Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J. Am. Chem. Soc.* **2005**, *127*, 6940-6941.

(236) Klahn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J. Phys. Chem. B* **2005**, *109*, 15645-15650.

(237) Doshi, U.; McGowan, L. C.; Ladani, S. T.; Hamelberg, D. Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5699-5704.

(238) McGowan, L. C.; Hamelberg, D. Conformational Plasticity of an Enzyme during Catalysis: Intricate Coupling between Cyclophilin A Dynamics and Substrate Turnover. *Biophys. J.* **2013**, *104*, 216-226.

(239) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.

(240) Ferrer, S.; Tuñón, I.; Martí, S.; Moliner, V.; Garcia-Viloca, M.; González-Lafont, À.; Lluch, J. M. A Theoretical Analysis of Rate Constants and Kinetic Isotope Effects Corresponding to Different Reactant Valleys in Lactate Dehydrogenase. *J. Am. Chem. Soc.* **2006**, *128*, 16851-16863.

(241) Sanchez-Martinez, M.; Marcos, E.; Tauler, R.; Field, M.; Crehuet, R. Conformational Compression and Barrier Height Heterogeneity in the N-Acetylglutamate Kinase. *J. Phys. Chem. B* **2013**, *117*, 14261-14272.

(242) Lodola, A.; Sirirak, J.; Fey, N.; Rivara, S.; Mor, M.; Mulholland, A. J. Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths. *J. Chem. Theor. Comp.* **2010**, *6*, 2948-2960.

(243) Roca, M.; Messer, B.; Hilvert, D.; Warshel, A. On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13877-13882.

(244) Benkovic, S. J.; Hammes, G. G.; Hammes-Schiffer, S. Free-Energy Landscape of Enzyme Catalysis. *Biochemistry* **2008**, *47*, 3317-3321.

(245) Zhang, Y. K.; Kua, J.; McCammon, J. A. Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies. *J. Phys. Chem. B* **2003**, *107*, 4459-4463.

(246) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catalysis* **2015**, *5*, 5617-5626.

(247) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562-12566.

(248) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*.

(249) Petersen, L.; Ardevol, A.; Rovira, C.; Reilly, P. J. Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi alpha-Mannosidase II: A QM/MM Metadynamics Investigation. *J. Am. Chem. Soc.* **2010**, *132*, 8291-8300.

(250) Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. Catalytic Itinerary in 1,3-1,4-β-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not Yet Fully Developed at the Oxocarbenium Ion-like Transition State. *J. Am. Chem. Soc.* **2011**, *133*, 20301-20309.

(251) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218-6226.

(252) Warshel, A.; Florián, J. The Empirical Valence Bond (EVB) Method. In *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd: 2002; Chapter.

(253) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471-2474.

(254) Senn, H. M.; Thiel, S.; Thiel, W. *J. Chem. Theory Comput.* **2005**, *1*, 494.

(255) Tafipolsky, M.; Amirjalayer, S.; Schmid, R. Ab initio parametrized MM3 force field for the metal-organic framework MOF-5. *J. Comput. Chem.* **2007**, *28*, 1169-1176.

(256) Tamames, J. A. C.; Ramos, M. J. Metals in proteins: cluster analysis studies. *J. Mol. Model.* **2011**, *17*, 429-442.

(257) Siegbahn, P. E. M. Quantum chemical studies of manganese centers in biology. *Curr. Opin. Chem. Biol.* **2002**, *6*, 227-235.

(258) Yocum, C. F.; Pecoraro, V. L. Recent advances in the understanding of the biological chemistry of manganese. *Curr. Opin. Chem. Biol.* **1999**, *3*, 182-187.

(259) Stich, T. A.; Lahiri, S.; Yeagle, G.; Dicus, M.; Brynda, M.; Gunn, A.; Aznar, C.; DeRose, V. J.; Britt, R. D. Multifrequency pulsed EPR studies of biologically relevant manganese(II) complexes. *Appl. Magn. Reson.* **2007**, *31*, 321-341.

(260) Donini, O. A. T.; Kollman, P. A. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180-4188.

(261) Dal Peraro, M. D.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations. *J. Struct. Biol.* **2007**, *157*, 444-453.

(262) Marcial, B. L.; Sousa, S. F.; Barbosa, I. L.; Dos Santos, H. F.; Ramos, M. J. Chemically Modified Tetracyclines as Inhibitors of MMP-2 Matrix Metalloproteinase: A Molecular and Structural Study. *J. Phys. Chem. B* **2012**, *116*, 13644-13654.

(263) Lee, C. W.; Chakravorty, D. K.; Chang, F. M. J.; Reyes-Caballero, H.; Ye, Y. Z.; Merz, K. M.; Giedroc, D. P. Solution Structure of Mycobacterium tuberculosis NmtR in the Apo State: Insights into Ni(II)-Mediated Allostery. *Biochemistry* **2012**, *51*, 2619-2629.

(264) Sindhikara, D. J.; Roitberg, A. E.; Merz, K. M. Apo and Nickel-Bound Forms of the Pyrococcus horikoshii Species of the Metalloregulatory Protein: NikR Characterized by Molecular Dynamics Simulations. *Biochemistry* **2009**, *48*, 12024-12033.

(265) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. Quantum Mechanically Derived AMBER-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle. *J. Comput. Chem.* **2012**, *33*, 119-133.

(266) Comba, P.; Remenyi, R. Inorganic and bioinorganic molecular mechanics modeling - the problem of the force field parameterization. *Coordin. Chem. Rev.* **2003**, *238*, 9-20.

(267) Borgstahl, G. E. O.; Parge, H. E.; Hickey, M. J.; Beyer, W. F.; Hallewell, R. A.; Tainer, J. A. The Structure of Human Mitochondrial Manganese Superoxide-Dismutase Reveals a Novel Tetrameric Interface of 2 4-Helix Bundles. *Cell* **1992**, *71*, 107-118.

(268) Lah, M. S.; Dixon, M. M.; Pattridge, K. A.; Stallings, W. C.; Fee, J. A.; Ludwig, M. L. Structure-Function in Escherichia-Coli Iron Superoxide-Dismutase - Comparisons with the Manganese Enzyme from Thermus-Thermophilus. *Biochemistry* **1995**, *34*, 1646-1660.

(269) Koepke, J.; Olkhova, E.; Angerer, H.; Muller, H.; Peng, G. H.; Michel, H. High resolution crystal structure of Paracoccus denitrificans cytochrome c oxidase: New insights into the active site and the proton transfer pathways. *BBA-Bioenergetics* **2009**, *1787*, 635-645.

(270) Fedorov, A. A.; Fedorov, E. V.; Gerlt, J. A.; Burley, S. K.; Almo, S. C., Crystal structure of Glucarate dehydratase from Burkholderia cepacia complexed with magnesium. 2010.

(271) Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T. R.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A. M. Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4831-4836.

(272) Sakai, A.; Fedorov, A. A.; Fedorov, E. V.; Schnoes, A. M.; Glasner, M. E.; Brown, S.; Rutter, M. E.; Bain, K.; Chang, S.; Gheyi, T.; Sauder, J. M.; Burley, S. K.; Babbitt, P. C.; Almo, S. C.; Gerlt, J. A. Evolution of Enzymatic Activities in the Enolase Superfamily: Stereochemically Distinct Mechanisms in Two Families of cis,cis-Muconate Lactonizing Enzymes. *Biochemistry* **2009**, *48*, 1445-1453.

(273) Williams, R.; Holyoak, T.; McDonald, G.; Gui, C.; Fenton, A. W. Differentiating a ligand's chemical requirements for allosteric interactions from those for protein binding. Phenylalanine inhibition of pyruvate kinase. *Biochemistry* **2006**, *45*, 5421-5429.

(274) Vetting, M. W.; Wackett, L. P.; Que, L.; Lipscomb, J. D.; Ohlendorf, D. H. Crystallographic comparison of manganese- and iron-dependent homoprotocatechuate 2,3-dioxygenases. *J. Bacteriol.* **2004**, *186*, 1945-1958.

(275) Neidhart, D. J.; Howell, P. L.; Petsko, G. A.; Powers, V. M.; Li, R. S.; Kenyon, G. L.; Gerlt, J. A. Mechanism of the Reaction Catalyzed by Mandelate Racemase .2. Crystal-Structure of Mandelate Racemase at 2.5-a Resolution - Identification of the Active-Site and Possible Catalytic Residues. *Biochemistry* **1991**, *30*, 9264-9273.

(276) Lawrence, M. C.; Pilling, P. A.; Epa, V. C.; Berry, A. M.; Ogunniyi, A. D.; Paton, J. C. The crystal structure of pneumococcal surface antigen PsaA reveals a metal-binding site and a novel structure for a putative ABC-type binding protein. *Struct. Fold. Des.* **1998**, *6*, 1553-1561.

(277) Kleywegt, G. J.; Hoier, H.; Jones, T. A. A re-evaluation of the crystal structure of chloromuconate cycloisomerase. *Acta Crystallogr. D* **1996**, *52*, 858-863.

(278) Carrasco, R.; Morgenstern-Badarau, I.; Cano, J. Two proton-one electron coupled transfer in iron and manganese superoxide dismutases: A density functional study. *Inorg. Chim. Acta* **2007**, *360*, 91-101.

(279) Georgiev, V.; Borowski, T.; Siegbahn, P. E. M. Theoretical study of the catalytic reaction mechanism of MndD. *J Biol. Inorg. Chem.* **2006**, *11*, 571-585.

(280) Michel, H.; Behr, J.; Harrenga, A.; Kannt, A. Cytochrome C oxidase: Structure and spectroscopy. *Annu. Rev. Bioph. Biom.* **1998**, *27*, 329-356.

(281) Wooll, J. O.; Friesen, R. H. E.; White, M. A.; Watowich, S. J.; Fox, R. O.; Lee, J. C.; Czerwinski, E. W. Structural and functional linkages between subunit interfaces in mammalian pyruvate kinase. *J. Mol. Biol.* **2001**, *312*, 525-540.

(282) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

(283) Becke, A. D. Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.* **1996**, *104*, 1040-1046.

(284) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988**, *37*, 785-789.

(285) Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*.

(286) Blomberg, M. R. A.; Siegbahn, P. E. M. A quantum chemical approach to the study of reaction mechanisms of redox-active metalloenzymes. *J. Phys. Chem. B* **2001**, *105*, 9375-9386.

(287) Comba, P.; Remenyi, R. A new molecular mechanics force field for the oxidized form of blue copper proteins. *J. Comput. Chem.* **2002**, *23*, 697-705.

(288) Dunning, T. H., Jr.; Hay, P. J. *Modern Theoretical Chemistry*. Plenum Press: New York, 1976; Vol. 3.

(289) Ruth, K.; Tullmann, S.; Vitze, H.; Bolte, M.; Lerner, H. W.; Holthausen, M. C.; Wagner, M. Copper complexes of mono- and ditopic [(methylthio)methyl]borates: Missing links and linked systems en route to copper enzyme models. *Chem. - Eur. J.* **2008**, *14*, 6754-6770.

(290) Waller, M. P.; Braun, H.; Hojdis, N.; Buhl, M. Geometries of second-row transition-metal complexes from density-functional theory. *J. Chem. Theor. Comp.* **2007**, *3*, 2234-2242.

(291) Yoshikai, N.; Zhang, S. L.; Yamagata, K.; Tsuji, H.; Nakamura, E. Mechanistic Study of the Manganese-Catalyzed [2+2+2] Annulation of 1,3-Dicarbonyl Compounds and Terminal Alkynes. *J. Am. Chem. Soc.* **2009**, *131*, 4099-4109.

(292) Suarez, D.; Brothers, E. N.; Merz, K. M. Insights into the structure and dynamics of the dinuclear zinc beta-lactamase site from Bacteroides fragilis. *Biochemistry* **2002**, *41*, 6615-6630.

(293) Yao, L. S.; Sklenak, S.; Yan, H. G.; Cukier, R. I. A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase. *J. Phys. Chem. B* **2005**, *109*, 7500-7510.

(294) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.

(295) Merz, K. M.; Banci, L. Binding of bicarbonate to human carbonic anhydrase II: A continuum of binding states. *J. Am. Chem. Soc.* **1997**, *119*, 863-871.

(296) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10*. 2008.

(297) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Modell.* **2006**, *25*, 247-260.

(298) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(299) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(300) van der Spoel, D.; van Maaren, P. J. The origin of layer structure artifacts in simulations of liquid water. *J. Chem. Theor. Comp.* **2006**, *2*, 1-11.

(301) Rulisek, L.; Ryde, U. Structure of reduced and oxidized manganese superoxide dismutase: A combined computational and experimental approach. *J. Phys. Chem. B* **2006**, *110*, 11511-11518.

(302) Harding, M. M. The geometry of metal-ligand interactions relevant to proteins. *Acta Crystallogr. D* **1999**, *55*, 1432-1443.

(303) Harding, M. M. The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. *Acta Crystallogr. D* **2000**, *56*, 857-867.

(304) Rietsch, A.; Beckwith, J. The genetics of disulfide bond metabolism. *Annu. Rev. Genet.* **1998**, *32*, 163-184.

(305) Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A. Disulfide Bonds and Protein Folding. *Biochemistry* **2000**, *39*, 4207-4216.

(306) Bošnjak, I.; Bojović, V.; Šegvić-Bubić, T.; Bielen, A. Occurrence of protein disulfide bonds in different domains of life: a comparison of proteins from the Protein Data Bank. *Protein Eng., Des. Sel.* **2014**, *27*, 65-72.

(307) Sevier, C. S.; Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 836-847.

(308) Hogg, P. J. Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* **2003**, *28*, 210-214.

(309) Fernandes, P. A.; Ramos, M. J. Theoretical Insights into the Mechanism for Thiol/Disulfide Exchange. *Chem. - Eur. J.* **2004**, *10*, 257-266.

(310) Rabenstein, D. L.; Weaver, K. H. Kinetics and Equilibria of the Thiol/Disulfide Exchange Reactions of Somatostatin with Glutathione. *J. Org. Chem.* **1996**, *61*, 7391-7397.

(311) Iozzi, M. F.; Helgaker, T.; Uggerud, E. Influence of External Force on Properties and Reactivity of Disulfide Bonds. *J. Phys. Chem. A* **2011**, *115*, 2308-2315.

(312) Rothwarf, D. M.; Scheraga, H. A. Equilibrium and Kinetic Constants for the Thiol Disulfide Interchange Reaction between Glutathione and Dithiothreitol. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7944-7948.

(313) David, C.; Foley, S.; Mavon, C.; Enescu, M. Reductive unfolding of serum albumins uncovered by Raman spectroscopy. *Biopolymers* **2008**, *89*, 623-634.

(314) Singh, R.; Whitesides, G. M. Thiol-disulfide interchange. In *Sulphur-Containing Functional Groups*, Patai, S.; Rappoport, Z., Eds. John Wiley & Sons, Inc.: Chichester, UK. , 1993; Chapter 13, pp 633-658.

(315) Edman, J. C.; Ellis, L.; Blacher, R. W.; Roth, R. A.; Rutter, W. J. Sequence of Protein Disulfide Isomerase and Implications of Its Relationship to Thioredoxin. *Nature* **1985**, *317*, 267-270.

(316) Carvalho, A. T. P.; Swart, M.; van Stralen, J. N. P.; Fernandes, P. A.; Ramos, M. J.; Bickelhaupt, F. M. Mechanism of Thioredoxin-Catalyzed Disulfide Reduction. Activation of the Buried Thiol and Role of the Variable Active-Site Residues. *J. Phys. Chem. B* **2008**, *112*, 2511-2523.

(317) Li, W.; Gräter, F. Atomistic Evidence of How Force Dynamically Regulates Thiol/Disulfide Exchange. *J. Am. Chem. Soc.* **2010**, *132*, 16790-16795.

(318) Levin, L.; Zelzion, E.; Nachliel, E.; Gutman, M.; Tsfadia, Y.; Einav, Y. A Single Disulfide Bond Disruption in the β3 Integrin Subunit Promotes Thiol/Disulfide Exchange, a Molecular Dynamics Study. *Plos One* **2013**, *8*, e59175.

(319) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618-622.

(320) Maurice, D.; Head-Gordon, M. Analytical second derivatives for excited electronic states using the single excitation configuration interaction method: theory and application to benzo[a]pyrene and chalcone. *Mol. Phys.* **1999**, *96*, 1533-1541.

(321) Head-Gordon, M.; Rico, R. J.; Oumi, M.; Lee, T. J. A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. *Chem. Phys. Lett.* **1994**, *219*, 21-29.

(322) Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910-1918.

(323) Van Voorhis, T.; Head-Gordon, M. Two-body coupled cluster expansions. *J. Chem. Phys.* **2001**, *115*, 5033-5040.

(324) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51-57.

(325) Heyd, J.; Scuseria, G. E. Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.* **2004**, *121*, 1187-1192.

(326) Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. Tests of functionals for systems with fractional electron number. *J. Chem. Phys.* **2007**, *126*, 154109.

(327) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109.

(328) Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals. *J. Chem. Phys.* **2006**, *125*, 074106.

(329) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.

(330) Peverati, R.; Truhlar, D. G. Exchange-Correlation Functional with Good Accuracy for Both Structural and Energetic Properties while Depending Only on the Density and Its Gradient. *J. Chem. Theor. Comp.* **2012**, *8*, 2310-2319.

(331) Peverati, R.; Truhlar, D. G. An improved and broadly accurate local approximation to the exchange-correlation density functional: The MN12-L functional for electronic structure calculations in chemistry and physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 13171-13174.

(332) Peverati, R.; Truhlar, D. G. Screened-exchange density functionals with broad accuracy for chemistry and solid-state physics. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187-16191.

(333) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and

Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics†. *J. Phys. Chem. A* **2008**, *112*, 12868-12886.

(334) Schwabe, T.; Grimme, S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398-4401.

(335) Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *J. Phys. Chem. C* **2010**, *114*, 20801-20808.

(336) Zhang, I. Y.; Luo, Y.; Xu, X. XYG3s: Speedup of the XYG3 fifth-rung density functional with scaling-all-correlation method. *J. Chem. Phys.* **2010**, *132*, 194105.

(337) Zhang, I. Y.; Luo, Y.; Xu, X. Basis set dependence of the doubly hybrid XYG3 functional. *J. Chem. Phys.* **2010**, *133*, 104105.

(338) Zhang, Y.; Xu, X.; Goddard, W. A. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4963-4968.

(339) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theor. Comp.* **2010**, *6*, 107-126.

(340) Ondarza, R. Enzyme regulation by biological disulfides. *Biosci. Rep.* **1989**, *9*, 593-604.

(341) Fass, D. Disulfide Bonding in Protein Biophysics. *Annu. Rev. Biophys.* **2012**, *41*, 63-79.

(342) Zhao, Y.; Truhlar, D. G. Density Functional Theory for Reaction Energies: Test of Meta and Hybrid Meta Functionals, Range-Separated Functionals, and Other High-Performance Functionals. *J. Chem. Theor. Comp.* **2011**, *7*, 669-676.

(343) Dumont, É.; Laurent, A. D.; Assfeld, X.; Jacquemin, D. Performances of recently-proposed functionals for describing disulfide radical anions and similar systems. *Chem. Phys. Lett.* **2011**, *501*, 245-251.

(344) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787-1799.

(345) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(346) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456-1465.

(347) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the ospersion interaction. *J. Chem. Phys.* **2005**, *122*.

(348) Bartlett, R. J.; Purvis, G. D. Many-Body Perturbation-Theory, Coupled-Pair Many-Electron Theory, and Importance of Quadruple Excitations for Correlation Problem. *Int. J. Quantum Chem.* **1978**, *14*, 561-581.

(349) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Electron Correlation Theories and Their Application to Study of Simple Reaction Potential Surfaces. *Int. J. Quantum Chem.* **1978**, *14*, 545-560.

(350) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic Configuration-Interaction - a General Technique for Determining Electron Correlation Energies. *J. Chem. Phys.* **1987**, *87*, 5968-5975.

(351) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. An Efficient Reformulation of the Closed-Shell Coupled Cluster Single and Double Excitation (CCSD) Equations. *J. Chem. Phys.* **1988**, *89*, 7382-7387.

(352) Scuseria, G. E.; Schaefer, H. F. Is Coupled Cluster Singles and Doubles (CCSD) More Computationally Intensive Than Quadratic Configuration-Interaction (QCISD). *J. Chem. Phys.* **1989**, *90*, 3700-3703.

(353) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639-9646.

(354) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243-252.

(355) Truhlar, D. G. Basis-set extrapolation. *Chem. Phys. Lett.* **1998**, *294*, 45-48.

(356) Zhao, Y.; Truhlar, D. G. Infinite-basis calculations of binding energies for the hydrogen bonded and stacked tetramers of formic acid and formamide and their use for validation of hybrid DFT and ab initio methods. *J. Phys. Chem. A* **2005**, *109*, 6624-6627.

(357) Varandas, A. J. C. Extrapolating to the one-electron basis-set limit in electronic structure calculations. *J. Chem. Phys.* **2007**, *126*, 244105.

(358) Angelucci, F.; Dimastrogiovanni, D.; Boumis, G.; Brunori, M.; Miele, A. E.; Saccoccia, F.; Bellelli, A. Mapping the Catalytic Cycle of Schistosoma mansoni Thioredoxin Glutathione Reductase by X-ray Crystallography. *J. Biol. Chem.* **2010**, *285*, 32557-32567.

(359) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.

(360) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669-681.

(361) Hobza, P.; Sponer, J. Toward true DNA base-stacking energies: MP2, CCSD(T), and complete basis set calculations. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.

(362) Jurecka, P.; Hobza, P. On the convergence of the (Delta E-CCSD(T)-Delta E-MP2) term for complexes with multiple H-bonds. *Chem. Phys. Lett.* **2002**, *365*, 89-94.

(363) Daabkowska, I.; Jurecka, P.; Hobza, P. On geometries of stacked and H-bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level. *J. Chem. Phys.* **2005**, *122*.

(364) Ditchfie.R; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.

(365) Hehre, W. J.; Ditchfie.R; Pople, J. A. Self-Consistent Molecular-Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J. Chem. Phys.* **1972**, *56*, 2257-2261.

(366) Oliveira, E. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Mechanism of Formation of the Internal Aldimine in Pyridoxal 5 '-Phosphate-Dependent Enzymes. *J. Am. Chem. Soc.* **2011**, *133*, 15496-15505.

(367) Gesto, D. S.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Unraveling the Enigmatic Mechanism of L-Asparaginase II with QM/QM Calculations. *J. Am. Chem. Soc.* **2013**, *135*, 7146-7158.

(368) Maranzana, A.; Giordana, A.; Indarto, A.; Tonachini, G.; Barone, V.; Causa, M.; Pavone, M. Density functional theory study of the interaction of vinyl radical, ethyne, and ethene with benzene, aimed to define an affordable computational level to investigate stability trends in large van der Waals complexes. *J. Chem. Phys.* **2013**, *139*.

(369) Sekharan, S.; Mooney, V. L.; Rivalta, I.; Kazmi, M. A.; Neitz, M.; Neitz, J.; Sakmar, T. P.; Yan, E. C. Y.; Batista, V. S. Spectral Tuning of Ultraviolet Cone Pigments: An Interhelical Lock Mechanism. *J. Am. Chem. Soc.* **2013**, *135*, 19064-19067.

(370) Varandas, A. J. C. Accurate Determination of the Reaction Course in HY_2 (sic) Y + YH (Y = O, S): Detailed Analysis of the Covalent- to Hydrogen-Bonding Transition. *J. Phys. Chem. A* **2013**, *117*, 7393-7407.

(371) Varandas, A. J. C. Odd-hydrogen: An account on electronic structure, kinetics and role of water in mediating reactions with atmospheric ozone. Just a catalyst or far beyond? *Int. J. Quantum Chem.* **2014**, *118*, 1327-1349.

(372) Viegas, L. P.; Varandas, A. J. C. Coupled-cluster reaction barriers of HO_2+H_2O+O_3: An application of the coupled-cluster//Kohn–Sham density functional theory model chemistry. *J. Comput. Chem.* **2014**, *35*, 507-517.

(373) Wheeler, S. E.; Houk, K. N. Integration Grid Errors for Meta-GGA-Predicted Reaction Energies: Origin of Grid Errors for the M06 Suite of Functionals. *J. Chem. Theor. Comp.* **2010**, *6*, 395-404.

(374) Goerigk, L.; Reimers, J. R. Efficient Methods for the Quantum Chemical Treatment of Protein Structures: The Effects of London-Dispersion and Basis-Set Incompleteness on Peptide and Water-Cluster Geometries. *J. Chem. Theor. Comp.* **2013**, *9*, 3240-3251.

(375) Minenkov, Y.; Singstad, A.; Occhipinti, G.; Jensen, V. R. The accuracy of DFT-optimized geometries of functional transition metal compounds: a validation study of catalysts for olefin metathesis and other reactions in the homogeneous phase. *Dalton Trans.* **2012**, *41*, 5526-5541.

(376) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theor. Comp.* **2013**, *9*, 1918-1931.

(377) Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747-2757.

(378) Valdes, H.; Pluhackova, K.; Hobza, P. Phenylalanyl-Glycyl-Phenylalanine Tripeptide: A Model System for Aromatic-Aromatic Side Chain Interactions in Proteins. *J. Chem. Theor. Comp.* **2009**, *5*, 2248-2256.

(379) Kruse, H.; Goerigk, L.; Grimme, S. Why the Standard B3LYP/6-31G*Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem. *J. Org. Chem.* **2012**, *77*, 10824-10834.

(380) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. Determination of the Delta pKa between the active site cysteines of thioredoxin and DsbA. *J. Comput. Chem.* **2006**, *27*, 966-975.

(381) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. Theoretical Study of the Unusual Protonation Properties of the Active Site Cysteines in Thioredoxin. *J. Phys. Chem. B* **2006**, *110*, 5758-5761.

(382) Carvalho, A. P.; Fernandes, P. A.; Ramos, M. J. Similarities and differences in the thioredoxin superfamily. *Prog. Biophys. Mol. Biol.* **2006**, *91*, 229-248.

(383) Carvalho, A. T. P.; Fernandes, P. A.; Swart, M.; van Stralen, J. N. P.; Bickelhaupt, F. M.; Ramos, M. J. Role of the Variable Active Site Residues in the Function of Thioredoxin Family Oxidoreductases. *J. Comput. Chem.* **2009**, *30*, 710-724.

(384) Wiita, A. P.; Ainavarapu, R. K.; Huang, H. H.; Fernandez, J. M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7222-7227.

(385) Wiita, A. P.; Perez-Jimenez, R.; Walther, K. A.; Grater, F.; Berne, B. J.; Holmgren, A.; Sanchez-Ruiz, J. M.; Fernandez, J. M. Probing the chemistry of thioredoxin catalysis with force. *Nature* **2007**, *450*, 124.

(386) Alegre-Cebollada, J.; Perez-Jimenez, R.; Kosuri, P.; Fernandez, J. M. Single-molecule Force Spectroscopy Approach to Enzyme Catalysis. *J. Biol. Chem.* **2010**, *285*, 18961-18966.

(387) Perdew, J. P. *Electronic Structure of Solids '91*. Akademie Verlag: Berlin, 1991; Vol. 11.

(388) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces - Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys. Rev. B* **1992**, *46*, 6671-6687.

(389) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. Atoms, Molecules, Solids, and Surfaces - Applications of the Generalized Gradient Approximation for Exchange and Correlation (Vol 46, Pg 6671, 1992). *Phys. Rev. B* **1993**, *48*, 4978-4978.

(390) Perdew, J. P.; Burke, K.; Wang, Y. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B* **1996**, *54*, 16533-16539.

(391) Burke, K.; Perdew, J. P.; Wang, Y. *Electronic Density Functional Theory: Recent Progress and New Directions*. Springer Plenum: New York, 1998.

(392) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.* **1998**, *108*, 664-675.

(393) Boese, A. D.; Martin, J. M. L. Development of density functionals for thermochemical kinetics. *J. Chem. Phys.* **2004**, *121*, 3405-3416.

(394) Viegas, L. P.; Branco, A.; Varandas, A. J. C. How Well Can Kohn-Sham DFT Describe the $HO_2 + O_3$ Reaction? *J. Chem. Theor. Comp.* **2010**, *6*, 2751-2761.

(395) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133-A1138.

(396) Slater, J. C. *The Self-Consistent Field for Molecular and Solids, Quantum Theory of Molecular and Solids*. McGraw-Hill: New York, 1974; Vol. 4.

(397) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - a Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200-1211.

(398) Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Phys. Rev. B* **1981**, *23*, 5048-5079.

(399) Perdew, J. P. Density-Functional Approximation for the Correlation-Energy of the Inhomogeneous Electron-Gas. *Phys. Rev. B* **1986**, *33*, 8822-8824.

(400) Becke, A. D.; Roussel, M. R. Exchange Holes in Inhomogeneous Systems - a Coordinate-Space Model. *Phys. Rev. A* **1989**, *39*, 3761-3767.

(401) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation-Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200-206.

(402) Devlin, F. J.; Finley, J. W.; Stephens, P. J.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields: A Comparison of Local, Nonlocal, and Hybrid Density Functionals. *J. Phys. Chem.* **1995**, *99*, 16883-16902.

(403) Gill, P. M. W. A new gradient-corrected exchange functional. *Mol. Phys.* **1996**, *89*, 433-445.

(404) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.

(405) Adamo, C.; Barone, V. Toward reliable adiabatic connection models free from adjustable parameters. *Chem. Phys. Lett.* **1997**, *274*, 242-250.

(406) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554-8560.

(407) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple (vol 77, pg 3865, 1996). *Phys. Rev. Lett.* **1997**, *78*, 1396-1396.

(408) Adamo, C.; Barone, V. Implementation and validation of the Lacks-Gordon exchange functional in conventional density functional and adiabatic connection methods. *J. Comput. Chem.* **1998**, *19*, 418-429.

(409) Ernzerhof, M.; Perdew, J. P. Generalized gradient approximation to the angle- and system-averaged exchange hole. *J. Chem. Phys.* **1998**, *109*, 3313-3320.

(410) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.* **1998**, *109*, 6264-6271.

(411) Rey, J.; Savin, A. Virtual space level shifting and correlation energies. *Int. J. Quantum Chem.* **1998**, *69*, 581-590.

(412) Schmider, H. L.; Becke, A. D. Optimized density functionals from the extended G2 test set. *J. Chem. Phys.* **1998**, *108*, 9624-9631.

(413) Van Voorhis, T.; Scuseria, G. E. A novel form for the exchange-correlation energy functional. *J. Chem. Phys.* **1998**, *109*, 400-410.

(414) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158-6170.

(415) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.* **1999**, *82*, 2544-2547.

(416) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. Adiabatic Connection for Kinetics. *J. Phys. Chem. A* **2000**, *104*, 4811-4815.

(417) Handy, N. C.; Cohen, A. J. Left-right correlation energy. *Mol. Phys.* **2001**, *99*, 403-412.

(418) Hoe, W. M.; Cohen, A. J.; Handy, N. C. Assessment of a new local exchange functional OPTX. *Chem. Phys. Lett.* **2001**, *341*, 319-328.

(419) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. Hybrid exchange-correlation functional determined from thermochemical data and ab initio potentials. *J. Chem. Phys.* **2001**, *115*, 9233-9242.

(420) Kormos, B. L.; Cramer, C. J. Adiabatic connection method for X^- + RX nucleophilic substitution reactions (X = F, Cl). *J. Phys. Org. Chem.* **2002**, *15*, 712-720.

(421) Toulouse, J.; Savin, A.; Adamo, C. Validation and assessment of an accurate approach to the correlation problem in density functional theory: The Kriger-Chen-Iafrate-Savin model. *J. Chem. Phys.* **2002**, *117*, 10465-10473.

(422) Boese, A. D.; Martin, J. M. L.; Handy, N. C. The role of the basis set: Assessing density functional theory. *J. Chem. Phys.* **2003**, *119*, 3005-3014.

(423) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* **2003**, *107*, 1384-1388.

(424) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*.

(425) Xu, X.; Goddard, W. A. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673-2677.

(426) Zhao, Y.; Truhlar, D. G. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions. *J. Phys. Chem. A* **2004**, *108*, 6908-6918.

(427) Dahlke, E. E.; Truhlar, D. G. Improved Density Functionals for Water. *J. Phys. Chem. B* **2005**, *109*, 15677-15683.

(428) Zhao, Y.; González-García, N.; Truhlar, D. G. Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods. *J. Phys. Chem. A* **2005**, *109*, 2012-2018.

(429) Zhao, Y.; Truhlar, D. G. Benchmark Databases for Nonbonded Interactions and Their Use To Test Density Functional Theory. *J. Chem. Theor. Comp.* **2005**, *1*, 415-432.

(430) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theor. Comp.* **2006**, *2*, 364-382.

(431) Zhao, Y.; Truhlar, D. G. Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States. *J. Phys. Chem. A* **2006**, *110*, 13126-13130.

(432) Zhao, Y.; Truhlar, D. G. Comparative DFT Study of van der Waals Complexes: Rare-Gas Dimers, Alkaline-Earth Dimers, Zinc Dimer, and Zinc-Rare-Gas Dimers. *J. Phys. Chem. A* **2006**, *110*, 5121-5129.

(433) Zhao, Y.; Truhlar, D. 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. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(434) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. W. Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry. *Phys. Rev. Lett.* **2009**, *103*.

(435) Peverati, R.; Truhlar, D. G. Communication: A global hybrid generalized gradient approximation to the exchange-correlation functional that satisfies the second-order density-gradient constraint and has broad applicability in chemistry. *J. Chem. Phys.* **2011**, *135*.

(436) Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theor. Comp.* **2012**, *8*, 4989-5007.

(437) Peverati, R.; Truhlar, D. G. M11-L: A Local Density Functional That Provides Improved Accuracy for Electronic Structure Calculations in Chemistry and Physics. *J. Phys. Chem. Lett.* **2012**, *3*, 117-124.

(438) Grodsky, N. B.; Soundar, S.; Colman, R. F. Evaluation by site-directed mutagenesis of aspartic acid residues in the metal site of pig heart NADP-dependent isocitrate dehydrogenase. *Biochemistry* **2000**, *39*, 2193-2200.

(439) Kim, T. K.; Lee, P.; Colman, R. F. Critical role of Lys(212) and Tyr(140) in porcine NADP-dependent isocitrate dehydrogenase. *J. Biol. Chem.* **2003**, *278*, 49323-49331.

(440) Lin, Y.; West, A. H.; Cook, P. F. Site-Directed Mutagenesis as a Probe of the Acid-Base Catalytic Mechanism of Homoisocitrate Dehydrogenase from Saccharomyces cerevisiae. *Biochemistry* **2009**, *48*, 7305-7312.

(441) Yang, B.; Zhong, C.; Peng, Y. J.; Lai, Z.; Ding, J. P. Molecular mechanisms of "off-on switch" of activities of human IDH1 by tumor-associated mutation R132H. *Cell Res.* **2010**, *20*, 1188-1200.

(442) Goncalves, S.; Miller, S. P.; Carrondo, M. A.; Dean, A. M.; Matias, P. M. Induced Fit and the Catalytic Mechanism of Isocitrate Dehydrogenase. *Biochemistry* **2012**, *51*, 7098-7115.

(443) Quartararo, C. E.; Hazra, S.; Hadi, T.; Blanchard, J. S. Structural, Kinetic and Chemical Mechanism of Isocitrate Dehydrogenase-1 from Mycobacterium tuberculosis. *Biochemistry* **2013**, *52*, 1765-1775.

(444) Dang, L.; White, D. W.; Gross, S.; Bennett, B. D.; Bittinger, M. A.; Driggers, E. M.; Fantin, V. R.; Jang, H. G.; Jin, S.; Keenan, M. C.; Marks, K. M.; Prins, R. M.; Ward, P. S.; Yen, K. E.; Liau, L. M.; Rabinowitz, J. D.; Cantley, L. C.; Thompson, C. B.; Heiden, M. G. V.; Su, S. M. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **2009**, *462*, 739-744.

(445) Lu, C.; Ward, P. S.; Kapoor, G. S.; Rohle, D.; Turcan, S.; Abdel-Wahab, O.; Edwards, C. R.; Khanin, R.; Figueroa, M. E.; Melnick, A.; Wellen, K. E.; O'Rourke, D. M.; Berger, S. L.; Chan, T. A.; Levine, R. L.; Mellinghoff, I. K.; Thompson, C. B. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **2012**, *483*, 474-478.

(446) Figueroa, M. E.; Abdel-Wahab, O.; Lu, C.; Ward, P. S.; Patel, J.; Shih, A.; Li, Y. S.; Bhagwat, N.; Vasanthakumar, A.; Fernandez, H. F.; Tallman, M. S.; Sun, Z. X.; Wolniak, K.; Peeters, J. K.; Liu, W.; Choe, S. E.; Fantin, V. R.; Paietta, E.; Lowenberg, B.; Licht, J. D.; Godley, L. A.; Delwel, R.; Valk, P. J. M.; Thompson, C. B.; Levine, R. L.; Melnick, A. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **2010**, *18*, 553-567.

(447) Xu, W.; Yang, H.; Liu, Y.; Yang, Y.; Wang, P.; Kim, S. H.; Ito, S.; Yang, C.; Wang, P.; Xiao, M. T.; Liu, L. X.; Jiang, W. Q.; Liu, J.; Zhang, J. Y.; Wang, B.; Frye, S.; Zhang, Y.; Xu, Y. H.; Lei, Q. Y.; Guan, K. L.; Zhao, S. M.; Xiong, Y. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of alpha-Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* **2011**, *19*, 17-30.

(448) Yang, H.; Ye, D.; Guan, K. L.; Xiong, Y. IDH1 and IDH2 Mutations in Tumorigenesis: Mechanistic Insights and Clinical Perspectives. *Clin. Cancer. Res.* **2012**, *18*, 5562-5571.

(449) Wise, D. R.; Ward, P. S.; Shay, J. E. S.; Cross, J. R.; Gruber, J. J.; Sachdeva, U. M.; Platt, J. M.; DeMatteo, R. G.; Simon, M. C.; Thompson, C. B. Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of alpha-ketoglutarate to citrate to support cell growth and viability. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 19611-19616.

(450) Leonardi, R.; Subramanian, C.; Jackowski, S.; Rock, C. O. Cancer-associated Isocitrate Dehydrogenase Mutations Inactivate NADPH-dependent Reductive Carboxylation. *J. Biol. Chem.* **2012**, *287*, 14615-14620.

(451) Jo, S. H.; Son, M. K.; Koh, H. J.; Lee, S. M.; Song, I. H.; Kim, Y. O.; Lee, Y. S.; Jeong, K. S.; Kim, W. B.; Park, J. W.; Song, B. J.; Huhe, T. L. Control of mitochondrial redox balance and cellular defense against oxidative damage by mitochondrial NADP(+)-dependent isocitrate dehydrogenase. *J. Biol. Chem.* **2001**, *276*, 16168-16176.

(452) Lee, S. M.; Koh, H. J.; Park, D. C.; Song, B. J.; Huh, T. L.; Park, J. W. Cytosolic NADP(+)-dependent isocitrate dehydrogenase status modulates oxidative damage to cells. *Free Radical Bio. Med.* **2002**, *32*, 1185-1196.

(453) Kim, S. Y.; Park, J. W. Cellular defense against singlet oxygen-induced oxidative damage by cytosolic NADP(+)-dependent isocitrate dehydrogenase. *Free Radical Res.* **2003**, *37*, 309-316.

(454) Gillet, N.; Levy, B.; Moliner, V.; Demachy, I.; de la Lande, A. Electron and Hydrogen Atom Transfers in the Hydride Carrier Protein EmoB. *J. Chem. Theor. Comp.* **2014**, *10*, 5036-5046.

(455) Xu, X.; Zhao, J. Y.; Xu, Z.; Peng, B. Z.; Huang, Q. H.; Arnold, E.; Ding, J. P. Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. *J. Biol. Chem.* **2004**, *279*, 33946-33957.

(456) Ceccarelli, C.; Grodsky, N. B.; Ariyaratne, N.; Colman, R. F.; Bahnson, B. J. Crystal structure of porcine mitochondrial NADP(+)-dependent isocitrate dehydrogenase complexed with Mn^{2+} and isocitrate - Insights into the enzyme mechanism. *J. Biol. Chem.* **2002**, *277*, 43454-43462.

(457) Bolduc, J. M.; Dyer, D. H.; Scott, W. G.; Singer, P.; Sweet, R. M.; Koshland, D. E.; Stoddard, B. L. Mutagenesis and Laue Structures of Enzyme Intermediates - Isocitrate Dehydrogenase. *Science* **1995**, *268*, 1312-1318.

(458) Aktas, D. F.; Cook, P. F. A Lysine-Tyrosine Pair Carries Out Acid-Base Chemistry in the Metal Ion-Dependent Pyridine Dinucleotide-Linked beta-Hydroxyacid Oxidative Decarboxylases. *Biochemistry* **2009**, *48*, 3565-3577.

(459) Grissom, C. B.; Cleland, W. W. Isotope Effect Studies of the Chemical Mechanism of Pig-Heart NADP Isocitrate Dehydrogenase. *Biochemistry* **1988**, *27*, 2934-2943.

(460) Rao, G. S. J.; Coleman, D. E.; Karsten, W. E.; Cook, P. F.; Harris, B. G. Crystallographic studies on Ascaris suum NAD-malic enzyme bound to reduced cofactor and identification of an effector site. *J. Biol. Chem.* **2003**, *278*, 38051-38058.

(461) Huang, Y. C.; Grodsky, N. B.; Kim, T. K.; Colman, R. F. Ligands of the Mn^{2+} bound to porcine mitochondrial NADP-dependent isocitrate dehydrogenase, as assessed by mutagenesis. *Biochemistry* **2004**, *43*, 2821-2828.

(462) Hurley, J. H.; Dean, A. M.; Sohl, J. L.; Koshland, D. E.; Stroud, R. M. Regulation of an Enzyme by Phosphorylation at the Active-Site. *Science* **1990**, *249*, 1012-1016.

(463) Hurley, J. H.; Dean, A. M.; Koshland, D. E.; Stroud, R. M. Catalytic Mechanism of $NADP^+$-Dependent Isocitrate Dehydrogenase - Implications from the Structures of Magnesium Isocitrate and $NADP^+$ Complexes. *Biochemistry* **1991**, *30*, 8671-8678.

(464) Leiros, H. K. S.; Fedoy, A. E.; Leiros, I.; Steen, I. H. The complex structures of isocitrate dehydrogenase from Clostridium thermocellum and Desulfotalea psychrophila suggest a new active site locking mechanism. *FEBS Open Bio* **2012**, *2*, 159-172.

(465) Vinekar, R.; Verma, C.; Ghosh, I. Functional relevance of dynamic properties of Dimeric NADP-dependent Isocitrate Dehydrogenases. *BMC Bioinformatics* **2012**, *13:S2*.

(466) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774-797.

(467) Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E. Orbital steering in the catalytic power of enzymes: Small structural changes with large catalytic consequences. *Science* **1997**, *277*, 202-206.

(468) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*, University of California: San Francisco, 2012.

(469) Sorin, E. J.; Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **2005**, *88*, 2472-2493.

(470) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712-25.

(471) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950-1958.

(472) Holmberg, N.; Ryde, U.; Bulow, L. Redesign of the coenzyme specificity in L-lactate dehydrogenase from Bacillus stearothermophilus using site-directed mutagenesis and media engineering. *Protein Eng.* **1999**, *12*, 851-856.

(473) Shao, J. Y.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theor. Comp.* **2007**, *3*, 2312-2334.

(474) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. 23. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654-3665.

(475) Brás, N. F.; Fernandes, P. A.; Ramos, M. J. QM/MM Study and MD Simulations on the Hypertension Regulator Angiotensin-Converting Enzyme. *ACS Catalysis* **2014**, *4*, 2587-2597.

(476) Calixto, A. R.; Bras, N. F.; Fernandes, P. A.; Ramos, M. J. Reaction Mechanism of Human Renin Studied by Quantum Mechanics/Molecular Mechanics (QM/MM) Calculations. *ACS Catalysis* **2014**, *4*, 3869-3876.

(477) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265-3269.

(478) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge-Densities. *Theor. Chim. Acta* **1977**, *44*, 129-138.

(479) Ritchie, J. P. Electron-Density Distribution Analysis for Nitromethane, Nitromethide, and Nitramide. *J. Am. Chem. Soc.* **1985**, *107*, 1829-1837.

(480) Ritchie, J. P.; Bachrach, S. M. Some Methods and Applications of Electron-Density Distribution Analysis. *J. Comput. Chem.* **1987**, *8*, 499-509.

(481) Dean, A. M.; Koshland, D. E. Kinetic Mechanism of Escherichia-Coli Isocitrate Dehydrogenase. *Biochemistry* **1993**, *32*, 9302-9309.

(482) Karlstrom, M.; Stokke, R.; Steen, I. H.; Birkeland, N. K.; Ladenstein, R. Isocitrate dehydrogenase from the hyperthermophile Aeropyrum pernix: X-ray structure analysis of a ternary enzyme-substrate complex and thermal stability. *J. Mol. Biol.* **2005**, *345*, 559-577.

(483) Wiklund, L. Carbon Dioxide Formation and Elimination in Man. *Ups. J. Med. Sci.* **1996**, *101*, 35-68.

(484) Hopkinson, B. M.; Dupont, C. L.; Allen, A. E.; Morel, F. M. M. Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3830-3837.

(485) Narayan, M.; Welker, E.; Wedemeyer, W. J.; Scheraga, H. A. Oxidative folding of proteins. *Acc. Chem. Res.* **2000**, *33*, 805-812.

(486) Kersteen, E. A.; Barrows, S. R.; Raines, R. T. Catalysis of protein disulfide bond isomerization in a homogeneous substrate. *Biochemistry* **2005**, *44*, 12168-12178.

(487) Malhotra, J. D.; Kaufman, R. J. The endoplasmic reticulum and the unfolded protein response. *Semin. Cell Dev. Biol.* **2007**, *18*, 716-731.

(488) Galligan, J. J.; Petersen, D. R. The human protein disulfide isomerase gene family. *Hum. Genomics* **2012**, *6*.

(489) Benham, A. M. The Protein Disulfide Isomerase Family: Key Players in Health and Disease. *Antioxid. Redox Sign.* **2012**, *16*, 781-789.

(490) Walker, K. W.; Gilbert, H. F. Effect of Redox Environment on the in-Vitro and in-Vivo Folding of Rtem-1 Beta-Lactamase and Escherichia Coli Alkaline-Phosphatase. *J. Biol. Chem.* **1994**, *269*, 28487-28493.

(491) Tu, B. P.; Weissman, J. S. Oxidative protein folding in eukaryotes: mechanisms and consequences. *J. Cell Biol.* **2004**, *164*, 341-346.

(492) Schroder, M.; Kaufman, R. J. The mammalian unfolded protein response. *Annu. Rev. Biochem.* **2005**, *74*, 739-789.

(493) Hotamisligil, G. S. Endoplasmic Reticulum Stress and the Inflammatory Basis of Metabolic Disease. *Cell* **2010**, *140*, 900-917.

(494) Darby, N. J.; Creighton, T. E. Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. *Biochemistry* **1995**, *34*, 16770-16780.

(495) Lappi, A. K.; Ruddock, L. W. Reexamination of the Role of Interplay between Glutathione and Protein Disulfide Isomerase. *J. Mol. Biol.* **2011**, *409*, 238-249.

(496) Ferrari, D. M.; Soling, H. D. The protein disulphide-isomerase family: unravelling a string of folds. *Biochem. J.* **1999**, *339*, 1-10.

(497) Tian, G.; Xiang, S.; Noiva, R.; Lennarz, W. J.; Schindelin, H. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **2006**, *124*, 61-73.

(498) Wang, C.; Li, W.; Ren, J. Q.; Fang, J. Q.; Ke, H. M.; Gong, W. M.; Feng, W.; Wang, C. C. Structural Insights into the Redox-Regulated Dynamic Conformations of Human Protein Disulfide Isomerase. *Antioxid. Redox Sign.* **2013**, *19*, 44-53.

(499) Lyles, M. M.; Gilbert, H. F. Mutations in the Thioredoxin Sites of Protein Disulfide-Isomerase Reveal Functional Nonequivalence of the N-Terminal and C-Terminal Domains. *J. Biol. Chem.* **1994**, *269*, 30946-30952.

(500) Darby, N. J.; Penka, E.; Vincentelli, R. The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *J. Mol. Biol.* **1998**, *276*, 239-247.

(501) Xu, S. L.; Sankar, S.; Neamati, N. Protein disulfide isomerase: a promising target for cancer therapy. *Drug Discov. Today* **2014**, *19*, 222-240.

(502) Klappa, P.; Hawkins, H. C.; Freedman, R. B. Interactions between protein disulphide isomerase and peptides. *Eur. J. Biochem.* **1997**, *248*, 37-42.

(503) Denisov, A. Y.; Maattanen, P.; Dabrowski, C.; Kozlov, G.; Thomas, D. Y.; Gehring, K. Solution structure of the bb' domains of human protein disulfide isomerase. *FEBS J.* **2009**, *276*, 1440-1449.

(504) Byrne, L. J.; Sidhu, A.; Wallis, A. K.; Ruddock, L. W.; Freedman, R. B.; Howard, M. J.; Williamson, R. A. Mapping of the ligand-binding site on the b ' domain of human PDI: interaction with peptide ligands and the x-linker region. *Biochem. J.* **2009**, *423*, 209-217.

(505) Wang, C.; Yu, J.; Huo, L.; Wang, L.; Feng, W.; Wang, C. C. Human Protein-disulfide Isomerase Is a Redox-regulated Chaperone Activated by Oxidation of Domain a '. *J. Biol. Chem.* **2012**, *287*, 1139-1149.

(506) Irvine, A. G.; Wallis, A. K.; Sanghera, N.; Rowe, M. L.; Ruddock, L. W.; Howard, M. J.; Williamson, R. A.; Blindauer, C. A.; Freedman, R. B. Protein Disulfide-Isomerase Interacts with a Substrate Protein at All Stages along Its Folding Pathway. *Plos One* **2014**, *9*.

(507) Hatahet, F.; Ruddock, L. W. Protein Disulfide Isomerase: A Critical Evaluation of Its Function in Disulfide Bond Formation. *Antioxid. Redox Sign.* **2009**, *11*, 2807-2850.

(508) Ali Khan, H.; Mutus, B. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. *Front. Chem.* **2014**, *2*.

(509) Raturi, A.; Mutus, B. Characterization of redox state and reductase activity of protein disulfide isomerase under different redox environments using a sensitive fluorescent assay. *Free Radical Bio. Med.* **2007**, *43*, 62-70.

(510) Cuozzo, J. W.; Kaiser, C. A. Competition between glutathione and protein thiols for disulphide-bond formation. *Nat. Cell. Biol.* **1999**, *1*, 130-135.

(511) Baker, K. M.; Chakravarthi, S.; Langton, K. P.; Sheppard, A. M.; Lu, H.; Bulleid, N. J. Low reduction potential of Ero1 alpha regulatory disulphides ensures tight control of substrate oxidation. *Embo J.* **2008**, *27*, 2988-2997.

(512) Chambers, J. E.; Tavender, T. J.; Oka, O. B. V.; Warwood, S.; Knight, D.; Bulleid, N. J. The Reduction Potential of the Active Site Disulfides of Human Protein Disulfide Isomerase Limits Oxidation of the Enzyme by Ero1 alpha. *J. Biol. Chem.* **2010**, *285*, 29200-29207.

(513) Wang, L.; Li, S. J.; Sidhu, A.; Zhu, L.; Liang, Y.; Freedman, R. B.; Wang, C. C. Reconstitution of human Ero1-L alpha/protein-disulfide isomerase oxidative folding pathway in vitro. Position-dependent differences in role between the a and a ' domains of protein-disulfide isomerase. *J. Biol. Chem.* **2009**, *284*, 199-206.

(514) Kemmink, J.; Darby, N. J.; Dijkstra, K.; Nilges, M.; Creighton, T. E. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear C-13/N-15 NMR spectroscopy. *Biochemistry* **1996**, *35*, 7684-7691.

(515) Kemmink, J.; Dijkstra, K.; Mariani, M.; Scheek, R. M.; Penka, E.; Nilges, M.; Darby, N. J. The structure in solution of the b domain of protein disulfide isomerase. *J. Biomol. NMR* **1999**, *13*, 357-368.

(516) Dijkstra, K.; Karvonen, P.; Pirneskoski, A.; Koivunen, P.; Kivirikko, K. I.; Darby, N. J.; van Straaten, M.; Scheek, R. M.; Kemmink, J. Assignment of H-1, C-13 and N-15 resonances of the a ' domain of protein disulfide isomerase. *J. Biomol. NMR* **1999**, *14*, 195-196.

(517) Lappi, A. K.; Lensink, M. F.; Alanen, H. I.; Salo, K. E. H.; Lobell, M.; Juffer, A. H.; Ruddock, L. W. A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. *J. Mol. Biol.* **2004**, *335*, 283-295.

(518) Walker, K. W.; Gilbert, H. F. Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* **1997**, *272*, 8845-8848.

(519) Schwaller, M.; Wilkinson, B.; Gilbert, H. F. Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase. *J. Biol. Chem.* **2003**, *278*, 7154-7159.

(520) Dyson, H. J.; Jeng, M. F.; Tennant, L. L.; Slaby, I.; Lindell, M.; Cui, D. S.; Kuprin, S.; Holmgren, A. Effects of buried charged groups on cysteine thiol ionization and reactivity in Escherichia coli thioredoxin: Structural and functional characterization of mutants of Asp 26 and Lys 57. *Biochemistry* **1997**, *36*, 2622-2636.

(521) Ellgaard, L.; Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *Embo Reports* **2005**, *6*, 28-32.

(522) Gruber, C. W.; Cemazar, M.; Heras, B.; Martin, J. L.; Craik, D. J. Protein disulfide isomerase: the structure of oxidative folding. *Trends Biochem. Sci.* **2006**, *31*, 455-464.

(523) Watanabe, M. M.; Laurindo, F. R. M.; Fernandes, D. C. Methods of measuring Protein Disulfide Isomerase activity: a critical overview. *Front. Chem.* **2014**, *2*.

(524) Westphal, V.; Darby, N. J.; Winther, J. R. Functional properties of the two redox-active sites in yeast protein disulphide isomerase in vitro and in vivo. *J. Mol. Biol.* **1999**, *286*, 1229-1239.

(525) Wilkinson, B.; Gilbert, H. F. Protein disulfide isomerase. *BBA-Proteins Proteom.* **2004**, *1699*, 35-44.

(526) Westphal, V.; Spetzler, J. C.; Meldal, M.; Christensen, U.; Winther, J. R. Kinetic analysis of the mechanism and specificity of protein-disulfide isomerase using fluorescence-quenched peptides. *J. Biol. Chem.* **1998**, *273*, 24992-24999.

(527) Darby, N. J.; Freedman, R. B.; Creighton, T. E. Dissecting the Mechanism of Protein Disulfide-Isomerase - Catalysis of Disulfide Bond Formation in a Model Peptide. *Biochemistry* **1994**, *33*, 7937-7947.

(528) Ruoppolo, M.; Freedman, R. B.; Pucci, P.; Marino, G. Glutathione-dependent pathways of refolding of RNase T-1 by oxidation and disulfide isomerization: Catalysis by protein disulfide isomerase. *Biochemistry* **1996**, *35*, 13636-13646.

(529) Bass, R.; Ruddock, L. W.; Klappa, P.; Freedman, R. B. A major fraction of endoplasmic reticulum-located glutathione is present as mixed disulfides with protein. *J. Biol. Chem.* **2004**, *279*, 5257-5262.

(530) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 13*, University of California: San Francisco, 2012.

(531) Lee, E. H.; Kim, H. Y.; Hwang, K. Y. The GSH- and GSSG-bound structures of glutaredoxin from Clostridium oremlandii. *Arch. Biochem. Biophys.* **2014**, *564*, 20-25.

(532) Kozlov, G.; Maattanen, P.; Thomas, D. Y.; Gehring, K. A structural overview of the PDI family of proteins. *FEBS J.* **2010**, *277*, 3924-3936.

(533) Darby, N. J.; Creighton, T. E. Functional-Properties of the Individual Thioredoxin-Like Domains of Protein Disulfide-Isomerase. *Biochemistry* **1995**, *34*, 11725-11735.

(534) Laurindo, F. R. M.; Pescatore, L. A.; Fernandes, D. D. Protein disulfide isomerase in redox cell signaling and homeostasis. *Free Radical Bio. Med.* **2012**, *52*, 1954-1969.

(535) Appenzeller-Herzog, C.; Ellgaard, L. In vivo reduction-oxidation state of protein disulfide isomerase: The two active sites independently occur in the reduced and oxidized forms. *Antioxid. Redox Sign.* **2008**, *10*, 55-64.