Insights into Epigenetic Modulation from Computational Studies

Kiran Kumar

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Abstract

Histone monomers are proteins found in the nucleosome, which contain tail domains that can be post-translationally modified as part of epigenetic modulation. The accumulation of these chemical groups on histone amino acid residues such as lysine, alter the flexibility of the nucleosome, influencing how transcriptional machinery binds to DNA. As a result, epigenetic regulation has large impacts on gene activation and repression, which can be exploited by cancerous cells to proliferate in hypoxic and dense tumours conditions. Subsequently, research to elucidate the chemical mechanisms driving the enzymes responsible for reading, writing, and erasing these post-translational modifications (PTMs) gained increasing attention in recent decades. In contribution to this field, the work presented here discusses the use of computational chemistry techniques, such as quantum mechanics (QM) and molecular dynamics (MD) simulations, to shed light on the underlying behaviour of the PTM substrates and epigenetic receptors. **Chapter 1** will introduce the relevant protein families, computational theory, and previous pioneering work in the field. In **Chapter 2**, an in-depth analysis using several *in silico* approaches to optimise a 5-isoxazolylbenzimidazole inhibitor **(Fig. 1, blue**) and bromodomain cation–π interaction is discussed, followed by details of a large-scale study exploring all the available crystallographic PDB data containing an analogous interaction. **Chapter 3** characterizes the enthalpic landscape of the rate-determining step for demethylation of epigenetic eraser Fe(II)-αKG substrates, such as DNA nucleobase 1-methyladenine, using a combination of *ab initio* and DFT calculations. **Chapters 4, 5, and 6** discuss the results of work performed in collaboration with experimental groups. MD simulations are employed in **Chapter 4** to understand the effects on recognition after altering trimethyllysine, the natural substrate of several epigenetic readers, by shortening and lengthening its side chain or inverting the backbone stereochemistry **(Fig. 1, green**). In **Chapter 5**, the wild type substrate is kept constant while the aromatic cage of the reader proteins is mutated and examined using MD simulations **(Fig. 1, orange**). A different application of modelling modified lysines is applied in **Chapter 6**, which discusses the results of a collaboration to understand the role of carbamyl-lysine in a novel degradation mechanism of carbapenem, an antibiotic of the last resort.

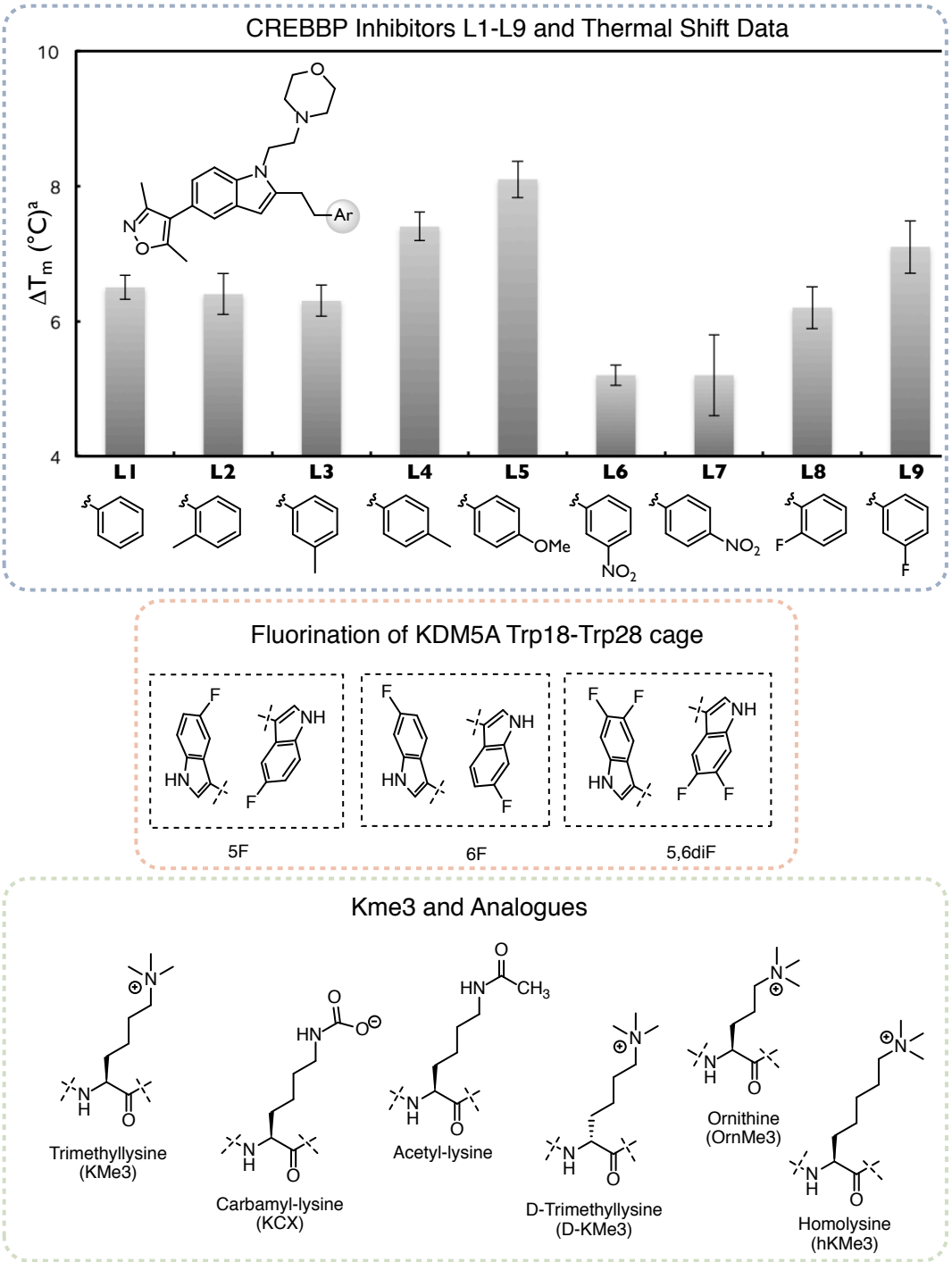


Figure 1: Examples of small-molecule CREBBP inhibitors **L1-L9** (blue), reader aromatic cage mutations via fluorination (orange), and KMe3 analogues (green) simulated in the investigations presented here.

Acknowledgements

First and foremost, thank you to my DPhil supervisor Prof. Robert Paton. I could not have achieved this goal without your guidance, vision, and commitment. You have always supported and encouraged me in all my interests and I am grateful to have been mentored by such an excellent supervisor. Thank you to my second supervisor Prof. Christopher Schofield for unhesitatingly stepping in to facilitate a critical time in my DPhil, and the opportunity to contribute to some exciting projects.

Dr. Akane Kawamura and Dr. Jasmin Mecinović, I have greatly enjoyed working with you and Roman Belle on a number of collaborations in the past few years. Dr. Fernanda Duarte, we could always rely on you to share your seemingly endless knowledge with us. Prof. (now President!) Jacquelyn Fetrow, thank you for believing in me those many years ago.

I am thankful to have worked with members of the Paton Group, past and present, who have made this the most enjoyable, challenging, and rewarding experience of my life. Wilian Cortopassi, I could always count on you for scientific discussions about biochemistry, advice, and of course laughs. Jacqueline Tan and Ruchuta Ardkhean, thanks for being amazing colleagues and loyal friends. Shinny Woo thanks for making me feel so welcome my first year at Oxford. Kelvin Jackson and Karrar Al-Ameed, I already miss the priceless banter.

I am blessed with an incredibly supportive family. Mom and Dad, thank you for all your love and all the sacrifices you made so I could always follow my dreams. Your unwavering support and encouragement allowed me to believe I was capable of anything. I am also grateful to my Nan and Denny Uncle for giving me a home away from home.

List of Publications

Publications describing work in this Thesis:

1. "Computational insights on bond dissociation enthalpies for C-H abstraction by Fe(II)/α-ketoglutarate enzymes" **Kumar, K.**; Tan, J. S. J.; Paton, R. S. **2019**, (*In preparation*).

2. "Mechanism and Molecular Dynamics of Non-Heme Demethylase Enzymes" Tan, J. S. J.; **Kumar, K.**; Nguyen, Q. N. N.; Rodeja-García, Y.; Paton, R. S. **2019**, (*In preparation*).

3. "Reading and erasing of the phosphonium analogue of trimethyllysine by epigenetic proteins" Kamps, J. J. A. G.; Belle, R.; Poater, J.; **Kumar, K.**; Pieters, B. J. G. E.; Claridge, T. D. W.; Paton, R. S.; Bickelhaupt, F. M.; Kawamura, A.; Schofield, C.; J. Mecinović, J. **2019**, (*In preparation*).

4. "Molecular recognition of trimethyllysine and trimethylthialysine by epigenetic reader proteins: a comparative study" Hintzen, J. C. J. and Al Temimi, A. H. K. and Poater, J. and **Kumar, K.**; Pieters, B. J. G. E.; Paton, R. S.; Bickelhaupt, F. M.; Mecinović, J. **2019**, (*Under Review*).

5. "The recognition of trimethyllysine by the KDM5A PHD3 finger is insensitive to fluorination of the aromatic cage" Pieters, B. J. G. E.; Wuts, M. H. M.; Poater, J.; **Kumar, K.**; White, P. B.; Kamps, J. J. A. G.; Sherman, W.; Pruijn, G. J. M.; Paton, R. S.; Beuming, T.; Bickelhaupt, F. M.; Mecinović, J. **2019**, (*Under Review*).

6. "Non-Hydrolytic β-Lactam Antibiotic Fragmentation by l,d-Transpeptidases and Serine β-Lactamase Cysteine Variants" Lohans, C. T.; Chan, H. H. T.; Malla, T.; **Kumar, K.**; Kamps, J. J. A. G.; van Groesen, E.; de Munnik, M.; Tooke, C. L.; Spencer, J.; Paton, R. S.; Brem, J.; Schofield, C. J. *Angew. Chem. Int. Ed.* **2019**, *58(7)*, 1990-1994.

7. "Recognition of shorter and longer trimethyllysine analogues by epigenetic reader proteins" Abbas H. K. Al Temimi; Belle, R.; **Kumar, K.**; Poater, J.; Betlem, P.; Pieters, B. J. G. E.; Paton R. S., Bickelhaupt, F. M.; Mecinović, J. *Chem. Commun.* **2018**, *54(19)*, 2409-2412.

8."Cation–π interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine" **Kumar, K.** and Woo, S. M.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. *Chem. Sci.* **2018**, *9(10)*, 2655-2665 (selected for Journal front cover).

9. "A New Mechanism for β-Lactamases: Class D enzymes degrade 1β-methyl carbapenems through lactone formation" Lohans, C. T.; van Groesen, E.; **Kumar, K.**; Tooke, C. L.; Spencer, J.; Paton, R. S.; Brem, J.; Schofield, C. J. *Angew. Chem. Int. Ed.* **2018**, *57(5)*, 1282-1285.

10. "Investigating D-lysine stereochemistry for epigenetic methylation, demethylation and recognition" Belle, R.; Abbas H. K. Al Temimi; **Kumar, K.**; Bas J. G. E. Pieters; Tumber. A.; Johansson, C.; Paton R. S., Brown T.; Schofield C. J.; Hopkinson, R. J.; Kawamura, A.; Mecinović, J. *Chem. Commun.* **2017**, *53*, 13264-13267.

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

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

Other publications

13. "An approach to functionally relevant clustering of the protein universe: active site profile-based clustering of protein structures and sequences" Knutson, S. T.; Westwood, B. M.; Leuthaeuser, J. B.; Turner, B. E.; Nguyendac, D.; Shea, G.; **Kumar, K.**, K.; Hayden, J. D.; Harper, A. F.; Brown; S. D.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C.; Fetrow, J. S. *Protein Sci.* **2017**, *26(4)*, 677-699.

Contents

Glossary of Abbreviations

Ac-CoA . . . . Acetyl coenzyme A

AdoMet . . . . S-Adenoysyl-L-methionine

AMBER . . . Assisted Model Building with Energy Refinement

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

BET . . . . . Bromodomain and Extraterminal Domain

CPCM . . . . Conductor-like Polarisable Continuum Model

CCSD . . . . . Coupled Cluster Singles Doubles Method

CCSD(T) . . . Coupled Cluster with Perturbative Estimate for Connected Triple

CREBBP . . . cAMP-response Element Binding Protein Binding Protein

DHQ . . . . . Dihydroquinoxalinone

DFT . . . . . Density Functional Theory

$\Delta \boldsymbol{T_m}$ . . . . . Melting Temperature Shift

DOT1L . . . . DOT1-Like Histone H3K79

ESP . . . . . Electrostatic Potential Surface

FAD . . . . . Flavin Adenine Dinucleotide

FDA . . . . . Food and Drug Administration

G4 . . . . . Gaussian-4 Theory

GAFF . . . . . General AMBER Force Field

GNAT . . . . . GCN5-related N-terminal acetyltransferases

H3 . . . . . Histone 3

H4 . . . . . Histone 4

HAT . . . . . Histone Lysine Acetyltransferase

HDAC . . . . . Histone Deacetylase

HF . . . . . Hartree-Fock

HIF . . . . . Hypoxia Inducible Factor

HKMT . . . . Lysine Histone Methyltransferase

IRC . . . . . Intrinsic Reaction Coordinate

ITC . . . . . Isothermal Titration Calorimetry

JMJD2A . . . JmjC Histone Demethylase

JmjC . . . . . Jumonji C Protein Family

H3 . . . . . Histone 3

K . . . . . Kelvin

kcal mol^{-1} . . . Kilocalorie per Mole

KIE . . . . . Kinetic Isotope Effect

kJ mol^{-1} . . . . Kilojoule per Mole

LSD . . . . . Lysine-specific Demethylase

MYST . . . . . Moz, Ybf2/Sas3, Sas2, Tip60

M06 . . . . . Minnesota '06 Functional Family

MM . . . . . Molecular Mechanics

MM-PBSA . . MM-Poisson Boltzman Surface Area

MM-GBSA . . MM-generalized Born Surface Area

MP2 . . . . . Møller-Plesset Second Order Perturbation Theory

NAD$^+$ . . . . . Nicotinamide Adenine Dinucleotide

NAMD . . . . Nanoscale Molecular Dynamics Program

NCI Plots . . Non-covalent Interaction Plots

ONIOM . . . . Our own N-layered Integrated Molecular Orbital and MM

PCM . . . . . Polarisable Continuum Model

PDB . . . . . Protein Data Bank

PES . . . . . Potential Energy Surface

PHD . . . . . Plant Homeodomain

PTM . . . . . Post-translational Modification

QM . . . . . Quantum Mechanics

QM/MM . . . Quantum Mechanics/Molecular Mechanics

RESP . . . . . Restrained Electrostatic Potential

RDS . . . . . Rate-determining Step

RMSD . . . . . Root Mean Squared Deviation

RMSE . . . . . Root Mean Squared Error

RMSF . . . . . Root Mean Square Fluctuation

SET . . . . . Su(var)3-9 Enhancer of Zeste Trithorax

TAF . . . . . TBP-associated Factor

TET . . . . . Ten-eleven Translocation

TIP3P . . . . . Transferable Intermolecular Potential with 3 Points

TS . . . . . Transition State (first order saddle point)

TPSS . . . . . The Exchange Functional of Tao, Perdew, Staroverov, and Scuseria

VMD . . . . . Visual Molecular Dynamics

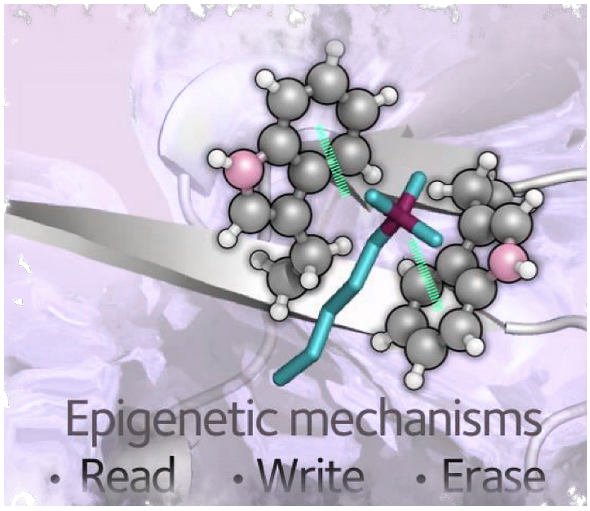
ω**B97XD** . . . Dispersion Corrected Density Functional of Head-Gordon

Epigenetics is a useful word if you don't know what's going on—if you do, you use something else.

— Adrian Bird, 2008

1 Introduction

Contents



Cortopassi W.A., **Kumar K.**, et al. Mechanisms of histone-modifying and reading enzymes: a computational perspective on the role of the protein environment *J. Mol. Graph. Mod.* **2016**, *67*, 69–84.

1.1 Epigenetics and Cancer

The term 'epigenetics' was first used in 1942 by embryologist Conrad Waddington who described it as a "whole complex of developmental processes" found between "genotype and phenotype."[1,2] Waddington's initial model of the 'epigenetic landscape' illustrated how the fate of the cell was determined by the presence or absence of certain genes.[3] In this metaphor, cellular development is represented as a ball rolling down a valley of diverging paths, the geography of which is dependent on its genetic composition.

Waddington's definition has since undergone its own evolution.[2] Modulation of gene expression is now understood to be regulated by chemical modifications to chromatin - DNA packaging complexes. The narrative of epigenetics today now includes discussion of *'reader'*, '*writer*', and '*eraser*' enzymes responsible for recognising, adding, and removing these markers, respectively.

The first post-translational modification (PTM) epigenetic marker to be discovered was the addition of methyl group(s) to DNA nucleobases. Following this, more than 85,000 PTMs of DNA and histone proteins have been identified and involve the addition of over 25 different chemical groups including: hydroxyl, acetyl, and phosphoryl groups.[4] Of these, regulation of methylation and acetylation of histone lysine residues located on unstructured N-terminal tail domains are the most common.[5–7]

Epigenetic processes are critical for cell differentiation and maintaining homeostasis. For example, under hypoxic conditions the cell will utilise complex epigenetic mechanisms to upregulate Hypoxia Inducible Factor (HIF) target genes to survive oxygen deprivation.[8] PTMs are able to modify the structure of chromatin, disrupting its rigid structure and allowing access for transcriptional machinery to operate. However, for the past several decades epigenetics has received increasing attention for its relevance in a wide range of diseases such as cancer, neurological disorders, and autoimmune diseases.[9] **Table 1.1** lists some of the various types of cancers that have been linked to specific histone PTMs.[6]

Table 1.1: Cancer types and implicated histone modification.[6]

Cancer type	Modification	Ref.
Lung adenocarcinoma	Global acetyl.	[10]
Non-small cell lung	Global H3 deacetyl.	[11]
Gastric	Global H3K9 tri-, dimethyl., decr. H3 acetyl.	[12, 13]
Prostate	H3K9, H3K18, H4K12 acetyl; H4K3,H3K4 dimethyl.	[14, 15]
Colorectal	Global H3K9 deacetyl.	[16]
Pancreatic	C/EPBα H3 acetyl.	[17]

The relevance of epigenetics in pathologies, such as cancer, arise due to the ability of cancerous cells to hijack these processes and over express certain genes optimal for tumouriogenesis. An example is overexpression of HIF-1α, a master regulator of genes crucial for adaptation to hypoxic conditions. An abundance of HIF-1α can facilitate proliferation of cancer cells found in dense tumours with low oxygen environments, and has even been directly correlated to diagnostic severity of numerous types of cancers listed in **Table 1.2**.[18]

Table 1.2: Tumour types linked to HIF-1α overexpression and affected outcome.[18]

Tumour type	Affected outcome	Ref.
Breast cancer, lymph-node negative	Mortality	[19]
Cervical cancer, early-stage	Mortality	[20]
Oligodendroglioma	Mortality	[21]
Oropharyngeal squamous cell carcinoma	Treatment failure, mortality	[22]
Esophageal cancer, early-stage, BCL2-positive	Treatment failure	[23]
Ovarian cancer, p53 mutant	Mortality	[24]

As a result, small-molecule epigenetic inhibitors have gained attention as promising anti-cancer therapeutics, with many going to clinical trials and four drugs already approved by the FDA.[25] Specifically, cutaneous T-cell lymphoma has been treated by romidepsin and vorinostat that target epigenetic erasers (**Fig. 1.1**, green).[26–28] Cytosine analogues decibitabine and 5-azacytidine block writers of DNA methylation and are used against myelodysplastic syndromes (**Fig. 1.1**, green).[29,30] Reader inhibitors, such as OTX015 which is used as a therapeutic for haematological malignancy, are in early phases of clinical trials (**Fig. 1.1**, purple).[31]

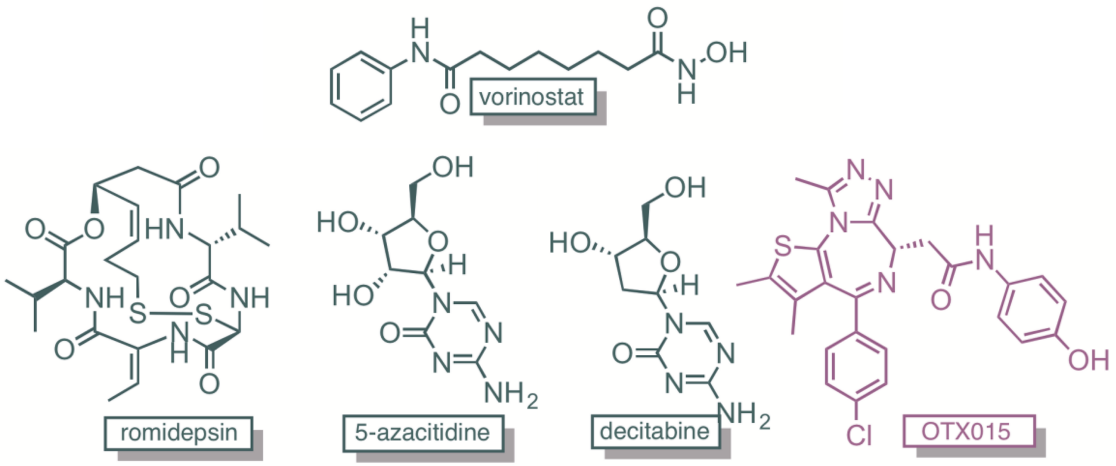


Figure 1.1: Vorinostat and romidepsin are FDA approved drugs that target epigenetic erasers while 5-azacytidine and decitabine are approved to inactivate writer proteins. Bromodomain reader inhibitor OTX015 is in early stages of clinical trials.

An overview of the classification and proposed mechanisms of readers, writers, and erasers will be discussed in subsequent sections.

1.2 Overview of Histone Lysine Modifiers

1.2.1 HATs and HKMTs

A major function of epigenetic writers is to catalyse addition of chemical groups such as methyls and acetyls to histone lysine side chains.

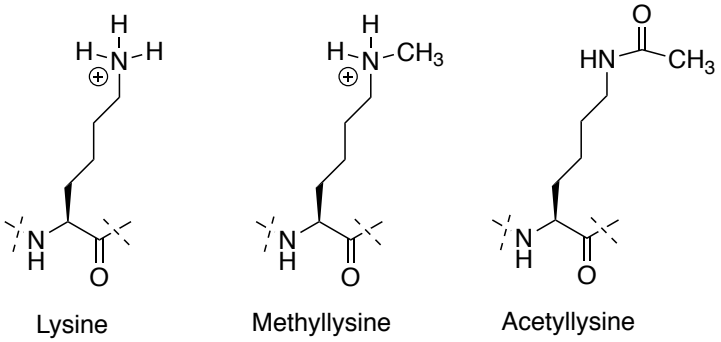


Figure 1.2: HKMT and HAT proteins post-translationally modify lysine to methyllysine and acetyllysine, respectively.

Histone lysine acetyltransferases (HATs) mechanistically utilise Acetyl coenzyme A (Ac-CoA) as a cofactor. HATs are further divided by sequence and structural similarities into three families: GCN5-related N-terminal acetyltransferases (GNAT), Moz, Ybf2/Sas3, Sas2, Tip60 (MYST), and p300/CREB-binding protein (CREBBP) family. Structural characteristics of these are outlined in **Table 1.3**.

Table 1.3: Properties of Histone lysine acetyltransferases (HAT) families.

Family	Description
GNAT	~160 residues contains HAT domain and bromodomain (reader)
MYST	~250 residues contains HAT, zinc finger, and N-terminal chromodomain (reader)
CREBBP	~500 residues contains HAT, zinc finger, and N-terminal bromodomain (reader)

Histone lysine methyltransferases (HKMTs) require S-Adenosyl-L-methionine (AdoMet) as a cofactor when catalysing methylation. These are grouped into the Su(var)3-9 Enhancer of Zeste Trithorax (SET) containing HKMTs and DOT1-Like Histone H3K79 (DOT1L) family that does not have a SET domain.

HATs and HKMTs have shown to be valuable drug targets for cancer treatment with several compounds undergoing clinical trials.[32,33]

1.2.2 HDACs and Demethylases

Epigenetic erasers that remove acetyl and methyl markers from histone lysine proteins are known as histone lysine deacetylases (HDACs) and histone lysine demethylases, respectively.

Two major families of HDACs are classical zinc dependent HDACs and Sirtuin, which require NAD^+ as a cofactor. These families are further divided into classes[34,35] outlined in **Table 1.4**.

Table 1.4: Histone lysine deacetylases (HDAC) classes and proteins.

Class	Proteins
Class I:	HDAC1, HDAC2, HDAC3, and HDAC8
Class II:	HDAC4, HDAC5, HDAC6, HDAC7, HDAC9, HDAC10
Class III:	SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6, SIRT7, Sir2
Class IV:	HDAC11

Lysine demethylases are divided into two major classes: amino oxidases Lysine-specific demethylase (LSD) and the Jumonji (JMJC) families[36] and are generally characterized by properties listed in **Table 1.5**:

Table 1.5: Lysine demethylase classes and descriptions.

Class	Description
LSD	LSD1 (KDM1A) and LSD2 (KDM1B) proteins Flavin adenine dinucleotide (FAD)-dependent demethylates mono-, dimethyllysine
JMJC	non-heme iron dependent active site demethylates all methylation states

An overview of the mechanistic differences of demethyation for JMJC and LSD proteins is shown in **Fig. 1.3**.[37–41]

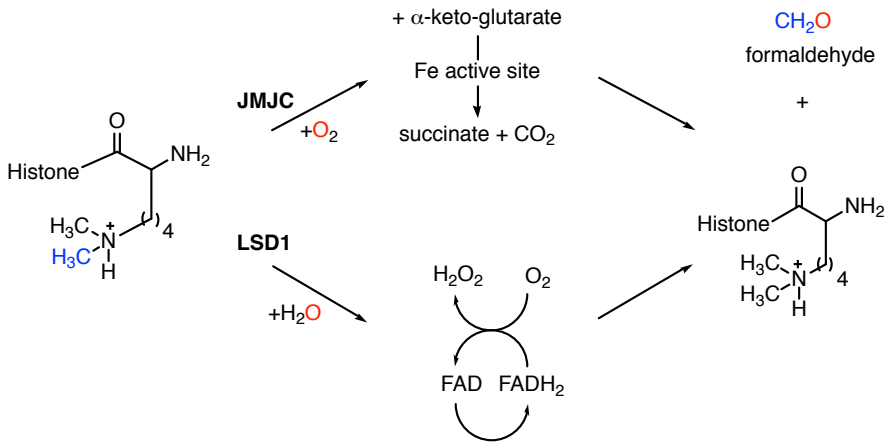


Figure 1.3: Histone lysine demethylation mechanism overview for JMJC and LSD1 protein families.

1.2.3 Tudor, PHD, Chromodomain, and Bromodomains

Recognition of epigenetic markers is crucial for regulation. Four primary classes of readers exist including the Tudor domains, The Plant Homeodomain (PHD) Fingers, Chromodomains, and Bromodomains that have varying secondary and tertiary structures as shown in **Fig. 1.4**.

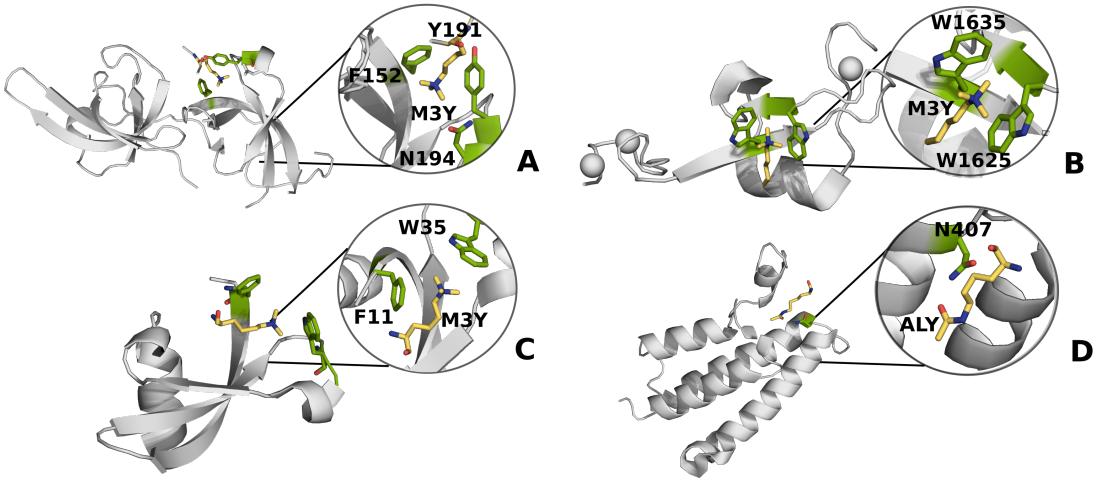


Figure 1.4: Structures of four major epigenetic read proteins: A) tudor domains (PDB: 3GL6), PHD fingers (PDB: 3DB3), chromodomains (PDB: 3I90), and D) bromodomains (PDB: 1E6I) with key residues highlighted that are important for binding trimethylated and acetylated lysine histone residues.

All of the readers are responsible for recognising N-states of methyllysine, apart from bromodomains that have acetyllysine as a substrate. Bromodomains are also the only class of readers not characterized by an aromatic cage active site. Other structural features are outlined in **Table 1.6**.

Table 1.6: Lysine demethylase classes and descriptions.

Reader	**Description**
Tudor Domain	Recognition of methyl groups ~50 residues Anti-parallel and bent conformation of β-sheets Phe-Tyr-Asn active site
PHD Fingers	Recognition of methyl groups α-helical globular fold and two-stranded β-sheet Trp-Trp active site
Chromodomains	Recognition of methyl groups C-terminal α-helix and three-stranded β-sheet Phe-Trp active site
Bromodomain	Recognition of acetyl groups Four α-helices and two loop regions Asn key residue

A wealth of information concerning the key players of epigenetics and their role in cancer has been established since Waddington's time; some of which has been discovered through innovative computational studies. The next section will discuss these previous works and their valuable contributions to both *in silico* technique developments and cancer research.

1.3 Computational Advancements

Previous computational works modelling epigenetic processes have provided important insight and elucidation of the chemical events driving epigenetic modulation. *In silico* techniques employed to investigate the behaviour of epigenetic proteins and their corresponding reactions include molecular dynamics (MD) simulations, quantum mechanical (QM) calculations, and hybrid methods such as quantum mechanics/molecular mechanics (QM/MM). Here, I will discuss these pioneering studies that were successful in advancing the design of selective and potent cancer therapeutics performed on specific reader, writer, and eraser protein families, as well as the challenges currently facing the field.

1.3.1 Understanding Histone Lysine Acetyltransferases

GNAT Proteins

Determining the reaction catalysed by the lysine acetyltransferase GNAT family proved challenging; while a sequential mechanism with a formation of a ternary complex was proposed from the experimental data,[42,43] the structure of the complex and catalytic mechanism were not discernable from experiment. By employing MD simulations and QM/MM techniques, Jiang et al. were able to investigate this mechanism further.[44] Snapshots obtained at 20 ns intervals from the simulation trajectory were then analysed using QM/MM to model the catalytic process. They identified a key glutamate residue acting as a general base for the deprotonation of a histone lysine with a formation of a "proton wire" mediated by a water molecule. Addition of the lysine to the carbonyl of the Ac-CoA was found to be the rate-determining step (RDS) and MD simulations revealed the importance of two GCN5 loops external from the active site for anchoring the substrate via formation of a cleft.

Jędrzejewski used homology modelling and classical MD simulations to analyse acetylation by Patt1, a novel human histone acetyltransferase that has shown to promote apoptosis in human hepatocellular carcinoma cell lines.[45] A similar finding, whereby a glutamate residue acts as a general base to activate lysine, was also presented. Furthermore, the aromatic groups of phenylalanine residues and Ac-CoA were found to form important stabilising π–π interactions.

MD simulations have also been used to study GNT5 proteins, specifically GNC5 with the non-histone HIV-1 integrase and the HIV-1 trans-activator.[46,47] These studies paralleled experimental data showing that the complexes could exist in solution following 200 ns of simulation and that acetylation by GNT5 proteins is also possible in non-histone environments.[48,49]

MYST Proteins

The enzymatic kinetics for the MYST family have been found to be similar to GNT5,[50–52] although less is known about the mechanistic details of acetylation by MYST proteins. To date, there have been no theoretical studies on this topic, although two mechanisms have been suggested. 1) A ping-pong mechanism,[53] where the acetyl group from the cofactor is first transferred to a cysteine via an acetyl-enzyme intermediate, and then to the histone lysine; 2) a sequential mechanism, where a ternary complex is formed.[54] A suggestion for future work in this field is to employ techniques such as QM/MM to compare the energetic profiles of a reaction involving glutamate acting as a general base against the formation of an acetyl-enzyme intermediate, which could provide further insight.

p300-CBP Family

Members of the p300-CBP family differ from the GNAT or MYST proteins in that they are suggested to undergo a unique "hit and run" mechanism without formation of a ternary complex.[55] In this case, the enzyme active site already contains acetylated Ac-CoA. The histone weakly associates to the surface, and immediately following transfer of the acetyl group to lysine, the histone is freed.[56]

100 ns MD simulation and QM/MM calculations performed by Zhang et al.[55] characterized binding as favoured by hydrogen bonds and hydrophobic interactions with the p300 loops. For the first step, a long-range proton transfer scheme involving specific tyrosine, aspartate, and water molecules was shown to have a low energy barrier. Finally, a different tyrosine residue was observed to be needed for re-protonation of the CoA leaving group to acetylate the histone lysine.

1.3.2 SET-domain Containing Methyltransferase Specificity

Recent interest has been shown in performing extensive computational investigations of histone lysine methyltransferases, with particular attention to understanding structural properties needed for product specificity.[57–59] Experimental crystal structures indicate the presence of a water molecule in the substrate-binding pocket that must be released preceding methylation.[60] Following 200 ns MD simulations performed by Hu and Yang of lysine methyltransferase 9 (SET9) in complex with the tumour suppressor p53, it was suggested that experimentally observed dimethylation by SET9 is possible when water molecules (responsible for the lysine deprotonation step) can leave the active site via a channel shown in **Fig. 1.5**.[61–63] This channel is characterized by a glycine, alanine, and two tyrosine residues.

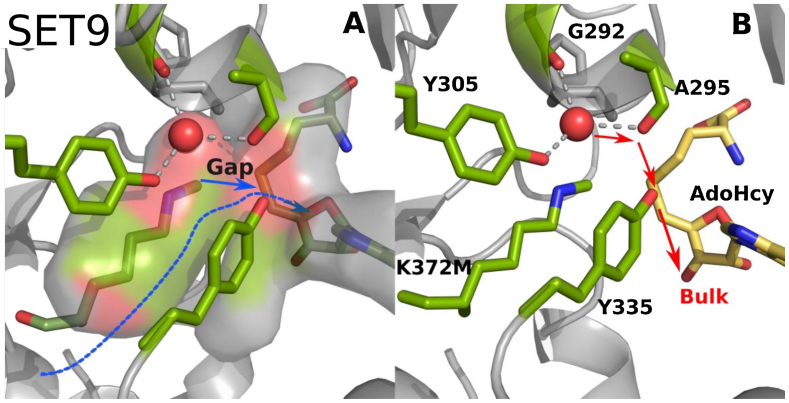


Figure 1.5: Two proposed water channels A) and B) found in lysine methylstransferase 9 (SET9). Computational work by Bai et al. suggests B) is the preferred channel.[61]

Qiao et al. used shorter 2 ns simulations to examine the methyltransferase Nuclear receptor binding SET Domain protein 1 (NSD1) bound to histone H3.[64] A novel regulatory loop connecting SET and a post-SET domain was found to control access of the histone lysine to the SET domain of NSD1, and therefore the degree of lysine methylation.

When used together, MM and QM/MM approaches have provided valuable insight into SET protein substrate selectivity.[65–67] Zhang performed QM/MM Hartree-Fock (HF)//Amber calculations and found that conformational changes of the cofactor binding site occur as a result of the first methyl transfer in SET7/9 enzymes and increases the energetic barrier for dimethylation.[68] This transfer was

also characterized as a concerted S_N2 reaction following QM/MM calculations,[65] where the barrier for the charge-separated TS structure was lowered by favourable electrostatic contributions from the protein environment from 30.9 to 22.5 kcal mol^{-1}.[67] The theoretical value was in good agreement with the experimental free energy of 20.9 kcal mol^{-1}.

Similarly to SET7/9, Chu et al. used QM/MM and MD and also found that SET8 showed a higher energetic cost for catalysing dimethylation over monomethylation due to the need for substrate reorientation after the initial deprotonation step.[69] They also applied these techniques to understand the effects of amino acid mutations on the energetics of a reaction. From these calculations they proposed that a Y334F variant of SET8 was more favourable for dimethylation, although a high activation barrier for trimethylation remained. This followed an earlier computational study by Xu *et al.*[70] that examined the inverse occurrence of trimethylase protein Defective In Methylation-5 (DIM-5) becoming a mono- or dimethylase following a F281Y mutation due to a destabilisation of the methylation transition state by 8 kcal mol^{-1} when comparing the third and first methyl transfer.

Chu et al.[71] also described reduced catalytic activity following a mutation after performing QM/MM on methyltransferase PR domain zinc finger protein 9 (PRDM9). A similar finding of an increased energetic barrier was calculated as a results of a Y276F mutation, corroborating experiments by Wu and co-workers.[72]

1.3.3 Protein Dynamics in Histone Deacetylases

Classical HDACs Family

MD simulations have been a popular choice for studying histone deacetylases as they contain dynamic loop regions and MD-based techniques are capable of quantifying the behaviour and flexibility of a large protein over time.

Specifically, the structural features concerning regulation of water and product release have been investigated. Following 20 ns of MD simulation, Haider *et al.* determined that a key HDAC8 arginine residue acts as a gate-keeper, via hydrogen bond formation with two conserved glycine backbone carbonyl groups, that can block a HDAC8 channel.[73] Kalyaanamoorthy and Chen also investigated the gate-keeping process further using MD simulations and mutagenesis experiments, which

yielded further identification of additional pathways for water and product release in HDAC1 and HDAC2 enzymes.[74,75] Thangapandian et al. performed 5 ns simulations of the apoforms of the native and mutated HDAC8, HDAC10, HDAC11. While subtle differences were found in these native and mutated active sites, non-conserved residues in the tunnel region were suggested to affect arrangement and dynamics.[76]

Arrar et al. performed 100 ns MD simulations using the AMBER ff99SB force field on epigenetic regulator inositol tetraphosphate (IP4) and HDAC3.[77] They found that conformational changes induced by interactions with IP4 and the deacetylase-activating domain (DAD) were necessary for creating an active conformation of HDAC3.

Kunze et al. performed long-time MD also using the AMBER ff99SB force field and crystallographic techniques to investigate the substrate entrance site of HDAC8.[78] They observed interactions between HDAC8 residues in two loops, L1 (residues 31–35) and L2 (83–108), controlled enzymatic activity by affecting substrate binding and positioning.

Several QM/MM studies were performed to further understand the deacetylation mechanism by HDAC proteins. One question of interest was elucidating which of the two active site histidine-aspartate dyads either acts as a base or is involved in proton transfer to the deacetylated histone. Wu et al. presented an alternative proton-shuttle mechanism for HDAC8, where a single neutral histidine residue accepts a proton from the active site water molecule and then transfers it to the amide nitrogen, facilitating the release of the acetyl group, shown in **Fig. 1.6**.[79] However, QM/MM studies by Chen *et al.* included an extended active site with aspartate residues, similar to previous QM studies by Vanommeslaeghe *et al.*[80,81] Findings from this study supported the previous model, with one histidine/aspartate dyad functioning as a general base for the reaction and one positively charged histidine protonating the epsilon nitrogen atom enabling collapse of the tetrahedral intermediate.[82] Furthermore, QM/MM calculations by Gleeson also supported this mechanism.[83]

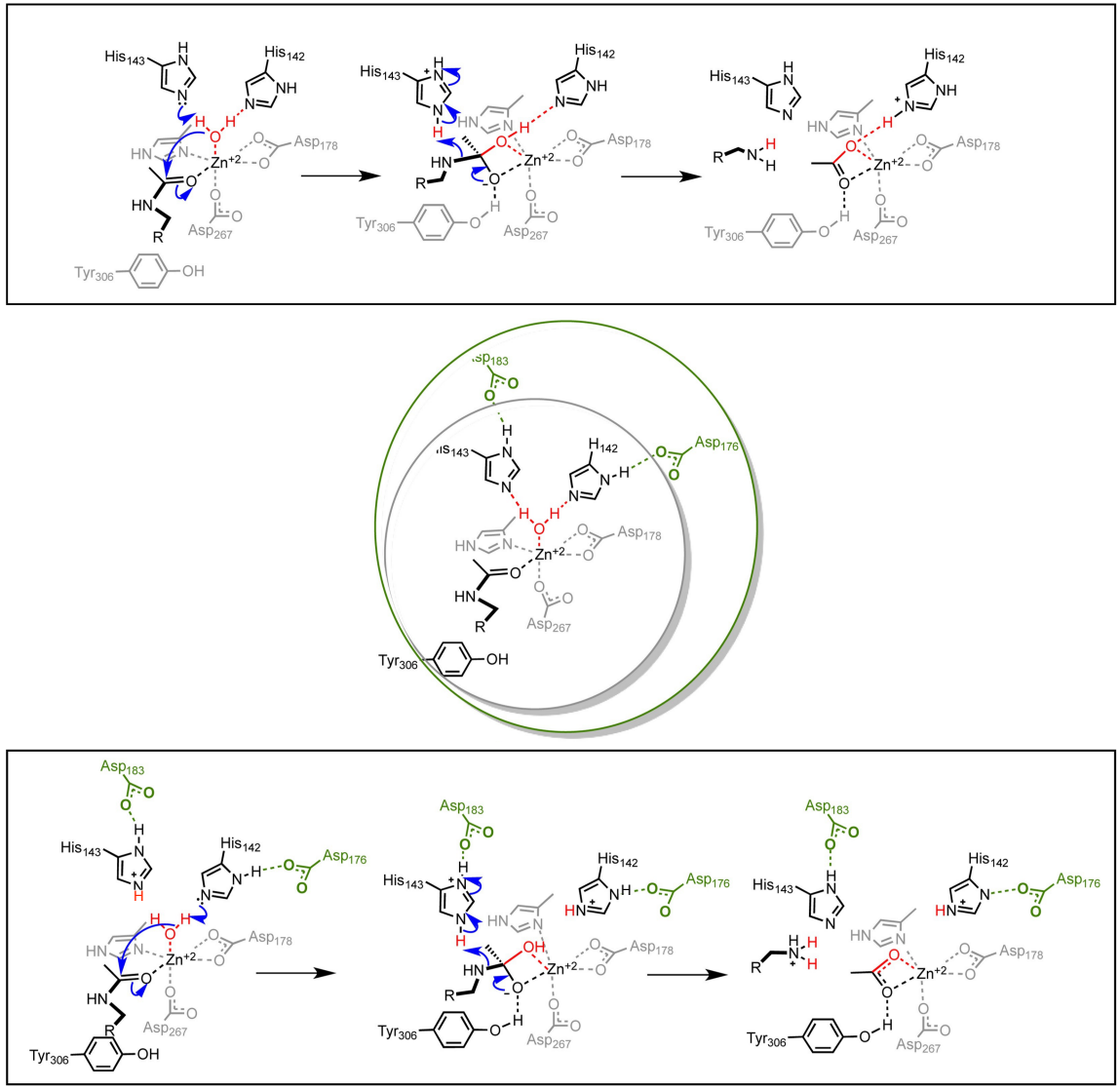


Figure 1.6: Scheme demonstrating the importance of inclusion in the active site in calculations: (Top) deacetylation in HDAC8 via the proposed proton shuttle mechanism is predicted using the smaller active site.[79] (Bottom) an extended active site model including two aspartate residues implicated the involvement of both histidines in the mechanism[80,81]

Sirtuin HDACs

MD simulations have been valuable for identifying broad conformational changes of Sirtuin HDAC proteins. Hsu et al. was able to explain allosteric stimulation of acetylation by Sir2 when bound to Sir4 from analysing 75 ns MD with the AMBER ff03 force field and SPC solvation.[84] They revealed that this phenomenon is caused by favourable interactions between a long loop of Sir4 and the catalytic domain of Sir2.

Sakkiah et al. performed relatively short simulations of 5 ns with the GROMOS96 force field and simple point charge (SPC) solvation.[85] From this they showed that the binding of the substrate and NAD^+ to the apoform of SIRT2, a Sir2 homolog, directs a phenylalanine towards an active site histidine, ultimately favouring the deacetylation. Guan et al. carried out 10 ns MD with the AMBER ff99SB force field to conclude that the nicotinamide product of the acetylation inhibits the reaction by competing with the cofactor NAD^+.[86]

Liang et al. were able to elucidate intrinsic details of deacetylation by the Sirtuin family.[87] Specifically, they suggested a S_N2 mechanism for the first step of the deacetylation by Sir2, after performing QM/MM calculations at the B3LYP//AMBER level of theory. Then, after performing 15 ns MD with AMBER ff99 and SPC solvation, they observed a histidine and a glutamine residue acting as both a base for the deprotonation step of the N-ribose and positioning of the cofactor NAD^+.

The full reaction profile of the deacetylation reaction by Sir2 was reported by Hu et al. at the B3LYP//AMBER level of theory.[88] They found that the first step, the nicotinamide cleavage, takes place through a concerted and highly dissociative displacement mechanism. The second step, highly dependent on a water molecule, is rate-determining and involves the formation of a bicyclic intermediate followed by its collapse. Following conformational analysis of MD-QM/MM simulations, Shi et al. reported that the cofactor-binding loop is very dynamic and crucial for allowing water molecules to come close to the intermediates.[89]

1.3.4 Revealing Insights into Demethylation

JMJC Family

Proteins in the JMJC family are characterized with having high sequence similarity but differences in residues surrounding the active site region that significantly influence substrate binding and recognition.[90,91]

Ulucan et al. studied JMJD2A complexed with varying lysine protonation states by carrying out 18 ns MD simulations with the AMBER ff03 force field and explicit TIP3P solvation.[91] A specific hydrogen bond between the side chain of the arginine residue and the backbone of the methylated lysine was found to be key for stabilising the binding of di- and trimethylated H3 lysine to JMJD2A.

A recent study performed by the Paton Group investigated O_2 binding to the active site of JMJD2A using 10 ns MD simulations with the AMBER ff99SB force field and QM/MM calculations with B3LYP//AMBER.[90] Inclusion of non-active site residues for computational calculations were found to be essential for an accurate energetic description of the binding process. For example, positively charged arginine and lysine residues had favourable electrostatic contributions for the first step of the demethylation reaction due to the charge transfer that occurs upon O_2 binding. Inclusion of the protein environment was found to be important in the calculations and resulted in a more favourable binding free energy by 6.7 kcal mol^{-1}.

1.3.5 LSD Family

A trend identified among eraser regulator proteins is the relationship between their flexible nature and how this facilitates their role in removing epigenetic PTMs - a pattern that has also been observed for the LSD family.

Baron et al. performed 50 ns MD of LSD1 bound to REST corepressor 1 (CoREST), a highly dynamic complex containing three domains: the amine oxidase (AO), the Tower, and the Swirm domains. They highlighted a conserved lysine that is important for diffusion of the O_2 into the active site.[92] Another study of this complex by Baron and Vellore utilised long-time microscale MD and found that the unbound LSD1/CoREST presents a gate at the histone interaction site formed by a triad of lysine residues, and complexation of LSD1/CoREST and histones affects the amino oxidase activity of LSD1.[93] A follow-up MD study by the same authors further identified that upon histone binding, global changes to the protein shape by an induced-fit mechanism dramatically affects the structural arrangements of the AO domain compared to the unbound LSD1/CoREST complex.[94]

QM/MM mechanistic studies have also been valuable for revealing insight into demethylation by LSD proteins. From QM/MM calculations, performed by Karasulu et al. at the B3LYP//CHARMM22 level of theory, three key residues and their

roles were identified: a lysine acting as a base; a tyrosine helping to position the reactants; and a tryptophan stabilising the water bridge that is important for the proton transfer step.[95]

Kong et al. utilised QM/MM and 30 ns MD simulations and also found that the tyrosine has a stabilising effect on the reaction, and mechanistically the process involves hydride transfer from the α-carbon of the lysine H3 to the flavin cofactor.[36] This mechanism has also been studied by Yu et al. with electrostatically embedded multiconfiguration molecular mechanics (EE-MCMM). This approximates the gradient in a typical QM/MM calculation by an analytic potential based on a constant valence bond coupling term, which facilitates MD umbrella sampling. They found that the pathway occurs via a concerted transfer of a hydrogen atom and electron, with an activation free energy of 10 kcal mol^{-1}. Although formally equivalent to hydride transfer, this description is preferred since the atomic charge on hydrogen remains positive throughout.[96]

1.3.6 Acetyllysine Readers

Computational mechanistic work on epigenetic reading regulators has been successful in using MM approaches to identify residues that are important for substrate binding and specificity.

Following 5–10 ns MD simulations with the CHARMM27 force field and TIP3P solvation, Pizzitutti et al. concluded that a proline residue on the bromodomain of Gcn5p was essential for its ability to recognise acetylated lysine on H4 by affecting the movement and structure of the ZA loop. They also performed complimentary experimental mutational studies and confirmed the importance of the proline residue.[97]

Caflisch and co-workers used microscale MD with the CHARMM27 force field to discover a "P-binding mode" where the acetylated lysine forms a hydrogen bond to the backbone carbonyl of a proline located in the vacant lipophilic tryptophan-proline-phenylalanine (WPF) shelf of the TBP-Associated Factor 1 (TAF1) bromodomain. During their simulations, they found the ligand reversibly interconverts with the well characterized "N-binding mode" in which the ligand is hydrogen bonded to a conserved asparagine residue. The "P-binding mode" was observed in 50% of MD snapshots while the "N-binding mode" was found slightly

less frequently in 40% of the snapshots.[98,99] Two experimental studies following Caflisch's computational investigations supported these findings by reporting a binding pose similar to the "P-binding mode" in other bromodomains.[100,101]

Steiner et al. carried out 1 µs MD with the CHARMM22 force field and found opening of the bromodomain active site was promoted via a displacement of the side chains belonging to the conserved residues of the acetyl lysine-binding site.[102] The frequency of the $-NH_2$ asparagine side chain group to point towards the bulk solvent was observed for significant amounts of simulation time for various receptors: TAF1L(2) (90% of the simulation time), TAF1(2) (81%), TAF1(1) (68%), Gcn5L2 (44%), ATAD2 (43%), BAZ2B (39%), and PB1(2) (38%).

In the instances for TAF1L (2) and TAF1 bromodomains, the conformation was stabilised by hydrogen bonds mediated by water bridges with a serine. Five of the analysed bromodomains had a conserved tyrosine residue oriented away from the binding site for significant amounts of time in proteins: BRDT(1) 100%, CREBBP (31%), BAZ2B (24%), SMARCA4 (21%), and BRD1(11%). They found that this is only reflected in two PDB structures (PDBs: 2RFJ and 2DKW) out of more than 50 available crystal and solution structures. Another bromodomain study identified flexibility for the formation of an induced fit-pocket for small-molecule binding with CREBBP proteins.[103]

1.3.7 Methylated Lysine Readers

Recognition of methylated lysines usually involves a hydrophobic cage, composed of two to four aromatic residues such as tryptophan, tyrosine, and phenylalanine, which form favourable cation–π interactions with the methylammonium ion.[104,105]

Di- and trimethylated lysines tend to be recognised by readers through a surface-groove mode, meaning that they are inserted in a wide and accessible binding site. Monomethylated lysines can also bind to these cages, but usually require other negatively charged residues to form H-bonds with the methyllysine.[104,106] Selectivity for monomethylated substrates is achieved by containing a binding site with a deep and size-sensitive protein cleft, known as the cavity-insertion recognition mode.[107]

When analysing past computational studies of methylated lysine readers two major themes emerge: firstly, an interest in understanding the flexibility of complexes formed between the reader regulators and the modified histones, and secondly,

identification of key residues, such as those in hydrophobic cages, for the recognition and selectivity of different methylation states (e.g. mono-, di- or trimethylated lysines) of the histones. Of these readers, the most work has been done on the chromodomain class of proteins.

Yu and co-workers performed 10 ns molecular dynamics with the AMBER ff03 force field and TIP3P solvation to evaluate the flexibility of the Drosophila heterochromatin-associated protein 1 (HP1) in the free and histone H3 bound states.[108] The free complex was much more flexible than when bound to methylated histones and trimethylated lysine interacted more favourably with the aromatic cage of HP1 than the monomethylated, thus providing a key feature for protein–protein recognition.

These observations are in good agreement with previous experimental studies showing the preference for H3K9Me3-HP1 over lower methylation states.[109,110] Stein et al. investigated H3 and the chromodomain *Drosophila melanogaster* protein polycomb complex stability with 4 ns MD simulations using the AMBER force field. Binding free energy calculations determined the more stable complex is formed with trimethylated lysine. They also showed that H3K27Me2 binds more weakly than H3K27Me1 to polycomb, suggesting that dimethylated lysine is less likely to bind to the hydrophobic pocket of the chromodomain. Furthermore, through Electrostatic Decomposition Analysis (EDA), they found that tyrosine and tryptophan residues formed the strongest interactions with the substrate. Other works explored the importance of these hydrophobic cage residues. Li et al. performed 4 ns MD simulations and showed that a hydrophobic cage, formed by a phenylalanine and a tryptophan, is important for the H3 binding to another chromodomain, the Chromobox Protein Homolog 6 Chromodomain (CBX6).[111] In a later study, Stein and Wang performed 2 ns MD with the AMBER ff03 force field and TIP3P solvation to conclude that hydrophobic residues tyrosine, phenylalanine and tryptophan, are fundamental for the stability of the chromodomain helicase DNA-binding Protein 1 with the methylated H3 and binding was more energetically favourable for trimethylated lysine over monomethylated lysine.[112]

Lu et al. studied the importance of charge independent and hydrophobic cage effects for H3 substrate recognition by the HP1 chromodomain with 5 ns MD using the AMBER ff99 force field.[113] They found that cation–π interactions play

an important role for the binding of the positively charged lysine compared to neutral substrates. Ozboyaci et al. studied the tudor domains of the demethylase JMJD2A using 25 ns MD with the AMBER ff03 force field and found that binding free energies are lower for the JMJD2A binding to H4 than to H3 and that binding is preferred for tri- over dimethylated histones.[114]

Grauffel et al. analysed the recognition of methylated H3 by PHD fingers and the most important energy contributions of individual residues using 10 ns MD with the CHARMM27 force field and TIP3P solvation.[115] They highlighted that a conserved tryptophan residue is crucial, but alone insufficient, for histone recognition by PHD finger-containing proteins. Additionally, the first four residues of H3, known as the Linear Interaction Motif, and not only the modified lysine residue itself were found to be responsible for significant energetic contributions to the binding and recognition of the methylated substrate K4Me3 to the PHD finger.

1.3.8 Outlook for Computational Epigenetics

The *in silico* studies outlined in this section have been valuable for quantifying and explaining mechanisms that drive writing, reading, and erasing epigenetic processes. They have also helped to establish the sequence of catalytic events and impacts of protein dynamics on binding. Specifically, the value of techniques such as:

- i) long-time MD simulations – to understand the impact of loop-dynamics on recognition of methylated and acetylated histone lysines;
- ii) QM/MM studies, which include the effects of dispersion in the active site and long-range electrostatic interactions from remote residues – to describe chemical events;
- iii) and energy decomposition schemes – for identifying key residues contributing to recognition and catalysis;

has been thoroughly demonstrated.

An importance on inclusion of the protein environment when modelling both binding and reactions has been emphasised by many of these works. One example is the role of loop-dynamics on reader enzymes capable of activating or deactivating

the protein depending on whether they are open or closed during substrate-binding. Another example are the single residues with the ability to act as a selective gate and block channels necessary for movement of substrates and water molecules. Future studies should emphasise inclusion and accurate representation of the surrounding biological environment, as well as continued refinement of parameters, such as MM forcefields. The possible insights gained by advancing these *in silico* approaches have the potential to advance the design of inhibitors, both in terms of efficiency and selectivity.

1.4 Principles of Computational Chemistry

Computational chemistry techniques can be valuable for studying chemical phenomena in a more cost and time efficient way than traditional experimental procedures. Examination via observation might be more difficult because of the short-lived or unstable properties of intermediates and transition states, where utilisation of computational calculations and models can provide missing insight.

A diverse array of *in silico* approaches exist that can generally be classified as empirical, semi-empirical, and *ab initio*. Empirical methods, such as molecular dynamics, neglect electronic effects, and rely on parametrisation from higher level calculations and experimental data. *Ab initio* approaches, such as the Hartree-Fock Method, depend solely on inputs from fundamental physical constants and solve the electronic structure problem from first principles. Semi-empirical methods utilise a simplified version of Hartree-Fock Theory and adjustable empirical parameters.

In the following subsections, I will describe the fundamental theory underlying these techniques and the cost-benefits of specific methods.

1.4.1 Molecular Dynamics Simulations

Classical molecular dynamics (MD) simulations provide the biochemist with a unique ability to view the behaviour of macromolecules as a function of time. These approaches have advanced greatly since their initial application in 1977 by McCammon et al. studying the bovine pancreatic trypsin inhibitor.[116] This simulation was performed for a relatively short period of time (9.2 ps) on the small system (987 atoms). Advances in analysis and visualisation software, more accurate force fields, and ability to run on more powerful computer/graphics processing units have allowed the timescale to reach microseconds for larger system sizes - fundamental capabilities needed for simulating drug/protein interactions, ion channels, cell membranes, and nucleic acids.[117]

Classical Equations

MD simulations utilise a ball and spring model to describe atoms and bonds within a given system. Nuclear movement within a molecule is approximated using numerical solutions to Newton's second law of motion to simulate the trajectory of particles, shown as:

$$\frac{F_i}{m_i} = \frac{d^2 r_i}{dt^2} \tag{1.1}$$

Here, F_i is the force on atom i and divided by its mass m_i is related to the second derivative of its position as a function of time. F_i can be solved by taking the derivative of $U_{total}(r_N)$, the potential energy of the system:

$$F_i = \sum F_{ij} = -\frac{d}{dr_i} U_{total(r_1, r_2 \ldots r_n)}, i = 1, 2 \ldots N \tag{1.2}$$

$U_{total}(r_N)$ is a function of the atomic coordinates of the total N number of atoms and $\sum F_{ij}$ is the summation of forces on atoms i and j along the interatomic axis defined as:

$$r_{ij} = |r_i - r_j| \tag{1.3}$$

Velocity Verlet Algorithm

Positions and velocities are computed using integration algorithms based on Taylor series expansions. The Velocity Verlet Algorithm is a time reversible integrator for MD simulations that computes position r and velocity v at the next timestep $(t+\delta t)$, as shown in the two equations below:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}\delta t^2 a(t) \tag{1.4}$$

$$v(t + \delta t) = v(t) + \frac{[a(t) + a(t + \delta t)]}{2}\delta(t) \tag{1.5}$$

The Velocity Verlet Algorithm is beneficial in terms of data storage as it eliminates the need to remember positions and velocities from previous timesteps and preserves linear and angluar momentum.[118]

MM Force Fields: the functional form

The general potential energy function used in molecular mechanics (MM) force fields is a summation of terms describing contributions from bonded and non-bonded interactions terms:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb} \tag{1.6}$$

The bond stretching term models small deviations from equilibrium, or reference values, which can be derived from empirical data, such as X-ray crystal structures, or QM calculations. The harmonic bond potential U_{bond} is defined as:

$$U_{bond} = \sum_{bonds} k_b(r - r_0)^2 \tag{1.7}$$

where k_b is the stretching constant, r is the bond length between two bonded atoms, and r_0 is the bond length for the minimum energy configuration of the molecule. The U_{angle} term describes the energy related to oscillations of three atoms about an equilibrium bond angle and utilises a harmonic potential:

$$U_{angle} = \sum_{angles} k_\theta(\theta - \theta_0)^2 \tag{1.8}$$

where k_θ is the valence angle force constant, θ is the valence angle among three bonded atoms, θ_0 is the reference valence angle. The $U_{dihedral}$ term describes the energy contribution from the torsional rotation of four atoms around a bond. Deviations in dihedral angles are large, this is due to the nature of the internal rotation barriers, which are relatively lower than in other interactions. Additionally, the torsional potential is periodic through a 360° rotation and therefore it is common to model this using a Fourier series:

$$U_{dihedral} = \sum_{dihedrals} k_\phi(1 + cos(n\phi - \delta)) \tag{1.9}$$

Where the phase angle is represented as δ and the following terms are provided by the parameter files: the dihedral angle among four bonded atoms, ϕ, the dihedral angle force constant, k_ϕ, and the multiplicity, n. The non-bonded interactions are accounted for in the final two terms and represent energetic effects as a result of van der Waals attraction, steric repulsion, and electrostatic interactions. The U_{vdW}

is represented using the Lennard-Jones potential, which describes both the steep repulsion due to the interaction of the electron clouds from both atoms and the weakly attractive dispersion forces from dipole-induced dipole interactions:

$$U_{vdW} = 4\varepsilon[(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^{6}] \tag{1.10}$$

Where r_{ij} is the distance between atoms i and j, σ is the minimum energy difference, and ε is the depth of the potential well. The behaviour given $\sigma = 1$, $\varepsilon = 1$ and $\sigma = 1$, $\varepsilon = 2$ is shown in **Fig. 1.7**. The electrostatic interaction is accounted

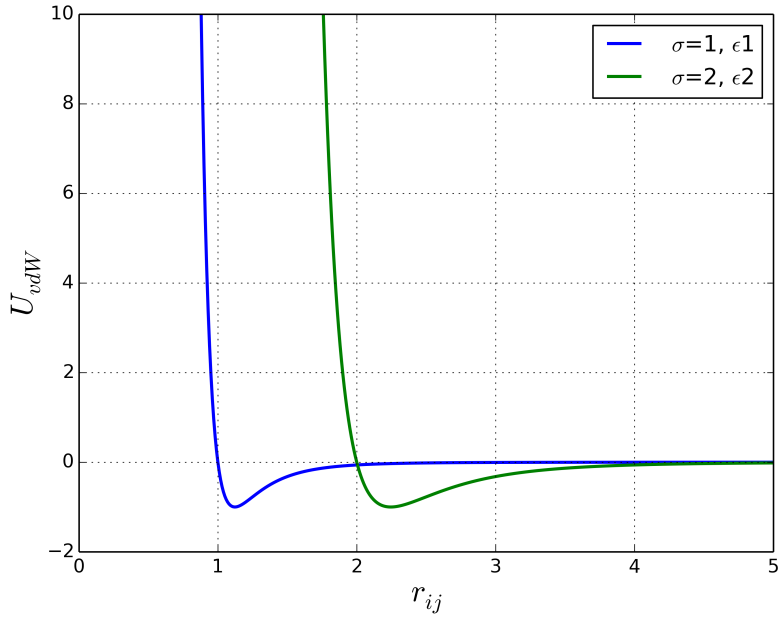


Figure 1.7: Lennard-Jones potential plot for $\sigma = 1$, $\varepsilon = 1$ (blue) and $\sigma = 1$, $\varepsilon = 2$ (green).

for by the $U_{Coulomb}$ term. This derives from the unequal distribution of charge in a molecule, and is modelled by designating point charges for each atom that sum to the formal charge of the molecule.

$$U_{Coulomb} = \frac{q_i q_j}{\varepsilon r_{ij}} \tag{1.11}$$

Where q_i and q_j are the atomic charges of atoms i and j, and ε is the permittivity of free space.

Explicit Solvation Models

Solvation models are used to understand solute-solvent interactions and quantify associated energies when considering solutes in the liquid phase. The simulated systems discussed in this Thesis employed explicit solvation models, where water molecules are explicitly added to the system of interest. As with the solute, the solvent interactions are treated using force fields that define parameters such as atom, bond, and angle size as well as fitted charges, based on experimental data or high level calculations. An extensively used series of models used in simulations are the TIPXP models, where X refers to the number of sites considered. The TIP3P, -4P, -5P, and -6P models are shown in **Fig. 1.8**. Specifically, the TIP3P is

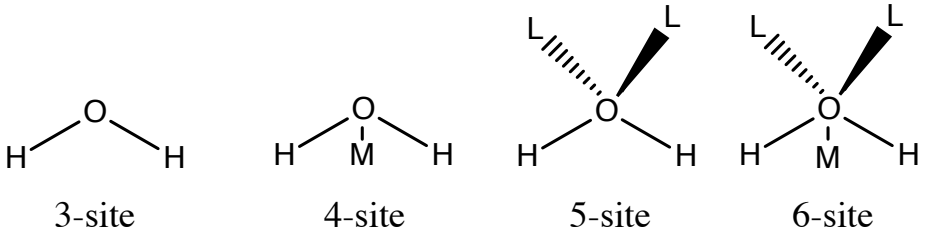


Figure 1.8: Structures of 3-, 4-, 5-, and 6-site solvation models where M is a dummy atom and L represents a lone pair.

commonly used when simulating protein systems and was employed in all simulations discussed in this Thesis. This model describes water molecules using three atoms each assigned with a point charge with a rigid geometry. TIP4P contains an extra dummy atom while TIP5P models the lone pair on oxygen with two extra sites. TIP6P is comprised of both a dummy atom and two lone pair sites. Incorporation of the dummy atom and lone pairs has shown to produce more accurate representations of the electrostatics and tetrahedral geometry of water, respectively, but with an expected increase in computational cost.

Periodic Boundary Conditions

The application of periodic boundary conditions (PBC) facilitates the calculation of bulk properties, necessary for studying proteins in solution, by minimising of edge effects in a finite system. These edge effects are important to consider due to the

geometry of the systems. For example, considering a simple cubic lattice of 1000 molecules, 488 would be located on the surface, and would otherwise experience different forces than non-surface atoms. When using PBC, the system is placed into a space-filling box, or unit cell, and then surrounded by translated copies of itself, to form a lattice. This eliminates artefacts fundamentally imposed by defining boundaries in the system. In the case of 3D PBC, If a particle travels outside of the simulation box it is reintroduced with the same velocity in the box by the opposite unit cell, as shown in **Fig. 1.9**. Some examples of space-filling

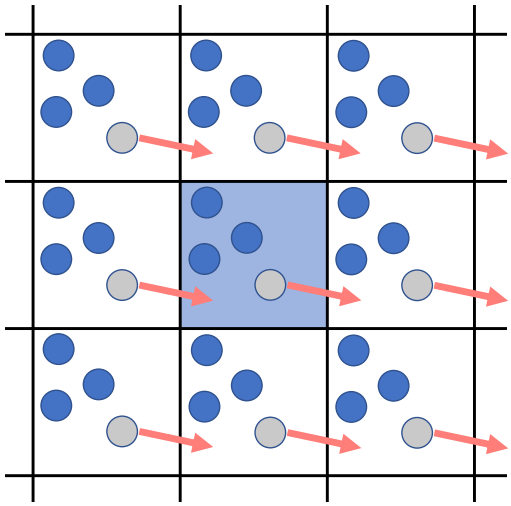


Figure 1.9: Visualisation of 2D periodic boundary conditions.

unit cells shapes are truncated octahedrons, rhombic dodecahedrons, and cubic cells. In this Thesis, all simulations employed a truncated octahedron, which is more spherical than cubic cells and therefore more appropriate for simulating globular biological macromolecules.

Neighbour Lists

The use of neighbour lists reduces computational costs by employing a distance cut-off, r_{cut}, when considering non-bonded interactions. Generally, these are expensive as pairwise calculations between all atoms is required to calculate the minimum image separations r_{ij}. If $r_{ij} > r_{cut}$, the potential $\upsilon\ r_{ij} \approx 0$. From this, it is justified to not perform the force calculation and move to the next atom, e.g. j. However,

performing the pairwise distance measurement for all distinct pairs, $\frac{1}{2}N(N-1)$ for N atoms, is still time consuming. The employment of a pairwise list of atoms, such as the Verlet List,[118] can expedite the program, where a potential cutoff sphere with radius r_{cut} around an atom is also surrounded by a larger sphere with radius r_{list}. Pairwise r_{ij} are computed for all pairs in the first step, but only those within the r_{list} sphere are retained and the corresponding forces calculated, as shown in **Fig. 1.10**. In the following MD steps only the listed pairs are checked until the list is reconstructed after a specified amount of steps. This can be automated by calculating the distance travelled by each since the previous update. Another

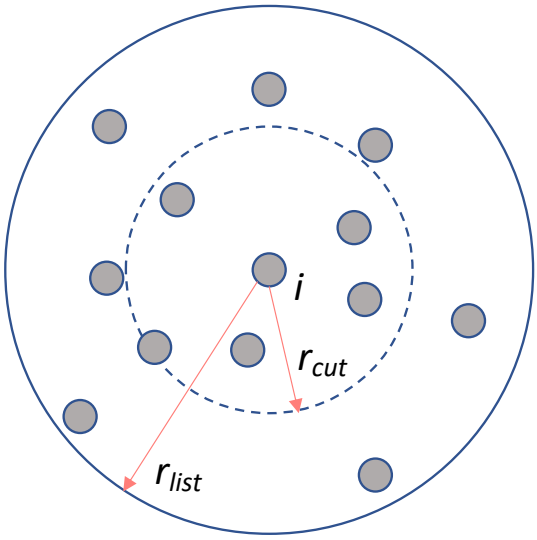


Figure 1.10: Schematic of Verlet List where atoms within r_{cut} (solid) and r_{list} are included in the Neighbour List.

technique employs dividing the unit cell into a lattice of n_{cell} x n_{cell} x n_{cell} cells, where the side of the cell $l_{cell} = \frac{L}{n_{cell}} > r_{cut}$. A list can be made at each step of the atoms belonging to which cell and linked lists can describe neighbouring atoms in other cells with corresponding force calculations. This method is advantageous for large systems with short-range forces.

Ensemble Generation

Simulations are characterized quantitatively by properties such as volume (V), pressure (P), total energy (E), temperature (T), number of particles (N), and

chemical potential (μ). The main justification of MD is the ergodic hypothesis which states that statistical ensemble averages are equal to time averages of the system. Depending on the ensemble type, different properties must be fixed or varied as outlined in **Table 1.7**. A microcanonical ensemble is requires the number of atoms, volume, and energy to remain constant. With NVT dynamics, known as a canonical ensemble, the temperature is maintained using a thermostat. The simplest method to fix temperature is the direct scaling of atomic velocities:

$$V^{new} = c_t V^{old} \tag{1.12}$$

where,

$$c_t = \sqrt{\frac{T_0}{T}} \tag{1.13}$$

Here the velocity of a particle, V^{old}, is scaled at each time step to a new velocity, V^{new}, to reach a set temperature, T_0. However, this eliminates the ability to account for small fluctuations in temperature typical of biological systems. The Berendsen thermostat instead tends towards a set temperature instead of resetting to a specified value by rescaling atomic velocities such that the rate of temperature change is proportional to the temperature difference. A system that maintains a constant temperature and pressure is known as an isothermic-isobaric ensemble. Here, a barostat and thermostat are required and operate in similar ways. In the Berendsen barostat, the size of the simulation cell is scaled to maintain constant pressure. Depending on the ensemble type, a distinction is made thermodynamically with

Table 1.7: Constants and associated equilibrium states for a given ensemble type.

Name	N	P	V	T	E	μ	Equil.
Canonical (NVT)	x		x	x			A min.
Micro-canonical (NVE)	x		x		x		S max.
Isothermal-isobaric (NPT)	x	x		x			G min.
Grand Canonical (VEμ)			x	x		x	(PV) max.

regards to Helmholtz free energy, defined as:

$$F = U - TS \tag{1.14}$$

and Gibbs free energy, defined as:

$$G = H - TS \tag{1.15}$$

Here, U is the internal energy, T is temperature, and S is entropy. The enthalpy H is further decomposed into two terms:

$$H = U + pV \tag{1.16}$$

where p and V are the pressure and volume, respectively. The Helmholtz and Gibbs free energies are appropriate when considering constant volume and pressure, respectively. Experimental techniques and physiological conditions specify isobaric conditions, therefore the Gibbs free energy gradient acts as a driving force for a system and is the focus of post-processing calculations.

Monte Carlo Simulations

Monte Carlo simulations are an alternative way to sample the configurational space of a given system. Here, the energy of the system is calculated for a starting state and after an arbitrary particle is feasibly moved to generate a new configuration. The lower energy state is carried forward and is used as the initial point for the next iteration. The higher energy state could be accepted as determined by the Metropolis criterion,[119] defined as:

$$p = e^{-\frac{\Delta E}{RT}} \tag{1.17}$$

ensuring aHere p is the probability of retaining the state, E is the potential energy, R is the ideal gas constant, and T is the temperature. If the configuration is rejected the initial configuration is again used in the next iteration. This affords the capability to sample higher-energy structures, this is important so that the simulation is not trapped at a local minimum.

Post-processing Techniques

Validation of the simulation is an important post-processing procedure, including checking the stability of the system and ensuring it is within reasonable agreement with empirical data. These include: calculating the energy fluctuations over simulation time, the root mean square deviation (RMSD) compared to the starting structure, the solvent accessible surface area (SASA), changes in secondary structure, and radius of gyration. To compare with NMR measurements, chemical shifts, coupling constants, and relaxation data can also be calculated. Conversion of positional fluctuations to B-factors can also be performed to validate against X-ray crystal structures. When evaluating binding affinities of ligands it is helpful to calculate binding free energies from the simulation trajectory data. Commonly used techniques include the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) or alchemical methods, which include free energy perturbation (FEP) and thermodynamic integration (TI). MM-PBSA is an approximate method that calculates the free-energy difference over a given time frame at set intervals. Here, the gas-phase binding energies are computed by the MM energy function, solute entropies by the quasi-harmonic approximation, the electrostatic term of solvation free energy by the PB equation, and the non-polar term by an approximation proportional to the SASA. The relative energies depend on the initial and final states from each snapshot, and are then averaged. In the case of alchemical calculations, the ligand is alchemically transformed into another chemical or a non-interacting dummy molecule through a number of intermediates in nonphysical stages. Subsequently, the FEP and TI approaches are much more computationally expensive than performing MM-PBSA calculations. One application is comparing binding energies of two ligands to the same protein. This employs a thermodynamic cycle, shown in **Fig. 1.11** where ligand A is transformed into ligand B in solution while bound and unbound to the enzyme E, using simulations. In the force field energy expression, this corresponds to adjusting van der Waals parameters, atomic charges, and all other parameters that are affected by the change in atom types.

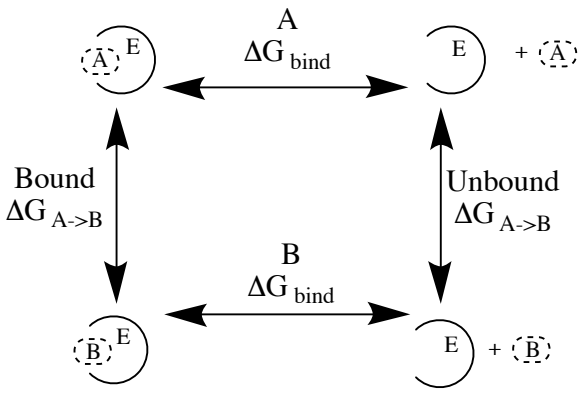


Figure 1.11: Thermodynamic cycle used in alchemical post-processing methods to compare binding affinities of ligands A and B to enzyme E in solution.

The advantage of this method is to eliminate the need to simulate both ligands leaving the enzyme pocket, which would incur a huge computational expense. Instead the ligand is transformed through intermediate states and the difference in binding energy, $\Delta\Delta G_{bind}$ is obtained using the equation:

$$\Delta\Delta G_{bind} = \Delta G^{B}_{bind} - \Delta G^{A}_{bind} = \Delta G^{bound}_{A->B} - \Delta G^{unbound}_{A->B} \tag{1.18}$$

If the free energy cost of transforming ligand A into ligand B is larger when unbound than bound, ligand A would therefore be predicted to bind more weakly to enzyme E than ligand B.

Limitations of MD

As with any theoretical technique, it is crucial to consider the restrictions facing the methodology employed in an investigation. For MD simulations, it is important to account for limitations such as:

- The appointment of force field. This is crucial to ensure that the simplified form of the potential represented by the chosen force field is appropriate for the given system. For example, certain force field libraries are parametrised such that they capable of simulating all elements, such as the Universal Force Field (UFF),[120] while others have only been designed to model biologically relevant systems, such as the Assisted Model Building with Energy Refinement (AMBER)[121] and Chemistry at HARvard Macromolecular Mechanics (CHARMM)[122] force fields.

- The size of the system directly affects the time scale of a process that can be simulated at the atomic level. Truncation of the system depending on the methodology employed may be required. For example, the GROMOS Software for (Bio)Molecular Simulation manual cites a limitation of 10^5 atoms that can reasonably be simulated.[123]

- MD simulations are reasonably capable of modelling the behaviour of macromolecules over timescales ranging from tens of picoseconds to hundreds of nanoseconds. With respect to the phenomenon of interest, it is important to identify if the simulation time scale can sample enough of the conformational space of the system required. For example, the Trpcage miniprotein, composed of 20 residues, folds in 4 μs *in vitro*.[124]

- Because MD simulations rely on Newton's Laws of Motion, covalent rearrangements, or other phenomenon that are not able to be described by a classical description of atomic motion, will not be captured with these methods alone.

These assumptions and limitations imposed by definition of MD techniques, should be considered when constructing and drawing conclusions from simulations.

1.4.2 Quantum Mechanics

Quantum mechanics (QM) based methods are able to study the electronic effects of molecules of interest. *Ab initio* calculations are computed from solutions to the electronic Schrödinger equation using approximations, whereas density functional theory (DFT) calculations utilise functionals of spatially dependent electron density. The underlying theory, as well as benefits and advantages of different approaches, are discussed in the following sections.

The Schrödinger Equation

To understand electronic structure of atoms and molecules, it is necessary to solve the time-independent and non-relativistic Schrödinger Equation:

$$\hat{\mathcal{H}}\Psi = E\Psi \tag{1.19}$$

Where Ψ is the wave function that describes the quantum state of an entire system of particles, and $\hat{\mathcal{H}}$ is the Hamiltonian operator that is made up of the kinetic and potential energy terms of the system. The square modulus of the wave function $|\Psi^2|$ is a probability density, with the probability of locating a particle within a volume element $d\tau$ proportional to the product $|\Psi^2|d\tau$. For the lowest energy of the system, the related eigenfunction, Ψ, of $\hat{\mathcal{H}}$ corresponds to the ground state.

The Hamiltonian

The Hamiltonian operator is composed of the sum of kinetic and potential energy operators for a system of particles, where N electrons and M nuclei are described by position vectors r_i and R_A, respectively.

$$\hat{\mathcal{H}} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \quad (1.20)$$

Here, r_{ij} and r_{iA} represent the distance between the ith electron and jth electron or the ith electron and Ath nucleus, respectively. M_A and A_A are the mass and atomic number of the Ath nucleus, respectively. The five terms above that comprise the Hamiltonian can also be written more simply:

$$\hat{\mathcal{H}} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{ne} + \hat{V}_{nn} \quad (1.21)$$

Where $\hat{T}_e$ and $\hat{T}_n$ describe the kinetic energy operators of the electrons and nuclei, respectively and $\hat{V}_{ee}$ $\hat{V}_{ne}$ $\hat{V}_{nn}$ represent the attractive Coulombic interactions between electrons, electrons and nuclei, and nuclei, respectively. However, an important approximation, known as the Born-Oppenheimer approximation, is implemented to simplify the Hamiltonian.

The Born-Oppenheimer Approximation

In the Born-Oppenheimer Approximation, nuclei are defined as stationary compared to their electrons. This relies on the principle that electrons essentially react instantaneously to variation in position of the nucleus, rather than velocity, allowing for a simplification of the Schrödinger equation.

This is due to the much higher mass of the nucleus ($1.67493 \cdot 10^{-27}$ kg) as compared to the electron ($9.10938 \cdot 10^{-31}$ kg). Subsequently, a neglect of the nuclei

kinetic energy term and nuclear-nuclear potential energy term is justified such that the electronic Hamiltonian, $\hat{H}_e$, becomes:

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} \tag{1.22}$$

From this, the electronic Schrödinger equation is formulated, which considers the motion of N electrons in the potential field of M stationary point charges, as shown below:

$$\hat{H}_e = \Psi_e(\mathbf{r};\mathbf{R}) = E_e\Psi_e(\mathbf{r};\mathbf{R}) \tag{1.23}$$

where Ψ_e is the electronic wave function. Before discussing different methodologies to solve for Ψ_e, it is important to note that the wave function must be antisymmetric with respect to the interchange of coordinates for any pair of electrons, to account for the asymmetry principle, because it is ultimately describing a multi-fermionic system.

$$\Psi(x_1, ...x_i, x_j, ..., x_N) = -\Psi(x_1, ...x_i, x_j, ..., x_N) \tag{1.24}$$

Utilising Slater determinants (SD) ensures the formulation of antisymmetric wave functions.

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \cdots & \psi_N(1) \\ \psi_1(2) & \psi_2(2) & \cdots & \psi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(N) & \psi_2(N) & \cdots & \psi_N(N) \end{vmatrix} \tag{1.25}$$

Here the columns and rows in the SD refer to the single electron wave functions and the electron indices, respectively. The spin orbitals ψ_i are the products of a spatial and spin function where,

$$\psi_i = \phi(\mathbf{r})\alpha(\omega) \tag{1.26}$$

or

$$\psi_i = \phi(\mathbf{r})\beta(\omega) \tag{1.27}$$

Hartree-Fock Method

The Hartree-Fock method treats pairwise electron-electron repulsion by considering the electron to be moving in a potential defined by the spherical average of the potential of the other electrons. The variation principle is employed to determine a set of spin orbitals ψ_i that yields the lowest electronic energy, and where the HF equations:

$$\hat{f}_i \psi_i = \varepsilon_i \psi_i \tag{1.28}$$

represent a set of one-particle eigenvalue problems. Here $\hat{f}_i$ is the Fock operator, described as the effective one-particle Hamiltonian, with eigenfunction orbital ψ_i, and energy ε_i.

$$\hat{f}_i = \hat{h}_i + \hat{V}_i^{HF} = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_j (\hat{J}_j(1) - \hat{K}_j(1)) \tag{1.29}$$

where

$$\hat{J}_j(1) \left|\psi_i(1)\right\rangle = \left\langle\psi_j(2)\right| \frac{1}{r_{12}} \left|\psi_j(2)\right\rangle \left|\psi_i(1)\right\rangle \tag{1.30}$$

$$\hat{K}_j(1) \left|\psi_i(1)\right\rangle = \left\langle\psi_j(2)\right| \frac{1}{r_{12}} \left|\psi_i(2)\right\rangle \left|\psi_j(1)\right\rangle \tag{1.31}$$

The terms $\hat{J}_j(1)$ and $\hat{K}_j(1)$ are the Coulomb operator and exchange operator (spin correlation), respectively. The self-consistent field (SCF) method is employed to iteratively solve the HF equations. It requires an input of the 3D coordinates of atomic nuclei, makes an initial guess of the 1-electron orbital, calculates the Coulomb and exchange integrals, forms the Fock Matrix by using the average potential calculated in the previous step, and solves the HF equations. The new orbitals are compared to the initial guess and the process continues until the electronic energy difference is below a certain cut-off.

A finite set of K known basis functions, $\phi_\mu(\mathbf{r})$, is used to convert the differential equations to matrix equations, thereby constructing the basis set. Here, the linear combination of atomic orbitals to give molecular orbitals (LCAO-MO) method can be understood, where atomic orbitals are constructed from the expansion of molecular orbitals in terms of basis functions:

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu(\mathbf{r}) \tag{1.32}$$

Substituting this into the HF equations:

$$\hat{f}_i \sum_{\mu=1}^{K} c_{\mu i}\phi_\mu(\mathbf{r}) = \varepsilon_i \sum_{\mu=1}^{K} c_{\mu i}\phi_\mu(\mathbf{r}) \tag{1.33}$$

Collecting the K equations into a single matrix for all the $c_{\mu i}$, the molecular orbital expansion coefficients, produces the Roothaan-Hall equations:

$$\mathbf{FC}=\mathbf{SC}\varepsilon \tag{1.34}$$

Where the **F** Fock matrix and **S** are defined as:

$$F_{\mu\upsilon} = \langle\phi_\mu|\,\hat{f}\,|\phi_\upsilon\rangle \tag{1.35}$$

$$S_{\mu\upsilon} = \langle\phi_\mu|\phi_\upsilon\rangle \tag{1.36}$$

Here, **C** is a K x K matrix relating to $c_{\mu i}$ and ε is a K x K diagonal matrix of ϵ_i. The Roothaan-Hall equations are subsequently solved using a set of K basis functions for a system of N electrons, yielding $2K$ spin orbitals and N occupied orbitals, and $2K$-N unoccupied orbitals. The ground state of the system is the SD based on the occupied spin orbitals.

Electron Correlation

In HF theory, individual electron-electron interactions are not treated explicitly. Instead, the interaction between electrons is determined by the movement of one electron in the mean field defined by the remaining electrons. As a result, this neglects instantaneous interactions between electrons, and does not consider that electrons will avoid moving very close to each other. This is commonly referred to as a neglect of electron correlation effects. By definition, calculations using the HF method will therefore always overestimate energies by approximately 1%. The correlation energy $\mathrm{E_{corr}}$ is defined as:

$$E_{corr} = E_{exact} - E_{HF} \tag{1.37}$$

where E_{exact} and E_{HF} are the exact and HF computed energies of they system, respectively. Correlation can be described as 'static' or 'dynamic.' Dynamic correlation refers to the avoidance of electrons with electrons that occupy the same

orbital. Static correlation arises due to repulsion felt by electrons in different orbitals. This is significant when considering systems for which a single SD is inadequate to represent multiple degenerate configurations, such as H_2 at the dissociation limit. Here, H_2 has two degenerate electronic configurations, $|\uparrow ... \downarrow\rangle$ and $|\downarrow ... \uparrow\rangle$ and therefore requires two SDs to represent the ground state.

To correct for correlation some approaches, such as Configuration Interaction (CI) and Coupled Cluster (CC), utilise the HF wave function as a foundation and improve the HF method by adding flexibility. In theory, electrons in the occupied orbitals in the HF reference determinant are moved into the virtual orbitals, therefore producing new SDs. To denote the number of excited electrons, this is referred to as singly, doubly, triply, etc. (Singles (S), Doubles (D), Triples (T)...) excited. This is visualised in **Fig. 1.12**. CI methods employ linear combinations of SD

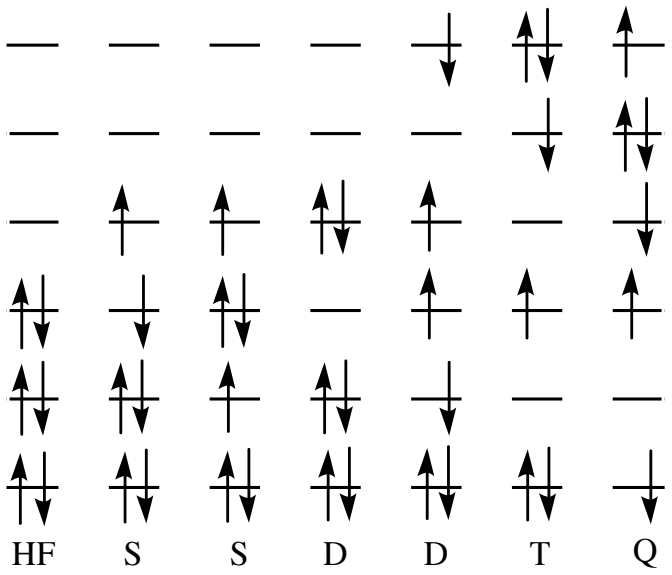


Figure 1.12: Electrons excited from HF wave function to generate excited Slater determinants.

to construct the wave function:

$$\Psi_{CI} = c_0\Psi_{HF} + \sum_S c_S\Psi_S + \sum_D c_D\Psi_D + \sum_T c_T\Psi_T + \sum_Q c_Q\Psi_Q + ... \tag{1.38}$$

The CI wave function is the produced by finding the expansion coefficients c_i using the linear variational method. The correlation energy for a given basis set is calculated by finding the difference between the HF energy and the full configuration interaction calculation (FCI) energy (the exact ground state energy), where the FCI is computed by considering all possible excitations. To reduce computational cost, only the SDs corresponding to double and triple excitations are computed, known as the configuration interaction including singles and doubles (CISD) or doubles and triples (CISDT) methods. Furthermore, the inner core electrons can be frozen to reduce computational cost. CISD is capable of recovering $\approx 90\%$ of the correlation energy.

Density Functional Theory

Density functional theory (DFT) is a subset of QM methods where models of molecular orbitals as a function of electron density $\rho(\mathbf{r})$ are used instead of individual one-electron wave functions. A major advantage of this approach is the dependence on fewer variables. When considering electron density, this can be described with three spatial coordinates for any size N electron system. In contrast, a wave function for a system containing N electrons will require $4N$ coordinates (three spatial and one spin variable for each electron). By definition, DFT is also advantageous because it considers electron correlation by definition.

The underlying theory justifying DFT approaches was first proven by Hohenberg and Kohn.[125] The Hohenberg-Kohn theorem asserts that the ground state electron density can uniquely determine $\hat{h}_e$, the electronic Hamiltonian, and subsequently the ground state wave function.

$$\rho_0 = N, Z_A, R_A \rightarrow \hat{H} \rightarrow \Psi_0 \rightarrow E_0 \tag{1.39}$$

The total energy can be described in three terms that are a functional of ρ

$$E[\rho] = E_{Ne}[\rho] + T[\rho] + E_{ee}[\rho] \tag{1.40}$$

Here $E_{\mathrm{Ne}}[\rho]$ is the potential energy due to the nucleus-electron attraction, $T_{\mathrm{ne}}[\rho]$ is the kinetic energy and $E_{\mathrm{ee}}[\rho]$ is the electron-electron interaction. The universal functional $F_{\mathrm{HK}}[\rho]$ is used to describe the system-independent terms:

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho] \tag{1.41}$$

Determination of the functional $F_{\mathrm{HK}}[\rho]$ would allow for the exact solution to the Schrödinger equation, however this has yet to be achieved. Initial attempts to calculate the kinetic energy term using only electron density proved difficult, but implementation of the Kohn-Sham (KS) DFT method were able to provide solutions to these challenges. The concept relies on a fictitious reference system of non-interacting electrons from which the density is constructed and the kinetic energy is calculated. One-electron Kohn-Sham orbitals φ_i are introduced and the same SCF methodology, discussed previously, can be employed to yield an equation relating the Kohn-Sham one-electron operator, $\mathbf{F}_i^{KS}$ and the electronic energy:

$$\mathbf{F}_i^{KS}\varphi_i = \varepsilon_i\varphi_i \tag{1.42}$$

where the functional form of $\mathbf{F}_i^{KS}$ is similar to the one-electron Fock operator but instead contains the potential term

$$\mathbf{F}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_A -\frac{Z_A}{|\mathbf{R_A} - r_i|} + \int \frac{\rho(\mathbf{r'})}{|\mathbf{r'} - \mathbf{r}|}dr' + V_{XC}(\mathbf{r}) \tag{1.43}$$

Optimising the $V_{XC}(\mathbf{r})$ term is necessary as it accounts for the non-Coulomb electron-electron repulsion and corrects for the difference in kinetic energy from the employment of fictitious orbitals. The exchange-correlation potential is related to the exchange-correlation energy as follows:

$$V_{XC} = \frac{\delta E_{XC}}{\delta \rho} \tag{1.44}$$

Density Functionals

A hierarchy of density functionals exist with respect to accuracy. This is visualised in Jacob's Ladder, where commonly used functions are grouped by the methodology underlying the treatment of empirical data, outlined in **Table 1.8**.

The local density approximation (LDA) method is the most basic approach, using only the density ρ at a point to determine the exchange-correlation energy at that point. Unrestricted LDA, known as local spin density approximation (LSDA) is used when α and β have unequal densities. The exchange-correlation functional, $E_{XC}^{LDA}[\rho]$ is based on a homogenous gas and is defined as:

$$E_{XC}^{LDA}[\rho] = \int \rho(\mathbf{r})\varepsilon_{XC}(\rho(\mathbf{r})d\mathbf{r} \tag{1.45}$$

Where the exchange-correlation energy per particle, $\varepsilon_{XC}(\rho(\mathbf{r})$, can be described as the sum of the exchange, $\varepsilon_X(\rho(\mathbf{r})$, and correlation $\epsilon_X(\rho(\mathbf{r})$ energies:

$$\varepsilon_{XC}(\rho(\mathbf{r}))d\mathbf{r} = \epsilon_X(\rho(\mathbf{r})) + \varepsilon_C(\rho(\mathbf{r})) \tag{1.46}$$

and $\varepsilon_X(\rho(\mathbf{r})$ has an analytical solution:

$$\varepsilon_X(\rho(\mathbf{r})) = \frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}(\rho(\mathbf{r}))^{\frac{1}{3}} \tag{1.47}$$

$\varepsilon_C(\rho(\mathbf{r})$ has been evaluated by high level quantum Monte Carlo calculations.[126] L(S)DA performs well when considering systems such as solids that a more uniform electron density throughout. Where L(S)DA struggles is representing molecular systems by overestimating bond strength due to a significant underestimation of exchange energy and a two-fold overestimation of correlation energies. An advancement on L(S)DA methods is achieved by considering the gradient of the electron density, $\nabla\rho$, known as generalised gradient approximation (GGA) methods. This is able to account for the inhomogeneity of the electron density:

$$E_{XC}^{GGA}[\rho_\sigma] = \int f(\rho_\sigma(\mathbf{r})), \nabla\rho_\sigma(\mathbf{r}))d\mathbf{r} \tag{1.48}$$

where

$$\sigma = \alpha, \beta \tag{1.49}$$

A well known GGA exchange functional is the B or B88 by Becke that offers a correction to the L(S)DA exchange energy.[127] This contains one empirically parametrised term that was fitted with data of exchange energies from six noble gases. Other examples include the popular Lee, Yang, and Parr (LYP) functional fitted using empirical data from helium, and has been known to accurately treat the self-interaction error.[128,129] Some noted challenges are computing properties such as reaction barriers and relative energies for different electronic states. The third 'rung' of Jacob's Ladder refers to meta-generalised gradient approximation (meta-GGA methods). The spin-dependent equation is defined as:

$$E_{XC}^{MGGA}[\rho_\sigma] = \int f(\rho_\sigma(\mathbf{r})), \nabla^2\rho_\sigma(\mathbf{r}), \tau_\sigma)d(\mathbf{r}) \tag{1.50}$$

Where the kinetic energy, τ_σ, and the second derivative of ρ is included. Two popular functionals from this class is the functional by Tao, Perdew, Staroverov, and Scuseria

Table 1.8: Jacob's Ladder of density functional approximations.

Level	Name	Variables	Examples
1	Generalized RPA	ρ, $\nabla\rho$, $\nabla^2\rho$, or τ, HF exchange, Virtual orbitals	OEP2
2	Hybrid-GGA	ρ, $\nabla\rho$, $\nabla^2\rho$, or τ, HF exchange	B3LYP, TPSSh, PBE0
3	Meta-GGA	ρ, $\nabla\rho$, $\nabla^2\rho$, or τ,	TPSS, B95,
4	GGA	ρ, $\nabla\rho$	BLYP, PBE, PW91
5	Local density	ρ	LDA, LSDA

(TPSS)[130] and B95 developed by Becke.[131] One benefit of the B95 method is the elimination of the self-interaction error since the correlation vanishes for one-electron systems. Even more success has been achieved through the development of hybrid functionals. These use combinations of exchange energy treatments from HF, LSDA, and GGA methods. An example is the B3PW91 functional by Becke that incorporates three empirical parameters, shown below as:

$$E_{XC}^{B3PW91} = (1-a)E_X^{LSDA} + aE_X^{HF} + bE_X^{B88} + E_C^{LSDA} + c\Delta E_C^{PW91} \tag{1.51}$$

Where a, b, and c are 0.20 (20% exact exchange), 0.72, and 0.81, respectively. Replacement of the E_C^{PW91} term yields the extensively used B3LYP functional, popular for its accuracy when applied to a diverse array of system.[132] The meta-GGA TPSS method, discussed previously, has also been improved (TPSSh) by incorporating around 10% of exact exchange.[133] Another widely used family of functionals are the Minnesota functionals by Truhlar and co-workers including: M05, M06, M06-2X, M06-HF containing 28%, 27%, 54%, and 100% HF exchange, respectively.[134–138] Overall, performance of hybrid-functionals is system-dependent with the regards to the amount of exchange that should be included. Examples of DFT functions employed in the projects presented in this Thesis include B3LYP, M06, and ωB97.

Basis Sets

A basis set is defined as a set of functions used to describe the orbitals in an atom or molecule. A 'complete' basis set requires an unknown molecular orbital (MO) to be expanded by an infinite number of functions, which is impossible to achieve in actual calculations. Subsequently, a finite basis set must be used and therefore acts as an example of an approximation introduced into *ab initio* calculations. Because description of the MO is then limited to the coordinate axes spanned by the finite basis set, larger basis sets are able to provide more accurate representations but are more computationally expensive.

There are two types of basis functions, or atomic orbitals (AO), Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs). STOs have the functional form:

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \tag{1.52}$$

Where N is a normalization constant and $Y_{l,m}$ are spherical harmonic functions, quantum numbers, n, l, m, r is the distance from the electron to the atomic nucleus, ζ is the Slater orbital exponent. The exponential dependence of the electron-nucleus distance parallels the decay observed for the orbitals of the hydrogen atom. From making linear combinations of STOs, radial nodes can be introduced to the radial part of STOs. GTOs have the radial form:

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}e^{-\zeta r^2} \tag{1.53}$$

or using Cartesian notation:

$$\chi_{\zeta,l_x,l_yl_z}(x,y,z) = Nx^{l_x}y^{l_y}z^{l_z}e^{-\alpha r^2} \tag{1.54}$$

Here α is the Gaussian orbital exponent and determines the contraction of the resulting function and the type of orbital is determined by the sum of l_x, l_y, and l_z (e.g. $l_x + l_y + l_z = 1$ is a p-orbital) Three primary differences exist when considering STOs and GTOs:

1. GTOs are considered inferior to STOs due to limitations imposed by the r^2 dependence in the exponential. GTOs have zero slope at the nucleus ($r = 0$), whereas STOs have a cusp (discontinuous behaviour), and subsequently are inferior at representing behaviour near the nucleus, as shown in **Fig.1.13**.

2. As an extension of point 1), GTOs also fall off too rapidly far from the nucleus ($r \to \infty$), causing the 'tail' of the wave function to be represented poorly. This is in contrast to the behaviour of STOs, as shown in **Fig. 1.13**.

3. Despite the limitations with GTOs, they are still preferred with regards to maximising computational efficiency. While GTOs and STOs can be used to form a complete basis set, and it is known that approximately three times as many GTOs as STOs are required to achieve a certain level of accuracy, the increase in number of basis functions is still overall compensated when considering the ease of which the integrals can be calculated.

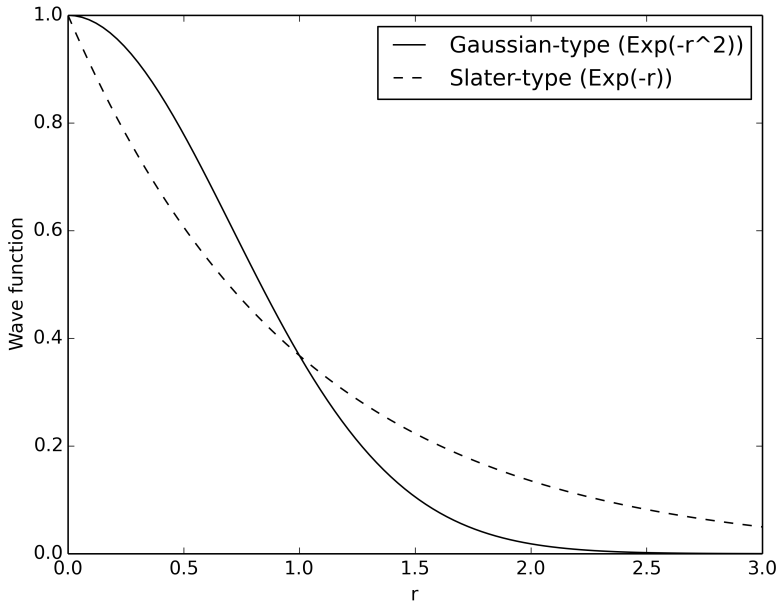


Figure 1.13: Comparison of Slater-type and Gaussian-type functions.

GTOs are grouped together to form contracted Gaussian functions (CGFs). A CGF is a linear combination of the original (or primitive) GTOs, centred on the same nuclei, to reproduce a STO as closely as possible.

$$\chi^{CGF} = \sum_i \alpha_i \chi_i^{GTO} \tag{1.55}$$

A CGF is defined by the contraction coefficients of the GTO and its individual exponents, and these CGFs make up the basis set. A 'minimal' basis set uses one function for each atomic orbital. For example, hydrogen and helium would be described with a single *s*-function (1s) and for the first row elements with two *s*-functions (1s, 2s) and one set of *p*-functions (2p). This is described as being of single = ζ (SZ) quality. To improve the accuracy of the basis sets the basis functions can be doubled (DZ), tripled (TZ), quadrupled (QZ), etc. Continuing the example, first row elements would be described with four *s*-functions (1s, 1s', 2s, 2s') and two sets of *p*-functions (2p, 2p').

Because valence shell electrons are responsible for certain physical phenomenon, such as chemical bonding, it is justified that these should be represented by more flexible basis sets, and the core regions with a more minimal approach. This is accomplished with the employment of split-valence basis sets, which reconcile the need to choose either less accurate SZ basis sets or more accurate but computationally expensive DZ and TZ basis sets.

A commonly used series of basis sets, which were also employed in calculations presented in this Thesis, were developed by Pople and co-workers. These are broadly described in the form: *n-ijk*G, where *n* represents the number of primitives used to describe the core region, and *ijk* denotes how many functions the valence orbitals are split into, and how many PGFs are used to represent them. A basis set with *ij* or *ijk* indicate a split valence or triple split valence, respectively. An example is the 6-31G basis set. From this it can be determined that the core orbitals are described by a single CGF formed by six primitive GTOs. Because there is no *k* term, the valence orbitals are being described by two functions, one constructed by three primitives and one uncontracted. An example of a triple split valence basis set is 6-311G, where a core orbital would be represented with a contraction of six primitive GTOs and the valence orbitals are split into three functions represented by three, one, and one primitive GTOs, respectively.

The addition of polarisation functions to basis sets accounts for the distortion of orbitals caused by proximity to adjacent atoms, as they are always of higher angular momentum *l* than the maximum value of the orbitals occupied in the atomic ground state configuration. For example, adding a set of *p*-functions to a helium atom. This is generally denoted after the 'G' with designations for heavy atoms and

hydrogen. Diffuse functions can also be added and are denoted before the 'G' with either '+' that corresponds to one set of diffuse *s*- and *p*-functions only on heavy atoms or '++' which is identical but with the addition of *s*-functions on hydrogen atoms. This is advantageous for modelling anionic, large, or excited electronic states where these functions have smaller exponents and thus more spatially shallow.

Considering the 6-311++G(2df,2pd) basis set, the nomenclature defines a triple split valence with diffuse *sp*-functions, two *d*-functions, and one *f*-function on heavy atoms and diffuse *s*-, two *p*-, and one *d*-functions on hydrogen atoms.

Correlation consistent basis sets

The correlation consistent basis sets were developed by Dunning and co-workers. These are designed to recover the correlation energy of the valence electrons and converge post-HF calculations systematically towards the complete basis set limit. They are denoted as cc-pVXZ, characterized by the cardinal number X, which refer to the final number of contracted functions. These include: cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z, and cc-pV6Z meaning correlation consistent polarised Valence Double/Triple/Quadruple/Quintuple/Sextuple Zeta, respectively. These basis sets are designed such that a complete shell of functions is added depending on the principle quantum number. For first row atoms, polarisation functions are added as 1d, 2d1f, 3d2f1g when moving from cc-pVDZ to cc-pVQZ, as shown in the **Table 1.9**. This is explained by considering that functions that contribute similar amounts of correlation energy are included at the same stage, independent of the function type. For example, consider that the energy lowering from the first *d*-function is large but the lowering from a second *d*-function is similar to the first *f*-function, and the lowering of a third *d*-function is similar to that from the second *f*- and first *g*-functions. The *s*- and *p*-basis are expanded accordingly. These can be modified with diffuse functions with the 'aug-' prefix. This adds one extra function with a smaller exponent for each angular momentum. For example, in the case of aug-cc-pVDZ there would be an additional *s*-, *p*-, and *d*-function. To recover core-core and core-valence electron correlation additional 'tight' functions (large exponents) can be introduced denoted as cc-pCVXZ. In the case of cc-pVDZ one tight *s*- and *p*- function would be added to heavy atoms.

Table 1.9: Description of contracted and primitive basis functions for the correlation consistent basis sets.

Basis	Hydrogen		First Row		Second Row	
	Contr.	Prim.	Contr.	Prim.	Contr.	Prim.
cc-pVDZ	$2s1p$	$4s$	$3s2p1d$	$9s4p$	$4s3p2d$	$12s8p$
cc-pVTZ	$3s2p1d$	$5s$	$4s3p2d1f$	$10s5p$	$5s4p3d1f$	$15s9p$
cc-pVQZ	$4s3p2d1f$	$6s$	$5s4p3d2f1g$	$12s6p$	$6s5p4d2$f1g	$16s11p$
cc-pV5Z	$5s4p3d2f1g$	$8s$	$6s5p4d3f2g1h$	$14s8p$	$7s6p5d4f2g1h$	$20s12p$
cc-pV6Z	$6s5p4d3f2g1h$	$10s$	$7s6p5d4f3g2h1i$	$16s10p$	$8s7p6d4f3g2h1i$	$21s14p$

Implicit Solvation Models

An accurate but cost-efficient method for modelling solvation is crucial for QM calculations representing molecules in biological systems. Explicit solvation techniques are very computationally expensive because they require the addition of many molecules to the system. Instead implicit solvation models are used, where the solute is placed in a cavity, where the solvent is represented as a polarisable continuum with a given dielectric constant. Here, an electric field is produced from the mutual polarisation of the solute and solvent, that interacts with the solute electronic density resulting in a self-consistent reaction field (SCRF). The goal is to calculate the solvation free energy of the solute, $\Delta G_{solvation}$:

$$\Delta G_{solvation} = \Delta G_{cav} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{elec} \tag{1.56}$$

The terms refer to the cavity solvation energy (ΔG_{cav}), dispersion effects (ΔG_{disp}), repulsive interactions (ΔG_{rep}), and electrostatic interactions (ΔG_{elec}). Size and shape of the cavities are important considerations. The polarisable continuum model (PCM), utilises multiple spheres centred at the solute nuclei atoms, which are based on van der Waals radii, to define a solvent excluding surface used to calculate the cavitation energy.[139] Another surface is defined using a spherical probe that traces the solvent excluding surface to calculate repulsion and dispersion. The electrostatics are calculated using the Poisson equation:

$$\nabla\varepsilon(r)\nabla\phi(r) = -4\pi\rho(r) \tag{1.57}$$

Here ε is the dielectric constant, ϕ is the electrostatic potential, and ρ is the charge distribution. Another widely used model is the solvation model based on density

(SMD).[140] With this model, the electrostatics are defined using the Poisson equation, but uses a set of parametrised radii to define the cavity. Both PCM and SMD models were used in calculations presented in this Thesis.

Potential Energy Surface

The potential energy surface (PES) maps in 3D the energy of a molecule as a function of its energy. The landscape of the PES visualises energetically favourable conformations of a molecule, such as a stationary point, were the gradient of the internal coordinates is zero. These refer to stable species such as the reactants, products, intermediates (minima), and transition states (saddle points) as shown in **Fig. 1.14**. The geometric optimisation calculations performed in this Thesis explore the PES until it converges on a stationary point.

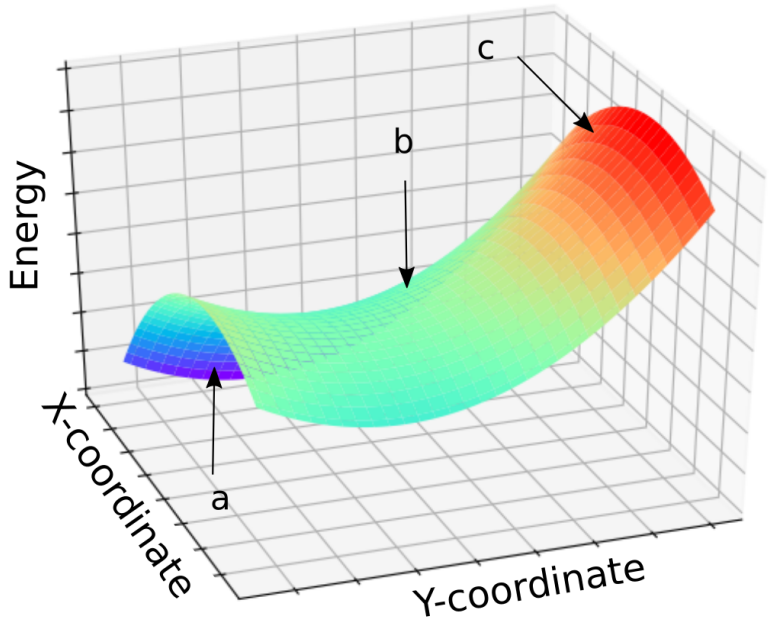


Figure 1.14: Potential energy surface (PES) maps the energy of the molecule as a function of its internal coordinates visualising a) global minimum, b) saddle point, c) and global maximum.

1.5 Contributions of Thesis

An analysis of the past work in the field of *in silico* approaches applied to epigenetic problems has underscored the value of computational techniques for revealing mechanistic insights and advancing drug design. In the following chapters, I will present the results of my research employing molecular mechanics and quantum mechanics calculations, in part performed in collaborations with other experimental groups, to understand:

- Strengths and limitations of MD-based techniques when compared to QM models for predicting binding strength of reader inhibitors.
- Understanding a specific interaction (e.g. a cation–π interaction between a small-molecule and arginine residue) in the context of all available crystal structures containing a similar non-covalent bond by developing a protocol to filter information contained in a large database.
- Investigative MD simulations on the effects of binding to epigenetic readers following mutations of the wild type trimethylated lysine substrate including: alterations of the stereochemistry, chain length, and fluorination of the aromatic cage.
- Methodological development from documentation of novel simulation parameters for post-translationally modified residues, mutated residues, and inhibitors achieved from building simulations.

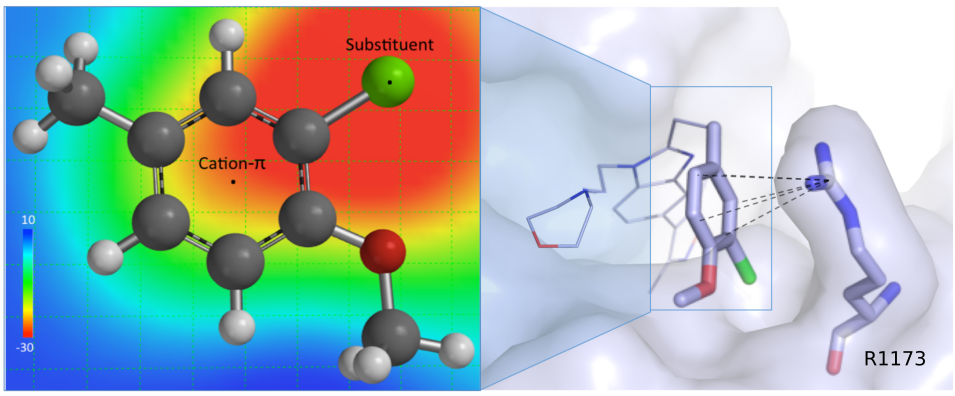
Growth for the sake of growth is the ideology of the cancer cell.

— Edward Abbey, 1977

2

CREBBP Bromodomain Inhibitor Development

Contents



Cortopassi W. A. and **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.

2.1 Introduction: CREBBP and known inhibitors

Epigenetic proteins have been implicated in a wide array of cancers as well as other diseases, such as male infertility and adult obesity.[141,142] Although they have gained increasing attention in recent years as valuable drug targets, only inhibitors targeting writer and eraser inhibitors have been clinically proven, while small-molecule inactivators of epigenetic readers are still in early stages of clinical trials.[143]

Studies examining readers, such as bromodomains, largely focus on one group - namely the bromodomain extra C-terminal domain (BET) family.[144] Non-BETs, such as the cAMP response element-binding protein (CREB) binding protein (CREBBP) family, are underexplored by comparison.[145–148] Drug design targeting non-BETs poses a greater challenge for medicinal chemists. This is due to the high sequence similarity found among the proteins in these families,[149] leading to difficulties achieving selectivity. Maximising the selectivity of an inhibitor is extremely important in order to prevent binding to other structurally similar enzymes, therefore preventing unwanted side effects.

In an experimental and computational collaboration between the Conway Group and Paton Group, the first potent inhibitors of non-BET inhibitors were successfully synthesized in 2014.[103] These target CREBBP bromodomains by mimicking acetylated lysines, their natural substrate. Of these compounds, *(R)*-**2**, a dihydroquinoxalinone (DHQ) derivative, was proven to be selective for CREBBP over seven Bromodomain-containing proteins (BCPs). Furthermore, their work revealed insight into a key driving force, specifically a stabilising cation–π interaction formed between DHQ derivatives and a positively charged CREBBP Arg1173 (R1173) residue.

Following this, Brennan and co-workers (Structural Genomics Consortium at the University of Oxford) synthesized a series of 5-isoxazolylbenzimidazole derivatives that are selective for CREBBP.[150] Evaluation of the X-ray crystal structure (PDB: 4NR7) revealed a stabilising electrostatic interaction between an aromatic 3-chloro-4-methoxyphenyl group and R1173 (**Fig. 2.1A**), similar to the previously investigated *(R)*-**2** inhibitor (**Fig. 2.1B**). A benzoaxazepine-based inhibitor, I-CBP112, synthesized by Fedorov et al. also forms a non-covalent interaction with R1173 that parallels those with the other two small-molecules

(**Fig. 2.1C**).[151] These three inhibitors were able to achieve selectivity for CREBBP bromodomains by exploiting a conserved positively charged Arg (R1173) residue unique to non-BET proteins. In fact, at this location on BET bromodomains, such as BRD4, there is a conserved *negatively* charged Asp residue (D145) present. **Fig. 2.1** highlights not only the stabilizing cation–π interaction between R1173 and selective CREBBP inhibitors, but also the significance of R1173 vs D145 active site residues for distinguishing between the otherwise structurally similar CREBBP and BRD4 bromodomains.

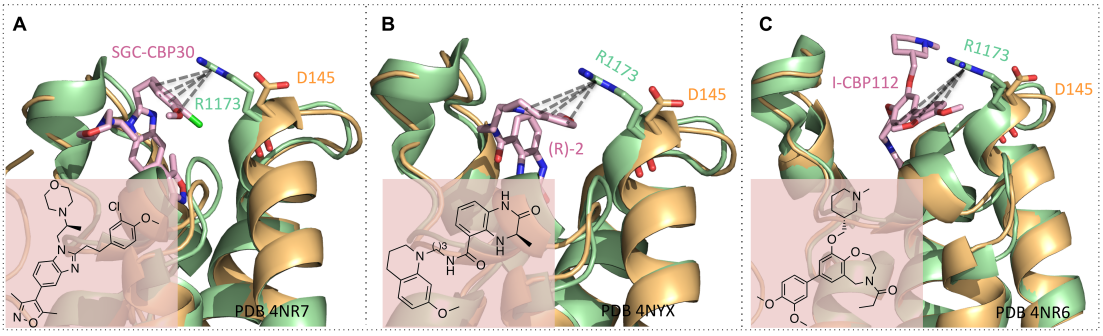


Figure 2.1: Electrostatics between a diverse set of ligands and CREBBP Arg1173 residue are potentially a driving force for selectivity of CREBBP bromodomains (green) over BRD4 (yellow, PDB: 3UVW) for a diverse set of inhibitors including: A) 5-isoxazolylbenzimidazole derivative (PDB: 4NR7), SGC-CBP30; B) dihydroquinoxalinone derivative (PDB: 4NYX), *(R)*-**2**; and C) benzoaxazepine-based inhibitor, I-CBP112 (PDB: 4NR6).

Having established the importance of R1173 for driving selectivity, the next step was to optimise interactions with this residue to design more potent CREBBP inhibitors. This Chapter will describe the computational techniques used to evaluate fifteen 5-isoxazolylbenzimidazole derivatives, referred to throughout as **L1-L15**. Attention was directed to the aromatic groups of these derivatives that were proposed to interact non-covalently to R1173. MD simulations of all CREBBP and inhibitor complexes were employed to study their behaviour and flexibility in a biological environment over time. Statistical models were generated by comparing MD simulation and QM calculation results against available data obtained using thermal shift assays. Thermal shift values (ΔT_m) are defined as changes in temperature required to denature a protein given a modification of conditions such as a mutation,

changes in pH, or the binding of a drug. When comparing binding strength among a series of ligands with the same protein, greater thermal stability, indicated by a higher melting point, is directly related to an increase in binding affinity.

This Chapter will also discuss how the best predictive model of binding affinity was successfully applied by the Brennan Group to facilitate the synthesis of four additional selective CREBBP inhibitors. Finally, a broader discussion of the importance of cation–π interactions among small-molecules inhibitors and charged receptor residues will be presented following a large-scale investigation of crystal structures in the Protein Data Bank (PDB), inspired by this work.

2.2 Methodology

MD simulation post-processing techniques: Poisson Boltzmann Surface Area (MM-PBSA), Generalized Born Surface Area (MM-GBSA), and electrostatic calculations will be discussed, as well as QM-based techniques such as evaluation of Electrostatic Potential Surface values (ESPs). Ligands **L1-L15** and corresponding experimental ΔT_m values[150] are shown in **Fig. 2.2**.

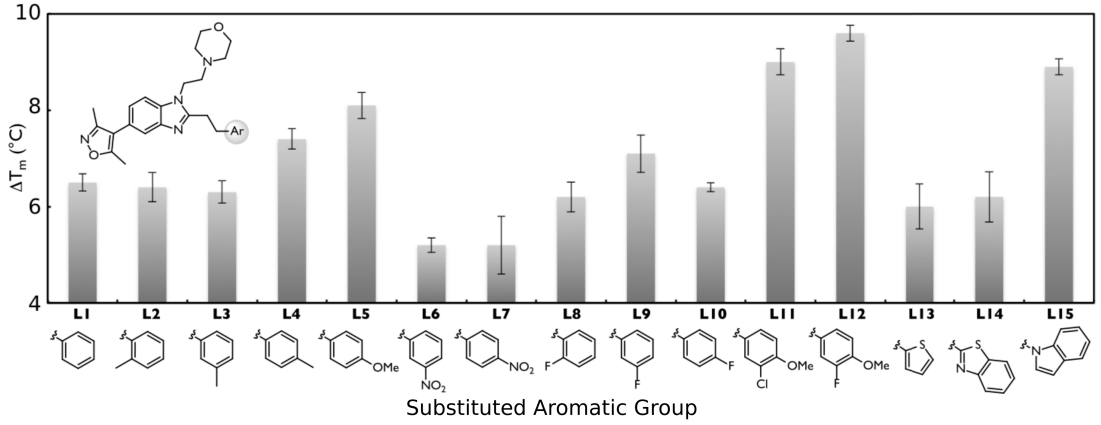


Figure 2.2: 5-isoxazolylbenzimidazole derivatives with variable aromatic (Ar) included in this study, referred to throughout as **L1-L15**, and corresponding CREBBP bromodomain binding assay ΔT_m data.[150]

2.2.1 Previous Computational Insights

In silico studies have proven to be valuable for elucidating key binding interactions for bromodomains. For example, while it was known that bromodomains are characterized by a bundle of four α-helices, two loops, and bind to histone acetylated lysine residues, Pizzitutti et al. were able to identify non-active site P371 and M372 residues to be necessary for binding following 5-10 ns MD simulations (**Fig. 2.3**).[97] The application of MM-PBSA and MM-GBSA to MD-derived ensembles has been previously performed for bromodomains with a diverse set of inhibitors and was similarly successful in calculating binding energies ($R^2 > 0.6$),[152,153] as well as ranking binding affinity of inhibitors within a set (Spearman rank-order correlation coefficient, $r_s > 0.6$).[154–157] Biggin et al. also used MD simulations and free energy perturbation (FEP) to compute absolute binding free energies of small-molecules bound to bromodomains with unprecedented accuracy of $\pm$ 1.0 kcal mol^{-1} of experimental values.[158]

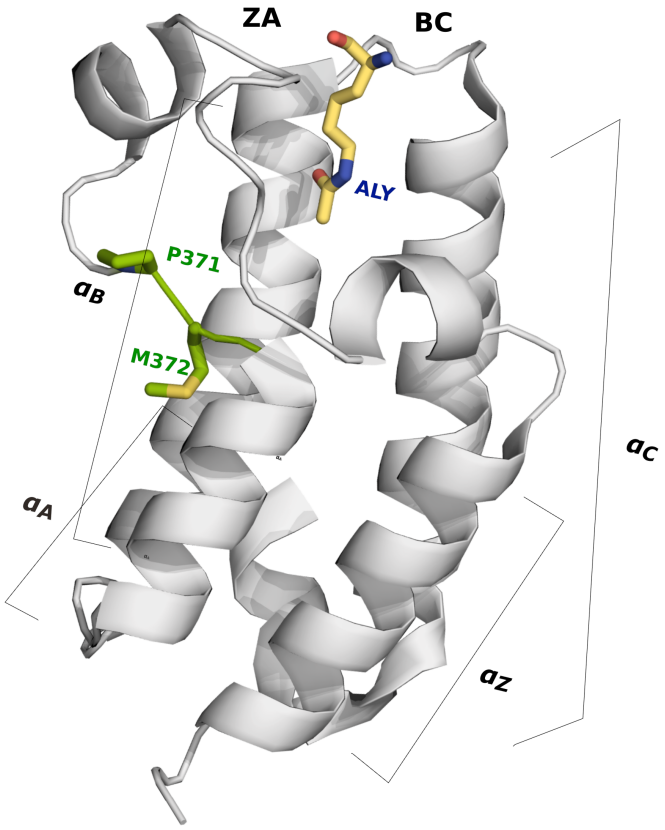


Figure 2.3: The bromodomain motif is composed of a left handed bundle of four α-helices (αZ, αA, αB, and αC) linked by two loops (ZA and BC). Its function is to recognise histone acetylated lysine residues (yellow). Non-active site residues P371 and M372 (green) are important for substrate binding.

2.2.2 Classical Methods

MD Simulation Parameters

All simulations were performed with the Amber12 package.[159] The RESP ESP charge Derive (RED) Server RESP-A1B methodology was used to determine ligand partial charges.[160] For this technique, B3LYP/cc-pVTZ molecular ESPs are computed involving the Connolly surface algorithm[161] with the integral equation formalism polarisable continuum solvent model (IEFPCM; $\epsilon = 4$, ether as solvent).[162] The force constant for dihedral angles was set to 3.6 kcal mol^{-1} to ensure planarity for cases in which the ligand contains methoxy or nitro substituents. The crystal structure PDB ID: 4NR7 was used as a template to create the CREBBP complexes.

Crystallographic waters were preserved and tLeap[159] was used to add hydrogens to the protein. To solvate and neutralize the system a 10 Å octahedron composed of ∼7100 water molecules and two Na^+ ions was created using the transferable intermolecular potential with 3 points (TIP3P) model.[163]

Preceding MD simulation, 1000 steps of steepest descent minimisation were performed while constraining heavy atoms then a second minimisation of 1000 steps for the entire system. This was followed by 1000 ps of equilibration and 5 ns of production. These were carried out using the NPT ensemble and Langevin thermostat at a constant temperature of 310 K. The SHAKE algorithm was used to constrain all bonds containing hydrogen atoms and allowed the use of a time step of 2 fs. The cut-off of 12 Å was used for long-range interactions. Analysis of the simulation required the use of Visual Molecular Dynamics (VMD) 1.9.2.[164]

The average root-mean-square-deviation (RMSD) values of Cα backbone atoms relative to crystal structures were calculated to evaluate the stability of the systems. Frequency calculations of cation–π interactions were performed by determining the distance between the centre of mass of the aromatic ring and the R1173 guanidinium ion over all frames. Instances in which the distance was 6 Å or shorter were defined as a cation–π event and recorded. This distance based cut-off was first established by Dougherty, showing that 99% of cation–π structures extracted from the PDB satisfying geometric and energetic-based criteria lie within this value.[165,166]

Classical Free Energy and Electrostatics Calculations

The MMPBSA.py script[167] distributed by AmberTools12[159] was used to determine binding free energies with both the MM-GBSA and MM-PBSA continuum solvation methods. These are examples of 'post-processing techniques' used to quantitatively analyse trajectories following completion of MD simulations. These methods have been used with increasing popularity to characterize protein–protein interactions and estimation of binding energies.[168] It was necessary to pursue both methods as their accuracy is system dependent, as established by previous studies where MM-GBSA was proven to perform better,[154,156,169,170] worse,[171,172] or the same[173–176] as MM-PBSA.

These implicit solvent models approximate solute-solvent interactions to calculate binding free energies, defined as the difference between the free energy of the complex

and the protein and ligand. The MM-PBSA approach is based on the Poisson Boltzmann equation, described in **Section 1.4.2**, and is considered a gold standard of continuum models but is computationally expensive. The electrostatics using the MM-GBSA approach are defined in **Eq. 2.1**:

$$\nabla G_{Born}^{GB} = (1 - \frac{1}{\epsilon}) \sum_{ij} \frac{q_i q_j}{\sqrt{(r_{ij}^2 \alpha_{ij}^2 exp(-D_{ij}))}} \tag{2.1}$$

where the summation runs over all pairs of atoms i and j, D_{ij}=$r_{ij}{}^2/2\alpha_{ij}{}^2$, α_i is the effective Born radius i, $q_i q_j$ are the electrostatic charges on particles i and j, and ϵ is the dielectric function. Parameters dictating calculation frequency must be defined. In this study, energy values were averaged over 100 frames, corresponding to snapshots taken at 50 ps intervals for 5 ns of simulation time. A salt concentration of 0.15 M was used to parallel physiological conditions. The interior (ϵ_{int}) and exterior (ϵ_{ext}) dielectric constants were set to 1 and 80, respectively. Default atomic radii were read from the original parameter files generated to run the MD production.

A third post-processing technique was used whereby the Coulombic electrostatic interactions between the aromatic groups of ligands **L1**–**L15** and arginine guanidinium were calculated at intervals of 10 ps over the total simulation time. This was accomplished using the NAMD Energy Plugin 1.4.[177] The average and standard deviation energy values were computed for each system and correlated against experimental ΔT_m values.

2.2.3 QM Methods

QM calculations were performed by Dr. Wilian Cortopassi to develop an optimal model to predict binding affinity for CREBBP inhibitors.

Electrostatic Surface Potential Calculations

DFT optimisations using the B3LYP/6-31G+(d) level of theory were performed using Gaussian09[178] on **L1-L15**, where the starting point geometry was taken from a representative crystal structure (PDB: 4NR7) shown in **Fig. 2.4**. ESP maps were generated with Spartan 14.[179]

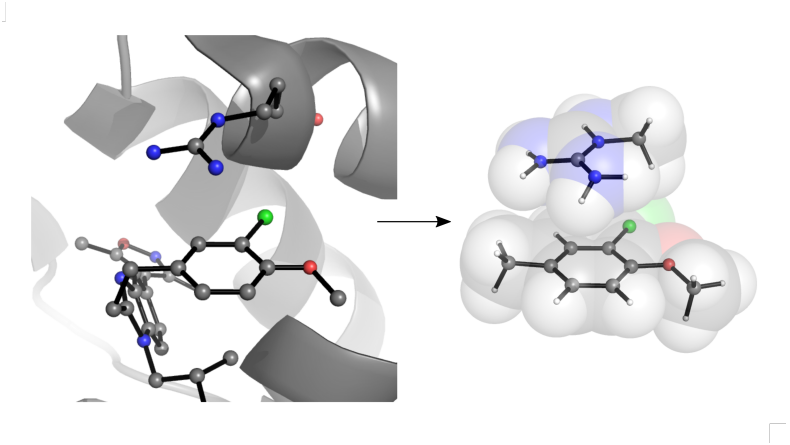


Figure 2.4: Visualisation of aromatic group and R1173 atoms included in QM calculations.

ESPs were defined as the potential experienced by a positive charge in different points around the ligand due to electron density as shown in **Eq. 2.2**,[180]

$$V(r) = \sum_{A} \frac{Z_A}{|r - R_A|} - \sum_{\mu v} P_{\mu v} \int \frac{\theta_\mu(r_1)\theta_v(r_1)}{|r - r_1|} dr_1 \tag{2.2}$$

where Z_A is the nuclear charge of atom A centred at position R_A, $P_{\mu v}$ is the density matrix, and the terms θ_μ and θ_v are the orbital basis functions.

Fig. 2.5 visualises the ESP calculation from a planar view and three dimensionally with reference to the surface space of the aromatic group. In order to mimic the non-covalent interaction, the ESP calculation was performed at 3.5 Å and 4.0 Å above the plane of the ligand (**Fig. 2.5A**), the approximate distance of R1173 to the chloromethoxybenzene of **L11** (PDB: 4NR7). This yielded similar results and ultimately a vertical distance of 3.5 Å was chosen. ESPs were calculated using B3LYP/6-31G(d,p) in diethylether.

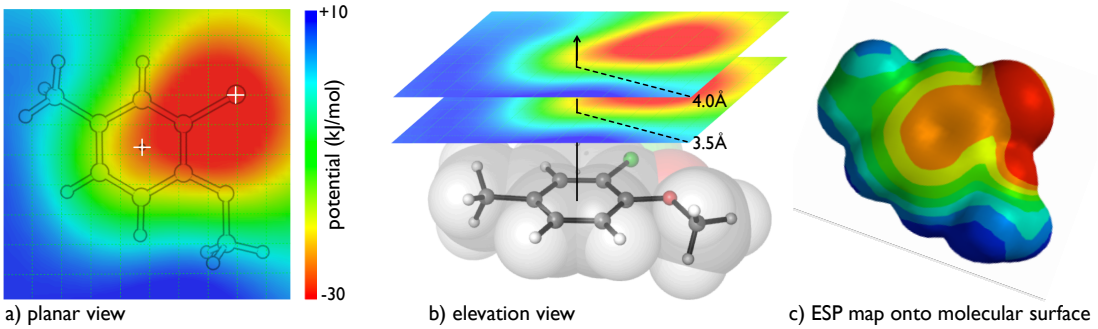


Figure 2.5: Visualisation of ESP method with a) the planar view looking down onto the molecule in 2D, b) the plane 3.5 Å and 4.0 Å above the molecule, and c) the ESP surface mapped onto the molecule.

2.3 Results and Discussion

2.3.1 Proving Relevance of Cation–π Interaction

Before proceeding to develop computational models describing a specific R1173 CREBBP interaction with the aromatic group of inhibitors **L1-L15**, its relevancy in the bound systems needed to be established. Two crystal structures of CREBBP in complex with inhibitors **L11** (PDB: 4NR7) and **L1** (PDB: 4NR5) were available for analysis. Upon inspection of the CREBBP:**L11** system, the chloromethoxyphenyl ring is observed to be in close contact with R1173 and satisfies the geometric criteria for a cation–π interaction (**Fig. 2.6A**). In contrast, the crystal structure of CREBBP:**L1** did not show a similar interaction between the ligand and residue (**Fig. 2.6A**). To investigate this beyond examination of static X-ray structures, long-time MD simulations of 100 ns were performed to further elucidate the dynamics of the protein-ligand interactions.

For the CREBBP:**L11** system the intermolecular distances between the guanidinium group of R1173 and the chloromethoxyphenyl group indicated a cation–π interaction was maintained for over 98% of the 100 ns simulation time (**Fig. 2.6C**). This substantiated the initial hypothesis that this interaction was important for ligand binding and stability. Interestingly, the same interaction was observed for a majority of simulation time for the CREBBP:**L1** system, despite that the crystal structure shows the phenyl group is oriented away from the guanidinium

group. Spontaneous formation occurs at time 2 ns and is maintained for 74% of the simulation (**Fig. 2.6C**). Another important metric is evaluation of RMSD, which indicates the stability of the entire system. The low average RMSD of 1.6 Å for CREBBP:**L11** and 1.8 Å for CREBBP:**L1**, with respect to the starting crystal structure, provides confidence in the legitimacy of the simulations (**Fig. 2.6B**).

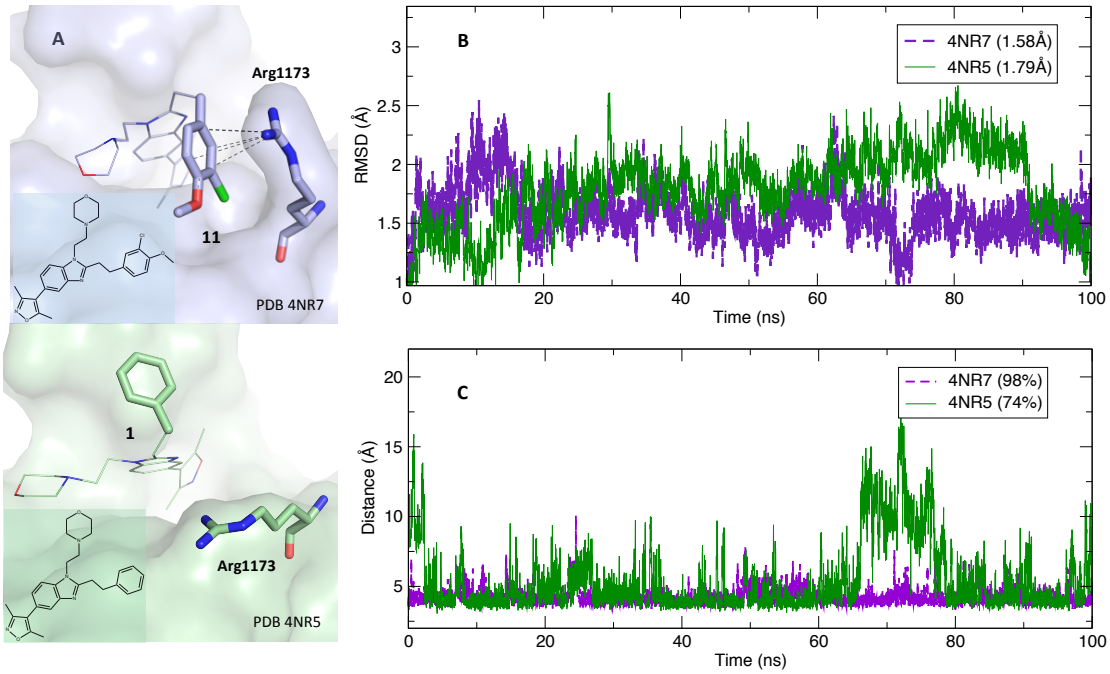


Figure 2.6: A) The CREBBP-**L1** complex active site shows no initial cation–π interaction (PDB: 4NR5) in the crystal structure, while the CREBBP-**L11** complex has a clear cation–π interaction between the ligand and the arginine. B) Following 100 ns of MD simulation, the two complexes are stable, as shown by their average RMSD values and C) both form active site cation–π interactions for a majority of simulation time, using a distance cut-off of 6.0 Å between the guanidinium and the unsaturated chain of the ligands.

These results indicated the cation–π interaction is important for a strong binder **L11** but also an inhibitor with weaker binding affinity such as **L1**, where this interaction was still observed for a majority of time despite a higher degree of flexibility. Therefore, further computational calculations were performed to quantify this interaction for the other 13 inhibitors bound to CREBBP.

2.3.2 MD Results for Full Inhibitor Series

Simulations of the remaining 13 inhibitors bound to CREBBP were performed for 5 ns. This was necessary to further establish relevancy of the cation–π interaction with all the 5-isoxazolylbenzimidazole ligands. This would substantiate that small model time-independent QM calculations performed on these systems were appropriate. Furthermore, because techniques such as MM-PBSA and MM-GBSA are post-processing methods, this necessitates simulation trajectories of the systems.

The cation–π distance and RMSD were similarly analysed for all 15 simulations, shown in **Fig. 2.7** and **Fig. 2.8**. Results indicate a strong interaction for all ligands. The most stable systems defined by simulations with the lowest RMSDs were **L7**, **L14**, and **L15**.

2.3.3 MM-GB/PBSA Results

Following 5 ns simulation of the 15 CREBBP complexes in water, binding free energies between arginine and aromatic group were analysed (**Fig. 2.9**). In parallel with experimental data, the strongest binding ligands in experiment have more negative binding free energies, and higher values for the weaker binding ligands (**Table A.2**). The quantitative ability to successfully rank ligands according to ΔT_m data did not vary in regards to the implicit solvation model (PB versus GB). MM-PBSA performed the best with $r_s = 0.59$ followed by MM-GBSA with $r_s = 0.53$. Ligands **L5**, **L11**, **L12** and **L15** are ranked between the top six strongest binders according to the MM-PBSA approach, in line with the thermal shift data. It is important to note the large error associated with both the MM-GBSA and MM-PBSA techniques that preclude the ability to distinguish between inhibitors that bind neither weakly nor strongly, e.g. those with ΔT_m values between 6.2 - 7.4 °C. We also analysed the electrostatic contribution to the interaction energy and frequency of observed cation–π interactions. Coulombic electrostatic energies were able to successfully rank the ligands when compared to experimental binding affinities ($r_s = 0.72$, **Fig. 2.10**). This is consistent with our QM studies showing that these interactions are electrostatically-dominated. However, it is important to also note the large error values associated with these calculations employed to analyse the given trajectories. As in the case with the MM-PBSA and -GBSA

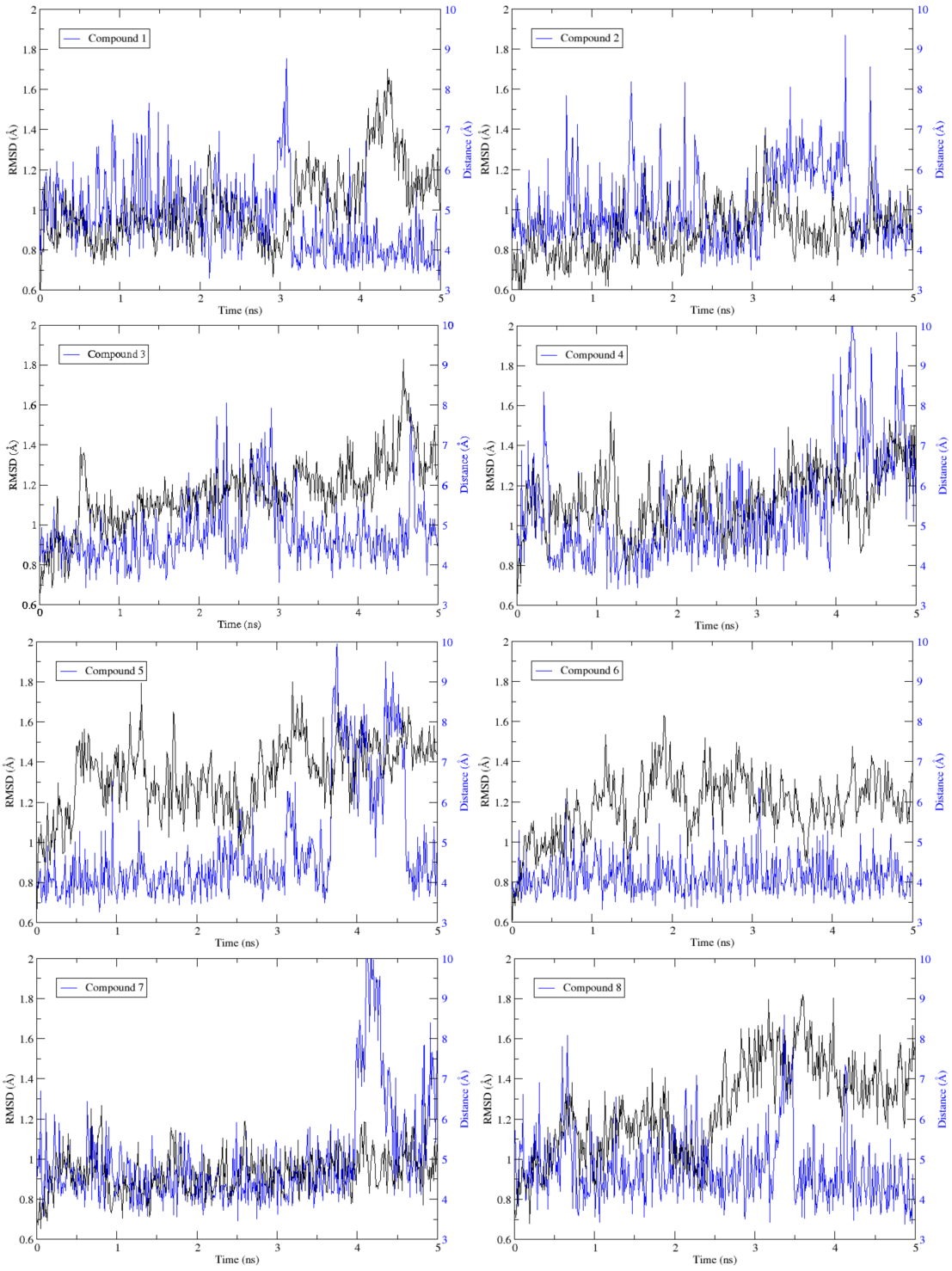


Figure 2.7: RMSD of CREBBP-ligand complexes (black) and distances between the substituted benzene of Ligands **L1-L8** and CREBBP R1173 (blue) over 5 ns of MD simulation.

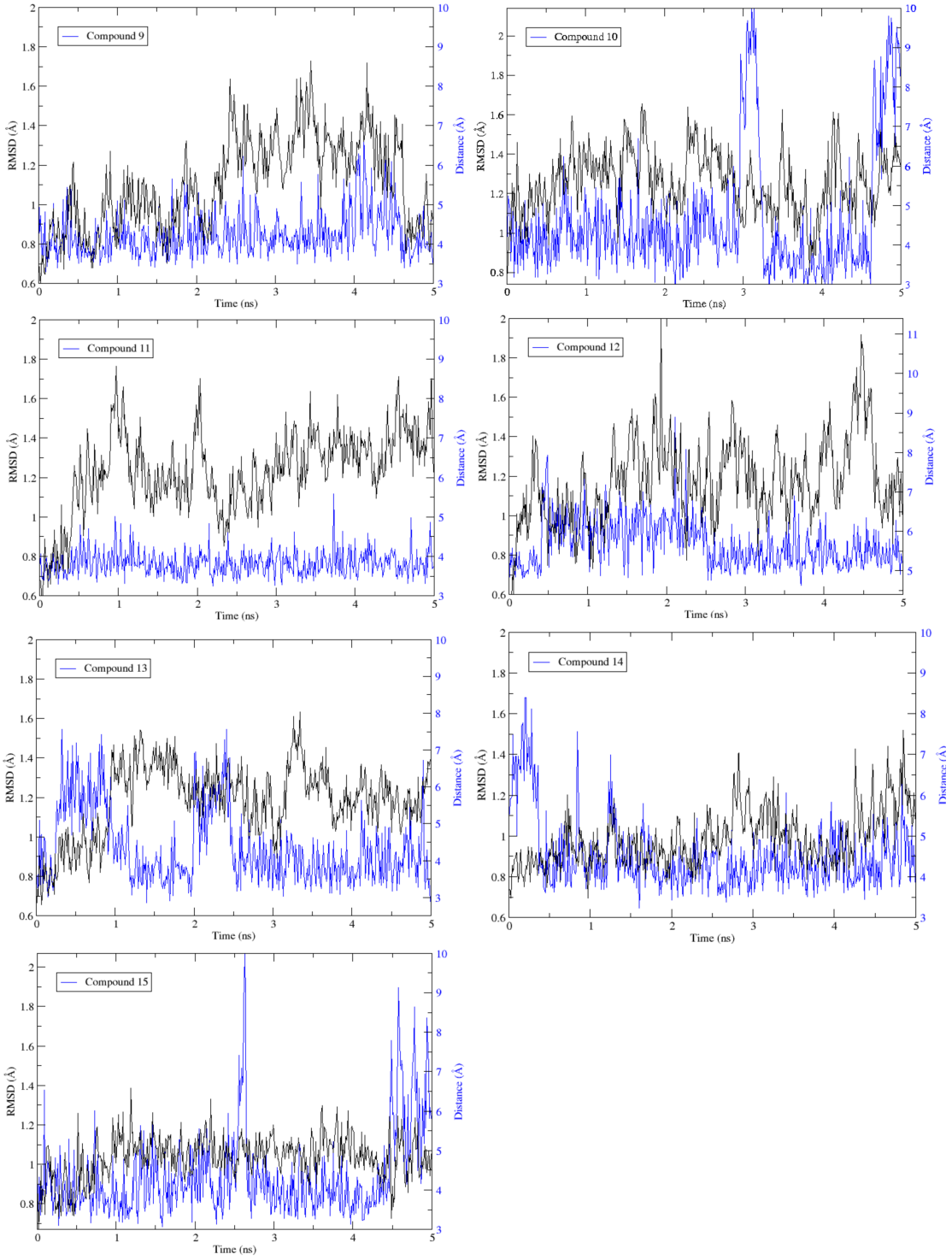


Figure 2.8: RMSD of CREBBP-ligand complexes (black) and distances between the substituted benzene of Ligands **L9-L15** and CREBBP R1173 (blue) over 5 ns of MD simulation.

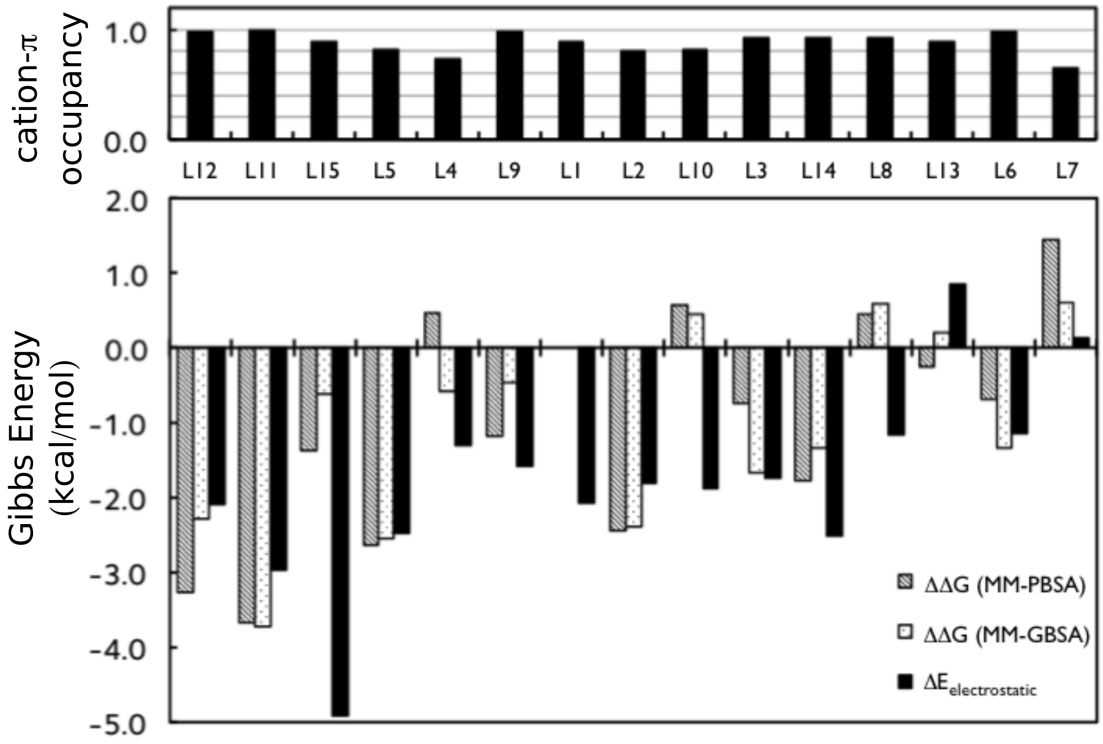


Figure 2.9: Occupancy of a cation–π interaction between ligands **L1-L15** and R1173, binding free energies, and electrostatics results following 5 ns of MD simulations. Ligands are ordered in decreasing experimental binding affinity.

calculations, this method is not able to distinguish among inhibitors that are neither the strongest nor most weakly binding. Rather, these quick calculations are able to broadly categorize the 15 ligands into three groups. In addition to the through-space electrostatic contributions associated with ring substituents identified in QM and ESP studies, these simulations highlight additional close-contacts of ring-substituents with the arginine (alkyl) side-chain. All the analysed ligands showed the existence of a close contact cation–π interaction for more than half of the simulation, shown in **Fig. 2.9**. **L6**, **L9**, **L11** and **L12** had occupancy values close to 1. The formation of this interaction for almost 100% of the simulated time does not necessarily mean strongest guanidinium–benzene interactions. All these compounds have heteroatoms in the meta-position, favouring –CH interactions with the alkyl chain of the arginine residue. **L7**, containing the inductively and mesomerically electron-withdrawing nitro group in the para-position, showed the

lowest occupancy (0.65), as expected from a weaker interaction with the guanidinium. This is consistent with our QM computations of ESP and interaction energies and shows that weakening the cation–π interaction may also result in different structural conformations of the protein in solution.

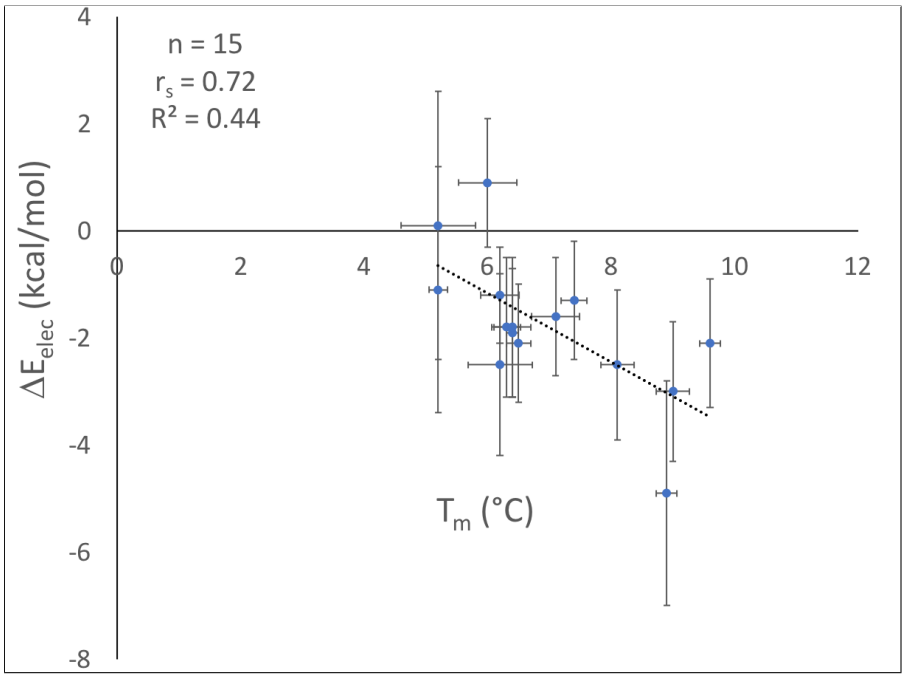


Figure 2.10: Plot of average Coulombic electrostatic energies against thermal shift data for aromatic group of ligands **L1**-**L15** over 5 ns of MD simulations.

2.3.4 ESP Results

The ESP model was the best predictor of binding affinity of all the computational methods utilised. Qualitatively, the basis of this model can be observed in **Fig. 2.11** where the ESP 3.5 Å above the surface of the molecule has been mapped and ordered by activity. From analysing the ESP maps of each ligand, **L15** containing an indole group appears to have the most negative ESP. It is important to note that the ESP depends on the electron density of the entire molecule, rather than from local areas.[167] For the least active ligands **L6** and **L7** the nitro group appear to have a very negative ESP, however the presence of these substituents cause a strong electron withdrawing effect as seen by the higher ESP above the centroid of the

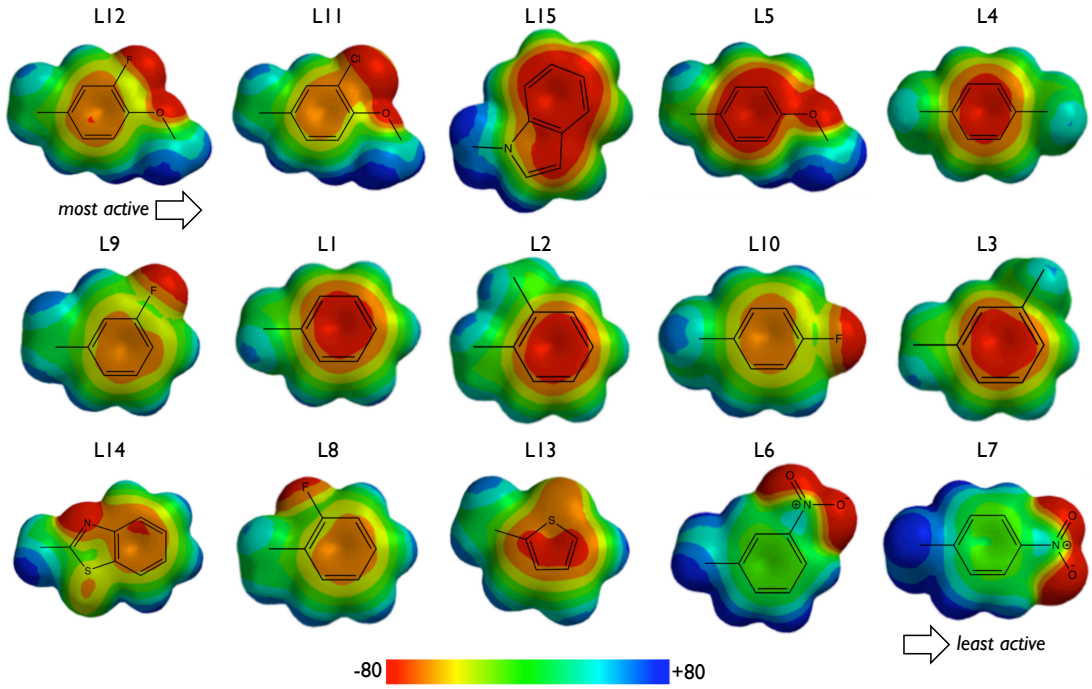


Figure 2.11: Electrostatic Potential Surfaces of the modified aromatic chains of **L1**-**L15** were computed using B3LYP/6-31G(d,p) level of theory in diethylether, showing areas of positive (red) and negative (blue) potential. The molecules are ordered from most active (highest ΔT_m) to least active (lowest ΔT_m). The same isosurface value 0.002 is used throughout and the same colour scale spans -80 to 80 kJ/mol.

aromatic ring. In contrast, the electron donating methoxy group in **L5** results in a negative ESP above both the aromatic ring and oxygen of the methoxy substituent.

The through-space electrostatic effect as first proposed by Wheeler and Houk[181] where substituents of the aromatic group can interact with the cation is especially relevant in these systems that have a cationic molecule rather than a single atom. This is considered when examining ligand **L12** that contains an electron withdrawing fluorine atom and yet was still shown experimentally to be the strongest binder. This is explained by considering an independent interaction of the substituent with the cation via through space electrostatics.[181]

When compared to benzene (**L1**), heterocycles thiazole (**L13**) and benzothiazole (**L14**) have higher (less negative) ESP values above their surfaces. This parallels the experimental findings where **L1** had a slightly higher ΔT_m value than the other

two ligands, further demonstrating that qualitative analysis of the ESP is valuable for understanding the overall strength of the cation–π interaction.

In general, ligands with a very negative ESP above the π-system (**L15**) or both the π-system and substituent (**L5**, **L11**, and **L12**) corresponded to favourable high ΔT_m values. A quantitative bivariate model to predict ΔT_m as a function of ESP above the aromatic ring and substituent was then generated by Dr. Wilian Cortopassi shown in **Eq. 2.3**.

$$\Delta T_m = -307.0 * ESP_{ring} - 238.74 * ESP_{substituent} + 2.58 \tag{2.3}$$

The final linear regression model had a strong correlation ($R^2 = 0.84$) to ΔT_m values and was able to successfully rank the data set ($r_s = 0.91$) with no clear outliers (RMSE = 0.52 °C).

2.3.5 Application of ESP Model

Because the ESP model, defined in **Eq. 2.3**, was the best predictor for binding affinity of all the computational approaches, it was then applied to a series of newly synthesised 7-membered ring DHQ derivatives by Dr. Michael Brand, working with Prof. Stuart Conway. Predicted ΔT_m values were compared against IC_{50} data for four molecules (**Fig. 2.12**). Interestingly, **L19** (IC_{50} = 1.0 μM) containing an electron withdrawing acetamide group and **L17** (IC_{50} = 1.6 μM) that has an electron donating methoxy group resulted in similar levels of binding affinity. When quantitatively examining the ESP surfaces, expectedly **L17** has a more negative ESP above the aromatic ring than **L19**, however it is also observed that the amide oxygen of **L19** is likely stabilising the electrostatic interaction with the R1173 guanidinium.

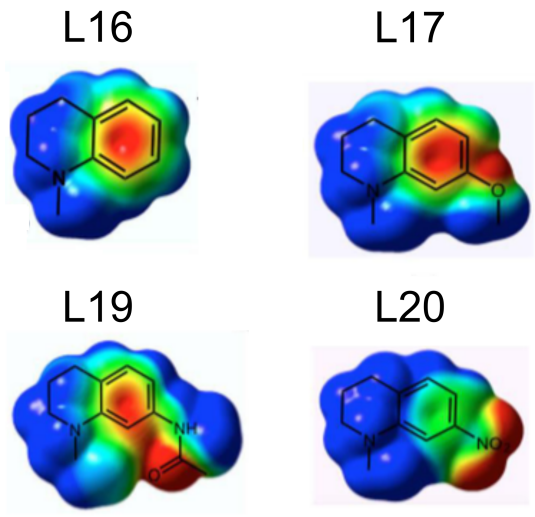


Figure 2.12: Ligands synthesized following application of theoretical bivariate ESP model. IC_{50} data: **L16** = 2.5 μM, **L17** = 1.6 μM, **L19** = 1.0 μM, **L20** = 3.1 μM

2.4 Large-scale Analysis of Cation–π Interactions

The potency and selectivity achieved by this conserved cation–π interaction between the inhibitors and CREBBP inspired a desire to further elucidate similarly powerful, but undescribed, interactions through a large-scale analysis of the PDB.

Previous studies have obtained valuable information concerning key cation–π interactions through analysis of a database containing crystallographic data. Specifically, past studies have revealed information concerning the occurrence and geometric characteristics of such interactions in biological systems, including protein–DNA complexes,[182] protein–protein interactions,[165,166,183–185] metal cation–π interactions,[186] and cation–π–cation interactions[187] using similar approaches. However, a large-scale analysis of aromatic ligand-protein cation–π interactions, such as the one driving selectivity between the 5-isoxazolylbenzimidazole inhibitors and CREBBP reader proteins and those described below in **Fig. 2.13**, had not yet been attempted. Furthermore, the applications of this study could provide insights for drug discovery; a study by Pitt et al. in 2009 found that over 40% of small molecule drugs launched in clinical trials or under development contained a (hetero)-aromatic ring.[188]

This section will discuss a large-scale analysis of available crystal structures and small model calculations undertaken to reveal the properties and characteristics of aromatic small-molecules and charged amino acid side chains: arginine (Arg), lysine (Lys), and histidine (His) interactions.

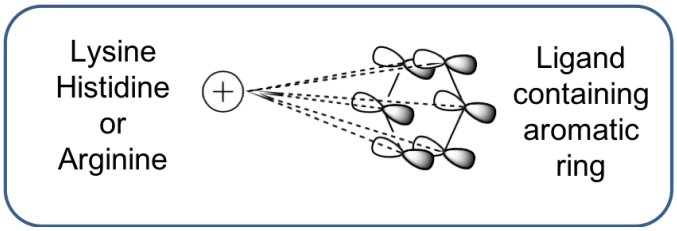


Figure 2.13: Description of specific cation–π interactions sought from the PDB.

Kumar K. and Woo S.M., et al. Cation–π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **2018**, *9(10)*, 2655-2665.

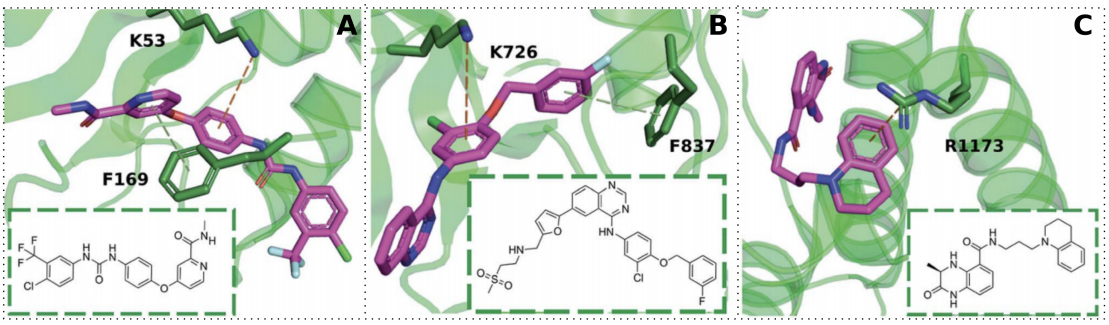


Figure 2.14: Aromatic ligands forming cation–π interactions with the target protein: (A) sorafenib complexed with Human p38 MAP kinase, (B) lapatinib in complex with ErbB4 kinase, (C) and a dihydroquinoxalinone CREBBP bromodomain inhibitor.)

2.4.1 Parsing Crystallographic Data

An empirical distribution of protein–ligand cation–π interactions was sought from the PDB.[189] The following workflow shown in **Table 2.1** was implemented in Python to select a high-resolution, non-redundant dataset.

Table 2.1: Stages and filtration criteria for obtaining a dataset of PDB structures that contain a small-molecule which forms a cation-π interaction with a lysine, arginine, and/or histidine residue of the receptor protein.

			Residues			
Stage	Filter	PDBs	His	Arg	Lys	Total
A	2.0 A resolution, containing ligand	30,053				30,053
B	Geometric Thresholds	3,248	4,959	3,141	930	9,030
C	Residue pKa >7.4	1,827	68	2,954	848	3,870
D	Non-redundant chain and polycyclic ligand	1,827	49	1,381	582	2,012

An initial search returned 30,053 X-ray crystal structures with resolution 2.0 Å or better and containing unique ligands (**Table 2.1**, stage A). Following an implementation of geometric rule-based criteria, 3,248 structures were found to contain 9,030 active sites with a Lys/Arg/His residue in proximity to an aromatic ligand (**Table 2.1**, stage B). Of the complexes identified, 4,954 involved histidine, however, these results also contained π–π interactions involving a neutral residue. The protonation state of each of the residues at physiological pH was computed with PROPKA (**Table 2.1**, stage C), drastically reducing the number of interactions involving histidine acting as a cation. In contrast, Arg-ligand and Lys-ligand

complexes retained 94% and 92% of their interactions from the previous stage. Finally, 40% of duplicate interactions were removed to yield a final data set of 2,012 interactions dominated by Arg (69%), Lys (29%), and to a lesser extent His (2%) residues (**Table 2.1**, stage D).

Finally, the prevalence of protein–ligand cation–π interactions could be evaluated in a non-redundant set of PDB structures. **Fig. 2.15** demonstrates the variety of complexes observed.

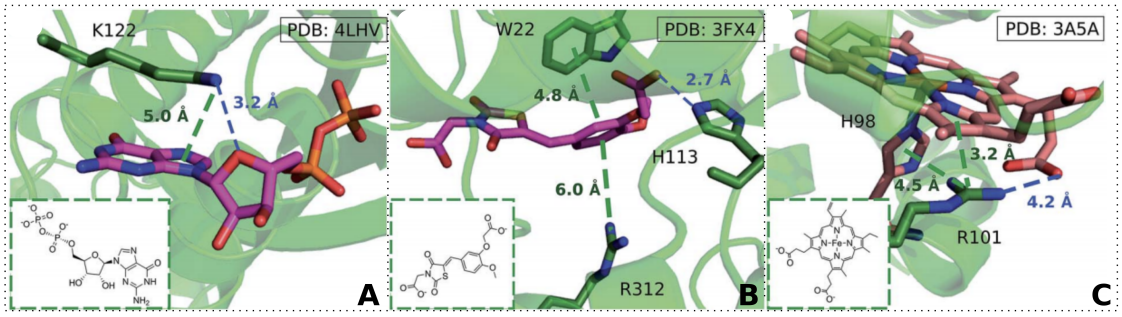


Figure 2.15: (A) GDP/GTP Lys-aromatic binding site selected from ammonium-PES $R_{x,y}$ R_z (2.0, 4.0) (ligand ID GDP). Examples of Arg-aromatic cation–π (B) long distance (ligand ID FX4) and (C) short distance bond (ligand ID HEM)

2.4.2 Small Model Profiling

When combined with profiling the energetics of the complexes using QM, the empirical and theoretical behaviour of these interactions could be explored. Structures of the small model ammonium, guanidinium, and imidazolium cations complexed with benzene were used to represent Lys, Arg, and His, shown in **Fig. 2.16**. High level DLPNO-CCSD(T) calculations were employed to analyse the distance and angular dependence of the calculations and then compared to the empirical data obtained from the PDB. In **Fig. 2.17**, the frequency of observed crystal structure complexes and the corresponding small model energetics, given a distance R_z in Å, are compared. For Arg and His complexes, the empirical data and theoretical data are in good agreement, with the highest distribution of complexes reflecting the minimum energy well. However, for interactions involving Lys the distance R_z at which the small model is most energetically favourable does not correspond with highest number of complexes observed empirically. Furthermore, interactions

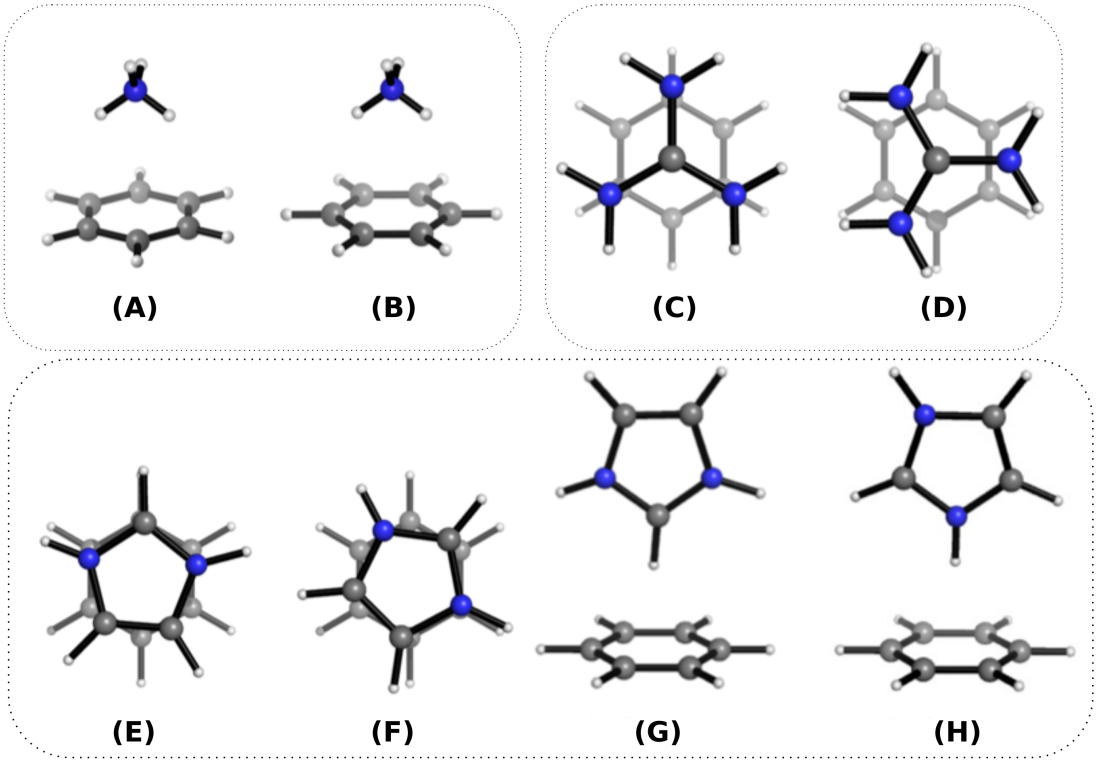


Figure 2.16: Cation–π complexes analysed in this work. Models of (A)/(B) lysine; (C)/(D), arginine; and (E)–(H), histidine side chains interacting with benzene.

involving Lys were the most weakened with the surrounding medium, decreasing to 19% and 7% of its gas-phase strength in ether and water models, respectively.

Results from symmetry adapted perturbation theory (SAPT) calculations were able to shed light on this phenomenon. SAPT provides a rigorous partitioning of the intermonomer interaction energy (E_{int}) into various physical contributions. These include short-range exchange–repulsion (E_{ee}), electrostatics (E_{ele}, e.g. charge–charge, charge–dipole, dipole–dipole, etc.), induction polarisation (E_{ind}, e.g. dipole/induced-dipole) and London dispersion forces (E_{disp}, e.g. instantaneous dipole/induced dipole). Of all the complexes, (A) and (B) representing Lys were the most dominated by electrostatics, while the other complexes show a mix of favourable electrostatics and dispersion(**Fig. 2.18**). When introducing a dielectric medium (4.2 = diethylether, 78.4 = water) that mimics the protein environment, Lys cation–π complexes become less favourable than Arg-arene, despite being 11.5 kcal mol^{-1} stronger in the gas-phase due to its electrostatically-dominated nature. In contrast,

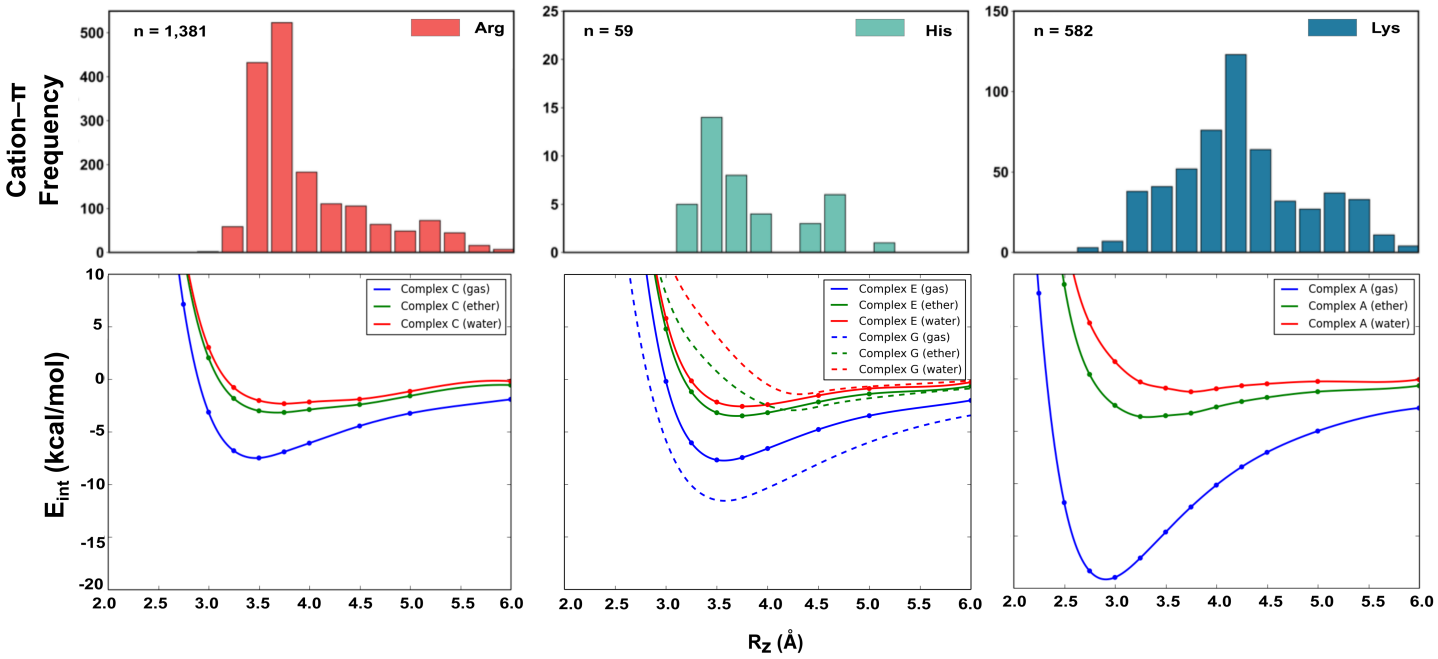


Figure 2.17: Normalized distance dependence of empirical interactions (bars) compared against DLPNO-CCSD(T)/aug-cc-pVTZ computed potential energy curves (kcal mol^{-1}) in the gas phase and with a dielectric constant of 4.2 (diethylether) and 78.4 (water). Solvation corrections were computed at the CPCM-MP2/cc-pVTZ level of theory.

the mixed-nature of Arg:arene interactions means it is much less affected by the surrounding medium. His:arene interactions also parallel this finding, where T-shaped complexes are weakened to 21% of the gas-phase value for $\epsilon = 4.2$ (3.0 kcal mol^{-1}) and by 9% for $\epsilon = 78.4$ (1.3 kcal mol^{-1}). For stacked histidine the reduction is smaller, to 44% (3.4 kcal mol^{-1}) and 32% (2.5 kcal mol^{-1}).

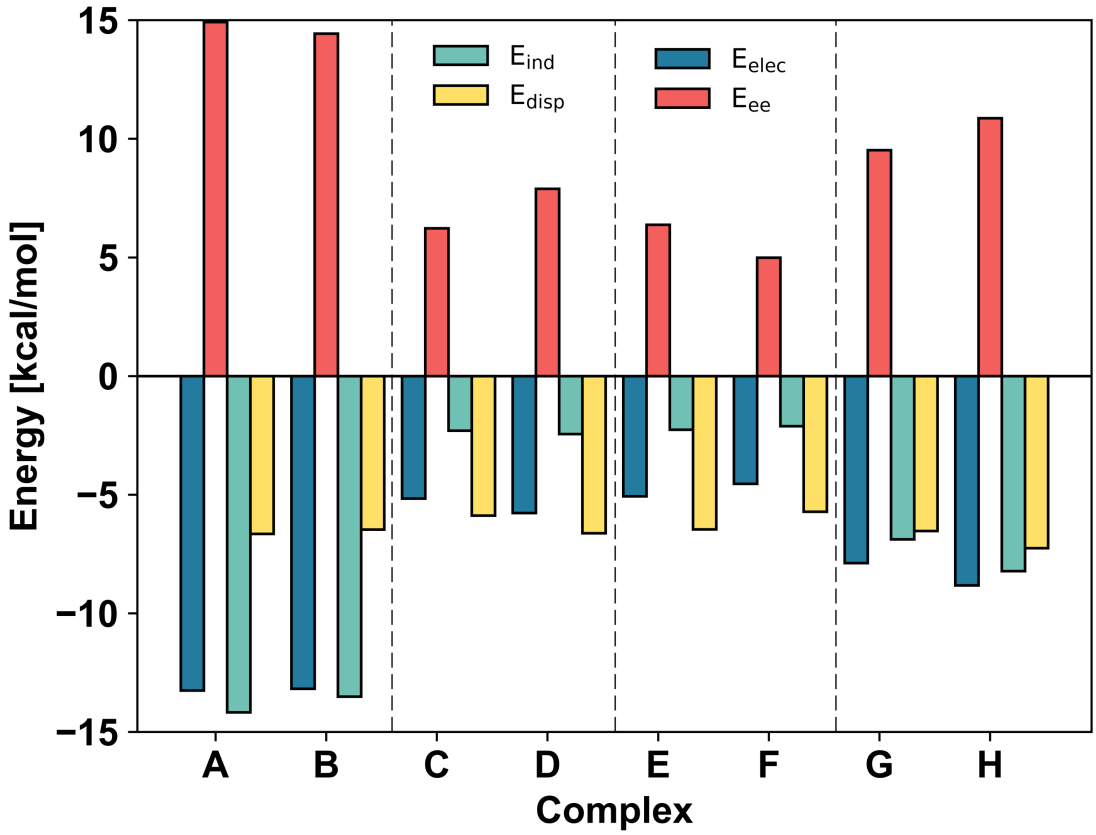


Figure 2.18: SAPT2+3/aug-cc-pVDZ decomposition of the interaction energy (in kcal/mol) into exchange repulsion (E_{ee}, red), electrostatics (E_{elec}, blue), induction-polarisation (E_{ind}, green) and London dispersion (E_{disp}, yellow) at the equilibrium separation.

2.5 Conclusion

First the occurrence of a cation–π interaction between an epigenetic 'reader' protein, CREBBP bromodomain, and fifteen small-molecule inhibitors using MD simulations was validated. Post-processing methods determined the binding free energies of these ligands and the average strength of the cation–π interaction with CREBBP R1173 to develop models capable of predicting binding affinity. MD simulation results were also used to validate the relevance of DFT-based models that accurately correlated ESP strength with experimental ΔT_m values.

Interest in this selective non-covalent Arg-arene interaction motivated a large-scale characterization of all similar interactions in the PDB that had not yet been

reported. Quantifying the properties of Lys/Arg/His and arene complexes provided valuable insight into drug design. Optimising Lys:small molecule interactions may prove difficult due to factors such as competitive hydrogen bonding interactions and solvent accessibility, which may also affect the cation–π interaction geometry. There were also significantly fewer His:arene interactions found in the PDB, although the stacked π–π^+ was calculated to be of similar strength to those involving Arg. This conclusion could signify that optimising unique His:small-molecule interactions could drive selectivity in other structurally similar protein families due to its scarcity, but strong interaction strength.

Together, these investigations are valuable for ligand optimisation as interest in developing potent and selective epigenetic inhibitors continues to increase.

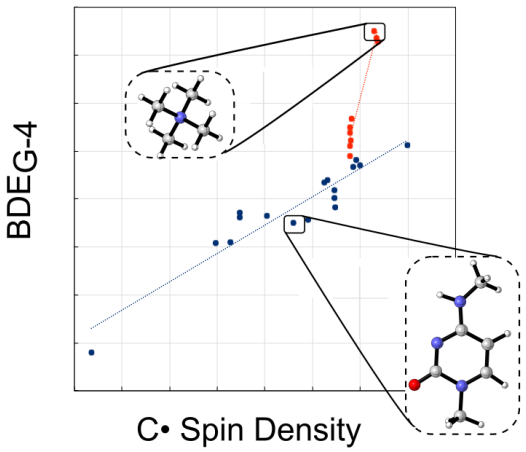
Our world is built on biology and once we begin to understand it, it then becomes a technology.

— Ryan Bethencourt

3

Fe(II)-αKG Demethylase Substrates; capturing the enthalpic landscape

Contents



Kumar K., Tan J., Paton R. S. Computational insights on bond dissociation enthalpies for

3.1 Introduction: Fe(II)-αKG protein families and substrates

In the previous chapter, the importance of elucidating acetyllysine recognition mechanisms by bromodomain readers was emphasised. Epigenetic erasers of PTMs (e.g. demethylases) also mediate the activation and repression of gene expression associated with the addition of PTMs by writer proteins. Although necessary for cellular development and *homeostasis*, regulation of demethylation can be successfully manipulated by cancerous cells to aid in tumour progression and metastasis.[190] Therefore, elucidating the inhibition of epigenetic erasers, such as non-heme Fe(II)/α-ketoglutarate dependent dioxygenases (Fe(II)/α-KG) demethylases, has also proved promising for the design of anti-cancer drugs.[191]

The Fe(II)/α-KG superfamily contains many protein subfamilies responsible for epigenetic demethylation of a diverse set of post-translationally modified substrates, shown in **Fig. 3.1**. N-methylated DNA nucleobases such as 3-methylthymine (3meT), 3-methylcytosine (3meC), N4-methylcytosine (4meC), N6-methyladenine (6meA), 1-methyladenine (1meA), 1-methylguanine (1meG), and N2-methylguanine (2meG) are demethylated by AlkB, a subfamily of DNA repair enzymes.[192] 5-methylcytosine (5meC) is repaired by Ten-Eleven Translocation (TET) proteins[193,194] in successive steps, via 5-hydroxymethylcytosine (5hmC), and 5-formylcytosine (5fC). Thymine Dioxygenase (THase) enzymes demethylate unmodified thymine (Thy) residues as part of the Thymidine Salvage Pathway, similarly to TET via 5-hydroxymethyluracil (5hmU), and 5-formyluracil (5fU).[195] In addition to nucleobases, arginine (Arg) and lysine (Lys) residues located on the N-terminal tails of histone proteins, are mono-, di- and in the case of lysine, tri-methylated. The Jumonji C domain (JMJD) proteins are able to demethylate all states of methylated Lys and some studies have found they also possess Arg demethylation functionality, although this is controversial.[196–200] Finally, asparagine (Asn) and proline (Pro) Hypoxia-Inducible Factor (HIF) residues are modified with the addition of an epigenetic hydroxyl marker for cellular survival during periods of low oxygen concentration by Prolyl Hydroxylase Domain (PHD) and Factor Inhibiting Hypoxia-Inducible Factor (FIH) proteins, respectively.[201,202]

C-H abstraction by Fe(II)/α-ketoglutarate enzymes, **2018** (*In preparation*).

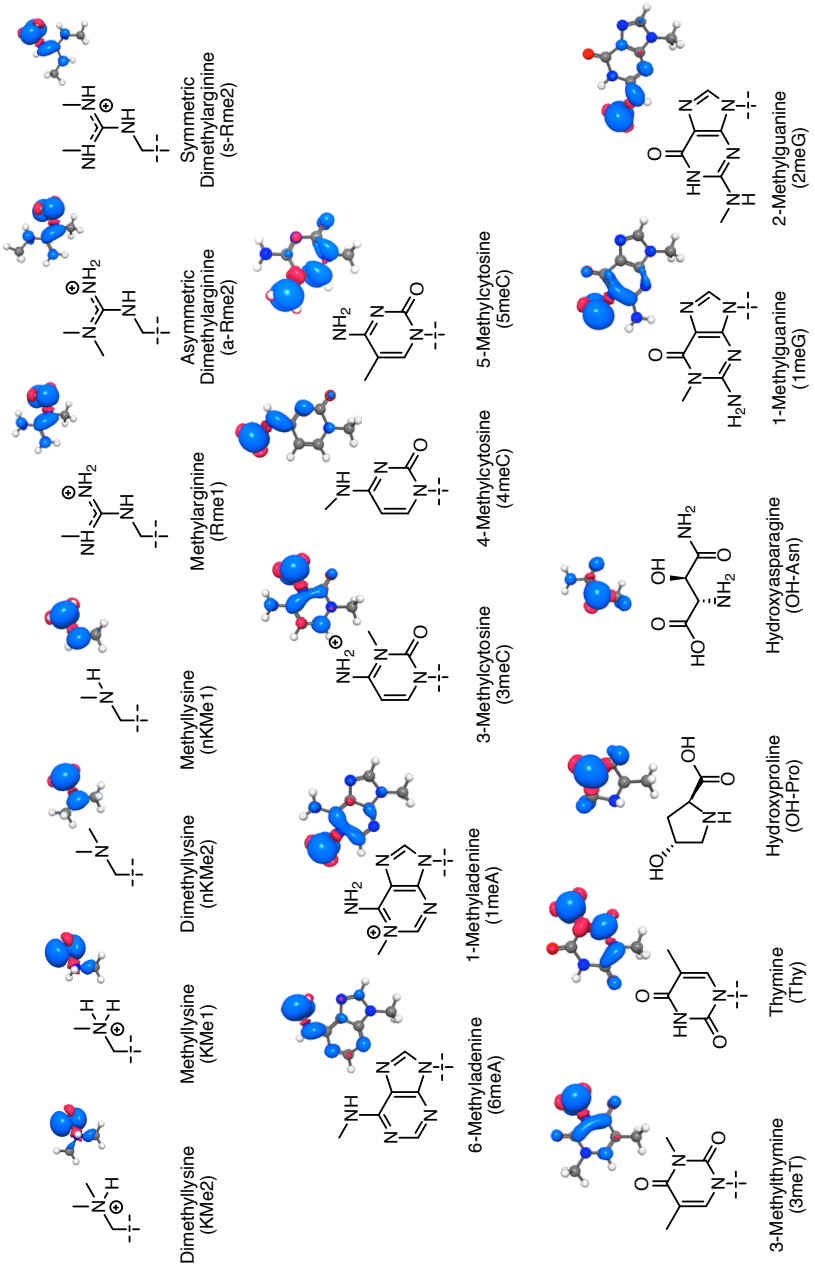


Figure 3.1: Fe(II)-αKG substrates and corresponding spin density plots following H-abstraction.

The Fe(II)/α-KG active site is characterized by an iron centre bound by two histidines and either an aspartic or glutamic acid with three bound waters.[203,204] Bidente binding of α-ketoglutarate occurs such that the keto and carboxylate group are opposite the aspartic/glutamic acid and histidine, respectively, displacing two of the water molecules. The presence of the primary substrate displaces a third water molecule and allows for O_2 binding. A high valent Fe(IV)-oxo species is formed by the attack of O_2 on α-ketoglutarate creating a bicyclic Fe(IV)-peroxohemiketal intermediate and subsequent oxidative decarboxylation.[203] The Fe(IV)-oxo performs a C−H abstraction from the substrate, yielding the hydroxylated intermediate via radical rebound,[205] shown in **Fig. 3.2**. For TET and THase enzymes, hydroxylation via H-abstraction of 5meC and Thy occurs three times in succession, shown in **Fig. 3.3**.

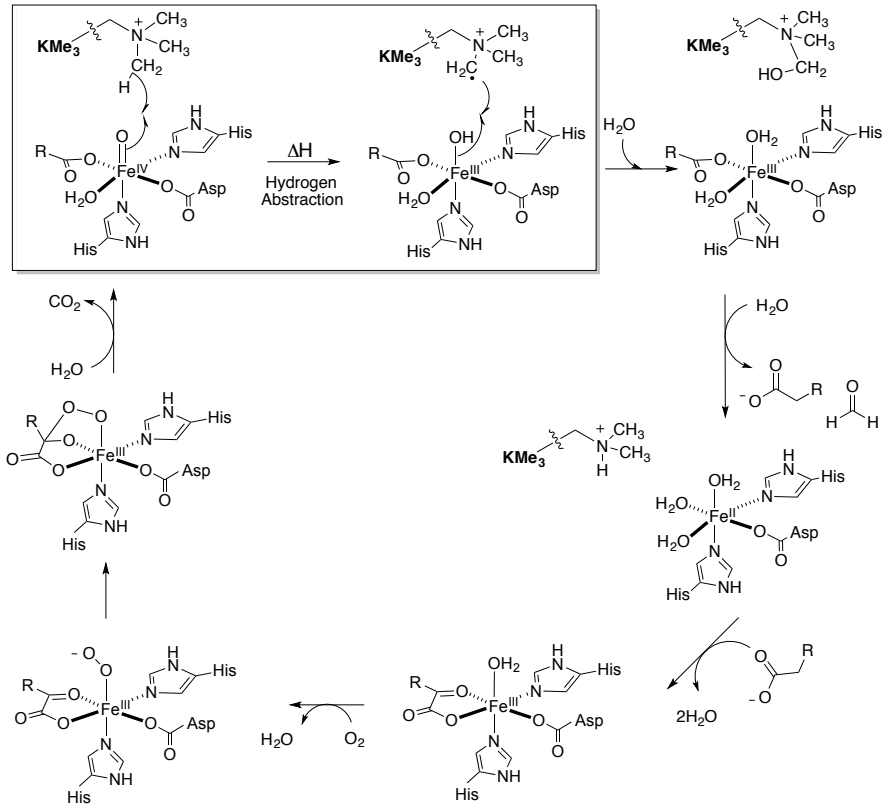


Figure 3.2: Proposed demethylation mechanism of trimethyllysine (Kme3) by Fe(II)-αKG protein JMJC.

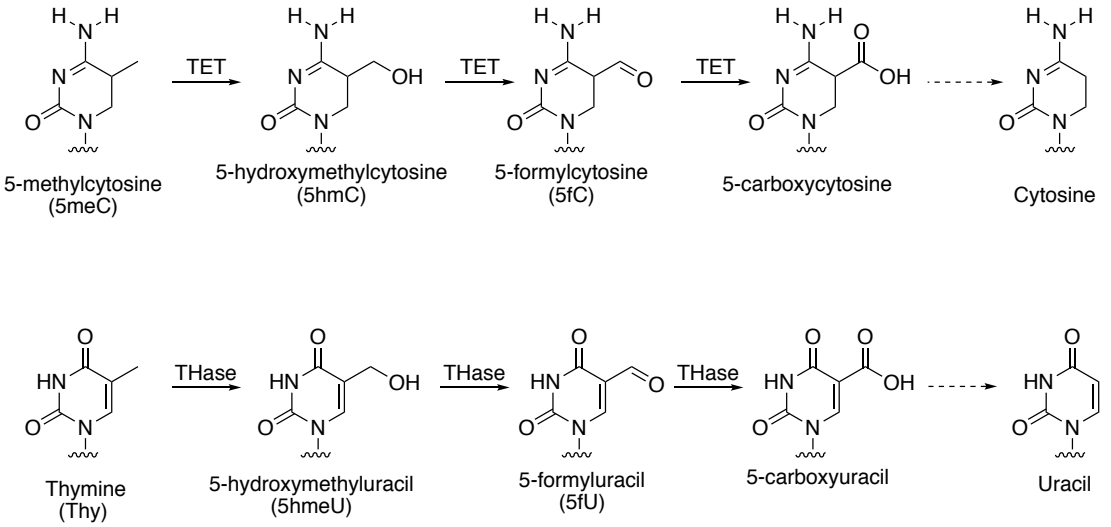


Figure 3.3: 5-methylcytosine and thymine (Thy) are demethylated successively following an H-abstraction step by Fe(II)-αKG enzymes TET and THase, respectively. The dotted arrows indicate the final step is catalysed by decarboxylase or DNA base repair proteins.

Alternatively, Cossío and co-workers have proposed distinct methylation state-dependent mechanisms, following QM calculations of JMJD demethylation, that suggest mono- and di- methyllysine may proceed via sequential N−H/C−H abstraction to yield an iminium intermediate.[206] Fang and Cisneros performed a comprehensive QM/MM study of the AlkB demethylation of 1meA and provide an in-depth discussion of an alternative OH rebound step coupled with a proton transfer to the OH^- ligand that results in a novel zwitterion intermediate.[207]

I directed my attention here, to the C−H abstraction from R−CH_3 shown in **Fig. 3.2**, which has been proposed to be the rate-determining step,[207–209] conserved across demethylation and hydroxylation mechanisms of post-translationally modified DNA nucleobases and residues catalysed by Fe(II)/α-KG enzymes.

Previous studies have analysed bond dissociation enthalpies (BDE) values using computational methods to provide insight into demethylase enzymes. A study by Hu et al. analyzed the C−H BDEs of 5meC, 5hmC, and 5fC models using high accuracy *ab initio* calculations, including G4, to explain differences in TET enzymatic activity.[210] de Visser et al. found homolytic $BDE_{C\text{-}H}$ values in active sites containing a heme iron are related to C−H abstraction barriers and obtained a correlation of $R^2 = 0.90$ with a model predicting cytochrome P450 C−H

hydroxylation barriers by employing UB3LYP/LACV3P++** level and zero point energy (ZPE) corrections.[211] Liu et al. used DFT to analyze the H-abstraction barrier by Fe(II)/α-KG AlkB DNA repair enzymes. The computationally predicted barrier of 83 kJ mol^{-1} was in agreement with the experimentally observed barrier of 87.4 kJ mol^{-1}.[209] Finally, Drew and Reynisson performed a large-scale analysis of over 586 C−H BDE values on major H-abstraction sites on cytochrome P450 drug-like compounds. A high correlation of $R^2 = 0.97$ (n = 100) was found when correlating UB3LYP/6-311+G(2df,p) and ZPE corrections with experimental values and favourable C−H BDE values were identified as the second most prominent factor in determining metabolic sites emphasising the effect on drug design.[212]

Utilising QM calculations, I have explored the intrinsic enthalpic energetic landscape of this important step to contribute towards a greater understanding of the relationship between methylation states, C−Me vs N−Me, and charge with enzymatic activity. I calculated $BDE_{C\text{-}H}$ values using **Eq. 3.1**, zero-point inclusive enthalpy energies, for proline, asparagine, and all possible states of lysine and arginine methylation as well as the N-methylated analogues of DNA nucleobases: adenine, guanine, cytosine, and thymine that are substrates of Fe(II)/α-KG subfamilies such as: JMJD, AlkB, TET, THase, PHD, and FIH.

$$BDE_{C\text{-}H} = H^{o}(R{-}CH_2^{\bullet}) + H^{o}(H^{\bullet}) - H^{o}(R{-}CH_3) \tag{3.1}$$

The overall enthalpy, where ΔH^{o} is defined in: **Eq. 3.2**

$$[Fe]{=}O + RCH_3 \xrightarrow{\Delta H^{o}} [Fe]{-}OH + RCH_2^{\bullet} \tag{3.2}$$

of each reaction involving the 26 model substrates was then determined following benchmarking six Density Functional Theory (DFT) methods: M06, M06L, M06-2X, ωB97XD, TPSS, and B3LYP.

3.2 Methodology: Benchmarking DFT functionals against G4

To achieve high-accuracy estimates of $BDE_{C\text{-}H}$ values we employed G4, an accurate composite *ab initio* method[213] in the gas phase. G4 theory involves geometry optimization and evaluation of zero-point energies (ZPEs) at the B3LYP/6-31G(2df,p) level. Correlation corrections with large basis sets, including multiple sets of polarisation functions, are added using Møller-Plesset (MP) perturbation theory up to fourth-order and with coupled cluster (CC) theory. G4 computed radical formation enthalpies are closest to values recommended in the Active Thermochemical Tables when compared to other composite *ab initio* thermochemistry calculations, such as Gaussian-3 (G3), W1BD and Complete Basis Set (CBS) methods CBS-APNO and CBS-QB3; based on 95% percent confidence intervals obtained for the G4 formation enthalpies of 38 open- and 45 closed-shell species (4.53 and 6.21 kJ mol^{-1}, respectively).[214] A 95% confidence interval for $BDE_{C\text{-}H}$ values of 7.7 kJ mol^{-1} is reasonably expected here. Favourable error cancellation may reduce this value further. T_1 Diagnostic calculations were performed on all G4 outputs, values less than 0.02 and 0.026 were accepted for calculations involving substrates before and after H-abstraction (**Table B.11**), respectively.[215–217] G4 results were used to benchmark $BDE_{C\text{-}H}$ values computed using six relatively inexpensive unrestricted density functional theory (DFT) calculations: M06, M06L, M06-2X, ωB97XD, TPSS, and B3LYP with the triple ζ-valence def2-TZVPP basis set for geometry optimisation and frequency calculations. DFT functional benchmarking was necessary to determine the optimal methodology for calculating accurate ΔH^o for each of the reactions that included the iron active site that is too computationally expensive to model with G4. Initial iron active site structures were taken from the crystal structure of JMJD2A (PDB: 2OQ6) and were optimised using M06/def2-TZVPP//M06/def2-QZVPP (**Table B.38**). The CO_2 was removed and replaced with a water molecule to yield a hexacoordinated iron(IV)-oxo model. A study by Cortopassi et al. performed hybrid QM/MM calculations for the triplet, quintet, and septet spin states of the iron-dioxygen bound complex and determined the quintet was the most stable, which we therefore elected to use for our calculations.[90] The iron active site following H-abstraction was modelled as [Fe]-OH with the quartet

spin state. Benchmarking was performed comparing DFT calculated O–H BDE values of four other metal complexes compared to experimental results found in the literature using M06-2X/def2-SVPP//M06-2X/def2-TZVPP. The theoretical and experimental values are listed in **Table 3.1**, where a high correlation of $R^2 = 0.93$ was achieved, shown in **Fig. B.2**. QM calculations were performed using Gaussian09.[178]

3.3 Analysis of Model Demethylase Substrates

3.3.1 Performance of DFT Methods

All DFT functionals were first paired with basis set def2-TZVPP and had high correlations, $R^2 = .95$ or better, when compared to Gaussian-4[213] (G4) values (**Fig. B.1**). G4 computed radical formation enthalpies are closest to values recommended in the Active Thermochemical Tables when compared to other composite *ab initio* thermochemistry calculations, such as Gaussian-3 (G3), W1BD and Complete Basis Set (CBS) methods CBS-APNO and CBS-QB3; based on 95% percent confidence intervals obtained for the G4 formation enthalpies of 38 open- and 45 closed-shell species (4.53 and 6.21 kJ mol^{-1}, respectively).[214] Diagnostic calculations were performed on all G4 outputs, values less than 0.02 and 0.026 were accepted for calculations involving substrates before and after H-abstraction (**Table B.11**), respectively.[216,217]

The hybrid meta-GGA functional M06 had the best correlation with G4 results and was therefore tested with two other basis sets def2-SVPP and def2-QZVPP (**Table 3.2**). The model substrates were subsequently optimised with M06/def2-QZVPP that was found to have the highest correlation, $R^2 = 0.97$, and the lowest error, RMSE = 1.3 kcal mol^{-1}, and was therefore selected for further calculations of the iron active site. Our results parallel another large-scale study of BDE calculations of biologically relevant molecules in which a Minnesota 06 functional better correlated with *ab initio* calculated values than B3LYP and ωB97XD.[222]

Table 3.1: Extrapolated $BDE_{O\text{-}H}$ values from linear regression constructed using experimental and computational $BDE_{O\text{-}H}$ values of metal complexes calculated at the M06-2X/def2-SVPP//M06-2X/def2-TZVPP level of theory.

	BDE (kcal mol^{-1})			
Complex	Exp.	Theo.	Extrap.[1]	Ref.[2]
$LCu^{III}OH$	90	84	90	[218]
$Fe^{IV}(O)TPFPP$	82	78	84	[219]
$Fe^{IV}(O)(TBP_8Cz^+)$	81	76	81	[220]
$Fe^{III}(PY5)(OH)\text{-}(Otf)_2$	80	74	79	[221]

[1]RMSE ± 1 kcal/mol

[2]Ref. experimental data

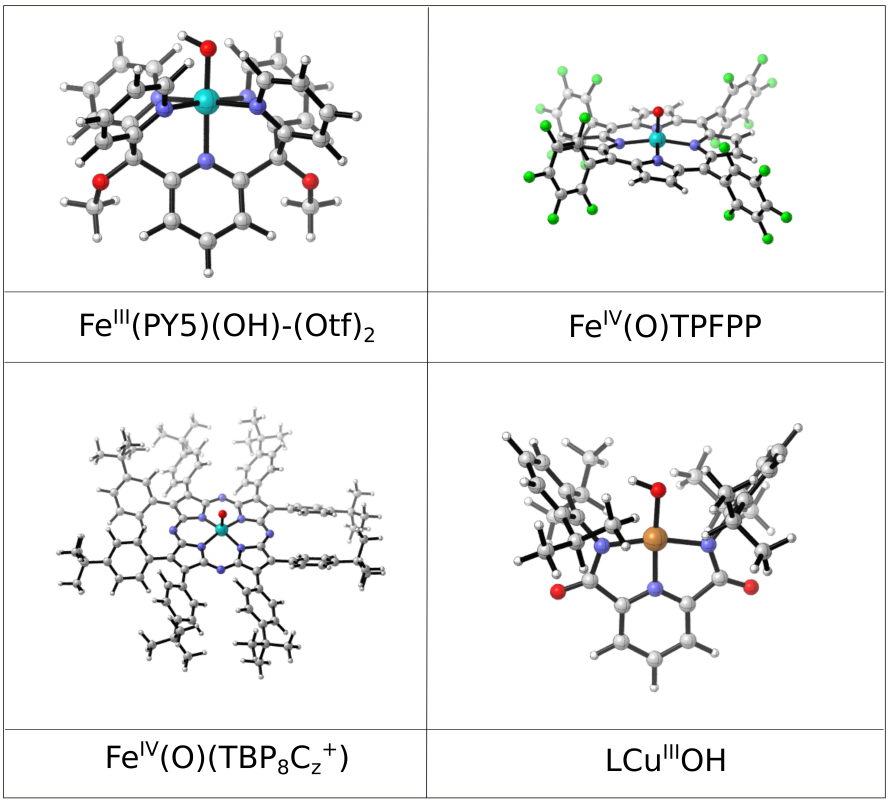
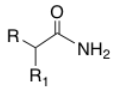
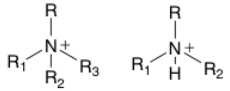
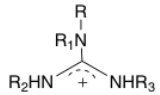
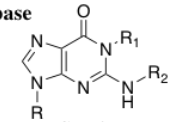
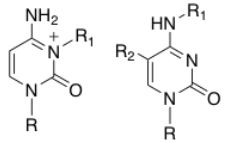
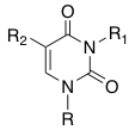
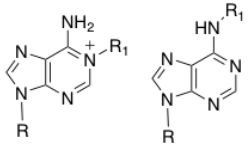


Table 3.2: Native and non-native Fe(II)/α-KG model substrates investigated here. BDE of C−H abstraction computed for 26 compounds, where R = Me.

					$BDE_{C-H}/\Delta BDE_{C-H}$ (kcal mol^{-1})		
	Model	**R_1**	**R_2**	**R_3**	**M06-2X**	**M06-2X/def2-QZVPP**	**M06-2X/def2-SVPP**
Residue  **Proline**	**Pro**	H	-	-	96.00/0.4	93.80/-1.8	93.27/-2.4
 **Asparagine**	**Asn**	H	-	-	92.56/-0.8	91.27/-2.1	89.96/-3.4
 **Lysine**	**Kme1 (left)**	CH_3	H	H	106.94/-0.5	106.94/-0.5	105.51/-2.0
	Kme2	CH_3	CH_3	H	106.61/-0.2	105.78/-1.0	104.50/-2.3
	Kme3	CH_3	CH_3	CH_3	106.29/-0.1	105.90/-0.5	105.12/-1.3
	nKme1 (right)	CH_3	H	-	91.57/-0.4	91.00/-1.0	90.67/-1.3
	nKme2	CH_3	CH_3	-	91.81/0.0	91.40/-0.4	91.33/-0.4
 **Arginine**	**ω-Rme1**	CH_3	H	H	97.45/0.6	96.42/-0.5	95.72/-1.2
	γ-Rme1	H	CH_3	H	96.14/0.6	95.31/-0.2	94.69/-0.8
	sRme2	H	CH_3	CH_3	97.06/1.1	96.06/0.3	95.20/-0.5
	aRme2	CH_3	H	CH_3	95.63/0.9	94.85/-0.1	94.05/-0.9
Nucleobase  **Guanine**	**1meG**	CH_3	H	-	95.64/1.5	94.70/0.6	95.64/1.5
	2meG	H	CH_3	-	93.08/2.1	92.09/1.1	91.84/0.9
 **Cytosine**	**3meC (left)**	CH_3	-	-	98.85/1.4	97.29/-0.2	96.98/-0.5
	4meC (right)	CH_3	H	-	91.67/1.5	90.34/0.2	89.96/-0.2
	5meC	H	CH_3	-	89.54/1.7	87.46/-0.4	87.09/-0.8
	5hmC	H	CH_2OH	-	89.24/1.0	86.76/-1.4	85.31/-2.9
	5fC	H	C=O	-	90.41/1.9	90.85/2.3	90.77/2.2
	f5Chydrate	H	$CH(OH)_2$	-	91.38/5.9	84.80/-0.7	85.84/0.3
 **Thymine**	**3meT**	CH_3	H	-	95.11/1.6	94.00/0.5	93.60/0.1
	Thy	H	CH_3	-	88.86/1.3	86.94/-0.6	86.03/-1.5
	5hmU	H	CH_2OH	-	90.17/2.1	86.77/-1.3	85.96/-2.1
	5fU	H	C=O	-	87.32/1.9	87.93/2.5	87.87/2.4
	f5Uhydrate	H	$CH(OH)_2$	-	73.81/-0.2	70.37/-3.6	67.90/-6.1
 **Adenine**	**1meA (left)**	CH_3	-	-	99.64/1.2	98.39/0.0	97.95/-0.5
	6meA (right)	CH_3	-	-	90.96/1.8	90.40/1.3	90.47/1.4
				RMSE (ΔBDE_{C-H})	**1.7**	**1.3**	**2.0**

Comparison of all DFT functionals with the def2-TZVPP basis set are shown in **Table 3.3**. M06 had the lowest RMSE of 1.5 kcal mol^{-1} followed by: ωB97XD (1.5 kcal mol^{-1}), M06-2X (1.7 kcal mol^{-1}), B3LYP (2.3 kcal mol^{-1}), TPSS (2.6 kcal mol^{-1}), and M06L (3.5 kcal mol^{-1}).

Analysis of $\Delta BDE_{C\text{-}H}$ values, defined here as the difference between G4 calculated $BDE_{C\text{-}H}$ (BDE_{G4}) (**Table B.1**) and DFT calculated $BDE_{C\text{-}H}$ (BDE_{DFT}) (**Table B.2 - B.9**) for a given substrate and functional, was used to identify trends among DFT functionals, such as overestimation of enthalpies or highlight model substrates that were more difficult to analyse using DFT. ωB97XD produced very accurate results, but failed to describe the larger model substrate 5fChydrate yielding a $\Delta BDE_{C\text{-}H}$ of 5.7 kcal mol^{-1}. 5fChydrate was poorly modeled by DFT, having the largest $\Delta BDE_{C\text{-}H}$ range of -7.8 to 5.9 kcal mol^{-1} observed by all functionals. Neglecting M06-2X, underestimation of 5fUhydrate $BDE_{C\text{-}H}$ values by all functionals was observed, ranging from -7.9 to -3.6 kcal mol^{-1}. Furthermore, an overall trend across all the functionals was an increase in error when predicting $BDEs_{C\text{-}H}$ for DNA nucleobase models rather than amino acids. Functionals B3LYP, TPSS, M06L produced lower $BDE_{C\text{-}H}$ predictions, or negative $\Delta BDEs_{C\text{-}H}$, than G4 for all substrates except 5fC and 5fU, for which they were subsequently better able to predict with a good ΔBDE range of -1.0 to 0.2 kcal mol^{-1} while the other functionals overestimated enthalpies by -1.0 to 2.5 kcal mol^{-1}.

Interestingly, the two worst performing, TPSS and M06L, are non-hybrid pure functionals, meaning they have no Hartree Fock (HF) exchange. This suggests the occurrence of a possible self-interaction error (SIE), a phenomenon in which there is spurious interaction with an electron and itself such that the exchange and Coulomb energy do not exactly cancel.[223] In contrast, neglecting 5fUhydrate, M06-2X which has a large amount of exchange (58%) overestimated rather than underestimated $BDEs_{C\text{-}H}$ by an average of 1.8 kcal mol^{-1} for all DNA nucleobase substrates.

3.3.2 Analysis of Bond Strengths

A large variance in bond strength was exhibited among the set of model Fe(II)/α-KG substrates ranging from 74.0 to 107.5 kcal mol^{-1}. The most noticeable difference was observed between the charged and neutral substrates, in which the positively charged molecules represent the most difficult for H-abstraction. These include

Table 3.3: $BDE_{C\text{-}H}$ and $\Delta BDE_{C\text{-}H}$ values for model amino acid and DNA nucleobases for DFT functionals M06, B3LYP, TPSS, M06L, and ωWB97XD with the def2-TZVPP basis set.

	BDE				
Model	M06	B3LYP	TPSS	M06L	ωB97XD
Pro	93.59/-2.1	93.72/-1.9	93.62/-2.4	91.46/-4.2	96.52/0.9
Asn	90.51/-2.8	89.53/-3.8	89.16/-4.2	86.91/-6.5	91.38/-2.0
Kme1	106.63/-0.9	105.64/-1.8	105.48/-2.0	105.09/-2.4	106.83/-0.7
Kme2	106.11/-0.7	105.06/-1.8	104.88/-1.9	104.54/-2.3	106.50/-0.3
Kme3	106.01/-0.4	104.71/-1.7	104.53/-1.9	104.20/-2.2	105.99/-0.4
nKme1	91.08/-0.9	88.81/-3.1	89.39/-2.6	89.47/-2.5	90.11/-1.8
nKme2	91.57/-0.2	88.69/-3.1	89.21/2.6	89.36/-2.4	90.54/-1.2
ω-Rme1	96.52/-0.4	94.59/-2.3	94.10/-2.8	94.38/-2.5	96.56/-0.3
γ-Rme1	95.28/-0.2	93.27/-2.2	93.02/-2.5	93.05/-2.5	95.27/-0.2
s-Rme2	96.13/0.0	94.23/-1.9	93.80/-2.3	94.01/-2.1	96.27/0.1
a-Rme2	94.95/0.4	92.76/-1.8	92.47/-2.1	92.81/-1.7	94.63/0.1
1meG	94.63/0.5	92.95/-1.1	93.32/-0.8	92.57/-1.5	95.00/0.9
2meG	92.09/1.1	89.72/-1.2	89.46/-1.5	89.52/-1.4	91.35/0.4
3meC	97.21/-0.3	95.76/-1.7	95.35/-2.1	95.10/-2.4	98.45/1.0
4meC	90.29/0.2	88.65/-1.5	88.27/-1.9	88.22/-1.9	90.91/0.8
5meC	86.72/-1.1	86.01/-1.8	86.02/-1.8	85.29/-2.6	89.11/1.3
5hmC	86.46/-1.8	85.88/-2.3	85.01/-3.2	83.38/-4.8	87.69/-0.5
5fC	90.75/2.2	88.79/0.2	89.11/0.6	88.52/0.0	89.57/1.0
5fChydrate	84.54/-1.0	81.25/-4.3	81.49/-4.0	77.68/-7.8	91.19/5.7
3meT	93.93/0.4	92.41/-1.1	92.04/-1.5	91.76/-1.8	93.74/0.2
Thy	86.79/-0.7	85.81/-1.7	85.87/-1.6	85.52/-2.0	87.60/0.1
5hmU	86.48/-1.6	85.76/-2.3	84.14/-3.9	82.96/-5.1	87.88/-0.2
5fU	87.86/2.4	85.51/0.1	85.63/0.2	85.28/-0.2	84.48/-1.0
5fUhydrate	69.81/-4.2	69.34/-4.6	67.55/-6.4	66.09/-7.9	70.97/5.7
1meA	97.68/-0.7	96.78/-1.6	97.12/-1.3	96.15/-2.3	98.99/0.6
6meA	89.88/0.8	88.24/-0.9	88.45/-0.7	87.18/-1.9	89.27/0.2
RMSE ($\Delta BDE_{C\text{-}H}$)	1.5	2.3	2.6	3.5	1.5

all 1^+ charged model substrates: Kme1, Kme2, Kme3, ω-Rme1, γ-Rme1, aRme1, sRme2, 1meA, and 3meC. A standout group that had significantly higher $BDEs_{C\text{-}H}$ was Kme1, Kme2, and Kme3 with 107.5, 106.8, and 106.4 kcal mol^{-1}, respectively. These displayed a striking difference of 8.0 kcal mol^{-1} in energy required for H-abstraction from the next most difficult substrate, 1meA (98.4 kcal mol^{-1}). In **Fig. 3.4** highlights the range of BDE values for a substrate from each family of proteins.

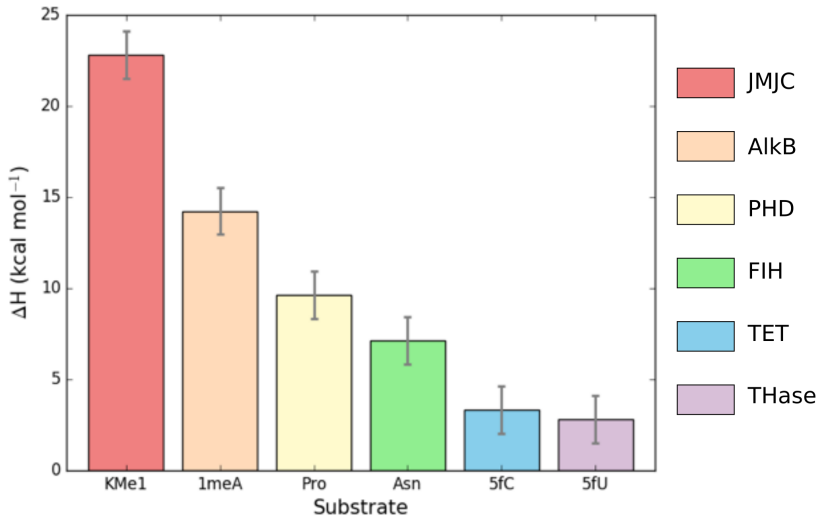


Figure 3.4: ΔH^o values calculated using M06/def2-TZVPP//M06/def2-QZVPP for amino acid and nucleobase H-abstraction by Fe(II)/α-KG enzymes with the quintet spin state. Bar colours correspond to receptor family. Error bars show RMSE vs G4 theory.

Analysis of the Mulliken spin density on C$\cdot$ compared to the M06/def2-QZVPP//G4 ΔH^o_{rxn} for each substrate yielded a good agreement when separated by charge shown in **Fig. 3.5**, with a correlation of $R^2 = 0.95$ and $R^2 = 0.89$ for the cationic and neutral substrates, respectively. The high correlation and lack of outliers observed when separating substrates by charge highlights the importance of radical stability and the effect on ease of abstraction for neutral over charged substrates.

More subtle differences in bond strengths were qualitatively analysed by generation of Mulliken spin density plots, shown in **Fig. 3.1**. Three substrates that exemplified this were Thy, 4meC, and Kme1, shown in **Fig. 3.6** These allow for variations in H-abstraction favourability to be rationalised by identifying areas of delocalisation for the unpaired electron. Lower $BDEs_{C\text{-}H}$ are associated with substrates that are capable of hyperconjugation with a neighbouring π-system or

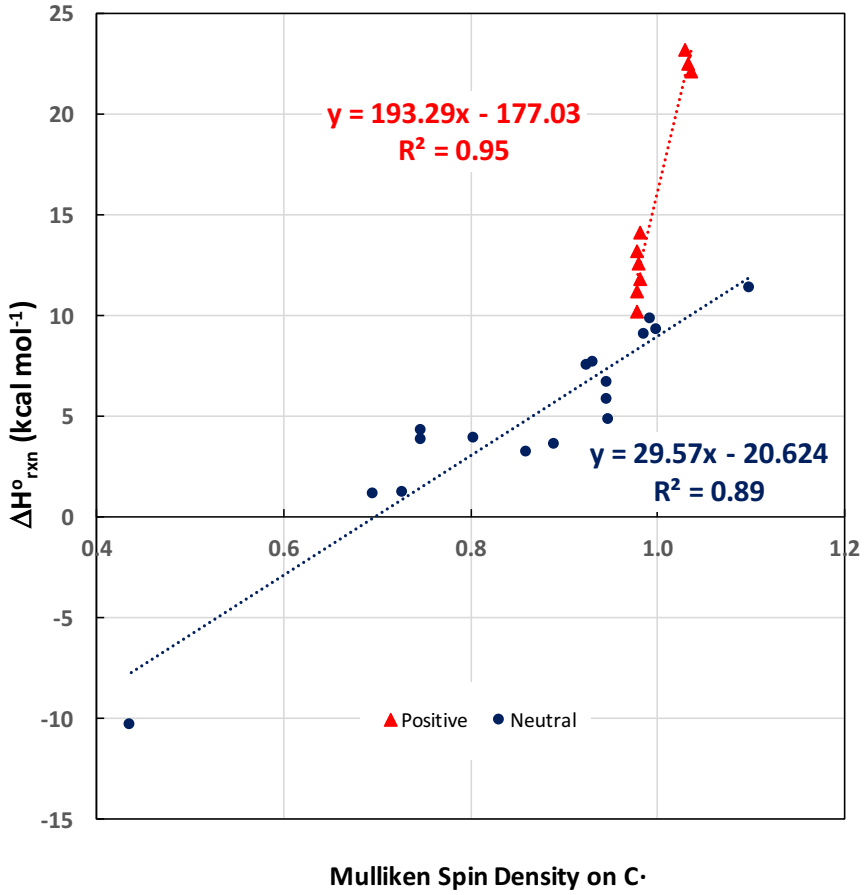


Figure 3.5: Correlation between the Mulliken spin density of C· and the ΔH^{o} reaction for Fe(II)/α-KG substrates.

favourable interaction with an adjacent lone-pair containing heteroatom or σ-bond, although less so. An example of this is Thy, which is methylated at the benzylic position and can delocalise the radical around the pyrimidine ring. The molecular orbital diagram shows how interaction of the 2p(C·) with the bonding and anti-bonding π orbitals of the π-acceptor group can yield a net two-electron stabilisation, favoured more so by the planarity of the molecule allowing for better overlap.[224] This produces three new orbitals, including a non-bonding and bonding orbital that becomes occupied by the radical and two electrons from the π-acceptor, respectively. Stabilisation can also occur due to a three electron interaction with a heteroatom. In the case of 4meC, the unpaired electron is next to a NH_2 substituent, a non-planar

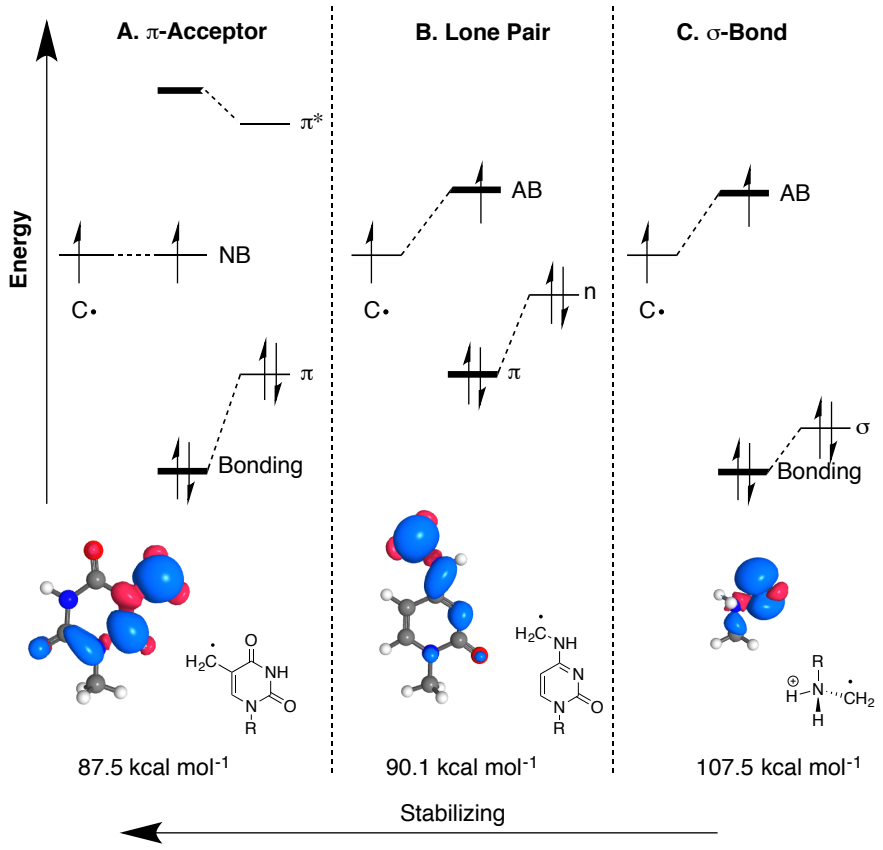


Figure 3.6: Molecular orbital diagram and spin density plots showing location of spin-up (blue) and spin-down (red) electron density that demonstrate radical stabilisation effects influenced by the presence of a π-acceptor, lone pair, and σ bond for thymine, 4-methylcytosine, and monomethyllysine radicals, (R = Me).

lone pair donor. The doubly occupied orbital containing the lone pair is stabilised as the unpaired electron moves into the anti-bonding orbital. Similarly, this occurs with a σ-bond interaction, however, the differences in energy between the new bonding and antibonding orbitals is much greater and therefore decreases stability, as observed with the N-methylated lysine and arginine model substrates.

H-abstraction of 1meG and 3meT is the most difficult of the modified neutral nucleobases. 1meG has its $CH_2\cdot$ positioned near a lone pair but ortho to the C6 carbonyl, an ortho/para deactivator, while the 3meT radical carbon centre is sandwiched between the C4 and C2 carbonyls. H-abstraction is energetically

more favourable by 3.1 kcal mol^{-1} for 2meG than 1meG, where the radical centre is stabilised by the guanidine substituent, which is located meta to the C6 carbonyl. In contrast, the C6 position is the site of demethylation for 6meA contains a methylated NH_2 group. 6meA is the only model substrate to exhibit a lone pair interaction with an imidazole nitrogen atom and has a much lower $BDE_{C\text{-}H}$ of 89.1 kcal mol^{-1} than the analogous positively charged 1meA, which requires 98.4 kcal mol^{-1}.

Pyrimidine substrates 5meC and Thy are hydroxylated three times in succession in the 5-methylcytosine demethylation pathway and Thymidine Salvage Pathway, respectively. This produces intermediates 5hmC, 5fC, 5hmU, and 5fU. Here we modeled the third intermediate in each pathway as the formyl (5fC and 5fU) and the hydrate (5fChydrate, 5fUhydrate). These yielded the lowest $BDE_{C\text{-}H}$ values of all 26 model substrates with a range of 74.0 to 88.6 kcal mol^{-1}. This is primarily due to delocalisation of the unpaired CH_2 electron around the aromatic pyrimidine ring. 5fUhydrate is stabilized by a H-bond between the C6 carbonyl and an -OH group, stabilising the structure and explaining the very low BDE of 74.0 kcal mol^{-1}. Furthermore the effects of induction caused by the presence of an amine group rather than a carbonyl at the C6 position can be analysed by examining the corresponding intermediate pairs of each pathway: 5meC/Thy, 5hmC/5hmU, 5fC/5fU, and 5fChydrate/5fUhydrate.

Our analysis of BDEs for C−H cleavage by Fe(II)/α-KG on model substrates has the potential to provide insights into enzymatic reactivity. Energy for H-abstraction of N-methylated lysine was notably greater than for all other substrates. Calculations of two theoretical substrates suggested this could be drastically lowered for Kme1 by 15.6 kcal mol^{-1} and for Kme2 by 14.6 kcal mol^{-1} if the charge could be neutralized by an adjacent residue (e.g. via a formation of a hydrogen bond). Deprotonation of the terminal amine has been proposed to occur preceding the methylation for histone lysines by a number of suggested molecules such as active site tyrosine residues, an ordered water, bulk solvent, or changes in pK_a. For multiple methylation reactions this must occur several times without loss of binding of the lysine substrate. Due to the high enthalpic cost of H-abstraction from N-methylated lysine, it is plausible that a similar interaction with the terminal hydrogens and active site environment is occurring as a precursor to demethylation of Kme1 and Kme2.

3.4 Conclusion

In summary, I have explored the enthalpic landscape of reactions catalysed by Fe(II)/α-KG enzymes via a comprehensive analysis of $BDE_{C\text{-}H}$values for 26 models of native and non-native substrates. Of six DFT methods M06-2X/def2-QZVPP had the lowest error when compared to G4 calculated $BDE_{C\text{-}H}$ values. Results of C−H bond strength indicated a $BDE_{C\text{-}H}$ range of 36.6 kcal mol^{-1}. All positively charged substrates such as N-methylated lysine and arginine, 1meA, and 3meC represented the substrates with the highest $BDE_{C\text{-}H}$values. Model substrates that had a hydrogen abstracted from the benzylic position were better able to stabilize the free radical through delocalisation over the ring, as observed from analysis of spin density plots. Subsequently, pyrimidine model substrates such as those involved in the Thymidine Salvage Pathway and demethylation of 5meC catalysed by THase and TET enzymes respectively, had the lowest $BDEs_{C\text{-}H}$. 5fUhydrate was further stabilised by an internal H-bond and had the lowest $BDE_{C\text{-}H}$. Thus far we have focused on innate substrate reactivity: future work will aim to compare dissociation energies with abstraction barriers for Fe(II)/α-KG substrates as well as consider the effects of enzyme environment and the kinetics of this step, in addition to thermodynamics.

Corpora non agunt nisi fixata

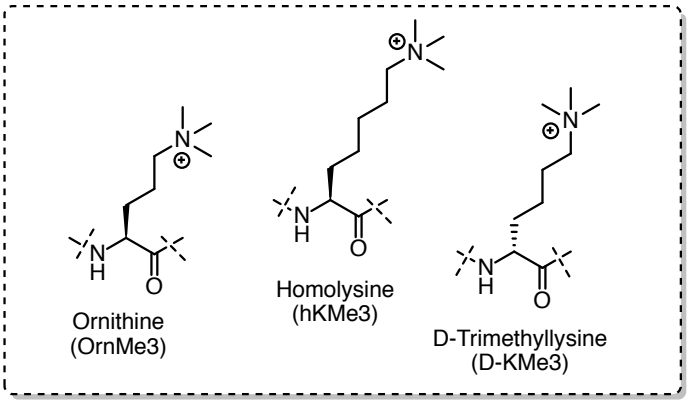
No compound is active unless it is bound by a receptor

— Paul Ehrlich, 1913

4

Exploring Selectivity by Epigenetic Reader Proteins

Contents



Belle R., Al Temimi A. H. K., **Kumar K.**, et al. Investigating D-lysine stereochemistry for epigenetic methylation, demethylation and recognition, *Chem. Commun.* **2017**, *53*.

Al Temimi A. H. K., Belle R., **Kumar K.**, et al. Recognition of shorter and longer trimethyllysine analogues by epigenetic reader proteins, *Chem. Commun.* **2018**, *54*.

4.1 Background: KMe3 analogues and reader proteins

Chapter 2 discussed an in-depth MD-based examination of a specific Arg:arene interaction for a class of bromodomains, acetyllysine readers. Also modulators of gene activation and repression are N-methyllysine readers such as: Plant Homeodomain (PHD) zinc fingers, tandem tudor domains (TTD), chromodomains (CD) and malignant brain tumour (MBT) domains.This chapter will analyse computational simulations performed along experimental research conducted by the Schofield, Kawamura, and Mecinović Group (Radboud University), to discover properties of selectivity by readers.

D-KMe3, Ornithine, and Homolysine

I first explored stereoselectivity by inverting the C_α backbone to analyse D-trimethyllysine (D-KMe3) vs L-lysine (referred to as L-KMe3, KMe3, or wild type throughout) binding. The chemical structures are shown in **Fig. 4.1**. D-amino acids have been found in ageing and diseased cells[225] and proposed to occur in tumour cells,[226] but have not yet been investigated in these systems. Next

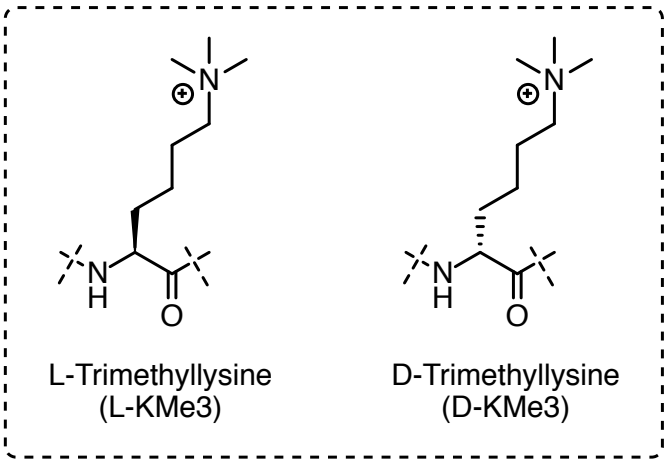


Figure 4.1: Epigenetic regulation of methylated histone L-lysines is well known, however modulation of methylated D-lysines has yet to be explored.

differences in binding based on chain length were analysed, namely methylated lysine one carbon shorter (ornithine, OrnMe3) and longer (homolysine, hKMe3) as

compared with the natural substrate, shown in **Fig. 4.2**. It was hypothesised that differences in chain length would affect the cation–π interactions strength necessary for recognition of methylated lysine by the aromatic cages of epigenetic readers.

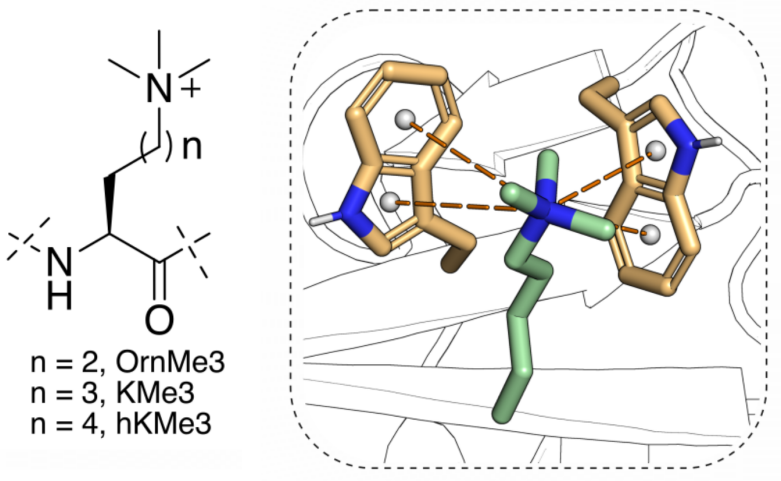


Figure 4.2: Trimethyllsine (KMe3) bound to PHD finger KMD5A aromatic cage and structures of trimethylated ornithine (OrnMe3) and homolysine (hKMe3).

Experimental isothermal titration calorimetry (ITC) data was used to calculate binding affinities for the three ligands with writer, reader, and eraser proteins. This technique is widely used in drug discovery because it not only provides binding affinities for small molecules, but a thermodynamic characterization of binding. The calorimeter is able to detect differences in heat even at a submillionth of a degree that occur as a result of two molecules interacting. Specifically, changes in power required to maintain the same isothermal conditions as the control with no ligand present are measured. Injections of the ligand solution are performed incrementally until the biomolecule is saturated, represented by peaks that eventually plateau to represent only the heats of dilution. A Wiseman plot is then constructed following integration of the peaks and allow the binding enthalpy (ΔH), binding constant (K_d), and entropy (ΔS) to be calculated.

Interestingly, it was found from experimental ITC data that only the reader class of enzymes recognised KMe3, D-KMe3, OrnMe3, and hKMe3, while demethylases and methyltransferases were only reactive to the wild type. Therefore, only reader complexes were chosen to be investigated computationally. Following MD

simulations with 5 reader proteins bound to a Histone 3 (H3) chain containing the wild type and modified KMe3 residues, I was able to quantify key protein interactions that stabilised binding and visualise the flexibility of the H3 chain in each system. By identifying favourable interactions between KMe3 analogues with each active site using computational approaches, these insights could facilitate the design of novel scaffolds to inhibit reader enzymes.

4.2 Methodology: Modelling D-Kme3, OrnMe3, and hKMe3

4.2.1 Simulating D-KMe3

Ten MD simulations were carried out for 10 ns each. A PDB structure for the model representing $TAF3_{PHD}$ (PDB: 2K17), $KDM4A_{TTD}$ (PDB: 2GFA), $KDM5A_{PHD}$ (PDB: 2KGI), $BPTF_{PHD}$ (PDB: 2F6J), and $SGF29_{TTD}$ (PDB: 3ME9) reader proteins were used as a template for building the reader: D-KMe3 systems. The L-KMe3 stereocenter was manually inverted to generate the D-KMe3 complex, with priority given to replicating the position of L-KMe3 in the original PDB structure.

Hydrogen atom addition was performed with tLEaP. Systems were solvated in a 10 Å truncated octahedron of TIP3P water molecules[163] that extended at least 10 Å from protein atoms and neutralised explicitly with either sodium or chloride counterions.AMBER12.[159] was used with the Amberff12SB force field to define protein partial charges. The force constants for bond, angle, and torsions for the atoms bonded to zinc derived from the Zinc AMBER Force Field (ZAFF) developed by the Merz Group.[227].

Atomic partial charges for KMe3 and D-KMe3 correspond to the Restrained Electrostatic Potential (RESP) charges calculated using HF/6-31G(d), shown in **Table C.1**.

The final systems were minimised in two steps. First 1000 steps of steepest descent and 1000 steps of conjugate gradient minimisation where the protein was held fixed by using position restraints with a force constant of 500 kcal mol^{-1} $Å^{-2}$. This was repeated without the position restraints. The system was then heated for 1 ns from 0 K to 310 K under constant volume periodic boundary conditions (NVT). Then 1 ns of equilibration under constant pressure and temperature (NPT)

was performed. Following this, 10 ns molecular dynamics simulations were then performed. Langevin thermostat[228] was used to simulate a constant temperature of 310 K with collision frequency of 1 ps^{-1}. The SHAKE algorithm[229] was turned on to constrain all bonds involving hydrogen and 2fs was defined as the time step for numerical integration. Isotropic position scaling was used to maintain the pressure of 1 atm (τ_p = 2 ps). The Particle Mesh Ewald summation method[230] was employed to enforce a 8.0 Å cutoff for non-bonded long-range and electrostatic interactions. Trajectories were saved at 20-ps intervals and visualised using VMD.[164] Electrostatic energies between the terminal KMe3 side chain $N_\epsilon{}^+$ atom and the π-system of surrounding aromatic cages were calculated with the NAMD Energy Plugin 1.4.[177] For these calculations, the π-system was defined for tryptophan, tyrosine, and phenylalanine residues as the ring (non-H) atoms. Energy values were measured every 20 ps and averaged over 10 ns.

4.2.2 Simulating OrnKMe3 and hKMe3

Ten MD simulations were carried out for 10 ns each. A PDB structure for the model representing $TAF3_{PHD}$ (PDB: 2K17), $KDM4A_{TTD}$ (PDB: 2GFA), $KDM5A_{PHD3}$ (PDB: 2KGI), $BPTF_{PHD}$ (PDB: 2F6J), and $SGF29_{TTD}$ (PDB: 3ME9) reader proteins were used as a template for building the reader:OrnMe3 and hKMe3 systems. The backbone of KMe3 was manually extended or shortened by a CH_2 to construct the hKMe3 and OrnMe3 ligands, respectively. Atomic partial charges for OrnKMe3 and hKMe3 are listed in **Table C.2** and were derived using the RESP methodology. Initially, $KDM5A_{PHD}$ was simulated for a longer timescale of 50 ns bound to OrnMe3 and hKMe3 to confirm equilibration could be reached for the reader:H3 systems. AMBER12[159] was used with the Amberff12SB force field to define protein partial charges. Hydrogen atom addition was performed with LEaP. Systems were solvated in a 10 Å truncated octahedral box of TIP3P water[163] and neutralised explicitly with either sodium or chloride counterions. were employed. The minimisation, equilibration, and production ensembles used the same methodology as the L-KMe3 and D-KMe3 systems.

4.3 Processing MD Simulation Results

4.3.1 Analysis of Readers Bound to D-KMe3

It is well understood that N-methylated L-lysine is a natural substrate of epigenetic readers, writers, and erasers, however the effects with respect to selectivity of the C_α backbone stereocenter had not been previously investigated. Following experimental and computational studies, it was concluded that D-KMe3 is recognised by the reader class of enzymes. To further understand binding of D-KMe3, MD simulations were performed and analysed using post-processing techniques to quantify key interactions.

Benchmarking MD parameters

To validate the simulations, average RMSD values were calculated relative to the starting structure of the simulation, for each of the 10 trajectories, listed in **Table 4.1**.

Table 4.1: Average RMSD and error of C_α for reader:L-KMe3 and reader:D-KMe3 MD simulations over 10 ns.

	L-KMe3 RMSD (Å)		D-KMe3 RMSD (Å)	
System	Reader	H3	Reader	H3
$BPTF_{PHD}$	5.85 ± 2.49	0.47 ± 0.20	4.98 ± 1.99	1.00 ± 0.66
$KDM4A_{TTD}$	2.08 ± 0.64	1.17 ± 0.27	2.68 ± 0.60	0.79 ± 0.24
$KDM5A_{PHD3}$	2.45 ± 0.58	1.16 ± 0.34	2.34 ± 0.41	2.97 ± 0.91
$SGF29_{TTD}$	1.28 ± 0.20	1.03 ± 0.45	1.56 ± 0.25	1.19 ± 0.42
$TAF3_{PHD}$	3.25 ± 0.82	3.24 ± 1.20	3.68 ± 0.79	2.74 ± 0.44

When compared to the wild type counterpart, systems with D-KMe3 did not differ significantly in RMSD. The receptor in the D-KMe3:$BPTF_{PHD}$ complex showed a high overall change from the crystal structure (4.98 ± 1.99), though this was similar to the D-KMe3:$BPTF_{PHD}$ complex (5.85 ± 2.49). Interestingly, the H3 residues containing L-KMe3 (0.47 ± 0.20) and D-KMe3 (1.00 ± 0.66) in these simulations were some of the lowest average RMSD values observed. Differences among the receptors can be explained by analsying their tertiary structures. **Fig. 4.3** shows the

globular and active site structure for each reader complex. $BPTF_{PHD}$ is observed to contain two major regions (**Fig. 4.3B**), a bundle of α-helices and a globular domain containing the active site that more closely resembles $KDM5A_{PHD3}$ (**Fig. 4.3A**) and $TAF3_{PHD3}$ (**Fig. 4.3C**), linked by a short loop region. The flexible linker loop allows for large global movements of the two domains, explaining the large average RMSD values observed for $BPTF_{PHD}$. However, local differences in movement at the $BPTF_{PHD}$:L-KMe3 and $BPTF_{PHD}$:D-KMe3 active site were shown to be reasonable from the low H3 RMSD values and qualitative observations of the trajectories.

The low RMSD values observed for simulations with $KDM4A_{TTD}$ (2.08 ± 0.64) and $SGF29_{TTD}$ (1.28 ± 0.20) can be rationalised by the predominance of stable α-helices and β-sheets, visualised in **Fig. 4.3D** and **Fig. 4.3E**, respectively.

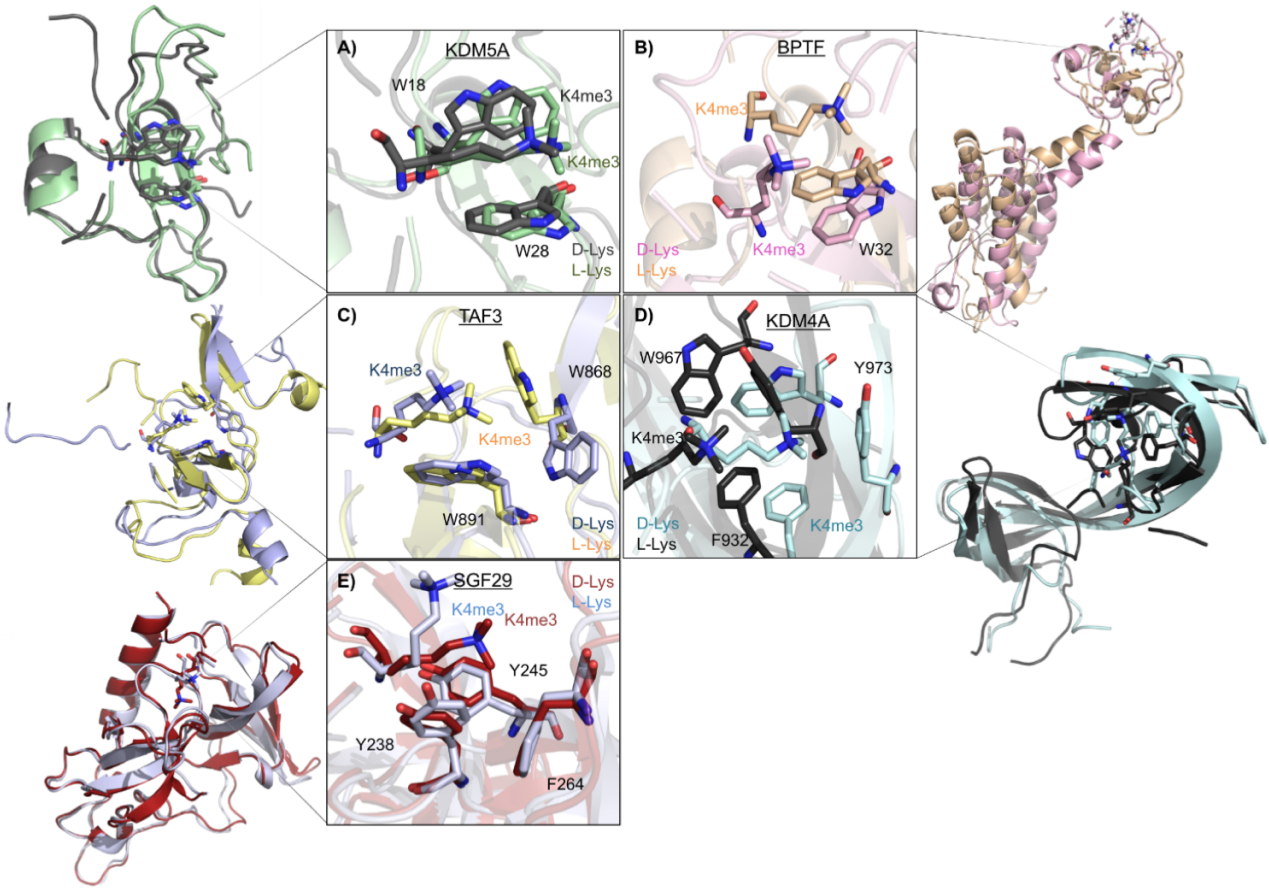


Figure 4.3: View of the histone binding site of the reader protein. The following crystal structures were used as starting point for the MD simulations: A) $KDM5A_{PHD3}$ (PDB: 2KGI), B) $BPTF_{PHD}$ (PDB: 2F6J), C) $TAF3_{PHD}$ (PDB: 2K17), D) $KDM4A_{TTD}$ (PDB: 2GFA), and E) $SGF29_{TTD}$ (PDB: 3ME9) complexed with Histone 3 chains containing D-, and L-H3K4me3 residues following MD minimisation and equilibration.

Trajectory Analysis

10 ns MD simulation trajectories were qualitatively and quantitatively analysed in order to further understand binding.

Fig. 4.4 visualises snapshots at time = 0 ns, 5 ns, and 10 ns of each simulation, where the substrate, aromatic cage, and H3 chain have been highlighted. Orientations of the two KMe3 stereoisomers within the aromatic pocket of the same reader are similar at 0 ns, with the exception of $SGF29_{TTD}$. $KDM5A_{PHD3}$ bound to both D- and L-KMe3 exhibit the most similar pose regarding placement of the W18–W28 aromatic cage and modified residues. In the case of $SGF29_{TTD}$, D-KMe3 more fully occupies the aromatic Y238–Y245–F264 cage than does L-KMe3; however, L-KMe3 adopts a more similar orientation to D-KMe3 by 5 ns that is maintained throughout the simulation. For the two H3-PHD3 reader complexes, $KDM5A_{PHD3}$ and $TAF3_{PHD3}$, major differences in the H3 chain backbone geometry were observed throughout the simulation when comparing the L- and D-systems, also paralleled by differences in average RMSD values.

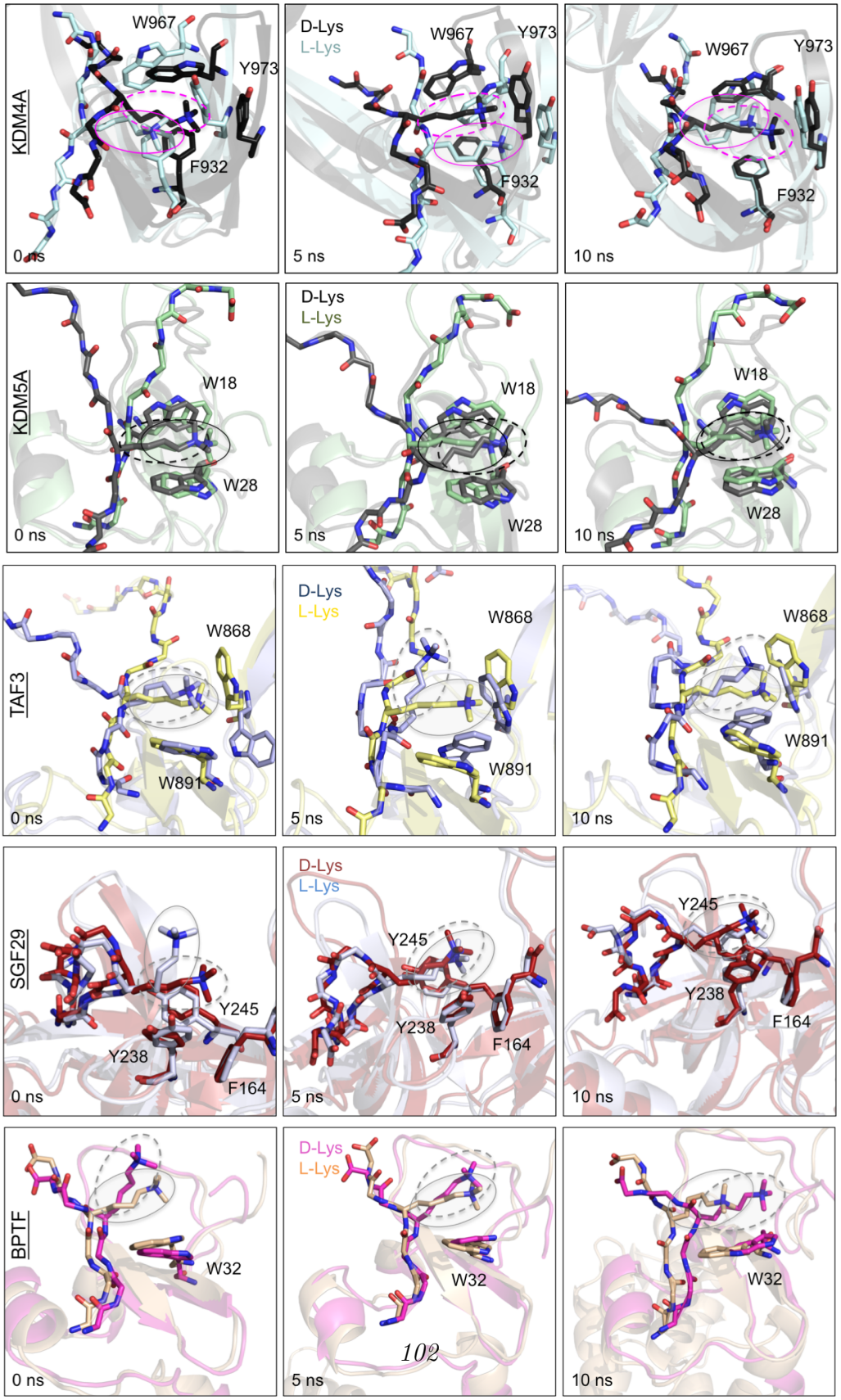


Figure 4.4: Snapshopts of readers complexed with L- and D-KMe3 at times 0 ns, 5 ns, and 10 ns. The substrate, aromatic cage side chain, and H3 backbone atoms are shown in sticks.

In order to characterize the occupancy of a cation-π interaction, the distance from the $N_\epsilon{}^+$ atom of the terminal lysine trimethylammonium group to the centroid of the π-system, as defined for tryptophan, tyrosine, and phenylalanine residues as the ring (non-H) atoms, was calculated. A geometric distance threshold of < 6 Å indicated a cation–π occurrence for a given frame. The distance vs time plots and frequency for each D-KMe3 simulation as compared to the wild type are shown in **Fig. 4.5-4.9**.

All the simulations maintain a stable cation–π interaction with at least one aromatic cage residue. In the case of KDM4A, the enzyme is characterized by a Phe-Trp-Tyr aromatic cage. Large movements are observed with respect to the F932 benzene face, with patterns more similar to the wild type for residues W967 and Y973 bound to D-KMe3 (**Fig. 4.5**). When comparing the $KDM5A_{PHD3}$ (**Fig. 4.6**) and $TAF3_{PHD3}$ (**Fig. 4.7**) systems, it is interesting to note that although they are both composed of Trp-Trp aromatic cages, the cation is much more dynamic in the $TAF3_{PHD3}$ W18–W28 cage. However, after $\sim$2 ns of simulation, a cation–π interaction is formed between the trimethylammonium ion with both $TAF3_{PHD3}$ indoles. The significant reorientation of H3 backbone for D-KMe3 bound to $KDM5A_{PHD3}$ and $TAF3_{PHD3}$ displayed in the snapshots of the simulations, shown in **Fig. 4.4**, are likely stabilised by these favourable interactions. For $SGF29_{TTD}$, this reader contains an aromatic cage composed of Tyr-Tyr-Phe residues. Cation–π interactions are observed for L-KMe3 and D-KMe3 with all three residues (**Fig. 4.8**). Finally, the Trp32 residue of $BPTF_{PHD}$ showed a dynamic interaction with D-KMe3 that is able to stabilise a cation–π interaction at $\sim$6 ns, which is maintained until the end of the simulation (**Fig. 4.9**)

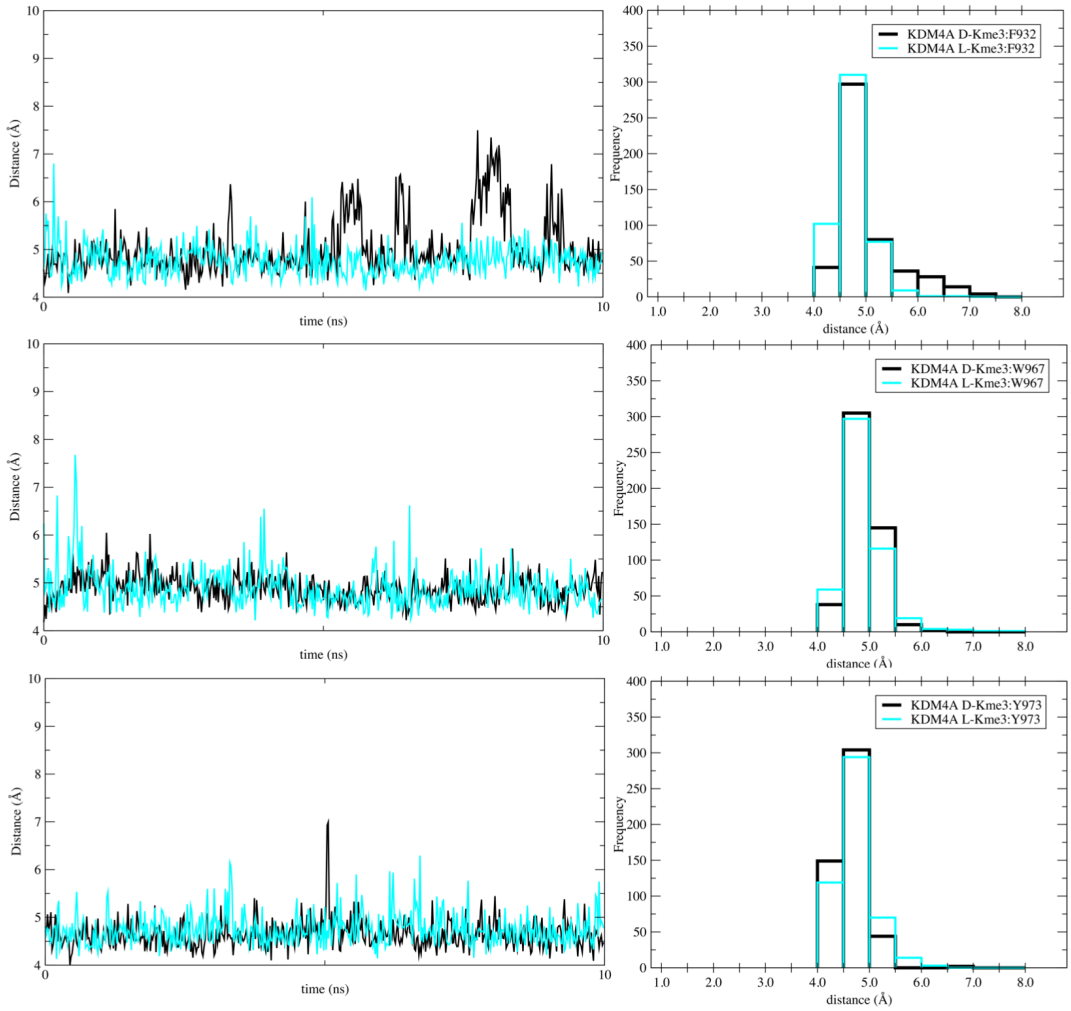


Figure 4.5: Distance vs time plots of $N_{\epsilon}{}^{+}$ atom of D-H3K4me3 and L-H3K4me3 to F932, W967, and Y973 side chains over 10 ns.

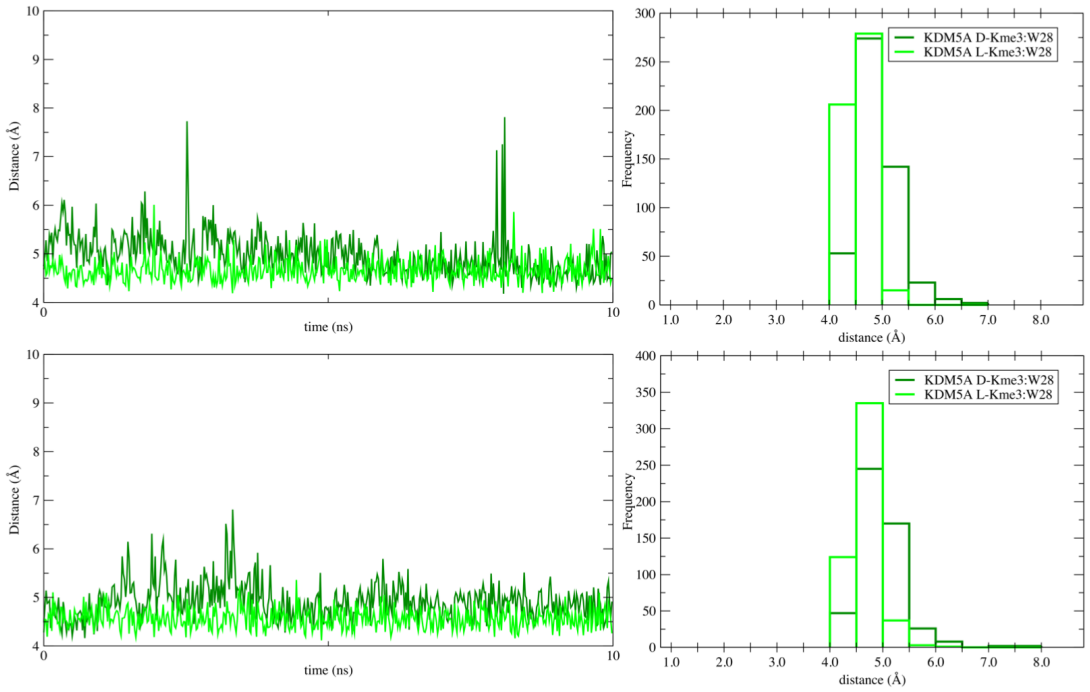


Figure 4.6: Distance vs time plots of $N_\epsilon{}^+$ atom of D-H3K4me3 and L-H3K4me3 to W18 and W28 side chains over 10 ns.

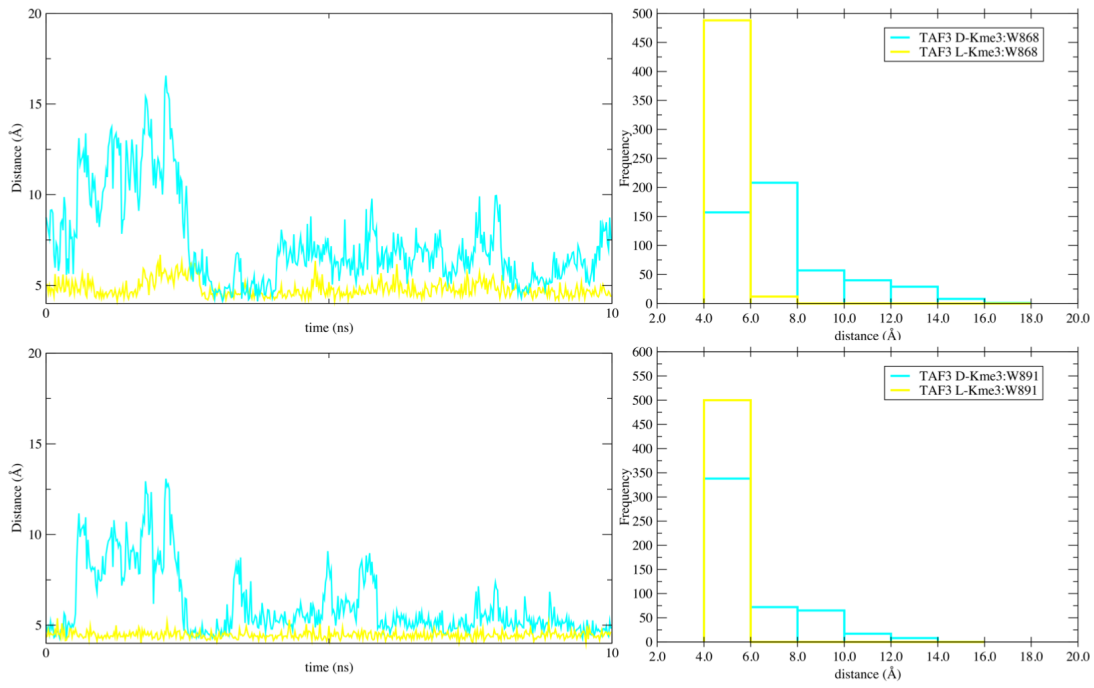


Figure 4.7: Distance vs time plots of $N_\epsilon{}^+$ atom of D-H3K4me3 and L-H3K4me3 to W868 and W891 side chains over 10 ns.

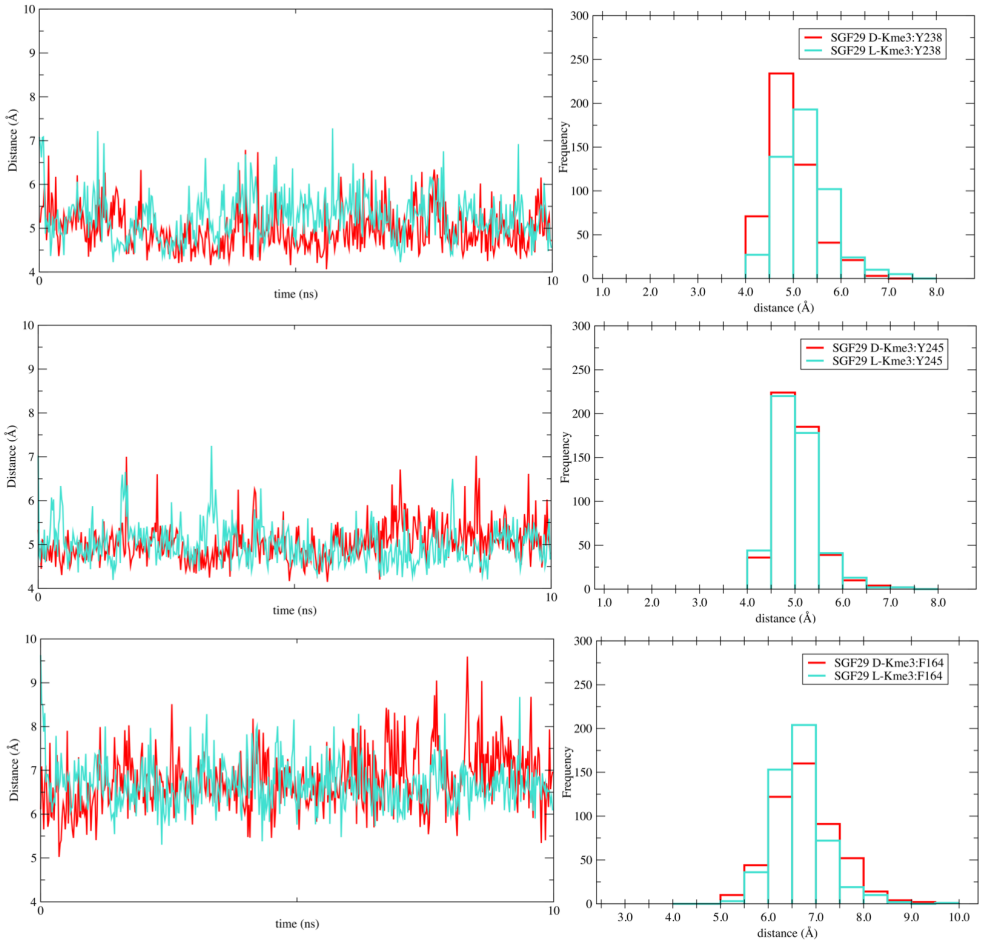


Figure 4.8: Distance vs time plots of $N_\epsilon{}^+$ atom of D-H3K4me3 and L-H3K4me3 to Y238, Y245, and F164 side chains over 10 ns.

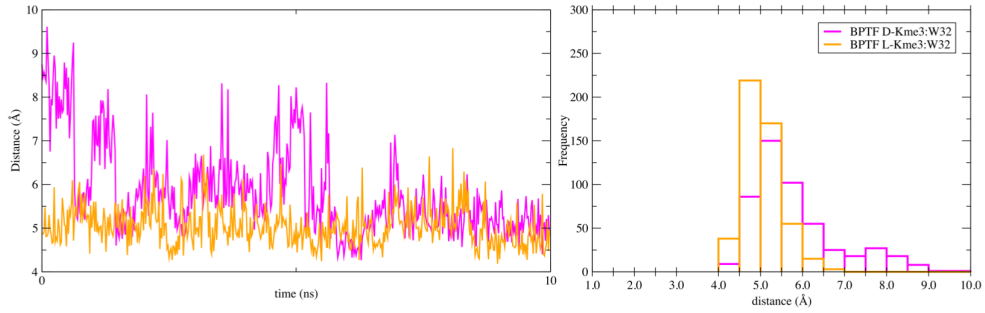


Figure 4.9: Distance vs time plots of $N_\epsilon{}^+$ atom of D-H3K4me3 and L-H3K4me3 to W32 side chain over 10 ns.

Overall, behaviour of the systems as a result of altering the stereochemistry of the substrate yielded similar results to the wild type following analysis by MM methods. Distance calculations between the $N_\epsilon{}^+$ atom of KMe3 and surrounding aromatic residues suggest the presence of favourable stabilising cation–π interactions for both stereoisomers. Examination of the H3 orientation for D-KMe3 complexed with readers $KDM5A_{PHD3}$ and $TAF3_{PHD}$ emphasise the flexibility of the histone backbone to potentially prioritise the cation–π interactions.

4.3.2 Investigating Length Selectivity

Section 4.3.1 discussed the employment of MD simulations to analyse differences in stereoselectivity between L-KMe3 and D-KMe3 for five reader proteins. I applied a similar method to elucidate the binding of OrnMe3 and hKMe3 to the same panel of enzymes.

Benchmarking MD Parameters

In order to sufficiently validate simulations involving OrnMe3 and hKMe3, long-time MD was performed to confirm stability. The necessity for a longer timescale was important because there were no representative crystal structures available of bound OrnMe3 and hKMe3 to a reader enzyme, and modelling the substrates required more extensive modifications *in silico* than inverting the C_α backbone. $KDM5A_{PHD3}$ was chosen for benchmarking because it was the least structured of all five enzymes, and therefore highly dynamic. $KDM5A_{PHD3}$ bound to H3 containing the two lysine analogues were observed over 100 ns. RMSD vs time plots of $KDM5A_{PHD3}$:hKMe3 and $KDM5A_{PHD3}$:OrnMe3 are shown in **Fig. 4.10** and **Fig. 4.11**, respectively.

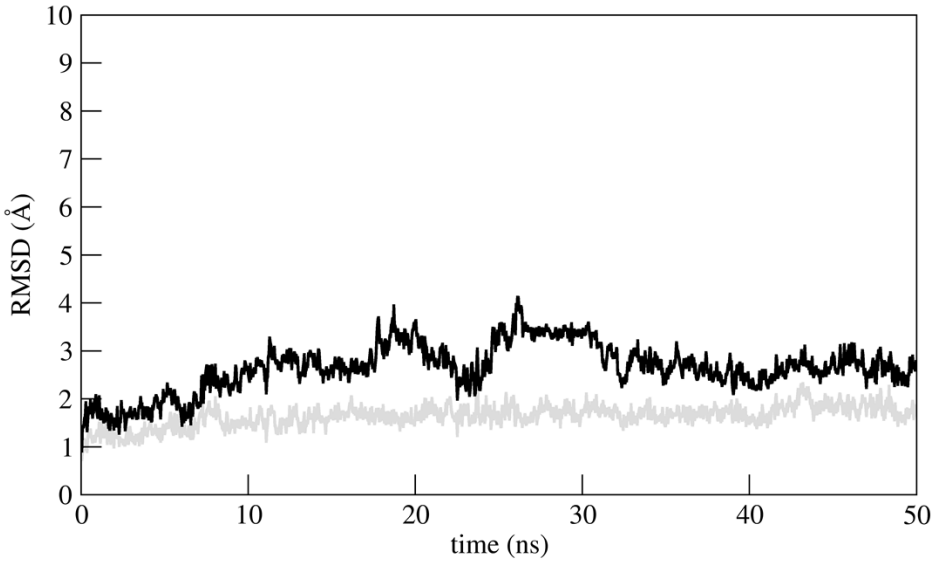


Figure 4.10: RMSD plot of KDM5A$_{PHD3}$ complexed with hKMe3 Cα atoms over 50 ns (grey, 1.64 ± 0.23), and including random coil KDM5A$_{PHD3}$ and H3 terminal residues 47-52, 58-61, and 23-25 (black, 2.65 ± 0.51).

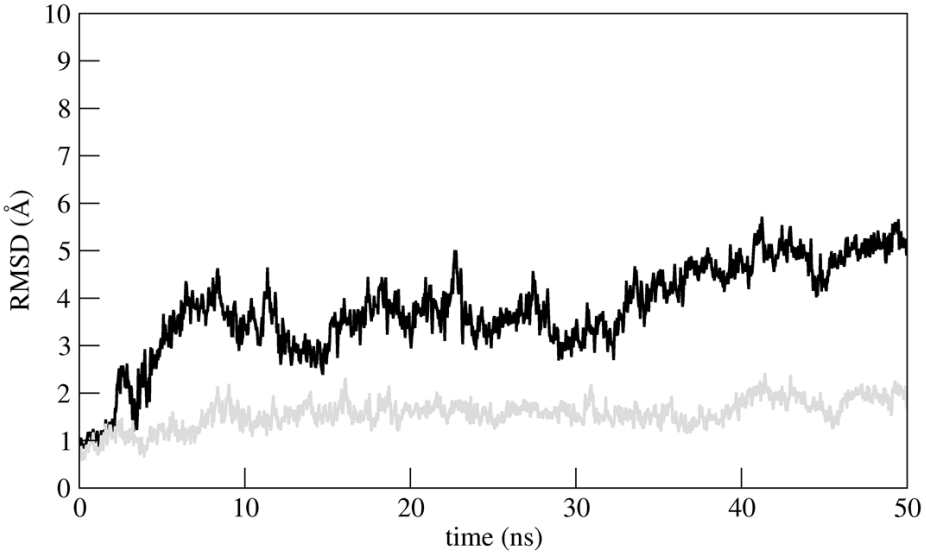


Figure 4.11: RMSD plot of KDM5A$_{PHD3}$ complexed with OrnKMe3 Cα atoms over 50 ns (grey, 1.56 ± 0.30), and including random coil KDM5A$_{PHD3}$ and H3 terminal residues 47-52, 58-61, and 23-25 (black, 3.76 ± 0.98).

Average RMSD values were lower and fluctuated less for KDM5A$_{PHD3}$:hKMe3 (2.65 ± 0.51) than KDM5A$_{PHD3}$:OrnKMe3 (3.76 ± 0.98) over time. Because KDM5A$_{PHD3}$ is composed primarily of unstructured loops, it was important to evaluate if large movements were occurring here or at the active site. Snapshots of the systems at time = 0 ns, 25 ns, and 50 ns were overlaid and presented in three representations, shown in **Fig. 4.12** and **Fig. 4.13**. These structures of KDM5A$_{PHD3}$ bound to H3 containing OrnMe3 and hKMe3 are visualised in three different representations. The RGB spectrum represents different properties. A) Is coloured by sequence for each structure, where red is the C-terminus of H3 and blue is the N-terminus of KDM5A$_{PHD3}$. B) Is coloured by secondary structure where red, yellow, and green represent a β-sheet, α-helix, and random coil, respectively. C) Is coloured by simulation time where the red, yellow, and blue structures show the protein at time 0 ns, 25 ns, and 50 ns, respectively. Red circles highlight highly flexible random coil regions of the protein and green circles highlight the rigid core containing the aromatic TRP2 cage and ligand. This analysis emphasises the stability of the active site core throughout the simulation, and therefore I proceeded with the parameters employed here to simulate the other four reader proteins.

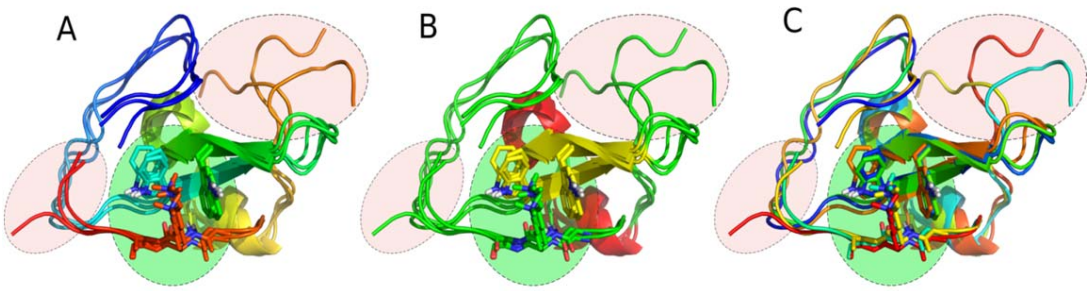


Figure 4.12: Overlaid structures of KDM5A$_{PHD3}$:hKMe3 at times 0 ns, 25 ns, and 50 ns where RGB spectrum represents different properties. A) Is coloured by sequence, C-terminus of H3 (red) to N-terminus of KDM5A$_{PHD3}$ (blue). B) Secondary structure where red, yellow, and green represent β-sheet, α-helix, and random coil, respectively. C) Is coloured by simulation time where red = 0 ns, yellow = 25 ns, and blue = 50 ns. Flexible random coil (red circles) and rigid core containing active site (green circles) are highlighted.

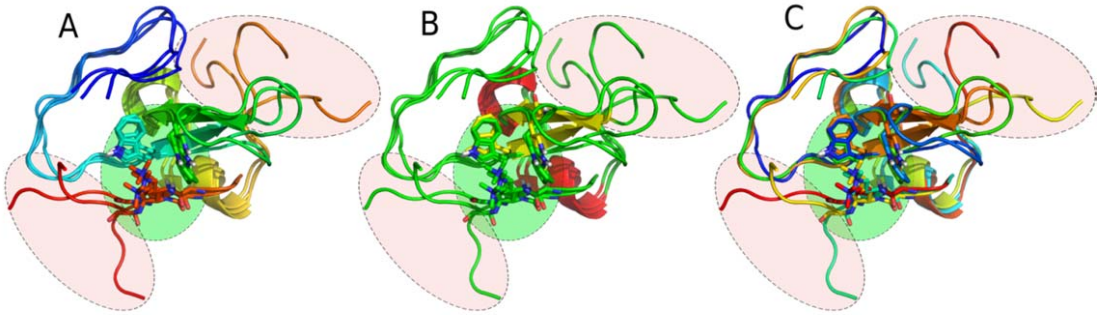


Figure 4.13: Overlaid structures of KDM5A$_{PHD3}$:OrnMe3 at times 0 ns, 25 ns, and 50 ns where RGB spectrum represents different properties. A) is coloured by sequence, C-terminus of H3 (red) to N-terminus of KDM5A$_{PHD3}$ (blue). B) Secondary structure where red, yellow, and green represent β-sheet, α-helix, and random coil, respectively. C) is coloured by simulation time where red = 0 ns, yellow = 25 ns, and blue = 50 ns. Flexible random coil (red circles) and rigid core containing active site (green circles) are highlighted.

10 ns simulations were performed on the remaining reader complexes. RMSD values of the receptors and H3 chains of these trajectories are listed in **Table 4.2**. Similarly to my investigation concerning stereoselectivity, BPTF$_{PHD}$ complexes were observed to have the highest average RMSD values, namely when bound to OrnMe3 (7.10 ± 1.97). However, the low movement of the H3 chain (0.79 ± 0.35) supports the qualitative appearance of stability at the active site. When comparing the two analogues, a trend of greater stability is observed for hKMe3 over OrnMe3 complexes. Large movements are observed in the simulations for H3 when hKMe3 is bound to TAF3$_{PHD3}$ (4.70 ± 1.35), however the receptor (3.15 ± 0.64) was slightly lower than the wild type (3.25 ± 0.82). Analysis of the substrate-aromatic cage distances are necessary to determine if this is prioritising cation–π interactions.

Trajectory Analysis

The binding poses of OrnMe3 and hKMe3 with respect to their aromatic cages are shown in **Fig. 4.14**.

Table 4.2: Average RMSD and error of C_α for reader:hKMe3 and reader:OrnMe3 MD simulations over 10 ns.

		RMSD (Å)	
		Reader	H3
$BPTF_{PHD}$	hKMe3	3.32 ± 0.93	0.71 ± 0.26
	KMe3	5.85 ± 2.49	0.47 ± 0.20
	OrnMe3	7.10 ± 1.97	0.79 ± 0.35
$KDM5A_{PHD3}$	hKMe3	1.77 ± 0.42	1.58 ± 0.44
	KMe3	0.25 ± 0.58	1.16 ± 0.34
	OrnMe3	2.77 ± 1.17	1.23 ± 0.43
$KDM4A_{TTD}$	hKMe3	2.20 ± 0.76	1.96 ± 0.18
	KMe3	2.08 ± 0.64	1.17 ± 0.27
	OrnMe3	3.68 ± 0.89	1.98 ± 0.47
$SGF29_{TTD}$	hKMe3	1.17 ± 0.14	0.57 ± 0.12
	KMe3	1.28 ± 0.20	1.03 ± 0.45
	OrnMe3	1.21 ± 0.14	1.61 ± 0.86
$TAF3_{PHD}$	hKMe3	3.15 ± 0.64	4.70 ± 1.35
	KMe3	3.25 ± 0.82	3.24 ± 1.20
	OrnMe3	3.87 ± 0.64	3.14 ± 0.84

Distance vs time graphs and frequency histograms for each $N_\epsilon{}^+$ to the centroid of each aromatic residue side chain are shown in **Fig. 4.15-4.19**. When analysing the trajectories, ligands bound to KDM4A were able to mimic the WT closely in terms of cation–π occupation. The other simulations, however, showed large deviances of the trimethylammonium ion from the aromatic cage pocket for both OrnMe3 and hKMe3. It was at first unclear how these interactions could be so dynamic, while still allowing for recognition by the readers. Specifically, simulations of $TAF3_{PHD}$ (**Fig. 4.17**) and $KDM5A_{PHD3}$ (**Fig. 4.16**) showed the positioning of $N_\epsilon{}^+$ to be highly dynamic relative to the aromatic cage. In the case of $SGF29_{TTD}$, hKMe3 maintained a distance of 8.5 - 10.0 Å away from residue F264, while OrnMe3 failed to interact with Y238 (**Fig. 4.18**). It became clear that consideration of the size in area of the aromatic cage could allow greater flexibility to allow maintenance of a cation–π interaction. In **Fig. 4.20**, the significant differences in the aromatic cage configuration for all five readers is visualised.

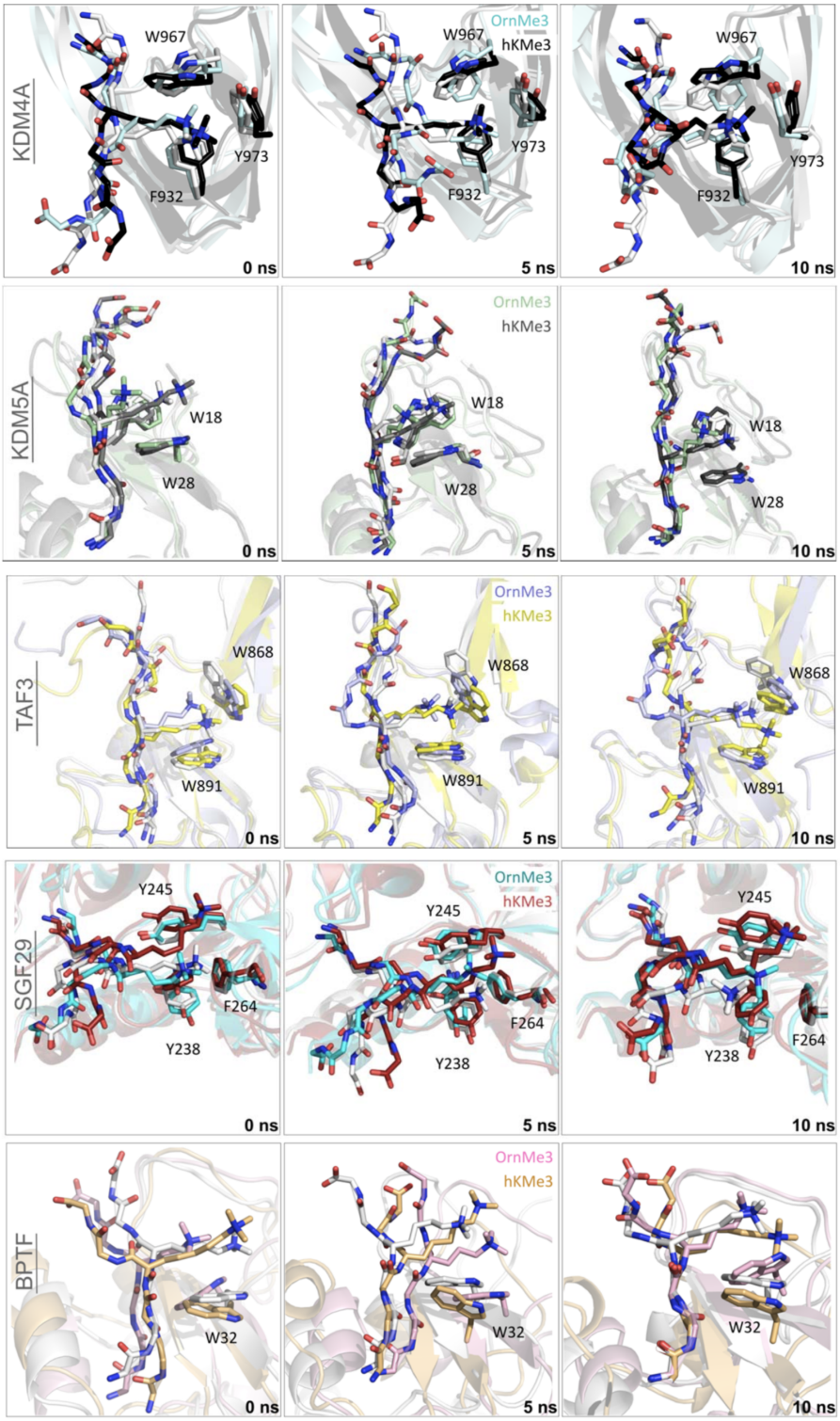
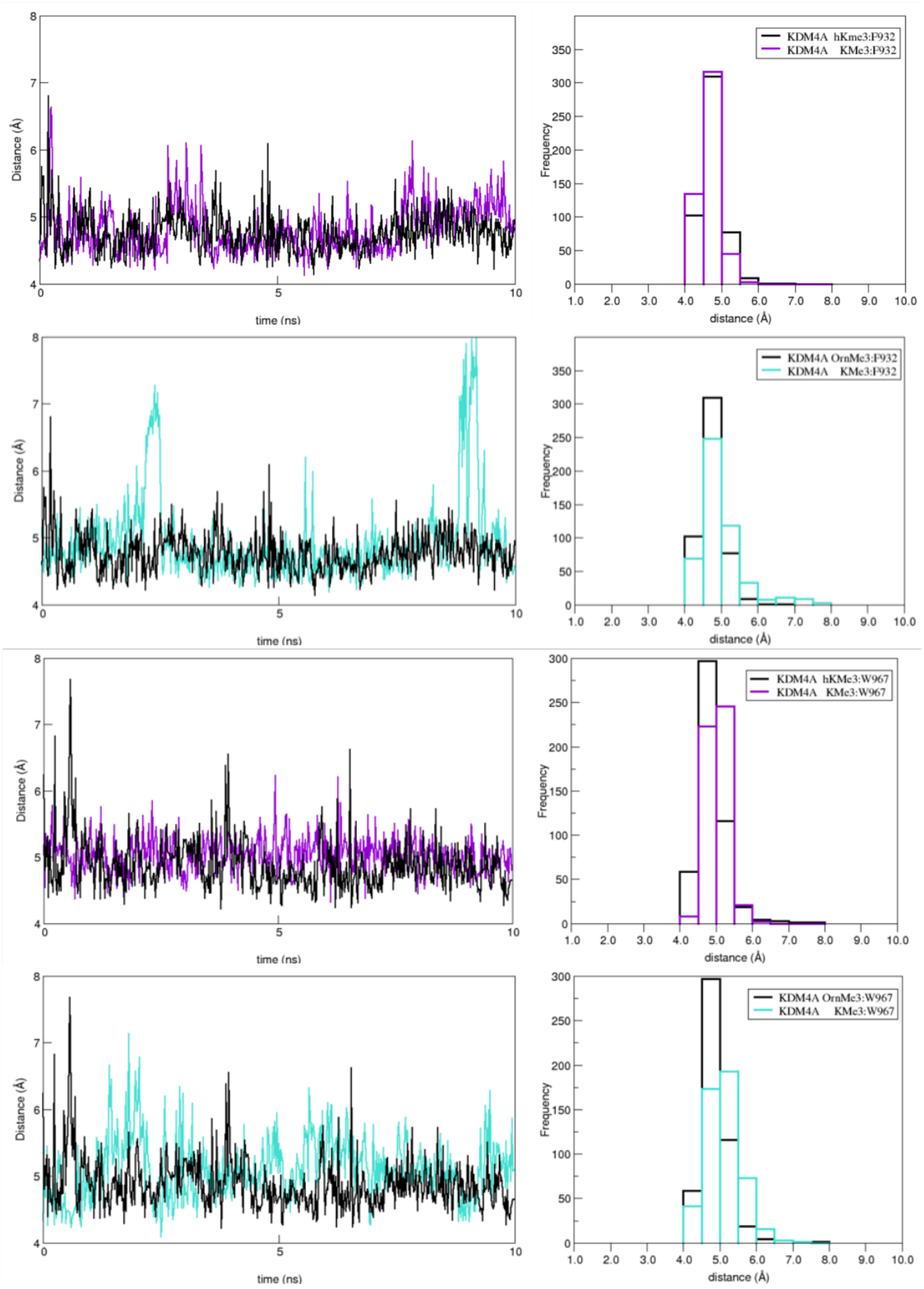


Figure 4.14: Snapshots of readers complexed with OrnMe3, hKMe3, and WT (white) at times 0 ns, 5 ns, and 10 ns. The substrate, aromatic cage side chain, and H3 backbone atoms are shown in sticks.


Distance (Å)
time (ns)
Frequency
distance (Å)
KDM4A hKme3:F932
KDM4A KMe3:F932
KDM4A OrnMe3:F932
KDM4A KMe3:F932
KDM4A hKMe3:W967
KDM4A KMe3:W967
KDM4A OrnMe3:W967
KDM4A KMe3:W967

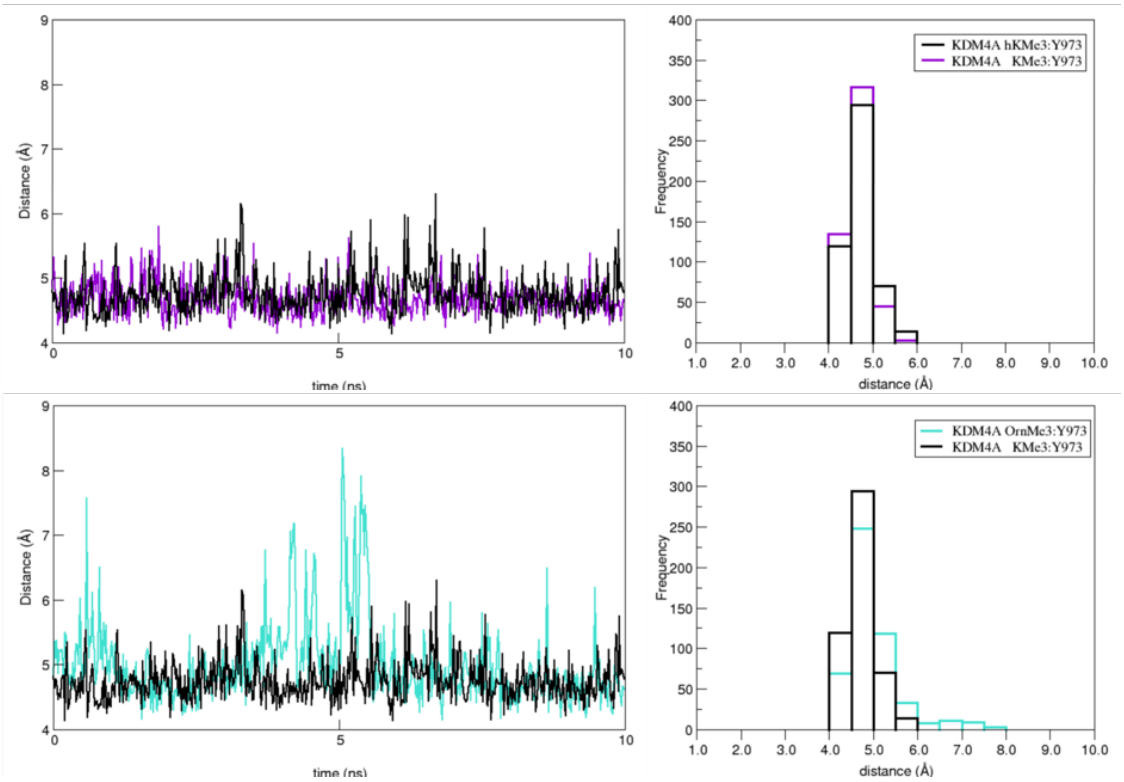


Figure 4.15: Distance vs time plots of $N_{\epsilon}{}^{+}$ atom of hKMe3 and OrnMe3 to F932, W967, and Y973 side chains over 10 ns.

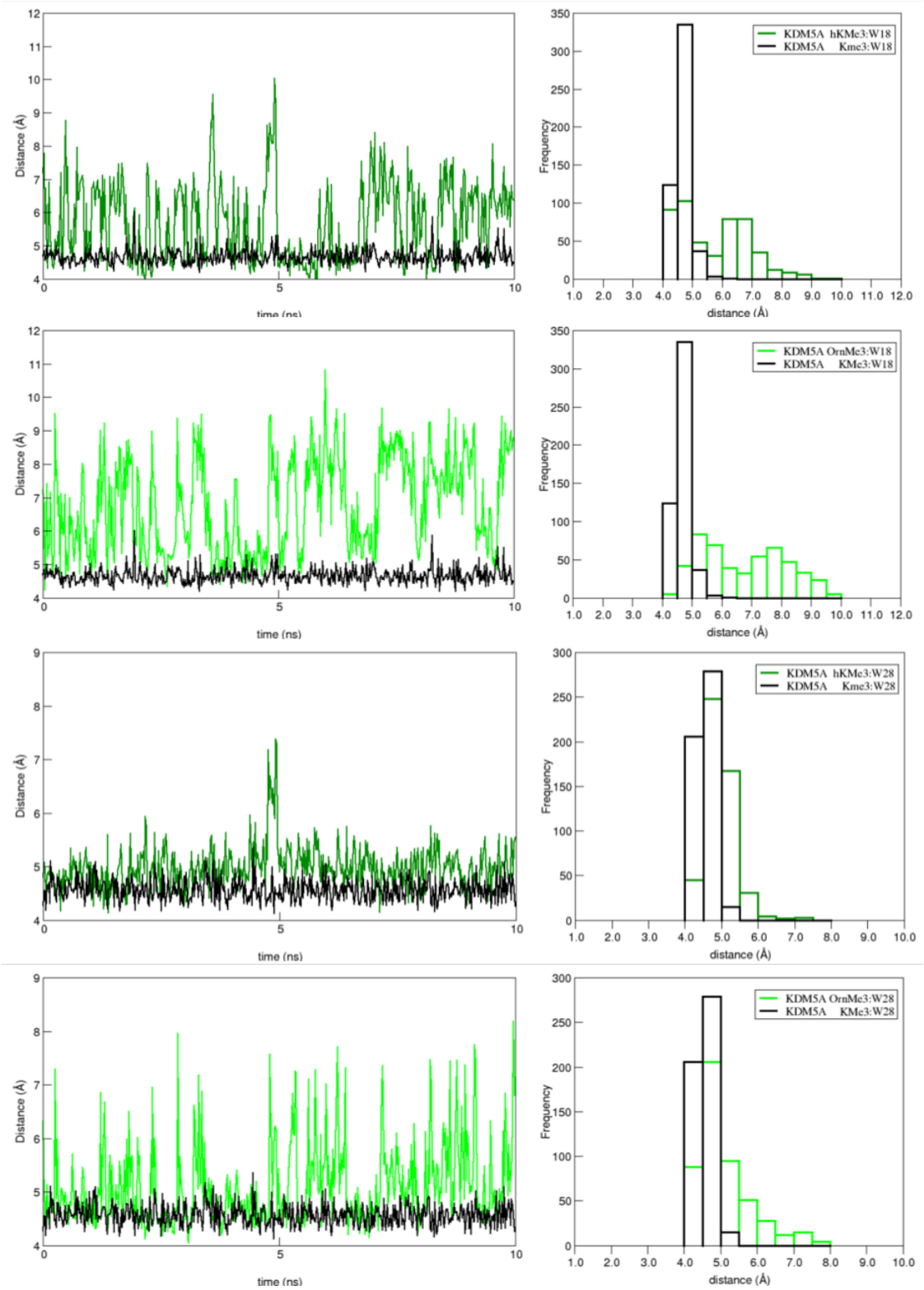


Figure 4.16: Distance vs time plots of $N_\epsilon{}^+$ atom of hKMe3 and OrnMe3 to W18 and W28 side chains over 10 ns.

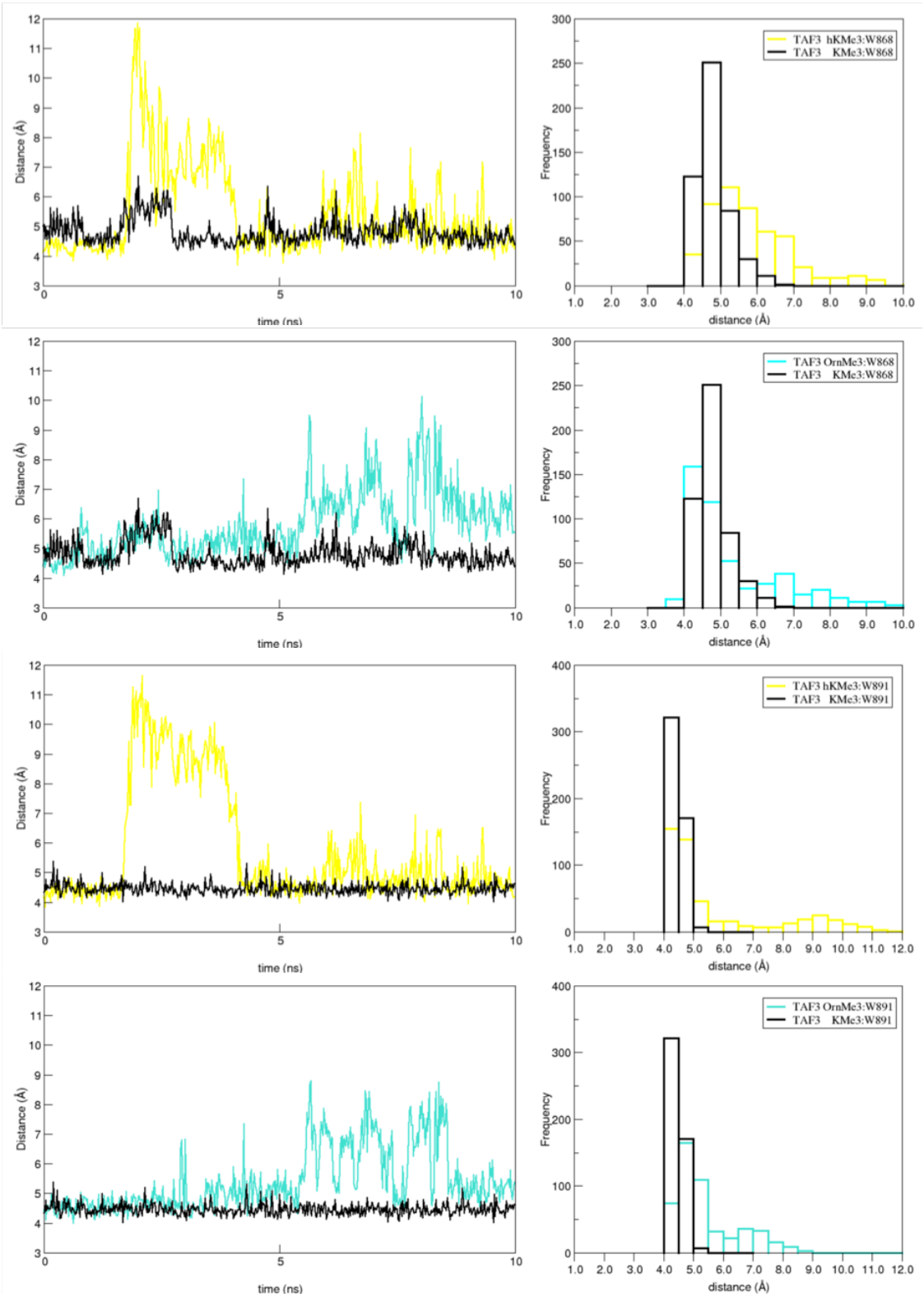
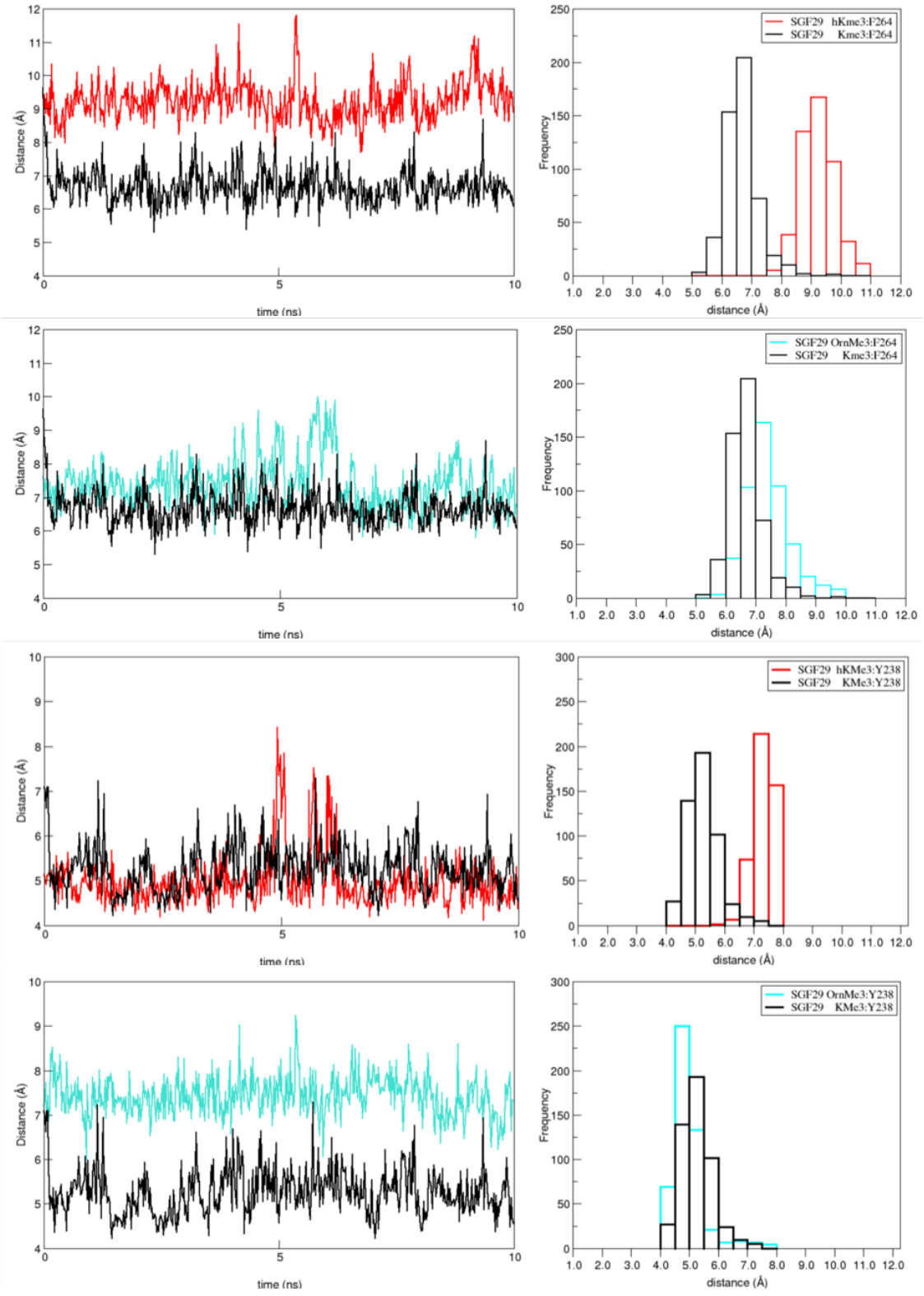


Figure 4.17: Distance vs time plots of $N_{\epsilon}{}^{+}$ atom of hKMe3 and OrnMe3 to W868 and W891 side chains over 10 ns.


Distance (Å)
time (ns)
Frequency
distance (Å)
SGF29 hKme3:F264
SGF29 Kme3:F264
SGF29 OrnMe3:F264
SGF29 Kme3:F264
SGF29 hKMe3:Y238
SGF29 KMe3:Y238
SGF29 OrnMe3:Y238
SGF29 KMe3:Y238

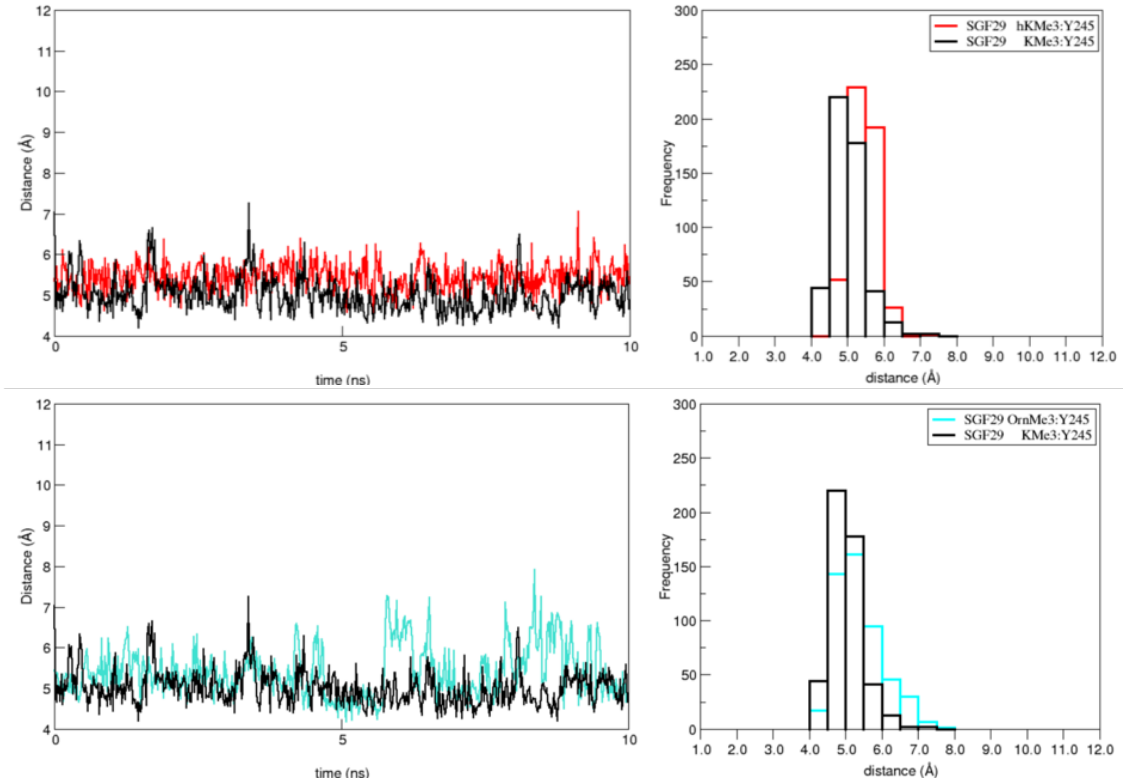


Figure 4.18: Distance vs time plots of $N_\epsilon{}^+$ atom of hKMe3 and OrnMe3 to Y238, Y245, and F164 side chains over 10 ns.

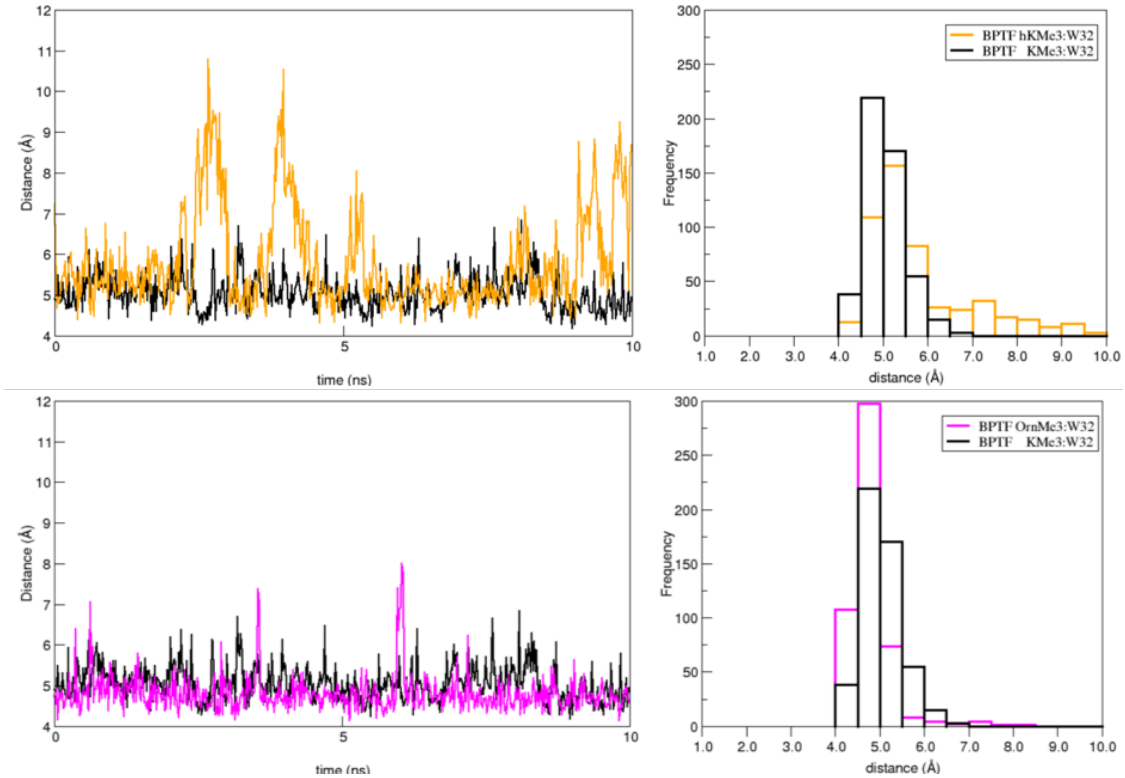


Figure 4.19: Distance vs time plots of $N_\epsilon{}^+$ atom of hKMe3 and OrnMe3 to W32 side chain over 10 ns.

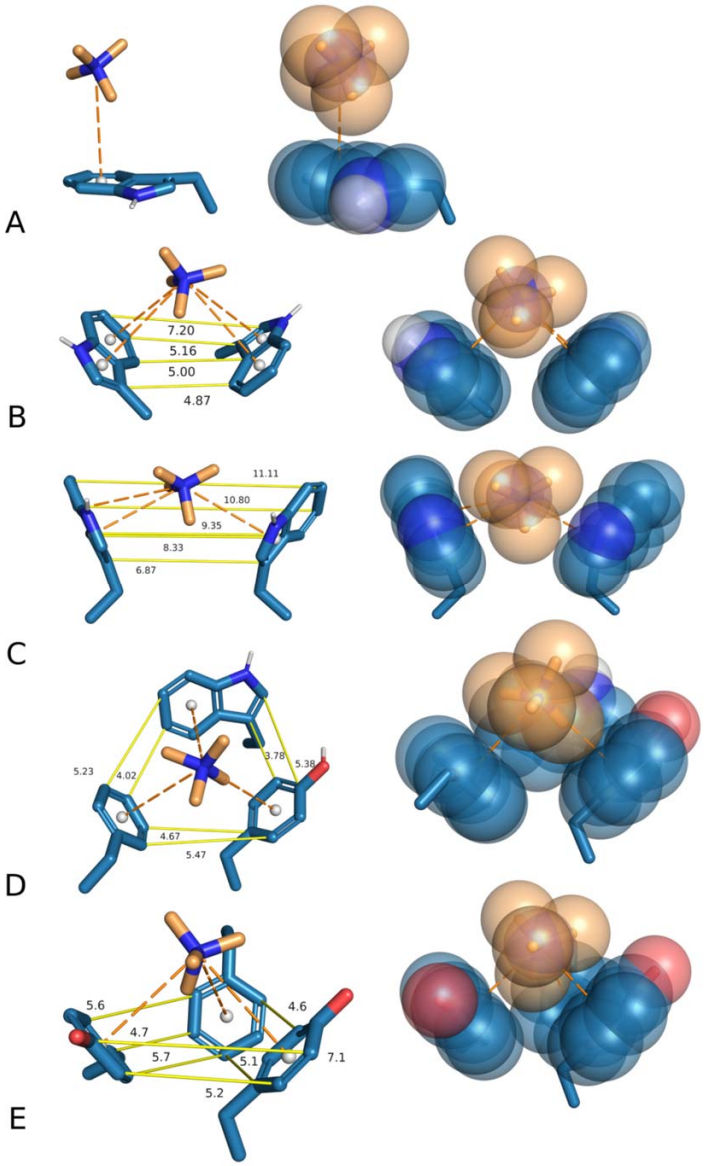


Figure 4.20: Schematic and van der Waals visualisation highlight the different cation-π interactions with aromatic cage of readers A) $BPTF_{PHD}$, B) $KDM5A_{PHD3}$, C) $TAF3_{PHD3}$, D) $KDM4A_{TTD}$, and E) $SGF29_{TTD}$. Distances in Å. Structures taken from crystal structures.

$KDM5A_{PHD3}$ is characterised by having an aromatic cage composed of two tryptophans. Interestingly, a slight preference for W28 over W18 is shown by both residues (**Fig. 4.16**). Reader $TAF3_{PHD}$ also contains dual adjacent facing tryptophan residues, however, unlike in $KDM5A_{PHD3}$ the side chains are positioned as mirror images rather than symmetrically (**Fig. 4.20** C). This creates a much wider conical shape that extends up to 11.1 Å, explaining the large deviation seen in the time vs distance plots (**Fig. 4.17**). $KDM4A_{TTD}$ and $SGF29_{TTD}$ have a Phe–Trp–Tyr and a Tyr–Phe–Tyr aromatic cage, respectively, creating a basket shape (**Fig. 4.20** D, E). The $KDM4A_{TTD}$ binding site appears to be less rigid than $SGF29_{TTD}$, where the latter is composed of equidistant substituted benzenes. This parallels the differences seen for $SGF29_{TTD}$ OrnMe3, which is observed to be short for Tyr238, but stable. Additionally, a similar pattern is shown by hKMe3 and $SGF29_{TTD}$ F264 (**Fig. 4.18**), where due to a longer chain it stabilises outside the centre of the aromatic cage.

4.4 Conclusion

Here, I discuss the results of collaborations with several experimental groups where we explored the effects of altering H3KMe3. The binding of D-KMe3, OrnMe3, and hKMe3, to epigenetic reader proteins was investigated. Experimental data indicated the ability of readers to recognise these analogues over writer and eraser proteins. MD simulations of $BPTF_{PHD}$, $TAF3_{PHD}$, $KDM5A_{PHD3}$, $KDM4A_{TTD}$, and $SGF29_{TTD}$ complexed with the three ligands and compared to the natural substrate, KMe3, revealed important factors necessary to maintain binding. The importance of the H3 peptide tail flexibility to accommodate a stereoisomer of KMe3 was revealed, specifically with $TAF3_{PHD}$ and $KDM5A_{PHD3}$. Binding of shorter and longer KMe3 produced trajectories where the lysine side chain was much more dynamic in comparison. Although occupancy in the aromatic cages for readers with dual Trp-Trp active sites did not meet the < 6.0 Å for a majority of the simulation time, experimentally they were still shown to bind OrnMe3 and hKMe3. This highlighted the importance of considering the large surface area of the mirroring indole side chains, and rigid basket structure of cages comprised of three residues. Overall, the flexibility of the H3 backbone atoms and geometric configuration of the reader active sites indicate a priority to form and maintain a stabilising cation–π interaction.

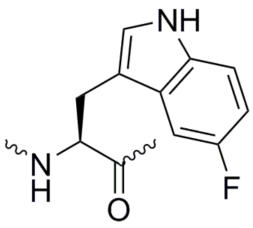
Any sufficiently advanced technology is indistinguishable from magic.

— Arthur C. Clark

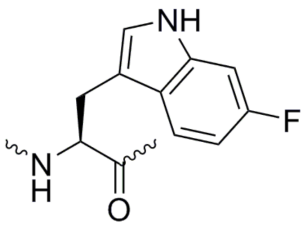
5

Fluorination of KDM5A$_{PHD3}$ Aromatic Cage

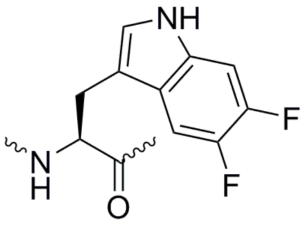
Contents



5F-Trp



6F-Trp



5,6diF-Trp

Pieters B. J. G. E., Wuts M. H. M., Poater J., **Kumar K.**, et al. The recognition of trimethyllysine by the KDM5A PHD3 finger is insensitive to fluorination of the aromatic cage, **2018**, (*Submitted*).

5.1 Introduction

Following investigations analysing reader:H3 complexes given shorter and longer KMe3 analogues, backbone stereochemistry, and a neutral derivative where the trimethyl ammonium is replaced by a tert-butyl group,[106] the importance of favourable cation–π interactions and release of high-energy water molecules for binding has been explored. However, the effects of modifications to the π-system contributed by the aromatic cage residue side chains had not yet been reported.

5.1.1 Fluorinating The Trp18-Trp28 Cage

The examination of cation–π interactions in biomolecular recognition of positively charged ligands using fluorinated tryptophans was pioneered by Dougherty and co-workers.[166,231–237] In our collaborative study with the Mecinović group at Radboud University, attenuations to the π-system and the results on binding were achieved through substitution of the KDM5A$_{\text{PHD3}}$ Trp18-Trp28 cage by fluorinated tryptophan residues. Additions of these electron withdrawing groups were hypothesised to significantly weaken cation–π interactions, but it remained to be seen if other favourable protein-ligand interactions could compensate to allow recognition. Here, three fluorinated substituted tryptophans were introduced to KDM5A$_{\text{PHD3}}$, shown in **Fig. 5.1**. Of all the epigenetic readers, KDM5A$_{\text{PHD3}}$ was the clear choice because a) the aromatic cage only contained two residues; b) no other tryptophan residues are present outside the active site in the entire domain; and c) KDM5A$_{\text{PHD3}}$ recognises H3K4me3 with submicromolar binding affinity. The choice of substitution by fluorine atoms to reduce the electron density of the indole rings was also important to minimise structural perturbations of the protein due to their comparability in size.

Work described in this Chapter necessitated computational skills described in both **Chapter 2** and **Chapter 4** of this Thesis. Here, I sought to model epigenetic reader KDM5A$_{\text{PHD3}}$ complexed with H3, however with modifications made to the aromatic cage instead of the KMe3 substrate. Effects on electrostatic interactions required accurate modelling of the different aromatic substituents, drawing on expertise gained from simulating the 15 **L1-L15** heteroaromatic inhibitors bound to CREBBP, five of which contained halogen atoms. The result was a comprehensive experimental and computational investigation to understand the effects of fluorination to the Trp-Trp reader aromatic cage.

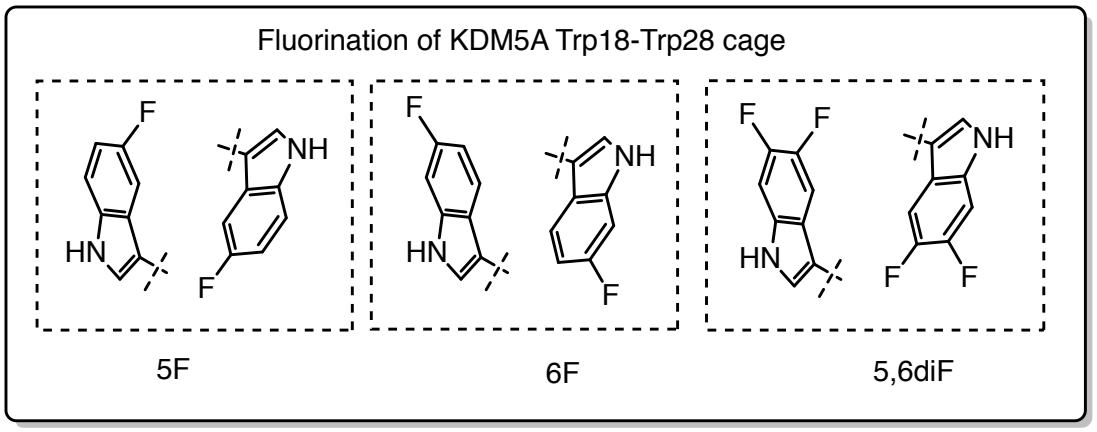


Figure 5.1: Structure of 5F, 6F, and 5,6diF KDM5A$_{PHD3}$ modified Trp18-Trp28 aromatic cages used to probe the strength of cation–π interactions.

5.2 Computationally Modelling A Modified Aromatic Cage

Four MD simulations were carried out for 100 ns each. A PDB structure for the model representing KDM5A PHD3 (PDB: 2KGI) was used as a template for building the reader:KMe3 systems. KDM5A$_{PHD3}$ residues Trp18 and Trp28 were manually modified to generate the 5F-Trp18/5F-Trp28, 6F-Trp18/6F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28 systems. AMBER12[159] was used with the Amberff12SB force field to define protein partial charges. Hydrogen atom addition was performed with LEaP. Systems were solvated in a 10 Å truncated octahedral box of TIP3P water[163] and neutralised explicitly with either sodium or chloride counterions. Atomic partial charges for 5F-Trp, 6F-Trp, and 5,6diF-Trp correspond to the RESP charges (see **D.1**). The structures for the modified residues were constructed manually. The final systems were minimised for 1,000 cycles of steepest-descent minimisation followed by 1,000 cycles of conjugate-gradient minimisation to remove close van der Waals contacts using the sander program in AMBER12. All simulations used a dielectric constant of 1.0, Particle Mesh Ewald summation[230] to calculate long-range electrostatic interactions, and bond-length constraints applied to all bonds to H atoms. Trajectories were saved at 20-ps intervals and visualised using VMD.[164] Electrostatic energies between the modified terminal Kme3 side chain N_ϵ atom and the π-system of surrounding aromatic cages were calculated with the NAMD Energy

Plugin 1.4.[177] For these calculations, the π-system was defined for tryptophan and modified tryptophans as the side chain indole ring (non-H) atoms. Energy values were measured every 20 ps and averaged over 10 ns.

5.3 Results and Discussion

5.3.1 ITC Data

Surprisingly, experimental ITC data determined no significant differences in the free energy of binding for all three fluorinated systems when compared to the wild type (**Table 5.1**). A similar decrease in $\Delta\Delta H^\circ$ of 1.0 ± 0.1 kcal mol^{-1} was observed for both 5F-Trp and 6F-Trp, however this was compensated by an increase in $-T\Delta\Delta S^\circ$ of –1.0 ± 0.1 kcal mol^{-1}.

Table 5.1: ITC data for epigenetic reader for substrate H3K4Me3 complexed with the epigenetic read protein KDM5A containing the wild type, 5F, 6F, and 5,6diF aromatic cage.

	K_d (nM)	ΔG° (kcal mol^{-1})	ΔH° (kcal mol^{-1})	$-T\Delta S^\circ$ (kcal mol^{-1})
WT-KDM5A	48 ± 8	–10.0 ± 0.1	–11.9 ± 0.1	–1.9 ± 0.1
5F-KDM5A	49 ± 12	–10.0 ± 0.1	–10.9 ± 0.1	–0.9 ± 0.2
6F-KDM5A	57 ± 17	–9.9 ± 0.2	–12.1 ± 0.1	–2.2 ± 0.2
5,6diF-KDM5A	70 ± 15	–9.8 ± 0.1	–11.7 ± 0.2	–1.9 ± 0.2

5.3.2 Stability of Simulations Over 100 ns

After experimentally determining that H3K4me3 binds to fluorinated PHD3 fingers of KDM5A, I carried out four long-time MD simulations to observe the flexibility of the reader and effects of fluorination on key interactions for binding. This included the wild type and three systems containing mutated Trp18-Trp28 aromatic cages: 5F-Trp18/5F-Trp28, 6F-Trp18/6F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28.

The average RMSD values for each simulation are listed in **Table 5.2**. 5,6diF-KDM5A (1.5 ± 0.7), 6F-KDM5A (1.7 ± 0.7), WT-KDM5A (2.2 ± 0.7), and 5F-KDM5A (2.6 ± 0.5) were of similar stability. In principle, it was surprising

that the wild type simulation was more stable than the mutated systems because it was simulated directly from the crystal structure. In this case it was important to consider the tertiary structure of KDM5A$_{\text{PHD3}}$. In **Fig. 5.2**, the global structural features of each system overtime can be qualitatively analysed where the RGB colouration corresponds to snapshots at 0 ns, 10 ns, 50 ns, and 100 ns, respectively. Because PHD3 domain of KDM5A is relatively small, the average RMSD is more easily skewed due to local movements in terminal or loop regions. Furthermore, the protein contains only two small β-sheets and one α-helix, but is otherwise predominately composed of extremely flexible unstructured loops.

Table 5.2: Average C_α RMSD values for 100 ns MD simulations.

Protein	RMSD (Å)
WT-KDM5A	2.2 ± 0.7
5F-KDM5A	2.6 ± 0.5
6F-KDM5A	1.7 ± 0.7
5,6diF-KDM5A	1.5 ± 0.7

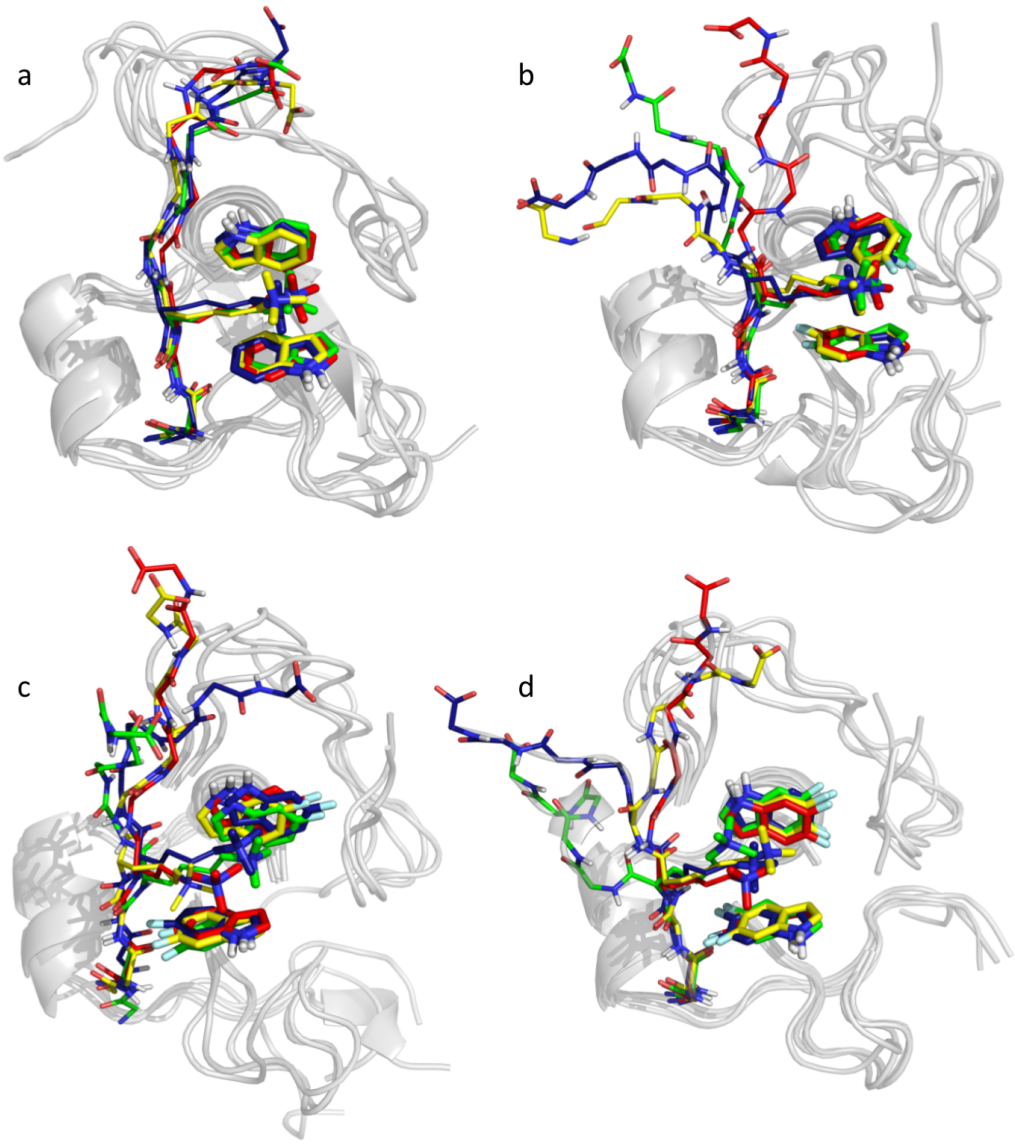


Figure 5.2: Visualisation of molecular dynamics simulation for a) wild type KDM5A$_{PHD3}$ bound to H3K4me3 (PDB: 2KGI), and mutated KDM5A$_{PHD3}$ containing b) 5F-Trp18/5F-Trp28, c) 6F-Trp18/6F-Trp28, and d) 5,6diF-Trp18/5,6diF-Trp28. Reader protein (grey, cartoon) accommodates the flexible H3 chain (RGB, lines) where the terminal amine of H3K4me3 occupies the Trp18-Trp28 aromatic cage (RGB, sticks). Red, yellow, green, and blue represent structures at 0 ns, 10 ns, 50 ns, and 100 ns, respectively.

From **Fig. 5.2** it is clear that overall movement of the protein complex is reasonable over 100 ns. Interestingly, the greatest difference that can be observed qualitatively is behaviour of the H3 residues. For KDM5A$_{PHD3}$ containing 5F-Trp18/5F-Trp28, and 5,6diF-Trp18/5,6diF-Trp28, behaviour of terminal H3 residues shows great flexibility, compared to the wild type simulation that shows little difference in the H3 backbone geometry. This dynamic movement is observed primarily for the H3 residues farther away from the KDM5A$_{PHD3}$ Trp-Trp cage. This contrasts with the placement of the KMe3 terminal trimethylammonium group, where lateral movement does not extend far beyond the indole side chains as observed in **Fig. 5.3**.

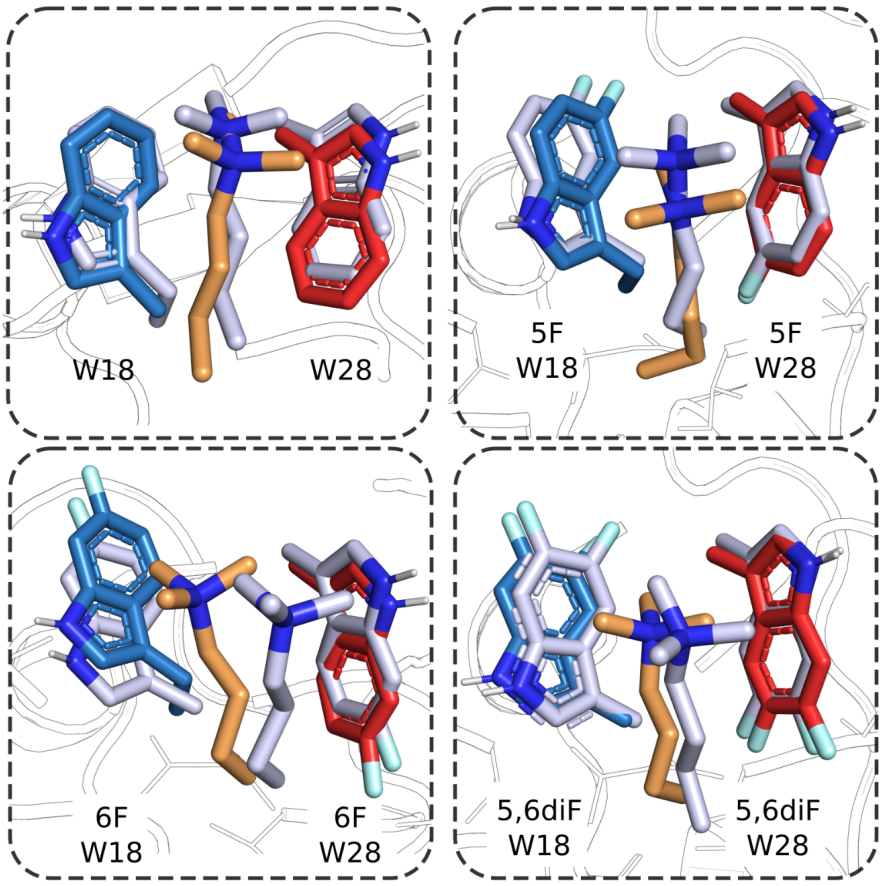


Figure 5.3: Visualisation above KMe3 at time 0 ns (light blue) and 100 ns (orange) relative to Trp18 (blue) and Trp28 (red) KDM5A$_{PHD3}$ residues.

5.3.3 Effects on KDM5A$_{\mathbf{PHD3}}$ Binding

To further understand the non-bonding favourable interactions, the distance over time from KMe3 to each of the aromatic ligands was computed to quantify the occupancy of present cation–π interactions as well as the average ΔE_{ele} values. In **Fig. 5.4**, we can evaluate the frequency of a cation–π interaction per aromatic cage residue as compared to the wild type. Non-covalent cation–π interactions are shown to be present for H3K4me3 and both Trp residues in each system using an established geometric cut-off of 6.0 Å. Interestingly, a bimodal distribution is observed with KMe3 and 6F-Trp18 around 5 Å and 7.5 Å. Kme3 and 5,6-Trp28 exhibit the closest behaviour to the wild type than the other five plots.

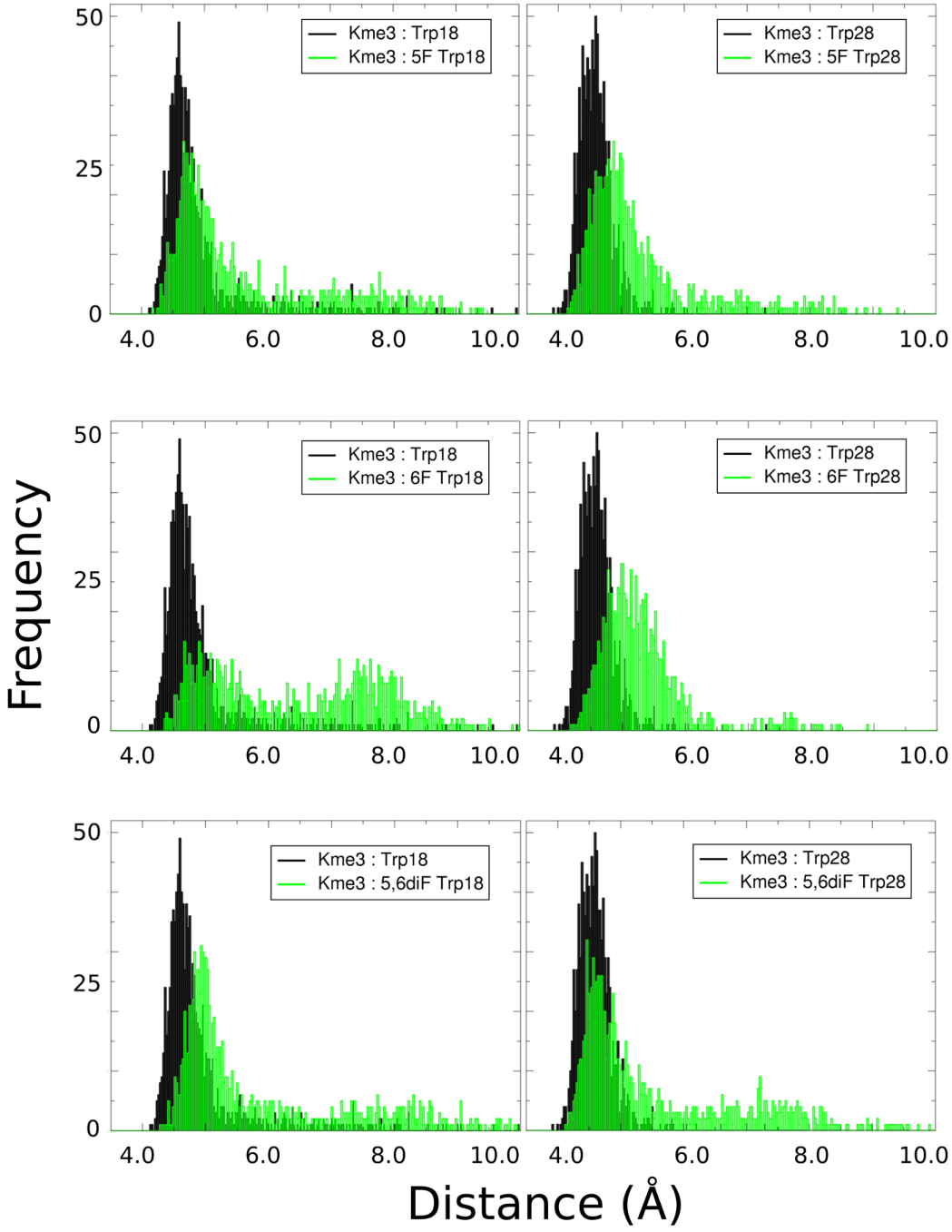


Figure 5.4: Frequency of cation–π interactions for each system given a < 6.0 Å geometric threshold. Black histograms represent the wild type.

Analysis of the distance vs time plots describes all the movement around the active site to be extremely dynamic, as shown in **Fig. 5.5 - 5.7**. Although large movements of the trimethylammonium relative to the side chains are observed, they are uniform throughout the simulation. This is in good agreement with the findings of KMe3 analogues and KDM5A$_{\mathrm{PHD3}}$ in **Chapter 4**, in which the geometry of the aromatic cage allows for a large area for binding. Combined with the flexibility of the H3 tail backbone, KDM5A$_{\mathrm{PHD3}}$ exhibits a unique ability to accommodate modifications to not only KMe3, but its own residues.

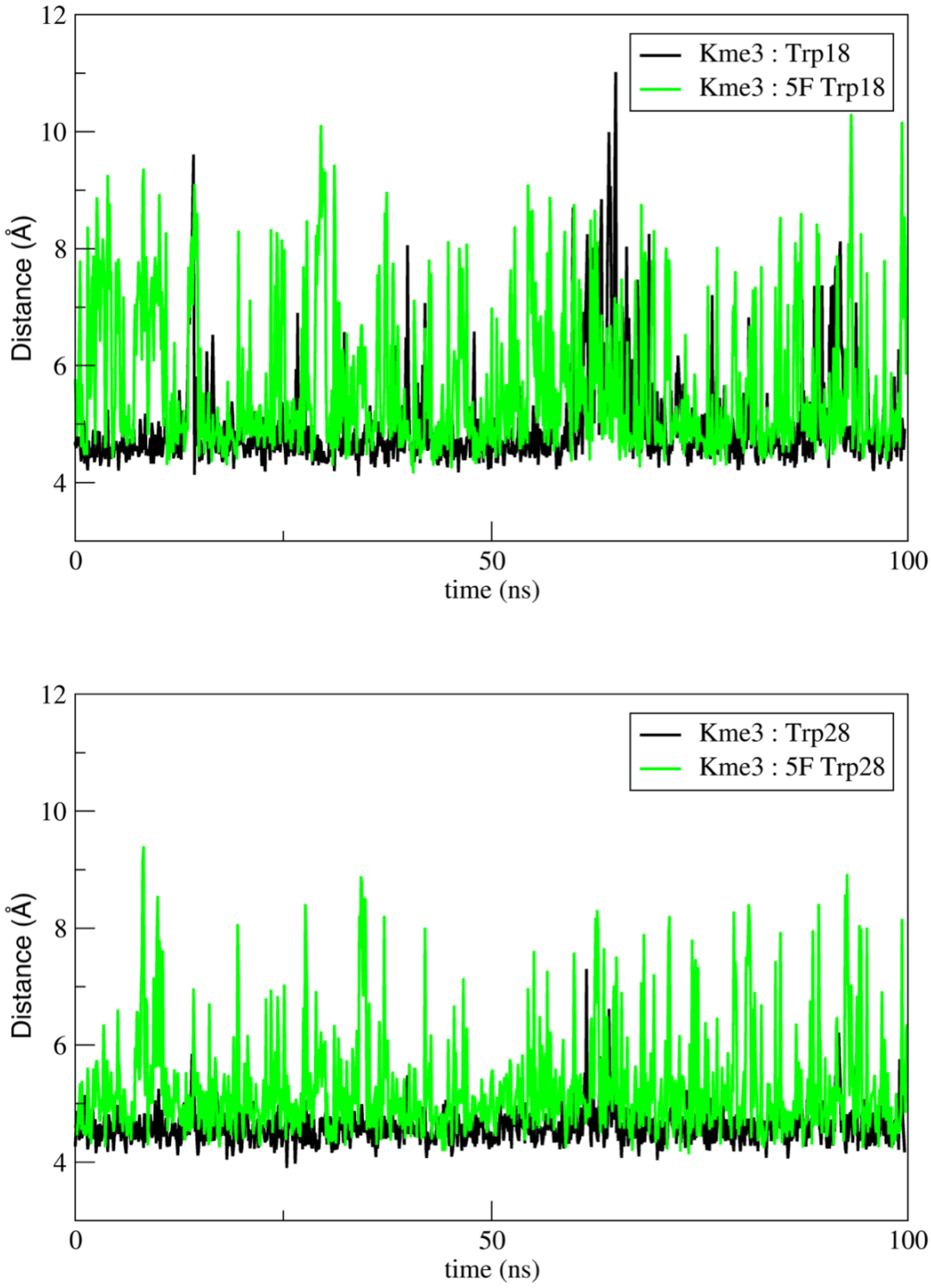


Figure 5.5: Distance vs time plots of N_ϵ^+ atom of KMe3 to Trp18, Trp28, 5F-Trp18, and 5F-Trp28 side chain center of mass over 100 ns.

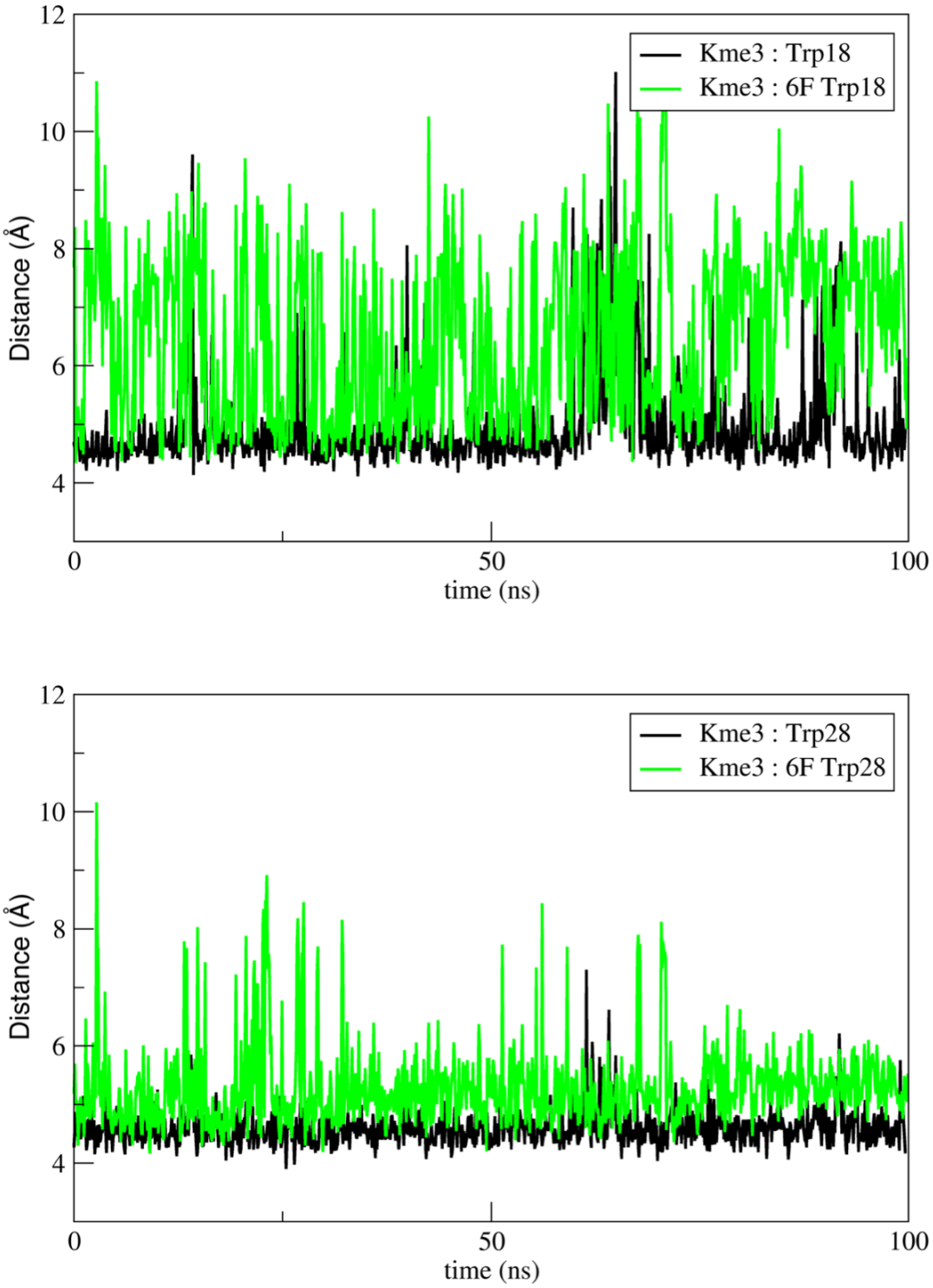


Figure 5.6: Distance vs time plots of $N_\epsilon{}^+$ atom of KMe3 to Trp18, Trp28, 6F-Trp18, and 6F-Trp28 side chain center of mass over 100 ns.

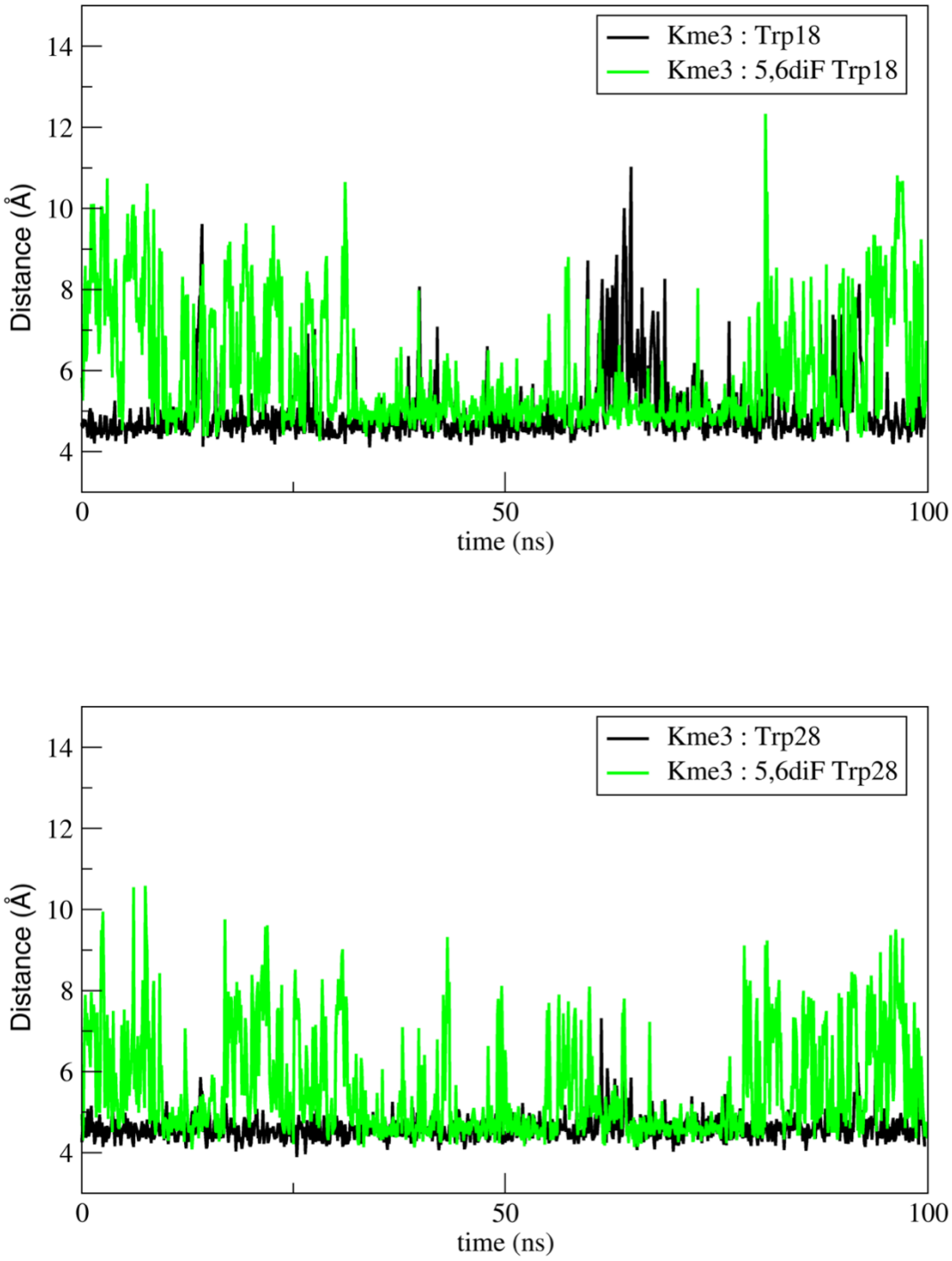


Figure 5.7: Distance vs time plots of $N_\epsilon{}^+$ atom of KMe3 to Trp18, Trp28, 5,6diF-Trp18, and 5,6diF-Trp28 side chain center of mass over 100 ns.

In all cases of fluorinated KDM5A$_{PHD3}$, H3K4me3 is able to occupy the aromatic cage throughout the simulation. Flexibility of the H3 chain to prioritise this interaction is demonstrated for the mutated systems. To quantify the strength of these energetically favourable cation–π interactions, average ΔE_{ele} values were calculated between the quaternary ammonium cations of H3K4Me3 to each aromatic side chain of Trp18-Trp28, as shown in **Fig. 5.8**. The 5,6diF system is the only simulation where the electrostatics are noticeably stronger for one tryptophan over the other, namely Trp28 over Trp18. This is paralleled with the distance vs time plots and of the interaction. This observation might suggest that because the Trp18-Trp28 residues are oriented in an anti-parallel conformation to each other, the KMe3 cation is stabilised by favourable through space interactions by an in-line electron rich five-membered ring.

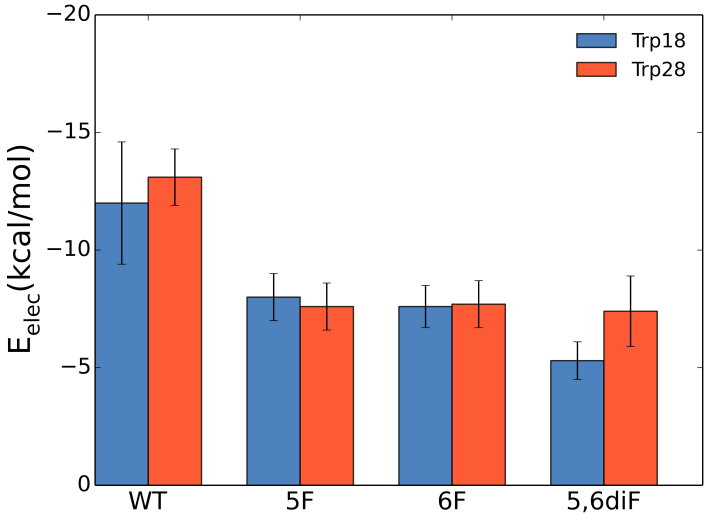


Figure 5.8: Average ΔE_{ele} values over 100 ns from terminal quaternary ammonium cation of H3K4me3 to each aromatic side chain in the Trp18-Trp28 cage.

Effects on ΔE_{ele} from fluorination on the Trp side chain suggest a general trend WT $>$ 5F $\approx$ 6F $>$ 5,6diF when comparing just the indole heavy atoms or with inclusion of the electronegative fluorine substituents (**Table D.2**). This trend agrees with results shown previously in work described in **Chapter 2**, where computational DFT calculations of the Electrostatic Potential Surface (ESP) values taken 4.0

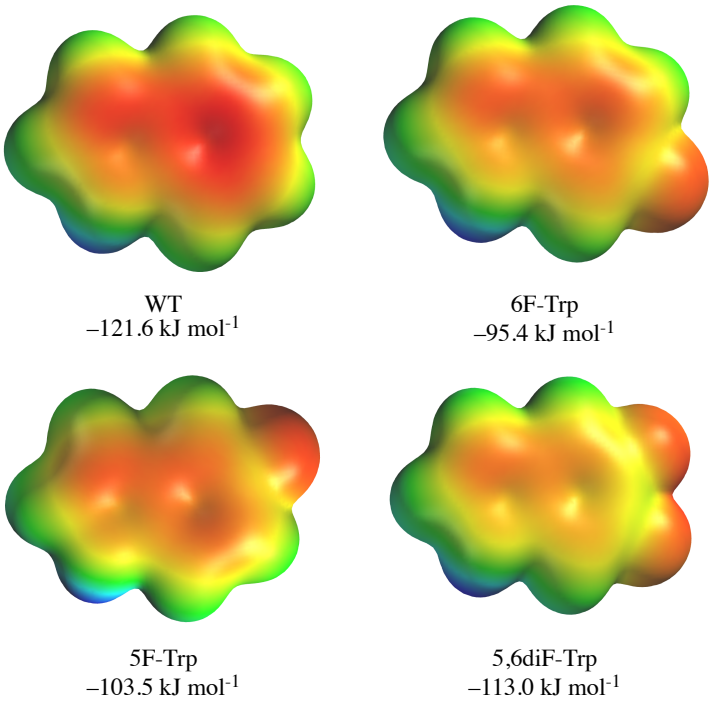


Figure 5.9: Lowest electrostatic surface potential (ESP) of wild type and fluorinated indoles.

Å above aromatic molecules showed that through-space electrostatic interactions between the cation and fluorinated benzene cause a decrease in binding affinity when compared to an indole. Furthermore, qualitative analysis of the ESP for WT, 5F, 6F, and 5,6diF also parallel these results, with the centroid of the indole ring and 5,6diF-Trp ring observed to be the most electron-rich and poor, respectively **(Fig. 5.9).**

Binding of H3K4me3 to 6F-KDM5A lead to differences in ΔE_{ele} values when comparing 6F-Trp18 and 6F-Trp28 and agrees with the bimodal distribution of distance between the cation and π-face. An overall preference for Trp28 over Trp18 is also observed for the other systems, except for KDM5A$_{PHD3}$ containing 5,6diF-Trp where this interaction is almost equal. The stronger interaction between the quaternary ammonium cation with Trp28 has been previously observed for D-H3K4Me3, OrnKMe3, and hKMe3.

Overall, our MD simulation results show that fluorination of the aromatic cage of KDM5A$_{PHD3}$ decreases ΔE_{ele}, but the key cation–π interactions between H3K4me3 and Trp18-Trp28 are prioritised through flexibility of the H3 chain.

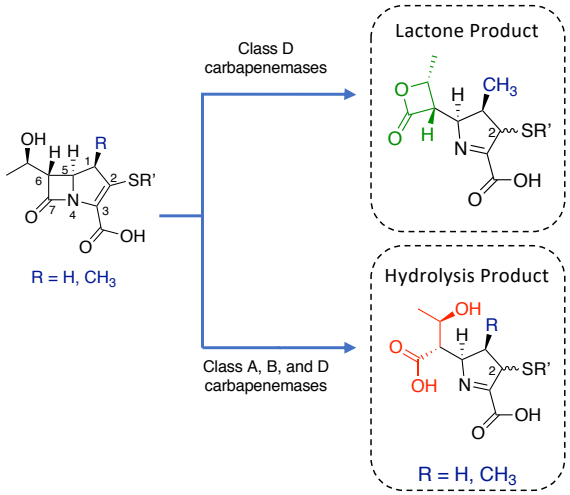
The significant problems we have cannot be solved at the same level of thinking with which we created them.

— Albert Einstein

6

Modified Lysines in Other Applications

Contents



Lohans C.T., van Groesen E., **Kumar K.**, et al. A New Mechanism for β-Lactamases: Class D enzymes degrade 1β-methyl carbapenems through lactone formation, *Angew. Chem. Int. Ed.* **2018**, *57(5)*, 1282-1285.

6.1 A New Mechanism for Degradation by D Class β-lactamases

In the previous chapters, applying computational techniques to provide solutions for epigenetic questions has been the primary goal. Parameters to build realistic simulations of modified aromatic cages or H3K4me3 analogues, such as D-KMe3, ornithine, and homolysine, bound to epigenetic reader proteins were developed and validated. Here, applications of these approaches outside of epigenetics will be discussed, specifically to reveal insights on a novel mechanism implicated in antibiotic resistance. This presented an interesting challenge to question the universal nature of these methodologies for investigating interactions between non-epigenetically post-translationally modified lysines and small-molecule inhibitors.

6.1.1 Background of Carbamyl-lysine

The opportunity to validate these approaches beyond the simulation of reader, writer, and eraser proteins was possible with consideration of carbamyl-lysine, shown in **Fig. 6.1** (Ligand ID KCX).

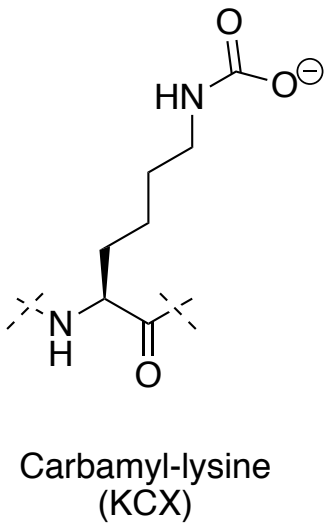


Figure 6.1: Structure of lysine analogue carbamyl-lysine (KCX).

Interestingly, KCX is catalytically active in carbapenemases, a family of enzymes that have been established as key targets for developing new drugs to combat

antimicrobial resistance. Carbapenem small-molecules are considered antibiotics of the last resort.[238,239] However, their potency has been threatened by β-lactamase proteins, a class of carbapenemases. β-lactamases are able to hydrolyse carbapenems in a two step process where an acyl-enzyme intermediate involving a serine residue is formed, following hydrolysis via a nucleophilic water molecule.[240] This established mechanism is shown in red in **Fig. 6.2**.

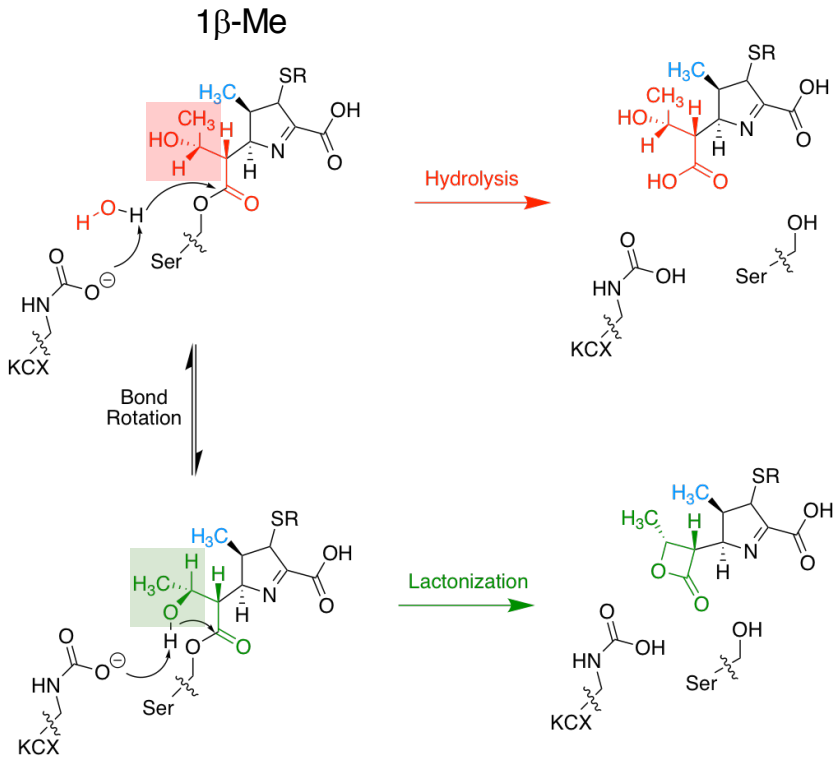


Figure 6.2: Proposed mechanisms for carbapenem 1β-Me hydrolysis and lactone formation. Steric hindrance of the hydroxyethyl group and 1β-Me group (blue) may cause a bond rotation that promotes lactonization. The carbamylated lysine residue (KCX) acts as a general base. The timing of the imine tautomerisation is not defined, but likely occurs (in part) prior to fragmentation of the acyl-enzyme complex. It is unclear based on the products observed whether a particular configuration at C-2 predominates.

From experimental studies undertaken by the Schofield Group, a new mechanism to degrade carbapenems by class D β-lactamases was observed. A lactonization mechanism was proposed by which the KCX also acts as a general base and the ester carbonyl of the acyl-enzyme intermediate is attacked by the hydroxyethyl hydroxy group of the carbapenem, shown in green in **Fig. 6.2**. The result is a carbapenem-lactone derivative product that could represent a novel scaffold for future lactamse inhibitors.

Representative crystal structures were not able to support positioning of the lysine to be favourable for reacting with the hydroxyethyl hydroxy group, although these did not contain a carbamylated lysine due to low pH or mutations.[241–243] Even more interestingly, replacement of the 1-βMe group (blue, **Fig. 6.2**) by a hydrogen atom yielded a complete absence of the lactonization and only the hydrolysis mechanism was observed. The effects of this pyrroline substitution and the role of KCX in lactonization were properties of this novel mechanism I sought to elucidate through longterm MD simulations and DFT calculations.

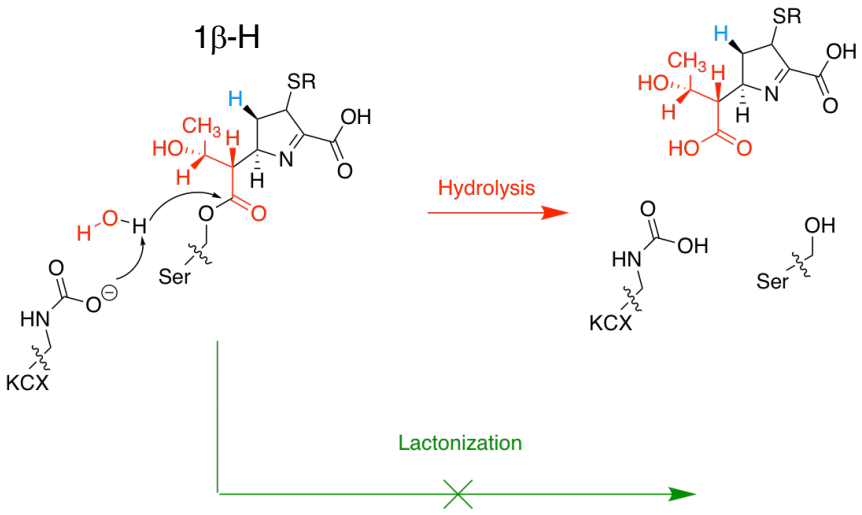


Figure 6.3: Proposed mechanisms for carbapenem 1β-H hydrolysis. The carbamylated lysine residue (KCX) acts as a general base. No lactone formation was observed for carbapenems bearing a 1β-H (blue) as opposed to a methyl, likely due to no bond rotation of the hydroxyethyl group.

6.1.2 Parameters for Simulating Bound Carbapenems

MD simulations were carried out for 100 ns each. Starting structures were taken from PDB: 3ISG[244] and used as a template for building the OXA-1:doripenem systems. AMBER12[159] was used with the Amberff12SB force field to define protein partial charges. Hydrogen atom addition was performed with LEaP. Systems were solvated in a 10 Å truncated octahedron of TIP3P water molecules[163] and were neutralized explicitly with either sodium or chloride counterions. Atomic partial charges for the 1β-Me and 1β-H doripenem-derived ligands and for the carbamylated lysine correspond to the Restrained Electrostatic Potential (RESP)[160] charges, listed in **Table E. 1**. The final systems were minimised for 1,000 cycles of steepest-descent minimisation, followed by 1,000 cycles of conjugate-gradient minimisation to remove close van der Waals contacts using the sander program in AMBER12. All simulations used a dielectric constant of 1.0, Particle Mesh Ewald summation[230] to calculate long-range electrostatic interactions, and bond-length constraints were applied to all bonds to H atoms. Trajectories were saved at 20 ps intervals and visualised using VMD.[164] Density functional theory (DFT) non-covalent interaction (NCI) plots[245] of the 1β-Me and 1β-H doripenem-derived ligands were generated from their wave functions and electron densities at the ωB97XD/def2-TZVP level of theory. Geometries were extracted from a snapshot taken at 40 ns of the MD simulation for the OXA-1:1β-Me doripenem system, manually modified to the 1β-H counterpart, and optimised. These plots allowed the qualitative visualisation of attractive or repulsive non-covalent interactions mapped onto a reduced density gradient (RDG) isosurface, where red and blue represent repulsive non-bonding and strong attractive interactions, respectively.

6.1.3 Behavior of 1β-Me vs 1β-H Binding

MD Trajectory Analysis

Activity of the 1β-Me and 1β-H derivatives of carbapenem with D class β-lactamases yielded different results, with lactonization observed experimentally only for 1β-Me. However, it was uncertain as to what was driving these differences in reactivity.

Following long-time MD simulations, insights into the binding poses of 1β-Me and 1β-H were able to be elucidated. However, initial analysis of the stability of the systems was necessary to validate the results. Stability of the simulations were evaluated by calculating the average RMSD of all backbone C_α atoms over time. Average RMSD values for the OXA-1:1β-Me system indicated stability throughout the simulation (1.89 $\pm$ 0.38, **Fig. 6.4**) as well as the OXA-1:1β-H system (1.90 $\pm$ 0.36, **Fig. 6.5**)

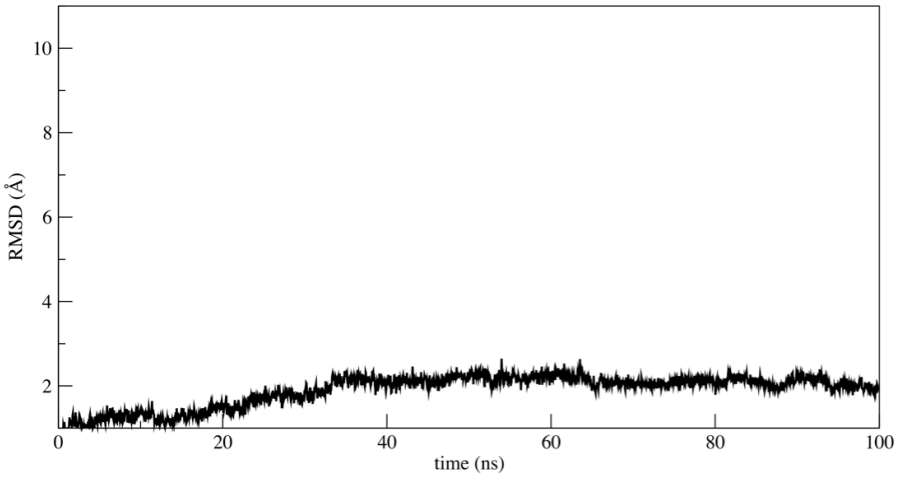


Figure 6.4: Plot of the RMSD (Å) of protein C_α atoms for the OXA-1:1β-Me doripenem acyl-enzyme complex over 100 ns of MD simulations (1.89 $\pm$ 0.38).

After determining the reliability of the simulations, I turned my attention to analysing the orientation of the KCX side chain. The proposed mechanism hypothesised the terminal carboxy group of the side chain reacts with either a hydrolytic water or the hydroxy group of the penem derivative in the hydrolysis and lactonization reactions, respectively.

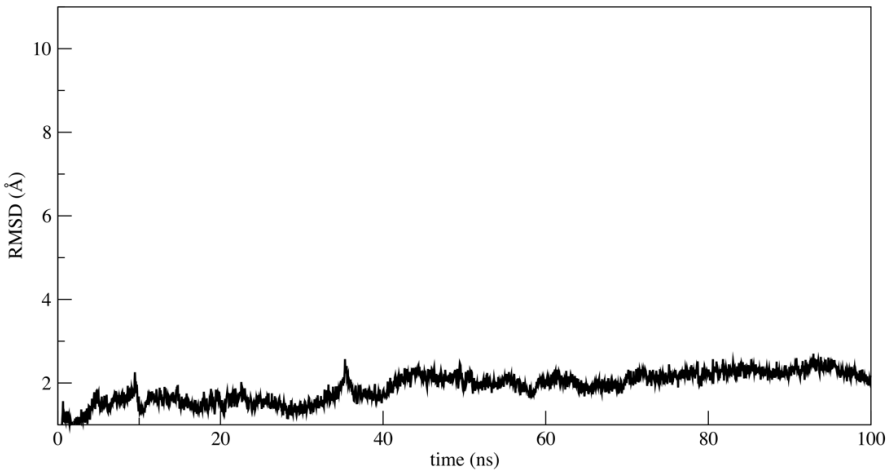


Figure 6.5: Plot of the RMSD (Å) of protein C_α atoms for the OXA-1:1β-H doripenem acyl-enzyme complex over 100 ns of MD simulations (1.90 ± 0.36).

Over 100 ns, the binding poses of KCX relative to the carbapenem in the OXA-1:1β-Me and OXA-1:1β-H were observed. When analysing the distance over time of the KCX carboxyl group and hydroxyethyl hydroxy group in each system, the corresponding poses necessary for either hydrolysis or lactonization can be observed in **Fig. 6.6**. For the OXA-1:1β-Me system, a stepwise pattern can be seen in the time vs distance plot, where the O···O−H distance oscillates regularly from ∼2 Å to ∼6 Å. Interestingly, this reflects the ability of 1β-Me to undergo both hydrolysis, whereby a larger O···O−H distance is necessary for movement of bulk waters into the active site, and lactonization. When compared to the poses shown by 1β-H, the distance vs time plots indicate initial movement toward the hydroxy group rotating away from the KCX side chain in the first 5 ns. From ∼40 ns onwards, it stabilises in this pose, corresponding to an orientation optimal for hydrolysis.

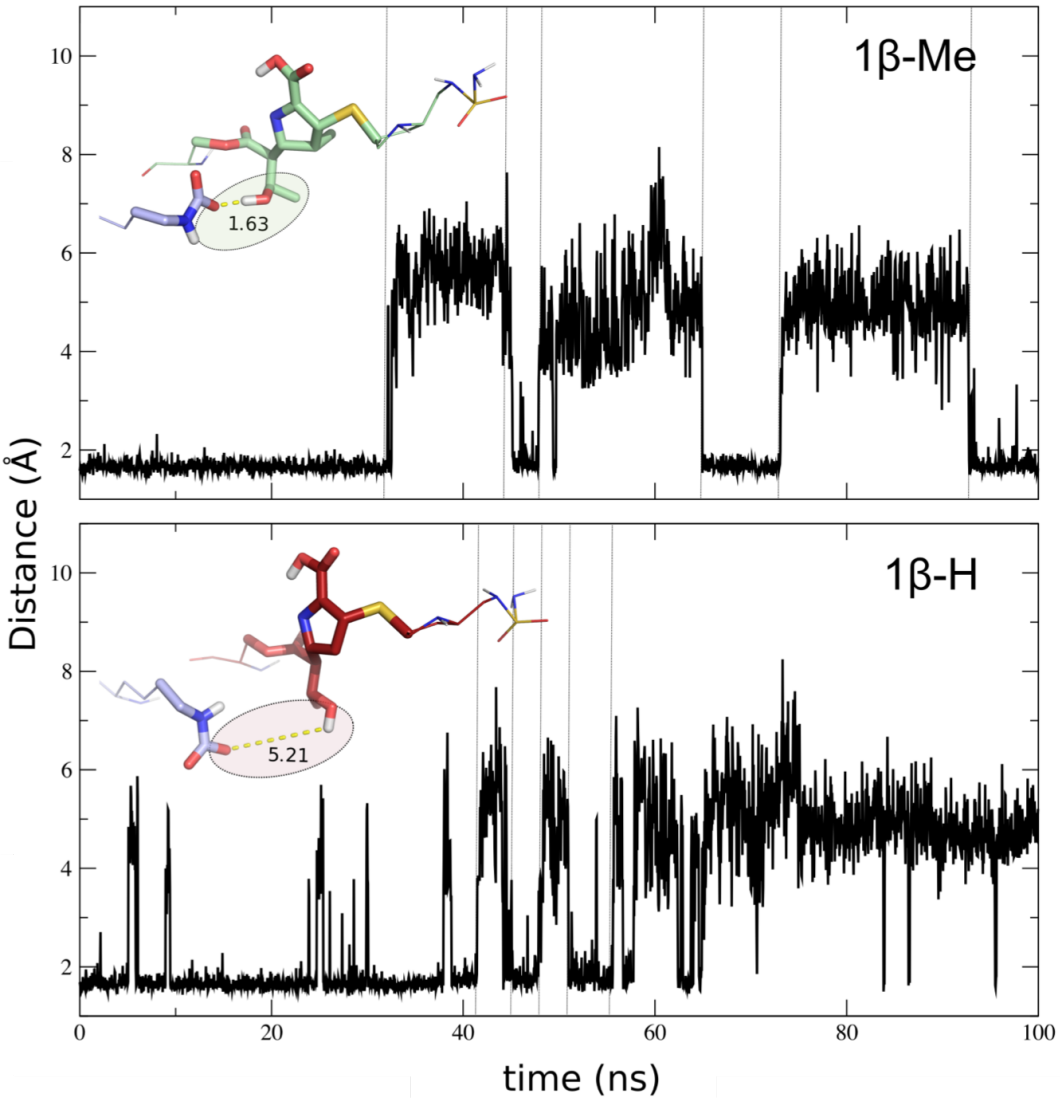


Figure 6.6: Representations of two major conformations of the hydroxyethyl group observed during MD simulation of the OXA-1 doripenem complex; the upper conformation appears to be more favourable for β-lactone formation. Note that the doripenem thioether side chain and serine backbone are represented as methyl groups.

In order to show that the O···O−H distances were a result of bond rotation described in **Fig. 6.2**, and not of a larger global movement of the ligand or receptor KCX residue, the dihedral C–C–C–O angle was also plotted over time. For OXA-1:1β-Me the stepwise pattern paralleling the distance vs time plot can be observed in **Fig. 6.7** as well as the stabilisation of the hydroxy group in OXA-1:1β-H in **Fig. 6.8**.

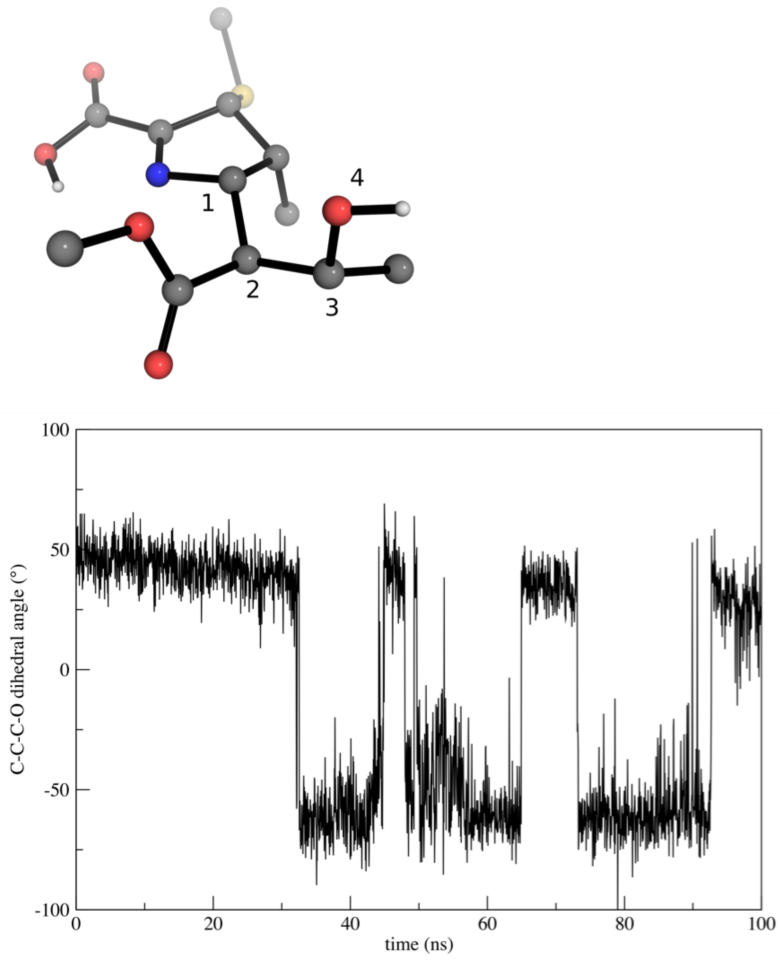


Figure 6.7: Dihedral angle of the doripenem (1β-Me) acyl-enzyme complex hydroxyethyl side chain C–C–C–O (atoms numbered 1-4) in the above structure over 100 ns. The atoms for which the dihedral angle is based are indicated on the structure shown above the plot. The doripenem sulfide side chain and the nucleophilic serine backbone are both represented as methyl groups.

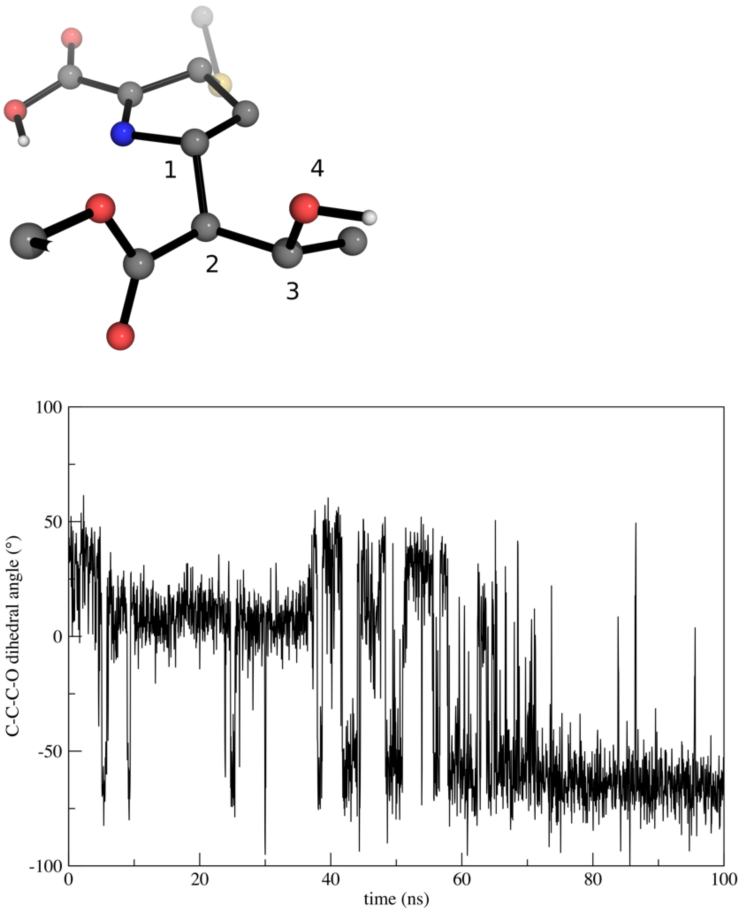


Figure 6.8: Dihedral angle of the doripenem (1β-H) acyl-enzyme complex hydroxyethyl side chain C–C–C–O (atoms numbered 1-4) in the above structure over 100 ns. The atoms for which the dihedral angle is based are indicated on the structure shown above the plot. The doripenem sulfide side chain and the nucleophilic serine backbone are both represented as methyl groups.

Visualisation of NCI Plots

MD simulations were valuable for visualising the binding poses of 1β-H and 1β-Me, however they did not explain why substitution of a proton in the 1β position of the pyrroline ring prevented lactonization. Non-covalent interaction (NCI) plots were generated for both carbapenem derivatives from geometries taken from a snapshot at time = 40 ns in the OXA-1:1β-Me system, where the wave function and electron density was calculated at the ωB97XD/def2-TZVP level of theory. The NCIPLOT program developed by Contreras-García et al.[246] analyses the

reduced density gradient (**S**) of the electron density (**r**) following single point energy calculations using a cut-off ρ. Low densities correspond to weak interactions, which are represented using RGB 3D NCI isosurfaces.

NCI plots indicate that the 1β-methyl group may interact sterically with the hydroxyethyl group in the acyl-enzyme complex, shown in the black box in **Fig. 6.9**, which is largely alleviated in the corresponding 1β-proton system, highlighted in the red box in **Fig. 6.9**. The increased sterics for 1β-methyl contribute towards the preferred binding pose, and subsequently explains the stepwise pattern observed in the distance and angle vs time plots.

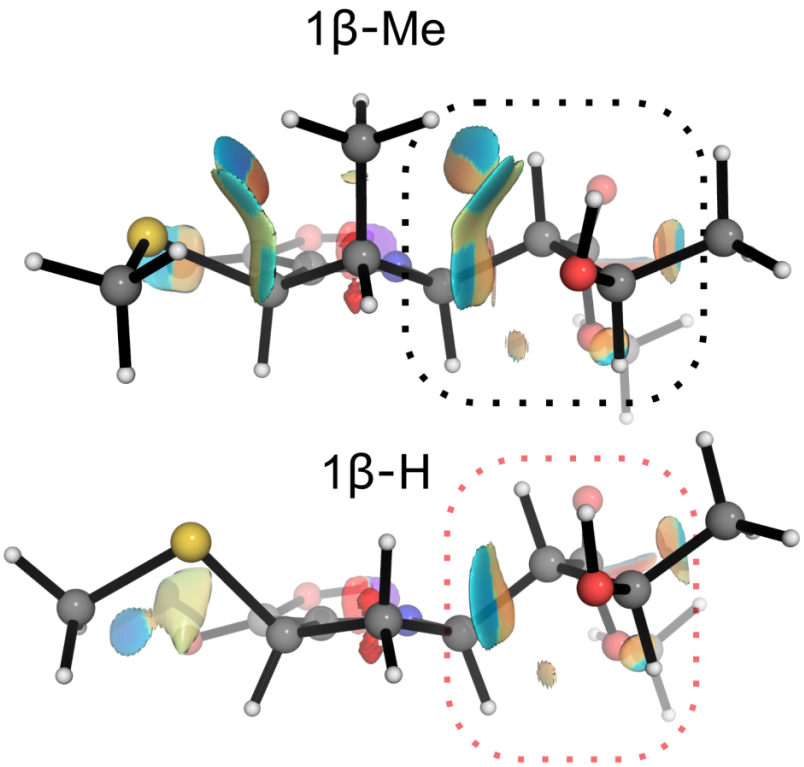


Figure 6.9: DFT non-covalent interaction isosurfaces showing unfavourable steric interactions between the carbapenem 1β-Me and hydroxyethyl groups, which are alleviated in the corresponding 1β-H system. Reduced density gradient isosurface s = 0.5, ρ(r) signλ_2 (e/au) colour scale runs from -0.02 (blue) to 0.02 (red).

Following long-time MD simulations, insights into this novel mechanism of carbapenem degradation unique to D class β-lactamases could be revealed. Furthermore, differences in reactivity following substitution of a proton to methyl group

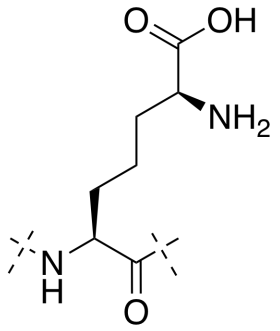
were rationalised by NCI plots that visualised the steric hindrance driving bond rotation of 1β-Me. The following section will discuss DFT approaches applied to understanding another non-hydrolytic carbapenem degradation mechanism also involved with a lysine analogue.

6.2 Non-Hydrolytic β-Lactam Antibiotic Fragmentation by l,d-Transpeptidases and Serine β-Lactamase Cysteine Variants

Like **Section 6.1**, another non-hydrolytic mechanism was observed experimentally. However, in this investigation the receptor carbapenemase was mutated rather than the inhibitor small-molecule. Here, I will describe a new computational model I developed, used to elucidate differences in reactivity that could also be applied to facilitating the development of novel scaffolds for antibiotic resistance.

6.2.1 Background of PBP/SBL and Ldt Enzymes

Bacterial peptidoglycan biosynthesis is one of the most clinically important antibiotic targets.[247–249] Peptidoglycan plays a critical structural role in the bacterial cell wall and is characterized by an extracellular network of polysaccharides and cross-linked peptides. In Gram-negative bacteria, crosslinks form between *meso*-diaminopimelate (*meso*-Dap), an epsilon-carboxy derivative of lysine (**Fig. 6.10**, and D-alanine residues. This formation is catalysed by penicillin-binding proteins (PBPs).



Diaminopimelic acid
(DAP)

Figure 6.10: Chemical structure of diaminopimelic acid (DAP), an analogue of lysine.

Lohans C.T., Chan H. H.T., Malla T., **Kumar K.**, et al. The nucleophilic cysteine of the l,d-transpeptidases enables non-hydrolytic mechanisms of β-lactam antibiotic degradation, *Angew. Chem. Int. Ed.* **2019**, *58(7)*, 1990-1994.

In other bacteria, the peptidoglycan is characterized by extensive cross-links between *meso*-DAP residues catalysed by the L,D-transpeptidases (Ldts).[250,251] While both transpeptidases require formation of an acyl-enzyme intermediate, the PBPs and serine β-lactamases (SBLs) utilise a nucleophilic serine residue for this reaction, whereas the Ldts have been shown to employ a nucleophilic cysteine. β-lactams have been shown to react curiously with Ldt enzymes. An example is penicillin, for which Ldts have been shown to be insensitive.[251–253] Interestingly, the carbapenem and penem classes of β-lactams are potent inactivators of Ldts,[254,255] and manifest antibiotic activity against *M. tuberculosis*.[256,257] The generic chemical structures of β-lactam classes: penicillin, cephalosporin, penem, and carbapenem are shown in **Fig. 6.11**.

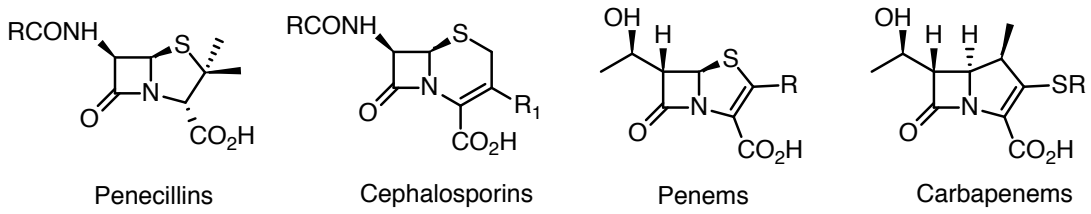


Figure 6.11: The generic chemical structures of β-lactam classes: penicillin, cephalosporin, penenem, and carbapenem.

Although acyl-enzyme complexes, analogous to the ones observed in **Section 6.1**, are formed with Ldts and PBPs/SBLs following binding of β-lactams, as shown below in **Fig. 6.12**, differences in reactivity have yet to be elucidated; specifically, occurrences of hydrolysis and/or fragmentation.

6.2.2 Serine to Cysteine Mutation

It was hypothesised that the serine/cysteine variation was responsible for differences observed in activity among Ldt and PBPs/SBLs, specifically the favourability of a respective ester vs thioester intermediate. Because the two classes of proteins are structurally very different, with a sequence similarity of <25% when comparing PBP5 from *Pseudomonas aeruginosa* and Ldt_{Mt2} from *M. tuberculosis*, shown in **Fig. 6.13**, it was unclear whether significant structural differences in the active sites or the point cysteine/serine residue difference was responsible.

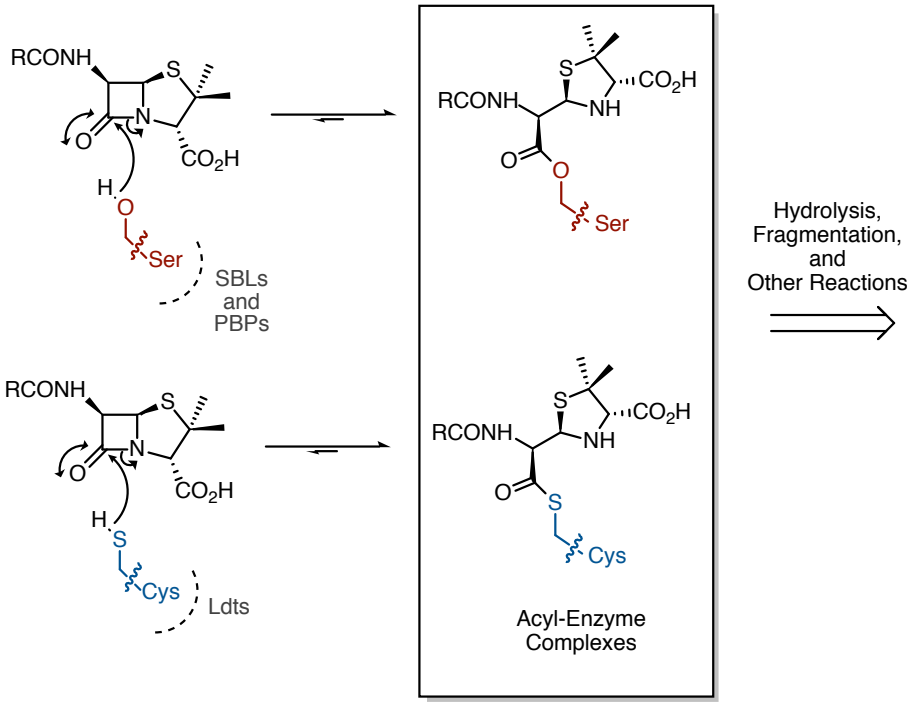


Figure 6.12: Reaction of nucleophilic serine for PBPs and serine β-lactamases and cysteine for Ldt enzymes with generic penicillin that both form an acyl-enzyme complex and can undergo fragmentation, hydrolysis, and other reactions.

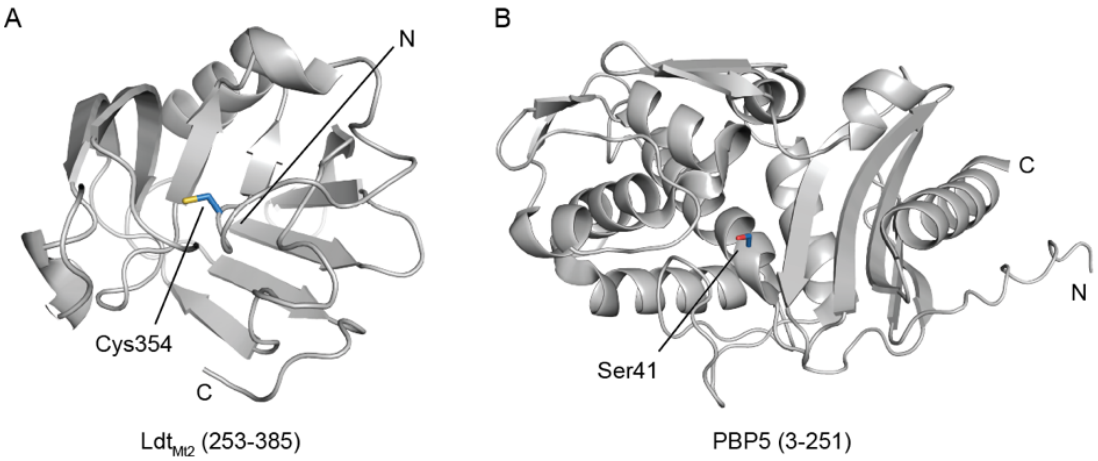


Figure 6.13: Comparison of the active sites of Ldt_{Mt2} and PBP5. Structures of (A) Ldt_{Mt2} from Mycobacterium tuberculosis (PDB: 3VYO) and (B) PBP5 from Pseudomonas aeruginosa (PDB: 4K91). The nucleophilic residues Cys354 (Ldt_{Mt2}) and Ser41 (PBP5) are shown in blue. N and C indicate the N-terminal and C-terminal.

In order to understand the degradation mechanisms, mutational serine to cysteine studies of SBL proteins, OXA-48 and KPC-2, were performed and are referred throughout as **OXA-48 S70C** and **KPC-2 S69C**, respectively. This section will discuss the first analysis of the products formed by this mutation and evaluation by an accurate pK_a descriptor based model.

6.2.3 Developing Model to Accurately Predict pK_a Values

DFT calculations were performed using Gaussian09.[178] A linear regression model was calculated comparing experimental pK_a values[258] of 11 compounds against pK_a values obtained at the M06-2X/6-31++g(d,p) level of theory with implicit polarisable continuum model (PCM) solvation in DMSO. The pK_a values were calculated from ΔG for acid dissociation in solution where

$$\Delta G = G(H^+) + G(A^-) - G(AH) \tag{6.1}$$

and

$$pK_a = \frac{\Delta G}{2.303RT} \tag{6.2}$$

A value of -273.3 kcal mol^{-1} was used for the H^+ solvation term[259] and the temperature was set at 298.15 K. A strong correlation coefficient was obtained (R^2 = 0.95). Experimental and theoretical pK_a values of the control dataset are shown in **Table 6.1**. This model was then used to calculate pK_a values of the predicted thioester and ester enolate fragmentation products of ampicillin and faropenem.

Table 6.1: Calculated ΔG values with experimentally and computationally derived pK_a values for model compounds.

Compound	ΔG (kcal mol^{-1})[a]	Expt. pK_a[b]	Comp. pK_a
$PhSCH_3$	330.0	39.0	41.6
CH_3CN	314.9	31.3	30.6
$PhCOCH_3$	307.8	24.7	25.3
CH_3CONEt_2	321.9	35.0	35.7
CH_3CO_2Et	314.9	29.5	30.5
CH_3COCH_3	310.6	26.5	27.3
$PhSO_2CH_3$	308.9	29.0	26.2
$PhCH_3$	329.9	16.4	13.4
$CH_2(CO_2Et)_2$	291.6	30.8	31.1
$PhCH_2SPh$	304.1	19.9	22.6
$PhCH_2COCH_3$	304.1	43.0	41.5

a. M06-2X/6-31++G(d,p) in DMSO
b. Ref [258]

6.2.4 Stability of Thioester and Ester Enolates

To investigate the importance of the cysteine/serine role in catalysis, SBL proteins OXA-48 and KPC-2 were experimentally mutated to contain a cysteine residue, namely **OXA-48 S70C** and **KPC-2 S69C**, by the Schofield Group. The SBL wild type, SBL mutants, and Ldt enzymes were then treated with a panel of β-lactam antibiotics, and the formed products were compared. The results are shown in **Table 6.2** below:

Table 6.2: Results of experimental treatment of three β-lactams and SBL wild type, SBL mutant, and Ldt enzymes. (Frag. = fragmentation, Hydrol. = hydrolysis).

	Antibiotics		
Enzyme	Ampicillin	Faropenem	Meropenem
OXA-48 or KPC-2	Hydrol.; No Frag.	Hydrol.; No Frag.	Hydrol.; No Frag.
OXA-48 S70C/ KPC-2 S69C	Hydrol.; Frag.	Frag.	-
Ldt_{Mt2}	Hydrol.; Frag.	Rapid Frag.	No Frag.

The importance of the serine/cysteine residue is highlighted from these mutational studies. Where the wild type SBL enzyme showed no fragmentation, **OXA-48 S70C** and **KPC-2 S69C** were shown to degrade ampicilin and faropenem with either a hydrolysis and fragmentation, or solely a fragmentation reaction. This was similar to Ldt_{Mt2}, despite their large overall tertiary structural differences. From this we directed our attention to developing a computational method to quantify the favourability of hydrolysis vs fragmentation from the stability of the formed intermediates. Analysis of the proposed fragmentation mechanism indicated the formation of a thioester and ester enolate for the catalytic cysteine and serine containing enzymes, respectively (**Fig. 6.14**).

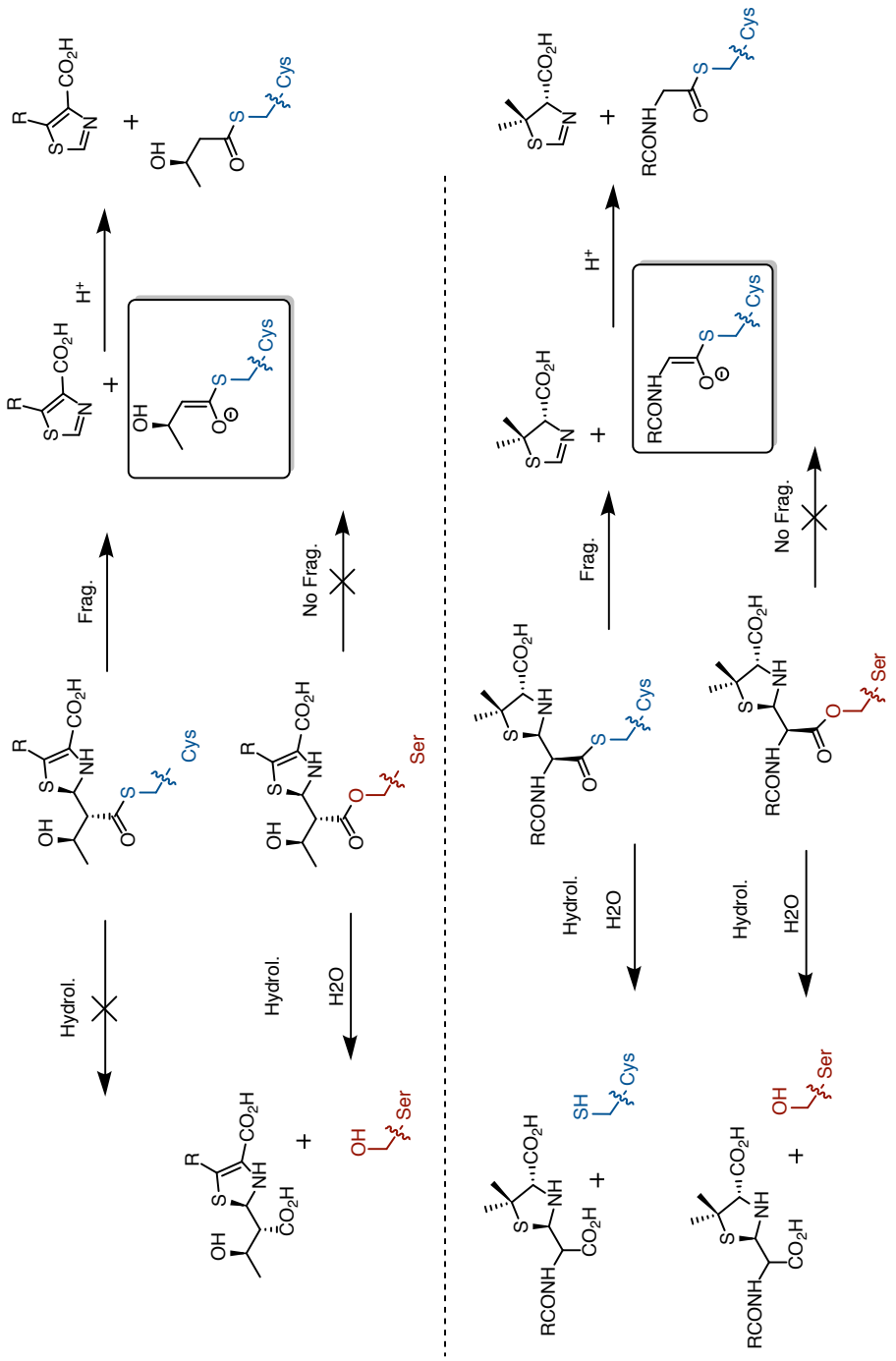


Figure 6.14: Fragmentation and hydrolysis mechanisms for ampicillin and faropenem by Ldt_{Mt2}. pK_a of thioester and ester enolate intermediates (box) calculated using DFT.

Favourability of the thioester/cysteine enolate over the ester/serine enolate was hypothesised to drive the reaction between fragmentation or hydrolysis. As stability of the enolate relates in part to the acidity of the α-position of the corresponding ester/thioester, a model to calculate accurate pK_a values would be valuable for predicting reactivity and selectivity of future lactamase inhibitors.

A dataset of 11 molecules with structural similarity to the ampicillin and faropenem fragments with experimental pK_a values was sought from the literature. These were then optimised at the M06-2X/6-31++G(d,p) level of theory in DMSO. I was able to obtain a high correlation ($R^2 = 0.95$) and low error (RMSE = 1.8) with single variate model (y = 1.028x - 0.733), shown in **Fig. 6.15**.

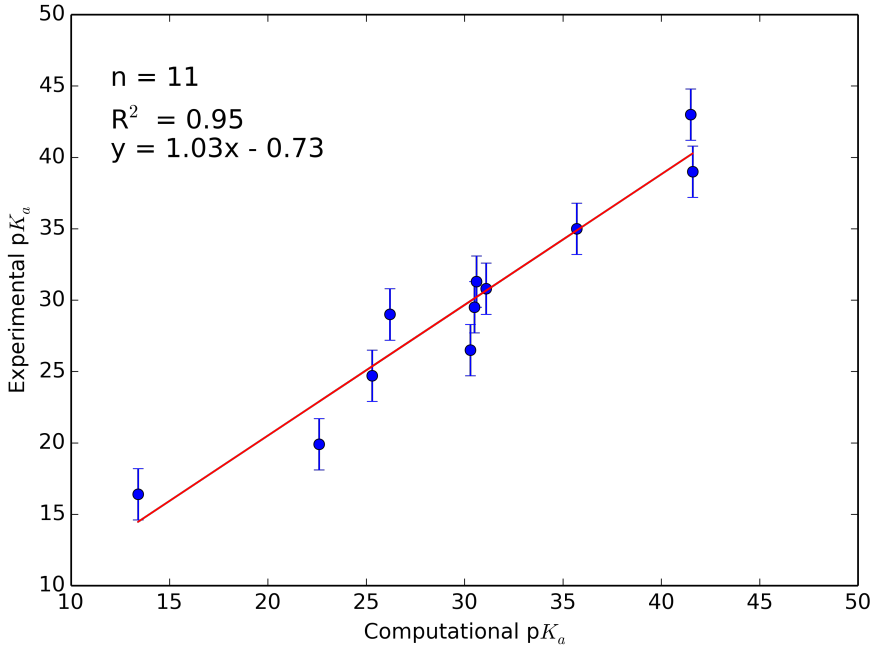
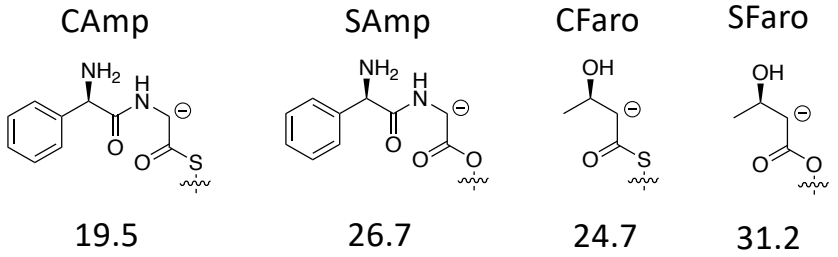


Figure 6.15: Plot of computationally vs experimental pK_a values.[258]

These calculated values are consistent with experimentally determined pK_a values for ethyl acetate (pK_a = 25.6)[260] and ethyl thioacetate (pK_a = 20.4 - 21.5).[261]

Table 6.3: Calculated ΔG and computationally derived pK_a values for models of the fragmentation-derived complexes. Values were obtained using the model shown in **Fig. 6.15** and DFT calculations for the ampicillin thioester (CAmp) and ester (SAmp) enolate models and faropenem thioester (CFaro) and ester (SFaro) enolate models, where the cysteine and serine side chains were replaced with methyl groups.



Compound	ΔG (kcal mol^{-1})[a]	Comp. pK_a	Extrapol. pK_a
CAmp	297.5	18.2	19.5
SAmp	307.4	25.3	26.7
CFaro	305.0	23.3	24.7
SFaro	313.8	29.7	31.2

a. M06-2X/6-31++g(d,p) in DMSO

The Ldts have emerged as antibiotic targets for a number of pathogenic bacteria, notably M. tuberculosis.[251,256] Results from this collaborative study, suggest that the different reactivities of the Ldts with β-lactam antibiotics, as compared to the PBPs and SBLs, is substantially governed by the conserved nucleophilic cysteine residue, which facilitates fragmentation reactions over proteins employing nucleophilic serine residues.

From development of an accurate model to calculate absolute pK_a values, these fragmentation reactions could be rationalised mechanistically by the relatively favourable involvement of a thioester enolate intermediate, as compared to an ester enolate. These results highlight a potential new approach for the development of antibiotics targeting the Ldts, in which β-lactams are rationally designed to favour C5-C6/C6-C7 fragmentation to give a stable acyl-enzyme complex. Applications of this model can also be used to further investigate links between the biological roles of nucleophilic hydrolases/transferases and the nature of the nucleophilic residue employed.

Concluding Remarks

This Thesis provides a discussion of computational chemistry approaches applied towards revealing insights of epigenetic processes. The goal of this work is to facilitate the design of future potent drugs capable of regulating activity of reader, writer, and eraser enzymes.

Electrostatic non-covalent interactions are prevalent in the binding of epigenetic readers and post-translationally modified histone residues. The study of a selective CREBBP bromodomain cation–π interaction with a series of 15 5-isoxazolylbenzimidazole derivatives using several *in silico* approaches is first discussed. I performed long-time molecular dynamic simulations that identified the importance of this interaction, otherwise not shown in the static X-ray crystal structures. Furthermore, simulations of all 15 complexes identified the ability of electrostatic energy calculations to predict activity within a set of inhibitors. This motivated a study to characterize similar cation–π interactions, where the cation is donated from a charged residue side chain and the π-system contribution derives from a small-molecule. All such interactions in available PDB data were documented and compared to energetic profiles of analogous small models. The druggability of targets containing arginine and histidine for aromatic small-molecules was found to be easier, while those involving lysine were harder to predict.

Epigenetic eraser proteins in the non-heme Fe(II)-αKG superfamily catalyse reactions with a diverse array of substrates. The enthalpic landscape of the proposed rate-determining hydrogen abstraction step was explored for 26 substrates including N-methyllysine and arginine, hydroxy-proline and asparagine, as well as methylated DNA nucleobases. The bond dissociation enthalpies of H-abstraction were calculated using high-level composite *ab initio* G4 and density functional theory calculations, where this step was found to be more unfavourable for charged substrates.

Experimental collaborations with the Kawamura Group, Schofield Group, and Mecinović Group provided empirical data that was extremely important for developing modelling tools of methyllysine and analogues. Specifically, this allowed the selectivity of five epigenetic readers for mutated trimethyllysine (KMe3) to be

investigated. Derivatives containing longer and shorter side chains and inverted backbone stereochemistry were explored. Because each reader was characterized with a distinct aromatic cage configuration, analysis of the effects on binding strength between the modified trimethylated lysine side chain and aromatic residues yielded diverse results. Specifically, readers KDM5A and TAF3, which contain Trp-Trp aromatic cages, were the most flexible in accommodating the mutated substrates.

Following this, mutations of the aromatic cage and its effects on binding were analysed. KDM5A residues Trp18 and Trp28 were fluorinated as 5F-Trp, 6F-Trp, and 5,6diF-Trp. Results from MD simulations and electrostatic potential surface (ESP) calculations agreed with experimental results that showed binding was not affected when compared to the wild type KMe3, despite the presence of electron withdrawing fluorine groups.

The methodological development of a robust technique to simulate lysine analogues was applied to a new biochemical system. In collaboration with the Schofield Group, the role of carbamyl-lysine and carbapenem analogues, molecules considered as antibiotics of the last resort, were explored using MD simulations and non-covalent interaction plots. From this, we confirmed that formation of a binding pose necessary for a novel degradation mechanism was due to steric hindrance caused by a 1β-C methylation of the inhibitor pyrroline ring.

Further development of computational techniques will prove valuable for future design of novel scaffolds for epigenetic targets. By reporting the parameters developed for the purpose of the investigations discussed here, these can continued to be utilised and the protocols accelerated. For example, while I focused on interactions involving H3 residue KMe3, the parameters developed for KMe3 and analogues could be directly applied for methyllysine located on the other seven histone monomer tails (e.g. Histone 4) that are highly relevant in epigenetic processes. Furthermore, given our published protocol of searching the PDB for specific interactions with small-molecules, attempts to identify another unique cation–π interaction capable of driving selectivity for an epigenetic enzyme may be useful for facilitating further potent inhibitor design. Finally, in addition to advancing the rational design of cancer therapeutics, other applications of the methodologies presented here have the potential to facilitate development of novel carbapenem scaffolds by assessing the stability of acyl-enzyme intermediates of other nucleophilic hydrolases associated with antibiotic resistance.

Appendices

Torture the data, and it will confess to anything

— Ronald Coase

A

CREBBP Bromodomain Inhibitor Development - Appendix

A.1 Ligand partial charges

Compounds L1-L15 partial charges used to simulate CREBBP-ligand complexes. Calculated using the RESP ESP charge Derive (RED) Server which employs the RESP-A1B methodology in which molecular electrostatic potentials (MEP) involving the Connolly surface algorithm is computed using the B3LYP functional, the integral equation formalism polarizable continuum solvent model (IEFPCM; ϵ=4, ether as solvent) and the cc-pVTZ basis set.

Ligand 1 Atom	X	Y	Z	RESP Charge
C	-2.06	2.75	2.47	-0.18
H	-1.42	2.6	3.31	0.13
C	-3.44	2.64	2.61	-0.08
H	-3.87	2.41	3.56	0.11
C	-4.25	2.85	1.51	-0.18
H	-5.32	2.78	1.61	0.13
C	-3.7	3.16	0.28	-0.1
H	-4.34	3.33	-0.57	0.11
C	-2.32	3.26	0.12	0.00
C	-1.51	3.06	1.23	-0.1
H	-0.44	3.14	1.13	0.11
C	-1.72	3.58	-1.24	-0.01
H	-0.78	4.10	-1.12	0.03
H	-2.39	4.24	-1.78	0.03
C	-1.48	2.33	-2.1	-0.02
H	-1.23	2.64	-3.11	0.04
H	-2.4	1.76	-2.17	0.04
C	-0.35	1.47	-1.6	0.33
N	-0.43	0.11	-1.47	-0.05
C	-1.57	-0.76	-1.67	-0.06
H	-2.19	-0.37	-2.47	0.08
H	-1.19	-1.72	-2.01	0.08
C	-2.41	-0.91	-0.4	-0.03
H	-2.73	0.08	-0.1	0.08
H	-1.78	-1.3	0.4	0.08
N	-3.58	-1.73	-0.64	-0.35
C	-3.28	-3.14	-0.82	-0.03
H	-2.5	-3.47	-0.13	0.08
H	-2.92	-3.33	-1.82	0.08
C	-4.52	-3.99	-0.58	0.09
H	-5.29	-3.71	-1.28	0.06
H	-4.29	-5.03	-0.73	0.06
O	-4.99	-3.87	0.74	-0.39
C	-4.8	-2.61	1.34	0.09
H	-5.64	-2.43	1.99	0.06
H	-3.9	-2.63	1.95	0.06
C	-4.69	-1.51	0.29	-0.03
H	-4.57	-0.55	0.79	0.08
H	-5.62	-1.46	-0.26	0.08
C	0.82	-0.3	-1.01	0.01
C	1.34	-1.54	-0.69	-0.19
H	0.77	-2.45	-0.77	0.14
C	2.65	-1.58	-0.27	-0.18
H	3.1	-2.52	-0.03	0.14
C	3.44	-0.41	-0.15	-0.09
C	2.89	0.82	-0.47	-0.13
H	3.46	1.72	-0.37	0.11
C	1.57	0.87	-0.91	0.3
N	0.8	1.95	-1.28	-0.61
C	4.84	-0.51	0.32	-0.09
C	5.98	-0.21	-0.34	0.29
C	6.27	0.28	-1.72	-0.3
H	6.74	1.25	-1.68	0.1
H	6.94	-0.41	-2.22	0.1
H	5.35	0.36	-2.29	0.1
O	7.03	-0.43	0.44	-0.1
N	6.61	-0.89	1.66	-0.46
C	5.33	-0.94	1.6	0.35
C	4.53	-1.38	2.79	-0.18
H	5.19	-1.56	3.63	0.07
H	3.8	-0.62	3.06	0.07
H	3.98	-2.29	2.57	0.07

Ligand 2 Atom	X	Y	Z	RESP Charge
C	-2.3	3.47	2.2	-0.22
C	-3.6	3	2.29	-0.16
H	-1.77	3.74	3.09	0.14
H	-4.07	2.92	3.26	0.13
C	-4.29	2.66	1.14	-0.1
H	-5.31	2.32	1.19	0.11
C	-3.65	2.78	-0.08	-0.18
H	-4.2	2.53	-0.98	0.12
C	-2.34	3.23	-0.19	0.02
C	-1.65	3.59	0.97	0.1
C	-0.24	4.15	0.94	-0.14
H	0.43	3.51	0.4	0.05
H	-0.22	5.13	0.46	0.05
H	0.14	4.28	1.94	0.05
C	-1.73	3.32	-1.58	-0.01
H	-0.79	3.85	-1.56	0.03
H	-2.4	3.9	-2.21	0.03
C	-1.51	1.96	-2.27	-0.05
H	-1.28	2.13	-3.32	0.05
H	-2.43	1.39	-2.24	0.05
C	-0.37	1.17	-1.69	0.32
N	-0.43	-0.18	-1.46	-0.05
C	-1.56	-1.08	-1.61	-0.04
H	-2.19	-0.74	-2.42	0.07
H	-1.17	-2.05	-1.91	0.07
C	-2.38	-1.19	-0.32	-0.03
H	-2.71	-0.19	-0.06	0.08
H	-1.73	-1.54	0.49	0.08
N	-3.54	-2.03	-0.5	-0.36
C	-3.22	-3.45	-0.64	-0.02
H	-2.42	-3.74	0.05	0.07
H	-2.88	-3.67	-1.64	0.07
C	-4.44	-4.3	-0.34	0.1
H	-5.23	-4.06	-1.04	0.05
H	-4.2	-5.35	-0.47	0.05
O	-4.88	-4.14	0.98	-0.4
C	-4.7	-2.86	1.53	0.1
H	-5.53	-2.67	2.2	0.05
H	-3.78	-2.85	2.12	0.05
C	-4.63	-1.8	0.45	-0.02
H	-4.51	-0.82	0.91	0.07
H	-5.58	-1.78	-0.09	0.07
C	0.82	-0.54	-0.99	0.01
C	1.36	-1.76	-0.59	-0.2
H	0.8	-2.67	-0.6	0.14
C	2.68	-1.75	-0.17	-0.16
H	3.14	-2.67	0.13	0.13
C	3.45	-0.57	-0.14	-0.07
C	2.88	0.63	-0.54	-0.13
H	3.44	1.55	-0.51	0.11
C	1.56	0.63	-0.97	0.29
N	0.78	1.68	-1.42	-0.59
C	4.86	-0.62	0.32	-0.13
C	5.99	-0.36	-0.36	0.29
C	6.26	0.03	-1.78	-0.28
H	6.73	1.01	-1.81	0.1
H	6.94	-0.69	-2.23	0.1
H	5.34	0.06	-2.34	0.1
O	7.04	-0.51	0.42	-0.1
N	6.63	-0.88	1.68	-0.47
C	5.35	-0.95	1.63	0.39
C	4.56	-1.31	2.85	-0.18
H	5.23	-1.42	3.7	0.06
H	3.83	-0.54	3.07	0.06
H	4.03	-2.24	2.7	0.06

Ligand 3 Atom	X	Y	Z	RESP Charge
C	-2.06	2.90	1.98	0.03
C	-1.17	2.87	3.19	-0.20
H	-0.27	2.28	3.01	0.07
H	-0.85	3.87	3.47	0.07
H	-1.68	2.44	4.05	0.07
C	-3.44	2.78	2.10	-0.11
H	-3.89	2.65	3.07	0.11
C	-4.25	2.84	0.98	-0.18
H	-5.32	2.77	1.08	0.13
C	-3.69	3.01	-0.28	-0.11
H	-4.33	3.07	-1.15	0.11
C	-2.31	3.13	-0.43	-0.05
C	-1.52	3.07	0.71	-0.09
H	-0.45	3.16	0.60	0.12
C	-1.70	3.28	-1.80	0.00
H	-0.77	3.83	-1.74	0.04
H	-2.37	3.86	-2.43	0.04
C	-1.44	1.94	-2.51	-0.07
H	-1.17	2.13	-3.55	0.06
H	-2.35	1.35	-2.52	0.06
C	-0.31	1.16	-1.89	0.32
N	-0.37	-0.17	-1.59	-0.06
C	-1.51	-1.07	-1.70	-0.02
H	-2.12	-0.79	-2.54	0.07
H	-1.12	-2.06	-1.91	0.07
C	-2.36	-1.07	-0.43	-0.01
H	-2.70	-0.05	-0.26	0.07
H	-1.74	-1.35	0.43	0.07
N	-3.52	-1.92	-0.57	-0.36
C	-3.20	-3.35	-0.56	-0.02
H	-2.43	-3.58	0.18	0.07
H	-2.83	-3.66	-1.53	0.07
C	-4.44	-4.17	-0.23	0.09
H	-5.21	-3.99	-0.97	0.05
H	-4.19	-5.22	-0.25	0.05
O	-4.93	-3.88	1.06	-0.39
C	-4.76	-2.56	1.49	0.09
H	-5.61	-2.30	2.10	0.05
H	-3.87	-2.49	2.11	0.05
C	-4.65	-1.60	0.31	-0.02
H	-4.54	-0.58	0.68	0.07
H	-5.57	-1.63	-0.26	0.07
C	0.86	-0.51	-1.07	0.01
C	1.40	-1.70	-0.59	-0.18
H	0.83	-2.61	-0.56	0.13
C	2.70	-1.67	-0.15	-0.17
H	3.15	-2.57	0.21	0.13
C	3.48	-0.49	-0.18	-0.10
C	2.92	0.69	-0.66	-0.11
H	3.49	1.60	-0.67	0.10
C	1.61	0.67	-1.11	0.28
N	0.84	1.69	-1.63	-0.61
C	4.88	-0.52	0.31	-0.12
C	6.02	-0.30	-0.37	0.30
C	6.32	0.02	-1.80	-0.30
H	6.79	0.99	-1.88	0.10
H	7.00	-0.72	-2.20	0.10
H	5.41	0.02	-2.38	0.10
O	7.06	-0.41	0.44	-0.10
N	6.63	-0.71	1.70	-0.47
C	5.35	-0.78	1.64	0.37
C	4.54	-1.08	2.87	-0.19
H	5.20	-1.14	3.73	0.07
H	3.81	-0.29	3.04	0.07
H	4.01	-2.01	2.76	0.07

Ligand 4 Atom	X	Y	Z	RESP Charge
C	2.09	-2.90	1.86	-0.20
H	1.47	-2.89	2.74	0.13
C	3.47	-2.76	1.99	0.15
C	4.11	-2.62	3.36	-0.18
H	5.10	-2.18	3.29	0.06
H	4.22	-3.60	3.83	0.06
H	3.51	-2.01	4.02	0.06
C	4.24	-2.78	0.84	-0.20
H	5.31	-2.68	0.90	0.13
C	3.64	-2.93	-0.41	-0.15
H	4.27	-2.95	-1.29	0.13
C	2.27	-3.07	-0.54	0.00
C	1.50	-3.06	0.62	-0.15
H	0.43	-3.16	0.55	0.13
C	1.63	-3.22	-1.91	0.01
H	0.71	-3.79	-1.83	0.03
H	2.30	-3.77	-2.56	0.03
C	1.31	-1.87	-2.58	-0.03
H	1.02	-2.05	-3.61	0.04
H	2.21	-1.26	-2.60	0.04
C	0.18	-1.13	-1.92	0.33
N	0.22	0.20	-1.61	-0.06
C	1.32	1.13	-1.74	-0.03
H	1.91	0.88	-2.61	0.07
H	0.90	2.11	-1.93	0.07
C	2.22	1.13	-0.50	-0.01
H	2.59	0.13	-0.36	0.07
H	1.62	1.39	0.37	0.07
N	3.35	2.02	-0.68	-0.36
C	3.00	3.44	-0.63	-0.03
H	2.24	3.63	0.13	0.07
H	2.58	3.75	-1.59	0.07
C	4.22	4.29	-0.33	0.10
H	4.96	4.14	-1.11	0.05
H	3.95	5.33	-0.33	0.05
O	4.76	4.00	0.93	-0.39
C	4.64	2.66	1.34	0.10
H	5.53	2.42	1.92	0.05
H	3.78	2.55	2.00	0.05
C	4.51	1.71	0.16	-0.03
H	4.45	0.69	0.51	0.07
H	5.42	1.78	-0.44	0.07
C	-1.01	0.49	-1.05	0.00
C	-1.57	1.66	-0.54	-0.19
H	-1.03	2.59	-0.52	0.14
C	-2.86	1.59	-0.07	-0.18
H	-3.32	2.48	0.32	0.14
C	-3.60	0.39	-0.08	-0.09
C	-3.02	-0.76	-0.58	-0.14
H	-3.56	-1.69	-0.59	0.11
C	-1.72	-0.70	-1.08	0.30
N	-0.94	-1.70	-1.62	-0.61
C	-4.98	0.38	0.45	-0.09
C	-6.14	0.13	-0.19	0.29
C	-6.47	-0.18	-1.61	-0.29
H	-6.92	-1.17	-1.69	0.10
H	-7.19	0.54	-1.99	0.10
H	-5.58	-0.15	-2.22	0.10
O	-7.15	0.20	0.66	-0.10
N	-6.69	0.50	1.91	-0.46
C	-5.42	0.61	1.81	0.35
C	-4.58	0.92	3.01	-0.18
H	-5.21	0.96	3.89	0.07
H	-3.82	0.16	3.14	0.07
H	-4.08	1.87	2.89	0.07

Ligand 5 Atom	X	Y	Z	RESP Charge
C	2.10	-2.87	1.36	-0.21
H	1.55	-2.97	2.27	0.12
C	3.49	-2.72	1.43	0.31
O	4.01	-2.70	2.67	-0.31
C	5.40	-2.61	2.84	0.01
H	5.58	-2.63	3.91	0.06
H	5.91	-3.45	2.37	0.06
H	5.79	-1.68	2.44	0.06
C	4.22	-2.61	0.26	-0.21
H	5.29	-2.50	0.28	0.12
C	3.57	-2.65	-0.97	-0.15
H	4.16	-2.57	-1.86	0.14
C	2.19	-2.79	-1.06	-0.04
C	1.47	-2.91	0.13	-0.15
H	0.41	-3.03	0.09	0.14
C	1.50	-2.81	-2.41	-0.02
H	0.62	-3.44	-2.37	0.04
H	2.17	-3.25	-3.15	0.04
C	1.09	-1.42	-2.91	-0.02
H	0.77	-1.50	-3.95	0.03
H	1.96	-0.77	-2.91	0.03
C	-0.04	-0.82	-2.13	0.34
N	-0.05	0.48	-1.67	-0.05
C	1.01	1.47	-1.74	-0.04
H	1.57	1.34	-2.66	0.07
H	0.53	2.44	-1.80	0.07
C	1.96	1.38	-0.55	-0.02
H	2.37	0.38	-0.53	0.07
H	1.39	1.50	0.37	0.07
N	3.04	2.33	-0.67	-0.36
C	2.63	3.71	-0.46	-0.04
H	1.90	3.79	0.36	0.08
H	2.16	4.11	-1.35	0.08
C	3.82	4.59	-0.11	0.10
H	4.54	4.56	-0.93	0.06
H	3.51	5.61	0.02	0.06
O	4.43	4.19	1.09	-0.40
C	4.39	2.81	1.36	0.10
H	5.30	2.55	1.89	0.06
H	3.55	2.60	2.03	0.06
C	4.25	1.99	0.09	-0.04
H	4.24	0.94	0.34	0.08
H	5.13	2.16	-0.53	0.08
C	-1.26	0.65	-1.04	0.00
C	-1.85	1.72	-0.38	-0.20
H	-1.35	2.67	-0.27	0.14
C	-3.11	1.54	0.14	-0.17
H	-3.59	2.36	0.64	0.14
C	-3.79	0.31	0.02	-0.08
C	-3.19	-0.75	-0.63	-0.14
H	-3.69	-1.70	-0.72	0.11
C	-1.92	-0.57	-1.17	0.32
N	-1.12	-1.46	-1.85	-0.63
C	-5.15	0.17	0.60	-0.10
C	-6.32	-0.06	-0.02	0.29
C	-6.71	-0.23	-1.45	-0.28
H	-7.11	-1.22	-1.62	0.10
H	-7.47	0.49	-1.71	0.10
H	-5.85	-0.08	-2.09	0.10
O	-7.30	-0.13	0.87	-0.10
N	-6.80	0.05	2.13	-0.47
C	-5.54	0.23	1.99	0.36
C	-4.66	0.44	3.18	-0.19
H	-5.24	0.35	4.09	0.07
H	-3.86	-0.29	3.20	0.07
H	-4.20	1.42	3.16	0.07

Ligand 6 Atom	X	Y	Z	RESP Charge
C	-0.91	4.30	0.79	-0.01
N	0.23	4.62	1.64	0.72
O	0.01	5.11	2.70	-0.44
O	1.32	4.37	1.22	-0.44
C	-2.18	4.65	1.23	-0.13
H	-2.31	5.13	2.17	0.14
C	-3.25	4.34	0.41	-0.15
H	-4.25	4.61	0.71	0.14
C	-3.04	3.72	-0.81	-0.08
H	-3.89	3.50	-1.44	0.12
C	-1.77	3.37	-1.24	-0.01
C	-0.69	3.67	-0.42	-0.06
H	0.31	3.41	-0.72	0.11
C	-1.57	2.68	-2.58	-0.02
H	-0.58	2.89	-2.96	0.04
H	-2.29	3.08	-3.29	0.04
C	-1.77	1.15	-2.52	-0.03
H	-1.88	0.77	-3.53	0.05
H	-2.70	0.93	-2.01	0.05
C	-0.63	0.42	-1.87	0.29
N	-0.75	-0.86	-1.42	-0.05
C	-1.93	-1.70	-1.39	-0.03
H	-2.60	-1.41	-2.19	0.07
H	-1.62	-2.72	-1.60	0.07
C	-2.68	-1.60	-0.06	-0.04
H	-2.93	-0.56	0.11	0.08
H	-2.01	-1.90	0.75	0.08
N	-3.90	-2.37	-0.08	-0.37
C	-3.68	-3.82	-0.10	-0.02
H	-2.86	-4.11	0.56	0.08
H	-3.43	-4.15	-1.10	0.08
C	-4.93	-4.55	0.35	0.10
H	-5.76	-4.32	-0.32	0.05
H	-4.77	-5.62	0.30	0.05
O	-5.28	-4.24	1.67	-0.39
C	-4.96	-2.94	2.09	0.10
H	-5.73	-2.63	2.79	0.05
H	-4.01	-2.95	2.62	0.05
C	-4.90	-1.98	0.92	-0.02
H	-4.68	-0.98	1.28	0.08
H	-5.88	-1.93	0.45	0.08
C	0.50	-1.22	-0.94	0.02
C	1.00	-2.38	-0.36	-0.20
H	0.39	-3.25	-0.20	0.14
C	2.33	-2.38	-0.01	-0.15
H	2.76	-3.26	0.43	0.13
C	3.16	-1.26	-0.20	-0.07
C	2.64	-0.11	-0.77	-0.11
H	3.25	0.77	-0.92	0.11
C	1.30	-0.10	-1.15	0.21
N	0.56	0.91	-1.73	-0.53
C	4.59	-1.32	0.20	-0.14
C	5.70	-1.27	-0.56	0.29
C	5.92	-1.17	-2.03	-0.30
H	6.45	-0.25	-2.27	0.10
H	6.52	-2.00	-2.37	0.10
H	4.97	-1.18	-2.56	0.10
O	6.78	-1.35	0.20	-0.10
N	6.41	-1.45	1.51	-0.47
C	5.13	-1.43	1.54	0.38
C	4.38	-1.49	2.83	-0.17
H	5.08	-1.51	3.66	0.06
H	3.73	-0.63	2.94	0.06
H	3.77	-2.38	2.88	0.06

Ligand 7

Atom	X	Y	Z	RESP Charge
C	1.89	-3.63	0.49	-0.17
H	1.26	-4.08	1.23	0.15
C	3.25	-3.49	0.72	0.04
C	4.09	-2.92	-0.22	-0.17
H	5.14	-2.84	-0.02	0.15
C	3.55	-2.48	-1.41	-0.08
H	4.20	-2.04	-2.15	0.12
C	2.18	-2.59	-1.67	0.01
C	1.37	-3.18	-0.71	-0.08
H	0.31	-3.26	-0.90	0.12
C	1.62	-2.08	-2.98	-0.02
H	0.70	-2.60	-3.22	0.04
H	2.33	-2.28	-3.77	0.04
C	1.34	-0.57	-2.96	0.01
H	1.15	-0.23	-3.98	0.04
H	2.22	-0.04	-2.61	0.04
C	0.14	-0.20	-2.13	0.30
N	0.00	1.03	-1.53	-0.08
C	0.96	2.11	-1.47	0.01
H	1.56	2.11	-2.37	0.06
H	0.40	3.04	-1.47	0.06
C	1.87	2.01	-0.25	-0.03
H	2.37	1.05	-0.28	0.07
H	1.26	2.02	0.66	0.07
N	2.88	3.05	-0.26	-0.37
C	2.35	4.38	-0.01	-0.02
H	1.57	4.37	0.76	0.07
H	1.90	4.78	-0.91	0.07
C	3.45	5.33	0.44	0.12
H	4.21	5.39	-0.33	0.05
H	3.05	6.32	0.60	0.05
O	4.02	4.92	1.66	-0.41
C	4.06	3.53	1.87	0.12
H	4.96	3.32	2.44	0.05
H	3.21	3.23	2.48	0.05
C	4.07	2.77	0.56	-0.02
H	4.12	1.70	0.77	0.07
H	4.96	3.03	0.00	0.07
C	-1.24	1.01	-0.92	0.02
C	-1.93	1.95	-0.16	-0.18
H	-1.52	2.92	0.07	0.14
C	-3.18	1.59	0.30	-0.18
H	-3.75	2.30	0.88	0.14
C	-3.74	0.33	0.02	-0.06
C	-3.04	-0.59	-0.73	-0.10
H	-3.44	-1.56	-0.94	0.10
C	-1.78	-0.24	-1.20	0.23
N	-0.88	-0.96	-1.97	-0.56
C	-5.09	0.00	0.54	-0.14
C	-6.23	-0.25	-0.14	0.30
C	-6.57	-0.28	-1.59	-0.30
H	-6.88	-1.27	-1.89	0.10
H	-7.39	0.40	-1.79	0.10
H	-5.71	0.02	-2.18	0.10
O	-7.21	-0.52	0.71	-0.10
N	-6.75	-0.45	1.99	-0.46
C	-5.51	-0.15	1.91	0.37
C	-4.68	0.00	3.15	-0.17
H	-5.27	-0.25	4.02	0.06
H	-3.82	-0.66	3.11	0.06
H	-4.31	1.01	3.25	0.06
N	3.80	-3.96	1.98	0.73
O	3.06	-4.47	2.76	-0.44
O	4.98	-3.81	2.16	-0.44

Ligand 8 Atom	X	Y	Z	RESP Charge
C	-2.23	3.17	2.27	-0.23
H	-1.62	3.31	3.15	0.16
C	-3.57	2.84	2.35	-0.14
H	-4.03	2.71	3.32	0.14
C	-4.31	2.68	1.19	-0.12
H	-5.36	2.44	1.25	0.12
C	-3.71	2.86	-0.04	-0.19
H	-4.29	2.75	-0.94	0.14
C	-2.36	3.18	-0.16	0.00
C	-1.65	3.34	1.03	0.24
F	-0.37	3.68	0.98	-0.18
C	-1.71	3.36	-1.51	0.00
H	-0.76	3.88	-1.41	0.03
H	-2.35	3.99	-2.12	0.03
C	-1.48	2.04	-2.27	-0.04
H	-1.22	2.27	-3.30	0.05
H	-2.39	1.47	-2.29	0.05
C	-0.34	1.23	-1.70	0.35
N	-0.41	-0.12	-1.47	-0.09
C	-1.55	-1.02	-1.63	-0.02
H	-2.18	-0.67	-2.44	0.07
H	-1.16	-1.98	-1.93	0.07
C	-2.37	-1.13	-0.35	-0.03
H	-2.71	-0.13	-0.08	0.08
H	-1.73	-1.48	0.46	0.08
N	-3.54	-1.97	-0.54	-0.36
C	-3.21	-3.39	-0.69	-0.03
H	-2.41	-3.69	0.00	0.08
H	-2.86	-3.60	-1.69	0.08
C	-4.43	-4.25	-0.40	0.10
H	-5.22	-4.00	-1.10	0.05
H	-4.18	-5.29	-0.54	0.05
O	-4.88	-4.10	0.92	-0.39
C	-4.70	-2.82	1.48	0.10
H	-5.53	-2.64	2.15	0.05
H	-3.79	-2.81	2.08	0.05
C	-4.63	-1.75	0.41	-0.03
H	-4.51	-0.78	0.87	0.08
H	-5.57	-1.73	-0.13	0.08
C	0.83	-0.49	-1.00	0.01
C	1.37	-1.70	-0.59	-0.18
H	0.80	-2.61	-0.60	0.13
C	2.68	-1.70	-0.17	-0.18
H	3.13	-2.63	0.13	0.14
C	3.46	-0.53	-0.14	-0.06
C	2.90	0.67	-0.55	-0.13
H	3.47	1.59	-0.52	0.11
C	1.58	0.69	-0.98	0.29
N	0.81	1.74	-1.43	-0.62
C	4.87	-0.59	0.32	-0.16
C	6.00	-0.34	-0.37	0.31
C	6.28	0.04	-1.78	-0.29
H	6.75	1.01	-1.82	0.10
H	6.95	-0.68	-2.23	0.10
H	5.36	0.07	-2.35	0.10
O	7.05	-0.50	0.43	-0.10
N	6.64	-0.86	1.68	-0.47
C	5.36	-0.91	1.63	0.38
C	4.57	-1.26	2.85	-0.18
H	5.23	-1.38	3.70	0.07
H	3.84	-0.48	3.07	0.07
H	4.02	-2.19	2.71	0.07

Ligand 9 Atom	X	Y	Z	RESP Charge
C	-2.01	3.06	1.91	0.15
F	-1.20	3.04	2.97	-0.18
C	-3.37	2.98	2.11	-0.11
H	-3.76	2.91	3.11	0.13
C	-4.19	2.99	1.00	-0.18
H	-5.26	2.94	1.13	0.14
C	-3.65	3.09	-0.27	-0.12
H	-4.31	3.11	-1.13	0.12
C	-2.27	3.16	-0.46	-0.05
C	-1.44	3.15	0.66	-0.09
H	-0.38	3.21	0.56	0.13
C	-1.69	3.24	-1.86	0.00
H	-0.75	3.77	-1.84	0.04
H	-2.37	3.80	-2.49	0.04
C	-1.47	1.86	-2.50	-0.06
H	-1.24	1.99	-3.55	0.06
H	-2.39	1.29	-2.45	0.06
C	-0.33	1.10	-1.88	0.31
N	-0.38	-0.23	-1.57	-0.05
C	-1.52	-1.14	-1.67	-0.04
H	-2.13	-0.87	-2.51	0.08
H	-1.12	-2.13	-1.87	0.08
C	-2.37	-1.12	-0.40	-0.04
H	-2.70	-0.10	-0.23	0.08
H	-1.74	-1.40	0.46	0.08
N	-3.53	-1.97	-0.53	-0.36
C	-3.22	-3.40	-0.52	-0.02
H	-2.44	-3.63	0.21	0.08
H	-2.85	-3.71	-1.50	0.08
C	-4.45	-4.21	-0.19	0.09
H	-5.23	-4.03	-0.93	0.06
H	-4.22	-5.27	-0.21	0.06
O	-4.94	-3.93	1.10	-0.39
C	-4.75	-2.61	1.54	0.09
H	-5.60	-2.35	2.16	0.06
H	-3.86	-2.55	2.15	0.06
C	-4.65	-1.64	0.36	-0.02
H	-4.53	-0.63	0.74	0.08
H	-5.58	-1.67	-0.19	0.08
C	0.86	-0.55	-1.06	0.02
C	1.40	-1.74	-0.57	-0.20
H	0.85	-2.65	-0.53	0.14
C	2.71	-1.69	-0.14	-0.17
H	3.17	-2.59	0.23	0.14
C	3.47	-0.51	-0.17	-0.07
C	2.91	0.66	-0.65	-0.12
H	3.46	1.58	-0.67	0.11
C	1.59	0.63	-1.10	0.27
N	0.81	1.64	-1.62	-0.59
C	4.87	-0.52	0.32	-0.13
C	6.02	-0.31	-0.37	0.30
C	6.31	-0.02	-1.80	-0.30
H	6.78	0.96	-1.90	0.10
H	7.00	-0.76	-2.19	0.10
H	5.40	-0.03	-2.38	0.10
O	7.06	-0.40	0.45	-0.10
N	6.63	-0.68	1.72	-0.47
C	5.35	-0.76	1.66	0.36
C	4.54	-1.03	2.89	-0.17
H	5.20	-1.09	3.75	0.06
H	3.81	-0.25	3.04	0.06
H	4.01	-1.97	2.79	0.06

Ligand 10

Atom	X	Y	Z	RESP Charge
C	1.95	-3.10	1.84	-0.20
H	1.33	-3.14	2.71	0.15
C	3.31	-2.98	1.98	0.27
F	3.84	-2.90	3.20	-0.20
C	4.16	-2.93	0.89	-0.20
H	5.22	-2.84	1.04	0.15
C	3.60	-3.00	-0.38	-0.13
H	4.25	-2.97	-1.24	0.13
C	2.23	-3.12	-0.57	-0.02
C	1.41	-3.18	0.56	-0.13
H	0.35	-3.27	0.44	0.13
C	1.64	-3.18	-1.96	0.00
H	0.71	-3.74	-1.96	0.03
H	2.33	-3.70	-2.62	0.03
C	1.38	-1.79	-2.57	-0.02
H	1.13	-1.90	-3.62	0.04
H	2.28	-1.20	-2.53	0.04
C	0.24	-1.07	-1.91	0.33
N	0.26	0.26	-1.59	-0.06
C	1.37	1.20	-1.70	-0.02
H	1.97	0.95	-2.56	0.07
H	0.95	2.18	-1.88	0.07
C	2.25	1.19	-0.45	-0.02
H	2.61	0.18	-0.30	0.07
H	1.64	1.44	0.42	0.07
N	3.40	2.06	-0.60	-0.37
C	3.05	3.48	-0.58	-0.02
H	2.28	3.69	0.17	0.07
H	2.66	3.79	-1.54	0.07
C	4.27	4.32	-0.26	0.11
H	5.03	4.16	-1.02	0.05
H	4.01	5.37	-0.27	0.05
O	4.79	4.04	1.01	-0.40
C	4.64	2.71	1.44	0.11
H	5.51	2.47	2.05	0.05
H	3.76	2.63	2.08	0.05
C	4.53	1.75	0.27	-0.02
H	4.45	0.74	0.63	0.07
H	5.45	1.81	-0.31	0.07
C	-0.97	0.55	-1.04	0.01
C	-1.53	1.71	-0.52	-0.20
H	-1.00	2.64	-0.49	0.15
C	-2.83	1.63	-0.06	-0.17
H	-3.30	2.51	0.33	0.14
C	-3.56	0.43	-0.09	-0.07
C	-2.98	-0.72	-0.60	-0.12
H	-3.52	-1.65	-0.62	0.11
C	-1.68	-0.65	-1.08	0.28
N	-0.89	-1.64	-1.63	-0.61
C	-4.95	0.40	0.43	-0.12
C	-6.10	0.16	-0.23	0.30
C	-6.42	-0.14	-1.65	-0.28
H	-6.86	-1.13	-1.74	0.10
H	-7.14	0.57	-2.03	0.10
H	-5.53	-0.10	-2.26	0.10
O	-7.13	0.22	0.61	-0.10
N	-6.68	0.51	1.87	-0.46
C	-5.41	0.62	1.78	0.35
C	-4.58	0.92	2.99	-0.16
H	-5.21	0.95	3.87	0.06
H	-3.82	0.16	3.13	0.06
H	-4.08	1.88	2.89	0.06

Ligand 11 Atom	X	Y	Z	RESP Charge
C	-2.08	-2.92	-0.51	-0.16
Cl	-1.14	-3.44	-1.87	-0.11
C	-3.46	-2.78	-0.63	0.38
O	-4.00	-3.06	-1.83	-0.29
C	-5.39	-3.00	-2.00	-0.01
H	-5.58	-3.29	-3.03	0.07
H	-5.90	-3.69	-1.34	0.07
H	-5.77	-2.00	-1.84	0.07
C	-4.18	-2.36	0.48	-0.18
H	-5.25	-2.26	0.42	0.12
C	-3.53	-2.10	1.68	-0.23
H	-4.12	-1.79	2.52	0.16
C	-2.16	-2.24	1.81	-0.03
C	-1.44	-2.65	0.69	-0.04
H	-0.38	-2.77	0.75	0.13
C	-1.47	-1.92	3.12	0.00
H	-0.58	-2.54	3.23	0.04
H	-2.13	-2.16	3.94	0.04
C	-1.05	-0.45	3.26	-0.05
H	-0.74	-0.26	4.28	0.05
H	-1.91	0.19	3.07	0.05
C	0.09	-0.07	2.35	0.31
N	0.15	1.10	1.64	-0.05
C	-0.88	2.12	1.50	-0.03
H	-1.43	2.21	2.43	0.07
H	-0.37	3.06	1.35	0.07
C	-1.84	1.81	0.36	-0.02
H	-2.28	0.83	0.55	0.08
H	-1.27	1.72	-0.58	0.08
N	-2.90	2.79	0.28	-0.36
C	-2.46	4.09	-0.20	-0.03
H	-1.73	3.99	-1.01	0.07
H	-1.98	4.64	0.60	0.07
C	-3.63	4.90	-0.71	0.09
H	-4.35	5.05	0.09	0.06
H	-3.30	5.88	-1.04	0.06
O	-4.26	4.30	-1.81	-0.39
C	-4.23	2.89	-1.82	0.09
H	-5.15	2.55	-2.29	0.06
H	-3.40	2.54	-2.43	0.06
C	-4.12	2.33	-0.41	-0.03
H	-4.13	1.25	-0.45	0.07
H	-4.99	2.63	0.16	0.07
C	1.36	1.08	0.98	0.02
C	1.97	1.97	0.11	-0.20
H	1.51	2.90	-0.19	0.15
C	3.22	1.63	-0.37	-0.17
H	3.73	2.31	-1.03	0.14
C	3.86	0.43	0.00	-0.08
C	3.22	-0.45	0.86	-0.13
H	3.68	-1.38	1.14	0.11
C	1.97	-0.11	1.36	0.27
N	1.14	-0.80	2.22	-0.59
C	5.20	0.12	-0.55	-0.12
C	6.38	-0.01	0.10	0.30
C	6.78	0.13	1.52	-0.30
H	7.16	-0.81	1.91	0.10
H	7.57	0.87	1.62	0.10
H	5.93	0.44	2.12	0.10
O	7.34	-0.31	-0.76	-0.10
N	6.82	-0.40	-2.03	-0.47
C	5.56	-0.15	-1.92	0.36
C	4.67	-0.18	-3.12	-0.17
H	5.24	-0.48	-3.99	0.07
H	3.85	-0.87	-2.97	0.07
H	4.24	0.80	-3.30	0.07

Ligand 12 Atom	X	Y	Z	RESP Charge
C	2.09	-2.92	0.90	0.09
F	1.38	-3.17	1.99	-0.18
C	3.47	-2.79	1.02	0.28
O	3.98	-2.94	2.26	-0.28
C	5.36	-2.89	2.44	-0.01
H	5.53	-3.06	3.50	0.07
H	5.86	-3.67	1.87	0.07
H	5.77	-1.92	2.17	0.07
C	4.19	-2.53	-0.13	-0.19
H	5.26	-2.44	-0.09	0.13
C	3.55	-2.41	-1.35	-0.22
H	4.14	-2.21	-2.23	0.16
C	2.18	-2.53	-1.47	-0.06
C	1.45	-2.80	-0.30	-0.14
H	0.38	-2.91	-0.33	0.16
C	1.48	-2.37	-2.80	0.00
H	0.60	-3.00	-2.85	0.04
H	2.15	-2.70	-3.59	0.04
C	1.07	-0.92	-3.11	-0.07
H	0.76	-0.86	-4.15	0.06
H	1.94	-0.27	-3.01	0.06
C	-0.07	-0.44	-2.26	0.32
N	-0.10	0.80	-1.67	-0.05
C	0.93	1.81	-1.63	-0.02
H	1.49	1.80	-2.56	0.07
H	0.44	2.77	-1.57	0.07
C	1.89	1.61	-0.46	-0.02
H	2.32	0.62	-0.55	0.07
H	1.32	1.62	0.48	0.07
N	2.96	2.58	-0.47	-0.36
C	2.53	3.93	-0.12	-0.03
H	1.80	3.92	0.69	0.08
H	2.05	4.41	-0.97	0.08
C	3.71	4.78	0.31	0.09
H	4.43	4.85	-0.51	0.06
H	3.38	5.78	0.54	0.06
O	4.33	4.27	1.46	-0.39
C	4.30	2.87	1.60	0.09
H	5.22	2.57	2.09	0.06
H	3.47	2.59	2.24	0.06
C	4.17	2.18	0.25	-0.03
H	4.18	1.11	0.39	0.08
H	5.05	2.43	-0.35	0.08
C	-1.32	0.87	-1.01	0.02
C	-1.92	1.85	-0.24	-0.20
H	-1.44	2.79	-0.03	0.14
C	-3.17	1.58	0.26	-0.17
H	-3.68	2.33	0.85	0.14
C	-3.83	0.36	0.02	-0.07
C	-3.21	-0.61	-0.75	-0.12
H	-3.69	-1.56	-0.94	0.11
C	-1.95	-0.35	-1.27	0.27
N	-1.13	-1.13	-2.06	-0.60
C	-5.18	0.12	0.58	-0.12
C	-6.36	-0.05	-0.05	0.30
C	-6.75	-0.05	-1.49	-0.30
H	-7.14	-1.02	-1.77	0.10
H	-7.53	0.69	-1.66	0.10
H	-5.90	0.18	-2.11	0.10
O	-7.32	-0.25	0.83	-0.10
N	-6.81	-0.21	2.10	-0.47
C	-5.55	0.01	1.97	0.36
C	-4.67	0.09	3.17	-0.17
H	-5.24	-0.11	4.07	0.07
H	-3.86	-0.63	3.10	0.07
H	-4.22	1.07	3.26	0.07

Ligand 13 Atom	X	Y	Z	RESP Charge
H	0.76	-4.03	-1.14	0.05
H	2.40	-4.17	-1.73	0.05
C	1.54	-2.23	-2.01	-0.11
H	1.35	-2.49	-3.05	0.08
H	2.45	-1.66	-1.99	0.08
C	0.38	-1.40	-1.52	0.32
N	0.41	-0.04	-1.42	-0.07
C	1.54	0.86	-1.62	-0.02
H	2.18	0.47	-2.40	0.07
H	1.14	1.80	-1.99	0.07
C	2.35	1.05	-0.34	-0.05
H	2.69	0.08	-0.01	0.08
H	1.70	1.45	0.44	0.08
N	3.51	1.89	-0.58	-0.36
C	3.18	3.29	-0.79	-0.03
H	2.39	3.62	-0.11	0.08
H	2.83	3.45	-1.80	0.08
C	4.40	4.17	-0.55	0.10
H	5.19	3.89	-1.24	0.05
H	4.15	5.21	-0.73	0.05
O	4.86	4.08	0.78	-0.39
C	4.69	2.83	1.39	0.10
H	5.53	2.68	2.06	0.05
H	3.78	2.84	2.00	0.05
C	4.61	1.71	0.37	-0.03
H	4.50	0.76	0.88	0.08
H	5.55	1.66	-0.17	0.08
C	-0.84	0.34	-0.98	0.03
C	-1.41	1.58	-0.69	-0.18
H	-0.86	2.50	-0.78	0.13
C	-2.73	1.59	-0.28	-0.19
H	-3.20	2.53	-0.06	0.14
C	-3.48	0.40	-0.15	-0.06
C	-2.89	-0.82	-0.44	-0.10
H	-3.44	-1.74	-0.33	0.10
C	-1.56	-0.84	-0.86	0.24
N	-0.77	-1.91	-1.21	-0.58
C	-4.89	0.47	0.30	-0.13
C	-6.01	0.13	-0.37	0.29
C	-6.27	-0.39	-1.74	-0.29
H	-6.72	-1.38	-1.69	0.10
H	-6.96	0.27	-2.26	0.10
H	-5.35	-0.45	-2.29	0.10
O	-7.07	0.34	0.40	-0.10
N	-6.68	0.82	1.62	-0.47
C	-5.40	0.91	1.57	0.37
C	-4.62	1.39	2.76	-0.15
H	-5.30	1.57	3.59	0.06
H	-3.88	0.66	3.05	0.06
H	-4.10	2.32	2.53	0.06
C	2.22	-3.28	0.20	-0.05
C	1.51	-3.12	1.34	-0.05
H	0.43	-3.17	1.36	0.13
C	2.32	-2.87	2.49	-0.32
H	1.92	-2.72	3.48	0.18
C	3.63	-2.85	2.21	-0.11
H	4.45	-2.71	2.89	0.17
S	3.92	-3.14	0.53	-0.05

Ligand 14

Atom	X	Y	Z	RESP Charge
C	1.40	-3.00	-1.98	-0.04
H	0.54	-3.57	-1.64	0.07
H	1.96	-3.61	-2.68	0.07
C	0.90	-1.74	-2.71	-0.07
H	0.54	-2.03	-3.69	0.07
H	1.73	-1.06	-2.87	0.07
C	-0.24	-1.07	-2.00	0.31
N	-0.27	0.27	-1.68	-0.03
C	0.75	1.28	-1.87	-0.02
H	1.34	1.04	-2.74	0.07
H	0.24	2.21	-2.08	0.07
C	1.66	1.42	-0.65	0.00
H	2.13	0.46	-0.48	0.07
H	1.05	1.63	0.24	0.07
N	2.69	2.41	-0.87	-0.38
C	2.18	3.78	-0.85	-0.02
H	1.43	3.92	-0.07	0.07
H	1.71	4.02	-1.80	0.07
C	3.32	4.77	-0.61	0.12
H	4.06	4.68	-1.40	0.05
H	2.93	5.78	-0.62	0.05
O	3.92	4.57	0.65	-0.40
C	3.95	3.24	1.10	0.12
H	4.87	3.11	1.66	0.05
H	3.12	3.06	1.78	0.05
C	3.90	2.25	-0.06	-0.02
H	3.95	1.24	0.33	0.07
H	4.77	2.40	-0.69	0.07
C	-1.50	0.47	-1.08	0.01
C	-2.11	1.60	-0.54	-0.21
H	-1.63	2.56	-0.53	0.14
C	-3.38	1.45	-0.02	-0.14
H	-3.88	2.30	0.39	0.13
C	-4.04	0.20	-0.02	-0.08
C	-3.41	-0.91	-0.55	-0.12
H	-3.88	-1.87	-0.55	0.11
C	-2.13	-0.77	-1.09	0.28
N	-1.31	-1.70	-1.68	-0.61
C	-5.40	0.09	0.56	-0.14
C	-6.56	-0.21	-0.05	0.29
C	-6.93	-0.52	-1.46	-0.30
H	-7.31	-1.53	-1.54	0.11
H	-7.70	0.16	-1.80	0.11
H	-6.06	-0.43	-2.10	0.11
O	-7.55	-0.21	0.84	-0.10
N	-7.06	0.09	2.07	-0.47
C	-5.80	0.28	1.93	0.38
C	-4.94	0.62	3.11	-0.17
H	-5.53	0.61	4.01	0.07
H	-4.13	-0.10	3.20	0.07
H	-4.49	1.60	2.99	0.07
C	2.26	-2.69	-0.79	0.27
N	1.81	-2.49	0.39	-0.50
C	2.83	-2.18	1.28	0.25
C	4.11	-2.15	0.73	0.09
C	5.22	-1.85	1.51	-0.17
H	6.21	-1.83	1.08	0.17
C	5.03	-1.59	2.85	-0.18
H	5.87	-1.35	3.47	0.14
C	3.75	-1.62	3.40	-0.14
H	3.62	-1.41	4.45	0.14
C	2.65	-1.91	2.63	-0.16
H	1.66	-1.94	3.05	0.13
S	4.00	-2.53	-0.97	-0.13

Ligand 15 Atom	X	Y	Z	RESP Charge
C	-1.72	2.89	-2.14	0.00
H	-0.82	3.46	-1.95	0.07
H	-2.32	3.45	-2.85	0.07
C	-1.35	1.54	-2.79	-0.10
H	-1.02	1.74	-3.81	0.07
H	-2.23	0.92	-2.85	0.07
C	-0.22	0.84	-2.09	0.33
N	-0.28	-0.45	-1.64	-0.05
C	-1.40	-1.38	-1.67	-0.01
H	-1.99	-1.20	-2.56	0.07
H	-0.99	-2.37	-1.77	0.07
C	-2.28	-1.26	-0.43	0.00
H	-2.66	-0.25	-0.38	0.07
H	-1.67	-1.41	0.46	0.07
N	-3.40	-2.18	-0.50	-0.38
C	-3.03	-3.57	-0.28	-0.02
H	-2.27	-3.66	0.50	0.07
H	-2.61	-4.00	-1.19	0.07
C	-4.24	-4.39	0.13	0.11
H	-4.99	-4.35	-0.66	0.05
H	-3.95	-5.43	0.26	0.05
O	-4.78	-3.95	1.34	-0.40
C	-4.69	-2.57	1.58	0.11
H	-5.58	-2.27	2.12	0.05
H	-3.83	-2.36	2.22	0.05
C	-4.57	-1.79	0.28	-0.02
H	-4.53	-0.73	0.50	0.07
H	-5.47	-1.95	-0.31	0.07
C	0.96	-0.71	-1.09	0.01
C	1.50	-1.84	-0.48	-0.18
H	0.95	-2.75	-0.35	0.13
C	2.81	-1.74	-0.04	-0.16
H	3.26	-2.60	0.43	0.14
C	3.57	-0.57	-0.18	-0.09
C	3.00	0.54	-0.79	-0.09
H	3.56	1.46	-0.90	0.10
C	1.69	0.46	-1.25	0.25
N	0.91	1.41	-1.88	-0.60
C	4.96	-0.52	0.31	-0.11
C	6.11	-0.37	-0.38	0.28
C	6.42	-0.21	-1.83	-0.28
H	6.88	0.75	-2.01	0.10
H	7.11	-0.98	-2.14	0.10
H	5.51	-0.29	-2.41	0.10
O	7.14	-0.38	0.45	-0.10
N	6.70	-0.54	1.74	-0.47
C	5.43	-0.63	1.67	0.36
C	4.61	-0.80	2.92	-0.15
H	5.25	-0.76	3.79	0.06
H	3.86	-0.01	2.99	0.06
H	4.09	-1.75	2.91	0.06
N	-2.48	2.76	-0.92	0.01
C	-2.00	2.83	0.38	0.07
C	-3.10	2.70	1.25	0.11
C	-4.27	2.55	0.42	-0.32
H	-5.28	2.45	0.74	0.18
C	-3.84	2.60	-0.86	-0.20
H	-4.41	2.55	-1.77	0.18
C	-2.89	2.75	2.63	-0.19
H	-3.72	2.66	3.31	0.13
C	-1.61	2.92	3.10	-0.17
H	-1.44	2.96	4.17	0.13
C	-0.53	3.04	2.22	-0.18
H	0.46	3.17	2.62	0.14
C	-0.70	3.00	0.86	-0.15
H	0.14	3.08	0.19	0.13

Table A.2: The MM-PB(GB)SA analysis was performed with an internal dielectric constant ϵ_{int} of 1; the MM-PBSA analysis was also repeated with $\epsilon_{\text{int}} = 4$, giving worse results: the correlation coefficient r_s drops from 0.59 to 0.38. This difference is due to the importance of electrostatics (and thus dielectric constant) for quantification of the strength of cation–π interactions. Both PB and GB surface area models give similar levels of correlation with experimental binding affinities, while consideration of just the electrostatic term improves this correlation. The intermolecular electrostatic term tends to correlate more strongly with QM-computed results than does the total classical intermolecular energy.

Ligand	MM-PBSA	MM-GBSA	Electrostatic	MM-PBSA $\epsilon_{int} = 4.0$
L1	-33.5 ± 2.7	-36.4 ± 4.6	-2.1 ± 1.1	-43.4 ± 2.6
L2	-35.9 ± 3.3	-38.9 ± 3.4	-1.8 ± 1.3	-48.1 ± 2.5
L3	-34.2 ± 3.1	-38.1 ± 2.6	-1.8 ± 1.3	-45.1 ± 2.8
L4	-33.0 ± 3.3	-37.0 ± 3.3	-1.3 ± 1.1	-46.4 ± 2.2
L5	-36.1 ± 2.7	-39.0 ± 2.8	-2.5 ± 1.4	-44.8 ± 2.6
L6	-34.2 ± 2.3	-37.8 ± 2.5	-1.1 ± 2.3	-44.8 ± 2.6
L7	-32.0 ± 2.5	-35.8 ± 2.6	0.1 ± 2.5	-46.4 ± 2.6
L8	-33.0 ± 2.8	-35.8 ± 3.4	-1.2 ± 0.9	-45.3 ± 2.7
L9	-34.6 ± 3.0	-36.9 ± 2.6	-1.6 ± 1.1	-44.4 ± 2.3
L10	-32.9 ± 2.5	-36.0 ± 2.6	-1.9 ± 1.2	-45.2 ± 2.2
L11	-37.1 ± 2.3	-40.2 ± 2.1	-3.0 ± 1.3	-43.9 ± 2.4
L12	-36.7 ± 2.5	-38.7 ± 2.4	-2.1 ± 1.2	-47.5 ± 2.4
L13	-33.7 ± 2.6	-36.2 ± 2.5	0.9 ± 1.2	-43.8 ± 2.2
L14	-35.2 ± 2.5	-37.8 ± 2.4	-2.5 ± 1.7	-46.6 ± 3.0
L15	-34.8 ± 2.4	-37.1 ± 2.4	-4.9 ± 2.1	-46.1 ± 2.4
	$r_s = 0.58$	$r_s = 0.53$	$r_s = 0.72$	$r_s = 0.38$

B

Fe(II)-αKG Demethylase Substrates - Appendix

DFT vs G4 $BDE_{C\text{-}H}$ models; Enthalpy values for model substrates; T_1 Diagnostic calculation values; XYZ Coordinates for model substrates

B.1 DFT vs G4 $BDE_{C\text{-}H}$ models

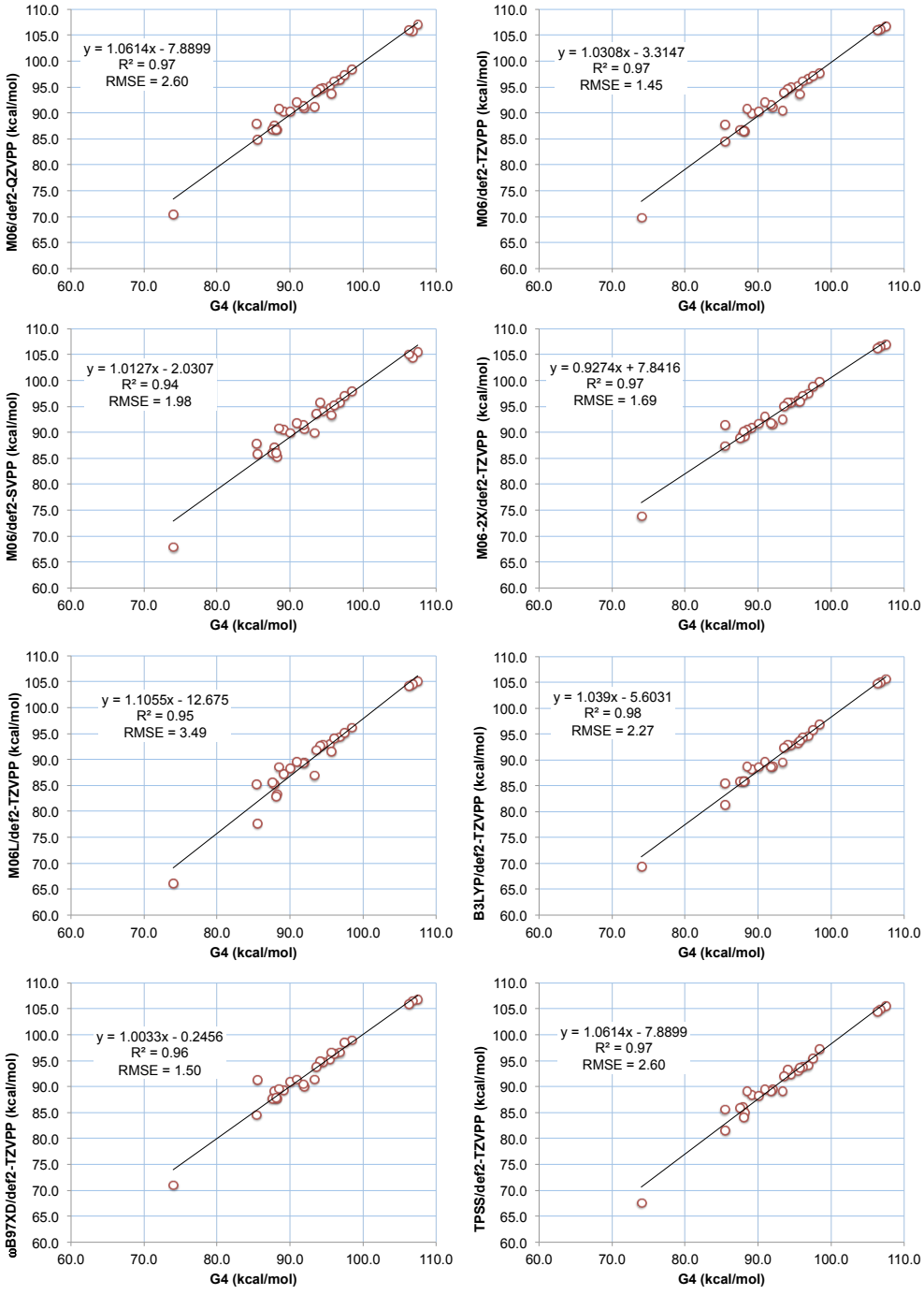


Figure B.1: Models for predictions of $BDE_{C\text{-}H}$ using DFT methods: M06/def2-QZVPP, M06/def2-TZVPP, M06/def2-SVPP, M06L/def2-TZVPP, M06-2X/def2-TZVPP, TPSS/def2-TZVPP, B3LYP/def2-TZVPP, and ωB97XD/def2-TZVPP.

B.2 Enthalpy values for model substrates

Table B.1: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **Gaussian-4** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o(\text{R-CH}_2\cdot + \text{H}\cdot) - H^o(\text{R-CH}_3)$

Functional **G4**	Substrate	$H^\bullet$ (a.u.) = -0.499060		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.39	-134.72	107.49
	Kme2	-174.67	-174.00	106.81
	Kme3	-213.95	-213.28	106.39
	ω-Rme1	-284.18	-283.53	96.88
	γ-Rme1	-284.18	-283.53	95.51
	sRme2	-323.45	-322.80	96.13
	aRme2	-323.45	-322.80	94.52
	nKme1	-135.04	-134.40	91.96
	nKme2	-174.31	-173.67	91.76
	Asn	-248.38	-247.73	93.36
	Pro	-251.69	-251.04	95.64
Nucleobases	1meA	-546.02	-545.36	98.40
	3meC	-473.66	-473.01	97.48
	6meA	-545.65	-545.01	89.11
	4meC	-473.29	-472.64	90.13
	1meG	-620.87	-620.22	94.10
	2meG	-620.86	-620.22	90.95
	3meT	-532.47	-531.82	93.54
	5meC	-473.30	-472.66	87.86
	5hmC	-548.50	-547.86	88.21
	5fChydrate	-623.71	-623.07	85.52
	5fC	-547.32	-546.68	88.55
	Thy	-493.19	-492.56	87.52
	5hmU	-568.39	-567.75	88.08
	5fUhydrate	-643.60	-642.99	73.99
	5fU	-567.19	-566.56	85.44

Table B.2: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional M06 and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **M06/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.49751		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.36	-134.70	106.63
	Kme2	-174.63	-173.96	106.11
	Kme3	-213.90	-213.23	106.01
	ω-Rme1	-284.15	-283.50	96.52
	γ-Rme1	-284.15	-283.50	95.28
	sRme2	-323.41	-322.76	96.13
	aRme2	-323.41	-322.76	94.95
	nKme1	-135.02	-134.37	91.08
	nKme2	-174.28	-173.63	91.57
	Asn	-248.36	-247.72	90.51
	Pro	-251.63	-250.99	93.59
Nucleobases	1meA	-545.97	-545.31	97.68
	3meC	-473.63	-472.98	97.21
	6meA	-545.75	-545.11	89.88
	4meC	-473.26	-472.62	90.29
	1meG	-620.83	-620.19	94.63
	2meG	-620.83	-620.19	92.09
	3meT	-532.44	-531.80	93.93
	5meC	-473.27	-472.64	86.72
	5hmC	-548.48	-547.85	86.46
	5fChydrate	-623.71	-623.07	84.54
	5fC	-547.31	-546.67	90.75
	Thy	-493.18	-492.55	86.79
	5hmU	-568.39	-567.76	86.48
	5fUhydrate	-643.62	-643.01	69.81
	5fU	-567.21	-566.57	87.86

Table B.3: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional M06 and basis set def2-QZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **M06/def2-QZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.497673		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.38	-134.71	106.94
	Kme2	-174.65	-173.98	105.78
	Kme3	-213.90	-213.23	105.90
	ω-Rme1	-284.15	-283.50	96.42
	γ-Rme1	-284.18	-283.53	95.31
	sRme2	-323.41	-322.76	96.06
	aRme2	-323.41	-322.76	94.85
	nKme1	-135.03	-134.39	91.00
	nKme2	-174.29	-173.65	91.40
	Asn	-248.38	-247.74	91.27
	Pro	-251.66	-251.01	93.80
Nucleobases	6meA	-545.64	-544.99	90.40
	1meA	-546.01	-545.35	98.39
	3meC	-473.67	-473.02	97.29
	4meC	-473.30	-472.65	90.34
	1meG	-620.88	-620.23	94.70
	2meG	-620.88	-620.24	92.09
	3meT	-532.48	-531.84	94.00
	5meC	-473.31	-472.67	87.46
	5hmC	-548.52	-547.89	86.76
	5fChydrate	-623.75	-623.12	84.80
	5fC	-547.35	-546.71	90.85
	Thy	-493.22	-492.58	86.94
	5hmU	-568.44	-567.80	86.77
	5fUhydrate	-643.67	-643.06	70.37
	5fU	-567.25	-566.61	87.93

Table B.4: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional M06 and basis set def2-SVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **M06/def2-SVPP**	Substrate	$H^\bullet$ (a.u.) = -0.495443		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.18	-134.52	105.51
	Kme2	-174.40	-173.74	104.50
	Kme3	-213.63	-212.96	105.12
	ω-Rme1	-283.80	-283.15	95.72
	γ-Rme1	-283.80	-283.15	94.69
	sRme2	-323.02	-322.37	95.20
	aRme2	-323.01	-322.37	94.05
	nKme1	-134.84	-134.20	90.67
	nKme2	-174.06	-173.42	91.33
	Asn	-248.06	-247.42	89.96
	Pro	-251.33	-250.68	93.27
Nucleobases	1meA	-545.34	-544.69	97.95
	3meC	-473.08	-472.43	96.98
	6meA	-544.97	-544.33	90.47
	4meC	-472.71	-472.07	89.96
	1meG	-620.13	-619.48	95.64
	2meG	-620.12	-619.48	91.84
	3meT	-531.83	-531.18	93.60
	5meC	-472.72	-472.08	87.09
	5hmC	-547.84	-547.21	85.31
	5fChydrate	-623.71	-623.07	85.84
	5fC	-546.68	-546.04	90.77
	Thy	-492.61	-491.98	86.03
	5hmU	-567.73	-567.10	85.96
	5fUhydrate	-642.86	-642.26	67.90
	5fU	-566.56	-565.92	87.87

Table B.5: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional M06-2X and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates.$BDE_{C\text{-}H} = H^o(\text{R-}CH_2\cdot + H\cdot) - H^o(\text{R-}CH_3)$

Functional/Basis Set **M06-2X/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.495778		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.40	-134.73	106.94
	Kme2	-174.68	-174.01	106.61
	Kme3	-213.96	-213.29	106.29
	ω-Rme1	-284.22	-283.57	97.45
	γ-Rme1	-284.22	-283.57	96.14
	sRme2	-323.49	-322.84	97.06
	aRme2	-323.49	-322.84	95.63
	nKme1	-135.05	-134.40	91.57
	nKme2	-174.32	-173.68	91.81
	Asn	-248.41	-247.77	92.56
	Pro	-251.71	-251.06	96.00
Nucleobases	1meA	-546.12	-545.46	99.64
	3meC	-473.75	-473.10	98.85
	6meA	-545.75	-545.11	90.96
	4meC	-473.38	-472.73	91.67
	1meG	-620.99	-620.34	95.64
	2meG	-620.99	-620.35	93.08
	3meT	-532.57	-531.92	95.11
	5meC	-473.39	-472.75	89.54
	5hmC	-548.61	-547.97	89.24
	5fChydrate	-623.84	-623.19	91.38
	5fC	-547.43	-546.79	90.41
	Thy	-493.29	-492.66	88.86
	5hmU	-568.51	-567.87	90.17
	5fUhydrate	-643.75	-643.13	73.81
	5fU	-567.32	-566.69	87.32

Table B.6: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional M06L and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **M06L/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.501055		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.45	-134.78	105.09
	Kme2	-174.74	-174.07	104.54
	Kme3	-214.03	-213.36	104.20
	ω-Rme1	-284.31	-283.66	94.38
	γ-Rme1	-284.31	-283.66	93.05
	sRme2	-323.60	-322.95	94.01
	aRme2	-323.60	-322.95	92.81
	nKme1	-135.10	-134.45	89.47
	nKme2	-174.38	-173.74	89.36
	Asn	-248.49	-247.85	86.91
	Pro	-251.79	-251.14	91.46
Nucleobases	1meA	-546.27	-545.62	96.15
	3meC	-473.89	-473.24	95.10
	6meA	-545.89	-545.25	87.18
	4meC	-473.51	-472.87	88.22
	1meG	-621.15	-620.51	92.57
	2meG	-621.15	-620.51	89.52
	3meT	-532.71	-532.07	91.76
	5meC	-473.52	-472.88	85.29
	5hmC	-548.75	-548.12	83.38
	5fChydrate	-623.99	-623.36	77.68
	5fC	-547.31	-546.67	88.52
	Thy	-493.43	-492.79	85.52
	5hmU	-568.66	-568.02	82.96
	5fUhydrate	-643.90	-643.29	66.09
	5fU	-567.47	-566.83	85.28

Table B.7: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional B3LYP and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **B3LYP/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.499794		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.48	-134.81	105.64
	Kme2	-174.78	-174.11	105.06
	Kme3	-214.08	-213.41	104.71
	ω-Rme1	-284.36	-283.71	94.59
	γ-Rme1	-284.36	-283.71	93.27
	sRme2	-323.65	-323.00	94.23
	aRme2	-323.65	-323.00	92.76
	nKme1	-135.13	-134.49	88.81
	nKme2	-174.42	-173.78	88.69
	Asn	-248.53	-247.89	89.53
	Pro	-251.84	-251.19	93.72
Nucleobases	1meA	-546.35	-545.69	96.78
	3meC	-473.96	-473.30	95.76
	6meA	-545.97	-545.33	88.24
	4meC	-473.58	-472.94	88.65
	1meG	-621.24	-620.59	92.95
	2meG	-621.24	-620.60	89.72
	3meT	-532.79	-532.14	92.41
	5meC	-473.59	-472.95	86.01
	5hmC	-548.83	-548.20	85.88
	5fChydrate	-624.09	-623.46	81.25
	5fC	-547.66	-547.02	88.79
	Thy	-493.50	-492.86	85.81
	5hmU	-568.74	-568.10	85.76
	5fUhydrate	-644.00	-643.39	69.34
	5fU	-567.55	-566.91	85.51

Table B.8: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional TPSS and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **TPSS/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.497506		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.50	-134.83	105.48
	Kme2	-174.80	-174.14	104.88
	Kme3	-214.11	-213.45	104.53
	ω-Rme1	-284.40	-283.75	94.10
	γ-Rme1	-284.39	-283.75	93.02
	sRme2	-323.70	-323.05	93.80
	aRme2	-323.70	-323.05	92.47
	nKme1	-135.14	-134.50	89.39
	nKme2	-174.44	-173.80	89.21
	Asn	-248.56	-247.92	89.16
	Pro	-251.87	-251.23	93.62
Nucleobases	1meA	-546.44	-545.79	97.12
	3meC	-474.03	-473.38	95.35
	6meA	-546.07	-545.43	88.45
	4meC	-473.65	-473.01	88.27
	1meG	-621.34	-620.69	93.32
	2meG	-621.34	-620.70	89.46
	3meT	-532.87	-532.23	92.04
	5meC	-473.66	-473.03	86.02
	5hmC	-548.91	-548.28	85.01
	5fChydrate	-624.17	-623.54	81.49
	5fC	-547.74	-547.10	89.11
	Thy	-493.57	-492.94	85.87
	5hmU	-568.82	-568.19	84.14
	5fUhydrate	-644.08	-643.48	67.55
	5fU	-567.63	-567.00	85.63

Table B.9: Enthalpies of R-CH_3, R-$CH_2\cdot$, and H$\cdot$ and corresponding $BDE_{C\text{-}H}$ (kcal/mol) calculated using **DFT functional ωB97XD and basis set def2-TZVPP** for 26 native and non-native Fe(II)/α-KG demethylase substrates. $BDE_{C\text{-}H} = H^o$(R-$CH_2\cdot$ + H$\cdot$) – H^o(R-CH_3)

Functional/Basis Set **ωB97XD/def2-TZVPP**	Substrate	$H^\bullet$ (a.u.) = -0.500306		$BDE_{C\text{-}H}$ (kcal/mol)
		R-CH_3 (a.u.)	R-$CH_2^\bullet$ (a.u.)	
Residues	Kme1	-135.43	-134.76	106.83
	Kme2	-174.72	-174.05	106.50
	Kme3	-214.01	-213.34	105.99
	ω-Rme1	-284.27	-283.61	96.56
	γ-Rme1	-284.26	-283.61	95.27
	sRme2	-323.55	-322.89	96.27
	aRme2	-323.55	-322.89	94.63
	nKme1	-135.08	-134.43	90.11
	nKme2	-174.36	-173.72	90.54
	Asn	-248.45	-247.80	91.38
	Pro	-251.76	-251.10	96.52
Nucleobases	1meA	-546.16	-545.50	98.99
	3meC	-473.80	-473.14	98.45
	6meA	-545.78	-545.14	89.27
	4meC	-473.42	-472.77	90.91
	1meG	-621.03	-620.38	95.00
	2meG	-621.03	-620.38	91.35
	3meT	-532.61	-531.96	93.74
	5meC	-473.43	-472.79	89.11
	5hmC	-548.65	-548.01	87.69
	5fChydrate	-623.88	-623.23	91.17
	5fC	-547.47	-546.83	89.57
	Thy	-493.33	-492.69	87.60
	5hmU	-568.55	-567.91	87.88
	5fUhydrate	-643.75	-643.13	70.97
	5fU	-567.32	-566.69	84.48

Table B.10: Energies (a.u.) of the iron active site models [F=O] and [F-OH] and corresponding enthalpy values calculated using DFT functional **M06**. **M06/def2-QZVPP** enthalpy values calculated using the thermal correction to enthalpy determined by **M06/def2-TZVPP** calculations, 0.362912 a.u. and 0.374727 a.u., respectively.

Basis Set	[Fe=O]		[Fe-OH]	
	E (a.u.)	H (a.u.)	E (a.u.)	H (a.u.)
def2-TZVPP	-2403.359901	-2402.996989	-2404.003649	-2403.628922
def2-QZVPP	-2403.248474	-2402.885562	-2403.892124	-2403.517397

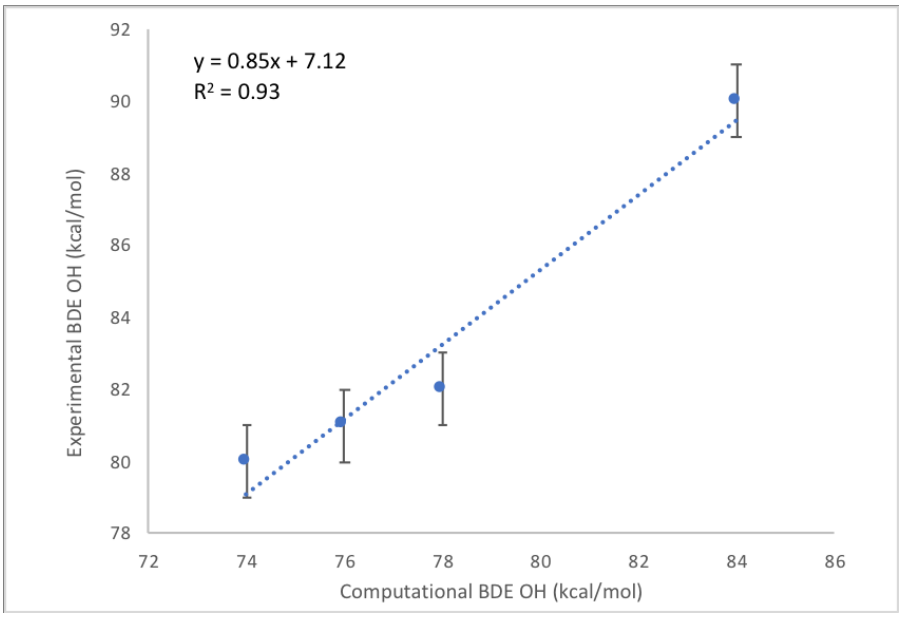


Figure B.2: Plot of computational vs experimental $BDE_{O\text{-}H}$ values for metal complexes.

B.3 T_1 Diagnostic

Table B.11: The diagnostic of Lee and co-workers, performed at CCSD/GTBas1 level of theory; as used in composite G4 calculations.

Substrate	R-CH_3 (a.u.)	*T1*	R-$CH_2^{\bullet}$ (a.u.)	*T1*
Kme1	-135.39	0.008	-134.72	0.009
Kme2	-174.67	0.008	-174.00	0.009
Kme3	-213.95	0.008	-213.28	0.009
ω-Rme1	-284.18	0.012	-283.53	0.016
γ-Rme1	-284.18	0.013	-283.53	0.016
sRme2	-323.45	0.012	-322.80	0.015
aRme2	-323.45	0.012	-322.80	0.016
nKme1	-135.04	0.008	-134.40	0.009
nKme2	-174.31	0.009	-173.67	0.009
Asn	-248.38	0.013	-247.73	0.017
Pro	-251.69	0.009	-251.04	0.010
1meA	-546.02	0.015	-545.36	0.019
3meC	-473.66	0.017	-473.01	0.020
6meA	-545.65	0.015	-545.01	0.022
4meC	-473.29	0.017	-472.64	0.019
1meG	-620.87	0.016	-620.22	0.018
2meG	-620.86	0.015	-620.22	0.018
3meT	-532.47	0.016	-531.82	0.020
5meC	-473.30	0.016	-472.66	0.023
5hmC	-548.50	0.016	-547.86	0.024
5fChydrate	-623.71	0.016	-623.07	0.022
5fC	-547.32	0.018	-546.68	0.019
Thy	-493.19	0.016	-492.56	0.024
5hmU	-568.39	0.016	-567.75	0.024
5fUhydrate	-643.60	0.016	-642.99	0.020
5fU	-567.19	0.017	-566.56	0.019

B.4 G4 optimised coordinates of model substrates

Table B.12: Coordinates of G4 optimised model 1meA and following H-abstraction (1meA ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	1meA				1meA ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-2.160466	-0.090398	-0.000016	N	-2.115827	-0.074844	0.006723
C	-2.172981	1.289784	0.000014	C	-2.112865	1.307261	0.019623
H	-3.105073	1.836879	0.000004	H	-3.037947	1.866271	0.024715
N	-0.978471	1.834797	-0.000036	N	-0.912288	1.836785	0.021791
C	-0.135719	0.756395	-0.000006	C	-0.079462	0.748671	0.01852
C	1.263713	0.699074	-0.000008	C	1.315176	0.687437	-0.0127
N	2.027891	1.790013	-0.000025	N	2.090362	1.772651	-0.034501
H	3.032838	1.774851	-0.000074	H	3.058472	1.750402	-0.308855
H	1.561671	2.686115	-0.00005	H	1.61773	2.666027	-0.065653
N	1.799205	-0.563899	0.000015	N	1.854315	-0.591447	-0.001055
C	0.98645	-1.685947	0.000026	C	1.00889	-1.714342	-0.031754
H	1.521555	-2.628284	0.000032	H	1.536827	-2.660163	-0.04662
N	-0.309921	-1.681213	0.000025	N	-0.281966	-1.684437	-0.03069
C	-0.845633	-0.445821	0.000013	C	-0.806632	-0.444188	0.006307
C	-3.31178	-0.991339	-0.000008	C	-3.279252	-0.961537	-0.000809
H	-3.291153	-1.622362	-0.8903	H	-2.923401	-1.986441	-0.102718
H	-3.291558	-1.621857	0.890653	H	-3.834282	-0.859888	0.934077
H	-4.221274	-0.390937	-0.000386	H	-3.927389	-0.714725	-0.843588
C	3.260611	-0.73658	0.000021	C	3.224262	-0.798927	0.053653
H	3.693522	-0.286181	-0.89726	H	3.581201	-1.811513	-0.048491
H	3.693522	-0.286068	0.897245	H	3.865917	-0.007174	0.404229
H	3.490311	-1.800647	0.000091				

Table B.13: Coordinates of G4 optimised model 1meG and following H-abstraction (1meG ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	1meG				1meG ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-2.260407	0.29983	-0.004888	N	2.219069	0.308213	0.011597
C	-2.604922	-1.040511	-0.003851	C	2.572251	-1.029262	-0.002347
H	-3.642516	-1.343621	-0.009028	H	3.610796	-1.328607	0.002118
N	-1.576192	-1.843749	0.004264	N	1.548125	-1.838756	-0.026241
C	-0.484295	-0.997521	0.007833	C	0.450156	-0.999984	-0.02184
C	0.91482	-1.30113	0.013335	C	-0.940838	-1.320643	-0.031865
O	1.476007	-2.377168	0.019112	O	-1.504985	-2.38975	-0.056098
N	1.703629	-0.075275	0.002609	N	-1.759304	-0.082995	0.026662
C	1.157181	1.183715	0.00914	C	-1.204628	1.188065	-0.023439
N	2.027078	2.249691	-0.053298	N	-2.086793	2.247317	-0.023995
H	2.923715	2.155648	0.395935	H	-2.926456	2.122511	-0.572203
H	1.562502	3.127827	0.124412	H	-1.612662	3.120803	-0.204848
N	-0.128904	1.445694	0.015969	N	0.079418	1.439964	-0.026783
C	-0.891063	0.331003	0.002258	C	0.849164	0.330978	0.001368
C	-3.146531	1.445875	-0.012984	C	3.097635	1.460709	0.061894
H	-2.99044	2.061412	0.877255	H	2.53867	2.333741	-0.277545
H	-2.969204	2.064424	-0.896996	H	3.455183	1.645498	1.080148
H	-4.17899	1.092215	-0.025843	H	3.956915	1.303653	-0.5944
C	3.153125	-0.246999	-0.048257	C	-3.125081	-0.268305	0.125567
H	3.57589	0.306576	-0.891008	H	-3.725658	0.529411	0.532999
H	3.628067	0.086557	0.882535	H	-3.452475	-1.294559	0.097816
H	3.336598	-1.313602	-0.167589				

Table B.14: Coordinates of G4 optimised model 2meG and following H-abstraction (2meG·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	2meG				2meG·		
Atom	X	Y	Z	Atom	X	Y	Z
N	2.148864	-0.790092	0.000803	N	2.101467	-0.792967	-0.006588
C	2.893064	0.377145	0.002181	C	2.843976	0.372909	-0.027545
H	3.973393	0.342809	0.003548	H	3.924274	0.339913	-0.039688
N	2.163633	1.459074	0.002226	N	2.114144	1.456326	-0.031581
C	0.863398	0.995897	-0.000068	C	0.815785	0.994782	-0.013744
C	-0.381316	1.70947	-0.001416	C	-0.428892	1.714797	0.012208
O	-0.637222	2.894341	-0.000311	O	-0.672264	2.900754	0.031878
N	-1.470903	0.765988	-0.002777	N	-1.520672	0.776755	0.027439
C	-1.373336	-0.596855	-0.003841	C	-1.418767	-0.583927	0.027702
N	-2.525981	-1.319358	-0.007813	N	-2.573633	-1.320423	0.031644
H	-2.385457	-2.314998	0.000846	H	-2.397256	-2.308093	0.144635
N	-0.225018	-1.238694	-0.002518	N	-0.276097	-1.234634	0.017595
C	0.83612	-0.399115	-0.000896	C	0.788678	-0.40047	0.002699
C	2.636187	-2.153815	0.001098	C	2.589543	-2.157198	0.002153
H	2.286535	-2.689197	0.887989	H	2.259052	-2.680076	0.903847
H	2.286095	-2.689494	-0.885415	H	2.220765	-2.703704	-0.870082
H	3.72734	-2.138277	0.00084	H	3.680506	-2.14071	-0.021304
H	-2.375273	1.215313	-0.004133	H	-2.425121	1.215414	0.127598
C	-3.864444	-0.778458	0.008417	C	-3.871625	-0.871242	-0.043941
H	-4.574249	-1.606557	0.007026	H	-4.643959	-1.623667	-0.011884
H	-4.059334	-0.175966	0.90581	H	-4.068794	0.081576	-0.514901
H	-4.073464	-0.162391	-0.876313				

Table B.15: Coordinates of G4 optimised model 3meC and following H-abstraction (3meC·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	3meC				3meC·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-0.663273	0.773915	0.000046	C	-0.691885	0.767377	0.023188
N	-1.418417	-0.401902	-0.000176	N	-1.370568	-0.441397	-0.004783
C	-0.839541	-1.620916	-0.000135	C	-0.726502	-1.630585	-0.044726
C	0.510847	-1.789253	0.000015	C	0.632867	-1.720698	-0.036767
C	1.329939	-0.637554	0.000034	C	1.400924	-0.542901	0.008872
N	0.745211	0.587687	-0.000221	N	0.753836	0.671011	-0.019727
H	-1.5158	-2.46734	-0.000201	H	-1.355895	-2.511122	-0.067717
O	-1.168798	1.863575	0.000316	O	-1.23739	1.834276	0.062841
N	2.659189	-0.730627	0.000353	N	2.736215	-0.568365	0.060819
H	3.264907	0.072924	-0.000376	H	3.281935	0.239612	0.31161
H	3.096811	-1.63847	0.000032	H	3.206472	-1.459253	0.10341
C	-2.881491	-0.223689	0.000026	C	-2.843029	-0.346228	0.010528
H	-3.180982	0.335857	0.887252	H	-3.166872	0.156047	0.923124
H	-3.181053	0.336897	-0.886505	H	-3.179576	0.233222	-0.850159
H	-3.349183	-1.207215	-0.000509	H	-3.253026	-1.354104	-0.030511
H	0.949266	-2.77611	0.000153	H	1.119872	-2.685062	-0.038063
C	1.562823	1.812801	-0.00024	C	1.417994	1.875828	-0.082118
H	2.183583	1.84921	0.899409	H	0.812529	2.763717	-0.01004
H	2.184403	1.84863	-0.899277	H	2.455091	1.89723	-0.372415
H	0.880719	2.6591	-0.000678				

Table B.16: Coordinates of G4 optimised model 3meT and following H-abstraction (3meT ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	3meT				3meT ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-1.120002	-0.977482	0.000103	N	1.144032	-0.909823	-0.000629
C	0.110519	-1.587537	0.000034	C	-0.072754	-1.548624	-0.000458
H	0.081607	-2.671521	0.000064	H	-0.018716	-2.630884	-0.000699
C	1.275463	-0.909871	-0.00001	C	-1.253716	-0.895131	-0.00027
C	1.238828	0.549534	-0.000002	C	-1.257768	0.555323	0.000314
O	2.234789	1.251671	0.000011	O	-2.257683	1.249904	0.001344
N	-0.055476	1.117264	-0.000032	N	0.037027	1.174271	-0.000642
C	-1.247644	0.4079	0.000037	C	1.257399	0.470249	-0.000153
O	-2.338015	0.947193	0.000066	O	2.337586	1.027533	0.000783
C	-2.362781	-1.743196	-0.000092	C	2.400861	-1.65523	0.000499
H	-2.959142	-1.503577	0.883359	H	2.992909	-1.40434	-0.88262
H	-2.958695	-1.50386	-0.883929	H	2.990921	-1.405355	0.88526
H	-2.117227	-2.805748	0.000128	H	2.172581	-2.721399	-0.000357
C	-0.183767	2.575583	-0.000049	C	0.098058	2.550264	-0.000833
H	-0.734815	2.902658	0.884385	H	-0.843166	3.07116	-0.001351
H	0.823098	2.986495	0.000043	H	1.075437	2.999282	-0.001324
H	-0.734669	2.902649	-0.884574	C	-2.583542	-1.591531	-0.000178
C	2.62205	-1.571405	-0.00001	H	-3.173954	-1.306824	0.876776
H	3.205504	-1.270965	-0.876396	H	-3.174164	-1.306366	-0.876834
H	3.205515	-1.270907	0.876349	H	-2.459714	-2.677831	-0.000494
H	2.526968	-2.660662	0.000018				

Table B.17: Coordinates of G4 optimised model 4meC and following H-abstraction (4meC ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	4meC				4meC ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.149039	0.88667	-0.000055	C	1.075083	0.8928	0.00026
N	-1.514148	-0.50699	0.000004	N	1.479477	-0.488608	0.000717
C	-0.573707	-1.473529	0.000055	C	0.567269	-1.484505	-0.002639
C	0.753545	-1.178615	0.000059	C	-0.768565	-1.230883	-0.003988
C	1.08724	0.213981	0.000039	C	-1.137906	0.147926	-0.002131
N	0.181539	1.176352	-0.000013	N	-0.265307	1.141995	-0.001192
H	-0.937179	-2.495977	0.000085	H	0.961469	-2.495101	-0.004789
O	-2.05143	1.702908	-0.000043	O	1.949683	1.738111	0.001489
N	2.384049	0.612509	0.000004	N	-2.451627	0.538246	-0.001028
C	-2.939293	-0.807621	0.000002	C	2.913061	-0.746989	0.002795
H	-3.415516	-0.371182	0.881013	H	3.377453	-0.295207	-0.876837
H	-3.415492	-0.371287	-0.881075	H	3.374447	-0.298175	0.885543
H	-3.077058	-1.890695	0.000064	H	3.082987	-1.825328	0.001356
H	1.494956	-1.962918	0.000087	H	-1.482852	-2.039235	-0.010891
H	2.509553	1.612084	0.000098	H	-2.541381	1.546199	-0.009935
C	3.539122	-0.255008	-0.000048	C	-3.590121	-0.219017	-0.004843
H	3.57317	-0.89737	-0.889224	H	-3.524776	-1.291938	0.053655
H	4.438451	0.36308	-0.000107	H	-4.535539	0.29648	0.023791
H	3.573269	-0.897357	0.889135				

Table B.18: Coordinates of G4 optimised model ω-Rme1 and following H-abstraction (ω-Rme1 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	ω-Rme1				ω-Rme1 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-1.339259	0.436337	0.020511	N	-1.267604	0.459387	0.043692
H	-1.799752	1.325889	0.12868	H	-1.706717	1.351696	0.211846
C	-0.003008	0.427823	0.024175	C	0.067231	0.425005	0.060757
C	-2.178859	-0.758127	-0.038843	C	-2.139258	-0.706196	-0.100652
H	-3.212367	-0.439129	-0.168853	H	-3.138107	-0.353657	-0.356994
H	-1.906529	-1.379991	-0.896795	H	-1.785169	-1.34612	-0.913143
H	-2.116274	-1.34388	0.884868	H	-2.208139	-1.282451	0.828889
N	0.66451	-0.725564	0.048506	N	0.711297	-0.756675	0.129138
H	0.131919	-1.567306	0.195259	H	0.163667	-1.55081	0.430862
N	0.667143	1.586832	-0.014189	N	0.784016	1.553644	-0.029021
H	0.190246	2.463036	-0.1466	H	0.338871	2.436957	-0.218211
H	1.666189	1.617636	0.094904	H	1.733308	1.576084	0.310226
C	2.119478	-0.847218	-0.034184	C	2.06243	-0.960266	-0.087085
H	2.608321	-0.462674	0.867221	H	2.45205	-1.941865	0.128123
H	2.365993	-1.903898	-0.131016	H	2.613857	-0.235579	-0.666374
H	2.499829	-0.327789	-0.918347				

Table B.19: Coordinates of G4 optimised model 5fC and following H-abstraction (5fC ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5fC				5fC ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.547716	0.78574	0.000061	C	-1.508743	0.78508	0.000189
N	-1.423926	-0.670365	-0.000129	N	-1.406041	-0.672009	-0.000133
C	-0.227555	-1.254731	-0.000115	C	-0.220714	-1.275599	-0.000174
C	0.9395	-0.52925	-0.000131	C	0.95646	-0.563969	-0.000207
C	0.780521	0.91555	-0.000134	C	0.821353	0.884959	-0.000211
N	-0.398374	1.506876	0.000086	N	-0.347974	1.491405	0.000139
H	-0.208487	-2.341663	-0.000109	H	-0.205414	-2.361005	-0.000192
O	-2.674685	1.232149	0.000435	O	-2.627059	1.250504	0.000487
N	1.883366	1.676969	-0.000585	N	1.933566	1.636273	-0.000817
H	1.772516	2.677143	0.000266	H	1.832361	2.637388	0.000554
H	2.788861	1.225323	0.000153	H	2.843	1.201772	0.000475
C	2.213907	-1.215724	0.000113	C	2.211486	-1.270585	0.000057
C	-2.669925	-1.429925	-0.000166	C	-2.663594	-1.413513	-0.000191
H	-3.260686	-1.173795	0.881446	H	-3.249839	-1.148443	0.881692
H	-3.260445	-1.174196	-0.882067	H	-3.24966	-1.148698	-0.882278
H	-2.43838	-2.496585	0.000085	H	-2.446777	-2.482984	-0.000038
H	2.133256	-2.32484	0.00006	O	3.332307	-0.846248	0.000599
O	3.313874	-0.684112	0.000413				

Table B.20: Coordinates of G4 optimised model 5fChyd and following H-abstraction (5fChyd ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5fChyd				5fChyd ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.9394	0.70226	0.03384	C	-1.977838	0.613272	0.047164
N	-1.6901	-0.7104	-0.0393	N	-1.584241	-0.746742	-0.05039
C	-0.4319	-1.1979	-0.0687	C	-0.269721	-1.113061	-0.108728
C	0.65708	-0.3832	-0.0601	C	0.761631	-0.165125	-0.049155
C	0.36768	1.02959	-0.0215	C	0.290152	1.225275	-0.001919
N	-0.845	1.52555	0.02459	N	-0.973435	1.55172	0.041333
H	-0.3352	-2.2794	-0.0891	H	-0.077144	-2.167392	-0.258856
O	-3.0943	1.07377	0.08409	O	-3.159693	0.898184	0.120723
N	1.42273	1.90521	0.0178	N	1.187911	2.235791	0.022315
H	1.16012	2.875	-0.0848	H	0.814401	3.169846	-0.006291
H	2.26457	1.60377	-0.4576	H	2.167342	2.066947	-0.12198
C	2.06596	-0.9194	0.00885	C	2.094826	-0.554226	0.008169
H	2.05504	-2.0134	-0.0917	C	-2.61479	-1.774849	-0.095771
C	-2.8616	-1.5786	-0.0531	H	-2.551473	-2.429084	0.78128
H	-3.4596	-1.4135	0.84579	H	-3.576851	-1.266469	-0.10091
H	-3.4901	-1.3468	-0.9159	H	-2.509169	-2.383913	-1.000166
H	-2.5356	-2.6193	-0.0987	O	3.140806	0.289995	-0.187427
O	2.88835	-0.3503	-1.0219	H	3.923807	-0.120207	0.202228
H	2.4645	-0.5293	-1.8698	O	2.496338	-1.844786	0.173044
O	2.70041	-0.6616	1.2206	H	1.872277	-2.289973	0.762611
H	2.56016	0.26862	1.44102				

Table B.21: Coordinates of G4 optimised model 5fU and following H-abstraction (5fU ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5fU				5fU ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-2.543105	-1.582439	0.000083	C	-2.542298	-1.559547	-0.000012
N	-1.363325	-0.717675	0.000271	N	-1.345635	-0.717132	0.00019
C	-0.099075	-1.206879	0.000047	C	-0.094801	-1.232859	-0.000019
H	-0.022578	-2.290522	-0.00013	H	-0.023779	-2.315177	-0.000139
C	1.028836	-0.444746	-0.000006	C	1.044694	-0.48427	-0.000084
C	0.893902	1.018338	0.000037	C	0.938568	0.983738	0.000093
O	1.772976	1.844426	0.00134	O	1.840279	1.784896	0.000385
N	-0.460077	1.431515	-0.001858	N	-0.405529	1.42196	-0.000834
H	-0.60748	2.432637	-0.001147	H	-0.539905	2.425282	-0.000407
C	-1.604912	0.670021	0.000088	C	-1.56241	0.676811	-0.000196
O	-2.728459	1.118547	0.000383	O	-2.677376	1.144132	0.000427
H	-2.213904	-2.621642	-0.000317	H	-2.231992	-2.604355	0.001598
H	-3.152286	-1.387223	0.884822	H	-3.147944	-1.35064	0.883889
H	-3.152502	-1.386468	-0.884329	H	-3.146305	-1.352862	-0.885586
C	2.32508	-1.141179	0.0001	C	2.316915	-1.198064	-0.000062
H	2.207515	-2.252951	0.000815	O	3.430105	-0.785389	0.000043
O	3.418069	-0.634147	-0.00056				

Table B.22: Coordinates of G4 optimised model 5fUhyd and following H-abstraction (5fUhyd ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5fUhyd				5fUhyd ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	2.539857	-1.847972	0.100933	C	2.634751	-1.786388	0.000454
N	1.534205	-0.79136	0.01113	N	1.593235	-0.766102	-0.00008
C	0.194065	-1.069899	-0.087326	C	0.260883	-1.171092	-0.000399
H	-0.069828	-2.119546	-0.099143	H	0.085233	-2.234592	0.003384
C	-0.759717	-0.119281	-0.166888	C	-0.78576	-0.219966	-0.000258
C	-0.369061	1.277812	-0.127681	C	-0.439904	1.190742	-0.000014
O	-1.136118	2.23216	-0.153182	O	-1.261315	2.11946	0.000478
N	1.009851	1.476588	-0.04912	N	0.906242	1.462371	-0.000046
H	1.328682	2.436034	-0.016819	H	1.179556	2.435219	-0.000161
C	2.008124	0.519567	0.030308	C	1.977013	0.553633	-0.000183
O	3.186949	0.789148	0.111032	O	3.126712	0.950631	-0.000383
H	2.033941	-2.812738	0.06424	H	2.545429	-2.416476	0.891346
H	3.246803	-1.769601	-0.727961	H	2.546323	-2.416708	-0.890355
H	3.099804	-1.760822	1.03477	H	3.599235	-1.282946	0.000802
C	-2.23623	-0.422918	-0.296747	C	-2.140166	-0.520855	-0.000159
O	-3.000455	0.278792	0.654137	O	-3.105394	0.385332	0.000247
H	-2.791323	1.217363	0.518672	H	-2.63092	1.2747	0.000289
H	-2.569485	-0.135286	-1.312892	O	-2.552869	-1.795456	-0.000427
O	-2.422357	-1.792448	-0.095385	H	-3.519161	-1.78926	-0.000383
H	-3.373355	-1.937059	-0.043348				

Table B.23: Coordinates of G4 optimised model 5hmC and following H-abstraction (5hmC ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5hmC				5hmC ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	1.647418	0.743526	0.041404	C	2.618281	-1.577183	-0.000735
N	1.469508	-0.682909	0.005477	N	1.453917	-0.699229	-0.000011
C	0.241597	-1.230483	-0.115003	C	0.188197	-1.251298	0.000148
C	-0.885364	-0.471963	-0.191701	H	0.165169	-2.33357	-0.000026
C	-0.670138	0.954147	-0.106081	C	-0.987346	-0.498736	0.000273
N	0.516991	1.509077	0.001639	C	-0.846036	0.977308	-0.000248
H	0.202215	-2.314988	-0.162874	O	-1.744163	1.789398	-0.001417
O	2.783382	1.168668	0.120941	N	0.484972	1.406312	0.001304
N	-1.755229	1.772828	-0.163498	H	0.61479	2.408977	0.000703
H	-1.577332	2.74624	0.029468	C	1.66106	0.670004	0.000514
H	-2.646992	1.383775	0.107795	O	2.757334	1.196553	0.000074
C	-2.245732	-1.065322	-0.411991	H	2.617415	-2.212554	0.890734
H	-2.173668	-2.162228	-0.441375	H	2.616216	-2.212689	-0.892134
C	2.678257	-1.493559	0.081034	H	3.506353	-0.948316	-0.001378
H	3.351418	-1.240041	-0.740906	C	-2.233043	-1.119883	0.000209
H	3.20842	-1.293196	1.015058	H	-2.292565	-2.205046	0.001044
H	2.405537	-2.549403	0.027993	O	-3.397979	-0.434347	0.000071
H	-2.665791	-0.739596	-1.370475	H	-4.127815	-1.060485	0.001209
O	-3.215062	-0.641651	0.556052				
H	-2.865492	-0.857745	1.428073				

Table B.24: Coordinates of G4 optimised model 5hmU and following H-abstraction (5hmU ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5hmU				5hmU ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	2.100719	-1.973059	-0.000101	C	2.618281	-1.577183	-0.000735
N	1.179939	-0.839645	0.000045	N	1.453917	-0.699229	-0.000011
C	-0.184654	-1.01181	0.000008	C	0.188197	-1.251298	0.000148
H	-0.529651	-2.037481	0.000021	H	0.165169	-2.33357	-0.000026
C	-1.063837	0.010784	-0.000028	C	-0.987346	-0.498736	0.000273
C	-0.56493	1.377332	-0.000007	C	-0.846036	0.977308	-0.000248
O	-1.251282	2.3829	-0.000086	O	-1.744163	1.789398	-0.001417
N	0.834815	1.45964	0.000137	N	0.484972	1.406312	0.001304
H	1.229559	2.390627	0.000104	H	0.61479	2.408977	0.000703
C	1.75618	0.428125	0.000043	C	1.66106	0.670004	0.000514
O	2.957334	0.601276	0.000002	O	2.757334	1.196553	0.000074
H	1.518191	-2.894486	-0.000954	H	2.617415	-2.212554	0.890734
H	2.742406	-1.942366	-0.883783	H	2.616216	-2.212689	-0.892134
H	2.741484	-1.943467	0.884304	H	3.506353	-0.948316	-0.001378
C	-2.55217	-0.185709	-0.000097	C	-2.233043	-1.119883	0.000209
H	-2.981991	0.316959	0.88036	H	-2.292565	-2.205046	0.001044
H	-2.981908	0.316625	-0.88079	O	-3.397979	-0.434347	0.000071
O	-2.83151	-1.575656	0.000162	H	-4.127815	-1.060485	0.001209
H	-3.785554	-1.688503	-0.000064				

Table B.25: Coordinates of G4 optimised model 5meC and following H-abstraction (5meC ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	5meC				5meC ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	1.213611	-0.846831	-0.000722	C	1.120199	-0.864852	-0.024574
N	1.231705	0.578548	0.000055	N	1.208227	0.557507	0.023753
C	0.087413	1.298431	0.000037	C	0.103983	1.33739	0.021004
C	-1.139182	0.701374	-0.000682	C	-1.182777	0.805965	-0.011886
C	-1.118278	-0.736551	-0.001469	C	-1.205577	-0.659498	0.005401
N	-0.028479	-1.447807	-0.00165	N	-0.136566	-1.41089	-0.005301
H	0.20503	2.377581	0.001185	H	0.266724	2.408921	0.03108
O	2.271823	-1.442094	-0.000096	O	2.153851	-1.505339	-0.06132
N	-2.348494	-1.479664	-0.001674	N	-2.409154	-1.2953	0.007996
H	-2.895705	-1.235379	0.821239	H	-2.367746	-2.290473	0.165808
H	-2.91864	-1.195554	-0.79569	H	-3.22177	-0.812662	0.350297
C	-2.411753	1.502828	0.000287	C	-2.328499	1.603182	-0.059717
H	-3.030561	1.297453	-0.881445	H	-2.247127	2.683193	-0.041743
H	-3.033044	1.290761	0.878639	H	-3.320203	1.192191	-0.186269
H	-2.194946	2.575049	0.00445	C	2.550296	1.128074	0.043268
C	2.545969	1.210976	0.001089	H	3.10491	0.748254	0.903941
H	3.109557	0.895628	0.882114	H	3.096134	0.837258	-0.857247
H	3.110287	0.897036	-0.87997	H	2.47498	2.215249	0.09858
H	2.423641	2.295279	0.00189				

Table B.28: Coordinates of G4 optimised model g-Rme1 and following H-abstraction (g-Rme1 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	g-Rme1				g-Rme1 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.619821	-0.000062	0.000091	N	-0.654927	0.128816	0.004687
C	0.710685	0.000002	0.000014	C	0.682786	-0.043975	-0.001935
C	-1.387513	-1.241897	-0.15281	C	-1.549602	-1.032966	-0.164477
H	-2.318647	-1.009351	-0.671224	H	-2.490338	-0.677623	-0.583211
H	-0.835032	-1.958443	-0.762756	H	-1.108196	-1.748126	-0.859428
H	-1.629861	-1.68416	0.819591	H	-1.753063	-1.509292	0.799271
N	1.398062	-1.145212	0.124644	N	1.210867	-1.266965	0.146137
H	0.971993	-1.975288	0.500073	H	0.687804	-2.020741	0.559646
N	1.397764	1.14543	-0.12448	N	1.487613	1.016439	-0.150582
H	0.971718	1.975048	-0.500967	H	1.14369	1.852373	-0.597762
H	2.388922	1.173577	0.049122	H	2.483669	0.931162	-0.028228
H	2.389096	-1.173454	-0.049654	H	2.191634	-1.42593	-0.019632
C	-1.387598	1.241802	0.152773	C	-1.221834	1.380272	0.175956
H	-0.834887	1.95852	0.762295	H	-2.284946	1.469271	0.025018
H	-1.630247	1.683705	-0.819696	H	-0.643224	2.160886	0.645375
H	-2.318527	1.009311	0.671574				

Table B.29: Coordinates of G4 optimised model Kme1 and following H-abstraction (Kme1 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	Kme1				Kme1 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.0000050	0.5417930	0.0000000	N	0.0715650	0.5317900	0.0000000
C	1.2651980	-0.2764280	0.0000000	C	1.3013520	-0.2539770	0.0000000
H	2.1198980	0.4002870	0.0000130	H	1.6477840	-0.6152610	0.9573610
H	1.2705110	-0.8984430	0.8949830	H	1.6477870	-0.6152570	-0.9573610
H	1.2705240	-0.8984220	-0.8949970	C	-1.2087800	-0.3012940	0.0000000
C	-1.2651920	-0.2764270	0.0000000	H	-1.2045680	-0.9202180	0.8962660
H	-1.2705260	-0.8984210	0.8949980	H	-2.0576140	0.3837470	0.0000050
H	-2.1199060	0.4002730	-0.0000200	H	-1.2045710	-0.9202090	-0.8962730
H	-1.2705060	-0.8984510	-0.8949770	H	0.0573980	1.1481460	0.8192580
H	0.0000010	1.1588780	0.8171930	H	0.0573980	1.1481480	-0.8192550
H	0.0000020	1.1588790	-0.8171920				

Table B.26: Coordinates of G4 optimised model a-Rme2 and following H-abstraction (a-Rme2 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	a-Rme2				a-Rme2 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.507699	1.565556	-0.120009	N	-0.413372	1.547464	-0.172045
H	0.173218	2.221757	-0.460652	H	0.290654	2.155213	-0.554838
C	-0.218689	0.26142	0.006909	C	-0.191119	0.235882	-0.005183
N	-1.222277	-0.617215	0.131787	N	-1.215028	-0.607491	0.152935
H	-0.993464	-1.53439	0.47785	H	-0.992853	-1.505198	0.558929
N	1.052695	-0.151624	0.002097	N	1.076045	-0.24313	-0.005542
C	-2.639554	-0.28605	-0.000902	C	-2.623612	-0.239076	0.034949
H	-2.814463	0.274092	-0.92325	H	-2.791609	0.313971	-0.892868
H	-3.201658	-1.217776	-0.061195	H	-3.212568	-1.155161	-0.006108
H	-3.011581	0.283451	0.858325	H	-2.967113	0.353489	0.890594
C	2.163722	0.795678	0.142968	C	2.212852	0.67545	0.187282
H	2.998434	0.280975	0.620091	H	3.039662	0.107248	0.612333
H	2.498474	1.173964	-0.829563	H	1.942967	1.468219	0.886119
H	1.877026	1.62867	0.787904	C	1.334331	-1.592279	-0.184063
C	1.391436	-1.570582	-0.147643	H	2.341846	-1.931974	-0.008491
H	2.37643	-1.643503	-0.609691	H	0.614186	-2.19974	-0.711355
H	1.423478	-2.084305	0.820035	H	-1.331413	1.934828	-0.032507
H	0.677348	-2.066189	-0.808688	H	2.538016	1.101342	-0.767149
H	-1.443775	1.903433	0.023711				

Table B.27: Coordinates of G4 optimised model 6meA and following H-abstraction (6meA ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	6meA				6meA ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	1.9619	-0.561718	-0.001284	N	1.955117	-0.509363	-0.001056
C	1.316809	-1.779184	-0.000995	C	1.364326	-1.753057	-0.000663
H	1.879854	-2.702312	-0.001754	H	1.964777	-2.651974	-0.001229
N	0.012128	-1.691453	-0.000509	N	0.054797	-1.719864	-0.000341
C	-0.234115	-0.326731	-0.000165	C	-0.249018	-0.370712	0.00002
C	-1.406741	0.459002	-0.000064	C	-1.448552	0.370096	0.000147
N	-2.666332	-0.028707	0.000184	N	-2.701394	-0.166023	0.000373
H	-3.382371	0.679401	0.000496	H	-3.423027	0.542718	0.000525
N	-1.279433	1.80179	-0.000071	N	-1.385407	1.717596	0.000057
C	-0.056129	2.33667	-0.000362	C	-0.188391	2.304465	-0.000335
H	-0.023434	3.423907	-0.00066	H	-0.199911	3.391721	-0.000704
N	1.125905	1.716648	-0.000386	N	1.020554	1.731164	-0.000442
C	0.970107	0.390734	-0.000306	C	0.923532	0.399836	-0.000257
C	3.388519	-0.306886	0.002045	C	3.370042	-0.193618	0.001378
H	3.669629	0.295441	-0.865576	H	3.624195	0.419159	-0.867112
H	3.680322	0.234078	0.906234	H	3.63813	0.359996	0.905105
H	3.918591	-1.260317	-0.033775	H	3.940066	-1.123624	-0.033795
C	-3.040984	-1.429282	0.000893	C	-3.062835	-1.485222	0.000741
H	-2.657291	-1.951473	-0.880887	H	-2.287813	-2.233129	0.000322
H	-2.658354	-1.950407	0.883801	H	-4.116714	-1.71017	0.000567
H	-4.13092	-1.490184	0.000303				

Table B.30: Coordinates of G4 optimised model Kme2 and following H-abstraction (Kme2 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	Kme2				Kme2 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	0.0000750	-0.0000150	-0.3395220	N	-0.0118630	0.0610230	-0.3369980
C	1.1917590	-0.8032810	0.1033460	C	-1.0735640	-0.9212570	0.1307720
H	1.0970880	-1.8179050	-0.2840230	H	-2.0433780	-0.5792290	-0.2298720
H	2.0967310	-0.3340840	-0.2831670	H	-0.8346280	-1.9048550	-0.2765440
H	1.2102660	-0.8164590	1.1937330	H	-1.0568970	-0.9408310	1.2203440
C	-1.2915730	-0.6304050	0.1032350	C	-0.2998000	1.4205070	0.1048410
H	-2.1230570	-0.0403340	-0.2827350	H	-0.0434020	1.6594300	1.1273640
H	-1.3116940	-0.6412340	1.1936460	C	1.3723980	-0.3808900	0.0762910
C	0.0997920	1.4337540	0.1032990	H	1.5523190	-1.3784240	-0.3251700
H	1.0275630	1.8579070	-0.2811560	H	2.0953270	0.3318330	-0.3197900
H	-0.7573420	1.9835820	-0.2859410	H	1.4116680	-0.3976530	1.1658750
H	0.0984020	1.4565270	1.1937120	H	-0.9548760	2.0163510	-0.5144040
H	-1.3384780	-1.6483240	-0.2842360	H	-0.0373010	0.0560580	-1.3602380
H	0.0001270	0.0000210	-1.3624580				

Table B.31: Coordinates of G4 optimised model KMe3 and following H-abstraction (KMe3 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	KMe3				KMe3 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.0001570	0.0000000	-0.0000950	N	0.0000010	-0.0009330	0.0498720
C	0.4552960	0.0000000	1.4348500	C	-1.2336920	0.7992740	-0.2966710
H	0.0701600	-0.8935470	1.9268710	H	-1.2263140	1.7187320	0.2883220
H	1.5453790	0.0000000	1.4582640	H	-1.2143090	1.0241040	-1.3630990
H	0.0701600	0.8935460	1.9268710	H	-2.1148680	0.2048640	-0.0542450
C	-1.5048630	0.0000000	-0.0492400	C	-0.0000320	-0.3078140	1.4776490
H	-1.8234830	0.0000000	-1.0919810	H	0.9530850	-0.5450540	1.9284950
H	-1.8744260	0.8933930	0.4549180	H	-0.9532020	-0.5448330	1.9284990
C	0.5248070	1.2292420	-0.6927560	C	-0.0000680	-1.2951360	-0.7453980
H	1.6143640	1.2188650	-0.6513740	H	-0.0001890	-1.0476900	-1.8084520
H	0.1898590	1.2199790	-1.7303440	H	0.8941860	-1.8618010	-0.4867460
H	0.1395070	2.1128220	-0.1831750	H	-0.8942500	-1.8618230	-0.4865490
C	0.5248070	-1.2292420	-0.6927560	C	1.2337900	0.7991440	-0.2966460
H	0.1898590	-1.2199780	-1.7303440	H	2.1149000	0.2046690	-0.0541380
H	1.6143640	-1.2188650	-0.6513750	H	1.2144900	1.0239150	-1.3630880
H	0.1395060	-2.1128220	-0.1831760	H	1.2264750	1.7186410	0.2882900
H	-1.8744260	-0.8933930	0.4549180				

Table B.32: Coordinates of G4 optimised model nKme1 and following H-abstraction (nKme1 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	nKme1				nKme1 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.0000030	0.5638540	-0.1490930	N	-0.0934240	0.4717000	-0.1379940
H	-0.0000760	1.3289950	0.5164730	H	-0.1124060	1.4120640	0.2315750
C	-1.2128990	-0.2219040	0.0203400	C	-1.2548480	-0.2533030	0.0964260
H	-2.0897620	0.4278410	-0.0584740	H	-2.1839560	0.3020950	0.0336030
H	-1.2786050	-0.9674530	-0.7810400	H	-1.2503720	-1.2793900	-0.2581190
H	-1.2723470	-0.7669780	0.9810600	C	1.1886380	-0.1861040	0.0330350
C	1.2129250	-0.2218990	0.0203480	H	1.2160280	-1.0852950	-0.5918100
H	1.2779480	-0.9682470	-0.7803270	H	1.9920350	0.4780870	-0.2976230
H	2.0897850	0.4276870	-0.0597480	H	1.3898920	-0.4930200	1.0715640
H	1.2729140	-0.7660060	0.9815770				

Table B.33: Coordinates of G4 optimised model nKme2 and following H-abstraction (nKme2 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	nKme2				nKme2 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.000109	0.000141	-0.366497	N	0	0.08092	-0.212039
C	-1.326747	0.40903	0.059767	C	1.220797	-0.660292	0.035307
H	-2.074604	-0.287095	-0.33358	H	1.370751	-0.888415	1.104214
H	-1.441277	0.442663	1.161722	H	1.199646	-1.6057	-0.516467
H	-1.552147	1.406487	-0.331252	H	2.07879	-0.077663	-0.311108
C	1.017725	0.944232	0.05972	C	0.000001	1.429942	0.106688
H	1.107217	1.023226	1.161675	H	-0.929949	1.960475	-0.066409
H	1.993391	0.642367	-0.334771	H	0.929951	1.960474	-0.06641
H	0.787004	1.940921	-0.330078	C	-1.220798	-0.66029	0.035307
C	0.309075	-1.353326	0.059657	H	-2.07879	-0.077661	-0.311108
H	1.286298	-1.652605	-0.332837	H	-1.199648	-1.605698	-0.516467
H	0.33636	-1.469348	1.161634	H	-1.370752	-0.888414	1.104214
H	-0.441795	-2.047219	-0.331903				

Table B.34: Coordinates of G4 optimised model Asn and following H-abstraction (Asn ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	Asn				Asn ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.976471	0.039939	0.002885	C	2.013578	-0.0328	-0.009857
H	-2.036362	0.677266	0.888233	H	1.898563	1.052402	-0.029678
C	-0.675518	-0.758627	-0.00361	C	0.679775	-0.679796	0.011624
H	-0.621734	-1.424431	0.867729	H	0.608312	-1.764214	0.04144
H	-0.624289	-1.413531	-0.883359	C	-0.541145	0.118091	0.004739
C	0.561549	0.140271	-0.000692	O	-0.533733	1.344829	-0.003006
O	0.504043	1.354162	-0.001491	N	-1.72283	-0.596937	0.047662
N	1.751197	-0.532843	0.00225	H	-2.561722	-0.068398	-0.132726
H	2.604601	0.000375	0.003401	H	-1.736545	-1.568849	-0.212972
H	1.804897	-1.53579	0.003214	H	2.599052	-0.349025	-0.884769
H	-2.840709	-0.630364	-0.004003	H	2.608767	-0.314967	0.870092
H	-2.034488	0.693576	-0.870537				

Table B.35: Coordinates of G4 optimised model Pro and following H-abstraction (Pro ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	Pro				Pro ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	-0.008676	-1.183215	0.110374	N	-0.09831	-1.167401	0.056234
C	-1.421539	-0.830856	-0.088031	C	-1.509279	-0.740065	-0.04369
C	0.738899	-0.013523	-0.378018	C	0.701025	-0.007463	-0.386618
H	-2.057577	-1.402948	0.592363	H	-2.122463	-1.255864	0.707383
H	-1.707982	-1.111301	-1.109158	H	-1.915889	-1.048164	-1.023364
C	-1.551912	0.7128	0.094414	C	-1.496269	0.754956	0.098305
C	-0.08857	1.222722	0.055934	C	-0.088889	1.254204	0.062309
H	0.725697	-0.069796	-1.476835	H	0.705792	-0.02519	-1.486003
C	2.185934	-0.016978	0.094036	C	2.136895	-0.085033	0.111437
H	-2.032471	0.963664	1.044347	H	-2.379983	1.378914	0.143182
H	-2.163409	1.155661	-0.697049	H	0.257981	1.58578	1.058006
H	0.235637	1.543077	1.053384	H	0.058436	2.111231	-0.608423
H	0.048274	2.075987	-0.615094	H	2.172394	-0.052726	1.20773
H	2.238811	0.04257	1.188561	H	2.614457	-1.012311	-0.218063
H	2.698703	-0.929693	-0.222676	H	2.726074	0.757215	-0.263825
H	2.730719	0.841772	-0.309918	H	0.110464	-1.326673	1.039286
H	0.167455	-1.271484	1.109445				

Table B.36: Coordinates of G4 optimised model s-Rme2 and following H-abstraction (s-Rme2 ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	s-Rme2				s-Rme2 ·		
Atom	X	Y	Z	Atom	X	Y	Z
N	1.250362	0.475626	0.029007	N	0.51226	-1.187424	0.064332
H	2.000074	-0.186842	0.136215	H	-0.110749	-1.954544	0.261094
C	-0.000089	0.000067	0.036647	C	-0.008023	0.043576	0.097425
C	1.599995	1.891894	-0.038305	C	1.929839	-1.496734	-0.114743
H	2.681925	1.969703	-0.139532	H	2.018776	-2.554507	-0.361669
H	1.14396	2.363667	-0.914026	H	2.342316	-0.917193	-0.944699
H	1.303977	2.426745	0.871013	H	2.508884	-1.309019	0.796645
N	-1.037278	0.84502	0.029221	N	0.81375	1.118257	0.110022
H	-0.838671	1.825812	0.134041	H	1.770211	0.95088	0.388897
N	-0.213219	-1.320686	0.029793	N	-1.330524	0.236591	0.078325
H	-1.16192	-1.639045	0.134633	H	-1.653514	1.142718	0.385143
C	-2.438543	0.439388	-0.03846	C	0.431991	2.429942	-0.112637
H	-2.75248	-0.087678	0.869364	H	1.1681	3.193736	0.07887
H	-3.047303	1.337637	-0.135479	H	-0.448844	2.62152	-0.707013
H	-2.619986	-0.188511	-0.91614	C	-2.32069	-0.824818	-0.081444
C	0.838763	-2.331305	-0.038664	H	-3.288552	-0.362078	-0.27332
H	0.365258	-3.307735	-0.134742	H	-2.065493	-1.449094	-0.941715
H	1.47209	-2.174689	-0.917216	H	-2.408235	-1.446186	0.817417
H	1.453269	-2.339056	0.868422				

Table B.37: Coordinates of G4 optimised model Thy and following H-abstraction (Thy ·). G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	Thy				Thy ·		
Atom	X	Y	Z	Atom	X	Y	Z
C	-2.437268	1.361429	-0.006945	C	2.46211	1.279355	0.000288
N	-1.172265	0.632873	0.012366	N	1.172559	0.593202	0.000194
C	0.041611	1.284135	0.004602	C	0.01071	1.329076	-0.000085
H	-0.02988	2.366279	0.005545	H	0.149807	2.403009	-0.000019
C	1.238539	0.660479	-0.000305	C	-1.261317	0.767453	-0.000111
C	1.25719	-0.800871	-0.000672	C	-1.34507	-0.716242	-0.000032
O	2.254694	-1.495854	-0.003667	O	-2.370798	-1.36121	0.000873
N	-0.019779	-1.382823	0.00466	N	-0.103652	-1.351726	-0.001008
H	-0.049663	-2.393666	0.001014	H	-0.123594	-2.362918	-0.000522
C	-1.251923	-0.757294	0.002858	C	1.167943	-0.797981	-0.000184
O	-2.308968	-1.353561	-0.004571	O	2.173608	-1.478234	0.000151
H	-2.238229	2.415385	0.190389	H	2.558709	1.905321	-0.891884
H	-2.932648	1.259406	-0.976922	H	2.557848	1.906645	0.891615
H	-3.104429	0.961856	0.758214	H	3.241309	0.520042	0.001156
C	2.554886	1.379486	-0.002939	C	-2.443964	1.490841	-0.000257
H	3.152112	1.103592	0.872327	H	-3.388021	0.964242	-0.000108
H	3.148151	1.104218	-0.881098	H	-2.443361	2.573865	-0.000443
H	2.414872	2.463709	-0.002347				

Table B.38: Coordinates of G4 optimised iron Fe=O and Fe-[OH] active sites. G4 uses UB3LYP/6-31G(2df,p) equilibrium structures.

	[Fe=O]				[Fe-OH]		
Atom	X	Y	Z	Atom	X	Y	Z
C	3.761474	0.009201	0.628263	C	3.600472	-1.277462	0.611921
N	3.356518	0.071142	1.944831	N	2.933574	-2.443698	0.916871
H	3.957835	0.180837	2.741603	H	3.359187	-3.320182	1.159579
C	2.014641	-0.06423	1.981014	C	1.606187	-2.205011	0.835993
H	1.422936	-0.069266	2.882384	H	0.829277	-2.929044	1.027821
N	1.542508	-0.200122	0.772357	N	1.386583	-0.963744	0.498549
C	2.613399	-0.160022	-0.080425	C	2.614968	-0.374966	0.354801
H	2.487651	-0.27887	-1.144254	H	2.720334	0.664437	0.083858
C	0.283778	-3.919363	-2.089362	C	-1.39198	-3.114138	-2.850053
H	0.884947	-3.43815	-2.860184	H	-0.577031	-2.84788	-3.52206
H	-0.304891	-4.727507	-2.516163	H	-2.251999	-3.465474	-3.414561
C	-0.605987	-2.912757	-1.405267	C	-1.768076	-1.94799	-1.971169
O	-1.787528	-3.163477	-1.206013	O	-2.946079	-1.686165	-1.766578
O	-0.00512	-1.817214	-1.085667	O	-0.756169	-1.30185	-1.501116
C	-0.241839	2.905476	-2.08706	C	0.761168	3.307134	-1.38853
N	-0.703192	1.95697	-2.976096	C	-1.195596	1.688962	3.879884
H	-0.953044	2.118701	-3.935203	H	-2.002933	1.697266	4.607173
C	-0.792003	0.777963	-2.329843	H	-0.285551	1.302727	4.338966
H	-1.143222	-0.13813	-2.775753	H	-0.981626	2.70806	3.555729
N	-0.413475	0.915787	-1.087733	O	-1.278068	-1.785078	1.053341
C	-0.070863	2.23241	-0.919241	O	-2.645689	0.270754	2.630609
H	0.256388	2.6044	0.03779	H	-1.032431	-3.922288	-2.211006
Fe	-0.541462	-0.605926	0.377161	C	1.17611	4.724131	-1.445531
C	-2.101712	0.932055	2.301944	H	2.22172	4.83317	-1.741244
O	-2.561598	-0.790559	-0.261051	H	0.570998	5.297151	-2.15123
O	-1.024169	0.869308	1.597597	H	1.059686	5.180992	-0.46472
C	-2.020858	1.918978	3.441126	C	5.074496	-1.169108	0.608812
H	-2.991565	2.039561	3.91529	H	5.5027	-1.387046	1.589693
H	-1.303463	1.554351	4.178006	H	5.533793	-1.851577	-0.109821
H	-1.654649	2.881103	3.083441	H	5.369801	-0.156969	0.338122
O	-0.632769	-1.745943	1.516035	H	-3.101762	0.322255	0.718525
O	-3.123444	0.286033	2.11425	H	-3.071043	-0.303005	-0.728066
H	0.975072	-4.330082	-1.352117	H	-1.815595	-1.503314	1.802407
C	-0.032096	4.320626	-2.45711	N	0.811375	2.46533	-2.480824
H	0.718982	4.432731	-3.242278	H	1.114304	2.716133	-3.404874
H	-0.953064	4.789111	-2.810973	C	0.356119	1.25428	-2.107733
H	0.310109	4.881109	-1.589191	H	0.273147	0.390204	-2.746711
C	5.177217	0.109167	0.21629	N	0.021653	1.274899	-0.843188
H	5.620383	1.065513	0.502682	C	0.266725	2.542301	-0.379959
H	5.786662	-0.683035	0.656592	H	0.068447	2.800438	0.646902
H	5.253186	0.020117	-0.865757	Fe	-0.716386	-0.312995	0.184291
H	-3.018685	-0.609925	0.588657	C	-1.560712	0.841369	2.687275
H	-2.583943	-1.747122	-0.477447	O	-2.749775	0.446623	-0.176811
				O	-0.647842	0.801378	1.783035

C

Selectivity by Epigenetic Readers - Appendix

C.1 Ligand partial charges and XYZ coordinates of L-KMe3 and D-KMe3

L-KMe3

Atom	X	Y	Z	Charge
N	-3.252	1.483	0.199	-0.9943
C	-3.165	0.204	-0.472	0.4440
C	-1.866	-0.521	-0.080	-0.2969
C	-0.621	0.257	-0.504	0.1166
C	0.672	-0.381	0.025	0.0526
C	1.868	0.461	-0.408	-0.2183
N	3.234	-0.002	0.067	0.2357
C	-4.344	-0.715	-0.183	0.3660
O	-5.155	-0.486	0.652	-0.4634
C	3.545	-1.375	-0.436	-0.3779
C	4.253	0.948	-0.477	-0.3779
C	3.320	0.011	1.560	-0.3779
H	-3.594	1.349	1.132	0.3760
H	-3.923	2.072	-0.255	0.3760
H	-3.147	0.375	-1.547	-0.0071
H	-1.850	-1.516	-0.521	0.0726
H	-1.864	-0.657	1.000	0.0726
H	-0.581	0.310	-1.590	0.0062
H	-0.709	1.273	-0.139	0.0062
H	0.756	-1.394	-0.353	0.0291
H	0.615	-0.439	1.107	0.0291
H	1931	0.496	-1.487	0.1332
H	1.761	1.474	-0.046	0.1332
H	-4.395	-1.630	-0.781	0.0133
H	3.443	-1.388	-1.511	0.1835
H	4.559	-1.623	-0.162	0.1835
H	2.868	-2.086	0.008	0.1835
H	4.208	0.936	-1.555	0.1835
H	5.235	0.641	-0.149	0.1835
H	4.041	1.942	-0.112	0.1835
H	3.055	0.995	1.919	0.1835
H	2.648	-0.724	1.969	0.1835
H	4.333	-0.224	1.851	0.1835

D-KMe3

Atom	X	Y	Z	Charge
N	-3.414	-0.900	1.164	-1.0164
C	-3.321	-0.605	-0.250	0.4854
C	-3.718	0.831	-0.528	0.3440
O	-4.108	1.587	0.299	-0.4506
H	-4.013	-0.243	1.624	0.3815
C	-1.898	-0.871	-0.778	-0.2915
H	-1.692	-1.930	-0.647	0.0611
H	-1.862	-0.677	-1.848	0.0611
C	-0.833	-0.049	-0.048	0.1206
H	-0.969	-0.187	1.018	0.0133
H	-0.978	1.010	-0.251	0.0133
C	0.594	-0.447	-0.451	0.0194
H	0.736	-1.502	-0.244	0.0386
H	0.724	-0.301	-1.518	0.0386
C	1.590	0.399	0.338	-0.2223
H	1.448	1.448	0.119	0.1366
H	1.440	0.256	1.400	0.1366
N	3.063	0.124	0.091	0.2428
C	3.427	0.371	-1.337	-0.3832
H	3.132	1.373	-1.607	0.1845
H	4.496	0.263	-1.446	0.1845
H	2.927	-0.344	-1.967	0.1845
C	3.852	1.059	0.954	-0.3832
H	4.905	0.881	0.797	0.1845
H	3.609	2.076	0.687	0.1845
H	3.601	0.882	1.988	0.1845
C	3.420	-1.280	0.460	-0.3832
H	3.118	-1.464	1.480	0.1845
H	2921	-1.967	-0.202	0.1845
H	4.489	-1.402	0.367	0.1845
H	-4.002	-1.209	-0.852	-0.0316
H	-3.625	1.142	-1.573	0.0272
H	-3.771	-1.820	1.325	0.3815

C.2 Chain length selectivity ligand partial charges

hMe3 Atom	X	Y	Z	Charge
N	3.844	1.498	-0.189	-0.993558
C	3.798	0.199	0.448	0.419383
C	2.550	-0.580	0.001	-0.233717
C	1.250	0.113	0.411	0.124829
C	0.009	-0.596	-0.138	-0.048834
C	-1.286	0.134	0.250	0.056981
C	5.031	-0.652	0.181	0.368966
O	5.874	-0.353	-0.599	-0.469423
H	4.176	1.397	-1.130	0.374567
H	4.511	2.088	0.271	0.374567
H	3.739	0.344	1.525	-0.002661
H	2.578	-1.586	0.416	0.058341
H	2.583	-0.688	-1.081	0.058341
H	1.191	0.154	1.496	-0.010380
H	1.277	1.136	0.058	-0.010380
H	-0.027	-1.619	0.230	0.028014
H	0.077	-0.655	-1.221	0.028014
H	-1.355	0.178	1.331	0.024116
H	-1.233	1.153	-0.116	0.024116
H	5.095	-1.588	0.745	0.008756
C	-2.490	-0.593	-0.341	-0.134217
H	-2.542	-1.604	0.038	0.111937
H	-2.407	-0.643	-1.418	0.111937
N	-3.853	0.019	-0.065	0.194742
C	-4.883	-0.844	-0.722	-0.376156
H	-4.826	-1.840	-0.309	0.184893
H	-5.862	-0.428	-0.540	0.184893
H	-4.691	-0.877	-1.784	0.184893
C	-3.952	1.397	-0.636	-0.376156
H	-3.272	2.054	-0.120	0.184893
H	-3.706	1.362	-1.686	0.184893
H	-4.963	1.753	-0.509	0.184893
C	-4.137	0.067	1.402	-0.376156
H	-5.149	0.412	1.548	0.184893
H	-4.025	-0.924	1.815	0.184893
H	-3.453	0.747	1.881	0.184893

OrnMe3				
Atom	X	Y	Z	Charge
N	-2.484	1.459	0.074	-0.996696
C	-2.467	0.129	-0.495	0.394184
C	-1.242	-0.652	0.013	-0.349661
C	0.068	0.042	-0.378	0.305879
C	1.251	-0.669	0.272	-0.196384
C	-3.720	-0.680	-0.184	0.393279
O	-4.532	-0.330	0.605	-0.461126
H	-2.906	1.429	0.984	0.383628
H	-3.063	2.068	-0.473	0.383628
H	-2.394	0.211	-1.577	0.005175
H	-1.260	-1.667	-0.375	0.094726
H	-1.310	-0.726	1.096	0.094726
H	0.163	0.030	-1.459	-0.019932
H	0.000	1.072	-0.061	-0.019932
H	1.346	-1.679	-0.104	0.124069
H	1.116	-0.722	1.344	0.124069
H	-3.824	-1.630	-0.718	0.013141
N	2.610	-0.024	0.065	0.178239
C	2.946	0.077	-1.387	-0.378967
H	2.863	-0.901	-1.837	0.187321
H	2.269	0.764	-1.868	0.187321
H	3.958	0.440	-1.485	0.187321
C	3.633	-0.886	0.734	-0.378967
H	4.608	-0.439	0.611	0.187321
H	3.397	-0.964	1.785	0.187321
H	3.621	-1.866	0.282	0.187321
C	2.659	1.338	0.683	-0.378967
H	2.387	1.262	1.725	0.187321
H	3.664	1.721	0.594	0.187321
H	1.976	1.995	0.171	0.187321

D

Fluorination of KDM5A$_{PHD3}$ Aromatic Cage - Appendix

D.1 Fluorinated TRP cage partial charges and XYZ coordinates of ligands 5F-Trp, 6F-Trp, 5,6diF-Trp.

5F-Trp Atom	X	Y	Z	Charge
N	2.808	0.398	1.397	-1.084269
C	2.418	-0.492	0.332	0.504578
C	3.635	-1.220	-0.195	0.419917
O	4.751	-0.835	-0.062	-0.478103
C	1.721	0.180	-0.882	-0.424242
C	0.423	0.849	-0.533	0.100781
C	0.205	2.177	-0.395	-0.242948
C	-0.847	0.210	-0.268	0.087437
N	-1.099	2.422	-0.059	-0.375718
C	-1.766	1.226	0.021	0.094860
C	-1.274	-1.124	-0.265	-0.349371
C	-3.099	0.954	0.318	-0.210229
C	-2.585	-1.368	0.027	0.362621
C	-3.505	-0.359	0.320	-0.301287
H	3.639	0.900	1.154	0.393316
H	1.742	-1.242	0.735	-0.009830
H	2.405	0.911	-1.304	0.113947
H	1.546	-0.565	-1.656	0.113947
H	0.893	2.989	-0.522	0.210990
H	-1.499	3.320	0.073	0.369296
H	-0.615	-1.943	-0.485	0.179767
H	-3.796	1.743	0.539	0.176683
H	-4.519	-0.631	0.541	0.191751
F	-3.025	-2.629	0.037	-0.223847
H	2.079	1.050	1.608	0.393316
H	3.424	-2.128	-0.769	-0.013363

6F-Trp Atom	X	Y	Z	Charge
N	2.879	0.357	1.445	-1.077819
C	2.568	-0.511	0.338	0.486763
C	3.837	-1.154	-0.178	0.431700
O	4.927	-0.720	0.007	-0.481065
C	1.871	0.168	-0.873	-0.406804
C	0.530	0.753	-0.539	0.112672
C	0.237	2.058	-0.364	-0.266157
C	-0.712	0.034	-0.328	-0.019942
N	-1.094	2.217	-0.055	-0.373504
C	-1.693	0.990	-0.032	0.199442
C	-1.077	-1.315	-0.377	-0.103281
C	-3.021	0.648	0.223	-0.445644
C	-2.380	-1.675	-0.131	-0.389942
C	-3.319	-0.685	0.164	0.437374
H	3.691	0.907	1.249	0.392747
H	1.922	-1.311	0.692	-0.010470
H	2.526	0.950	-1.246	0.109310
H	1.759	-0.555	-1.678	0.109310
H	0.878	2.913	-0.446	0.213168
H	-1.548	3.085	0.099	0.372937
H	-0.351	-2.073	-0.608	0.144826
H	-3.782	1.371	0.451	0.212242
H	-2.699	-2.700	-0.159	0.200781
F	-4.573	-1.068	0.398	-0.224650
H	2.112	0.965	1.653	0.392747
H	3.691	-2.048	-0.794	-0.016740

5,6diF-Trp

Atom	X	Y	Z	Charge
N	3.118	0.315	1.436	-1.086308
C	2.722	-0.547	0.351	0.503447
C	3.926	-1.311	-0.158	0.426946
O	5.051	-0.967	0.007	-0.478144
C	2.080	0.164	-0.871	-0.463851
C	0.799	0.875	-0.545	0.186917
C	0.626	2.207	-0.395	-0.321047
C	-0.500	0.279	-0.318	-0.001369
N	-0.681	2.493	-0.086	-0.304664
C	-1.388	1.324	-0.038	0.109248
C	-0.979	-1.037	-0.342	-0.309149
C	-2.741	1.103	0.225	-0.392988
C	-2.302	-1.251	-0.086	0.218428
C	-3.170	-0.192	0.195	0.274993
H	3.975	0.784	1.223	0.395594
H	2.012	-1.279	0.727	-0.009224
H	2.799	0.876	-1.267	0.122537
H	1.897	-0.564	-1.66	0.122537
H	1.342	2.997	-0.494	0.225782
H	-1.053	3.402	0.051	0.358256
H	-0.346	-1.878	-0.557	0.193691
H	-3.434	1.894	0.442	0.223891
F	-4.442	-0.476	0.435	-0.187665
F	-2.803	-2.481	-0.098	-0.188765
H	2.411	0.993	1.636	0.395594
H	3.698	-2.205	-0.748	-0.014687

D.2 Electrostatic values

Residue	Indole[b]	Indole + F[b]
Trpl8	–12.0 ± 2.6	-
Trp28	–13.1 ± 1.2	-
5F-Trp18	–8.0 ± 1.0	–8.9 ± 1.6
5F-Trp28	–7.6 ± 1.0	–8.1 ± 2.5
6F-Trp18	–7.6 ± 0.9	–8.8 ± 1.1
5F-Trp28	–7.7 ± 1.0	–9.0 ± 1.1
5,6diF-Trp18	–5.3 ± 0.8	–7.4 ± 0.8
5,6diF-Trp28	–7.4 ± 1.5	–7.7 ± 1.0

E

Antibiotics - Appendix

E.1 Partial charges and XYZ coordinates of ligands 1-βH, 1-βMe, and KCX.

1-βH Atom	X	Y	Z	Charge
N	-7.010	-2.001	-0.033	-1.192228
C	-6.554	-1.660	1.299	0.490890
C	-5.857	-0.309	1.438	-0.148876
O	-4.737	-0.277	0.557	-0.339410
C	-7.726	-1.759	2.257	0.400353
O	-7.923	-0.997	3.148	-0.503252
C	-3.008	-0.861	-3.742	-0.245295
N	8.447	-1.767	0.594	-0.879666
O	-5.768	1.017	-0.922	-0.609450
O	-1.595	3.009	2.093	-0.588902
O	6.475	-3.379	0.512	-0.572946
O	6.479	-1.383	1.970	-0.572946
O	0.543	2.536	1.912	-0.507325
O	-4.045	-2.035	-1.893	-0.728838
C	-4.814	0.397	-0.572	0.728838
C	3.670	2.768	-0.559	0.271010
C	6.196	0.246	-0.602	-0.085806
C	3.627	0.393	-0.537	-0.238950
N	-2.155	1.067	0.356	-0.689319
N	4.982	2.305	-0.112	-1.098078
N	6.273	-1.215	-0.562	-0.688926
S	1.223	1.702	-1.309	-0.443388
C	-0.575	2.372	1.558	0.685769
C	-0.950	1.382	0.484	0.244535
C	-3.942	-0.783	-2.539	0.296353
C	-0.965	0.001	-1.402	-0.420699
C	4.964	0.848	0.072	0.664960
C	2.721	1.598	-0.281	0.046984
C	-3.578	0.257	-1.458	-0.300498
C	-2.305	-0.002	-0.640	0.773666
C	0.025	0.656	-0.419	0.128393
S	6.830	-2.013	0.737	1.198859
H	-6.230	-2.219	-0.622	0.444500
H	-7.499	-1.235	-0.456	0.444500
H	-5.857	-2.438	1.601	0.005778
H	-5.462	-0.179	2.433	0.127505
H	-6.524	0.507	1.213	0.127505
H	-8.390	-2.606	2.069	0.033471
H	-3.452	-1.501	-4.499	0.064850
H	-2.862	0.116	-4.190	0.064850
H	-2.044	-1.278	-3.483	0.064850
H	8.862	-2.346	-0.110	0.415825
H	8.901	-1.860	1.482	0.415825
H	-4.239	-2.709	-2.532	0.459948
H	3.368	3.670	-0.042	0.068463
H	3.697	2.994	-1.622	0.068463
H	7.096	0.636	-0.148	0.121237
H	6.214	0.539	-1.644	0.121237
H	3.259	-0.524	-0.091	0.084659
H	3.725	0.233	-1.608	0.084659
H	5.263	2.768	0.726	0.433571
H	5.536	-1.727	-1.000	0.361682
H	-4.919	-0.487	-2.907	0.041484
H	-0.642	-0.986	-1.701	0.145773
H	4.982	0.589	1.125	-0.041732
H	2.413	1.631	0.753	0.064640
H	-3.486	1.215	-1.960	0.105885
H	-2.430	-0.925	-0.091	-0.076386
H	-2.412	2.680	1.725	0.438982
H	0.570	-0.084	0.154	0.078473
H	-1.038	0.613	-2.293	0.145773

1-βMe Atom	X	Y	Z	Charge
N	-6.311	2.552	0.439	-1.181961
C	-5.718	2.539	-0.884	0.482560
C	-5.263	1.174	-1.392	-0.147977
O	-4.333	0.627	-0.462	-0.310855
C	-6.683	3.182	-1.860	0.416651
O	-6.875	2.787	-2.964	-0.507074
C	-3.232	-0.432	3.903	-0.361217
C	-1.278	-2.661	2.040	-0.418167
N	7.962	2.414	0.016	-0.877754
O	-5.790	-0.838	0.346	-0.615833
O	-1.737	-2.331	-2.970	-0.601334
O	5.904	3.479	1.079	-0.578483
O	5.849	2.490	-1.186	-0.578483
O	0.425	-2.062	-2.690	-0.508815
O	-3.625	1.387	2.361	-0.725566
C	-4.704	-0.351	0.338	0.745821
C	3.766	-2.711	-0.900	0.275807
C	6.075	-0.221	0.214	-0.053808
C	3.541	-0.667	0.290	-0.243603
N	-2.167	-1.207	-0.596	-0.598822
N	4.968	-1.917	-1.148	-1.094118
N	6.030	1.063	0.915	-0.69432
S	1.366	-2.493	0.470	-0.399382
C	-0.675	-2.012	-2.258	0.699478
C	-0.988	-1.536	-0.861	0.313000
C	-3.928	0.036	2.631	0.309795
C	-0.998	-1.268	1.463	0.139133
C	4.812	-0.576	-0.570	0.668567
C	2.703	-1.695	-0.470	0.045052
C	-3.630	-0.769	1.342	-0.216047
C	-2.218	-0.624	0.754	0.317648
C	-0.002	-1.309	0.272	-0.14951
S	6.327	2.469	0.160	1.216147
H	-5.607	2.399	1.133	0.439132
H	-6.993	1.823	0.537	0.439132
H	-4.852	3.197	-0.844	0.005872
H	-4.737	1.266	-2.329	0.120353
H	-6.095	0.501	-1.518	0.120353
H	-7.201	4.061	-1.468	0.027242
H	-3.573	0.175	4.738	0.095595
H	-3.485	-1.460	4.133	0.095595
H	-2.156	-0.335	3.844	0.095595
H	-1.635	-3.348	1.279	0.137016
H	-2.022	-2.621	2.822	0.137016
H	-0.381	-3.082	2.474	0.137016
H	8.430	2.619	0.877	0.413426
H	8.273	2.997	-0.736	0.413426
H	-3.740	1.903	3.148	0.463042
H	3.465	-3.266	-1.779	0.063108
H	3.951	-3.428	-0.105	0.063108
H	6.931	-0.207	-0.446	0.110060
H	6.261	-0.985	0.958	0.110060
H	3.047	0.292	0.403	0.089277
H	3.771	-1.048	1.282	0.089277
H	5.173	-1.865	-2.123	0.430343
H	5.326	1.181	1.614	0.359158
H	-4.997	-0.065	2.796	0.037180
H	-0.609	-0.625	2.241	0.030167
H	4.665	0.168	-1.345	-0.051651
H	2.267	-1.253	-1.354	0.065968
H	-3.849	-1.808	1.553	0.101939
H	-2.028	0.435	0.633	0.019990
H	-2.530	-2.135	-2.477	0.434939
H	0.435	-0.324	0.144	0.140736

	KCX			
Atom	X	Y	Z	Charge
N	-2.078	-1.720	-0.342	-1.079360
C	-2.514	-0.438	0.191	0.365783
C	-1.595	0.701	-0.282	-0.406189
C	-0.170	0.595	0.267	0.261469
C	0.710	1.786	-0.109	-0.259857
C	2.184	1.567	0.255	0.465934
N	2.809	0.520	-0.515	-0.870780
C	-3.908	-0.181	-0.337	0.537326
O	-4.820	0.229	0.311	-0.568047
C	3.083	-0.766	0.057	1.076900
O	3.971	-1.395	-0.518	-0.858684
O	2.381	-1.078	1.026	-0.858684
H	-2.460	-2.476	0.193	0.393748
H	-1.082	-1.810	-0.297	0.393748
H	-2.567	-0.414	1.278	-0.028042
H	-2.042	1.643	0.029	0.096128
H	-1.572	0.704	-1.371	0.096128
H	-0.204	0.506	1.350	0.024758
H	0.323	-0.301	-0.080	0.024758
H	0.340	2.688	0.385	0.015124
H	0.645	1.965	-1.182	0.015124
H	2.260	1.343	1.314	-0.079913
H	2.716	2.503	0.081	-0.079913
H	-4.026	-0.394	-1.405	-0.025117
H	3.554	0.812	-1.106	0.347659

E.2 Coordinates of optimised molecules (M06-2X/6-31++G(d,p),DMSO)

	$PhSCH_3$				CH_3CN		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.833	1.275	-0.089	C	0.000	0.000	0.280
C	-0.466	1.131	-0.321	N	0.000	0.000	1.436
C	0.128	-0.133	-0.245	C	0.000	0.000	-1.181
C	-0.657	-1.251	0.050	H	0.000	1.028	-1.549
C	-2.027	-1.105	0.267	H	0.890	-0.514	-1.549
C	-2.616	0.157	0.203	H	-0.890	-0.514	-1.549
H	-2.287	2.260	-0.149				
H	0.141	1.998	-0.567				
H	-0.193	-2.231	0.112				
H	-2.632	-1.977	0.495				
H	-3.681	0.271	0.376				
S	1.878	-0.332	-0.552				
C	2.552	0.455	0.946				
H	2.215	-0.076	1.837				
H	3.639	0.390	0.874				
H	2.258	1.504	0.997				

	$PhCOCH_3$				CH_3CONEt_2		
Atom	X	Y	Z	Atom	X	Y	Z
C	-1.966	-1.135	0.000	C	2.136	-0.959	-0.300
C	-0.579	-1.224	0.000	H	1.680	-1.640	-1.023
C	0.204	-0.062	0.000	H	2.372	-1.529	0.603
C	-0.421	1.19	0.000	H	3.060	-0.559	-0.714
C	-1.812	1.279	0.000	C	1.230	0.213	0.021
C	-2.584	0.118	0.000	O	1.649	1.374	-0.078
H	-2.568	-2.038	0.000	N	-0.039	-0.044	0.419
H	-0.082	-2.189	0.000	C	-0.594	-1.387	0.582
H	0.169	2.101	0.000	C	-1.574	-1.765	-0.524
H	-2.291	2.252	0.000	H	-1.095	-1.422	1.556
H	-3.668	0.188	0.000	H	0.223	-2.106	0.625
C	1.696	-0.201	0.000	H	-1.929	-2.788	-0.376
O	2.212	-1.308	0.000	H	-2.444	-1.103	-0.524
C	2.544	1.047	0.000	H	-1.091	-1.700	-1.505
H	3.595	0.764	0.000	C	-0.925	1.092	0.681
H	2.327	1.657	-0.882	C	-1.424	1.771	-0.592
H	2.328	1.656	0.883	H	-0.383	1.813	1.299
				H	-1.767	0.713	1.268
				H	-2.067	2.618	-0.337
				H	-0.579	2.143	-1.176
				H	-1.999	1.078	-1.211

	CH_3CO_2Et				CH_3COCH_3		
Atom	X	Y	Z	Atom	X	Y	Z
C	2.133	-0.885	0.001	C	0.000	0.179	0.000
H	1.994	-1.584	-0.827	O	0.000	1.398	0.000
H	2.090	-1.457	0.931	C	1.284	-0.612	-0.003
H	3.097	-0.388	-0.083	H	2.137	0.047	-0.169
C	1.034	0.138	-0.004	H	1.393	-1.114	0.964
O	1.195	1.342	-0.003	H	1.250	-1.392	-0.769
O	-0.176	-0.434	-0.001	C	-1.284	-0.612	0.003
C	-1.311	0.455	0.008	H	-1.393	-1.114	-0.965
C	-2.556	-0.404	-0.004	H	-1.250	-1.392	0.769
H	-1.253	1.103	-0.870	H	-2.137	0.047	0.169
H	-1.255	1.081	0.903				
H	-3.442	0.236	0.003				
H	-2.587	-1.027	-0.901				
H	-2.589	-1.050	0.876				

	$PhSO_2CH_3$				$PhCH_3$		
Atom	X	Y	Z	Atom	X	Y	Z
C	-2.431	-1.157	0.000	C	-1.201	1.204	0.002
C	-1.041	-1.203	0.000	C	0.194	1.203	-0.009
C	-0.326	-0.004	0.000	C	0.913	0.002	-0.011
C	-0.968	1.231	0.000	C	0.197	-1.201	-0.009
C	-2.363	1.262	0.000	C	-1.197	-1.206	0.002
C	-3.091	0.074	0.000	C	-1.902	-0.002	0.008
H	-3.000	-2.081	0.000	H	-1.740	2.147	0.002
H	-0.518	-2.155	0.000	H	0.734	2.146	-0.017
H	-0.411	2.161	0.000	H	0.740	-2.143	-0.017
H	-2.876	2.217	0.000	H	-1.734	-2.150	0.002
H	-4.175	0.105	0.000	H	-2.988	-0.003	0.013
S	1.451	-0.164	0.000	C	2.421	0.001	0.009
O	1.849	-0.820	1.255	H	2.795	-0.103	1.033
O	1.849	-0.820	-1.255	H	2.823	-0.832	-0.574
C	2.116	1.487	0.000	H	2.822	0.932	-0.398
H	1.804	2.007	-0.905				
H	1.804	2.007	0.905				
H	3.198	1.344	0.000				

	$CH_2(CO_2Et)_2$				$PhCH_2SPh$		
Atom	X	Y	Z	Atom	X	Y	Z
C	4.468	-0.548	-0.378	C	4.064	-1.218	0.247
C	3.034	-0.170	-0.672	C	2.693	-1.153	0.489
H	5.056	-0.481	-1.296	C	2.016	0.068	0.419
H	4.899	0.129	0.363	C	2.731	1.225	0.092
H	4.526	-1.572	-0.001	C	4.103	1.163	-0.151
H	2.965	0.858	-1.036	C	4.772	-0.059	-0.074
H	2.593	-0.847	-1.410	H	4.580	-2.171	0.311
O	2.304	-0.288	0.569	H	2.141	-2.056	0.740
C	1.013	0.037	0.657	H	2.210	2.177	0.033
O	0.438	-0.093	1.715	H	4.649	2.068	-0.398
C	0.331	0.570	-0.590	H	5.840	-0.107	-0.262
H	0.808	1.504	-0.899	C	0.531	0.128	0.651
H	0.419	-0.155	-1.404	H	0.205	-0.661	1.333
C	-1.123	0.845	-0.283	H	0.234	1.099	1.054
O	-1.558	1.932	0.028	S	-0.334	-0.120	-0.956
O	-1.863	-0.256	-0.386	C	-2.030	-0.044	-0.395
C	-3.263	-0.119	-0.061	C	-2.608	1.187	-0.066
C	-3.896	-1.483	-0.209	C	-2.784	-1.216	-0.291
H	-3.702	0.615	-0.741	C	-3.928	1.240	0.378
H	-3.342	0.264	0.960	H	-2.024	2.098	-0.159
H	-4.961	-1.417	0.025	C	-4.109	-1.157	0.142
H	-3.789	-1.851	-1.233	H	-2.331	-2.169	-0.545
H	-3.435	-2.201	0.474	C	-4.681	0.069	0.480
				H	-4.371	2.197	0.636
				H	-4.691	-2.070	0.217
				H	-5.711	0.113	0.820

	$PhCH_2COCH_3$		
Atom	X	Y	Z
C	2.303	1.211	0.215
C	0.971	1.233	-0.196
C	0.300	0.046	-0.503
C	0.988	-1.166	-0.393
C	2.321	-1.193	0.017
C	2.982	-0.003	0.323
H	2.812	2.142	0.447
H	0.446	2.182	-0.281
H	0.476	-2.095	-0.632
H	2.842	-2.142	0.096
H	4.020	-0.021	0.640
C	-1.148	0.069	-0.911
H	-1.375	-0.752	-1.603
H	-1.397	0.992	-1.448
C	-2.126	-0.062	0.252
O	-1.758	-0.290	1.388
C	-3.583	0.114	-0.096
H	-4.209	-0.180	0.746
H	-3.764	1.167	-0.334
H	-3.841	-0.469	-0.984

References

(1) Deans, C.; Maggert, K. A. *Genetics* **2015**, *199*, 887–896.

(2) Deichmann, U. *Developmental Biology* **2016**, *416*, 249–254.

(3) Waddington, C. H., *The strategy of the genes: a discussion of some aspects of theoretical biology*; Allen and Unwin: London: 1957.

(4) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Scientific Reports* **2011**, *1*.

(5) Phillips, D. M. P. *Biochemical Journal* **1963**, *87*, 258–263.

(6) Andreoli, F.; Rio, A. D. *Computational and Structural Biotechnology Journal* **2015**, *13*, 358–365.

(7) Mersfelder, E. L.; Parthun, M. R. *Nucleic Acids Research* **2006**, *34*, 2653–2662.

(8) Perez-Perri, J.; Acevedo, J. M.; Wappner, P. *International Journal of Molecular Sciences* **2011**, *12*, 4705–4721.

(9) Moosavi, A.; Ardekani, A. M. *Iranian Biomedical Journal* **2016**, *20*, 246–258.

(10) Albertus, D.; Seder, C. W.; Chen, G.; Wang, X.; Hartojo, W.; Lin, L.; Silvers, A.; Thomas, D. G.; Giordano, T. J.; Chang, A. C.; Orringer, M. B.; Bigbee, W. L.; Chinnaiyan, A. M.; Beer, D. G. *Journal of Thoracic Oncology* **2008**, *3*, 1236 –1244.

(11) Gery, S.; Komatsu, N.; Kawamata, N.; Miller, C. W.; Desmond, J.; Virk, R. K.; Marchevsky, A.; McKenna, R.; Taguchi, H.; Koeffler, H. P. *Clinical Cancer Research* **2007**, *13*, 1399–1404.

(12) Park, Y. S.; Jin, M. Y.; Kim, Y. J.; Yook, J. H.; Kim, B. S.; Jang, S. J. *Annals of Surgical Oncology* **2008**, *15*, 1968–1976.

(13) Lee, S. H.; Kim, J.; Kim, W. H.; Lee, Y. M. *Oncogene* **2009**, *28*, 184–194.

(14) Cooper, C. S.; Foster, C. S. *British Journal of Cancer* **2009**, *100*, 240–245.

(15) Kikuno, N.; Shiina, H.; Urakami, S.; Kawamoto, K.; Hirata, H.; Tanaka, Y.; Majid, S.; Igawa, M.; Dahiya, R. *International Journal of Cancer* **2008**, *123*, 552–560.

(16) Liu, Y.; Hong, Y.; Zhao, Y.; Ismail, T. M.; Wong, Y.; Eu, K. W. *Cancer Investigation* **2008**, *26*, 575–582.

(17) Kumagai, T.; Akagi, T.; Desmond, J. C.; Kawamata, N.; Gery, S.; Imai, Y.; Song, J. H.; Gui, D.; Said, J.; Koeffler, H. P. *International Journal of Cancer* **2009**, *124*, 827–833.

(18) Semenza, G. L. *Trends in Molecular Medicine* **2002**, *8*, S62–S67.

(19) Bos, R.; Zhong, H.; Hanrahan, C. F.; Mommers, E. C. M.; Semenza, G. L.; Pinedo, H. M.; Abeloff, M. D.; Simons, J. W.; van Diest, P. J.; van der Wall, E. *Journal of the National Cancer Institute* **2001**, *93*, 309–314.

(20) Birner, P.; Schindl, M.; Obermair, A.; Plank, C.; Breitenecker, G.; Oberhuber, G. *Cancer Research* **2000**, *60*, 4693–4696.

(21) Birner, P.; Gatterbauer, B.; Oberhuber, G.; Schindl, M.; Rössler, K.; Prodinger, A.; Budka, H.; Hainfellner, J. A. *Cancer*, *92*, 165–171.

(22) Aebersold, D.; Burri, P.; Beer, K.; Laissue, J.; Djonov, V.; Greiner, R.; Semenza, G. *Cancer Research* **2001**, *61*, 2911–2916.

(23) Koukourakis, M.; Giatromanolaki, A.; Skarlatos, J.; Corti, L.; Blandamura, S.; Piazza, M.; Gatter, K.; Harris, A. *Cancer Research* **2001**, *61*, 1830–1832.

(24) Birner, P.; Schindl, M.; Obermair, A.; Breitenecker, G.; Oberhuber, G. *Clinical Cancer Research* **2001**, *7*, 1661–1668.

(25) Cortopassi, W. A.; Kumar, K.; Duarte, F.; Pimentel, A. S.; Paton, R. S. *Journal of Molecular Graphics and Modelling* **2016**, *67*, 69–84.

(26) VanderMolen, K. M.; McCulloch, W.; Pearce, C. J.; Oberlies, N. H. *The Journal Of Antibiotics* **2011**, *64*, 525–531.

(27) Mann, B. S.; Johnson, J. R.; Cohen, M. H.; Justice, R.; Pazdur, R. *The Oncologist* **2007**, *12*, 1247–1252.

(28) Syzf, M. *Nature Reviews Drug Discovery* **2015**, *14*, 461–474.

(29) 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. *Clinical Cancer Research* **2005**, *11*, 3604–3608.

(30) Saba, H. I. *Therapeutics and Clinical Risk Management* **2007**, *3*, 807–817.

(31) 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.; Rezaï, K. *Clinical Pharmacokinetics* **2016**, *55*, 397–405.

(32) Wapenaar, H.; Dekker, F. J. *Clinical Epigenetics* **2016**, *8*, 59.

(33) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. *Nature Reviews Drug Discovery* **2012**, *11*, 384–400.

(34) Yang, X.; Seto, E. *Oncogene* **2007**, *26*, 5310–5318.

(35) Marks, P.; Xu, W.-S. *Journal of Cellular Biochemistry*, *107*, 600–608.

(36) Kong, X.; Ouyang, S.; Liang, Z.; Lu, J.; Chen, L.; Shen, B.; Li, D.; Zheng, M.; Li, K. K.; Luo, C.; Jiang, H. *PLoS One* **2011**, *6*, 1–11.

(37) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. *Cell* **2004**, *119*, 941–953.

(38) Stavropoulos, P.; Blobel, G.; Hoelz, A. *Nature Structural & and Molecular Biology* **2006**, *13*, 626–632.

(39) Kooistra, S.; Helin, K. *Nature Reviews Molecular Cell Biology* **2012**, *13*, 297–311.

(40) Tsukada, Y.-I.; Fang, J.; Erdjument-Bromage, H.; Warren, M.; Borchers, C.; Tempst, P.; Zhang, Y. *Nature* **2006**, *439*, 811–816.

(41) Langley, G.; Brinkø, A.; Münzel, M.; Walport, L.; Schofield, C.; Hopkinson, R. *ACS Chemical Biology* **2016**, *11*, 755–762.

(42) Tanner, K.; Trievel, R.; Kuo, M.-H.; Howard, R.; Berger, S.; Allis, C.; Marmorstein, R.; Denut, J. *Journal of Biological Chemistry* **1999**, *274*, 18157–18160.

(43) Sternglanz, R.; Schindelin, H. *Proceedings of the National Academy of Sciences* **1999**, *96*, 8807–8808.

(44) Jiang, J.; Lu, J.; Lu, D.; Liang, Z.; Li, L.; Ouyang, S.; Kong, X.; Jiang, H.; Shen, B.; Luo, C. *PLoS One* **2012**, *7*, e25444.

(45) Jedrzejewski, R.; Kazmierkiewicz, R. *Journal of Molecular Modeling* **2013**, *19*, 5533–5538.

(46) Fenza, A. D.; Rocchia, W.; Tozzini, V. *Proteins* **2009**, *76*, 946–958.

(47) Quy, V.; Pantano, S.; Rossetti, G.; Giacca, M.; Carloni, P. *Biology (Basel)* **2012**, *1*, 277–296.

(48) Mujtaba, S.; He, Y.; Zeng, L.; Farooq, A.; Carlson, J. E.; Ott, M.; Verdin, E.; Zhou, M.-M. *Molecular Cell* **2002**, *9*, 575–586.

(49) Terreni, M.; Valentini, P.; Liverani, V.; Gutierrez, M. I.; Di Primio, C.; Di Fenza, A.; Tozzini, V.; Allouch, A.; Albanese, A.; Giacca, M.; Cereseto, A. *Retrovirology* **2010**, *7*, 18.

(50) Smith, B.; Denu, J. *Biochimica et Biophysica Acta* **2009**, *1789*, 45–57.

(51) Roth, S.; Denu, J.; Allis, C. *Annual Review of Biochemistry* **2001**, *70*, 81–120.

(52) Cyr, A.; Domann, F. *Antioxidants & Redox Signaling* **2011**, *15*, 551–589.

(53) Yan, Y.; Harper, S.; Speicher, D.; Marmorstein, R. *Nature Structural & Molecular Biology* **2002**, *9*, 862–869.

(54) Berndsen, C.; Albaugh, B.; Tan, S.; Denu, J. *Biochemistry* **2007**, *46*, 623–629.

(55) Zhang, X.; Ouyang, S.; Kong, X.; Liang, Z.; Lu, J.; Zhu, K.; Zhao, D.; Zheng, M.; Jiang, H.; Liu, X.; Marmorstein, R.; Luo, C. *Journal of Physcial Chemistry B* **2014**, *118*, 2009–2019.

(56) Liu, X.; Wang, L.; Zhao, K.; Thompson, P.; Hwang, Y.; Marmorstein, R.; Cole, P. *Nature* **2008**, *451*, 846–850.

(57) Dillon, S.; Zhang, X.; Trievel, R.; Cheng, X. *Genome Biology* **2005**, *6*, 227.

(58) Zhang, X.; Wen, H.; Shi, X. *Acta Biochimica et Biophysica Sinica* **2012**, *44*, 14–27.

(59) Kaniskan, H.; Jin, J. *ACS Chemical Biology* **2015**, *10*, 40–50.

(60) Chuikov, S.; Kurash, J.; Wilson, J.; Xiao, B.; Justin, N.; Ivanov, G.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S.; Barlev, N.; Reinberg, D. *Nature* **2004**, *432*, 353–360.

(61) Bai, Q.; Shen, Y.; Yao, X.; Wang, F.; Du, Y.; Wang, Q.; Jin, N.; Hai, J.; Hu, T.; Yang, J. *PLoS One* **2011**, *6*, 1–9.

(62) Kwon, T.; Chang, J.; Kwak, E.; Lee, C.; Joachimiak, A.; Kim, Y.; Lee, J.; Cho, Y. *EMBO Journal* **2003**, *22*, 292–303.

(63) Huang, J.; Perez-Burgos, L.; Placek, B. J.; Sengupta, R.; Richter, M.; Dorsey, J. A.; Kubicek, S.; Opravil, S.; Jenuwein, T.; Berger, S. L. *Nature* **2006**, *444*, 629–632.

(64) Qiao, Q.; Li, Y.; Chen, Z.; Wang, M.; Reinberg, D.; Xu, R. *Journal of Biological Chemistry* **2011**, *286*, 8361–8368.

(65) Hu, P.; Zhang, Y. *Journal of the American Chemical Society* **2006**, *128*, 272–1278.

(66) Guo, H.; Guo, H. *Proceedings of the National Academy of Sciences* **2007**, *104*, 8797–8802.

(67) Wang, S.; Hu, P.; Zhang, Y. *The Journal of Physical Chemistry B* **2007**, *111*, 3758–3764.

(68) Hu, P.; Wang, S.; Zhang, Y. *Journal of the American Chemical Society* **2008**, *130*, 3806–3813.

(69) Chu, Y.; Yao, J.; Guo, H. *PLoS One* **2012**, *7*, 1–9.

(70) Xu, Q.; Chu, Y.; Guo, H.; Smith, J.; Guo, H. *Chemistry* **2009**, *15*, 12596–12599.

(71) Chu, Y.; Sun, L.; Zhong, S. *Journal of Molecular Modeling* **2015**, *21*, 125.

(72) Wu, H.; Mathioudakis, N.; Diagouraga, B.; Dong, A.; Dombrovski, L.; Baudat, F.; Cusack, S.; DeMassy, B.; Kadlec, J. *Cell Reports* **2013**, *5*, 13–20.

(73) Haider, S.; Joseph, C.; Neidle, S.; Fierke, C.; Fuchter, M. *Bioorganic & Medicinal Chemistry Letters* **2011**, *21*, 2129–2132.

(74) Kalyaanamoorthy, S.; Chen, Y. *Journal of Computational Chemistry* **2013**, *34*, 2270–2283.

(75) Kalyaanamoorthy, S.; Chen, Y. *Journal of Chemical Information and Modeling* **2012**, *52*, 589–603.

(76) Thangapandian, S.; John, S.; Lee, Y.; Arulalapperumal, V.; Lee, K. *PLoS One* **2012**, *7*, 1–11.

(77) Arrar, M.; Turnham, R.; Pierce, L.; de Oliveira, C.; McCammon, J. *Protein Science* **2013**, *22*, 83–92.

(78) Kunze, M.; Wright, D.; Werbeck, N.; Kirkpatrick, J.; Coveney, P.; Hansen, D. *Journal of the American Chemical Society* **2013**, *135*, 17862–17868.

(79) Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. *Journal of the American Chemical Society* **2010**, *132*, 9471–9479.

(80) Vanommeslaeghe, K.; Proft, F. D.; Loverix, S.; Tourwe, D.; Geerlings, P. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 3987–3992.

(81) Chen, K.; Zhang, X.; Wu, Y.-D.; Wiest, O. *Journal of the American Chemical Society* **2014**, *136*, 11636–11643.

(82) Finnin, M.; Donigian, J.; Cohen, A.; Richon, V.; Rifkind, R.; Marks, P.; Breslow, R.; Pavletich, N. *Nature* **1999**, *401*, 188–193.

(83) Gleeson, D.; Gleeson, M. P. *Medicinal Chemistry Communication* **2015**, *6*, 477–485.

(84) Hsu, H.-C.; Wang, C.-L.; Wang, M.; Yang, N.; Chen, Z.; Sternglanz, R.; Xu, R.-M. *Genes & Development* **2013**, *27*, 64–73.

(85) Sakkiah, S.; Arooj, M.; Cao, G.; Lee, K. *PLoS One* **2013**, *8*, 1–11.

(86) Guan, X.; Lin, P.; Knoll, E.; Chakrabarti, R. *PLoS One* **2014**, *9*, 1–18.

(87) Liang, Z.; Shi, T.; Ouyang, S.; Li, H.; Yu, K.; Zhu, W.; Luo, C.; Jiang, H. *Journal of Physical Chemistry B* **2010**, *114*, 11927–11933.

(88) Hu, P.; Wang, S.; Zhang, Y. *Journal of the American Chemical Society* **2008**, *130*, 16721–16728.

(89) Shi, Y.; Zhou, Y.; Wang, S.; Zhang, Y. *Journal of Physical Chemistry Letters* **2013**, *4*, 491–495.

(90) Cortopassi, W. A.; Simion, R.; Honsby, C. E.; França, T. C. C.; Paton, R. S. *Chemistry: A European Journal* **2015**, *21*, 18865–18865.

(91) Ulucan, O.; Keskin, O.; Erman, B.; Gursoy, A. *PLoS One* **2011**, *6*, 1–9.

(92) Baron, R.; Binda, C.; Tortorici, M.; McCammon, J.; Mattevi, A. *Structure* **2011**, *19*, 212–220.

(93) Baron, R.; Vellore, N. *Proceedings of the National Academy of Sciences* **2012**, *109*, 12509–12514.

(94) Vellore, N.; Baron, R. *BMC Biophysics* **2013**, *6*, 15.

(95) Karasulu, B.; Patil, M.; Thiel, W. *Journal of the American Chemical Society* **2013**, *135*, 13400–13413.

(96) Yu, T.; Higashi, M.; Cembran, A.; Gao, J.; Truhlar, D. *Journal of Physical Chemistry B* **2013**, *117*, 8422–8429.

(97) Pizzitutti, F.; Giansanti, A.; Ballario, P.; Ornaghi, P.; Torreri, P.; Ciccotti, G.; Filetici, P. *Journal of Molecular Recognition* **2006**, *19*, 1–9.

(98) Spiliotopoulos, D.; Caflisch, A. *Israel Journal of Chemistry* **2014**, *54*, 1084–1092.

(99) Magno, A.; Steiner, S.; Caflisch, A. *Journal of Chemical Theory and Computation* **2013**, *9*, 4225–4232.

(100) Ferguson, F.; Fedorov, O.; Chaikuad, A.; Philpott, M.; Muniz, J.; Felletar, I.; Von Delft, F.; Heightman, T.; Knapp, S.; Abell, C.; Ciulli, A. *Journal of Medicinal Chemistry* **2013**, *56*, 10183–10187.

(101) Lucas, X.; Wohlwend, D.; Hügle, M.; Schmidtkunz, K.; Gerhardt, S.; Schüle, R.; Jung, M.; Einsle, O.; Günther, S. *Angewandte Chemie International Edition* **2013**, *52*, 14055–14059.

(102) Steiner, S.; Magno, A.; Huang, D.; Caflisch, A. *FEBS Letters* **2013**, *587*, 2158–2163.

(103) Rooney, T. P. C. et al. *Angewandte Chemie International Edition* **2014**, *53*, 6126–6130.

(104) *Epigenetics for Drug Discovery*; Carey, N., Ed.; Drug Discovery; The Royal Society of Chemistry: 2016.

(105) Gao, C.; Herold, J.; Kireev, D. *Journal of Computational Chemistry* **2012**, *33*, 659–665.

(106) Kamps, J. J.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Bickelhaupt, F. M.; Li, H.; Mecinović, J. *Nature Communications* **2015**, *6*.

(107) Taverna, S.; Li, H.; Ruthenburg, A. J.; Allis, C.; Patel, D. *Nature Structural & Molecular Biology* **2007**, *14*, 1025–1040.

(108) Jiang, J. Y.; Zeng, M.; Zhang, N.; Yu, Q. *International Journal of Quantum Chemistry* **2009**, *109*, 1135–1147.

(109) Jacobs, S.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080–2083.

(110) Fischle, W.; Boo, S.; Dormann, H.; Ueberheide, B.; Garcia, B.; Shabanowitz, J.; Hunt, D.; Funabiki, H.; Allis, C. *Nature* **2005**, *438*, 1116–1122.

(111) Li, N.; Stein, R.; He, W.; Komives, E.; Wang, W. *Molecular and Cellular Proteomics* **2013**, *12*, 2750–2760.

(112) Stein, R.; Wang, W. *Journal of Molecular Biology* **2011**, *406*, 527–541.

(113) Lu, Z.; Lai, J.; Zhang, Y. *Journal of the American Chemical Society* **2009**, *131*, 14928–14931.

(114) Ozboyaci, M.; Gursoy, A.; Erman, B.; Keskin, O. *PLoS One* **2011**, *6*.

(115) Grauffel, C.; Stote, R.; Dejaegere, A. *Biochimica et Biophysica Acta* **2015**, *1850*, 1026–1040.

(116) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.

(117) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. *Advances and Applications in Bioinformatics and Chemistry* **2015**, *8*, 37–47.

(118) Verlet, L. *Physical Review* **1967**, *159*, 98–103.

(119) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953**, *21*, 1087–1092.

(120) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

(121) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.

(122) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187–217.

(123) Van Gunsteren, W.; Billeter, S.; Eising, A.; Hünenberger, P.; Krüger, P.; Mark, A.; Scott, W.; Tironi, I. *Biomolecular simulation: The GROMOS96 manual and user guide* **1996**, 1–1042.

(124) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Structural Biology* **2002**, *9*, 425–430.

(125) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, 864–871.

(126) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566–569.

(127) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(128) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chemical Physics Letters* **1989**, *157*, 200–206.

(129) Lee, C.; Yang, W.; Parr, R. G. *Physical Review B* **1988**, *37*, 785–789.

(130) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(131) Becke, A. D. *The Journal of Chemical Physics* **1996**, *104*, 1040–1046.

(132) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.

(133) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *The Journal of Chemical Physics* **2003**, *119*, 12129–12137.

(134) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2006**, *2*, 364–382.

(135) Zhao, Y.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2006**, *110*, 5121–5129.

(136) Zhao, Y.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2006**, *110*, 13126–13130.

(137) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *The Journal of Chemical Physics* **2005**, *123*, 161103.

(138) Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

(139) Miertus, S.; Scrocco, E.; Tomasi, J. *Chemical Physics* **1981**, *55*, 117–129.

(140) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2009**, *113*, 6378–6396.

(141) Okada, Y.; Scott, G.; Ray, M.; Mishina, Y.; Zhang, Y. *Nature* **2007**, *450*, 119–123.

(142) Uemura, M.; Yamamoto, H.; Takemasa, I.; Mimori, K.; Hemmi, H.; Mizushima, T.; Ikeda, M.; Sekimoto, M.; Matsuura, N.; Doki, Y.; Mori, M. *Clinical Cancer Research* **2010**, *16*, 4636–4646.

(143) Brand, M.; Measures, A. M.; Wilson, B. G.; Cortopassi, W. A.; Alexander, R.; Hoss, M.; Hewings, D. S.; Rooney, T. P.; Paton, R. S.; Conway, S. J. *ACS Chemical Biology* **2015**, *10*, 22–39.

(144) Filippakopoulos, P.; Knapp, S. *Nature Reviews Drug Discovery* **2014**, *13*, 337–356.

(145) Pan, C.; Mezei, M.; Mujtaba, S.; Muller, M.; Zeng, L.; Li, J.; Wang, Z.; Zhou, M. M. *Journal of Medicinal Chemistry* **2007**, *50*, 2285–2288.

(146) Gerona-Navarro, G.; Yoel, R.; Mujtaba, S.; Frasca, A.; Patel, J.; Zeng, L.; Plotnikov, A. N.; Osman, R.; Zhou, M. M. *Journal of the American Chemical Society* **2011**, *133*, 2040–2043.

(147) 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. *Chemical Biology* **2011**, *118*, 531–541.

(148) Galdeano, C.; Ciulli, A. *Future Medicinal Chemistry* **2016**, *8*, 1655–1680.

(149) Hewings, D. S. et al. *Journal of Medicinal Chemistry* **2013**, *56*, 3217–3227.

(150) Hay, D. A. et al. *Journal of the American Chemical Society* **2014**, *136*, 9308–9319.

(151) Picaud, S. et al. *Cancer Research* **2015**, *75*, 5106–5119.

(152) 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. *Molecular BioSystems* **2015**, *11*, 1295–1304.

(153) Muvva, C.; Singam, E. R.; Raman, S. S.; Subramanian, V. *Molecular BioSystems* **2014**, *10*, 2384–2397.

(154) Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. *Journal of Computational Chemistry* **2011**, *32*, 866–877.

(155) Rapp, C.; Kalyanaraman, C.; Schiffmiller, A.; Schoenbrun, E. L.; Jacobson, M. P. *Journal of Chemical Information and Modeling* **2011**, *51*, 2082–2089.

(156) Oehme, D. P.; Brownlee, R. T.; Wilson, D. J. *Journal of Computational Chemistry* **2012**, *33*, 2566–2580.

(157) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V. *Journal of Chemical Theory and Computation* **2014**, *10*, 1228–1241.

(158) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. *Chemical Science* **2016**, *7*, 207–218.

(159) Case, D. A.; et al. *Amber 12* **2012**, *1*.

(160) Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Physical Chemistry Chemical Physics* **2010**, *12*, 7821–7839.

(161) Connolly, L. *Science* **1983**, *221*, 709–713.

(162) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *The Journal of Chemical Physics* **1992**, *96*, 6796–6806.

(163) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(164) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*.

(165) Minoux, H.; Chipot, C. *Journal of the American Chemical Society* **1999**, *121*, 10366–10372.

(166) Gallivan, J. P.; Dougherty, D. A. *Proceedings of the National Academy of Sciences* **1999**, *96*, 9459–9464.

(167) Miller, B. R.; McGee T. D., J.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314–3321.

(168) Genheden, S.; Ryde, U. *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

(169) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 1940–1951.

(170) Wang, J.; Hou, T. *Journal of Chemical Information and Modeling* **2012**, *52*, 1199–1212.

(171) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *Journal of Medicinal Chemistry* **2006**, *49*, 6596–6606.

(172) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. *The Journal of Physical Chemistry B* **2013**, *117*, 8408–8421.

(173) Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. *Physical Chemistry Chemical Physics* **2014**, *16*, 22035–22045.

(174) Giulio, R.; Del, R. A.; Gianluca, D.; Miriam, S. *Journal of Computational Chemistry* **2010**, *31*, 797–810.

(175) Sun, H. Y.; Li, Y. Y.; Tian, S.; Xu, L.; Hou, T. J. *Physical Chemistry Chemical Physics* **2014**, *16*, 16719–16729.

(176) Genheden, S. *Journal of Computer-Aided Molecular Design* **2011**, *25*, 1085–1093.

(177) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(178) Frisch, M. J. et al. Gaussian 09 Revision D.01., Gaussian Inc. Wallingford CT 2009.

(179) Wavefunction, Inc. Irvine, CA.

(180) Johnson, B. G.; Gill, P. M. W.; Pople, J. A.; Fox, D. J. *Chemical Physics Letter* **1993**, *206*, 239–246.

(181) Wheeler, S. E.; Houk, K. N. *Journal of the American Chemical Society* **2009**, *131*, 3126–3127.

(182) Wintjens, R.; Lievin, J.; Rooman, M.; Buisine, E. *Journal of Molecular Biology* **2000**, *302*, 395—410.

(183) Flocco, M. M.; Mowbray, S. L. *Journal Molecular Biology* **1994**, *235*, 709–717.

(184) Gallivan, J. P.; Dougherty, D. A. *Journal of the American Chemical Society* **1999**, *122*, 870–874.

(185) Crowley, P. B.; Golovin, A. *Proteins* **2005**, *59*, 231–239.

(186) Reddy, A. S.; Vijay, D.; Sastry, G. M.; Sastry, G. N. *The Journal of Physical Chemistry B* **2006**, *110*, 2479–2481.

(187) Pinheiro, S.; Soteras, I.; Gelpi, J. L.; Dehez, F.; Chipot, C.; Luque, F. J.; Curutchet, C. *Physical Chemistry Chemical Physics* **2017**, *19*, 9849–9861.

(188) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. *Journal of Medicinal Chemistry* **2009**, *52*, 2952–2963.

(189) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235–242.

(190) Baxter, E.; Windloch, K.; Gannon, F.; Lee, J. S. *Cell & Bioscience* **2014**, *4*, 1–11.

(191) Cortopassi, W. A.; Kumar, K.; Paton, R. S. *Organic & Biomolecular Chemistry* **2016**, *14*, 10926–10938.

(192) Aravind, L.; Koonin, E. V. *Genome Biology* **2001**, *2*.

(193) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *2*, 1300–1303.

(194) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930–935.

(195) Liu, C.-K.; Hsu, C.-A.; Abbott, M. T. *Archives of Biochemistry and Biophysics* **1973**, *159*, 180–187.

(196) Chang, B.; Chen, Y.; Zhao, Y.; Bruick, R. K. *Science* **2007**, *318*, 444–447.

(197) Mantri, M.; Loik, N. D.; Hamed, R. B.; Claridge, T. D. W.; McCullagh, J. S. O.; Schofield, C. J. *ChemBioChem* **2011**, *12*, 531–534.

(198) Han, G.; Li, J.; Wang, Y.; Li, X.; Mao, H.; Liu, Y.; Chen, C. D. *Journal of Cellular Biochemistry* **2012**, *113*, 1663–1670.

(199) Webby, C. J. et al. *Science* **2009**, *325*, 90–93.

(200) Walport, L. J.; Hopkinson, R. J.; Chowdhury, R.; Schiller, R.; Ge, W.; Kawamura, A.; Schofield, C. J. *Nature Communication* **2016**, *7*, 11974.

(201) Zhang, N.; Fu, Z.; Linke, S.; Chicher, J.; Gorman, J. J.; Visk, D.; Haddad, G. G.; Poellinger, L.; Peet, D. J.; Powell, F.; Johnson, R. S. *Cell Metabolism* **2010**, *11*, 364–378.

(202) Fong, G.-H.; Takeda, K. *Cell Death and Differentiation* **2008**, *15*, 635–641.

(203) Kovaleva, E. G.; Lipscomb, J. D. *Nature Chemical Biology* **2008**, *4*, 186–193.

(204) Que, L. *Nature Structural and Molecular Biology* **2000**, *7*, 182–184.

(205) Martinez, S.; Hausinger, R. P. *J. Bio. Chem.* **2015**, *290*, 20702–20711.

(206) Alberro, N.; Torrent-Sucarrat, M.; Arrastia, I.; Arrieta, A.; Cossío, F. P. *Chemistry: A European Journal* **2017**, *23*, 137–148.

(207) Fang, D.; Cisneros, G. A. *Journal of Chemical Theory and Computation* **2014**, *10*, 5136–5148.

(208) Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P. *Chemistry: A European Journal* **2014**, *20*, 435–446.

(209) Liu, H.; Llano, J.; Gauld, J. W. *The Journal of Physical Chemistry B* **2009**, *113*, 4887–4898.

(210) Hu, L. et al. *Nature* **2015**, *527*, 118–122.

(211) De Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. *Journal of the American Chemical Society* **2004**, *126*, 8362–8363.

(212) Drew, K. L.; Reynisson, J. *European Journal of Medicinal Chemistry* **2012**, *56*, 48–55.

(213) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *The Journal of Chemical Physics* **2007**, *126* 084108.

(214) Somers, K. P.; Simmie, J. M. *The Journal of Physical Chemistry A* **2015**, *119*, 8922–8933.

(215) Lee, T. J.; Taylor, P. R. *International Journal of Quantum Chemistry* **1989**, *36*, 199–207.

(216) Lee, T. J.; Rendell, A. P.; Taylor, P. R. *The Journal of Physical Chemistry* **1990**, *94*, 5463–5468.

(217) Sun, H.; Law, C. K. *The Journal of Physical Chemistry A* **2010**, *114*, 12088–12098.

(218) Donoghue, P. J.; Tehranchi, J.; Cramer, C. J.; Sarangi, R.; Solomon, E. I.; Tolman, W. B. *Journal of the American Chemical Society* **2011**, *133*, 17602–17605.

(219) Jeong, Y. J.; Kang, Y.; Han, A.-R.; Lee, Y.-M.; Kotani, H.; Fukuzumi, S.; Nam, W. *Angewandte Chemie International Edition* **2008**, *47*, 7321–7324.

(220) Cho, K.; Leeladee, P.; McGown, A. J.; DeBeer, S.; Goldberg, D. P. *Journal of the American Chemical Society* **2012**, *134*, 7392–7399.

(221) Goldsmith, C. R.; Stack, T. D. P. *Inorganic Chemistry* **2006**, *45*, 6048–6055.

(222) Kim, S.; Chmely, S. C.; Nimlos, M. R.; Bomble, Y. J.; Foust, T. D.; Paton, R. S.; Beckham, G. T. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2846–2852.

(223) Perdew, J. P.; Zunger, A. *Physical Review B* **1981**, *23*, 5048–5079.

(224) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *The Journal of Physical Chemistry A* **2001**, *105*, 6750–6756.

(225) Fujii, N. *Biological and Pharmaceutical Bulletin* **2005**, *28*, 1585–1589.

(226) Fisher, G. H. *Experientia Supplementum* **1998**, *85*, 109–118.

(227) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. *Journal of Chemical Theory and Computation* **2010**, *6*, 2935–2947.

(228) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers*, *32*, 523–535.

(229) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *Journal of Computational Physics* **1977**, *23*, 327 –341.

(230) Wang, H.; Dommert, F.; Holm, C. *The Journal of Chemical Physics* **2010**, *133*, 034117.

(231) Dougherty, D. A. *Science* **1996**, *271*, 163–168.

(232) Dougherty, D. A. *Accounts of Chemical Research* **2013**, *46*, 885–893.

(233) Dougherty, D. A. *Journal of Nutrition* **2007**, *137*, 11504S–1508S.

(234) Xiu, X.; Puskar, N.; Shanata, J.; Lester, H.; Dougherty, D. *Nature* **2009**, *458*, 534–537.

(235) Cashin, A. L.; Petersson, E. J.; Lester, H. A.; Dougherty, D. A. *Journal of the American Chemical Society* **2005**, *127*, 350–356.

(236) Dougherty, D. A. *Journal of Organic Chemistry* **2008**, *73*, 3667–3673.

(237) Tavares, X. D. S.; Blum, A. P.; Nakamura, D. T.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. *Journal of the American Chemical Society* **2012**, *134*, 11474–11480.

(238) Patel, S.; Tariot, P.; Hamill, R. *Journal of Geriatric Psychiatry and Neurology* **1991**, *4*, 48–51.

(239) Queenan, A.; Bush, K. *Clinical Microbiology Reviews* **2007**, *20*, 440–458.

(240) Jeon, J.; Lee, J.; Lee, J.; Park, K.; Karim, A.; Lee, C.; Jeong, B.; Lee, S. *International Journal of Molecular Sciences* **2015**, *16*, 9654–9692.

(241) Smith, C.; Antunes, N.; Stewart, N.; Toth, M.; Kumarasiri, M.; Chang, M.; Mobashery, S.; Vakulenko, S. *Chemistry & Biology* **2013**, *20*, 1107–1115.

(242) Schneider, K.; Ortega, C.; Renck, N.; Bonomo, R.; Powers, R.; Leonard, D. *Journal of Molecular Biology* **2011**, *406*, 583–594.

(243) June, C.; Muckenthaler, T.; Schroder, E.; Klamer, Z.; Wawrzak, Z.; Powers, R.; Szarecka, A.; Leonard, D. *Protein Science* **2011**, *406*, 583–594.

(244) Schneider, K.; Karpen, M.; Bonomo, R.; Leonard, D.; Powers, R. *Biochemistry* **2009**, *48*, 11840–11847.

(245) Johnson, E.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A.; Yang, W. *Journal of the American Chemical Society* **2010**, *132*, 6498–6506.

(246) Saleh, G.; Gatti, C.; Lo Presti, L.; Contreras-García, J. *Chemistry – A European Journal* **2012**, *18*, 15523–15536.

(247) Schneider, T; Sah, l. *International Journal of Medical Microbiology* **2010**, *300*, 161–169.

(248) Müller, A.; Klöckner, A.; Schneider, T. *Nature Product Reports* **2017**, *34*, 909–932.

(249) Bugg, T.; Braddick, D.; Dowson, C.; Roper, D. *Trends in Biotechnology* **2011**, *29*, 167–173.

(250) Lavollay, M.; Arthur, M.; Fourgeaud, M.; Dubost, L.; Marie, A.; Veziris, N.; Blanot, D.; Gutmann, L.; Mainardi, J. *Journal of Bacteriology* **2008**, *190*, 4360–4366.

(251) Gupta, R.; Lavollay, M.; Mainardi, J.; Arthur, M.; Bishai, W.; Lamichhane, G. *Nature Medicine* **2010**, *16*, 466–469.

(252) Mainardi, J.; Fourgeaud, M.; Hugonnet, J.; Dubost, L.; Brouard, J.; Ouazzani, J.; Rice, L.; Gutmann, L.; Arthur, M. *The Journal of Biological Chemistry* **2005**, *280*, 38146–38152.

(253) Biarrotte-Sorin, S.; Hugonnet, J.; Delfosse, V.; Mainardi, J.; Gutmann, L.; Arthur, M.; Mayer, C. *Journal of Molecular Biology* **2006**, *359*, 533–538.

(254) Mainardi, J.-L.; Hugonnet, J.-E.; Rusconi, F.; Fourgeaud, M.; Dubost, L.; Moumi, A. N.; Delfosse, V.; Mayer, C.; Gutmann, L.; Rice, L. B.; Arthur, M. *Journal of Biological Chemistry* **2007**, *282*, 30414–30422.

(255) Steiner, E.; Schneider, G.; Schnell, R. *The FEBS Journal* **2017**, *284*, 725–741.

(256) Kumar, P.; Kaushik, A.; Lloyd, E.; Li, S.; Mattoo, R.; Ammerman, N.; Bell, D.; Perryman, A.; Zandi, T.; Ekins, S.; Ginell, S.; Townsend, C.; Freundlich, J.; Lamichhane, G. *Nature Chemical Biology* **2017**, *13*, 54–61.

(257) Dhar, N.; Dubée, V.; Ballell, L.; Cuinet, G.; Hugonnet, J.; Signorino-Gelo, F.; Barros, D.; Arthur, M.; McKinney, J. *Antimicrobial Agents and Chemotherapy* **2015**, *59*, 1308–1319.

(258) Zhang, X. M.; Bordwell, F. G.; Puy, M. V. D.; Fried, H. E. *Journal of Organic Chemistry* **1993**, *58*, 3060–3066.

(259) Kelly, C.; Cramer, C.; Truhlar, D. *The Journal of Physical Chemistry B* **2007**, *111*, 408–422.

(260) Amyes, T. L.; Richard, J. P. *Journal of the American Chemical Society* **1996**, *118*, 3129–3141.

(261) Amyes, T. L.; Richard, J. P. *Journal of the American Chemical Society* **1992**, *114*, 10297–10302.