

**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 November 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**

List the significant and substantial inputs made by others to the research, work and writing

represented and/or reported in the research report. These would include significant

contributions to: the conception and design of the project; non-routine technical work; analysis

and interpretation of research data; drafting significant parts of the work or critically revising it

so as to contribute to the interpretation.

*Elizabeth Krenske*:

*Alan Mark*:

*Thomas Lee*:

*Martin Stroet*:

*Romain Lepage*:

Romain provided useful suggestions to aid in the location of some transition state structures.

*Yuk Ping Chin*:

Yuk Ping introduced convenient ways for conversion of molecular coordinate file formats and provided tutorial for some functionalities of GaussView, MacroModel and NCIPLOT programs.

*David Poger*:

David provided critical advices regarding the construction of missing residues in X-ray crystal structures of proteins for simulations and specification of simulation parameters for molecular dynamics studies of proteins under biological condition.

This research was supported by a University of Queensland School of Chemistry and Molecular Biosciences Honours Scholarship.

List the significant and substantial inputs made by others to the research, work and writing represented and/or reported in the thesis. These could include significant contributions to: the conception and design of the project; non-routine technical work; analysis and interpretation of research data; drafting significant parts of the work or critically revising it so as to contribute to the interpretation.

**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 Associate Professor Elizabeth Krenske and Professor Alan Mark for the supervisions throughout the year. I also wish to thank Dr. Martin Stroet and Dr. Thomas Lee for their guidance in MD simulations on Wiener, providence of codes that enables the setup of systems containing non-standard amino acids and simulation of GROMOS systems using graphics processing unit-accelerated AMBER machineries. We are grateful to Jack Taunton and his colleagues for making the co-crystal structure of BTK with inhibitor molecule 3 from their experimental work publicly available. Computer resources were 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.

**Signature of Author: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_**

Acknowledgements recognise those who have been instrumental in the completion of the project. Acknowledgements should include any professional editorial advice received including the name of the editor and a brief description of the service rendered.

free text section for you to record your acknowledgment and gratitude for the more general

academic input and support of your supervisor and colleagues; financial support from grants and

scholarships; and the non-academic support you have received during the course of your

candidature

**Principal Supervisor Agreement**

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

**Signature of Principal Supervisor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Date: \_\_\_\_\_\_\_\_\_**

**Keywords**

bruton’s tyrosine kinase, kinetics, thermodynamics, reversible covalent inhibitors, michael acceptors, cyanoacrylamides, quantum mechanics, molecular dynamics, quantum mechanics/molecular mechanics, density functional theory

# SUMMARY

The abstract should outline the main approach and findings of the thesis and must be between 300 and 800 words.

# TABLE OF CONTENTS

[1 SUMMARY 6](#_Toc20407229)

[2 TABLE OF CONTENTS 7](#_Toc20407230)

[3 LIST OF ABBREVIATIONS 10](#_Toc20407231)

[4 INTRODUCTION 12](#_Toc20407232)

[4.1 Background and Significance 12](#_Toc20407233)

[4.1.1 Covalent Drugs. 12](#_Toc20407234)

[4.1.2 Residence Time and its Relationship to Pharmacological Effects. 13](#_Toc20407235)

[4.1.3 Discovery of Reversible Covalent Inhibitors of Bruton’s Tyrosine Kinase. 14](#_Toc20407236)

[2.1.4 Questions Emerging from Previous Work. 18](#_Toc20407237)

[2.1.5 Computational Studies of BTK. 19](#_Toc20407238)

[4.2 Objectives 21](#_Toc20407239)

[5 DETERMINATION OF THE INTRINSIC REACTIVITY OF THE COVALENT INHIBITORS. 22](#_Toc20407240)

[5.1 Methods 22](#_Toc20407241)

[5.1.1 Rationale for QM Methods Chosen 22](#_Toc20407242)

[5.1.2 Relationship between Parameters of Interest 23](#_Toc20407243)

[5.1.3 Conformational Sampling 24](#_Toc20407244)

[5.1.4 Calculation of Gibbs Free Energy 25](#_Toc20407245)

[5.1.5 Noncovalent Interactions Analysis 27](#_Toc20407246)

[5.2 Conformational Analysis 29](#_Toc20407247)

[5.2.1 Reactants 29](#_Toc20407248)

[5.2.2 Transition State Structures 29](#_Toc20407249)

[5.2.3 Products 31](#_Toc20407250)

[5.3 Benchmarking of Functionals and Basis Sets 33](#_Toc20407251)

[5.4 Single Point Calculations 37](#_Toc20407252)

[5.5 Rationalisation of the Predicted Intrinsic Reactivities 39](#_Toc20407253)

[5.5.1 Reactant Lowest Unoccupied Molecular Orbital Energies 39](#_Toc20407254)

[5.5.2 -Carbon Charges 39](#_Toc20407255)

[5.5.3 Distortion/Interaction Analysis 43](#_Toc20407256)

[6 MOLECULAR DYNAMICS SIMULATIONS OF THE INHIBITED PROTEIN 46](#_Toc20407257)

[6.1 Methods 46](#_Toc20407258)

[6.1.1 Structure Preparation 46](#_Toc20407259)

[6.1.2 Simulation Setup 47](#_Toc20407260)

[6.1.3 Trajectory Analysis 48](#_Toc20407261)

[6.2 Stability of Simulated BTK 49](#_Toc20407262)

[6.2.1 RMSD of Protein Backbones from X-ray Crystal Structure 49](#_Toc20407263)

[6.2.2 Hydrogen Bond Analysis 50](#_Toc20407264)

[6.3 Interactions between Cyanoacrylamides Inhibitors and BTK Active Site Residues 51](#_Toc20407265)

[6.3.1 Distance of Cys481 Sulfur Atom from Electrophilic Carbon on Ligands 51](#_Toc20407266)

[6.3.2 Dihedral Rotations about C=C-C=O Bonds 53](#_Toc20407267)

[6.3.3 Hydrogen Bond Analysis 54](#_Toc20407268)

[6.3.4 Cluster Analysis 54](#_Toc20407269)

[6.4 Identification of Potential Base Species 55](#_Toc20407270)

[6.4.1 Distance from Charged Residues 55](#_Toc20407271)

[6.4.2 Intramolecular Proton Transfer 55](#_Toc20407272)

[7 INVESTIGATION OF THE EFFECT OF BINDING SITE RESIDUES ON THE REACTIVITY. 56](#_Toc20407273)

[7.1 Concept of QM/MM 56](#_Toc20407274)

[7.2 Hybrid QM/MM Studies of Different Kinases 58](#_Toc20407275)

[7.3 Future Direction 59](#_Toc20407276)

[8 CONCLUSION 60](#_Toc20407277)

[9 REFERENCES 62](#_Toc20407278)

[10 APPENDIX 71](#_Toc20407279)

[10.1 Benchmarking of Force Fields for Conformational Sampling 71](#_Toc20407280)

[10.2 Failure of MacroModel to Locate Stable s-*cis* Conformer 71](#_Toc20407281)

[10.3 Comparison of CPU Time between Different Methods 73](#_Toc20407282)

[10.4 Identification of the Most Relevant HOMO for Electrophilicity Index Computation 74](#_Toc20407283)

[10.5 Investigation on TS5 Methylthiolate Distortion 76](#_Toc20407284)

[10.6 QM Conformational Analysis on Results from Method G 77](#_Toc20407285)

[10.6.1 Reactant LUMO Energies 77](#_Toc20407286)

[10.6.2 -Carbon Charges 77](#_Toc20407287)

[10.6.3 Distortion/Interaction Analysis 79](#_Toc20407288)

[10.7 Details of MD Parameters Used 80](#_Toc20407289)

[10.7.1 Preparation of GROMOS System 80](#_Toc20407290)

[10.7.2 Simulation of AMBER System 81](#_Toc20407291)

[10.8 Programming Scripts Written 81](#_Toc20407292)

[10.8.1 Gaussian Job Generation and Submission 81](#_Toc20407293)

[10.8.2 Management and Modification of Files and Directories 82](#_Toc20407294)

[10.8.3 Post-Calculation Correction, Tabulation, and Visualisation of QM Calculation Results 82](#_Toc20407295)

[10.8.4 Automated Analysis of MD Trajectories 82](#_Toc20407296)

[10.8.5 Preparation of MD Systems and Visualisation of MD Trajectory Analysis Results 83](#_Toc20407297)

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| **ATB** | Automated Topology Builder |
| **BTK** | Bruton’s tyrosine kinase |
| **CPU** | Central processing unit |
| **Cys** | Cysteine |
| **DFT** | Density functional theory |
| **EGFR** | Epidermal growth factor receptor |
| **ESP** | Electrostatic potential |
| **FDA** | Food and Drug Administration |
| **GBSA** | Generalised Born |
| **GGA** | Generalised Born model augmented with the hydrophobic solvent accessible surface area |
| **HF** | Hartree-Fock |
| **ifp** | Interaction function parameter |
| **ITK** | Interleukin-2-inducible T-cell kinase |
| **LUMO** | Lowest unoccupied molecular orbitals |
| **MAD** | Mean absolute deviation |
| **MD** | Molecular dynamics |
| **MM** | Molecular mechanics |
| **MOPAC** | Molecular Orbital Package |
| **MTLMS** | Mixed torsional/low-mode sampling |
| **mtb** | Molecular topology building block |
| **NCI** | Noncovalent interactions |
| **PDB** | Protein Data Bank |
| **PT2** | Second-order perturbation correlation term |
| **PVED** | parity-violating energy difference |
| **QM** | Quantum mechanics |
| **QSAR** | Quantitative structure-activity relationship |
| **RMSD** | Root-mean-square deviation |
| **RT** | Residence time |
| **SAR** | Structure-affinity relationships |
| **SKR** | Structure-kinetic relationships |
| **TS** | Transition state |
| **TXK** | T-cell X chromosome kinase |
| **vdW** | Van der Waals |

# INTRODUCTION

## Background and Significance

### Covalent Drugs.

The ability of an inhibitor molecule to modify the activity of a target enzyme is largely dependent on the strength of the interaction formed between the inhibitor and the enzyme. Covalent inhibitors are molecules that inhibit their target proteins by forming covalent attachments to them.[1](#_ENREF_1) They complement conventional (noncovalent) inhibitors by enabling achievement of much higher binding affinities to their targets.[2](#_ENREF_2) Such capability opens up the possibility for lower drug dosages and dose frequencies in the treatment of diseases, potentially allowing covalent drugs to attain favourable safety profiles, provided that an acceptable target selectivity is achieved.[3](#_ENREF_3)

The structures of several common covalent drugs are shown in Figure 1. As depicted, covalent inhibitors typically bear a reactive electrophilic functional group, called a warhead, which forms a covalent bond rapidly with a nucleophilic residue at the target site following proper positioning of the molecule through noncovalent binding to the target protein. The specificity of the inhibition could be enhanced by targeting noncatalytic residues,[4](#_ENREF_4) which are usually distinct across different enzymes.[3](#_ENREF_3),[5](#_ENREF_5) One important class of warheads used in covalent drugs are compounds containing α,β-unsaturated carbonyls, also known as Michael acceptors. The mechanism of action of Michael acceptors as covalent drugs is exemplified in Scheme 1. Once the inhibitor enters the binding site, the nearby deprotonated cysteine thiol group undergoes conjugate addition to the electrophilic Michael acceptor. For both thiol addition and elimination, a base species acts as a catalyst by deprotonating the thiol or adduct, increasing their reactivities and thus lowering the activation barrier for the formation and dissociation of the S-C bond, respectively.



**Figure 1.** Structures of well-known covalent inhibitors. The bond-forming functional groups are highlighted in red.

**Scheme 1. General Reaction Scheme for Thiol Addition to a Michael Acceptor.**



Despite the advantages of well-known covalent inhibitors such as aspirin,[6](#_ENREF_6) penicillin,[7](#_ENREF_7) and fosfomycin,[8](#_ENREF_8) most covalent drugs were rarely designed deliberately as covalent inhibitors, due to fear of side effects.[3](#_ENREF_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,[9](#_ENREF_9),[10](#_ENREF_10) drug-induced liver injury,[11](#_ENREF_11),[12](#_ENREF_12) and multiple haematological disorders.[13](#_ENREF_13) Reversible inhibitors are normally favoured over irreversible ones in drug design as they are less likely to form permanent adducts with off-target proteins containing homologous or highly-reactive residues.[14-16](#_ENREF_14) On top of that, reversible inhibition also allows fine tuning of a critical parameter, the importance of which has emerged over the last decade, known as the drug-target residence time (RT), ,[17](#_ENREF_17) which is defined as the reciprocal of off-rate constant , as shown in equation (1). This feature allows for tailoring of duration of effect by reversible drugs of varying RT.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

### 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, and the half-maximal inhibitory concentration, IC50.[17](#_ENREF_17) However, unlike an *in vitro* setting, the dynamic 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.[18](#_ENREF_18),[19](#_ENREF_19) Due to the fundamental differences between open and closed systems, binding kinetics parameters, especially RT, have grown important as optimisation measures for drug candidates alongside binding affinity and potency.[17-23](#_ENREF_17) 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.[17](#_ENREF_17),[18](#_ENREF_18),[22](#_ENREF_22) For example, Swinney reported that therapeutics with long RT demonstrated good clinical efficacy and potentially involved non-equilibrium conditions in their mechanism of action.[24](#_ENREF_24) That being said, it should be underlined that a long RT is not always required or desired. While sustained target engagement is required for the treatment of cancers,[21](#_ENREF_21) allergies,[21](#_ENREF_21) hypertension,[25](#_ENREF_25) and hormone-dependent diseases,[26](#_ENREF_26) 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,[21](#_ENREF_21),[27](#_ENREF_27) such as thrombosis[17](#_ENREF_17),[28](#_ENREF_28) and antipsychosis.[18](#_ENREF_18) In fact, long RT may even be contraindicated for these clinical indications due to the convolution of the effect of drug occupancy at the same site on both efficacy and toxicity.[21](#_ENREF_21) As such, the ability to tune the drug-target RT has been a major focus within the covalent drug discovery community.

### Discovery of Reversible Covalent Inhibitors of Bruton’s Tyrosine Kinase.

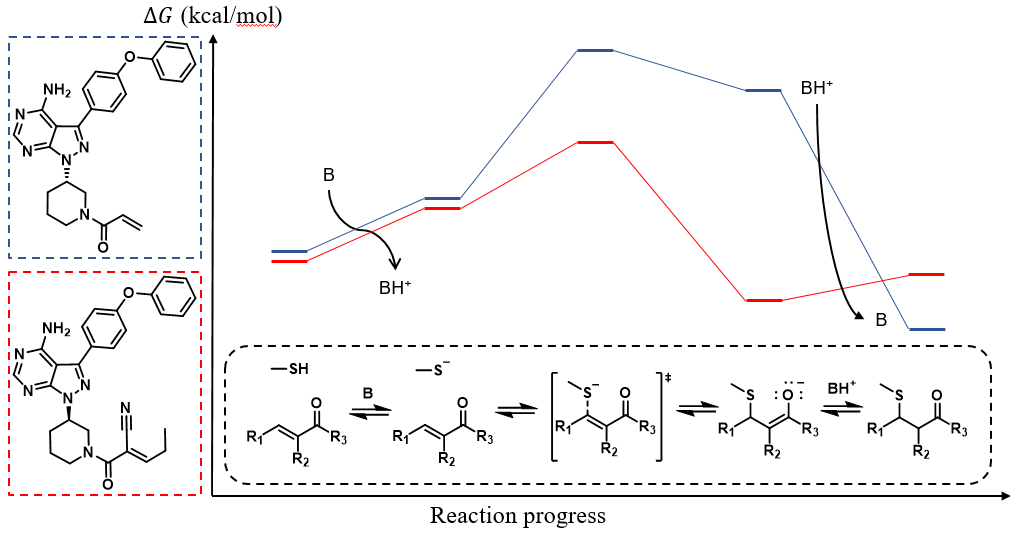
In order to counteract the abovementioned dangers of irreversible binding of covalent inhibitors to enzymes, new classes of covalent inhibitors that engage their targets reversibly have been designed. In this context, Taunton et al. reported an important discovery of reversible covalent inhibitors targeting noncatalytic cysteine residues in Bruton’s Tyrosine Kinase (BTK).[4](#_ENREF_4) BTK is a member of the Tec tyrosine kinase family which participates in immune function regulation through B-cell development[29](#_ENREF_29),[30](#_ENREF_30) and its inhibition has been shown to allow treatment of various cancers[31](#_ENREF_31),[32](#_ENREF_32) and autoimmune diseases.[33](#_ENREF_33),[34](#_ENREF_34) Ibrutinib[35](#_ENREF_35) (Figure 2) was the first BTK inhibitor approved by the Food and Drug Administration (FDA) as a medicament for chronic lymphocytic leukaemia,[31](#_ENREF_31) and is used to treat mantle cell lymphoma[32](#_ENREF_32) as a second-line treatment.[36](#_ENREF_36) Ibrutinib consists of a kinase-recognition scaffold linked to an acrylamide warhead, which is capable of forming covalent bonds with appropriate nucleophiles.



**Figure 2.** Structures of ibrutinib and acalabrutinib. The Michael acceptor moieties and the kinase-recognition scaffolds are highlighted in red and blue, respectively.

However, ibrutinib does not have optimal selectivity. It led to permanent inhibition of analogous kinase targets[34](#_ENREF_34),[37](#_ENREF_37) such as epidermal growth factor receptor (EGFR), interleukin-2-inducible T-cell kinase (ITK) and T-cell X chromosome kinase (TXK), causing severe adverse events including atrial fibrillation, major haemorrhage, and arthralgia.[32](#_ENREF_32),[38](#_ENREF_38) To date, acalabrutinib[39](#_ENREF_39) is the only other approved inhibitor of BTK. Even though it exhibits better kinase selectivity compared to ibrutinib,[40](#_ENREF_40),[41](#_ENREF_41) the irreversible nature of its binding interaction[41](#_ENREF_41) naturally triggers concerns regarding the off-target effects that are yet unknown when applied chronically as typically required for the treatment of autoimmune diseases.[42](#_ENREF_42) Accordingly, this prompted the exploration of reversible inhibition of BTK as an alternative solution.

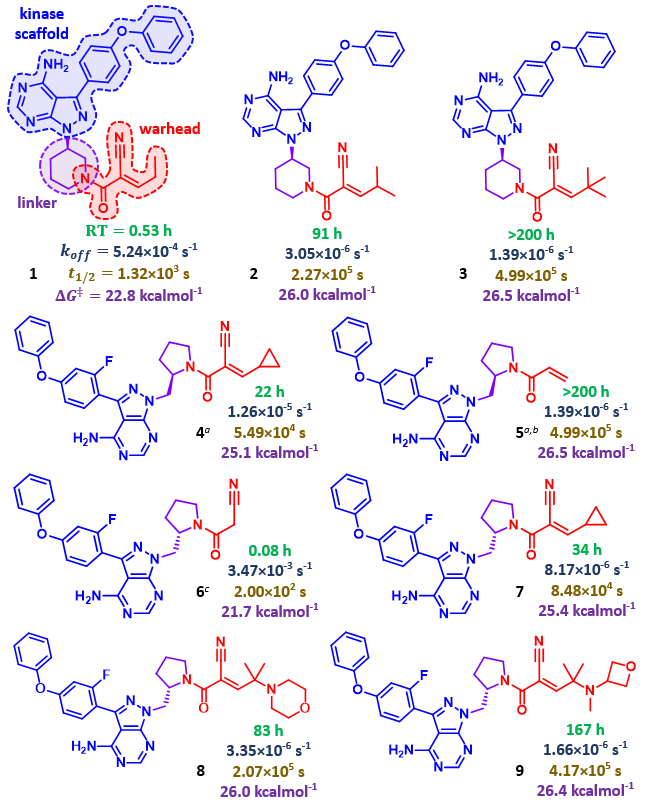
The inhibitor molecules investigated by Taunton et al. resembled ibrutinib except that cyanoacrylamides were used as warheads in place of acrylamides and the absolute configurations of the stereogenic centres were inverted. This rationale is based on the observations in their earlier chemical experiments, where the installation of the nitrile groups at the position of the acrylamide warheads resulted in rapidly reversible Michael additions.[5](#_ENREF_5) Quantum mechanical (QM) calculations revealed that this is because the electron-withdrawing substituent both stabilises the anionic transition state (TS) and destabilises the neutral adduct, thus decreasing the for binding and accelerating the reverse reaction, elimination of thiol, as illustrated in Figure 3.[43](#_ENREF_43)

****

**Figure 3.** Representative free energy profiles for reversible (red) and irreversible (blue) Michael additions.

By varying the electronic and steric environment around the warhead using different -substituents on the cyanoacrylamides, Taunton et al. discovered inhibitors exhibiting RT ranging from just minutes to one week. Some examples are presented in Chart 1. A remarkable outcome was the extrapolation of long RT to durable *in vivo* pharmacodynamic inhibition, where **9** demonstrated sustained BTK occupancy after the clearance of unbound drugs from the systemic circulation in rodent models. As commented by Copeland in a recent review,[44](#_ENREF_44) the achievements of Taunton’s group unlocked the possibility to systematically approach the holy grail of covalent inhibitor design, i.e. matching the biological requirement of targeted proteins to ligand-receptor binding interactions through rational tuning of the structural features of the inhibitors. Following their pioneering work, the incorporation of structure-kinetic relationships (SKR) apart from the traditional structure-affinity relationships (SAR) in cell-based assays has been encouraged owing to the awareness raised concerning the relevance of binding kinetics in drug discovery.[23](#_ENREF_23),[44-55](#_ENREF_44) Many research groups have followed up on these ideas to advance the exploration of kinase inhibition by covalent means.[56-65](#_ENREF_56)

Chart 1. Structures and Binding Data of Reversible Covalent Inhibitors of BTK Designed by Taunton et al.[4](#_ENREF_4)



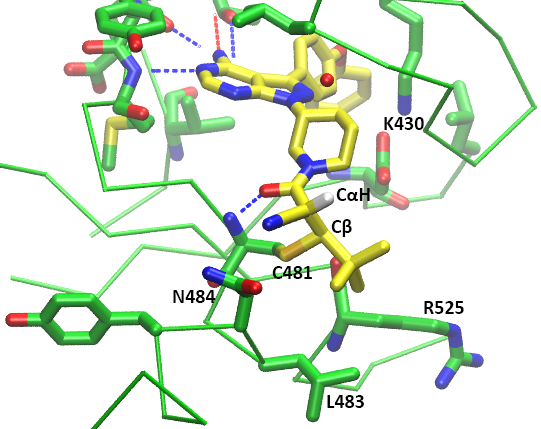
*a***4** and **5** have R linker configurations instead of S. *b***5** contains an acrylamide electrophile instead of cyanoacrylamide. *c***6** does not contain Michael acceptor.

### 2.1.4 Questions Emerging from Previous Work.

What remains unclear regarding these reversible cyanoacrylamides is how their different structural components govern the reversibility of their binding to BTK. Furthermore, **9** had also been found to show exceptional selectivity for BTK over other kinases; the molecular basis for the selectivity is yet to be uncovered.

In a computational study, Taunton et al. reported[66](#_ENREF_66) an inverse correlation between the -elimination rates of the thiol adducts and the proton affinities of the intermediate enolate carbanions, suggesting the possibility of using computed proton affinities to predictably tune the intrinsic reversibility of the thiol addition. In spite of this, examination of the experimental data in Chart 1, especially the comparison between the RT of enantiomers **4** and **7**, reveals that acidity is not the only important criterion; other factors such as steric hindrance, the impact of substituents on the overall electrophilicity, and interactions with the BTK binding site, must also contribute to determining their overall behaviour. Unravelling the factors affecting the RT of the inhibitors is expected to be useful in understanding not only BTK inhibition but also for designing inhibitors of related cysteine-containing enzymes. In fact, 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 on a co-crystal structure of BTK covalently bonded to **3** as depicted in Figure 4. They observed that the C hydrogen was solvent-exposed and two hydrophobic patches were identified in the vicinity of the tert-butyl group of **3**. Taunton et al. thus suggested that the C hydrogen, which needs to be removed as the first step of thiol elimination, was shielded from the participating base and that the hydrophobic interactions contributed to the stabilisation of the product. The minimal overlap between the C=O bond and the C-H bond was also proposed to have reduced the acidity of the hydrogen both thermodynamically and kinetically.

However, the species that acted as the base was not identified. It should also be pointed out that X-ray structures merely present an average, static picture of the dynamic, structurally diverse ensembles in the crystals.[67](#_ENREF_67),[68](#_ENREF_68) This limitation recommends careful interpretation of X-ray structures and incorporation of other measures to better understand the protein dynamics.[69](#_ENREF_69),[70](#_ENREF_70) These effects are yet to be examined by simulations at an atomistic level.



**Figure 4.** Co-crystal structure of a reversible covalent BTK inhibitor containing a cyanoacrylamide bound to BTK at 2.2-Å resolution.[4](#_ENREF_4) The covalent bond between Cβ and Cys481, the acidic CαH hydrogen, and the hydrogen bonds possibly contributing to geometry stabilisation indicated correspond to those proposed to help stabilise the geometry.

### 2.1.5 Computational Studies of BTK.

Numerous computational studies have been performed on BTK, in many cases using molecular dynamics (MD) simulations. For instance, combination of MD simulations and 3D quantitative structure-activity relationship (QSAR) models allowed highly efficient screening of noncovalent BTK inhibitors based on binding affinities.[71-75](#_ENREF_71) Related MD studies have helped to improve the understanding of the biological functions of different domains of BTK,[76](#_ENREF_76),[77](#_ENREF_77) the features essential for maintenance of its inactive state such as the conservation of an isoleucine residue,[78](#_ENREF_78) and the interactions of BTK with various phosphatidylinositols,[77](#_ENREF_77),[79](#_ENREF_79) which are its natural substrates. To date, only a single study associated with BTK has utilised QM calculations.[71](#_ENREF_71) This work employed density functional theory (DFT) to construct 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.[80](#_ENREF_80),[81](#_ENREF_81) Fortunately, this suboptimal choice was attenuated by the fact that only the geometry of the ligands were optimised, allowing the avoidance of observing 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, these studies disclose many possibilities of gaining atomistic insight into BTK inhibition using appropriate computational tools.

## Objectives

The ultimate goal of the present work is to elucidate the factors that affect the overall kinetics and thermodynamics of thiol additions in the context of the binding of BTK Cys481 to cyanoacrylamide-containing Michael acceptors. The intention is to perform the first ever study of BTK binding site environmental effects on the reactivities of these inhibitor molecules, with the hope that a deeper understanding of the binding interactions will enable more rational design of BTK inhibitors in the future. Such an understanding would hopefully also facilitate the design of reversible covalent inhibitors for other kinases. Given the importance of the tunability of *in vivo* RT, it is believed that these findings will contribute to the advancement of the still maturing field of protein kinase covalent inhibitor discovery.[47](#_ENREF_47),[82](#_ENREF_82),[83](#_ENREF_83) These goals were approached by first performing QM calculations to determine the intrinsic reactivities of different cyanoacrylamide warheads towards a model thiol, followed by MD studies to explore the impact of residues near BTK binding site on the reactivities of the inhibitors.

# DETERMINATION OF THE INTRINSIC REACTIVITY OF THE COVALENT INHIBITORS.

## Methods

5 inhibitors were selected to be studied in detail to reveal the structural factors driving the intrinsic reactivities of the warheads. **3** was chosen due to the availability of the crystal structure and **1** was chosen to be compared with it. The enantiomer pairs **4** and **7** could potentially demonstrate the effect of the stereogenic centre on the linkers. The acrylamide **5** was chosen as a control as it exhibited irreversible BTK inhibition while **9** was needed for the study on kinase selectivity. As the computational expenses grow exponentially with increasing number of atoms, the scaffolds of the inhibitors were truncated as their effects on the reactivity are likely to be minimal. The truncated inhibitors will be differentiated from the original molecule by adding the prefix **R** to the compound number, with a special case of **R47** for the truncated inhibitors **4** and **7** as the simplification renders them identical.

### Rationale for QM Methods Chosen

The first computational studies of thiol-Michael additions were performed using *ab initio* methods by Kollman et al.[84](#_ENREF_84) Subsequently, several related studies[85-87](#_ENREF_85) have been conducted using DFT methods instead due to their more favourable computational cost to accuracy ratio.[88](#_ENREF_88) In this regard, a critical remark was made by Smith et al.[80](#_ENREF_80) who noted that several popular DFT functionals returned spurious results for thiol additions due to delocalisation errors, where an artificial spreading of the electron density of several DFT functionals results from the inconsistent correction of the self-repulsion in the dominant Coulomb functional.[89](#_ENREF_89) In fact, the observation of artefactual charge-transfer complexes stabilised by the excessively delocalised electron density distributions had led the researchers unaware of the problem to propose alternative mechanisms for thiol addition, including a concerted water-mediated addition[90](#_ENREF_90) and a direct 1,2-addition.[91](#_ENREF_91) That being said, it is beyond question that proper employment of QM methods is critical to gaining reliable insights into the intrinsic properties of the reactions. It is important that only QM methods that are known to perform reliably for thiol Michael additions be used and special attention ought to be given to the mitigation of the delocalisation errors. The development of range-separated functionals,[92-95](#_ENREF_92) 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, has proved to be effective in correcting the aforesaid error and is therefore investigated in this project.[80](#_ENREF_80)

### Relationship between Parameters of Interest

The application of statistical mechanics to QM systems allows the derivation of thermodynamic quantities at particular standard states including Gibbs free energy, from the partition functions constructed.[96](#_ENREF_96) The activation barrier () and Gibbs free energy change () for the addition of a thiol to a Michael acceptor (or elimination) can therefore be estimated from QM calculations on reactants, TS and products. The rate constant for thiol elimination (off-rate), , can be computed using equation (2), known as the Eyring equation:[97](#_ENREF_97)

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where is absolute temperature, is the elimination activation barrier calculated by DFT, , and are Boltzmann’s constant, Planck’s constant and gas constant, respectively, and is a transmission coefficient, which reflects the fraction of the molecules overcoming the activation barrier that proceeds to the product without recrossing the TS. Assuming that the no-recrossing assumption of the TS theory[97](#_ENREF_97) is held perfectly, a value of 1 is often assigned to . The calculated results are compared with the experimentally measured half-life, , which is related to by a factor of for first order reactions as shown in equation (3):

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

The , , and values are calculated for thiol additions to Taunton’s cyanoacrylamide warheads (Chart 1). To minimise the computation required for the expensive QM calculations, the relatively unimportant structural features of the inhibitors to the study of the warhead reactivities such as the kinase recognition scaffold were truncated.

### Conformational Sampling

The reactivity of a particular molecule is highly dependent on its geometry. Since molecules at non-zero temperature constantly undergo changes in their molecular geometries, the measured properties are therefore always 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):

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where is the expected number of particles in the single-particle microstate , is the total number of particles in the system, is the energy of microstate , is gas constant, is the equilibrium temperature of the system, and is the total number of microstates. As such, the relevant conformers for the key species involved along the reaction coordinates, 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 MM to enable efficient exploration of the large conformational space. This was done using mixed torsional/low-mode sampling (MTLMS) method in MacroModel,[98](#_ENREF_98) which had been proven to be the best performing sampling method for flexible compounds.[99](#_ENREF_99) The algorithm applies Monte-Carlo sampling to the torsional degrees of freedom while the energy minima are searched for along the directions of the low-frequency vibrational mode eigenvectors. An initial structure of each reactant, product, and TS structure involved was built using Avogadro. A pair of diastereomers were generated for each product as 2 chiral centres were generated during the thiol additions except **R5**, which is not prochiral. Despite the weak neutral current mechanism that generates the parity-violating energy difference (PVED) even between enantiomer pairs,[100](#_ENREF_100) the difference is diminutive and thus enantiomers were assumed to have equivalent in this work. For TS, both *syn* (C-S--C=C dihedral angle < 90º) and *anti* conformations for the thiol addition from both *Re* and *Si* faces were built, resulting in a total of 4 conformations for each species. In the cases where the orientation of the conformers returned from the conformational sampling was reversed, manual adjustment was applied to give the desired orientation. The structures were then QM optimised and used as the input geometry for the conformational sampling. The initial QM geometry optimisation is especially important for TS to approximate the length of the S-C bonds prior to the searching of conformers.

The molecular interactions were described using the OPLS3e force field.[101](#_ENREF_101) The force field was recommended by the developer due to its more extensive parameterisation and reduced parameter transferability errors. A benchmarking was conducted and discussed in the Appendix. 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 first 15 conformers with a mean absolute deviation (MAD) of at least 0.5 Å from previously located energy minima within an energy window of 3 kcal/mol were stored from a sampling of 1000 maximum steps and 100 steps per rotatable bond. 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 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. The parameters chosen were mostly the default values of MacroModel conformational sampling job configuration, except for the range of the energy window.

### Calculation of Gibbs Free Energy

The MM-optimised structures were then reoptimised using QM methods from which the thermodynamic parameter of interest, , could be obtained from equation (5):

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

where and are absolute temperature and entropy, respectively, while is enthalpy, which is computed from the addition of pressure-volume work, to the internal thermal energy, as shown in equation (6):

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

The geometry reoptimisation aimed to rank the energetic stability of the conformers to identify the most stable conformation. However, the large number of conformers to be optimised inevitably limited the options of QM methods to those that could capture the essential components of thiol addition systems yet are reasonably low in both time and computing resource expenses. From a previous benchmarking done on similar thiol addition system,[43](#_ENREF_43) the M06-2X/6-31+G(d) method had been found to yield rather accurate values of , and therefore had been employed here. Once the most stable conformer for each species was identified, values at a higher level of theory could be computed by only calculating the values of the fixed geometries as shown in equation (7):

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

where and correspond to higher and lower level of theory, respectively, the assumption being negligible change in over the course of reaction.

To choose a suitable combination of functional and basis set prior to the single point energy (SPE) calculations, a benchmarking was carried out for 10 different methods as shown in Table 1. The combinations SCS-MP2/6-31+G(d), B2PLYP-D/6-31+G(d) and M06-2X/6-311G(2d,p) hadperformed well in a solution phase study of thiol addition in comparison to high level ab initio CBS-QB3 calculations,[102](#_ENREF_102) while ωB97X-D/aug-cc-pVTZreturned high accuracy in a gas phase study of thiol addition in comparison to CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ calculations.[80](#_ENREF_80) M06-2X/6-311+G(d,p) was found to best describe the thermodynamic parameters in a solution phase study of thiol addition in comparison to experimental data.[43](#_ENREF_43) Combinations of different functionals with the largest basis set, namely 6-311G(2d,p), were tested. It is expected that the performance of DFT methods (M06-2X and ωB97X-D) do not scale with the size of the basis set, while the accuracy of SCS-MP2 calculations should increase following convergence to complete basis set. The double hybrid functional B2PLYP-D involves both the second-order perturbation correlation term (PT2) and HF exchange, therefore could potentially perform better with larger basis sets (depending on the offset due to divergence from the optimum basis set?). The recently popularised range-separated functional ωB97X-D, which was found to describe thiol additions to small molecules accurately, was also tested with different basis sets.

**Table 1. Combinations of functionals and basis sets chosen for benchmarking.**

|  |  |  |
| --- | --- | --- |
| Method | Functional | Basis set |
| **A** | B2PLYP-D | 6-311G(2d,p) |
| **B** | M06-2X | 6-311G(2d,p) |
| **C** | SCS-MP2 | 6-311G(2d,p) |
| **D** | ωB97X-D | 6-311G(2d,p) |
| **E** | ωB97X-D | aug-cc-pVTZ |
| **F** | ωB97X-D | 6-311G(d,p) |
| **G** | ωB97X-D | 6-311+G(d,p) |
| **H** | B2PLYP-D | 6-31+G(d) |
| **I** | M06-2X | 6-311+G(d,p) |
| **J** | M06-2X | 6-311G(d,p) |

Gaussian 16[103](#_ENREF_103) was used to carry out all DFT calculations on ultrafine integration grids. Harmonic vibrational frequency calculations indicated whether stationary points were minima or saddle points while providing unscaled zero-point energy and thermal corrections. values are reported at a standard state of 1 mol/L and 25 ºC. The molecular orbitals computed in Gaussian 16 were visualised using GaussView 6.[104](#_ENREF_104) Programming scripts were written to enable batch generation and submission of QM calculation jobs, efficient tabulation of data, and data visualisation. The description of the codes is included in the Appendix.

### Noncovalent Interactions Analysis

Noncovalent interactions (NCI) between and within the molecules were visualised as isosurface in real space using NCIplot,[105](#_ENREF_105) where promolecular densities are analysed to identify the strength and type of NCI. The types of interaction were distinguished based on the sign of the second eigenvalue of the electron-density Hessian matrix, while the strength was assessed based on the density. A colour scale of -3 to 3 was used to colour the constructed surfaces.

## Conformational Analysis

The most stable conformers of the key species were tabulated in Table 2.

**Table 2. Most Stable Conformers.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Inhibitor | Reactants | TS Structures | Intermediates | Products |
| **R1** |  |  |  |  |
| **R3** |  |  |  |  |
| **R47** |  |  |  |  |
| **R5** |  |  |  |  |
| **R9** |  |  |  |  |

### Reactants

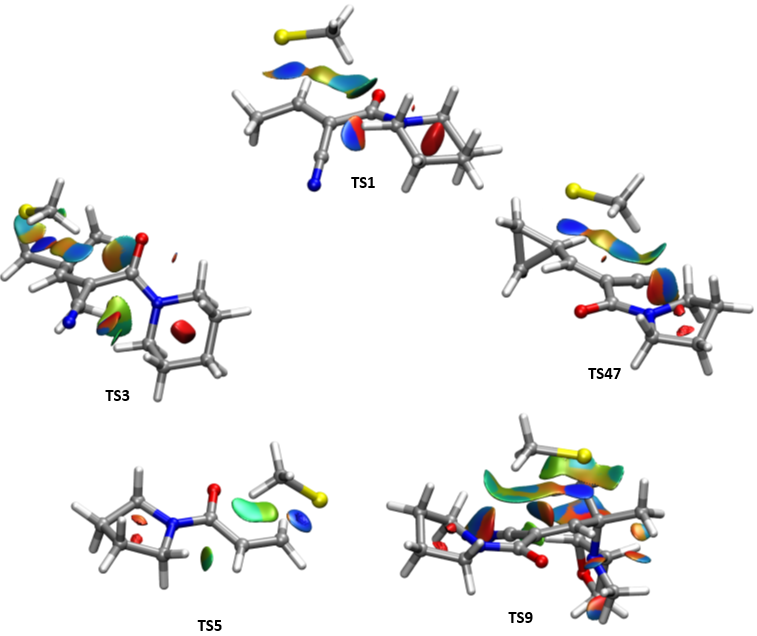
For the reactants, the conformational sampling conducted by MacroModel did not manage to locate any s-*cis* conformer for **R1** within an energy window of 5.02 kcal/mol. An s-*cis* conformer was thus built manually in Avogadro and geometry-optimised using QM. Surprisingly, the optimised s-*cis* conformation was more stable than all other conformers found from the MM-guided sampling. The reporting of this finding would hopefully serve as a warning against using standard chemistry programs as black boxes and encourage consistent inspections of generated data with chemical intuition. Such issue implies that data visualisation is likely one of the most important aspects currently in the field of computational chemistry due to its still early stage of methodology development. An investigation into the reasons behind the apparent failure of MacroModel to locate the global minima was conducted. Further discussions on this issue could be found in the Appendix. For the reactants of **R1**, **R47**, and **R5**,the most stable conformers have s-*cis* geometries. The most stable conformer for **R3** has a C=C-C=O dihedral angle of 90º while an s-*trans* conformer of **R9** was predicted to be the global minima by MM. The NCI within the molecules are shown in Figure 5. The underlying reason for the s-*cis* conformations to be more favourable for R1, R47 and R5 could be understood to be the stabilising C-H··· interactions between the electron cloud of the nitrile group and methyl protons of the .



**Figure 5.** Visualisation of the NCI of the reactants. A blue-green-red scale was used to indicate the strength and type of the interactions, with blue signifying strong attractive interactions while red indicates strong nonbonded overlap.

### Transition State Structures

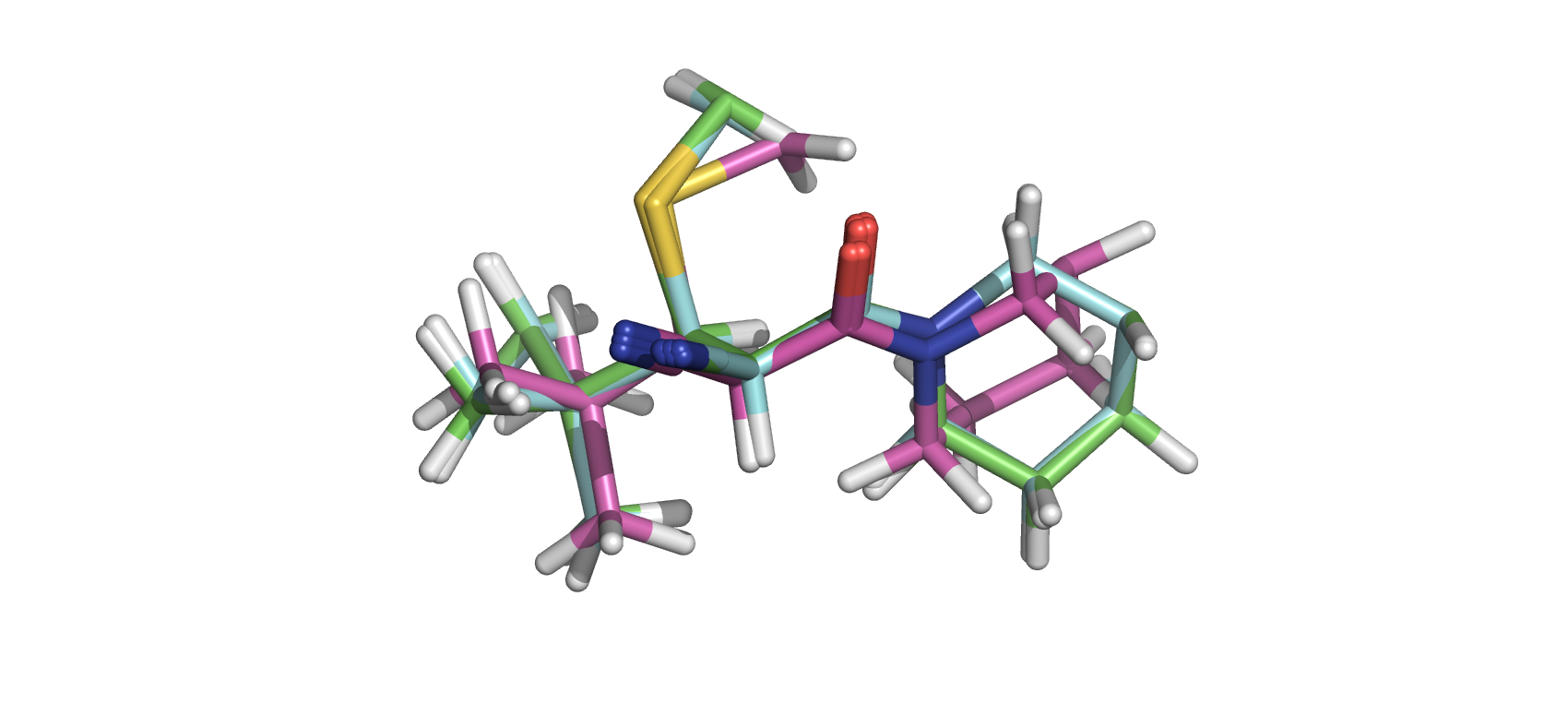
All of the most stable TS conformations located were in *syn* form, that is, the S-methyl bonds of the methanethiols were approximately in alignment with the C=C bonds of the Michael acceptors. An investigation into the NCI within the TS conformations revealed the existence of stabilising C-H··· interactions between the methanethiols and the cross-conjugated system spanning across the C=C double bond to the carbonyl and nitrile groups as shown in Figure 5. It is interesting to note that cross-conjugation is presence in all of the Michael acceptors investigated except **R5**. The C=O and C=C bonds were conjugated with two other electron clouds that don’t delocalise into each other. One would expect the centre of such cross-conjugations to be more concentrated in electron density, which could very well has led to the observation of the relatively more stable *syn* conformations.



**Figure 5.** Visualisation of the NCI of the TS. A blue-green-red scale was used to indicate the strength and type of the interactions, with blue signifying strong attractive interactions while red indicates strong nonbonded overlap.

### Products

The similarity between overlapping region of the most stable geometry of **P3** and the conformations observed in the crystal structure of the homodimer BTK protein, which consists of 40 atoms, was compared by root mean squared deviation (RMSD). The superposition of the 2 crystal structure geometries of the thiol adducts in BTK chains A and B returns an RMSD of 0.093 Å while the RMSDs from the best alignment with the ground state geometry are 1.295 Å and 1.266 Å, respectively. The QM optimised thiol adduct was found to have opposite chirality to the experimental structures, and therefore a more relevant geometry was generated using the *Mirror Molecule* function in MacroModel for the comparison under the assumption of equivalency of the values of enantiomers. Surprisingly, the QM calculations of the of the crystal structure geometries revealed that the conformations are actually astonishingly higher in relative to the ground state geometry by 110.2 kcal/mol and 110.9 kcal/mol, respectively. Prior to rationalisation of the findings, it should be kept in mind that the crystal structure conformations were solved experimentally through X-ray diffractions, which suffers from the loss of dynamic information as an average structure was summarised from the ensemble of distinct structures. The extremely high of the crystal structure conformations could then be comprehended when one considers the possibility of obtaining an unrealistic overall conformation due to the averaging of relatively different geometries. Nevertheless, the difference of 2 orders of magnitude tells that the excluded interactions between the inhibitor and its surrounding are extremely stabilising, such that penalties for distortions from its most favourable geometry worth about 100 kcal/mol could be compensated for.

****

**Figure 6.** Superposition of the ground state structure (pink) and crystal structure conformations in chains A (green) and B (blue).

## Benchmarking of Functionals and Basis Sets

The of a set of cyanoacrylamides with available experimental thermodynamics data were calculated as an attempt to decide the best combinations of functionals and basis sets for the description of thiol addition reaction. The predictions of the addition for entacapone (**15**)are collectively inaccurate, potentially due to the failure of taking the partial ionisation of the nitrocatechol group under the experimental conditions (PBS, 1-2% DMSO) into account.[43](#_ENREF_43) The uniform direction of the prediction error of the stability of the thiol adduct of **15** adds to the credibility of the hypothesis that the experimental results were obtained with negatively-charged nitrocatechol groups. **15** is thus excluded as an outlier for the statistical metrics to be recomputed.

The corrected benchmarking results pointed toward Method I as the most accurate method in terms of prediction of the addition for the cyanoacrylamides, while Method A performed the worst despite being the second most expensive method. The comparisons of relative central processing unit (CPU) time taken by each method for the same calculation are discussed in the Appendix. Methods C, G and I were able to predict the addition barrier to within chemical accuracy, namely 1 kcal/mol. It is arguable that the performance of Method Cis comparable to Methods G and I considering the magnitude of the uncertainties in the QM calculations stemming from the usage of implicit solvents. The great performance of Method C validated the prediction that *ab initio* methods converge to the exact solution as the basis set tends toward full configuration interaction.

In contrast to the reduction of magnitude of error corresponding to complete basis set extrapolation, the usage of larger basis set for DFT methods such as Methods B, and D-F had been found to result in greater error on average. The relatively better performance of Methods G and I in this case should not be attributed to accurate description of the underlying physics of the system, but merely because the smaller basis sets resulted in some error cancellations that resulted in outcomes similar to the experimental data. That said, 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 if high accuracy is deemed to be more important than time expenditure.

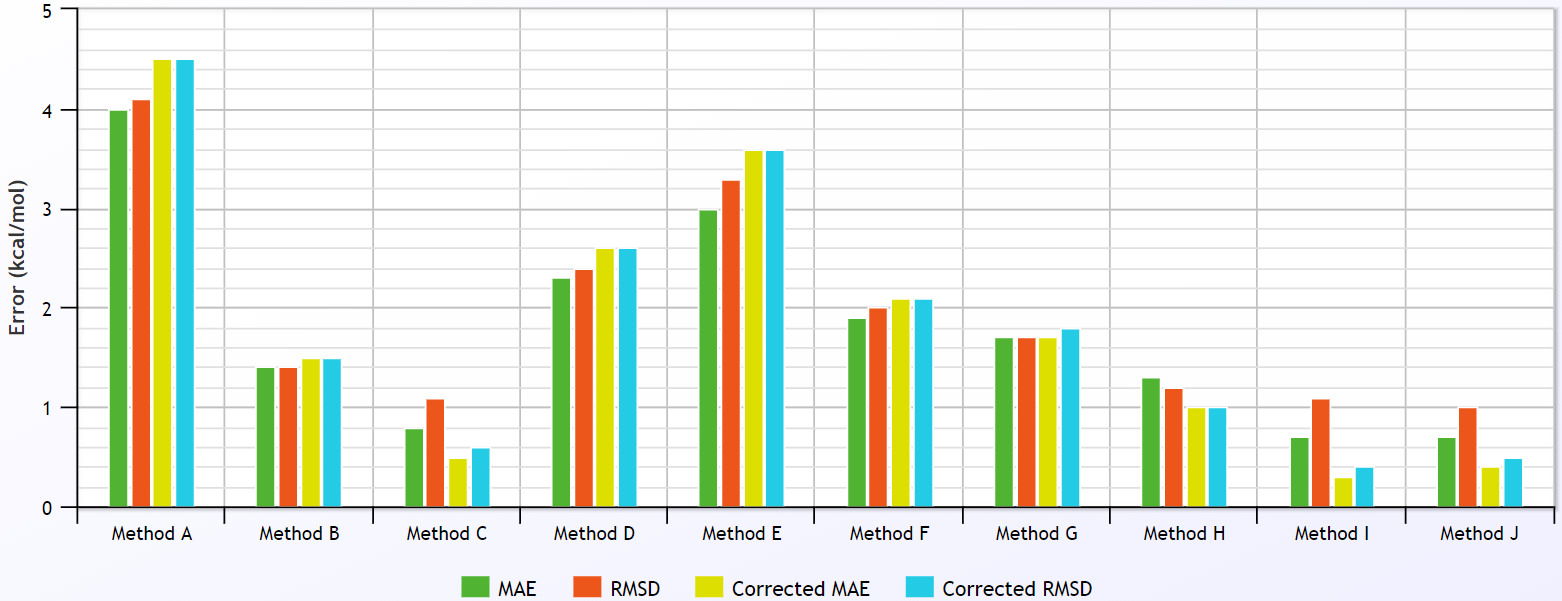
Taken together, Method I is chosen for SPE calculation of thiol additions to the cyanoacrylamides investigated by Taunton’s group. However, with regards to quantum mechanics/molecular mechanics (QM/MM) calculations, Method J is preferred for practicality purposes given that it only requires about 2/3 of the number of computations required by Method I with minimal impact on the accuracy.

**Table 3. Combinations of functionals and basis sets chosen for benchmarking.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Michael acceptor | Calc.(kcal/mol) for Methods | | | | | | | | | |
| A | B | C | D | E | F | G | H | I | J | |
| **10**  = -2.8*a* | 1.7 | -1.6 | -3.3 | -0.1 | 0.6 | -0.5 | -1.0 | -1.9 | -2.9 | -2.5 | |
| **11**  = -3.9*a* | 0.6 | -2.4 | -4.5 | -1.4 | -0.5 | -2.0 | -2.4 | -3.2 | -3.8 | -3.5 | |
| **12**  = -3.0*a* | 1.6 | -1.3 | -3.4 | -0.2 | 0.8 | -0.8 | -1.1 | -1.8 | -2.8 | -2.4 | |
| **13**  = -2.6*a* | 1.6 | -1.6 | -3.5 | -0.4 | 0.5 | -1.0 | -1.3 | -1.9 | -3.0 | -2.7 | |
| **14**  = -2.0*a* | 2.6 | -0.2 | -1.7 | 0.9 | 2.0 | 0.5 | 0.2 | -0.6 | -1.5 | -1.2 | |
| **15**  = -3.2*a* | -1.7 | -4.4 | -5.7 | -3.8 | -2.8 | -4.4 | -4.8 | -5.3 | -5.8 | -5.4 | |
| MAD | 4.0 | 1.4 | 0.8 | 2.3 | 3.0 | 1.9 | 1.7 | 1.3 | 0.7 | 0.7 | |
| RMSD | 4.1 | 1.4 | 1.1 | 2.4 | 3.3 | 2.0 | 1.7 | 1.2 | 1.1 | 1.0 | |
| Corrected*b* MAD | 4.5 | 1.5 | 0.5 | 2.6 | 3.6 | 2.1 | 1.7 | 1.0 | 0.3 | 0.4 | |
| Corrected*b* RMSD | 4.5 | 1.5 | 0.6 | 2.6 | 3.6 | 2.1 | 1.8 | 1.0 | 0.4 | 0.5 | |

*a*Experimental values in kcal/mol. *b*Compound **15** was excluded for the computation of corrected metrics due to discrepancies in the modelled environment compared to the experimental conditions.

**Chart 2. Statistical Metrics from Benchmarking Results**.



## Calculations of Thermodynamic Quantities

The values were obtained from values of the ground state species calculated using Method I. The predicted values were compared with the experimental values as tabulated in Table 4. Due to the lack of knowledge about the base catalyst, the computed values are based on a model base. The absolute values could therefore not be compared with the experimental results. However, the predicted trend matched surprisingly well with the ranking of values calculated based on the experimentally measured RT even prior to the inclusion of the environmental effects, with the only dissimilarity being the interchange between the ranking of **R1** and **R47**.

**Table 4. Comparison between experimental results and calculated thermodynamic and kinetic parameters.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Michael acceptor | Expt. rank | Calc. *a* (kcal/mol) | Pred. rank*b* | Calc.(kcal/mol) | Calc.(kcal/mol) |
| **R1** = 22.8*c* | 5 | 11.8 | 4 | -4.0 | 7.8 |
| **R3**  > 26.5*c* | 2*d* | 14.8 | 2 | -2.0 | 12.8 |
| **R47**  = 25.1*c* | 4 | 11.2 | 5 | -2.1 | 9.0 |
| **R5**  >> 26.5*c* | 1*d* | 25.8 | 1 | -9.0 | 16.8 |
| **R9**  = 26.4*c* | 3 | 14.6 | 3 | -3.6 | 11.0 |

*a*The values were calculated assuming methylthiolate being the participating base for thiol elimination. *b*Ranked in descending order in terms of magnitude of . *c*Experimental values in kcal/mol. *d*Despite the same magnitude of rounded , no recovery of **5** from its covalent adductwas observed experimentally over 200 hours, in contrast to the gradual recovery of **3**.

That said, a closer examination on the relative magnitude of the predicted values revealed that the intrinsic reactivities of the warheads do not account for the overall reactivities of the Michael acceptors as much as one would expect from the seemingly excellent agreement in the trend of elimination rate. The trend shown in Chart 3 suggests that the factors uncaptured by the intrinsic reactivities exert a combination of stabilising or/destabilising effects on each of the inhibitors to different extents. These effects could only be decomposed using cheaper computational techniques that allow dynamic studies of the inhibitors with the inclusion of the binding site residues that collectively contribute to the environmental effects.

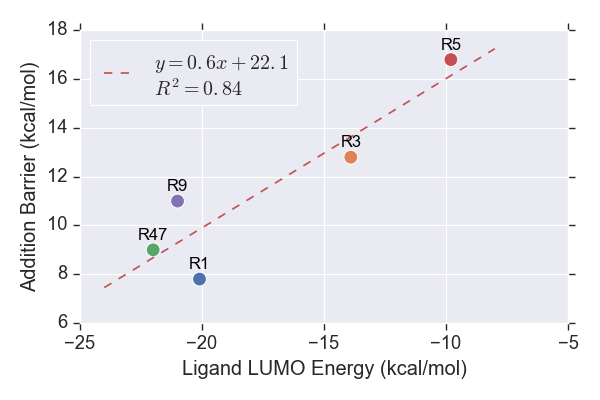
**Chart 3. Trends of Predicted and Experimental Elimination Barriers**

## Rationalisation of the Predicted Intrinsic Reactivities

The correlation between the predicted addition barriers and several properties of the molecular systems were investigated in attempt to decompose the components contributing to the calculated values. A statistical measure, value was computed from linear regression analysis of each of these properties. It should be stressed that the value simply indicates the extent to which the independent variable, which is the calculated in this case, is explained by the variation of a dependent variable in a regression model. The value does not allow conclusions regarding the quality and reliability of the underlying model, biasness of the data, nor choice of regression types to be made.

### Reactant Lowest Unoccupied Molecular Orbital Energies

The predicted addition barriers are first compared to the lowest unoccupied molecular orbital (LUMO) energies of the reactants. As seen in Figure 7, despite the rather high correlation shown by the inhibitors **R3**, **R47**,and **R5**, the other 2 Michael acceptors with more the simplest and the most elaborated warheads do not seem to agree with the expected linear relationship. This might indicate the presence of other factors in the determination of the intrinsic reactivities of the compounds.



**Figure 7.** LUMO energies of the reactants.

### -Carbon Charges

There has been a focus on discovering well-defined chemical concepts in order to allow qualitative understanding and quantitative prediction of chemical reactivity. The molecular responses during a reaction are thought to be dependent on the changes in the number of electrons and the external potential felt by the electrons. These concepts are summarised as conceptual DFT (CDFT). The investigation in this project focuses on some global descriptors, which measure the overall susceptibility of a molecular system to either electrophilic or nucleophilic attacks.

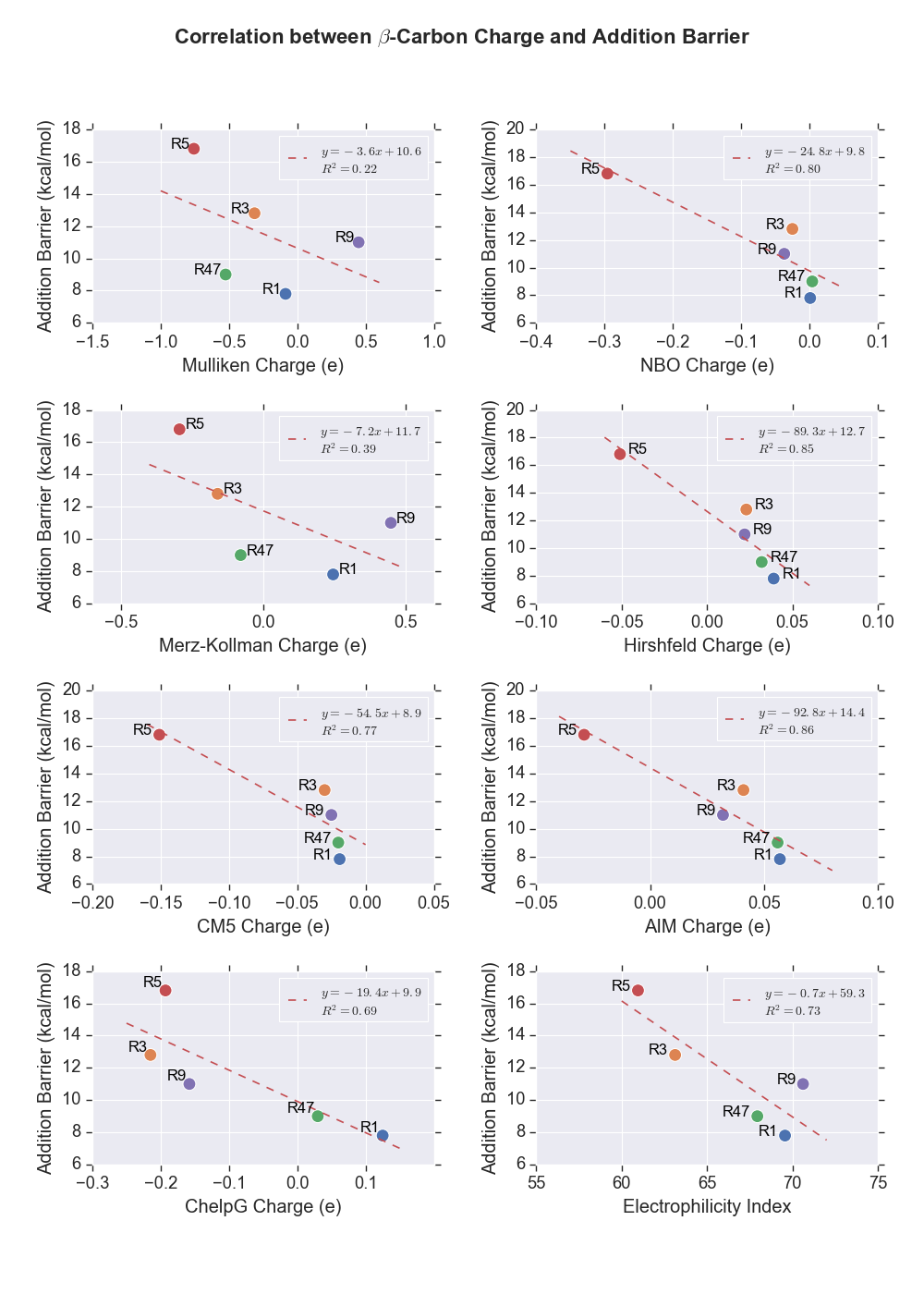
It is intuitive to expect an explicit relationship between the electron density distribution near the reacting site and the addition barrier. The speculation is in agreement with the concept of a global descriptor, namely electrophilicity index, , devised by Robert Parr[106](#_ENREF_106) as shown in equation (4):

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where and are electronegativity and chemical hardness, respectively. is defined as the extent of lowering of energy at the maximal uptake of electrons. For comparisons across a series of similar compounds, a relatively smaller electron density around a particular atom corresponds to greater hardness as defined by the Pearson or hard-soft acid-base (HSAB) concept.[107](#_ENREF_107) The electrophilicity of an atom is thus anticipated to be inversely proportional to its electron density.

Partial charge is an intuitive approach to approximate the electron density distribution. An attempt was thus made to find the correlation between the addition barrier and the TS -carbon partial charges along with the . The charge is inversely proportional to electron density and thus proportional to the electrophilicity of the -carbon. The energies for the systems were modelled using the finite differences (FD) approach. The values are computed based on Koopman’s theorem,[108](#_ENREF_108) where electron affinity (EA) and ionisation potential (IP) are simply approximated by the negatives of the LUMO and HOMO energies, respectively. However, care should be taken when the orbitals were chosen as the LUMO might not exhibit significant overlapping with the first HOMO in some cases, and hence is unlikely to participate in the electron donation to LUMO in the reaction. The orbitals were thus visualised in search for the most relevant HOMO. Further details regarding the process could be found in the Appendix. The relationship between the computed values and the addition barriers is plotted in Figure 8 along with the -carbon partial charge calculated from 7 distinct charge models.

All of the charge schemes showed the inversely proportional relationship with the addition barrier as expected. In agreement with the notoriety of Mulliken population analysis regarding its unreliability as it allows unphysical negative electronic populations and blatantly violates Pauli Exclusion Principle,[109](#_ENREF_109) Mulliken charges exhibited the lowest value (0.22) followed by Merz-Kollman charges (0.39). All other charge schemes showed relatively good correlation (>0.69) with the charges computed from the quantum theory of atoms in molecules (QTAIM) returning the highest value (0.86). The addition barriers are described rather well by the values ( = 0.73). An interesting observation is that an exponential-like relationship was observed for natural bond orbital (NBO), Hirshfeld and Charge Model 5 (CM5) charge schemes.



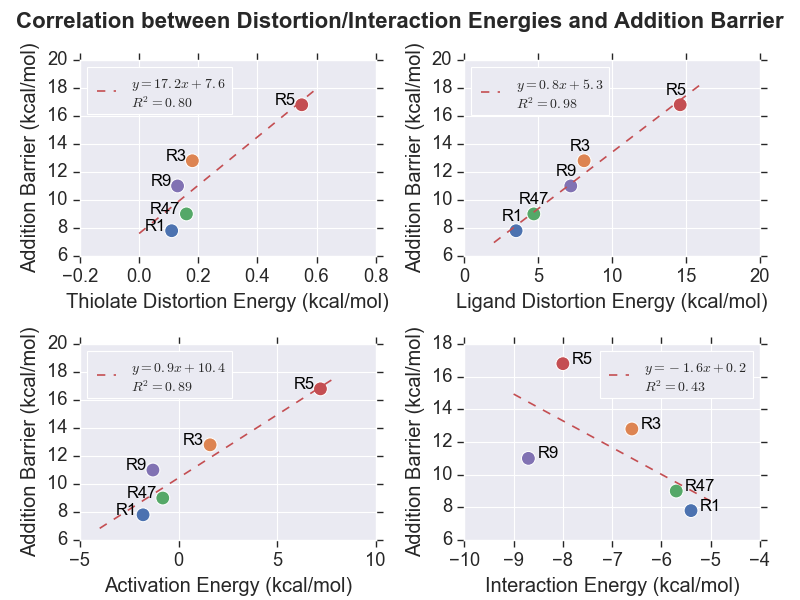
**Figure 8.** Partial charges of -carbon and of the Michael acceptors.

### Distortion/Interaction Analysis

The distortion/interaction model[110](#_ENREF_110) was used to analyse the reactivity trend among the investigated cyanoacrylamide Michael acceptors. The activation energies for a reaction, , could be understood as the sum of energies required to distort the reactants into TS geometries, and the interaction energies between the molecules in TS structures, , as shown in equation (5):

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

The correlation plots of each parameter with the addition barrier were shown in Figure 9.

****

**Figure 9.** Plots showing the relationship between distortion energies of thiolate (upper left), and cyanoacrylamides (upper right), activation energies (lower left) and interaction energies (lower right).

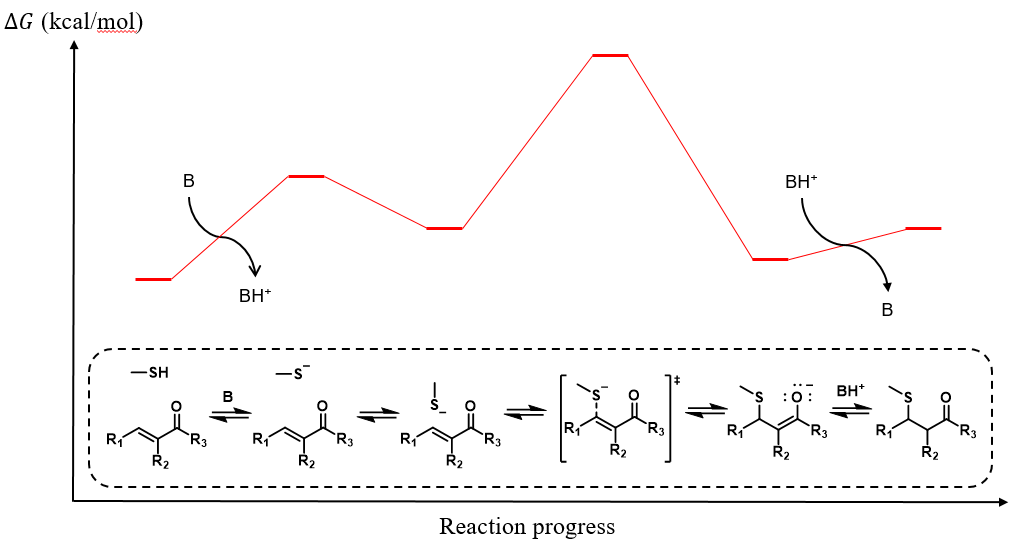
The amount of the strain that the methylthiolate anion had to overcome to achieve TS, is largely similar except for **R5**, which seems to have undergone a greater distortion overall. The hypothesis was tested by comparing the similarities of the geometries of each methylthiolate in TS relative to its ground state structure. The results which are tabulated in Table 5 provides strong evidence to support the speculation. A more detailed investigation on the distortion of methylthiolate anion in **TS5** was discussed in the Appendix.

**Table 5.** RMSD of methylthiolates in TS for each inhibitor from the ground state structure.

|  |  |
| --- | --- |
| Transition State | RMSD (Å) |
| **TS1** | 0.0026 |
| **TS3** | 0.0042 |
| **TS47** | 0.0023 |
| **TS5** | 0.0082 |
| **TS9** | 0.0026 |

Despite the positive correlation observed ( = 0.80), the relative miniscule magnitudes of the most probably did not influence the addition barriers in any significant way. In contrast, the distortion energies of the Michael acceptors, , which dominated the unfavourable components showed noticeably high correlation as evidenced by the value of 0.98 with the values.

The values also exhibited a rather strong correlation with the predicted addition barriers ( = 0.89). Following some intrinsic reaction coordinate (IRC) studies, the negative values are found to be due to the formation of charge complexes prior to the TS structures. As illustrated in the energy profile in Figure 10, the negligence of the entropic effect, which could be up to 12 kcal/mol for organic small molecules, has resulted in the relatively lower potential energy of the TS relative to the infinitely separated and diluted reactants. That said, the strong positive correlation observed indicates that the effects of the work done and the change in entropy, , on the addition barriers of the inhibitors investigated are uniform enough such that the trend of , which omits these terms, is approximately the same as the trend of .



**Figure 10.** Illustrative potential energy profile.

A rather low value of 0.43 was returned from the linear regression analysis on the interaction energies. A closer look at the data suggests that **R9** is highly likely an outlier as a better linear fit could be obtained with the data point excluded. For further analysis, the interaction energy could be further dissected into individual components, including electrostatic, polarisation, exchange repulsion, and charge transfer interactions.[111](#_ENREF_111) The independent investigations on each aspect of these noncovalent interactions might allow pinpointing of the component that contributed most significantly to thiol-Michael additions.

# INCORPORATION OF THE ENVIRONMENTAL EFFECTS

## Methods

All production MD simulations were conducted using AMBER18 *pmemd.cuda* program[112-114](#_ENREF_112) installed on a Dell EMC-manufactured high-performance computer named Wiener, which is hosted at the University of Queensland. The preparation of the systems and energy minimisations were carried out using a series of programs of GROMOS++ version 1.4.0[115](#_ENREF_115) Twenty-four systems in total (covalently bound BTK with and without thiol adduct deprotonation, noncovalently bound BTK with and without cysteine thiol deprotonation) were considered for the study.

### Structure Preparation

All protein topology parameters were obtained from GROMOS11 54A7 force field[116](#_ENREF_116) in the form of interaction function parameter (ifp) and molecular topology building block (mtb) files from the Automated Topology Builder (ATB) version 3.0 web server.[117](#_ENREF_117),[118](#_ENREF_118) ATB was also used to generate the topology parameters of ligands **1**, **3**, **4**, **5**, **7**, and **9** in the form of unreacted compound and covalently bonded form with a cysteine residue. The geometry was optimised at the B3LYP/6-31G\* level of theory in water using polarizable continuum model (PCM). The electrostatic potential (ESP) was then calculated using the optimised geometry, from which the charges were obtained from least-squares fitting. The details of the parameter assignment protocol was described in their first paper.[117](#_ENREF_117)

The simulated systems were built based on the X-ray crystal structure of a BTK dimer inhibited by 2 inhibitors **3** (4YHF)made publicly available on the Protein Data Bank (PDB) by Taunton’s group. The missing residues at the ends of both residue chains A and B were built in as a linear chain using PyMOL.[119](#_ENREF_119) The N terminals were left as NH3+ while C terminals were left as COO-. Swiss PDB Viewer was employed to add the missing atoms through reconstruction of the side chains.[120](#_ENREF_120) Molecules other than BTK and **3** such as SO42- anions and ethylene glycol were present due to the buffers added during protein crystallisation. They were thus purged from the PDB file. Modification of the covalently bonded inhibitors **3** to the reactant state was done by removing the extra proton acquired from thiol addition and adjusting the orbital hybridisation. was used to generate the topology files. The topology files of the residue chain A, chain B, and **3** were generated separately using the *make\_top* program and then combined using the *com\_top* program. The protonation state of the each residue was decided based on the predicted values made by PROPKA version 3.1.[121](#_ENREF_121),[122](#_ENREF_122) A residue is protonated if its value is equal to or lower than the physiological pH value, which is 7.4. The residue and atom names of **3** in the PDB file were then updated to match the mtb file obtained from the ATB. The *pdb2g96* program was used to generate a coordinate file of the system in GROMOS96 format followed by the positional adjustment of the hydrogen atoms by the *gch* program.

The protein and ligands were first energy minimised in vacuum over 5000 steps using steepest descend 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 charges of the periodic systems are negative overall. The neutralisation was achieved by replacing the water molecule with the highest potentials with sodium ions. The solvated system was then energy minimised again over 5000 steps. The strained bonds, angles, improper dihedrals, and dihedral angles were identified from the *check\_top­* program. A positional restrained energy minimisation was then carried out by first applying a harmonic potential to the coordinates of the protein structure in place, with the unfavourable degrees of freedom excluded. The details of the unmentioned parameters were included in the Appendix.

The GROMOS system topologies (.top) and coordinates (.cnf) were converted into AMBER-compatible format (.prmtop and .mdcrd, respectively) utilising the topology converter on the ATB web server. The residue name for water solvent was specified to be H2O to enable the application of the SETTLE constraint algorithm.[123](#_ENREF_123)

### Simulation Setup

The systems were first equilibrated as a 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 systems were then heated up to 310.15 K and maintained constant with a Berendsen thermostat.[124](#_ENREF_124) A 2 ns isobaric-isothermal (NPT) equilibration with a pressure of 1 atm maintained using Berendsen barostat[124](#_ENREF_124) was then carried out to adjust the box density followed by another 12 ns NVT equilibration.

All-atom NVT simulations was carried out for 100 ns for each system. Berendsen algorithm was applied to maintain simulation temperature of 310.15 K and pressure of 1 atm with coupling time constants of 0.1 and 0.5 respectively. The isothermal compressibility of the system was set to 45.75×10-6 bar-1. The time constants were chosen such that the resonance between the temperature-pressure couplings could be avoided. Periodic boundary conditions (PBC) were imposed to avoid edge effects. Long-range Coulomb interactions were handled using the particle mesh Ewald (PME) method.[125](#_ENREF_125) The SHAKE algorithm[126](#_ENREF_126) with tolerance of 10-5 was utilized to constrain the bond involving hydrogen atoms. Despite this, a time step of 1 fs was used due to the relatively high temperature of simulation. Both the Coulomb and the vdW interactions were truncated at 1.4 nm. A dielectric constant of 1.0 was used. The coordinate trajectories were recorded every 1 ps. The SETTLE algorithm was employed to constrain the geometry of water molecules for higher throughput.

### Trajectory Analysis

The results were analysed using the CPPTRAJ program.[127](#_ENREF_127) The systems were visualised using visual molecular dynamics (VMD)[128](#_ENREF_128) analysis toolkit. Figures are plotted using Python scripts written as appended.

## Identification of Potential Base Species

### Distance from Charged Residues

Distance analyses are carried out in attempt to search for potential base species. There is a convention in the field that assumes that the basic residues are unable to act as basic catalysts as they are often protonated at physiological pH.[129](#_ENREF_129) However, there has been growing discoveries on the role of basic residues as biological basic catalysts lately.[130](#_ENREF_130),[131](#_ENREF_131) Taking these into considerations, all charged residues, including histidines, lysines, arginines, glutamates, and aspartates are nominated as likely candidates of the participating base in the thiol addition/elimination. Considering that it took more than 200 hours for **3** to dissociate from BTK experimentally, it is expected that a relatively significant change in the conformation is needed prior to the extraction of proton and further advancement of the reverse reaction. Figure 9 showed that the 100 ns simulations managed to capture the moments when the guanidine group of a particular arginine and the acetate group of a particular glutamate approaches the critical C H. The distances were maintained at about 4-6 Å throughout the simulations.

**Figure 12.** Distance between basic (histidine, lysine and arginine) and acidic (aspartate and glutamate) residues of interest from the C proton of the covalently bound inhibitors and the thiol proton of the unreacted Cys481.

**Figure 12.** Distance between basic (histidine, lysine and arginine) and acidic (aspartate and glutamate) residues of interest from the C proton of the covalently bound inhibitors and the thiol proton of the unreacted Cys481.

### Intramolecular Proton Transfer

An investigation into the possibility of elimination through either 6-membered or 4-membered intramolecular proton transfer was conducted. The search for 6-membered intramolecular proton transfer TS is to no avail. However, the failure by no means indicates the impossibility for the reaction paths to be taken.

The comparison of the energy barriers with the conventional base catalysed addition/elimination mechanism revealed that they are much less likely to occur.

## Construction of Energy Profile

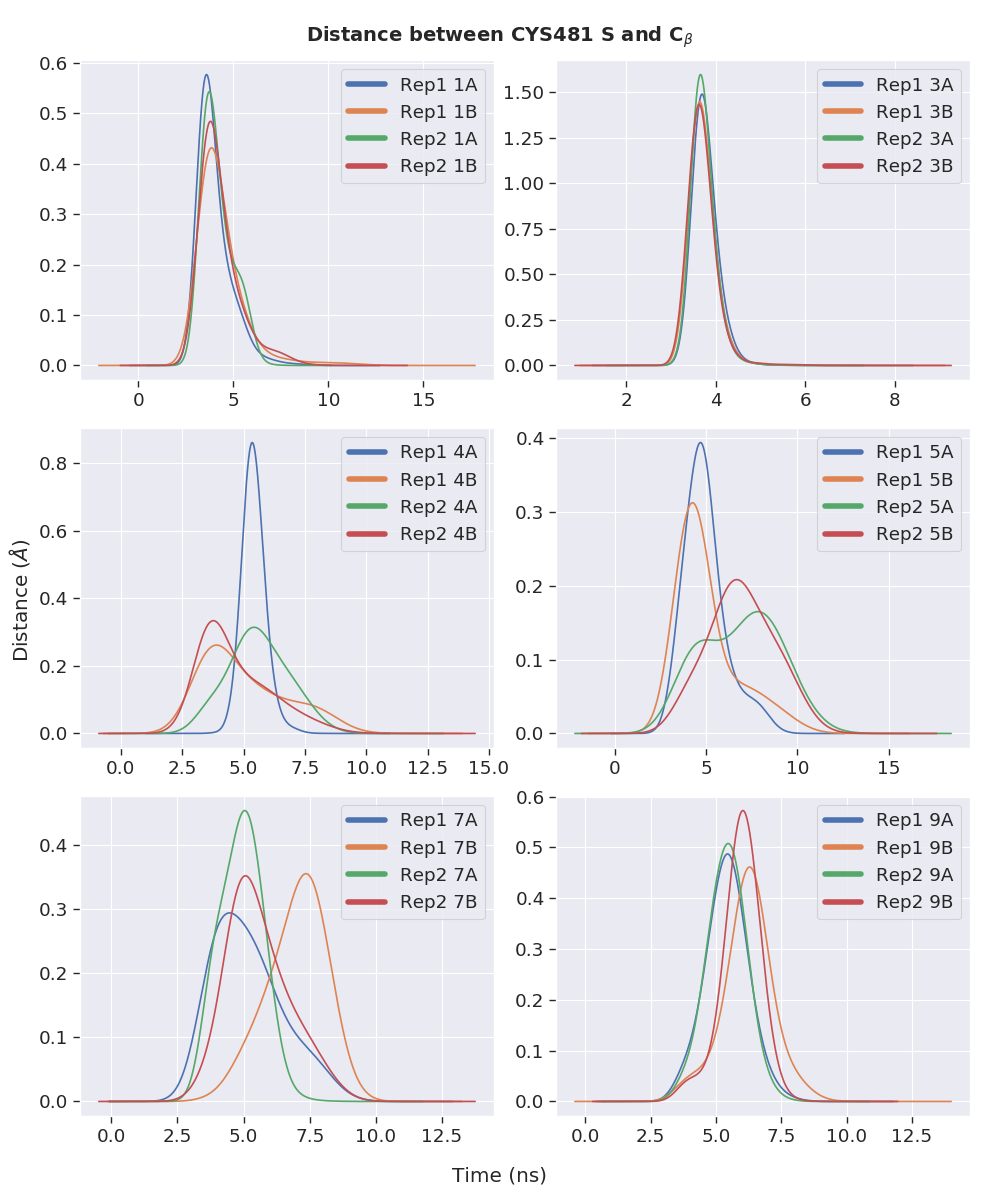
Putting together the information, an energy profile of thiol-Michael addition to the **R3** is constructed as shown in Figure.

**Figure 12.** Distance between basic (histidine, lysine and arginine) and acidic (aspartate and glutamate) residues of interest from the C proton of the covalently bound inhibitors and the thiol proton of the unreacted Cys481.

## Interactions between Cyanoacrylamides Inhibitors and BTK Active Site Residues

### Distance of Cys481 Sulfur Atom from Electrophilic Carbon on Ligands

The distances between the atoms participating in the Michael addition throughout the simulations are presented as distribution plots. Learning from the QM calculation of the TS of the Michael addition, it is expected that the reaction happens when the S-C distance is about.

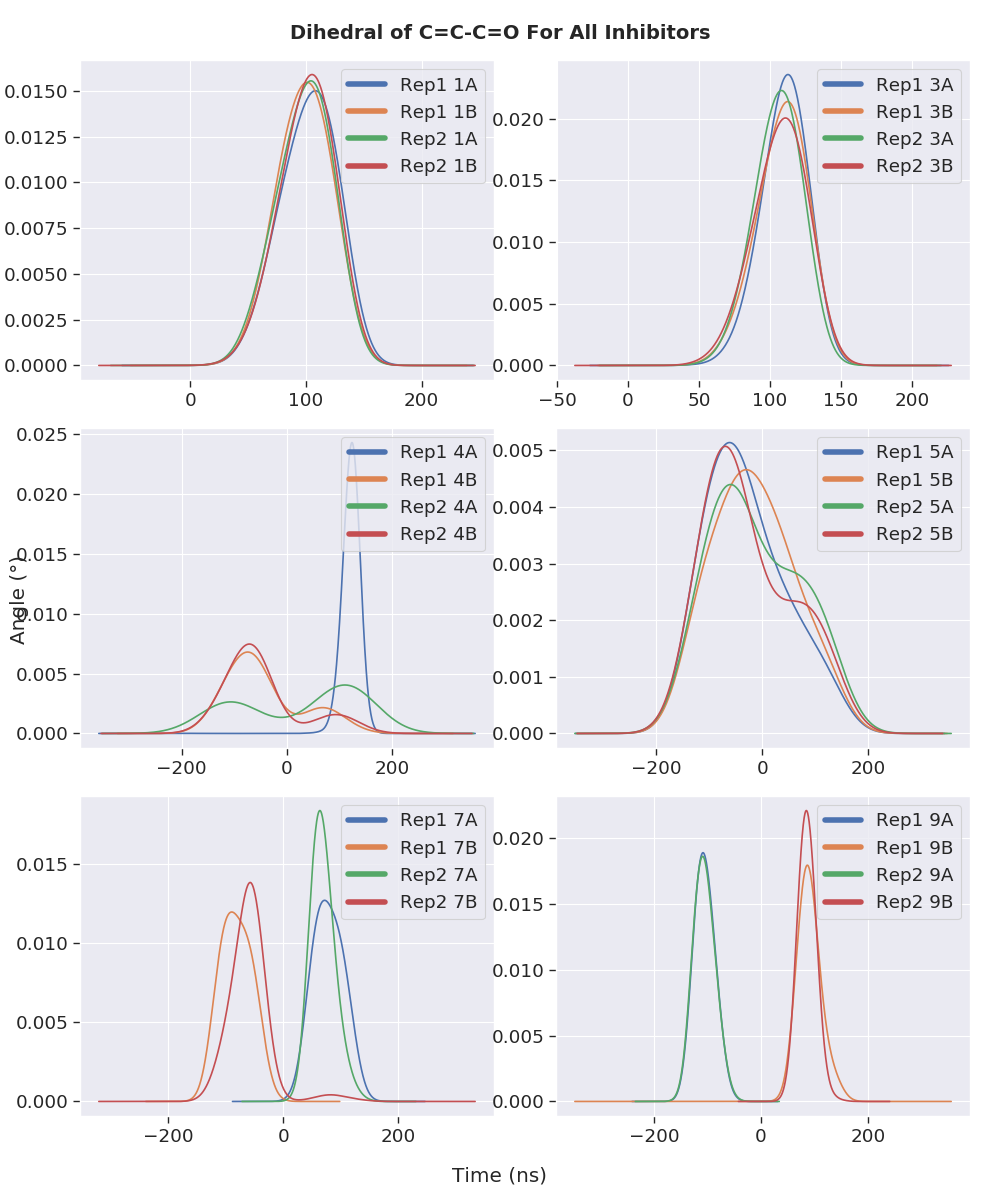


**Figure 12.** Distance between the -C of the inhibitors and the sulfur atom of the Cys481 thiolate anions.

### Dihedral Rotations about C=C-C=O Bonds

The distributions of the rotations of the bond connecting carbonyl C and the -C for all noncovalently bound inhibitors are plotted as shown in Figure 13. It is seen that for **1** and **3**, the 2 planes are maintained at about 100°. It is interesting to see that the mode of the distributions of chain A ligands for **4**, **7**, and **9** have opposite signs.

For **5**, while the angle favoured negative values, the distributions are skewed to the right.



**Figure 13.** Distributions of the C=C-C=O dihedral angles throughout the 100 ns MD simulations of noncovalently bound BTK inhibitors.

### Hydrogen Bond Analysis

**Table 5.** RMSD of methylthiolates in TS for each inhibitor from the ground state structure.

|  |  |
| --- | --- |
| Transition State | RMSD (Å) |
| **TS1** | 0.0026 |
| **TS3** | 0.0042 |
| **TS47** | 0.0023 |
| **TS5** | 0.0082 |
| **TS9** | 0.0026 |

### Cluster Analysis

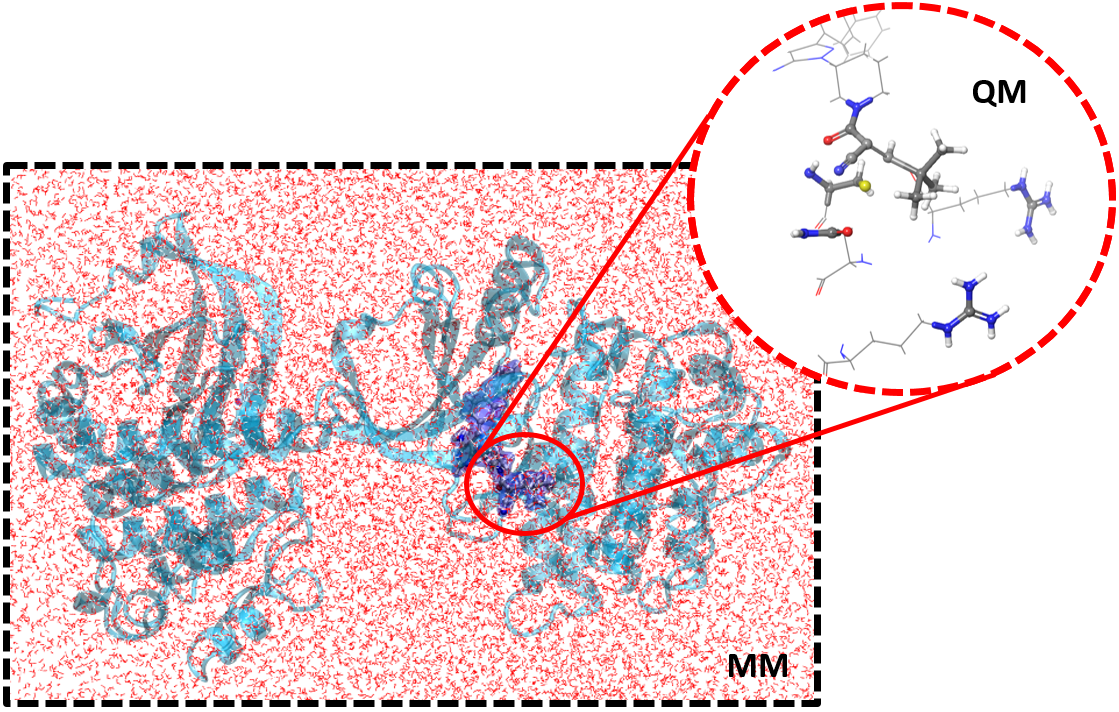
**Figure 15.** Conformations of different structural clusters located throughout the simulations.

**Figure 15.** Conformations of different structural clusters located throughout the simulations.

# FUTURE DIRECTION – DETAILED INVESTIGATION OF THE EFFECT OF BINDING SITE RESIDUES ON THE REACTIVITY.

## Concept of QM/MM

Despite their usefulness in describing chemical reactions, pure QM calculations are restricted to systems containing up to a few hundred atoms. The modelling of biomolecular systems where reactions occur within a protein environment often requires inclusion of thousands of atoms. This necessitates different computational techniques such as hybrid QM/MM methods for practicality purposes. As illustrated in Figure 10, 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 MM potential functions, commonly known as force fields, which allow reduced computation through simplification of the underlying physics.



**Figure 10.** Illustration of the QM/MM concept.

Such an approach allows the electrostatic and steric effects of the protein on the reaction to be accounted for while ensuring accurate representation of the electronic structure in the vicinity of the reacting atoms. The multiple strengths of these hybrid methods are that the simulation cost and accuracy with the entire system included explicitly in the calculations are balanced and that they complement experimental data, allowing different mechanistic proposals to be supported or contradicted.

Ever since the pioneering study conducted by Warshel and Levitt on the mechanism of the lysozyme reaction, QM/MM methods have received extensive use in recent years as a versatile tool for the study of enzymatic reaction mechanisms,[132-138](#_ENREF_132) calculation of spectroscopic properties,[139](#_ENREF_139),[140](#_ENREF_140) prediction of p*K*a values[141](#_ENREF_141) and investigation of electronically excited states.[142](#_ENREF_142) The value of QM/MM methods in the field of computational enzymology can be appreciated from the multiple mechanistic studies conducted on a diverse range of kinases, including the more recent reports on isopentenyl phosphate kinase,[135](#_ENREF_135) mevalonate kinase,[134](#_ENREF_134) and *N*-acetyl-L-glutamate kinase,[133](#_ENREF_133) just to name a few. As such, it is anticipated that a similar study on BTK would provide insightful details about the inhibition mechanism of BTK of use to future research.

## Location of TS Structures

A QM/MM system could be built for each inhibitor based on a snapshot of the MD simulations of the covalently and noncovalently bound BTK for more detailed studies on the thiol eliminations and additions, respectively. The frame should be selected such that the positions of the reactive moieties are close to the TS geometries. This would allow for calculations of reaction profiles for the Michael reactions. Some potential investigations that could be conducted include the verification of the participating base, the illumination of the roles of specific residues near the binding site, and the location of TS structures. Once the TS structures of the Cys481 thiol additions/eliminations to the inhibitors areidentified, the comparisons with the geometries throughout the MD simulations would indicate the closeness of the geometries to the TS. The variations in reaction for different warheads could also be examined through structural modification of the warheads. The calculated kinetic parameters could be compared to the experimentally measured RT of the inhibitors to extent of inclusion of the environmental factors.

## Selectivity of Inhibitor for BTK

The experimentally observed selectivity of **9** for BTK over other kinases could potentially be studied following the examination of the binding site effects. Employing the same techniques, QM/MM studies of the binding of **9** to other kinases such as EGFR, ITK and TXK could be constructed readily from the widely available X-ray crystal structures of the unbound proteins. The bound forms could be prepared by selecting a snapshot from the MD simulations of the kinases with a selected inhibitor in proximity for QM/MM analysis. The structural features responsible for the apparent kinase selectivity of the inhibitorcould then be uncovered from a comparison among the kinases. The approach developed in the study of the selectivity of **9** would serve as the foundation for future studies of the selectivities of other inhibitors, providing a broader evaluation of the factors that influence selectivity.

# CONCLUSION

.

# 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. 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.

5. 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.

6. Tóth, L.; Muszbek, L.; Komáromi, I. *J. Mol. Graphics Modell.* **2013,** *40*, 99-109.

7. Wright, P. M.; Seiple, I. B.; Myers, A. G. *Angew. Chem. Int. Ed.* **2014,** *53*, 8840-8869.

8. Kahan, F. M.; Kahan, J. S.; Cassidy, P. J.; Kropp, H. *Ann. N.Y. Acad. Sci.* **1974,** *235*, 364-386.

9. Uetrecht, J. *Chem. Res. Toxicol.* **2008,** *21*, 84-92.

10. Uetrecht, J. *Chem. Res. Toxicol.* **2009,** *22*, 24-34.

11. Chen, M.; Suzuki, A.; Borlak, J.; Andrade, R. J.; Lucena, M. I. *J. Hepatol.* **2015,** *63*, 503-514.

12. Yuan, L.; Kaplowitz, N. *Clin. Liver Dis.* **2013,** *17*, 507-518.

13. Mintzer, D. M.; Billet, S. N.; Chmielewski, L. *Adv. Hematol.* **2009,** *2009*, 1-11.

14. Barf, T.; Kaptein, A. *J. Med. Chem.* **2012,** *55*, 6243-6262.

15. Mah, R.; Thomas, J. R.; Shafer, C. M. *Bioorg. Med. Chem. Lett.* **2014,** *24*, 33-39.

16. Kalgutkar, A. S.; Dalvie, D. K. *Expert Opin. Drug Discov.* **2012,** *7*, 561-581.

17. Copeland, R. A.; Pompliano, D. L.; Meek, T. D. *Nat. Rev. Drug Discov.* **2006,** *5*, 730-739.

18. Tummino, P. J.; Copeland, R. A. *Biochemistry* **2008,** *47*, 5481-5492.

19. Zhang, R.; Monsma, F. *Expert Opin. Drug Discov.* **2010,** *5*, 1023-1029.

20. Guo, D.; Hillger, J. M.; Ijzerman, A. P.; Heitman, L. H. *Med. Res. Rev.* **2014,** *34*, 856-892.

21. Copeland, R. A. *Expert Opin. Drug Discov.* **2010,** *5*, 305-310.

22. Lu, H.; Tonge, P. J. *Curr. Opin. Chem. Biol.* **2010,** *14*, 467-474.

23. 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.

24. Swinney, D. C. *Nat. Rev. Drug Discov.* **2004,** *3*, 801-808.

25. Kahan, T.; Eliasson, K. *Am. J. Hypertens.* **1999,** *12*, 1188-1194.

26. Kohout, T. A.; Xie, Q.; Reijmers, S.; Finn, K. J.; Guo, Z.; Zhu, Y.-F.; Struthers, R. S. *Mol. Pharmacol.* **2007,** *72*, 238-247.

27. Bauer, R. A. *Drug Discov. Today* **2015,** *20*, 1061-1073.

28. 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.

29. 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.

30. 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.

31. Hallek, M.; Shanafelt, T. D.; Eichhorst, B. *Lancet* **2018,** *391*, 1524-1537.

32. 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.

33. 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.

34. 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.

35. Brown, J. R. *Curr. Hematol. Malig. Rep.* **2013,** *8*, 1-6.

36. Peterson, D.; Schwartz, J. *J. Adv. Pract. Oncol.* **2014,** *5*, 348-354.

37. 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.

38. 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.

39. 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.

40. Wu, J.; Zhang, M.; Liu, D. *J. Hematol. Oncol.* **2016,** *9*, 1-4.

41. 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.

42. Chandrashekara, S. *Indian J. Pharmacol.* **2012,** *44*, 665-671.

43. Krenske, E. H.; Petter, R. C.; Houk, K. N. *J. Org. Chem.* **2016,** *81*, 11726-11733.

44. Copeland, R. A. *Nat. Rev. Drug Discov.* **2015,** *15*, 87-95.

45. Guo, D.; Heitman, L. H.; Ijzerman, A. P. *ChemMedChem* **2015,** *10*, 1793-1796.

46. Tonge, P. J. *ACS Chem. Neurosci.* **2017,** *9*, 29-39.

47. Ferguson, F. M.; Gray, N. S. *Nat. Rev. Drug Discov.* **2018,** *17*, 353-377.

48. De Cesco, S.; Kurian, J.; Dufresne, C.; Mittermaier, A. K.; Moitessier, N. *Eur. J. Med. Chem.* **2017,** *138*, 96-114.

49. 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.

50. Zhan, P.; Pannecouque, C.; De Clercq, E.; Liu, X. *J. Med. Chem.* **2016,** *59*, 2849-2878.

51. 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.

52. Vauquelin, G. *Br. J. Pharmacol.* **2016,** *173*, 2319-2334.

53. 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.

54. Fisher, J. F.; Mobashery, S. *MedChemComm* **2016,** *7*, 37-49.

55. Willemsen-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.

56. Gupta, V.; Carroll, K. S. *Chem. Sci.* **2016,** *7*, 400-415.

57. 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.

58. 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.

59. Chaikuad, A.; Koch, P.; Laufer, S. A.; Knapp, S. *Angew. Chem. Int. Ed.* **2018,** *57*, 4372-4385.

60. Gupta, V.; Carroll, K. S. *Chem. Commun.* **2016,** *52*, 3414-3417.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. *J. Am. Chem. Soc.* **2014,** *136*, 12624-12630.

67. Chruszcz, M.; Wlodawer, A.; Minor, W. *Biophys. J.* **2008,** *95*, 1-9.

68. Kruschel, D.; Zagrovic, B. *Mol. Biosyst.* **2009,** *5*, 1606-1616.

69. Henzler-Wildman, K.; Kern, D. *Nature* **2007,** *450*, 964.

70. Kuzmanic, A.; Pannu, N. S.; Zagrovic, B. *Nature Communications* **2014,** *5*, 1-10.

71. Sakthivel, S.; Habeeb, S. K. M. *J. Biomol. Struct. Dyn.* **2018,** *36*, 4320-4337.

72. Ratzon, E.; Bloch, I.; Nicola, M.; Cohen, E.; Ruimi, N.; Dotan, N.; Landau, M.; Gal, M. *ACS Omega* **2017,** *2*, 4398-4410.

73. Balasubramanian, P. K.; Balupuri, A.; Kang, H.-Y.; Cho, S. J. *BMC Syst. Biol.* **2017,** *11*, 1-11.

74. Balasubramanian, P. K.; Balupuri, A.; Cho, S. J. *Arch. Pharmacal Res.* **2016,** *39*, 328-339.

75. Bavi, R.; Kumar, R.; Choi, L.; Woo Lee, K. *PLOS ONE* **2016,** *11*, 1-19.

76. Sultan, M. M.; Denny, R. A.; Unwalla, R.; Lovering, F.; Pande, V. S. *Sci. Rep.* **2017,** *7*, 1-11.

77. 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).

78. 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.

79. 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).

80. Smith, J. M.; Jami Alahmadi, Y.; Rowley, C. N. *J. Chem. Theory Comput.* **2013,** *9*, 4860-4865.

81. Kruse, H.; Goerigk, L.; Grimme, S. *J. Org. Chem.* **2012,** *77*, 10824-10834.

82. Fedorov, O.; Müller, S.; Knapp, S. *Nat. Chem. Biol.* **2010,** *6*, 166-169.

83. Botta, M. *ACS Med. Chem. Lett.* **2014,** *5*, 270-270.

84. Thomas, B. E.; Kollman, P. A. *J. Org. Chem.* **1995,** *60*, 8375-8381.

85. Wang, C.; Qi, C. *Tetrahedron* **2013,** *69*, 5348-5354.

86. Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. *ChemMedChem* **2010,** *5*, 869-880.

87. Carlqvist, P.; Svedendahl, M.; Branneby, C.; Hult, K.; Brinck, T.; Berglund, P. *ChemBioChem* **2005,** *6*, 331-336.

88. Martell, J. M.; Beaton, P. T.; Holmes, B. E. *J. Phys. Chem. A* **2002,** *106*, 8471-8478.

89. Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008,** *321*, 792-794.

90. Pereira, S. R.; Vasconcelos, V. M.; Antunes, A. *FEBS J.* **2013,** *280*, 674-680.

91. Mulliner, D.; Wondrousch, D.; Schüürmann, G. *Org. Biomol. Chem.* **2011,** *9*, 8400-8412.

92. Chai, J.-D.; Head-Gordon, M. *PCCP* **2008,** *10*, 6615-6620.

93. Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004,** *393*, 51-57.

94. Livshits, E.; Baer, R. *PCCP* **2007,** *9*, 2932-2941.

95. Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. *J. Chem. Phys.* **2007,** *126*, 1-9.

96. Wereszczynski, J.; McCammon, J. A. *Q. Rev. Biophys.* **2012,** *45*, 1-25.

97. Eyring, H. *J. Chem. Phys.* **1935,** *3*, 107-115.

98. *MacroModel, version 9.7*, Schrödinger: LLC, New York, NY, 2009.

99. Chen, I. J.; Foloppe, N. *Biorg. Med. Chem.* **2013,** *21*, 7898-7920.

100. Barron, L. D. Fundamental Symmetry Aspects of Molecular Chirality. In *New Developments In Molecular Chirality* Mezey, P. G., Ed.; Springer, Dordrecht: 1991; Vol. 5, pp 1-55.

101. 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.

102. Krenske, E. H.; Petter, R. C.; Zhu, Z.; Houk, K. N. *J. Org. Chem.* **2011,** *76*, 5074-5081.

103. 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. 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.

104. Dennington, R.; Keith, T. A.; Millam, J. M. *GaussView, Version 6*, Semichem Inc., Shawnee Mission, KS, 2016.

105. 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.

106. Parr, R. G.; Szentpály, L. v.; Liu, S. *J. Am. Chem. Soc.* **1999,** *121*, 1922-1924.

107. Pearson, R. G. *J. Am. Chem. Soc.* **1963,** *85*, 3533-3539.

108. Koopmans, T. *Physica* **1934,** *1*, 104-113.

109. GÓMEZ-JERIA, J. S. *Journal of the Chilean Chemical Society* **2009,** *54*, 482-485.

110. Bickelhaupt, F. M.; Houk, K. N. *Angew. Chem. Int. Ed.* **2017,** *56*, 10070-10086.

111. Morokuma, K.; Kitaura, K. Energy Decomposition Analysis of Molecular Interactions. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials: Reactivity, Structure, Scattering, and Energetics of Organic, Inorganic, and Biological Systems*, Politzer, P.; Truhlar, D. G., Eds.; Springer US: Boston, MA, 1981; pp 215-242.

112. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2013,** *9*, 3878-3888.

113. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2012,** *8*, 1542-1555.

114. Le Grand, S.; Götz, A. W.; Walker, R. C. *Comput. Phys. Commun.* **2013,** *184*, 374-380.

115. 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.

116. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J.* **2011,** *40*, 843.

117. 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.

118. Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. *J. Chem. Theory Comput.* **2018,** *14*, 5834-5845.

119. *The PyMOL Molecular Graphics System, Version 2.0*, Schrödinger: LLC, New York, NY, 2015.

120. Guex, N.; Peitsch, M. C. *Electrophoresis* **1997,** *18*, 2714-2723.

121. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011,** *7*, 2284-2295.

122. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011,** *7*, 525-537.

123. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992,** *13*, 952-962.

124. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81*, 3684-3690.

125. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995,** *103*, 8577-8593.

126. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977,** *23*, 327-341.

127. Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013,** *9*, 3084-3095.

128. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996,** *14*, 33-38.

129. Gerlt, J. A.; Kozarich, J. W.; Kenyon, G. L.; Gassman, P. G. *J. Am. Chem. Soc.* **1991,** *113*, 9667-9669.

130. Highbarger, L. A.; Gerlt, J. A.; Kenyon, G. L. *Biochemistry* **1996,** *35*, 41-46.

131. Guillén Schlippe, Y. V.; Hedstrom, L. *Arch. Biochem. Biophys.* **2005,** *433*, 266-278.

132. Lu, J.; Zhang, Z.; Ni, Z.; Shen, H.; Tu, Z.; Liu, H.; Lu, R. *Comput. Biol. Chem.* **2014,** *52*, 25-33.

133. McClory, J.; Hu, G.-X.; Zou, J.-W.; Timson, D. J.; Huang, M. *J. Phys. Chem. B* **2019,** *123*, 2844-2852.

134. McClory, J.; Lin, J.-T.; Timson, D. J.; Zhang, J.; Huang, M. *Org. Biomol. Chem.* **2019,** *17*, 2423-2431.

135. McClory, J.; Timson, D. J.; Singh, W.; Zhang, J.; Huang, M. *J. Phys. Chem. B* **2017,** *121*, 11062-11071.

136. Montenegro, M.; Garcia-Viloca, M.; Lluch, J. M.; González-Lafont, À. *PCCP* **2011,** *13*, 530-539.

137. Shi, T.; Lu, Y.; Liu, X.; Chen, Y.; Jiang, H.; Zhang, J. *J. Phys. Chem. B* **2011,** *115*, 11895-11901.

138. van der Kamp, M. W.; Mulholland, A. J. *Biochemistry* **2013,** *52*, 2708-2728.

139. Salvadori, E.; Di Valentin, M.; Kay, C. W. M.; Pedone, A.; Barone, V.; Carbonera, D. *PCCP* **2012,** *14*, 12238-12251.

140. Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramírez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. *Chem. Rev.* **2018,** *118*, 4071-4113.

141. Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005,** *109*, 17715-17733.

142. Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. *Chem. Rev.* **2017,** *117*, 13502-13565.

# APPENDIX

## Benchmarking of Force Fields for Conformational Sampling

A benchmarking of 6 force fields was carried out to assess the similarities between them. The MM2 and MM3 force fields are not included due to their relatively inaccurate GB/SA solvation energies compared to other force fields. The conformational searches were conducted on 3 different species, consisting of a reactant, a thiol adduct, and a TS. The same settings as described in Section 2.1 are employed here and the results are tabulated in Table S1. The MMFF family force fields returned rather consistent results, with MMFFs outputting 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 the reactant and TS but managed to search a wider conformational space of the more flexible thiol adduct compared to the other force fields except AMBER force field. OPLS3e force field, being “the most extensively parametrized force field” and hence “recommended for all applications” by Schrödinger, was eventually chosen for the conformational sampling.

**Table S1. Conformational sampling of R1 using different force fields.**

|  |  |  |  |
| --- | --- | --- | --- |
| Force field | Number of conformers | | |
| R1 | P1\_SR | TS1\_R\_anti |
| 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 |

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

## Failure of MacroModel to Locate Stable s-*cis* Conformer

**Table S2. Conformational sampling of QM optimised R1 using different parameters.**

|  |  |  |  |
| --- | --- | --- | --- |
| Sampling Method | Cutoff Range | Implicit Solvent | Number of s-cis Conformers |
| Monte Carlo Multiple Minimum (MCMM) | Normal | Water | 0 |
| None | 4 |
| Extended | Water | 0 |
| None | 8 |
| None | Water | 0 |
| Octanol | 0 |
| None | 5 |
| Systematic Pseudo Monte Carlo (SPMC) | Normal | Water | 0 |
| None | 4 |
| Extended | Water | 0 |
| None | 4 |
| None | Water | 0 |
| Octanol | 0 |
| None | 4 |
| Mixed Torsional/Low-Mode Sampling (MTLMS) | Normal | Water | 0 |
| None | 4 |
| Extended | Water | 0 |
| None | 4 |
| None | Water | 0 |
| Octanol | 0 |
| None | 4 |
| Low-Mode Sampling (LMS) | Normal | Water | 0 |
| None | 4 |
| Extended | Water | 0 |
| None | 4 |
| None | Water | 0 |
| Octanol | 0 |
| None | 4 |
| Large Scale Low-Mode Sampling (LSLMS) | Normal | Water | 0 |
| None | 285 |
| Extended | Water | 0 |
| None | 223 |
| None | Water | 0 |
| Octanol | 0 |
| None | 237 |
| Mixed Torsional/Large Scale Low-Mode Sampling (MTLSLMS) | Normal | Water | 0 |
| None | 5 |
| Extended | Water | 0 |
| None | 4 |
| None | Water | 0 |
| Octanol | 0 |
| None | 4 |

## Comparison of CPU Time between Different Methods

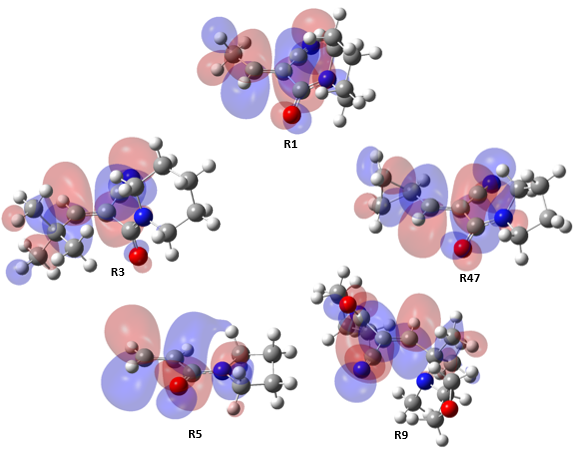
The CPU time taken for each combination of functional and basis set to calculate the SPE of **10**, which consists of 23 atoms, was tabulated in Table S3. Despite being known to return equivalent accuracy, pairing of aug-cc-pVTZ with the same functional as 6-311+G(d,p) resulted in more than twentyfold longer CPU time for the same calculation. This is rationalised by the fact that aug-cc-pVTZ employs more than twice the amount of basis functions compared to 6-311+G(d,p). Considering the equivalence of the calculation accuracy, there is no obvious reason for researcher studying similar chemical system to utilise the aug-cc-pVTZ basis set (in combination with ωB97X-D). Calculations using ωB97X-D generally require slightly less computations compared to M06-2X. Comparison between Methods B and J (M06-2X) and Methods D and F (ωB97X-D) showed that the addition of an extra d polarization function amplifies the CPU time by about 150%.

**Table S3. CPU time taken to compute SPE of compound 10 for each method.**

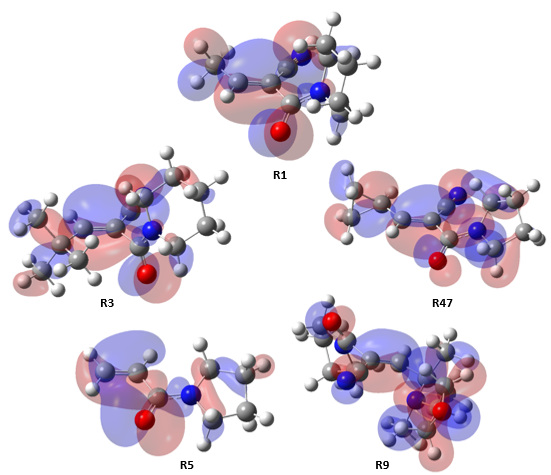
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | Functional | Basis set | CPU Time (min) | Ratio | Number of basis functions |
| **A** | B2PLYP-D | 6-311G(2d,p) | 58 | 3.1 | 454 |
| **B** | M06-2X | 6-311G(2d,p) | 30 | 1.6 | 454 |
| **C** | SCS-MP2 | 6-311G(2d,p) | 61 | 3.2 | 454 |
| **D** | ωB97X-D | 6-311G(2d,p) | 28 | 1.5 | 454 |
| **E** | ωB97X-D | aug-cc-pVTZ | 789 | 41.5 | 1039 |
| **F** | ωB97X-D | 6-311G(d,p) | 19 | 1.0 | 374 |
| **G** | ωB97X-D | 6-311+G(d,p) | 37 | 1.9 | 438 |
| **H** | B2PLYP-D | 6-31+G(d) | 26 | 1.4 | 334 |
| **I** | M06-2X | 6-311+G(d,p) | 37 | 1.9 | 438 |
| **J** | M06-2X | 6-311G(d,p) | 21 | 1.1 | 374 |

## Identification of the Most Relevant HOMO for Electrophilicity Index Computation

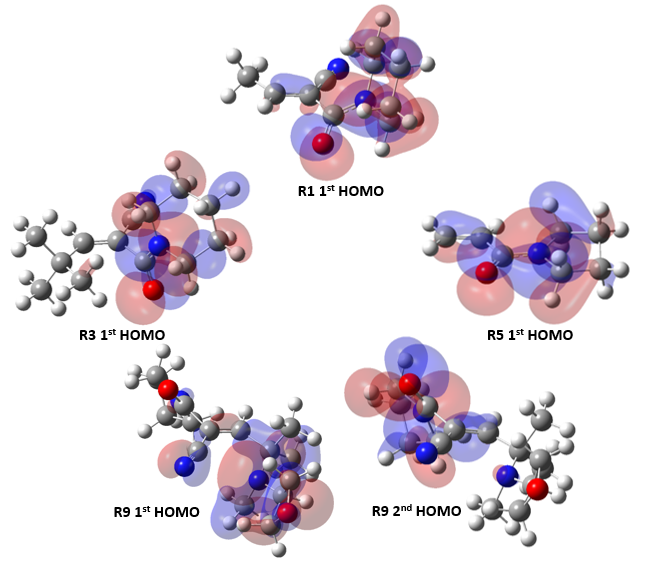
The chosen LUMOs and HOMOs of the reactants are visualised as shown in Figure S1 and S2 to ensure that the molecular orbitals chosen for the calculation of based on Koopman’s theorem are appropriate. The 2nd HOMO is chosen for **R1**, **R3**, and **R47** while the 3rd HOMO of **R9** is chosen due to their much greater coefficients on the electrophilic carbon of interest compared to their previous HOMOs. The HOMOs deemed unsuitable are shown in Figure S3.



**Figure S1.** Visualisation of the LUMO of the truncated Michael acceptors.



**Figure S2.** Visualisation of the chosen HOMOs of the truncated Michael acceptors.



**Figure S3.** Visualisation of the inappropriate HOMOs of the truncated Michael acceptors.

## Investigation on TS5 Methylthiolate Distortion

The distortion of the methylthiolate in TS from its ground state geometry could be seen most clearly from the superposition of the anion in **TS5** with the optimised structure. An attempt was made to explain the phenomenon observed by listing all degrees of freedom of the methylthiolate in ground state and **TS5** as shown in Table S4. The S-C bond of the methylthiolate was found to be elongated in the TS structure. The S-C-H angle with H corresponding to the proton attracted towards the cloud electrons of **R5** is reduced by 3.7° while the angle between the other 2 protons has deviated from the even angle of 107.6° in the reactant state and became more spread out by 1.4°. We propose that the presence of the nitrile group on the position of the other Michael acceptors lowers the cloud electron density near the site of interaction. The smaller extent of their methylthiolate distortions could then be explained by the weaker CH- attractive interactions.

**Table S4.** All degrees of freedom of the methylthiolate anion as reactant (left) for addition to **R5** and in the corresponding TS (right).

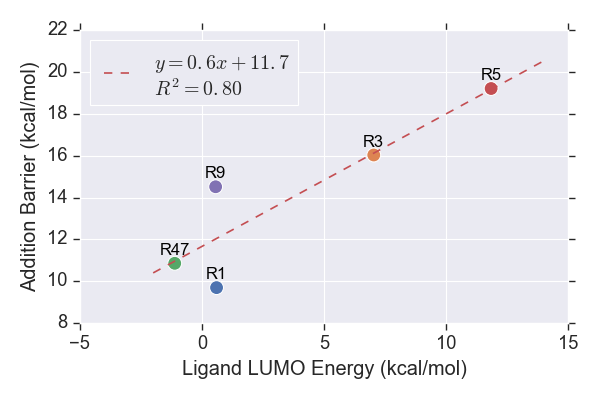
|  |  |  |  |
| --- | --- | --- | --- |
| Molecule | | Reactant | Transition State |
| Bond distance | S2-C1 (Å) | 1.84 | 1.81 |
| C1-H3 (Å) | 1.09 | 1.09 |
| C1-H4 (Å) | 1.10 | 1.10 |
| C1-H5 (Å) | 1.10 | 1.09 |
| Bond angles | S2-C1-H3 (°) | 111.4 | 111.0 |
| S2-C1-H4 (°) | 111.3 | 112.1 |
| S2-C1-H5 (°) | 111.3 | 107.6 |
| H3-C1-H4 (°) | 107.6 | 109.0 |
| H3-C1-H5 (°) | 107.6 | 108.5 |
| H4-C1-H5 (°) | 107.6 | 108.6 |

## QM Conformational Analysis on Results from Method G

The QM calculations were carried out using Method G to investigate the difference in the analysis results between the ωB97X-D and M06-2X functionals.

### Reactant LUMO Energies

The reactants LUMO energies calculated using Method G explains the variation addition barrier slightly worse than Method I.



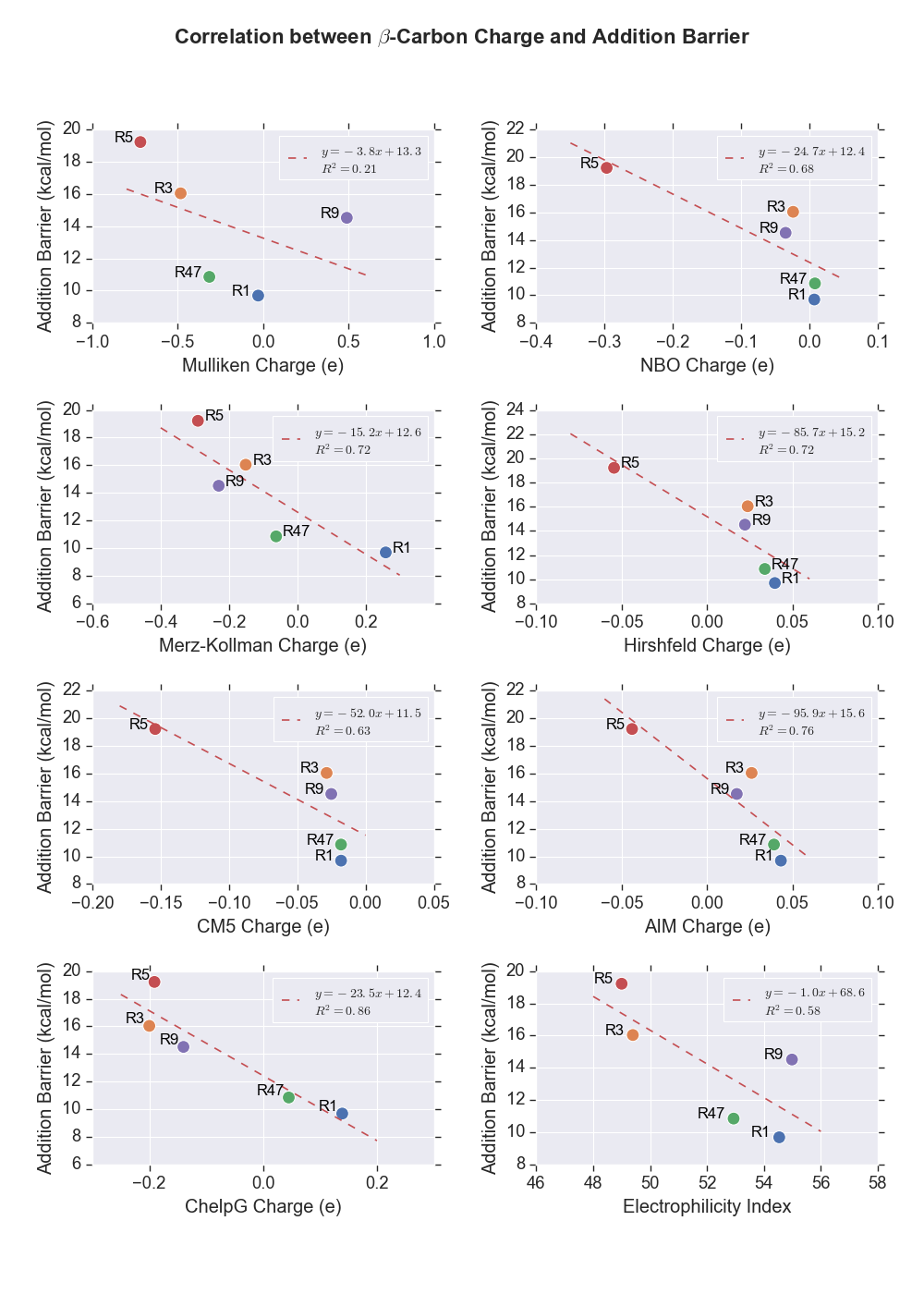
**Figure 7.** LUMO energies of the reactants.

### -Carbon Charges

The HOMOs chosen for the calculation of are roughly similar to Method I, except that the 3rd and 4th HOMOs are chosen for **R5** and **R9** instead of the 2nd and 3rd. The reason for the discrepancy in the energy hierarchy of the molecular orbitals is yet to be inspected.

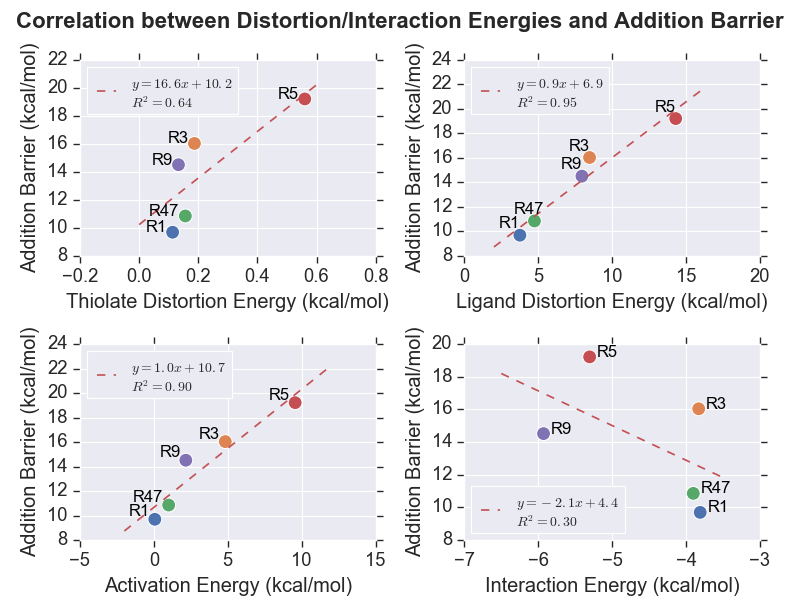
It is interesting to note that the description power of the addition barriers by was reduced to about 50% when the functional was changed to ωB97X-D. Despite the slightly poorer performance for most of the charge schemes, the Merz-Kollman and ChelpG charges calculated using Method G unexpectedly returned significantly higher values than calculations of Method I. It is possible that range-separated functionals are more suitable for computation of the 2 charge schemes. However, further investigations would be required for definitive conclusions to be drawn.

The correlation trends showed by the components in the distortion/interaction analysis are similar to Combination I. The methylthiolate in **TS5** is still found to exhibit a greater distortion compared to the other TSs. The ligand distortion energy and activation energy returned high values of as seen from the analysis on results from Combination G. The interaction energy remains lacking in the predictive power of the calculated addition barriers. Furthermore, the previously seen hypothetical ideal linear fit through the rest of the data points excluding **R9** is now unclear.



**Figure 8.** Partial charges of -carbon and of the Michael acceptors computed using Method G.

### Distortion/Interaction Analysis

****

**Figure 9.** Plots showing the relationship between distortion energies of thiolate (upper left), and cyanoacrylamides (upper right), activation energies (lower left) and interaction energies (lower right) computed using Method G.

## Details of MD Parameters Used

### Preparation of GROMOS System

A threshold of 0.1% was specified for the *gch* program, which repositions 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 as the *pmemd.cuda* program could not handle non-rectangular PBC just yet.

A cutoff of 0.8 was specified for the Coulomb potential calculation for the *ion* program as it searches for the water molecule with highest potential to be replaced with Na+ ions. It should be noted that the replacement of water molecule with the ion could only be done one at a time as the subsequent ions would be inserted at exactly the same coordinates as the first ion. The reason for the phenomenon was not investigated due to time constraint.

A harmonic force constant of 2.5×104 Nm-1 was used to restrain the positions of specified solute atoms.

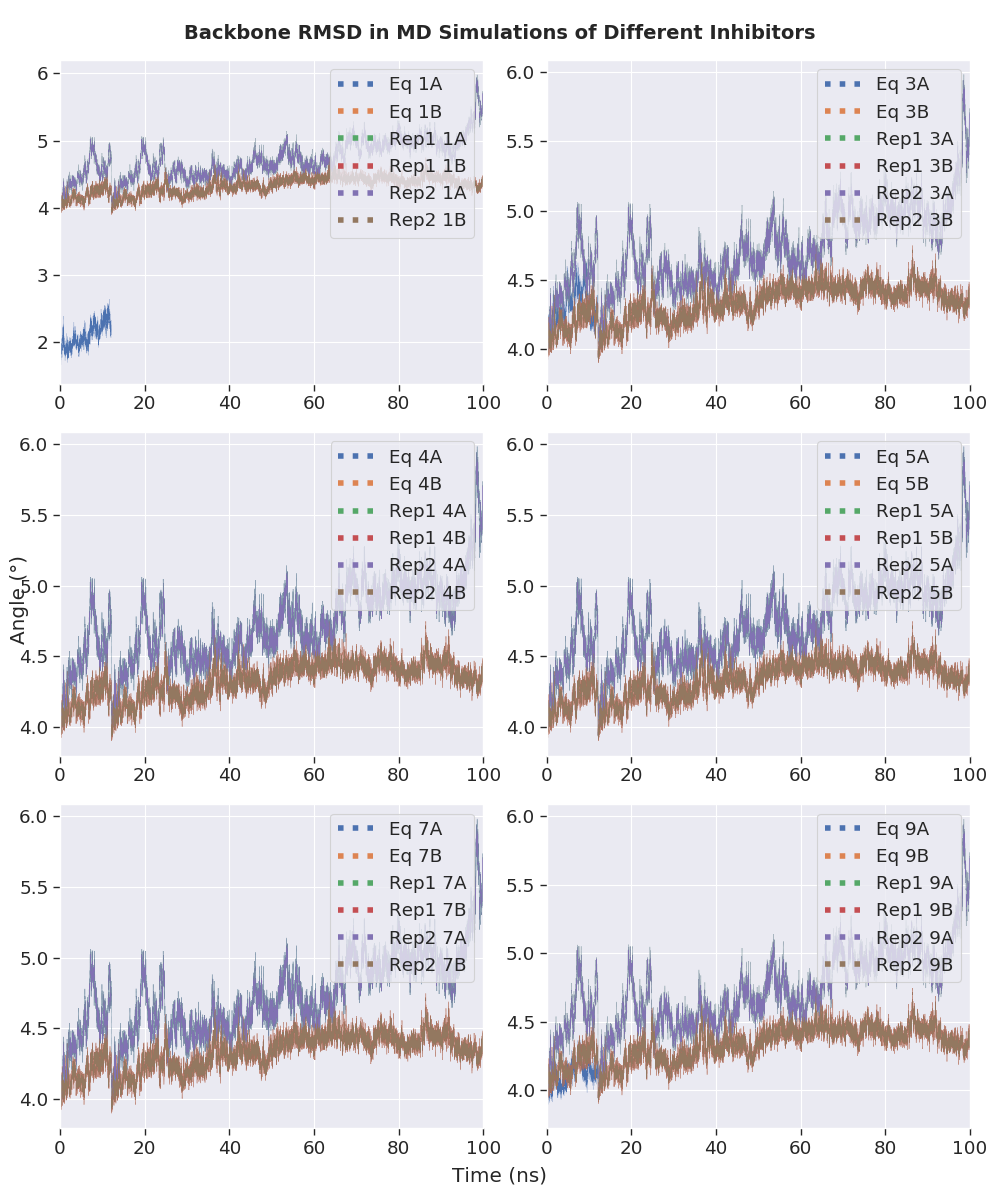
### Simulation of AMBER System

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 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 to prevent incompatibility with the Lennard-Jones parameters of the atoms in the systems built, which were parameterised without them. A cubic spline function was used as switching function of Coulomb forces. The skin distance for neighbour lists is set to 2 Å.

## Stability of Simulated BTK

### RMSD of Protein Backbones from X-ray Crystal Structure

The deviations of the protein backbones from the experimentally measured structure are plotted as shown in Figure 11. Due to the protocol of the preparation of the simulation systems which involves multiple energy minimisations, the initial structure for the simulations are already deviated for about 4 Å from the X-ray crystal structure from which the systems were built. However, the fluctuations of the RMSD throughout the MD simulations are relatively small.



**Figure 11.** RMSD of BTK protein backbones from the X-ray crystal structure throughout the 100 ns MD simulations.

### Hydrogen Bond Analysis

**Figure 12.** Distributions of the angles of the hydrogen bonds to the inhibitors throughout the 100 ns MD simulations of noncovalently bound BTK inhibitors.

## 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>.

### 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 raijin\_sub and rcc\_sub in Bash directory allows automatic submission of HPC jobs by looping through all directories.

### Management and Modification of Files and Directories

Functions in the Python file admin.py in QM/run\_gaussian directory allow easy name-changing of multiple files or directories simultaneously, generation of new directories according to existing files, grouping them correspondingly, and splitting concatenated coordinates files containing multiple molecules into files containing exclusive data of individual molecule. The rpname() function in the .bashrc file ease the name modification of entries in terminals.

### 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, -carbon charges, and distortion-interaction analysis). The interconversions between and , and , , and RT values were done using calculation.py in QM directory.

### Automated Analysis of MD Trajectories

Provided the template input files for the CPPTRAJ program, 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.

### 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 former provides functionality to map the overlapping atoms between 2 molecules, which could potentially be of use for other usage in the future. The plotting of the analysis of MD results was done using mdtraj\_analysis.py in the MD directory.