Epigenetic Mechanisms *in silico*: Understanding Demethylation and Rational Design of Bromodomain Inhibitors

Thesis submitted for the degree of

Doctor of Philosophy

Department of Chemistry

February 2017

Wilian Augusto Cortopassi Coelho

Linacre College

University of Oxford

Contents

Abstract

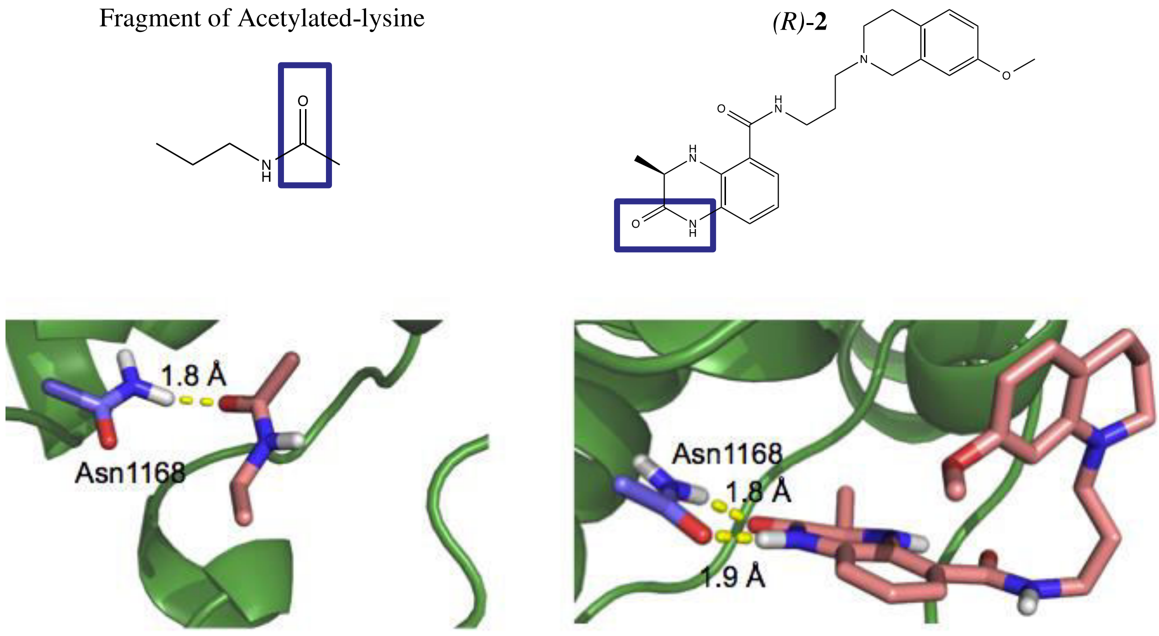
"The map is not the territory."

Alfred Korzybski

Histone octamer proteins are crucial for DNA packaging and storage in the confined space of the cell nucleus. Epigenetic post-translational modifications to specific histone monomers facilitate changes in chromatin flexibility that are necessary for access by transcription machinery, and therefore have control over gene expression. Targeting enzymes that regulate these powerful modifications has been established as a promising strategy in the treatments of some diseases, such as cancer, male infertility and adult obesity. The work presented in this thesis aims to shed light on the mode of action of epigenetic proteins, and in so doing, aid in the design of more potent and selective small molecules that target them (*e.g.* Jumonji C (JmjC) and bromodomain containing-proteins). We have used a diverse array of computational techniques such as quantum mechanics (QM), classical Molecular Dynamics (MD) simulations, and hybrid Newtonian molecular mechanics/quantum mechanics (QM/MM) approaches. Following an introduction to these computational chemistry techniques (**Chapter 1**) and a description of the methodology used in this thesis (**Chapter 2**), we then present our studies on: (a) improving Bromodomain-containing proteins inhibition (**Chapters 3**, **4**) and (b) providing novel insights into the lysine demethylation catalysed by JmjC proteins (**Chapters 5**, **6**).

In **Chapter 3**, we have investigated a series of dihydroquinoxalinone (DHQ) derivatives, analogues of *(R)*-**2** **(Figure A.1)**, previously proven inactivators of the epigenetic molecular target CREBBP bromodomain, resulting in the first potent

inhibitors of a bromodomain outside the Bromomodomain and extra-terminal (BET) family. Through QM and MD calculations, I established the significance of cation-π interactions for the binding of CREBBP inhibitors. In **Chapter 4**, we then developed a novel electrostatic model for the quantification of cation-π interactions in biological environments, inspired by a good fit between experimental binding affinity data of a series of fifteen 5-isoxazolylbenzimidazole derivatives and electrostatic potential (ESP) surfaces. This model has been prospectively applied to newly synthesised DHQ derivatives. Our approach could be used for the development of force fields and docking scoring functions with increased reliabilities to reproduce cation-π interactions.


Fragment of Acetylated-lysine

(R)-**2**

Figure A.1. *(R)*-**2** mimicking Acetylated-lysine by H-bonding an asparagine.

Together with the understanding of DHQ binding in CREBBP, I have investigated one of the most common but poorly understood epigenetic processes: histone lysine demethylation by a JmjC protein, JMJD2A (**Figure A.2**). In **Chapter**

5, in the first QM/MM studies reported for this system, we describe the importance of the protein environment on the binding of molecular O_2, while in **Chapter 6** we analyse the PES of the elementary steps associated with enzymatic demethylation: co-factor/O_2 activation, C-H abstraction and hydroxyl-rebound. Insights into this reaction mechanism and energetic contributions have been used to the role of specific JMJD2A residues in this process, which can be used to accelerate the design of effective drug molecules.

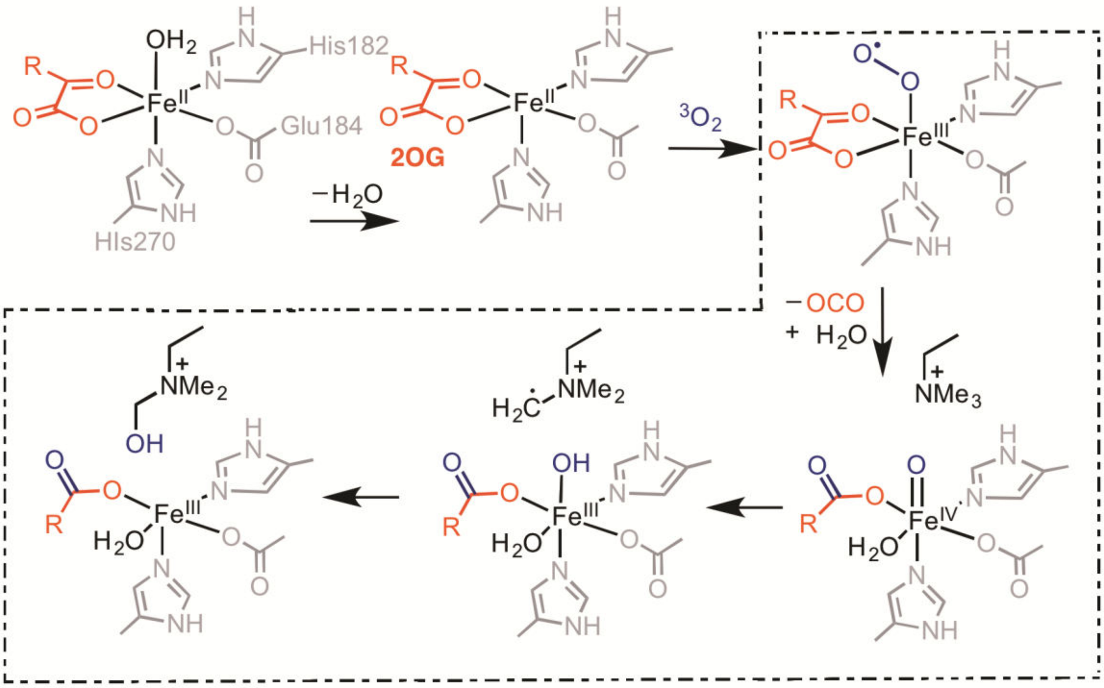


Figure A.2. Computed mechanism for lysine demethylation via a Fe^{IV}–oxo complex in JMJD2A. The O_2 binding happens preferentially on the quintet PES, with O_2 in a so-called “end-on” conformation and oxidation of Fe^{II} to Fe^{III}. The co-factor/O_2 activation, C-H abstraction and hydroxyl-rebound were also computed on the quintet PES, with the hydroxyl-rebound step shown here to be rate-determining.

Acknowledgements

"You have to do it by yourself, and you can't do it alone."

Martin Rutte

I would like to thank Prof. Robert Paton for his guidance during my DPhil, not only for his academic support, but also for our constant discussions about career. Under his supervision, I have been in touch with various different fields of research, from programming and physical chemistry to organic chemistry and biology. He has also motivated me to participate in important chemistry-related conferences and I am truly thankful for his patience and dedication towards my DPhil. Prof. Stuart Conway and Michael Brand have also guided me towards the application of my computational models to biological data, and I really appreciate their interesting insights during my DPhil.

I could always count on the support of my colleagues from Prof. Paton's group. I would particularly like to mention Chornsby, Sami Jaffar, Kelvin Jackson, Robert Simion (EnzBond), Fernanda Duarte and Kiran Kumar, who have been around for a pint at University Club, enjoyable conversations and, of course, interesting scientific discussions.

Uncountable were the times I had difficulties with initially setting up my calculations, and then I emailed the support staff for two pieces of software I have constantly used in my thesis: Gaussian and Amber. I really appreciate the helpful answers from Dr. Fernando Clemente (Gaussian) and Jason Swails (Amber). Along the same line, I would like to thank Prof. Hajime Hirao and Prof. Francois Vanquelef

(RED server) for kindly answering my questions related to problems in running QM/MM and MD simulations, respectively.

My previous supervisors from Brazil, Prof. Tanos Franca, Prof. Andre Pimentel and Prof. Antoniana Krettli, have been in touch with me since the very first day I arrived in Oxford, offering constant advice and highly valuable scientific background knowledge. Dr. Daniel Branco, Ana Luiza, Andre Street, Rodrigo Anunciato and Eduardo Pontes have been outstanding mentors even before I came to the UK, and also provided immeasurable guidance by showing me different ways of achieving my life goals.

The author of this thesis also acknowledges the funding agency CAPES under the *Science Without Borders* scheme for supporting this research project, as well as the use of the EPSRC UK National Service for Computational Chemistry Software (CHEM870), the University of Oxford Advanced Research Computing (ARC) facility (10.5281/zenodo.22558) and the dirac cluster (EP/L015722/1) in carrying out this work.

A special acknowledgment goes to my family, to my two sisters and my mother for getting worried when I did not call them back at the weekend, for hearing news every week about my life in the UK and for always welcoming me back home with lots of Christmas gifts. I am lucky to have had such great support from all of them.

Glossary of abbreviation and acronyms

Å	angstrom(s), 10^{-10} m
Ac	acetyl, $COCH_3$
ALY	acetylated lysine
ASM	active site model
Atm	atmospheres (1 atm = 1.01 x 10^5 Pascal)
au	atomic units
aug	augmented
B3LYP	hybrid density functional incorporating Becke exchange with Lee-Yang-Parr correlation
cc	correlation consistent
CCSD(T)	coupled cluster with perturbative estimate for connected triples
CPCM	conductor-like polarizable continuum model
CREBBP	cAMP-response element binding protein binding protein
DHQ	dihydroquinoxalinone
DFT	density functional theory
ΔT_m	melting temperature shifts
double-ζ	double zeta
ECP	effective core potential
ESP	electrostatic potential surface
FAD	flavin adenine dinucleotide
GAFF	general Amber force field
GGA	generalised gradient approximation
GTO	Gaussian-type orbital
HF	Hartree-Fock

IC_{50}	half maximal inhibitory concentration
IRC	intrinsic reaction coordinate
ISC	intersystem crossing
JMJD2A	a member of the Jumonji C family of proteins
K	Kelvin
kcal mol^{-1}	kilocalorie per mole
JmjC	Jumonji C
KIE	kinetic isotope effect
LANL	Los Alamos National Laboratory
LDA	local density approximation
M06-2X	Minnesota 06 density functional of Truhlar
M-Lysine	methylated lysine
MM	molecular mechanics
MM-PBSA	MM-Poisson Boltzman surface area
MM-GBSA	MM-generalized Born surface area
Me	methyl group, CH_3
MEP	molecular electrostatic potential
MP2	Møller-Plesset second order perturbation theory
M3Y	trimethylated lysine
NAD^+	(nicotinamide adenine dinucleotide)$^+$
ONIOM	our own N-layered integrated molecular orbital and MM
OMe	methoxy group, $-OCH_3$
PCM	polarizable continuum model
PDB	protein data bank
PES	potential energy surface

Ph	phenyl group, C_6H_5
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
RESP	restrained electrostatic potential
r_s	Spearman coefficient
R^2	coefficient of determination
RMSD	root mean squared deviation
RMSE	root mean square error
RMSF	root mean square fluctuation
SCRF	self-consistent reaction field polarizable continuum model
STO	Slater-type orbital
single-ζ	single zeta
tauD	taurine dioxygenase
TS	transition state (first order saddle point)
TPSS	the exchange functional of Tao, Perdew, Staroverov, and Scuseria
triple-ζ	triple zeta
uB3LYP	unrestricted B3LYP
UFF	Universal Force Field
ωB97XD	dispersion corrected density functional of Head-Gordon
ZPE	Zero-point (vibrational) energy

Author’s declaration

This thesis is an account of the work carried out by the author in the Department of Chemistry, University of Oxford, under the supervision of Professor Robert S. Paton. No part of this thesis has previously been submitted for a degree in this University or elsewhere. The work of other authors has been freely drawn and is duly acknowledged in the text. A list of references is given at the end of each chapter.

Publications, conferences and awards

Publications describing work in this thesis

Citations from *SciFinder*® on 21/12/16.

[1] "*Cation-π interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity*" Cortopassi, W.A.; Kiran, K.; Paton, R.S. *Org. Bio. Chem.* **2016**. *14*, 10926-10938. *Citations*: 0.

[2] "*Mechanisms of histone-modifying and reading enzymes: a computational perspective on the role of the protein environment*" Cortopassi, W.A.; Kiran, K.; Duarte, F.; Paton, R.S. *J. Mol. Graph. Model.* **2016**. *67, 69-84. Citations: 2.*

[3] "*Dioxygen binding in the active site of histone demethylase JMJD2A and the role of protein environment*" Cortopassi, W. A.; Simion, R.; Hornsby, C. E.; Costa Franca, T. C.; Paton, R. S. *Chem. Eur. J.* **2015**, *21*, 18983-18992 (rated Hot Paper in Chemical Biology and selected for Journal front cover). *Citations: 3.*

[4] "*Small molecule inhibitors of bromodomain-acetyl-lysine interactions" Brand,* M.; Measures, A.; Wilson, B.; Cortopassi, W. A.; Alexander, R.; Hoss, M.; Hewings, D. S.; Paton, R. S.*;* Conway, S. J. *ACS Chem. Bio.* **2015**, *10*, 22-39. *Citations: 35.*

[5] "*A series of potent CREBBP bromodomain ligands reveals an induced fit pocket stabilized by a cation-π interaction"* Rooney, T. P. C.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin,S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Müller-Knapp, S.; Knapp, S.; Brennan, P. E.; Conway, S. J. *Angew. Chem. Int. Ed.* **2014***, 126, 26,* 6240-6244. *Citations: 33.*

Awards and Honours

2016: Selected for the *Future Leaders in Chemistry* Program, SciFinder, Columbus/Pennsylvania.

2016: *Scholarship for Scientific Excellence* from the Chemical Information Division of the American Chemical Society at the San Diego ACS meeting.

2016: Travel Grant Scholarship from the Royal Society of Chemistry.

2015: Travel Grant Scholarship from the German Chemical Society.

2015: Poster prize at the UK's National Service for Computational Chemistry Software (NSCCS) meeting at Imperial College, London.

Conference Presentations

---

[6] Poster: "Prediction and quantification of cation-π interactions in ligand-bromodomain binding: Using quantum chemistry to capture electronic effects" *251st American Chemical Society National Meeting & Exposition,* **2016**, San Diego.

[5] Oral: "Role of the protein environment for the O_2 binding in the active site of histone demethylase JMJD2A" *251st American Chemical Society National Meeting & Exposition,* **2016**, San Diego, US.

[4] Oral: "The role of cation-π interactions for ligand-bromodomain binding" *Young Modellers' Forum*, **2015**, Greenwich, UK.

[3] Oral: "Cation-π Interactions in Ligand-Bromodomain Binding: Quantification and Predictions with Quantum Chemistry" *10th European Conference on Computational Chemistry*, **2015**, Fulda, Germany.

[2] Poster: "Cation-π interactions play a key role in the binding of new CREBBP bromodomain inhibitors*" National Service for Computational Chemistry Software Meeting*, **2015**, London, UK.

[1] Oral: "Cation-π interactions play a key role in the binding of new CREBBP bromodomain inhibitors" *7th Brazilian Symposium on Medicinal Chemistry*, **2014**, Campos do Jordão, Brazil.

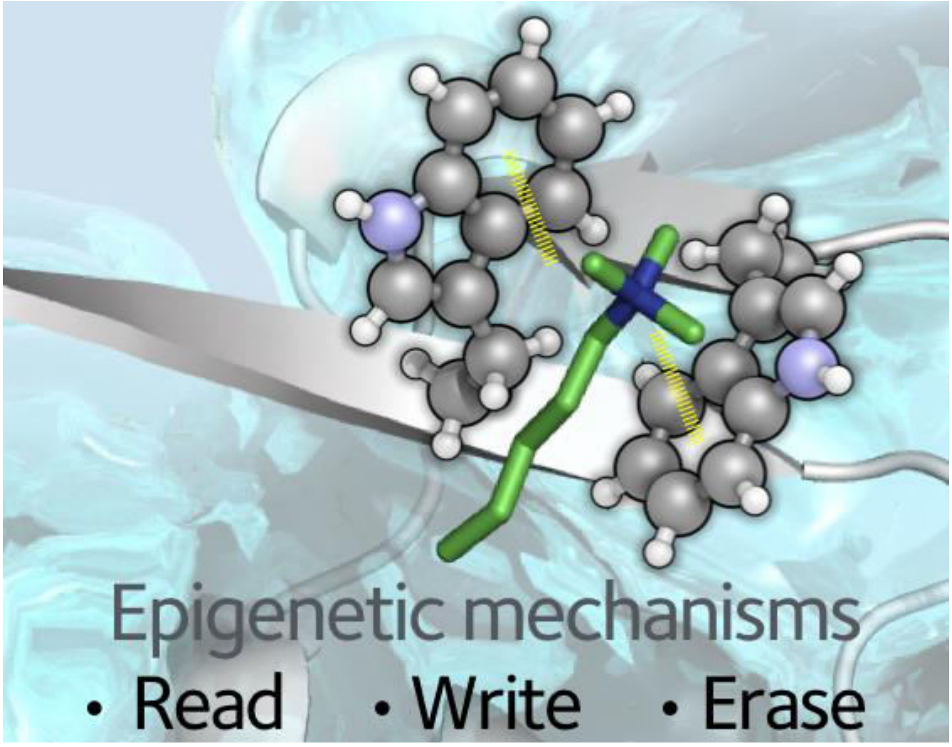
Public Awareness in Science

Cancer Facts, 2015 – YouTube bilingual channel that consists of short animated videos in simple language explaining what cancer is and how it can be treated. https://goo.gl/lO8Uau

Chapter 1

Introduction

"Man is born free, and everywhere he is in chains."
Jean-Jacques Rousseau



Cortopassi WA, Kumar K, Duarte F, Pimentel AS, Paton RS. Mechanisms of histone lysine-modifying enzymes: A computational perspective on the role of the protein environment, *J Mol Graph Mod* **2016**, 67, 69–84.

1.1 Overview

Chromatin remodeling and histone modifications are epigenetic changes that regulate several biological processes, such as DNA replication and repair. The cell nucleus contains chromatin, a combination of DNA and proteins.[1] Histones are proteins responsible for packaging DNA into the confined space of the nucleus, giving strength during mitosis and controlling gene expression.[1] Targeting enzymes that regulate histone modifications is a promising strategy for treatments of diseases related to cell function, such as cancer,[2] male infertility and adult obesity,[3] among others. Therefore, understanding the chemical mechanisms of these modifications at the molecular level is a key area of research, and a combination of computational and experimental work has the potential to deliver important mechanistic insights.

In this thesis, we will focus on epigenetic modifications involving proteins that add (writers), remove (erasers) and recognize (readers) chemical modifications in lysines of histones. More specifically, we will be studying the removal of methyl groups from lysine residues and the recognition of acetyl-lysine groups in modified histones. Most of these changes take place at histone tails, but occasionally in their globular domains. While DNA methylation is probably the most well-known epigenetic modification,[4] other epigenetic modifications have been shown to be equally important.

For a better understanding of the context in which these modifications are included, the subsequent sections will provide an overview of the importance of investigating epigenetic mechanisms for drug discovery.

1.2 Lysine histone modifications – Structural and mechanistic aspects

Interest in developing new medicines targeting epigenetic proteins has increased in recent years, in particular after the successful results obtained with four already approved drugs by the Food and Drug Administration (FDA) regulatory agency (**Figure 1.1**). Among those, 5-azacytidine and decibitabine, two cytosine analogues that block epigenetic writers (DNA methyltransferases), have been approved for treating myelodysplastic syndromes.[5] Romidepsin and vorinostat, inhibitors of epigenetic erasers (histone deacetylases), are clinically used against cutaneous T-cell lymphoma.[6] Compounds targeting the epigenetic reading process are yet to be approved, but some in the early-stage of clinical trials, including OTX015, have shown promising results against haematological malignancies.[7]

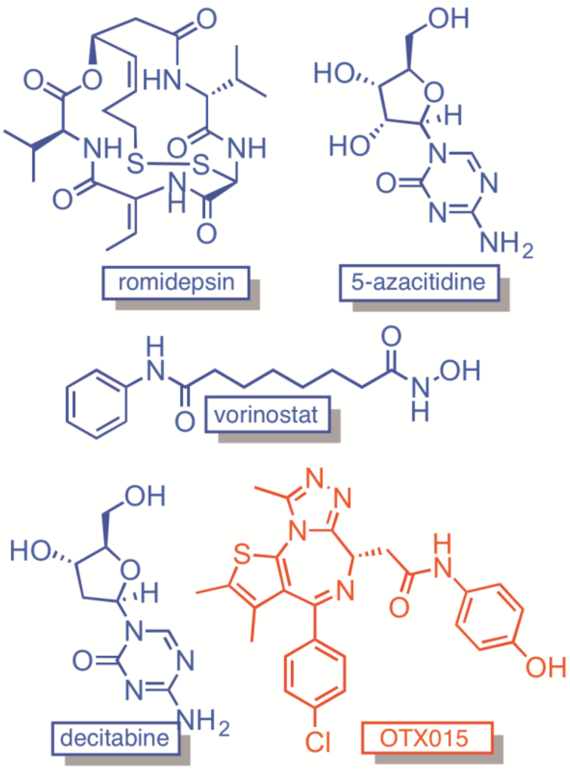


Figure 1.1. FDA-approved drugs (in blue) targeting epigenetic proteins: romidepsin, 5-azacytidine, vorinostat and decibitabine. The bromodomain inhibitor OTX015 (in red) is at an early stage of clinical trials.

In this section, we will introduce the histone writing, erasing and reading regulators, in terms of their structural and mechanistic features. We will also briefly comment on the success of small molecule inhibitors currently under clinical trials.

Histone writing regulators, such as histone lysine acetyltransferases (HAT) and methyltransferases (HKMT), add acetyl and methyl groups, respectively, to the ε-amino group of a lysine of the histone.

HATs are usually classified depending mostly on their structures, but also on their sequence and functional similarities:[8] (i) the GCN5-related N-terminal acetyltransferases (GNAT); (ii) the Moz, Ybf2/Sas3, Sas2, Tip60 family (MYST) and (iii) the p300/CREB-binding protein (CBP/CREBBP) families. Each of these families shares a common mechanism, changing the exact sequence of substrate- and cofactor-binding (**Figure 1.2**).

GNAT proteins contain approximately 160 residues, including both a bromodomain, which is responsible for the recognition/reading process of acetylated lysines, as well as a HAT domain. GNAT proteins follow a sequential mechanism with the formation of an Ac-CoA/HAT/substrate ternary complex. This complex is formed upon the nucleophilic attack of the activated lysine at the carbonyl group of Ac-CoA.

In the MYST family, proteins have a HAT domain containing approximately 250 residues, an N-terminal chromodomain, responsible for the recognition/reading process of methylated lysines, as well as a zinc finger. Proteins from the MYST family could also undergo the sequential mechanism, but some studies report an alternative consisting of the transfer of the acetyl group firstly to a cysteine residue, and then to the substrate – the so-called “ping pong” mechanism.

Finally, CBP proteins have a HAT domain containing approximately 500 residues, a bromodomain and a zinc finger.[9] An alternative mechanism to the previously described for MYST proteins is observed in CBP proteins, where the ternary complex is not stable, and instantly releases the acetylated histone – in the so-called “hit and run” mechanism.[9]

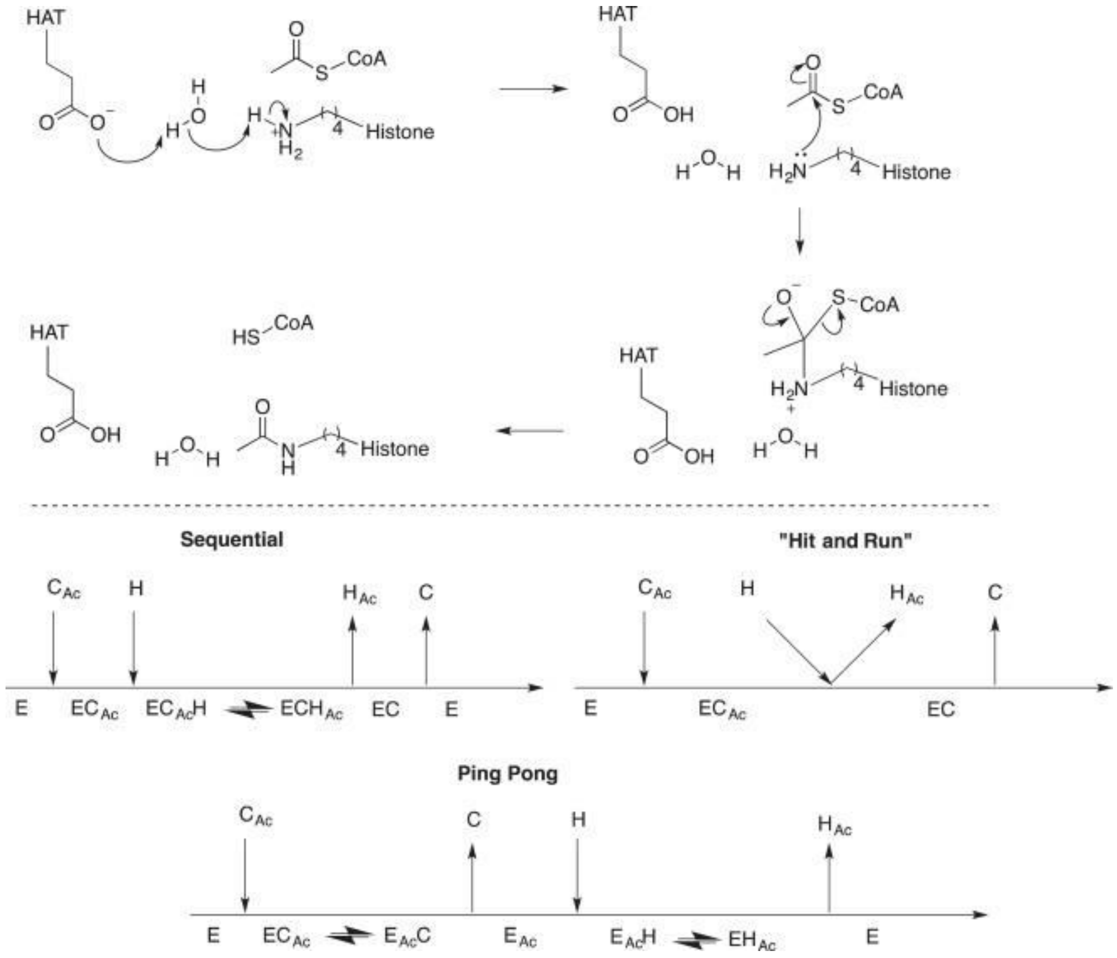


Figure 1.2. Top: Sequential mechanism for histone lysine acetylation in GNAT proteins. Bottom: suggested kinetics for HATs. For the sequential mechanism, the acetylated cofactor (C_{Ac}) and the histone (H) bind to the active site of the enzyme (E) forming a ternary complex ($EC_{Ac}H$). Following the acetyl group (Ac) transfer to the lysine, the acetylated histone (H_{Ac}) and cofactor (C) are then released. In the hit and run mechanism, histone acetylation occurs as a bimolecular step without forming a stable ternary complex. In the ping pong mechanism, the acetyl group is transferred to the enzyme ($E_{Ac}C$), followed by the release of the cofactor (C). After this, histone binding (forming $E_{Ac}H$) and acetylation (forming EH_{Ac}) occur.[10]

HKMT proteins are cofactor-dependent. They use S-adenosyl-L-methionine (AdoMet) for the donation of a methyl group to the lysine of the histones (**Figure 1.3**). Classification of the HKMTs relies on the presence or absence of the Su(var)3-9 Enhancer of Zeste Trithorax (SET) domain. Lysine methyltransferases containing the SET domain are classified as SET HKMTs, while proteins from the DOT1-Like Histone H3K79 (DOT1L) family do not contain this domain. For SET HKMTs, the proposed mechanism begins with the AdoMet binding, after which a tryptophan residue deprotonates the histone lysine. This is then followed by nucleophilic attack by the neutral N-atom resulting in transfer of the methyl group from AdoMet to the lysine residue. Proteins belonging to the DOT1L family may follow a similar mechanism, although currently it is not fully understood how the lysine deprotonation happens in these proteins.

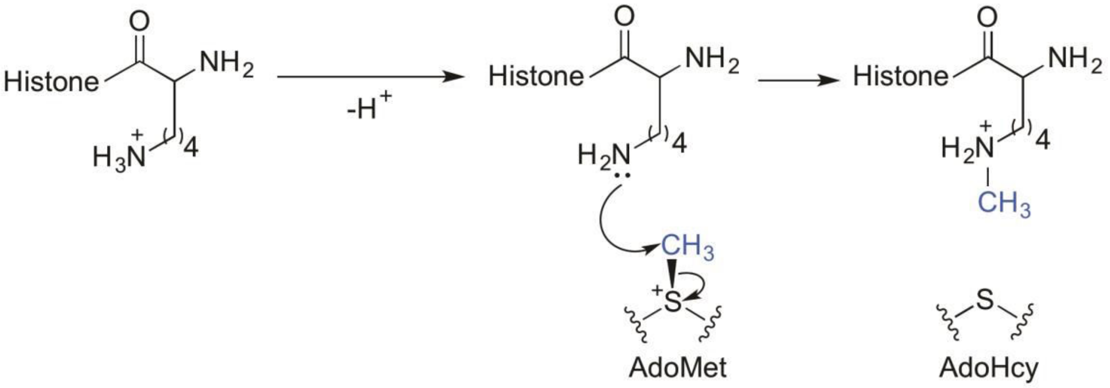


Figure 1.3. Mechanism of methylation by histone lysine methyltransferases.

These families of writing regulators discussed so far incorporate a considerable number of druggable proteins, and some compounds are currently undergoing clinical trials for a large number of applications, mainly related to cancer. Curcumin, a natural phenol present in turmeric, is an HAT inhibitor currently in testing as a treatment against cancer, inflammation, cardiovascular and metabolic diseases, although the origin of its efficacy is not completely clear.[11] Pinometostat, a HKMT inhibitor, is another example of a potent compound which is in *in-human* studies, for the treatment of acute leukemia.[12]

Histone lysine erasing regulators are responsible for the removal of acetyl (histone deacetylases - HDACs) and methyl (demethylases) groups. The former are involved in more than 350 clinical trials, most of them focusing on haematological malignancies.[13]

Two major protein families are responsible for the deacetylation of lysines (**Figure 1.4**). Classical HDACs contain active sites that are zinc-dependent, while Sirtuin proteins use (nicotinamide adenine dinucleotide)$^+$, NAD$^+$, as their cofactor. For classical HDAC proteins, the process is initiated by polarization of the carbonyl group, which activates the rate-determining nucleophilic attack by a zinc-bound water molecule at the electrophilic carbon. This is followed by hydrogen abstraction, with the release of the acetyl group as the final step. Proteins belonging to the Sirtuin family initiate the reaction by the nucleophilic attack of the acetylated lysine on the ribose ring of the NAD$^+$ cofactor, releasing the nicotinamide and forming a positively charged imidate intermediate. A nucleophilic attack of the imidate is promoted by surrounding active site histidine residues, forming a bicyclic intermediate. With the hydrolysis, there is the net transfer of an acetyl group to the ribose of the cofactor, resulting in the deacetylation of the histone lysine.

Demethylases (**Figure 1.4**) are also classified according to the cofactor they use. Jumonji demethylases contain a α-ketoglutarate cofactor and a non-haem iron-dependent active site. Lysine-specific demethylase 1 (LSD) proteins use flavin adenine dinucleotide (FAD) as a redox cofactor. In LSD proteins, the reaction is initiated by α-CH cleavage at the methylated lysine, in the form of a hydride transfer to the FAD cofactor. The cofactor is then recovered in the presence of O_2, forming as a subproduct hydrogen peroxide. The next step involves hydrolysis of the imine intermediate with the concomitant release of formaldehyde and the demethylated histone lysine residue.[14] In Jumonji demethylases, the reaction is initiated by the binding of O_2 to the iron active site. The next step, oxygen activation, involves the release of -CO_2 and cleavage of the peroxo O-O bond, forming an iron-oxo intermediate. This species is responsible for C-H oxidation of the substrate, which is proposed to occur in two distinct steps: C-H abstraction followed by the radical rebound of the hydroxyl group to the histone lysine. The hemiaminal product is then thermodynamically unstable with respect to the release of formaldehyde and amine. Differences in these mechanisms are responsible for changes in the recognition process of different methylation states by these two families of demethylases. While JmjC proteins are able to recognize all possible methylation states, LSD proteins, which depend on the formation of the iminium intermediate, are unable to catalyse demethylation of trimethylated lysines.

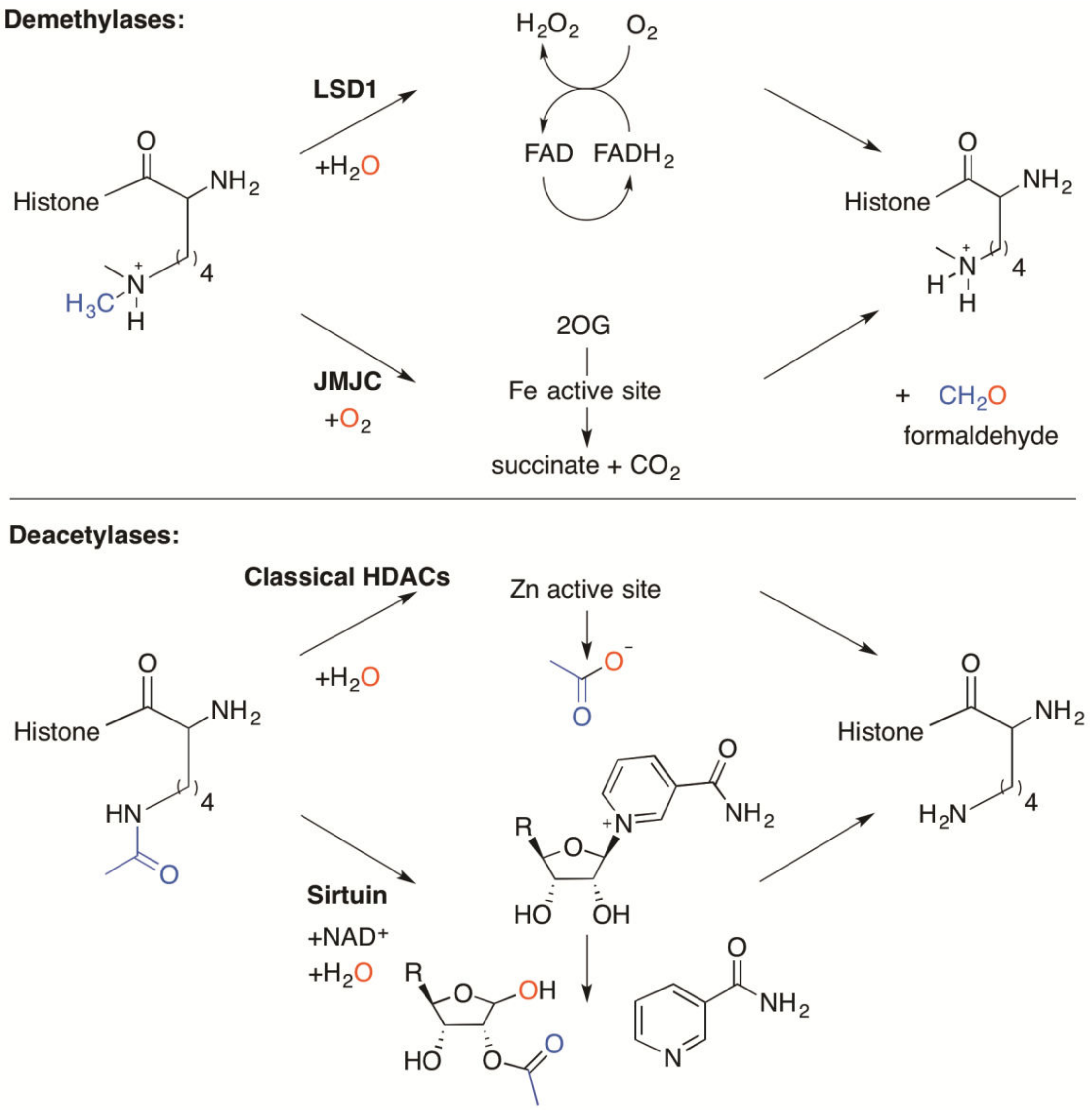


Figure 1.4. Overview of histone lysine demethylation and deacetylation mechanisms.[10]

Finally, histone reading regulators (**Figure 1.5**) bind to modified histones, through either methyl (Tudor domains, The Plant Homeodomain (PHD) fingers and chromodomains) or acetyl groups (bromodomains). Apart from the previously described OTX015, other bromodomain inhibitors are also in the early-stage of clinical trials. For example, both GlaxoSmithkline-sponsored compound, GSK525762, and Bayer-sponsored compound, BAY1238097, are currently in clinical trials for the treatment of brain tumours.[15]

Tudor domains are characterized by an anti-parallel and bent conformation of β-sheets, with approximately 50 amino acids. PHD fingers are protein domains usually consisting of two-stranded β-sheet and a globular fold defined by α-helices. Chromodomains are characterized by the presence of three β-sheets packed against a C-terminal α-helix. The structure of bromodomain containing proteins presents four helices (αA, αB, αC and αZ,) as well as two loop regions connecting αZ-αA (ZA) and αB-αC (BC). The binding site of bromodomains is predominantly hydrophobic and, in most of the cases, contains a conserved tyrosine or asparagine directly binding to the acetylated lysine.

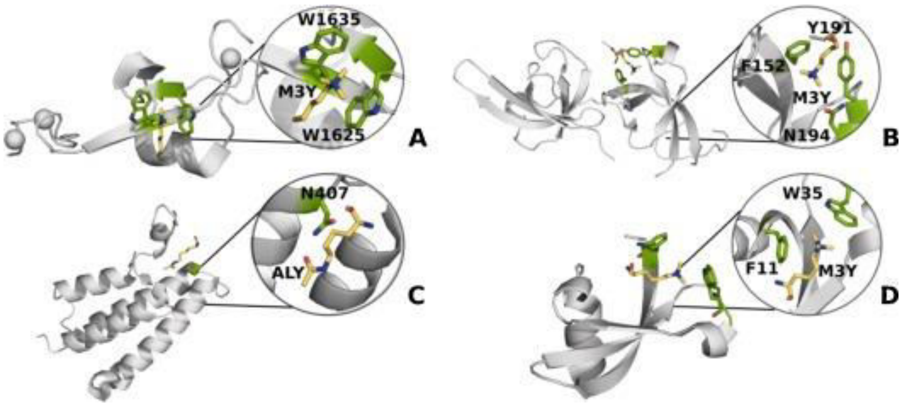


Figure 1.5. General structures of four major epigenetic reader proteins: A) PHD fingers (PDB ID: 3DB3), B) tudor domains (PDB ID: 3GL6), C) bromodomains (PDB ID: 1E6I) and D) chromodomains (PDB ID: 3I90) shown with select key residues necessary for binding with trimethylated (M3Y) and acetylated lysine (ALY) histone residues.[10]

1.3 Understanding of PTMs using computation

Histone modifications involve a wide range of proteins that can regulate chromosome organization and the expression of genes related to several diseases.[3] A full understanding of the epigenetic mechanisms, *e.g.* lysine methylation and acetylation, taking place at the protein-protein interface would provide important insights on the role of such modifications for biological changes in the chromosome. Experimental data alone may not be sufficient to elucidate the molecular basis for substrate recognition, nor the chemical events taking place inside proteins. In this context, computational modelling can provide important insights on structural features driving the kinetics and thermodynamics of the process. In this section, we will provide some examples in which the use of computational modelling was crucial for such mechanistic understanding.

The underlying computational methodologies are described in more detail in the following chapter. However, it is worth emphasising the different domains of applicability of classical *vs.* quantum methodologies in broad terms at this point. Classical simulations use a molecular mechanics (*i.e.* force field) representation of the system, and therefore are applicable to the study of processes involving structural (*e.g.* conformational) changes. In contrast, quantum mechanics (QM) calculations are able to describe changes in electronic structure associated with chemical reactions. Thus QM or hybrid QM/MM calculations are used to probe mechanistic steps including the calculation of transition structures, which are inaccessible using purely classical techniques.

For the deacetylation reaction in proteins from the classical HDAC family, QM/MM calculations were performed to understand which of the two active site aspartate-histidine diads is involved in the proton transfer to the acetylated histone and which one acts as a base in the process (**Figure 1.6**). Chen[16] showed that one of the aspartate-histidine diads functions as a general base for the deacetylation reaction, and the other diad contains a positively charged histidine that protonates the ε-nitrogen, thus enabling the collapse of the tetrahedral intermediate. They performed QM/MM calculations at the M05-2X/6-31G*:Amber level of theory using an extended active site including aspartates, similar to previously reported QM studies by Vanommeslaeghe.[17,18] An alternative mechanism is proposed by Wu[19] showing a proton-shuttle for HDAC8, in which a single neutral histidine accepts a proton from the water molecule in the active site and then transfers it to the amide nitrogen, facilitating the release of the acetyl group. Their calculations were performed at the B3LYP/6-31G*:Amber level of theory.

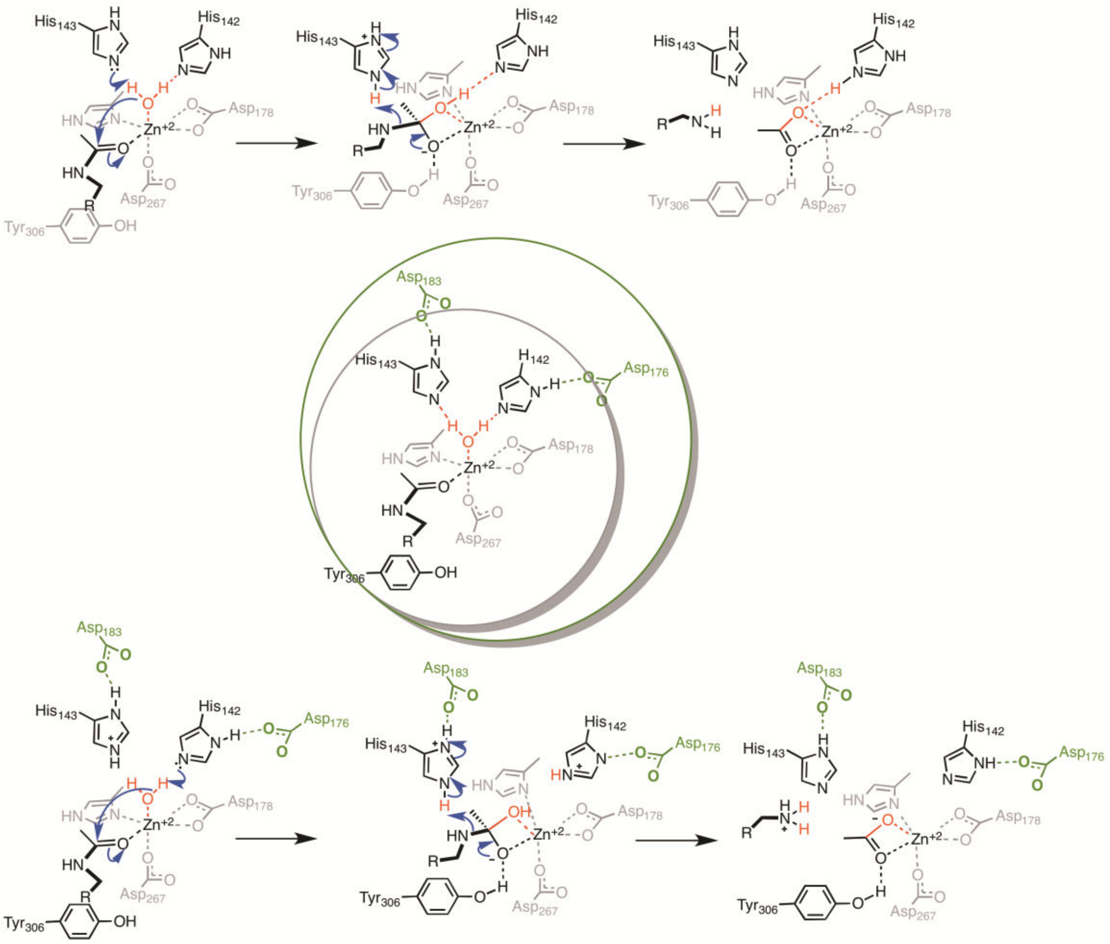


Figure 1.6. The active site matters: at the top, deacetylation in HDAC8 via the proposed proton shuttle mechanism is predicted using the smaller active site.[19] At the bottom, an extended active site model including two aspartate residues implicated the involvement of both histidines in the mechanism.[17,10]

The use of computational modelling has also been important for underlying the basis for recognition of acetylated lysines by bromodomains. Pizzitutti[20] performed 5-10 ns MD simulations with TIP3P solvation and the CHARMM27 force field on the bromodomain of Gcn5p (**Figure 1.7**). They highlighted the importance of a proline for affecting the ZA loop movement and structure in a way that the protein is able to recognize the histone acetylated, corroborating their experimental mutation studies – the mutation of this residue interferes in the histone-Gcn5p complexation.

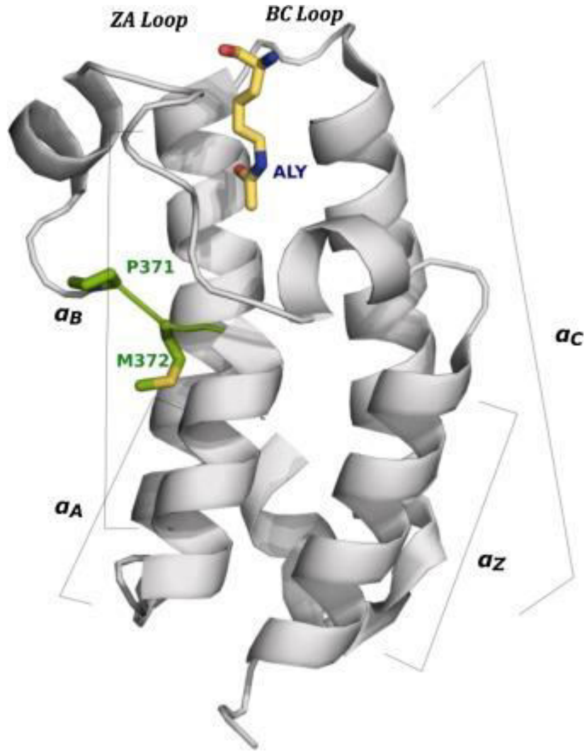


Figure 1.7. Gcn5p bromodomain. Shown here is the residue proline (P371, green), which although not in the binding site of the acetylated lysine is important for substrate recognition. This residue affects the dynamic and structure of the ZA loop (PDB ID: 1E6I).[10]

Caflisch[21] and co-workers have performed microseconds of MD with the CHARMM27 force field, showing a new "P-binding mode" for the acetylated lysine-

TBP-Associated Factor 1 (TAF1) complexation. They highlighted the importance of a proline residue, located in the vacant lipophilic WPF shelf of TAF1, for the complex stabilization, by hydrogen bonding the acetylated lysine. This differs from the well-characterized "N-binding mode" involving an asparagine residue and currently awaits further experimental studies.[21]

In this section, just few cases are showed where the use of computational modelling techniques, such as QM, QM/MM and MD, was crucial for the understanding of epigenetics processes. In particular for the deacetylation reaction in classical HDACs and the recognition of acetylated lysines by bromodomains. The reader is referred to recent reviews[10] where a more detailed summary of the different computational studies in this area is provided.

1.4 Non-BET Bromodomains as promising anticancer drugs

CREB-binding protein (CREBBP) bromodomains are epigenetic "reader" proteins that recognize acetylated histone lysine residues and their selective inhibition has been suggested as a promising way of designing more effective anticancer drugs.[22] Recent bromodomain crystal structures reported in the literature provide a structural basis for the design of new inhibitors of epigenetic reader enzymes.[23] Although bromodomain targeted therapies have focused on proteins from the bromodomain and extra terminal (BET) family,[23] few molecules have been reported to have nanomolar binding affinity for bromodomains that are not part of this family, *e.g.* CREBBP, but with relatively low binding affinities or selectivities.[23-24] Computational insights can help us to understand important interactions such as hydrogen bonding interactions with water and other small molecules, or cation-π interactions not immediately apparent from X-ray structures. Through our investigation, we hope to be able to compare drug binding in different bromodomain's active sites, with the intention of this leading to the development of selective drugs based on differences in protein primary and secondary structure.

1.5 Uncovering mechanistic details of *Jumonji C* demethylases

Despite a recent rise in interest surrounding histone demethylation by non-haem iron centres, detailed insights into the mechanism (**Figure 1.8**) have remained elusive. In particular, the structural basis for substrate selectivity within the active site, and the dependence on C-H bond strength in the hydrogen abstraction step, are still poorly understood. Furthermore, the mechanistic details of demethylation and the potential involvement of intersystem crossing (ISC) have yet to be fully understood for these important epigenetic processes. We have implemented QM/MM techniques[25] in our attempts to uncover the steps surrounding demethylation, and to evaluate the contribution of individual amino acids on structures, energetics and dynamics along the reaction coordinate. Computational insights into the mode of action of the most promising residues for demethylation with JmjC proteins will no doubt prove to be invaluable tools in the future testing of enzyme mutants and in the rational design of small-molecule inhibitors.

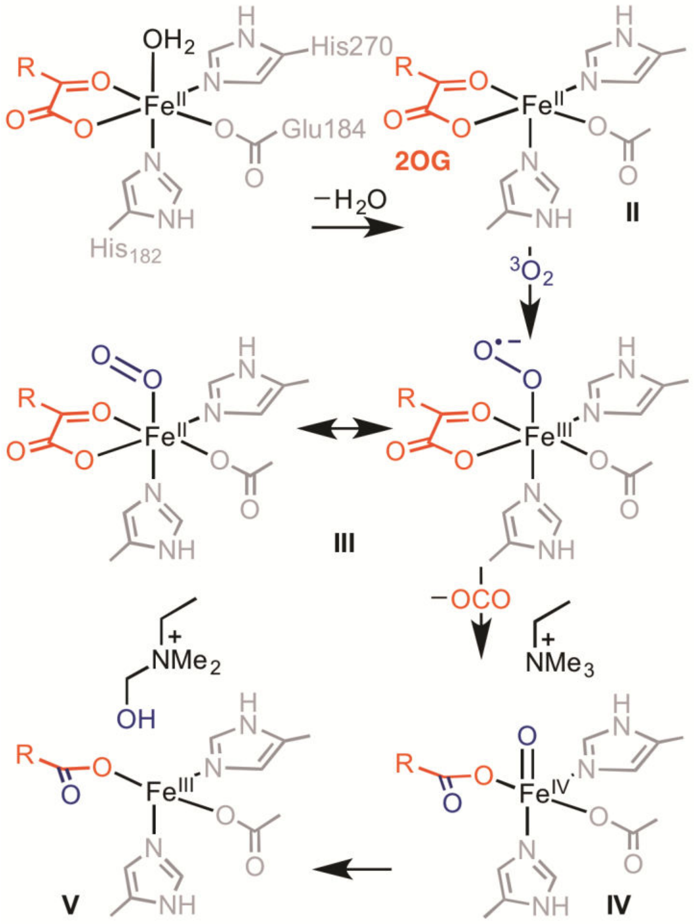


Figure 1.8. The X-ray structure of JMJD2A, a member of the JmjC family, with cofactor analogue and axial water molecule; proposed mechanism of oxygen activation and lysine demethylation by an iron^{IV}–oxo. JMJD2A contains a catalytic domain with Fe^{II} and a 2-oxoglutarate (2OG) cofactor that activates molecular oxygen to form a high-valent iron^{IV}-oxo species, which has been proposed to be the active species of these enzymes. After O_2 binding, 2OG is converted to succinate with the release of carbon dioxide. The resulting iron^{IV}-oxo complex is capable of carrying out a hydrogen atom abstraction from the methyl group of the methylated lysine of the histone, generating formaldehyde and eventually demethylating the histone.[25]

1.6 Aims of this thesis

Recent discoveries have shown the value of computational chemistry techniques for uncovering mechanistic aspects of epigenetic chemical reactions and small-molecule recognition. In the next chapters, applications of state-of-the-art *in silico* approaches are presented for:

a) Evaluation of important intrinsic interactions in biological environments (*e.g.* hydrogen bonds and cation-π interactions) in the CREBBP bromodomain. Through a collaborative work with the group of Prof. Stuart Conway, this thesis demonstrated how computational models are powerful tools for predicting binding affinity of inhibitors in CREBBP.

b) Determining the full PES of the demethylation reaction in JMJD2A. The use of cutting edge computational techniques, previously benchmarked against spectroscopy studies and multi-configuration QM methods in proteins with similar active sites, was essential for computing the thermodynamic spontaneity of the O_2 binding step. Thus, these approaches were also adopted for getting a more accurate description of the following steps in JMJD2A demethylation.

We hope these insights will facilitate the design of more potent and selective inhibitors of epigenetic targets, as well as the exploration of protein mutants.

1.7 References for Chapter 1

1. Wang, F.; Higgins, J. M., Histone modifications and mitosis: countermarks, landmarks, and bookmarks. *Trends Cell Biol* **2013,** *23* (4), 175-84.

2. Uemura, M.; Yamamoto, H.; Takemasa, I.; Mimori, K.; Hemmi, H.; Mizushima, T.; Ikeda, M.; Sekimoto, M.; Matsuura, N.; Doki, Y.; Mori, M., Jumonji domain containing 1A is a novel prognostic marker for colorectal cancer: in vivo identification from hypoxic tumor cells. *Clin Cancer Res* **2010,** *16* (18), 4636-46.

3. Okada, Y.; Scott, G.; Ray, M. K.; Mishina, Y.; Zhang, Y., Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* **2007,** *450* (7166), 119-23.

4. (a) Drohat, A. C.; Coey, C. T., Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA. *Chem Rev* **2016,** *116* (20), 12711–729; (b) Rasmussen, K. D.; Helin, K., Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* **2016,** *30* (7), 733-50; (c) Erdmann, A.; Halby, L.; Fahy, J.; Arimondo, P. B., Targeting DNA methylation with small molecules: what's next? *J Med Chem* **2015,** *58* (6), 2569-83.

5. (a) Kaminskas, E.; Farrell, A.; Abraham, S.; Baird, A.; Hsieh, L. S.; Lee, S. L.; Leighton, J. K.; Patel, H.; Rahman, A.; Sridhara, R.; Wang, Y. C.; Pazdur, R.; Fda, Approval summary: azacitidine for treatment of myelodysplastic syndrome subtypes. *Clin Cancer Res* **2005,** *11* (10), 3604-8; (b) Saba, H. I., Decitabine in the treatment of myelodysplastic syndromes. *Ther Clin Risk Manag* **2007,** *3* (5), 807-17.

6. (a) VanderMolen, K. M.; McCulloch, W.; Pearce, C. J.; Oberlies, N. H., Romidepsin (Istodax, NSC 630176, FR901228, FK228, depsipeptide): a natural product recently approved for cutaneous T-cell lymphoma. *J Antibiot (Tokyo)* **2011,** *64* (8), 525-31; (b) Mann, B. S.; Johnson, J. R.; Cohen, M. H.; Justice, R.; Pazdur, R., FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **2007,** *12* (10), 1247-52; (c) Szyf, M., Prospects for the development of epigenetic drugs for CNS conditions. *Nat Rev Drug Discov* **2015,** *14* (7), 461-74.

7. Odore, E.; Lokiec, F.; Cvitkovic, E.; Bekradda, M.; Herait, P.; Bourdel, F.; Kahatt, C.; Raffoux, E.; Stathis, A.; Thieblemont, C.; Quesnel, B.; Cunningham, D.; Riveiro, M. E.; Rezai, K., Phase I Population Pharmacokinetic Assessment of the Oral Bromodomain Inhibitor OTX015 in Patients with Haematologic Malignancies. *Clin Pharmacokinet* **2015**, *55* (3), 397-405.

8. Pandey, R.; Muller, A.; Napoli, C. A.; Selinger, D. A.; Pikaard, C. S.; Richards, E. J.; Bender, J.; Mount, D. W.; Jorgensen, R. A., Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res* **2002,** *30* (23), 5036-55.

9. Marmorstein, R., Structure of histone acetyltransferases. *J Mol Biol* **2001,** *311* (3), 433-44.

10. (a) Cortopassi, W. A.; Kumar, K.; Duarte, F.; Pimentel, A. S.; Paton, R. S., Mechanisms of histone lysine-modifying enzymes: A computational perspective on the role of the protein environment. *J Mol Graph Model* **2016,** *67*, 69-84; (b) Smith, B. C.; Denu, J. M., Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta* **2009,** *1789* (1), 45-57; (c) Vellore, N. A.; Baron, R., Epigenetic molecular recognition: a biomolecular modeling perspective. *ChemMedChem* **2014,** *9* (3), 484-94.

11. Wapenaar, H.; Dekker, F. J., Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clin Epigenetics* **2016,** *8*, 59.

12. Morera, L.; Lubbert, M.; Jung, M., Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin Epigenetics* **2016,** *8*, 57.

13. Nervi, C.; De Marinis, E.; Codacci-Pisanelli, G., Epigenetic treatment of solid tumours: a review of clinical trials. *Clin Epigenetics* **2015,** *7*, 127.

14. Kong, X.; Ouyang, S.; Liang, Z.; Lu, J.; Chen, L.; Shen, B.; Li, D.; Zheng, M.; Li, K. K.; Luo, C.; Jiang, H., Catalytic mechanism investigation of lysine-specific demethylase 1 (LSD1): a computational study. *PLoS One* **2011,** *6* (9), e25444.

15. Wadhwa, E.; Nicolaides, T., Bromodomain Inhibitor Review: Bromodomain and Extra-terminal Family Protein Inhibitors as a Potential New Therapy in Central Nervous System Tumors. *Cureus* **2016,** *8* (5), e620.

16. Chen, K.; Zhang, X.; Wu, Y. D.; Wiest, O., Inhibition and mechanism of HDAC8 revisited. *J Am Chem Soc* **2014,** *136* (33), 11636-43.

17. Vanommeslaeghe, K.; De Proft, F.; Loverix, S.; Tourwe, D.; Geerlings, P., Theoretical study revealing the functioning of a novel combination of catalytic motifs in histone deacetylase. *Bioorg Med Chem* **2005,** *13* (12), 3987-92.

18. Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P., Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **1999,** *401* (6749), 188-93.

19. Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y., A proton-shuttle reaction mechanism for histone deacetylase 8 and the catalytic role of metal ions. *J Am Chem Soc* **2010,** *132* (27), 9471-9.

20. Pizzitutti, F.; Giansanti, A.; Ballario, P.; Ornaghi, P.; Torreri, P.; Ciccotti, G.; Filetici, P., The role of loop ZA and Pro371 in the function of yeast Gcn5p bromodomain revealed through molecular dynamics and experiment. *J Mol Recognit* **2006,** *19* (1), 1-9.

21. Magno, A.; Steiner, S.; Caflisch, A., Mechanism and Kinetics of Acetyl-Lysine Binding to Bromodomains. *J Chem Theory Comput* **2013,** *9* (9), 4225-32.

22. Fu, L. L.; Tian, M.; Li, X.; Li, J. J.; Huang, J.; Ouyang, L.; Zhang, Y.; Liu, B., Inhibition of BET bromodomains as a therapeutic strategy for cancer drug discovery. *Oncotarget* **2015,** *6* (8), 5501-16.

23. Rooney, T. P.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin, S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Muller, S.; Knapp, S.; Brennan, P. E.; Conway, S. J., A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation-pi interaction. *Angew Chem Int Ed Engl* **2014,** *53* (24), 6126-30.

24. (a) Patel, J.; Pathak, R. R.; Mujtaba, S., The biology of lysine acetylation integrates transcriptional programming and metabolism. *Nutr Metab* **2011,** *8*, 12; (b) Borah, J. C.; Mujtaba, S.; Karakikes, I.; Zeng, L.; Muller, M.; Patel, J.; Moshkina, N.; Morohashi, K.; Zhang, W.; Gerona-Navarro, G.; Hajjar, R. J.; Zhou, M. M., A small molecule binding to the coactivator CREB-binding protein blocks apoptosis in cardiomyocytes. *Chem Biol* **2011,** *18* (4), 531-41.

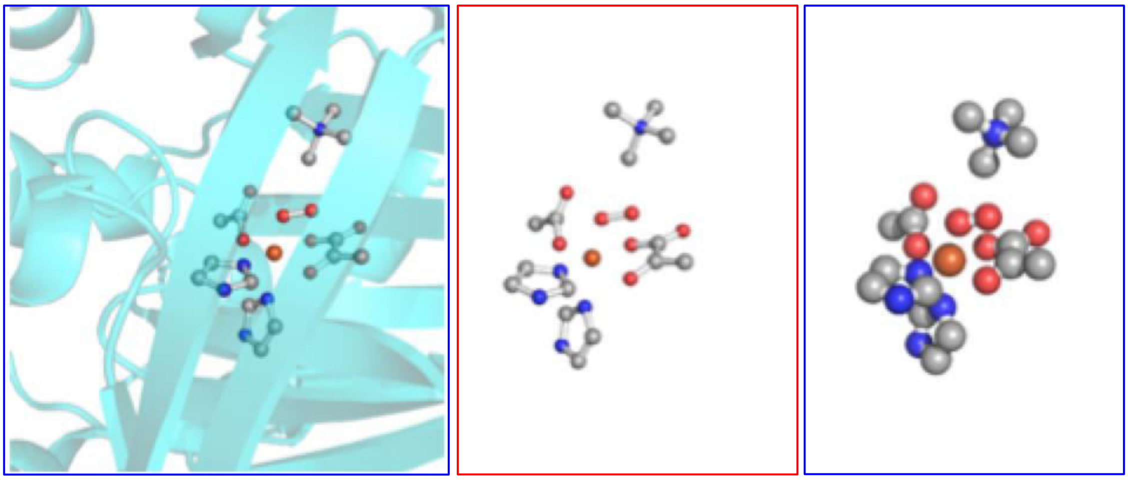
25. Cortopassi, W. A.; Simion, R.; Honsby, C. E.; Franca, T. C.; Paton, R. S., Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem Eur J* **2015**, *21* (52), 18983-92.

Chapter 2

Quantification of binding energies and energy potential surfaces in epigenetic targets

"If a man never contradicts himself, the reason must be that he virtually never says anything at all."
Erwin Shrödinger

Our own **N**-layered **I**ntegrated molecular **O**rbital and molecular **M**echanics
ONIOM - Gaussian®



$$E^{\text{low level}}_{\text{real system}} - E^{\text{low level}}_{\text{small model}} + E^{\text{high level}}_{\text{small model}}$$

2.1 Overview

Ascertaining the mechanistic details of reactions taking place at the protein-protein interface of histones or epigenetic macromolecules often proves to be an arduous task for Experimental techniques. The derivation of parameters for the description of transition states (TS) and the thermodynamics of the reaction path is essential in the development of accurate models but is all too often fraught with difficulty. In drug design, problems arise in the development of models to predict binding affinities accurately, even before the synthesis of potential novel protein inhibitors. As previously described in **Chapter 1**, implementation of computational techniques, such as long-time Molecular Dynamics (MD) and Quantum Mechanics (QM) in the computation of potential energy surfaces (PES) can shed light on the physicochemical properties of a system at the atomic level that may not be clear from simple analysis of well-established experimental data nascent from crystal structures or other spectroscopic techniques. Computation of bond lengths, angles, dihedrals, molecular vibrations as well as analysis of important small molecule-protein interactions (such as hydrogen bonds, π-π and cation-π interactions) as a function of time are of great value in the understanding of phenomena in biological environments.

This Chapter describes the theoretical background underpinning Quantum and Molecular Mechanics techniques. These calculations have been used in our MD simulations, geometry optimizations and single point energy calculations for elucidating the mechanistic details of epigenetic proteins.

2.2 Molecular Mechanics

Computational Chemists often implement classical molecular mechanics (MM) approximations in their study of the dynamics of small molecules, taking advantage of the technique's significantly lower computational cost compared with more elaborate QM or semi-empirical methods. This is not to say that MM approximates are certainly less accurate: high-quality classical calculations can be obtained with suitable parameters. At the same time, the lower computational cost enables the study of longer time-scales than would be possible using more expensive alternatives. MD techniques use classical mechanics and empirical parameters to quantify interactions at the interatomic level. The energies computed in MM calculations are derived differently dependent upon the various types of the interaction between the atoms (bonded or non-bonded).

$$E_{pair} = E_{bonded} + E_{nonbonded} \qquad \textbf{\textit{Eq. 2.1}}$$

Atoms, which are treated distinctly according to their formal hybridization and connectivity (the so-called atom-type), are represented by well-defined spherical radii, and bonds are usually represented using harmonic potentials:

$$E_{bonded} = \sum_{bonds} k_r\left(r - r_{eq}\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{v_n}{2}[1 + \cos(n\phi - \gamma)] \qquad \textbf{\textit{Eq. 2.2}}$$

The non-bonded interactions, such as van der Waals (vdW), which includes both repulsive (*i.e.* steric) and attractive (*i.e.* dispersive) terms, and electrostatics, are computed as:

$$E_{nonbonded} = \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \qquad \textbf{\textit{Eq. 2.3}}$$

Parameters for these equations (A_{ij} and B_{ij}) are usually derived from experimental reference data, although QM benchmark calculations may also be used, and they are included in force fields. Force fields contain expected values for chemical and physicochemical interactions based on similar molecules to the ones under analysis. VdW parameters, bond lengths, angles, and dihedrals can be based on the geometries resulting from higher-level calculations using QM or *ab initio* techniques, and also from spectroscopic data; charges are usually defined using QM calculations, *e.g.* point charges that can reproduce the electrostatic potential surface (ESP) of a given molecule.[1]

The accuracy of these calculations relies on the accuracy of the parameter set, and their *transferability* to systems, which may not be identical to those originally used during the parameterization process. For example force fields well parameterized to describe amino acid conformational preferences are more well-suited to the study of proteins than say, a force field focussed on the description of small organic systems. Thus the appropriate choice of method is important, given that different biological systems may require different parameterization for good reproducibility of experimental data.[2]

MD simulation is an iterative process, where forces are computed given initial atomic positions and velocities before these vectors are updated using classical mechanics. For each of the simulated steps, energies are calculated according to ***Eq. 2.1***, ***Eq. 2.2*** and ***Eq. 2.3***, and the output trajectory is analysed.

2.3 Quantum Mechanics

With sufficiently accurate parameters, MM techniques are able to accurately describe many small molecule and protein binding dynamics.[3] These techniques are therefore inherently *empirical* since they require explicit parameterization with prior data. However, in cases involving formation or cleavage of bonds, *e.g.* for the understanding of chemical reaction paths, development of accurate force fields is still a challenge. In this thesis, we have focussed on both approaches: studies involving ligand binding in bromodomains were treated at the MM level, while analysis of the demethylation reaction in histone demethylases required a more accurate description of the electronic structure, and so we used QM. QM calculations may in principle be *ab initio*, in the sense that no parameterization is required. Practical implementations of QM, such as density functional theory (DFT), may involve the use of some global parameters. In this case, they are not completely free of empiricism, however, these parameters are at least system independent.

Calculations at the QM level in the present work used the solution of the time independent Schrödinger equation:

$$H\Psi = E\Psi \qquad \textbf{\textit{Eq. 2.4}}$$

Where H represents the Hamiltonian, that, when applied to the wave function Ψ, provides the total energy for the analysed system.

While different wave functions can solve **Eq. 2.4**, the variational principle[8] states that the ground state is better represented by the wave function with the lowest energy associated to it.

Finding the exact solution for the Schrödinger equation would result in an accurate description of the molecule at the QM level, and would be excellent for prediction of chemical properties. This is, however, a challenge for multi-electron systems. Approximate methods are therefore needed, and the Born-Oppenheimer[4] approximation allows us to separate the analysis of nuclei and electrons. This is based on the assumption that the atomic nuclei are approximately three orders of magnitude heavier than electrons, and it is thus reasonable to solve the time independent

Schrödinger equation for the electronic structure for a molecule with stationary nuclei.

$$\Psi = \Psi_n \Psi_e$$ ***Eq. 2.5***

For calculating the wave function related to the electronic part, we turn back to **Eq. 2.4** and apply the Born-Oppenheimer approximation:

$$H_e \Psi_e = E_e \Psi_e$$ ***Eq. 2.6***

With the Hamiltonian operator defined as:

$$H_e = -\frac{1}{2}\sum_{i}^{n} \nabla_i^2 - \sum_{I}^{N}\sum_{i}^{n} \frac{Z_l}{r_{Ii}} + \sum_{i<j}^{n} \frac{1}{r_{Ii}} + V_{nuc}$$ ***Eq. 2.7***

Solving **Eq. 2.6** will give us the electronic wave functions for specific nuclei coordinates. When we find the electronic wave function for all possible nucleic positions, we will have a PES.

The Slater determinant is now introduced (**Eq. 2.8**),[5] which describes the electronic wave function as a determinant of one-electron wave functions. When using the Slater determinant, we ensure the Pauli exclusion principle[6] is followed, in contrast to the Hartree method,[7] in which a wave function is separated into a product of functions reliant on one electron.

$$\Psi_{(R_1, R_2 .. R_n)} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(e_1) & \phi_2(e_1) & \phi_n(e_1) \\ \phi_1(e_2) & \phi_2(e_2) & \phi_n(e_2) \\ \cdots & \cdots & \cdots \\ \phi_1(e_n) & \phi_2(e_n) & \phi_n(e_n) \end{vmatrix} = |\phi_1, \phi_2 \ldots \phi_n|$$ ***Eq. 2.8***

The Linear Combination of Atomic Orbitals (LCAO) (**Eq. 2.9**)[8] states that these one-electron wave functions can be defined as the sum of one-electron functions; these are often referred to as basis sets.

$$\emptyset_i = \sum_{\mu}^{k} c_{i\mu} \chi_\mu$$ ***Eq. 2.9***

Solving the Schrödinger equation now relies on defining an appropriate LCAO that would result in the minimum energy for the ground state – according to the variational principle.[8]

The Hartree-Fock method[9] uses the Slater determinant to find the multi-electron wave function. The major pitfall of this method is that it does not explicitly consider the electronic correlation, and while the Pauli exclusion principle is followed, electron-electron repulsion is not clearly taken into account. One of the ways to solve this problem is by defining the multi-electron wave function not only dependent on a single determinant (as in the case of HF) but as a linear combination of various determinants, an approach known as configuration interaction (CI).[10] Another alternate approach, which is extensively used in this thesis, is the use of density functional theory (DFT).[11] Based on the Hohenberg-Kohn theorems,[12] DFT methods rely on the assumption that the ground state energy depends on the electron density (*i.e.* the energy is a function of this density). Then, the ground state energy is the minimum value that this energy can assume, according to the variational principle. The challenge remains in defining the most accurate functional that could describe the total energy of the multi-electron system under analysis.

With DFT, energy can be described as a functional of the electron density:

$$E[\rho(\boldsymbol{r})] = T_e[\rho(\boldsymbol{r})] + V_{nu}[\rho(\boldsymbol{r})] + V_{el}[\rho(\boldsymbol{r})] + E_{XC}[\rho(\boldsymbol{r})]$$

Eq. 2.10

The exchange-correlation term, a problem previously described in our discussion of the Hartree-Fock method, can be separated as follows:

$$E_{XC}[\rho(\boldsymbol{r})] = E_X[\rho(\boldsymbol{r})] + E_C[\rho(\boldsymbol{r})] = \int \rho(\boldsymbol{r})\, \boldsymbol{\varepsilon}_{\boldsymbol{X}}[\rho(\boldsymbol{r})]d\boldsymbol{r} + \int \rho(\boldsymbol{r})\, \boldsymbol{\varepsilon}_{\boldsymbol{c}}[\rho(\boldsymbol{r})]d\boldsymbol{r}$$

Eq. 2.11

Different functionals vary in their way of calculating the exchange-correlation term. Some employ the local-density approximations (LDA),[11] in which the exchange-correlation term depends on the electron density at each point in space.

However, LDA methods consider the electron density to be uniform. More advanced methods, such as the generalized gradient approximation (GGA),[13] take into consideration the non-homogeneity of the electron density by calculating its gradient at each point in space to evaluate the exchange-correlation. One of the most popular GGA density functionals is B3LYP,[14] but this is also hybrid: it incorporates some HF exchange. The parameters a, b and c (**Eq. 2.12**) were optimized by fitting against reference atomization energies.

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSDA} + \mathrm{a}E_X^{HF} + b\Delta E_X^{B} + (1-c)E_C^{LSDA} + cE_C^{LYP}$$

Eq. 2.12

Other functionals are classified as meta-GGA,[15] such as the Minnesota family of functionals, which includes the hybrid M06-2X.[16] Apart from the magnitude of the density and its gradient, these methods also take into consideration the Laplacian (second derivative) of the density. By using higher derivatives of the electron density, these functionals account for local changes in the electron density, and are thus considered to be “semi-local”. Capturing these effects leads to an improved description of electron correlation at medium-range and thus the ability to a superior description of nonbonding (*e.g.* van der Waals) interactions *vs.* GGA functionals.

Another important factor to consider when using QM methods for optimization and energy calculations is the role of the solvation. Due to the increased system size caused by explicitly adding solvent molecules and the sheer number of possible configurations of such a system, and consequently the high cost of such calculations, the use of implicit solvation models is often considered a reasonable alternative.

The replacement of the explicit solvent is usually done by applying a surrounding dielectric constant around a cavity containing the molecules under analysis. In the self-consistent reaction field (SCRF) method,[17] the free energy of solvation is calculated as:

$$\Delta G_{solvation} = \Delta G_{cav} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{elec} \qquad \textbf{\textit{Eq. 2.13}}$$

With the ΔG_{cav} being the free energy associated with creating this cavity, ΔG_{disp} the free energy from dispersion effects, and ΔG_{rep} and ΔG_{elec} accounting for repulsive and electrostatic interactions respectively between the solvent and the chemical system of interest.

2.4 Potential Energy Surface

We have discussed approaches to solving the Schrödinger equation and finding energy minima associated with a ground state; by evaluating all the possible atomic coordinates and finding electronic wave functions for fixed nuclei, we can derive a PES, as previously discussed in **Section 2.3**. To confirm that a system under scrutiny is indeed a ground state, it is important to evaluate the second derivative of the Hessian matrix.[8] A ground state exists when the calculated eigenvalues (force constants) are all positive. In cases where one, and only one, of the force constants is negative, it is representative of a transition state (a saddle point on the PES – **Figure 2.1**). Both ground states and transition states are stationary points on the PES, for which the gradient of the energy is zero.

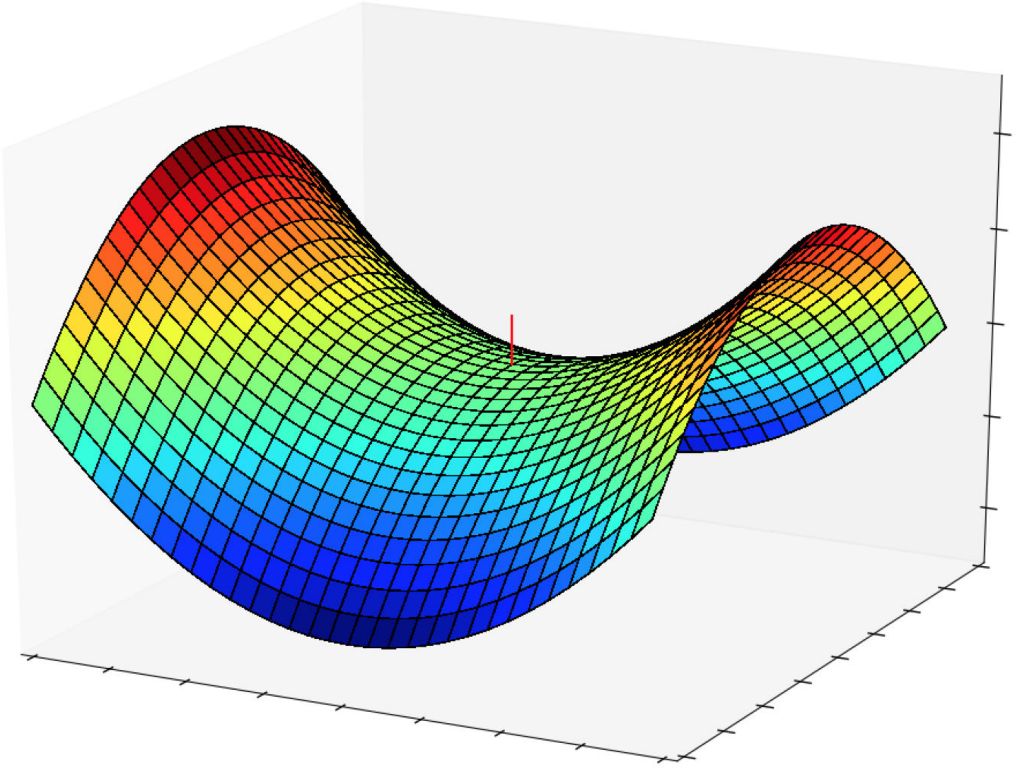


Figure 2.1. Illustration showing the location of the saddle point (indicated by the red line) on a fictitious potential energy surface plot. Thanks to Dr. Kelvin E. Jackson for allowing the use of this Figure from his thesis (Dr. Kelvin Jackson, DPhil Thesis, Oxford University, 2016).

2.5 Quantum Mechanics/Molecular Mechanics

Our own **N**-layered **I**ntegrated molecular **O**rbital and molecular **M**echanics (ONIOM)[18] combines QM and MM approaches to deal with large systems, such as reactions taking place in protein environments, at reduced computational cost relative to pure QM. Atoms directly involved with the reaction are usually treated with QM (here defined as the "high level") while the remaining atoms are treated at the MM level (here defined as the "low level"). This technique consists of separating a system into different layers. In this thesis we have exclusively used two different layers. Amino acids and other molecules involved in chemical changes in the active site that were considered essential for the computation of the PES are defined as the model system. The entire protein environment, including the substrates and cofactors, is defined as the real system. The ONIOM approach computes energies using **Eq. 2.14**:

$$E_{ONIOM} = E_{real}^{low} + E_{model}^{high} - E_{model}^{low}$$

Eq. 2.14

When considering just part of residues in the model system, it is necessary to create link atoms. Where a covalent bond crosses between different layers of the QM/MM calculations, the electronics are treated as if a hydrogen atom had replaced the rest of the molecule.

An additional approach that can be useful in ONIOM calculations is the use of electronic embedding (EE),[19] which enables both the steric and electrostatic properties of the entire molecule to be taken into account when modelling processes in the high layer. EE works by incorporating the partial charges of the protein environment, treated at the MM level, into the model system, treated at the QM level, thereby providing a better description of the electrostatic interaction between these two regions and allowing the QM wave function to be polarized.

2.6 Supporting Information for Chapter 2

In this section, we will present some methods developed by our group and used in this present thesis for thorough analysis of ONIOM QM/MM energies. These methods were applied for the understanding of the dioxygen binding in the active site of the demethylase JMJD2A,[20] *discussed further in* ***Chapter 5****, and could be easily extended to other protein environments.*

2.6.1 Energy Decomposition Analysis

The objective with this energy decomposition analysis (EDA) discussion is to dissect the total binding energy of a small molecule (dioxygen) to a macromolecular receptor (the JMJD2A enzyme, PDB ID: 2OQ6) into individual components derived from fundamental physical interactions. Electrostatic and vdW contributions to the overall energy were also separated into contributions from each residue in the low level of the QM/MM calculation and atoms in the active site, following an approach proposed by Hirao.[21] This approach is implemented using Python, post-processing the results of QM/MM calculations performed with the ONIOM approach implemented.

Electrostatic energies are defined using the Coulomb potential between two point charges:

$$E_{electrostatic} = \sum_{pairs} \frac{s_{pairs}^{el}\ q_1 q_2}{R^{pairs}} \qquad \textbf{\textit{Eq. 2.15}}$$

s_{pairs}^{el} scales 1,4-interactions by a factor of 1/1.2; 1,2- and 1,3-electrostatic interactions are neglected, while all others are unscaled. q_1 and q_2 are the Mulliken partial charges placed on the atoms and R^{pairs} is the distance between them.

While vdW energies were calculated as follows:

$$E_{vdW} = \sum_{pairs} s_{pairs}^{vdW} \left[\frac{A_{pairs}}{r_{pairs}^{12}} - \frac{B_{pairs}}{r_{pairs}^{6}}\right] \qquad \textbf{\textit{Eq. 2.16}}$$

The vdW scale factor, s_{pairs}^{vdW}, scales 1-4 interactions by 0.5, while 1,2- and 1,3-interactions are neglected and all others are unscaled. A_{pairs} and B_{pairs} are taken from the standard Amber nonbonding parameters.

A script has been written called ONIOM_EDA.py that performs the EDA using both equations described above. An example of the application of this script is shown below:

There are 3 different input files required by ONIOM_EDA.py:

1) PDB files of the structures before (before.pdb) and after (after.pdb) the ligand

binding: These files are used to extract the residue number for each atom.

2) The Gaussian ".log" file of the optimized structure before (before.log) and after (after.log) the ligand binding: These files are used to extract Mulliken charges and atomic positions.
3) The Gaussian ".com" file used for QM/MM calculations for the structures before (before.com) and after (after.com) the ligand binding: these files are used to extract information pertaining to the bonding, atom type assignment, and which layer (high or low) each atom belongs to.

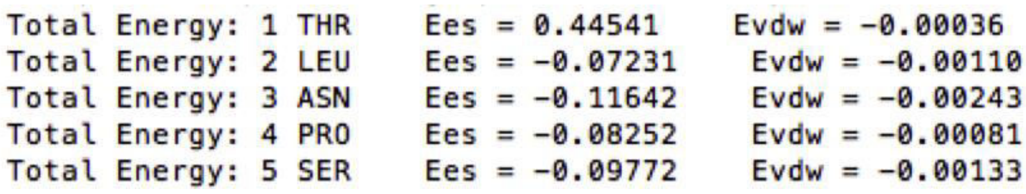
With these 6 files, the next step is to run ONIOM_EDA.py:

```
python ONIOM_EDA.py before.com before.log before.pdb > EDA_before.txt
```

and:

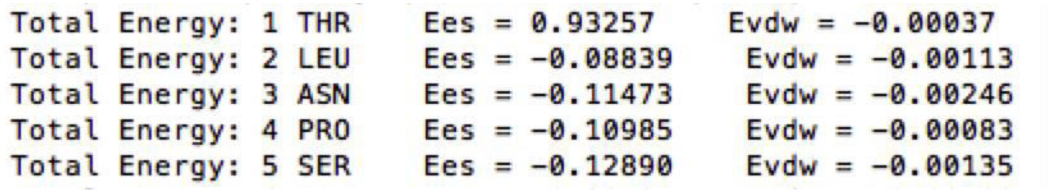
```
python ONIOM_EDA.py after.com after.log after.pdb > EDA_after.txt
```

The two output files "EDA_before.txt" and "EDA_after.txt" allow us to compare electrostatic and vdW energy differences between the unbound and the bound system, respectively.



```
Total Energy: 1 THR     Ees = 0.44541     Evdw = -0.00036
Total Energy: 2 LEU     Ees = -0.07231     Evdw = -0.00110
Total Energy: 3 ASN     Ees = -0.11642     Evdw = -0.00243
Total Energy: 4 PRO     Ees = -0.08252     Evdw = -0.00081
Total Energy: 5 SER     Ees = -0.09772     Evdw = -0.00133
```

Figure 2.2 Electrostatic and vdW interactions of the JMJD2A protein before O_2 binding.



```
Total Energy: 1 THR    Ees = 0.93257    Evdw = -0.00037
Total Energy: 2 LEU    Ees = -0.08839    Evdw = -0.00113
Total Energy: 3 ASN    Ees = -0.11473    Evdw = -0.00246
Total Energy: 4 PRO    Ees = -0.10985    Evdw = -0.00083
Total Energy: 5 SER    Ees = -0.12890    Evdw = -0.00135
```

Figure 2.3. Electrostatic and vdW interactions of the JMJD2A protein after O_2 binding.

Through comparison of the energy differences per residue, we are now able to see the most important residues for O_2 binding in JMJD2A:

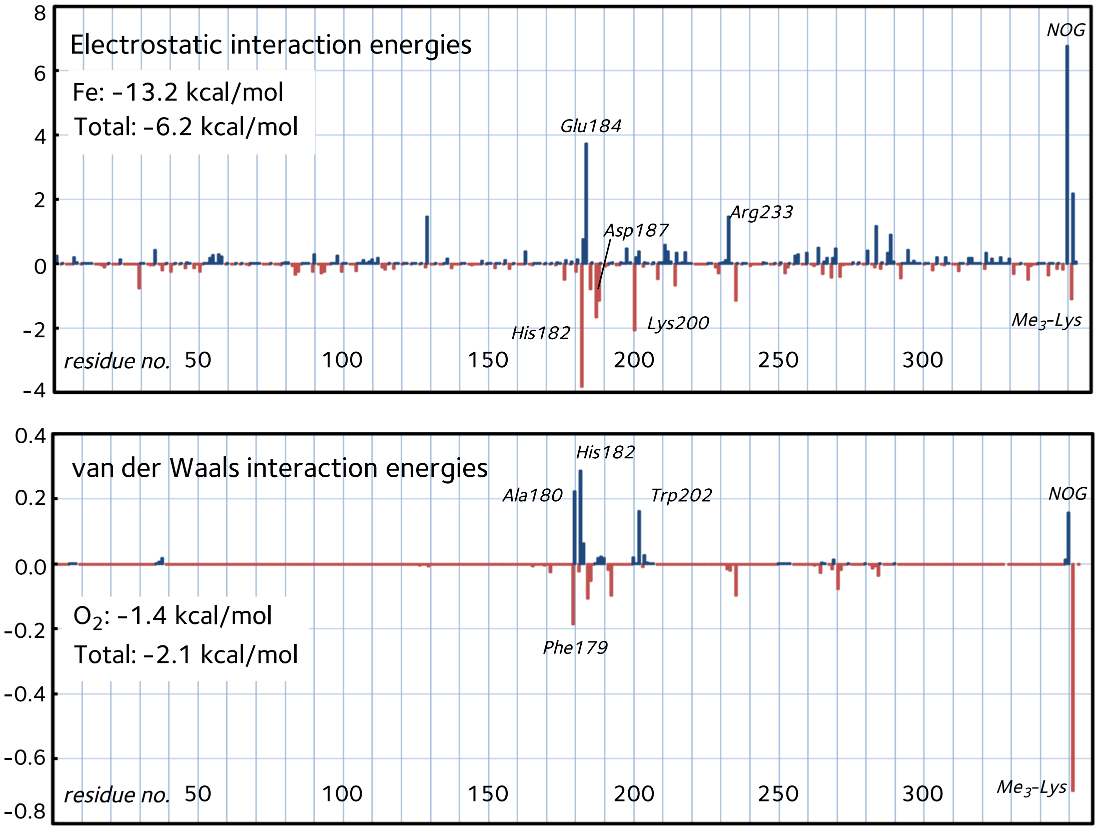


Figure 2.4. Per-residue energy decomposition analysis (kcal mol^{-1}) of O_2 binding in JMJD2A based on a QM/MM description.[20]

2.6.2 Gibbs Free Energy

The objective in this discussion is to calculate the Gibbs Free Energy for the dioxygen binding in the JMJD2A enzyme (PDB ID: 2OQ6) using the free-rotor approximation to estimate the vibrational entropy. The default method used by Gaussian 09[22] is the rigid rotor-harmonic oscillator (RRHO) approximation, for which the vibrational entropy tends to infinity. This leads to potentially large errors associated with numerical noise in the computation of low frequencies, which are poorly described by the harmonic approximation anyway. Grimme has shown that computations of binding thermochemistry are improved by adopting the quasi-RRHO approach in which these low modes are treated as a free rotation rather than a vibration.[23] We implemented Grimme's approach in Python, which can be applied automatically to the output from Gaussian calculations. More details about the partition functions q_x, the entropy S_x, and the energy E_x can be found in the Ochterski's tutorial.[23]

Contributions from translation:

The partition function corresponding to the translation of the molecule is calculated as:

$$q_t = \left[\frac{2\pi m k_b T}{h^2}\right]^{\frac{3}{2}} V \qquad \textbf{\textit{Eq. 2.17}}$$

With T representing temperature, m being the mass, V the volume and h the Planck constant.

Also:

$$R = N_A k_b \qquad \textbf{\textit{Eq. 2.18}}$$

With N_A being the Avogadro constant and R the universal gas constant.

The translational entropy is then calculated as:

$$S_t = R(\ln q_t + 1 + 3/2)$$ ***Eq. 2.19***

And, finally, translational energy is defined as:

$$E_t = 3/2RT$$ ***Eq. 2.20***

Contributions from rotational motion:

For the rotational partition function:

$$q_r = \frac{\pi^{1/2}}{\sigma_r}\left[\frac{T^{3/2}}{\theta_{r,x}\theta_{r,y}\theta_{r,z}}\right]$$ ***Eq. 2.21***

Where

$$\theta_r = \frac{h^2}{8\pi^2 I k_b}$$ ***Eq. 2.22***

With *I* representing the moment of inertia.

The rotational entropy is defined as:

$$S_r = R(\ln q_r + 3/2)$$ ***Eq. 2.23***

Eq. 2.23 is only valid for polyatomic molecules. For monoatomic:

$$S_r = 0$$ ***Eq. 2.24***

For diatomics, such as the dioxygen molecule:

$$S_r = R(\ln q_r + 1)$$ ***Eq. 2.25***

Finally, rotational energy is calculated as follows:

$$E_r = 3/2RT$$ ***Eq. 2.26***

Contributions from vibrational motion:

The vibrational partition function is defined as:

$$q_v = \prod_K \frac{1}{1 - e^{-\frac{\theta_{v,k}}{T}}}$$ ***Eq. 2.27***

With K representing each vibrational mode and:

$$\theta_{v,k} = \frac{hv_k}{k_b}$$

Eq. 2.28

The vibrational entropy is calculated as:

$$s_v = R\sum_K \frac{\frac{\theta_{v,K}}{T}}{\frac{\theta_{v,K}}{T} - 1} - \ln\left(1 - e^{-\frac{\theta_{v,K}}{T}}\right)$$

Eq. 2.29

The above equation represents the vibrational entropy for a harmonic oscillator with frequency v_k. However, if v_k tends to zero, s_v tends to infinity. This script then presents a solution to fix this common issue, by using the quasi-harmonic approximation with a free-rotor that redefines the entropy for a low-frequency mode, as described by Grimme.[24]

$$s'_v = R\left[\frac{1}{2} + \ln\left[\frac{8\pi^3\theta''_v k_b T}{h^2}\right]^{1/2}\right]$$

Eq. 2.30

$$\theta''_v = \frac{\theta'_v B_{AV}}{\theta'_v + B_{AV}}$$

Eq. 2.31

Where B_{AV} is a limiting factor equal to 10^{-34} kg m^2 to very small values of frequencies and:

$$\theta'_v = \frac{h}{8\pi^2 v_k}$$

Eq. 2.32

For a continuous interpolation between the harmonic and the free-rotor approximations, we use the Head-Gordon damping function, also following Grimme:[24]

$$S = f(v_k)S_v + [1 - f(v_k)]S'_v$$

Eq. 2.33

Where:

$$\mathrm{f(v_k)} = \frac{1}{1 + \frac{\mathrm{v_0^4}}{\mathrm{v_k}}}$$ ***Eq. 2.34***

And, finally, the vibrational energy is calculated as:

$$\mathrm{E_v} = \mathrm{R} \sum\nolimits_{\mathrm{K}} \mathrm{e}^{\frac{\theta_{\mathrm{v,K}}}{\mathrm{T}}} \left[\frac{1}{2} + \frac{1}{\frac{\theta_{\mathrm{v,K}}}{\mathrm{T}} - 1}\right]$$ ***Eq. 2.35***

Where:

$$\frac{\mathrm{e}^{\frac{\theta_{\mathrm{v,K}}}{\mathrm{T}}}}{2}$$

is the zero point vibrational energy.

Contributions from electronic motion

The electronic partition function is defined as:

$$\mathrm{q_e} = \omega_0$$ ***Eq. 2.36***

Where ω is the degeneracy of the energy level, and the electronic motion entropy calculated as:

$$\mathrm{S_e} = \mathrm{Rlnq_e}$$ ***Eq. 2.37***

Finally, the electronic motion energy:

$$\mathrm{E_e} = 0$$ ***Eq. 2.38***

Free energy calculations

The total energy is then the sum of the individual contributions:

$$\mathrm{E_{total}} = \mathrm{E_v} + \mathrm{E_e} + \mathrm{E_r} + \mathrm{E_t}$$ ***Eq. 2.39***

The enthalpy is obtained after thermal corrections:

$$\mathrm{H} = \mathrm{E_{total}} + \mathrm{k_bT}$$ ***Eq. 2.40***

And the Gibbs free energy is:

$$G = H - TS_{total}$$ ***Eq. 2.41***

Where:

$$S_{total} = S_v + S_e + S_r + S_t$$ ***Eq. 2.42***

For this work, we used GoodVibes.py to calculate the binding energy of the bound and unbound non-haem active site of JMJD2A.

a) For the active site without O_2:

```
GoodVibes.py freq_before.log -f 50 -t 298 -c 1 -s 1
```

Where f represents the cut-off frequency (cm^{-1}), t being the temperature (K), c for the concentration ($mol\ L^{-1}$) and, finally, s for the scale factor.

b) For the active site with O_2:

```
GoodVibes2.py freq_after.log -f 50 -t 298 -c 1 -s 1
```

c) For the O_2:

```
GoodVibes.py freq_oxygen.log -f 50 -t 298 -c 1 -s 1
```

By comparing the calculated values of G:

$$\Delta G = G_{bound} - (G_{unbound} + G_{dioxygen})$$ ***Eq. 2.43***

It is possible to conclude that free energy corrections contribute to +12.8 kcal mol^{-1} for the dioxygen binding process in JMJD2A.

2.6.3 Pairwise dispersion analysis

The objective with this discussion is to dissect the dispersion energy due to dioxygen binding in the JMJD2A enzyme (PDB ID: 2OQ6). The dispersion effects were taken into account using the D3-dispersion proposed by Grimme with zero-damping at short range[24], using the optimal B3LYP parameters S_6=1.000, SR_6=1.261, S_8=1.703.

$$E_{DFT+dispersion} = E_{DFT} + E_{dispersion} \qquad \textbf{\textit{Eq. 2.44}}$$

Where:

$$E_{dispersion} = -s_6 \sum_{pairs} \frac{C_6^{pairs}}{R_{pairs}^6} f\left(R_{pairs}, sr_6\right) - s_8 \sum_{pairs} \frac{C_8^{pairs}}{R_{pairs}^8} f\left(R_{pairs}\right) \qquad \textbf{\textit{Eq. 2.45}}$$

How to use PairWiseTable.py:

Create a ".txt" file that specifies the pairs of atoms you would like to calculate the dispersion energies for. Open the ".log" file and save as a ".com" file. Then, use the command line:

```
python pairwiseTable.py file.com example.txt -d -p S6 S8 S9 SR6 SR8 SR9
```

By using the script described above, it is possible to find that the total dispersion energies between the dioxygen and the active site of JMJD2A is close to -5 kcal mol^{-1}.

2.7 References for Chapter 2

1. Singh, U. C.; Kollman, P. A., An approach to computing electrostatic charges for molecules. *J Comp Chem* **1984,** *5* (2), 129-145.

2. Hopfinger, A. J.; Pearlstein, R. A., Molecular mechanics force-field parameterization procedures. *J Comp Chem* **1984,** *5* (5), 486-499.

3. (a) Wang, L. P.; Martinez, T. J.; Pande, V. S., Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J Phys Chem Lett* **2014,** *5* (11), 1885-91; (b) Coveney, P. V.; Wan, S., On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys Chem Chem Phys* **2016**, *18* (44), 30236-30240.

4. Max Born, J. R. O., *Nuclear and Electronic Motion*. Annalen der Physik: **1927**, *84*.

5. Slater, J. C., The Theory of Complex Spectra. *Phys Rev* **1929,** *34* (10), 1293-1322.

6. Pauli, W., Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik* **1925,** *31* (1), 765-783.

7. (a) Hartree, D. R., 1 - The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion A2 - MARCH, N.H. In *Self-Consistent Fields in Atoms*, Pergamon: **1975**, 167-193; (b) Hartree, D. R., The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math Proc Camb Phil Soc* **1928,** *24* (01), 89-110.

8. (a) Marx, D.; Hutter, J., *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, **2009**; (b) Leach, A., *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, **2001**.

9. Fock, V., Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **1930,** *61* (1), 126-148.

10. Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, **1989**.

11. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev* **1965,** *140* (4A), A1133-A1138.

12. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys Rev* **1964,** *136* (3B), B864-B871.

13. (a) Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **1996,** *77* (18), 3865-3868; (b) Perdew, J. P.; Burke, K.; Ernzerhof, M.,

Generalized gradient approximation made simple (vol 77, pg 3865, 1996). *Phys Rev Lett* **1997,** *78* (7), 1396-1396.

14. (a) Becke, A. D., Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J Chem Phys* **1993,** *98* (7), 5648-5652; (b) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter* **1988,** *37* (2), 785-789; (c) 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* (8), 1200-1211.

15. Staroverov, V. N.; Scuseria, G. E.; Tao, J. M.; Perdew, J. P., "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes". *J Chem Phys* **2004,** *121* (22), 11507-11507.

16. Zhao, Y.; Truhlar, D. G., 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* (1), 215-241.

17. Miertus, S.; Scrocco, E.; Tomasi, J., Electrostatic Interaction of a Solute with a Continuum - a Direct Utilization of Abinitio Molecular Potentials for the Prevision of Solvent Effects. *Chem Phys* **1981,** *55* (1), 117-129.

18. Dapprich, S.; Komáromi, 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 1 Dedicated to Professor Keiji Morokuma in celebration of his 65th birthday. 1. *J Mol Struct-theochem* **1999,** *461* (1-2), 1-21.

19. Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J., Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J Chem Theo Comp* **2006,** *2* (3), 815-826.

20. Cortopassi, W. A.; Simion, R.; Honsby, C. E.; Franca, T. C.; Paton, R. S., Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem Eur J* **2015,** *21* (52), 18869.

21. Hirao, H., The effects of protein environment and dispersion on the formation of ferric-superoxide species in myo-inositol oxygenase (MIOX): a combined ONIOM(DFT:MM) and energy decomposition analysis. *J Phys Chem B* **2011,** *115* (38), 11278-85.

22. Frisch, M. J. et al, Gaussian 09, Revision B.01. Wallingford CT, **2009**.

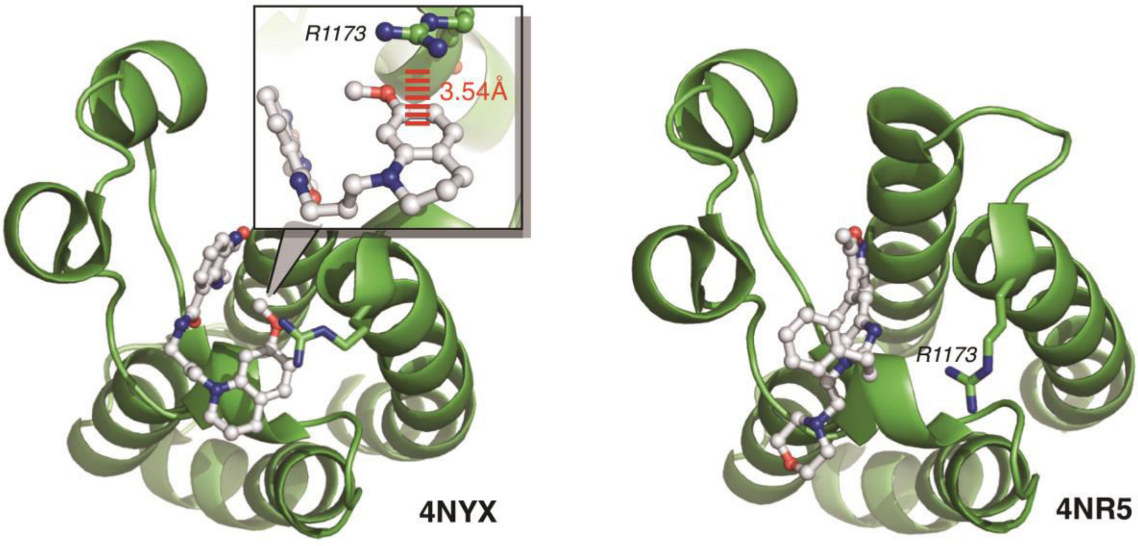
23. Ochterski, J. W. Thermochemistry in Gaussian. Acessed in 02/11/2016. http://www.gaussian.com/g_whitepap/thermo.htm (accessed 02/11/2016).

24. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011,** *32* (7), 1456-65.

Chapter 3

The opening of the active site of CREBBP by *(R)*-2 and a 7-membered ring analogue

“We are more often treacherous through weakness than through calculation.”
François VI de la Rochefoucault



Rooney TP, Filippakopoulos P, Fedorov O, Picaud S, Cortopassi WA, Paton RS, Conway SJ et al, A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation-π interaction, *Angew Chem Int Ed* **2014**, 53, 6126-30.

3.1 Overview

A series of dihydroquinoxalinone (DHQ) derivatives synthesized by Conway and co-workers were found to be potent inhibitors of the epigenetic molecular target CREBBP bromodomain by mimicking acetylated-lysine, resulting in the first potent inhibitors of a bromodomain outside the bromo and extra terminal (BET) family.[1] Using docking studies, Quantum Mechanics (QM) and Molecular Dynamics (MD) simulations, we were able to evaluate important interactions between these derivatives and the active site of the CREBBP. This allowed us firstly to identify the binding site of the DHQ derivatives, and consequently to determine the strength and stability of the most important protein-ligand interactions.

Enantiopure compounds (*R*)-**2**, (*R*)-**5**, (*R*)-**1** and (*R*)-**6** (**Figure 3.1**) were designed as inhibitors of the CREBBP bromodomain, with (*R*)-**2** showing the highest biological activity and being able to inhibit the binding of CREBBP to chromatin in U2OS cells.[1] We then turned our attention to the analyses of important interactions between (*R*)-**2** and the binding site of the CREBBP bromodomain. Computational simulations confirmed that the crystallographically observed cation-π interaction is an important component of ligand binding in the solution phase. The establishment of this interaction upon the binding of (*R*)-**2** mediates the opening of the binding site. An arginine residue (R1173) that is located near the protein-solvent interface is repositioned parallel to the tetrahydroquinoline ring of (*R*)-**2**, a phenomenon that is not observed in the crystal structure of CREBBP bromodomain bound to acetylated lysine.

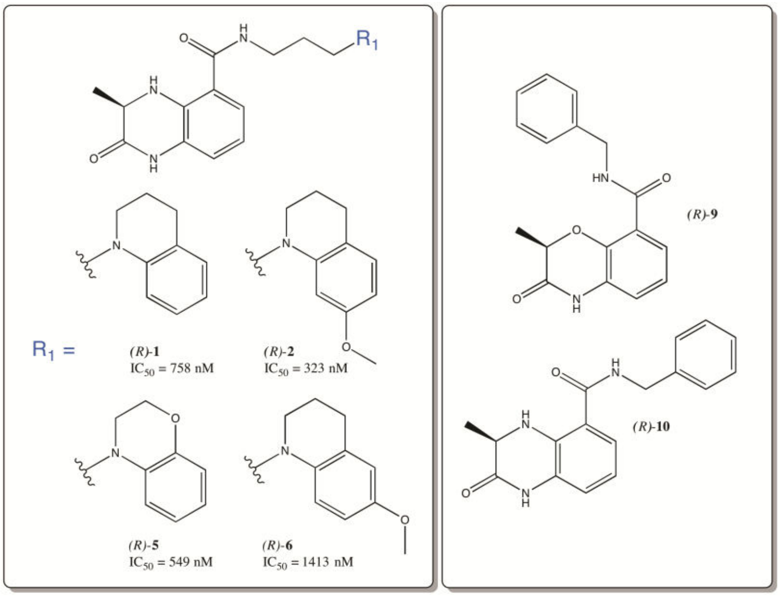


Figure 3.1. Chemical structures of compounds *(R)*-**1**, *(R)*-**2**, *(R)*-**5** and *(R)*-**6** and their respective AlphaScreen Peptide Displacement Assay data for inhibition of CREBBP.[1] *(R)*-**9** and *(R)*-**10** are model compounds for computational analysis of conformational stability in benzoxazinone and DHQ derivatives, respectively.

3.2 Methodology for docking and MD simulations

The structure of *(R)*-**2** was extracted from the Protein Data Bank (PDB Code: 4NYX)[1] and submitted to conformational search using DFTBA methods with Gaussian 09 revision D.[2] This was performed using an *in*-house Python implementation of the Multiple Minimum Monte Carlo (MMMC) search from Macromodel, to trigger external Gaussian optimizations. All low-energy conformations of *(R)*-**2** were then submitted to docking calculations. The 7-membered analogue (**Figure 3.9**), a benzodiazepinone derivative, was obtained after modifications of *(R)*-**2** structure and the same methodology was used to obtain its conformers.

AutoDock Vina 1.1.2[3] was used for docking of *(R)*-**2** to the protein structure. The size of the docking grid was 13.3 Å × 16.7 Å × 14.5 Å, centred on the inhibitor, with all other parameters maintained as default values. The low value of the root-mean-square deviation (RMSD), equal to 0.89 Å, between the pose with the lowest docking energy and the X-ray crystal structure of *(R)*-**2** suggested that this starting point is in accordance with structural data available from experiment.

Further geometry optimization of the ligand was done in the B3LYP-D3/6-31+G(d) level of theory using Gaussian 09 revision D. The R.E.D. (RESP and ESP Derivation) server version 4.0[4] was used: 1) for geometry optimization of ligand (*R*)-**2** and the 7-membered ring analogue and 2) for RESP charge derivation, using the methodology previously proposed by Duan and coworkers.[5] Molecular electrostatic potentials (MEP) derived from the Connolly Surface Algorithm were computed with the B3LYP exchange and correlation density functional with a triple-ζ valence polarized cc-pVTZ basis set,[4,6] employing an IEFPCM continuum solvation model to describe the relatively hydrophobic protein site (diethylether, $\rho = 4$). The electrostatic potentials of the inhibitor were saved and used in the charge fitting process.

Molecular mechanics parameters for inhibitor (*R*)-**2** and the 7-membered analogue were taken from the General Amber Force Field (GAFF)[6] using the RESP atomic charges previously obtained. Hydrogens were removed from amino acid residues using the *MolProbity* Server[7] and added using *tLeap*.[7] Glutamate and aspartate residues were assigned as negatively charged and lysine and arginine as positively charged. Minimization and MD calculations were performed using the Amber Force Field 99SB within Amber version 12. Crystallographic waters were not removed, while the protein was further solvated by a box of TIP3P water molecules.[8] Simulations were carried out in cubic boxes. Two counter-cations (Na^+) were added to equilibrate the system. Energy minimization was performed in two steps. Firstly, we used steepest descent followed by conjugate gradients with position of the inhibitor restrained (PR). Secondly, the same minimization methodology was performed without PR. The minimized complex was then subjected to 500 ps of equilibration and 120 ns of production MD simulation in the NPT ensemble using a Langevin thermostat to simulate a constant temperature at 310 K ($\tau_T = 0.1$ ps). Isotropic position scaling was used to maintain the pressure of 1 atm ($\tau_p = 2$ ps). MD simulation was carried out using 1 fs of integration time and a nonbonding cut-off of 8 Å, with the Shake algorithm turned on to constrain bonds involving hydrogen. Snapshots were obtained at intervals of 20 ps in producing plots of the geometric variation during the simulation. Trajectory visualization and graphical analysis were generated using Visual Molecular Dynamics (VMD).[9]

3.3 Results and discussion

The coordinates of (*R*)-**2** were taken from the crystal structure and Density Functional Theory (DFT) calculations were performed at the CPCM-wB97XD/TZVP level of theory to search for the lowest energy conformation in solution. A model of (*R*)-**2**, here named *(R)*-**10** – **Figure 3.1**, was used in our calculations. It is possible to see that the most stable conformer in solution presents the hydrogen bond between the -NH of the dihydroquinoxalinone and the amide carbonyl oxygen atom, with a $C_{carbonyl}$–C_{aryl} dihedral angle of $\theta = 30°$ and 330°. The crystal structure of (*R*)-**2** bound to CREBBP (PDB ID: 4NYX) shows an angle of $\theta = 2°$, with the same hydrogen bond as shown in our QM studies. Curiously, the crystal structure of a benzoxazinone derivative, *(R)*-**3**, presented an internal hydrogen bond between the -NH of the amide and the oxygen of the benzoxazinone that directs the side chain of this compound in the opposite direction compared to *(R)*-**2**, resulting in clashes with the protein environment and water molecules from the ZA water channel. This would explain the low affinity of *(R)*-**3**. When calculating the lowest energy conformation of *(R)*-**3** by using *(R)*-**9** as a model (**Figure 3.1**), we found a $C_{carbonyl}$–C_{aryl} dihedral angle of θ = 210°, as observed in the crystal structure.

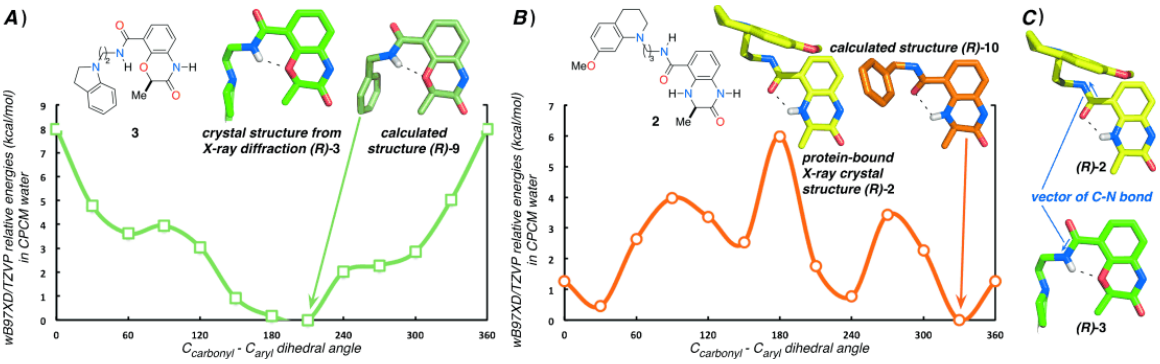


Figure 3.2. A plot of CPCM-ωB97XD/TZVP relative energetics (kcal mol^{-1}) in implicit water for rotation about the $C_{carbonyl}$–C_{aryl} bond of the model compounds A) *(R)*-**9** (green) and B) *(R)*-**10** (orange). C) The reversed internal hydrogen bond changes the amide bond vector in benzoxazinone and dihydroquinoxalinone analogues (blue arrows).[1]

Additional calculations of *(R)*-**2** were performed in a cubic 12 Å x 12 Å x 12 Å box of TIP3P water molecules using the General Amber Force Field (GAFF) with RESP atomic charges. These 500 ps MD simulations showed that the hydrogen bond previously observed in our DFT calculations were also present in explicit solution (**Figure 3.3**).

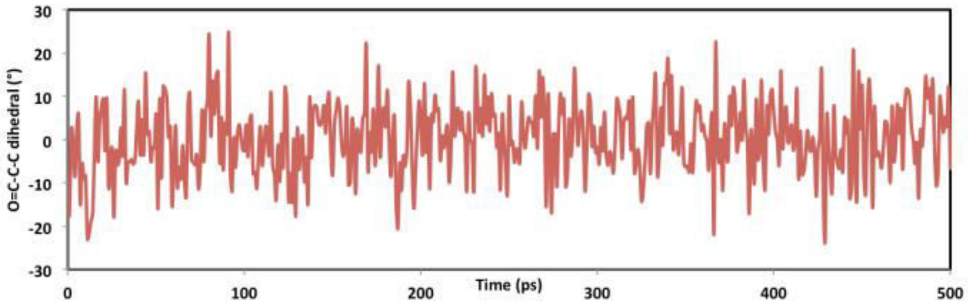


Figure 3.3. Impact of solvent on the conformational bias of inhibitor *(R)*-**2**. To verify the conformational preference of dihydroquinoxalinone amides obtained from DFT in implicit solvent (**Figure 3.2**, *(R)*-**10**), we performed a molecular dynamics simulation using the full structure of *(R)*-**2** (500 ps) in explicit water.[1]

To evaluate the strength of the cation-π interaction, a truncated model was used consisting of the trimethyl-guanidinium of the CREBBP arginine R1173 and the side chain of the DHQ derivative *(R)*-**2** (**Figure 3.4**). The energy was computed at the counterpoise and dispersion corrected B3LYP-D3/aug-cc-pVTZ as well as M06-2X/ aug-cc-pVTZ level of theory with the SMD[10] solvation model (diethylether, ρ = 4). The counterpoise-correction reduces the effects of basis set superposition error (BBSE), which can overestimate complexation energies. The use of the M06 functional was here considered since it takes into account weak dispersive interactions by including the kinetic energy density in the exchange correlation functional. Coordinates of non-hydrogen atoms were fixed, while hydrogens were allowed to move, since they are not well-defined crystallographically. The complexation energy was calculated to be -3.2 kcal mol^{-1} and -4.6 kcal mol^{-1} using the M06-2X and B3LYP-D3 functionals, respectively, and are in good agreement with previous reported energies for cation-π interactions involving either lysines (-3.3 ± 1.5 kcal mol^{-1}) or arginines (-2.9 ± 1.4 kcal mol^{-1}).[11] By way of comparison, the intermolecular hydrogen-bond between two water molecules is estimated in the gas-phase as 4.7-5.0 kcal mol^{-1}. Therefore, the cation-π interaction here is predicted to be comparable in magnitude.

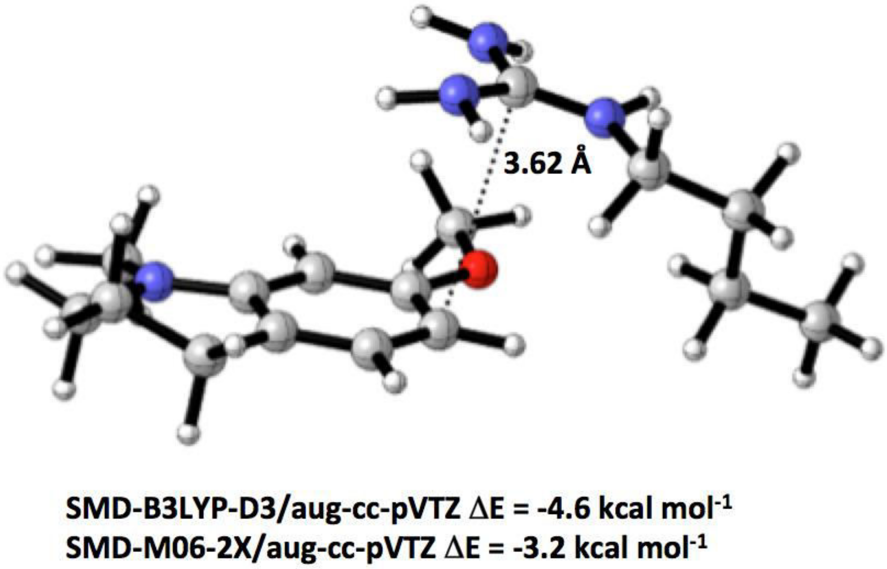


Figure 3.4. Counterpoise and dispersion corrected complexation energies of a truncated model of *(R)*-**2** in CREBBP at the B3LYP-D3/aug-cc-pVTZ and M06-2X/ aug-cc-pVTZ levels of theory.[1]

Our calculations showed that the presence of intra and intermolecular hydrogen bonds were maintained during the course of the MD simulation, and resulted in the stability of the ligand-protein complex. Two hydrogen bonds formed with an asparagine residue N1168 (**Figure 3.5**), mimicking the acetylated lysine recognition in most of the bromodomains, and an intramolecular hydrogen bond between the –NH of the DHQ ring and the oxygen of the amide (**Figure 3.6**) were largely maintained during the simulation length indicative of a relatively strong ligand stability in the binding site.

The formation of a cation-π interaction between the arginine residue R1173 and the tetrahydroquinoline ring of (*R*)-**2** (**Figure 3.7**) was also observed, revealing an induced fit-pocket that is maintained during 100 ns of molecular dynamics (MD) simulation. Applying a distance cut-off of 6 Å, as proposed by Dougherty,[12] we found that this cation-π interaction is present for more than 40% of the simulation time, suggesting that this interaction is indeed present (and important) in the solution phase. At the beginning of the MD trajectory in explicit water, R1173 is oriented away from *(R)*-**2** into bulk solvent such that the distance between the centre of mass of its side chain and the guanidinium central C atom is around 12 Å. However, during the course of the 100 ns MD simulation, the cation-π interaction is observed to form (**Figure 3.7**). The conformation of the guanidinium was also evaluated in the presence of the ligand, as proposed by Kaznessis.[13] Two planes were defined: one containing the guanidinium atoms and another one containing the aromatic ring of the DHQ derivative *(R)*-**2**. An angle between the two planes of 0-45° and 135-180° was considered in the parallel conformation, otherwise, it was considered T_{shaped} conformation. During the 100 ns of the MD, it is possible to observe this interaction is predominantly in the parallel conformation (approximately 70% of the simulation), corroborating the crystal structure.

The average Root Mean Square Deviation (RMSD) was stabilized at around 1.2 Å during the simulation, while temperature, volume and water density were almost constant at around 310 K, 53,000 $Å^3$ and 1.1 g cm^{-3}, respectively. The highest fluctuations of the residues (**Figure 3.8**) correspond to the ZA-loop, BC-loop, and termini, as also established by Caflisch for the CREBBP/acetyl-lysine complex during repeated μs time scale MD simulation in explicit solvent.[14]

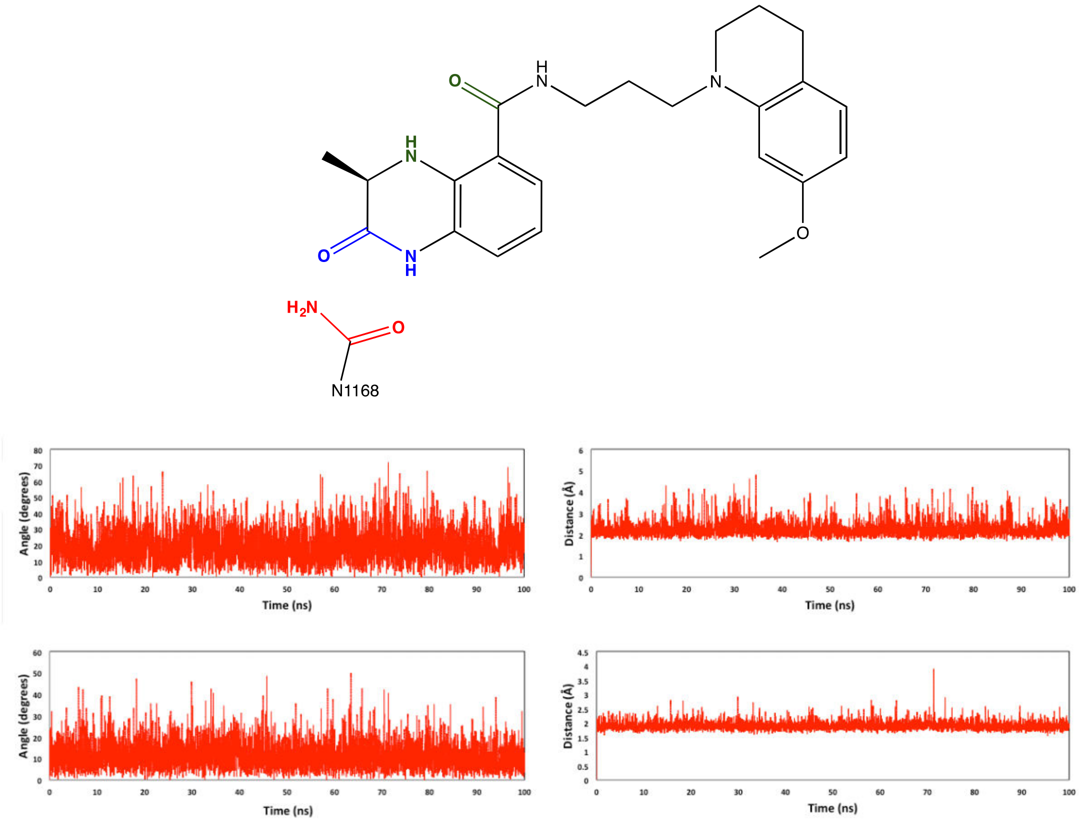


Figure 3.5. Angle (°) and distance (Å) between the **–NH** of the ligand and the **O** atom of Asn1168 (top), and between the **–NH** of N1168 and the **O** atom of the ligand (bottom).[1]

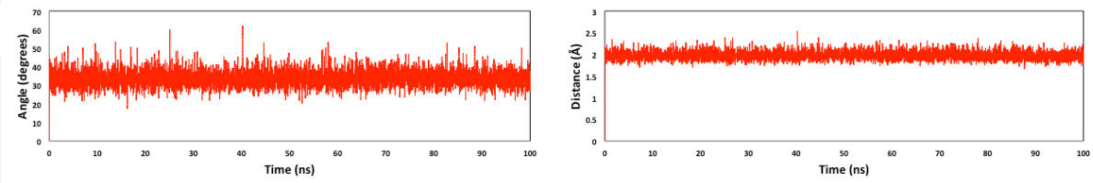


Figure 3.6. Angle (°) and distance (Å) between the **–NH** and the **O** atom of the intramolecular hydrogen bond during simulation of *(R)*-**2** bound to CREBBP.[1]

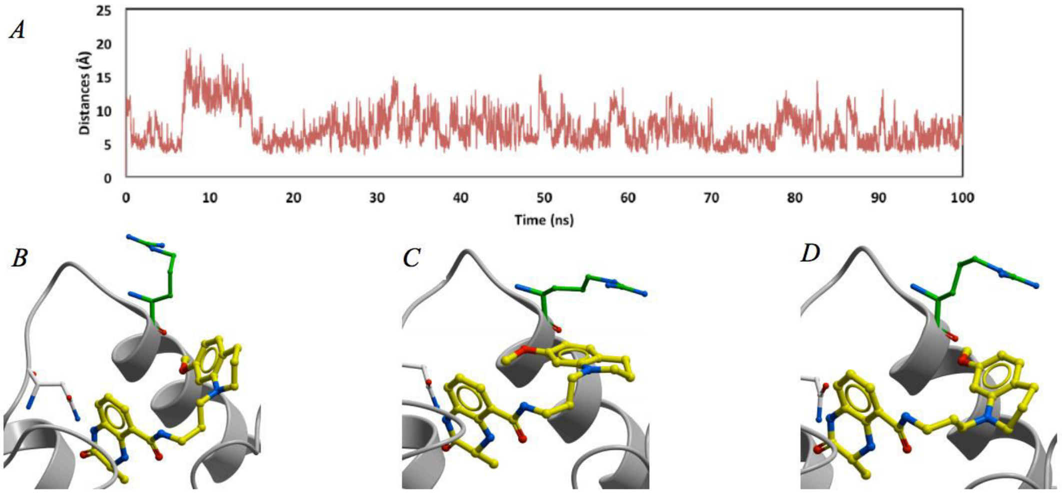


Figure 3.7. A) Distance (Å) between the guanidium C atom in R1173 and the centre of mass of the aromatic ring of *(R)*-**2** during the MD simulation with the CREBB bromodomain (PDB ID: 4NYX). (B) Cation-π interaction formed during the MD. (C) Parallel cation-π interaction. (D) T_{shaped} cation-π interaction.[1]

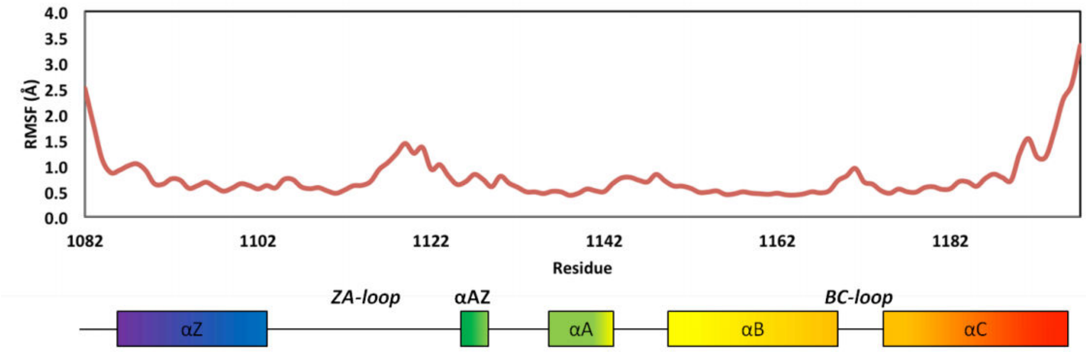


Figure 3.8. Per-residue Root Mean Square Fluctuations (RMSF) of the Cα atoms of the protein.[1]

Apart from evaluating the stability of crystal structures containing DHQ derivatives with the Amber force field,[15] the opening of the active site by (*R*)-**2** and a 7-membered ring analogue was also investigated starting from the apoenzyme structure (PDB ID: 3DWY),[16] to analyse how this behaviour is related to the nanomolar biological activities reported by Conway.[17] Two independent 120 ns MD simulations were performed using the Amber ff99SB force field (a more detailed description of the MD methodology can be found in the **Section 3.2**). The 6-membered ring compound (*R*)-**2** can easily aromatise upon 2-electron oxidation and loses the important amide functionality of the DHQ system. The 7-membered analogue (**Figure 3.9**) is less prone to oxidation since there is no driving force towards aromaticity, making this compound metabolically more stable, and thus a more suitable candidate for CREBBP inhibition *in vivo*.

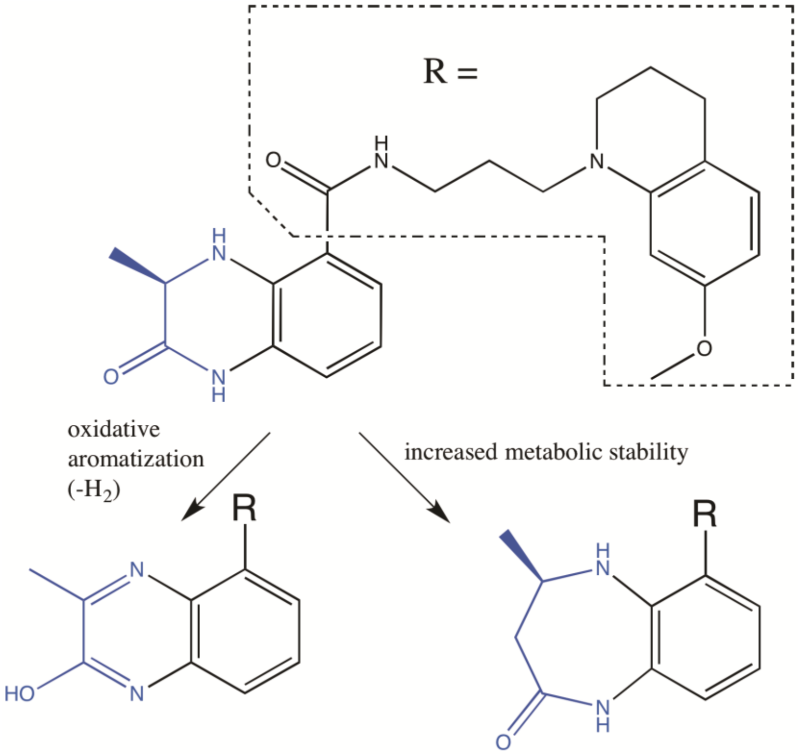


Figure 3.9. Replacement of the 6-membered ring of the DHQ derivative *(R)*-**2** by a benzodiazepinone derivative.

MD simulations with the 7-membered ring analogue and *(R)*-**2** in CREBBP stabilized after 40 ns and 20 ns, respectively, at around 2 Å (**Figure 3.10**). It was possible to observe that both ligands *(R)*-**2** and the 7-membered ring analogue were able to open the active site upon formation of the cation-π interaction (**Figure 3.11**). In addition, important interactions between *(R)*-**2** and the CREBBP bromodomain maintained during our calculations are similar to the ones observed in the crystal structure with this ligand bound (PDB ID: 4NYX). The 7-membered ring analogue was predicted to have a similar binding mode than *(R)*-**2**, including the intramolecular hydrogen bond and other two intermolecular hydrogen bonds with the asparagine residue N1168 (**Figures 3.10 and 3.11**). It was also predicted the cation-π interaction between this ligand and the arginine R1173, as previously reported in the literature for *(R)*-**2**.[1]

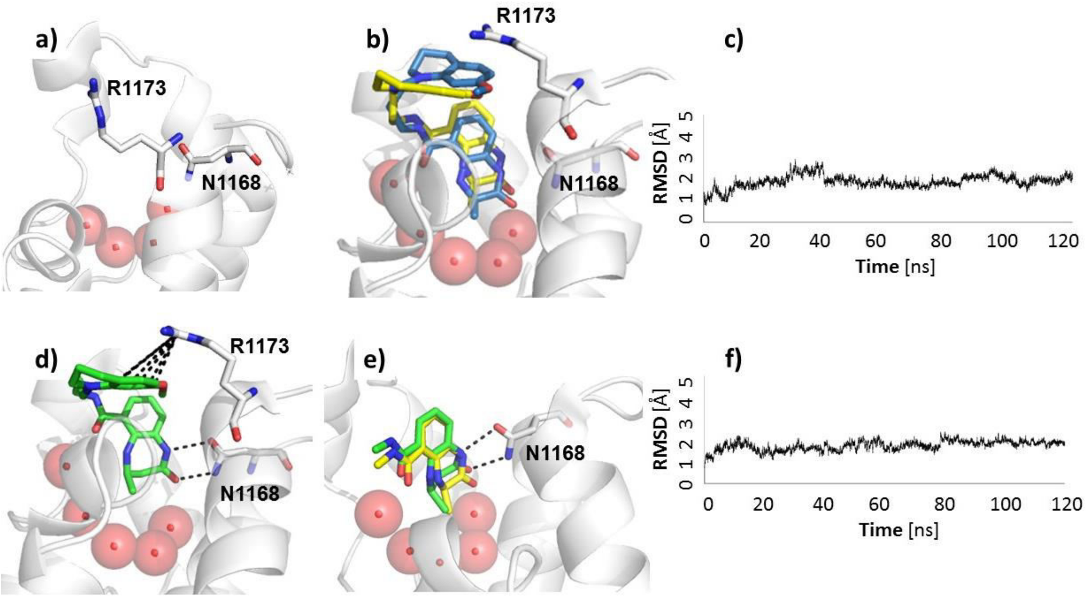


Figure 3.10. The *apo* crystal structure of the CREBBP bromodomain (PDB ID: 3P1C) highlighting the key amino acids N1168 and R1173 (carbon = grey). (a) The conserved water molecules in the binding pocket are shown as red spheres. (b) Overlaid structures of the MD (carbon = blue) and crystal structure of (*R*)-**2** (carbon = yellow). Root mean square deviation (RMSD) over 120 ns for 7 membered analogue (c) and (*R*)-**2** (f). (d) Simulated structure of 7-membered analogue by MD (carbon = green) and the overlaid 6-membered (carbons = yellow) and 7-membered (carbon = green) DHQ headgroup are shown (e). Thanks to Dr. Michael Brand for kindly allowing the re-use of this Figure from his thesis (Dr. Michael Brand, DPhil thesis, Oxford University, 2017).

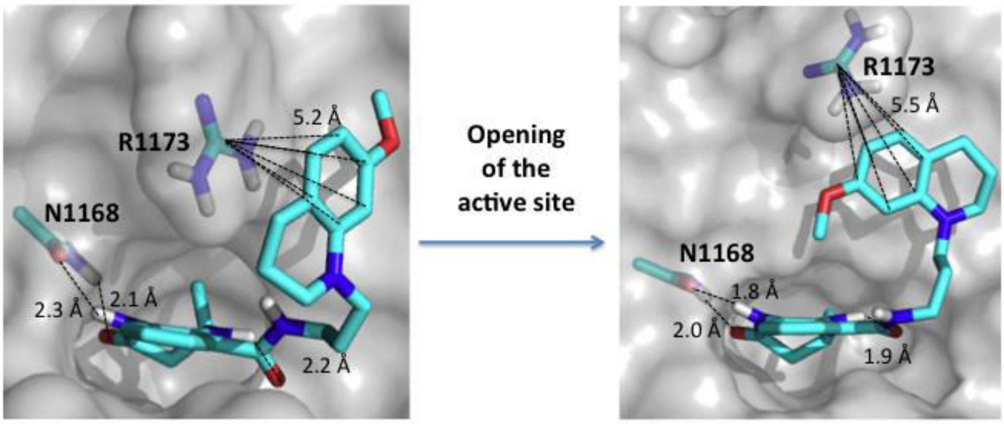


Figure 3.11. Predicted opening of the active site of CREBBP by the 7-membered ring analogue. On the left, complex structure following docking of the ligand in the binding site of the CREBBP bromodomain (PDB ID: 3DWY). On the right, ligand/protein complex obtained following 120 ns MD simulation.

3.4 Conclusion

DFT studies to investigate the conformational preference of several DHQ derivatives were combined with MD simulations to investigate their binding to the CREBBP bromodomain. These computational techniques were used to understand the ability of the DHQ derivative *(R)*-**2** to create an induced-fit pocket upon binding. It is shown in this Chapter that this behaviour is driven by the formation of a cation-π interaction in the CREBBP brodomomain binding site following 100 ns MD simulation. This interaction was maintained during the MD simulation and corroborates the crystal structure of the CREBBP bromodomain with this ligand bound. DFT calculations show that this interaction is of a comparable magnitude to *e.g.* hydrogen bonding interactions regularly considered as important for protein:ligand binding. Intermolecular hydrogen bonds with an asparagine residue N1168 mimic the native substrate, while an intramolecular hydrogen bond is also an important source of conformational stability for this compound in the protein complex. Then similar modelling techniques were applied in a predictive fashion: a 7-membered ring compound, a benzodiazepinone derivative, was considered for further improvements in the stability of the inhibitor, since this compound would be less prone to oxidation than its 6-membered ring analogue *(R)*-**2**. Our MD simulations showed that changing the DHQ ring size would not affect the ability of the ligand to bind in the CREBBP bromodomain. However, given its better pharmacological properties, the 7-membered ring analogue is suggested to be a more potential CREBBP inhibitor.

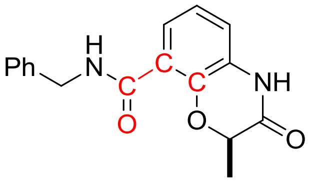
3.5 Supporting Information for Chapter 3

Supporting Table S1: Coordinates of the truncated model used for evaluation of the strength of the cation-π interaction energy.

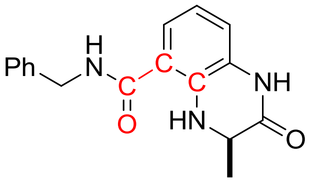
Atom	**X**	**Y**	**Z**
C(Fragment=2)	11.20700	38.51700	20.37700
C(Fragment=2)	10.77900	39.84200	19.75100
C(Fragment=2)	11.28501	41.08801	20.47401
C(Fragment=2)	11.03604	42.35093	19.69004
C(Fragment=2)	11.74906	43.16897	17.44897
N(Fragment=2)	10.86301	44.14001	17.48401
N(Fragment=2)	11.86589	42.34809	18.47600
N(Fragment=2)	12.56700	43.06700	16.42400
O(Fragment=1)	10.40500	40.46500	15.74100
N(Fragment=1)	7.78700	43.50200	13.08300
C(Fragment=1)	8.35300	43.09700	11.82900
C(Fragment=1)	6.59600	44.35000	12.97600
C(Fragment=1)	6.34600	45.12400	14.24900
C(Fragment=1)	6.21500	44.11100	15.36500
C(Fragment=1)	7.33100	43.15300	15.42800
C(Fragment=1)	7.59800	42.47200	16.57500
C(Fragment=1)	8.62199	41.54100	16.64998
C(Fragment=1)	9.39501	41.31000	15.56000
C(Fragment=1)	9.13100	41.99101	14.35101
C(Fragment=1)	8.12100	42.89800	14.27700
C(Fragment=1)	11.27900	40.22100	14.70400
H(Fragment=2)	9.68233	39.88944	19.71001
H(Fragment=2)	11.10333	39.86220	18.70109
H(Fragment=2)	10.79578	41.18557	21.44912
H(Fragment=2)	12.35998	41.00239	20.69300
H(Fragment=2)	11.26766	43.24319	20.28517
H(Fragment=2)	9.98368	42.39465	19.38212
H(Fragment=2)	12.67436	41.74099	18.49601
H(Fragment=2)	10.22423	44.20254	18.26145
H(Fragment=2)	10.49098	44.49951	16.61495
H(Fragment=2)	13.05913	42.20479	16.24563
H(Fragment=2)	12.50152	43.70923	15.64906
H(Fragment=2)	10.86159	38.43327	21.41291
H(Fragment=1)	12.03275	39.52124	15.07356
H(Fragment=1)	10.78835	39.76182	13.83483
H(Fragment=1)	11.78219	41.13825	14.35438
H(Fragment=1)	9.72031	41.76308	13.47369
H(Fragment=1)	8.81177	40.98001	17.55989
H(Fragment=1)	6.97564	42.65061	17.44947
H(Fragment=1)	6.09950	44.61761	16.33227
H(Fragment=1)	5.26911	43.56648	15.21908
H(Fragment=1)	7.18533	45.80610	14.43476
H(Fragment=1)	5.44029	45.72882	14.15252
H(Fragment=1)	6.75028	45.03561	12.13711
H(Fragment=1)	5.70955	43.74162	12.73164
H(Fragment=1)	9.44476	43.22145	11.80967
H(Fragment=1)	8.13620	42.04477	11.58657
H(Fragment=1)	7.94587	43.71427	11.02860
H(Fragment=2)	10.79372	37.67294	19.81885
H(Fragment=2)	12.29790	38.41308	20.38368

Supporting Table S2: Conformational energy profiles for a simplified model benzoxazinone and dihydroquinoxalinone. CPCM-wB97XD/TZVP//wB97XD/TZVP energies are calculated for rotation about the O=C-C=C dihedral angle for molecules (*R*)-**9** and (*R*)-**10**.

Dihedral Angle (°)	(*R*)-**9** E (Hartree)	(*R*)-**10** E (Hartree)
0	-992.6666449	-972.8184132
30	-992.6717798	-972.8204279
60	-992.6736122	-972.8168118
90	-992.6730923	-972.8149650
120	-992.6745265	-972.8192018
150	-992.6779346	-972.8176322
180	-992.6791168	-972.8108877
210	-992.6793833	-972.8164043
240	-992.6761782	-972.8150656
270	-992.6757483	-972.8140823
300	-992.6748365	-972.8162256
330	-992.6713736	-972.8197029
360	-992.6666449	-972.8184132



(*R*)-**9**



(*R*)-**10**

3.6 References for Chapter 3

1. Rooney, T. P.; Filippakopoulos, P.; Fedorov, O.; Picaud, S.; Cortopassi, W. A.; Hay, D. A.; Martin, S.; Tumber, A.; Rogers, C. M.; Philpott, M.; Wang, M.; Thompson, A. L.; Heightman, T. D.; Pryde, D. C.; Cook, A.; Paton, R. S.; Muller, S.; Knapp, S.; Brennan, P. E.; Conway, S. J., A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation-pi interaction. *Angew Chem Int Ed* **2014,** *53* (24), 6126-30.

2. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision B.01. Wallingford CT, **2009**.

3. Trott, O.; Olson, A. J., Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **2010,** *31* (2), 455-461.

4. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y., R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **2011,** *39* (Web Server issue), W511-7.

5. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **2003,** *24* (16), 1999-2012.

6. 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 Model* **2006,** *25* (2), 247-260.

7. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (1), 12-21.

8. Sun, L.; Siepmann, J. I.; Schure, M. R., Conformation and solvation structure for an isolated n-octadecane chain in water, methanol, and their mixtures. *J Phys Chem B* **2006,** *110* (21), 10519-25.

9. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

10. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* **2009,** *113* (18), 6378-96.

11. Gallivan, J. P.; Dougherty, D. A., Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **1999,** *96* (17), 9459-64.

12. Dougherty, D. A., Cation-pi interactions involving aromatic amino acids. *J Nutr* **2007,** *137* (6), 1504S-1508S; discussion 1516S-1517S.

13. Khandelia, H.; Kaznessis, Y. N., Cation-pi interactions stabilize the structure of the antimicrobial peptide indolicidin near membranes: molecular dynamics simulations. *J Phys Chem B* **2007,** *111* (1), 242-50.

14. Steiner, S.; Magno, A.; Huang, D.; Caflisch, A., Does bromodomain flexibility influence histone recognition? *FEBS letters* **2013,** *587* (14), 2158-63.

15. Case, D. A.; Darden, T. A.; Cheatham, T. E.; 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.; Goetz, 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**.

16. Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S.,

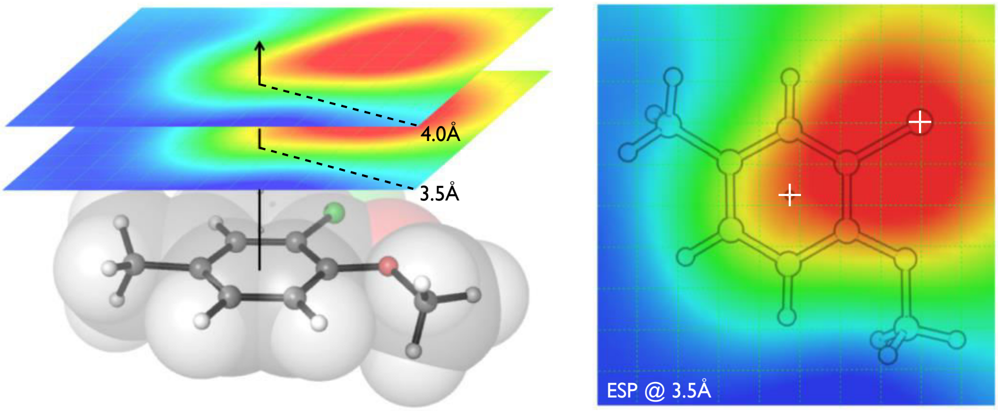
Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012,** *149* (1), 214-31.

17. Shafaat, H. S.; Pantazis, D. A.; Neese, F.; Lubitz, W.; Cox, N.; Griese, J. J.; Graslund, A.; Hogbom, M.; Roos, K.; Siegbahn, P. E., Probing the electronic structure of the heterobimetallic Mn/Fe cofactor of the R2lox proteins using advanced EPR techniques. *J Biol Inorg Chem* **2014,** *19*, S311-S311.

Chapter 4

Cation-π interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity based on QM descriptors

"Before enlightenment — chop wood, carry water. After enlightenment — chop wood, carry water."
Zen Buddhist Proverb



Cortopassi WA, Kumar K, Paton RS. Cation-π interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity, *Org Bio Chem* **2016**, 14, 10926-10938.

4.1 Overview

In the previous chapter, we discussed the importance of hydrogen bond (both intra- and intermolecular) and cation-π intermolecular interactions for the stability of dihydroquinoxalinone derivatives (DHQ) bound to the CREBBP bromodomain. The cation-π interaction formed between the aromatic side-chain of ligand *(R)*-**2** and the R1173 guanidinium group of CREBBP was maintained during 100 ns MD simulation in explicit solvent: counterpoise and dispersion-corrected Density Functional Theory (DFT) calculations estimated the strength of the cation-π interaction to be in the region of 3.2-4.7 kcal mol^{-1}, comparable in strength with some hydrogen-bonds.

Subsequent to our initial collaborative work with the Conway group, Brennan and co-workers (Structural Genomics Consortium, University of Oxford) described the synthesis of a series of selective 5-isoxazolylbenzimidazole derivatives, which form the same interaction as *(R)*-**2** (observed in a bound X-ray structure) in the binding site of CREBBP, and showed selectivity for this target.[1] This interaction is of particular interest from the point of view of optimizing selectivity in bromodomain targeting by small-molecules: compared to other bromodomain containing proteins, such as the Bromo and Extra-Terminal (BET) family, which shares a high degree of sequence homology, this arginine residue is unique to CREBBP. In **Figure 4.1**, it is possible to see close contacts between inhibitors and a conserved arginine residue involved in cation-π interactions as driving forces in binding selectivity for CREBBP bromodomains. This Figure emphasises the overall similarity between bromodomain structures (CREBBP *vs.* BRD4), making selectivity a significant challenge in the development of small-molecule chemical probes. Nonetheless, we also highlight an important difference in side-chain functionality between positively-charged Arg

(R1173) in CREBBP and negatively-charged Asp (D145) in BRD4, which provides the basis for discrimination between the two targets.

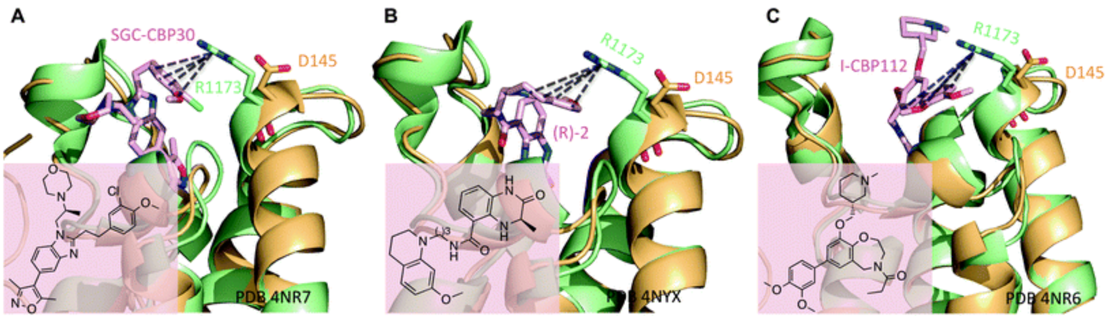


Figure 4.1. Ligand binding in the CREBBP bromodomain (green) over BRD4 (yellow, PDB ID: 3UVW) for a diverse set of inhibitors including (A) 5-isoxazolylbenzimidazole derivative, SGC-CBP30 (**L11**); (B) dihydroquinoxalinone derivative, *(R)-**2***; and (C) benzoaxazepine-based inhibitor, I-CBP112.[2]

This observation raises the prospect of the design of small molecules, which by forming a cation-π interaction, could be both potent and selective CREBBP inhibitors. In this Chapter, we analyse different computational techniques in their ability to correlate binding affinities between fifteen 5-isoxazolylbenzimidazole derivatives (**Chart 4.1**) and the CREBBP: (1) Poisson Boltzmann Surface Area (MM-PBSA) and Generalized Born Surface Area (MM-GBSA) scoring functions that incorporated the entire ligand; (2) QM-complexation energies and (3) Electrostatic Potential Surface values (ESPs). A key motivation in the development of the models discussed in this chapter has been to quantitatively rationalize experimental trends in a manner that allows for rapid and reliable predictions to be made.

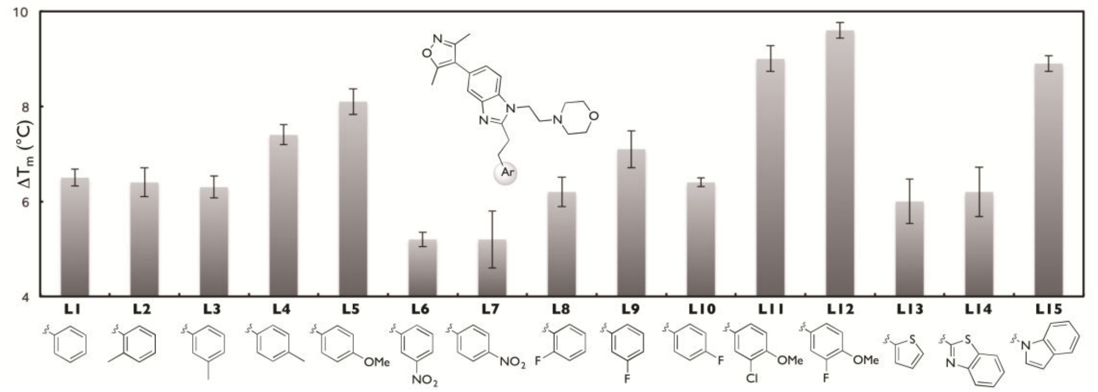


Chart 4.1. CREBBP bromodomain binding assay for fifteen 5-isoxazolylbenzimidazole derivatives with variable aromatic (Ar) groups; ΔT_m data (Thermal Stability Shift Assay - Differential Scanning Fluorimetry) taken from ref[1].[2]

We have constructed a model to predict biological activities based on the electrostatic potential of the aromatic rings of 5-isoxazolylbenzimidazoles and the positively charged arginine. A simple linear model using ESP values captures variations in the cation-π interaction strength, giving the best correlation ($R^2 = 0.84$, $n = 15$) with experimental binding affinities.

4.2 Methodology

QM-complexation energies. QM calculations were performed using Gaussian 09.[3] Coordinates for the guanidinium-inhibitor complex were taken from PDB ID: 4NR7. Only atoms from the aromatic group of the 5-isoxazolylbenzimidazole derivatives were selected for a better description of the cation-π interaction (**Figure 4.2**). Non-hydrogen guanidinium atoms as well as two carbon atoms from the inhibitor were fixed to mimic the protein environment. The positions of all other atoms were optimized with the meta-generalized gradient approximation (GGA) functional TPSS with a triple-ζ valence def2-TZVPP basis set. Boys-Bernadi counterpoise corrections[4] were used for quantification of the complexation energies comparing different DFT descriptions (the hybrid B3LYP and the meta-GGA functionals M06-2X and TPPS) and basis sets. The M06-2X functional of Truhlar and Zhao[5] has been used recently for accurate calculations of $C_6H_5X\cdots Na^+$ binding energies relative to the benchmark coupled cluster singles, doubles and triples (CCSD(T)) values.[6] In all calculations, Grimme's D3 density-independent correction were used for atom-pairwise dispersion with a Becke-Johnson damping function at short range.[7] The use of dispersion corrections with the B3LYP functional is encouraged since this fails to describe London (van der Waals) dispersion Interactions.[8] The inclusion of dispersion energies in cation-π interactions has been evaluated by Kim and co-workers.[9] Although such contribution has a negligible effect on Na^+/π and Li^+/π complexes, since alkali metal cations have very low polarizability, they are important for accurate energy estimation of complexation between organic cations and π systems. In these cases, based on the size of the systems involved and their greater polarizability compared with metal cations, it is anticipated that dispersive forces could play a greater role than in the simpler systems which have previously been studied.

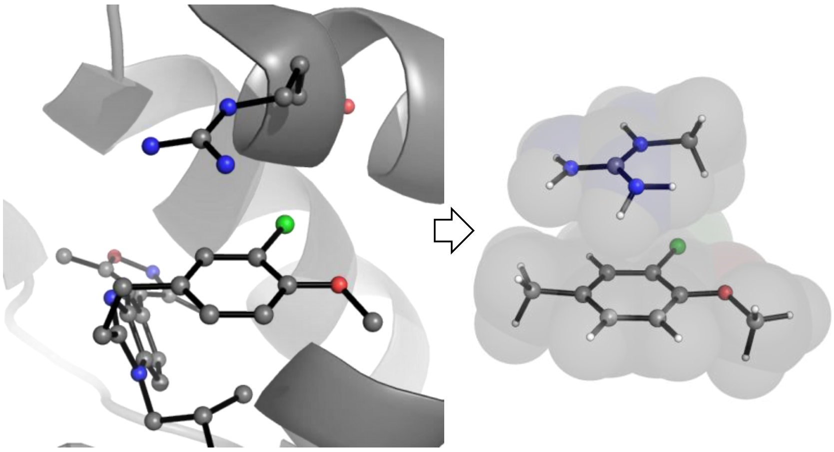


Figure 4.2. Cluster model of the aromatic:R1173 interaction taken from PDB 4NR7 used in QM studies.[2]

Quantitative and qualitative analysis of Electrostatic Potentials. Initial coordinates for the aromatic groups of ligands **L1-L15** were taken from the PDB ID: 4NR7 (**Figure 4.2**), which would take into account the preferred conformation of these molecules in the CREBBP environment. The geometries of the unsaturated substituents were then fully optimized using the B3LYP/6-31G+(d) level of theory. The choice of functional is relatively unimportant for these optimizations, since the isolated substituents' geometries are insensitive, with negligible changes using different methods. The ESPs were computed from B3LYP/6-31G(d,p) densities and compared to M06-2X/6-31G(d,p) analysis. Both qualitatively, and quantitatively, there is negligible difference in the ESPs obtained from both methods, which is to be expected since the isolated systems are well-described by DFT with little functional dependency. The ESPs represent the potential experienced by a positive charge in different points surrounding the analysed structure due to the electron density and can be calculated as:[10]

$$V(r) = \sum_{A} \frac{Z_A}{|r - R_A|} - \sum_{\mu v} P_{\mu v} \int \frac{\theta_{\mu}(r_1)\, \theta_{v}(r_1)}{|r - r_1|} dr_1$$

Eq. 4.1

where Z_A is the nuclear charge of atom A centred at R_A, $P_{\mu v}$ is the corresponding element of an appropriate density matrix, and θ_{μ} and θ_{v} are orbital basis functions. It is important to bear in mind that ESP values at a point in space consider the effects of the whole molecule's charge distribution. Although closer functional groups will have a dominant effect (since the electrostatic potential decay as the inverse of distance in Eq. 4.1), remote groups still contribute directly, and indirectly through polarization.

We also investigated the effect of a larger basis set size, performing B3LYP calculations with diffuse basis functions with polarization functions on all atoms

(B3LYP/6-31G++(d,p)). These calculations were performed using the Solvation Model Density (SMD)[11] implicit solvation model with diethylether to mimic the hydrophobic environment of the binding site of CREBBP.[12] The low dielectric constant of this solvent is commonly chosen to simulate the protein interior, and we reasoned it may have some effect on substrate polarization which could influence the ESP values obtained.

From the DFT calculations, a quantitative model was constructed in which the ESP was analysed above the aromatic ring of each ligand (**Figure 4.3**). A plane of best fit and ring-centroid were determined considering the heavy atom coordinates of the aromatic ring. A parallel plane at a distance of 3.5 Å above the centroid was considered for computing the ESP. This is an approximate distance between the carbon of the guanidinium of arginine and the aromatic ring of the chloro-methoxybenzene of **L11** (PDB ID: 4NR7). First, we implemented a (univariate) single-variable model based on the ESP value above the centroid. However, this approach was insufficient to give a good quantitative reproduction of the binding affinities as judged by a poor correlation-coefficient between the model and experimental values. While such a model has been successfully applied to correlate the complexation energies with ESP values of small aromatic compounds with monoatomic ions,[13] evidently this approach fails for larger systems such as the guanidinium cation. We reasoned that this failure arises from an inability to take into account the larger shape of residues such as arginine, which has a positive charge distributed not just at a single site, but delocalized across the π-system. In the spirit of *Occam's razor,* (*i.e.* the simplest model to account for the observed data is preferable) the new model here presented was then constructed aiming to contain the minimum number of variables (two) that could describe the cation-π interaction. We found that such an approach could be used to give good correlation with experimental binding data for ligands **L1-L15**.

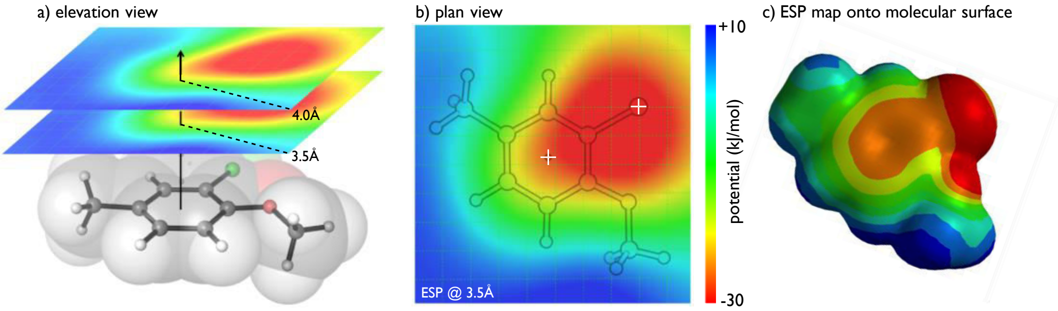


Figure 4.3. 2D ESP horizontal slice spanning -30 – 10 kJ mol^{-1} generated at 3.5 Å and 4.5 Å above the aromatic group of **L11** in a) elevation view and b) plan view; c) ESP map onto the molecular surface of the aromatic group of **L11** spanning -80 – 80 kJ mol^{-1}.[2]

ESP isosurfaces were generated with Spartan 14.[14] We analysed 2D isosurfaces at a distance of 3.5 Å (**Figure 4.3**) above the centroid of each aromatic ring. Repetition of this analysis at a distance of 4.0 Å (**Figure 4.4**) results in qualitatively similar results. The colour at each point on these surfaces reflects the potential, which is experienced by a point positive charge. The two extremes of the scale are represented by red – corresponding to an attractive potential of 30 kJ mol^{-1} – and by blue – corresponding to a repulsive potential of 10 kJ mol^{-1}, and intermediate values are described by the rainbow scale. These values approximately correspond to local regions of negative (attractive) and positive (repulsive) charge associated with a molecule’s functional groups, however, they also reflect electrostatic interactions with the entire distribution of charge around each molecule.

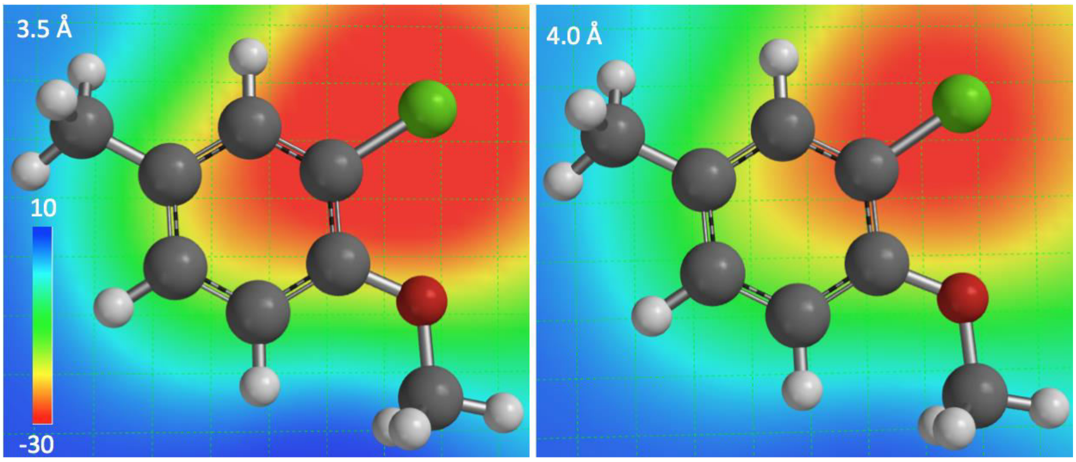


Figure 4.4. 2D ESP isosurfaces were generated for planes 3.5 Å and 4.0 Å perpendicular distance above the aromatic systems. ESP values for **L1-L15** were used for model fitting. Colour scale spans -30 to 10 kJ mol^{-1}.[2]

As seen in **Figure 4.4,** regions with the lowest values of electrostatic potentials (in red) are found towards the chloro- and methoxy-substituents, not directly above the aromatic rings. Interestingly, regions closer to the chlorine presented more negative ESP values than in the proximity of the oxygen. By analysis of the crystal structure of the CREBBP bromodomain bound to an analogue of **L11** (PDB 4NR7), it is possible to see the guaninidium lying parallel to the aromatic ring, in the direction of the chlorine atom. These observations led us to develop a bivariate model using the ESPs evaluated at two different points in space: above the aromatic ring, and secondly, above the substituent with the most negative ESP (**Figure 4.4**). The resulting equation obtained after applying a linear regression with these two variables (**Eq. 4.2**) was used to predict ΔT_m values for comparison with experimental values:

$$\Delta\Delta T_m = a \cdot ESP_{ring} + b \cdot ESP_{substituent}$$

Eq. 4.2

ESP values are used in Hartree: the units of the regression coefficients (a,b) in **Eq. 4.2** are therefore °C/Hartree.

The bivariate model was constructed using the *Real Statistics* plugin for Microsoft Excel. Here the objective is to optimize coefficients (b_1, b_2 and c) to produce a linear model for experimental thermal shift data (y), given ESP values above the aromatic ring (x_1) and above the substituent (x_2) for each ligand **L1-L15**.

$$\hat{y} = b_1 x_1 + b_2 x_2 + c$$

Eq. 4.3

By following the Method of Least Squares, we can rewrite this equation as:

$$\hat{y} - b_0 = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2)$$

Eq. 4.4

The best correlation corresponds to finding the best-fit line to:

$$\hat{y} - \bar{y} = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2)$$

Eq. 4.5

With b_1 and b_2 being calculated as the solutions of the following two equations system based on the covariance (cov):

$$cov(y, x_j) = \sum_{m=1}^{2} b_m \,.\, cov(x_m\,, x_j)$$

Eq. 4.6

The quality of the model was then based on the analysis of the standard R^2 between the sum of squares of defined as follow:

$$R^2 = \frac{SS_{reg}}{SS_{reg} + SS_{res}}$$

Eq. 4.7

where the sum of squares ($SS_{reg}\ and\ SS_{res}$) $is\ defined\ as$:

$$SS_{reg} = \sum(\hat{y} - \bar{y})^2$$

Eq. 4.8

and

$$SS_{res} = \sum(y - \hat{y})^2$$

Eq. 4.9

The choice of functional and basis set has a negligible effect on the correlation obtained constructing a model as outlined above. In each case the regression coefficients of the two ESP descriptors are very similar in magnitude (**Table 4.1**): the ratio of b_1:b_2 is in the narrow region of 55:45 – 57:43, showing the near equal significance of both ESP descriptors in the overall model performance. An alternative model, using the minimum value of the ESP (*i.e.* the most attractive part of the potential) rather than a value directly above one of the substituents led to a very small improvement (an increase in R^2 by 0.01). Based on the simplicity of the original model, there is no reason to adopt this more complex implementation.

Table 4.1. Theoretical ΔT_m calculated after using multiple linear regression with ESP values at 3.5 Å above the aromatic ring and above the substituent with the most negative ESP. Different methodologies were considered: 1) M06-2X/6-31G(d,p) in diethylether; 2) B3LYP/6-31G(d,p) in gas-phase; 3) B3LYP/6-311++G(d,p) in diethylether.

Ligand	**Expt. ΔT_m**	**Theoretical ΔT_m**			
		M06-2X	**Gas-phase**	**6-311++G(d,p)**	**ESP***
1	6.5	6.3	6.4	6.4	6.3
2	6.4	7.0	6.9	6.9	6.8
3	6.3	6.8	6.7	6.8	6.9
4	7.4	7.0	7.0	7.1	7.1
5	8.1	8.6	8.5	8.5	8.5
6	5.2	4.8	4.8	4.8	4.8
7	5.2	4.9	4.9	4.9	5.0
8	6.2	6.9	6.9	7.0	6.9
9	7.1	7.2	7.2	7.2	7.1
10	6.4	7.2	7.2	7.3	7.2
11	9.0	9.2	9.3	9.1	9.2
12	9.6	8.5	8.5	8.6	8.6
13	6.0	6.1	6.3	6.1	6.2
14	6.2	6.1	6.0	6.0	6.0
15	8.9	7.9	7.8	7.8	7.9
		$R^2 = 0.83$ (57/43)**	$R^2 = 0.82$ (56/44)	$R^2 = 0.82$ (55/45)	$R^2 = 0.84$ (55/45)

*Consideration of the most negative ESP value near the substituent.

**Ratio between the coefficients related to the ESP values above the π-system and the substituent, respectively.

Molecular Dynamics simulations. These calculations were performed by Kiran Kumar. MD simulations are important for the study of the protein-ligand dynamics along the time. They are dependent on parameters available in force fields, as previously discussed in **Section 2.2** of **Chapter 2**, and are often competitive with more time-consuming QM calculations for analysis of weak non-bonded interactions.[15] For this reason, MD calculations were performed to investigate the temporal flexibility of the arene:arginine interactions in CREBBP. The Amber12[16] package was used to simulate all fifteen systems. Generation of the small-molecule force field with AmberFF12SB was performed with Antechamber.[17] The

MMPBSA.py script[18] distributed with AmberTools12[16] was used to compute binding free energies with both the Generalized Born and Poisson-Boltzmann surface area continuum solvation methods (MM-GBSA and MM-PBSA). While Biggin and co-workers[19] have showed Free Energy Perturbation (FEP) as a state-of-the-art methodology for accurate prediction of binding free energies in bromodomains, MM-GBSA and MM-PBSA have shown to reasonably correlate with experimental binding values to different proteins ($R^2 > 0.6$) and are considerably less time consuming than the FEP analysis.[20] More details of the MD methodology can be found in ref[21].

4.3 Results and discussion

Following our observation that a cation-π interaction between the DHQ derivative *(R)*-**2** and the guanidinium of R1173 of CREBBP was maintained during 100 ns of MD, dynamical behaviour of 5-isoxazolylbenzimidazole derivatives was also analysed in the CREBBP bromodomain binding site. Compound **L11** was able to form this cation-π interaction with R1173 for most of the simulated time (98%), showing the importance of this interaction for drug binding and recognition (**Figure 4.5**). Compound **L1**, which presented lower values of ΔT_m compared to *(R)*-**2**, was also able to form the cation-π interaction during the simulation; however, it was observed to be less frequent (74%). Both simulations (**L11**:PDB 4NR7 and **L1**:4NR5) presented Root Mean Square Deviation (RMSD) values lower than 2 Å, indicative of relative stability of these systems.

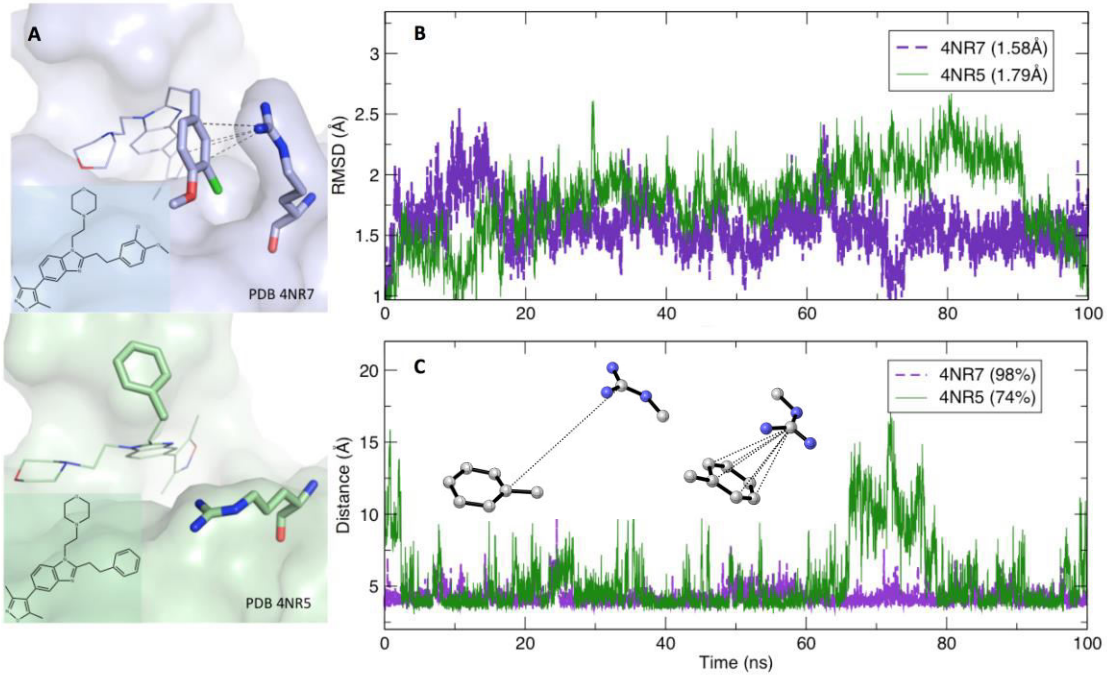


Figure 4.5. (A) The X-ray crystal structure of the CREBBP-**L11** complex shows a clear cation-π interaction; that of the CREBBP-**L1** complex does not; (B) following 100 ns MD simulation the two complexes are stable, as shown by average RMSD values; (C) both ligands form active site cation–π interactions for a majority of simulation time, using a distance cut-off of 6 Å[22] between guanidinium and aromatic group.[2]

The formation of the cation-π interaction for Compound **L1** for most of the simulated time, even when not shown in the crystallographic data (PDB 4NR5), led us to quantify binding energies for **L1**-**L15** by using PDB 4NR7 as a starting point and replacing **L11** by the other ligands. Short-time MD simulations (5 ns) were performed and binding energies of the fifteen 5-isoxazolylbenzimidazole derivatives in CREBBP in water were calculated using MM-PBSA and MM-GBSA scoring functions (**Figure 4.6**). The potential of these calculated binding energies to rank the best binders in CREBBP was evaluated by analysis of the Spearman coefficient (r_s).

Different solvation models (PB vs GB) presented similar coefficient values (PB: r_s = 0.59 and GB: r_s = 0.53), with PB showing a slightly better ranking prediction. MM-PBSA scoring functions were able to successfully rank four (**L5**, **L11**, **L12**, **L15**) out of the six strongest binders.

Cation-π interactions have been well described by electrostatic models,[13] so we turned our attention to the analysis of Coulombic electrostatics following short-time MD simulations. This approach resulted in better ranking prediction ($r_s = 0.72$) than MM-PBSA and MM-GBSA scoring functions, corroborating the idea that cation-π interactions gain important insights from electrostatic-based models.

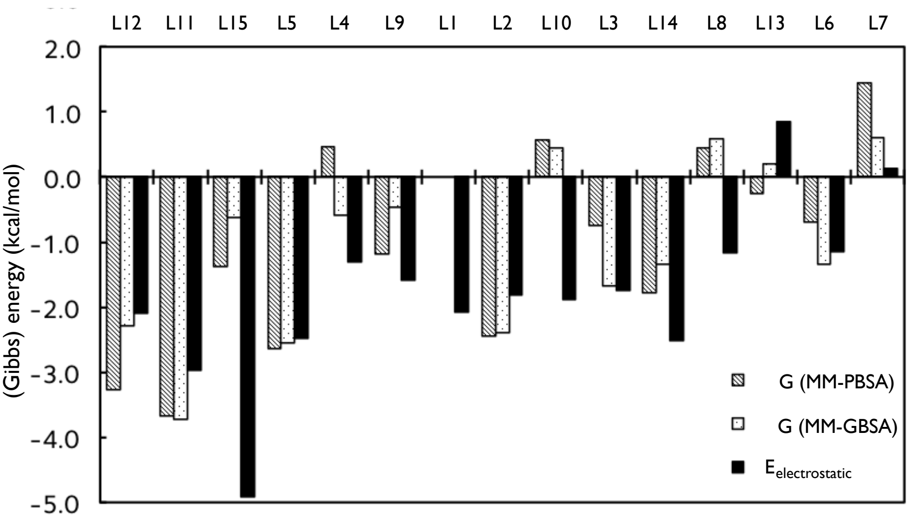


Figure 4.6. Binding free energies and arginine electrostatics calculated for ligands L1-L15 using MM-PBSA and MM-GBSA analysis over 5 ns of MD simulation. The ligands are ordered from strongest to weakest experimental affinity.[2]

Previous work has shown the ability of QM to quantify the interaction strength between functionalized aromatic rings and (mainly Na^+) cations.[23] In this work, counterpoise and dispersion-corrected DFT calculations were performed for DHQ-guanidinium complexation and potential of this technique was analysed to rank experimental binding affinities.

Coordinates of the guanidinium and the chloro-methoxyphenyl ring of **L11** in CREBBP (PDB ID: 4NR7) were used as a starting point for our QM calculations. Non-hydrogen atoms of guanidinium and link atoms between the modified benzene and the rest of the molecule were fixed (**Figure 4.2**). The interaction energy was computed from the difference between the counterpoise-corrected energy[4] of the complex and those of the separated guanidinium and arene. The correlation between QM and experimental data did not change substantially when testing different DFT functionals (TPSS, B3LYP-D3, M06-2X) and basis sets (triple- and quadruple-ζ) (**Table 4.2**), with a better performance for the M06-2X/cc-pVQZ level of theory (r_s = 0.79 and R^2 = 0.64, **Chart 4.2**). Analysis of complexation energies was able to successfully rank the four strongest binders (**L5**, **L11**, **L12** and **L15**). Interestingly, the indole-containing ligand (**L15**) presented the strongest interaction. **L15** has a very negative ESP around its very well conjugated system, comparable to the one observed by the tryptophan amino acid.[22] The other strongest binders **L5**, **L11** and **L12** have a methoxyphenyl in their structures, while the weakest binders contain a nitrophenyl. This raises the question as to whether the binding energies are controlled by π electron donation or withdrawing group (i.e. resonance/mesomeric effects). However, the addition of inductively-withdrawing halogens in **L11** (Cl) and **L12** (F) results in a further increase in experimental binding affinity and computed interaction energy. For example, a more favourable QM binding energy was obtained for **L11** (-13.4 kcal mol^{-1}) compared to **L5** (-12.6 kcal mol^{-1}).

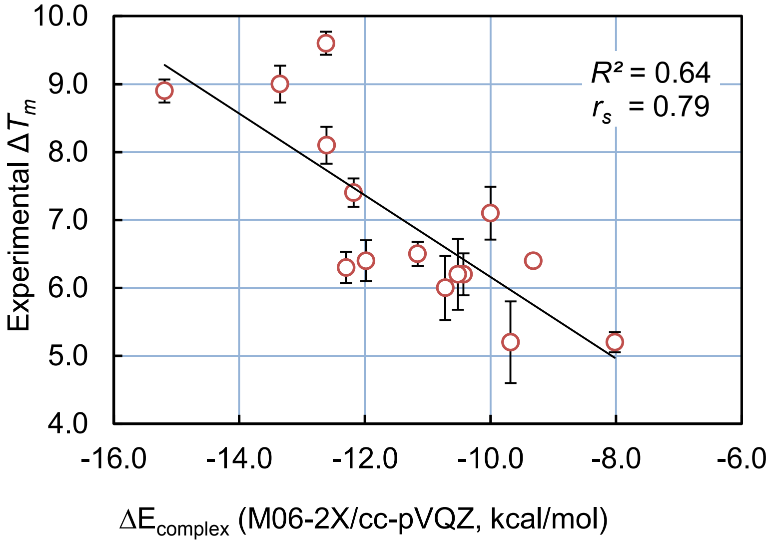


Chart 4.2. Correlation between experimental ΔT_m and counterpoise-corrected M06-2X/cc-pVQZ complexation energies; experimental error bars shown.[2]

Table 4.2. Counterpoise-corrected complexation energies (kcal mol^{-1}).

Ligand	**TPSS-D3/def2-TZVPP (TPSS-D3/cc-pVQZ)**	**B3LYP-D3/def2-TZVPP**	**M06-2X/def2-TZVPP**
1	-11.20 (-11.24)	-11.52	-11.10
2	-12.06 (-12.11)	-12.43	-11.93
3	-12.35 (-12.38)	-12.66	-12.26
4	-12.24 (-12.27)	-12.54	-12.14
5	-12.97 (-13.00)	-13.11	-12.54
6	-8.93 (-9.01)	-9.69	-9.57
7	-7.81 (-7.92)	-8.21	-7.87
8	-10.17 (-10.21)	-10.58	-10.36
9	-9.88 (-9.93)	-10.16	-9.92
10	-9.50 (-9.55)	-9.68	-9.24
11	-13.25 (-13.27)	-13.8	-13.3
12	-12.45 (-12.48)	-12.73	-12.54
13	-10.75 (-10.77)	-11.15	-10.67
14	-12.28 (-10.81)	-11.35	-10.39
15	-14.77 (-14.82)	-15.48	-15.12
	R^2 = 0.57 (0.63)	R^2 = 0.61	R^2 = 0.63

Houk and Wheeler[13] and Dougherty[24] observed a strong correlation between the QM-computed ESP values above substituted benzenes and the interaction energies of the cation-π complexes. For a better understanding of the role of electrostatics on the experimental binding affinity data, B3LYP/6-31G(d,p) ESP calculations were performed. An implicit conductor-like polarizable continuum model (CPCM)[25] description of diethyl ether was used to mimic the hydrophobic protein environment in these calculations, although gas-phase results were not noticeably different. Analysis of the ESP surfaces (**Figure 4.7**) allowed us to confirm that the indole group has the most negative electrostatic potential, which would be expected to interact most strongly with a positively charged (*e.g.* guanidinium) group, as observed in our QM-complexation energies analysis. It is worth noting that while a negative ESP value often results from an electron-rich region, the ESP depends not just on the local electron density but upon the entire molecule.[18] The least active compounds, the ligands containing the nitrophenyl (**L6** and **L7**), present a very negative ESP surface around the nitro group. However, the ESP above the aromatic ring is the least negative compared to all the other compounds, a result of the electron withdrawing effect. The presence of the electron donating methoxy group in the strong binder **L5** results in a very negative ESP not only above the ring, but also surrounding the oxygen of the methoxy group.

As previously suggested by Wheeler and Houk,[13] substituents can have a through space electrostatic effect on cations, especially in this case, in which the cation under investigation is no longer an atom, but indeed a molecule that crosses the borders of the benzene. The presence of a fluorine atom reduced the ESP above the aromatic ring, but **L12** is the strongest binder experimentally, showing that substituents would also have an independent way of interacting with cations, by through space electrostatics.[13] Finally, heterocyclic thiazole (**L13**, ΔT_m = 6.0 °C) and benzothiazole (**L14**, ΔT_m = 6.2 °C) have a less negative ESP than the benzene (**L1**, ΔT_m = 6.5 °C), which results in slightly lower values of binding affinities. These results suggest that it is possible to have an idea about the overall strength of the cation-π interaction by taking a general look at the ESP surfaces (**Figure 4.7**). Most potent ligands have either a combination of very negative electron density above the π-system and one substituent (**L5**, **L11**, **L12**) or simply above a bicyclic π-system (**L15**).

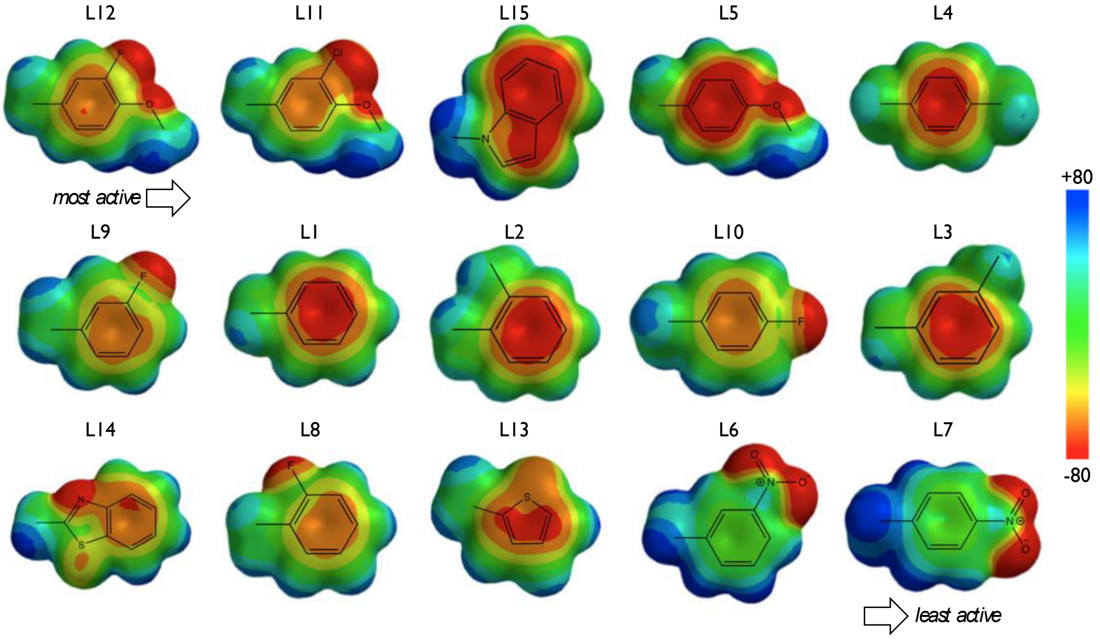


Figure 4.7. B3LYP/6-31G(d,p) molecular ESP maps (isosurface value 0.002) for aromatic/heteroaromatic groups of **L1-L15,** ordered according to experimental binding affinity for CREBBP.[2]

Table 4.3. Correlation of experimental affinity data against QM computations. [a]M06-2X/cc-pVQZ complexation energies; [b]B3LYP/6-31G* ESP values 3.5 Å above the ring-centre and substituent with the most negative value. For ligands **L14** and **L15**, the ring with the most negative potential above it was considered; [c]Predicted thermal shift data obtained after bivariate linear regression (**Eq. 4.10**) using ESP values; [d]Model error relative to experimental ΔT_m values; [e]For comparison, these ligands were considered as unsubstituted conjugated π-systems.

Ligand	$\Delta E_{complex}$[a] (kcal/mol)	ESP_{ring}[b] (Hartree)	$ESP_{subst.}$ (Hartree)	Model[c] ΔT_m (°C)	Error[d] (°C)
L1	-11.16	-0.0122	0	6.3	-0.2
L2	-11.98	-0.0128	-0.0014	6.9	+0.5
L3	-12.30	-0.0126	-0.0011	6.7	+0.4
L4	-12.18	-0.0127	-0.0020	7.0	-0.4
L5	-12.61	-0.0134	-0.0074	8.5	+0.4
L6	-9.68	0.0083	-0.0104	4.8	-0.4
L7	-8.02	0.0142	-0.0117	4.9	-0.3
L8	-10.43	-0.0086	-0.0070	6.9	+0.7
L9	-10.00	-0.0082	-0.0086	7.2	+0.1
L10	-9.32	-0.0080	-0.0089	7.2	+0.8
L11	-13.35	-0.0094	-0.0158	9.3	+0.3
L12	-12.62	-0.0099	-0.0124	8.6	-1.0
L13	-10.72	-0.0118	N/A[e]	6.2	+0.2
L14	-10.52	-0.0113	N/A[e]	6.1	-0.1
L15	-15.19	-0.0175	N/A[e]	8.0	-0.9
	r_s = 0.79	**r_s = 0.47**		**r_s = 0.91**	

The qualitative description of the ESP showed that compounds with the most favourable binding affinities had either an aromatic group with a very negative ESP above its π-system (**L15**), or a combination of a negative ESP above both the aromatic ring and the substituents (**L5**, **L11** and **L12**). We then tried to quantify the ligand-arginine interaction by analysing the ESP of the unsaturated group. Considering just the ESP 3.5 Å above the aromatic rings was not sufficient for successfully rank the strongest binders (r_s = 0.47, **Table 4.3**). However, when considering two different points, one above the aromatic ring and another one above the substituent with the most negative potential, a linear regression model was built (**Eq. 4.10**) being able to rank the inhibitors according to ΔT_m values with good accuracy ($r_s = 0.91$, $R^2 = 0.84$) with no clear outliers (RMSE = 0.52 °C) (**Chart 4.3**).

A higher regression coefficient (-307.7) was given to the ESP above the π-system compared to above the substituent (-238.7) (**Eq. 4.10**), while negative signs attributed to these variables reflect the additive nature of the calculated electrostatic contribution. This highlights the importance of electrostatic effects coming from the π-system, although through-space electrostatics from the substituents still presents a significant contribution to the overall ESP values. This is in accordance with previous studies by Wheeler[8, 26] and Dougherty[27] outlining that cation-π interactions are more affected by electrostatic interactions with substituents than π-resonance effects.

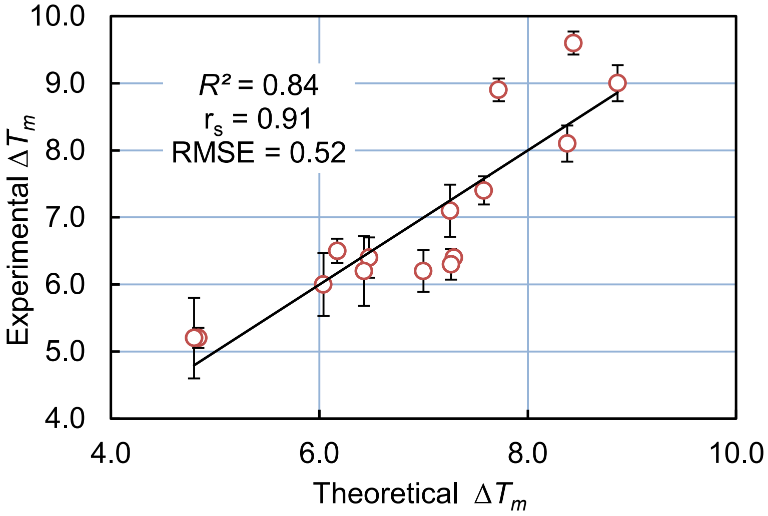


Chart 4.3. Correlation between experimental and predicted ΔT_m using a parameterized electrostatic model (**Eq. 4.10**).[2]

$$\Delta T_m = -307.70 * ESP_{ring} - 238.74 * ESP_{substituent} + 2.58$$

Eq. 4.10

The sensibility of this model was tested with different DFT functionals, solvation models and basis sets (**Table 4.1**), leading only to small changes in the quality of the ESP model. The correlation between dipole moment and binding affinities was also analysed, given recent work by Wheeler and co-workers[28] showing that binding affinities between small heterocycles correlate well with theoretical data computed at the B97D/def2-TZVPP level of theory. However, Wheeler[28] also found a low correlation for stacking of fused rings. In our analysis, dipole moment showed a very poor correlation (**Table 4.4**, $R^2 < 0.15$) with the experimental binding affinities in CREBBP for both small (**L1-L13**) and fused ring systems (**L14-L15**).

Table 4.4. B3LYP/6-31G* dipole moment (Debye) of the aromatic/heteroaromatic groups of Ligands **L1-L15**.

Ligand	**Dipole Moment**
1	0.3447
2	2.4238
3	0.7305
4	0.5882
5	0.0827
6	0.7799
7	0.3544
8	1.1869
9	5.0183
10	1.6946
11	5.3473
12	1.1795
13	1.8249
14	2.4212
15	2.8971
	$R^2 = 0.13$

4.4 Prospective validation

The use of a bivariate ESP model (**Eq. 4.10**) successfully correlated with ΔT_m of fifteen 5-isoxazolylbenzimidazole derivatives in CREBBP. The same model was then applied to a series of newly synthesised 7-membered ring DHQ derivatives by Dr. Michael Brand, working with Prof. Stuart Conway. The predicted ΔT_m is then compared with IC_{50} data in the CREBBP bromodomain ($R^2 = 0.73$, n=4, **Figure 4.8**).

The most intriguing result comes out with the analysis of the ligand containing the electron withdrawing group acetamide (**L19,** IC$_{50}$ = 1.0 μM). This compound has a lower electrostatic potential above the π-system than the ligand containing the electron donating methoxy group (**L17,** IC$_{50}$ = 1.6 μM). However, the amide oxygen improves the electrostatic interaction with the guanidinium of R1173. By the use of the bivariate model, this would result in similar binding affinities in CREBBP comparing **L17** with **L19**, corroborating the IC$_{50}$ data. Given the increasing number of CREBBP inhibitors being reported in literature,[29] models based on the structural analysis of small molecules such as the one presented in this work may be valuable tools for the understanding of selectivity and potency of these compounds.

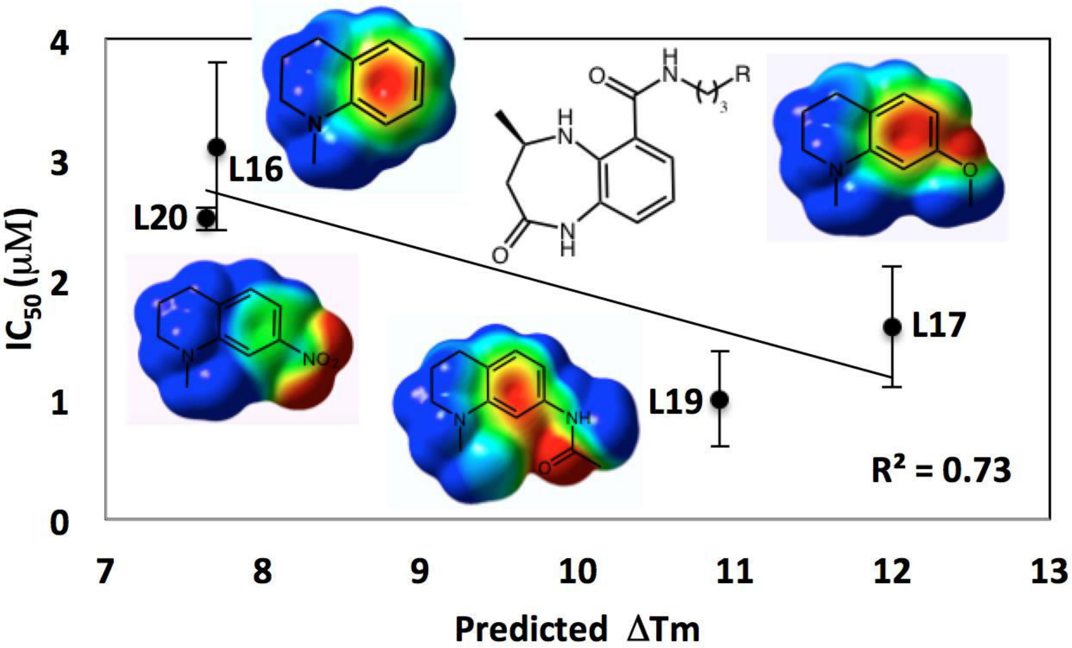


Figure 4.8. Comparison between predicted ΔT_m and experimental IC$_{50}$ in CREBBP.

4.5 Conclusion

The observation that cation-π interactions contribute to CREBBP binding, led us to develop a quantitative model to correlate structure with affinity for a library of 5-isoxazolylbenzimidazoles, which combine an isoxazole hydrogen-bonding group and an aromatic side-chain. In theory, such functionalities have the potential to engage in a cation-π interaction. In contrast to most screening approaches, which utilize scoring or energy functions based on molecular mechanics potentials, this work has successfully validated the use of electronic structure theory (in the form of the DFT-computed electrostatic potential) to rank binding affinities. DFT calculations for entire protein systems remain out of reach, and so we utilize the QM description provided by DFT to compute the electronic properties of ligands in a non-empirical way. This approach does not attempt to calculate *absolute* binding affinity, which is notoriously challenging for theory, but rather *relative* changes in binding affinity associated with modifications of a single protein:ligand interaction. Our tool would be useful for ligand optimization, rather than initial screening of different ligand scaffolds. By taking the systematic modification of these aromatic substituents, a quantitative model is here formulated that relates the electrostatic potential above or below the aromatic side-chain and experimentally determined binding affinities. In this model, the size of the guanidinium cation is taken into account, which would require the analysis of the electrostatic potential not only above the ring of the ligand, but also in a point away from the ring centre.

This new model relies on the relatively modest cost for the computation of electrostatic potentials using DFT, has been trained against literature data (R^2=0.84, n=15) of binding affinities and successfully tested against newly synthesized DHQ derivatives. The use of similar ESP values is suggested as an alternative for scaffold-hopping instead of methodologies depending only on steric parameters.

4.6 Supporting Information Chapter 4

Ligand 1: Arg complex

Counterpoise corrected energy = -516.9725
BSSE energy = 0.0003
Sum of monomers = -516.9547
Complexation energy = -11.41 kcal/mol (raw)
Complexation energy = -11.20 kcal/mol (corrected)

Atom	X	Y	Z
C	3.254757	-1.196359	-0.260889
N	2.760614	0.160226	-0.464438
C	1.863896	0.744207	0.319794
N	1.356313	0.091733	1.356060
N	1.460071	1.970973	0.052844
C	-1.147831	-1.407212	-1.268060
C	-1.332192	-2.118981	-0.081304
C	-1.839744	-1.461388	1.041383
C	-2.148519	-0.100506	0.975254
C	-1.461309	-0.047988	-1.330251
C	-1.970640	0.628007	-0.213451
C	-2.253308	2.091646	-0.249558
H	2.415105	-1.896804	-0.226695
H	3.855476	-1.277696	0.651073
H	3.286274	0.753771	-1.092933
H	1.661272	-0.854445	1.534927
H	0.399320	0.286437	1.638416
H	0.794882	2.439073	0.650695
H	1.670108	2.412151	-0.830149
H	-2.009768	-2.008861	1.963781
H	-2.557419	0.400382	1.850161
H	-1.318596	0.494510	-2.261763
H	-2.044402	2.568590	0.714235
H	-1.672656	2.597066	-1.027640
H	3.878224	-1.458990	-1.114187
H	-3.309925	2.288840	-0.468328
H	-1.105377	-3.179680	-0.037953
H	-0.772170	-1.913425	-2.152411

Ligand 2: Arg complex

Counterpoise corrected energy = -556.3155
BSSE energy = 0.0004
Sum of monomers = -556.2964
Complexation energy = -12.33 kcal/mol (raw)
Complexation energy = -12.06 kcal/mol (corrected)

Atom	X	Y	Z
C	3.388361	0.924579	0.699348
N	2.804280	-0.389883	0.460428
C	1.968574	-0.655771	-0.535171
N	1.612074	0.299673	-1.381970
N	1.474281	-1.871199	-0.668462
C	-1.074611	1.209204	1.298680
C	-1.048644	2.220615	0.336597
C	-1.448683	1.938075	-0.968984
C	-1.856173	0.641815	-1.295915
C	-1.490931	-0.090337	0.986215
C	-1.889551	-0.381543	-0.336065
C	-2.251371	-1.768685	-0.746518
H	2.595784	1.673694	0.784966
H	4.094224	1.205310	-0.089241
H	3.229100	-1.176059	0.934887
H	1.976766	1.231722	-1.244136
H	0.675061	0.275235	-1.778925
H	0.871079	-2.103084	-1.443889
H	1.585789	-2.568324	0.052416
H	-1.463386	2.719520	-1.722716
H	-2.184463	0.421236	-2.309442
H	-2.418762	-1.835248	-1.824432
H	-1.473808	-2.492543	-0.468867
H	3.920897	0.888491	1.648561
H	-3.165996	-2.115549	-0.250245
H	-0.744626	3.225763	0.612551
H	-0.783050	1.434289	2.321623
C	-1.521517	-1.157864	2.049743
H	-2.532824	-1.558428	2.180765
H	-0.883042	-2.011478	1.786599
H	-1.185913	-0.765210	3.011965

Ligand 3: Arg complex

Counterpoise corrected energy = -556.3155
BSSE energy = 0.0004
Sum of monomers = -556.2958
Complexation energy = -12.58 kcal/mol (raw)
Complexation energy = -12.35 kcal/mol (corrected)

Atom	X	Y	Z
C	-3.233765	1.116135	0.285129
N	-2.781468	-0.084026	-0.408461
C	-1.933693	-0.961671	0.112515
N	-1.439311	-0.773341	1.327873
N	-1.564754	-2.013619	-0.592213
C	1.115883	1.476361	-0.495752
C	1.310697	1.668536	0.877903
C	1.810986	0.636140	1.673139
C	2.106577	-0.606078	1.109042
C	1.418745	0.222068	-1.043606
C	1.918714	-0.828864	-0.265742
C	2.150411	-2.181573	-0.848885
H	-2.371230	1.710202	0.601042
H	-3.865408	0.874965	1.146450
H	-3.308001	-0.372158	-1.222985
H	-1.713183	0.050094	1.844891
H	-0.497076	-1.099573	1.529127
H	-0.926457	-2.695050	-0.208178
H	-1.752496	-2.073119	-1.581983
H	1.978408	0.801605	2.733485
H	2.502657	-1.403829	1.732951
H	1.855100	-2.977254	-0.155698
H	1.607952	-2.313835	-1.790120
H	-3.813979	1.712306	-0.417585
H	3.211924	-2.347315	-1.070065
H	1.262613	0.060971	-2.109066
H	1.089568	2.635778	1.321569
C	0.622564	2.599466	-1.374246
H	1.467746	3.151657	-1.800942
H	0.027645	2.222220	-2.211575
H	0.020232	3.316113	-0.808682

Ligand 4: Arg complex

Counterpoise corrected energy = -556.3151
BSSE energy = 0.0003 (Hartrees)
Sum of monomers = -556.2956
Complexation energy = -12.45 kcal/mol (raw)
Complexation energy = -12.24 kcal/mol (corrected)

Atom	X	Y	Z
C	-3.039264	1.607780	-0.257029
N	-2.861686	0.174835	-0.459681
C	-2.116237	-0.594205	0.323287
N	-1.472763	-0.070831	1.357324
N	-1.997816	-1.880476	0.057253
C	1.257563	0.830971	-1.254874
C	1.600716	1.506419	-0.076386
C	1.964545	0.731537	1.033509
C	1.981884	-0.662867	0.964885
C	1.278054	-0.562422	-1.325132
C	1.645656	-1.339327	-0.219365
C	1.593620	-2.829140	-0.253901
H	-2.063460	2.100706	-0.215878
H	-3.612562	1.822119	0.650906
H	-3.511420	-0.286601	-1.083186
H	-1.564223	0.918328	1.540058
H	-0.574462	-0.465693	1.625724
H	-1.440774	-2.480026	0.648559
H	-2.283590	-2.259035	-0.833609
H	2.249219	1.224286	1.960031
H	2.278882	-1.236374	1.840532
H	1.007062	-1.054229	-2.256560
H	1.284191	-3.246207	0.710817

H	0.911809	-3.191830	-1.029758
H	-3.581250	2.004372	-1.114253
H	2.577321	-3.261671	-0.474594
H	0.979603	1.404683	-2.135794
C	1.624111	3.013725	-0.017977
H	1.407971	3.380441	0.989664
H	2.614170	3.393754	-0.295143
H	0.903479	3.453465	-0.713893

Ligand 5: Arg complex

Counterpoise corrected energy = -631.5662
BSSE energy = 0.0004
Sum of monomers = -631.5455
Complexation energy = -13.23 kcal/mol (raw)
Complexation energy = -12.97 kcal/mol (corrected)

Atom	X	Y	Z
C	1.873828	-2.864558	-0.139980
N	2.486051	-1.589936	-0.495588
C	2.358077	-0.483453	0.225157
N	1.635960	-0.489611	1.336787
N	2.933737	0.632065	-0.179228
C	-1.480109	0.128878	-1.078950
Atom	X	Y	Z
C	-1.913165	-0.196327	0.212286
C	-1.612085	0.668431	1.276458
C	-0.878240	1.827075	1.042822
C	-0.740812	1.296992	-1.289581
C	-0.424190	2.169656	-0.245506
C	0.426465	3.377154	-0.449906
H	0.797783	-2.731474	0.006117
H	2.326844	-3.298360	0.757642
H	3.223122	-1.610677	-1.188574
H	1.196973	-1.351219	1.629606
H	1.115265	0.349014	1.586620
H	2.839361	1.482268	0.357342
H	3.283630	0.727294	-1.121217
H	-1.980372	0.426035	2.268400
H	-0.673008	2.496806	1.875619
H	-0.410953	1.530254	-2.299120
H	1.069590	3.570666	0.416315
H	1.058385	3.280954	-1.338657
H	2.019576	-3.551773	-0.972129
H	-0.177503	4.282185	-0.590369
H	-1.718097	-0.509309	-1.922070
O	-2.618093	-1.319162	0.534047
C	-3.078245	-2.144306	-0.546628
H	-3.664364	-2.934136	-0.079714
H	-3.706149	-1.568093	-1.234169
H	-2.234715	-2.582027	-1.094862

Ligand 6: Arg complex

Counterpoise corrected energy = -721.6023
BSSE energy = 0.0007
Sum of monomers = -721.5881
Complexation energy = -9.36 kcal/mol (raw)
Complexation energy = -8.93 kcal/mol (corrected)

Atom	X	Y	Z
C	-3.173837	0.031853	0.792022
N	-2.536290	-0.994099	-0.024692
C	-1.456455	-1.668628	0.348541
N	-0.887682	-1.424349	1.520634
N	-0.935421	-2.570294	-0.460629
C	0.653319	1.238951	-0.247430
C	1.254543	1.728255	0.907776
C	2.342469	1.021264	1.419775
C	2.799012	-0.136251	0.785864
C	1.111300	0.103604	-0.907906
C	2.194954	-0.610287	-0.390894
C	2.666655	-1.841564	-1.087259
H	-2.547353	0.924007	0.891700
H	-3.415571	-0.374997	1.777151
H	-2.747953	-0.984039	-1.014333
H	-1.103786	-0.581469	2.032455
H	-0.099169	-1.964758	1.843635
H	-0.043627	-3.000204	-0.266888

H	-1.454577	-2.925662	-1.250323
H	2.837830	1.375977	2.317418
H	3.646029	-0.673517	1.203935
H	2.806155	-2.675980	-0.389659
H	1.970454	-2.148587	-1.872445
H	-4.102524	0.320878	0.303242
H	3.639228	-1.678442	-1.565523
H	0.612798	-0.213894	-1.816164
H	0.875793	2.627117	1.378827
N	-0.562544	1.906294	-0.762581
O	-1.185213	1.334076	-1.669852
O	-0.911279	2.959365	-0.230784

Ligand 7: Arg complex

Counterpoise corrected energy = -721.5991
BSSE energy = 0.0005
Sum of monomers = -721.5866
Complexation energy = -8.13 kcal/mol (raw)
Complexation energy = -7.81 kcal/mol (corrected)

Atom	X	Y	Z
C	-1.239442	-3.100944	0.163147
N	-2.185846	-2.001103	0.306681
C	-2.179867	-0.917934	-0.459390
N	-1.275269	-0.782833	-1.418900
N	-3.064125	0.037765	-0.249589
C	0.763703	0.333406	1.268639
C	1.659890	0.298976	0.205104
C	1.688719	1.290700	-0.771980
C	0.727766	2.297850	-0.722442
C	-0.165675	1.367099	1.318070
C	-0.228872	2.339038	0.306945
C	-1.358028	3.311982	0.264151
H	-0.212089	-2.730096	0.237243
H	-1.388674	-3.637403	-0.779668
H	-3.005494	-2.163913	0.876461
H	-0.577452	-1.501303	-1.551225
H	-1.022404	0.139890	-1.749323
H	-3.161549	0.812627	-0.888487
H	-3.712185	-0.002630	0.522617
H	2.430208	1.244323	-1.561380
H	0.710725	3.055497	-1.500995
H	-0.873950	1.409403	2.140818
H	-1.811906	3.350473	-0.732633
H	-2.130836	3.061949	0.995268
H	-1.403121	-3.794015	0.986940
H	-1.020625	4.332012	0.483732
H	0.797648	-0.442021	2.024718
N	2.558456	-0.864826	0.068468
O	3.568013	-0.736946	-0.617546
O	2.199488	-1.910409	0.632024

Ligand 8: Arg complex

Counterpoise corrected energy = -616.2621
BSSE energy = 0.0005
Sum of monomers = -616.2459
Complexation energy = -10.51 kcal/mol (raw)
Complexation energy = -10.17 kcal/mol (corrected)

Atom	X	Y	Z
C	3.323259	1.171583	0.349475
N	2.805711	-0.190964	0.308827
C	1.921679	-0.617872	-0.583688
N	1.449905	0.209257	-1.505805
N	1.494644	-1.864806	-0.537824
C	-0.838998	1.212313	1.230964
C	-1.232683	2.185943	0.310337
C	-1.954687	1.816638	-0.825806
C	-2.290813	0.477013	-1.037087
C	-1.186010	-0.108057	0.984445
C	-1.923297	-0.525595	-0.127309
C	-2.226887	-1.968999	-0.346255
H	2.494002	1.883389	0.395489
H	3.963019	1.387953	-0.512432
H	3.297883	-0.891789	0.847535
H	1.786613	1.159915	-1.549480
H	0.525652	0.055800	-1.891799

H	0.848195	-2.214781	-1.229638
H	1.617722	-2.426463	0.292442
H	-2.272395	2.570564	-1.538333
H	-2.865142	0.196556	-1.916179
H	-2.553806	-2.150264	-1.372942
H	-1.362640	-2.604647	-0.127836
H	3.912656	1.281354	1.258513
H	-3.026143	-2.312296	0.320602
H	-0.990506	3.228025	0.492237
H	-0.291285	1.463655	2.133097
F	-0.765003	-1.062132	1.867988

Ligand 9

Counterpoise corrected energy = -616.2608
BSSE energy = 0.0006 (Hartrees)
Sum of monomers = -616.2450 (Hartrees)
Complexation energy = -10.24 kcal/mol (raw)
Complexation energy = -9.88 kcal/mol (corrected)

Atom	X	Y	Z
C	3.249932	-1.042566	0.233303
N	2.768271	0.221887	-0.310024
C	1.861129	0.983676	0.287575
N	1.330654	0.609660	1.443306
N	1.470228	2.106519	-0.283161
C	-0.897134	-1.381514	-0.666557
C	-1.295837	-1.875275	0.568187
C	-2.046728	-1.032086	1.391190
C	-2.366607	0.263958	0.983330
C	-1.217666	-0.101962	-1.104136
C	-1.963604	0.745197	-0.273767
C	-2.236478	2.154518	-0.677326
H	2.409827	-1.723491	0.398270
H	3.812422	-0.895038	1.160989
H	3.289280	0.627592	-1.076031
H	1.605432	-0.273036	1.849849
H	0.395806	0.919801	1.684811
H	0.814558	2.723970	0.172067
H	1.743793	2.342600	-1.225182
H	-2.387162	-1.394049	2.356186
H	-2.950192	0.904192	1.639014
H	-0.886432	0.223778	-2.085484
H	-2.038735	2.852910	0.143781
H	-1.638467	2.450082	-1.544006
H	3.907758	-1.496967	-0.505754
H	-3.288777	2.295216	-0.951120
F	-0.139087	-2.168797	-1.475481
H	-1.043435	-2.888528	0.860297

Ligand 10: Arg complex

Counterpoise corrected energy = -616.2592
BSSE energy = 0.0004
Sum of monomers = -616.2441
Complexation energy = -9.77 kcal/mol (raw)
Complexation energy = -9.50 kcal/mol (corrected)

Atom	X	Y	Z
C	-3.124544	1.455139	-0.250561
N	-2.860330	0.035469	-0.452483
C	-2.067574	-0.685793	0.329612
N	-1.455021	-0.123127	1.361691
N	-1.871630	-1.962732	0.064370
C	1.178855	0.932222	-1.272475
C	1.501938	1.580171	-0.086720
C	1.954401	0.888619	1.031067
C	2.060111	-0.502035	0.953071
C	1.294914	-0.457671	-1.326196
C	1.731451	-1.200535	-0.220849
C	1.770608	-2.690794	-0.254012
H	-2.182287	2.010054	-0.220564
H	-3.701109	1.634126	0.662904
H	-3.479478	-0.465064	-1.076827
H	-1.604826	0.857343	1.552898
H	-0.554677	-0.484532	1.660891
H	-1.306218	-2.539158	0.670083
H	-2.164682	-2.369365	-0.811426
H	2.226125	1.430104	1.930888

H	2.420556	-1.051824	1.819167
H	1.474510	-3.125407	0.707090
H	1.119421	-3.093032	-1.036054
H	-3.699694	1.814970	-1.102317
H	2.780670	-3.064166	-0.462227
H	0.854407	1.508568	-2.132251
H	1.043954	-0.972319	-2.250077
F	1.364580	2.927389	-0.016095

Ligand 11: Arg complex

Counterpoise corrected energy = -1091.2019
BSSE energy = 0.0007
Sum of monomers = -1091.1808
Complexation energy = -13.69 kcal/mol (raw)
Complexation energy = -13.25 kcal/mol (corrected)

Atom	X	Y	Z
C	-1.975456	-2.694968	-0.503301
N	-2.639045	-1.451879	-0.128570
C	-2.417333	-0.288213	-0.726130
N	-1.546108	-0.203126	-1.721485
N	-3.050760	0.791854	-0.311949
C	0.820459	0.001152	0.937344
C	1.762940	-0.071617	-0.105637
O	2.417336	-1.248566	-0.267693
C	3.487792	-1.287951	-1.228590
C	1.942978	1.065592	-0.905477
C	1.193483	2.222031	-0.679321
C	0.091231	1.162373	1.165301
C	0.255520	2.296511	0.358130
C	-0.622233	3.487464	0.544369
H	-0.890326	-2.578207	-0.429525
H	-2.265487	-3.016322	-1.508950
H	-3.420105	-1.511752	0.510379
H	-1.014375	-1.023124	-1.976930
H	-1.077438	0.679222	-1.892949
H	-2.889627	1.689659	-0.743754
H	-3.601925	0.785809	0.533085
H	3.911016	-2.287092	-1.143231
H	4.246308	-0.536242	-0.989010
H	3.107871	-1.127887	-2.243953
H	2.672800	1.051153	-1.706685
H	1.353319	3.082514	-1.323432
H	-0.617172	1.176130	1.988104
H	-1.078382	3.806414	-0.400805
H	-1.420773	3.286434	1.264465
C	0.552881	-1.393581	1.945275
H	-2.274722	-3.462102	0.208769
H	-0.063557	4.352623	0.920184

Ligand 12: Arg complex

Counterpoise corrected energy = -730.8504
BSSE energy = 0.0008
Sum of monomers = -730.8306
Complexation energy = -12.93 kcal/mol (raw)
Complexation energy = -12.45 kcal/mol (corrected)

Atom	X	Y	Z
C	-2.040209	-2.754727	-0.000375
N	-2.661495	-1.446518	0.168847
C	-2.371403	-0.389868	-0.579276
N	-1.466311	-0.481514	-1.543337
N	-2.970795	0.761643	-0.346741
C	0.838550	0.020495	1.123537
C	1.776571	-0.257487	0.115555
O	2.362618	-1.482872	0.144173
C	3.436810	-1.715067	-0.785053
C	2.007680	0.743493	-0.835971
C	1.307552	1.953146	-0.782599
C	0.156567	1.222632	1.189585
C	0.374026	2.220656	0.225924
C	-0.457395	3.458366	0.208099
H	-0.952730	-2.666927	0.082569
H	-2.318480	-3.212378	-0.955255
H	-3.482385	-1.393285	0.757093
H	-0.976014	-1.355124	-1.672647
H	-0.939521	0.344969	-1.805454

H	-2.761739	1.578859	-0.900938
H	-3.532152	0.903641	0.479573
H	3.809526	-2.710183	-0.549339
H	4.230146	-0.973631	-0.647520
H	3.072962	-1.687020	-1.818595
H	2.735430	0.584029	-1.623248
H	1.505331	2.702961	-1.543766
H	-0.549035	1.372029	2.001775
H	-0.868209	3.653825	-0.789592
H	-1.286456	3.394645	0.918959
H	-2.384501	-3.396775	0.808717
H	0.125130	4.347205	0.477792
F	0.576606	-0.940996	2.044284

Ligand 13: Arg complex

Counterpoise corrected energy = -837.7292
BSSE energy = 0.0004
Sum of monomers = -837.7121
Complexation energy = -11.02 kcal/mol (raw)
Complexation energy = -10.75 kcal/mol (corrected)

Atom	X	Y	Z
C	3.272676	-1.190683	-0.233070
N	2.729857	0.135766	-0.501046
C	1.870085	0.756027	0.296589
N	1.448156	0.170802	1.408660
N	1.416723	1.950340	-0.031105
C	-2.310264	1.965465	-0.094741
H	2.456625	-1.905596	-0.092487
H	3.933069	-1.190322	0.640363
H	3.190773	0.689921	-1.210885
H	1.785893	-0.753931	1.634036
H	0.509007	0.372618	1.739744
H	0.805780	2.461409	0.588555
H	1.566598	2.335769	-0.951558
H	-2.434213	2.395289	0.901857
H	-1.524398	2.508902	-0.627667
H	3.846473	-1.501997	-1.104629
H	-3.242349	2.114592	-0.647915
C	-2.080801	-0.327089	1.152974
C	-1.175056	-1.869397	-0.335704
C	-1.635852	-1.665800	0.939628
H	-2.501941	0.029436	2.087393
H	-1.680959	-2.447227	1.689436
S	-1.302592	-0.433592	-1.283990
C	-1.975708	0.467632	0.032384
H	-0.822519	-2.786675	-0.786307

Ligand 14: Arg complex

Counterpoise corrected energy = -1007.5333
BSSE energy = 0.0005
Sum of monomers = -1007.5137
Complexation energy = -12.57 kcal/mol (raw)
Complexation energy = -12.28 kcal/mol (corrected)

Atom	X	Y	Z
C	2.852846	-2.033901	0.163170
N	3.046038	-0.631908	-0.187520
C	2.452992	0.374417	0.441549
N	1.624138	0.137644	1.448626
N	2.672140	1.613749	0.047415
C	-0.544928	3.336064	-0.712813
H	1.787522	-2.282474	0.131877
H	3.273782	-2.264877	1.147351
H	3.809963	-0.405477	-0.810248
H	1.436820	-0.813233	1.732457
H	0.905610	0.811652	1.684565
H	2.225447	2.389832	0.512462
H	3.072452	1.804312	-0.859615
H	-0.092948	3.718089	0.208485
H	0.164947	3.387626	-1.538912
H	3.361994	-2.640167	-0.584414
H	-1.404381	3.972697	-0.941311
C	-0.974721	1.867245	-0.541145
C	-0.952251	-0.332773	-0.839358
C	-1.747768	-0.300402	0.335219
C	-0.642655	-1.564666	-1.432367
C	-2.232296	-1.472433	0.922455
C	-1.119675	-2.730618	-0.843785
H	-0.057148	-1.586266	-2.345785
C	-1.903497	-2.685176	0.322440
H	-2.852189	-1.443201	1.811898
H	-0.901220	-3.690881	-1.299784
H	-2.273059	-3.608763	0.755524
N	-0.520490	0.906945	-1.290803
S	-1.958666	1.357473	0.836202

Ligand 15: Arg complex

Counterpoise corrected energy = -648.6199
BSSE energy = 0.0005
Sum of monomers = -648.5963
Complexation energy = -15.10 kcal/mol (raw)
Complexation energy = -14.77 kcal/mol (corrected)

Atom	X	Y	Z
C	3.254473	0.006745	-0.623892
N	2.511800	-0.764395	0.365915
C	1.472069	-1.534594	0.072505
N	1.045482	-1.633511	-1.178602
N	0.847640	-2.186188	1.034184
C	-2.790916	-1.375244	1.017433
H	2.570053	0.664884	-1.166387
H	3.800171	-0.645747	-1.313156
H	2.920120	-0.859057	1.286014
H	1.518078	-1.089920	-1.887378
H	0.049973	-1.746297	-1.353870
H	0.183469	-2.913577	0.817896
H	1.041957	-2.004118	2.007019
H	-3.503720	-1.984536	0.464988
H	-1.986452	-1.992744	1.422853
H	3.968904	0.633941	-0.093215
H	-3.297005	-0.846081	1.824855
C	-1.232890	0.585953	0.355956
C	-2.306703	-0.426309	-1.343608
C	-0.750489	1.143170	-0.867068
C	-0.731687	0.973509	1.603387
C	-1.448321	0.478058	-1.930089
H	-3.032377	-1.089504	-1.793480
C	0.235184	2.145535	-0.811983
C	0.258980	1.950572	1.620009
H	-1.117902	0.554465	2.527695
H	-1.376386	0.683526	-2.988758
C	0.730739	2.535076	0.426416
H	0.588373	2.621626	-1.722492
H	0.653402	2.292799	2.571526
H	1.479124	3.319290	0.482780
N	-2.205267	-0.343588	0.035462

4.7 References for Chapter 4

1. Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Muller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E., Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J Am Chem Soc* **2014,** *136* (26), 9308-19.

2. Cortopassi W.A.; Kumar K.; Paton R.S., Cation-π interactions in CREBBP bromodomain inhibition: an electrostatic model for small-molecule binding affinity and selectivity, *Org Bio Chem* **2016**, *14*, 10926-10938.

3. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision B.01. Wallingford CT, **2009**.

4. (a) Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors (Reprinted from Molecular Physics, vol 19, pg 553-566, 1970). *Mol Phys* **2002,** *100* (1), 65-73; (b) Simon, S.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J Chem Phys* **1996,** *105* (24), 11024-11031.

5. Zhao, Y.; Truhlar, D. G., 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* (1-3), 215-241.

6. Pople, J. A.; Headgordon, M.; Raghavachari, K., Quadratic Configuration-Interaction - a General Technique for Determining Electron Correlation Energies. *J Chem Phys* **1987,** *87* (10), 5968-5975.

7. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011,** *32* (7), 1456-65.

8. Raju, R. K.; Bloom, J. W.; An, Y.; Wheeler, S. E., Substituent effects on non-covalent interactions with aromatic rings: insights from computational chemistry. *Chemphyschem* **2011,** *12* (17), 3116-30.

9. Tsuzuki, S.; Yoshida, M.; Uchimaru, T.; Mikami, M., The origin of the cation/pi interaction: The significant importance of the induction in Li+ and Na+ complexes. *J Phys Chem A* **2001,** *105* (4), 769-773.

10. Johnson, B. G.; Gill, P. M. W.; Pople, J. A.; Fox, D. J., Computing molecular electrostatic potentials with the PRISM algorithm. *Chem Phys Lett* **1993,** *206* (1), 239-246.

11. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* **2009,** *113* (18), 6378-96.

12. Mujtaba, S.; Zeng, L.; Zhou, M. M., Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **2007,** *26* (37), 5521-7.

13. Wheeler, S. E.; Houk, K. N., Substituent effects in cation/pi interactions and electrostatic potentials above the centers of substituted benzenes are due primarily to through-space effects of the substituents. *J Am Chem Soc* **2009,** *131* (9), 3126-7.

14. *Spartan'14*, Wavefunction, Inc. Irvine, CA.

15. Paton, R. S.; Goodman, J. M., Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *J Chem Inf Mod* **2009,** *49* (4), 944-955.

16. Case, D. A.; Darden, T. A.; Cheatham, T. E.; 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.; Goetz, 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**.

17. 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 Model* **2006,** *25* (2), 247-260.

18. Miller, B. R., 3rd; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* **2012,** *8* (9), 3314-21.

19. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knappcd, S.; Biggin, P. C., Accurate calculation of the absolute free energy of binding for drug molecules. *Chem Sci* **2016,** *7* (1), 207-218.

20. (a) Ran, T.; Zhang, Z. M.; Liu, K. J.; Lu, Y.; Li, H. F.; Xu, J. X.; Xiong, X.; Zhang, Y. M.; Xu, A. Y.; Lu, S.; Liu, H. C.; Lu, T.; Chen, Y. D., Insight into the key interactions of bromodomain inhibitors based on molecular docking, interaction fingerprinting, molecular dynamics and binding free energy calculation. *Mol Biosyst* **2015,** *11* (5), 1295-1304; (b) Oehme, D. P.; Brownlee, R. T.; Wilson, D. J., Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease. *J Comput Chem* **2012,** *33* (32), 2566-80; (c) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V., Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *J Chem Theory Comput* **2014,** *10* (3), 1228-1241.

21. 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* (39), 9375-9386.

22. Dougherty, D. A., Cation-pi interactions involving aromatic amino acids. *J Nutr* **2007,** *137* (6 Suppl 1), 1504S-1508S; discussion 1516S-1517S.

23. (a) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. *Proc Natl Acad Sci U S A* **1998,** *95* (21), 12088-93; (b) Gallivan, J. P.; Dougherty, D. A., A computational study of cation-pi interactions vs salt bridges in aqueous media: Implications for protein engineering. *J Am Chem Soc* **2000,** *122* (5), 870-874.

24. Mecozzi, S.; West, A. P.; Dougherty, D. A., Cation-pi interactions in simple aromatics: Electrostatics provide a predictive tool. *J Am Chem Soc* **1996,** *118* (9), 2307-2308.

25. (a) 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* (11), 1995-2001; (b) 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* (6), 669-81.

26. Wheeler, S. E., Understanding substituent effects in noncovalent interactions involving aromatic rings. *Acc Chem Res* **2013,** *46* (4), 1029-38.

27. Dougherty, D. A., The cation-pi interaction. *Acc Chem Res* **2013,** *46* (4), 885-93.

28. An, Y.; Doney, A. C.; Andrade, R. B.; Wheeler, S. E., Stacking Interactions between 9-Methyladenine and Heterocycles Commonly Found in Pharmaceuticals. *J Chem Inf Mod* **2016,** *56* (5), 906-914.

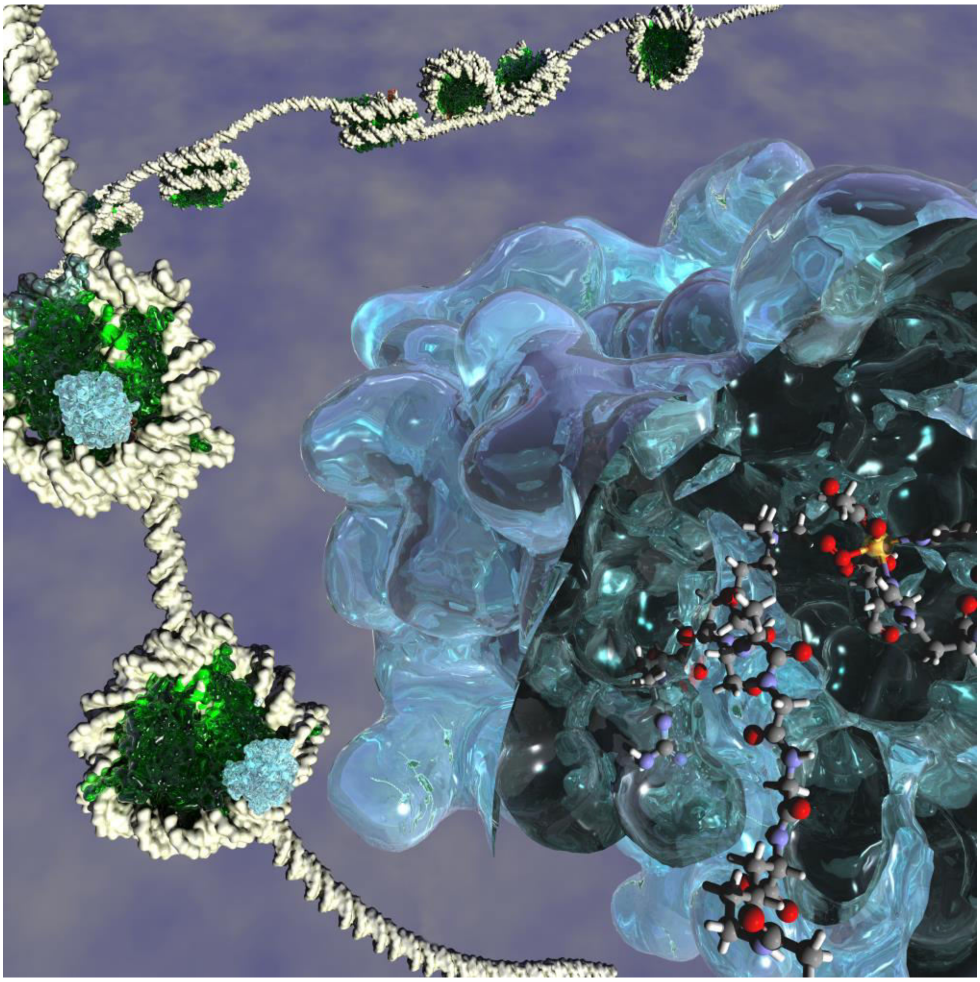
29. Taylor, A. M.; Cote, A.; Hewitt, M. C.; Pastor, R.; Leblanc, Y.; Nasveschuk, C. G.; Romero, F. A.; Crawford, T. D.; Cantone, N.; Jayaram, H.; Setser, J.; Murray, J.; Beresini, M. H.; de Leon Boenig, G.; Chen, Z.; Conery, A. R.; Cummings, R. T.; Dakin, L. A.; Flynn, E. M.; Huang, O. W.; Kaufman, S.; Keller, P. J.; Kiefer, J. R.; Lai, T.; Li, Y.; Liao, J.; Liu, W.; Lu, H.; Pardo, E.; Tsui, V.; Wang, J.; Wang, Y.; Xu, Z.; Yan, F.; Yu, D.; Zawadzke, L.; Zhu, X.; Zhu, X.; Sims, R. J., 3rd; Cochran, A. G.; Bellon, S.; Audia, J. E.; Magnuson, S.; Albrecht, B. K., Fragment-Based Discovery of a Selective and Cell-Active Benzodiazepinone CBP/EP300 Bromodomain Inhibitor (CPI-637). *ACS Med Chem Lett* **2016,** *7* (5), 531-6.

Chapter 5

O_2 binding in JMJD2A: The role of the protein environment

"Begin at the beginning... and go on till you come to the end: then stop."

Lewis Carroll, Alice's Adventures in Wonderland



Cortopassi WA, Simion R, Hornsby CE, França TCC, Paton RS "Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment", *Chem Eur J* **2015**, 21, 18983–18992. Selected for front cover.

5.1 Overview

JMJD2A, a member of the Jumonji C (JmjC) family, catalyses the demethylation of di- and trimethylated lysines in histone tails and has been reported as a promising molecular target for the development of new anticancer drugs. The mechanistic details of demethylation, involving multiple electronic states, still await full elucidation for these fundamental epigenetic processes. In this Chapter, we studied the initial step of the histone demethylation in JMJD2A, showing the O_2 binding in demethylase proteins as energetically favourable after Quantum Mechanics/Molecular Mechanics (QM/MM) calculations. We considered all possible spin states and different conformations of O_2, using QM, MM and QM/MM approaches.

Despite a recent rise in interest surrounding histone demethylation by non-haem iron centres, detailed insights into the mechanism of this process (**Figure 5.1**) have remained elusive.[1] In particular, the structural basis for substrate selectivity (*i.e.* the recognition and demethylation of mono-, di- and tri-methyl lysine residues) within the active site, and the dependence on C-H bond strength in the hydrogen abstraction step, are still poorly understood. Furthermore, the mechanistic details of demethylation and the potential involvement of intersystem crossing (ISC) between different spin states on the potential energy surface have yet to be fully understood for these important epigenetic processes.

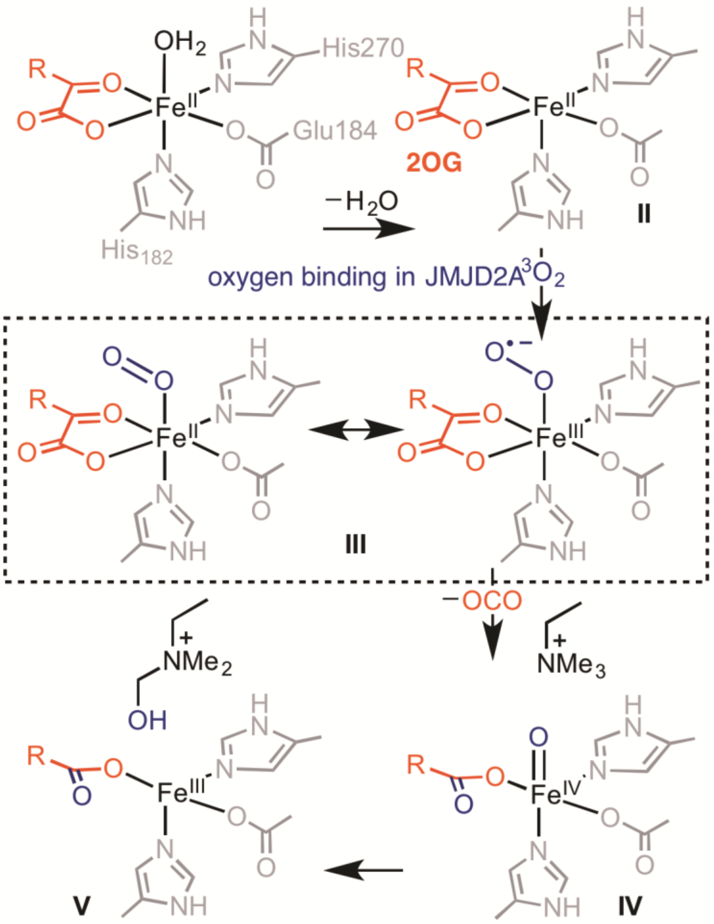


Figure 5.1. Postulated mechanism of O_2-activation and lysine demethylation *via* an Fe^{IV}–oxo intermediate in JMJD2A.[1]

Here we describe calculations to develop an understanding of the mechanism for O_2 binding in histone demethylases; the most stable reaction pathway along potential energy surface proceeds via the quintet spin configuration with the O_2 molecule in a so-called *bent end-on* conformation and the oxidation of Fe^{II} to Fe^{III}. This process is enabled by electrostatic energy contributions due to the proximity of His182 and the methylated lysine, supporting a proposed mechanism where the substrate binds the active site before O_2. Glu184 and the oxoglutarate cofactor 2OG were deleterious for O_2 binding. The inclusion of dispersion effects and free energy corrections were essential to obtaining a favourable final energy of -1.3 kcal mol^{-1} for the O_2 binding in the overall quintet spin state. The inclusion of the full protein structure surrounding the active site treated at the MM level of theory stabilized O_2 binding by -6.7 kcal mol^{-1} compared to the use of QM active site models only, highlighting the importance of the protein environment.

QM/MM techniques[1] have been implemented in our attempts to uncover the steps surrounding demethylation, and to evaluate the contribution of individual amino acids on structures, energetics and dynamics along the reaction coordinate. Computational insights into the mode of action of the most promising residues for demethylation with JmjC proteins will no doubt prove to be invaluable tools in the future testing of enzyme mutants and in the rational design of small-molecule inhibitors.

5.2 Methodology

QM calculations. Density Functional Theory (DFT) calculations were performed on an active site model (ASM) of JMJD2A for the understanding of the electronic structure and geometry changes associated with O_2 binding (proceeding from **II** to **III** in **Figure 5.1**). The quantum chemical analysis of a relatively small but well-chosen cluster from a redox-active metalloenzyme using crystallographic restraints has been pioneered by Siegbahn.[2] Coordinates of the active site were extracted from the crystal structure of JMJD2A (PDB 2OQ6). Our ASM includes the Fe-centre and its coordinating residues His170, His182 and Glu184 along with a truncated model of the 2OG oxoglutarate cofactor. These residues provide a chemical model to capture geometric and electronic properties of the non-haem binding site. Geometry optimization and single point energy calculations were performed with the unrestricted B3LYP (uB3LYP) functional using *Gaussian09*.[3] A double-ζ valence 6-31G(d) basis set was employed for all atoms except Fe, for which the modified

LANL2DZ[4] effective core potential (ECP) and associated double-ζ valence basis set was used with an additional f-type polarization function with exponent of 2.462.[5] Single point energies were evaluated for stationary points with a larger triple-ζ valence polarized 6-311+G(d,p) basis set used for all non metal atoms and with the LANL2TZ(f) ECP and decontracted triple-ζ valence basis set for Fe.[5] A conductor-like polarizable continuum model (CPCM) with ε=4.3 (diethyl ether as solvent)[6] was applied to mimic the dielectric effects of the relatively hydrophobic protein interior.[7] The B3LYP functional incorporates 20% exact HF exchange energy and has been used in the study of oxygen binding to metalloenzymes with good reproducibility of experimental results.[8] The coordinates of the active site were extracted from the crystal structure of JMJD2A at 2Å resolution (PDB 2OQ6).[9] Considering that JMJD2A functions as a monomer,[10] just the active site of the B chain was used in this work. The partitioning of higher and lower layers in QM/MM calculations is shown in **Figure 5.2**. The link atoms between the active site and the protein were kept fixed. In all DFT calculations, we used Grimme's D3-dispersion correction[11] with zero damping at short range.

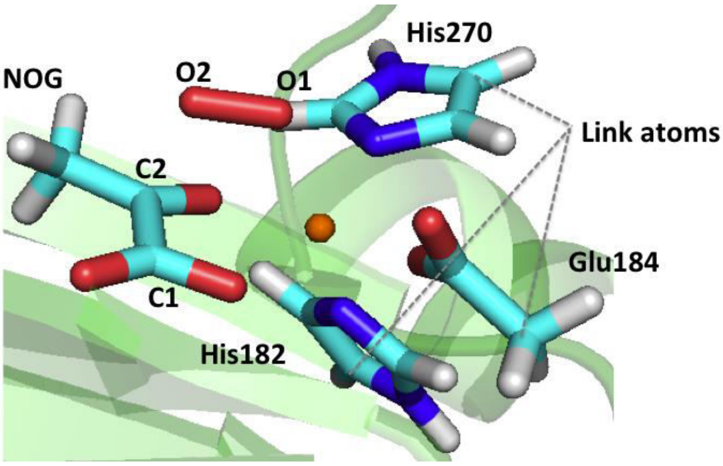


Figure 5.2. Representation of the model of the active site of JMJD2A used in our QM/MM calculations.[1]

Classical MD calculations. To obtain a stabilized structure for QM/MM calculations, the B3LYP-optimized active site (without oxygen) was combined with the surrounding protein atoms at their crystallographic positions. This structure was subjected to 10 ns of molecular dynamics (MD) simulation using the NPT ensemble with the AMBER ff99SB protein force field.[12] The MD simulations were carried out in cubic boxes with dimensions 5.85 nm × 5.85 nm × 5.85 nm (an initial volume of 200.2 nm^3) containing 3541 water molecules and *four* counter-ions (Cl^-) added with the standard *addIons* tool of tLeap,[13] followed by a visual inspection of the complex. Hydrogens were removed from the residues using the MolProbity Server[14] and new hydrogens were added using tLeap.[13] Glutamate and aspartate residues were assigned as negatively charged while lysine and arginine as positive. Histidine and cysteine protonation states were assigned based on a visual inspection of their local environment and comparing with a prediction performed by the MolProbity Server.[14] As a result, His182, His240 and His270 were protonated in the N_δ position, while the other histidine residues were protonated in the N_ε position. Cys234, Cys306 and Cys308 were assigned as being negatively charged. The minimizations were performed in two steps. First, we used steepest descent followed by a conjugate gradient optimization, keeping the active site position restrained (PR). Secondly, the same minimization methodology was performed without PR. The iron centre was treated using the non-bonded methodology with RESP atomic charges derived from QM calculations.[15]

QM/MM calculations. The relaxed system was then used as a starting point for QM/MM calculations to evaluate the importance of the protein environment upon O_2 binding. For this work, calculations were performed using the two-layered hybrid QM/MM optimizations implemented using the “our own N-layered integrated molecular orbital and molecular mechanics" (ONIOM) protocol with electronic embedding.[16] The full system for QM/MM calculations consists of 6071 atoms, 348 residues, a truncated model of the oxuglutarate cofactor to reduce the size of the active site during the QM calculations, trimethylated peptide, zinc, iron and 97 water molecules. The inclusion of dioxygen adds a further two atoms. For the QM region, geometry optimization and single point energy calculations were performed with the unrestricted B3LYP functional and at the same level of theory as previously described for the ASM.[3] For the MM region, the AMBER force field was used.

5.3 Results and discussion

The energetics and geometry changes were analysed for the first step of the demethylation reaction (**Figure 5.1**) by using a relatively small but well-chosen ASM of JMJD2A in our DFT calculations, based on similar studies previously reported by Siegbahn[2] for redox-active metalloenzyme.

Four possible spin states for the ASM were considered in this study: singlet, triplet, quintet and septet, in both end-on and side-on conformations of the dioxygen. Singlet has been reported as the most unstable spin multiplicity,[17] and singlet side-on conformations were not studied in this work. We also did not perform calculations for the triplet side-on configuration due to the difficulties of achieving a true minimum for this state, a problem previously reported in the literature.[17b] For the quintet end-on and septet side-on structures, the iron was oxidized to Fe^{III} and the O_2 was reduced to O_2^-, corroborating previous calculations[17a] and experimental data suggesting that in high spin states Fe^{II} oxidizes to Fe^{III}.[18]

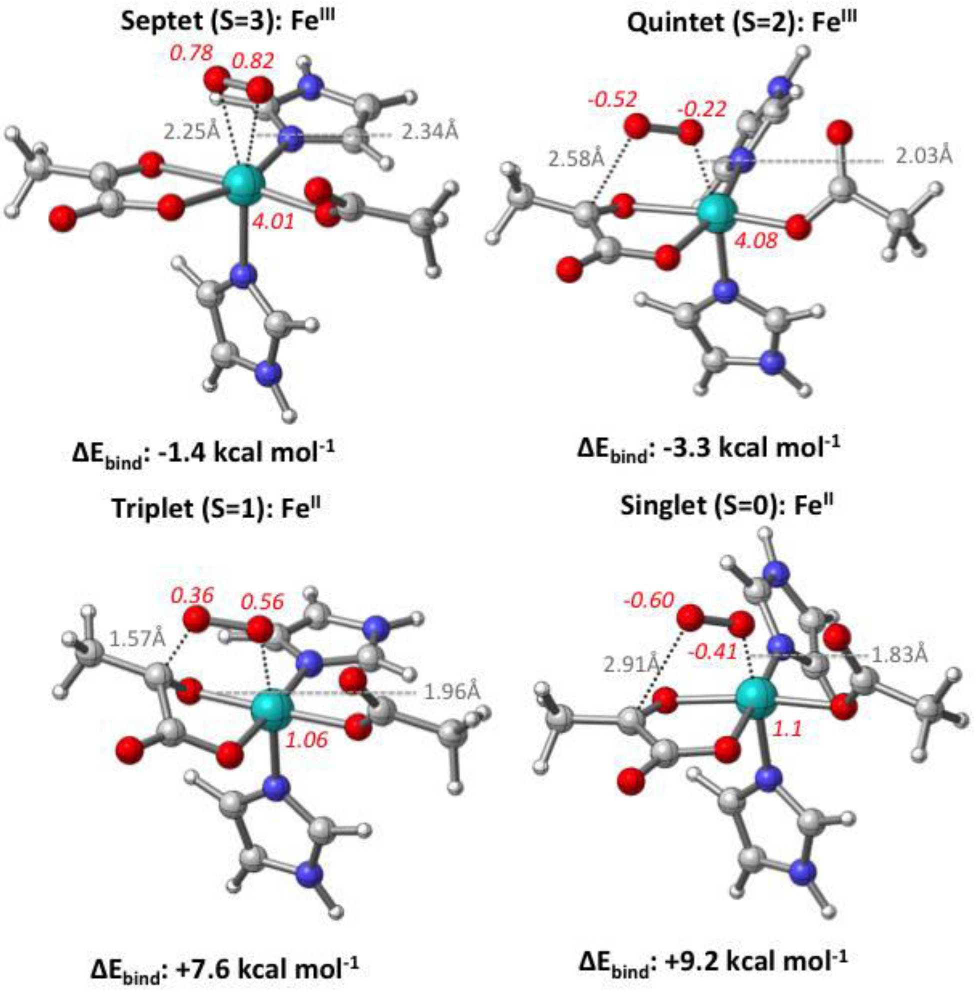


Figure 5.3. Optimized CPCM-UB3LYP-D3 geometries of the O_2 adducts formed at the JMJD2A active site; binding energetics (valence triple-ζ); Mulliken spin densities shown in red.[1]

The singlet and triplet overall spin states were found without charge transfer between the iron and the dioxygen, corroborating previous calculations.[17] The singlet spin state was the least stable configuration, with an energy of +9.2 kcal mol^{-1} higher than the reference complex, suggesting that during the dioxygen binding step the highest overall spins may better describe the active site, corroborating experimental work.[18] The most stable complex was the quintet end-on when computed with 4 out of 5 density functionals (**Figure 5.4**). The same trend was previously reported,[19] showing that $E_{singlet}$ > $E_{triplet}$ > $E_{quintet}$, except for with BP86, a Generalized Gradient Approximation (GGA) functional, in which the triplet complex presented the lowest energy. In our attempt to locate a minimum for septet end-on and quintet side-on conformations, we tested different initial positions for the dioxygen, as well as changing the size of the active site, taking into account the methylated lysine and some neighbouring residues. However, these changes did not affect our results and we did not find a minimum for these complexes.

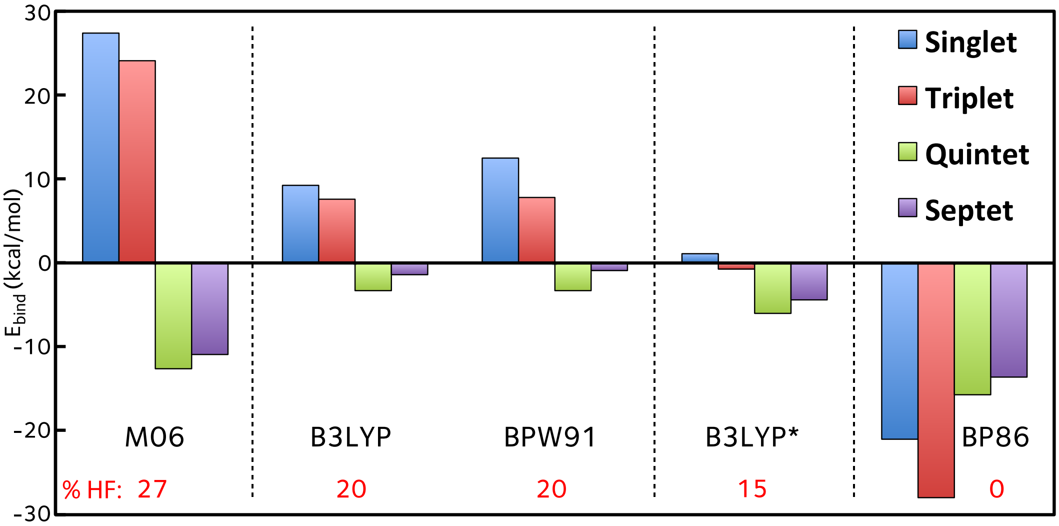


Figure 5.4. D3-corrected density functional comparison of binding energies (kcal mol^{-1}) to the active site of JMJD2A computed for different spin states (valence triple-ζ); percentage of exact exchange included. The proportion of HF exchange is decreased to 15 % for the B3LYP* functional (from 20 % in B3LYP).[1]

The bond length of O_2 in all complexes that showed a spin density higher than 4.0 increased compared to free O_2 (1.21 Å), indicating the reduction of the formal O-O bond order.

Although some calculations have been performed for similar active sites,[17a,20] there is no data in the literature regarding the O_2 binding in the active site of proteins from the JmjC family. Comparison with a previously reported in which they worked with AlkB proteins,[17a] differences in the position of the residues and ligands can be observed (**Figure 5.5**). The presence of charged residue Arg210, which interacts *via* a hydrogen bond with Asp133 in AlkB, may reduce the negative character of the carboxyl group, and thus facilitate the O_2 binding. In JMJD2A, Arg210 is replaced by the neutral residue Asn284, which does not alter the O_2 binding process in our QM calculations. Also, since the O_2 binding process is not favourable according to QM-computed energies of ASM,[17b] slight differences in the coordinates of these residues may considerably affect the QM results.

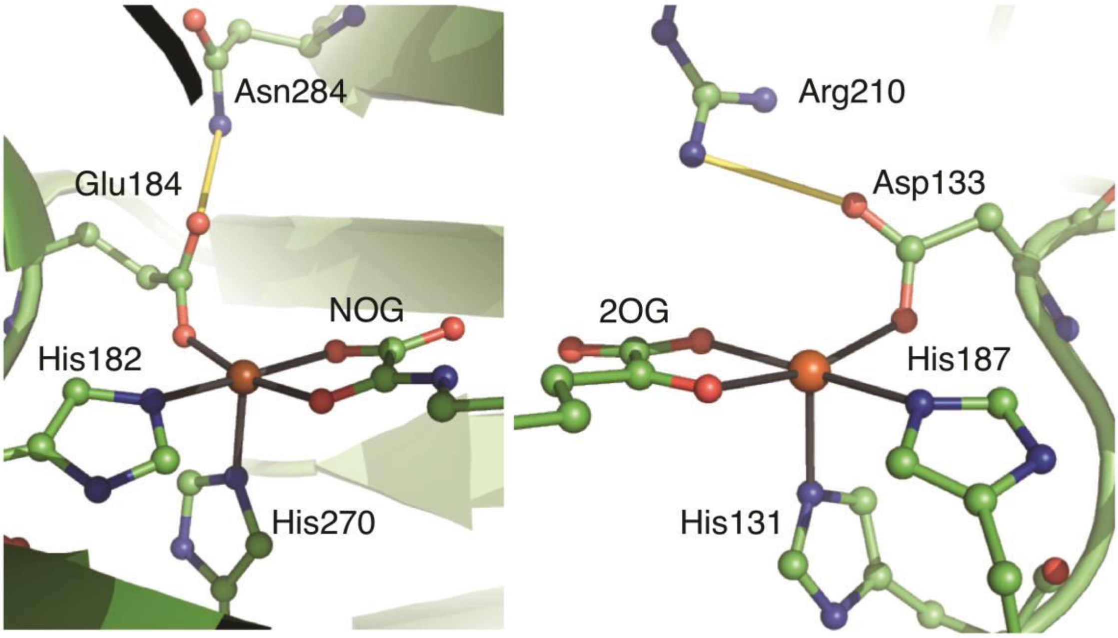


Figure 5.5. The non-haem active sites of JMJD2A (PDB ID: 2OQ6) - left, and AlkB (PDB ID: 2FD8) - right. Active site water molecules are not shown (*N*-oxalylglycine, NOG, is an unreactive analog of 2OG).[1]

A problem that may arise from QM calculations of inorganic complexes in which numerous spin states are considered is the spin contamination.[21] This can affect not only the energy values, but also the geometries and spin densities, which are of great importance in this study. One way of verifying spin contamination in a QM calculation using Gaussian is by checking the expectation value of the total spin $<S^2>$, which is logged in the output file. Without contamination, the $<S^2>$ should be equal to $s \times (s +$

1), where s is half the total number of unpaired electrons. If the <S^2> is different from $s \times (s + 1)$ by less than 10%, the spin contamination can be neglected.[22] All output files were checked after geometry optimization and SPC; no spin contamination was observed, with the exception of the open shell singlet, for which we noticed spin contamination with the triplet spin state.[22] This observed spin contamination is not expected to have a significant role in our understanding for the dioxygen binding since the singlet spin state is not likely to be competitive in this reaction.

By considering just the ASM, we were unable to find a favourable (negative) O_2 binding energy for JMJD2A. These calculations were then expanded for the inclusion of non-active site residues, in our attempt to investigate the role of the protein environment for this step of the reaction. To do this, the ONIOM approach was used in Gaussian, which was previously discussed in **Section 2.5** of **Chapter 2**. MD simulations were performed with the coordinates of the optimized active site preceding the binding of O_2, providing a reasonable starting point for our QM/MM calculations.

The Root Mean Square Deviation (RMSD) analysis can give an idea of how much the 3D structure fluctuates over time, and it allows local fluctuations to be monitored. The temporal RMSD calculations were carried out on all protein atoms of each complex to give 500 frames every 20 ps during the 10 ns of simulation. In **Figure 5.6**, it is possible to see that the system achieved equilibration during the first 3 ns of the simulation, showing an average RMSD around 1.1 Å. Interestingly, fluctuations in this value never exceeded 2.0 Å for the protein. This result suggests stabilization of the system during the MD simulation.

The 2D Root Mean Square Fluctuation (RMSF) per residues was also calculated, as illustrated in **Figure 5.6**. The analysis of the RMSF values gives an idea of the dynamics of the protein during the simulation. The regions that moved the most during the MD simulations (*i.e.* had the largest RMSF values) correspond to the N- and C-terminals. The key result from this section of study is that there was no residue in close proximity to the active site that presented a high fluctuation, suggesting that the new active site derived from our QM calculations did not substantially affect the protein environment.

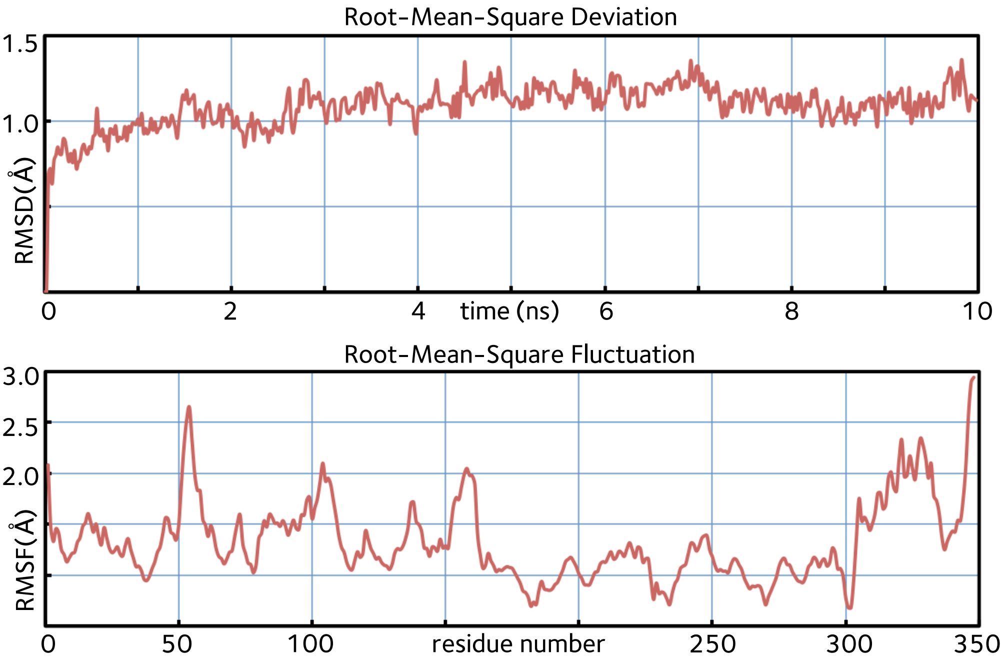


Figure 5.6. Top: Temporal RMSD values relative to the JMJD2A crystal structure. Bottom: RMSF profiles of residues during 10 ns MD simulation.[1]

The total energy during the MD was constant at around 33,000 Joules. The temperature, volume and water density were almost constant at around 310 K, 160,000 Å^3 and 1.1 g cm^{-3}, respectively.

It is also important to analyse distances between the iron and coordinating water molecules present in the crystallographic structure. Water molecules close to iron are crucial when looking at the specificity for the tri-methylated peptide in the active site of JMJD2A.[9, 23] Before the MD simulation, this distance was 2.15 Å. However, after 4 ns of MD simulation, no water molecule within a 2.5 Å radius of iron was observed. This means that water molecules are dissociated during the MD simulation, which is essential before the subsequent steps involving oxygen binding and activation (**Figure 5.1**).[9, 23]

With the optimised structure of the JMJD2A taken from our MD simulations, we were then able to analyse the O_2 binding in its non-haem active site by performing ONIOM QM/MM calculations (**Figure 5.7**). Only calculations with the quintet end-on and side-on complexes showed charge transfer, indicating the oxidation of Fe^{II} to Fe^{III}, and thus corroborating experimental data.[18] As previously observed for the septet end-on complex through our QM analysis, the dioxygen does not bind to the active site in this spin state. Similar results were observed for QM and QM/MM calculations in the quintet and triplet potential energy surfaces (PES).

In all optimized structures considered in this work, the protein environment enabled favourable O_2 binding. The energy contribution coming from the low level of theory (MM) varies from -6.0 kcal mol^{-1} to -12.3 kcal mol^{-1} in the quintet side-on and the triplet complexes (**Figure 5.7**) respectively, highlighting the importance of the protein environment for these calculations. The quintet end-on complex presents the lowest energy compared to the other analysed spin states, suggesting that the O_2 binding may happen in the quintet PES, with oxidation of Fe. This is in accordance with experimental data, which shows that this process in JMJD2A occurs in high spin states.[18]

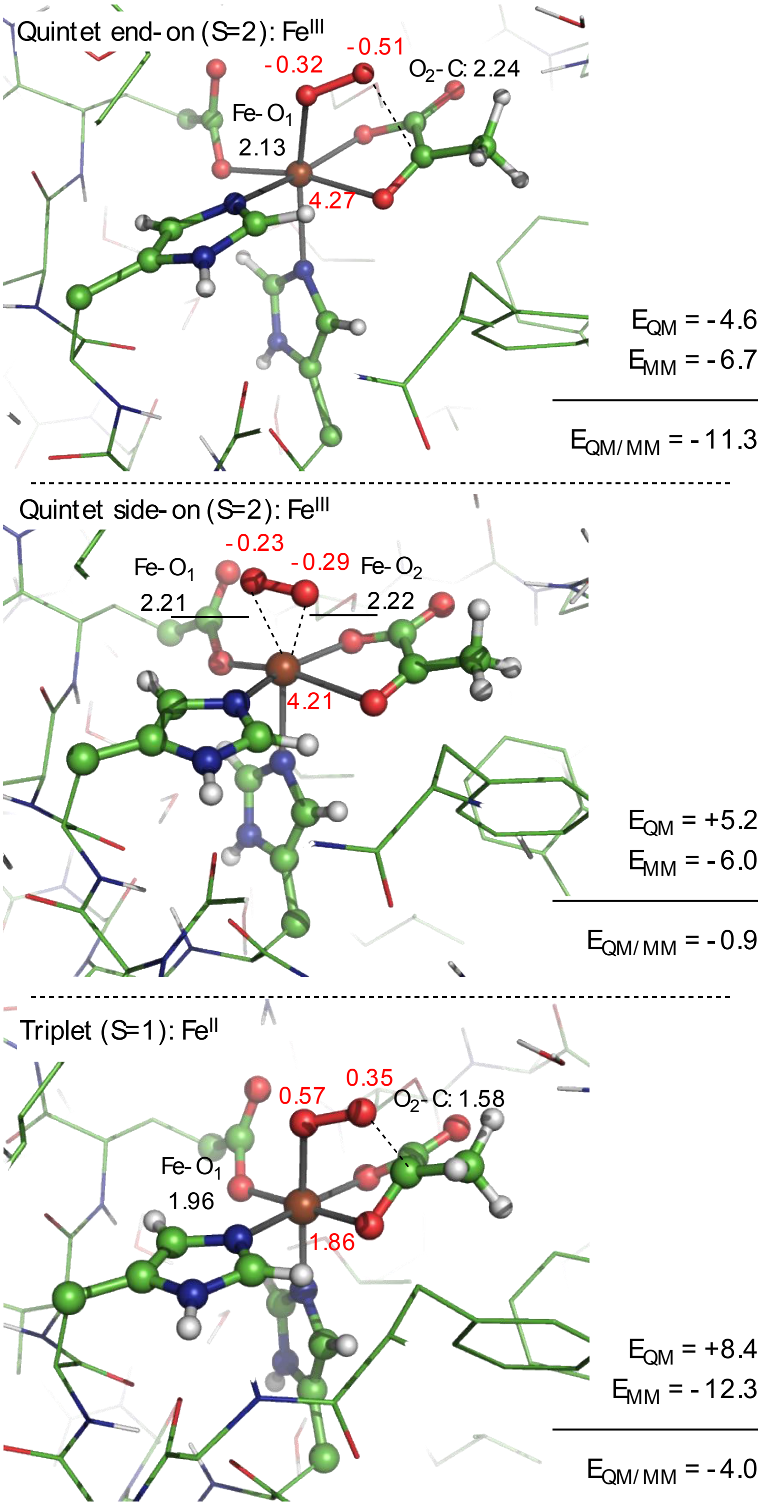


Figure 5.7. QM/MM (UB3LYP-D3/6-31G(d):AMBER) optimized JMJD2A active site in different spin states with high level region shown in bold. TZ-corrected ONIOM energies quoted relative to separated triplet O_2 and quintet enzyme in kcal·mol^{-1}. Distances in Å; Mulliken spin densities in red.[1]

The resulting QM/MM protein structure obtained after the dioxygen binding is similar to the crystal structure of JMJD2A preceding the dioxygen binding (PDB 2OQ6) with an RMSD of 1.27 Å computed using the Visual Molecular Dynamics (VMD)[24] software (**Figure 5.8**). In addition, the active site of the quintet end-on complex superimposes well with the active site of the crystallographic structure of JMJD2A (**Figure 5.9**).

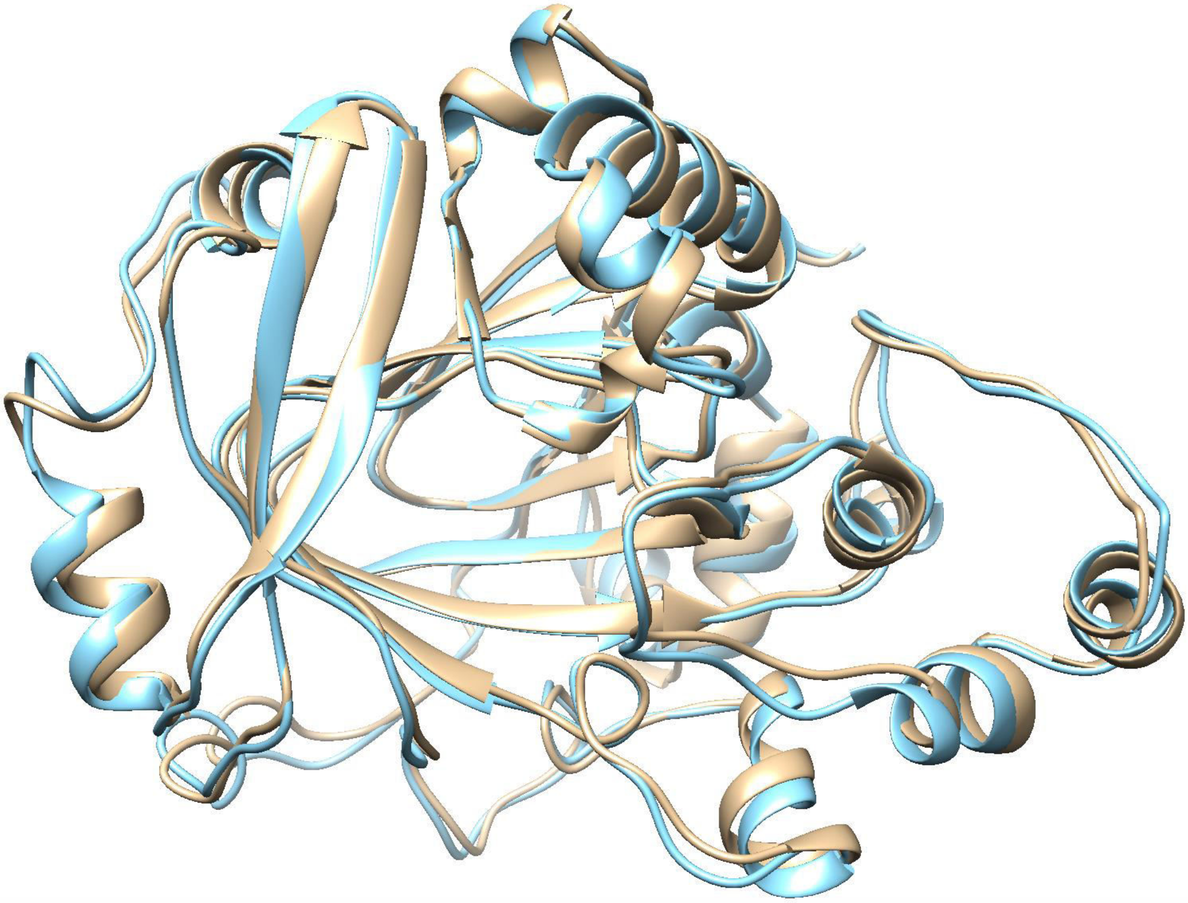


Figure 5.8. Superposition of the chain B of JMJD2A (PDB ID: 2OQ6), in blue, with the protein obtained after QM/MM calculations in the quintet overall spin state in the end-on conformation, in gold.

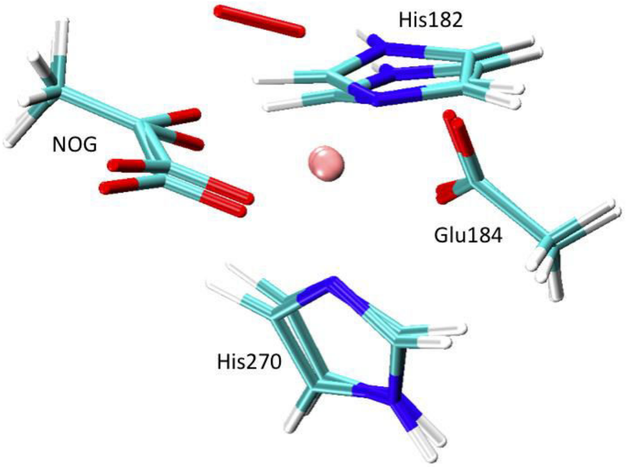


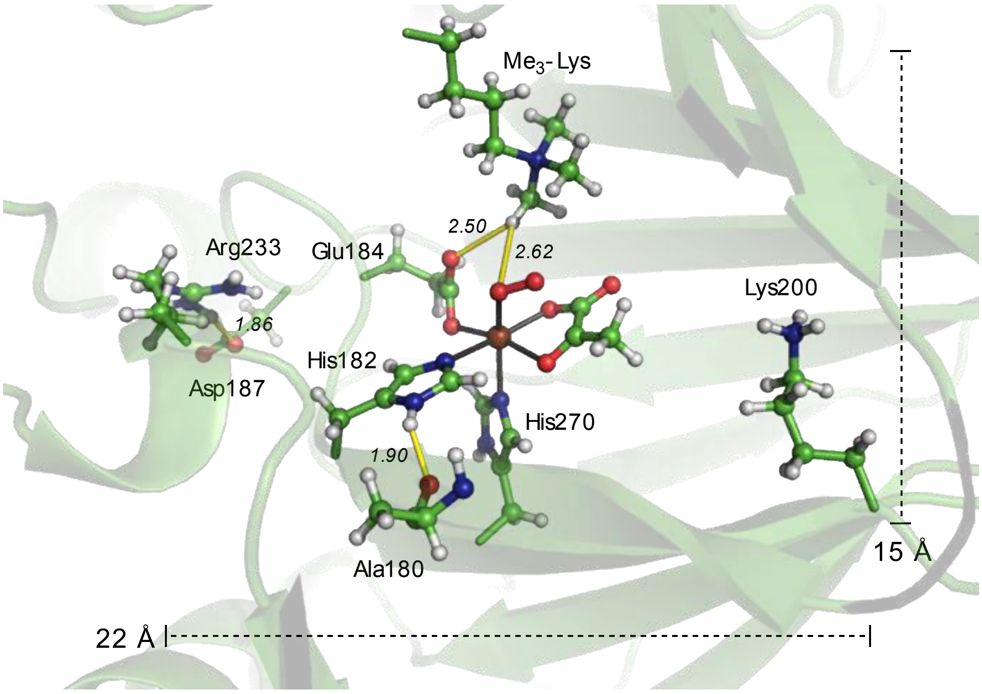
Figure 5.9. Superposition of the active site of chain B of JMJD2A (PDB ID: 2OQ6) with the active site obtained after QM/MM calculations in the quintet overall spin state in the end-on conformation. O_2 is not shown in the crystal structure.[1]

As with our previous QM study of this spin state, no spin contamination was found when analysing the QM/MM results here.

O_2 binding in JMJD2A happens in the quintet PES. Following QM and QM/MM calculations, our results suggest that the complexation of O_2 in this active site happens in the quintet overall spin state.

Although we must heed when applying Kohn-Sham DFT calculations to systems with multi-reference character, our choice of the unrestricted B3LYP functional for optimizations follows work by Neese who found good agreement between spectroscopic and multi-reference calculations.[25] Modifications to the amount of HF exchange in this functional have been shown to affect the O_2 binding considerably.[26] In this work, we found that reducing the HF exchange from 20% to 15% stabilizes the O_2 binding process by -2.9 kcal mol^{-1} (**Figure 5.10**), corroborating previous calculations with other non-haem iron enzymes.[26a]

When using the ASM in our QM calculations, we found an overall energy of -3.3 kcal mol^{-1} for the O_2 binding in the quintet spin state. In our ONIOM QM/MM calculations, this energy was lowered to -14.1 kcal mol^{-1} (**Figure 5.11**), with -6.7 kcal mol^{-1} coming from the low level of theory (MM), highlighting the importance of non-active site residues for active sites containing non-haem iron.[1,17b,26a] Free energy corrections were computed at the UB3LYP-D3/6-31G(d),LANL2DZ level of theory (a detailed description of this methodology can be found in **Section 2.6.2** of **Chapter 2**). This correction was responsible for +12.8 kcal mol^{-1} (**Figure 5.10**), corroborating previous values reported in the literature.[17b, 26a, 27] The corrected O_2 binding energy is then predicted to be -1.3 kcal mol^{-1}, which is still favourable for the O_2 binding, evidencing the importance of the protein environment when trying to understand this process.



	B3LYP:AMBER	B3LYP*:AMBER
$\Delta E_{QM/MM}$	-11.3	-14.1
ΔE_{QM}	-4.6	-7.4
ΔE_{MM}	-6.7	-6.7
ZPE, enthalpic, entropic corrections	+12.8	
$\Delta G^0_{QM/MM}$ *(298K, 1 mol/l)*	**+1.5**	**-1.3**

Figure 5.10. B3LYP*:AMBER optimized JMJD2A quintet active site. TZ-corrected ONIOM energies quoted relative to separated triplet O_2 and quintet enzyme in kcal•mol^{-1}. Residues close to the active site which contribute favourably to O_2 binding are shown.[1]

Our results show that Grimme's D3-dispersion correction[11] lowers the energy of O_2 binding by 5 kcal mol^{-1}, and is thus of great importance for this process, as previously described for the MIOX protein and the Cytochrome P450 enzymes.[26a,28] Just the interaction between the 2OG oxoglutarate cofactor and O_2 is responsible for around -2 kcal mol^{-1}.

The ONIOM QM/MM energies were deconstructed at the low level of theory (MM) into their Coulombic electrostatic and vdW contributions by using a methodology similar to Hirao,[26a] described in **Section 2.6.1** of **Chapter 2**. The sum of vdW and electrostatic energies for all protein residues (-8.1 kcal mol^{-1}) was slightly different from the QM/MM ONIOM computed energies (-6.7 kcal mol^{-1}). This difference comes from other bonded energies that were not considered in our analysis, *e.g.* due to stretching, bending, torsion and out-of-plane torsion.[26a, 29]

The analysis of the per-residue contribution showed that only few residues had a large impact on the dioxygen binding (**Figure 5.11**). The largest per-residue electrostatic energy comes from Glu184, with a positive value of +3.7 kcal mol^{-1}, indicating that this residue reduces the favourability of the first step of the reaction. Other charged residues (Arg233 and Asp187) also had a significant role in this process and interestingly they both interact with Glu184 *via* hydrogen bonds. His182 is shown to be favourable for the O_2 binding, while the charged residue Asp129, which interacts with the methylated substrate *via* hydrogen bonding, is deleterious. The substrate M-Lysine facilitates the O_2 binding by contributing with approximately -2 kcal mol^{-1} when considering electrostatic and vdW energies as well as dispersion effects. Interestingly, electrostatics (-6.2 kcal mol^{-1}) is shown to have a bigger role (three times in magnitude) for the energetics of the dioxygen binding than vdW interactions (-2.1 kcal mol^{-1}), which is due to the weaker distance dependence of the former (electrostatics $\propto r^{-1}$; vdW $\propto r^{-6}$).

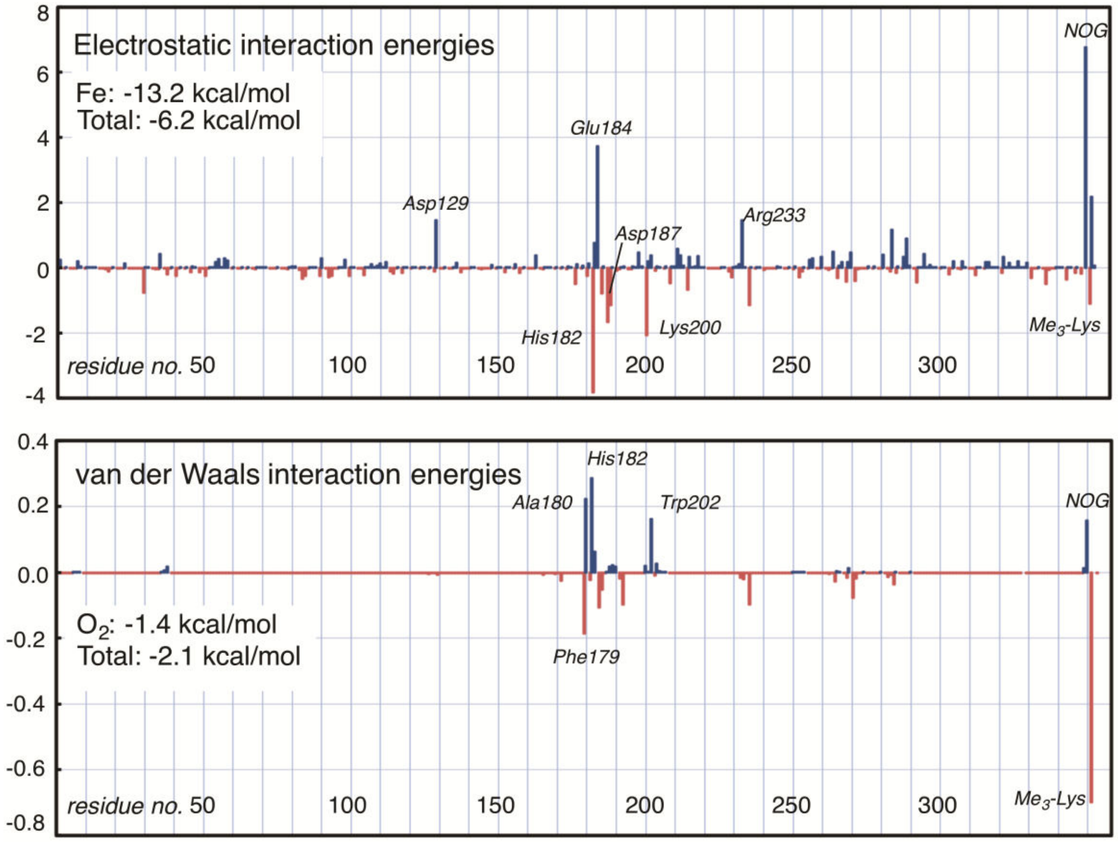


Figure 5.11. Per-residue energy decomposition analysis (kcal mol^{-1}) of O_2 binding in JMJD2A based on a QM/MM description.[1]

5.4 Conclusion

The mechanism of O_2 binding in the histone demethylase JMJD2A was analysed, finding that this happens preferentially on the quintet PES, with O_2 in a so-called "end-on" conformation and oxidation of Fe^{II} to Fe^{III}. This binding mode is favoured by sizable electrostatic energy contributions and dispersion effects due to His182 and M-lysine, which suggests that the substrate must be bound in the active site before O_2 binding. The inclusion of dispersion effects and free energy corrections was essential in obtaining the favourable final energy of -1.3 kcal mol^{-1} for the O_2 binding in the quintet overall spin state. The QM/MM calculations stabilized the O_2 binding by -6.7 kcal mol^{-1} compared to the QM active site models. This highlights the importance of the protein environment, since “small model” calculations are shown to be insufficient in capturing these important contributions to this step of the reaction. Since Coulombic electrostatic interactions operate over a longer distance than other non-bonding interactions (they decay with the inverse of distance), our findings should be considered more generally: models of any biochemical process for which there is considerable charge separation occurring in the active site may also display large energetic contributions from remote, particularly those charged, residues.

5.5 Supporting Information for Chapter 5

5.5.1 DFT Calculations

Stationary points

All DFT-computed stationary points were fully optimized using *Gaussian 09* revision D.01[1] with default convergence criteria for the optimizations and grid spacing for numerical integrations. All minima were verified by the presence of zero imaginary frequencies, following calculation of harmonic vibrations.

Active Site Model

Table S5.1: Computed structural data and Mulliken spin densities for the active site model of JMJD2A (PDB ID: 2OQ6) after DFT (UB3LYP-D3) optimization. Comparison is drawn with previously obtained data for the active site of non-haem AlkB.

Structure	**Distance (Å)**				**Angle (°)**	**Mulliken spin density**			**<S²>**
	$Fe-O_1$	$Fe-O_2$	O_1-O_2	O_2-C_2	$Fe-O_1-O_2$	O_1	O_2	Fe	
Septet η^1 (*JMJD2A*)[a]	2.37	3.19	1.23	3.16	122	0.86	0.91	4.00	12.00
Septet η^1 (*AlkB*)[b]	2.37	3.24	1.24	2.92	125	0.85	0.89	3.98	-
Septet η^2 (*JMJD2A*)[a]	2.25	2.34	1.26	4.33	78	0.78	0.82	4.10	12.00
Septet η^2 (*AlkB*)[b]	2.33	2.40	1.25	3.86	77	0.86	0.83	3.99	-
Quintet η^1 (*JMJD2A*)[a]	2.04	2.92	1.28	2.58	121	-0.22	-0.52	4.24	6.01
Quintet η^1 (*AlkB*)[b]	2.25	3.10	1.25	2.83	122	-0.16	-0.36	4.12	-
Quintet η^2 (*JMJD2A*)[a]	2.19	2.25	1.29	4.21	76	0.65	0.75	2.57	6.01
Quintet η^2 (*AlkB*)[b]	2.19	2.25	1.28	3.79	76	-0.30	-0.24	4.14	-
Triplet η^1 (*JMJD2A*)[a]	1.96	2.72	1.33	1.57	110	0.56	0.35	1.06	2.00
Triplet η^1 (*AlkB*)[b]	2.30	3.15	1.22	3.19	124	-0.92	-0.99	3.78	-
Singlet η^1 (*JMJD2A*)[a]	1.83	2.73	1.26	2.91	123	-0.41	-0.59	1.10	0.37
Singlet η^1 (*AlkB*)[b]	1.90	2.74	1.29	2.60	117	-0.51	-0.58	1.15	-

[a] this work; [b] as previously reported for *AlkB* at the B3LYP/LACVP(d) level of theory (H. Liu, J. Llano and J. W. Gauld, *J Phys Chem B* **2009**, *113*, 4887-4898.)

Table S5.2. Absolute energies (Hartree) for O_2 binding obtained using different density functionals.

	MO6	**B3LYP**	**B3PW91**	**B3LYP***	**BP86**
$^3O_2$	-150.290884	-150.3704108	-150.3093513	-149.556835	-150.3800457
Quintet active site	-1146.040277	-1146.6967	-1146.31881	-1139.605503	-1146.775459
O_2-bound singlet η^1	-1296.263999	-1297.026834	-1296.581997	-1289.143832	-1297.186445
O_2-bound triplet η^1	-1296.292783	-1297.055002	-1296.615748	-1289.163508	-1297.200075
O_2-bound quintet η^1	-1296.351298	-1297.072474	-1296.633465	-1289.171925	-1297.181964
O_2-bound septet η^2	-1296.348461	-1297.069431	-1296.629606	-1289.169353	-1297.172005

Single point energy calculations were performed with the 6-311+G(d,p) basis set and LANL2TZ(f) effective core potential.

Table S5.2*. Relative energies (kcal mol^{-1}) for O_2 binding obtained using different density functionals.

	MO6	**B3LYP**	**B3PW91**	**B3LYP***	**BP86**
O_2-bound singlet η^1	27.4	9.2	12.5	1.1	-21
O_2-bound triplet η^1	24.1	7.6	7.8	-0.73	-27.97
O_2-bound quintet η^1	12.6	-3.3	-3.3	-6	-15.7
O_2-bound septet η^2	-10.9	-1.4	-0.9	-4.4	-13.6

Single point energy calculations were performed with the 6-311+G(d,p) basis set and LANL2TZ(f) effective core potential.

Table S5.3. Basis set effects for O_2 binding obtained with the B3LYP* functional. The proportion of HF exchange is decreased to 15 % for the B3LYP* functional (from 20 % in B3LYP).

	6-311++G(d,p)	def2TZVPP	def2QZVPP
$^3O_2$	-149.556835	-149.575933	-149.584224
Quintet active site	-1139.605503	-1139.723935	-1139.798731
O_2-bound quintet η^1	-1289.171925	-1289.309952	-1289.390194

5.5.2 QM/MM Calculations

RESP calculations

Table S5.4. Coordinates of the reference active site after QM and RESP charges calculated using the R.E.D. Server.

Atom	Coordinates			RESP charges	Mulliken charges
	X	Y	Z		
C	-40.164985	-53.858006	8.407989	-0.2340	0.010
N	-41.315071	-53.78879	7.645573	-0.1206	-0.524
C	-39.237045	-54.480589	7.615508	-0.0537	-0.014
C	-41.060409	-54.350831	6.441412	-0.0267	0.230
N	-39.807797	-54.780044	6.392365	-0.1264	-0.461
C	-35.065016	-57.201983	5.780005	-0.2511	-0.529
C	-36.581129	-57.034265	5.73238	0.6820	0.605
O	-37.342333	-58.006676	5.70512	-0.6423	-0.522
O	-36.959426	-55.788306	5.69213	-0.5616	-0.624
C	-37.657999	-51.146011	4.399006	-0.2340	0.011
N	-36.448613	-51.74885	4.685207	-0.1206	-0.527
C	-38.547729	-52.169049	4.214872	-0.0537	-0.028
C	-36.633571	-53.087882	4.672409	-0.0267	0.248
N	-37.896245	-53.372706	4.388783	-0.1264	-0.464
Fe	-38.640416	-55.360811	4.684913	0.6265	0.712
C	-39.389956	-56.045656	1.960795	0.5656	0.539
C	-40.600752	-55.163403	2.379801	0.4257	0.374
O	-39.354653	-56.465887	0.806153	-0.5874	-0.492
O	-38.537271	-56.238056	2.905199	-0.5500	-0.589
O	-40.631689	-54.707527	3.520131	-0.4122	-0.415
H	-38.206256	-54.735046	7.809321	0.1303	0.189
H	-41.772803	-54.409985	5.634348	0.1501	0.211

H	-34.637891	-56.578071	6.572266	0.0620	0.158
H	-34.641927	-56.857247	4.829044	0.0620	0.172
H	-39.600895	-52.136016	3.980034	0.1303	0.172
H	-35.881848	-53.82399	4.908226	0.1501	0.217
H	-35.576342	-51.274719	4.874632	0.3253	0.376
H	-42.199394	-53.394818	7.935987	0.3253	0.376
H	-34.787299	-58.246701	5.938024	0.0620	0.157
C	-41.673639	-54.910763	1.367509	-0.2691	-0.529
H	-41.229563	-54.458546	0.47386	0.0983	0.204
H	-42.095548	-55.869163	1.043954	0.0983	0.205
H	-42.454048	-54.268032	1.779377	0.0983	0.176
H	-40.120979	-53.477251	9.416163	0.2022	0.189
H	-37.764371	-50.073565	4.348191	0.2022	0.185

Single point energy calculations were performed with the unrestricted B3LYP functional and a double-ζ valence 6-31G(d) basis set was employed for all atoms except Fe, for which the modified LANL2DZ[4] effective core potential (ECP) and associated double-ζ valence basis set was used with an additional f-type polarization function with exponent of 2.462[5].

Table S5.5. Coordinates of the zinc site after QM and RESP charges calculated using the R.E.D. Server.

Atom		Coordinates		RESP charges	Mulliken charges
	X	Y	Z		
C	-41.917006	-55.896986	22.730998	-0.0795	-0.563
H	-42.694931	-56.478878	23.234711	0.0529	0.185
H	-42.404981	-55.175485	22.066099	0.0529	0.153
H	-41.340872	-55.344376	23.480221	0.0529	0.149
S	-40.78076	-57.016538	21.78006	-0.7154	-0.516
Zn	-42.399841	-58.42747	20.656228	0.6606	0.613
N	-43.652215	-57.101578	19.384761	-0.0749	-0.457
C	-43.725158	-57.01037	18.01021	-0.2624	-0.019
H	-43.056239	-57.582207	17.384334	0.2051	0.180
C	-44.738957	-56.156044	17.662054	-0.1275	-0.002
H	-45.115611	-55.82417	16.707099	0.1717	0.177
N	-45.282786	-55.72237	18.856364	-0.2107	-0.523
H	-46.061491	-55.087322	18.961906	0.3198	0.368
C	-44.602243	-56.316835	19.866737	0.1343	0.223
H	-44.829468	-56.192375	20.914504	0.0926	0.208
S	-41.66391	-60.128451	19.09312	-0.7154	-0.503
C	-39.951029	-59.679986	18.538991	-0.0795	-0558
H	-39.940933	-58.7283	17.997188	0.0529	0.164
H	-39.576516	-60.461776	17.870198	0.0529	0.151
H	-39.268052	-59.59703	19.390779	0.0529	0.172
S	-44.291115	-58.959801	22.095932	-0.7154	-0.530
C	-43.457008	-59.630984	23.601957	-0.0795	-0.560
H	-44.157544	-59.637385	24.444089	0.0529	0.145
H	-42.592757	-59.01441	23.871227	0.0529	0.178
H	-43.1083	-60.656044	23.437216	0.0529	0.165

Single point energy calculations were performed with the unrestricted B3LYP functional and a double-ζ valence 6-31G(d) basis set was employed for all atoms except Fe, for which the modified LANL2DZ ECP and associated double-ζ valence basis set was used with an additional f-type polarization function with exponent of 2.462. Charges of nitrogen, sulphur and zinc atoms were further adjusted to neutralize the system.

Table S5.6. Comparison of the energies (Hartree) from ONIOM (B3LYP:AMBER) calculations.

	Low-Model	High-Model	Low-Real
Reference B3LYP	-0.100724	-1146.750765	-8.426061
Reference B3LYP*	-0.100724	-1139.661558	-8.426061
Quintet end-on B3LYP	0.001166	-1297.128462	-8.334816
Quintet end-on B3LYP*	0.001166	-1289.230099	-8.334816
Oxygen B3LYP	-	-150.3703078	-
Oxygen B3LYP*	-	-149.5566497	-
Triplet end-on B3LYP	3.610415	-1297.107736	-4.734610
Quintet side-on B3LYP	-0.070058	-1297.112853	-8.405019

Table S5.7. Free energy corrections (hartree) for the O_2 binding during QM/MM calculations.

Corrections (hartree)						
	Energy	**ZPE**	**Enthalpy**	**T*ΔS**	**ΔG**	**Dif.**
C= 1 mol L^{-1}						
Quintet	-1305.0504	7.1468	-1297.4555	0.8461	-1298.3016	6.7488
Reference	-1154.7240	7.1396	-1147.1376	0.8448	-1147.9825	6.7416
Oxygen	-150.3200	0.0038	-150.3130	0.0203	-150.3332	-0.0132
Ref+Oxy						6.7284
Correction						0.0204
Correction (kcal mol^{-1})						12.8036
C= 50 mol L^{-1}						
Quintet	-1305.0504	7.1468	-1297.4555	0.8424	-1298.2979	6.7525
Reference	-1154.7240	7.1396	-1147.1376	0.8411	-1147.9788	6.7453
Oxygen	-150.3200	0.0038	-150.3130	0.0166	-150.3295	-0.0095
Ref+Oxy						6.7358
Correction						0.0167
Correction (kcal mol^{-1})						10.4881
C= 100 mol L^{-1}						
Quintet	-1305.0504	7.1468	-1297.4555	0.8813	-1298.3368	6.7136
Reference	-1154.7240	7.1396	-1147.1376	0.8802	-1148.0179	6.7061
Oxygen	-150.3200	0.0038	-150.3130	0.0159	-150.3289	-0.0088
Ref+Oxy						6.6973
Correction						0.0163
Correction (kcal mol^{-1})						10.2207

Cartesian coordinates:

Singlet η^1 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.637226
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.052363
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.267391
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.344161
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.42703

C	-40.164993	-53.858001	8.407993
N	-41.090613	-54.811141	8.032812
C	-39.188009	-53.897943	7.447239
C	-40.66655	-55.386799	6.885377
N	-39.515903	-54.851319	6.502601
C	-35.065004	-57.201998	5.780003
C	-36.551255	-57.029077	5.458573
O	-37.224853	-58.005456	5.119995
O	-36.979597	-55.799756	5.555200
C	-37.658003	-51.146001	4.399004
N	-36.416476	-51.73554	4.545413
C	-38.562403	-52.169148	4.502912
C	-36.592443	-53.06482	4.726406
N	-37.883357	-53.355382	4.705609
Fe	-38.617056	-55.246563	4.727279
C	-38.699546	-55.378015	2.015279
C	-40.069642	-54.857446	2.502971
O	-38.52379	-55.536393	0.809526
O	-37.855512	-55.535746	2.975806
O	-40.184523	-54.55576	3.701522
H	-38.285658	-53.315502	7.357323
H	-41.197075	-56.168853	6.366422
H	-34.759299	-56.53995	6.595386
H	-34.485573	-56.931406	4.889013
H	-39.639062	-52.146505	4.447050
H	-35.806136	-53.788567	4.865340
H	-35.524233	-51.262182	4.510093
H	-41.936258	-55.051617	8.531284
H	-34.852064	-58.242214	6.034314
C	-41.173034	-54.665888	1.525034
H	-40.810961	-54.066993	0.682915
H	-41.437648	-55.646866	1.111452
H	-42.044002	-54.204706	1.993878
O	-39.936833	-57.302548	3.704799
O	-39.402099	-56.834313	4.761346
H	-40.277828	-53.265245	9.301782
H	-37.777538	-50.085545	4.243492

Triplet η^1 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.658224
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.055002
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.257926
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.377479
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.453426

C	-40.165002	-53.857991	8.407996
N	-38.930801	-54.336447	8.797527
C	-40.251848	-54.102019	7.064971
C	-38.308653	-54.844015	7.711048
N	-39.087339	-54.713495	6.647729
C	-35.065002	-57.201985	5.779992
C	-36.460804	-57.013045	5.195563
O	-37.047321	-57.919681	4.619162
O	-36.928171	-55.795841	5.405681
C	-37.657996	-51.146024	4.399012
N	-36.417795	-51.693876	4.659993
C	-38.530317	-52.199068	4.370033
C	-36.559190	-53.032115	4.784241
N	-37.829795	-53.362742	4.615473
Fe	-38.599534	-55.177597	4.753777
C	-39.364274	-55.713333	2.252574
C	-40.575514	-55.535752	3.225336
O	-39.500989	-55.960618	1.063201
O	-38.251806	-55.545666	2.905674
O	-40.271277	-54.602143	4.143518
H	-41.046889	-53.892671	6.367951
H	-37.323142	-55.278997	7.709499
H	-35.075441	-57.003334	6.857685
H	-34.366507	-56.491639	5.323697
H	-39.591930	-52.217588	4.186831
H	-35.762315	-53.724408	4.999393
H	-35.544747	-51.189432	4.730047
H	-38.554873	-54.325994	9.735772
H	-34.715346	-58.220554	5.601119
C	-41.946395	-55.455346	2.601306
H	-42.007592	-54.535610	2.012667
H	-42.111564	-56.301937	1.930736
H	-42.716691	-55.435224	3.377143
O	-40.555653	-56.909186	3.979171
O	-39.547324	-56.897131	4.844287
H	-40.854013	-53.408297	9.105333
H	-37.800150	-50.087239	4.251319

Quintet η^1 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.662405
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.072474
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.253775
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.382301
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.468706

C	-40.16499	-53.858021	8.408003
N	-39.839026	-55.165614	8.72091
C	-39.860008	-53.710671	7.083563
C	-39.35671	-55.763547	7.609953
N	-39.353148	-54.902144	6.600974
C	-35.065013	-57.202001	5.780012
C	-36.57337	-57.020757	5.789273
O	-37.313252	-57.709353	6.489507
O	-36.973786	-56.053388	4.997116
C	-37.657998	-51.145978	4.398984
N	-36.485343	-51.759558	4.79445
C	-38.559585	-52.157842	4.209957
C	-36.698979	-53.092804	4.83861
N	-37.949122	-53.363743	4.492348
Fe	-38.725258	-55.367603	4.59709
C	-39.860348	-55.434975	2.003467
C	-40.991668	-54.872394	2.897133
O	-40.082217	-55.651092	0.817927
O	-38.738335	-55.582008	2.637834
O	-40.689194	-54.429501	4.011319
H	-39.96313	-52.847395	6.446286
H	-38.988803	-56.776681	7.545428
H	-34.588088	-56.294865	6.169648
H	-34.712037	-57.339936	4.75308
H	-39.591391	-52.116551	3.898379
H	-35.963622	-53.832773	5.111623
H	-35.611359	-51.296021	5.002564
H	-39.941604	-55.605935	9.624941
H	-34.774888	-58.05728	6.393248
C	-42.37307	-54.795496	2.33497
H	-42.363712	-54.15903	1.442663
H	-42.68105	-55.792487	2.004241
H	-43.071524	-54.40376	3.076607
O	-40.801917	-57.268349	3.834005
O	-39.955288	-56.983112	4.756714
H	-40.577281	-53.178821	9.13741
H	-37.734838	-50.076267	4.284018

Quintet η^2 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.630554
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.030528
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.253182
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.352487
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.436099

C	-40.164971	-53.858002	8.407974
N	-41.200597	-54.689383	8.029812
C	-39.213582	-53.989473	7.433542
C	-40.868285	-55.289529	6.870236
N	-39.665377	-54.882097	6.476708

C	-35.065025	-57.201974	5.780007
C	-36.404117	-57.013008	5.053526
O	-36.781211	-57.844740	4.232161
O	-37.034235	-55.910652	5.384295
C	-37.658004	-51.146023	4.399018
N	-36.460272	-51.821774	4.267694
C	-38.602794	-52.115896	4.602083
C	-36.703354	-53.146532	4.390566
N	-37.994489	-53.358193	4.594373
Fe	-38.826089	-55.333378	4.701843
C	-39.124755	-55.069718	1.917508
C	-40.402101	-54.334583	2.386164
O	-38.869310	-55.128541	0.720046
O	-38.416038	-55.574993	2.877696
O	-40.585392	-54.200695	3.592227
H	-38.249628	-53.515396	7.347839
H	-41.499564	-55.986251	6.343803
H	-35.198582	-57.116865	6.864125
H	-34.361502	-56.416739	5.477829
H	-39.666533	-52.011131	4.746757
H	-35.951599	-53.917362	4.341693
H	-35.553951	-51.405289	4.102745
H	-42.065653	-54.834469	8.532948
H	-34.640373	-58.177561	5.535299
C	-41.343690	-53.817296	1.345302
H	-40.803798	-53.154881	0.659788
H	-41.706947	-54.655042	0.738101
H	-42.182655	-53.293494	1.806962
O	-39.466583	-57.461824	5.051205
O	-40.413436	-56.821220	4.451444
H	-40.193890	-53.275780	9.315029
H	-37.723235	-50.071166	4.337796

Septet η^1 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.666309
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.069335
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.252168
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.387283
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.478503

C	-40.164956	-53.858040	8.407984
N	-41.349184	-53.906828	7.699132
C	-39.222507	-54.414127	7.585431
C	-41.099405	-54.471845	6.496688
N	-39.817267	-54.790656	6.395940
C	-35.065049	-57.201970	5.780027
C	-36.574208	-57.081365	5.589336
O	-37.296945	-58.076336	5.493236
O	-36.990441	-55.847395	5.513883
C	-37.657995	-51.145990	4.398989
N	-36.456215	-51.789995	4.612248
C	-38.591211	-52.138334	4.272375
C	-36.688770	-53.121486	4.615990
N	-37.975700	-53.365770	4.412461
Fe	-38.727450	-55.409071	4.624492
C	-39.314377	-55.802895	1.811442
C	-40.579865	-55.038511	2.290433
O	-39.225641	-56.092719	0.620884
O	-38.472251	-56.044803	2.754953
O	-40.646133	-54.699239	3.470754
H	-38.167778	-54.573384	7.749901
H	-41.844792	-54.616269	5.730423
H	-34.728468	-56.554452	6.596577
H	-34.559538	-56.865856	4.866824
H	-39.653344	-52.058007	4.096292
H	-35.942963	-53.878967	4.796645
H	-35.556574	-51.347864	4.742383
H	-42.251038	-53.582999	8.021529
H	-34.776387	-58.235513	5.984668
C	-41.656576	-54.742147	1.294247
H	-41.232401	-54.184084	0.452327
H	-42.032042	-55.683405	0.876261
H	-42.471282	-54.181060	1.755974
O	-40.777040	-57.750605	3.902365
O	-40.226444	-57.228607	4.867797
H	-40.109263	-53.453835	9.406395
H	-37.731603	-50.070722	4.350117

Septet η^2 active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.665639
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.069431
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.263227
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.375162
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1296.466468

C	-40.165000	-53.858007	8.408004
N	-41.336752	-53.791688	7.680600
C	-39.241727	-54.417501	7.566615
C	-41.100832	-54.296340	6.449551
N	-39.837847	-54.684788	6.348464
C	-35.065001	-57.202005	5.780003
C	-36.486800	-57.137017	5.219558
O	-37.016547	-58.133803	4.730929
O	-37.030424	-55.947270	5.294823
C	-37.657999	-51.145989	4.398992
N	-36.445438	-51.784686	4.573541
C	-38.590972	-52.145323	4.338930
C	-36.672680	-53.117661	4.617095
N	-37.964995	-53.369610	4.476805
Fe	-38.818053	-55.414045	4.588204
C	-39.327872	-55.440244	1.723004
C	-40.537413	-54.611753	2.233040
O	-39.218454	-55.626585	0.513752
O	-38.544693	-55.842565	2.662653
O	-40.618487	-54.372429	3.435931
H	-38.201507	-54.647210	7.740462
H	-41.839867	-54.352650	5.665962
H	-35.023363	-56.756684	6.779927
H	-34.391520	-56.622623	5.136914
H	-39.659703	-52.073411	4.209487
H	-35.913544	-53.869449	4.763449
H	-35.541456	-51.338898	4.650829
H	-42.223720	-53.434831	8.009539
H	-34.716245	-58.236056	5.819997
C	-41.547049	-54.135630	1.236602
H	-41.046531	-53.533575	0.470135
H	-41.973850	-54.998916	0.713116
H	-42.335162	-53.558805	1.724239
O	-39.659471	-57.448866	5.381962
O	-40.472163	-56.933898	4.562500
H	-40.102414	-53.517695	9.429490
H	-37.737601	-50.072332	4.331240

Unbound quintet active site model (ASM):

E[UB3LYP-D3/6-31G(d),LANL2DZ] = -1146.346869
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.052363
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 0.247728
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1146.075088
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1146.696700

C	-40.164985	-53.858006	8.407989
N	-41.315071	-53.788790	7.645573
C	-39.237045	-54.480589	7.615508
C	-41.060409	-54.350831	6.441412
N	-39.807797	-54.780044	6.392365
C	-35.065016	-57.201983	5.780005
C	-36.581129	-57.034265	5.732380
O	-37.342333	-58.006676	5.705120
O	-36.959426	-55.788306	5.692130
C	-37.657999	-51.146011	4.399006
N	-36.448613	-51.748850	4.685207
C	-38.547729	-52.169049	4.214872
C	-36.633571	-53.087882	4.672409
N	-37.896245	-53.372706	4.388783
Fe	-38.640416	-55.360811	4.684913
C	-39.389956	-56.045656	1.960795
C	-40.600752	-55.163403	2.379801
O	-39.354653	-56.465887	0.806153
O	-38.537271	-56.238056	2.905199
O	-40.631689	-54.707527	3.520131
H	-38.206256	-54.735046	7.809321
H	-41.772803	-54.409985	5.634348
H	-34.637891	-56.578071	6.572266

H	-34.641927	-56.857247	4.829044
H	-39.600895	-52.136016	3.980034
H	-35.881848	-53.823990	4.908226
H	-35.576342	-51.274719	4.874632
H	-42.199394	-53.394818	7.935987
H	-34.787299	-58.246701	5.938024
C	-41.673639	-54.910763	1.367509
H	-41.229563	-54.458546	0.473860
H	-42.095548	-55.869163	1.043954
H	-42.454048	-54.268032	1.779377
H	-40.120979	-53.477251	9.416163
H	-37.764371	-50.073565	4.348191

5.6 References for Chapter 5

1. (a) Cortopassi W.A.; Simion R.; Hornsby C.E.; França T.C.C.; Paton R.S., Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem Eur J* **2015**, *21*, 18983–18992. (b) Fang, D.; Lord, R. L.; Cisneros, G. A., Ab initio QM/MM calculations show an intersystem crossing in the hydrogen abstraction step in dealkylation catalyzed by AlkB. *J Phys Chem B* **2013,** *117* (21), 6410-20.

2. (a) 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* (39), 9375-9386; (b) Siegbahn, P. E. M., Quantum chemical studies of redox-active enzymes. *Faraday Discuss* **2003,** *124*, 289-296; (c) Siegbahn, P. E. M.; Borowski, T., Modeling enzymatic reactions involving transition metals. *Accounts Chem Res* **2006,** *39* (10), 729-738.

3. Frisch, M. J. T., 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.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox. D. J. , Gaussian 09. *Gaussian, Inc.* **2009**.

4. Couty, M.; Hall, M. B., Basis sets for transition metals: Optimized outer p functions. *J Comput Chem* **1996,** *17* (11), 1359-1370.

5. (a) Roy, L. E.; Hay, P. J.; Martin, R. L., Revised basis sets for the LANL effective core potentials. *J Chem Theor Comp* **2008,** *4* (7), 1029-1031; (b) Ehlers, A. W.; Bohme, M.; Dapprich, S.; Gobbi, A.; Hollwarth, A.; Jonas, V.; Kohler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G., A Set of F-Polarization Functions for Pseudo-Potential Basis-Sets of the Transition-Metals Sc-Cu, Y-Ag and La-Au. *Chem Phys Lett* **1993,** *208* (1-2), 111-114.

6. 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* (11), 1995-2001.

7. Walter, P.; Metzger, J.; Thiel, C.; Helms, V., Predicting where small molecules bind at protein-protein interfaces. *PloS one* **2013,** *8* (3), e58583.

8. 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* (7), 3601-3658.

9. Ng, S. S.; Kavanagh, K. L.; McDonough, M. A.; Butler, D.; Pilka, E. S.; Lienard, B. M.; Bray, J. E.; Savitsky, P.; Gileadi, O.; von Delft, F.; Rose, N. R.; Offer, J.; Scheinost, J. C.; Borowski, T.; Sundstrom, M.; Schofield, C. J.; Oppermann, U., Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **2007,** *448* (7149), 87-91.

10. Chen, Z.; Zang, J.; Whetstine, J.; Hong, X.; Davrazou, F.; Kutateladze, T. G.; Simpson, M.; Mao, Q.; Pan, C. H.; Dai, S.; Hagman, J.; Hansen, K.; Shi, Y.; Zhang, G., Structural insights into histone demethylation by JMJD2 family members. *Cell* **2006,** *125* (4), 691-702.

11. 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* (15), 154104.

12. Wang, J.; 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* (9), 1157-74.

13. Case, D. A.; Darden, T. A.; Cheatham, T. E.; 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.; Goetz, 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**.

14. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 1), 12-21.

15. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y., RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **2011,** *39*, W511-W517.

16. (a) 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 Theory Comput* **2006,** *2* (3), 815-826; (b) 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* (9), 1170-1179; (c) 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.

17. (a) Liu, H.; Llano, J.; Gauld, J. W., A DFT study of nucleobase dealkylation by the DNA repair enzyme AlkB. *J Phys Chem B* **2009,** *113* (14), 4887-98; (b) Lundberg, M.; Morokuma, K., Protein environment facilitates O2 binding in non-heme iron enzyme. An insight from ONIOM calculations on isopenicillin N synthase (IPNS). *J Phys Chem B* **2007,** *111* (31), 9380-9.

18. Geng, C.; Ye, S.; Neese, F., Analysis of reaction channels for alkane hydroxylation by nonheme iron(IV)-oxo complexes. *Angew Chem Int Ed* **2010,** *49* (33), 5717-20.

19. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P., Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. *Chem Eur J* **2014,** *20* (2), 435-46.

20. Fang, R.; Chen, F.; Big-Dong, Z.; Hu, D.; Barbera, A. J.; Clark, E. A.; Fang, J.; Yang, Y.; Mei, P.; Rutenberg, M.; Li, Z.; Zhang, Y.; Xu, Y.; Yang, H.; Wang, P.; Simon, M. D.; Zhou, Q.; Li, J.; Marynick, M. P.; Li, X.; Lu, H.; Kaiser, U. B.; Kingston, R. E.; Xu, Y.; Shi, Y. G., LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol Cell* **2013,** *49* (3), 558-70.

21. Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd, Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J Comput Chem* **2012,** *33* (2), 119-33.

22. Young, D., *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. New York, **2001**.

23. Ulucan, O.; Keskin, O.; Erman, B.; Gursoy, A., A comparative molecular dynamics study of methylation state specificity of JMJD2A. *PLoS One* **2011,** *6* (9), e24664.

24. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

25. Ye, S. F.; Price, J. C.; Barr, E. W.; Green, M. T.; Bollinger, J. M.; Krebs, C.; Neese, F., Cryoreduction of the NO-Adduct of Taurine:alpha-Ketoglutarate Dioxygenase (TauD) Yields an Elusive {FeNO}(8) Species. *J Am Chem Soc* **2010,** *132* (13), 4739-4751.

26. (a) Hirao, H., The effects of protein environment and dispersion on the formation of ferric-superoxide species in myo-inositol oxygenase (MIOX): a combined ONIOM(DFT:MM) and energy decomposition analysis. *J Phys Chem B* **2011,** *115* (38), 11278-85; (b) Blomberg, L. M.; Blomberg, M. R.; Siegbahn, P. E., A theoretical study on the binding of O(2), NO and CO to heme proteins. *J Inorg Biochem* **2005,** *99* (4), 949-58.

27. Wirstam, M.; Lippard, S. J.; Friesner, R. A., Reversible dioxygen binding to hemerythrin. *J Am Chem Soc* **2003,** *125* (13), 3980-3987.

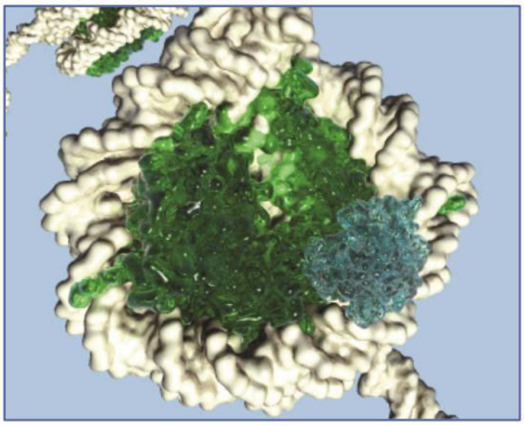
28. Lonsdale, R.; Harvey, J. N.; Mulholland, A. J., Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J Chem Theo Comp* **2012,** *8* (11), 4637-4645.

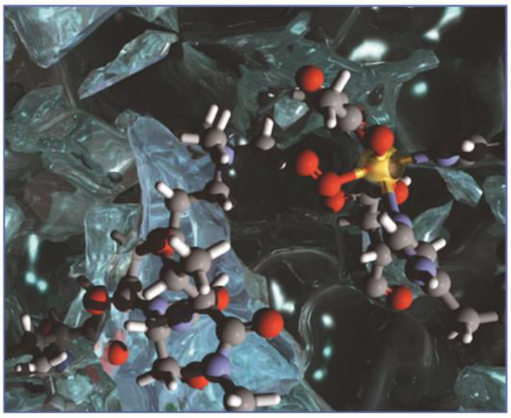
29. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J., Combining quantum mechanical methods with molecular mechanics methods in ONIOM. *Abstr Pap Am Chem S* **2006,** *232*, 408-408.

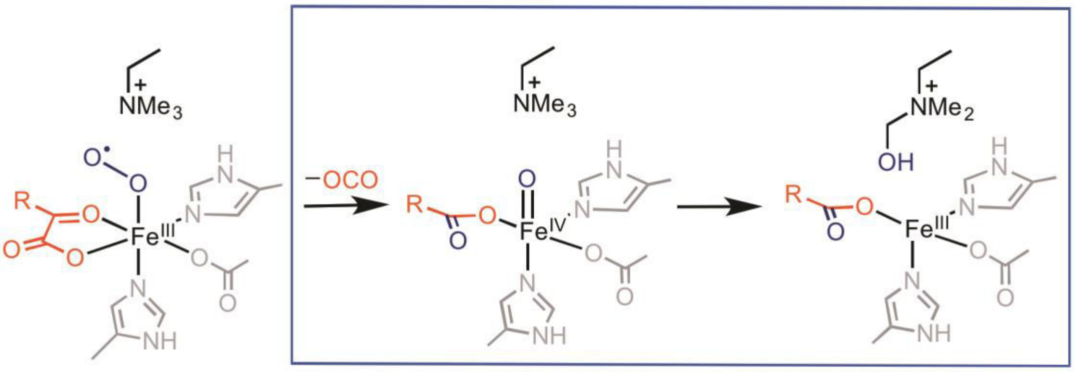
Chapter 6

Evaluating the reaction mechanism of trimethyllysine demethylation by JMJD2A

> *"Almost every wise saying has an opposite one, no less wise, to balance it."*
>
> *George Santayana*







6.1 Overview

Following the understanding of the dioxygen binding in the active site of the demethylase JMJD2A, we turned our attention to perform the first quantum mechanics/molecular mechanics (QM/MM) study of the subsequent steps of the demethylation of the trimethylated lysine: oxygen activation, H-abstraction and –OH rebound (**Figure 6.1**).

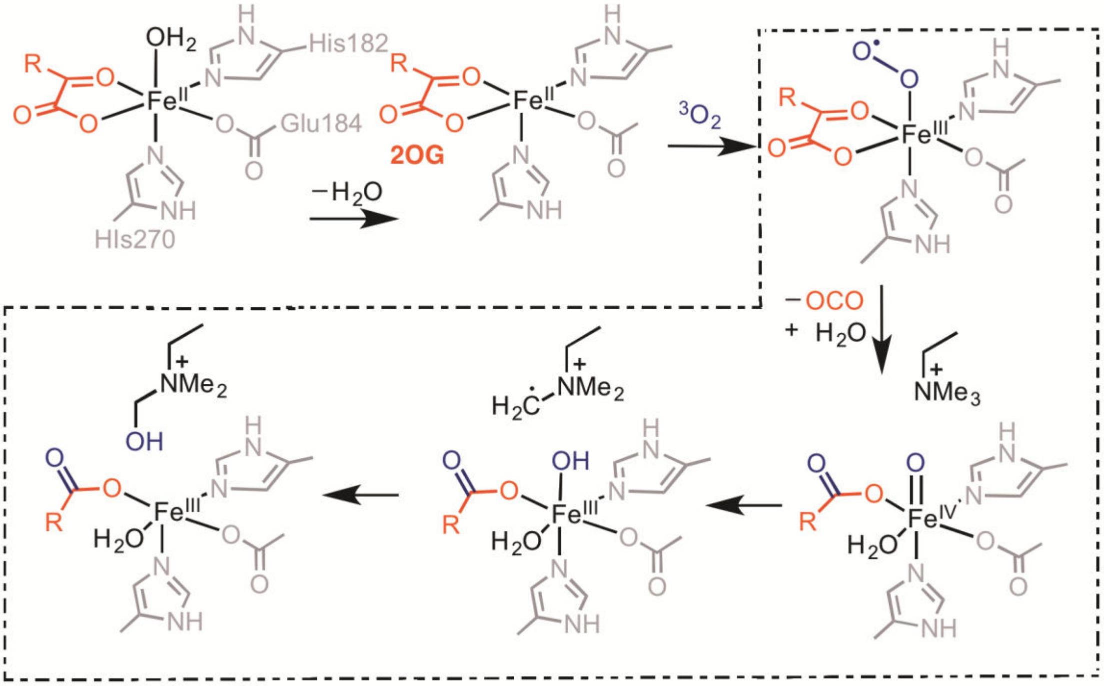


Figure 6.1. Postulated mechanism of O_2 activation and lysine demethylation *via* a Fe^{IV}–oxo intermediate in JMJD2A. The dimethylated lysine and formaldehyde are released following fragmentation of the unstable hemiaminal (step not shown).

The postulated mechanism for non-haem iron dioxygenases begins with the dioxygen binding to the active site in the quintet energy potential surface, resulting in the oxidation of Fe^{II} to Fe^{III}.[1] The favourability of this step in JMJD2A is due to the electrostatic role of charged residues surrounding the active site, making this step highly dependent on the protein environment.[2] This mechanism of binding was proposed by us[2] and corroborates previous studies with the DNA demethylase Alkb, which promotes the demethylation of nucleobases.[1] Subsequent steps have been proposed to occur in JMJD2A as follows, although at the time of initiating this work, had yet to be computed: the next step involves the formation of a high-valent Fe^{IV}-oxo reactive intermediate that is responsible for the H-abstraction of the trimethylated substrate, forming a radical $CH_2^{\cdot}$. Finally, the hydroxyl –OH group is transferred to this radical, and due to the instability of the hemiaminal intermediate, the demethylation proceeds via the release of formaldehyde. Previous calculations have shown that the demethylation reaction in non-haem iron dependent oxygenases happens in both triplet and quintet spin states,[3] although the later corroborates Electron Paramagnetic Resonance (EPR) characterisation of non-haem Fe^{IV}-oxo complexes in high spin states. Also, computational data for AlkB, responsible for demethylation of DNA bases, support the quintet overall spin state for the protein reactivity.[4]

This chapter will consist of the energy potential surface evaluation of the demethylation of trimethylated lysine in JMJD2A. Our aim is to characterize the intermediates and transition states, as well as to investigate the rate-determining step in this enzyme. Recent theoretical work at the DFT level of theory has been performed by Cossío and co-workers[5] for the investigation of the H-abstraction and –OH rebound steps of this reaction. This work provides a comparison of the behaviour of different lysine methylation states undergoing CH-oxidation, and develops a novel hypothesis to account for the reactivity of mono- and di-methylated lysines. However, previous studies have used QM-only models of the active site, in which the substrate is freely able to move. Therefore, questions still remain regarding the structural and energetic contributions of the protein environment along the reaction coordinate for demethylation in JMJD2A. Apart from gaining structural insights into the reactivity profile of this reaction, a better understanding of the full chemical profile will give us an idea which steps should be the focus for the design of more selective JMJD2A inhibitors. In addition, knowing the contribution of residues, particularly those remote from the immediate coordination of the metal centre, opens possibilities of mutations that are not obvious from crystal structures, with potential of altering the rate of the reaction.

6.2 Using QM/MM for modelling non-haem iron dependent oxygenases

Reactions in non-haem iron dependent oxygenases usually involve multiple steps with structural modifications in the protein environment that are not always seen in small computational models containing a few active site residues. For the AlkB demethylase, for instance, consideration of an arginine close to the active site is necessary for describing the rotation of the Fe^{IV}-oxo group preceding the oxygen activation. In 2013, Cisneros and co-workers[6] studied the H-abstraction in AlkB following the binding of the dioxygen in the vacant site trans to His187. They used the BP86/6-311G(d) level of theory for Fe and O, and BP86/6-31G(d) for C, H and N. The assumption of binding in this vacant site would not be realistic since it would result in steric clashes between the dioxygen and other residues surrounding the active site. De Visser and co-workers[4] then evaluated the potential energy surface for the H-abstraction by considering first the binding of dioxygen trans to another histidine, His131, and, following the oxygen activation, the rotation of the Fe^{IV}-oxo group trans to His187. They performed QM/MM calculations using UB3LYP/6-31+G(d) for C, N, H and O, and double-ζ LACVP basis set on Fe. The isomerisation energy difference is -6.0 kcal mol^{-1} and is, therefore, essential for the correct evaluation of the subsequent H-abstraction in AlkB. This work emphasises the structural role of the enzyme in controlling the available geometries in the active site, which may not be captured using computational models limited to the primary residues only.

Paneth and Geronimo[7] evaluated computationally the C–H hydroxylation by nitrobenzene 1,2-dioxygenase (NBDO), an enzyme responsible for the oxidative degradation of nitrobenzene. They performed QM/MM calculations using the ONIOM method with the B3LYP/LACVP*:Amber level of theory to compare two possible mechanisms for the C–H hydroxylation in NBDO: 1) a concerted mechanism with Fe^{III}–OOH in which the C–H bond should be aligned with the O–O bond for good orbital overlap and 2) H-abstraction by HO–Fe^{V}=O involving a sideway attack of the substrate. The transition state (TS) for the first mechanism is sterically affected by the inclusion of non-active site residues, which do not allow the perfect orbital overlap. On the other hand, the protein environment slightly reduces the energy barrier for the second proposed mechanism.

The inclusion of the protein environment has been shown to have minimal effects in some cases. Paneth and Geronimo[7] showed that QM/MM calculations reduce slightly the energy barriers for the transition states in NBDO catalysed reaction. De Visser and co-workers[8] performed calculations to understand the chemical properties of the Fe^{IV}-oxo species in the TauD protein using ONIOM with the UB3LYP/LANL2DZ:UFF level of theory. They showed that spin densities and bond lengths obtained with QM/MM calculations on Fe^{IV}-oxo complexes in TauD are comparable to QM-only calculations, suggesting that the inclusion of the protein environment would not have a large effect on determining these electronic properties.

Finally, our study[2] on the potential energy surface for the dioxygen binding in JMJD2A at the UB3LYP/6-311+G(d,p):Amber level of theory for C, N, O, H and

UB3LYP/LANLTZ(f) for Fe showed that appreciation of the protein environment was important to make this process exergonic. A similar approach is then used in this work to investigate if protein residues not directly included in the active site would have an important effect on determining the rate-determining step of the demethylation reaction in JMJD2A.

6.3 Investigating the rate-determining step of the demethylation reaction in JMJD2A

6.3.1 Oxygen activation

Experimental and computational efforts have focussed previously on understanding of the energy potential surface of the oxygen activation following the dioxygen binding in non-haem iron oxygenases – such as TauD[9] and AlkB[4], which catalyse the hydroxylation of taurine and the demethylation of nucleobases as substrates, respectively.

Mossbauer and EPR spectroscopy studies of TauD suggest the presence of high-valent Fe^{IV} species in the quintet spin state.[10] Further kinetic isotope effect (KIE) studies with deuterated substrates provide evidence that this intermediate abstracts hydrogen from the substrate taurine.[11] Que and co-workers built a chemical model to get insight on the oxygen activation of non-haem iron oxygenases, consisting of $[Fe(TpPh_2)(BF)]$ ($TpPh_2$=hydrotris(3,5-diphenylpyrazol-1-yl)borate, BF=benzoylformate) (**Figure 6.2**). Through isotopic labelling experiments, they showed the decarboxylation of the BF ligand, suggesting the nucleophilic attack of the distal O atom derived from the oxygen on C2 of the α-ketoacid in non-haem iron oxygenases.[12]

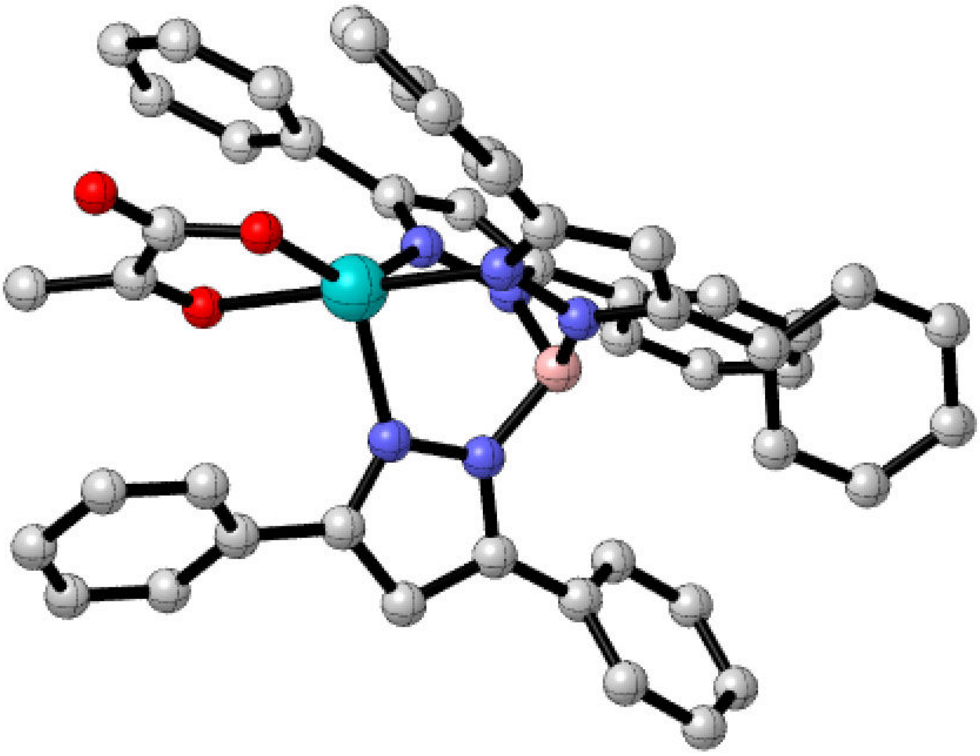


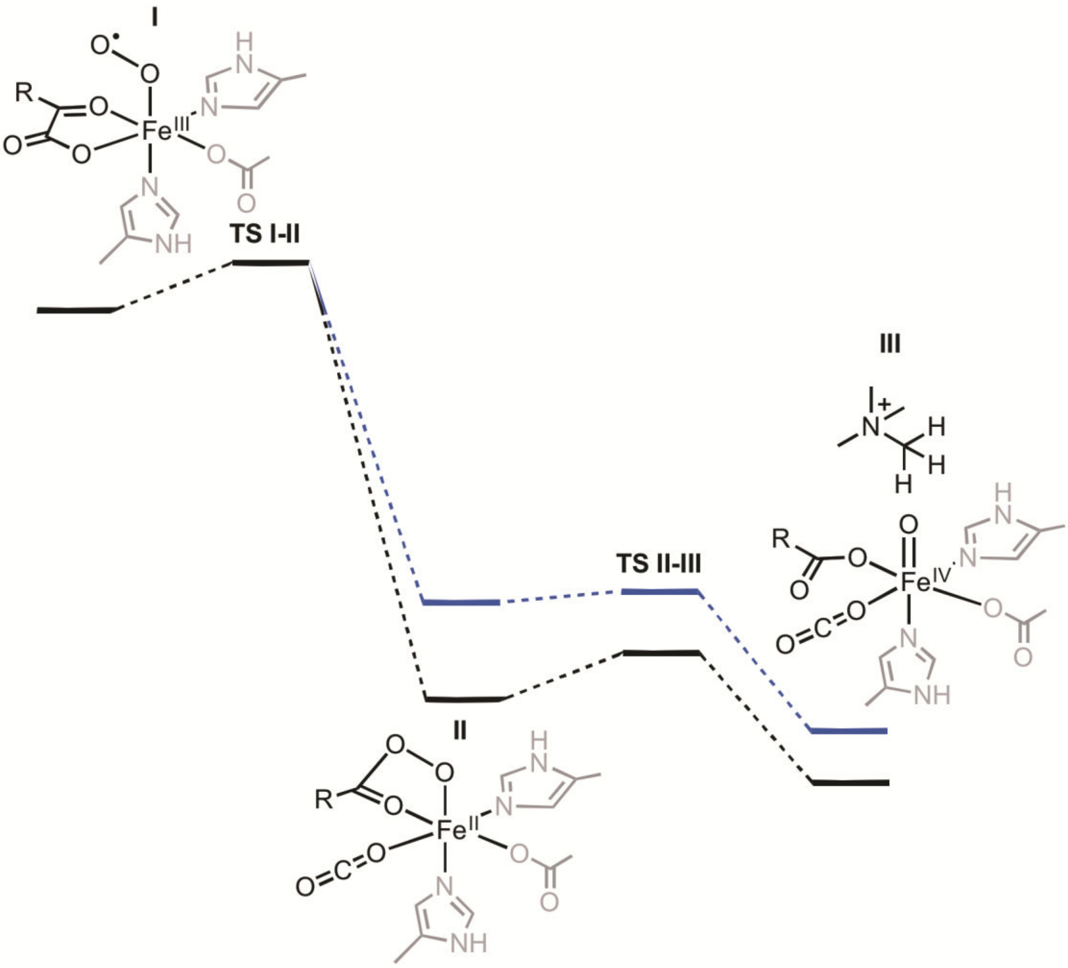
Figure 6.2. Representation of [Fe(TpPh$_2$)(BF)] (TpPh$_2$=hydrotris(3,5-diphenylpyrazol-1-yl)borate, BF=benzoylformate). Hydrogens are omitted for better visualization.

DFT calculations for non-haem active sites similar to that found in JMJD2A have been performed to derive insight into the rate-determining step of the oxygen activation. Most of these results suggest that the reaction happens in the overall quintet spin state. Siegbahn and Borowski[13] used B3LYP/LACV3P** to propose a mechanism that starts with the nucleophilic attack of the distal O to the ketoglutarate cofactor, forming a bicyclic compound – this being the rate-determining step. Following this, there is the cleavage of the O-O bond to derive the Fe^{IV}-oxo complex. In related work by de Visser[14] using the UB3LYP/6-311+G(d) for C, O, N and LACV3P+(d) for Fe, the rate-determining step was suggested to be the cleavage of the peroxo-bond. Solomon[15] used UB3LYP/6-311+G(2d,p) with the LANL2DZ effective core potential basis set for Fe and suggested that this reaction could also involve triplet-quintet crossover. More recently, Neese,[9] using triple-ζ TZVP basis sets on Fe, O, and N, and the SV(P) basis sets on the remaining atoms, and Borowski,[16] using UB3LYP/LACV3P+(d) for Fe and 6-311G(d) for the other atoms, showed that this reaction is more likely to happen in high spin states (quintet or septet) and the nucleophilic attack of the distal oxygen to the oxoglutarate cofactor is the rate-determining step.

Given computational and experimental studies suggest the favorability of the quintet energy potential surface for the oxygen activation in non-haem iron oxygenases,[15-16] we investigated the role of the protein environment for these steps in JMJD2A in the quintet spin state. We performed QM/MM calculations to evaluate the role of the protein environment at the low level of theory (Amber), using as a starting point the complex with the dioxygen bound into the active site, following our previous calculations.[17]

Calculations were performed using the two-layered hybrid QM/MM optimizations implemented in the Our own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM) scheme with electronic embedding, similar to the methodology described in **Section 5.2**, **Chapter 5**. The output file from the dioxygen binding was used as a starting point for the oxygen activation studies in this chapter. For the QM region, a double-ζ valence 6-31G(d) basis set was employed for all atoms except Fe, for which the modified LANL2DZ[18] effective core potential (ECP) and associated double-ζ valence basis set was used with an additional f-type polarization function with exponent of 2.462.[19] Single point energies were evaluated for stationary points with a larger triple-ζ valence polarized 6-311+G(d,p) basis set used for all non metal atoms and with the LANL2TZ(f) ECP and decontracted triple-ζ valence basis set for Fe.[19] A conductor-like polarizable continuum model (CPCM) with $\varepsilon=4.3$

(diethyl ether as solvent)[20] was applied to mimic the dielectric effects of the relatively hydrophobic protein interior.[21] For the MM region, the Amber force field was used, and our QM/MM results for this process are shown in **Figure 6.3**.



	I	TS I-II	II	TS II-III	III
$E_{rel\text{-}QM}$	**0**	**5.6**	**-47.6**	**-41.4**	**-61.5**
$E_{rel\text{-}MM}$	**0**	**0.8**	**-14.3**	**-13.7**	**-14.2**
$E_{rel\text{-}QM/MM}$	**0**	**6.4**	**-61.9**	**-55.1**	**-75.7**
$G_{rel\text{-}QM/MM}$ (298K, 1 mol L^{-1})	**0**	**6.9**	**-63.0**	**-57.0**	**-76.6**
$G_{rel\text{-}QM}$* (298K, 1 mol L^{-1})	**0**	**7.1**	**-47.5**	**-45.1**	**-70.9**

Figure 6.3. Free energy profile for the oxygen activation in JMJD2A in the quintet spin state at the UB3LYP-D3/6-311G+(d,p):Amber; Fe:LANL2TZ(f) ECP level of theory. *In blue, free energy profile for the oxygen activation in TauD in the quintet spin state at the UB3LYP-D3/cc-pVTZ; Fe:LANL2DZ(f) level of theory in diethylether.[16] All values in kcal mol^{-1}.

Our QM/MM energy profile indicates the nucleophilic attack of the distal oxygen, with the formation of the bicyclic complex, is the harder of the two elementary steps involved in oxygen activation, corroborating previous studies.[4] The free energy barrier for the TS of this nucleophilic attack is 6.9 kcal mol^{-1} (**Figure 6.3**), and the other TS barrier for peroxo-cleavage is slightly lower (6.0 kcal mol^{-1}). The protein environment was shown to have a relatively small effect, which is smaller than we observed previously for the dioxygen binding. The only significant energetic role was observed for formation of **Intermediate II** (-14.3 kcal mol^{-1}), coming mainly from electrostatic interactions (**Figure 6.3** – the electrostatic effect will be discussed later in this chapter). However, even without this correction, this step was shown to be highly exergonic (-47.6 kcal mol^{-1}), therefore predicting irreversibility for both steps. Given the differences in the position of the atoms for the active site of JMJD2A compared to TauD, the precise energy values vary for each of these steps in these proteins. However, the full reaction path remains qualitatively similar, and the resulting ferric complex is -76.6 kcal mol^{-1} lower in energy than the starting structure with dioxygen bound. Based on such favorable computed steps, and barriers substantially lower than required for room-temperature reactivity, we can predict with a high degree of confidence that oxygen activation, once bound at the active site, should not limit the overall turnover of this enzyme.

Analysis of the structural arrangements of the complex along the reaction coordinate has been performed in **Figure 6.4**. This shows that the dioxygen needs to move closer to the cofactor to allow the nucleophilic attack (1.32 Å compared to 2.24 Å of the dioxygen bound complex). After the O-O cleavage, the Fe-O distance is shortened to 1.62 Å in the ferryl complex (**Figure 6.4**) – this distance was shown to be 2.13 Å in the dioxygen bound complex. In the final intermediate, the carboxylate group of the oxoglutarate cofactor also moves away from the Fe^{IV}-oxo group (O-O distance changes from 1.29 Å to 3.02 Å). Interestingly, our calculations considering non-active site protein residues did not present significant structural differences when compared to QM calculations performed by Borowski and co-workers with a truncated model of TauD at the UB3LYP-D3/cc-pVTZ; Fe:LANL2DZ(f) level of theory, with differences in distance values not exceeding 0.2 Å (**Figure 6.4**).

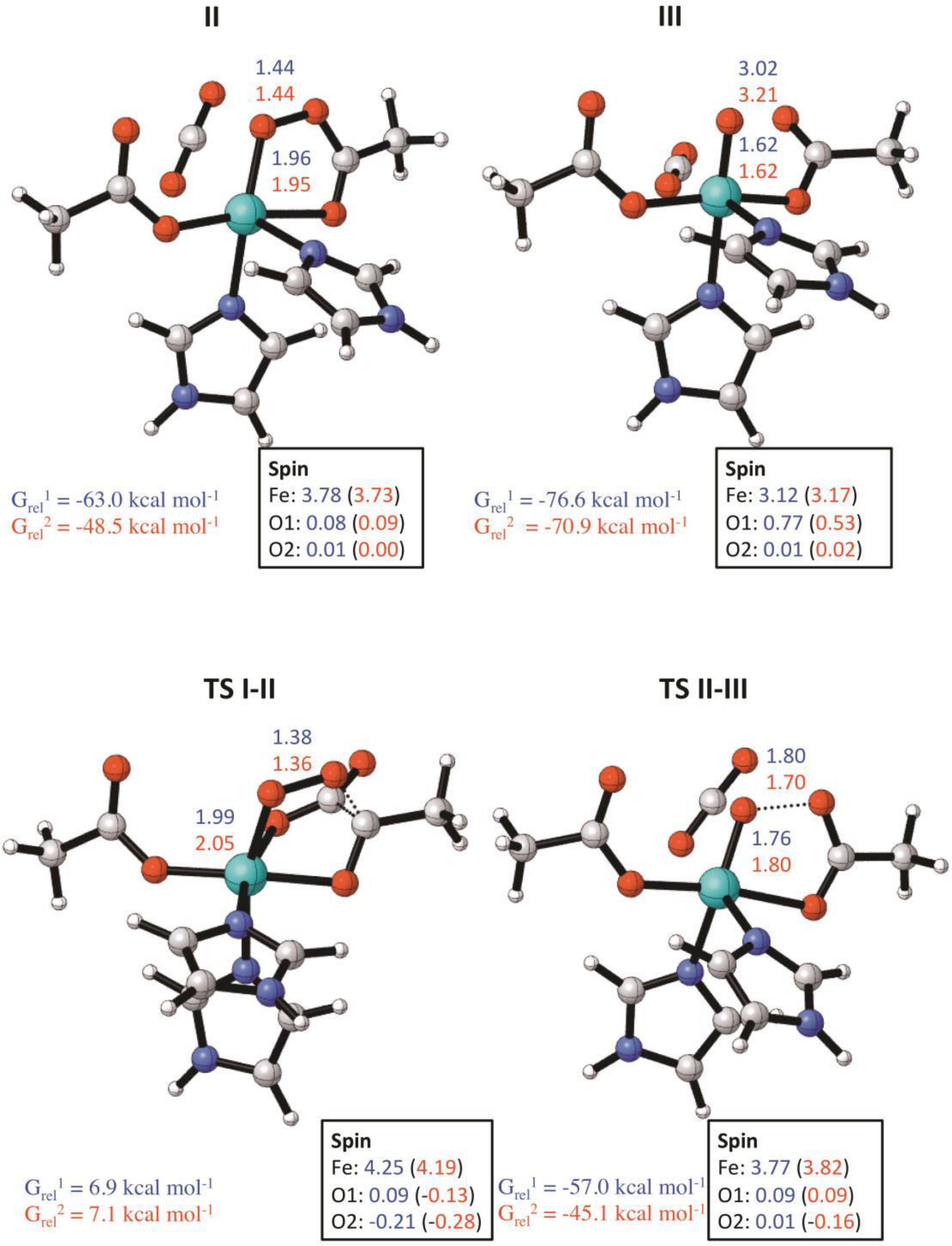


Figure 6.4. Small model representation of QM/MM calculations (the protein environment is ommited for a better visualization of the active site). UB3LYP-D3/6-311G+(d,p):Amber; Fe:LANL2TZ(f) intermediates and transition structures for the oxygen activation in JMJD2A. Blue: UB3LYP-D3/6-311+G(d,p):Amber ; Fe:LANL2TZ(f) ECP. Red: UB3LYP-D3/6-31G+(d):Amber ; Fe: LACVP ECP.[4] Red: ΔG:UB3LYP-D3/cc-pVTZ Fe:LANL2DZ(f) ECP (TauD).[22] Distances in Å are shown near the bonds.

6.3.2 H-Abstraction

On the basis of computational data, C-H abstraction has been proposed as the rate-determining step of the demethylation reaction in AlkB.[4] Recent computational studies by Cossío and co-workers have focussed on the investigation of this step (and the subsequent radical-rebound) in proteins of the Jumonji C family by using a truncated model active site of JMJD2A.[5] They found that the demethylation of the trimethylated lysine is likely to happen in the quintet energy potential surface, although the triplet and quintet states are very close to each other along the entire reaction coordinate, suggesting the possibility of spin crossover. Determining the potential energy surface of the C-H abstraction step is important not only to investigate if this is the rate-determining step for the lysine demethylation, but also to understand the chemical basis for substrate selectivity displayed by these proteins for the different methylated histone substrates. For AlkB, it has been proposed that this reaction happens in the quintet spin state.[4] In JMJD2A, this mechanism consists of the H-abstraction of the methylated lysine by the Fe^{IV}-oxo complex. This results in the formation of the methylated alkyl radical Lys-$CH_2^{\cdot}$ and Fe^{III}–OH. Studies performed by Neese and co-workers[3] using UB3LYP/def2-TZVPP showed that the TS barrier for the H-abstraction is lower in the septet energy potential surface (+13.9 kcal mol^{-1}) than in the quintet (+24.6 kcal mol^{-1}) or triplet (33.0 kcal mol^{-1}) surfaces. However, a septet spin state would not allow the formation of the singlet ethanol product in TauD. Additionally, the septet reactant $Fe^{IV}=O$ is +15.1 kcal mol^{-1} higher in energy than its correspondent quintet, and thus it seems unlikely the reaction would start in the septet surface.

Interestingly, de Visser[4] showed that the H-abstraction step in AlkB differs from the proposed mechanism for most of the non-haem iron oxygenases, by using

UB3LYP/6-311+G(d):Amber for C, O, N and LACV3P+(d) for Fe. Usually, this abstraction happens in a linear way, with the substrate attacking from the top, the so-called *σ-pathway*. However, due to a rotation of the $Fe^{IV}=O$ mediated by an arginine residue, and a consequent change in the orbital shapes (**Figure 6.5**), the σ-pathway is disfavoured in comparison with the *π-pathway*.

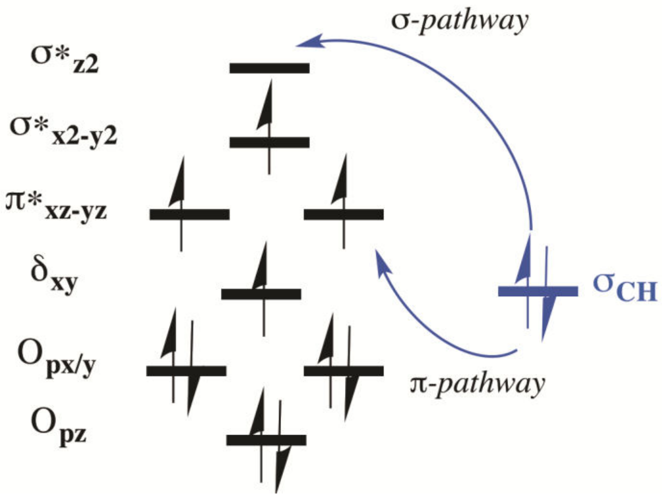


Figure 6.5. Electronic diagram of the *σ* and *π-pathways* for the H-abstraction of methylated nucleobases by Alkb. In black, orbitals for the $Fe^{IV}=O$ complex. In blue, orbital representation of the C-H bond from the substrate.

Our calculations were performed considering the quintet overall spin state and the output from our previously studied oxygen activation steps. As observed for AlkB, we demonstrated that the H-abstraction happens in a non-linear way, with the Fe-O-H angle close to 120 degrees, corresponding to the *π-pathway*. We could not locate TS for the *σ-pathway*. This is due to the restraints of the protein environment, which makes the Fe-O-C to be at around 125 degrees (131.4 degrees in AlkB), a way further than the ideal 180 degrees required for the *σ-pathway*.

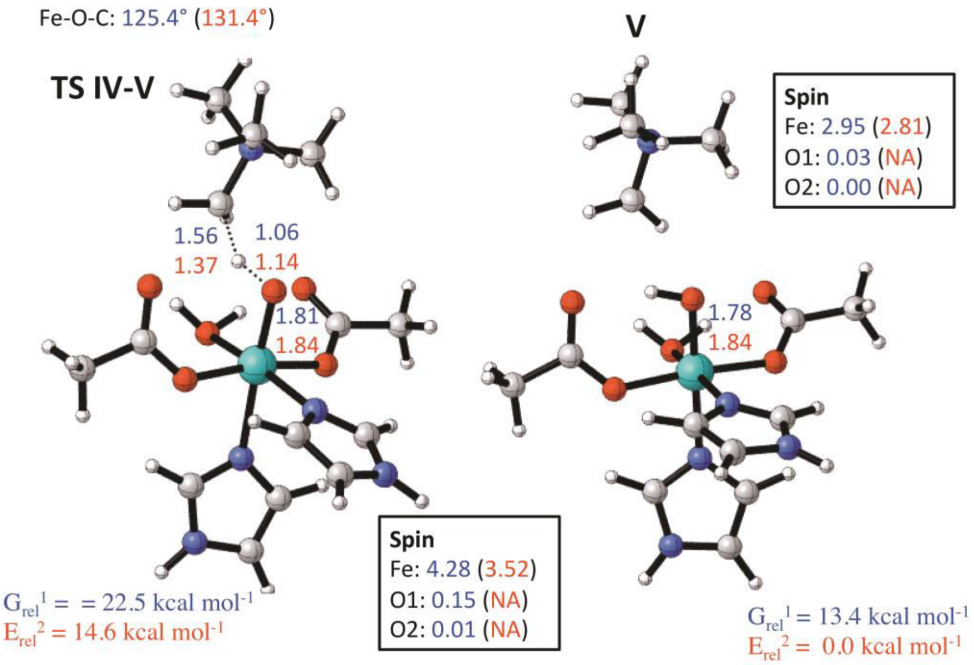
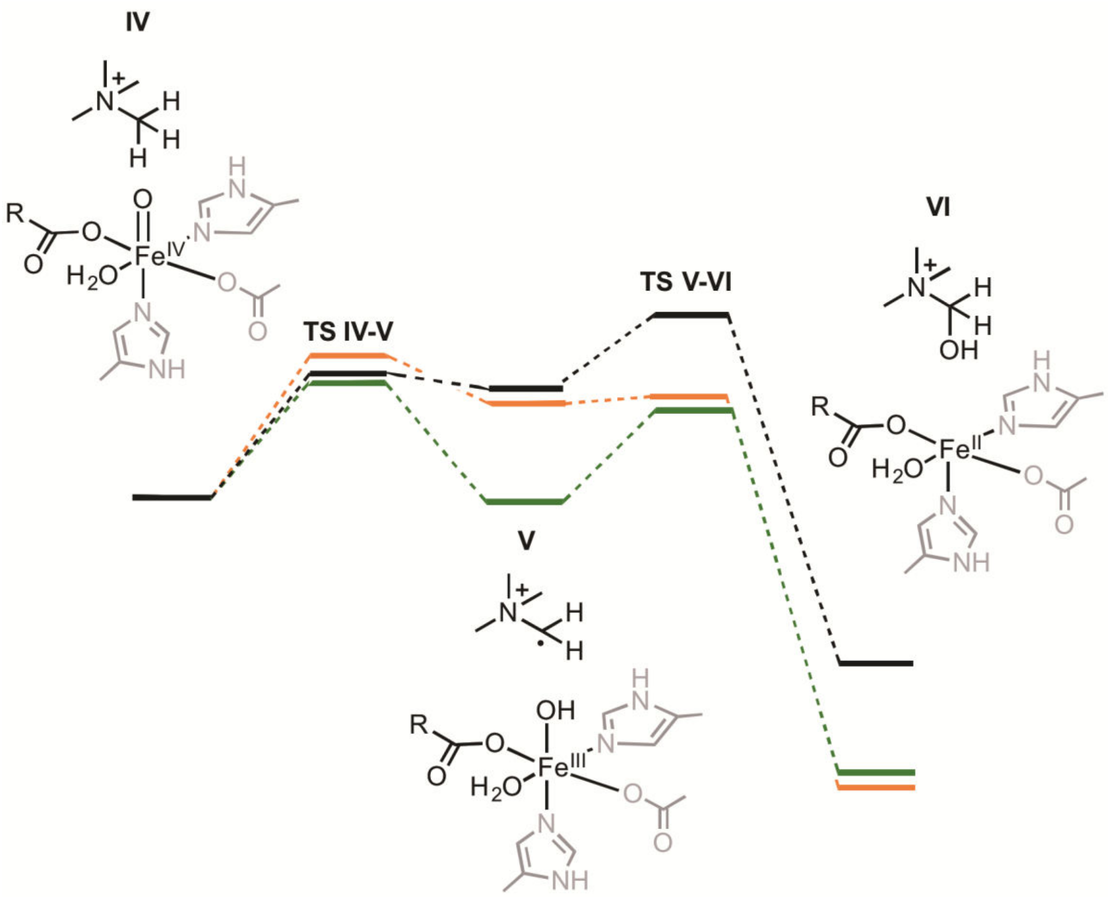


Figure 6.5. Small model representation of the QM/MM calculations (the protein environment is ommited for a better visualization of the active site). UB3LYP-D3/6-311G+(d,p):Amber; Fe:LANL2TZ(f) intermediates and transition structures for the H-abstraction in JMJD2A. [1]Blue: UB3LYP-D3/6-311+G(d,p):Amber ; Fe:LANL2TZ(f) ECP. [2]Red: UB3LYP-D3/6-31G+(d):Amber ; Fe: LACVP ECP.[4] Distances in Å are shown near the bonds.

The geometry of the C-H abstraction TS is somewhat uncommon, in the sense that there is a deviation from the ideal angle of 180 degrees for the Fe-O-C angle (as judged from model QM-calculations). However, consultation of the Intrinsic Reaction Coordinate (IRC) confirmed that the hydrogen does indeed transfer from the methylated lysine to the iron centre, forming the Fe–OH complex, expected for this transition state. The activation barrier is consistent with values previously obtained in related reactions, and in model studies, and so the geometric distortion is likely a consequence of the enzyme restricting the orientation of the substrate. Evidently, this does not lead to a substantial disfavouring of this step relative to an undistorted model.

The impact of the amount of HF exchange was evaluated in our calculations. The decrease in this amount from the standard B3LYP 20% to 15% (B3LYP*) is shown to increase the energy by 5.1 kcal mol^{-1}. Free energy corrections have a considerable effect in lowering the energy of the TS (-5.3 kcal mol^{-1}).

Comparison of the H-abstraction in Jumonji C and AlkB proteins (**Figure 6.6**) shows similarities in terms of the geometric features of the reaction in both active sites, although higher energy differences are observed for JMJD2A. We then examined if the biological environment of JMD2A has an important role to make this reaction possible. Our calculations showed that the consideration of non-active site residues treated at the low level of theory with the Amber force field slightly increases the energy (+2.4 kcal mol^{-1} - **Figure 6.6**). A notable effect upon the linear C-H abstraction TS was found, in which the restriction of the substrate by the enzyme environment leads to a distortion from the idealized linearity – although this does not lead to a substantial increase in the barrier for this step. However, it is not discarded in this present work the importance of these residues for the favourability of this reaction, and we hypothesise that dynamical behaviour, that are still a challenge to take into account in QM/MM ONIOM calculations, may also help to reduce the high observed free energy related to this step (+18.0 kcal mol^{-1}). Cossío and co-workers[5] showed the H-abstraction TS barrier to be +22.5 kcal mol^{-1} by using a truncated model active site of JMJD2A.



	IV	TS IV-V	V	TS V-VI	VI
$E_{rel\text{-}QM}$	0	20.9	19.4	18.7	-42.7
$E_{rel\text{-}MM}$	0	2.4	0.5	14.2	13.5
$E_{rel\text{-}QM/MM}$	0	23.3	19.9	32.9	-29.2
$G_{rel\text{-}QM/MM}$ (298K, 1 mol L^{-1})	0	18.0	16.4	29.6	-28.9
$G_{rel\text{-}QM}$* (298K, 1 mol L^{-1})	0	22.5	13.4	15.9	-50.8
$E_{rel\text{-}QM/MM}$**	0	14.6	0.0	12.8	-53.4

Figure 6.6. Free energy profile for the H-Abstraction and –OH rebound in JMJD2A in the quintet spin state at the UB3LYP-D3/6-311G+(d,p):Amber; Fe:LANL2TZ(f) ECP level of theory. *In green, free energy profile for JMJD2A in the quintet spin state at the UB3LYP-D3/6-311+G(2d,2p) Fe:LANL2DZ ECP level of theory in diethylether.[5] **In red, free energy profile for AlkB in the quintet spin state at the UB3LYP-D3/6-31G+(d):Amber; Fe:LACVP ECP level of theory in diethylether. All values in kcal mol^{-1}.

6.3.3 –OH rebound

The subsequent step of the demethylation reaction following the H-abstraction is the hydroxyl rebound from the iron centre to the lysine alkyl radical. As we can observe from **Figure 6.7**, the –OH group hydrogen bonds the carboxylate of the glutamate. These similar results were observed by de Visser and co-workers for the demethylation reaction in AlkB.[4]

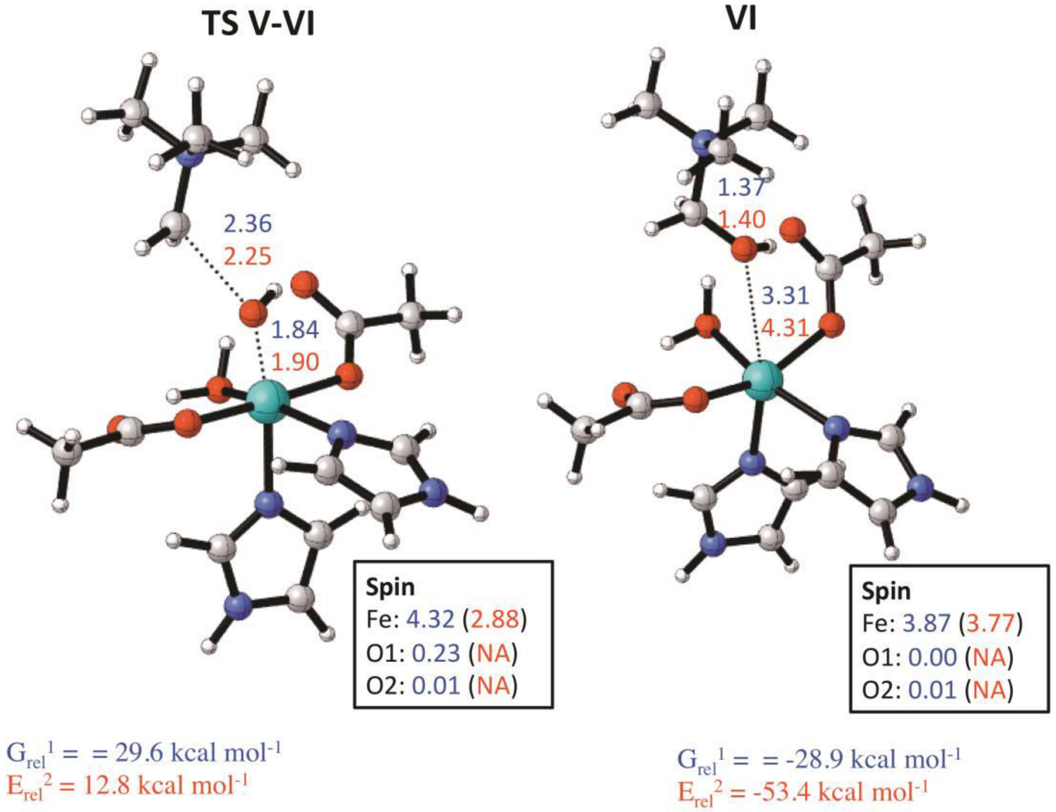


Figure 6.7. Small model representation of the QM/MM calculations (the protein environment is ommited for a better visualization of the active site). UB3LYP-D3/6-311G+(d,p):Amber; Fe:LANL2TZ(f) intermediates and transition structures for the –OH rebound in JMJD2A. Blue: UB3LYP-D3/6-311+G(d,p):Amber ; Fe:LANL2TZ(f) ECP. Red: UB3LYP-D3/6-31G+(d):Amber ; Fe: LACVP ECP.[4] Distances in Å are shown near the bonds.

The radical rebound step is often found to proceed with relatively small barriers (< 5 kcal mol^{-1}) when studied in model systems of the active site. Previously, however, de Visser calculated the TS barrier of –OH rebound in the AlkB enzyme with QM/MM using the UB3LYP-D3/6-31G+(d):Amber; Fe: LACVP level of theory to be about +12.8 kcal mol^{-1}. This higher value observed by de Visser was explained by the need of breaking a network of hydrogen bonds not only between the hydroxyl group and an active site aspartate residue, but also with an arginine residue, which interacts directly with the OH group. An analogous arginine or another hydrogen bonding residue is not present in the JMJD2A active site, and so this explanation is not applicable in our case. Our QM only calculations show relatively small electronic energy values for the TS barrier of –OH rebound (-0.7 kcal mol^{-1}, **Figure 6.6**), as observed in other DFT models on similar iron-dependent active sites.[1] However, the inclusion of the protein environment increased the TS barrier to +13.0 kcal mol^{-1}, thus suggesting this as the rate-determining step of the demethylation reaction. In AlkB, despite the high energy barrier of the rebound, the TS corresponding to this step is marginally lower (by 1.8 kcal mol^{-1}) than the TS for H-abstraction. It is important to notice that energy values reported by de Visser and co-workers did not include free energy corrections, which, according to our calculations, could reduce the H-abstraction TS barrier by up to 5 kcal mol^{-1}, while not having an important role for the –OH rebound step. By reducing the amount of the Hartree Fock exchange to 15% in the B3LYP functional (B3LYP*), the –OH rebound step was found to be the rate-determining step of the reaction by de Visser and co-workers.[4]

Using a truncated active site model, Cossío and co-workers[5] found the TS of the H-abstraction to be 6.6 kcal mol^{-1} higher than the TS energy of the rebound step, and then suggesting the former as the rate-determining step of the reaction. However, given the protein environment contributions can increase the TS corresponding to the –OH rebound by more than 10 kcal mol^{-1} according to our calculations, this could change considerably the kinetics of the reaction.

A clear trend emerges from QM/MM studies including the whole enzyme: the radical rebound step is substantially more difficult than predicted by model studies. As with C-H abstraction, it is likely that the protein restricts the substrate position and orientation, whereas it is completely free to move when a small model is used. Our finding that the rebound step is actually higher than C-H abstraction step should be interpreted more cautiously, however, since we have shown that these barriers are sensitive to the level of theory used.

The high energy value (+14.2 kcal mol^{-1}) of the low level of theory for the –OH rebound should be considered carefully. This value could be a result of the chosen point charge model derived from QM calculations preceding the O_2 binding step,[17] and different models may have a significant impact on the electrostatic energies. Also, the flat Potential Energy Surface corresponding to the QM region of our QM/MM calculations indicate that changes in the size of the active site treated at the high level of theory would also have the potential of changing the kinetics of the reaction.

The product following –OH rebound is highly stable, with an overall free energy of -28.2 kcal mol^{-1} (**Figure 6.6**). Very negative values for this step of the reaction have already been reported by other studies. De Visser and co-workers[4] showed that this final complex is -53.4 kcal mol^{-1} lower in energy in AlkB, while Cossío and co-workers showed this step to be -50.8 kcal mol^{-1} by using a truncated model active site of JMJD2A.

6.4 The role of the protein environment for evaluating the reaction mechanism of trimethyllysine demethylation by JMJD2A

Following our calculations on the oxygen activation, H-abstraction and –OH rebound steps, the potential energy surface for the demethylation reaction in JMJD2A has been determined (**Figure 6.8**). For the oxygen activation, inclusion of the protein environment stabilizes intermediates **II** and **III** by around 14.0 kcal mol^{-1}. QM calculations with a truncated model of TauD performed by Borowski and co-workers[22] at the UB3LYP-D3/cc-pVTZ Fe:LANL2DZ(f) level of theory suggest the intermediate **III** to be at around 6.0 kcal mol^{-1} higher in energy than ours, although still suggesting irreversibility for the oxygen activation steps. Interestingly, non-active residues treated at the low level of theory (Amber) had an opposite effect on the H-abstraction and –OH rebound in our calculations. The energies for the H-abstraction were slightly increased by considering the protein environment, while a more significant role was observed during the –OH rebound, with a positive energy contribution by up to +14.2 kcal mol^{-1}. Due to this increase in these energies, our calculations suggest the –OH rebound as the rate-determining step.

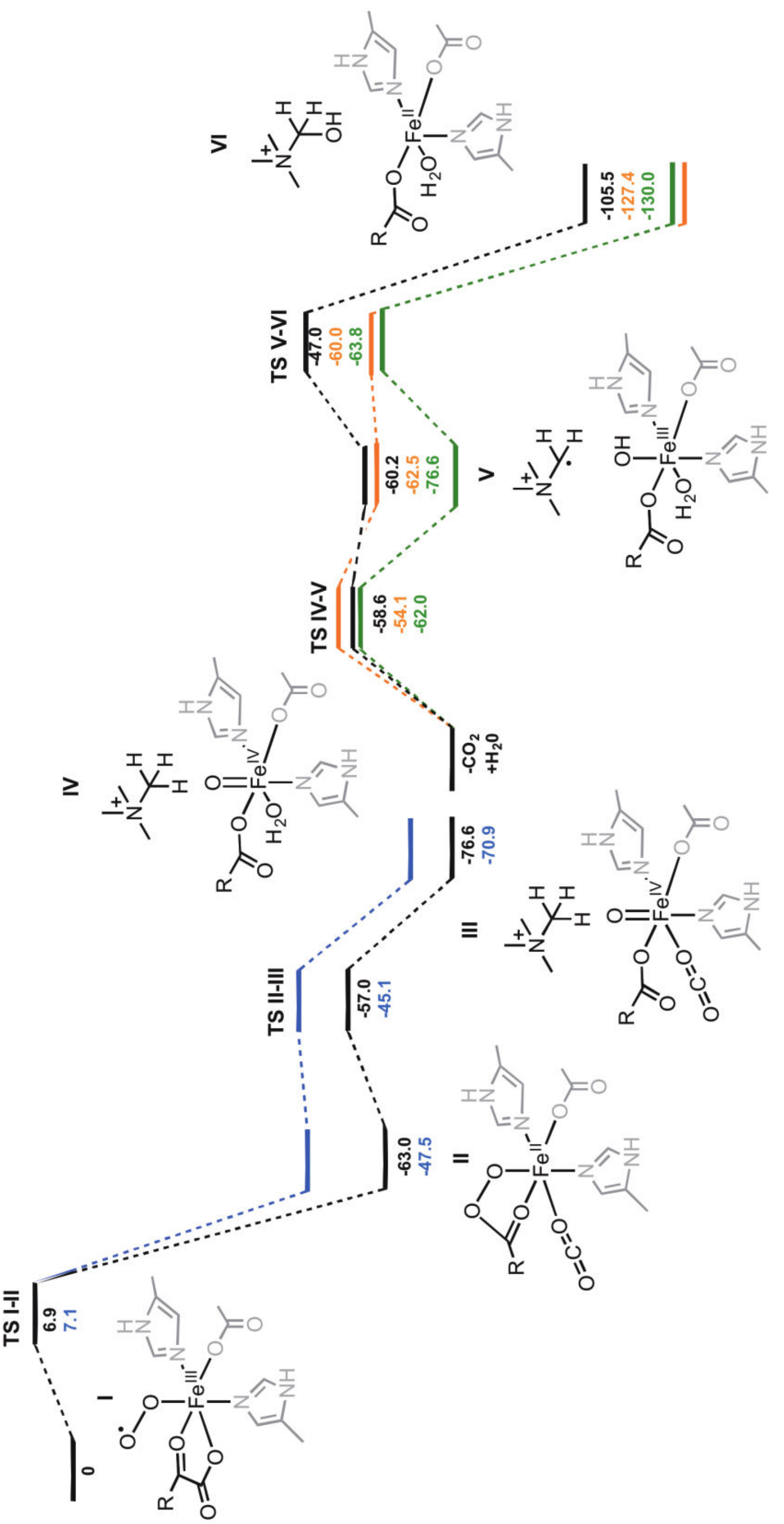


Figure 6.8. Energy reaction profile for the demethylation reaction in JMJD2A. Black: ΔG:UB3LYP-D3/6-311+G(d,p)/Amber Fe:LANL2DZ(f) ECP. Blue: ΔG:UB3LYP-D3/cc-pVTZ Fe:LANL2DZ(f) ECP (TauD).[22] Orange: ΔG:UB3LYP-D3/6-311+G(2d,2p) Fe:LANL2DZ ECP (JMJD2A).[5] Green: ΔE:UB3LYP-D3/6-311+G(2d,2p) Fe:LANL2DZ ECP (AlkB)[4]. All values in kcal mol^{-1}.

6.5 Energy decomposition analysis: Choosing point charges – a challenge for decomposing ONIOM QM/MM energies

Our studies for the dioxygen binding in JMJD2A showed the importance of considering the protein environment for understanding the chemical reaction profile of the demethylation reaction. By doing so, we were able to describe the favourability of this process, being -2.0 kcal mol^{-1} lower in energy than the unbound iron active site. We then decomposed the energy contribution from non-active site residues to the total energy potential surface considering: 1) Coulombic electrostatics and 2) vdW energies (For a better description of these calculations, please see our discussion in **Section 2.6.1** of **Chapter 2**). We found that charged residues in the vicinity of the active site had an important role for determining the overall energy, and suggested that mutations of these residues may have an impact in the occurrence of the dioxygen binding step. For the subsequent steps of the demethylation reaction, we performed an energy decomposition analysis for the step of the demethylation reaction that showed to be the most affected by the protein environment: the nucleophilic attack of the distal oxygen to the oxoglutarate cofactor (**Figure 6.8**). For this step, -14.3 kcal mol^{-1} came from non-active site residues treated at the low level of theory using ONIOM (**Figure 6.3**).

Amber vdW parameters were taken from the Amber ff12SB force field.[23] RESP atomic point charges for these regions of the protein were obtained from the R.E.D. (Resp and Esp Derivation) server version 4.0[24] using the B3LYP functional, the integral equation formalism polarizable continuum solvent model (IEFPCM; ε=4, ether as solvent) and the cc-pVTZ basis set. The other residues were assigned standard Amber charges following B3LYP/cc-pVTZ//HF/6-31G** calculations.[25] For decomposing the electrostatic energies, we took the Mulliken charges obtained after each step of the reaction. Although Mulliken population analysis is highly dependent of the choice of basis sets and can be inaccurate for intermolecular interaction energies,[26] our predicted energy contribution coming from the non-QM region for the dioxygen binding in JMJD2A was −8.1 kcal mol^{-1} kcal mol^{-1}, close to the calculated ONIOM QM/MM contribution −6.7 kcal mol^{-1}.[2] This energy difference was attributed to other bonded energies evaluated during QM/MM calculations but not considered in our analysis, for example, due to stretching, bending, torsion and out-of-plane torsion.

Hirao[27] suggested an alternative way of decomposing the energy for the ONIOM low level of theory. Their proposed methodology consists of multiple iterations with an updated set of Merz–Kollman (MK) charges, until the convergence of energy (<0.025 kcal mol^{-1}) is achieved. The redefinition of charges, however, is more expensive than the use of Mulliken charges following one-step optimization, especially for multi-steps reactions. Due to the higher cost of the method suggested by Hirao, we turned to the use of Mulliken charges following one-step optimization.

Electrostatics and vdW contributions of residues outside of the QM region were determined by using **Equations 2.15** and **2.16** **(Chapter 2)**. The total energy (+22.2 kcal mol^{-1} for electrostatics and -1.9 kcal mol^{-1} for vdW) calculated by the use of these equations (**Figure 6.9 and 6.10**) was not representative of the favourable protein environment contribution following our ONIOM QM/MM calculations (-14.3 kcal mol^{-1}). We then analysed if changing Mulliken by MK charges would make these values closer. However, we again failed in describing these electrostatics contributions (MK charges resulted in + 21.8 kcal mol^{-1} for electrostatic energies – **Figure 6.10**). The observed discrepancies in MM-ONIOM energies and electrostatic calculations using the standard Coulombic equation with point charges are due to difficulties to represent polarization effects in ONIOM electronic embedding methods, especially in systems with high changes in the geometries during a reaction path.[28] In the electronic embedding scheme, the electrostatic interactions between atoms from the active site and other residues are not computed classically as pairwise Coulomb interactions between point charges. Instead, these interactions are included in the self-consistent field (SCF) energy as a result of the protein environment charges onto the model system SCF density. So, any point charge assignment could result in an inaccurate representation of the electrostatic interactions with the electron density. For the oxygen activation, it is possible to see the release of the $–CO_2$ after the nucleophilic attack of the oxygen to the oxoglutarate cofactor. This is a considerable geometric change, electrostatically taken into account by the use of electronic embedding, although difficult to reproduce by using point charge models following one-step optimization.

The decomposition of vdW energies (-1.9 kcal mol^{-1}) is not affected by the choice of the point charge model; neither by polarization effects, and the use of **Equation 2.16** for these energy contributions is accurate. The low value due to vdW interactions (**Figure 6.11**) suggests that most of the protein environment energy contributions comes from electrostatics. This shows that, as previously observed for the dioxygen binding, these processes affect the charges of the residues in the active site, and then charged surrounding residues would play an important role in the determination of the energy potential surface. Even giving inaccurate predictions of the electrostatic contributions per residue, we can observe most of the amino acids that presented significant contributions for this step are charged (Glu184, Lys200, Arg233, Glu281). This suggests that mutations of such residues may affect the profile of the reaction, although due to inaccuracy of our energy decomposition analysis method it is difficult to predict which mutations would favour/disfavour the reaction.

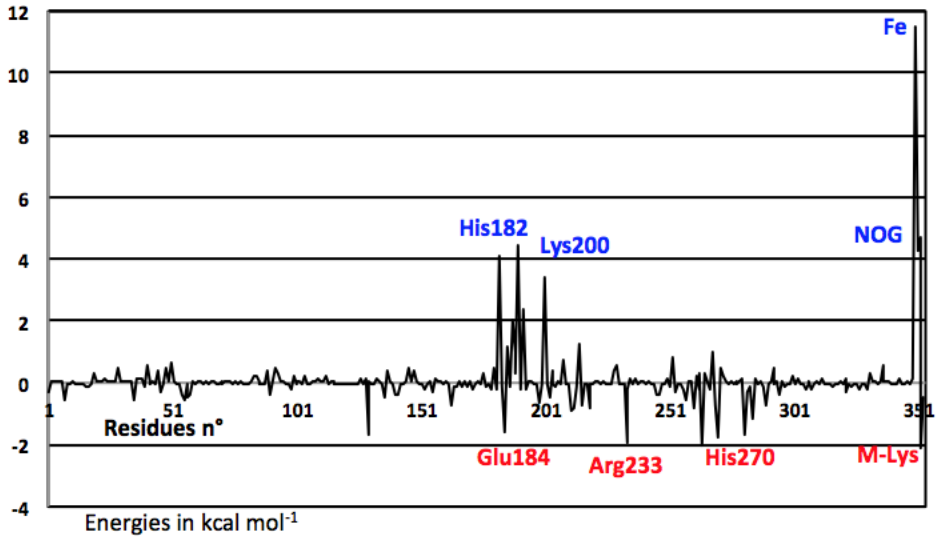


Figure 6.9. Electrostatic per-residue energy decomposition analysis of the nucleophilic attack of the distal oxygen to the oxoglutarate cofactor in JMJD2A based on a QM/MM description using Mulliken charges.

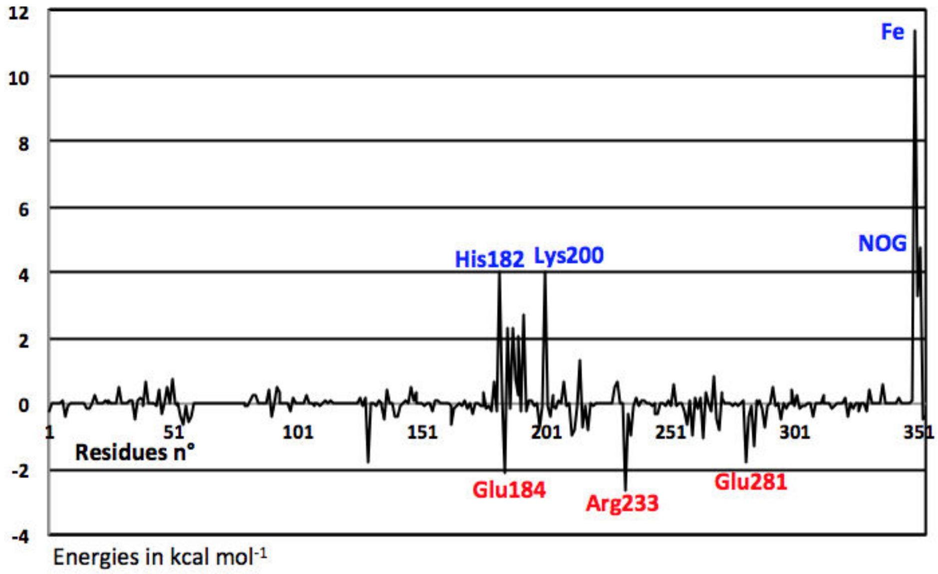


Figure 6.10. Electrostatic per-residue energy decomposition analysis of the nucleophilic attack of the distal oxygen to the oxoglutarate cofactor in JMJD2A based on a QM/MM description using MK charges.

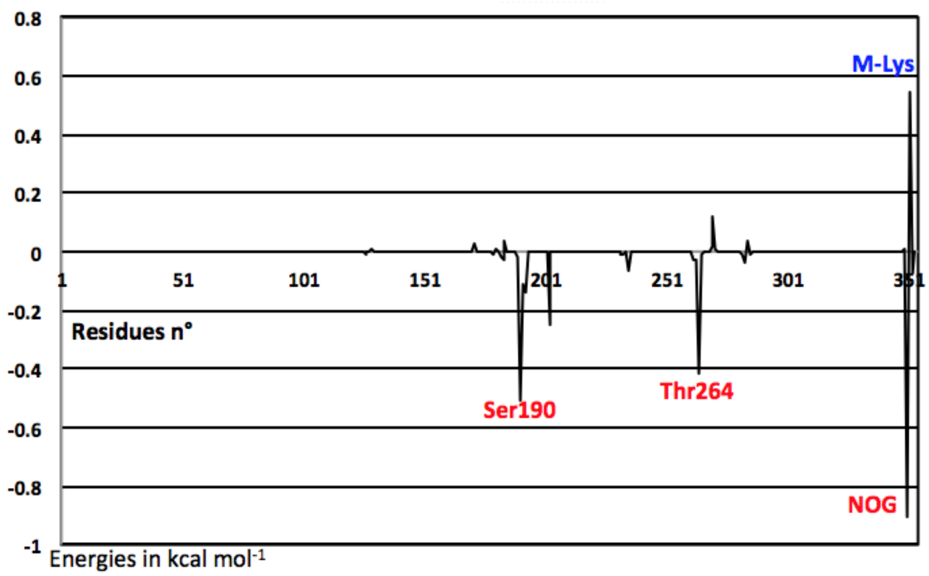


Figure 6.11. vdW per-residue energy decomposition analysis of the nucleophilic attack of the distal oxygen to the oxoglutarate cofactor in JMJD2A.

6.6 Conclusion

The full reaction profile of the demethylation reaction in proteins from the Jumonji C family is yet to be fully understood. Here, we provide insights from the first QM/MM study of this process. The rate-determining step has been studied as well as the importance of the protein environment for the overall determination of the energy potential surface. In contrast with other non-haem iron-dependent enzymes such as AlkB, the –OH rebound step is shown here to be rate-determining. We interpret this data cautiously, since the barriers show sensitivity to several modelling assumptions, such as the choice of the charge point model for the calculations at the low level of theory. Nevertheless, taking our and previous QM/MM studies into account from AlkB, a general trend emerges in which the radical rebound step is decidedly more difficult than is expected based on geometrically-unrestricted model studies. The positioning of the substrate, which is captured by our full model has an appreciable effect upon the reacting geometries at the active site. We also observe deviation from linearity in the CH abstraction TS as a result of this effect.

Contributions coming from the protein environment did not have a significant role for the H-abstraction step. However, they lowered the energy of the oxygen activation steps by up to -14.3 kcal mol^{-1}, while increasing the energy for the TS of the –OH rebound by approximately +13.0 kcal mol^{-1}. While we were unable to provide an accurate energy decomposition analysis by using Mulliken or Merz–Kollman point charges following one-step optimizations, our calculations suggest that charged residues in the vicinity of the active site have a high contribution for the energy potential surface. Mutations of such charged residues may affect the favourability of the demethylation reaction in proteins from the Jumonji C family.

6.7 Supporting information for Chapter 6

Cartesian coordinates:

Quintet $^{-2}$ O_2 binding:

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1305.0504
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.0278
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1468
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1297.4555
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1298.3411
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7092
Low-model (Amber) = 0.0012
Low-real (Amber) = -8.3348
Protein environment = -8.3360

C	25.609	26.077	26.930
N	24.913	27.105	26.319
H	24.011	27.054	25.847
C	25.592	28.251	26.539
H	25.300	29.225	26.184
N	26.683	28.002	27.246
C	26.716	26.646	27.499
H	27.536	26.205	28.039
C	30.363	27.404	30.891
H	29.459	27.431	31.463
H	31.118	27.976	31.389
C	30.108	27.995	29.507
O	31.012	28.143	28.697
O	28.839	28.281	29.262
C	24.294	29.743	30.579
N	25.193	29.122	31.426
H	25.003	28.851	32.392
C	26.387	29.069	30.795
H	27.285	28.635	31.203
N	26.300	29.636	29.599
C	24.993	30.059	29.446
H	24.679	30.545	28.536
Fe	27.900	29.498	28.143
C	28.716	32.115	27.509
C	27.489	31.837	26.624
O	29.399	33.116	27.342
O	28.902	31.183	28.407
O	26.713	30.925	26.987
C	27.145	32.803	25.537
H	28.055	33.232	25.115
H	26.554	32.314	24.760
H	26.550	33.620	25.967
O	28.832	29.422	26.225
O	28.956	30.516	25.560
H	30.692	26.391	30.792
H	23.259	29.938	30.772
H	25.337	25.043	26.953

TS I -> II (Oxygen activation)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1305.0363
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.0188
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1459
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1297.4434
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1298.3262
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7101
Low-model (Amber) = 11.8664
Low-real (Amber) = 3.5317
Protein environment = -8.3347
1 Imaginary Frequency = -380.82

C	25.58800	25.95800	26.91900
N	24.88000	26.99500	26.33700
H	23.97100	26.95500	25.87700
C	25.56500	28.13600	26.55700
H	25.27000	29.11800	26.22300
N	26.66800	27.87500	27.23700
C	26.70600	26.51800	27.47300
H	27.54000	26.06800	27.98500
C	30.32000	27.38700	30.89200
H	29.39700	27.41800	31.43200
H	31.05800	27.96000	31.41500
C	30.11400	27.97500	29.49800
O	31.05200	28.12600	28.72800
O	28.85600	28.24800	29.19800
C	24.33800	29.72500	30.54600
N	25.23600	29.10300	31.39500
H	25.04700	28.84000	32.36300
C	26.42700	29.03600	30.75900
H	27.32300	28.59800	31.16900
N	26.34200	29.59400	29.56000
C	25.03800	30.03000	29.40900
H	24.73300	30.52000	28.49800
Fe	27.88500	29.40200	28.03800
C	28.84600	32.09500	27.55600
C	27.73200	31.47500	26.33000
O	29.36500	33.16500	27.32400
O	28.93400	31.23400	28.47300
O	26.82700	30.79300	26.97200
C	27.30600	32.58600	25.39600
H	28.18200	33.10200	25.00200
H	26.72300	32.16900	24.57000
H	26.68800	33.29500	25.95200
O	28.89400	29.46600	26.32700
O	28.74500	30.68300	25.68900
H	30.64900	26.37300	30.80700
H	23.30500	29.92800	30.73900
H	25.31400	24.92300	26.93300

Intermediate II (Oxygen activation)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1305.1256
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.1037
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1469
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1297.5297
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1298.4181
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7075
Low-model (Amber) = 35.1234
Low-real (Amber) = 26.7645
Protein environment = -8.3589

C	25.61000	26.02000	26.79300
N	24.89700	27.04500	26.19900
H	23.96700	27.00000	25.78200
C	25.60700	28.18600	26.35800
H	25.28400	29.16000	26.02800
N	26.73800	27.94200	26.99700
C	26.75800	26.59100	27.27800
H	27.58800	26.16700	27.81800
C	30.37500	27.45000	31.02400
H	29.48600	27.41800	31.61900
H	31.12200	28.02600	31.52900
C	30.06100	28.09500	29.67000
O	30.99500	28.43000	28.94000
O	28.79400	28.18800	29.38200
C	24.26300	29.68200	30.55800
N	25.19400	29.17800	31.44800
H	25.02000	28.97300	32.43200
C	26.38400	29.11200	30.80500
H	27.30200	28.75300	31.24500
N	26.26500	29.55200	29.56100
C	24.94100	29.90900	29.38900
H	24.60000	30.29900	28.44400
Fe	27.88500	29.40600	28.09500
C	29.27000	32.35800	28.86400
C	27.23100	31.63000	26.13200
O	29.71100	33.15500	28.13100
O	28.78500	31.61800	29.63000
O	26.50500	30.98100	26.88800
C	26.74700	32.70500	25.19100
H	27.56200	33.37900	24.91900
H	26.36700	32.24700	24.27100
H	25.93600	33.26100	25.66300
O	29.06800	30.48600	26.96300
O	28.53600	31.46300	26.04000
H	23.22500	29.85900	30.74500
H	25.32400	24.99200	26.86100
H	30.73700	26.45500	30.87000

TS II-> III (Oxygen activation)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1305.1108
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.0937
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1452
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1297.5169
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1298.4044
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7064
Low-model (Amber) = 50.0621
Low-real (Amber) = 41.7042
Protein environment = -8.3579
1 Imaginary Frequency = -476.52

C	25.62900	26.02100	26.82300
N	24.93500	27.05200	26.21300
H	24.01100	27.01200	25.78200
C	25.65000	28.18700	26.38100
H	25.34800	29.16600	26.04600
N	26.76600	27.93100	27.04400
C	26.77000	26.58100	27.33100
H	27.58800	26.15300	27.88600
C	30.33400	27.43900	30.99700
H	29.44200	27.41400	31.58700
H	31.07600	28.02700	31.49700
C	30.02500	28.06100	29.63100
O	30.95700	28.34800	28.88500
O	28.74800	28.18400	29.36100
C	24.28900	29.71100	30.57100
N	25.21100	29.18000	31.45600
H	25.02900	28.95300	32.43400
C	26.40800	29.13100	30.82700
H	27.31900	28.75900	31.26700
N	26.30100	29.60900	29.59400
C	24.97900	29.97400	29.41700
H	24.65400	30.39500	28.47900
Fe	27.85900	29.42900	28.09700
C	29.24400	32.47200	28.89300
C	27.29800	31.61700	26.16600
O	29.76800	33.18300	28.12700
O	28.67800	31.82600	29.68800
O	26.56400	30.91500	26.90900
C	26.69200	32.69000	25.27600
H	25.86100	33.17400	25.79200
H	27.44500	33.42900	24.99500
H	26.31400	32.22800	24.35800
O	29.13800	30.31200	27.27300
O	28.56800	31.51600	26.06200
H	30.70000	26.44300	30.86200
H	23.24800	29.88000	30.75100
H	25.33300	24.99500	26.88500

Intermediate III (Oxygen Activation)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1305.1452
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1297.1259
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1470
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1297.5494
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1298.4373
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7079
Low-model (Amber) = 77.9865
Low-real (Amber) = 69.6278
Protein environment = -8.3587

C	25.60500	26.15900	26.87900
N	24.92600	27.20400	26.27800
H	24.01500	27.16700	25.82000
C	25.63300	28.33700	26.50600
H	25.35500	29.32700	26.17800
N	26.72100	28.05900	27.19900
C	26.72600	26.70700	27.44400
H	27.53200	26.26100	28.00200
C	30.33900	27.44100	31.02000
H	29.43000	27.41600	31.58400
H	31.07600	28.00300	31.55600
C	30.08100	28.10100	29.66300
O	31.02200	28.31600	28.91300
O	28.80600	28.35200	29.43300
C	24.36900	29.68700	30.56900
N	25.26500	29.11500	31.45400
H	25.06900	28.88100	32.42700
C	26.46700	29.04100	30.83700
H	27.36400	28.64600	31.28600
N	26.38700	29.54200	29.61300
C	25.08100	29.95300	29.42900
H	24.78300	30.40800	28.49800
Fe	27.96500	29.51500	28.15700
C	29.15500	32.36100	29.65600
C	27.60500	31.98000	26.70200
O	29.39200	33.49600	29.78000
O	28.90400	31.21700	29.63700
O	27.00800	30.99300	27.33500
C	26.80600	32.52100	25.52500
H	27.42100	33.20400	24.93500
H	26.45100	31.70300	24.88900
H	25.93300	33.06700	25.89800
O	29.12700	29.50900	27.03300
O	28.67700	32.49500	27.02000
H	30.69300	26.44200	30.87000
H	25.31300	25.13000	26.89900
H	23.33100	29.88200	30.73900

Intermediate III (Extended QM – inclusion of M-lysine)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1407.2250
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1399.2542
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1687
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1399.6104
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1400.4913
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7337
Low-model (Amber) = -0.0121
Low-real (Amber) = -8.2635
Protein environment = -8.2513

C	-2.86400	-5.05800	-1.89400
N	-3.53700	-4.02300	-2.51900
H	-4.44200	-4.07200	-2.98700
C	-2.85000	-2.88000	-2.27800
H	-3.12100	-1.89600	-2.62700
N	-1.78000	-3.14400	-1.55300
C	-1.76600	-4.49600	-1.30000
H	-0.96800	-4.92500	-0.71700
C	1.87800	-3.84100	2.21700
H	0.97600	-3.93900	2.78500
H	2.60100	-3.29500	2.78600
C	1.58300	-3.09300	0.91600
O	2.51100	-2.55500	0.30400
O	0.32800	-3.06100	0.56900
C	-4.03700	-1.51500	1.73900
N	-3.14000	-2.09800	2.61600
H	-3.34100	-2.34900	3.58400
C	-1.93600	-2.15300	2.00000
H	-1.04200	-2.56600	2.43600
N	-2.01300	-1.62800	0.78600
C	-3.32200	-1.22000	0.60800
H	-3.61900	-0.74800	-0.31500
Fe	-0.45000	-1.72300	-0.69400
C	-0.87200	0.82900	-2.13700
O	-1.44800	-0.20800	-1.62100
C	-1.65900	1.52600	-3.23500
H	-1.03200	2.25900	-3.74600
H	-2.04000	0.79800	-3.95700
H	-2.51700	2.04200	-2.79100
C	4.77200	-2.07900	-3.64800
H	5.03600	-2.76400	-2.87000
H	3.91700	-2.44900	-4.17400
N	4.44800	-0.72900	-3.04400
C	3.61700	-0.97100	-1.80600
H	2.80100	-1.65300	-2.03300
H	4.24900	-1.39800	-1.03500
H	3.20500	-0.01300	-1.48600
C	3.60700	0.09800	-3.97300
H	3.44000	1.07000	-3.50600
H	4.12600	0.21300	-4.91900
H	2.65200	-0.40800	-4.11300
C	5.68400	0.03800	-2.66000
H	6.35900	-0.63000	-2.12400
H	6.17800	0.43300	-3.55000
H	5.36600	0.85900	-2.02100
O	0.62700	-1.88100	-1.89100
O	0.22700	1.30900	-1.78300
O	0.65400	-0.20800	0.32300
H	1.58200	-0.51600	0.38900
H	0.65600	0.49600	-0.39500
H	2.26300	-4.81300	1.99000
H	-5.07800	-1.33100	1.91000
H	-3.14600	-6.09000	-1.87700
H	5.59300	-1.97800	-4.32700

TS (H-Abstraction)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1407.1608
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1399.2208
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1612
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1399.5530
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1400.4355
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7253
Low-model (Amber) = 43.0812
Low-real (Amber) = 34.8340
Protein environment = -8.2472

1 Imaginary Frequency = -1021.05

C	25.74300	26.36000	26.81500
N	25.09300	27.40900	26.18800
H	24.16500	27.38600	25.76400
C	25.86100	28.51800	26.34800
H	25.61300	29.50600	25.99400
N	26.96100	28.22500	27.02000
C	26.90600	26.88100	27.32200
H	27.70100	26.41700	27.88200
C	30.43300	27.44700	30.93600
H	29.50700	27.41500	31.47300
H	31.15200	28.00400	31.50000
C	30.21300	28.12200	29.58200
O	31.19400	28.39800	28.88100
O	28.97500	28.34100	29.26100
C	24.52900	29.81300	30.40200
N	25.42900	29.22300	31.27000
H	25.23100	28.95400	32.23300
C	26.63100	29.17200	30.64600
H	27.52700	28.75600	31.07500
N	26.55400	29.70700	29.43800
C	25.24300	30.11400	29.26900
H	24.93700	30.59200	28.35100
Fe	28.27000	29.69800	27.99400
C	27.79400	32.23700	26.46800
O	27.28700	31.15800	26.96100
C	26.99700	32.86600	25.33600
H	27.56400	33.67700	24.87200
H	26.74300	32.11300	24.58400
H	26.06300	33.27200	25.73700
C	32.93900	29.17000	25.22200
H	33.20900	28.54600	26.04900
H	32.09900	28.74400	24.71400
N	32.57300	30.55700	25.73200
C	31.82100	30.44300	27.00100
H	30.58600	29.49100	27.08600
H	32.37800	29.91900	27.76800
H	31.36200	31.39100	27.26000
C	31.69600	31.28800	24.75000
H	31.52900	32.29800	25.13100
H	32.19300	31.31400	23.78500
H	30.74600	30.75700	24.69800
C	33.81300	31.39200	26.03100
H	34.51600	30.77800	26.59300
H	34.26300	31.73700	25.09600
H	33.49500	32.24600	26.62200
O	29.58400	29.58300	26.75200
O	28.83600	32.80700	26.86800
O	29.21600	31.36700	29.04300
H	30.15900	31.31900	29.30700
H	29.19000	32.04400	28.30100
H	30.79200	26.45100	30.78400
H	23.49000	29.99700	30.57700
H	25.40600	25.34700	26.88900
H	33.76500	29.24500	24.54700

After Abstraction

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1407.1954
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1399.2227
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1643
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1399.5837
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1400.4678
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7277
Low-model (Amber) = 106.2197
Low-real (Amber) = 97.9691
Protein environment = -8.2506

C	-2.81500	-4.99300	-1.86000
N	-3.46900	-3.95200	-2.49500
H	-4.37000	-3.99500	-2.97200
C	-2.76800	-2.81700	-2.24200
H	-3.02600	-1.83000	-2.59400
N	-1.70900	-3.08500	-1.50100
C	-1.71900	-4.43800	-1.25200
H	-0.93700	-4.88400	-0.66000
C	1.94800	-3.84100	2.21100
H	1.05900	-3.89700	2.80500
H	2.69200	-3.27900	2.73600
C	1.63200	-3.15300	0.88500
O	2.55200	-2.91800	0.08200
O	0.38100	-2.87600	0.69200
C	-4.06000	-1.46200	1.74500
N	-3.17200	-2.07300	2.61100
H	-3.38200	-2.35100	3.57000
C	-1.95900	-2.10000	2.01300
H	-1.06700	-2.52600	2.43900
N	-2.02200	-1.52900	0.81900
C	-3.33100	-1.12200	0.63400
H	-3.61600	-0.62400	-0.27900
Fe	-0.43100	-1.54900	-0.54800
C	-0.80100	0.87500	-2.18800
O	-1.36600	-0.14800	-1.63700
C	-1.52800	1.43600	-3.40000
H	-0.90700	2.17800	-3.90700
H	-1.78900	0.63200	-4.09500
H	-2.45500	1.91700	-3.07300
C	4.74900	-2.08100	-3.63900
H	5.01900	-2.75700	-2.85500
H	3.88400	-2.45200	-4.14800
N	4.44100	-0.71200	-3.04500
C	3.69200	-0.89900	-1.80200
H	1.55900	-2.21000	-1.34400
H	4.05300	-1.63200	-1.10200
H	2.85400	-0.24000	-1.62800
C	3.56500	0.09400	-3.97000
H	3.41200	1.07500	-3.51900
H	4.05800	0.18300	-4.93100
H	2.60900	-0.42200	-4.06100
C	5.70100	0.07100	-2.72500
H	6.38800	-0.58200	-2.18800
H	6.16200	0.43600	-3.64700
H	5.39600	0.91100	-2.10600
O	0.85700	-1.67000	-1.77300
O	0.23200	1.45300	-1.78200
O	0.54100	0.13500	0.47200
H	1.49700	0.02100	0.65700
H	0.52800	0.75900	-0.31800
H	2.31400	-4.82800	2.02200
H	-5.10300	-1.29000	1.90800
H	-3.10500	-6.02300	-1.84600
H	5.56000	-1.99500	-4.33100

TS (OH rebound)

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1407.1612
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1399.2238
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1653
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1399.5493
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1400.4320
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7292
Low-model (Amber) = -0.0739
Low-real (Amber) = -8.3027
Protein environment = -8.2288
1 Imaginary Frequency = -215.0774

C	-2.79300	-5.07700	-1.98800
N	-3.46300	-4.03700	-2.61000
H	-4.39700	-4.06300	-3.02000
C	-2.70200	-2.92400	-2.48600
H	-2.96800	-1.94100	-2.83900
N	-1.58200	-3.20500	-1.83800
C	-1.62100	-4.54700	-1.51900
H	-0.80300	-4.99100	-0.97700
C	1.88500	-3.78400	2.33500
H	2.72700	-3.29500	2.77800
H	1.21800	-4.11300	3.10400
C	1.15600	-2.80900	1.41300
O	1.06000	-1.64500	1.88800
O	0.60900	-3.17300	0.32100
C	-4.04000	-1.64600	1.63500
N	-3.17300	-2.29300	2.49400
H	-3.36900	-2.54200	3.46400
C	-1.98500	-2.42400	1.86200
H	-1.12200	-2.86800	2.32700
N	-2.03900	-1.88300	0.65100
C	-3.32300	-1.39300	0.49400
H	-3.59900	-0.89300	-0.42100
Fe	-0.43800	-1.71900	-0.76700
C	-0.87800	0.72300	-2.46400
O	-1.43200	-0.33700	-1.99200
C	-1.76600	1.60000	-3.33300
H	-1.19300	2.43200	-3.74600
H	-2.19000	1.01100	-4.15200
H	-2.59500	1.98900	-2.73600
C	4.63800	-2.09200	-3.55300
H	4.93600	-2.79800	-2.80600
H	3.78500	-2.46800	-4.07900
N	4.28100	-0.77400	-2.88900
C	3.41900	-1.04600	-1.71800
H	0.98700	-0.61800	-2.16400
H	3.65600	-1.93100	-1.14900
H	3.05300	-0.15000	-1.24300
C	3.48800	0.11700	-3.81800
H	3.21300	1.01900	-3.27000

H	4.10100	0.36700	-4.68200
H	2.58900	-0.42200	-4.11500
C	5.51800	-0.01600	-2.42300
H	6.16900	-0.70100	-1.88300
H	6.04200	0.39500	-3.29100
H	5.18100	0.79500	-1.78000
O	1.10100	-1.51000	-1.76200
O	0.31200	1.08100	-2.25900
O	-0.00300	0.09200	0.32600
H	0.59900	-0.34200	0.99200
H	0.52100	0.61300	-0.31600
H	-5.06300	-1.39400	1.82200
H	-3.12500	-6.09000	-1.89300
H	2.22000	-4.62800	1.76800
H	5.44200	-1.93900	-4.24100

After rebound

E[UB3LYP-D3/6-31G(d),LANL2DZ:Amber] = -1407.2854
E[UB3LYP-D3/6-311+G(d,p),LANL2TZ(f)]= -1399.3222
ZPE[UB3LYP-D3/6-31G(d),LANL2DZ] = 7.1711
H[UB3LYP-D3/6-31G(d),LANL2DZ] = -1399.6676
G[UB3LYP-D3/6-31G(d),LANL2DZ] = -1400.5513
G-E[UB3LYP-D3/6-31G(d),LANL2DZ] = 6.7341
Low-model (Amber) = 0.0066
Low-real (Amber) = -8.2233
Protein environment = -8.2298

C	-2.89400	-5.23100	-1.99400
N	-3.55300	-4.18000	-2.60300
H	-4.47500	-4.19600	-3.03100
C	-2.79100	-3.06800	-2.46000
H	-3.04800	-2.08800	-2.83100
N	-1.68500	-3.35300	-1.79600
C	-1.73100	-4.69800	-1.50100
H	-0.91200	-5.14300	-0.96000
C	1.88200	-3.79000	2.26100
H	2.73000	-3.28900	2.67900
H	1.24400	-4.13000	3.05100
C	1.11300	-2.82800	1.36600
O	0.87200	-1.72300	1.91700
O	0.68600	-3.15200	0.19800
C	-4.24700	-1.68400	1.77900
N	-3.31100	-2.12800	2.68900
H	-3.47300	-2.31900	3.67700
C	-2.11300	-2.20000	2.06000
H	-1.19700	-2.47100	2.55900
N	-2.23000	-1.83500	0.79200
C	-3.56500	-1.51300	0.60200
H	-3.92200	-1.18700	-0.36300
Fe	-0.58500	-1.84200	-0.71200
C	-0.34500	0.30300	-2.81000
O	-1.04100	-0.63900	-2.31000
C	-1.09900	1.40000	-3.53300
H	-0.46300	2.27600	-3.67700
H	-1.40700	1.02800	-4.51800
H	-2.00300	1.66800	-2.98100
C	4.84500	-2.08400	-3.66200
H	5.15800	-2.77600	-2.90800
H	3.98700	-2.47300	-4.16900
N	4.49500	-0.77000	-3.02100
C	3.42700	-1.09500	-1.88500
H	1.65600	-0.96800	-2.58600
H	3.97500	-1.72500	-1.17900
H	3.22000	-0.10700	-1.45900
C	3.83600	0.17400	-3.98100
H	3.40300	1.00100	-3.42000
H	4.57800	0.54600	-4.68300
H	3.03800	-0.36200	-4.48400
C	5.67200	-0.09800	-2.40100
H	6.21400	-0.80200	-1.77200
H	6.34000	0.27000	-3.18400
H	5.31500	0.74400	-1.80800
O	2.34500	-1.70600	-2.40200
O	0.92100	0.36800	-2.75300
O	0.23400	0.03500	0.09900
H	0.64900	-0.41000	0.88600
H	0.95600	0.35100	-0.47700
H	2.21200	-4.62800	1.68300
H	-3.22300	-6.24700	-1.92200
H	-5.28800	-1.50900	1.95500
H	5.63900	-1.93800	-4.36400

6.8 References for Chapter 6

1. Liu, H.; Llano, J.; Gauld, J. W., A DFT study of nucleobase dealkylation by the DNA repair enzyme AlkB. *J Phys Chem B* **2009,** *113* (14), 4887-98.

2. Cortopassi, W. A.; Simion, R.; Honsby, C. E.; Franca, T. C.; Paton, R. S., Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem Eur J* **2015**, *21*(52), 18983-92.

3. Ye, S. F.; Neese, F., Nonheme oxo-iron(IV) intermediates form an oxyl radical upon approaching the C-H bond activation transition state. *Proc Natl Acad Sci U S A* **2011,** *108* (4), 1228-1233.

4. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P., Quantum mechanics/molecular mechanics study on the oxygen binding and substrate hydroxylation step in AlkB repair enzymes. *Chem Eur J* **2014,** *20* (2), 435-46.

5. Alberto, N.; Torrent-Sucarrat, M.; Arrastia, I.; Arrieta, A.; Cossío, F. P., Two-State Reactivity of Histone Demethylases Containing Jumonji-C Active Sites: Different Mechanisms for Different Methylation Degrees. *Chem Eur J* **2017**, *23*, 137-148.

6. Cisneros, G. A., DFT study of a model system for the dealkylation step catalyzed by AlkB. *Interdiscip Sci* **2010,** *2* (1), 70-7.

7. Geronimo, I.; Paneth, P., A DFT and ONIOM study of C-H hydroxylation catalyzed by nitrobenzene 1,2-dioxygenase. *Phys Chem Chem Phys* **2014,** *16* (27), 13889-99.

8. Godfrey, E.; Porro, C. S.; de Visser, S. P., Comparative quantum mechanics/molecular mechanics (QM/MM) and density functional theory calculations on the oxo-iron species of taurine/alpha-ketoglutarate dioxygenase. *J Phys Chem A* **2008,** *112* (11), 2464-8.

9. Ye, S. F.; Riplinger, C.; Hansen, A.; Krebs, C.; Bollinger, J. M.; Neese, F., Electronic Structure Analysis of the Oxygen-Activation Mechanism by Fe-II- and alpha-Ketoglutarate (alpha KG)-Dependent Dioxygenases. *Chem Eur J* **2012,** *18* (21), 6555-6567.

10. (a) Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M., Jr.; Krebs, C., The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **2003,** *42* (24), 7497-508; (b) Krebs, C.; Price, J. C.; Baldwin, J.; Saleh, L.; Green, M. T.; Bollinger, J. M., Jr., Rapid freeze-quench 57Fe Mossbauer spectroscopy: monitoring

changes of an iron-containing active site during a biochemical reaction. *Inorg Chem* **2005,** *44* (4), 742-57.

11. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Jr., Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). *J Am Chem Soc* **2003,** *125* (43), 13008-9.

12. Mehn, M. P.; Fujisawa, K.; Hegg, E. L.; Que, L., Jr., Oxygen activation by nonheme iron(II) complexes: alpha-keto carboxylate versus carboxylate. *J Am Chem Soc* **2003,** *125* (26), 7828-42.

13. (a) Borowski, T.; Bassan, A.; Siegbahn, P. E., 4-Hydroxyphenylpyruvate dioxygenase: a hybrid density functional study of the catalytic reaction mechanism. *Biochemistry* **2004,** *43* (38), 12331-42; (b) Borowski, T.; Bassan, A.; Siegbahn, P. E., Mechanism of dioxygen activation in 2-oxoglutarate-dependent enzymes: a hybrid DFT study. *Chem Eur J* **2004,** *10* (4), 1031-41.

14. de Visser, S. P., Can the peroxosuccinate complex in the catalytic cycle of taurine/alpha-ketoglutarate dioxygenase (TauD) act as an alternative oxidant? *Chem Com* **2007,** (2), 171-173.

15. Diebold, A. R.; Brown-Marshall, C. D.; Neidig, M. L.; Brownlee, J. M.; Moran, G. R.; Solomon, E. I., Activation of alpha-keto acid-dependent dioxygenases: application of an {FeNO}7/{FeO2}8 methodology for characterizing the initial steps of O2 activation. *J Am Chem Soc* **2011,** *133* (45), 18148-60.

16. Wojcik, A.; Radon, M.; Borowski, T., Mechanism of O2 Activation by alpha-Ketoglutarate Dependent Oxygenases Revisited. A Quantum Chemical Study. *J Phys Chem A* **2016,** *120* (8), 1261-74.

17. Cortopassi, W. A.; Simion, R.; Honsby, C. E.; Franca, T. C.; Paton, R. S., Dioxygen Binding in the Active Site of Histone Demethylase JMJD2A and the Role of the Protein Environment. *Chem Eur J* **2015,** *21* (52), 18869.

18. Couty, M.; Hall, M. B., Basis sets for transition metals: Optimized outer p functions. *J Comput Chem* **1996,** *17* (11), 1359-1370.

19. (a) Roy, L. E.; Hay, P. J.; Martin, R. L., Revised basis sets for the LANL effective core potentials. *J Chem Theor Comp* **2008,** *4* (7), 1029-1031; (b) Ehlers, A. W.; Bohme, M.; Dapprich, S.; Gobbi, A.; Hollwarth, A.; Jonas, V.; Kohler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G., A Set of F-Polarization Functions for Pseudo-Potential Basis-Sets of the Transition-Metals Sc-Cu, Y-Ag and La-Au. *Chem Phys Lett* **1993,** *208* (1-2), 111-114.

20. 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* (11), 1995-2001.

21. Walter, P.; Metzger, J.; Thiel, C.; Helms, V., Predicting where small molecules bind at protein-protein interfaces. *PloS one* **2013,** *8* (3), e58583.

22. Wójcik, A.; Radoń, M.; Borowski, T., Mechanism of O2 Activation by α-Ketoglutarate Dependent Oxygenases Revisited. A Quantum Chemical Study. *J Phys Chem A* **2016,** *120* (8), 1261-1274.

23. Case, D. A.; Darden, T. A.; Cheatham, T. E.; 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.; Goetz, 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**.

24. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y., R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res* **2011,** *39* (Web Server issue), W511-7.

25. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **2003,** *24* (16), 1999-2012.

26. Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J., Solvated ensemble averaging in the calculation of partial atomic charges. *J Comput Chem* **2001,** *22* (11), 1125-37.

27. Hirao, H., The effects of protein environment and dispersion on the formation of ferric-superoxide species in myo-inositol oxygenase (MIOX): a combined ONIOM(DFT:MM) and energy decomposition analysis. *J Phys Chem B* **2011,** *115* (38), 11278-85.

28. 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 Theory Comput* **2006,** *2* (3), 815-26.

Concluding remarks

This thesis discusses the use of state-of-the-art computational chemistry techniques for understanding epigenetic processes – chemical modifications to the structure of DNA and histone proteins that affect gene expression without changing DNA sequence. I used computation to probe the chemistry involved in these enzyme-catalysed mechanisms to study and design compounds capable of inhibiting these biological targets.

Firstly, I investigated the role of intrinsic interaction (*e.g.* hydrogen bonds and cation-π interactions) for the design of small molecules targeting the CREBBP bromodomain. In collaboration with Prof. Stuart Conway, I verified the dynamics of these interactions over time following MD simulations and showed that binding stability for a series of dihydroquinoxalinone derivatives (DHQ) in this protein is highly dependent on such interactions. The potential for the formation of a cation-π interaction, characterized by a positively charged group positioned favourably above the plane of π-bonding system, directly influenced the ability of DHQ derivatives to bind to the CREBBP receptor. As a continuation of this work, I then developed a novel computational model to predict the strength of these cation-π interactions in an attempt to improve selectivity on CREBBP over the other bromodomains. This work showed that analysis of Electrostatic Potential Surfaces ESP of substituted unsaturated chains interacting with a positively charged guanidinium correlates well with experimental binding affinities of a series of 5-isoxazolylbenzimidazole derivatives, subsequently leading to an improved qualitative understanding of the interaction in the CREBBP active site.

Computational chemistry techniques were also used in this thesis to uncover the Potential Energy Surface (PES) for the demethylation reaction catalysed by the proteins from the Jumonji C (JmjC) family. Although it is a significant epigenetic process, mechanistic details of demethylation in JMJD2A, a JmjC protein able to demethylase di- and tri-methylated histone lysines, had yet to be reported. In the initial part of this work, we have evaluated the first chemical step of histone demethylation by JMJD2A, the O_2 binding, happening preferentially in the quintet PES with a bent end-on configuration. We demonstrated the dramatic effect of the protein environment, dispersion and free energy corrections upon dioxygen binding using quantum mechanics/molecular mechanics (QM/MM) calculations. This methodology was further used for determining the full reaction profile of the demethylation reaction: oxygen activation, H-abstraction and -OH rebound. It was showed that the nucleophilic attack of the distal oxygen, with the formation of the bicyclic complex, is the harder of the two elementary steps involved in oxygen activation. Furthermore, in contrast to previous QM/MM studies on the DNA demethylase AlkB, the -OH rebound was found to be the rate-determining step for the demethylation reaction in JMJD2A. However, this data should be cautiously interpreted since TS barriers showed sensitivity to several modelling assumptions, such as the choice of the charge point model for the calculations at the low level of theory.

The work described in this thesis demonstrated the importance of cation-π interactions for the selectivity of small molecules on CREBBP, as well as the role of of the protein environment for computing energetics and structural determination of intermediates and TS for the demethylation reaction by JmjC proteins. Future opportunities in this work include the use of this novel ESP-based model for prediction of cation-π interactions in a wide range of proteins and computation of PES for different substrates in the active site of JmjC proteins. The computational insights here presented are powerful tools for rational design of inhibitors targeting epigenetic regulators and exploration of protein mutants.