



THE UNIVERSITY OF QUEENSLAND

A U S T R A L I A

Molecular Modelling of Covalent Inhibition of Bruton's Tyrosine Kinase by Cyanoacrylamides

Jonathan Yik Chang Ting

*A Research Report submitted for the degree of Bachelor of Advanced Science (Honours) at
The University of Queensland in October 2019
School of Chemistry & Molecular Biosciences*

Declaration by author

This research report is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my report.

I have clearly stated the contribution of others to my research report as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my report. The content of my report is the result of work I have carried out since the commencement of my honours research project and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my research report, if any, have been submitted to qualify for another award.

I acknowledge that copyright of all material contained in my research report resides with the copyright holder(s) of that material.

Statement of Contributions to Jointly Authored Works Contained in the Research Report

No jointly-authored works.

Statement of Contributions by Others to the Research Report as a Whole

No contribution by others.

Statement of Parts of the Research Report or Submitted to Qualify for the Award of Another

Degree

None.

Published Works by the Author Incorporated into the Research

None.

Additional Published Works by the Author Relevant to the Research Report or but not Forming Part of it

None

Acknowledgements

I thank Assoc. Prof. Elizabeth Krenske and Prof. Alan Mark for the supervisions throughout the year. I also wish to thank Dr. Martin Stroet and Dr. Thomas Lee for their patience in answering my questions and guidance provided in molecular dynamics simulations on Wiener, providence of codes that enabled the setup of systems containing non-standard amino acids, simulation of GROMOS systems using graphics processing unit-accelerated AMBER machineries, and the analysis of the results. I appreciate the advice given by Dr. David Poger on the simulation of biomolecular systems under physiological conditions and the assistance provided by Yuk Ping Chin and Dr. Romain Lepage in the finding of transition state structures. I am grateful to Jack Taunton and his colleagues for making the co-crystal structure of BTK with inhibitor **3** from their experimental work publicly available. I thank my other colleagues in both laboratories for making my Honours research experience such interesting and exciting. Computer resources provided by

the National Facility of the National Computational Infrastructure (Australia) through the National Computational Merit Allocation Scheme and by the University of Queensland Research Computing Centre are greatly appreciated. I am grateful for the financial support given by the School of Chemistry and Molecular Biosciences through the Honours scholarship.

Signature of Author: _____ **Date:** _____

Principal Supervisor Agreement

I have read the final report and agree with the student's declaration.

Signature of Principal Supervisor: _____ **Date:** _____

ABSTRACT

Computational studies on the reactivities of cyanoacrylamide-based Bruton's Tyrosine Kinase (BTK) inhibitors have been performed through quantum mechanical (QM) calculations and molecular dynamics (MD) simulations in this work. The aim was to understand the factors that determine the kinetics and thermodynamics of thiol-Michael addition or elimination in the enzyme active site. Benchmarking results of ten combinations of functionals and basis sets showed that the density functional theory method M06-2X/6-31+G(d)//M06-2X/6-311+G(d,p) best predicted the energetics of thiol-Michael additions to cyanoacrylamides. An analysis of the low-energy conformations of the Michael acceptors (MAs) revealed the essential role played by the bulkiness of the warhead in determining the most favourable geometry. The investigation of the noncovalent interactions within the molecules provided evidence that the *syn* geometries observed in the transition states (TSs) of the thiol-Michael additions are stabilised by weak attractive intermolecular interactions. The QM calculations correctly reproduced the rankings of the elimination barriers and also revealed correlations between the addition barriers with electronic descriptors including the lowest unoccupied molecular orbital energy of the MA, the earliness/lateness of the TS, and the distortion energy of the MA. Analysis of MD simulation trajectories has led to the proposal that one or more arginines may act as the base that catalyses the Michael reactions in the BTK active site. The reaction profiles computed with a model arginine as the base gave better agreement with the experimentally observed trend in thiol elimination barriers, with only one inhibitor (**R5**) not falling close to the trend line. The formation of enolate intermediate was found to be the rate determining step for the retro-Michael additions except **P3**. The comparison of Boltzmann averaged energies reported that the binding of the inhibitors **1** and **3** to BTK induces 3.4-4.1 kcal/mol and 5.6-5.8 kcal/mol of strain energy, respectively, which

would be compensated for by interactions with the active site residues. The flexibility of the inhibitors to rotate about the C=C-C=O bond of the inhibitor within the binding site were identified as potential influences on the rates of thiol addition. Taking these environmental effects into account, the irreversible acrylamide inhibitor **5** was predicted to be the most kinetically reactive while the cyanoacrylamide inhibitor **3** is the most inert overall.

TABLE OF CONTENTS

ABSTRACT	v
TABLE OF CONTENTS.....	VII
LIST OF ABBREVIATIONS	IX
1 INTRODUCTION	1
1.1 BACKGROUND AND SIGNIFICANCE.....	1
1.1.1 <i>Covalent Drugs</i>	1
1.1.2 <i>Residence Time and its Relationship to Pharmacological Effects</i>	3
1.1.3 <i>Discovery of Reversible Covalent Inhibitors of Bruton's Tyrosine Kinase</i>	4
1.1.4 <i>Questions Emerging from Previous Work</i>	8
1.1.5 <i>Previous Computational Studies of BTK</i>	10
1.2 OBJECTIVES.....	11
2 QUANTUM CHEMICAL CHARACTERISATION OF THE INTRINSIC REACTIVITIES OF COVALENT INHIBITORS... 	12
2.1 METHODS.....	12
2.1.1 <i>Choice of Michael Acceptors</i>	12
2.1.2 <i>Properties of Interest</i>	13
2.1.3 <i>Conformational Sampling</i>	14
2.1.4 <i>Factors Influencing Choice of QM Methods</i>	15
2.1.5 <i>Calculation of Gibbs Free Energy</i>	16
2.2 RESULTS AND DISCUSSION.....	18
2.2.1 <i>Benchmarking of Functionals and Basis Sets for Thiol Additions to Cyanoacrylamides</i>	18
2.2.2 <i>Conformational Analysis of Critical Point Species along the Reaction Coordinates</i>	20
2.2.3 <i>Calculations of Thermodynamic Quantities of Thiol-Michael Additions of MeSH to the Acrylamide Inhibitors</i>	24
2.2.4 <i>Rationalisation of the Calculated Thiol-Michael Addition Barriers</i>	27
2.2.5 <i>Comparison between Predicted and Observed Kinetic and Thermodynamic Parameters</i>	31
3 INVESTIGATIONS OF THE BINDING OF COVALENT INHIBITORS TO BTK	33
3.1 METHODS.....	33
3.1.1 <i>General Aspects</i>	33
3.1.2 <i>Preparation of the System</i>	33

3.1.3	<i>Setup for MD Simulations</i>	35
3.1.4	<i>Trajectory Analysis</i>	35
3.2	RESULTS AND DISCUSSION	36
3.2.1	<i>Identification of the Potential Base Catalyst</i>	36
3.2.2	<i>Conformational Changes that Occur Upon Binding of Cyanoacrylamides 1 and 3 to BTK</i>	43
3.2.3	<i>Trajectory Analysis of BTK Noncovalently Bound to the Inhibitors</i>	44
4	CONCLUSIONS AND FUTURE DIRECTIONS	48
4.1	CONCLUSIONS	48
4.2	FUTURE DIRECTIONS	49
5	REFERENCES	51
6	APPENDIX	60
6.1	BENCHMARKING OF FORCE FIELDS FOR CONFORMATIONAL SAMPLING	60
6.2	DETAILS OF PARAMETERS USED	61
6.2.1	<i>Conformational Sampling</i>	61
6.2.2	<i>NCI Plots</i>	61
6.2.3	<i>Preparation of GROMOS Simulation Systems</i>	61
6.2.4	<i>MD Simulation using AMBER Programs</i>	62
6.3	PROGRAMMING SCRIPTS WRITTEN	62
6.3.1	<i>Gaussian Job Generation and Submission</i>	63
6.3.2	<i>Management and Modification of Files and Directories</i>	63
6.3.3	<i>Post-Calculation Correction, Tabulation, and Visualisation of QM Calculation Results</i>	63
6.3.4	<i>Automated Analysis of MD Trajectories</i>	63
6.3.5	<i>Preparation of MD Systems and Visualisation of MD Trajectory Analysis Results</i>	64
6.4	PLOTTED FIGURES FROM DISTANCE ANALYSES	64
6.4.1	<i>Potential Base for Thiol Additions</i>	64
6.4.2	<i>Potential Base for Thiol Eliminations</i>	66

LIST OF ABBREVIATIONS

AMBER	Assisted Model Building with Energy Refinement
Arg	Arginine
ATB	Automated Topology Builder
BTK	Bruton's Tyrosine Kinase
CM5	Charge Model 5
CPCM	Conductor-like Polarisable Continuum Model
CPU	Central Processing Unit
Cys	Cysteine
DFT	Density Functional Theory
EA	Electron Affinity
GBSA	Generalised Born model augmented with the hydrophobic solvent accessible Surface Area
GGA	Generalised Gradient Approximation
GROMOS	GROningen MOlecular Simulation
HF	Hartree-Fock

HOMO	Highest Occupied Molecular Orbitals
HSAB	Hard-Soft Acid-Base
IP	Ionisation Potential
LUMO	Lowest Unoccupied Molecular Orbitals
MA	Michael Acceptor
MAD	Mean Absolute Deviation
MD	Molecular Dynamics
MM	Molecular Mechanics
MTLMS	Mixed Torsional/Low-Mode Sampling
NBO	Natural Bond Orbital
NCI	NonCovalent Interaction
OPLS	Optimized Potentials for Liquid Simulations
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PES	Potential Energy Surface
PME	Particle Mesh Ewald

QM	Quantum Mechanics
QTAIM	Quantum Theory of Atoms in Molecules
RDS	Rate Determining Step
RMSD	Root-Mean-Square Deviation
RT	Residence Time
SAR	Structure-Affinity Relationships
SCS	Spin-Component-Scaled
SKR	Structure-Kinetic Relationships
SPC	Simple Point-Charge
SPE	Single Point Energy
TS	Transition State
vdW	van der Waals
VMD	Visual Molecular Dynamics

1 INTRODUCTION

1.1 Background and Significance

1.1.1 Covalent Drugs.

The ability of a small molecule, such as an inhibitor, to modify the activity of a target enzyme is largely dependent on the strength of the interaction formed between the molecule and the enzyme. Covalent inhibitors are molecules that inhibit their target proteins by forming covalent attachments to them.¹ They complement conventional (noncovalent) inhibitors by enabling the achievement of much higher binding affinities to their targets.² Such capability opens up the possibility for lower drug dosage and dose frequencies in the treatment of disease, potentially allowing covalent drugs to have safety profiles, provided that an acceptable target selectivity is achieved.³

The structures of several common covalent drugs are shown in Figure 1.1. As illustrated by these examples, covalent inhibitors typically bear a reactive electrophilic functional group, called a warhead, which forms a covalent bond with a nucleophilic residue,⁴ such as cysteines, serines, and threonines, at the target site. The proper positioning of the molecule is enabled by the noncovalent interactions with the target protein, which also contribute to the overall binding interaction. The selectivity of inhibition can be enhanced by targeting noncatalytic residues,⁵ which often varies across different enzymes.^{3,6}

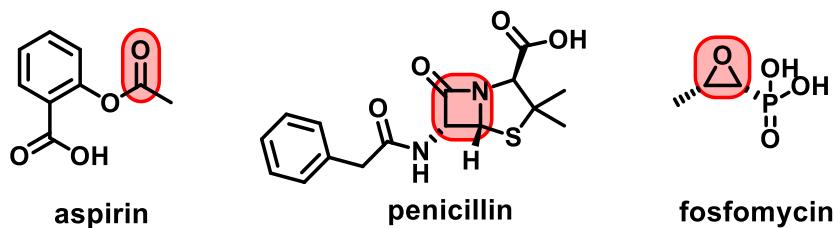
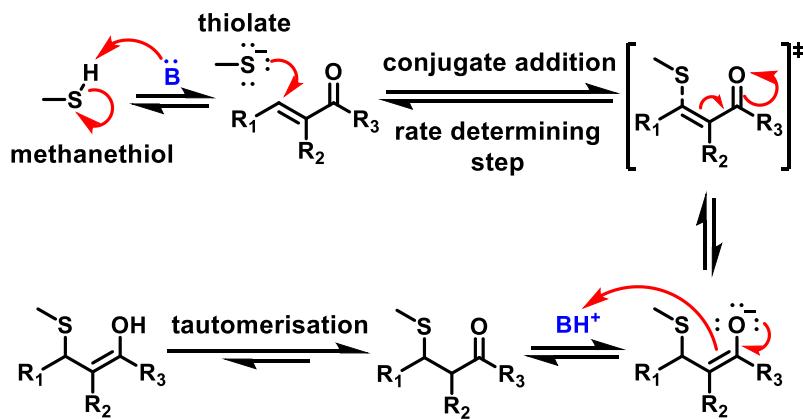


Figure 1.1. Structures of several well-known covalent inhibitors. The bond-forming functional groups are highlighted in red.

One important class of warheads used in covalent drugs are α,β -unsaturated carbonyl compounds, also known as Michael acceptors (MAs). The mechanism of action of a MA as a covalent drug involves conjugate addition and is exemplified in Scheme 1.1 with methanethiol being the nucleophile. Once the inhibitor enters the binding site, a deprotonated thiol group adds to the β position of the MA. This forms an enolate intermediate which is then protonated to form the corresponding thiol adduct. For both thiol addition and elimination, a base species acts as a catalyst by deprotonating the thiol or adduct. The stepwise first-order elimination from a conjugate base (E1cB) mechanism was supported by previous theoretical and experimental studies.^{7,8}

Scheme 1.1. General Reaction Scheme for Thiol Addition to a Michael acceptor.



Despite the advantages of well-known covalent inhibitors such as aspirin,⁹ penicillin,¹⁰ and fosfomycin,¹¹ few covalent drugs were designed deliberately to act as covalent inhibitors, due to the risk of side effects.³ The formation of covalent bonds often, albeit not always, corresponds to an equilibrium lying far towards the covalent adduct, leading to essentially irreversible enzyme modification, which has risks including severe idiosyncratic drug reactions such as toxic epidermal necrolysis,^{12,13} drug-induced liver injury,^{14,15} and haematological disorders.¹⁶ Reversible inhibitors are normally favoured over their irreversible counterparts in drug design as they are less likely to form permanent adducts with off-target proteins containing homologous or highly-reactive

residues.¹⁷⁻¹⁹ Furthermore, reversible inhibition also allows fine tuning of a critical parameter known as the drug-target residence time (RT), τ ,²⁰ the importance of which has emerged over the last decade. RT is defined as the reciprocal of the off-rate constant k_{off} , as shown in equation (1). The varying of this feature allows for tailoring of the duration of the effect of a reversible drug.

$$\tau = 1/k_{off} \quad (1)$$

1.1.2 Residence Time and its Relationship to Pharmacological Effects.

Traditionally, the optimisation of drug-target interactions has been determined by the equilibrium-derived binding parameters, such as the equilibrium dissociation constant, K_d and the half-maximal inhibitory concentration, IC_{50} .²⁰ However, unlike *in vitro* experiments, the dynamic blood flow *in vivo* induces fluctuations in the concentration of unbound inhibitors over time. This prevents the system from arriving at equilibrium and thus renders the physiological environment an open system.^{21,22} Due to the fundamental differences between open and closed systems, binding kinetic parameters, especially RT, have grown in importance as optimisation measures for drug candidates alongside binding affinity and potency.²⁰⁻²⁶ Considering the fact that the efficacy of a drug originates from its interaction with its physiological target, a strong correlation between the length of time a drug remains bound to its target and its clinical efficacy is expected.^{20,21,25} For example, Swinney reported that therapeutics with long RTs demonstrated good clinical efficacy and potentially involved non-equilibrium conditions in their mechanism of action.²⁷ However, it should be emphasised that a long RT is not always required or desired. While sustained target engagement is required for the treatment of cancers,²⁴ allergies,²⁴ hypertension,²⁸ and hormone-dependent diseases,²⁹ proteins that undergo fast turnover through resynthesis do not benefit much from the prolonged duration of action of long RT drugs. Besides, these inhibitors are unsuited for therapeutic applications where rapid dissociation is preferred,^{24,30} such as thrombosis^{20,31} and

antipsychosis.²¹ In fact, long RTs may even be a contradiction for these clinical indications due to the convolution of the effect of drug occupancy at the same site on both efficacy and toxicity.²⁴ As such, the ability to tune the drug-target RT has been a major focus within the covalent drug discovery community.

1.1.3 Discovery of Reversible Covalent Inhibitors of Bruton's Tyrosine Kinase.

In order to counteract the abovementioned dangers of irreversible binding of covalent inhibitors to enzymes, new classes of covalent inhibitors that engage their targets reversibly have been designed. In this context, Taunton et al. reported an important discovery by designing reversible covalent inhibitors that targeted noncatalytic cysteine residues in Bruton's Tyrosine Kinase (BTK).⁵ BTK is a member of the Tec tyrosine kinase family which participates in immune function regulation through B-cell development.^{32,33} Its inhibition has been utilised for treatment of various cancers^{34,35} and autoimmune diseases.^{36,37} Ibrutinib³⁸ (Figure 1.2) was the first BTK inhibitor approved by the US Food and Drug Administration as a medicament for chronic lymphocytic leukaemia,³⁴ and is also used to treat mantle cell lymphoma³⁵ as a second-line treatment.³⁹ Ibrutinib consists of a kinase-recognition scaffold (blue) linked to an acrylamide warhead (red) that is capable of forming covalent bonds with appropriate nucleophiles.

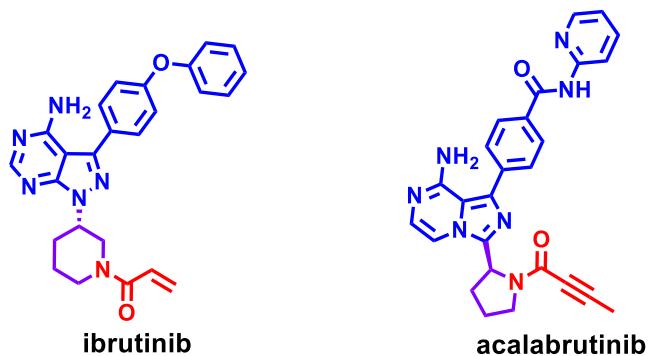


Figure 1.2. Structures of ibrutinib and acalabrutinib. The MA moieties, linker, and the kinase-recognition scaffolds are highlighted in red, purple, and blue, respectively.

However, ibrutinib does not have optimal selectivity. It led to permanent inhibition of analogous kinase targets^{37,40} such as epidermal growth factor receptor, interleukin-2-inducible T-cell kinase and T-cell X chromosome kinase, causing severe adverse events including atrial fibrillation, major haemorrhage, and arthralgia.^{35,41} To date, acalabrutinib⁴² is the only other approved inhibitor of BTK. Even though it exhibits better kinase selectivity compared to ibrutinib,^{43,44} the irreversible nature of its binding interaction⁴⁴ naturally triggers concerns regarding its as-yet unknown off-target effects. This is especially important for the treatment of autoimmune diseases, which typically requires long-term medication.⁴⁵ Accordingly, this prompted the exploration of reversible inhibition of BTK as an alternative solution.

The inhibitor molecules investigated by Taunton et al. resembled ibrutinib except that cyanoacrylamides were used in place of acrylamides as warheads. This modification was based on observations made in earlier chemical experiments, where the installation of the nitrile groups at the α position of acrylamides resulted in rapidly reversible Michael additions.⁶ Quantum mechanical (QM) calculations revealed that the reversibility arises because the electron-withdrawing α -CN substituent both stabilises the anionic transition state (TS) and destabilises the neutral adduct, thus decreasing the change in free energy, ΔG for binding and accelerating the reverse reaction (elimination of thiol) as illustrated in Figure 1.3.⁴⁶

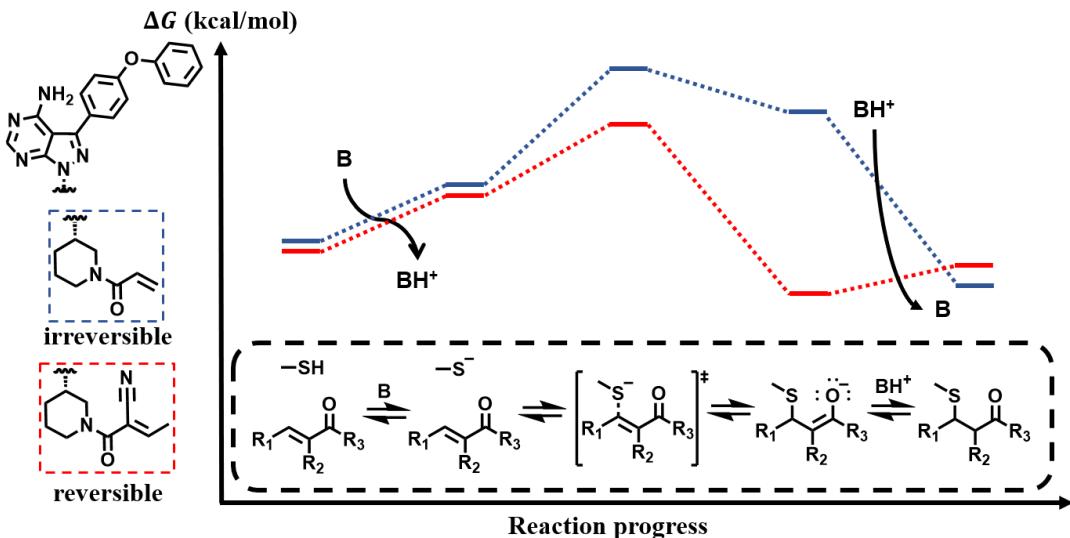
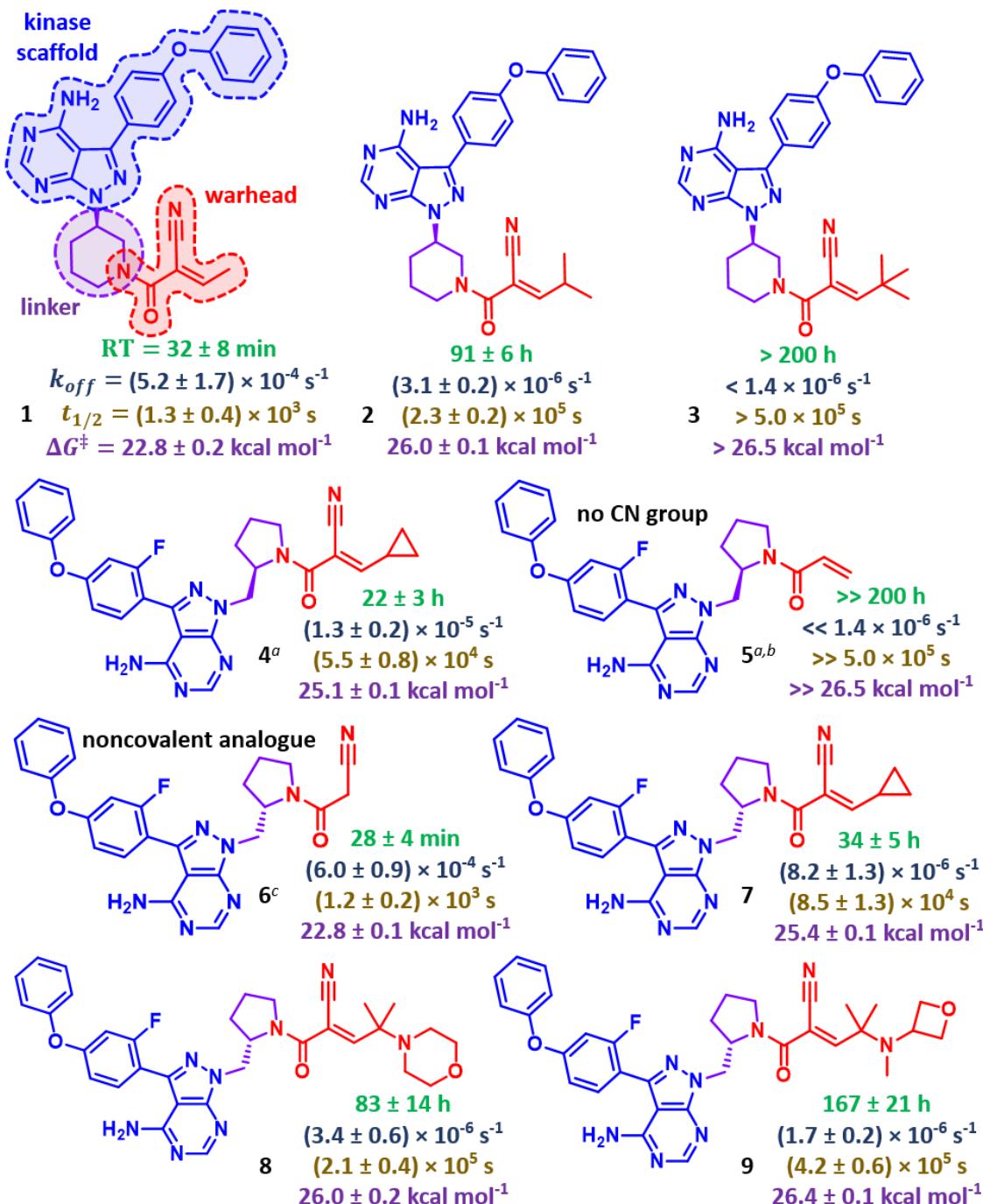


Figure 1.3. Representative ΔG profiles for irreversible and reversible Michael additions.

By varying the electronic and steric environment around the warhead using different β -substituents on the cyanoacrylamides, Taunton et al. developed inhibitors exhibiting RTs ranging from just minutes to one week. Key examples are presented in Chart 1.1. Significantly, long RT was found to translate into durable *in vivo* pharmacodynamic inhibition, where the longest-lived inhibitor, **9**, demonstrated sustained BTK occupancy even after clearance of unbound drugs from the systemic circulation in rodent models. In a recent review article, Copeland⁴⁷ commented that the achievements of Taunton's group unlocked the possibility of systematically approaching the holy grail of covalent inhibitor design, i.e. addressing the requirement to specifically inhibit targeted proteins to treat diseases through rational tuning of the structural features of the inhibitors and their binding interactions. Following Taunton's pioneering work, the incorporation of structure-kinetic relationships (SKR), distinct from the traditional structure-affinity relationships (SAR), into cell-based assays has been encouraged owing to the increasing awareness of the importance of binding kinetics in drug discovery.^{26,47-58} Many research groups have followed up on these ideas to advance the exploration of kinase inhibition by covalent means.⁵⁹⁻⁶⁸

Chart 1.1. Structures and Binding Data of Reversible Covalent Inhibitors of BTK Designed by Taunton et al.⁵



^a**4** and **5** have *R* linker configurations instead of *S*. ^b**5** contains an acrylamide electrophile instead of cyanoacrylamide. ^c**6** does not contain a MA and was included for comparison.

1.1.4 Questions Emerging from Previous Work.

It is unclear how the different structural components of cyanoacrylamides govern the reversibility of their binding to kinases such as BTK. The overall reversibility of binding depends on a combination of multiple effects, including the overall ΔG for the addition and the rate constants for binding and release. These properties are also fundamentally important to the selectivity of an acrylamide across different kinases. For example, **9** has also been found to show exceptional selectivity for BTK over other kinases, yet the molecular basis for the selectivity is yet to be uncovered.

In a computational study, Taunton et al. discovered⁶⁹ a strong inverse correlation between the β -elimination rates of thiol adducts and the proton affinities of the intermediate enolate carbanions ($R^2 = 0.96$). This suggested that it may be possible to use computed proton affinities to predictably tune the intrinsic reversibility of acrylamide thiol additions. In spite of this, examination of the experimental data in Chart 1, especially the comparison between the RT of diastereomers **4** and **7**, reveals that acidity is not the only important criterion that determines the β -elimination rates; other factors such as steric hindrance, the impact of substituents on the overall electrophilicity, and, importantly, interactions with the BTK binding site, must also contribute to determining the overall behaviour of a given acrylamide. Unravelling the factors that determine the RT of acrylamide inhibitors is expected to be useful for understanding not only BTK inhibition but also for designing inhibitors of related enzymes containing noncatalytic cysteines and predicting the selectivity of a given inhibitor against a range of kinases.

Prior to their computational work, Taunton's group had already proposed several other hypotheses to explain the observed trends of drug-target RT based on the examination of a co-crystal structure of BTK, with its Cys481 covalently bonded to **3** as depicted in Figure 1.4. They

identified two hydrophobic patches in the vicinity of the *tert*-butyl group of **3** and claimed that the hydrophobic interactions contributed to the stabilisation of the covalent adduct. The alignment between the C=O π bond and the C α -H bond was perpendicular, which led Taunton et al. to propose that the acidity of the hydrogen was lowered, both thermodynamically and kinetically.⁵ Furthermore, it was suggested that the C α hydrogen, which needs to be removed for the elimination of thiol, was shielded from potential base species as a result of the orientation of the capping group and piperidine amide to minimise torsional strains and steric clashes. However, the base catalyst was unidentified. If the base is an external one, the solvent-exposed C α hydrogen would in fact be readily accessible to it, leaving the argument of Taunton et al. on shaky ground.

It should be noted that X-ray structures merely present an average, static representation of what is really a dynamic, structurally diverse ensemble in the crystal.^{70,71} This limitation means that the X-ray structure should be interpreted with some caution and other measures should be employed to better understand the protein dynamics.^{72,73} Simulations at an atomistic level would accomplish this goal.

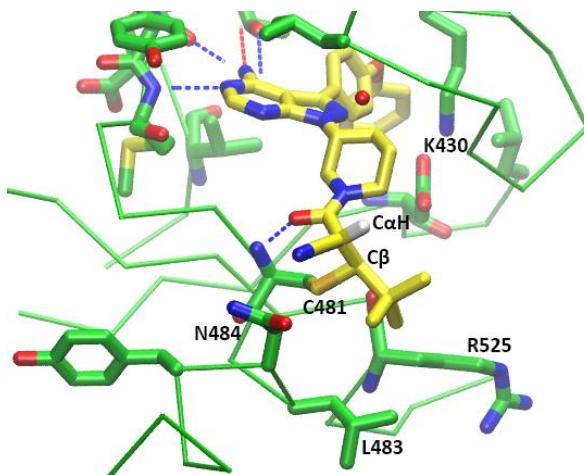


Figure 1.4. Crystal structure of BTK bound to the reversible covalent inhibitor **3** (2.2 Å resolution).⁵ The covalent bond between C β and Cys481, the acidic C α hydrogen, and the hydrogen bonds that possibly contribute to stabilising the covalent adduct are indicated.

1.1.5 Previous Computational Studies of BTK.

Numerous computational studies have been performed on BTK, in many cases using molecular dynamics (MD) simulations. For instance, a combination of MD simulations and 3D quantitative structure-activity relationship models have been used to screen noncovalent BTK inhibitors based on binding affinities.⁷⁴⁻⁷⁸ Related MD studies have helped to improve the understanding of the biological functions of different domains of BTK,^{79,80} the features involved in BTK changing from an inactive to an active state⁸¹ and the interactions of BTK with its natural substrates, phosphatidylinositols.^{80,82} To date, only a single study associated with BTK has utilised QM calculations.⁷⁴ That work employed density functional theory (DFT) to calculate the electronic excitation energy of BTK-targeting inhibitors. The functional and basis set combination chosen for the geometry optimisations of the detected hits was the B3LYP/6-31G*, which is notorious for its poor treatment of London dispersion.^{83,84} Fortunately, this suboptimal choice was somewhat attenuated by the fact that the study only focused on the structures of the inhibitors and not their TSs, thus avoiding problems associated with spurious charge-transfer complexes. The limitations of the standard DFT methods in the modelling of thiol additions will be discussed further in the following methodology section. Taken together, however, the body of computational work reported thus far illustrates the many possibilities for gaining atomistic insight into BTK inhibition using appropriate computational tools.

1.2 Objectives

The ultimate goal of the present work is to elucidate the factors that affect the kinetics and thermodynamics of thiol additions in the context of the binding of cyanoacrylamide-containing MAs to Cys481 of BTK. The intention is to perform the first ever study of BTK binding site environmental effects on the reactivities of these inhibitor molecules. The goal was approached in the following way:

- performing QM calculations to determine the intrinsic reactivities of different cyanoacrylamide warheads towards a model thiol, and
- conducting MD studies to explore the impact of residues near BTK active site on the binding of the inhibitors.

It is hoped that a deeper understanding of the binding interactions will enable more rational design of BTK inhibitors in the future. Such an understanding could also facilitate the design of reversible covalent inhibitors for other kinases. Given the importance of the tunability of *in vivo* RT, the results of these computations are anticipated to contribute to the advancement of the still maturing field of protein kinase covalent inhibitor discovery.^{50,85,86}

2 QUANTUM CHEMICAL CHARACTERISATION OF THE INTRINSIC REACTIVITIES OF COVALENT INHIBITORS.

The following chapter aims to study the intrinsic reactivities of the acrylamide inhibitors investigated by Taunton through QM calculations. The topics covered include conformational analysis of the relevant species, benchmarking of QM methods against similar compounds, calculation of thermodynamic and kinetic parameters using the chosen method, and investigation on the correlations between some molecular properties and the calculated addition barriers.

2.1 Methods

2.1.1 *Choice of Michael Acceptors*

From Taunton's set of inhibitors, six examples were selected for detailed study to reveal how structural factors determine the intrinsic reactivities of acrylamide warheads (Figure 2.1).

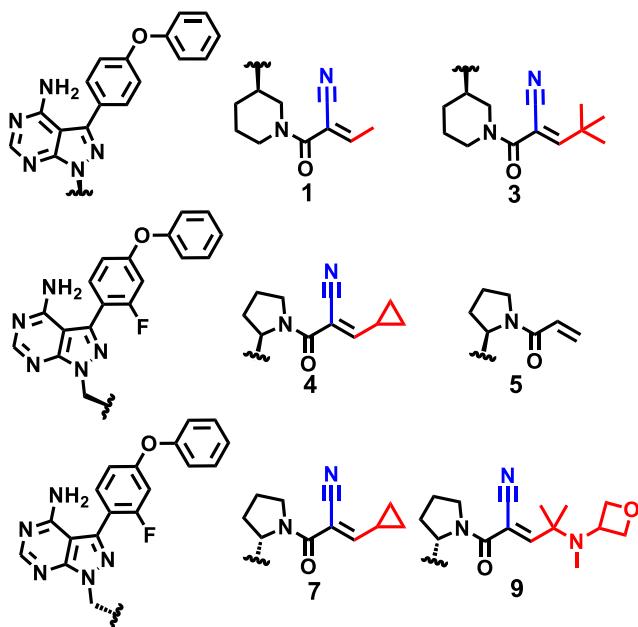


Figure 2.1. The acrylamide inhibitors chosen for detailed study.

First, acrylamide **3** was chosen due to the availability of its co-crystal structure with BTK. The related acrylamide **1**, with a methyl group in place of the *tert*-butyl group of **3**, was chosen to

be compared with **3**. The pair of diastereomers **4** and **7** were next chosen to explore the effect of the different configurations of the stereogenic centre in the linkers. Inhibitor **9** was chosen as it displayed the longest RT out of those studied by Taunton. Finally, the acrylamide **5**, lacking a cyano group, was included as a control, as it exhibited irreversible BTK inhibition. As the computational expenses grow exponentially with increasing number of electrons, the scaffolds of the inhibitors were truncated to piperidine groups. Their effects on the intrinsic reactivity are likely to be minimal, due to their remoteness from the electrophilic centre. The truncated inhibitors will be differentiated from the original molecule by adding the prefix **R** to the compound number, e.g. **R3** is the model for **3**. As a special case, the truncated forms of **4** and **7** are identical and are labelled **R4(7)**. The thiol was modelled as methanethiol (MeSH) for the sake of simplicity.

2.1.2 Properties of Interest

The application of statistical mechanics to QM systems allows the derivation of thermodynamic quantities including Gibbs free energy, G from the partition functions.⁸⁷ The free energy barrier (ΔG^\ddagger) and free energy change (ΔG) for a thiol-Michael addition or elimination can therefore be estimated from QM calculations on reactants, TS and products. The rate constant for thiol elimination (off-rate), k_{off} , can be computed using equation (2), the Eyring equation.⁸⁸

$$k_{off} = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G_{reverse}^\ddagger}{RT}} \quad (2)$$

where κ is a transmission coefficient, which reflects the fraction of the molecules overcoming the activation barrier that proceeds to the product without recrossing the TS, k_B is Boltzmann's constant, T is absolute temperature, h is Planck's constant, $\Delta G_{reverse}^\ddagger$ is the elimination activation barrier calculated by DFT, and R is gas constant, respectively. Assuming that the no-recrossing assumption of the TS theory⁸⁸ is held, a value of 1 is typically assigned to κ . The calculated results

can be compared with the experimentally measured half-life, $t_{1/2}$, which is related to k_{off} by a factor of $\ln(2)$ for first order reactions as shown in equation (3):

$$t_{1/2} = \tau \ln(2) \quad (3)$$

The ΔG , ΔG^\ddagger , k_{off} and $t_{1/2}$ values were calculated for thiol additions to Taunton's cyanoacrylamide warheads (Figure 2.1).

2.1.3 Conformational Sampling

The reactivity of a molecule is highly dependent on its geometry. Since molecules at any temperature constantly undergo changes in their molecular geometries, e.g. through vibrations, measured properties are an average over the ensemble of all possible states. The properties are predominantly determined by the lowest energy conformers as the conformational population follows a Boltzmann distribution as shown in equation (4):

$$\frac{N_i}{N_{\text{total}}} = \frac{e^{-E_i/RT}}{\sum_{k=1}^M e^{-E_k/RT}} \quad (4)$$

where N_i is the expected number of particles in the single-particle microstate i , N_{total} is the total number of particles in the system, E_i is the energy of microstate i , R is gas constant, T is the equilibrium temperature of the system, and M is the total number of microstates. As such, the most important conformers for the key species involved in a reaction, namely reactants, TS structures and products, need to be identified prior to the calculation of the thermodynamic parameters.

The conformational searching essential to the identification of the ground state conformers was guided by molecular mechanics (MM) to enable efficient exploration of the large conformational space. This was done for each species using the mixed torsional/low-mode sampling (MTLMS) method in the MacroModel software,⁸⁹ which has been demonstrated to be an effective sampling method for flexible compounds.⁹⁰ For thiol adducts, all possible

diastereomers were constructed and the lowest energy diastereomer was used for analysis. For TSs, *syn* (C-S-C=C dihedral angle < 90°) and *anti* (C-S-C=C dihedral angle > 90°) conformations for thiolate attack on both *Re* and *Si* faces of the MA were calculated. The molecular interactions were described using the Optimized Potentials for Liquid Simulations Version 3 Extended (OPLS3e) force field.⁹¹ Benchmarking of other force fields was conducted and is discussed in Appendix 6.1. The implicit Generalised Born model augmented with the hydrophobic solvent accessible surface area (GBSA) was employed to simulate infinitely diluted aqueous solution. The S-C bonds to be formed in TS were constrained during the samplings. Full details are included in Appendix 6.2.1.

2.1.4 Factors Influencing Choice of QM Methods

Historically, the first computational studies of thiol additions to acrolein using *ab initio* methods were performed by Kollman et al. in 1995.⁹² Subsequently, several related studies⁹³⁻⁹⁵ have been conducted using DFT methods instead due to their more favourable computational cost-to-accuracy ratio.⁹⁶ A critical observation was made by Smith et al.⁸³ who noted that several popular DFT functionals give spurious results for thiol-Michael additions, due to delocalisation errors, where the inconsistent correction of the self-repulsion in the dominant Coulomb functional leads to an artificial spreading of the electron density.⁹⁷ In fact, artefactual charge-transfer complexes stabilised by this excessively delocalised electron density had led other researchers unaware of the problem to propose alternative mechanisms for thiol addition.^{98,99} Proper selection of a QM method that could mitigate the delocalisation errors is therefore critical to gaining reliable insights into the intrinsic properties of thiol-Michael reactions.

To choose a suitable combination of functional and basis set for the single point energy (SPE) calculations, a benchmarking study of ten methods was carried out, as shown in Table 2.1. Range-separated functionals,¹⁰⁰⁻¹⁰³ which incorporate a greater proportion of Hartree-Fock (HF)

exchange at long-range while maintaining the typical generalised gradient approximation (GGA) functional at short distance, have been proven to be effective in correcting the aforesaid error. This was witnessed in a gas phase study of thiol additions to electron deficient olefins, where ω B97X-D/aug-cc-pVTZ returned highly accurate predictions compared to CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ results.⁸³ The methods SCS-MP2/6-31+G(d), B2PLYP-D/6-31+G(d) and M06-2X/6-311G(2d,p) had also performed well in a solution phase study of thiol additions to acrylamides, as judged by comparison to high level *ab initio* CBS-QB3 calculations.¹⁰⁴ Later, M06-2X/6-311+G(d,p) was found to best describe the energetics of thiol addition to cyanoacrylamides in solution compared to experimental data.⁴⁶ Combinations of these functionals with different basis sets were thus tested. It was expected that the performance of the DFT methods (M06-2X and ω B97X-D) would not necessarily scale with the size of the basis set, while the *ab initio* method (SCS-MP2) should increase in accuracy as the complete basis set limit is approached.

Table 2.1. Combinations of functionals and basis sets chosen for benchmarking.

Method	Functional	Basis set	Method	Functional	Basis set
A	SCS-MP2	6-311G(2d,p)	F	ω B97X-D	6-311+G(d,p)
B	B2PLYP-D	6-31+G(d)	G	ω B97X-D	aug-cc-pVTZ
C	B2PLYP-D	6-311G(2d,p)	H	M06-2X	6-311G(d,p)
D	ω B97X-D	6-311G(d,p)	I	M06-2X	6-311G(2d,p)
E	ω B97X-D	6-311G(2d,p)	J	M06-2X	6-311+G(d,p)

2.1.5 Calculation of Gibbs Free Energy

After conformational sampling, low-energy structures within an energy window of 5 kcal/mol were then geometry optimised quantum mechanically and finally ranked according to

their QM energies. The thermodynamic parameters of interest, enthalpy, H , entropy, S and free energy, G , were obtained from the QM vibrational frequencies using the harmonic oscillator approximation. The geometry reoptimisation aimed to rank the energetic stability of the conformers to identify the most stable conformation. Due to the large number of conformers to be optimised, the geometry optimisations were performed using a single QM method known to be capable of capturing the essential components of thiol addition systems at reasonably low computing resource cost. In a previous benchmarking done on related thiol additions to acrylamides,⁴⁶ the M06-2X/6-31+G(d) method had been found to yield rather accurate values of G . This method was therefore employed here for geometry optimisations. Once the most stable conformer for each species was identified at this level, G values at a higher level of theory were then obtained by computing the single point (fixed geometry) E values and adding the thermochemical correction from the lower-level calculation, as shown in equation (5):

$$G_{HL} = E_{HL} + (G_{LL} - E_{LL}) \quad (5)$$

where HL and LL correspond to higher and lower levels of theory, respectively.

The Gaussian 16¹⁰⁵ software was used to carry out all DFT calculations. The solution phase energies were calculated by employing conductor-like polarisable continuum model (CPCM)¹⁰⁶ implicit solvent model due to its better performance in a previously reported benchmarking effort.⁴⁶ The “UltraFine” integration grids was used to avoid large errors in computations of G values as warned by Bootsma and Wheeler.¹⁰⁷ Harmonic vibrational frequency calculations indicated whether stationary points were local minima or first-order saddle points and provided unscaled zero-point energy and thermal corrections. G values are reported at a standard state of 1 mol/L and 25 °C. Programming scripts were written to enable batch generation and submission of job scripts, efficient tabulation of data, and data visualisation, as described in Appendix 6.3.

2.2 Results and Discussion

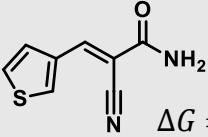
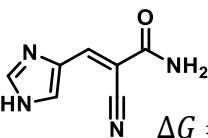
The following subsections discuss the results obtained from the benchmarking conducted, analysis on the most stable conformers, thermodynamic parameters calculated and their correlation with certain features of the molecular systems.

2.2.1 Benchmarking of Functionals and Basis Sets for Thiol Additions to Cyanoacrylamides

First, to determine the best method to compute the energetics of acrylamide thiol additions, the ΔG values for a set of known thiol additions having experimentally available thermodynamic data were calculated using different combinations of functionals and basis sets. The experimental ΔG values of these reference thiol additions were calculated from the equilibrium constant determined using nuclear magnetic resonance.⁶ The benchmarking results are shown in Table 2.2 along with the central processing unit (CPU) time taken for the calculation of a test case.

Table 2.2. Benchmarking calculations of ten different QM methods for the prediction of the ΔG values for addition of thiol (MeSH) to Michael acceptors.

Michael acceptor	Errors of Calculated ΔG (kcal/mol) for Methods A-J									
	A	B	C	D	E	F	G	H	I	J
10 	-0.5	+0.9	+4.5	+2.3	+2.7	+1.8	+3.4	+0.3	+1.2	-0.1
11 	-0.6	+0.7	+4.5	+1.9	+2.5	+1.5	+3.4	+0.4	+1.5	+0.1
12 	-0.4	+1.2	+4.6	+2.2	+2.8	+1.9	+3.8	+0.6	+1.7	+0.2

13		$\Delta G = -2.6^a$	-0.9	+0.7	+4.2	+1.6	+2.2	+1.3	+3.1	-0.1	+1.0	-0.4
14		$\Delta G = -2.0^a$	-0.3	+1.4	+4.6	+2.5	+1.1	+2.2	+4.0	+0.8	+1.8	+0.5
<hr/>												
Mean absolute deviation (MAD)		0.5	1.0	4.5	2.1	2.6	1.7	3.6	0.4	1.5	0.3	
RMSD		0.6	1.0	4.5	2.1	2.6	1.8	3.6	0.5	1.5	0.4	
CPU Time Taken (min) ^b		61	26	58	19	28	37	789	21	30	37	

^aExperimental ΔG values in kcal/mol. ^bCPU time taken for SPE calculation of **10**. The lowest values of the statistical measures and CPU time taken are shown in bold.

Methods A (SCS-MP2/6-311G(2d,p)), H (M06-2X/ 6-311G(d,p)) and J (M06-2X/ 6-311+G(d,p)) were able to predict the ΔG to MAD and RMSD of within 1 kcal/mol while Methods C (B2PLYP-D/6-311G(2d,p)) and G (ω B97X-D/aug-cc-pVTZ) overestimated the ΔG values by 4 kcal/mol and 3 kcal/mol, respectively. The performance results of the other methods are in between these extremes. The benchmarking outcome points toward Method J as the most accurate method in terms of prediction of the addition ΔG for the cyanoacrylamides. Method C performed the worst despite being the third most expensive method in terms of CPU time. It is noteworthy that the inclusion of the diffuse functions for heavy atoms reduces the error of the DFT calculations by 0.2-0.5 kcal/mol, and therefore is encouraged to be included for future studies on similar systems.

Among the three most accurate methods, the CPU time taken for Method J was about 1.5 times longer than Method H. However, since high accuracy is deemed to be more important than time expenditure for this work, Method J was determined to be the method of choice for calculating the thermodynamic parameters of thiol additions to cyanoacrylamides.

2.2.2 Conformational Analysis of Critical Point Species along the Reaction Coordinates

The most stable conformers of the reactants, TSs, intermediates, and products for the Michael addition of MeSH to each model inhibitor are compared. Some comments about the conformational trends and the reasons for these trends are made in the following subsections.

2.2.2.1 Reactants

For **R1**, **R4(7)**, and **R5**, the most stable conformers have s-*cis* geometries while the s-*trans* conformer was most stable for **R9**. The most stable conformer for **R3** is neither s-*cis* nor s-*trans*, but has a C=C-C=O dihedral angle of 90°. The non-planar optimum dihedral angle of 90° in **R3** is thought to minimise the destabilising clashes between the *tert*-butyl group and the C=O group or nitrogen heterocycles. For **R9**, the greater stability of the s-*trans* conformer compared to the s-*cis* was unexpected. Further investigations revealed that the s-*cis* conformers have lower potential energies, but higher *H* and *G* values than the s-*trans* conformer. The other conformers of **R1** and **R4(7)** are higher in energy within 1.4 kcal/mol of the most stable conformers and thus mixtures of s-*cis* and s-*trans* conformers would be expected to be present in solution at body temperature. On the other hand, the low-energy conformers of **R3**, **R5**, and **R9** are largely similar to their most stable conformers and therefore only one predominant form would be sampled under physiological conditions. The trend thus indicates that the preferred conformation is highly dependent on the bulkiness of the β substituent.

Interestingly, the conformational sampling algorithms of MacroModel failed to locate any s-*cis* conformer for **R1** within an energy window of 5.0 kcal/mol when in fact, the manually built s-*cis* conformer was found to be the most stable conformer. This finding reveals a shortcoming in the MacroModel software and should serve as a warning against using standard chemistry

programs as black boxes. It is suspected that the unusually high energies of the *s-cis* conformers are due to the overfitting of the dihedral parameters in OPLS3e.

To understand the conformational preferences, the noncovalent interactions (NCIs) present in the molecules were visualised using the NCIplot software¹⁰⁸ which plots the isosurface of reduced-density gradient and colours it according to electron density (Figure 2.2). The types of the interactions are distinguished based on the sign of the second eigenvalue of the electron density Hessian matrix and are colour-coded, with blue, green and red signifying strong attractive interactions, weak attractive interactions and strong nonbonded overlap, respectively. The *s-cis* conformers are stabilised by the attractive interactions between the α -CN substituent of the MAs and nearby protons on the nitrogen heterocycle. The π electron cloud of the CN group in **R1** and **R4(7)** allows C-H $\cdots\pi$ interactions¹⁰⁹ to be formed while a weaker dispersive interaction¹¹⁰ is found in **R5**. More importantly, the *s-trans* conformers of **R5** are likely more destabilized due to steric clashes.

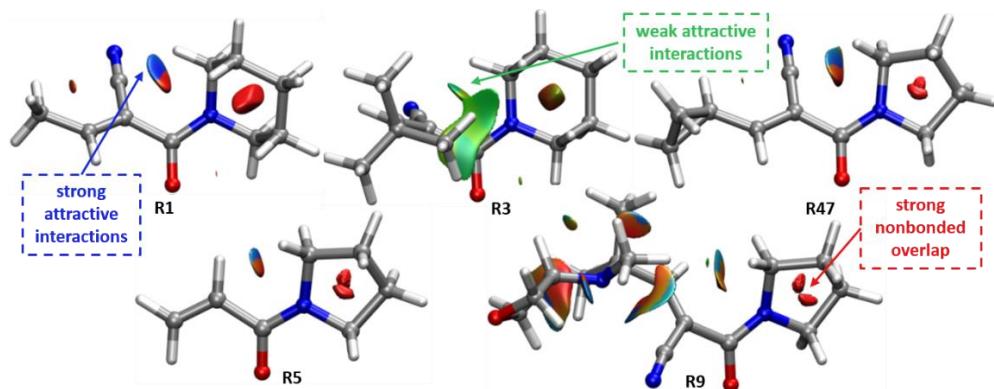


Figure 2.2. Visualisation of the NCIs of the reactants.

2.2.2.2 Transition State Structures

The calculated TS geometries for the additions of MeS^- to the MAs are shown in Figure 2.3. All of the most stable TS conformations had a *syn* geometry, that is, the S-Me bond of the

thiolate is approximately aligned with the C=C bond of the MA. This phenomenon was previously attributed to the repulsion between sulphur lone pairs and the π electron clouds, along with the electrostatic attraction between MeS^- protons and C=O oxygen.¹⁰⁴ An investigation into the NCI within the TS conformations has indeed verified the existence of stabilising C-H $\cdots\pi$ interactions between the MeS^- protons and the π electron system (Figure 2.3).

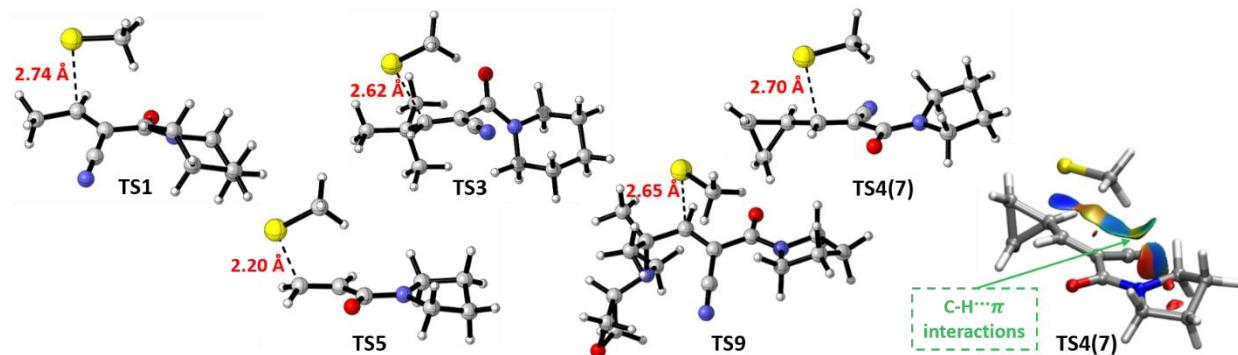


Figure 2.3. Structures of the TSs and the visualisation of the NCI in **TS4(7)** (lower right).

2.2.2.3 *Intermediates*

The most stable geometries of the intermediates of the Michael additions are shown in Figure 2.4. The α -carbon of all intermediate species is sp^2 hybridised as expected. An interesting observation is that one of the methyl protons of the methylthiolate groups is positioned to be relatively closer to the conjugated π electron system, presumably to provide electrostatic stabilisation. The closest CH-C α distance for each intermediate species is found to be 2.7 Å except for **I5** which is 0.1 Å further away. The NCIs in the molecules revealed the existence of stabilising C-H $\cdots\pi$ interactions between the S-Me protons and the π cloud electrons, as exemplified by **I4(7)** (Figure 2.4).

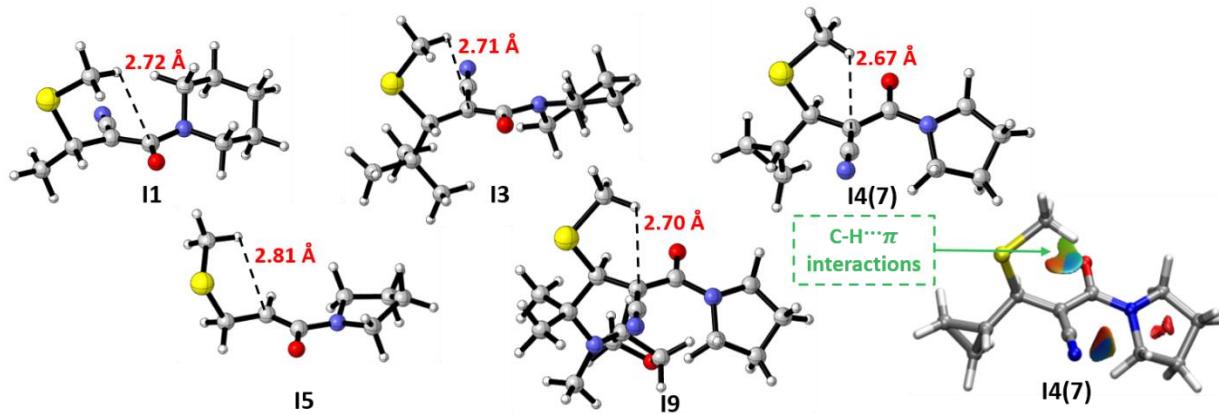


Figure 2.4. Structures of the intermediates and the visualisation of the NCI in **I4(7)** (lower right).

2.2.2.4 Covalent Adducts

The most stable structures of the thiol adducts are shown in Figure 2.5. There are no special interactions between the S-Me groups of most inhibitors and the π electron clouds of the C=O and CN groups, except in **P3**, where the S-Me protons approach the C=O oxygen to interact with the lone pair electrons. At least 6 conformers of each thiol adduct have energies within 1.4 kcal/mol of the lowest energy structures, hence the solution would comprise a mix of these conformations. The most stable geometry of **P3** was next compared to the conformation of **3** observed in the BTK co-crystal structure⁵ by computing the positional root-mean-square deviation (RMSD) of the core atoms (Figure 2.6). The A and B chains of BTK each contained a bound **3**, which superimpose upon each other with an RMSD of 0.09 Å. In comparison, the RMSDs between the bound **3** and the calculated structure of **P3** are much larger, 1.30 Å and 1.27 Å, respectively. An alignment of the calculated and crystallographic structures is shown in Figure 2.6.

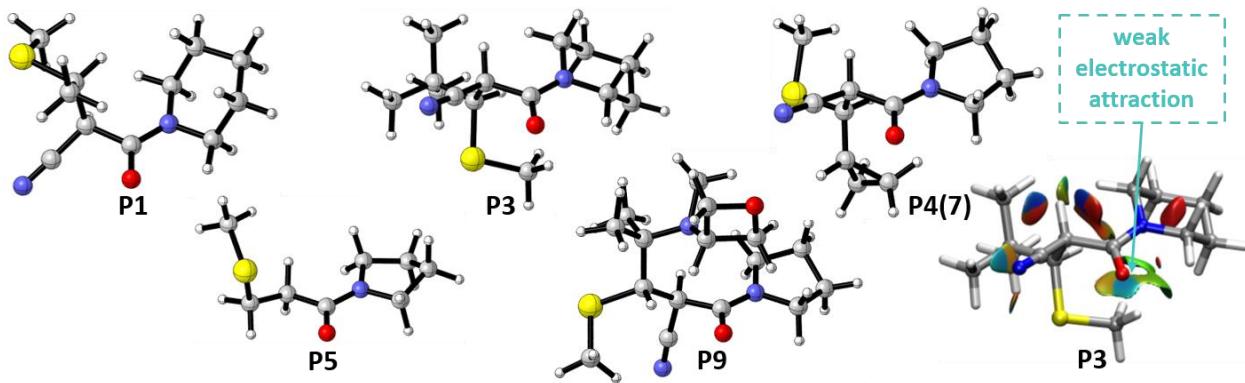


Figure 2.5. Structures of the thiol adducts and the visualisation of the NCI in **P3** (lower right).

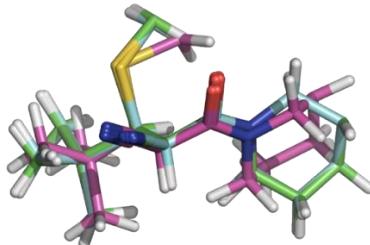


Figure 2.6. Superposition of core atoms of calculated **P3** (pink) and the two conformations of 3 present in the co-crystal structure conformations with BTK (A chain, green, and B chain, blue).

2.2.3 Calculations of Thermodynamic Quantities of Thiol-Michael Additions of MeSH to the Acrylamide Inhibitors

The calculated thermodynamic and kinetic parameters for MeSH additions to the MAs are shown in Table 2.3, along with the corresponding kinetic parameters for dissociation of the inhibitors from their covalent BTK adducts. The G values were obtained using Method J. A comparison between the predicted ΔG_{rev}^\ddagger values and the experimental values is given in Table 4. The identity of the base catalyst in the BTK binding reactions is not known with certainty. For the sake of convenience, methylthiolate was used as a model base to compute the ΔG_{rev}^\ddagger values. Further investigation on the importance of the choice of base will be considered in the next chapter.

Table 2.3. Comparison between calculated and experimental thermodynamic and kinetic parameters for thiol additions to Michael acceptors.

Michael acceptor	Calc. ΔG^a	Calc. $\Delta G^\ddagger a$	Calc. $\Delta G_{rev}^\ddagger a,b$	Expt. $\Delta G_{rev}^\ddagger a$
R1	-4.0	7.8	11.8	22.8 ± 0.2
R3	-2.0	12.8	14.8	>26.5
R4(7)	-2.1	9.0	11.2	25.1 ± 0.1 (4) 25.4 ± 0.1 (7)
R5	-9.0	16.8	25.8	$>>26.5$
R9	-3.6	11.0	14.6	26.4 ± 0.1

^aValues were reported in kcal/mol. ^bValues were calculated using methylthiolate as model base.

The experimental ΔG_{rev}^\ddagger values were calculated from the experimentally measured $t_{1/2}$ using equations (2) and (3). The $t_{1/2}$ values were computed from the dissociation curves tracking the percentage of BTK bound by the inhibitors over 108 hours. As exemplified in Chart 1.1 (page 20), the errors in the experimental ΔG_{rev}^\ddagger values are within 0.2 kcal/mol except for **3** and **5** which exhibited irreversible binding. Despite the same magnitude of the rounded ΔG_{rev}^\ddagger values, no recovery of **5** from its covalent adduct was observed experimentally, in contrast to the gradual recovery of **3**. Consequently, the data as a whole should be interpreted as semi-quantitative (Inhibitors **3**, **5**, and **9** have large ΔG_{rev}^\ddagger , **4** and **7** have medium ΔG_{rev}^\ddagger , while **1** has a small ΔG_{rev}^\ddagger). The relative magnitude and ranking of the ΔG_{rev}^\ddagger values are thus more important than the absolute barriers in the comparison between the calculations and experimental data.

The calculated and experimental ΔG_{rev}^\ddagger values showed identical rankings (Figure 2.7), with only the ranks **R1** and **R4(7)** being reversed in the calculations. This is unexpected to an extent because the QM model systems do not include any of the environmental effects of the BTK binding site. However, a closer examination on the relative magnitudes of the predicted and experimental

ΔG_{rev}^\ddagger values reveals that although the ranks follow approximately the same overall trend as the experiments, the calculated reactivities for some of the MAs differ more from the experimentally-observed reactivities than do others. This indicates that the intrinsic reactivities of the warheads do not fully account for the observed experimental behaviour. The factors not captured in the QM model systems likely comprise a combination of stabilising/destabilising effects that influence different adducts and TSs to different extent. These effects can potentially be determined through simulations of the full BTK-inhibitor systems, which will give an indication of how the binding site residues contribute to these environmental effects. The investigation would thus be continued in the next chapter.

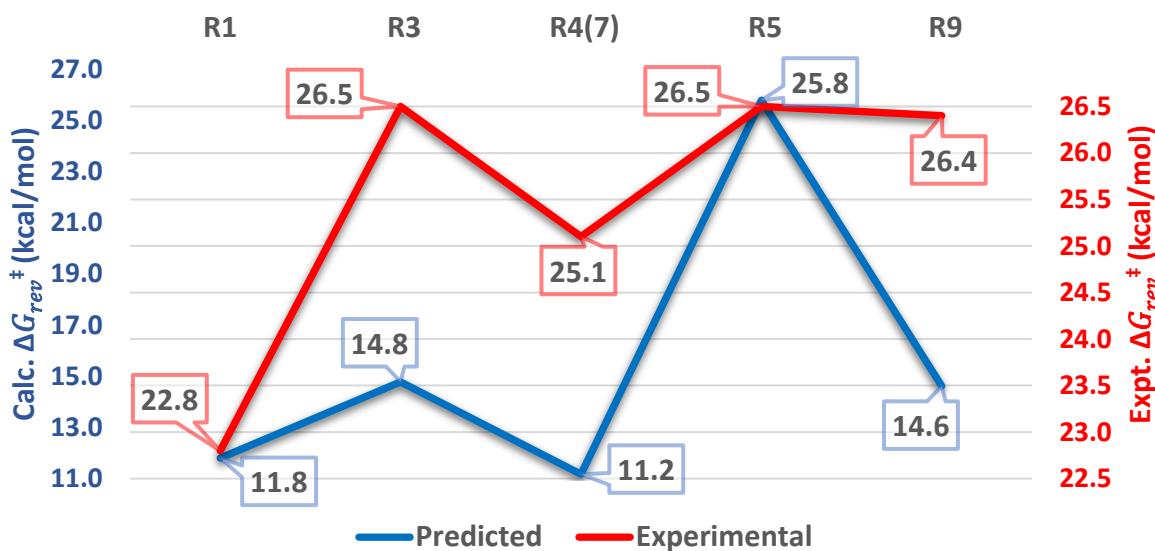


Figure 2.7. Comparison between calculated (primary axis) and experimental (secondary axis) ΔG_{rev}^\ddagger values. Experimental data for **7** ($\Delta G_{rev}^\ddagger = 25.4$ kcal/mol) is omitted for simplicity.

2.2.4 Rationalisation of the Calculated Thiol-Michael Addition Barriers

In order to uncover factors that influence the calculated ΔG^\ddagger values, analyses were performed to identify any correlations between ΔG^\ddagger and certain electronic descriptors of the MAs.

2.2.4.1 Lowest Unoccupied Molecular Orbital Energies

The calculated ΔG^\ddagger values were first compared to the lowest unoccupied molecular orbital (LUMO) energies of the reactants (Figure 2.8). Most inhibitors fall close to the linear relationship, except two MAs, **R1** with the simplest warhead and **R9** with the most elaborate warhead. Overall, a reasonably strong correlation was observed ($R^2 = 0.84$), although it is slightly weaker than what was observed previously in among acrylamide-containing compounds ($R^2 = 0.93$).¹¹¹ It is expected that lower LUMO energies of the inhibitors would correspond to faster thiol additions as the HOMO-LUMO energy gap with the nucleophilic thiolate will be smaller.

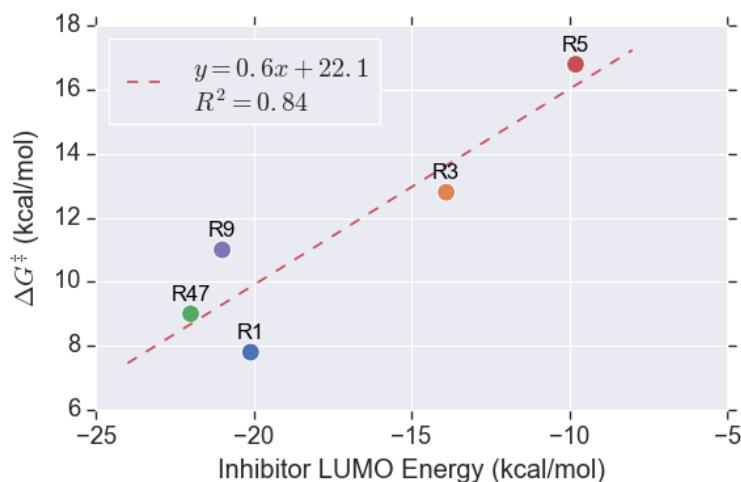


Figure 2.8. Plot of calculated ΔG^\ddagger against LUMO energy of the acrylamide inhibitor.

2.2.4.2 Earliness or Lateness of the Transition States

The distance between the sulfur atom of the methylthiolates and the electrophilic β -carbon of the MA indicates the earliness or lateness of the TS along the reaction coordinate. This could

serve as a measure of how advanced the bond-forming interaction between the reactant molecules must be for the addition to proceed. A longer S-C distance implies that the TS is achieved earlier in the reaction path. The calculated ΔG^\ddagger values are plotted against the S-C distance in Figure 2.9. A strong negative correlation ($R^2 = 0.88$) is found, indicating that the strength of the interaction between the reactants has a strong influence on the addition barrier.

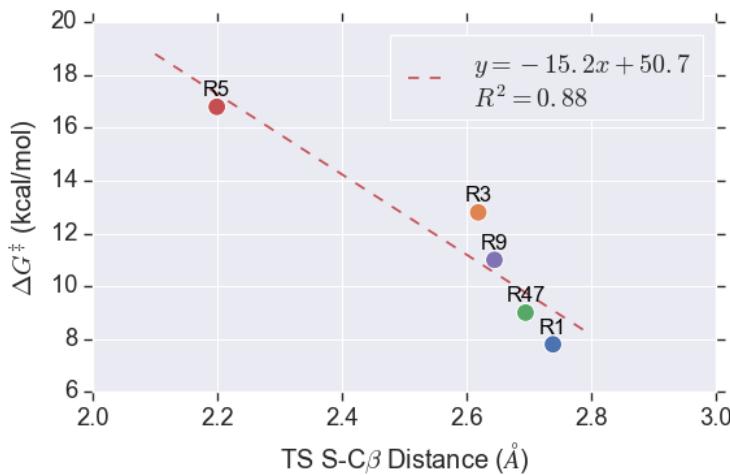


Figure 2.9. Plot of calculated ΔG^\ddagger against the S-C distance in the TSs.

2.2.4.3 Charge on the β -Carbon

Conceptual DFT is next utilised to calculate global descriptors that measure the overall susceptibility of a molecular system to electrophilic attacks. The electrophilicity index, ω , a global descriptor devised by Parr¹¹² to describe the extent of energy lowering upon uptake of electron density, is defined by equation (6):

$$\omega = \frac{\chi^2}{2\eta} \quad (6)$$

where χ is electronegativity and η is chemical hardness, which refers to the electron density concentration around a particular atom in line with the hard-soft acid-base (HSAB) concept.¹¹³ Both of the variables are calculated from the ionization potential (IP) and electron affinity (EA) of

the molecule of interest, which are approximated to be the negatives of the LUMO and the highest occupied molecular orbital (HOMO) energies respectively according to Koopmans' theorem.¹¹⁴

The electrophilicity of an atom in a molecule is anticipated to be inversely proportional to its electron density, which is commonly represented by partial atomic charges. An attempt was thus made to explore the correlation between the addition barriers and the partial atomic charge on the β -carbon of the MA along with ω (Figure 2.10). Seven different charge calculation schemes were examined: Mulliken population analysis, natural bond orbital (NBO), Merz-Kollman, Hirshfeld, Charge Model 5 (CM5), quantum theory of atoms in molecules (QTAIM), and ChelpG. All of the charge schemes showed an inverse relationship with the ΔG^\ddagger values as expected, with the QTAIM charges returning the highest R^2 value (0.86). The R^2 values of other charge schemes range from 0.22 to 0.85. Overall, no exceptional correlation was found between charges and ΔG^\ddagger values. The correlation between the barrier and ω ($R^2 = 0.73$) is weaker than what was observed among unsubstituted acrylamide-containing compounds ($R^2 = 0.87$).¹¹¹

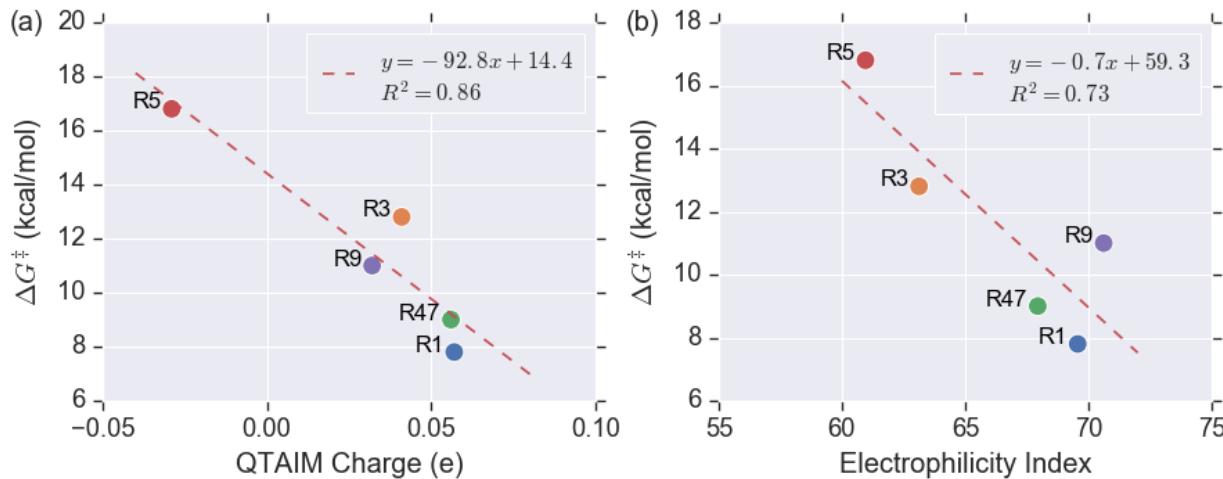


Figure 2.10. Plots of calculated ΔG^\ddagger against (a) QTAIM charges of β -carbon and (b) electrophilicity indices of the MAs.

2.2.4.4 Distortion/Interaction Analysis

The distortion/interaction model of reactivity¹¹⁵ was used to analyse the reactivity trend across the investigated cyanoacrylamide MAs. In this model, the activation energy for a reaction, ΔE^\ddagger , is understood as the sum of (i) the energies required to distort the reactants into their TS geometries, ΔE_d^\ddagger , and (ii) the interaction energy between the distorted reactants in the TS complex, ΔE_{int}^\ddagger , as shown in equation (7):

$$\Delta E^\ddagger = \Delta E_d^\ddagger + \Delta E_{int}^\ddagger \quad (7)$$

Plots showing the correlations between the computed ΔG^\ddagger and the distortion and interaction energies are shown in Figure 2.11.

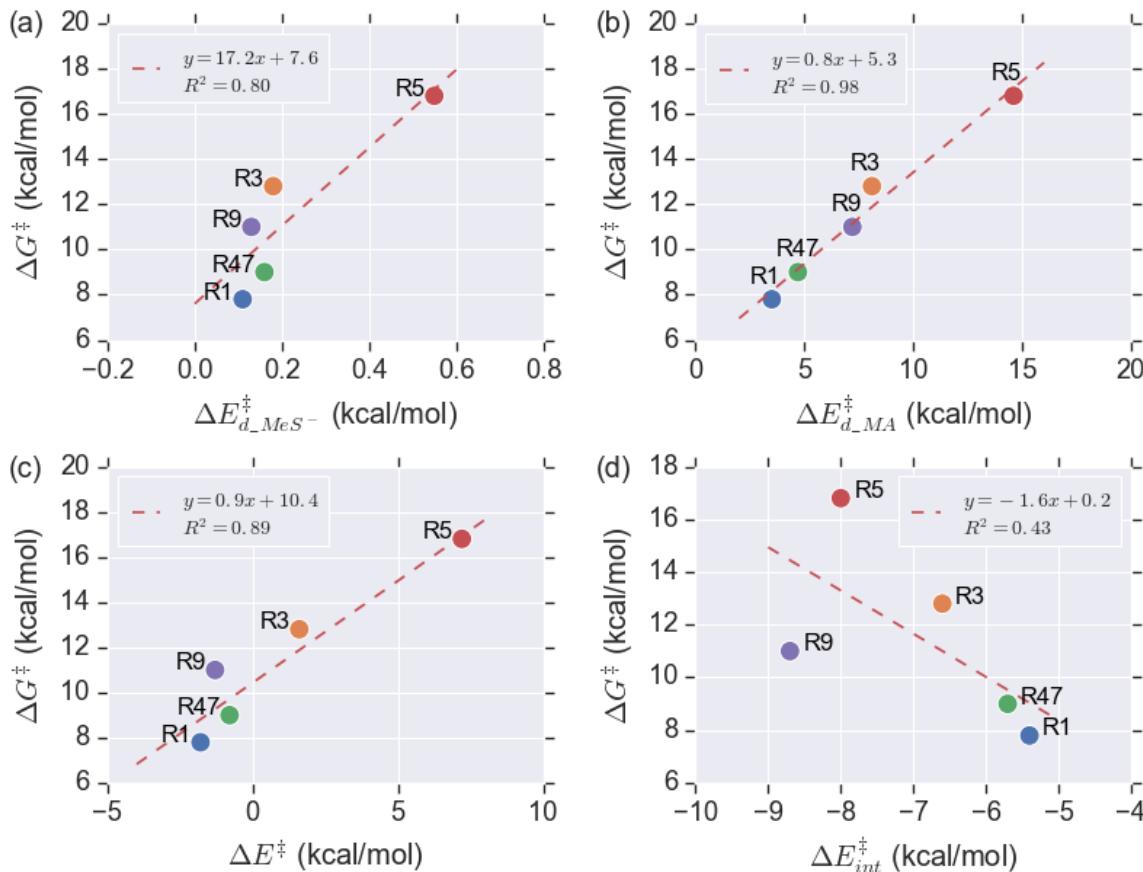


Figure 2.11. Plots showing the relationship between the computed ΔG^\ddagger and (a) the thiolate distortion energies, $\Delta E_{d_MeS^-}^\ddagger$, (b) the inhibitor distortion energies, $\Delta E_{d_MA}^\ddagger$, (c) ΔE^\ddagger , and (d) ΔE_{int}^\ddagger .

The amount of strain that the methylthiolate anion has to overcome to achieve the TS is small and varies little across the five acrylamides. One acrylamide, **R5**, has undergone a greater distortion to reach the TS compared to the other four. Despite the positive correlation observed ($R^2 = 0.80$), the relative minuscule magnitudes of the $\Delta E_{d_Mes^-}^\ddagger$ do not influence the ΔG^\ddagger in any major way. A rather weak correlation ($R^2 = 0.43$) was found from the regression analysis on ΔE_{int}^\ddagger .

In contrast, the $\Delta E_{d_MA}^\ddagger$ values show a very strong correlation with the addition ΔG^\ddagger values, as evidenced by the R^2 value of 0.98. The ΔE^\ddagger values also exhibited a rather strong correlation with the predicted ΔG^\ddagger values ($R^2 = 0.89$). The strong correlation implies that it is possible to use ΔE^\ddagger as a surrogate for ΔG^\ddagger , which would be of great use later in the analysis of the full BTK-inhibitor complexes. Interestingly, some of the reactions displayed negative ΔE^\ddagger values. This result was attributed to the formation of low-energy vdW complexes prior to the formation of TS.

The investigations above have demonstrated that the intrinsic addition barriers for the thiol-Michael additions do correlate with certain molecular properties, with $\Delta E_{d_MA}^\ddagger$ being the most prominent one. The trends observed intuitively indicates that $\Delta E_{d_MA}^\ddagger$ is smaller when the LUMO energy of the MA is lower as the TS could be achieved earlier (longer S-C distance).

2.2.5 Comparison between Predicted and Observed Kinetic and Thermodynamic Parameters

Overall, the irreversible acrylamide inhibitor **R5** exhibits much lower kinetic reactivity compared to the cyanoacrylamides. This is consistent with the understanding from previous work that the installation of electron-withdrawing groups at the α position of acrylamides enhances their reactivities kinetically at the cost of destabilisation of the corresponding thiol adducts.⁴⁶ The methyl- and *tert*-butyl-substituted cyanoacrylamides **R1** and **R3** are almost always on the opposite extremes in the kinetic reactivity trends predicted for all the properties tested among the

cynoacrylamides, with **R1** being the most reactive and **R3** being the most inert. This is in agreement with the experimental observation where **R3** had a longer RT than **R1**. The values predicted for **R9**, which had a long but reversible duration of action in the experiment, are largely similar to those of **R3**, being slightly less reactive in most cases. The reactivity differences between certain pairs of the inhibitors (**1** vs **3**, **3** vs **9**, and **5** vs the cynoacrylamides) are thus predicted successfully.

Compared to the experimentally observed kinetic reactivities in the reverse reaction, the rankings of the elimination rates predicted by the QM calculations are reproduced qualitatively, but the relative differences in magnitude of elimination barriers between the inhibitors do not agree well. It was therefore anticipated that the study on the dynamics of the full BTK-inhibitor systems, reported in the next chapter, would shed some light on the latter aspects.

The repetitive tasks of generation of QM job scripts and tabulation of data required immense alertness and are deemed to be the biggest challenges faced during this phase of the study. This was mitigated through the automation of repetitive tasks through programming scripts, which allowed significantly increased efficiency in the management and analysis of the large volume of data.

3 INVESTIGATIONS OF THE BINDING OF COVALENT INHIBITORS TO BTK

This chapter describes MD simulations of the binding of the acrylamide inhibitors to BTK. Systems with different bound forms (BTK with neutral/anionic noncovalently/covalently bound inhibitors) were simulated to study the conformational changes along the reaction coordinates. The key properties of interest in the simulations are the identity of the base catalyst, the strain induced in the inhibitors upon binding, and factors that influence the addition and elimination barriers.

3.1 Methods

3.1.1 *General Aspects*

Production MD simulations were conducted using Assisted Model Building with Energy Refinement Version 18 (AMBER18) *pmemd.cuda* program¹¹⁶⁻¹¹⁸ installed on a Dell EMC high-performance computer. The preparation of the systems and energy minimisations were carried out using GROMOS++ programs version 1.4.0.¹¹⁹ Twelve systems in total were considered. For each of inhibitors **1** and **3** four systems were studied, namely, (i) a noncovalent complex with neutral Cys481 thiol, (ii) a noncovalent complex with deprotonated Cys481 thiolate, (iii) a covalently bound anionic inhibitor-BTK adduct (enolate), and (iv) a neutral covalent adduct. For each of inhibitors **4**, **5**, **7**, and **9**, noncovalently bound BTK complexes with deprotonated Cys481 were studied.

3.1.2 *Preparation of the System*

Protein topology parameters were obtained from the Groningen Molecular Simulation Version 11 (GROMOS11) 54A7 force field¹²⁰ in the form of interaction function parameter (ifp) and molecular topology building block files from the Automated Topology Builder (ATB) version 3.0 web server.^{121,122} ATB was also used to generate the topology parameters of inhibitors **1**, **3**, **4**, **5**, **7**, and **9** both in the form of the unreacted compounds and of adducts covalently bound to Cys481.

The geometries were optimised at the B3LYP/6-31G*¹²³ level of theory in water using PCM.^{124,125} The electrostatic potential was then calculated using the optimised geometry, and the charges were obtained from the electrostatic potential by least-squares fitting.¹²¹

The systems simulated were built based on the X-ray co-crystal structure of BTK with inhibitor **3** (PDB accession number: 4YHF⁵). The structure consisted of the BTK dimer covalently bound to two molecules of **3**. Missing residues were built using PyMOL.¹²⁶ The N termini were modelled as NH₃⁺ and the C termini as COO⁻. Swiss PDB Viewer was employed to add missing atoms through reconstruction of the side chains.¹²⁷ Molecules other than BTK or **3** such as SO₄²⁻ anions and ethylene glycol present due to the buffers used during protein crystallisation were removed. The warhead of **3** was modified to give the other inhibitors. Conversion of the covalently bound inhibitors back to the reactant state was achieved by removing the extra proton acquired during thiol addition and adjusting the orbital hybridisation. The topology files of BTK monomers and inhibitors were generated separately using the *make_top* program and combined using the *com_top* program. The protonation state of each residue was based on predicted pK_a values obtained using PROPKA version 3.1.^{128,129} A residue was protonated if its pK_a value was equal to or lower than physiological pH (7.4). The residue and atom names of the inhibitor in the PDB file were then updated to match the names assigned by the ATB. The *pdb2g96* program was used to generate a coordinate file of the system in GROMOS96 format followed by the addition of hydrogen atoms by the *gch* program.

The protein and inhibitors were first energy minimised in vacuum over 5000 steps using steepest descent method with a tolerance of 0.01 kJ/mol before solvation in a rectangular box of simple point-charge (SPC) water with a minimum distance of 14 Å between the protein and the wall using the *sim_box* program. For the systems involving deprotonated species, the overall

charge of the system was -2. Neutralisation was achieved by progressively replacing water molecules with the highest potential with sodium ions. The solvated system was energy minimised again for 5000 steps before a positionally restrained energy minimisation was carried out by applying a harmonic potential to the coordinates of the protein structures. Full details of the parameters used are included in Appendix 6.2. The GROMOS topologies and coordinates were converted into AMBER-compatible format using the topology converter on the ATB web server.

3.1.3 Setup for MD Simulations

The systems were first equilibrated in the canonical (NVT) ensemble for 100 ps. The velocities of the atoms were initialised based on the sampling from Boltzmann distribution at 298.15 K. The temperature of the system was then increased to 310.15 K and maintained with a Berendsen thermostat.¹³⁰ A 2 ns isobaric-isothermal (NPT) equilibration with a pressure of 1 atm maintained using a Berendsen barostat¹³⁰ was then carried out. This was followed by 12 ns NVT equilibration. All-atom NVT simulations was carried out for 100 ns for each system. The simulation temperature of 310.15 K and pressure of 1 atm were maintained using coupling time constants of 0.1 and 0.5 respectively. Periodic boundary conditions (PBC) were imposed to avoid edge effects. Long-range Coulomb interactions were handled using the particle mesh Ewald (PME) method.¹³¹ The SHAKE algorithm¹³² with tolerance of 10^{-5} was utilized to constrain the bond involving hydrogen atoms in the protein. The time step and dielectric constant were set to 1 fs and 1.0, respectively. Both Coulomb and vdW interactions were truncated at 1.4 nm. The SETTLE algorithm was employed to constrain the geometry of water molecules.¹³³

3.1.4 Trajectory Analysis

The trajectories and structural properties of the simulated systems were analysed using the CPPTRAJ program¹³⁴ and plotted using Python scripts written as required during the project. The

systems were visualised using the Visual Molecular Dynamics (VMD)¹³⁵ analysis toolkit and UCSF Chimera software.¹³⁶

The Boltzmann averaged energies of inhibitors **1** and **3** as sampled by the MD simulations and by MM conformational searches were calculated using equation (4) on page 14. Fifteen structures of the inhibitors were obtained from the simulation trajectories based on their coordinates RMSD using an unsupervised machine learning algorithm known as k-means clustering. Each of the geometry was then optimised using the OPLS3e force field prior to the calculation of the Boltzmann averaged energy.

3.2 Results and Discussion

3.2.1 *Identification of the Potential Base Catalyst*

The base catalyst plays an important role in the mechanism of thiol-Michael reactions that proceed via the conventional mechanism shown in Scheme 1.1 (page 2). At the beginning of this project, the identity of the base(s) involved in deprotonation of Cys481 in the forward addition reaction and in deprotonation of the covalent adduct in the elimination reaction, which is not necessarily the same, was unknown. An analysis of the MD trajectories was thus performed to search for potential base species that might participate in the thiol addition or elimination.

3.2.1.1 *Distance of Protons to be Extracted from Potentially Basic Residues*

A common understanding assumes that the “basic” amino acid residues (histidines, lysines, and arginines) are unable to act as basic catalysts as they tend to be protonated at physiological pH.¹³⁷ However, there is growing evidence that these kinds of residues may act as biological base catalysts. Schlippe and Hedstrom¹³⁸ have reported solvent accessible arginines adjacent to carboxylate groups acting as general bases while Kenyon et al.¹³⁹ found that the pK_a of lysines could be reduced significantly due to spatial proximity to ammonium groups. Taking this into

consideration, when searching for possible participating bases, the candidates considered were histidines, lysines, arginines, glutamates, and aspartates.

The pK_a of Cys481 in the minimum energy geometries sampled in the noncovalently bound BTK by inhibitors **1** and **3** were calculated by PROPKA and estimated to be 8.9 and 9.6, respectively, indicating that it is unlikely to be deprotonated in the noncovalent adducts, which are formed before the covalent binding takes place. Possible candidates for the base catalyst of the forward addition reaction were identified by measuring the distance of the above-mentioned residues from the thiol proton of Cys481 during the simulations of the noncovalently bound BTK-inhibitor systems. The nitrogen atoms of the guanidine groups of Arg146 and Arg424 were observed to approach as close as 2-3 Å from the thiol protons briefly during the 100 ns simulations (Figure 3.1). This gives rise to the prospect that one or both of these arginines could be the base that deprotonates Cys481 to catalyse the thiol additions. Relevant data regarding the distances measured are recorded in Appendix 6.4.

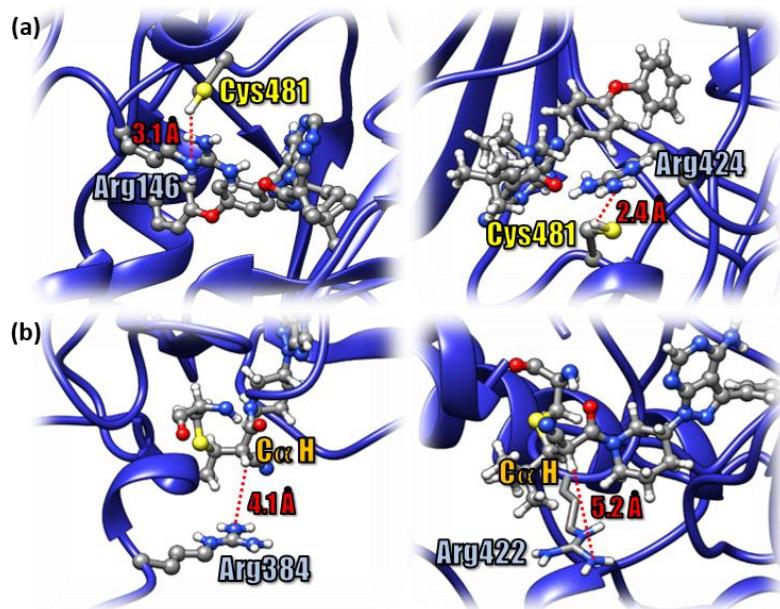


Figure 3.1. Snapshots of the simulations capturing the closest distances between arginines potentially acting as base catalysts and (a) Cys481 thiol proton and (b) $C\alpha$ of the covalent adducts.

The potential base catalyst for the elimination of thiol from the covalently bound adducts of **1** and **3** was then studied by looking at how close the charged moieties of the above amino acid types approached the C α proton of the bound inhibitors during the MD simulations. The closest interactions for both systems involved residues Arg384 and Arg422 and were about 4-5 Å. These distances are relatively far away from the expected geometry of a deprotonation TS (expected B-H-C distance of about 3 Å). The water molecules nearby do not exhibit orientations that would indicate they could assist in the shuttling of the C α proton to the arginines. Considering that it took more than 200 hours for **3** to dissociate from BTK experimentally, and that there is no obvious candidate for an internal base within the binding site, it is reasonable to hypothesise that the conformation observed in the X-ray co-crystal structure of the bound adduct must undergo a significant change before the elimination can take place. Multiple thermodynamically accessible states of BTK have been reported in the literature. The X-ray structure of BTK solved by Taunton et al. is most similar to the “C-helix-out” inactive conformation found by Silvian et al.¹⁴⁰ and is maintained in most of the MD simulations. However, the conformation which allows Arg384 to approach the C α proton to 4.1 Å was sampled from a replicate of BTK covalently bound by **1** which folded into a conformation resembling the apo form.¹⁴¹ Longer simulations would be desirable to sample the BTK conformation where C α proton extraction is possible. Techniques such as simulated annealing could potentially help to overcome the barrier to conformational interconversion.

Alternatively, the base for thiol elimination could be an external one from the solution, rather than internally from the protein. This would require the proton to be abstracted to remain solvent exposed for an extended period. Considering that the BTK dimers were maintained at the “C-helix-out” state in most of the simulations, the C α protons would meet this criterion.

3.2.1.2 Transition States for $\text{C}\alpha$ Proton Abstractions from the Covalent Adducts

On the basis of the abovementioned search for the base catalyst(s), QM calculations were next performed to determine the activation energies for the deprotonation of the covalent adducts by a model Arg base. The calculated TS structures are shown in Figure 3.2. The breaking C-H bonds of all inhibitors except **P9** are perpendicular to the C=O bond, enabling stabilisation through conjugation between the developing carbanion and the π^* orbital of the C=O bond. The ideal geometry is likely inaccessible for **TSD9** due to steric hindrance. However, its C-H bond was found to be in plane with the C=O bond, which allows stabilisation from negative hyperconjugation, where the electron pair of the breaking bond interacts with the low-lying empty C-O σ^* orbital.

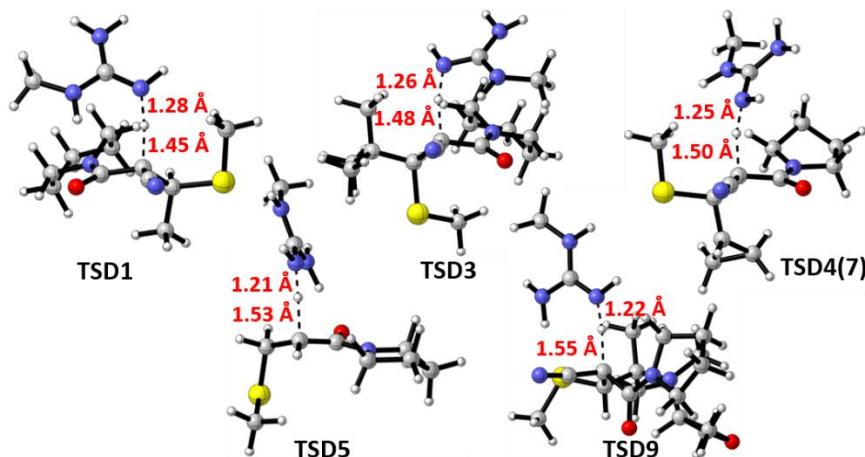


Figure 3.2. Structures of the thiol adduct deprotonation TSs.

3.2.1.3 Construction of Complete Energy Profiles

Putting all of the information together, the energy profiles for the arginine-catalysed thiol-Michael additions to the truncated MAs are shown in Figure 3.3.

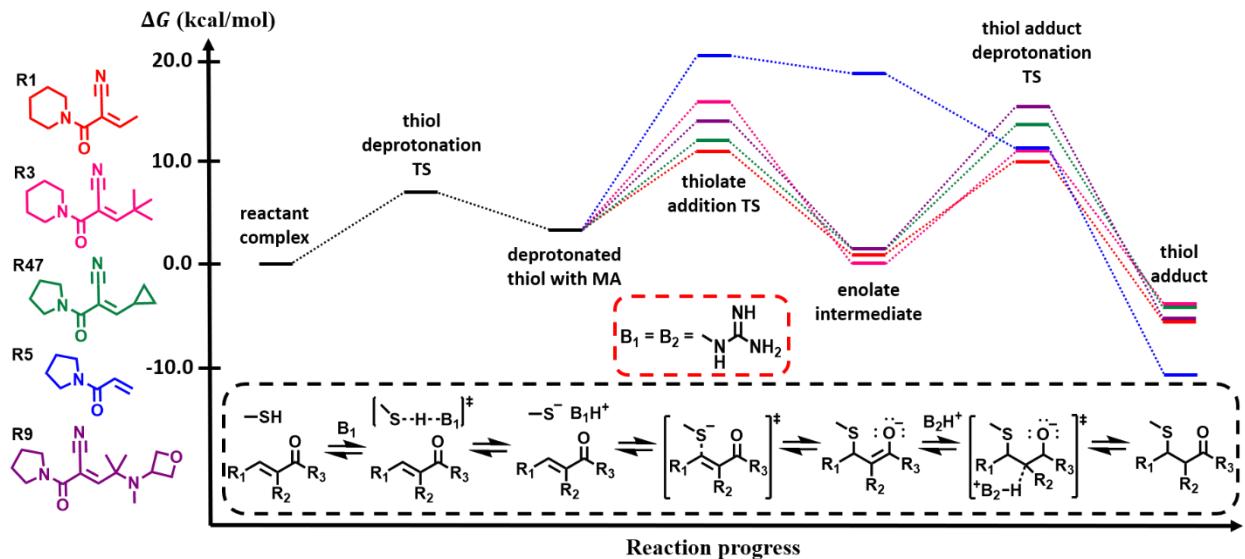


Figure 3.3. Energy profiles of thiol-Michael additions of methanethiols to acrylamides **R1**, **R3**, **R4(7)**, **R5**, and **R9**, calculated at M06-2X/6-311+G(d,p) level of theory.

The energy profiles shown differ qualitatively from previous computational studies on cyanoacrylamide thiol-Michael additions,⁴⁶ in that the energies of the intermediate enolate species derived from cyanoacrylamides **R1**, **R3**, **R4(7)**, **R5**, and **R9** appear to be higher than the energies of the products. This result is, however, consistent with the experimental findings, as the observed products were the neutral thiol adducts, not the enolate intermediates. The difference in the energetics between this and previous work can be attributed to the choice of the base species (1-methylguanidine here and DBU previously). The stabilities of the enolate ions (plus BH^+) depend on the pK_a of the model base. The absolute magnitudes of the kinetic parameters derived from the energy profiles (Figure 3.3) are only relevant if the “true” base is 1-methylguanidine, but their relative magnitudes can still be expected to be qualitatively correct. In contrast, the thermodynamic parameters of the overall equilibrium are not affected by the identity of the base. The thiol adducts for the four cyanoacrylamides are found to be 3.9–5.9 kcal/mol more stable than the reactants while the product of the thiol addition to the acrylamide (**R5**) is much more stabilised (-10.8 kcal/mol).

An unanticipated result is that the enolate formation TSs of **P4(7)** (green) and **P9** (purple) are higher in energy than their thiolate addition TSs **TS4(7)** and **TS9**. This is possibly due to greater strain, as indicated by their relatively later TSs (longer breaking C-H bond and shorter forming N-H bond) compared to the other cyanoacrylamides. In agreement with the recent report of Wang et al,⁸ the inclusion of 1-methylguanidine as the base reveals the formation of enolate intermediate to be the rate determining step (RDS) for the thiol elimination of all MAs investigated except **P3**.

With the RDSs established, the predicted ΔG_{rev}^\ddagger values are plotted against the experimental values in Figure 3.4. The trend now matches the experimentally observed trend to a greater extent, compared with the analysis shown above in Figure 2.7 based on using the thiolate as the model base. **R3** is now the only outlier. The incorporation of 1-methylguanidine as the catalyst has greatly improved the agreement between the calculated elimination barrier and the observed RTs. As proposed, the thiol elimination of **3** could involve a different conformation of BTK, which could induce a relatively large magnitude of strain in the adduct deprotonation TS and hence account for the apparently much higher barrier than the predicted ΔG_{rev}^\ddagger .

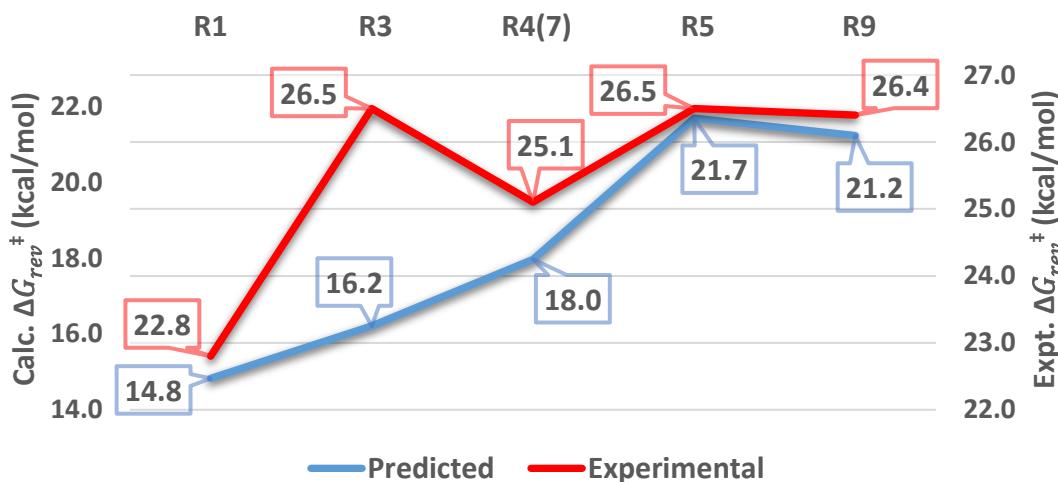


Figure 3.4. Comparison between calculated and experimental ΔG_{rev}^\ddagger values. The experimental data point for **7** ($\Delta G_{rev}^\ddagger = 25.4$ kcal/mol) is omitted for simplicity.

The possibility that thiol elimination takes place through alternative base free mechanisms such as six-membered and four-membered intramolecular proton transfers shown in Scheme 3.1 was also considered. In comparison with the conventional base-catalysed elimination mechanism, the ΔG_{rev}^\ddagger values for the intramolecular proton transfers were found to be much higher (Chart 3.1). The six-membered TSs are at least 8.1 kcal/mol higher in energy than the TSs of the RDS in the Arg-catalysed pathway and the four-membered TSs are at least 16.7 kcal/mol higher in energy than the six-membered TSs. This observation indicates that the rate of intramolecular proton transfers would be much slower than the E1cB mechanism at the experimental temperature (37 °C).

Scheme 3.1. Thiol Elimination via Intramolecular Proton Transfers

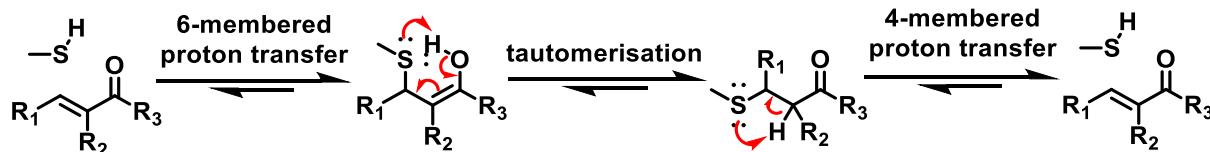
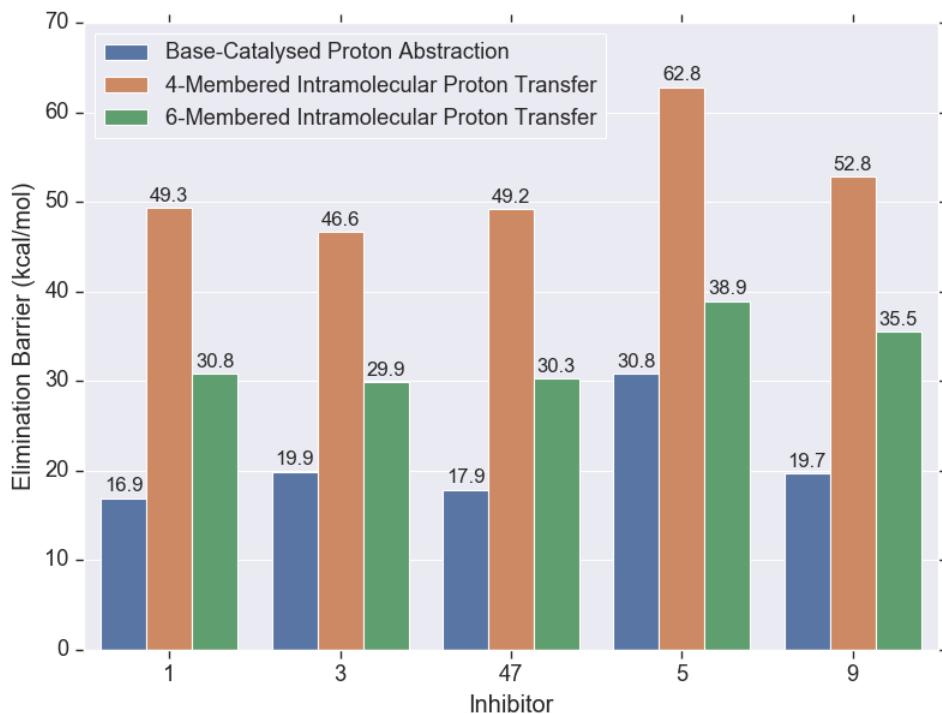


Chart 3.1. Activation Barriers for Different Thiol Elimination Mechanisms



3.2.2 Conformational Changes that Occur Upon Binding of Cyanoacrylamides 1 and 3 to BTK

3.2.2.1 Molecular Strain Induced Upon Binding

The distortion/interaction analysis in Chapter 2 had revealed a strong correlation between the magnitude of the barrier for thiol addition and the distortion energy of the MA. In light of this correlation, the conformations of the MAs in the MD simulations were analysed in order to determine how much internal strain is introduced upon binding to BTK. A key question was how the structural differences between the warheads of the inhibitors affect the distortion of the bound inhibitors from their most stable structures. The MD simulations revealed that only the warhead regions of the inhibitors were flexible in the bound adducts. The scaffold and linker moieties have a more or less fixed geometry, due to various interactions with the active site residues in the binding pocket.

Figure 3.5 shows the comparison between the conformations of **1** and **3** sampled from the simulations of noncovalently bound BTK and those sampled from MM conformer searches on the free inhibitors. It is seen that prior to binding to BTK, the inhibitors have to “unfold” from their most stable geometries into a significantly different conformation in which certain stabilising intramolecular interactions have been lost (e.g. C-H $\cdots\pi$ interactions between the methyl protons of the capping group and the *para*-substituted phenyl group of **3**). The Boltzmann averaged energies of the free inhibitors and the bound forms differ by 3.4-4.1 kcal/mol and 5.6-5.8 kcal/mol for **1** and **3**, respectively, which could easily be compensated for by the interaction with the residues. The geometry constrained exerted on Cys481 by the other residues was found to hinder the inhibitor from attaining the hypothetical ideal position of the S-Me group calculated in Section 2.2.2.4 (Figure 2.6).

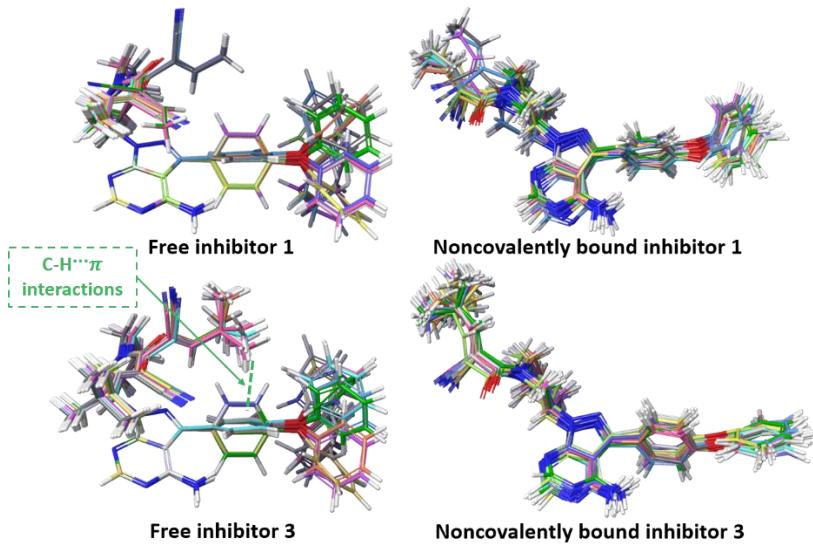


Figure 3.5. Superposition of fifteen conformers from conformational sampling of unbound conformers (left) and MD simulations of noncovalently bound BTK (right) for inhibitors **1** (upper) and **3** (lower).

3.2.3 *Trajectory Analysis of BTK Noncovalently Bound to the Inhibitors*

The trajectories of the noncovalently bound BTK-inhibitor complexes were analysed to better understand the factors which might affect the ability to form the covalent bond to Cys481. The features investigated were the distance between the atoms involved in C-S bond formation and the rotational barrier about the C=C-C=O bond of the inhibitor.

3.2.3.1 *Distance of Cys481 Sulfur from the Electrophilic Carbon*

The distance between the atoms that participate in the covalent bond formation event is a crucial indicator of the possibility of achieving the TS geometries. Figure 3.6 shows the distribution of the distances between the Cys481 sulfur and the electrophilic carbon of the inhibitors. The average S-C distances vary from 4 to 8 Å, with the distance between the thiolate sulphur and β -carbon of inhibitor **3** being the shortest on average (3.8 Å). Based on the QM calculations, the hypothetical ideal S-C distances in the TS are about 2.7 Å (**R1**), 2.6 Å (**R3**), 2.7

\AA (**R4(7)**), 2.2 \AA (**R5**), and 2.6 \AA (**R9**). None of the simulations contained any frames in which the S-C distance was such close. Nevertheless, the fluctuations of the distance give an indication of how easily the two reactive moieties can approach each other within the binding site, which would be expected to correspond to the likelihood of attaining the TS geometry. In agreement with the observations made in the investigation on the torsional barriers, the fluctuation in the S-C distance throughout the simulation is the smallest for **3** (about 1 \AA) and largest for **5** (about 7 \AA). Inhibitors **1**, **4**, and **7** exhibited fluctuations of about 4 \AA , while **9** deviates for about 3 \AA from its average S-C distance throughout the simulations. Inhibitor **5** is thus predicted to have the fastest thiol addition while **3** has the slowest addition rate. Since the forward addition rate constants have not yet been reported experimentally, these predictions await experimental testing.

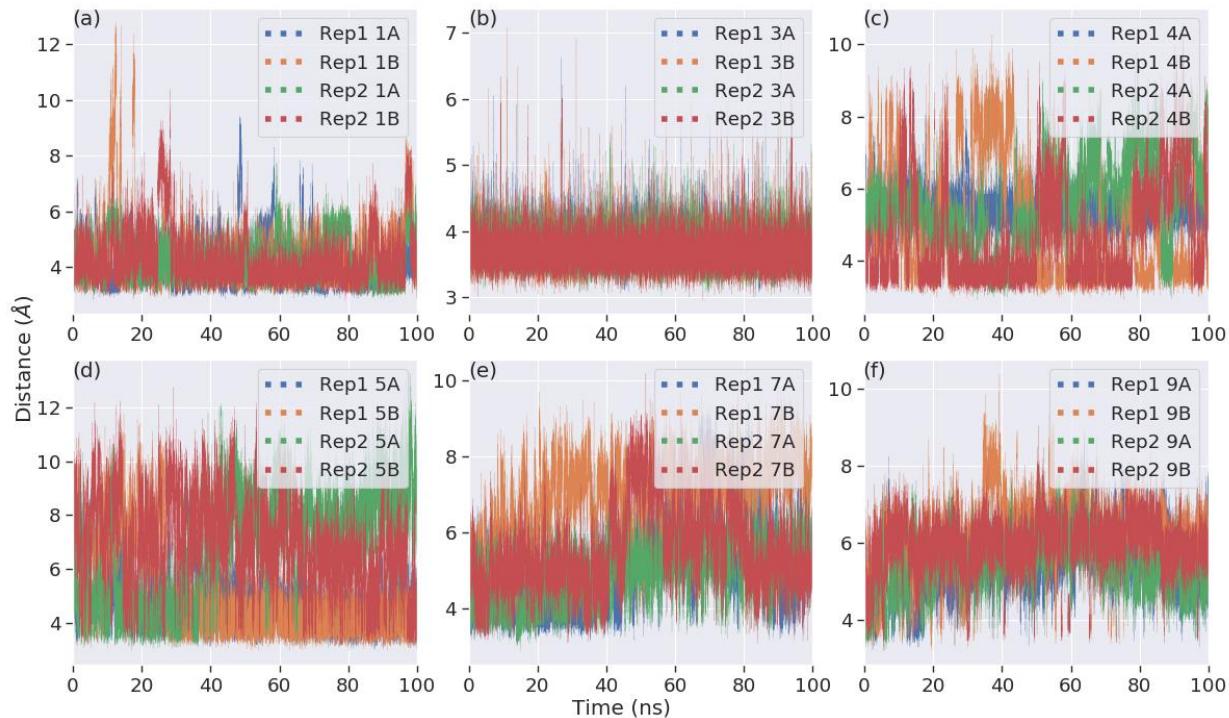


Figure 3.6. Plots of the distance between the Cys481 thiolate sulfur and β -carbon of the inhibitors as a function of time. A and B denote which BTK monomer each inhibitor is bound to while the number in front refers to the identity of the inhibitor.

3.2.3.2 Torsion of the C=C-C=O Bonds

In order to form a covalent bond, the nucleophilic thiolate of Cys481 has to approach the electrophilic carbon of the warhead. One feature identified as being important to this process was the ability of the warhead to rotate about the C=C-C=O bond. Flexibility in this torsion could be expected to increase the likelihood of the inhibitor of achieving a geometry resembling the TS for thiol addition by allowing faster sampling of different conformations. The torsion angles of the bond between the C=O carbon and the α -carbon were measured for all noncovalently bound inhibitors and are presented as distribution plots in Figure 3.7.

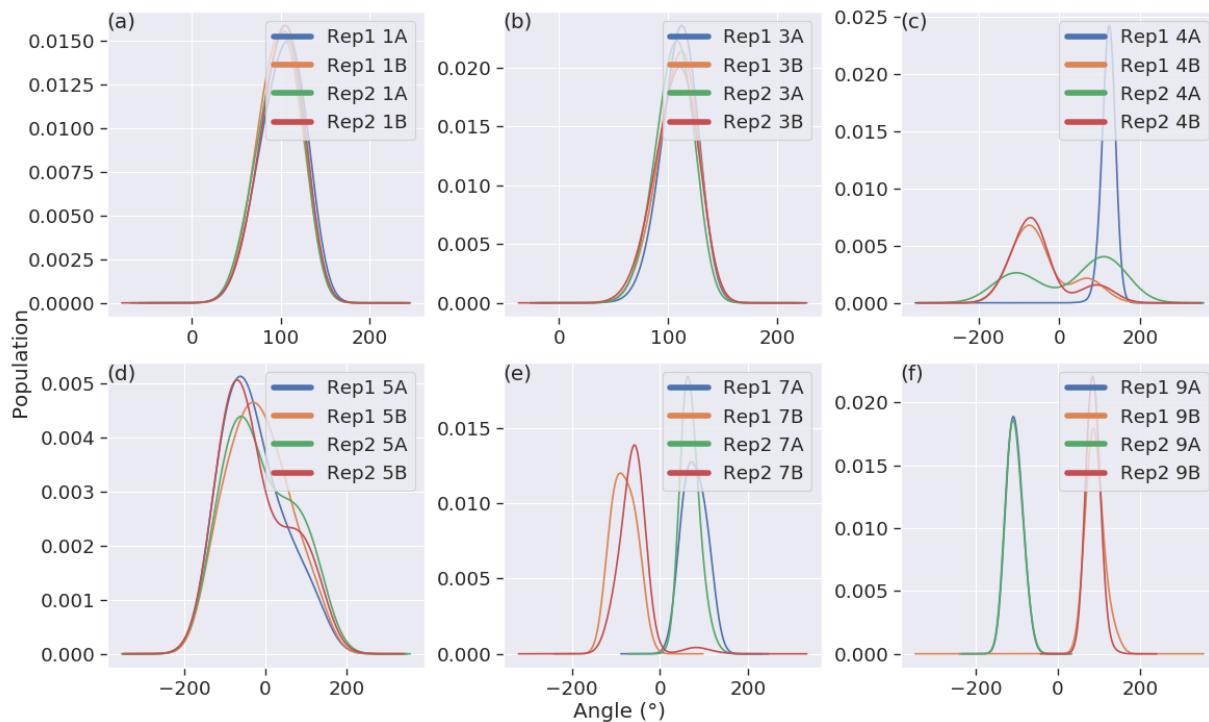


Figure 3.7. Distributions of the C=C-C=O dihedral angles throughout the 100 ns MD simulations of noncovalently bound BTK inhibitors. A and B denote which BTK monomer each inhibitor is bound to while the number in front refers to the identity of the inhibitor.

The dihedral angle of **3** was found to vary the least among the inhibitors, followed by **9**.

Upon visualisation, the dihedral rotation in **3** was found to be hindered by steric clashes between

the *tert*-butyl groups of **3** and the *iso*-propyl group of a nearby leucine throughout the simulation. In contrast, for **9**, the torsional barrier arises from the interactions of the polar functional groups (oxetane and nitrile) with the solvent, namely a hydrophobic effect. On the other hand, for inhibitor **5**, the torsion angle distribution spans a range of 180°, suggesting a small dihedral rotation barrier in the binding site, as expected from its least bulky warhead. Inhibitor **5** might thus be expected to undergo relatively faster thiol addition due to its greater flexibility within the binding site to form an interaction with Cys481. Conversely, the relatively smaller spread in the distributions of **3** and **9** suggests that the thiol additions might be slower for these inhibitors.

During this phase of study, an approach was developed to combine the strength of GROMOS force fields (which has been validated for biomolecular systems) and AMBER simulation machineries (which allow higher throughput). The simulation of covalently bound BTKs also required the inclusion of non-standard amino acids, which required extra care in their parameterisations.

4 CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions

Computational studies on the reactivities of cyanoacrylamide-based BTK inhibitors have been performed to understand the factors that determine the kinetics and thermodynamics of thiol-Michael addition and elimination in the enzyme active site. The DFT hybrid functional M06-2X/6-31+G(d)//M06-2X/6-311+G(d,p) was found to perform the best in predicting the energetics of thiol-Michael additions to cyanoacrylamides. The most stable geometries of the MAs are largely determined by the bulkiness of the warhead while the TS structures prefer *syn* geometries due to stabilising interactions. The rankings of the elimination barriers were predicted qualitatively by the QM calculations. Correlations between the addition barriers with electronic descriptors including the LUMO energy of the MA, the earliness/lateness of the TS, and the distortion energy of the MA were found. Arginine is proposed to be the base catalyst that participates in the Michael reactions in the BTK active site following the analysis of MD simulation trajectories. The inclusion of 1-methylguanidine as the model base in the computation of the energy profiles improved the agreement between the trends of calculated and experimental thiol elimination barriers, with only inhibitor **5** falling far from the trend line. Strain energy energies ranging from 3.4 to 5.8 kcal/mol were found to be induced upon the noncovalent binding of the inhibitors **1** and **3** to BTK. Several geometrical features related to the flexibility of the inhibitors within the binding site were identified as potential influences on the thiol addition rate. The acrylamide **5** was predicted to undergo the fastest thiol addition to Cys481 while cyanoacrylamide **3** was expected to have the highest addition barrier in the BTK active site.

4.2 Future Directions

QM calculations and MD simulations were employed in this project to provide complementary information into the binding of the cyanoacrylamides to BTK. To obtain further insightful details about the reaction pathways for addition and elimination beyond those reported here, an approach that accurately represent the electronic structure in the vicinity of the reacting atoms while taking the electrostatic and steric effects of the protein on the reaction into account would be required. Hybrid QM/MM methods were designed to circumvent the limitations of pure QM calculations (which are expensive and only possible for systems up to a few hundred atoms) and of MM force fields (which cannot describe bond-breaking/forming reactions) for large systems such as enzymes, and have received extensive use in recent years as a versatile tool for the study of enzymatic reaction mechanisms.¹⁴²⁻¹⁴⁸ The basic principle of these methods is to treat the chemically active region at QM level whereas the interactions within the protein surroundings or the explicit solvent molecules are represented using force fields (Figure 4.1).

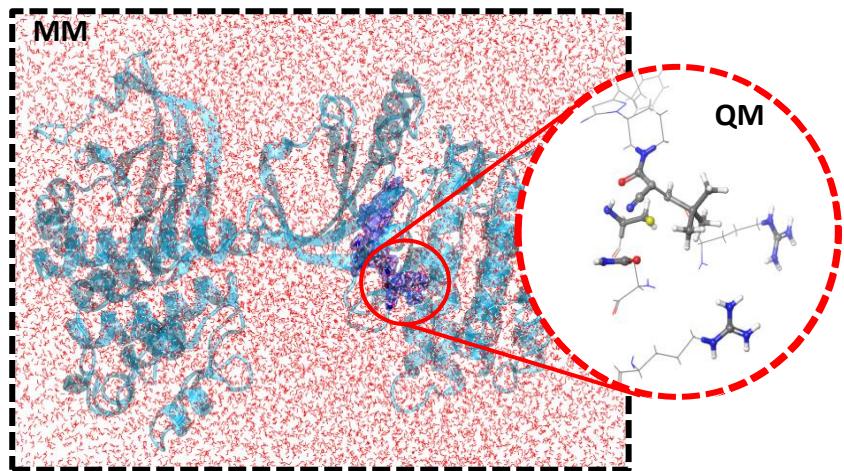


Figure 4.1. Illustration of the QM/MM concept.

The kinetic reactivity of the acrylamide inhibitors has been found to be highly dependent on the choice of the base catalyst and therefore the validation of its identity is paramount. The MD

simulations suggested that several arginine residues may play this role, especially in the forward addition reaction, but none was close enough to the C α to allow a confident assignment to be made for the thiol elimination. The current MD simulations should thus be extended to sample the conformational change that was proposed to be required before the deprotonation of the thiol adduct is possible. Various metaheuristic techniques could be applied to accelerate the sampling rate. Once a snapshot of the MD simulations resembling the TS geometries is obtained, the QM/MM studies could be conducted to construct more accurate energy profiles for the reactions. The comparison of the constructed reaction profiles with experimental data could potentially allow for the verification of the participating base. The study could also shed some light on the roles of specific residues near the binding site following further MD analysis on the interactions between the inhibitors with the active site residues.

Finally, the types of computational approaches developed here for BTK could be extended to investigate why **9** displayed selectivity for binding to BTK over other kinases. The QM/MM systems necessary for such a study could be constructed readily from the available X-ray crystal structures of the unbound proteins. Studies of the selectivities of **9** and other inhibitors would provide detailed information about the factors that influence the kinase selectivity of MAs, which is crucial for their safe use as covalent drugs.

5 REFERENCES

1. Potashman, M. H.; Duggan, M. E. *J. Med. Chem.* **2009**, *52*, 1231-1246.
2. Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. *J. Med. Chem.* **2009**, *52*, 225-233.
3. Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. *Nat. Rev. Drug Discov.* **2011**, *10*, 307-317.
4. Powers, J. C.; Asgian, J. L.; Ekici, Ö. D.; James, K. E. *Chem. Rev.* **2002**, *102*, 4639-4750.
5. Bradshaw, J. M.; McFarland, J. M.; Paavilainen, V. O.; Bisconte, A.; Tam, D.; Phan, V. T.; Romanov, S.; Finkle, D.; Shu, J.; Patel, V.; Ton, T.; Li, X.; Loughhead, D. G.; Nunn, P. A.; Karr, D. E.; Gerritsen, M. E.; Funk, J. O.; Owens, T. D.; Verner, E.; Brameld, K. A.; Hill, R. J.; Goldstein, D. M.; Taunton, J. *Nat. Chem. Biol.* **2015**, *11*, 525-531.
6. Serafimova, I. M.; Pufall, M. A.; Krishnan, S.; Duda, K.; Cohen, M. S.; Maglathlin, R. L.; McFarland, J. M.; Miller, R. M.; Frödin, M.; Taunton, J. *Nat. Chem. Biol.* **2012**, *8*, 471-476.
7. Kim, Y.; Mohrig, J. R.; Truhlar, D. G. *J. Am. Chem. Soc.* **2010**, *132*, 11071-11082.
8. Chen, J.; Jiang, X.; Carroll, S.; Huang, J.; Wang, J. *Org. Lett.* **2015**, *17*, 5978-5981.
9. Tóth, L.; Muszbek, L.; Komáromi, I. *J. Mol. Graphics Modell.* **2013**, *40*, 99-109.
10. Wright, P. M.; Seiple, I. B.; Myers, A. G. *Angew. Chem. Int. Ed.* **2014**, *53*, 8840-8869.
11. Kahan, F. M.; Kahan, J. S.; Cassidy, P. J.; Kropp, H. *Ann. N.Y. Acad. Sci.* **1974**, *235*, 364-386.
12. Utrecht, J. *Chem. Res. Toxicol.* **2008**, *21*, 84-92.
13. Utrecht, J. *Chem. Res. Toxicol.* **2009**, *22*, 24-34.
14. Chen, M.; Suzuki, A.; Borlak, J.; Andrade, R. J.; Lucena, M. I. *J. Hepatol.* **2015**, *63*, 503-514.
15. Yuan, L.; Kaplowitz, N. *Clin. Liver Dis.* **2013**, *17*, 507-518.
16. Mintzer, D. M.; Billet, S. N.; Chmielewski, L. *Adv. Hematol.* **2009**, *2009*, 1-11.
17. Barf, T.; Kaptein, A. *J. Med. Chem.* **2012**, *55*, 6243-6262.
18. Mah, R.; Thomas, J. R.; Shafer, C. M. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 33-39.
19. Kalgutkar, A. S.; Dalvie, D. K. *Expert Opin. Drug Discov.* **2012**, *7*, 561-581.
20. Copeland, R. A.; Pompliano, D. L.; Meek, T. D. *Nat. Rev. Drug Discov.* **2006**, *5*, 730-739.
21. Tummino, P. J.; Copeland, R. A. *Biochemistry* **2008**, *47*, 5481-5492.

22. Zhang, R.; Monsma, F. *Expert Opin. Drug Discov.* **2010**, *5*, 1023-1029.
23. Guo, D.; Hillger, J. M.; Ijzerman, A. P.; Heitman, L. H. *Med. Res. Rev.* **2014**, *34*, 856-892.
24. Copeland, R. A. *Expert Opin. Drug Discov.* **2010**, *5*, 305-310.
25. Lu, H.; Tonge, P. J. *Curr. Opin. Chem. Biol.* **2010**, *14*, 467-474.
26. Vilums, M.; Zweemer, A. J. M.; Yu, Z.; de Vries, H.; Hillger, J. M.; Wapenaar, H.; Bollen, I. A. E.; Barmare, F.; Gross, R.; Clemens, J.; Krenitsky, P.; Brussee, J.; Stamos, D.; Saunders, J.; Heitman, L. H.; Ijzerman, A. P. *J. Med. Chem.* **2013**, *56*, 7706-7714.
27. Swinney, D. C. *Nat. Rev. Drug Discov.* **2004**, *3*, 801-808.
28. Kahan, T.; Eliasson, K. *Am. J. Hypertens.* **1999**, *12*, 1188-1194.
29. Kohout, T. A.; Xie, Q.; Reijmers, S.; Finn, K. J.; Guo, Z.; Zhu, Y.-F.; Struthers, R. S. *Mol. Pharmacol.* **2007**, *72*, 238-247.
30. Bauer, R. A. *Drug Discov. Today* **2015**, *20*, 1061-1073.
31. Bryant, J.; Post, J. M.; Alexander, S.; Wang, Y.-X.; Kent, L.; Schirm, S.; Tseng, J.-L.; Subramanyam, B.; Buckman, B.; Islam, I.; Yuan, S.; Sullivan, M. E.; Snider, M.; Morser, J. *Thromb. Res.* **2008**, *122*, 523-532.
32. Smith, C. I.; Baskin, B.; Humire-Greiff, P.; Zhou, J. N.; Olsson, P. G.; Maniar, H. S.; Kjellén, P.; Lambris, J. D.; Christensson, B.; Hammarström, L. *J. Immunol.* **1994**, *152*, 557-565.
33. Mohamed, A. J.; Yu, L.; Bäckesjö, C.-M.; Vargas, L.; Faryal, R.; Aints, A.; Christensson, B.; Berglöf, A.; Vihinen, M.; Nore, B. F.; Edvard Smith, C. I. *Immunol. Rev.* **2009**, *228*, 58-73.
34. Hallek, M.; Shanafelt, T. D.; Eichhorst, B. *Lancet* **2018**, *391*, 1524-1537.
35. Wang, M. L.; Rule, S.; Martin, P.; Goy, A.; Auer, R.; Kahl, B. S.; Jurczak, W.; Advani, R. H.; Romaguera, J. E.; Williams, M. E.; Barrientos, J. C.; Chmielowska, E.; Radford, J.; Stilgenbauer, S.; Dreyling, M.; Jedrzejczak, W. W.; Johnson, P.; Spurgeon, S. E.; Li, L.; Zhang, L.; Newberry, K.; Ou, Z.; Cheng, N.; Fang, B.; McGreivy, J.; Clow, F.; Buggy, J. J.; Chang, B. Y.; Beaupre, D. M.; Kunkel, L. A.; Blum, K. A. *New Engl. J. Med.* **2013**, *369*, 507-516.
36. Di Paolo, J. A.; Huang, T.; Balazs, M.; Barbosa, J.; Barck, K. H.; Bravo, B. J.; Carano, R. A. D.; Darrow, J.; Davies, D. R.; DeForge, L. E.; Diehl, L.; Ferrando, R.; Gallion, S. L.; Giannetti, A. M.; Gribling, P.; Hurez, V.; Hymowitz, S. G.; Jones, R.; Kropf, J. E.; Lee, W. P.; Maciejewski, P. M.; Mitchell, S. A.; Rong, H.; Staker, B. L.; Whitney, J. A.;

- Yeh, S.; Young, W. B.; Yu, C.; Zhang, J.; Reif, K.; Currie, K. S. *Nat. Chem. Biol.* **2010**, *7*, 41-50.
37. Honigberg, L. A.; Smith, A. M.; Sirisawad, M.; Verner, E.; Loury, D.; Chang, B.; Li, S.; Pan, Z.; Thamm, D. H.; Miller, R. A.; Buggy, J. J. *Proc. Natl. Acad. Sci.* **2010**, *107*, 13075-13080.
38. Brown, J. R. *Curr. Hematol. Malig. Rep.* **2013**, *8*, 1-6.
39. Peterson, D.; Schwartz, J. J. *Adv. Pract. Oncol.* **2014**, *5*, 348-354.
40. Lanning, B. R.; Whitby, L. R.; Dix, M. M.; Douhan, J.; Gilbert, A. M.; Hett, E. C.; Johnson, T. O.; Joslyn, C.; Kath, J. C.; Niessen, S.; Roberts, L. R.; Schnute, M. E.; Wang, C.; Hulce, J. J.; Wei, B.; Whiteley, L. O.; Hayward, M. M.; Cravatt, B. F. *Nat. Chem. Biol.* **2014**, *10*, 760-767.
41. Byrd, J. C.; Furman, R. R.; Coutre, S. E.; Flinn, I. W.; Burger, J. A.; Blum, K. A.; Grant, B.; Sharman, J. P.; Coleman, M.; Wierda, W. G.; Jones, J. A.; Zhao, W.; Heerema, N. A.; Johnson, A. J.; Sukbuntherng, J.; Chang, B. Y.; Clow, F.; Hedrick, E.; Buggy, J. J.; James, D. F.; O'Brien, S. *New Engl. J. Med.* **2013**, *369*, 32-42.
42. Byrd, J. C.; Harrington, B.; O'Brien, S.; Jones, J. A.; Schuh, A.; Devereux, S.; Chaves, J.; Wierda, W. G.; Awan, F. T.; Brown, J. R.; Hillmen, P.; Stephens, D. M.; Ghia, P.; Barrientos, J. C.; Pagel, J. M.; Woyach, J.; Johnson, D.; Huang, J.; Wang, X.; Kaptein, A.; Lannutti, B. J.; Covey, T.; Fardis, M.; McGreivy, J.; Hamdy, A.; Rothbaum, W.; Izumi, R.; Diacovo, T. G.; Johnson, A. J.; Furman, R. R. *New Engl. J. Med.* **2015**, *374*, 323-332.
43. Wu, J.; Zhang, M.; Liu, D. *J. Hematol. Oncol.* **2016**, *9*, 1-4.
44. Barf, T.; Covey, T.; Izumi, R.; van de Kar, B.; Gulrajani, M.; van Lith, B.; van Hoek, M.; de Zwart, E.; Mittag, D.; Demont, D.; Verkaik, S.; Krantz, F.; Pearson, P. G.; Ulrich, R.; Kaptein, A. *J. Pharmacol. Exp. Ther.* **2017**, *363*, 240-252.
45. Chandrashekara, S. *Indian J. Pharmacol.* **2012**, *44*, 665-671.
46. Krenske, E. H.; Petter, R. C.; Houk, K. N. *J. Org. Chem.* **2016**, *81*, 11726-11733.
47. Copeland, R. A. *Nat. Rev. Drug Discov.* **2015**, *15*, 87-95.
48. Guo, D.; Heitman, L. H.; Ijzerman, A. P. *ChemMedChem* **2015**, *10*, 1793-1796.
49. Tonge, P. J. *ACS Chem. Neurosci.* **2017**, *9*, 29-39.
50. Ferguson, F. M.; Gray, N. S. *Nat. Rev. Drug Discov.* **2018**, *17*, 353-377.
51. De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A. K.; Moitessier, N. *Eur. J. Med. Chem.* **2017**, *138*, 96-114.
52. Renaud, J.-P.; Chung, C.-w.; Danielson, U. H.; Egner, U.; Hennig, M.; Hubbard, R. E.; Nar, H. *Nat. Rev. Drug Discov.* **2016**, *15*, 679-698.

53. Zhan, P.; Pannecouque, C.; De Clercq, E.; Liu, X. *J. Med. Chem.* **2016**, *59*, 2849-2878.
54. Fan, Q.; Aksoy, O.; Wong, R. A.; Ilkhanizadeh, S.; Novotny, C. J.; Gustafson, W. C.; Truong, A. Y.-Q.; Cayanan, G.; Simonds, E. F.; Haas-Kogan, D.; Phillips, J. J.; Nicolaides, T.; Okaniwa, M.; Shokat, K. M.; Weiss, W. A. *Cancer Cell* **2017**, *31*, 424-435.
55. Vauquelin, G. *Br. J. Pharmacol.* **2016**, *173*, 2319-2334.
56. Hilimire, T. A.; Bennett, R. P.; Stewart, R. A.; Garcia-Miranda, P.; Blume, A.; Becker, J.; Sherer, N.; Helms, E. D.; Butcher, S. E.; Smith, H. C.; Miller, B. L. *ACS Chem. Biol.* **2016**, *11*, 88-94.
57. Fisher, J. F.; Mobashery, S. *MedChemComm* **2016**, *7*, 37-49.
58. Willemse-Seegers, N.; Uitdehaag, J. C. M.; Prinsen, M. B. W.; de Vetter, J. R. F.; de Man, J.; Sawa, M.; Kawase, Y.; Buijsman, R. C.; Zaman, G. J. R. *J. Mol. Biol.* **2017**, *429*, 574-586.
59. Gupta, V.; Carroll, K. S. *Chem. Sci.* **2016**, *7*, 400-415.
60. Watterson, S. H.; De Lucca, G. V.; Shi, Q.; Langevine, C. M.; Liu, Q.; Batt, D. G.; Beaudoin Bertrand, M.; Gong, H.; Dai, J.; Yip, S.; Li, P.; Sun, D.; Wu, D.-R.; Wang, C.; Zhang, Y.; Traeger, S. C.; Pattoli, M. A.; Skala, S.; Cheng, L.; Obermeier, M. T.; Vickery, R.; Discenza, L. N.; D'Arienzo, C. J.; Zhang, Y.; Heimrich, E.; Gillooly, K. M.; Taylor, T. L.; Pulicicchio, C.; McIntyre, K. W.; Galella, M. A.; Tebben, A. J.; Muckelbauer, J. K.; Chang, C.; Rampulla, R.; Mathur, A.; Salter-Cid, L.; Barrish, J. C.; Carter, P. H.; Fura, A.; Burke, J. R.; Tino, J. A. *J. Med. Chem.* **2016**, *59*, 9173-9200.
61. Horton, J. R.; Engstrom, A.; Zoeller, E. L.; Liu, X.; Shanks, J. R.; Zhang, X.; Johns, M. A.; Vertino, P. M.; Fu, H.; Cheng, X. *J. Biol. Chem.* **2016**, *291*, 2631-2646.
62. Chaikuad, A.; Koch, P.; Laufer, S. A.; Knapp, S. *Angew. Chem. Int. Ed.* **2018**, *57*, 4372-4385.
63. Gupta, V.; Carroll, K. S. *Chem. Commun.* **2016**, *52*, 3414-3417.
64. Yang, W.; Hosford, S. R.; Dillon, L. M.; Shee, K.; Liu, S. C.; Bean, J. R.; Salphati, L.; Pang, J.; Zhang, X.; Nannini, M. A.; Demidenko, E.; Bates, D.; Lewis, L. D.; Marotti, J. D.; Eastman, A. R.; Miller, T. W. *Clin. Cancer Res.* **2016**, *22*, 2250-2260.
65. De Lucca, G. V.; Shi, Q.; Liu, Q.; Batt, D. G.; Beaudoin Bertrand, M.; Rampulla, R.; Mathur, A.; Discenza, L.; D'Arienzo, C.; Dai, J.; Obermeier, M.; Vickery, R.; Zhang, Y.; Yang, Z.; Marathe, P.; Tebben, A. J.; Muckelbauer, J. K.; Chang, C. J.; Zhang, H.; Gillooly, K.; Taylor, T.; Pattoli, M. A.; Skala, S.; Kukral, D. W.; McIntyre, K. W.; Salter-Cid, L.; Fura, A.; Burke, J. R.; Barrish, J. C.; Carter, P. H.; Tino, J. A. *J. Med. Chem.* **2016**, *59*, 7915-7935.

66. Brameld, K. A.; Owens, T. D.; Verner, E.; Venetsanakos, E.; Bradshaw, J. M.; Phan, V. T.; Tam, D.; Leung, K.; Shu, J.; LaStant, J.; Loughhead, D. G.; Ton, T.; Karr, D. E.; Gerritsen, M. E.; Goldstein, D. M.; Funk, J. O. *J. Med. Chem.* **2017**, *60*, 6516-6527.
67. Dalton, S. E.; Dittus, L.; Thomas, D. A.; Convery, M. A.; Nunes, J.; Bush, J. T.; Evans, J. P.; Werner, T.; Bantscheff, M.; Murphy, J. A.; Campos, S. *J. Am. Chem. Soc.* **2018**, *140*, 932-939.
68. Johnson, A. R.; Kohli, P. B.; Katewa, A.; Gogol, E.; Belmont, L. D.; Choy, R.; Penuel, E.; Burton, L.; Eigenbrot, C.; Yu, C.; Ortwine, D. F.; Bowman, K.; Franke, Y.; Tam, C.; Estevez, A.; Mortara, K.; Wu, J.; Li, H.; Lin, M.; Bergeron, P.; Crawford, J. J.; Young, W. B. *ACS Chem. Biol.* **2016**, *11*, 2897-2907.
69. Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. *J. Am. Chem. Soc.* **2014**, *136*, 12624-12630.
70. Chruszcz, M.; Wlodawer, A.; Minor, W. *Biophys. J.* **2008**, *95*, 1-9.
71. Kruschel, D.; Zagrovic, B. *Mol. Biosyst.* **2009**, *5*, 1606-1616.
72. Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.
73. Kuzmanic, A.; Pannu, N. S.; Zagrovic, B. *Nature Communications* **2014**, *5*, 1-10.
74. Sakthivel, S.; Habeeb, S. K. M. *J. Biomol. Struct. Dyn.* **2018**, *36*, 4320-4337.
75. Ratzon, E.; Bloch, I.; Nicola, M.; Cohen, E.; Ruimi, N.; Dotan, N.; Landau, M.; Gal, M. *ACS Omega* **2017**, *2*, 4398-4410.
76. Balasubramanian, P. K.; Balupuri, A.; Kang, H.-Y.; Cho, S. *J. BMC Syst. Biol.* **2017**, *11*, 1-11.
77. Balasubramanian, P. K.; Balupuri, A.; Cho, S. *J. Arch. Pharmacal Res.* **2016**, *39*, 328-339.
78. Bavi, R.; Kumar, R.; Choi, L.; Woo Lee, K. *PLOS ONE* **2016**, *11*, 1-19.
79. Sultan, M. M.; Denny, R. A.; Unwalla, R.; Lovering, F.; Pande, V. S. *Sci. Rep.* **2017**, *7*, 1-11.
80. Lu, D.; Jiang, J.; Liang, Z.; Sun, M.; Luo, C.; Shen, B.; Hu, G. Molecular Dynamic Simulation to Explore the Molecular Basis of Btk-PH Domain Interaction with Ins(1,3,4,5)P4 *Sci. World J.* [Online], **2013**, *2013*, 1-10. <http://dx.doi.org/10.1155/2013/580456> (accessed 2019/05/05).
81. Boyken, S. E.; Chopra, N.; Xie, Q.; Joseph, R. E.; Wales, T. E.; Fulton, D. B.; Engen, J. R.; Jernigan, R. L.; Andreotti, A. H. *J. Mol. Biol.* **2014**, *426*, 3656-3669.
82. Wang, Q.; Vogan, E. M.; Nocka, L. M.; Rosen, C. E.; Zorn, J. A.; Harrison, S. C.; Kuriyan, J. Autoinhibition of Bruton's Tyrosine Kinase (Btk) and Activation by Soluble

Inositol Hexakisphosphate *eLife* [Online], **2015**, 4, 1-31. <https://doi.org/10.7554/eLife.06074> (accessed 2019/05/05).

83. Smith, J. M.; Jami Alahmadi, Y.; Rowley, C. N. *J. Chem. Theory Comput.* **2013**, *9*, 4860-4865.
84. Kruse, H.; Goerigk, L.; Grimme, S. *J. Org. Chem.* **2012**, *77*, 10824-10834.
85. Fedorov, O.; Müller, S.; Knapp, S. *Nat. Chem. Biol.* **2010**, *6*, 166-169.
86. Botta, M. *ACS Med. Chem. Lett.* **2014**, *5*, 270-270.
87. Wereszczynski, J.; McCammon, J. A. *Q. Rev. Biophys.* **2012**, *45*, 1-25.
88. Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107-115.
89. *MacroModel*, version 9.7, Schrödinger: LLC, New York, NY, 2009.
90. Chen, I. J.; Foloppe, N. *Bioorg. Med. Chem.* **2013**, *21*, 7898-7920.
91. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. *J. Chem. Theory Comput.* **2019**, *15*, 1863-1874.
92. Thomas, B. E.; Kollman, P. A. *J. Org. Chem.* **1995**, *60*, 8375-8381.
93. Wang, C.; Qi, C. *Tetrahedron* **2013**, *69*, 5348-5354.
94. Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. *ChemMedChem* **2010**, *5*, 869-880.
95. Carlqvist, P.; Svedendahl, M.; Branneby, C.; Hult, K.; Brinck, T.; Berglund, P. *ChemBioChem* **2005**, *6*, 331-336.
96. Martell, J. M.; Beaton, P. T.; Holmes, B. E. *J. Phys. Chem. A* **2002**, *106*, 8471-8478.
97. Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792-794.
98. Pereira, S. R.; Vasconcelos, V. M.; Antunes, A. *FEBS J.* **2013**, *280*, 674-680.
99. Mulliner, D.; Wondrousch, D.; Schüürmann, G. *Org. Biomol. Chem.* **2011**, *9*, 8400-8412.
100. Chai, J.-D.; Head-Gordon, M. *PCCP* **2008**, *10*, 6615-6620.
101. Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51-57.
102. Livshits, E.; Baer, R. *PCCP* **2007**, *9*, 2932-2941.
103. Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. *J. Chem. Phys.* **2007**, *126*, 1-9.
104. Krenske, E. H.; Petter, R. C.; Zhu, Z.; Houk, K. N. *J. Org. Chem.* **2011**, *76*, 5074-5081.
105. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. C.; Aquilante, L. *GAUSSIAN 09*, Gaussian, Inc., Wallingford CT, 2009.

- H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*, Wallingford, CT, 2016.
106. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669-681.
107. Andrea N., B.; Steven, W. *Popular Integration Grids Can Result in Large Errors in DFT-Computed Free Energies*; 2019.
108. Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.
109. Fan, M. F.; Lin, Z. Y.; McGrady, J. E.; Mingos, D. M. P. *J. Chem. Soc.-Perkin Trans. 2* **1996**, *4*, 563-568.
110. Danovich, D.; Shaik, S.; Neese, F.; Echeverría, J.; Aullón, G.; Alvarez, S. *J. Chem. Theory Comput.* **2013**, *9*, 1977-1991.
111. Jackson, P. A.; Widen, J. C.; Harki, D. A.; Brummond, K. M. *J. Med. Chem.* **2017**, *60*, 839-885.
112. Parr, R. G.; Szentpály, L. v.; Liu, S. *J. Am. Chem. Soc.* **1999**, *121*, 1922-1924.
113. Pearson, R. G. *J. Am. Chem. Soc.* **1963**, *85*, 3533-3539.
114. Koopmans, T. *Physica* **1934**, *1*, 104-113.
115. Bickelhaupt, F. M.; Houk, K. N. *Angew. Chem. Int. Ed.* **2017**, *56*, 10070-10086.
116. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.
117. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 1542-1555.
118. Le Grand, S.; Götz, A. W.; Walker, R. C. *Comput. Phys. Commun.* **2013**, *184*, 374-380.
119. Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2011**, *7*, 3379-3390.

120. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011**, *40*, 843.
121. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 4026-4037.
122. Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. *J. Chem. Theory Comput.* **2018**, *14*, 5834-5845.
123. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
124. Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117-129.
125. Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622-6633.
126. *The PyMOL Molecular Graphics System, Version 2.0*, Schrödinger: LLC, New York, NY, 2015.
127. Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714-2723.
128. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284-2295.
129. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525-537.
130. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
131. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
132. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
133. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952-962.
134. Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
135. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33-38.
136. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
137. Gerlt, J. A.; Kozarich, J. W.; Kenyon, G. L.; Gassman, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 9667-9669.
138. Guillén Schlippe, Y. V.; Hedstrom, L. *Arch. Biochem. Biophys.* **2005**, *433*, 266-278.
139. Highbarger, L. A.; Gerlt, J. A.; Kenyon, G. L. *Biochemistry* **1996**, *35*, 41-46.
140. Marcotte, D. J.; Liu, Y.-T.; Arduini, R. M.; Hession, C. A.; Miatkowski, K.; Wildes, C. P.; Cullen, P. F.; Hong, V.; Hopkins, B. T.; Mertsching, E.; Jenkins, T. J.; Romanowski, M. J.; Baker, D. P.; Silvian, L. F. *Protein Sci.* **2010**, *19*, 429-439.
141. Mao, C.; Zhou, M.; Uckun, F. M. *J. Biol. Chem.* **2001**, *276*, 41435-41443.

142. Lu, J.; Zhang, Z.; Ni, Z.; Shen, H.; Tu, Z.; Liu, H.; Lu, R. *Comput. Biol. Chem.* **2014**, *52*, 25-33.
143. McClory, J.; Hu, G.-X.; Zou, J.-W.; Timson, D. J.; Huang, M. *J. Phys. Chem. B* **2019**, *123*, 2844-2852.
144. McClory, J.; Lin, J.-T.; Timson, D. J.; Zhang, J.; Huang, M. *Org. Biomol. Chem.* **2019**, *17*, 2423-2431.
145. McClory, J.; Timson, D. J.; Singh, W.; Zhang, J.; Huang, M. *J. Phys. Chem. B* **2017**, *121*, 11062-11071.
146. Montenegro, M.; Garcia-Viloca, M.; Lluch, J. M.; González-Lafont, Á. *PCCP* **2011**, *13*, 530-539.
147. Shi, T.; Lu, Y.; Liu, X.; Chen, Y.; Jiang, H.; Zhang, J. *J. Phys. Chem. B* **2011**, *115*, 11895-11901.
148. van der Kamp, M. W.; Mulholland, A. *J. Biochemistry* **2013**, *52*, 2708-2728.

6 APPENDIX

6.1 Benchmarking of Force Fields for Conformational Sampling

A benchmarking of seven force fields was carried out to assess the similarities between them. The MM2 and MM3 force fields were not included due to their relatively inaccurate GBSA solvation energies compared to other force fields. The conformational searches were conducted on three different species, consisting of a inhibitor **R1** and its corresponding thiol adduct (**P1**) and TS (**TS1**). The same settings as described in Section 2.1.3 were employed here and the results are tabulated in Table 6.1. The MMFF family force fields returned rather consistent results, with MMFFs locating slightly more conformers. The conformational samplings using the AMBER94 force field did not return any conformers, possibly due to the lack of extensions for organic molecules. The OPLS3e force field returned the least conformers for **R1** and **TS1** but managed to search a wider conformational space of the more flexible thiol adduct **P1** compared to the other force fields, inferior only to the AMBER force field. OPLS3e force field was eventually chosen for the conformational sampling as recommended by Schrödinger, who claims that the force field has more extensive parameterisation and reduced parameter transferability errors.

Table 6.1. Number of conformers found from the conformational sampling using different force fields.

Force field	Number of conformers		
	R1	P1	TS1
AMBER	20	98	- ^a
AMBER94	- ^a	- ^a	- ^a
MMFF	8	34	43
MMFFs	8	38	44

OPLS	12	42	- ^a
OPLS_2005	14	38	61
OPLS3e	4	51	35

^aMacroModel reported that the conformational searches completed successfully but no structure is returned, indicating the inability to handle the atom types or functional groups present in the molecules due to the absence of suitable parameters.

6.2 Details of Parameters Used

6.2.1 Conformational Sampling

The partial atomic charges were assigned based on the force field and an extended cutoff for non-bonded interactions was used. This corresponds to cutoff distances of 8.0 Å, 20.0 Å, and 4.0 Å, for van der Waals (vdW), electrostatic and hydrogen bonds respectively. The probability of a torsion rotation or molecule translation, minimum and maximum distance for low-mode move were set to be 0.5, 3.0, and 6.0, respectively. The conformers with a MAD of at least 0.5 Å from previously located energy minima within an energy window of 5 kcal/mol were stored from a sampling of 1000 maximum steps and 100 steps per rotatable bond. The parameters chosen were mostly the default values of MacroModel conformational sampling job configuration, except for the range of the energy window.

6.2.2 NCI Plots

The NCIs plotted were calculated using promolecular densities of the molecules. A colour scale of -3 to 3 was used.

6.2.3 Preparation of GROMOS Simulation Systems

A threshold of 0.1% was specified for the *gch* program, which adjusts the positions the H atoms for which the connecting bond deviates from the optimal distance by a percentage greater than the threshold. The initial and maximum time steps for energy minimisations were

set to 10 fs and 50 fs, respectively. All bond lengths were constrained using SHAKE algorithm with tolerance of 10^{-4} . The `@rotate` flag was specified to rotate the solute to direct the largest atom-atom distance between any 2 solute molecules along z-axis, and the largest atom-atom distance in the perpendicular plane points along y-axis prior to solvation. The default value of 2.3 Å was used for the minimum solvent to solute distance for the `sim_box` program. A rectangular PBC was used. A cutoff of 0.8 was specified for the Coulomb potential calculation for the `ion` program, which searches for the water molecule with highest potential to be replaced with Na^+ ions. A harmonic force constant of $2.5 \times 10^4 \text{ Nm}^{-1}$ was used to restrain the positions of specified solute atoms.

6.2.4 MD Simulation using AMBER Programs

Apart from the first NVT simulation during equilibration phase and production phase, the initial velocity of all simulations were obtained from the previous runs. Otherwise, the velocity of each atom is sampled from Boltzmann distribution at the specified temperature as mentioned in the main text. The isothermal compressibility of the system was set to $45.75 \times 10^{-6} \text{ bar}^{-1}$. The energy and coordinates were written out every 1000 and 2500 steps, respectively for simulations of less than 10 ns. The frequency of the storage of coordinates is halved for systems that run for 10 ns and above. The center of mass motion was removed every 1000 steps. All molecules were wrapped back into the box during the simulations to avoid the storage failure of large coordinates. No continuum correction was applied to energy and pressure in vdW interactions. A cubic spline function was used as switching function of Coulomb forces. The skin distance for neighbour lists is set to 2 Å.

6.3 Programming Scripts Written

All codes written by the author for the project were made publicly available on GitHub at <https://github.com/Jon-Ting/Honours>.

6.3.1 Gaussian Job Generation and Submission

The Python scripts gaussian.py and settings.py in QM/run_gaussian directory work in conjunction with each other to generate Gaussian input files and submission files compatible with High-Performance Computing (HPC) clusters using PBS Pro workload management system (Raijin, Tinaroo, Awoonga, Argon). The Bash scripts *rajin_sub* and *rcc_sub* in Bash directory allows automatic submission of HPC jobs by looping through all directories.

6.3.2 Management and Modification of Files and Directories

Functions in the Python file admin.py in QM/run_gaussian directory allow simultaneous modification of the names of multiple files or directories, generation of new directories according to existing files, grouping according to naming patterns, and splitting concatenated coordinates files containing multiple molecules into files containing exclusive data of individual molecule. The *rpname()* function in the *.bashrc* file eases the name modification of entries in terminals.

6.3.3 Post-Calculation Correction, Tabulation, and Visualisation of QM Calculation Results

The *SCS_corr()* function in Gaussian.py was written to carry out spin-component-scaled (SCS) correction. The QM data of interest were extracted and tabulated in Excel sheets using tabulate.py in QM/run_gaussian directory. The *plot_fig.py* in QM/visual directory obtains the settings for each figure from *plot_config.py* and generates a graph consisting of one or more subplots for each properties analysed (reactant LUMO energies, β -carbon charges, and distortion-interaction analysis). The interconversions between ΔG and K , and ΔG_{rev}^\ddagger , k_{off} , $t_{1/2}$ and RT values were done using *calculation.py* in QM directory.

6.3.4 Automated Analysis of MD Trajectories

The *post_amber_md()* function in the *.bashrc* file which utilises other functions (*mkvmdtop()* and *find_min_geom()*) was written in multiple programming languages (Bash, Sed and Awk) to automatically carry out the trajectory analysis, which includes stripping out solvent molecules and ions, setting the BTK dimers as the centre of the simulated box, reimaging the solute molecules, calculating RMSD relative to starting structure, generating VMD-compatible topology file, extracting the system properties (temperature, pressure, density, potential and kinetic energies), locating the geometry with minimum potential energy, calculating RMSD with respect to the most stable structure and generating an average structure.

6.3.5 Preparation of MD Systems and Visualisation of MD Trajectory Analysis Results

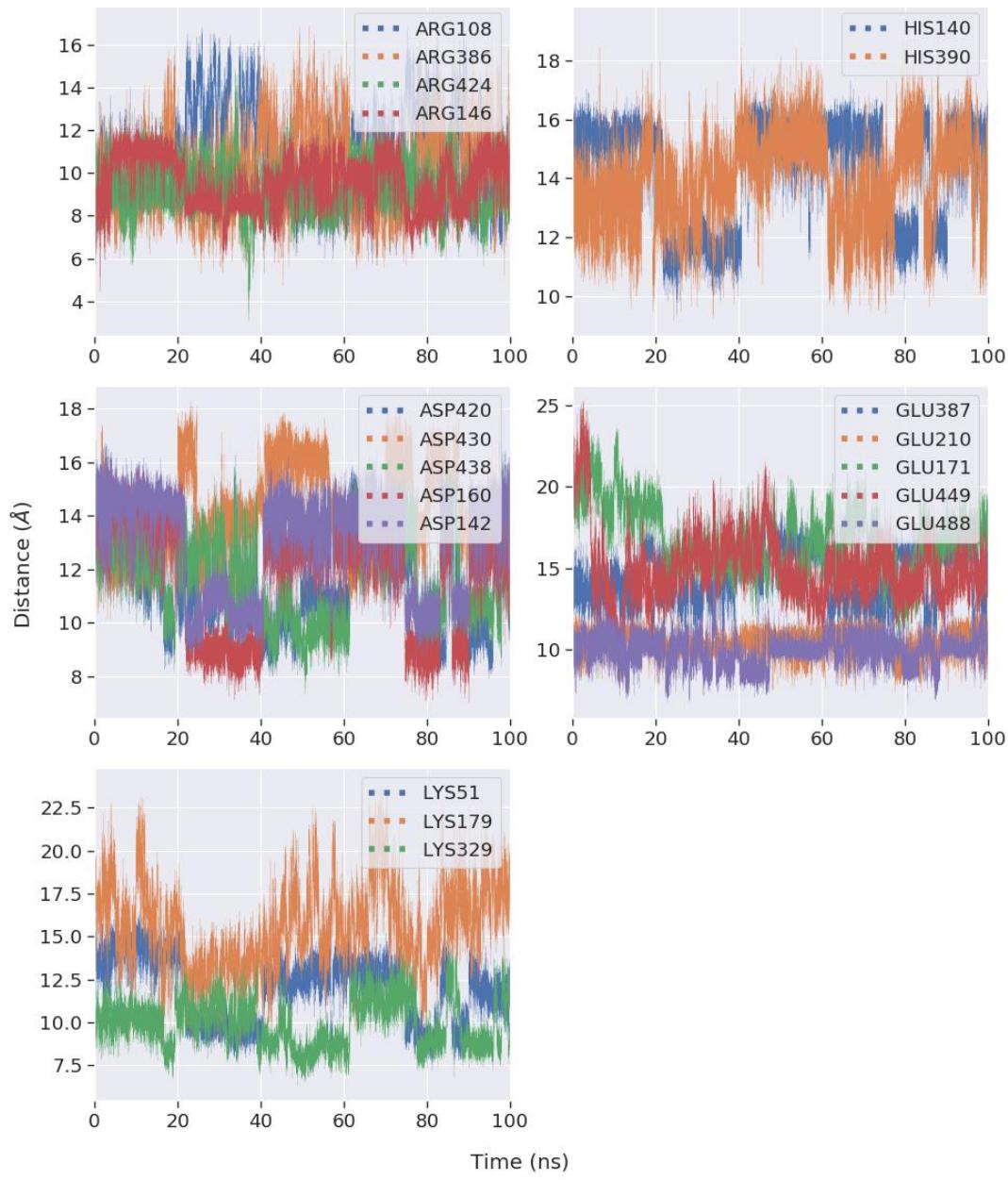
The Python files *prep_mtb.py* and *sum.py* were written to ensure the parameters of the non-standard amino acids, especially the total charges of the corresponding systems are correct. The plotting of the analysis of MD results was done using *base_id.py*, *SC_bond_dist.py*, *bb_rmsd.py*, *hbond_analysis.py*, and *lig_dihedral.py* in the MD directory.

6.4 Plotted Figures from Distance Analyses

In an effort to search for species with the potential of acting as base catalyst for the thiol additions and eliminations, the distances of the basic (histidine, lysine and arginine) and acidic (aspartate and glutamate) residues from the protons to be abstracted as functions of simulation time are measured. Figure 6.1 and 6.2 showed the trajectories of the charged residues that have approached the protons of interest to within 10 Å at any point of the simulations.

6.4.1 Potential Base for Thiol Additions

(a)



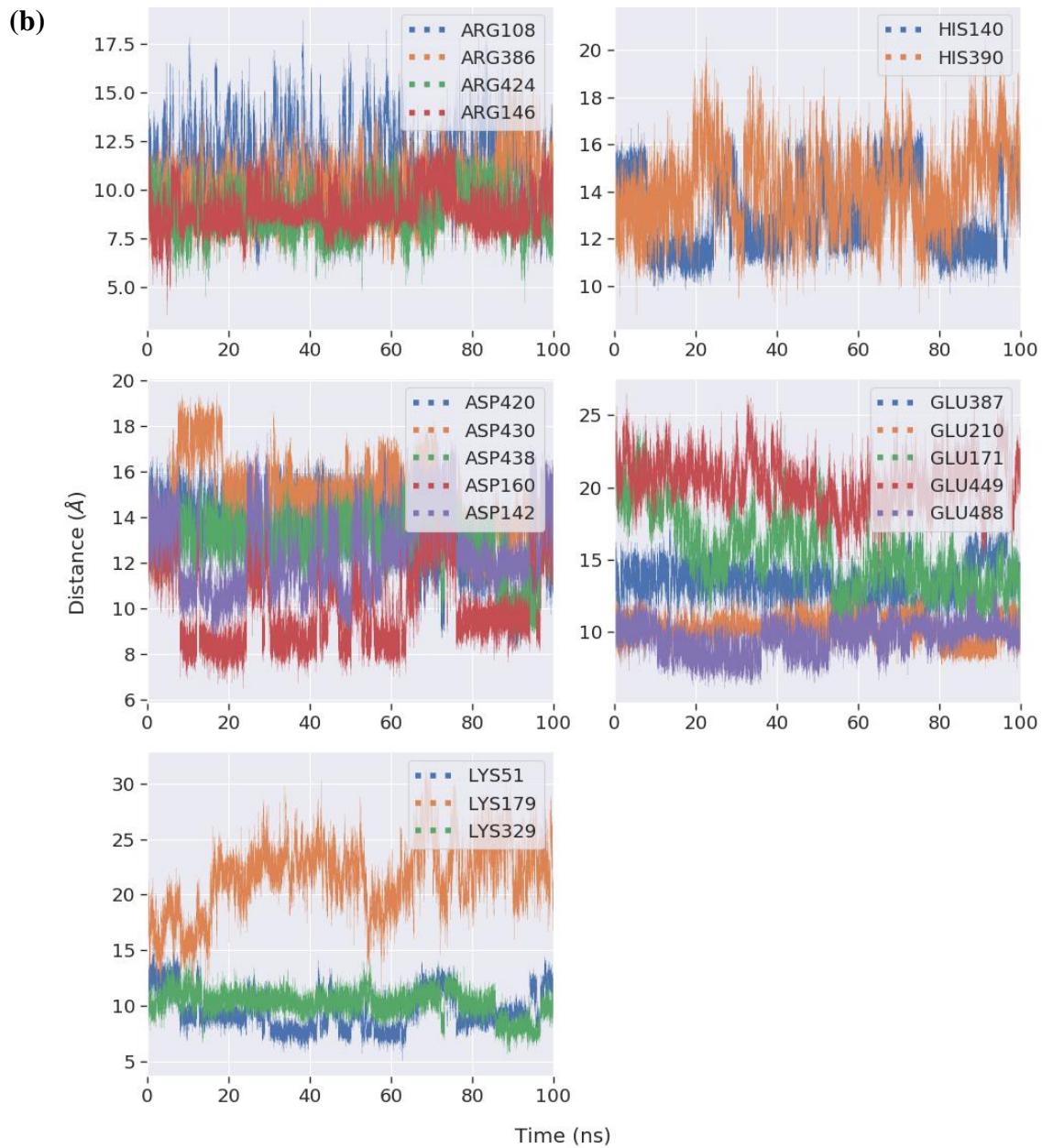
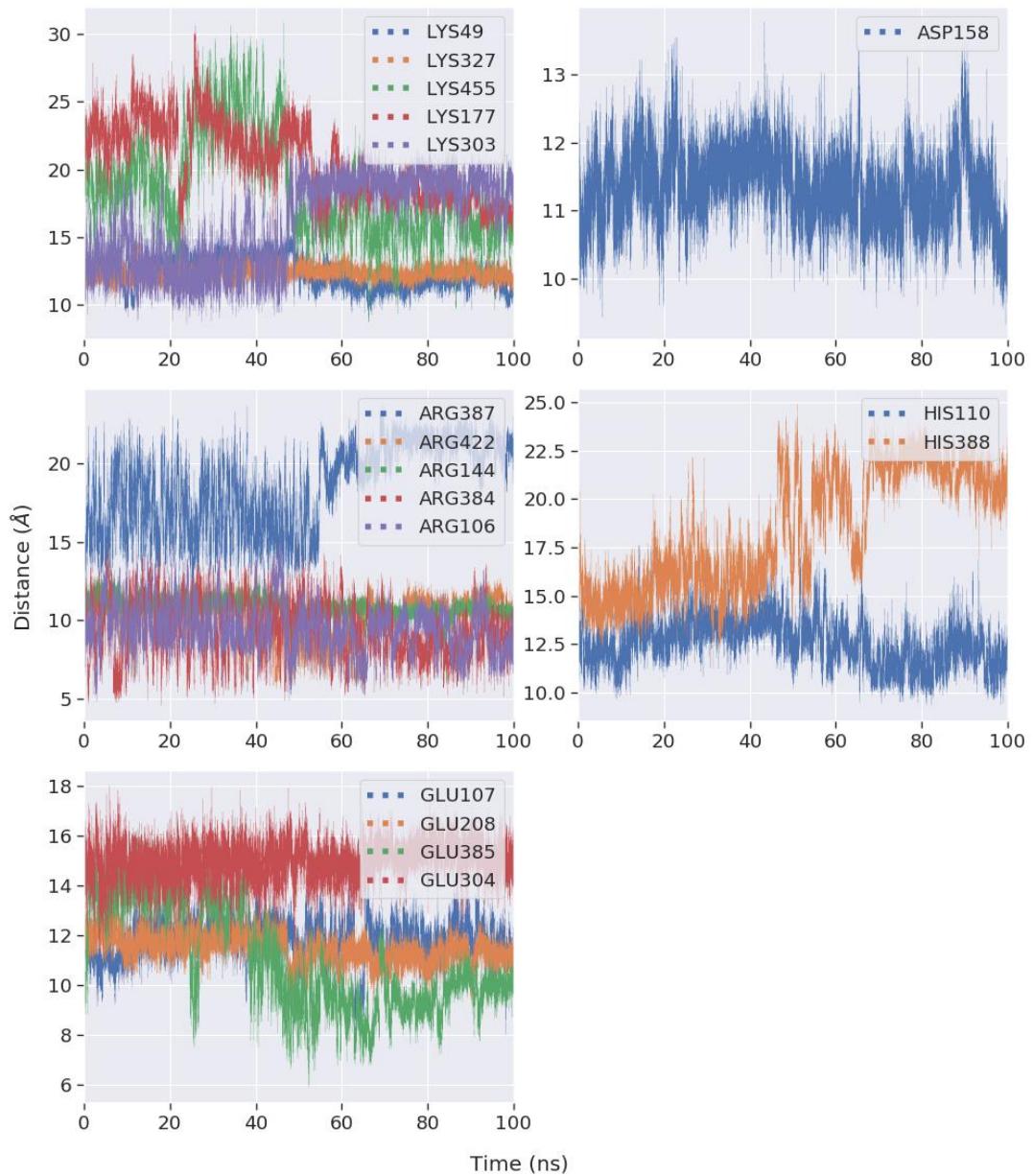


Figure 6.1. Distance between charged residues of interest and the thiol proton of the unreacted Cys481 of the noncovalently bound BTK containing inhibitors **3** in replicates (a) 1 and (b) 2.

6.4.2 Potential Base for Thiol Eliminations

(a)

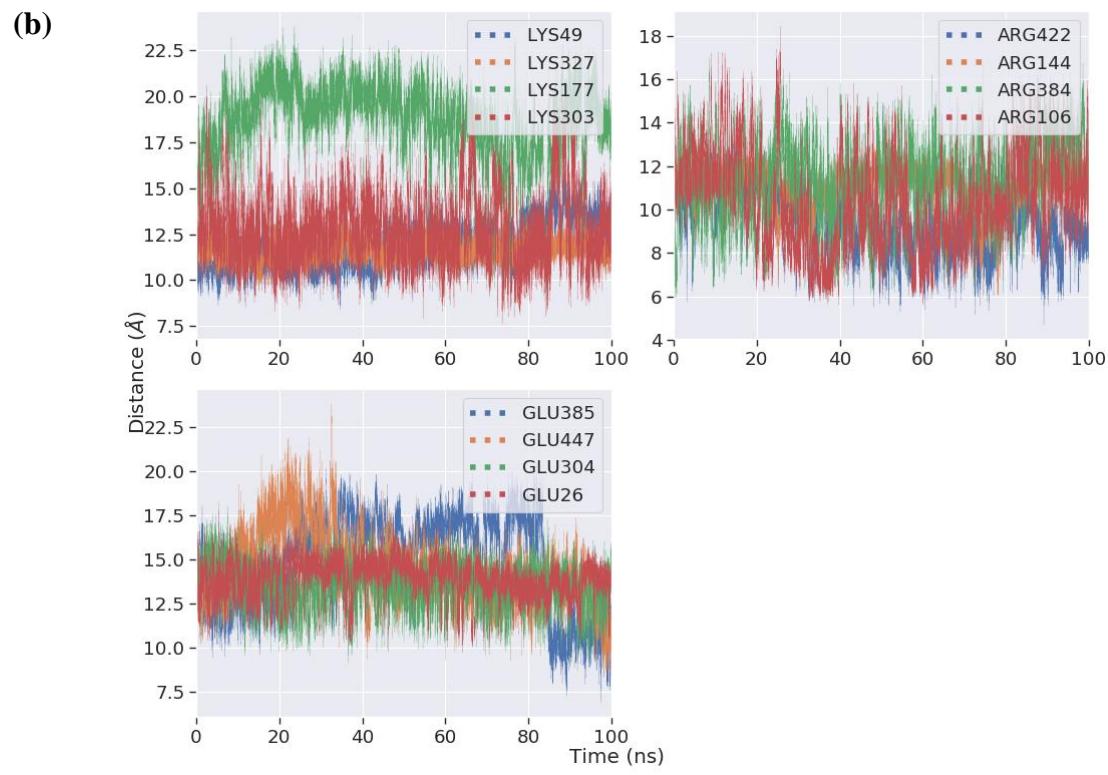


Figure 6.2. Distance between charged residues of interest and the $\text{C}\alpha$ proton of the covalently bound (a) **1** and (b) **3**.