



**This electronic thesis or dissertation has been  
downloaded from Explore Bristol Research,  
<http://research-information.bristol.ac.uk>**

*Author:*  
**Voice, Angus**

*Title:*  
**Modelling the reactivity of cysteine targeting covalent inhibitors**

**General rights**

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

**Take down policy**

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact [collections-metadata@bristol.ac.uk](mailto:collections-metadata@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

# Modelling the reactivity of cysteine targeting covalent inhibitors

Angus Voice



A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the School of Chemistry, Faculty of Science.

# Abstract

Targeted covalent inhibitors (TCI) are compounds that are designed to exert their therapeutic effect through the formation of a covalent bond with a biological target. In spite of several perceived benefits of pursuing a covalent mechanism of action, concerns about off-target reactivity and selectivity remain. The most efficient way to accurately characterise covalent reactivity by computational approaches can be challenging, and the mechanism of chemical inhibition of high profile drug targets for some protein kinases remains an unanswered question. Computational modelling of covalent reactivity provides an attractive approach for investigating the determinants of reactivity between a covalent inhibitor and its target and can aid in the design of safer and more selective covalent drugs.

Simple ligand only reactivity metrics of covalent reactivity including proton affinity (PA) and reaction energy calculations were investigated with quantum mechanical (QM) methods. However, limitations in the predictive power of these methods for pharmaceutical lead-like compounds were identified, which led to a focus on reactivity assessments ‘in situ’. A benchmarking study of the most appropriate semi-empirical and density functional QM methods confirmed that specific methods should be used to accurately model thiol reactivity.

A comprehensive reactivity study was carried out for the covalent inhibitor ibrutinib that targets a non-catalytic cysteine residue in Bruton’s tyrosine kinase. Constant pH molecular dynamics simulations were used to calculate the  $pK_a$  of Cys481 in BTK and identified the neutral thiol group to be the most likely protonation state at physiological pH. Combined quantum mechanics/molecular mechanics (QM/MM) calculations in combination with umbrella sampling simulations were used to assess chemical inhibition pathways in BTK and led to the identification of a novel mechanistic pathway in BTK that is distinct from other protein kinases.

# Acknowledgements

I would like to express thanks to my supervisor, Professor Adrian Mulholland, whose help and guidance has led to a productive and enjoyable project.

I would also like to thank Janssen Pharmaceuticals, whose support and funding made this project possible. In particular I would like to acknowledge Herman van Vlijmen and Gary Tresadern, whose dedication, interest and engagement in the project has fuelled my ambition to pursue a career in drug discovery.

A very special thank you goes to Rebecca Twidale, whose unwavering help and support both in and out of the office has kept me going during the hardest of times. I really couldn't have done it without you.

The current and ex-members of the Centre for Computational Chemistry also deserve my gratitude. In particular, I would like to thank Marc van der Kamp, Eric Lang and Kara Ranaghan for numerous useful discussions and helping me to find my feet in the world of computational chemistry. Wise counsel from Natalie Fey has taught me many valuable lessons, for which I am eternally grateful.

Finally, I wish to acknowledge the love and emotional support I've received from my mum, Emma, and the rest of my family.



## Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Angus Voice

University of Bristol

December 2020

# Table of Contents

<b>Chapter 1 An introduction to covalent inhibitor drug design .....</b>	<b>1</b>
1.1 Background.....	1
1.2 Computational approaches to TCI design .....	7
1.2.1 Covalent docking and free energy calculations .....	7
1.2.2 QM methods: warhead reactivity .....	9
1.2.3 Cysteine reactivity.....	10
1.2.4 QM/MM mechanistic modelling.....	12
1.3 Covalent inhibitors of protein kinases.....	14
1.3.1 Bruton's tyrosine kinase (BTK) .....	15
1.3.2 BTK inhibitor modelling studies .....	17
1.4 Aims .....	18
<b>Chapter 2 Limitations of ligand-only approaches for predicting the reactivity of covalent inhibitors.....</b>	<b>20</b>
2.1 Abstract .....	20
2.2 Introduction.....	21
2.3 Methods .....	25
2.4 Results .....	27
2.5 Discussion .....	33
2.6 Conclusions .....	37
<b>Chapter 3 Investigating cysteine reactivity with classical Molecular Dynamics and Constant pH Molecular Dynamics Simulations .....</b>	<b>39</b>
3.1 Introduction.....	39
3.2 Methods .....	46
3.3 Results and discussion .....	47
3.3.1 Classical MD simulations .....	47
3.3.2 CpHMD .....	55
3.4 Conclusions .....	64
3.5 Suggestions for further work.....	65

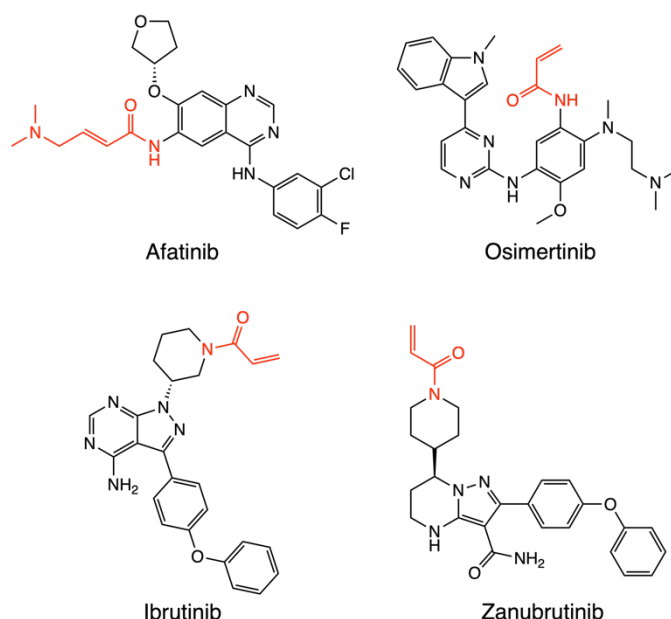
<b>Chapter 4 Evaluation of methods and convergence testing for modelling thio-Michael reactivity in biological systems .....</b>	<b>67</b>
4.1 Introduction .....	68
4.2 Methods .....	73
4.3 Benchmarking thiol addition with 1D potential energy scans .....	74
4.4 Benchmarking thiol addition by investigating alternative mechanisms .....	81
4.5 Benchmarking thiol addition in protein environments with umbrella sampling .....	87
4.6 Conclusions .....	98
4.7 Suggestions for further work .....	100
 <b>Chapter 5 Investigation of the mechanism of covalent inhibition of Bruton's tyrosine kinase by ibrutinib using QM/MM calculations .....</b>	 <b>103</b>
5.1 Introduction .....	103
5.2 Methods .....	108
5.3 Results .....	109
5.3.1 Mechanism 1: direct thiol addition and covalent keto product formation .....	109
5.3.2 Mechanism 2: solvent assisted thiol addition and covalent keto product formation .....	111
5.3.3 Mechanism 3: direct thiol addition and enol product formation .....	114
5.3.4 Mechanism 4: solvent assisted thiol addition and enol product formation .....	120
5.3.5 Solvent assisted tautomerisation .....	123
5.4 Comparisons with experimental kinetic measurements .....	127
5.5 Conclusions .....	129
5.6 Suggestions for further work .....	132
 <b>Appendix 1 .....</b>	 <b>135</b>
A1.1 Boltzmann averaging of $\Delta\Delta G_{PA}$ values from multiple low energy conformers .....	145
A1.2 Effects of using alternative optimisation strategies .....	147
A1.3 Effect of using an alternative DFT functional for PA calculation .....	148
A1.4 Quasi Harmonic Oscillator approximation .....	149
 <b>References .....</b>	 <b>151</b>

# Chapter 1

## An introduction to covalent inhibitor drug design

### 1.1 Background

Covalent inhibitors suppress their biological target through the formation of a covalent bond.<sup>1</sup> This is in contrast to conventional inhibitors that typically associate with their target by forming non-covalent, reversible interactions. Covalent mechanisms of action are routinely observed in biology, including in post translational modifications of proteins<sup>2</sup> and in cell signalling processes.<sup>3</sup> Covalent inhibitors differ in that they are specifically designed to contain a reactive group that is able to form a covalent bond with their target. Recently, this has led to interest in actively pursuing covalent compounds as pharmaceutical leads and the emergence of targeted covalent inhibitor (TCI) design.<sup>4</sup> TCIs contain a weakly electrophilic group, often referred to as a warhead, that can undergo covalent bond formation with a suitable nucleophilic target. The most utilised target is the thiol side chain of cysteine residues, given its potent nucleophilicity and low abundance in the proteome.<sup>5</sup> However, serine, lysine, tyrosine, histidine, arginine and threonine can all be targeted by covalent modifying agents, but often present significant challenges compared to cysteine targeting.<sup>6,7</sup> Numerous electrophilic warheads can be incorporated to inhibitor scaffolds, including acrylamides, sulfonyl fluorides, haloketones,  $\alpha$ -ketoamides, epoxides, alkynyl benzoxazines and dihydroquinazolines.<sup>8-10</sup> Acrylamides are by far the most common cysteine targeting warhead, and are found in several marketed covalent drugs that target protein kinases such as osimertinib, afatinib, ibrutinib and zanubrutinib (Figure 1.1).<sup>11</sup>



**Figure 1.1** Cysteine targeting covalent inhibitors osimertinib, afatinib, ibrutinib and zanubrutinib that are approved for the treatment of various cancers. Osimertinib and afatinib target epidermal growth factor receptor (EGFR) kinase, whereas ibrutinib and zanubrutinib target Bruton's tyrosine kinase (BTK). The acrylamide warhead in each drug is highlighted in red.

In spite of the renewed interest in TCIs in the last 10 – 15 years, drugs that act through a covalent mechanism of action have been around for over 100 years. Aspirin is a well-known drug that was marketed in 1888 and is the earliest example of a covalent drug. It exerts its biological action through the acetylation of the nucleophilic side chain of serine-530 in Prostaglandin endoperoxide synthase-1.<sup>12</sup> Halting the formation of prostaglandins stops pain signals being transmitted in the body, which makes aspirin an effective painkiller. Another milestone in the history of covalent inhibitors was the discovery of penicillin in the late 1920s. Penicillin was the first effective treatment for bacterial infections as it contains a  $\beta$ -lactam ring. This strained four-membered  $\beta$ -lactam ring covalently reacts with serine-36 in bacterial DD-transpeptidase, which stops bacteria from being able to synthesize cell walls.<sup>13</sup> Omeprazole, a commonly prescribed proton pump inhibitor (PPI) was the first drug of its kind to treat acid reflux and peptic ulcers. It works by forming a covalent disulfide bond with the sulfhydryl group of  $H^+/K^+$  ATPase.<sup>14</sup> Each of these covalent drugs were significant developments for various treatments and were often the first of their kind,

meaning previously untreatable conditions could be targeted by these drugs. This highlights the importance that covalent drugs have had on the therapeutic landscape since their inception. This continues today, but a significant difference between the earliest covalent inhibitors and those produced today are that they are designed to have a covalent mechanism of action, whereas the mechanism of early covalent inhibitors was often discovered retrospectively.<sup>15</sup>

Modern TCI drugs aim to capitalise on the advantageous properties of covalent inhibitors with respect to non-covalent drugs. Many drug candidates fail in clinical trials due to toxicity and poor efficacy.<sup>16</sup> Covalent inhibitors have the potential to overcome these issues as they exhibit many improved pharmacodynamic (PD) properties compared to their non-covalent counterparts. Potency, a function of binding affinity and target engagement is a desirable drug characteristic that improves drug efficacy.<sup>17</sup> Non-covalent drugs that form weaker, non-bonded interactions are considered to have a maximal possible binding affinity, estimated at around 10 pM in a recent study.<sup>18</sup> Potency can be increased by increasing molecular size, but this comes with the drawback of poor solubility, membrane permeability and metabolic stability.<sup>19</sup> In contrast, the formation of a covalent bond between an inhibitor and its target contributes to a strong binding free energy and results in highly potent compounds with a modest size that greatly exceed the binding affinity limits of non-covalently binding drugs.<sup>20</sup> Furthermore, much longer drug-target residence times can be achieved with covalent inhibitors, as target occupancy depends on the re-synthesis rate of the target rather than the pharmacokinetics (PK) of the drug.<sup>21</sup> Optimisation of residence times has been reported to be an important and undervalued strategy in drug design, particularly for covalent inhibitors.<sup>22,23</sup> Reversible covalent inhibitors have been sought that have long residence times but minimise the potential for off target effects.<sup>22</sup> A consequence of high potency and long residence time is the need for lower dosages of a drug, which aids in reducing toxicity and the potential for off target effects as lower concentrations of the drug are required to produce the same biological affect.<sup>11,24</sup> The

high potency that can be achieved through covalent bond formation also means that binding sites can be targeted that are usually difficult to target with non-covalent inhibitors. For example, the shallow and solvent exposed binding pocket in the cysteine protease Cathepsin K is often deemed undruggable, but has been successfully targeted by covalent inhibitors.<sup>25</sup> Drug resistance that arises from mutations of a target's binding pocket poses significant problems in drug discovery as it can dramatically reduce the potency of covalent and non-covalent drugs. Covalent inhibitors can theoretically overcome resistance mechanisms, provided the nucleophilic amino acid target of a covalent inhibitor remains unchanged and covalent bond formation can still occur. However, resistance can still be a problem for TCI's as their mechanism of action is heavily dependent on a single amino acid. This is a particular problem in protein kinases such as epidermal growth factor receptor (EGFR) and Bruton's tyrosine kinase (BTK), where the C797S mutation in EGFR and the C481S mutation in BTK have been documented as sources of resistance in these systems.<sup>26,27</sup> Changing the mechanism of action can help overcome resistance. For example, advances in treating the T970M mutation in EGFR have been made through the development of covalent inhibitors that specifically target the mutant kinase over the wild type (WT).<sup>28</sup>

Although there are numerous advantageous properties that can be profited from by pursuing a covalent mechanism of action, in the past there has been a reluctance by pharmaceutical companies to design drug candidates that contain reactive electrophilic groups that are able to covalently modify their target.<sup>29</sup> This is because electrophilic compounds can react indiscriminately with off target species, resulting in toxicity and safety related issues being associated with these types of compound.<sup>1,4,24</sup> Covalent modification of proteins and enzymes can result in an immune response as the covalent adduct is recognised as foreign. This is known as idiosyncratic drug toxicity.<sup>30</sup> A notable example is the covalent  $\beta$ -lactam antibiotic penicillin, which is known to cause an idiosyncratic drug reaction in some members of the population as a result of the  $\beta$ -lactam reacting with free amino and thiol groups in the body.<sup>31</sup> An

additional concern arising from the use of covalent drugs is the toxicity associated with non-covalent compounds that are metabolised in the body to form reactive electrophilic intermediates.<sup>32</sup> Paracetamol, a widely used pain killer, undergoes metabolism by cytochrome P450 enzymes to form the highly reactive N-acetyl-p-benzoquinone imine (NAPQI) intermediate that reacts with live proteins and results in hepatotoxicity.<sup>32</sup> However, TCI drugs are specifically designed so that they are selective for their target (through optimisation of non-covalent binding interactions) and contain only weakly electrophilic warheads that are designed to react with a specific target nucleophile in contrast to drug metabolites that have high levels of intrinsic reactivity.<sup>1</sup> These factors help TCI drugs to have an improved safety profile over reactive electrophiles that can irreversibly bind with biological macromolecules in an unselective fashion.<sup>33</sup> Safety concerns surrounding covalent inhibitors has led to them being reserved for scenarios where there is a strong medical need, for example if there are no appropriate treatments currently available or the biological target is difficult to inhibit with conventional strategies. This has resulted in an immense effort to produce covalent therapeutics that target protein kinases as they play a vital role in many disease processes including cancer. The ATP binding pocket of protein kinases is well conserved across the protein kinome, presenting a significant challenge to developing selective non-covalent inhibitors to target specific kinases.

Careful design principles must be employed in TCI design to minimise the potential for off target effects and alleviate toxicity concerns. Care must be taken as there are subtle differences between covalent and non-covalent drugs that must be taken into account. A commonly used measure of a reversible drug's potency is the half maximal inhibitory concentration ( $IC_{50}$ ).<sup>34</sup> This refers to the equilibrium concentration of an inhibitor that is required to achieve 50% blockade of the biological target. However, covalent inhibitors exhibit non-equilibrium binding as covalent bond formation means the inhibitor does not have to compete with endogenous ligands.<sup>15,35</sup> As a result, covalent inhibitor  $IC_{50}$  values are time dependant and are unsuitable



potency metrics. A more robust measure of the potency of a covalent inhibitor is to assess the  $k_{\text{inact}}/K_i$  ratio (Scheme 1), and is the recommended measure for assessing the potency of covalent inhibitors.<sup>1,4,35,36</sup>  $K_i$  is an inhibitory equilibrium constant and is a measure of the reversible, non-covalent association of a covalent inhibitor with its target. The inactivation rate,  $k_{\text{inact}}$ , is a kinetic first order rate constant that defines the rate of covalent bond formation. Experimental determination of the  $k_{\text{inact}}/K_i$  ratio is more difficult than performing  $\text{IC}_{50}$  measurements, so there is a tendency for  $\text{IC}_{50}$  values to still be used to rank the potency of covalent inhibitors.<sup>35</sup> In addition, the interplay between PD and PK can be difficult to predict for covalent inhibitors.<sup>37</sup> PD, defined as the effect a drug has on the body, is usually dependent on the PK (the kinetics of absorption, distribution, metabolism and excretion of the drug) for a conventional non-covalent inhibitor. The PD effect of a given covalent inhibitor is also dependent on the resynthesis rate of the protein target, as the inhibitor will remain bound to its target even after any free drug is excreted from the body.<sup>35</sup> This is a desirable characteristic for covalent inhibitors, as a rapid clearance of a covalent drug will reduce the potential for off-target effects.<sup>37</sup> On the other hand, if the excretion rate of a covalent drug is too high, for example through reaction with glutathione, it can result in reduced drug availability.<sup>38</sup> Careful consideration of PD and PK properties are therefore required in the design of covalent inhibitors, especially when optimising electrophilic warhead reactivity.



**Scheme 1.1.** Process of inhibition of a covalent inhibitor with its enzyme target. In the first step, the inhibitor reversibly binds with its target to form a non-covalent complex. The binding affinity for this process is denoted by  $K_i$ . If the lifetime of the non-covalent complex is sufficiently long, a reaction occurs between the inhibitor and its target to form the covalently bound E-I complex. The rate of covalent complex formation is given by  $k_{\text{inact}}$ .

## 1.2 Computational approaches to TCI design

Several computational and experimental tools are available to aid in the design of covalent inhibitors. Activity based protein profiling (ABPP) is a powerful experimental technique that has been used in the identification of biological targets that could be amenable to covalent targeting inhibitors.<sup>24,39</sup> ABPP uses a covalent probe that contains a reactive electrophilic warhead that is designed to react with nucleophilic sites in proteins.<sup>40</sup> It is attached to a tag molecule, often a fluorescent compound or biotin, that enables identification of the protein target.<sup>41</sup> This technique can be specific to a particular enzyme class such as serine hydrolases,<sup>42</sup> cysteine proteases<sup>43</sup> and protein kinases<sup>44</sup> through the incorporation of a binding group that increases selectivity for a target.<sup>24</sup> Experimental approaches are also used to determine covalent reactivity, with the most common method being the glutathione reactivity assay (Section 1.2.2). In addition to experimental approaches, computational methods have been used successfully in covalent drug design. Computational tools are routinely used in drug design but are often developed for the design of non-covalent drugs.<sup>45</sup> However, the nature of covalent bond formation between an inhibitor and its target makes computational approaches that incorporate quantum mechanics particularly useful in covalent inhibitor design as chemical change can be modelled at the atomistic level.<sup>46</sup> Molecular modelling methods can be used to calculate binding affinities and reactivity rates of covalent drugs with their target through the calculation of  $K_i$  and  $K_{inact}$ . The following will provide a summary of some of the state-of-the-art computational techniques used for  $K_i$  and  $k_{inact}$  calculations in the context of covalently binding inhibitors.

### 1.2.1 Covalent docking and free energy calculations

Docking methods are widely used in the field of computer aided drug design (CADD) and enable huge libraries of compounds to be virtually docked against a protein or enzyme target. Alternative binding poses can be evaluated and are ranked

by docking scores that predict the binding affinity of the docked compound with its target. Most molecular docking programs are designed to predict binding poses and affinities of non-covalent ligands, but the resurgence of interest in covalent drug design has led to the development of several docking programs specifically for covalent docking including AutoDock4,<sup>47</sup> CovDock,<sup>48</sup> GOLD,<sup>49</sup> and ICM-Pro.<sup>50</sup> A recent study found that these six programs were successful in predicting binding poses within 2.0 Å of the experimental bind pose in up to 60% of cases, and that higher success rates were found for Michael acceptor warheads.<sup>51</sup> No docking program performs significantly better than others for all protein targets however, which has led to the development of protocols for choosing the most appropriate covalent docking tool for the required task based upon precision, generality and robustness.<sup>52</sup>

Free energy calculations also provide a means to assess the binding free energy of covalent drugs. Free energy perturbation (FEP) is a well-established technique for ranking the binding affinity of non-covalent ligands and is the gold standard used in the pharmaceutical industry. However, FEP uses molecular mechanics (MM) force fields (FF) which do not allow for chemical change. The use of FEP to rank the binding affinity of covalent compounds is therefore limited to optimisation of the non-covalent interactions in a drug discovery project but can still be useful. Modification to a typical FEP protocol, for example by adapting the thermodynamic cycle to include covalent interactions, has seen success in predicating the non-covalent binding affinity of 9  $\alpha$ -ketoamide inhibitors of calpain cysteine protease, where a correlation coefficient of  $R^2=0.70$  between experimental and predicted binding free energies was observed.<sup>53</sup> FEP was also found to outperform other approaches in predicting the potency of 10 covalent nitrile inhibitors of human cathepsin-L.<sup>54</sup> Both covalent docking and free energy methods such as FEP are evidently useful tools in covalent drug design and can give reasonable estimates of  $K_i$ , the non-covalent binding affinity of a covalent drug (Scheme 1.1). However, these methods are unable to provide estimates of covalent reactivity.

### 1.2.2 QM methods: warhead reactivity

Quantum mechanics (QM) methods have been used extensively to predict covalent warhead reactivity. Computational models that aim to predict the reactivity of covalent warheads with cysteine typically rely on estimating GSH  $t_{1/2}$ . This is the half-life of the reaction of an electrophilic compound with glutathione (GSH). The glutathione reactivity assay (GSH  $t_{1/2}$ ) is a widely used model for predicting drug toxicity in the pharmaceutical industry, but also serves as a convenient way of comparing warhead reactivity of covalent inhibitors with cysteine. In a typical kinetic glutathione assay the reactivity of a covalent warhead is measured in solution with an excess of glutathione to approximate pseudo first order kinetics.<sup>55</sup> The concentration of the covalent warhead is reduced as the reaction proceeds and this can be tracked by NMR measurements or liquid chromatography coupled with mass spectrometry.<sup>6,56</sup> The rate constants can then be determined by plotting the natural log of the warhead concentration against time. An early example of predicting GSH  $t_{1/2}$  reactivity used QM transition state (TS) calculations on a series of 22  $\alpha$ ,  $\beta$ -unsaturated compounds with methylthiolate (a computational surrogate for GSH).<sup>57</sup> A strong correlation ( $r^2=0.90$ ) was observed between predicted and experimental kinetic rate constants ( $k_{\text{GSH}}$ ) of GSH adduct formation. In addition, Mulliner *et al.* used TS calculations to predict  $k_{\text{GSH}}$  of 35  $\alpha$ ,  $\beta$ -unsaturated compounds with methylthiolate and found a strong correlation ( $r^2=0.96$ ) between their calculated  $k_{\text{GSH}}$  values and experimental GSH reactivity data.<sup>58</sup> A similar quantum chemical protocol was employed by Flanagan *et al.* to predict the reactivity rates of 16 acrylamide containing electrophiles with methyl thiolate.<sup>56</sup> The calculated kinetic reaction barrier ( $\Delta G^\ddagger$ ) was strongly correlated ( $R^2=0.92$ ) to the experimental GSH reactivity rate (GSH  $t_{1/2}$ ). As a result of using computationally intensive QM TS calculations for GSH  $t_{1/2}$  predictions, the reactivity of only a relatively small numbers of compounds can be compared. This limits the utility of these QM approaches to small data sets, and the comparison of structurally similar compounds. As a result, the reactivity predictions cannot reliably be generalised to compounds with diverse chemical functionality. Furthermore, it can be difficult to

locate TSs for thiol addition reactions using QM methods due to the flat nature of the potential energy surface surrounding the TS.<sup>59</sup> To address this, Lonsdale *et al.* used reaction energies ( $\Delta E_{\text{react}}$ , the difference in energy between reactant and covalent adduct states) as a reactivity metric for GSH reactivity.<sup>60</sup>  $\Delta E_{\text{react}}$  values of a diverse set 46 electrophilic compounds that contained different reactive warheads were calculated at the M06-2X/6-31+G(d,p) level. A reasonable correlation of  $R^2=0.69$  was observed between predicted reaction energy and experimental GSH  $t_{1/2}$ . The authors cited potential errors in the computational methodology and conformational effects as possible explanations.

Proton affinity (PA) calculations are an additional reactivity metric that can be used to predict covalent warhead reactivity that avoid the use of computationally intensive QM TS optimisations. PA calculations have been shown to be a good predictor of thiol elimination rates by Krishnan *et al.*, who found a strong correlation between the calculated PA of a set of eight acrylonitrile based inhibitors and the elimination rate ( $\log k$ ) from  $\beta$ -mercaptoethanol (BME).<sup>61</sup> However, as detailed in Chapter 2 of this thesis, a thorough investigation in to the use of PA calculations as a predictor for covalent warhead reactivity found significant limitations in the approach.<sup>62</sup> More recently, the electrophilicity index has been investigated as a covalent warhead reactivity metric.<sup>63</sup> When combined with a machine learning model named BReactive, a reasonable correlation ( $R^2=0.85$ ) between activation energy (predicted by BReactive) and the experimental GSH reactivity was observed for acrylamide compounds.<sup>64</sup> The predictive power of the model was reduced when used for compounds that contain a chloroacetamide warhead ( $R^2=0.69$ ).

### 1.2.3 Cysteine reactivity

Computational models have also been used to assess the reactivity of the thiol group of cysteine residues that are targeted by covalent warheads. The predominant

strategy is to estimate the  $pK_a$  of the cystine thiol in question as the thiol acidity indicates how readily the thiol group can lose a proton to form the more nucleophilic thiolate anion.<sup>65</sup> For a detailed discussion of the methods employed to predict cysteine reactivity, see Chapter 3 of this thesis. The most robust and widely used methods are molecular mechanical FF methods such as constant pH molecular dynamics (CpHMD) simulations and thermodynamic integration (TI).<sup>66</sup> An explicit solvent all atom replica exchange approach was found to outperform simpler implicit solvent empirical  $pK_a$  prediction models including PROPKA, MCCE and H++ when the  $pK_a$ s of 18 cysteine residues were assed with these methods.<sup>65</sup> Explicit solvent CpHMD simulations are able to account for the coupling of protonation states from titratable amino acid side chains in the vicinity of the cysteine residue, and have seen reasonable success in predicting cystine  $pK_a$  values. However, they can struggle to predict acidic shifted  $pK_a$  values, as demonstrated in a recent investigation of the protonation state in cystine proteases.<sup>67</sup> An implicit solvent continuous CpHMD implementation developed by Jana Shen *et al.* has successfully predicted the  $pK_a$ s of cystine residues with acidic shifted  $pK_a$ s.<sup>68</sup> Machine learning models have also been developed for cysteine reactivity predictions. Zhang *et al.* built a supported vector machine (SVM) learning model to predict cysteine  $pK_a$  values with an AUC accuracy of 0.73, using descriptors that included reactivity, binding affinity of ligands, position and amino acid sequence.<sup>69</sup> Specific sequence based cysteine  $pK_a$  prediction models help to improve accuracy, with one sequence based machine learning model that has been reported with a high accuracy (98%) and precision (95%) for predicting cysteine reactivity.<sup>70</sup> In spite of the accuracy afforded by many of the aforementioned computational cysteine  $pK_a$  prediction tools, further investigation is needed as a result of the complex nature of accurately modelling sulfur. This is especially important for FF methods including TI and CpHMD as the parameters used to describe the bonded and non-bonded interactions can sometimes be a poor description of sulfur chemistry. For example, the AMBER FF uses the same non bonded van der Waals parameters to describe a sulfur atom in a cysteine thiol and the sulfur of a negatively charged cysteine thiolate anion.<sup>71</sup>

The result is a poor description of the hydration structure of sulfur containing species in the AMBER FF. This can translate to inaccurate  $pK_a$  predictions, given the close relationship between sulfur hydration number and cysteine  $pK_a$ .

#### 1.2.4 QM/MM mechanistic modelling

When QM methods are combined with MM methods, mechanistic modelling and reactivity predictions can be carried out in situ, and reactivity assessments for an inhibitor, or series of inhibitors can be made when bound to its protein target.<sup>72</sup> QM approaches are widely used for mechanistic modelling of enzymatic reactions<sup>73</sup> and have been useful in modelling the reaction mechanisms of covalent inhibitors with their target protein.<sup>74–76</sup> In a typical QM/MM reaction simulation, the system of interest is partitioned into two subsections, one that is treated with MM methods and one that is treated with QM methods.<sup>77,78</sup> Several approaches can be used when covalent bonds cross the boundary between the QM and MM regions, with one of the most popular being the link atom method.<sup>79</sup> This is especially important for modelling covalent inhibitors, as the reactive electrophilic warhead of the inhibitor must be included in the QM region, in addition to the amino acid side chain of the nucleophilic amino acid residue that is targeted by the inhibitor. One way of evaluating the total energy in combined QM/MM methods is to use an additive approach, as shown in Equation 1.1. Here, the total QM/MM energy is a sum of the energy of the QM region ( $E_{QM}$ ), the energy of the MM region ( $E_{MM}$ ) and the interaction energy between the QM and MM regions ( $E_{QM/MM}$ ).<sup>80</sup>

**Equation 1.1** Additive QM/MM energy expression.

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM}$$

The use of QM/MM methods for modelling covalent inhibitors are discussed extensively in Chapter 5 of this thesis, but some notable examples will briefly be

highlighted here. The majority of covalent inhibitor QM/MM studies are performed on protein kinase and cysteine protease targets, where a cysteine residue is the target of covalent modification in both cases. In cysteine proteases, the cysteine operates in a catalytic dyad with a nearby histidine residue to cleave peptide bonds. Understanding the molecular mechanism of this peptide bond cleavage is essential to design covalent inhibitors that can irreversibly inactivate the cysteine protease and are essential in viral replication. The global pandemic of 2019/2020 caused by the Sars-CoV-2 coronavirus has highlighted the vital importance of the development of new antiviral drugs that target cysteine proteases.<sup>81</sup> A QM/MM mechanistic study was able to determine the molecular mechanism of proteolysis in the SARS-CoV-2 main protease ( $M^{\text{pro}}$ ).<sup>76</sup> QM/MM protocols have also been used to study covalent inhibitors that target rhodesain, cruzain and cathepsin L cysteine proteases.<sup>72,82,83</sup> Another high-profile disease area that has benefit from mechanistic QM/MM studies is in the development of anti-cancer drugs. EGFR is one example of a receptor tyrosine kinase, where overexpression of this transmembrane protein has been linked to diseases such as Alzheimer’s and a variety of tumours.<sup>84</sup> Several covalent inhibitors of EGFR have been approved for the treatment of lung cancer such as osimertinib and afatinib.<sup>85</sup> A QM/MM modelling study of the mechanism of inhibition EGFR by N-(4-anilinoquinazolin-6-yl) acrylamide was carried out and found that the rate limiting step was S-C bond formation and highlighted the importance of an active site aspartate residue that participates in the inhibition mechanism.<sup>74</sup> At the time of writing, a mechanistic QM/MM study was submitted to ChemRxiv of a cyano-acrylamide inhibitor that targets BTK.<sup>86</sup> This study modelled an ionic mechanism for the inhibition of the active site cysteine residue and compliments the detailed reactivity study in Chapter 5 of this work that examines multiple mechanistic pathways for BTK inhibition by covalent drug ibrutinib. It highlights that reaction mechanisms of covalent drugs can change when subtle differences are made to the reactive warhead.



### 1.3 Covalent inhibitors of protein kinases

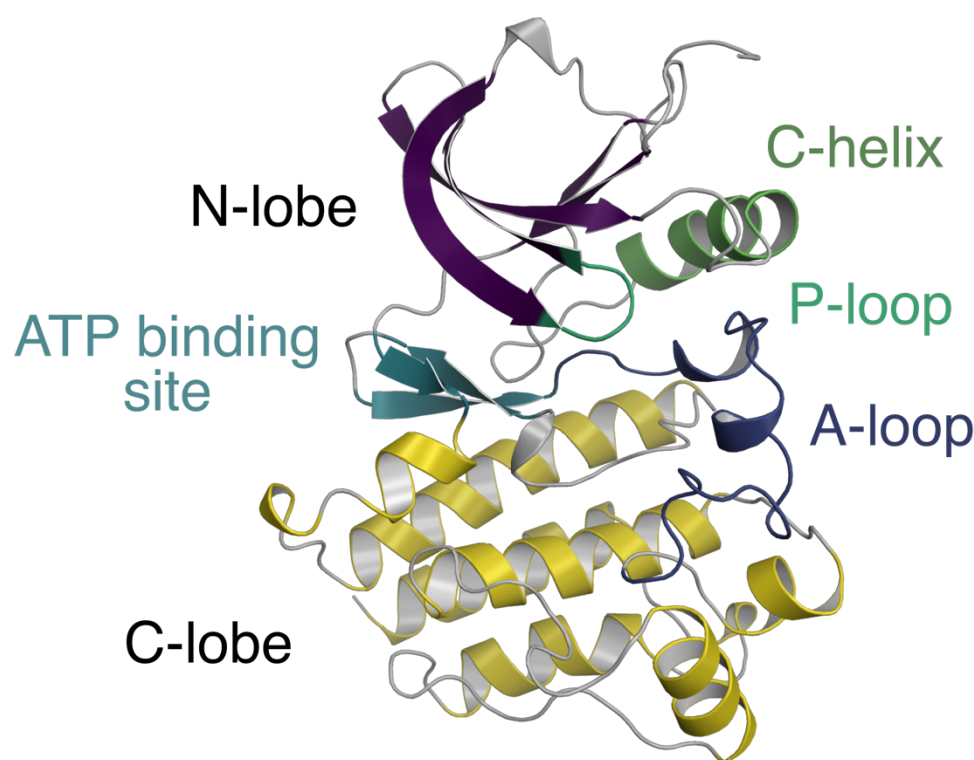
Many of the aforementioned examples of covalent inhibitors are compounds that target protein kinases. Protein kinases carry out phosphorylation of other proteins by transferring a phosphate group from adenosine triphosphate (ATP) and covalently attaching it to serine, threonine or tyrosine. There are approximately 500 protein kinases that make up 2% of human genes, making protein phosphorylation an essential component in all areas of human physiology.<sup>87</sup> The major functions of protein kinases are in the regulation of cell signalling pathways. Many diseases are linked to the overexpression or misregulation of protein kinases, making them important drug targets in drug discovery.<sup>88</sup> However, the high degree of structural similarity observed between protein kinase families means selective targeting of a particular protein kinase presents a significant challenge.<sup>89</sup> The proposed advantages of using targeted covalent inhibitors in drug discovery (increased selectivity, potency and ability to overcome resistance mechanisms)<sup>1,4</sup> make them ideal candidates for targeting protein kinases. Several covalent drugs that target protein kinases have been approved for the treatment of cancer, where there is often an unmet medical need for such therapeutics and resistance to current anti-cancer drugs can routinely occur.<sup>90,91</sup> BTK, a non-receptor tyrosine kinase, was chosen as the target for this work. The first covalent inhibitor targeting BTK was co-developed by Pharmacyclics and Johnson and Johnson,<sup>92</sup> the parent company of Janssen Pharmaceuticals who sponsored this work. BTK has not been studied as extensively as related kinase EGFR, which has been the focus of several computational modelling studies.<sup>74,93–96</sup> This could in part be due the complex chemistry and questions surrounding the mechanism of covalent modification of the cysteine residue C481 in BTK. BTK therefore represents an important and interesting target to study, and where increased understanding of the covalent binding of electrophiles to this target will be important for the design of new covalent drugs.

### 1.3.1 Bruton’s tyrosine kinase (BTK)

BTK belongs to the TEC family of non-receptor tyrosine protein kinases. The other four members of this family are bone marrow expressed kinase (BMX), interleukin-2-inducible T cell kinase (ITK), resting lymphocyte kinase (RLK) and tyrosine kinase expressed in hepatocellular carcinoma (TEC).<sup>97</sup> There are five protein interaction domains in BTK: a pleckstrin homology (PH) domain; a TEC homology (TH) domain, two SRC homology (SH) domains SH1 and SH2; and a kinase domain that possesses enzymatic activity (Figure 1.2A).<sup>98</sup> The focus of this work is on the kinase domain, as it is in the ATP binding site of the kinase domain where ibrutinib binds to BTK. There are several structures of the BTK kinase domain that are deposited in the protein data bank (PDB). Two structures in particular are used in this work, as they show an ibrutinib analogue non-covalently bound to the ATP binding site in BTK (PDB 5P9I)<sup>99</sup> and ibrutinib covalently bound to C481 in BTK (PDB 5P9J).<sup>99</sup> These structures show typical kinase structure consisting of a  $\beta$ -sheet dominated N-lobe and  $\alpha$ -helical C-lobe (Figure 1.2B). The activation loop (A-loop) controls substrate binding and connects each lobe.<sup>100</sup>

BTK plays an important function in several cell signalling pathways that are vital for normal B cell development. In addition, increased activity of BTK has been linked to its role in the survival and proliferation of tumours in B cell malignancies such as various types of leukaemia.<sup>98</sup> BTK is therefore a promising drug target and several BTK inhibitors have been developed as a result. The only approved drugs that target BTK are covalent kinase inhibitors that irreversibly bind to a cysteine residue, C481, situated on the edge of the ATP binding pocket. Ibrutinib has been approved for the treatment of mantle cell lymphoma, chronic lymphocytic leukaemia and Waldenström’s macroglobulinemia.<sup>101</sup> It contains an electrophilic acrylamide group that forms a covalent bond with C481.<sup>102</sup> Acalabrutinib contains a butynamide warhead and is approved for the treatment of mantle cell lymphoma in patients who have previously been treated with one other therapy and was designed to be a more

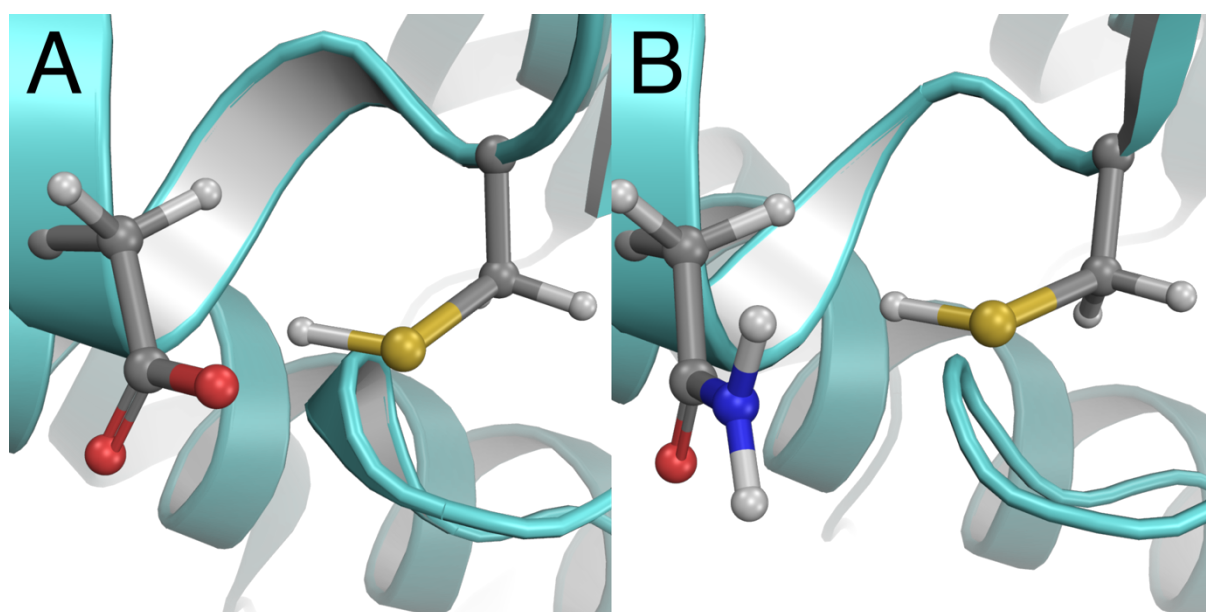
selective covalent BTK inhibitor than ibrutinib.<sup>103</sup> Zanubrutinib is another acrylamide containing covalent BTK inhibitor used for the treatment of MCL.<sup>104</sup> However, there are reports of resistance to these drugs as a result of a C481S mutation in the BTK kinase domain.<sup>27</sup> Non-covalent inhibitors have been investigated that target either the ATP binding pocket or the SH2/3 domain of BTK and work by stabilising an inactive kinase conformation and thus inhibiting its function.<sup>102,105</sup> However, no such inhibitors have yet been approved to target BTK although some are in clinical trials.<sup>106</sup> It is evident that there is a strong clinical need for improved BTK inhibitors, and further investigation into the mechanism of action of covalent drugs that target BTK are required to help in discovery and design of future therapies that can target BTK.



**Figure 1.2** The kinase domain in BTK from crystal structure 5P9J.<sup>99</sup> The characteristic N and C lobes of the kinase are highlighted, in addition to the C-helix, P-loop and A-loop which all perform vital roles in the functioning of BTK. The ATP binding site is also highlighted as this is the site that covalent inhibitors such as ibrutinib and zanubrutinib bind to BTK.

### 1.3.2 BTK inhibitor modelling studies

There are only a handful of computational studies that have been carried out on BTK to investigate its suitability as a drug target. A combined molecular docking and MD study was performed on a series of 2,5-diaminopyrimidine covalent inhibitors of BTK.<sup>107</sup> Important hydrogen bonding sites between gatekeeper residues Thr474, Met477 and active site residues Leu408 and Arg525 and the inhibitors were identified as important components of the binding potency. Only transient hydrogen bonds between C481 and the inhibitor structures were observed early on in the MD simulations. These interactions were confirmed by a separate, but similar study that used a combination of MD and virtual screening to investigate BTK inhibitors.<sup>108</sup> This study also used density functional theory calculations to assess the HOMO-LUMO gap of five of the most potent inhibitors predicted from docking studies as a way of assessing reactivity. However, similar frontier orbital energies were observed for each compound. More recently, extensive MD simulations totalling over 1 millisecond were used in combination with Markov state modelling to identify the accessible conformational states in the BTK kinase domain.<sup>100</sup> Several new potentially druggable sites were identified, and the interconversion between conformational states was found to be modulated by the protonation state of an aspartate residue (Asp539).



**Figure 1.3** (A) Active site of EGFR kinase from crystal structure 5YU9<sup>109</sup> showing the cysteine residue targeted by covalent inhibitors (C797) and the i+3 aspartate residue (D800) that is reported to act as a catalytic base in EGFR. (B) Active site of BTK from crystal structure 5P9J showing the C481 cysteine residue targeted by covalent inhibitors and the i+3 asparagine residue (N484).

The molecular mechanism of BTK inhibition by ibrutinib and related electrophile containing drugs is currently unknown. Modelling studies of covalent modification by an acrylamide inhibitor have been carried out for EGFR, which identified an active site aspartate residue (Asp800) was crucial for the formation of the covalently bound EGFR-drug complex as it acts as a catalytic base to deprotonate the cysteine thiol prior nucleophilic attack.<sup>74</sup> However, there is no equivalent aspartate in the analogous i+3 position in BTK,<sup>110</sup> which instead contains an asparagine residue (Figure 1.3), which raises questions about the precise mechanism of inhibition in BTK. At the time of writing, a preprint was published that investigated the covalent binding of a reversible cyanoacrylamide inhibitor to BTK.<sup>86</sup> However, the study did not address the question of how the C481 thiol is deprotonated and only modelled a single ionic mechanism for covalent modification. Knowledge of the mechanistic pathway and reactivity determinants in this important drug target would aid in the design of future covalent inhibitors of BTK and could even shed light on resistance mechanisms.

## 1.4 Aims

The main aims of this work are twofold:

Firstly, to establish appropriate computational methods for understanding and predicting covalent reactivity. An important aspect of this is to investigate which quantum mechanical methods are capable of accurately modelling sulfur containing compounds by testing a range of semi empirical and density functional methods and comparing these to high level QM methods. In addition, the most suitable techniques for modelling different aspects of covalent reactivity will be assessed. These include

approaches for modelling covalent warhead reactivity using ligand-based methodologies and assessing cysteine reactivity in a protein environment.

Secondly, to establish the mechanisms of covalent modifications of cysteine residues in protein kinases so that meaningful reactivity predictions can be made for a series of covalently binding inhibitors. In order to achieve this goal, enhanced sampling methods will be used in conjunction with QM/MM methods to elucidate the mechanism of covalent binding of the covalent drug ibrutinib to its protein target BTK. Once the mechanism is known for this important drug target, the rate determining step can be modelled with the most appropriate QM methods and reactivity predictions can be made that will allow those in the drug design community to tune covalent reactivity.

## Chapter 2

# Limitations of ligand-only approaches for predicting the reactivity of covalent inhibitors

The work presented in this chapter is a publication entitled “Limitations of Ligand-Only Approaches for predicting the Reactivity of Covalent Inhibitors”, and was published in The Journal of Chemical Information and Modeling on 9<sup>th</sup> September 2019.<sup>62</sup> All calculations were set up and run by myself, with the exception of the  $\Delta E_{\text{react}}$  values (Figure 2.7) that were set up by myself, but run by Gary Tresadern of Janssen Pharmaceuticals due to computer resource limitations in Bristol at the time. I am the first author and performed the data analysis and wrote the manuscript with assistance from Gary Tresadern, and the other authors approved the manuscript before submission. The additional information contained in the supporting information can be found in Appendix 1.

### 2.1 Abstract

Covalent inhibition has undergone a resurgence and is an important modern-day drug design and chemical biology approach. To avoid off-target interactions, and to fine tune reactivity, the ability to accurately predict reactivity is vitally important for the design and development of safer and more effective covalent drugs. Several ligand-only metrics have been proposed that promise quick and simple ways of determining covalent reactivity. In particular, we examine proton affinity and reaction energies calculated with the density functional B3LYP-D3/6-311+G(d,p)//B3LYP-D3/6-31G(d) method to assess the reactivity of a series of  $\alpha,\beta$ -unsaturated carbonyl compounds that form covalent adducts with cysteine. We demonstrate that, whilst these metrics correlate well with experiment for a diverse range of covalent fragments,

these approaches fail for predicting the reactivity of drug-like compounds. We conclude that ligand-only metrics such as proton affinity and reaction energies do not capture determinants of reactivity in situ and fail to account for important factors such as conformation, solvation and intramolecular interactions.

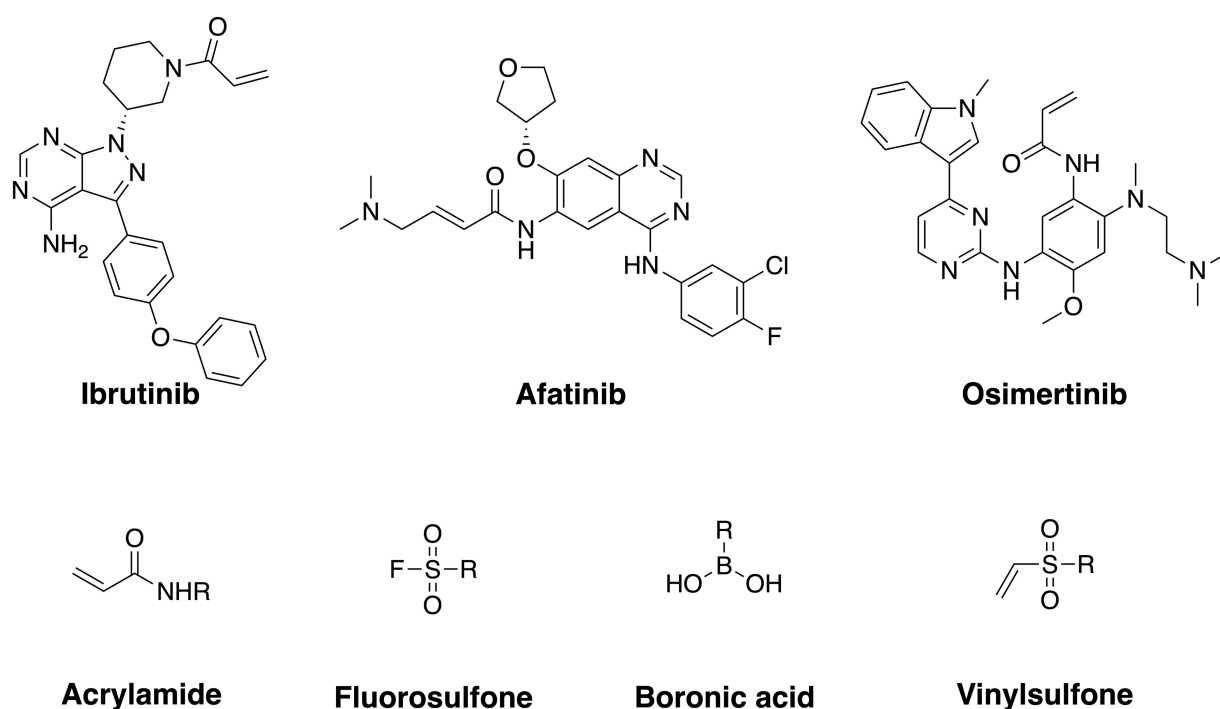
## 2.2 Introduction

Predicting covalent reactivity is an important goal in targeted covalent inhibitor (TCI) design. In contrast to conventional reversible inhibitors, TCIs form a covalent attachment with their target, resulting in several advantageous properties. These include increased potency, selectivity, and residence time all leading to distinct pharmacodynamic properties.<sup>1</sup> The half-life of the covalent complex (which may be long, compared with the clearance of the free TCI and synthesis rate of the target protein) can offer pharmacokinetic advantages. Efficacy can be extended beyond what would be expected due to the half-life of the free drug alone in plasma. There are now several recently marketed covalent kinase inhibitors, for example (Figure 2.1).<sup>111</sup> However, toxicity concerns are an important consideration when pursuing a covalent mechanism of action in drug design, owing to the intrinsic reactivity of TCIs.<sup>24</sup> In addition, clearance and cross-reactivity/selectivity can be problematic.<sup>4</sup> A typical approach is the addition of a reactive covalent warhead onto a potent reversible inhibitor of the target, in the expectation that selectivity will be improved by the optimized fit of the reversible molecule for the particular protein target.<sup>112</sup> Covalent warheads can target various amino acids including cysteine, serine, tyrosine and lysine.<sup>113</sup> Amongst these, acrylamides are often used to target cysteine residues; acrylamide reactivity can be modulated via the amine and  $\beta$ -carbon substituents of the motif.<sup>9</sup>

As the field of TCIs has blossomed over recent years, some model/simplified experimental and computational approaches have been proposed to assist in predicting optimal reactivity. Experimental approaches have been used to estimate the general reactivity of covalent fragments with sulfur-containing species representing cysteine,



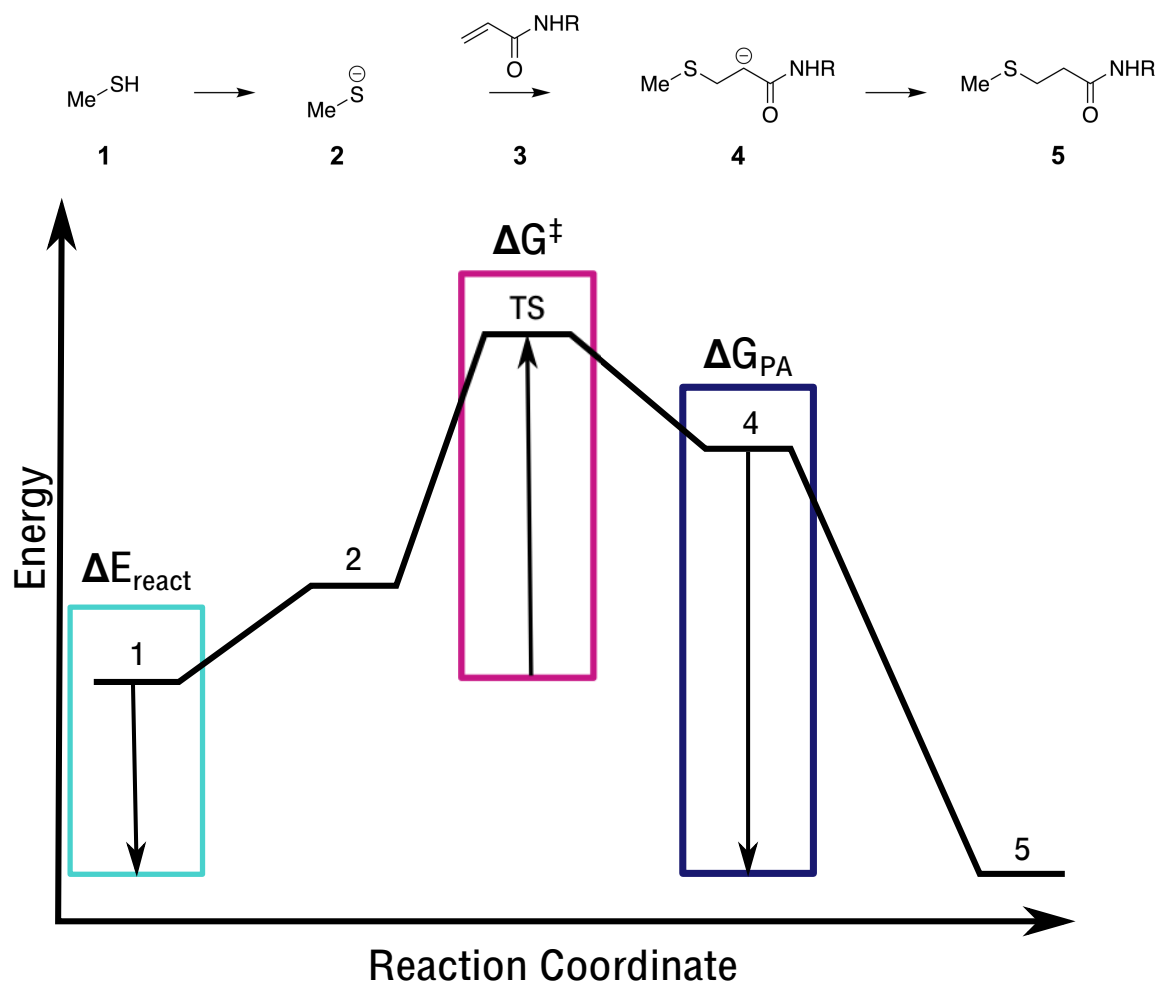
typically glutathione (GSH).<sup>114</sup> This is often reported as GSH  $t_{1/2}$ , the half-life of the reaction of a covalent compound with glutathione. The GSH trapping assessment is readily accessible within drug discovery due to its legacy as a model for reactive toxicity, particularly of metabolites.<sup>115</sup> Therefore, such data is relatively easy to generate compared to protein binding reactivity or kinetics and serves as a convenient metric for comparing warhead reactivity with cysteine in TCI design.<sup>55</sup>



**Figure 2.1** Recently approved acrylamide containing covalent kinase inhibitors<sup>111</sup> and alternative covalent warheads. Ibrutinib targets BTK and is used as a treatment for multiple B cell cancers. Afatinib and osimertinib both target EGFR and are treatments for non-small cell lung carcinoma.

With regard to computational chemistry methods, modelling the reaction in the protein-ligand complex with QM/MM methods has shown success<sup>74,83,116–118</sup> but remains complex and relatively time-consuming for drug discovery. Instead, one approach has been to reduce the covalent reaction between acrylamides and their protein target to a simple ligand-only reactivity metric. Ligand-only methods have shown some success in predicting reactivity trends for small covalent fragments.<sup>61,93,119</sup> Examples of these methods include using calculated proton affinity (PA), reaction

energies ( $\Delta E_{\text{react}}$ ), kinetic barrier heights ( $\Delta G^\ddagger$ ), electrophilicity index and NMR chemical shifts to predict covalent reactivity. Recent work from Houk *et al.* elaborated the full reaction mechanism for a small set of  $\alpha,\beta$ -unsaturated carbonyl compounds illustrating that both kinetics and thermodynamics are important determinants of covalent reactivity of Michael acceptors with methyl thiolate.<sup>120</sup> Complementary to this, Lonsdale *et al* used density functional theory (DFT) to calculate reaction energies of a diverse set of covalent fragments. Whilst they set out calculating some QM transition state (TS) barriers, most of their work focused on calculating the energy difference between reactants and adduct equivalent to species 1 and 5 ( $\Delta E_{\text{react}}$  in Figure 2.2). They found these to correlate well with GSH reactivity ( $R^2=0.69$ ).<sup>119</sup> Furthermore, Flanagan *et al* found that calculated reaction barrier heights ( $\Delta G^\ddagger$ ) correlate strongly with GSH reactivity ( $R^2=0.92$ ).<sup>56</sup>  $\Delta G^\ddagger$  is the energy difference between species 1 and the transition state in Figure 2.2, and reflects the overall reactivity rate (GSH  $t_{1/2}$ ) of the reaction. PA provides an estimation of the free energy difference between the covalent-thiol adduct and its corresponding conjugate base (4 and 5 respectively,  $\Delta G_{\text{PA}}$  in Figure 2.2). Krishnan *et al* found that PA values correlate well with  $\beta$ -elimination rates of covalent fragments from sulfur containing species, which vary based on the acidity of the  $\alpha$ -carbon adjacent to the carbonyl group of the acrylamide.<sup>61</sup> More recently, the electrophilicity index, derived from QM calculations has been shown to correlate well with experimental data. However, the authors note the difficulty in reactivity prediction using this method for larger compounds and non-terminal acrylamides.<sup>63</sup>



**Figure 2.2** A schematic view of thiol addition to a generic  $\alpha,\beta$ -unsaturated carbonyl compound. The energy profile highlights the chemical species commonly used in ligand-only reactivity metrics: reaction energy ( $\Delta E_{\text{react}}$ ), reaction barrier height ( $\Delta G^\ddagger$ ) and proton affinity ( $\Delta G_{\text{PA}}$ ).

Proton affinities and reaction energies ( $\Delta E_{\text{react}}$ ) are particularly attractive reactivity metrics, as their computational prediction is relatively straightforward from quantum mechanical (QM) calculations.<sup>121</sup> PA and  $\Delta E_{\text{react}}$  are reactivity metrics that provide insight into the thermodynamic stability of the reaction. Although the TS is not explicitly used, PA and  $\Delta E_{\text{react}}$  calculations give a good approximation to overall experimental reactivity as the enolate intermediate 4 that results from thio-Michael addition between cysteine and an acrylamide (Figure 2.2) is close in energy and geometry to the TS.<sup>119</sup> When considering an appropriate computational protocol to predict reactivity for drug molecules, the ability to consistently and reliably generate

results for various input molecules is crucial because such an approach needs to be applicable to libraries of hundreds of analogues. The automated QM calculation of transition states presents a challenge because of the difficulty to reach geometrical convergence for saddle point structures whereas calculating PA and  $\Delta E_{\text{react}}$  is more straightforward. However, new tools for TS searches are emerging and have been applied to thio-Michael reactivity, although the authors note the difficulty due to the typically flat nature of the potential energy surface (PES) around the TS.<sup>122</sup> In addition, it is difficult to accurately represent solvation changes with continuum solvation models, particularly for reactions involving charged intermediates, although good results can be achieved with explicit treatment of solvent.<sup>123</sup>

Our aim here is to investigate if two common ligand-only reactivity metrics, PA and  $\Delta E_{\text{react}}$ , can accurately and reliably be used to predict biologically relevant covalent reactivity. We use a large number of compounds ranging from small covalent fragments, to large drug-like molecules to demonstrate how the simple ligand-only reactivity trends based on PA and  $\Delta E_{\text{react}}$  compare favourably with GSH reactivity for similar small fragments but break down for larger drug like molecules.

## 2.3 Methods

A previous study used QM calculations to calculate proton affinities with the B3LYP functional and the 6-311+G(d) basis set.<sup>61</sup> Here we use a modified protocol, appropriate for proton affinity calculations of larger compounds. We performed gas phase geometry optimizations at the B3LYP-D3/6-31G(d) level of theory, followed by single point energies in solution using B3LYP-D3/6-311+G(d,p) level of theory. This has been shown to be sufficient for the calculation of proton affinity values for simple organic molecules of this type.<sup>61,124,125</sup> The inclusion of dispersion corrections to B3LYP has been recommended in QM calculations<sup>80</sup>, particularly when modelling cysteine reactivity.<sup>126</sup> Although B3LYP has known limitations in modelling reactivity<sup>46,77,127,128</sup>, it provides an acceptable balance of accuracy and speed in this context for predicting

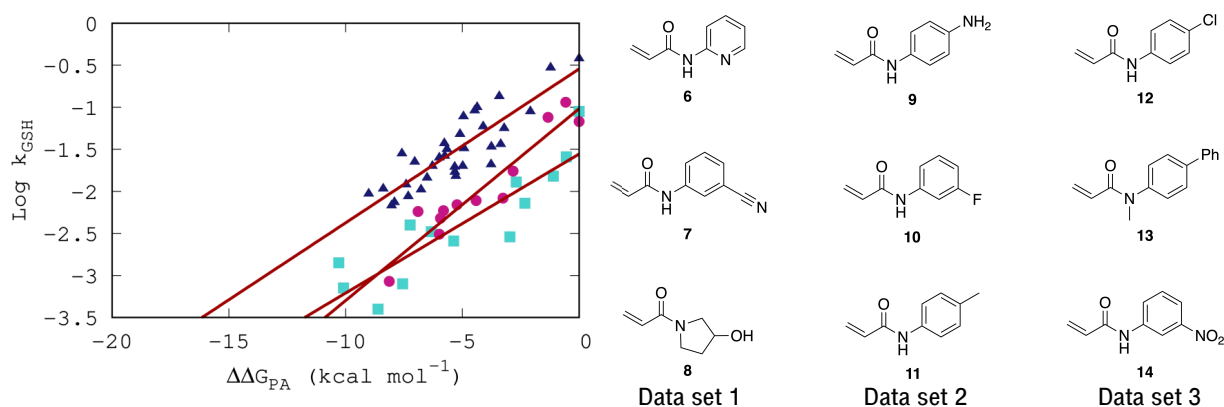
relative reactivities. We have also tested the M06-2X functional, but found it made no appreciable difference to the results (Appendix 1, A1.4). We included solvation by using the Poisson-Boltzmann Finite (PBF) element implicit water solvation model using a probe radius of 1.4 Å,<sup>129,130</sup> which has been shown to reproduce experimental solvation energies for anionic species with good accuracy.<sup>131</sup> All QM calculations were performed with Jaguar v8.5-13.<sup>132</sup>

For larger, drug-like molecules, conformational sampling becomes important. An assessment of the variation of PA with respect to the conformation of each molecule was performed. Six representative molecules were chosen. A conformational search was performed on each using the Macromodel tool in Maestro v2018-3.<sup>133</sup> The mixed torsional/low-mode sampling approach was used with default settings (maximum number of steps of 1000, 100 steps per rotatable bond and minimum and maximum distances for a low mode move of 3.0 and 6.0 Å respectively), using water as a solvent. An energy window of 5 kcal/mol was applied to discard higher energy conformations. This search resulted in over 2100 conformations for the 6 molecules combined. Two distinct conformations per molecule were extracted: the lowest energy conformation; and the next lowest energy with a root mean square deviation (RMSD) > 2.0 Å with respect to the first. This RMSD cut-off is typically used to distinguish structurally different conformations of drug-like molecules. These conformations were submitted to QM minimization at the B3LYP-D3/6-311+G(d,p) level of theory again using the PBF solvation model. All outputs were submitted to frequency calculations to confirm stationary points as minima, and to obtain zero-point energy for thermal and entropic corrections. These two minima were then used as inputs for the PA calculations as described above. PAs were calculated according to Figure 2.2. This involved calculating the difference in energy of the two states, 4 and 5 to obtain  $\Delta G_{PA}$ . These  $\Delta G_{PA}$  values were then normalized relative to the most acidic compound within each data set, providing what we refer to as  $\Delta \Delta G_{PA}$  (kcal mol<sup>-1</sup>). Reaction energy ( $\Delta E_{react}$ ) values were calculated by taking the difference in energy between states **1** and **5**. Since we

already had optimized structures for state **5**, state **1** was generated simply by breaking the S-C bond to minimize towards the local reactant energy minimum. Geometry optimization was performed at the B3LYP-D3/6-31G(d) level in the gas phase, followed by single point energy calculations in aqueous solution at the B3LYP-D3/6-311+G(d,p) level of theory.

## 2.4 Results

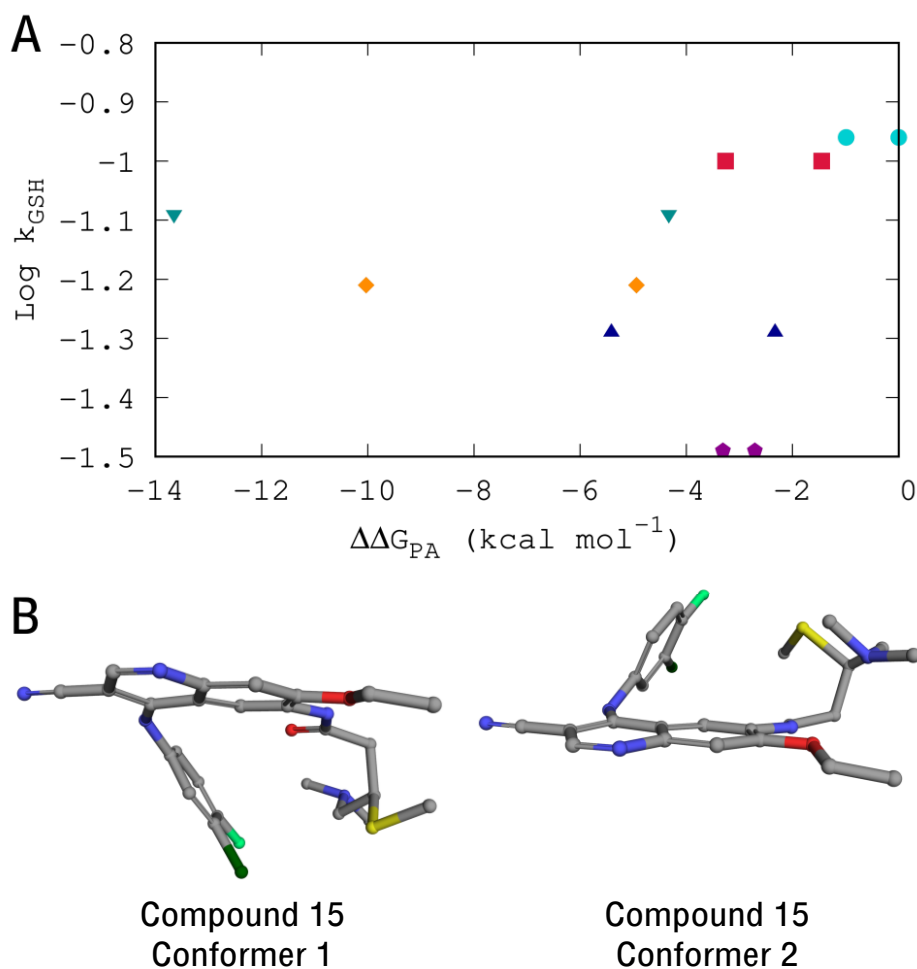
An online search was performed to identify data sets that contained small molecule compounds bearing acrylamide functionality and associated GSH reactivity data. The fragments identified had a molecular weight range of 133 to 271 g mol<sup>-1</sup>, and between 1 and 6 rotatable bonds. Using the method outlined above, we calculated proton affinity values for three data sets containing a total of 37 unique compounds. For each data set, a plot of the relative proton affinity ( $\Delta\Delta G_{PA}$ ) against the GSH reactivity rate (log k) reveals a strong correlation: data set 1:  $R^2=0.79$ ; data set 2:  $R^2=0.90$ ; and data set 3:  $R^2=0.75$  (Figure 2.3). A more negative  $\Delta\Delta G_{PA}$  value corresponds to a more acidic compound and as expected these acidic compounds also have a faster GSH reactivity rate.



**Figure 2.3** Plot of GSH reactivity vs relative proton affinity ( $\Delta\Delta G_{PA}$ ) for data set 1,<sup>56</sup>  $R^2=0.79$  turquoise squares, data set 2,<sup>93</sup>  $R^2=0.90$  magenta circles, data set 3,<sup>134</sup>  $R^2=0.75$  blue triangles. Representative structures from each data set are shown. There is a positive correlation observed between  $\Delta\Delta G_{PA}$  and log  $k_{GSH}$  for each data set. More acidic compounds (represented by a more positive  $\Delta\Delta G_{PA}$  value) have faster experimental reactivity rates.

To test our approach for larger, more conformationally complex compounds, we performed a literature search for data sets that contained ‘drug-like’ compounds, bearing structural similarity to already approved covalent drugs on the market that also had published GSH reactivity data. The compounds had higher MW and increased flexibility, MW in the range 352 to 659 g mol<sup>-1</sup>, and number of rotatable bonds from 6 to 13. We retrieved three data sets that contained 6, 11, and 9 compounds respectively (Table A1, Appendix 1). The first data set was based on a combination of anilinoquinoline and anilinoquinazoline EGFR inhibitors<sup>93</sup>, the second on a series of bis-anilino-pyrimidine EGFR inhibitors<sup>93</sup> and the third were based on the chemical scaffold of pyrazolopyrimidine BTK inhibitors containing an inverted cyanoacrylamide warhead.<sup>22</sup>

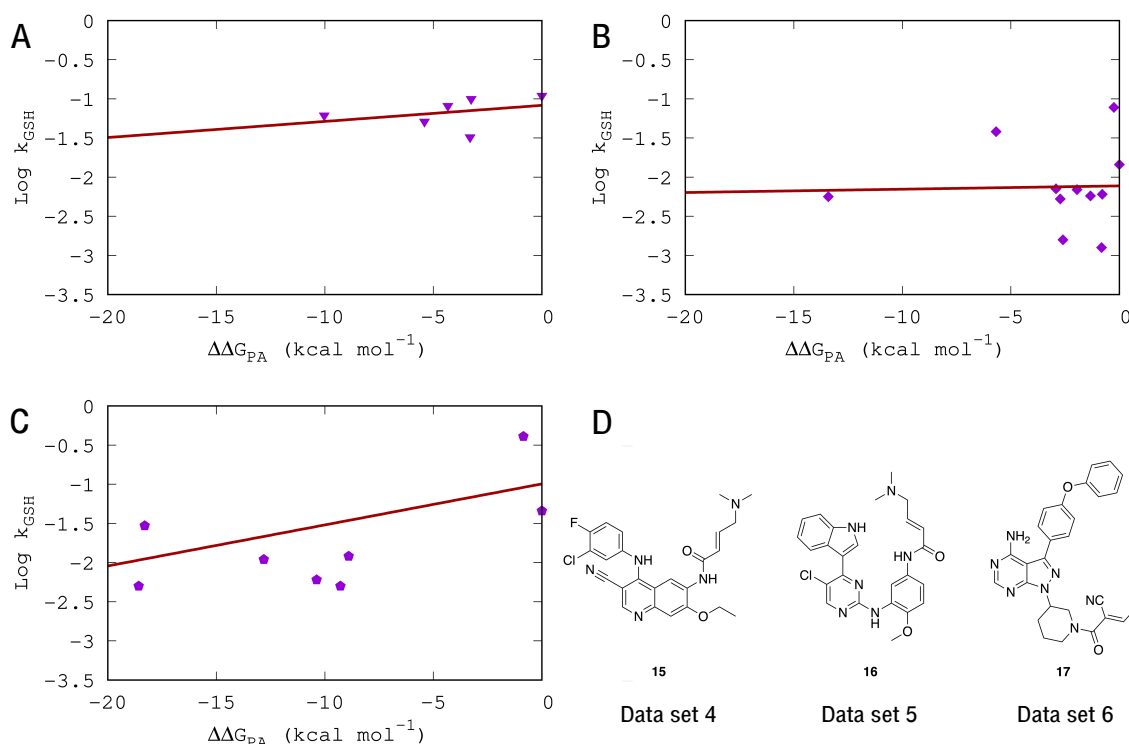
For these larger molecules, conformational sampling could be important. Therefore, we performed a test for the conformational dependence of PA. We took the 6 compounds from data set 4, and identified the lowest energy conformation, and also the next lowest energy but structurally different conformation ( $>2$  Å RMSD to the first, see methods) for each molecule. These conformations were then used to calculate the PA. The different conformations show significantly different PAs. Figure 2.4 shows there are large variations between the  $\Delta\Delta G_{PA}$  calculated for the two alternative low energy conformations of each molecule. There are some examples such as compound **15** (Figure 2.4) where the PA for the two different conformations varies by as much as 9.3 kcal mol<sup>-1</sup>, whereas others are closer. In addition, we calculated a Boltzmann weighted  $\Delta\Delta G_{PA}$  value for all compounds in data set 4, averaged over 4 low energy conformers per molecule (Appendix 1, A1.2). However, we find that there is no significant improvement in correlation between PA and experimental reactivity when using this approach (Figure A1, Appendix 1).



**Figure 2.4** (A) Plot of  $\Delta\Delta G_{PA}$  vs  $\log(k_{GSH})$  for the two conformers of each compound from data set 4, each pair of conformers for the same molecule is represented by different shape labels. (B) Structures of the lowest energy conformer and alternative low energy conformer with  $\text{RMSD} > 2 \text{ \AA}$  for compound 15. Hydrogen atoms are omitted for clarity. These conformers exhibit a large variation in calculated  $\Delta\Delta G_{PA}$ , caused by a rotation of the acrylamide side chain.

The strong conformational dependence of the PA suggested that it would not be an ideal metric for large compounds. We investigated this by performing PA calculations using the method outlined above for all the compounds in these new data sets, but only using the lowest energy conformation. This involved a total of 25 unique drug-like compounds across the three data sets. For each data set, the plot of the relative proton affinity ( $\Delta\Delta G_{PA}$ ) against the GSH reactivity rate ( $\log k$ ) revealed no correlation: data set 1:  $R^2=0.12$ ; data set 2:  $R^2=0.00$ ; and data set 3:  $R^2=0.14$  (Figure 2.5).





**Figure 2.5** Plots of GSH reactivity vs relative proton affinity ( $\Delta\Delta G_{PA}$ ) for (A) data set 4,<sup>93</sup>  $R^2=0.12$ , (B) data set 5,<sup>93</sup>  $R^2=0.00$  (C) data set 6,<sup>22</sup>  $R^2=0.14$ . (D) Representative structures from each data set. There is no correlation observed between  $\Delta\Delta G_{PA}$  and  $\log k_{GSH}$  for each data set. More acidic compounds (represented by a more positive  $\Delta\Delta G_{PA}$  value) would be expected to have faster experimental reactivity rates, but this is not observed by the large molecules in these sets.

As shown in Figure 2.3, the calculated PA for small fragments appears to be predictive for reactivity with GSH. Therefore, we examined it as a surrogate for predicting the reactivity of the larger molecules. We chose all the molecules from data set 4 and converted these to small fragments, so that only the core scaffold from the parent structure remained but each fragment was unique (Figure 2.6). These fragments are analogous to the types of compounds we studied in data sets 1-3. This approach is similar to the truncation algorithm used by Palazzesi *et al*,<sup>63</sup> but we had to modify the approach slightly to avoid duplication of fragments. The calculated PAs for the lowest energy conformation of the full molecules and the fragments are shown in Table 2.1. The results suggest that the fragment calculations are less predictive for the reactivity of larger molecules and do not correlate with experiment. For instance, the most reactive molecule identified from experiment (**F22**) was predicted to be the least

reactive based on the fragment  $\Delta\Delta G_{\text{PA}}$  calculation. Thus, even for the simpler case of GSH compared with protein reactivity, additional parameters beyond the local atomic and electronic environment are important.

**Table 2.1** Calculated  $\Delta\Delta G_{\text{PA}}$  values at the B3LYP-D3/6-311+G(d,p)// B3LYP-D3/6-31G(d) level for fragments and the corresponding parent compound, and experimental reactivity ranking. The predicted relative rank order from calculations is shown in parentheses. <sup>a</sup>Experimental ranking is based on most reactive being assigned ranking of 1, and least reactive a ranking of 6.

Fragment	Fragment $\Delta\Delta G_{\text{PA}}$	‘Parent’ $\Delta\Delta G_{\text{PA}}$	Experimental ranking <sup>a</sup>
<b>F15</b>	−11.6 (4)	−4.3 (3)	4
<b>F18</b>	−7.0 (5)	0.0 (6)	6
<b>F19</b>	−15.2 (1)	−5.4 (2)	2
<b>F20</b>	−13.2 (3)	−3.3 (5)	5
<b>F21</b>	−13.3 (2)	−10.0 (1)	3
<b>F22</b>	0.0 (6)	−3.3 (4)	1

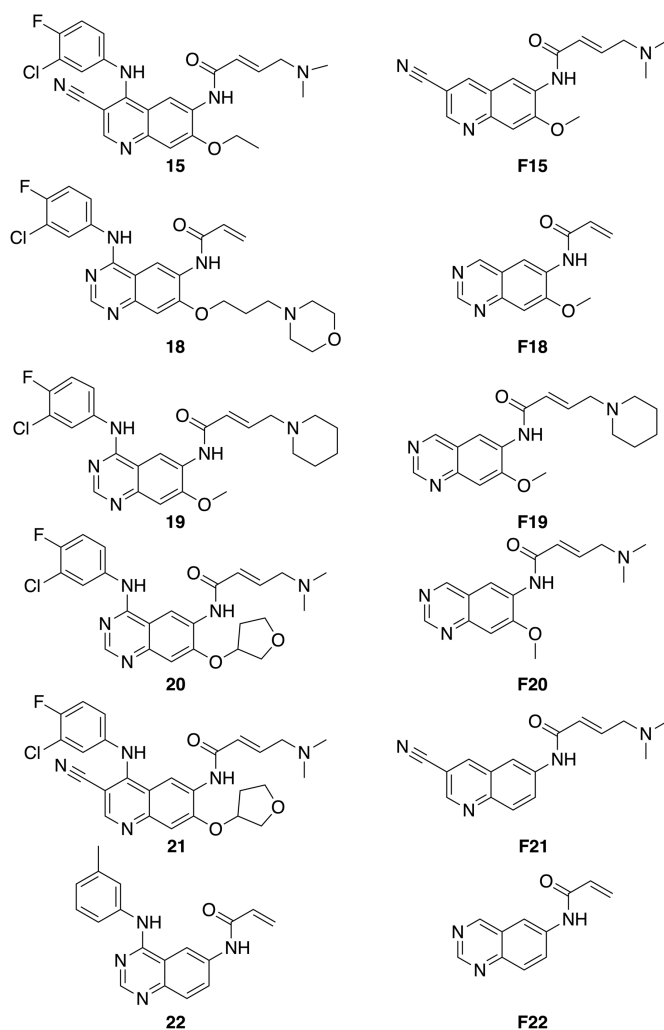
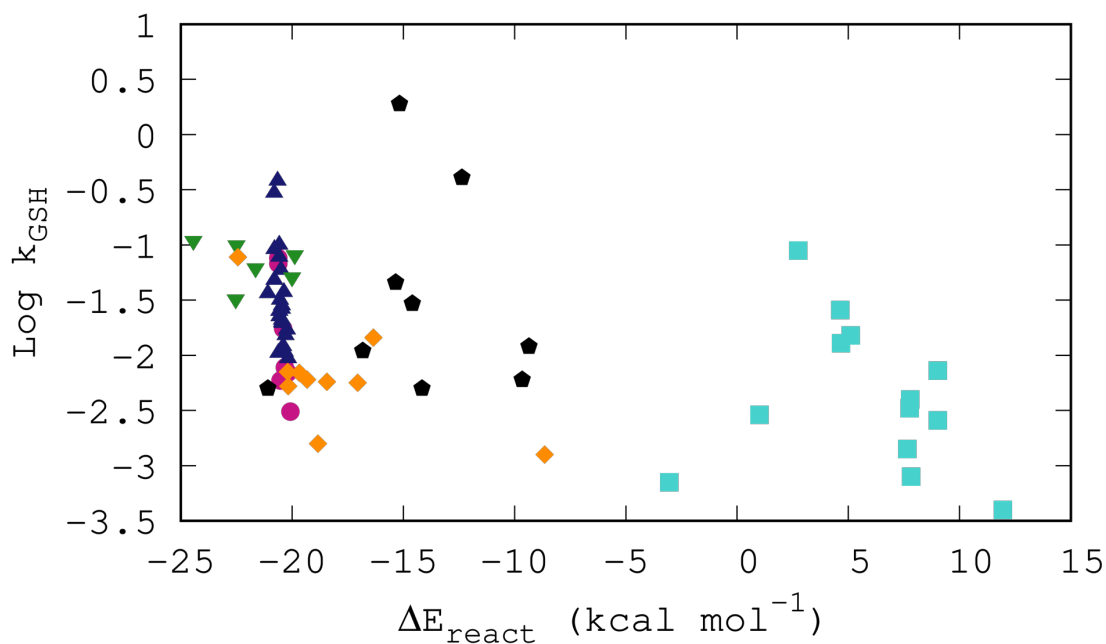


Figure 2.6 Fragments of drug-like compounds from data set 4 used here to calculate fragment  $\Delta\Delta G_{PA}$  values.

We also investigated how calculated  $\Delta E_{\text{react}}$  values, a commonly used alternative computational reactivity metric, correlate with the experimental reactivity.  $\Delta E_{\text{react}}$  values were calculated for all 68 compounds from data sets 1-6 (see methods). Generally, a poor correlation between  $\Delta E_{\text{react}}$  and experimental GSH reactivity was observed (Figure 2.7). Across all the data sets,  $\Delta\Delta G_{PA}$  generally performed better as a reactivity metric than calculated  $\Delta E_{\text{react}}$  values, although neither did as well as would be needed for a drug discovery lead optimization setting. In short, four of the six data sets of close analogues showed no appreciable correlation ( $R^2 < 0.5$ ).



**Figure 2.7** Plot of GSH reactivity vs reaction energy,  $\Delta E_{\text{react}}$  for data set 1,  $R^2=0.05$  turquoise squares; data set 2,  $R^2=0.68$  magenta circles; data set 3,  $R^2=0.33$  blue triangles; data set 4,  $R^2=0.09$  green triangles; data set 5,  $R^2=0.50$  orange diamonds; data set 6,  $R^2=0.00$  black pentagons.

## 2.5 Discussion

The calculation of PA values for small covalent fragments bearing  $\alpha,\beta$ -unsaturated functionality initially appears to be a promising covalent reactivity metric. For compounds of this type, we observe that increased acidity at the  $\alpha$ -carbon of the acrylamide group correlates with faster GSH reactivity rates, as expected. This reflects the stability of the enolate intermediate (compound 4 in Figure 2.2) in the covalent thio-Michael reaction mechanism, which leads to faster reaction. The resulting plots show that  $\Delta\Delta G_{\text{PA}}$  values correlate strongly with experimental GSH reactivity for small fragment compounds. The results are similar to previously published work<sup>61</sup> that found proton affinity to correlate with  $\beta$ -elimination rates with  $R^2=0.96$ .

For the larger and more complex compounds, no correlation is observed between  $\Delta\Delta G_{\text{PA}}$  values and GSH reactivity. To understand this result, we investigated the effect

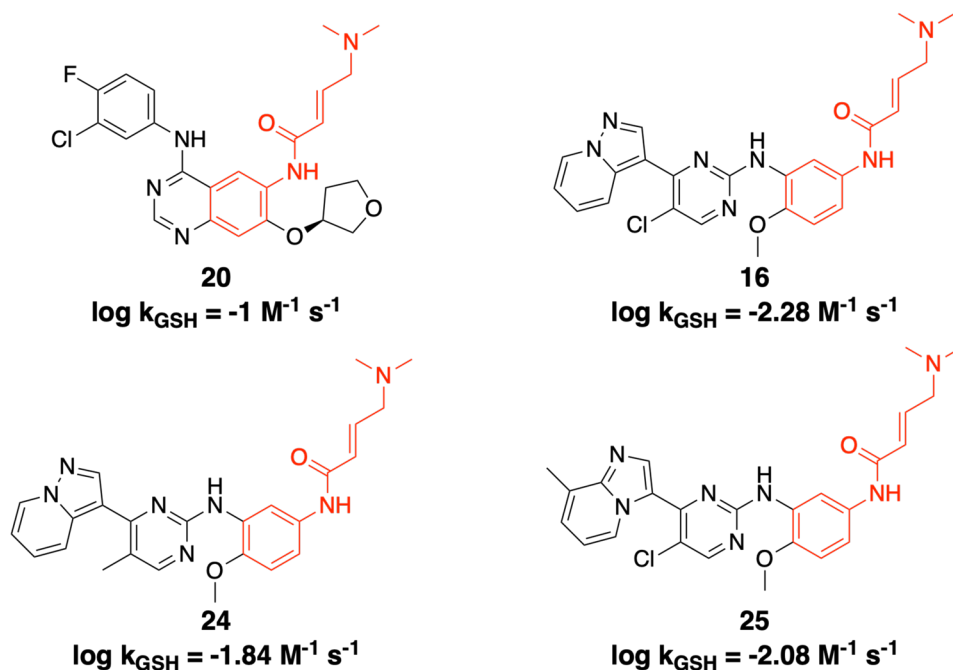
of conformational differences in our protocol by examining in detail the 6 compounds in data set 4. Two sources of potential error were identified, the first being that different input conformations significantly impact the calculated PA for each compound. The difference in PA between conformer pairs obtained from the conformational search (see Methods) can be as large as 9.3 kcal mol<sup>-1</sup> (compound **15**, data set 4, Figure 2.4A and 2.5). In this particular case, there was a rotation of the substituted acrylamide-thiol adduct leading to an additional NH-N interaction in one of the conformers that was not present in the other (Figure 2.4B). The second source of possible error resulted from the difference in conformation between the neutral and ionized species. After a geometry optimization, conformational changes in the ionized species were often observed, arising due to the change in hybridization (sp<sup>3</sup> to sp<sup>2</sup>) of the alpha-carbon upon deprotonation causing a change in intramolecular interactions. Rotations of flexible groups attached to aromatic rings and movements of the acrylamide-thiol adduct were also observed, suggesting the ionized species were optimized to alternate local minima from the neutral species. We used more stringent convergence criteria during our optimizations in order to address this, however, it did not significantly change the results (Figure A2, Appendix 1). These sources of error result in an overestimation of the calculated PA, suggesting an inaccurate comparison of the relative change in energy associated with proton affinity with the inclusion of conflicting conformational effects, and therefore explain the lack of correlation between PA and experimental reactivity.

An additional factor that could lead to errors in the calculation of  $\Delta\Delta G_{PA}$  values is the contribution to the free energy from low energy vibrational modes. These low energy modes are particularly important for the large, flexible compounds in data set 4, 5 and 6, and are difficult to calculate accurately and can therefore lead to errors in the free energy calculation.<sup>135</sup> To investigate the affects that these modes have on our free energy values, we used the Quasi Harmonic Oscillator (QHO) approximation<sup>136</sup> to scale all frequencies under 100 cm<sup>-1</sup> up to 100 cm<sup>-1</sup> (Table A2, Appendix 1). These

calculations were performed using GoodVibes, a Python script from the Paton lab.<sup>137</sup> Whilst some small changes to the  $\Delta G_{\text{PA}}$  values were observed, we conclude that these changes are small and do not significantly improve the correlation between  $\Delta\Delta G_{\text{PA}}$  and experimental GSH reactivity, therefore offering no improvement to the predictive power of ligand-only methods for the larger molecules.

The  $\Delta E_{\text{react}}$  calculations also showed generally poor correlation with GSH reactivity. This is in contrast to previous work by Lonsdale and colleagues who reported a correlation of  $R^2=0.69$ . Their study combined diverse molecules and warheads.<sup>119</sup>  $\Delta E_{\text{react}}$  values may offer reasonable reactivity prediction for compounds that differ greatly in reactivity, but will generally not predict reactivity within a chemical series, as shown by our results. It is also likely that reaction energy calculations are also more susceptible to conformational effects than  $\Delta\Delta G_{\text{PA}}$ , due to the greater variability in geometry optimization of methyl thiolate relative to the reactive warhead.

Although previous studies have reported success in using a fragment based approach to improve correlation between ligand-only reactivity metrics and experimental reactivity,<sup>63</sup> we do not find it useful in the context of PA. Inspection of the experimental data and molecular structures casts further doubt on using a fragment approach to predict the reactivity of larger molecules. For instance, compounds **16** (data set 5)<sup>93</sup> and **20** (data set 4)<sup>93</sup> all contain essentially the same fragment (a dimethylamine substituted phenylacrylamide group), but their experimental reactivity differs by approximately 1.3 log units (Figure 2.8). This variation in reactivity highlights that modifications to compounds distal from the site of reaction can have a significant impact on reactivity (conformational effects will also contribute).



**Figure 2.8** Four compounds that contain the same dimethylamine substituted phenylacrylamide fragment (highlighted in red) but have a wide range of experimental GSH reactivity, covering 1.3 log units.<sup>93</sup>

All other factors being equal, the change in reactivity caused by a small structural modification may be captured by PA calculations. Indeed, there could be scenarios in lead optimization where analogues are so similar that this method delivers results that translate into useful design prioritization. PA and reaction energies seem to be attractive reactivity metrics due to their simplicity. For PA, only geometry optimizations of the covalent thiol adduct and its conjugate base are required, negating the need to control the position of methyl thiolate during transition state searches and calculations of reaction energies. However, we see here that, for many common modifications in the medicinal chemistry data sets, further effects are at play. The calculation of PA, reaction energies and reaction barrier heights all require careful consideration of conformational variations in each compound as it is important to ensure that differences in the energy calculation do not arise from competing conformational effects. Lonsdale *et al.* found that the calculation of reaction energy barriers through transition state optimizations to be particularly challenging for large ‘drug like’ compounds and opted to calculate reaction energies instead.<sup>119</sup>

For biological reactivity, PA calculations require careful consideration that input geometries represent local minima corresponding to the noncovalently or covalently bound conformation, as opposed to global energy minima adopted in solution.<sup>138</sup> Reactivity of covalent fragments with GSH in solution is likely to be less constrained, but still dominated by a small number of conformations. Although conformational searches and low energy conformers were selected for our calculations, it is possible that these low energy conformers are not representative of the conformations that react with GSH in solution. Even Boltzmann weighted PA values, averaged over 4 low energy conformers that would be expected to dominate reactivity did not improve reactivity predictions (Appendix 1, Figure A1). In the protein environment it is easy to envisage that the reaction proceeds from the bound conformation of the ligand placed optimally for reactivity with the cysteine in the protein pocket. Replicating these reactive conformations is difficult to achieve for PA calculations and in part explains the lack of predictive ability of ligand-only approaches and indicates the need for treatment of biological reactivity in situ in biological targets.

## 2.6 Conclusions

In summary, calculated proton affinity values can serve as a useful reactivity metric for small molecule fragment compounds. However, we find proton affinity to be a poor reactivity metric for larger, ‘drug-like’ compounds. Calculated reaction energies are generally worse as a reactivity metric than proton affinities for both fragments and drug-like compounds. Our work shows that ligand-only approaches such as the calculation of reaction energies, barrier heights and proton affinity values are insufficient for reliable covalent reactivity prediction. These methods oversimplify covalent determinants of reactivity, and do not account for factors that will affect the reactivity of a covalent drug ‘in-situ’ (e.g. effects of the environment within a protein target). Calculated PA and reactivity are conformation dependent; experimental studies of ligand-only reactivity may also be confounded by conformational effects.



More detailed methods are needed to predict reactivity of covalent drugs in a protein environment, which include the (e.g. electrostatic) effects of the environment, the bound conformation and solvation of the covalent inhibitor in the target site; such methods, capable of modelling reactivity within proteins, including combined QM/MM techniques.<sup>139,140</sup>

## Chapter 3

# Investigating cysteine reactivity with classical Molecular Dynamics and Constant pH Molecular Dynamics Simulations

### 3.1 Introduction

Non-catalytic cysteine (Cys) residues in protein kinases are typically targeted by covalent kinase inhibitors owing to potential gains in selectivity.<sup>141</sup> This is because Cys residues are found in relatively low abundance in the proteome compared to the other amino acids and possess a nucleophilic thiol side chain.<sup>142</sup> The thiol side chain in Cys residues can be deprotonated to form a negatively charged thiolate, which is generally regarded to be the more reactive form of Cys.<sup>143</sup> The tendency of the thiol group to lose a proton can be used as a measure of Cys reactivity and is quantified by measuring or predicting the  $pK_a$  of the desired Cys residue. Cys  $pK_a$  and reactivity is determined by numerous factors including steric and electronic factors of the surrounding protein environment, how well solvated the Cys residue is and the stability of the thiolate anion.<sup>144</sup> The  $pK_a$  of a Cys residue free in solution has been measured as 8.6 from potentiometric titration experiments.<sup>145</sup> The  $pK_a$  of non-catalytic Cys residues in proteins can range between 7.4 to 9.1,<sup>146</sup> and some catalytic Cys residues have reported  $pK_a$ s as low as 2.<sup>147</sup> The lower the  $pK_a$ , the more reactive the Cys is because the increase in probability of a deprotonated Cys. The large shifts in Cys  $pK_a$  makes it challenging to accurately characterise, although there are a number of computational and experimental methods available.

Methods to experimentally determine Cys  $pK_a$  are not the focus of this work, but they serve as essential comparisons for any computational  $pK_a$  prediction model. Accurate determination of amino acid side chain  $pK_a$ s in protein environments by

experiment is extremely difficult but can be achieved through the use of spectroscopic methods by measuring reaction rates of the thiol and NMR spectroscopy.<sup>148</sup> Computational  $pK_a$  prediction methods use a thermodynamic cycle to calculate the deprotonation energy of the amino acid side chain in solution, and then use this to evaluate the acid dissociation constant ( $K_a$ ) and subsequently the  $pK_a$ .<sup>149</sup> The simplest methods are based on implicit solvation models and include environmental electrostatic effects from numerical solutions of the Poisson-Boltzmann equation.<sup>149</sup> Examples include the H++<sup>150</sup> and MCCE<sup>151</sup> methods. Arguably the most popular method for the rapid prediction of titratable groups in proteins is the empirical PROPKA model.<sup>152</sup> This uses a number of scoring functions and descriptors to capture the electrostatic environment that are parameterised on a large number of experimental  $pK_a$ s.<sup>153</sup> Although these methods can have a high level of accuracy, they are usually parametrised specifically for aspartate (Asp) and glutamate (Glu) residues. The performance for Cys residues is usually worse, as shown in a study from Awoonor-Williams *et al.*<sup>66</sup> They took a test set of 18 Cys residues from protein kinases where an experimental  $pK_a$  and crystal structure were available. Several  $pK_a$  prediction methods were tested, including PROPKA, H++ and MCCE. These were compared with a replica exchange thermodynamic integration (RETI) approach using the CHARMM36 and AMBER ff99SB-ILDNP force fields. The electrostatic and empirical models (PROPKA, H++ and MCCE) were found to perform poorly in comparison to explicit solvent models with RMSDs of 3.4-4.7 for electrostatic and empirical models compared to 2.4-3.2 for explicit solvent models.<sup>66</sup> Variations in the force field (FF) used was also important, with the CHARMM FF performing better than the AMBER FF. The major limitations of these methods are the use of approximate implicit solvation models to describe the surrounding environment, and that they only provide a  $pK_a$  estimate on static structures. Proteins are highly dynamic, and both solvation and conformation can have a large impact on  $pK_a$  values.

More robust  $pK_a$  prediction methods that account for protein dynamics involve FF methods such as molecular dynamics (MD) simulations. In classical MD simulations, the potential energy (PE) of the system is described by a FF. The general form of the AMBER FF is shown in Equation 3.1.<sup>154</sup>

**Equation 3.1.** PE function in the AMBER FF.<sup>154</sup>

$$V_{total} = \sum_{Bonds} k_r (r - r_{eq})^2 + \sum_{Angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{Dihedrals} \frac{V_n}{2} [1 + \cos(n\theta - \gamma)] + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{A_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

The PE function shown in Equation 3.1 is comprised of all the bond stretching, angle bending, dihedral torsions, non-bonded attractive and repulsive (Lennard-Jones) and electrostatic coulombic interaction terms that are present in the system of interest. To track how the PE evolves with time as atoms move in a classical system, the equations embodied by Newton's second law of motion ( $F=ma$ ) are integrated. The PE is linked to the forces through the relation in Equation 3.2.<sup>155</sup>

**Equation 3.2** Newton's second law of motion as a function of PE (V) and distance (R).<sup>155</sup>

$$F = -\frac{\partial V}{\partial R}$$

To run an MD simulation, the initial position of the atoms of the system are chosen (usually from a protein crystal structure), and initial velocities are assigned randomly from a Boltzmann distribution. The PE is then calculated according to Equation 3.1, and Newton's equations are then used to predict new positions and velocities after a short time interval,  $dT$ . The energy and gradient at the new positions can be evaluated, and these are used to predict new velocities and acceleration vectors.<sup>155</sup> The whole process is then repeated for each MD step. The parameters required by Equation 3.1 are defined in an FF and specified for protein atoms, water

molecules and small organic molecules. The protonation states of all ionisable residues must be assigned during system set up, with a common program used to predict these being PROPKA. The protonation states are then fixed for the duration of the simulation.<sup>156</sup> This is a poor description of the protonation state of certain amino acid side chains, particularly if the  $pK_a$  of the amino acid side chain is close to the pH at which the protonation states were determined. Although classical MD simulations do not provide a means to directly calculate  $pK_a$  values of titratable residues, snapshots from MD simulations can be used as input for empirical implicit solvent models such as PROPKA to estimate how the  $pK_a$  changes over the course of a trajectory.<sup>157</sup>

A more robust method for examining protonation state distributions from MD simulations is to use constant pH molecular dynamics (CpHMD) simulations. Several implementations of the CpHMD method exist, including a continuous protonation state model,<sup>158,159</sup> a discrete protonation state model<sup>160</sup> and a non-equilibrium molecular dynamics/Monte Carlo (ne-MD/MC) model.<sup>161</sup> The discrete protonation state model developed by Mongan *et al.* for the AMBER force field will be discussed briefly as it is the model used in this work. In discrete protonation state CpHMD simulations, the MD simulation is run at a fixed pH value, and the protonation states of titratable residues are allowed to change over the course of the simulation. The transition free energy (FE) of the protonation state change is evaluated in implicit solvent according to a Monte-Carlo (MC) energy criterion (Equation 3.3).<sup>160</sup>

**Equation 3.3** Calculation of the transition FE between protonation states in the discrete protonation state CpHMD method in AMBER.<sup>160</sup>

$$\Delta G = k_B T (pH - pK_{a,ref}) \ln 10 + \Delta G_{elec} - \Delta G_{elec,ref}$$

If the transition FE is calculated below a certain threshold value, the protonation change is allowed, and MD is continued with the new protonation state. MD can be carried out in both implicit<sup>160</sup> and explicit<sup>162</sup> solvent, although the

protonation state change is always carried out in implicit solvent. In explicit solvent CpHMD, MD is propagated as usual, and if a protonation state change is accepted, a solvent relaxation step follows and the MD run is continued.<sup>162</sup> Based on the distribution of protonation states sampled for a given residue, the most likely protonation state can be estimated at the chosen simulation pH. If CpHMD simulations are run at a range of pH values, a titration curve can be produced to give an accurate  $pK_a$  estimate for a particular residue. There are also replica exchange (RE) versions of CpHMD simulations that can be run in implicit or explicit solvent where exchanges are made between different pH replicas. CpHMD coupled with pH replica exchange dynamics (pH-REMD), will hereafter be abbreviated as CpHREMD. These simulations have been reported to improve sampling of the protonation state distribution and aid in convergence of CpHMD simulations.<sup>162</sup> When tested on Hen Egg White Lysosome (HEWL), CpHREMD simulations predicted more accurate  $pK_a$  values for Asp, Glu and histidine (His) residues when compared to implicit solvent models (root-mean-square-error (RMSE) of 0.92 and 1.32 for explicit and implicit solvent respectively). Unfortunately, HEWL does not contain any Cys residues so benchmarks for the performance of CpHMD for prediction of Cys  $pK_a$ s is sparse. The CpHREMD method accurately reproduces the  $pK_a$  of a model Cys compound and a Cys containing pentapeptide.<sup>162</sup>

Due to the resurgence of interest in targeting Cys residues by covalent inhibitors, several studies have assessed the accuracy of CpHMD methods for predicting Cys reactivity. The first study, by Awoonor-Williams *et al.* used two different implementations of CpHMD and compared the results to a RETI approach.<sup>65</sup> The first CpHMD implementation they tested is the discrete protonation state CpHMD method implemented in AMBER that uses the replica exchange technique in explicit solvent.<sup>162</sup> The other implementation is a ne-MD/MC approach and is implemented in NAMD.<sup>163</sup> The Cys  $pK_a$  of EGFR kinase and JAK3 kinase were computed using each implementation. The Cys797 in EGFR was calculated as 13.5 and 11.5 from the

CpHREMD and ne-MD/MC methods respectively. For JAK3 kinase, the Cys909  $pK_a$  was 12.7 and 11.1 from CpHREMD and ne-MD/MC. These results compare well with the RETI results, which found  $pK_a$ s of 13.0 and 11.1 for the Cys residues in EGFR and JAK3 kinase.<sup>65</sup> For JAK3 kinase, the ne-MD/MC method is closest to the TI method, but for EGFR the CpHREMD method is in closest agreement with TI results. Overall, the results are in reasonable agreement given the different implementations of CpHMD and different force fields used by each technique. In spite of the consistency between the CpHMD and RETI methods, the scarcity of experimental Cys  $pK_a$ s in protein kinases makes a detailed assessment of the accuracy of these methods difficult. There is no experimental  $pK_a$  estimate for JAK3 kinase. The experimental EGFR Cys797  $pK_a$  of 5.5 reported by Truong *et al.*<sup>164</sup> significantly deviates from the higher  $pK_a$  estimates from all of the computational models. A study by Liedl *et al.* examined Cys  $pK_a$  in papain, a cysteine protease using CpHMD simulations.<sup>67</sup> This study was unable to reproduce the acidic-shifted Cys25  $pK_a$  in papain of 3.3. Their CpHMD simulations employed various implicit solvent CpHMD methods available within AMBER, including a new continuous CpHMD method implemented by Jana Shen.<sup>165</sup> The continuous CpHMD model assigns a continuous titration coordinate to each titratable residue that describes the titration behaviour of a fictitious particle.<sup>158</sup> The implementation in AMBER uses a slightly modified Born radius for sulfur of 2.0 Å compared with the original value of 1.8 Å and the GB-Neck2<sup>166</sup> implicit solvent model.<sup>167</sup> This method improved the  $pK_a$  prediction of Cys25 in papain with a predicted  $pK_a$  of 9.8.<sup>67</sup> This compares to a  $pK_a$  estimate of 14 by the implicit solvent CpHMD method using the GB-OBC<sup>168</sup> solvent model.<sup>67</sup> However, all methods failed to predict the acidic  $pK_a$  shift for Cys in this cysteine protease, and the authors note that no improvement was observed by using CpHREMD in implicit solvent.<sup>67</sup> There appears to be a tendency of the computational  $pK_a$  prediction methods to predict Cys  $pK_a$ s that are higher than expected. The solvent model used to assess protonation state changes is clearly important, as are the parameters used to describe the sulfur in the thiol group of the Cys residue.

The aim of this chapter is to investigate the reactivity of the Cys residue, C481, in Bruton’s tyrosine kinase (BTK) using classical MD and CpHMD simulations, specifically the explicit solvent CpHREMD method in AMBER.<sup>162</sup> Classical MD simulations will enable the conformational dynamics of C481 in BTK to be examined, along with its interaction with the covalent inhibitor ibrutinib. This will reveal any potential reactive conformations that will be important starting points for Chapter 5, where a detailed reactivity study between C481 and ibrutinib is presented. CpHREMD simulations will give an indication of the likely protonation state of C481 and will help to inform its reactivity. Although there is no experimental  $pK_a$  estimate of C481, the results can be compared with other studies that have used computational techniques to predict  $pK_a$ s of Cys residues in other protein kinases. Any deviations from expected behaviour can be explored, with a particular focus on the parameters that are used to describe sulfur in a Cys thiol and thiolate in the AMBER force field. The surrounding protein environment can have a marked effect on Cys  $pK_a$ , depending on the properties of surrounding amino acid residues.<sup>39</sup> For kinases in particular, the residue that occupies the i+3 position relative to the Cys is particularly important.<sup>110</sup> Most i+3 residues in kinases are either Asp or asparagine (Asn). A more acidic Asp in the i+3 position has the effect of shifting the Cys  $pK_a$  to higher values than its solution value of 8.6.<sup>145</sup> For example, the  $pK_a$  of Cys797 in EGFR with an i+3 Asp was estimated to be 11.1.<sup>65</sup> Conversely, BTK contains an Asn residue in the i+3 position. Due to its neutral side chain  $pK_a$  of C481 is more acidic and calculated to be 10.4 in the same study.<sup>65</sup> This would make the C481 thiol side chain neutral at physiological pH. This is at odds with many studies that assume the starting point of covalent modification of Cys is the thiolate anion. The combination of classical MD and CpHMD simulations will hopefully be able to explain the intrinsic Cys reactivity in BTK.



## 3.2 Methods

Structural coordinates of the covalently bound BTK/ibrutinib complex were taken from PDB 5P9I.<sup>99</sup> Hydrogens atoms were added in situ using the protein preparation wizard in Maestro,<sup>133</sup> where PROPKA3.1<sup>152,153</sup> was used to assign protonation states of titratable amino acid side chains and optimise the hydrogen bonding network. For the non-covalently bound ibrutinib inhibitor, RESP charges were generated using the RedServer,<sup>169</sup> and any missing FF parameters were generated using antechamber, distributed the AMBER2018 package. The Solvate program, developed by H. Grubmuller and V. Groll, was used to solvate the system by creating a solvation shell of 5 Å of TIP3P water around the protein using 8 gaussians. The AmberTools program tleap was then used to create a truncated octahedral box of solvent around the system with a padding of 5 Å and a closeness of 0.75. Na<sup>+</sup> and Cl<sup>-</sup> ions were added to create a salt concentration of 0.1 M.<sup>170</sup>

The system was then subject to minimisation, then heating from 0 to 293.15 K over 75 ps with a weak 5 kcal mol<sup>-1</sup> restraint on backbone CA atoms using Langevin dynamics and a collision frequency of 5 ps<sup>-1</sup>. Equilibration was then performed in the NPT ensemble, using Langevin dynamics and a Monte Carlo barostat to maintain the pressure at 1.01325 bar with a pressure relaxation time of 1 ps. The weak backbone restraints were gradually released during this equilibration phase.

The equilibrated structure was used as an input for classical MD simulations and CpHMD simulations. For the classical MD simulations, 500 ns of production MD in the NPT ensemble was run according to the same pressure and regulation settings as the equilibration phase. The CpHMD simulations were run in the NVT ensemble, and 8 pH replicas were chosen to span the pH range 9.0 – 16.0 in steps of 1.0 pH unit. For each pH replica, 100 ns of pH-REMD were run, resulting in a total simulation time of 800 ns. Protonation state changes were attempted every 200 fs, and 100 steps of solvent relaxation dynamics was performed after each successful protonation state change. The

cphstats program distributed in AMBER 18 was used to analyse the simulations, including determination of protonation state statistics for C481.<sup>170</sup> The  $pK_a$  of C481 was calculated by fitting CpHMD simulation data to the Hill equation (Equation 3.4).

**Equation 3.4** The Hill equation used to calculate  $pK_a$  from CpHMD simulation data.

$$f_d = \frac{1}{1 + 10^{n(pK_a - pH)}}$$

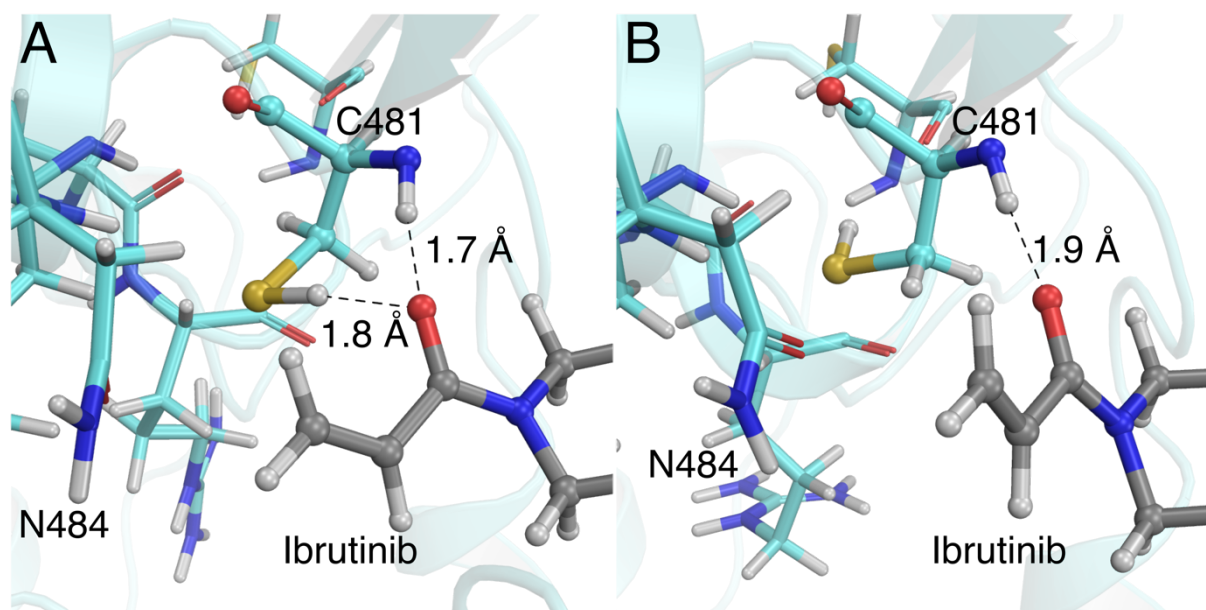
From Equation 3.4,  $f_d$  corresponds to the deprotonated fraction of the residue of interest and  $n$  is the hill coefficient. When the ionisable residue of interest does not interact with other ionisable residues,  $n = 1$  and the titration curve is described by the Henderson-Hasselbalch equation.<sup>171</sup> However, for interacting residues  $n \neq 1$  and this can indicate cooperativity or poor sampling in CpHMD simulations.<sup>162</sup>

## 3.3 Results and discussion

### 3.3.1 Classical MD simulations

Visual inspection of the 500ns classical molecular dynamics simulation of BTK with ibrutinib non-covalently bound in the ATP binding pocket shows that C481 predominately adopts a conformation where it does not interact with the acrylamide group of ibrutinib (Figure 3.1(B)). Very occasionally however, the Cys residue undergoes a small conformational change and does form a weak interaction with ibrutinib (Figure 3.1(A)). The difference in both of these conformations is best characterised by a dihedral angle, defined by the N, CA, CB and S atoms of the Cys residue. Throughout the simulation, the dihedral angle is centred around a value of approximately  $175^\circ$  (Figure 3.3A) when C481 is oriented away from ibrutinib. Although it is not immediately obvious from Figure 3.3A, the C481 thiol group does change orientation and interact with the inhibitor, causing a dihedral angle shift to approximately  $60^\circ$  (Figure 3.1B). However, this conformation is rarely sampled in the

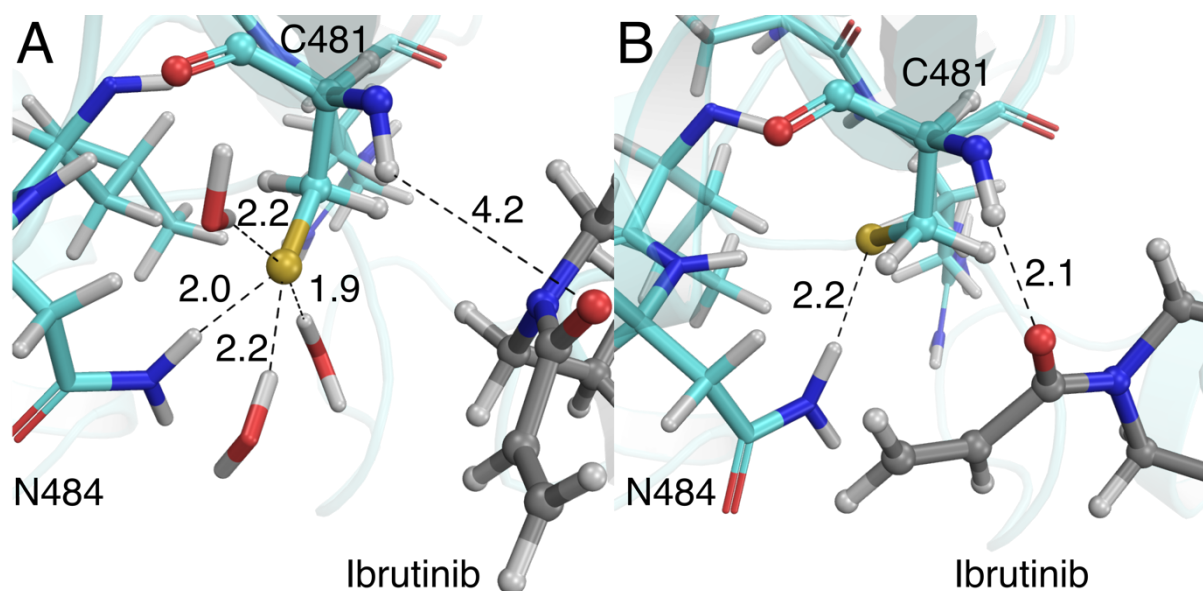
MD simulation and is observed only 0.03% of the time so is not visible in the histogram plot in Figure 3.5A, which shows the distribution of this dihedral angle.



**Figure 3.1** Representative structures of the BTK active site from classical MD simulations showing a reactive C481 thiol conformation (A) and an unreactive C481 thiol conformation (B).

When C481 is oriented away from the inhibitor, the distances between the sulfur atom of the Cys thiol and the  $\beta$ -carbon in the acrylamide group of ibrutinib are quite large. The histogram in Figure 3.5B shows that the S-C distance is centred around a value of approximately 6.5 Å throughout the MD trajectory. Conversely, in the few frames where C481 rotates and points towards the inhibitor, the S-C distance reduces to approximately 4 Å. The variation in S-C distance is largely dominated by the orientation of C481, and not the flexibility of ibrutinib, as the acrylamide warhead is maintained in position by a hydrogen bond between the acrylamide carbonyl oxygen atom and the backbone N-H of C481 (Figure 3.1). The S-C distance is important for covalent modification as it defines the reaction coordinate for the S-C bond formation step in a thio-Michael addition reaction. The large S-C distances and orientation of C481 observed throughout the MD trajectory indicate that the Cys mostly adopts an unreactive conformation. This is supported by the observed changes in Cys orientation

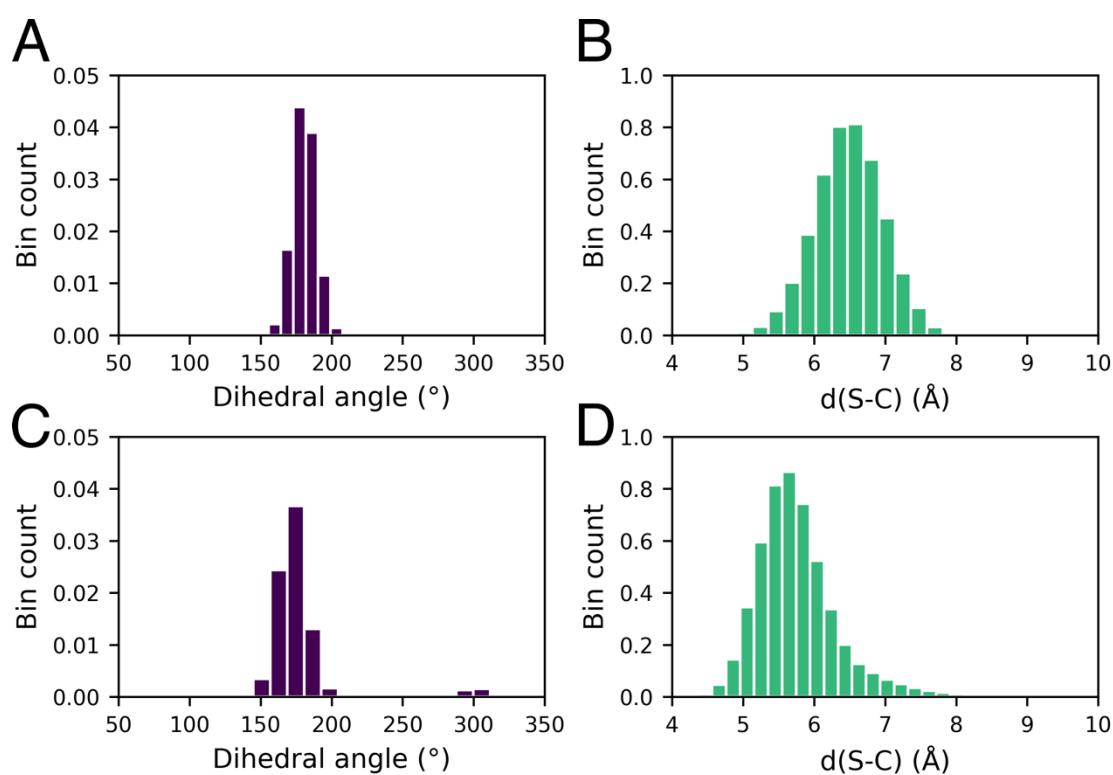
characterised by the dihedral angle (Figure 3.3) in the MD trajectory and also crystal structures of the non-covalently bound and covalently bound forms of ibrutinib. In the crystal structure of an ibrutinib analogue (PDB 5P9I)<sup>99</sup> that is non-covalently bound in the ATP binding pocket of BTK, the C481 dihedral angle measures 173°, indicative of the unreactive conformation observed during MD simulations. In addition, the crystal structure of ibrutinib covalently bound to C481 (PDB 5P9J)<sup>99</sup> shows that the dihedral angle is 62°, the same angle observed in MD simulations where the Cys thiol is oriented directly towards the inhibitor. The orientation of the Cys is therefore likely to be a crucial factor in the reactivity of C481 with covalent warheads, as it needs to be oriented towards the inhibitor for the reaction to occur.



**Figure 3.2** Representative structures of the BTK active site from classical MD simulations showing reactive Cys thiolate conformation (A) and an unreactive Cys thiolate conformation (B).

There is also the possibility that the Cys thiolate is actually the reactive species, instead of the thiol group. This depends largely on the Cys  $pK_a$ , which is investigated in subsequent sections of this chapter. To see if the Cys thiolate adopts similar reactive and unreactive conformations to the thiol, 500 ns of MD was performed on the non-covalently bound BTK/ibrutinib system where C481 was modelled as a negatively

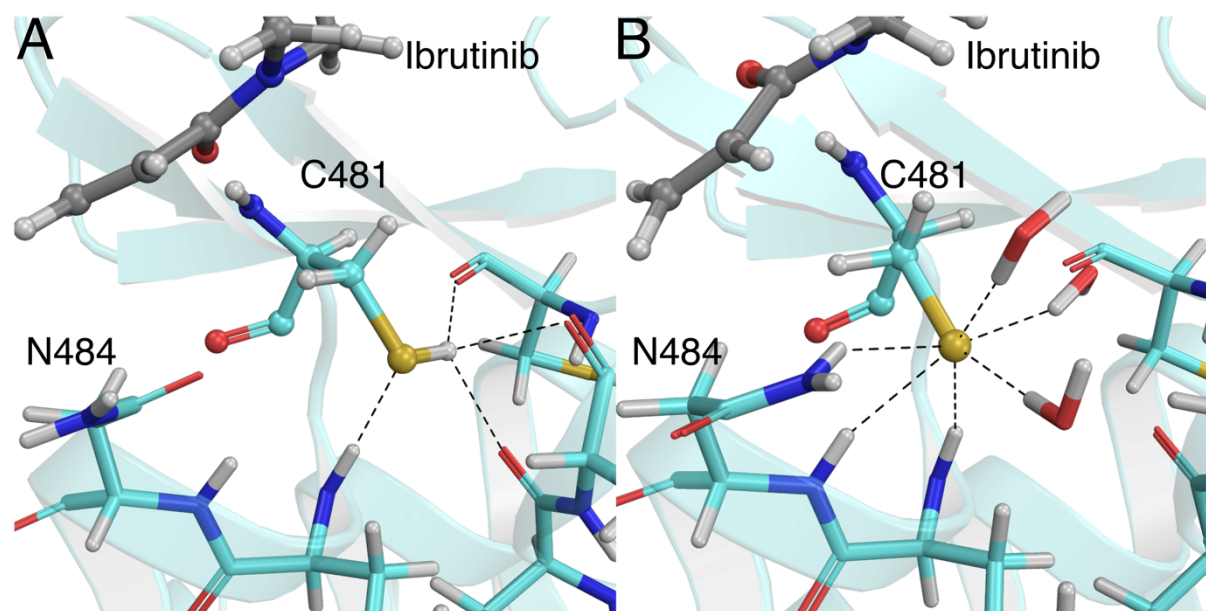
charged Cys thiolate. This is how several reactive Cys are modelled, especially in cysteine proteases.<sup>72,83,172</sup> MD simulations of the Cys thiolate using standard AMBER FF parameters show that, like the Cys thiol, the Cys thiolate also adopts a reactive and unreactive conformation as defined by the Cys dihedral angle. (Figure 3.2). In the unreactive conformation, the thiolate points away from the inhibitor and towards the same pocket as the thiol group, forming interactions with backbone N-H groups of nearby residues. In the reactive conformation, the thiolate points in the direction of the inhibitor, and interacts with the i+3 Asn residue, N484, and 3 water molecules.



**Figure 3.3** Histograms showing the distribution of dihedral angle (A) and S-C distance (B) for the C481 thiol, and the distribution of dihedral angle (C) and S-C distance (D) for the C481 thiolate.

However, when C481 adopts this orientation, the hydrogen bond that holds the inhibitor in place breaks and the inhibitor moves away from C481 (Figure 3.2). This could be the result of electrostatic repulsion between the anionic thiolate and the acrylamide group in ibuprofen, or steric bulk of the water molecules that move into the pocket to stabilise the thiolate. This suggests that when the thiolate is oriented towards

the inhibitor it is not actually a reactive conformation after all. However, this could be an artefact of the classical force field that would not allow any changes in polarisation or electronics that may be induced by the thiolate-inhibitor interaction.



**Figure 3.4** Active site of BTK showing the most stable conformations adopted by the Cys thiol (A) and Cys thiolate (B) over the course of a 500 ns MD simulation.

The C481 thiolate explores a greater number of conformations than the C481 thiol, as shown by the additional cluster in the dihedral angle around  $300^\circ$ . This corresponds to a conformation where the thiolate group points towards the inhibitor, but towards the piperidine ring rather than the acrylamide group. At dihedral angles of  $300^\circ$  the inhibitor also moves away from C481, resulting in large S-C distances of approximately 6 Å due to the presence of water molecules stabilising the thiolate. The S-C distance sampled by the thiolate is on average slightly shorter than the thiol distance (Figure 3.3). These slightly shorter S-C distances appear when the thiolate is oriented away from the inhibitor however, making it unlikely that the thiolate is the reactive state of C481 in BTK. A closer look at the C481 conformations that are predominately observed in MD simulations reveals why the C481 is rarely observed to adopt a reactive conformation, where the thiol group is oriented towards the inhibitor. The pocket that C481 interacts with is comprised of a number of residues that both

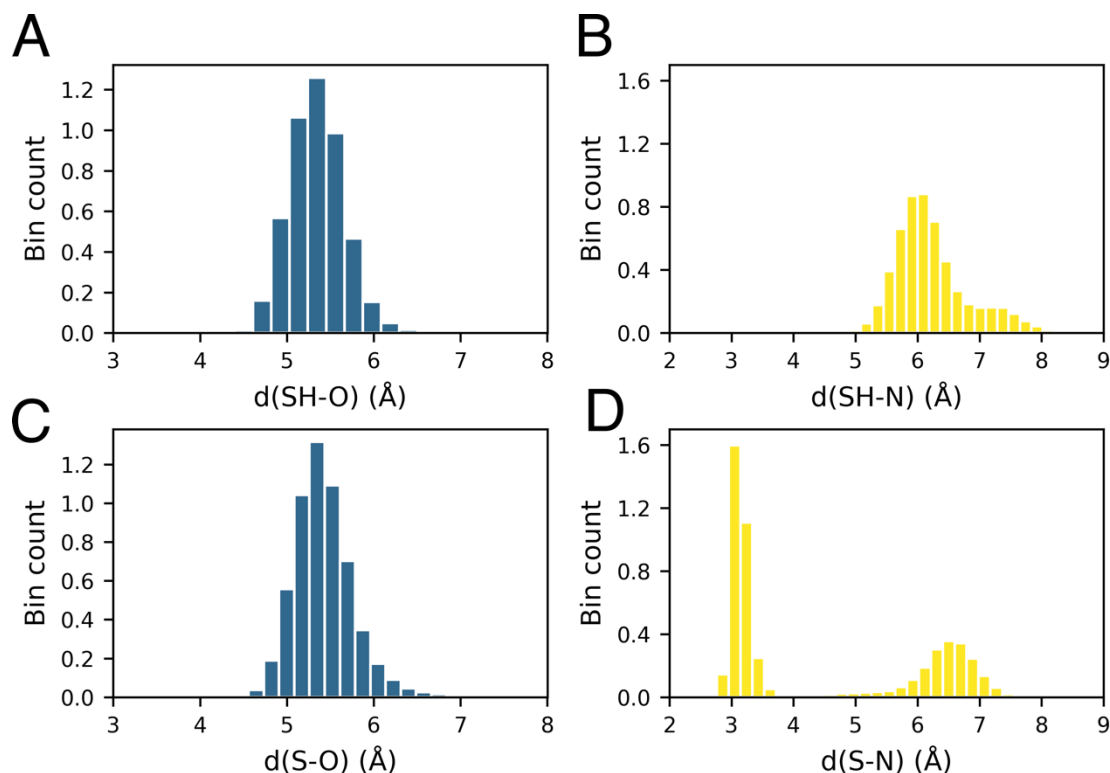
the Cys thiol and thiolate can interact with. The Cys thiol donates hydrogen bonds to the backbone carbonyl oxygen atoms of Ala525, Arg526 and Cys527. It also accepts a hydrogen bond from the backbone N-H of Leu482 (Figure 3.4A). This means the thiol is stabilised in this conformation by four hydrogen bonding interactions, as opposed to only two when the thiol is oriented towards the inhibitor. For the thiolate, the Cys changes position slightly in the pocket so that it can interact with the backbone N-H groups of Leu483 and N484, the amide side chain of N484 and three water molecules. When the thiolate is oriented towards ibrutinib however, two of the hydrogen bonds to the backbone N-H groups break, so it is only stabilised by four hydrogen bonding interactions instead of six.

**Table 3.1** Amount of time that reactive distances are observed between C481 and possible proton acceptor sites in BTK.

	% of frames thiol	% of frames thiolate
Optimal Cys orientation	0.03	0.72
S-C distance < 4.0	0.01	0.02
SH---O=C distance < 4.0	0.09	0.07
SH---ASN484 distance < 4.0	0.03	60.78

The classical MD simulations of the C481 thiol and thiolate both indicate that the principal conformation adopted by C481 is unreactive, as it points away from the inhibitor. This is illustrated by the distribution of the Cys dihedral angle and the S-C distances over the MD trajectory (Figure 3.3). However, there is also a proton transfer step prior to S-C bond formation. It is unclear if any neighbouring residues act as a proton acceptor for C481 in BTK, analogous to the i+3 Asp residue in EGFR kinase.<sup>74</sup> Potential proton acceptors in the vicinity of C481 include the amide side chain of N484, the acrylamide group of ibrutinib itself and a water molecule. Simple distances between relevant atoms (Table 1 and Figure 3.5) show that interactions between the thiol and possible proton acceptor sites do not routinely form over the course of the MD trajectory. The thiolate forms regular interactions with the amide side chain of an

N484, but this is in a stabilising capacity and is unlikely to be in a proton acceptor role. Asn has been reported to act as a proton shuttle, but only when in a basic imine resonance form that can be stabilised by a metal ion.<sup>173</sup>

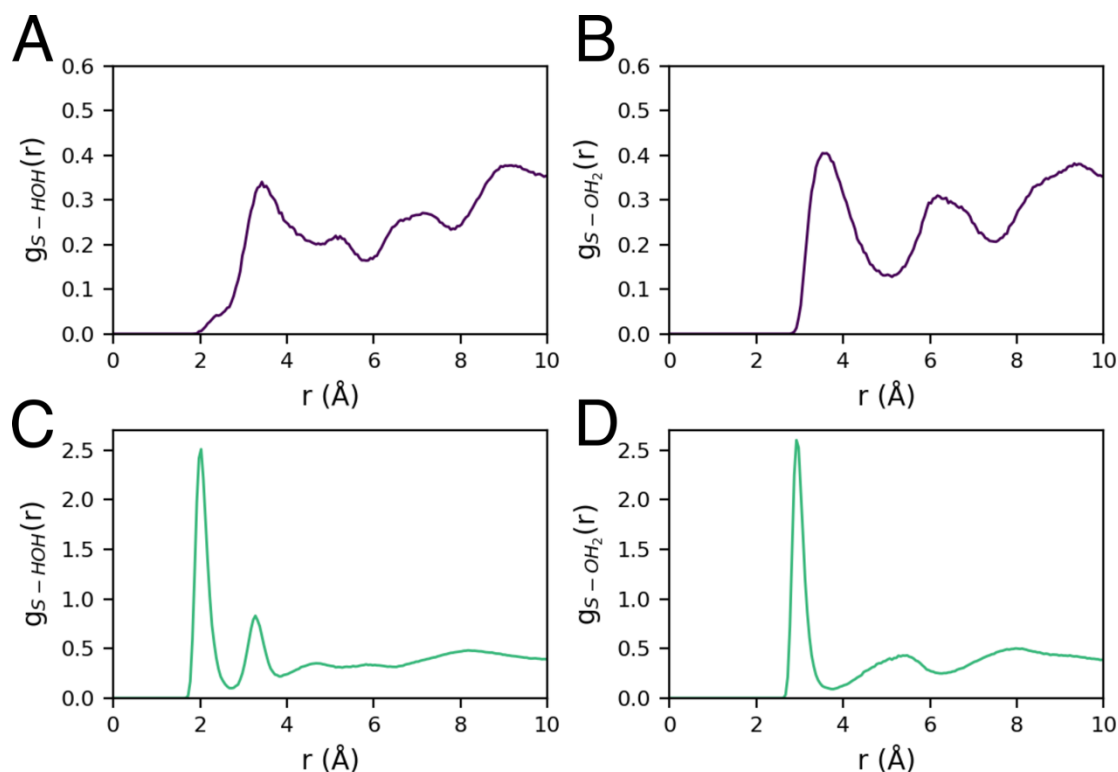


**Figure 3.5** Histograms showing the distribution of distances between C481 and possible proton acceptors. Ibrutinib carbonyl oxygen atom and thiol group (A), amide side chain of N484 and thiol group (B), Ibrutinib carbonyl oxygen atom and thiolate sulfur (C), amide side chain of N484 and thiolate sulfur (D).

The interaction between the C481 thiol/thiolate side chain and water is clearly important. Although the C481 residue is situated on the edge of the ATP binding pocket which appears to be exposed to the bulk solvent, the hydration number calculated along the MD trajectory shows the thiol is quite poorly solvated when ibrutinib is bound, with an average hydration number of 3.2. This is quite low when compared with other kinases, where hydration numbers have been found to range from 3 to 6.<sup>65</sup> C481 is situated in a small pocket and, when the inhibitor is bound, is effectively shielded from water (Figure 3.1), thus helping to explain the low hydration



number. The thiolate appears to be better solvated than the thiol in MD simulations, with an average hydration number of 3.4.



**Figure 3.6** RDFs for the thiol S---HOH interaction (A), thiol S---OH<sub>2</sub> interaction (B), thiolate S---HOH interaction (C), and thiolate S---OH<sub>2</sub> interaction (D).

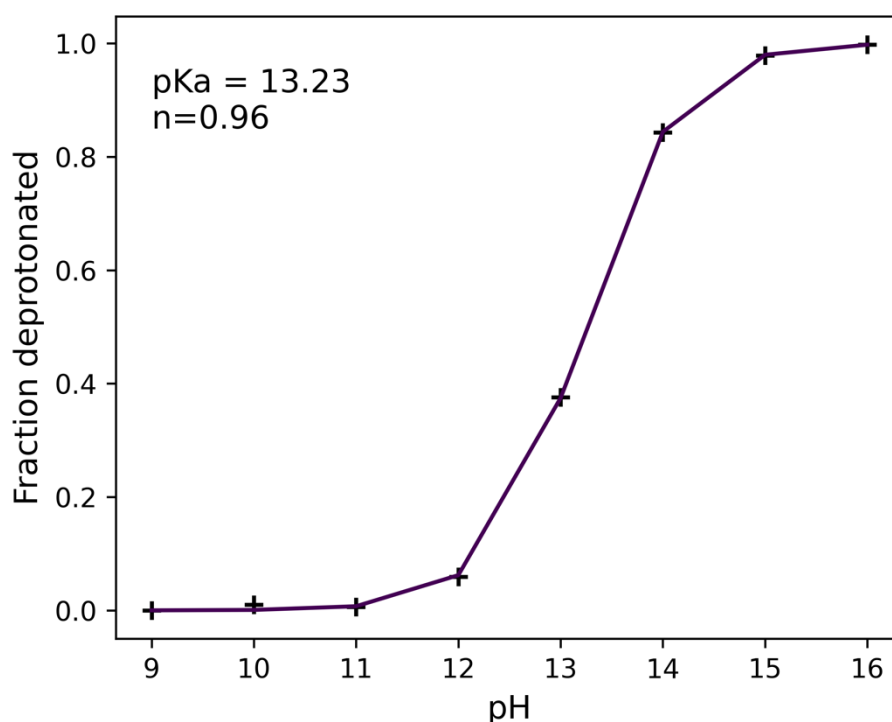
Water molecules clearly play an important role in stabilising the negatively charged thiolate and were observed to move towards the pocket to stabilise the thiolate and even displaced the acrylamide group of ibrutinib in some cases (Figures 3.2 and 3.4). The average hydration numbers calculated by counting the water molecules no further than 3.5 Å away from the C481 thiol group agree reasonably well with the radial distribution functions (RDF) calculated for the Cys thiol and thiolate along the MD trajectory (Figure 3.6). The RDF for the thiol group shows an initial solvation shell centred approximately 3.5 Å away from the sulfur atom. The corresponding integration trace estimates this solvation shell to be comprised of 2.5 water molecules for the S---OH<sub>2</sub> interaction. For the thiolate, the first solvation shell is centred approximately 3 Å away from the S<sup>-</sup> anion and is comprised of 3.2 water molecules.

RDFs are expected to tend toward 1 at large distances ( $r$ ). This is observed in the RDFs presented in Figure 3.6, but only at very large distances of approximately 30 Å because of the presence of the protein. Cys residues with lower hydration numbers are typically associated with higher  $pK_a$  values, although other factors including the electrostatic environment surrounding the Cys are also important. It might be expected that C481 has a relatively high  $pK_a$  due to its low hydration number. Cys  $pK_a$  cannot be measured by classical MD simulations, and so constant pH Molecular Dynamics (CpHMD) simulation must be used.

### 3.3.2 CpHMD

To characterise C481  $pK_a$ , CpHREMD simulations were carried out in explicit solvent. Initially, CpHREMD simulations were run at 8 pH replica values ranging between 5.0 and 12.0. Because the pH of a free Cys residue in solution is reported to be 8.6, the pH replica range was chosen to centre around this value. The pH of C481 in BTK was predicted to be higher than expected at 13.0. As a pH replica of 13 was not explicitly sampled in the CpHREMD simulations, the simulations were repeated at pH replicas ranging between 9.0 and 16.0. This time, a pH of 13.2 was predicted for C481 by CpHREMD simulations in explicit solvent (Figure 3.7). Examination of the RMSD, based on  $\alpha$ -carbons only for each pH replica, shows that the BTK kinase domain appears to be relatively stable, even at high pH values (Figure 3.8). Slightly larger RMSD values are observed at higher pH (especially at a pH of 16.0), indicating larger deviations from the crystal structure, which is to be expected due to the significant deviation from physiological pH. The overall stability of BTK across the pH range is evident from the similarity of the histogram plots and is consistent with experimental studies that have shown kinases retain their secondary and tertiary structure at pH values much higher than physiological pH.<sup>174</sup> The convergence of the CpHREMD simulations at each pH replica were assessed by plotting the cumulative average of the protonated fraction of C481 over the course of the 50 ns MD trajectories

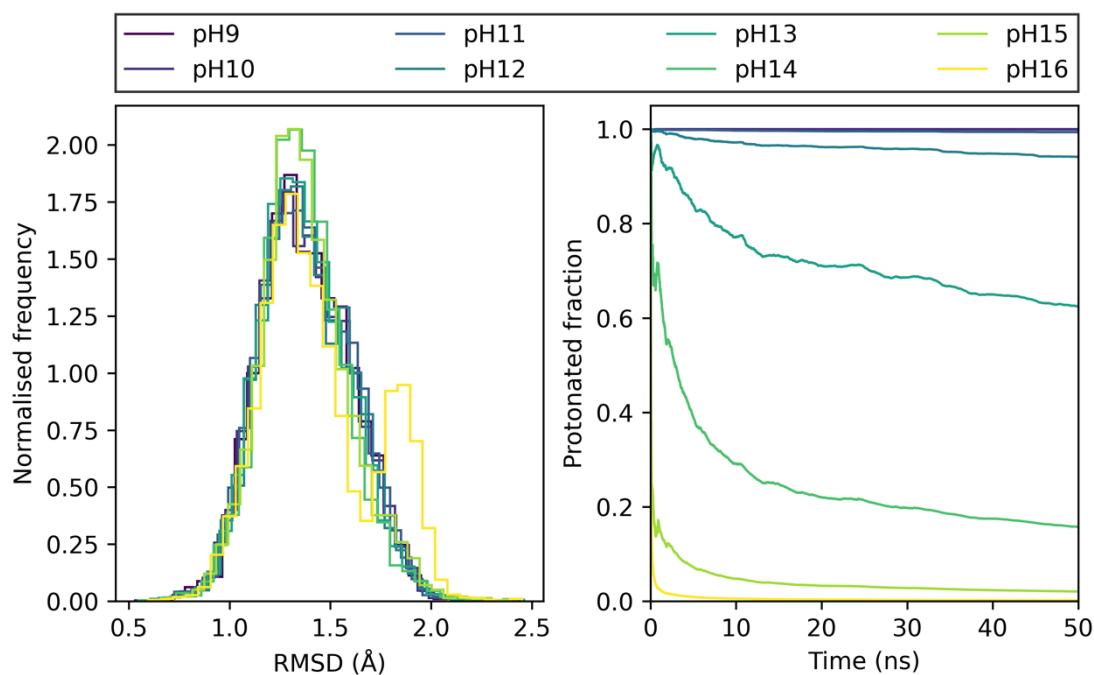
(Figure 3.8), a convergence metric previously used for CpHMD simulations.<sup>171</sup> These plots indicate that there is little variation in the protonation state of C481 after 50 ns of sampling has been performed and that the simulations are therefore well converged. This amount of sampling is consistent with a previous study that used CpHREMD to measure Cys  $pK_a$ s in EGFR and JAK3 kinase.<sup>65</sup>



**Figure 3.7** Titration curve of C481 in BTK from CpHREMD, with 50ns of MD performed per pH replica. The C481  $pK_a$  was calculated using the Hill equation (Equation 3.4), resulting in a  $pK_a$  estimate for C481 of 13.2 and a Hill coefficient of 0.96.

Sampling issues are unlikely to be the cause of the unexpectedly high C481  $pK_a$  prediction. Benchmarking of CpHREMD on a model Cys residue in solution, and the pentapeptide chain Ala-Cys-Phe-Cys-Ala shows that CpHREMD is capable of reproducing experimental  $pK_a$ s of Cys in these environments.<sup>162</sup> There appears to be an issue with calculating pH shifts from the reference compound in protein-like environments such as protein kinases. This has been noted in a recent study, where CpHMD simulations were used to determine Cys  $pK_a$  in the cysteine protease, papain.<sup>67</sup> Cysteine proteases are interesting, as the Cys  $pK_a$  shift is acidic in nature,

and experimental  $pK_a$  estimates give a Cys25  $pK_a$  of 3.3.<sup>175</sup> The aforementioned study calculated the  $pK_a$  of this residue to be 14.0 using the CpHREMD protocol used in this work. The predicted Cys  $pK_a$  is very close to the value predicted for C481 in BTK and very far from the experimental  $pK_a$  (3.3). The authors of the study also tested the GBNeck-2 solvent model with the discrete CpHMD implementation in implicit solvent, and found an improved  $pK_a$  estimate of 8.5, much closer to the model compound  $pK_a$ . In addition, the continuous CpHMD method in implicit solvent method was tested, which has been reported to perform well for Cys  $pK_a$  estimates, particularly for acidic-shifted Cys  $pK_a$ s.<sup>68</sup> However, the continuous CpHMD method was unable to successfully predict an acidic shifted  $pK_a$  for Cys25 in papain, and instead predicted a  $pK_a$  of 9.8.<sup>67</sup> These results suggest that the solvent model, and Born radii used in the CpHMD simulations are important for accurately predicting Cys  $pK_a$ .



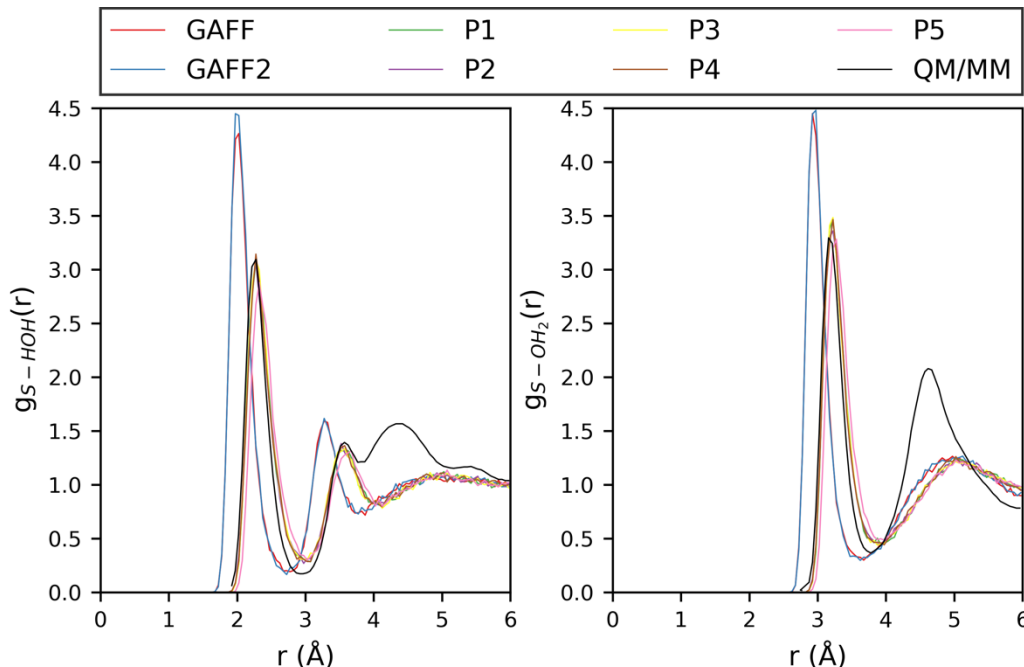
**Figure 3.8** Histograms of  $\alpha$ -carbon RMSD at each pH replica (left hand plot) and cumulative average of the protonated fraction of C481 across the 50 ns trajectory at each replica (right hand plot). The RMSD plot suggests the kinase domain structure is remarkably stable at high pH, and the cumulative average of the protonated fraction shows suggests that the simulations are well converged after 50 ns of simulation time.

**Table 3.2** Modified LJ parameters for thiolate ( $\text{S}^-$ ) in the AMBER FF. Parameter sets 1-5 were generated using linear regression on current AMBER parameters and an estimated upper limit parameter set 4.

Parameter set	$\epsilon$ (kcal mol $^{-1}$ )	$R_0$ ( $\text{\AA}$ )
GAFF <sup>176</sup>	0.2500	2.0000
GAFF2 <sup>176</sup>	0.2824	1.9825
AMBER FF14SB <sup>177</sup>	0.2500	2.0000
1	0.3836	2.2232
2	0.3773	2.2268
3	0.4022	2.2150
4	0.38975	2.2209
5	0.4500	2.2500

To improve predictions of Cys  $\text{p}K_{\text{a}}$  estimates from CpHMD methods, the underlying FF parameters used to describe a Cys residue were investigated in more detail, with a specific focus on the non-bonded Lennard-Jones (LJ) parameters. One potential source of error in the CpHMD method is the use of the same van der Waals (vdW) parameters for both a thiol and thiolate sulfur atom when calculating transition FEs between protonation states for a given amino acid residue. Although this does not seem to be a problem for the majority of titratable amino acids, as accurate  $\text{p}K_{\text{a}}$ s are routinely predicted for systems containing Asp, Glu, His and tyrosine (Tyr) residues, Cys  $\text{p}K_{\text{a}}$ s are generally much harder to predict. However, accurate FF parameters are especially important for Cys given the large change in vdW radius between a thiol and a negatively charged thiolate anion. In the AMBER FF, exactly the same vdW radius is used for sulfur in a thiol group and a thiolate anion.<sup>71</sup> This is a poor description of a sulfur anion and inaccurate LJ parameters will impact upon the correct description of sulfur-water interactions in proteins. This in turn will affect CpHMD simulation in explicit solvent and could contribute to the overestimate of Cys  $\text{p}K_{\text{a}}$  that is characteristic of this method. Different LJ parameters (Table 2) for thiolate in the AMBER FF were tested to examine the effect of thiolate radius and well depth on the

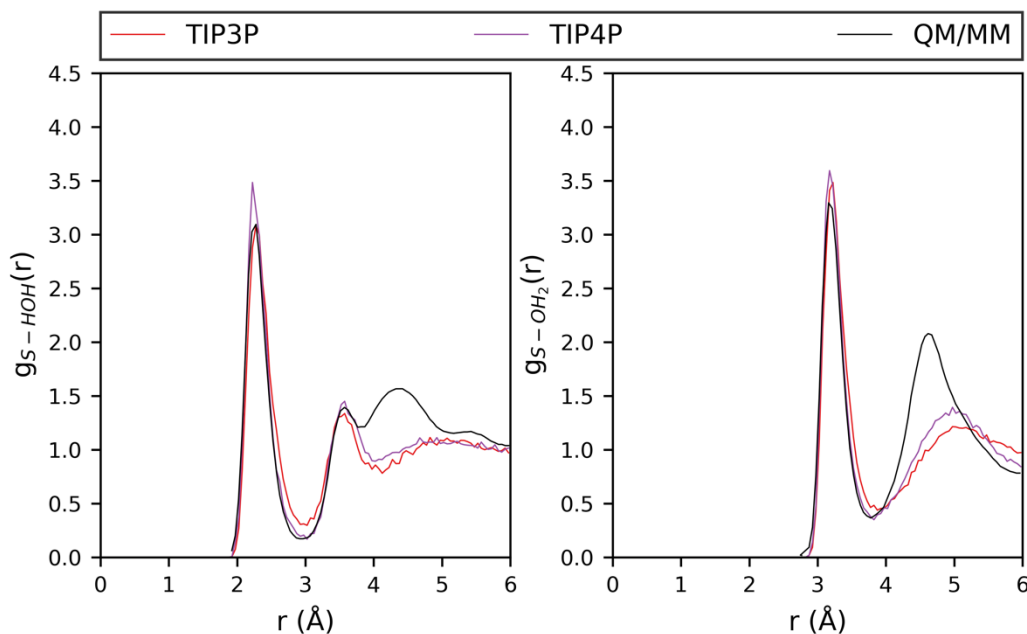
thiolate-water interactions in classical MD simulations of methyl thiol in explicit solvent (Figure 3.9).



**Figure 3.9** Radial distribution functions for methyl thiolate with varying sulfur LJ parameters (Table 2). The RDFs are calculated from 50 ns of classical MD simulations of methylthiolate in explicit solvent using the TIP3P water model. The QM/MM data has been extracted from Figure 2 in ref 70 for comparison.<sup>71</sup>

The radial distribution function calculated at the QM/MM level for thiolate in water has been reported by Awoonor-Williams *et al.*<sup>71</sup> and shows a broad first peak at distances of 2.3 Å and peak intensity of approximately 3.1 for the S<sup>-</sup>---HOH interaction. This corresponds to the first solvation shell surrounding thiolate and comprises six water molecules for methyl thiolate in water.<sup>71</sup> The second shell shows a slight splitting into two peaks and is the result of the second solvation sphere being made up of both QM and MM waters. The S<sup>-</sup>---OH<sub>2</sub> interaction gives a peak at 3.2 Å and peak intensity of 3.3. In comparison, the RDF produced using AMBER FF parameters underestimates the both the distance and strength of the S<sup>-</sup>---HOH interaction.<sup>71</sup> An upper limit parameter set of  $\epsilon = 0.45$  kcal mol<sup>-1</sup> and  $R_0 = 2.25$  Å for sigma and epsilon were estimated initially. After comparing the standard AMBER GAFF and GAFF2

parameters and parameter set 5, it appeared that the relationship between the modified parameters and the change in RDF followed an approximately linear relationship. A simple linear regression model was then used to further optimise the parameters, both for the S---OH<sub>2</sub> and S---HOH interaction. Modified parameter set 3 (Table 2) gives the closest agreement with the QM/MM RDF for the S---HOH interaction previously reported Awoonor-Williams *et al.*<sup>71</sup> The peak intensity and centre of the first peak is now much closer to the QM/MM values compared with the original AMBER FF14SB thiol parameters. The second solvation shell is still poorly described by the parameters when compared to the QM/MM results, however. This could partly be attributed to the split in the second solvation shell from the QM/MM simulations as it is comprised of a mixture of QM and MM waters. Another explanation is the FF used to describe the MM water interactions. The RDFs in Figure 3.9 make use of the TIP3P water model to describe interactions in the explicit solvent in the MD simulations of methyl thiolate that were used to produce the RDFs. The TIP3P water model can accurately reproduce certain characteristics of liquid water, including the density and heat of vaporisation at 25°C and 1 atm.<sup>178</sup> However, it struggles to predict a second solvation shell in RDFs for liquid water.<sup>179</sup> Other water models have been developed that include additional interaction points (sites) with examples that include TIP4P and TIP5P.<sup>180,181</sup> There are also variations to three-site models that add polarization correction (SPC/E) and use different bonded parameters such as alternative HOH angles of 109.5 degrees (SPC).<sup>182</sup> As the four-site TIP4P model is able to reproduce the second solvation shell observed in liquid water and is compatible with the AMBER FF14SB it was tested in conjunction with the newly derived parameter set 3 for thiolate (Figure 3.10).



**Figure 3.10** RDFs for methyl thiolate comparing the TIP3P water model with the TIP4P water model. The same P3 LJ parameter set were used in both cases (Table 2). The RDFs are calculated from 50 ns of classical MD simulations of methyl thiolate in explicit solvent using the TIP3P and TIP4P water model. The QM/MM data has been extracted from Figure 2 in ref 70 for comparison.<sup>71</sup>

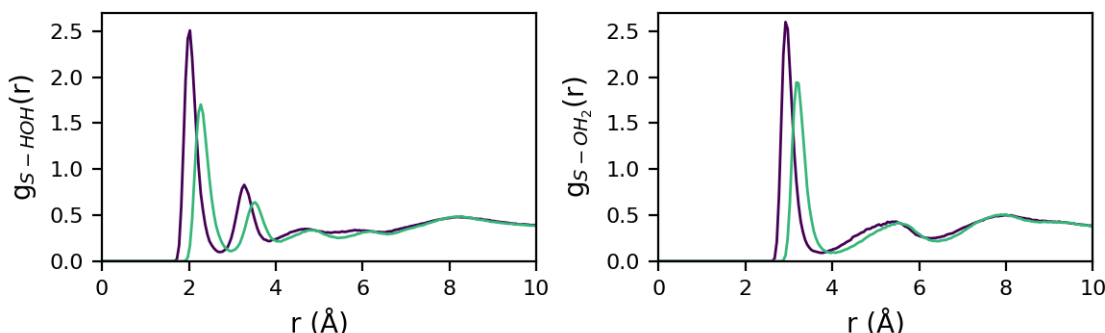
When the MD simulations and RDFs are reproduced using the TIP4P water model, a negligible difference in the RDF is observed when the TIP3P model is used compared with the TIP4P model. Although there is a slightly more pronounced peak in the second solvation shell with TIP4P for the  $S^- \cdots HOH$  interaction, it is more similar in shape and height to the TIP3P RDF compared to the QM/MM RDF. The results in Figure 3.10 suggest there is no significant advantage of using the TIP4P water model and, given that it is more computationally intensive than TIP3P, it was not considered further.<sup>178</sup> Unfortunately, the improved thiolate LJ parameters cannot be used in the current implementation of CpHMD, as only partial charges are allowed to change when the transition FEs between protonation states are calculated. To use the improved LJ parameter set for CpHMD in AMBER, rewriting of the AMBER source code would be required, in addition to extensive testing.



The neglect of changing the vdW parameters between protonation states when calculating transition FEs is significant and has been estimated to contribute up to 20% of the transition FE (the largest contribution arises from the electrostatic component).<sup>183</sup> Underestimating the thiolate vdW radius will lead to an overestimation of the  $\Delta G_{\text{elec}}$  component (Equation 3), which in turn will lead to an increase in the transition FE. An artificially high transition FE will lead to fewer protonation state changes being accepted which will ultimately result in a  $pK_a$  estimate that is too high. Part of this problem can be addressed by modifying the effective Born radius for a given element, used in the calculation of the electrostatic Generalised Born (GB) solvent contribution to  $\Delta G_{\text{elec}}$ . This has shown promise for improving the  $pK_a$  predictions of Asp and Glu residues in CpHMD simulations. An overestimation of the effective oxygen radius due to the inclusion of dummy protons in carboxylate residues has contributed to inaccurate  $pK_a$  shifts.<sup>184</sup> Modification of the carboxylate effective radii from 1.4 to 1.3Å has improved CpHMD predictions for Asp and Glu amino acids.<sup>184</sup> The effect of modifying the effective Born radius of sulfur has improved  $pK_a$  predictions for Cys residues in the continuous CpHMD implementation in AMBER by Harris *et al.*<sup>68</sup> They report an RMSE for their method of 0.95, using continuous CpHREMD in implicit solvent and increasing the effective Born radius for sulfur from its default value of 1.8 Å to 2.0 Å. The use of accurate effective radii is particularly important when simulations are performed in implicit solvent models.<sup>185</sup> Unfortunately, the continuous CpHMD method in implicit solvent was not tested on the protein kinase BTK, as the continuous CpHMD method was published and implemented in AMBER after the results presented in this chapter were completed.<sup>68</sup> Although altering the effective Born radius for anionic sulfur in a Cys thiolate has led to improved  $pK_a$  predictions for Cys in the continuous CpHMD method,<sup>167</sup> discussions with the developers of the discrete CpHMD method in AMBER highlighted doubts that this alone would address the issues in predicting accurate Cys  $pK_a$ s. Although allowing the LJ parameters to change between neutral and deprotonated Cys is likely to improve  $pK_a$  predictions for Cys,<sup>183</sup> it is a significant amount of work that requires modification

of the underlying AMBER code and may occur in the future in collaboration with the developers.

The modified vdW parameters were tested in classical MD simulations of ibrutinib non-covalently bound to the kinase domain in BTK, and the effect of increased vdW radius for thiolate was examined by comparing the RDFs (Figure 3.11) and sulfur hydration number for the modified and unmodified parameters. There is very little change in hydration number between the parameter sets, calculated by counting the nearest number of water molecules within a distance of 3.5 Å over the trajectory. Average hydration numbers of 3.2 and 3.1 for the standard and modified parameters were observed respectively, which are in excellent agreement with the computed RDFs.



**Figure 3.11** RDFs showing the hydration structure around deprotonated C481 in BTK from classical MD simulations using standard AMBER FF14SB LJ parameters (purple) and modified parameters (green) to describe the vdW interaction of the Cys thiolate.

The RDFs for the S-HOH thiolate interaction show that the peak in  $g(r)$  that corresponds to the first solvation shell around the thiolate is shifted to slightly larger distances for the modified thiol parameter set as expected. The integration trace indicated a small reduction in the number of waters in the first solvation shell, with 3.2 with standard parameters and 3.1 with modified parameters. The peak in the RDF is shifted to larger distances ( $r$ , (Å)) when modified parameters are used. This is expected as a larger thiolate radius will result in a slightly weaker hydrogen bond to

water molecules. Overall, there are only subtle differences evident in the MD simulations that use the standard and modified parameters for sulfur in a Cys thiolate. It appears that the increased vdW radius for the thiolate does have a small effect on the interactions of sulfur. This is a promising result and could be useful in CpHMD simulations if the transition FEs between protonation states are calculated with separate vdW parameter sets for sulfur. The current implementation of CpHMD method does not allow for this but increases in the effective sulfur radius have seen promising results so far.

### 3.4 Conclusions

Classical MD simulations and CpHMD simulations have highlighted some of the challenges in accurately modelling sulfur reactivity. The chemistry of sulfur is complex, owing to its size, electronegativity and large variations in the  $pK_a$  of the thiol group. Capturing these complexities in simple MM forcefields is difficult, and the work presented in this Chapter has highlighted some of these issues. One major problem is the LJ parameters that are used to describe sulfur. The AMBER FF uses the same vdW parameters for sulfur in a neutral thiol group and an anionic thiolate group. This is clearly an oversimplification and it has an impact on the length, and therefore strength, of interactions that the sulfur atom can form with neighbouring residues and water molecules. This is of particular significance in CpHMD simulations, where the transition FE between the thiol and thiolate is calculated to estimate the distribution of protonation states during an MD simulation. If the interaction strength is overestimated for the thiolate because the vdW radius is underestimated, the transition FE will be overestimated. The result is a predicted  $pK_a$  that is higher than expected, as fewer protonation state attempts are accepted in the CpHMD framework. Although the charges for thiol and thiolate are different in the AMBER FF and are allowed to change, they are still fixed values and do not fully capture the large polarizability of sulfur. The simulations presented in this chapter support the hypothesis that the current implementation of CpHMD in AMBER favours an alkaline shifted Cys  $pK_a$ .

The modified LJ parameters provide a possible way to address this shortcoming and could be incorporated in future CpHMD simulations.

In spite of the observed limitations in the AMBER FF for sulfur, a number of important reactivity determinants of the C481 residue in BTK have been discovered. It appears unlikely that the  $pK_a$  of C481 is sufficiently low enough for the Cys residue to exist as a thiolate. The consequence of this is the need for a proton acceptor in the vicinity of C481. The suitable candidates include water, the i+3 asparagine (although this is unlikely) and the inhibitor itself. The classical MD simulations show that the Cys mainly adopts an unreactive conformation throughout the trajectory, and rarely interacts with any of the proton acceptors mentioned. Instead, C481 predominantly interacts with the backbone carbonyl oxygen atoms of Ala525, Arg526 and Cys527. Although there are few reactive conformations, they serve as a useful starting point for exploring reactivity in more detail in Chapter 5. The observation of some reactive conformations is encouraging and suggests that any FE penalty for forming a reactive conformation is low. It is difficult to give an estimate of this energy cost without performing additional simulations, such as umbrella sampling on the Cys dihedral angle. Nevertheless, the orientation of C481 in BTK appears to be crucial to its reactivity.

### 3.5 Suggestions for further work

The work carried out in this Chapter has clearly shown that there are limitations of using CpHMD simulations to predict Cys  $pK_a$ , and thus reactivity. For the discrete protonation state CpHMD model implemented in AMBER, the FF14SB parameters used to describe sulfur interactions have been identified as a potential source of error. In particular, the non-bonded LJ parameters for sulfur in a Cys thiol group are the same as the LJ parameters for sulfur in a Cys thiolate group. An improved LJ parameter set for a Cys thiolate was optimised in this work by investigating the effect that changing the LJ parameters had on the resulting RDF of methyl thiolate in

classical MD simulations. These parameters could be further modified and improved by ensuring they can reproduce other chemical properties such as density, heats of vaporization and hydration energies, which are common measures of vdW parameter accuracy.<sup>186</sup> The discrete protonation state implementation of CpHMD in AMBER does not allow the LJ parameters to be varied when the transition FE between protonation states is calculated. Rewriting the AMBER CpHMD code to allow changes of vdW parameters would allow the modified thiolate parameters developed in this work to be used for Cys thiolate in CpHMD simulations. It is likely this is a large amount of work but could be done in collaboration with the developers of the AMBER CpHMD code. The result of such a collaboration would hopefully lead to improvements in  $pK_a$  predictions for Cys containing biomolecules. Extensive testing of the modified thiolate parameters would be required in combination with the discrete CpHMD method. This could be achieved by calculating  $pK_a$  values for Cys residues in proteins that span a wide pH range and checking that the predictions from the discrete CpHMD model in combination with improved LJ thiolate parameters can accurately reproduce the experimental  $pK_a$  values. If the results are satisfactory, the method can be used to predict the C481  $pK_a$  in BTK, for which there is currently not an experimentally known  $pK_a$  value.

## Chapter 4

# Evaluation of methods and convergence testing for modelling thio-Michael reactivity in biological systems

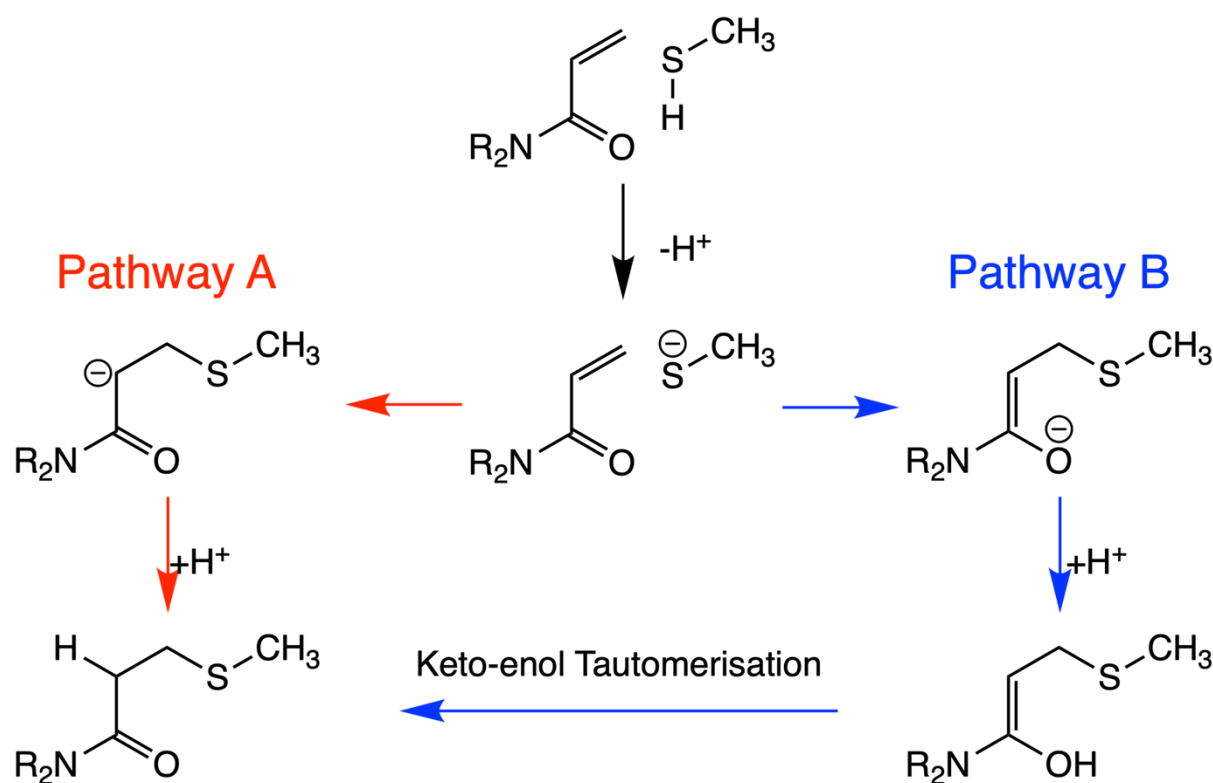
The aim of this chapter is to provide a robust and detailed benchmarking study for modelling thiol addition reactions with quantum chemical methods. Several molecular modelling techniques are used to investigate mechanisms of thiol addition including 1D potential energy scans, transition state (TS) optimisations in combination with intrinsic reaction coordinate (IRC) calculations, and umbrella sampling simulations. Several levels of quantum mechanical (QM) theory are tested that range from semi-empirical methods, density functional methods and post Hartree-Fock (HF) methods to assess which methods give the most accurate results when compared with higher levels of theory. This is essential for modelling biological thiol reactivity in-situ, as producing a free energy (FE) surface with density functional methods are impractical and too costly for modelling thiol reactivity in enzyme targets. Convergence tests are carried out for the weighted histogram analysis method (WHAM) for generating FE surfaces in biomolecular systems to ensure adequate sampling has been carried and accurate FE estimates are obtained. These results form the basis of a computational protocol tested and applied later in this work, that can be used to reliably and practically model thiol reactivity in biologically relevant systems that have applications in drug design.

Some of the work contained within this Chapter has formed the basis of a manuscript that was published in RSC Chemical Science on 28<sup>th</sup> January 2020 as a just accepted manuscript.<sup>187</sup>

## 4.1 Introduction

Conjugate addition, also known as thio-Michael addition of thiols to  $\alpha$ ,  $\beta$ -unsaturated compounds is well described in the literature.<sup>188–192</sup> Sulfur containing nucleophiles are described as ‘soft’ nucleophiles as a result of their high polarizability.<sup>193</sup> As a result, sulfur nucleophiles tend to preferentially react with soft electrophiles such as  $\alpha$ ,  $\beta$ -unsaturated compounds through conjugate, or 1,4-addition, reactions.<sup>194</sup> This is usually attributed to optimal orbital overlap between the LUMO of the  $\beta$ -carbon of the electrophile and the HOMO of the sulfur nucleophile, and is in contrast to direct addition where reactivity is dominated by electrostatic factors.<sup>195</sup> The mechanism of thio-Michael addition is usually assumed to proceed in a three-step base catalysed process.<sup>120,190,195–197</sup> First, deprotonation of the thiol group occurs and results in the formation of a thiolate anion. A conjugate addition reaction then occurs between the sulfur atom of the negatively charged thiolate anion and the  $\beta$ -carbon of the  $\alpha$ ,  $\beta$ -unsaturated carbonyl compound. The resulting carbanion (enolate) intermediate is then re-protonated to form the thioether product (Figure 4.1). Other mechanisms have been proposed, including a nucleophile catalysed mechanism, whereby the thiol group undergoes conjugate addition with an  $\alpha$ ,  $\beta$ -unsaturated compound to form the carbanion intermediate in the absence of an external base.<sup>190</sup> The precise mechanism by which thiol-Michael addition occurs depends on the reaction conditions,  $pK_a$  of the thiol group, solvation and nature of the electrophile.<sup>190</sup>

Computational modelling has been used to investigate the thiol-Michael addition reaction. One of the first modelling studies was performed in 1995, in which Thomas *et al.* used QM calculations to investigate the addition of sulfur and oxygen nucleophiles to acrolein, the first mechanistic step in the reaction of thymidylate synthase.<sup>198</sup> Restricted HF theory and Moller-Plesset (MP) methods were used to model the addition of bisulfide and methylthiolate to acrolein, and found stable intermediates corresponding to the carbanion in both cases.



**Figure 4.1** Two possible thio-Michael addition mechanisms.<sup>199</sup> In pathway A, thiol deprotonation is followed by nucleophilic attack, resulting in an enolate (or carbanion) intermediate.<sup>120</sup> The final step is re-protonation of the enolate intermediate and formation of a thio-ether adduct. In pathway B, nucleophilic attack results in an oxyanion that forms an enol intermediate when protonated. Keto-enol tautomerisation yields the keto product.<sup>199</sup>

Since the emergence of density functional theory (DFT), which includes electron correlation with little extra computational cost compared to HF theory, more recent studies of thio-Michael addition have generally used DFT rather than the HF and post-HF methods used by Thomas *et al.* Some of these studies have modelled the reaction mechanism using a base-catalysed route. For example, Carlqvist *et al.* modelled the thiol addition reaction of acrolein and methane thiol in a two-step process at the B3LYP/6-31+G(d) level.<sup>200</sup> Their model aimed to replicate the reaction that takes place in the active site of the enzyme *Candida antarctica lipase-B*. Three proximal active site residues were included in their computational cluster model including the base that is proposed to deprotonate the cysteine thiol (histidine), and two additional residues that are assumed to form important hydrogen bonding interactions with the



acrolein substrate along the reaction path (glutamine and threonine). A transition state (TS) corresponding to the simultaneous proton transfer and nucleophilic attack step was found, as well as a TS corresponding to the re-protonation of the enolate intermediate. This intermediate linking both TSs was found to be the most stable structure along the reaction path. This was attributed to the large stabilisation of the oxyanion hole through hydrogen bonding interactions with glycine and threonine residues, and from charged interactions from the positively charged imidazole group of the histidine base. A further study carried out by Paasche *et al.* in 2010 used computational modelling of the thiol-Michael addition reactions between a number of substituted  $\alpha$ ,  $\beta$ -unsaturated compounds with methyl thiolate.<sup>199</sup> This work compared a number of reaction pathways for the addition reaction, including the typical base catalysed route (Figure 4.1) and the possibility of protonation of the enolate occurring at the carbonyl oxygen and subsequent ketonisation or tautomerization steps to form the keto product of the reaction. The potential energy surfaces produced by successive geometry optimisations at the BLYP/TZVP level along the proton transfer and nucleophilic attack reaction coordinates for acrolein and methyl thiolate revealed no stable enolate intermediate, but instead spontaneous protonation of the carbonyl oxygen occurred to form an enol product (Figure 4.1, Pathway B). The calculations predicted that a keto-enol tautomerisation step is unlikely, due to the high energy of the enol intermediate. Interestingly, substituted  $\alpha$ ,  $\beta$ -unsaturated compounds were found to react via subtly different mechanisms depending on the electrophile.<sup>199</sup> Other studies have also used different mechanisms to model thiol-Michael addition, including an acid-catalysed route<sup>58</sup> and a water-mediated proton transfer from the thiol to the carbonyl oxygen of the inhibitor.<sup>201</sup> Both of these studies used the B3LYP density functional to model the reaction.

There appears to be a significant variation in the way thiol-Michael addition reactions are modelled computationally as different studies have predicted different mechanisms. These variations could arise from the mechanistic pathway assumed to

operate in each case as a result of the specific system studied, for example whether the reaction takes place in solution or in a protein environment. It could also be the result of using different computational techniques to model the mechanism, and/or the limitations of certain DFT functionals. Engels *et al.*<sup>199</sup> used an ammonia molecule and ammonium ion to mimic the behaviour of the acid/base catalyst that deprotonates the thiol and re-protonates the enolate intermediate. The optimal placing of these species so that facile proton transfer can occur is an oversimplification, especially if the reactivity of conjugate thiol addition is designed to mimic reactivity in a protein, as was the intention of the study. The conformations of catalytic acid/base residues are important determinates of reactivity in these environments.<sup>202</sup> Furthermore, a simple 1D reaction coordinate where the distance between the proton and its acceptor site is varied is a poor description of the complete proton transfer because it neglects the dynamics and bond cleavage in the donor group. The Carlqvist *et al.* study<sup>200</sup> predicted the enolate intermediate to be the most stable species along the reaction path, making the overall reaction slightly endothermic. Although this was argued to be the result of extra stabilisation of the oxyanion hole by nearby charged residues and hydrogen bonding interactions, it could also be the result of the density functional used. In fact, all of the aforementioned studies use B3LYP to model the reaction, with no justification for its use other than it being a good compromise between speed and accuracy.<sup>199</sup> No benchmarking against other density functionals was carried out in that work, despite the well documented limitations of B3LYP as a result of delocalisation error<sup>203</sup> and its tendency to underestimate barrier heights for reactions in simple organic systems compared to other popular density functional methods.<sup>204</sup>

Delocalisation error has previously been reported to cause problems in the calculation of charge transfer reactions,<sup>205</sup> dissociation of radical complexes,<sup>206</sup> the calculation of reaction barriers<sup>207</sup> and other scenarios where complexes are formed with fractional charges.<sup>208</sup> It has recently been recognised as a particular issue for the calculation of stable complexes in thio-Michael addition reactions due to the diffuse

electron density of the thiolate anion and enolate intermediate formed in the reaction. Smith *et al.* have investigated this effect by modelling the sulfur-carbon bond formation step in the addition of methyl thiolate and substituted variants of methyl vinyl ketone.<sup>127</sup> They examined several density functional methods including pure functionals, hybrid functionals and range-separated hybrid functionals. Potential energy (PE) scans of the S-C distance showed that only the range-separated hybrid functionals predicted a stable enolate intermediate, whereas hybrid functionals such as B3LYP gave a flat PE profile, with no stable intermediate. This helps to explain why multiple mechanisms of thiol-Michael addition have been reported using B3LYP, and therefore casts some doubt on their validity. The investigation into alkene substituent effects by Smith *et al.* also confirms why the Carlqvist *et al.* study did find an enolate intermediate despite using B3LYP.

The interest in using  $\alpha$ ,  $\beta$ -unsaturated compounds as covalent inhibitors of cysteine containing targets has grown in recent years. Specifically, acrylamide warheads have been extensively studied as potential drug compounds.<sup>9,209–211</sup> Reliable modelling of this type of reactivity requires the use of a density functional that can accurately predict the energetics of the reaction, including the enolate intermediate formed during thio-Michael addition (Figure 4.1). The aim of this work was to validate and build upon the findings of Smith *et al.* to investigate the most appropriate density functional method for accurately modelling acrylamide (rather than methyl vinyl ketone) reactivity with methyl thiolate. Given that modelling thiol-Michael reactivity in an enzyme environment is the ultimate goal of this work, in addition to testing density functional methods, a number of semi-empirical methods were also examined. Despite advances in computing power, modelling thio-Michael reactivity in large systems with QM/MM molecular dynamics and DFT QM methods on a reasonable time scale can be prohibitive.<sup>76</sup> Semi-empirical approaches are therefore often used as the QM method when modelling reactivity in large biological systems.<sup>73,77,212</sup>

## 4.2 Methods

Here, five density functional methods were tested, including B3LYP<sup>213-215</sup>, CAM-B3LYP<sup>216</sup>, M06-2X<sup>217</sup>,  $\omega$ B97X-D<sup>218</sup> and LC- $\omega$ PBE<sup>219,220</sup>. The four semi-empirical methods tested were AM1<sup>221</sup>, PM3<sup>222</sup>, PM6<sup>223</sup> and SCC-DFTB (also known as DFTB2).<sup>224</sup> Improved DFTB3<sup>225</sup> energies were also calculated by carrying out single point energy calculations at the DFTB3 level on DFTB2 geometries (denoted DFTB3//DFTB2). Results for these methods were compared against MP2 optimised geometries. The correlation consistent aug-cc-pVTZ<sup>226,227</sup> basis set was employed for all calculations. Relaxed PE scans were performed in the gas phase with Gaussian16<sup>228</sup> for all the above mentioned density functionals and semi-empirical methods, where the S-C distance between the  $\beta$ -carbon of acrylamide and the sulfur atom in methyl thiolate was varied in steps of 0.1 Å between distances of 1.6 and 3.0 Å. To investigate reaction pathways for thiol addition, IRC calculations were carried out starting from optimised transition state structures at different levels of QM theory in Gaussian16. Transition state structures were obtained from performing a transition state search using the quadratic synchronous transit (QST) approach implemented in Gaussian16.<sup>228</sup>

The umbrella sampling simulations were carried out in AMBER18.<sup>170</sup> The full details of the molecular dynamics simulation settings can be found in Chapter 3, Section 3.2 of this thesis. For the umbrella sampling, 2D FE surfaces were produced where the first reaction coordinate (RC1) corresponded to a proton transfer from the thiol group to the carbonyl oxygen atom of the inhibitor ( $d[\text{H}_{\text{C481-O}_{\text{ibrutinib}}}] - d[\text{S}_{\text{ibrutinib-H}_{\text{ibrutinib}}}]$ ). The second reaction coordinate (RC2) corresponded to the variation of the S-C distance ( $d[\text{S}_{\text{C481-C}\beta_{\text{ibrutinib}}}]$ ). In each RC window, up to 25 ps of sampling was performed, with the first 5 ps regarded as equilibration and subsequently discarded for analysis. Convergence was then assessed after 4, 8, 12, 16 and 20 ps of sampling. Harmonic biasing potentials of 100 kcal mol<sup>-1</sup> Å<sup>2</sup> and 200 kcal mol<sup>-1</sup> Å<sup>2</sup> were investigated to maintain the umbrella sampling restraints in reach RC window.

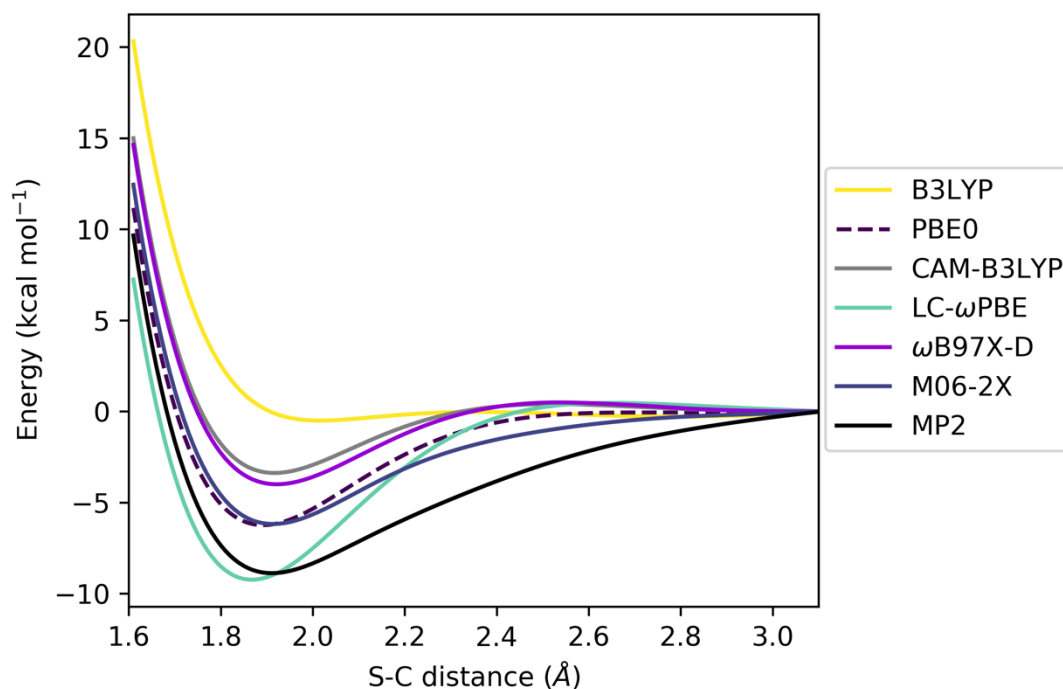
WHAM<sup>229,230</sup> was used to analyse the umbrella sampling simulations and produce the FE surfaces.

### 4.3 Benchmarking thiol addition with 1D potential energy scans

The results for S-C bond formation between methyl thiolate and acrylamide, resulting in an enolate intermediate are consistent with Smith *et al.*<sup>127</sup> For the density functional methods, a stable enolate intermediate is predicted for five out of the six functionals tested, with only B3LYP failing to predict a stable carbanion species (Figure 4.2). Table 1 contains the predicted S-C bond length corresponding to the enolate intermediate for each density functional, and shows they range from 1.86 to 1.91 Å. This is in good agreement with S-C bond lengths found in computational reactivity modelling studies of sulfur reactivity that include high level CCSD(T)//MP2 calculations.<sup>127,231</sup> The density functional  $\omega$ B97X-D gives the closest agreement with the high level MP2/aug-cc-pVTZ geometries, which is also in closest agreement to the experimental value and suggests that benchmarking against this method is accurate and appropriate in this case.

The reason for the apparent absence of a stable enolate intermediate for the B3LYP functional is attributed to delocalisation error.<sup>203</sup> This error is caused by electron-electron repulsion in the Coulomb functional<sup>127</sup> and leads to the artificial lowering of the energies of delocalised systems. In theory, this effect should be mitigated by the exchange-correlation functional. However, due to the inexact nature of this functional, incomplete cancellation of the Coulomb energy and the correlation energy results in the overstabilisation of states with delocalised electrons.<sup>232</sup> This effect has been characterised by the curvature of the energy as a function of fractional charge.<sup>233,234</sup> The exact relationship between these quantities should be linear<sup>235</sup>, but for many DFT functionals, including B3LYP, this is not the case. Rowley *et al.* demonstrated that for thio-Michael addition, the addition of an electron to methyl

vinyl ketone exhibits a significantly curved relationship between fractional charge and energy for B3LYP.<sup>127</sup> For the range-separated functional  $\omega$ B97X-D, however, the relationship was linear.



**Figure 4.2** PE scan of the nucleophilic attack step in the thiol-Michael addition between methyl thiolate and acrylamide at the DFT/aug-cc-pVTZ level. The post HF method MP2 is shown for comparison. All the methods tested show a stable enolate intermediate, apart from B3LYP/aug-cc-pVTZ.

Delocalisation error has been addressed via the introduction of range-separated hybrid functionals.<sup>236</sup> These functionals use a standard error function to split the Coulomb operator  $r_{12}^{-1}$  into two parts, a long-range part and a short-range part (Equation 4.1).<sup>237</sup> Different exchange functionals are then used for each part to remove the delocalisation error. A standard approach is for a density exchange functional to be used for the short-range part, and HF exchange to be used for the long-range part.<sup>238</sup> The  $\omega$  parameter is used to decide when to switch between the partitioned coulomb operator, and often contains empirical parameters.<sup>239</sup> The inclusion of a range correction in the Coulomb functional explains why the  $\omega$ B97X-D, CAM-B3LYP and

LC- $\omega$ PBE density functionals all predict a stable enolate intermediate (Figure 4.2), as they are all examples of range corrected functionals. The hybrid functionals PBE0 and M06-2X also predict stable enolate intermediates on account of the high exact exchange component in these functionals.<sup>217,240</sup> The functional that gives the closest agreement to MP2 for the PE profiles in Figure 4.2 is  $\omega$ B97X-D, in good agreement with previous work.<sup>127,231</sup> Therefore, this density functional was chosen to model thiol-Michael addition in the remainder of this thesis.

**Table 4.1** Predicted S-C bond length of the enolate intermediate formed from the addition of methyl thiolate and acrylamide.

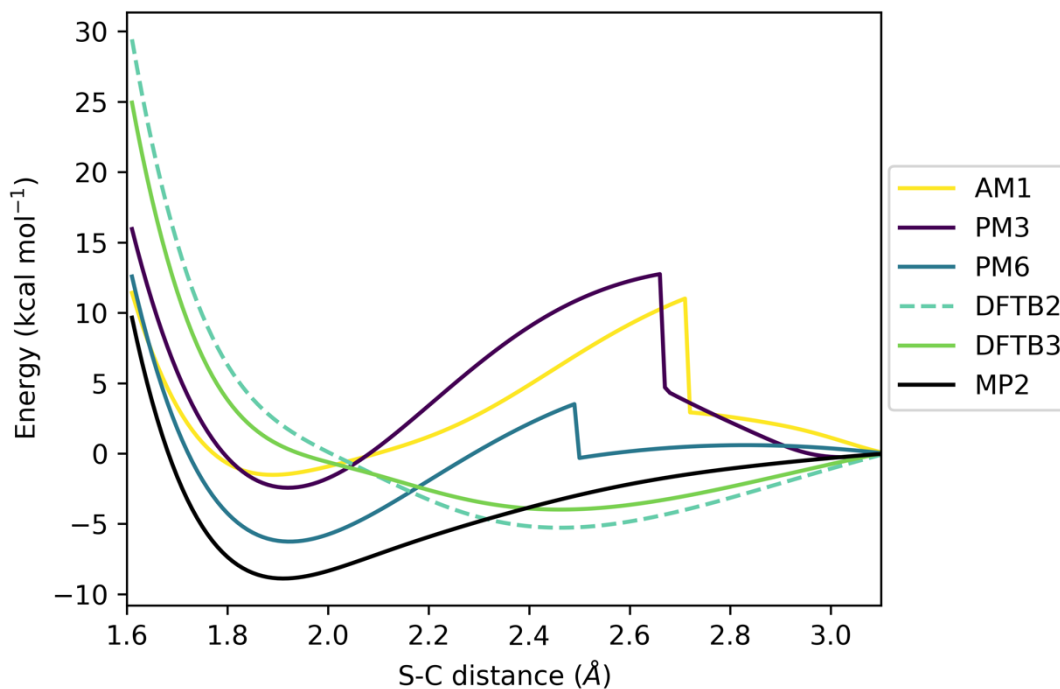
QM method	S-C distance (Å)
MP2	1.90
$\omega$ B97X-D	1.91
PBE0	1.88
LC- $\omega$ PBE	1.86
M06-2X	1.90
CAM-B3LYP	1.91
B3LYP	N/A
AM1	1.88
PM3	1.91
PM6	1.91
DFTB2	2.45

**Equation 4.1** Partitioning of the Coulomb operator into short- and long-range parts in range-separated functionals.<sup>239</sup>

$$\frac{1}{r_{12}} = \frac{1 - [\alpha + \beta \operatorname{erf}(\omega r_{12})]}{r_{12}} + \frac{[\alpha + \beta \operatorname{erf}(\omega r_{12})]}{r_{12}}$$

For the semi-empirical methods, the PE scans along the S-C reaction coordinate exhibit different behaviour to the DFT and MP2 methods. Semi-empirical methods employ a number of approximations in their formalism to reduce computational cost. However, this is often at the detriment of accuracy. Nevertheless, semi-empirical QM methods are still widely used and are routinely applied to problems involving large numbers of atoms. The assumptions and approximations made in semi empirical methods depend on the type of method they descend from. However, most use a minimal set of basis functions, only explicitly treat the valence electrons and ignore the core electrons.<sup>241</sup> The main assumption in these methods is the Zero Differential Overlap (ZDO) approximation, which ignores three and four centre two-electron coulomb repulsion integrals.<sup>242</sup> For AM1, PM3, PM6 and PM7, it is perhaps no surprise that all these methods predict a stable enolate intermediate (Figure 4.3), given that they are all based on the HF method. Although the major assumption in these models is the ZDO approximation, they still inherit some exact exchange from HF. In addition, the exchange-correlation behaviour is somewhat accounted for in the many empirical parameters used in these methods, which are typically derived from experiment.<sup>243</sup> PM6 is in closest agreement to the MP2 PE profile, consistent with other reports that have also found this method to perform well for modelling thiol reactivity.<sup>231</sup> The SCC-DFTB (also known as DFTB2) and DFTB3 methods are based on density functional theory and share many approximations with the ZDO methods including consideration of valence electrons only, use of a minimal basis set and neglect of 3 and 4 centre integrals. The remaining two-electron integrals are then parameterised against all electron DFT calculations, rather than experiment.<sup>243</sup> Additional second-order charge correction terms can be added, resulting in DFTB2.<sup>224</sup> If third-order charge corrections terms are used, the resulting model is known as DFTB3.<sup>225</sup> Both DFTB2 and DFTB3 perform poorly for the S-C bond formation step, exhibiting very broad energy minima at an S-C distance of approximately 2.5 Å that do not correspond to the expected enolate intermediate.





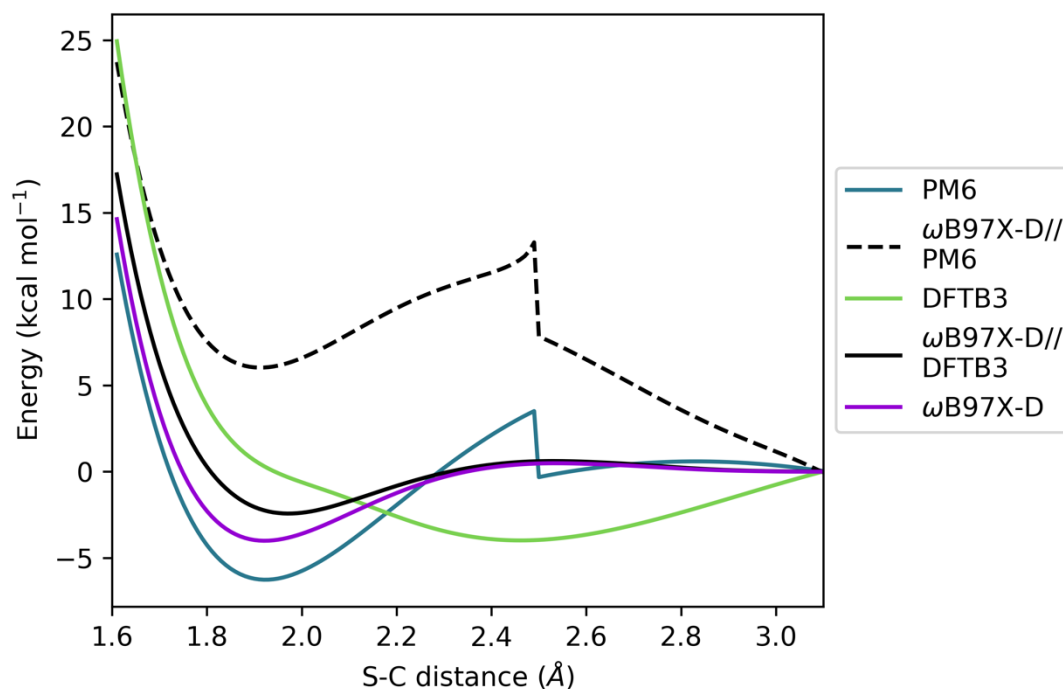
**Figure 4.3** PE scans of the nucleophilic attack step in thiol-Michael addition between methyl thiolate and acrylamide computed using semi-empirical QM methods. The post HF method MP2 is shown for comparison. AM1, PM3 and PM6 show minima corresponding to a stable enolate intermediate, but DFTB3 and DFTB3 show very broad surfaces with no potential energy minimum for the enolate intermediate.

The poor performance of DFTB energies is well documented, even in the latest 3-OB parameter set that was designed specifically to improve upon the energetics of sulfur and phosphorus containing compounds.<sup>244</sup> This parameterisation study notes the energies predicted for sulfur compounds can be inaccurate, but the geometries are much more accurate. To test this, single point energy calculations were carried out using the  $\omega$ B97X-D functional in combination with the aug-cc-pVTZ basis set on geometries obtained from optimisations at the DFTB2 and PM6 level (denoted  $\omega$ B97X-D/aug-cc-pVTZ//DFTB2 and  $\omega$ B97X-D/aug-cc-pVTZ//PM6, Figure 4.4). The results show that the potential energy curve for S-C bond formation at the  $\omega$ B97X-D/aug-cc-pVTZ//DFTB2 level is similar to the  $\omega$ B97X-D/aug-cc-pVTZ energy curve. This indicates that the geometries predicted by DFTB methods are reliable and are suitable

input for single point energy calculations with higher levels of theory. This efficient approach delivers results consistent with the higher level DFT methods and can therefore be used in modelling thiol-Michael addition reactions. For  $\omega$ B97X-D/aug-cc-pVTZ//PM6, the single point energies suggest the PM6 geometries are in poor agreement with  $\omega$ B97X-D/aug-cc-pVTZ geometries, mainly as a result of the observed isomerisation of the acrylamide group observed at large S-C distances.

Most of the semi-empirical methods show erratic energetics at S-C distances of approximately 2.7 Å or greater (Figure 4.3). Inspection of the geometries reveals that these unexpected energy changes are the result of spontaneous cis/trans isomerisation of the olefin group in acrylamide. These conformational changes do not spontaneously occur in the higher level DFT and MP2 PE scans, and so are likely to be an artefact of the semi-empirical QM family. The barrier to this cis/trans isomerism of acrylamide has been estimated at the MP2/aug-cc-pVTZ level to be 4.2 kcal mol<sup>-1</sup>.<sup>245</sup> Although the cis isomer is predicted to be more stable by 1.0 kcal mol<sup>-1</sup>, as the barrier is so low this isomerisation can occur easily. It is likely that the semi-empirical methods underestimate this barrier and explains why this spontaneous isomerisation is observed. However, in a recent study that examined the best methods to use to model thiol addition, 1D PE scans of S-C bond formation was carried out for a number of covalent warheads. Acrylamide was just one compound looked at in the study, and, although the results are consistent with the scans presented in Figure 4.2, the study did not observe any inconsistencies in geometry at S-C bond lengths greater than 2.7 Å. This could be the result of initial starting conformations of methyl thiolate and acrylamide, but as this study did not provide any coordinates of optimised structures at the semi-empirical QM level it is difficult to be certain. However, our geometries at the MP2/aug-cc-pVTZ do agree with the structures reported by Awoonor-Williams *et al.*<sup>231</sup> In addition, coordinates of the optimised geometries reported by Lonsdale *et al.*,<sup>119</sup> who examined methods of characterising covalent warhead reactivity at the M06-2X/6-311+G(d,p) level are in good agreement with the optimised structures found in

this work, at the same level of theory. All of the reported structures show the importance of the methyl thiolate group being situated above the double bond of acrylamide.



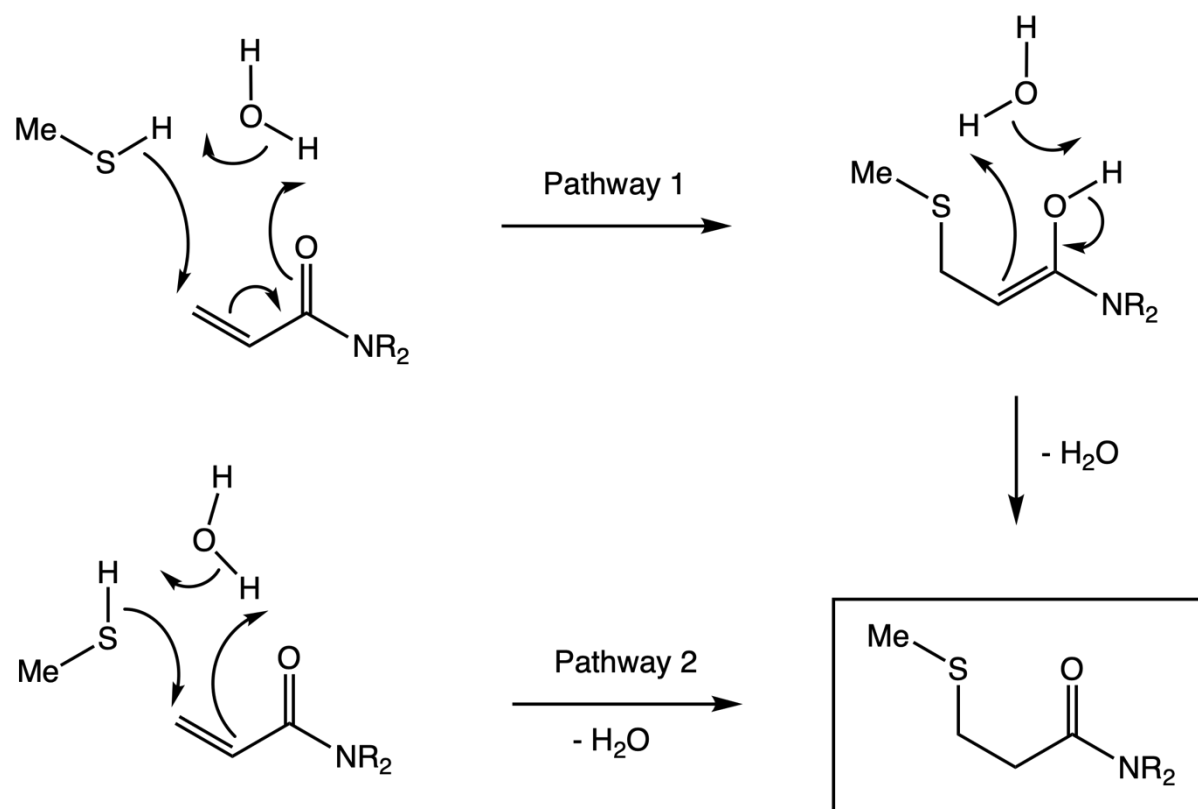
**Figure 4.4** PE scans of the nucleophilic attack step in thiol-Michael addition between methyl thiolate and acrylamide computed with  $\omega$ B97X-D/aug-cc-pVTZ on DFTB2 and PM6 geometries. These PE scans show that the energies calculated with DFTB3 are not accurate, but the geometries predicted by this method are reliable as the single point energies calculated at the  $\omega$ B97X-D/aug-cc-pVTZ//DFTB2 level are in close agreement with the  $\omega$ B97X-D/aug-cc-pVTZ/aug-cc-pVTZ profile. The  $\omega$ B97X-D/aug-cc-pVTZ//PM6 profile shows that the PM6 geometries are in poor agreement with the  $\omega$ B97X-D geometries, particularly for large S-C distances.

The results of the 1D PE scans for methyl thiolate and acrylamide indicate that the most appropriate DFT functional to use for modelling thiol reactivity is  $\omega$ B97X-D. This is supported by other reports where methods of modelling thiol addition have been studied.<sup>127,231</sup> The semi-empirical method that performs best, and is in closest agreement with MP2, is PM6. The performance of PM6 is also supported by other

studies<sup>231</sup>, however the density functional tight binding methods DFBT2 and DFTB3 are also popular choices for modelling thiol addition in protein environments.<sup>74,246,247</sup>

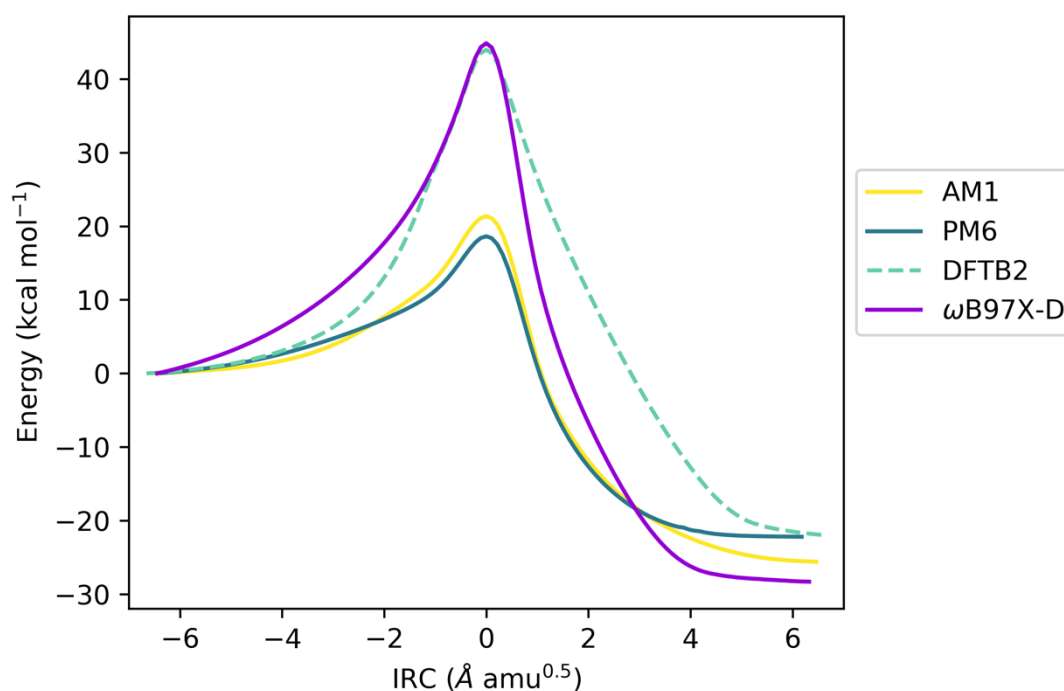
#### 4.4 Benchmarking thiol addition by investigating alternative mechanisms

A range of different QM methods were also used to explore the pathways of thio-Michael addition when mechanisms that differ from the generally accepted mechanism that proceeds through an enolate or carbanion intermediate (Figure 4.1).<sup>120,127</sup> Thio-Michael addition can follow alternative pathways in some systems depending on the nature of the electrophile and the environment surrounding the reaction centre.<sup>199</sup> The possibility of a concerted mechanism, where direct proton transfer occurs from the thiol group to the acrylamide group, was investigated.



**Figure 4.5** Alternative mechanisms for thiol-Michael addition that do not proceed through an enolate intermediate.<sup>201</sup> Instead, water acts as a proton shuttle to form an enol (pathway 1) that undergoes a solvent assisted tautomerisation to the keto product. The second step in pathway 1 shows the water assisted proton transfer directly to the keto product.

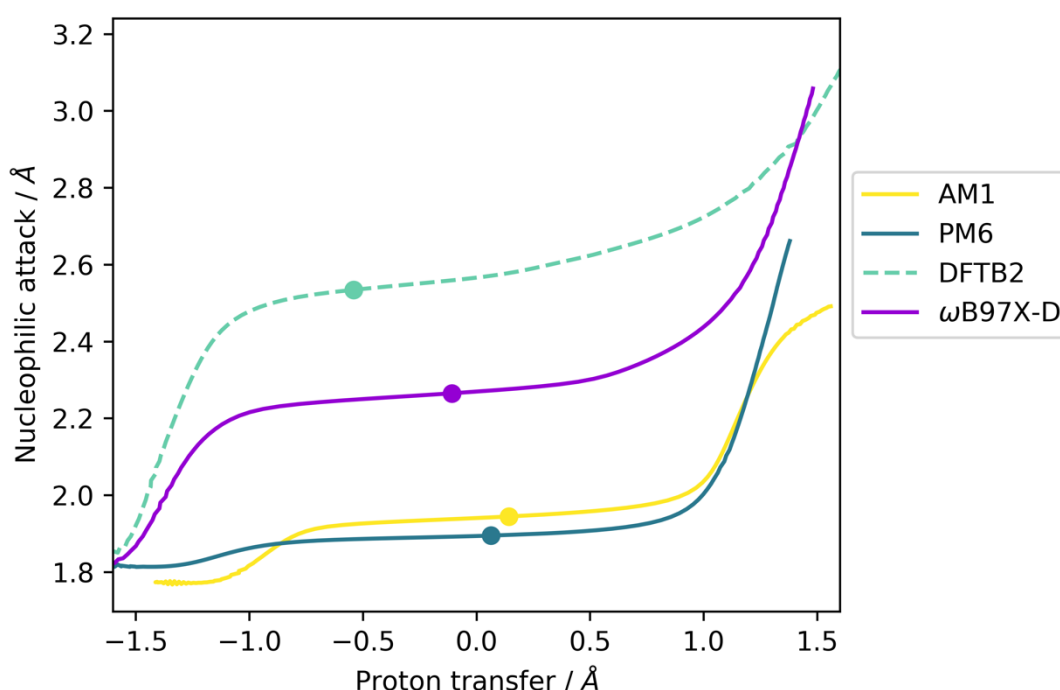
Testing multiple different QM methods for a concerted type mechanism of thiol addition is important to test whether there is consistency between the geometries, and therefore the pathway, for different QM methods. This will be especially important when investigating reactivity in a protein environment (Chapter 5). In order to test for a concerted mechanism, a model system was built that comprised of methyl thiol and acrylamide. Two pathways are possible if a direct proton transfer from the thiol to the acrylamide is assumed to take place. One possibility is where the proton transfers directly to the  $\alpha$ -carbon of the acrylamide group, and the other involves proton transfer from the thiol to the carbonyl group of the inhibitor. (Figure 4.5).



**Figure 4.6** IRC calculations with different QM methods for the pathway 2 mechanism in which the thiol proton transfers directly to the  $\alpha$ -carbon of acrylamide. The  $\omega$ B97X-D/aug-cc-pVTZ pathway predicts a very high barrier, consistent with this reaction being forbidden according to the Woodward-Hoffman rules for pericyclic reactions.<sup>248</sup> DFTB2 energies are in close agreement with the DFT level, whereas both AM1 and PM6 significantly underestimate the reaction barrier.

To investigate how these mechanistic pathways differ in both energy and geometry, transition state optimisations were carried out in Gaussian16<sup>228</sup> on an initial

approximation of the TS structure. This was followed by an IRC calculation to produce an estimate of the pathways that link the reactant, TS and product of both mechanistic pathways. The reaction coordinates corresponding to the S-C bond formation and the proton transfer were extracted from each pathway and plotted to assess the dependence of the pathway on the QM method used to model the reaction. Plots of the energy and pathway for the mechanism in which the thiol proton is transferred to the  $\alpha$ -carbon are shown in Figure 4.6 and Figure 4.7.



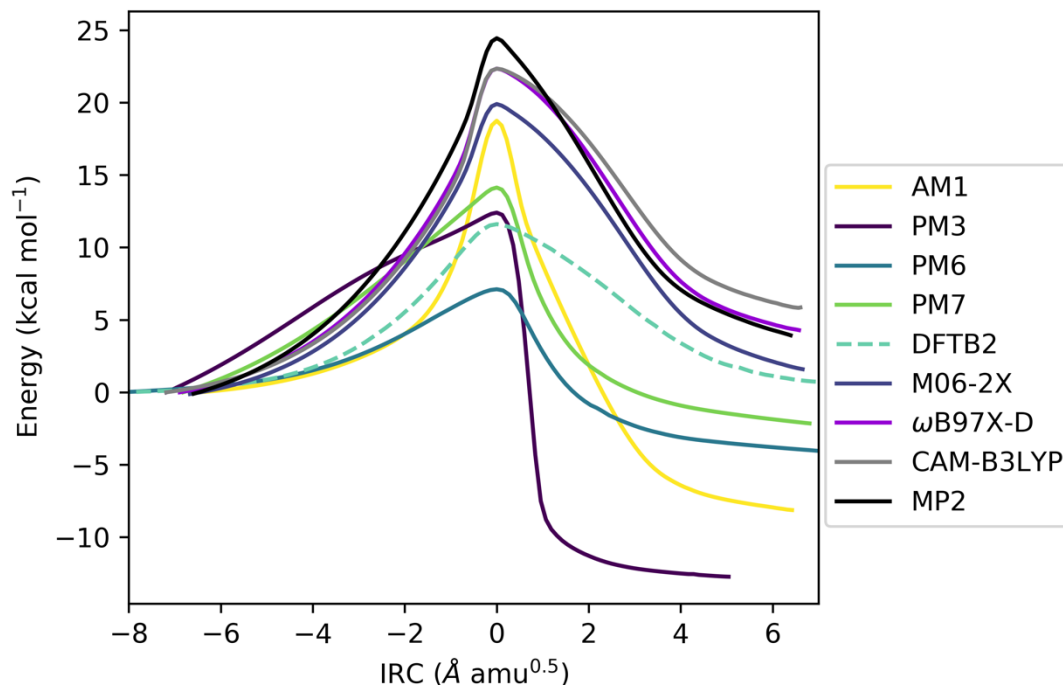
**Figure 4.7** Nucleophilic attack and proton transfer distances along the reaction path predicted by IRC calculations at different QM methods for the pathway 2 mechanism where the thiol proton transfers directly to the  $\alpha$ -carbon of acrylamide. The locations of TSs for each method are shown as circles on each pathway. The semi-empirical pathways are generally in poor agreement with the  $\omega$ B97X-D/aug-cc-pVTZ pathway that suggests an asynchronous concerted mechanism where the proton transfer occurs slightly before S-C bond formation. The AM1 and PM6 pathways seem to underestimate the S-C barrier and predict the S-C bond formation step to happen before PT occurs.

For pathway 2 (Figures 4.6 and 4.7), only a handful of QM methods were tested as a result of the very high energy barrier predicted for the reaction. 1D PE scans were

carried out at the  $\omega$ B97X-D/aug-cc-pVTZ level. Three semi-empirical methods were tested in order to see how comparable the energies and geometries were in comparison with  $\omega$ B97X-D. The  $\omega$ B97X-D/aug-cc-pVTZ profile gives a very high energy barrier of 44.86 kcal mol<sup>-1</sup> for this mechanistic pathway. The high barrier is in agreement with the work of Awoonor-Williams *et al.*, who predicted a barrier of 62.5 kcal mol<sup>-1</sup> for this reaction between MVK and methyl thiol at the CCSD(T)/aug-cc-pVTZ// $\omega$ B97X-D/aug-cc-pVTZ level of theory.<sup>231</sup> The high barrier is the result of the reaction being forbidden by symmetry rules.<sup>231</sup> For a  $4\pi$  electron pericyclic reaction, optimal orbital overlap only occurs if the reaction occurs in a conrotatory manner (orbitals rotate in the same direction).<sup>248</sup> However, for direct thiol addition, the reaction proceeds in a disrotatory manner (orbitals rotate in opposite directions), so it is not thermally allowed. The DFTB2 energies are in close agreement with the  $\omega$ B97X-D/aug-cc-pVTZ PE profile. This is surprising, considering the relatively poor performance of this semi-empirical method for the 1D PE scans (Figure 4.3) that describe the reaction between methyl thiolate and acrylamide. This apparent discrepancy could be the result of modelling a different reaction that does not result in a carbanion intermediate, which is known to be poorly described by the DFT functional that SCC-DFTB is parameterised on. The reaction pathway plot shows that AM1 and PM6 predict qualitatively different reaction paths for the reaction, where S-C bond formation occurs before transfer of the thiol proton. The result is pronounced differences in the TS structures, which could have an impact on reactivity and the energetics. However, this mechanism is not a feasible pathway given the high barrier to reaction and that it breaks orbital symmetry rules. Nevertheless, neither AM1 nor PM6 should be used to model thiol reactivity on account of the prediction of reaction pathways that differ from more accurate density functional methods.

The energy and reaction pathway plots for the first step of pathway 1 (Figures 4.8 and 4.9), where the thiol proton is transferred to the carbonyl carbon, show a different trend. It is clear from Figure 4.8 that the barrier to this reaction is much

lower than the barrier for pathway 2. Therefore, additional QM methods were tested that have already been shown to perform well for modelling thiol addition in the PES scans for methyl thiolate and acrylamide.

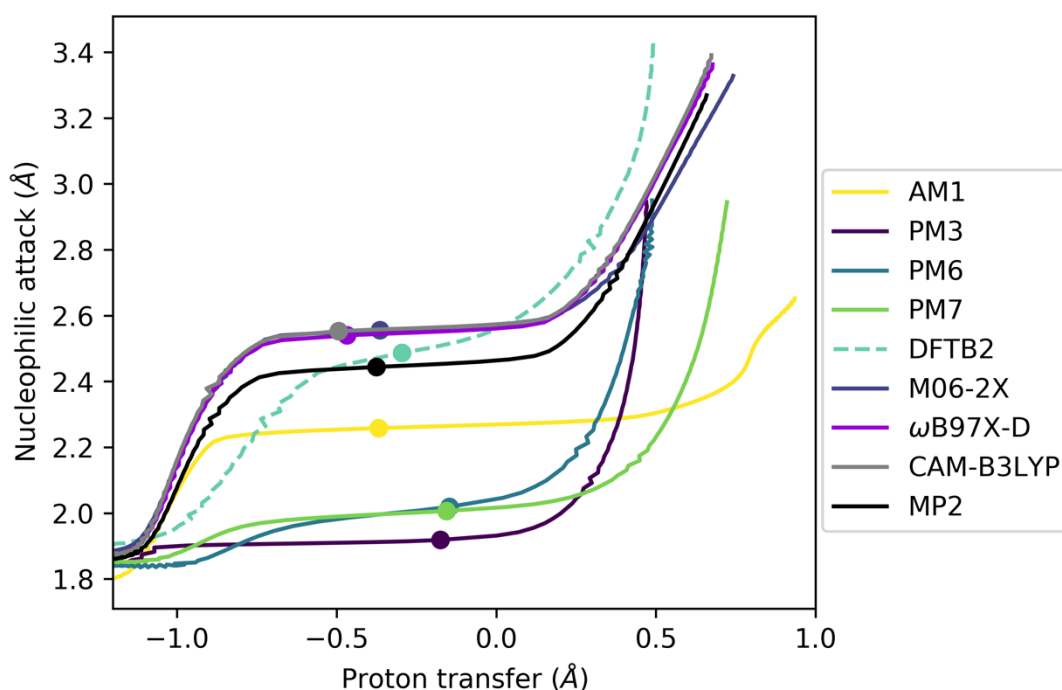


**Figure 4.8** IRC calculations at different QM levels for the pathway 1 mechanism where the thiol proton transfers to the carbonyl oxygen of acrylamide to form an enol. The MP2 IRC has a barrier of 25 kcal mol<sup>-1</sup>. The  $\omega$ B97X-D and CAM-B3LYP methods are in good agreement, with barriers around 23 kcal mol<sup>-1</sup>, whereas M06-2X performs slightly worse by predicting a barrier of 19 kcal mol<sup>-1</sup>. Nevertheless, the results indicate that this is a feasible reaction pathway compared to pathway 2. The semi-empirical methods generally underestimate this reaction barrier compared to higher level methods.

All of the density functional methods give PE barriers between 19 and 23 kcal mol<sup>-1</sup>. These all underestimate the barrier when compared with MP2, which predicts a PE barrier of 25 kcal mol<sup>-1</sup> for this reaction. The semi empirical methods AM1, PM3, PM6, PM7 and DFTB2 all predict substantially different barriers for this reaction, ranging between 6 kcal mol<sup>-1</sup> (PM6) and 18 kcal mol<sup>-1</sup> (AM1). DFTB2 massively underestimates the barrier (11 kcal mol<sup>-1</sup>) compared with the density functional methods and MP2. Although the energetics of this reaction may be described poorly by the semi-empirical methods, the geometries along the pathway are described well,



particularly for DFTB2 (Figure 4.4). AM1 also gives a reasonable description of the geometries along the pathway when compared with the density functional methods, that all predict the proton transfer to occur prior to nucleophilic attack. In other words, the TS, and therefore the rate-limiting step, appears to correspond to proton transfer, rather than C-S bond formation. The PM3, PM6 and PM7 methods all predict a completely different order of reactive events, with the C-S bond formation occurring prior to nucleophilic attack.



**Figure 4.9** Nucleophilic attack and proton transfer reaction coordinates along the reaction path predicted by IRC calculations at different QM methods for the pathway 1 mechanism where the thiol proton transfers to the carbonyl oxygen of acrylamide to form an enol. The location of TSs for each method are shown as circles on each pathway. Again, the semi empirical pathways are generally in poor agreement with the higher-level pathways, with the exception of DFTB2. DFTB2 predicts a concerted reaction pathway in good agreement with the higher-level methods tested, in particular MP2. AM1 and PM6 underestimate the S-C barrier and predict this step to happen before proton transfer occurs.

The preference of semi-empirical methods predicting sequential rather than concerted reaction mechanisms has been observed before in a similar computational study of the catalytic mechanism of hairpin ribosome, where semi-empirical and DFT

QM/MM reaction pathways were compared.<sup>249</sup> This is in contrast with the higher-level methods, and experimental *pH* dependence studies that suggest the thiolate, rather than the thiol is the reactive species in reactions between Michael acceptors and thiol containing peptides in solution.<sup>250,251</sup> Correct descriptions of the energetics and geometries along the reaction path are important, so both of these factors should be taken into account when selecting the most appropriate method to model thiol reactivity.

## 4.5 Benchmarking thiol addition in protein environments with umbrella sampling

The goal of this work is to model thiol-Michael reactivity in protein environments, specifically in the enzyme Bruton’s tyrosine kinase (BTK) as it contains a cysteine residue that can be targeted by covalent inhibitors. Although thio-Michael reaction mechanisms have been reported for cysteine proteases<sup>76,252,253</sup> and protein kinases,<sup>74,211</sup> the mechanisms have invoked an active site residue to function as a catalytic base to assist in the initial deprotonation of the thiol group and subsequent re-protonation of the enolate intermediate following nucleophilic attack. In BTK, there is no obvious candidate to carry out this catalytic role and so it is important to consider mechanistic pathways that do not require an external base. In order to investigate different mechanistic pathways in situ, QM/MM umbrella sampling MD simulations can be carried out.<sup>80</sup> The remainder of this chapter explains, tests and validates the protocol (including the convergence of calculated FEs) used to carry out the mechanistic studies in BTK described in later chapters.

Umbrella sampling is a popular enhanced sampling technique used in combination with molecular dynamics simulations to obtain FE changes between two different thermodynamic states in a biochemical process.<sup>254</sup> Given that reactive events such as bond breaking and forming are associated with significant energy barriers, they are not usually observable on the timescale of a typical MD trajectory. Umbrella

sampling overcomes this problem by forcing the simulation along a particular reaction coordinate using a biasing potential. Harmonic biasing potentials are often used as a result of their simplicity. The reaction coordinate is typically chosen to be a distance, angle or dihedral angle and is divided into a series of windows. The biasing potential restrains the simulation to sample a distribution of values around each reaction coordinate value. If sufficient sampling is achieved, the ensemble average reaches the limit where all of phase space has effectively been sampled: the system is ergodic. At this limit, one can approximate the ensemble average to be equal to the time average for infinite sampling of a particular reaction coordinate. This means that by counting the number of times the reaction coordinate is sampled in a given interval for a given amount of time, the FE change associated with the reaction coordinate in question can be estimated according to Equation 4.2.<sup>255</sup>

**Equation 4.2** FE change in reaction coordinate window  $i$  as a function of the reaction coordinate,  $\xi$ .<sup>255</sup>

$$A_i(\xi) = -k_B T \ln P_i^b(\xi) - \omega_i(\xi) + F_i$$

$A_i(\xi)$  is the FE in window  $i$  as a function of the reaction coordinate.  $A_i(\xi)$  corresponds to the Helmholtz FE if the umbrella sampling simulations are carried out in the NVT ensemble, and the Gibbs FE if the simulations are carried out in the NPT ensemble.  $P_i^b(\xi)$  is the biased probability distribution in window  $i$  as a function of the reaction coordinate,  $\omega_i(\xi)$  is the bias potential in window  $i$ , and  $F_i$  is the FE shift in window  $i$  and is dependent on the FE, as shown by Equation 4.3.<sup>255</sup>

**Equation 4.3.** FE shift in reaction coordinate window  $i$  as a function of the reaction coordinate,  $\xi$ .<sup>255</sup>

$$F_i = -k_B T \ln \left[ \sum_{i=1}^N e^{-\frac{(A_i(\xi) + \omega_i(\xi))}{k_B T}} \right]$$

The FE shift  $F_i$  is dependent on the unknown FE  $A_i(\xi)$ , so  $F_i$  cannot be obtained directly by sampling from molecular dynamics simulations. Instead, equations 2 and 3 can be solved iteratively until self-consistency is achieved. This is the basis of the Weighted Histogram Analysis Method (WHAM), which aims to minimise the uncertainty of the unbiased probability distribution.<sup>229</sup> Code required to analyse umbrella sampling simulations using the WHAM method is readily available and one notable example is distributed by the Grossfield research group<sup>230</sup> and works seamlessly with umbrella sampling outputs from the AMBER suite of programs.<sup>170</sup> Another method that can be used for analysis of umbrella sampling simulations is Umbrella Integration (UI).<sup>256</sup> Instead of averaging the unbiased probability distribution in the WHAM method, UI averages the mean force because this does not depend on the FE shift,  $F_i$ . Kastner *et al.* report that UI can be advantageous compared with WHAM as it converges more efficiently and reduces the statistical error in the FE from MD simulations as the iterative procedure used in WHAM is avoided.<sup>257</sup>

Umbrella sampling provides a simple way to predict FE changes by carrying out molecular dynamics simulations on biomolecular systems, and as a result has been used extensively to predict FE reaction barriers,<sup>258</sup> ligand binding/unbinding free energies<sup>259,260</sup> and conformational effects in enzyme catalysed reactions.<sup>261,262</sup> It has also been used to explore mechanistic pathways in thiol containing systems, particularly in protein kinases and cysteine proteases. Cysteine proteases are enzymes that cleave peptides, and the catalytic mechanism is a thiol addition where the first step is thiol deprotonation usually by a proximal histidine residue. The His/Cys pair is referred to as a catalytic dyad. Nucleophilic attack occurs on the carbonyl carbon of the peptide substrate to form a thioester. This thioester bond is then hydrolysed by a water molecule. Cysteine proteases are promising drug targets due to their roles in many disease processes. Arafet *et al.* used umbrella sampling simulations and the semi-empirical dispersion corrected AM1-d model to investigate the catalytic mechanism of cruzain cysteine protease.<sup>252</sup> They found the mechanism to proceed via a stepwise

acylation step, where a negatively charged thiolate attacks the carbonyl carbon of a peptide bond. This is proceeded by a concerted deacylation step where a positively charged histidine activates a nearby water molecule to attack the peptide and form the final products. In a follow up study by the same group, an additional reaction pathway was found using AM1-d energies corrected at the M06-2X/6-31+G(d,p) level.<sup>263</sup> In this new pathway, the reaction was found to proceed via a neutral cysteine/histidine catalytic dyad, rather than the charged Cys<sup>-</sup>/His<sup>+</sup> ion pair reported in the previous study. Reaction mechanisms of epoxy ketone inhibitors<sup>264</sup>, dipeptidyl nitroalkanes<sup>82</sup> and dipeptidyl nitrile inhibitors<sup>253</sup> of cysteine proteases have also been carried out using semi-empirical methods including AM1-d and PM6, starting from the Cys<sup>-</sup>/His<sup>+</sup> ion pair, rather than neutral thiol and histidine. More recently, there has been a huge amount of interest in the mechanism of cysteine proteases of the SARS-Cov-2 virus due to the global health pandemic. The catalytic mechanism of the SARS-Cov-2 main protease (M<sup>pro</sup>) has been modelled with umbrella sampling using semi-empirical AM1-d energies corrected at the M06-2X/6-31+G(d,p) level. Unlike other cysteine proteases, no stable Cys<sup>-</sup>/His<sup>+</sup> ion pair was predicted at this level of theory, and instead the stable species was predicted to be a thiohemiketal intermediate.<sup>76</sup> However, another study did find a stable intermediate for the Cys<sup>-</sup>/His<sup>+</sup> ion pair of the SARS-Cov-2 main protease at the B3LYP-D3/6-31+G(d) level.<sup>265</sup> This could be a result of the limitations of B3LYP for modelling thiol reactivity as already explored in this work (Section 4.3), and in previous studies,<sup>127,231</sup> or the different method used to explore the FE surface for this reaction. It is clear that there are still some doubts about how the thiol addition mechanism proceeds in cysteine proteases, including which parts are step wise or concerted, and whether a charged or neutral catalytic Cys/His ion pair is the reactive species. This highlights the importance of benchmarking QM methods and of testing convergence in umbrella sampling simulations for thiol addition to ensure the methods used give accurate and reliable energetics and structures for the system of interest.

Thiol addition of covalent inhibitors to protein kinases has also been modelled using umbrella sampling techniques. The addition mechanism of an acrylamide inhibitor of Epidermal Growth Factor Receptor (EGFR) kinase has been modelled with QM/MM umbrella sampling at the DFTB2 level.<sup>74</sup> The mechanism of thiol addition in kinases broadly follows the mechanism depicted in Figure 4.1, where thiol deprotonation is followed by a nucleophilic attack step and then re-protonation of the resulting enolate intermediate. An aspartate residue, D800, situated near to the reactive cysteine (C797) was found to deprotonate the cysteine thiol group prior to covalent bond formation with the acrylamide inhibitor. The protonated aspartate then acts as a general acid to re-protonate the enolate intermediate. Further investigations using QM/MM techniques on an EGFR mutant (L718Q) that developed resistance to the lung cancer drug osimertinib revealed that the mutation does not affect the energetics of C797 alkylation, but instead stabilises a non-reactive conformation of the drug in the EGFR binding site.<sup>211</sup> Both of these studies employ the semi-empirical DFTB2 Hamiltonian that has been shown to predict poor energetics for thiol-Michael addition reactions. However, the results reported in this chapter regarding DFTB methods suggest that the geometries, and in some cases even the energetics along the reaction path, predicted by this method are consistent with range-separated DFT functionals (Figure 4.4). Furthermore, the umbrella sampling method calculates FEs based upon a probability distribution of sampled values along a particular reaction coordinate, and the geometries are therefore a vitally important aspect of calculating accurate FEs from umbrella sampling simulations. In spite of the warnings against the use of DFTB for modelling thiol reactivity,<sup>231</sup> the results reported in this chapter suggest it is an accurate method to use for modelling thiol addition when used in combination with umbrella sampling simulations.

Testing convergence in umbrella sampling is non-trivial, and there are several factors that should be analysed in order to check that convergence has been reached. The vast majority of publications that use umbrella sampling to investigate biological

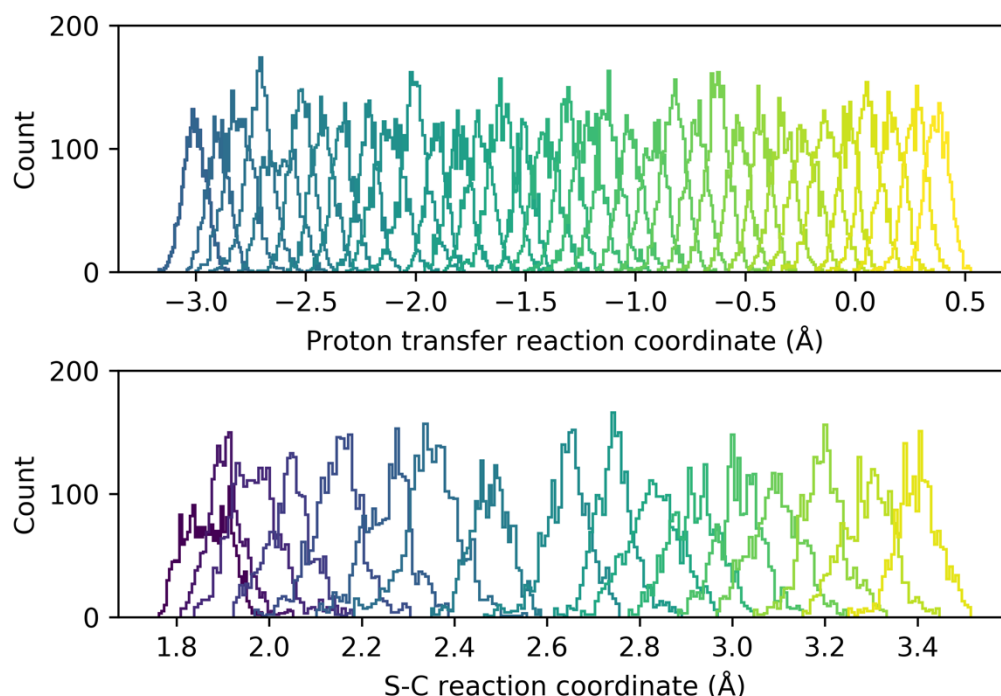
processes such as conformational changes, reactive events and binding/unbinding events often fail to provide detail on how the convergence of their simulations was assessed. However, the generally accepted protocol is to check the convergence of the WHAM equations, the overlap between the histograms of reaction coordinate sampling in neighbouring umbrella sampling windows and the dependence of the FE as a function of the amount of sampling carried out in each reaction coordinate window. Convergence tests of umbrella sampling simulations for thio-Michael addition were carried out by defining a 2D reaction coordinate corresponding to proton transfer and nucleophilic attack. The mechanistic pathway followed for thio-Michael addition in these convergence tests is discussed in more detail in Chapter 5 of this thesis. The reaction coordinates were chosen to include important steps in the mechanism, including proton transfer from the thiol to from the more reactive thiolate species and nucleophilic attack between the thiolate and acrylamide inhibitor.

**Table 4.2** Effects of tolerance value on FE from WHAM analysis. FEs from WHAM analysis using increasingly stringent convergence tolerance criteria. Beyond tolerances of  $1 \times 10^{-8}$ , the FE barrier has converged to approximately 18.37 kcal mol<sup>-1</sup>.

Tolerance value	Free energy (kcal mol <sup>-1</sup> )
$1 \times 10^{-2}$	18.99
$1 \times 10^{-4}$	18.69
$1 \times 10^{-6}$	18.38
$1 \times 10^{-8}$	18.37
$1 \times 10^{-10}$	18.37
$1 \times 10^{-13}$	18.37

Equations 2 and 3 must be solved self-consistently in order to obtain the FE in each umbrella sampling window along the predefined reaction coordinate. Testing the convergence of the WHAM equations is one important component of ensuring the FE and the resulting potential of mean force (PMF), or FE profile/surface, has converged.<sup>266</sup> The implementation of WHAM distributed by the Grossfield lab<sup>230</sup> uses

a convergence tolerance or ‘tol’ keyword in the input to measure the convergence of the WHAM equations. If the value of  $F_i$  changes by a number less than the value of tol, the simulation has reached convergence. Values of tol such as  $1 \times 10^{-5}$  are recommended in the WHAM documentation.<sup>230</sup> The dependence of the FE on different values of tol can be seen in Table 2. Using a tol value of greater than  $1 \times 10^{-8}$  suggests that the FE has converged in the TS window in a 2-dimensional FE surface of thio-Michael addition.



**Figure 4.10** Umbrella sampling histograms along the proton transfer and nucleophilic attack (S-C formation) reaction coordinates after 2 ps of sampling in combination with a biasing potential of  $100 \text{ kcal mol}^{-1} \text{ \AA}^2$  in each umbrella sampling window. The histograms indicate poor sampling, particularly for the nucleophilic attack reaction coordinate where there is poor histogram overlap at S-C distances of approximately  $2.5 \text{ \AA}$  and the histograms are not uniform in shape or height.

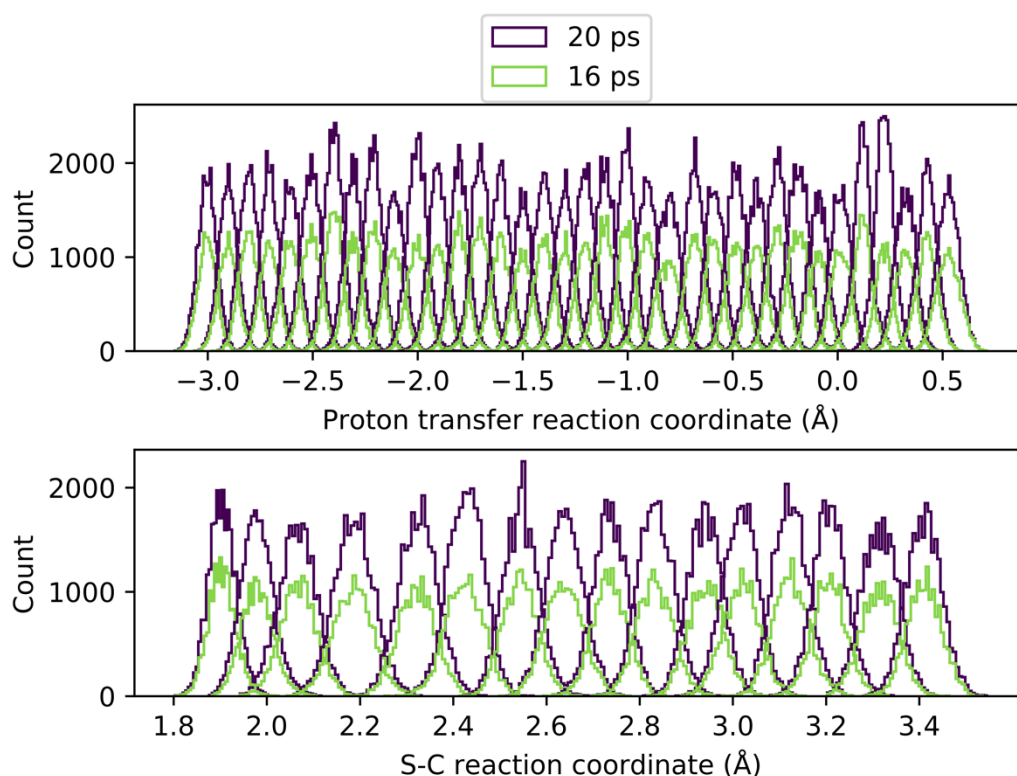
The WHAM approach uses umbrella sampling histograms that count the number of times the reaction coordinate is sampled in a particular umbrella sampling window.<sup>229,267</sup> The histogram shape and overlap are another important test of convergence in umbrella sampling simulations, as it shows whether sufficient sampling



along the reaction coordinate of interest has been achieved. If there is not enough overlap between the histograms of neighbouring umbrella sampling windows, this can lead to inaccuracies in the resulting FE curve from the biased umbrella sampling simulations. Choosing a suitable biasing potential and spacing between each umbrella sampling window sampled is essential for generating smooth FE curves. If a weak biasing potential is used, adequate sampling may not be achieved in the high energy regions of the PMF (Figure 4.10). However, if strong biasing potentials are used, adequate sampling of high energy regions may be achieved to the detriment of good overlap between neighbouring histograms. In these scenarios, additional simulations at smaller intervals along the reaction coordinate must be carried out to ensure sufficient overlap is achieved between neighbouring histograms.

A biasing potential, or force constant of  $100 \text{ kcal mol}^{-1} \text{ \AA}^2$  in combination with 2 ps of sampling was initially used for both the proton transfer and nucleophilic attack reaction coordinates. However, inspection of the histograms along the proton transfer RC (sampled from -3.0 to 0.5  $\text{\AA}$  in steps of 0.1  $\text{\AA}$ ) and the nucleophilic attack RC (sampled from 1.8 to 3.5  $\text{\AA}$  in steps of 0.1  $\text{\AA}$ ) showed that there was relatively poor histogram overlap in the nucleophilic attack reaction coordinate (Figure 4.10). At S-C distances of approximately 2.5  $\text{\AA}$ , there is particularly poor overlap between neighbouring histograms which could lead to an overestimation of the FE in this region. The midpoints of the histograms are also not always centred over their respective reaction coordinate values, indicating that the biasing potential is indeed too low. This is particularly evident at S-C values around 2.3  $\text{\AA}$  as this corresponds to the TS of S-C bond formation. In addition, the histograms along the S-C reaction coordinate are not uniform in size, another indication that there is not enough sampling in each window. The harmonic restraint is not sufficient for some umbrella sampling windows to sample close to their constrained reaction coordinate distance, and instead drift to lower or higher values. The proton transfer reaction coordinate histograms exhibit much better overlap and are a more uniform shape and size. This is not surprising,

given that the S-C reaction coordinate corresponds to the movement of the heavier sulfur and carbon atoms, compared to the proton transfer reaction coordinate. Increasing the biasing potential to  $200 \text{ kcal mol}^{-1} \text{ \AA}^2$  and increasing the amount of sampling in each umbrella sampling window to 20 ps successfully improved the sampling in the high energy regions, whilst maintaining good overlap between neighbouring histograms along the entire reaction coordinate.

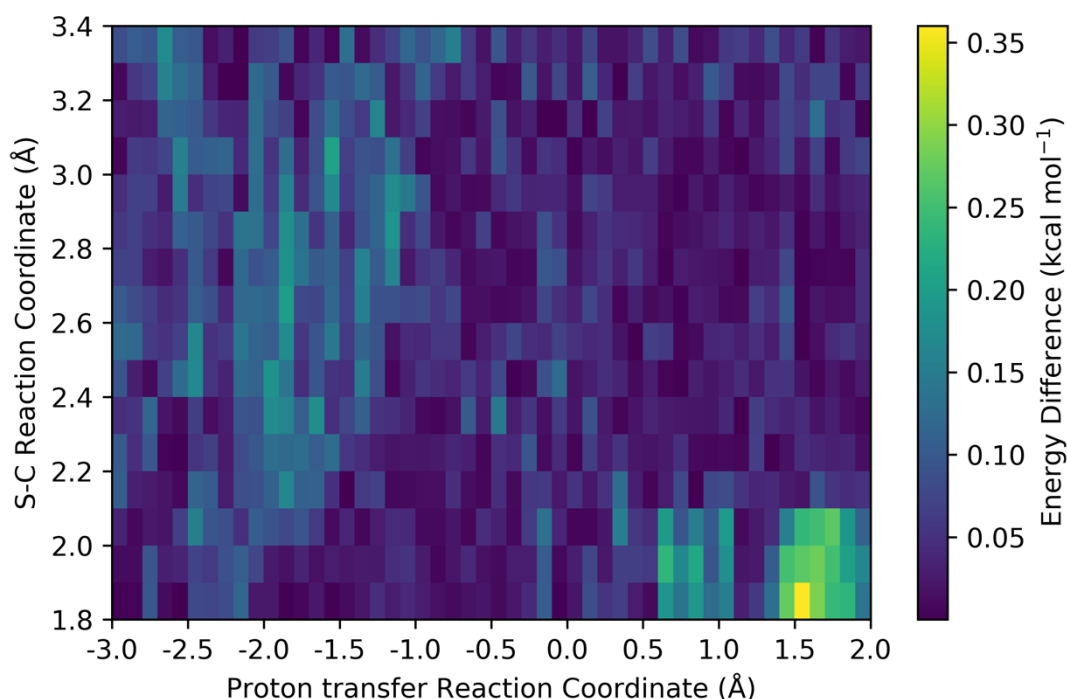


**Figure 4.11** Umbrella sampling histograms along the proton transfer and nucleophilic attack (S-C formation) reaction coordinates after 16 and 20 ps of sampling in combination with a biasing potential of  $200 \text{ kcal mol}^{-1} \text{ \AA}^2$  in each RC window. The histograms indicate good sampling of both reaction coordinates after 16 ps as there is good histogram overlap between neighbouring reaction coordinate values and the histograms are of uniform shape and height.

In Figure 4.11, the histograms along the S-C and PT reaction coordinates have been plotted after 20 ps of sampling per window has been performed and after 16 ps of sampling per window (80% of the data). For both 16 ps and 20 ps of sampling, the histograms along both reaction coordinates show good overlap, and are of a uniform

shape and height. The difference in height of the histograms between 20 ps and 16 ps simulations is due to the increase in sampling. The histograms show that the sampling and overlap is good enough at 16 ps when used in combination of a biasing potential of  $200 \text{ kcal mol}^{-1} \text{ \AA}^2$  for the simulations to be well converged in terms of the amount of sampling in each umbrella sampling window.

There is now confidence that convergence of the WHAM equations has been achieved, and that a sufficient amount of sampling in combination with an appropriate biasing potential has led to umbrella sampling histograms that show good overlap and uniform shapes. The final test for convergence in umbrella sampling simulations is how the energy changes with the different amounts of sampling.



**Figure 4.12** The difference in FE when comparing the FE surfaces produces from 20 ps and 16 ps sampling. Changes in FE are low for the whole surface and are consistently below  $0.4 \text{ kcal mol}^{-1}$ . The largest differences are those in the high energy regions of the FE surface, at S-C distances of  $1.8 \text{ \AA}$  and proton transfer distances of  $1.6 \text{ \AA}$ .

WHAM analysis was performed on simulation data from umbrella sampling along nucleophilic attack and proton transfer reaction coordinates that describe thio-Michael addition. The difference in FE was calculated between the FE surface from 20 ps of sampling and FE surface from 16 ps of sampling. This difference was plotted as a colour map (Figure 4.12) for the full FE surface. It is evident that the FE change between 20 ps and 16 ps is less than  $0.40 \text{ kcal mol}^{-1}$  across the surface. The energy change for the majority of umbrella sampling windows is actually much lower, approximately  $0.10 \text{ kcal mol}^{-1}$  for large areas of the surface. The area with the biggest changes in energy between 16 ps and 20 ps of sampling (approximately  $0.3 \text{ kcal mol}^{-1}$ ) is the bottom right-hand corner of the surface. This area corresponds to a very high energy region on the FE surface where the S-C bond has formed prior to proton transfer. The minimum energy path for this reaction suggests the reaction proceeds in a stepwise manner, where the proton transfer occurs prior to nucleophilic attack. Energy changes along this path are more important and are generally lower than  $0.25 \text{ kcal mol}^{-1}$ . This is well below the  $1.00 \text{ kcal mol}^{-1}$  normally quoted for chemical accuracy, so the energies along the path are well converged after 20 ps of sampling.

The barrier heights predicted from WHAM analysis at various levels of sampling also show very little change, even after 4 ps of sampling (Table 4.3). However, the pathway changed slightly and did not remain constant until after 8 ps of sampling per window, at which point the energy converged to  $18.30 \text{ kcal mol}^{-1}$ . The number of additional erroneous energy minima on the surface also decreased as the amount of sampling increased. Although the energies and histograms suggest that sampling seems to have converged after only 8 ps, it is unlikely that performing additional sampling in each umbrella sampling window will be detrimental. In fact, the majority of umbrella sampling studies that investigate thiol addition use between 10 ps<sup>72</sup> and 50 ps<sup>211</sup> of production sampling per window, with biasing potentials ranging from  $100 \text{ kcal mol}^{-1} \text{ \AA}^2$ <sup>211</sup> to nearly  $600 \text{ kcal mol}^{-1} \text{ \AA}^2$ <sup>263</sup>. It is also customary to carry out an equilibration in each window, where the data collected during this period is not included in the

WHAM analysis. As most studies do not provide detailed information on how convergence was checked, it is difficult to know how much sampling is appropriate. The analysis presented in Figures 4.11 and 4.12 showed that after 20 ps of sampling with a biasing potential of 200 kcal mol<sup>-1</sup> Å<sup>2</sup> and a 5 ps equilibration time in each umbrella sampling window, the FE is well converged. Running an extra 10 ps of sampling along the predicted minimum energy path helps to maintain sensible geometries and guarantee that no high energy and unrealistic structures are sampled.

**Table 4.3** Change in barrier height with increased sampling per reaction coordinate window in QM/MM umbrella sampling simulations. As sampling increases, the barrier height converges.

Amount of sampling per window (ps)	Barrier Height (kcal mol <sup>-1</sup> )
4	18.6
8	18.3
12	18.4
16	18.3
20	18.4

## 4.6 Conclusions

The extensive benchmarking performed in this chapter for the modelling of thio-Michael addition reactions has allowed the definition of a robust, practical protocol for modelling sulfur reactivity in biological systems. Stepwise 1D PE scans of S-C bond formation and 2D concerted IRC calculations involving a proton transfer step were carried out with several density functional and semi-empirical methods. These reaction pathways were compared with high level MP2 geometries and energies. Limitations with the popular density functional B3LYP were found: they result from a delocalisation error causing a flat PE profile and no energy minimum for the enolate intermediate following S-C formation.<sup>127</sup> These limitations can be overcome by using range separated density functionals, of which  $\omega$ B97X-D was found to perform

particularly well. Of the semi-empirical methods tested, AM1, PM3 and PM6 predict a stable enolate intermediate. However, all of these methods do not predict the correct energies and geometries at large S-C distances due to spontaneous cis/trans isomerisation of the olefinic bond in acrylamide. The popular density functional tight binding methods DFTB2 and DFTB3 performed poorly in predicting accurate energetics for S-C bond formation. In spite of this, the geometries predicted by these methods are presumably close to those that would be seen with  $\omega$ B97X-D because the resulting single point energy calculations with this DFT method using the DFTB3 geometries led to very similar energetics to the full DFT method alone. On the other hand, PM6 performs rather poorly for predicting accurate energies and geometries, despite being recommended for modelling sulfur reactivity by some studies. This holds true for the 1D PE scans of S-C formation, and 2D IRC scans of S-C bond formation and proton transfer. This is a particularly important finding, as semi-empirical methods are routinely used to model sulfur reactivity in enzymatic reactions using the umbrella sampling technique. Currently, there appears to be no general consensus on the best semi-empirical quantum method to use for investigating sulfur reactivity, with PM6, AM1 and DFTB methods all being used to model reactions of cysteine proteases and protein kinases in the literature. By convergence testing for thiol addition of acrylamide containing compounds to a cysteine thiol, an accurate QM/MM protocol has been established that yields converged FEs. Testing the convergence of the WHAM equations, scrutinising histogram overlap across a FE surface, and inspecting the change in FE as a function of sampling led to the following protocol: 5 ps of equilibration in each umbrella sampling window, which is discarded for analysis, followed by 20 ps of production sampling with an umbrella biasing potential of 200 kcal mol<sup>-1</sup> Å<sup>2</sup>. The semi-empirical method DFTB3 is used to describe the QM part of the system. A further 10 ps of sampling along the minimum FE path is then carried out to increase the sampling in this region for 2D FE surfaces. As this protocol has been shown to give accurate FEs and is computationally feasible in terms of time

required, it was chosen to explore alternative reaction mechanisms for the addition of covalent inhibitor ibrutinib to its protein target BTK in Chapter 5 of this thesis.

## 4.7 Suggestions for further work

The testing of different semi-empirical and density functional QM methods in this chapter has shown that no semi-empirical method can accurately reproduce the correct energetics and structures in thiol addition reactions. Although the semi-empirical tight binding DFTB2/3 methods result in geometries and pathways for thiol addition that are in good agreement with range-corrected density functional methods, there is clearly room for improvement. A new semi-empirical method based upon the self-consistent density functional tight binding scheme, named GFN-xTB was recently published by Grimme et al.<sup>268</sup> This method performs slightly better than DFTB3-D3 for cysteine conformational energies from the CYCONF<sup>269</sup> benchmarking set. An extension of GFN-xTB, called GFN2-xTB was recently released, which is a more robust and physically sound method than GFN-xTB due to the native inclusion of electrostatic interactions and exchange correlations effects, amongst other differences.<sup>270</sup> GFN2-xTB has been shown to perform well for a range of molecular properties including structures, noncovalent interaction energies and even barrier heights, despite not being parameterised for this purpose. As a result of the novelty of the GFN-xTB and GFN2-xTB methods, there are only a handful of examples in the literature of its use in QM/MM simulations,<sup>271,272</sup> and no direct test of its applicability for modelling sulfur reactivity. An extension of this work is therefore the inclusion of GFN2-xTB in the benchmarking of methods for sulfur reactivity. It has the potential to provide superior molecular properties to currently available semi-empirical methods at a fraction of the computational cost of density functional methods. This is useful in a drug design context, where molecular properties of large numbers of compounds can be compared in lead optimisation scenarios.

The analysis and combination of data from umbrella sampling simulations and generation of FE profiles presents another opportunity for further work. Alternative methods for analysing umbrella sampling simulations other than WHAM are available, including an extension of WHAM known as the dynamic histogram analysis method (DHAM).<sup>273</sup> DHAM overcomes the erroneous FE predictions made by WHAM if insufficient equilibration has been performed in each umbrella sampling window by using a Markov state model to obtain the FE along the reaction coordinate. The use of a Markov state model provides the user with kinetic data that can be used to assess the convergence in each reaction coordinate window. DHAM has been reported to give accurate FE estimates in scenarios where WHAM fails,<sup>274</sup> and so could be compared with WHAM for obtaining FE simulations from modelling thiol addition reactions with QM/MM methods. As already mentioned in section 4.5, umbrella integration is another method that can be used for analysing biased molecular dynamics simulations.<sup>255</sup> The method can reduce the statistical error when compared with WHAM<sup>256</sup> and would be another method that can be compared with WHAM to generate FEs from umbrella sampling simulations. A potential limitation of using umbrella integration and DHAM however, is the accessibility and implementation of each method in the proposed QM/MM protocol. Improvements to the umbrella sampling technique could also be explored with the use of replica exchange umbrella sampling (REUS). In REUS, replica exchange attempts are made between different umbrella sampling potentials.<sup>275</sup> This allows extensive sampling of conformational space to be performed compared with conventional umbrella sampling and can therefore help to improve convergence in umbrella sampling simulations.<sup>276</sup>

Finally, alternative embedding schemes could be investigated for improving the accuracy in QM/MM reaction simulations. The work contained within this chapter has demonstrated that there is a large variation in the accuracy of barrier heights that are predicted for modelling thiol addition reactions when using density functional and semi-empirical QM methods. The projector embedding approach, where a



wavefunction method is embedded within a density functional method, has been proposed as a robust way to remove the functional dependence on barrier height in QM/MM simulations.<sup>128</sup> By embedding correlated ab initio methods in DFT to model the Claisen rearrangement of chorismite to prephenate in chorismate mutase, the variation in barrier heights was reduced from 13 kcal mol<sup>-1</sup> to 0.3 kcal mol<sup>-1</sup> compared to using DFT alone.<sup>277</sup> Alternatively, embedded mean field theory (EMFT) could be explored as an embedding approach. In EMFT, one density functional method can be embedded within another.<sup>278</sup> The potential advantage of this embedding scheme in QM/MM reactivity simulations is that the QM region could be portioned into two subsections, one that's treated with a higher level DFT method and one that's treated with a lower-level DFT method. Furthermore, the partition between each density functional can cross multiple and polar bonds without compromising on accuracy.<sup>279</sup> For the reaction between BTK and ibrutinib this would be advantageous as a higher-level density functional such as  $\omega$ B97X-D could be used to treat a small subset of reactive atoms within the QM region, such as the thiol side chain of C481 and the reactive atoms in the acrylamide group of ibrutinib. The remainder of ibrutinib can then be treated with a lower-level method such as LDA with a minimal basis set. This has the potential to result in gains in speed to be made over pure DFT QM/MM, whilst retaining the accuracy of  $\omega$ B97X-D.

## Chapter 5

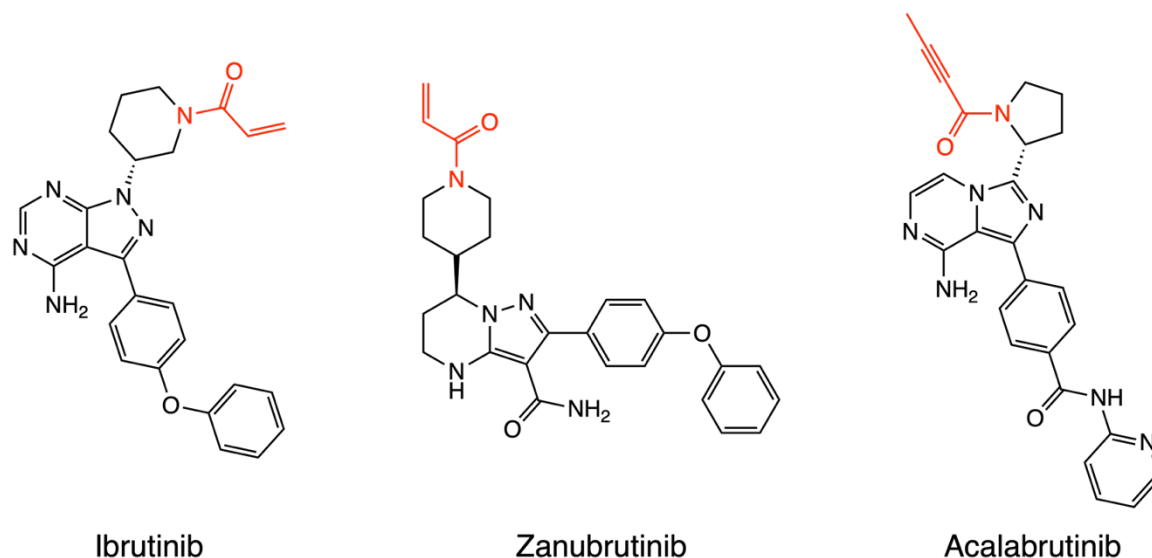
# Investigation of the mechanism of covalent inhibition of Bruton’s tyrosine kinase by ibrutinib using QM/MM calculations

Some of the work contained within this Chapter has formed the basis of a manuscript that was published in RSC Chemical Science on 28<sup>th</sup> January 2021 as a just accepted manuscript.<sup>187</sup>

### 5.1 Introduction

Covalent inhibition is an active area of drug discovery research because of the advantages of pursuing a covalent mechanism of action compared with conventional non-covalent reversible binding. These include complete target blockage, increased selectivity and duration of action.<sup>1,4,35</sup> Recent years have seen the approval of several new marketed covalent drugs targeting protein kinases, owing to their essential roles in protein phosphorylation and cell signalling processes.<sup>17,210</sup> Bruton’s tyrosine kinase (BTK) is a particularly attractive drug target for blood cancers and autoimmune diseases due to its function in signal transduction in the B-cell antigen receptor (BCR) pathway.<sup>280,281</sup> Ibrutinib, acalabrutinib and zanubrutinib are drugs that are approved for the treatment of B-cell cancers including mantle cell lymphoma (MCL) and chronic lymphocytic leukaemia (CLL).<sup>282</sup> Each drug contains an electrophilic Michael acceptor warhead that covalently modify a cysteine residue (C481) in the kinase domain of BTK (Figure 5.1). Utilising warheads of this type to target poorly conserved cysteine residues is a common technique to develop covalent inhibitors in drug discovery.<sup>9</sup> Despite the massive investments made to discover and develop these drugs, the detailed mechanism of covalent binding to BTK is unknown. Understanding the precise mechanism will help in the design of improved covalent drugs targeting BTK, and

other covalent drug targets. Furthermore, the ability to rationally tune covalent reactivity could lead to safer, reversible covalent drugs that have fewer side effects.<sup>6,119</sup>



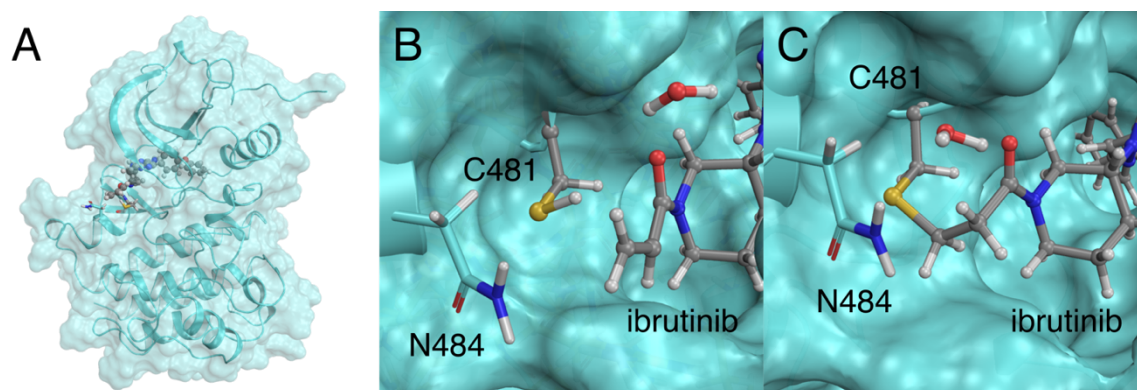
**Figure 5.1** Covalent kinase inhibitors ibrutinib, zanubrutinib and acalabrutinib that target the protein kinase BTK and are approved for the treatment of various B-cell cancers. Ibrutinib and zanubrutinib contain an electrophilic acrylamide warhead, whereas acalabrutinib contains a butynamide warhead. The electrophilic warhead in each drug is highlighted in red.

The addition of a Michael acceptor warhead, such as an acrylamide group, to a thiol side chain is typically modelled in three steps. First, deprotonation of the cysteine thiol occurs to form a thiolate anion, followed by nucleophilic attack of the thiolate on the electrophile to form an enolate intermediate, and finally re-protonation of the enolate to form a covalent thiol-adduct.<sup>127</sup> Krenske *et al.* used quantum mechanics (QM) calculations to study the fundamentals of this addition reaction. They used small molecules containing the electrophilic warhead and a simplified methyl thiolate cysteine-mimic. The calculations were performed in a continuum solvent and using density-functional theory (DFT). They found the rate determining step to be carbon-sulfur bond formation, with a barrier of 14 kcal mol<sup>-1</sup> for the addition of methyl acrylate to methyl thiol.<sup>120</sup> Sulfur reactivity has also been studied in proteins using combined Quantum Mechanics/Molecular Mechanics (QM/MM) methods. Effort has focused on cysteine proteases and protein kinases, given their function in disease processes. The reaction between several covalent nitrile inhibitors and the cysteine protease rhodesain has been investigated using QM/MM simulations at the semi-

empirical PM6 level.<sup>72</sup> The protocol was found to be a useful predictor of reversible covalent binding affinity and was in good agreement with experimental data. The mechanism of covalent modification of C797 by an acrylamide warhead in epidermal growth factor receptor (EGFR) kinase has been elucidated by QM/MM modelling at the self-consistent-charge density-functional-based tight-binding (SCC-DFTB) level.<sup>74,116</sup> These results identified an active site aspartate residue, D800, in the i+3 position relative to C797 that acts as a catalytic base to deprotonate the cysteine thiol. Desolvation of the thiolate anion prior to nucleophilic attack was found to be an important reactivity determinant.

The ATP binding pocket where ibrutinib binds in BTK contains an asparagine (Asn) residue, N484, rather than an aspartate (Asp) or equivalent proton acceptor in the important i+3 position (Figure 5.2). When comparing kinases with cysteine at the same position as C481, different  $pK_a$ -modifying amino acids are to be found at the i+3 position, with the majority being either Asp or Asn.<sup>110</sup> It is understood the surrounding hydrophobic microenvironment can lead to a less reactive and therefore higher cysteine  $pK_a$ .<sup>39</sup> The acidic  $pK_a$  of a free cysteine thiol in solution is 8.6.<sup>283</sup> A nearby more acidic i+3 sidechain such as Asp leads to an upshift in cysteine  $pK_a$ . The  $pK_a$  of C797 in EGFR (which contains an i+3 Asp residue), was estimated to be 11.1  $pK_a$  units in a recent study.<sup>65</sup> Due to its hydrogen bonding capability and neutral sidechain, the same study has shown the i+3 Asn in BTK results in a slightly more acidic Cys  $pK_a$  of 10.4, yet it would still be majority neutral at physiological pH. Unfortunately, there are very few experimentally measured protein cysteine  $pK_a$ 's and no reported experimental  $pK_a$  for C481 in BTK. A  $pK_a$  value of 5.5 has been measured for C797 in EGFR from titration experiments.<sup>164</sup> Although this is a lower  $pK_a$  than otherwise expected given the acidic microenvironment surrounding C797 in EGFR, the stable thiolate has been ascribed to the stabilising effect of the dipole created by the residues comprising the  $\alpha$ -helix on which C797 resides.<sup>164</sup> Some doubt has been cast on this value from the molecular modelling community as the MM forcefields should partially account for the

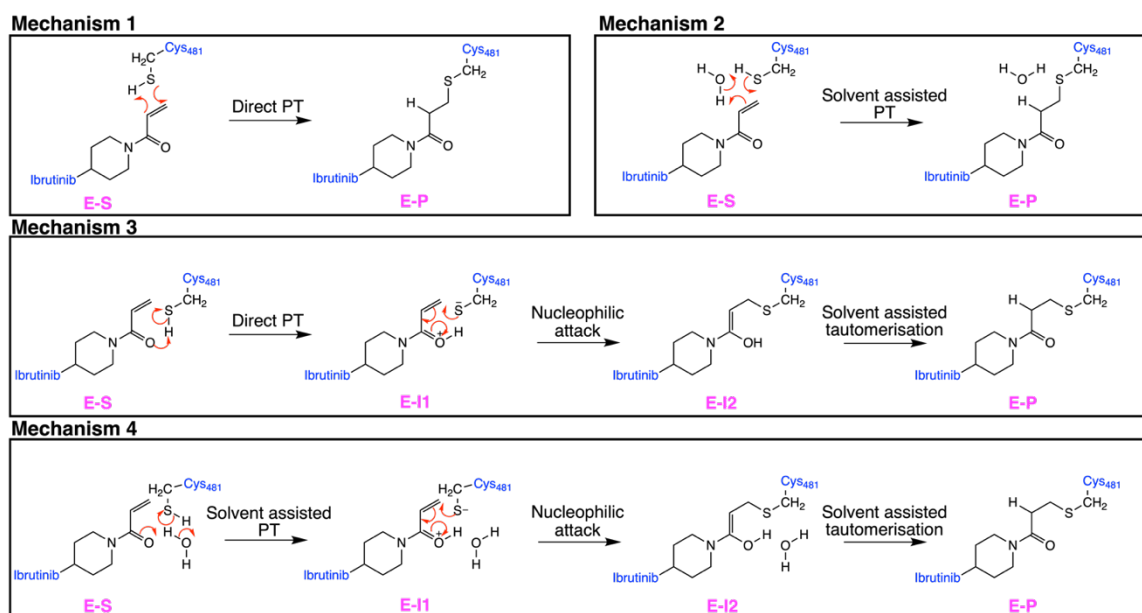
stabilising  $\alpha$ -helix dipole.<sup>65</sup> Clearly there is a substantial difference in cysteine  $pK_a$  prediction between calculation and experiment that warrants further investigation. The topic of cysteine thiol  $pK_a$  is covered extensively in Chapter 3 of this thesis, where the  $pK_a$  of the C481 thiol side chain was calculated to be 13.2 using explicit solvent constant pH molecular dynamics simulations. The fact still remains that the i+3 asparagine in BTK is likely to lead to an upshift in cysteine  $pK_a$  compared with solution and that there is unlikely to be a significant proportion of thiolate at physiological pH. Although Asn performs a catalytic role in some enzymes, for example in GTPase activation<sup>18</sup> and in some protein splicing reactions<sup>19,20</sup>, the amide side chain of Asn is only weakly basic and is therefore unlikely to perform a proton accepting role analogous to the Asp in EGFR. It has been shown that Asn is capable of performing such a role, but only in imine form when it is bound to a metal ion.<sup>173</sup> There are no metal ions, or other residues in the vicinity of Asn in BTK to allow the formation of a basic imine that could then act as a base to participate in covalent modification.



**Figure 5.2** Binding mode of ibrutinib in the active site of BTK. The acrylamide warhead is positioned in close proximity to C481, but with an asparagine in the i+3 position rather than an aspartate residue as in EGFR, it is unclear what the precise mechanism of covalent inhibition is in BTK.

The absence of a proximal acidic/basic residue and the high  $pK_a$  of C481 suggests ionisation is unlikely. This leaves the opportunity for a direct attack mechanism between Michael acceptor warheads and the C481 thiol side chain in BTK, of which there are 4 possible pathways. One possible pathway (Mechanism 1, Figure

5.3) is the direct transfer of the thiol proton to the  $\alpha$ -carbon of the acrylamide warhead to afford the covalently bound keto adduct in a single step. Alternatively, this pathway could occur in a solvent assisted manner where the thiol proton transfers to the  $\alpha$ -carbon of the inhibitor via a nearby water molecule (Mechanism 2, Figure 5.3). A multi-step mechanism is also possible, whereby the thiol proton transfers to the carbonyl oxygen atom of the acrylamide inhibitor to form an enol intermediate (Mechanism 3, Figure 5.3). This could also occur in a solvent assisted pathway, where a water molecule acts as a proton shuttle, analogous to the Asp residue in EGFR (Mechanism 4, Figure 5.3). Mechanisms 3 and 4 are then proceeded by a tautomerisation step to yield the keto product, which is likely to be solvent assisted as direct keto-enol tautomerisation pathways are generally higher in energy than water assisted pathways.<sup>284</sup> To determine the reaction mechanism of covalent addition of ibrutinib to BTK, all four pathways were studied in detail with a QM/MM umbrella sampling protocol, which is outlined below.



**Figure 5.3** Possible covalent inhibition pathways in BTK in the absence of an appropriate amino acid residue that can deprotonate the cystine thiol. Mechanism 1 has been previously studied for a model system comprising methylvinyl ketone,<sup>231</sup> and Mechanism 3 has been studied in the covalent binding of microcystins to cysteine residues.<sup>201</sup>

## 5.2 Methods

A QM/MM umbrella sampling protocol was used to explore mechanisms 1-4 (Figure 5.3) in the kinase domain of BTK. The justification for the umbrella sampling protocol used in this mechanistic study is explained in detail in Chapter 4 of this thesis, and full details of the system set-up and molecular dynamics settings can be found in Chapter 3, Section 3.2. The structural coordinates of the covalently bound BTK/ibrutinib complex were taken from PDB 5P9J,<sup>99</sup> and the mechanistic pathways were followed in the reverse direction. All energies are quoted assuming the reaction between BTK and ibrutinib proceeds in the forwards direction. For some of the reverse pathways, difficulty with obtaining a productive reaction pathway was encountered and so, in these rare cases, the reaction was investigated in the forwards direction starting from the reactant complex (ibrutinib non-covalently bound to BTK. Structural coordinates were taken from PDB 5P9I,<sup>99</sup> and the same procedure outlined in Chapter 3, Section 3.2 was used to set up and carry out the classical MD simulations. The reactant complex snapshots for the reaction simulations were selected based on the definition of a reactive conformation. This conformation included an S-C distance that was below 3.5 Å and a cysteine orientation where the S-H group pointed towards the inhibitor. Optimal cysteine orientation was characterized by a dihedral angle between the N, CA, CB and S atoms of C481 between 20° and 120°. For the pathways involving solvent assistance, snapshots containing a water molecule in close proximity to the covalent warhead was also required.

Briefly, the umbrella sampling protocol consisted of generating a full free energy (FE) surface at the DFTB3 level with 25ps of sampling per umbrella sampling window to get an approximate minimum energy pathway (MEP) for the reaction. The first 5 ps of sampling was discarded as equilibration. A further 10 ps of sampling, giving a total of 30 ps, was carried out along the MEP. Some of the FE surfaces presented were intended to be exploratory surfaces, and so less sampling was performed. This has been indicated in the appropriate places in the following sections.

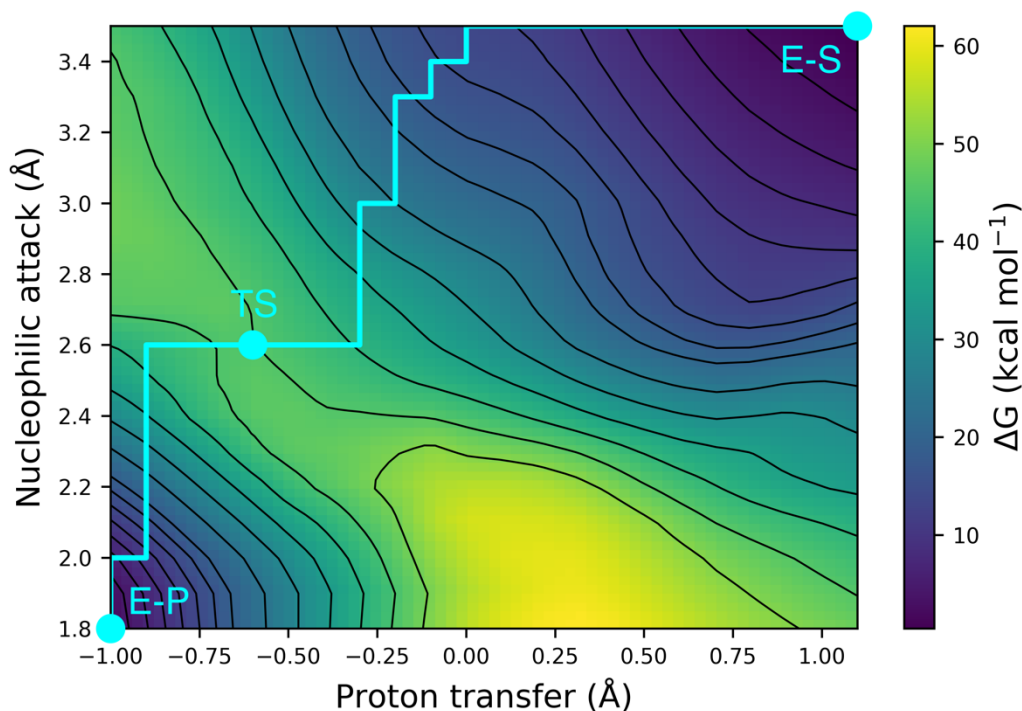
## 5.3 Results

### 5.3.1 Mechanism 1: direct thiol addition and covalent keto product formation

One mechanistic possibility for addition of a Michael acceptor warhead to C481 in BTK is a direct addition where the thiol proton transfers directly to the  $\alpha$ -carbon of the Michael acceptor warhead (Figure 5.3, Mechanism 1). We investigated this pathway according to the computational procedure outlined above to produce a FE surface for the reaction (Figure 5.4) starting from the covalently bound BTK/ibrutinib complex. Reaction coordinate (RC) 1 was defined as the proton transfer of the thiol hydrogen atom to the  $\alpha$ -carbon of the inhibitor ( $d[\text{H}_{\text{C481}}-\text{C}\alpha_{\text{ibrutinib}}] - d[\text{S}_{\text{C481}}-\text{H}_{\text{C481}}]$ ) and RC2 represents the sulfur-carbon bond distance between the thiol group and the  $\beta$ -carbon of the inhibitor ( $d[\text{S}_{\text{C481}}-\text{C}\beta_{\text{ibrutinib}}]$ ). The minimum energy pathway suggests a concerted reaction between the proton transfer and nucleophilic attack reaction coordinates, with a reaction barrier of 47.7 kcal mol<sup>-1</sup>, and a covalent complex energy of 6.4 kcal mol<sup>-1</sup> higher than the unbound state.

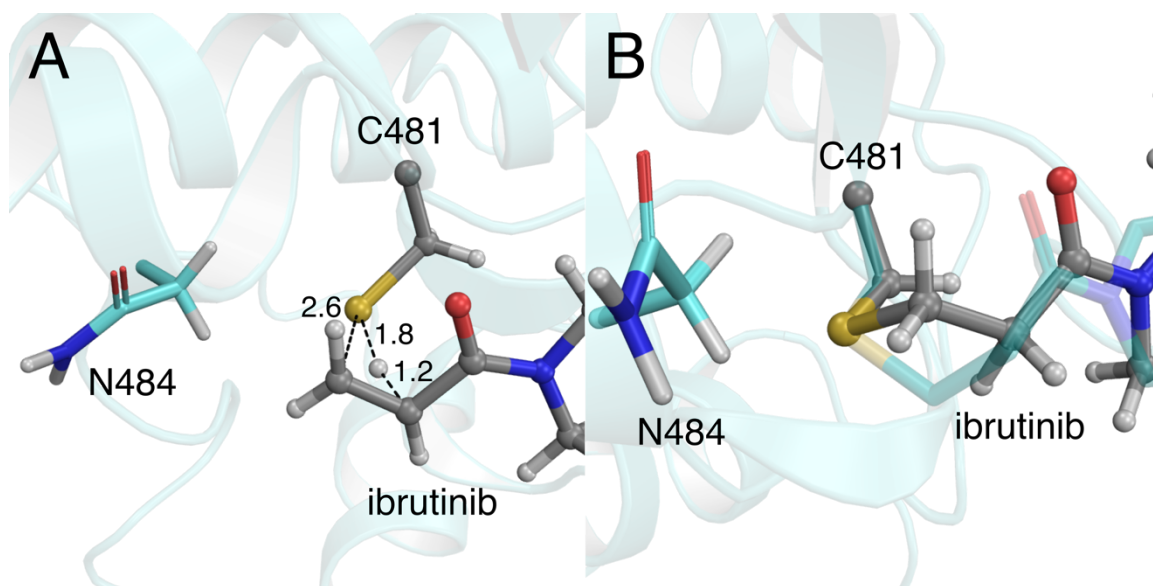
The energetics and geometries of this pathway are consistent with gas phase TS optimisation and IRC calculations carried out in the gas phase on a model system, based upon truncating the QM region defined in the umbrella sampling simulations. This reaction is an example of an electrocyclic reaction, a type of pericyclic reaction where one pi-bond is converted to a sigma-bond by passing through a cyclic transition state.<sup>285</sup> According to the generalized Woodward-Hoffman selection rules for pericyclic reactions,<sup>248</sup> this is an example of a reaction consisting of  $4\pi$  electrons and requires the reaction to proceed through conrotation to give optimal orbital overlap if it is thermally allowed. However, this reaction actually proceeds in a disrotatory fashion and is therefore thermally forbidden and thus explains the high reaction barrier.





**Figure 5.4** FE surface produced at the DFTB3 level, with 30 ps of sampling per window along the MEP for the direct addition of ibrutinib to C481 in BTK. The high reaction barrier of 47.7 kcal mol<sup>-1</sup> is consistent with the reaction being thermally forbidden.

A recent study from Awoonor-Williams *et al.* who looked at this direct addition mechanism for thiol addition also found a high barrier of 65.2 kcal mol<sup>-1</sup> for the reaction at the CCSD(T)/aug-cc-pVTZ//ωB97X-D/aug-cc-pVTZ level of theory consistent with the reaction being thermally forbidden.<sup>231</sup> Furthermore, the proton has to travel a large distance of approximately 4 Å in the starting snapshots analysed, indicating a sub-optimal geometry that hinders this reaction taking place, as well as inefficient orbital overlap. As a result, this mechanism is highly unlikely to be operative for thiol addition between ibrutinib and C481 in BTK and was not investigated further.

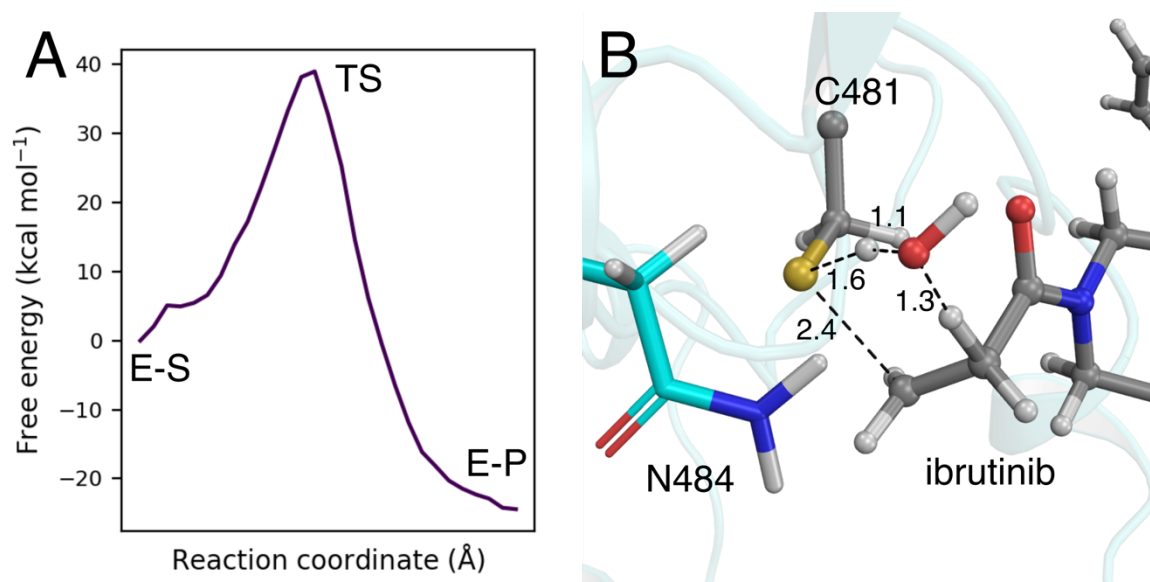


**Figure 5.5** (A) Approximate transition state, taken from the final frame of the highest energy reaction coordinate window ( $RC1 = 2.6 \text{ \AA}$ ,  $RC2 = 0.6 \text{ \AA}$ ) along the reaction pathway of Mechanism 1. (B) Representative snapshot of the covalently bound BTK/ibrutinib keto product (EP) from Mechanism 1. The geometry around the newly formed S-C bond differs from the crystal geometry (PDB 5P9J,<sup>99</sup> shown in transparent green), and results in a high energy product state.

### 5.3.2 Mechanism 2: solvent assisted thiol addition and covalent keto product formation

A solvent assisted pathway where the C481 thiol proton transfers to the  $\alpha$ -carbon of the acrylamide warhead of ibrutinib (Figure 5.3, Mechanism 2) is potentially a more feasible pathway than the direct addition (Mechanism 1) discussed in Section 5.3.1. This is based on the assumption that the thiol proton has less distance to travel to the  $\alpha$ -carbon of the acrylamide group in ibrutinib if it can be assisted via a water molecule, and the reaction would not have to proceed via a strained four-membered ring transition state, as is the case for Mechanism 1. Instead, Mechanism 2 would proceed through a more stable pseudo six-membered ring transition state to result in the covalently bound keto product. To investigate Mechanism 2, a 2D QM/MM FE surface for the reaction was produced. The first reaction coordinate described the proton transfer from ibrutinib to water and S-C formation as a linear combination of three distances. There are, therefore, six atoms that define this reaction coordinate

( $d[\text{O}_{\text{wat}}\text{-H}_{\text{ibrutinib}}] - d[\text{C}\alpha_{\text{ibrutinib}}\text{-H}_{\text{ibrutinib}}] - d[\text{S}_{\text{C481}}\text{-C}\alpha_{\text{ibrutinib}}]$ ). The second reaction coordinate was chosen to describe the proton transfer between the thiol group in C481 and a nearby water molecule ( $d[\text{S}_{\text{C481}}\text{-H}_{\text{wat}}] - [\text{O}_{\text{wat}}\text{-H}_{\text{wat}}]$ ).



**Figure 5.6** (A) Approximate FE profile for Mechanism 2, where solvent assisted thiol addition occurs to result in the keto product. The approximate FE barrier is 39.0 kcal mol<sup>-1</sup> and 2 ps of sampling was carried out in reach umbrella sampling window. (B) Approximate transition state for Mechanism 2. The structure is the final frame from the highest energy reaction coordinate window along the reaction path.

The transformation of the covalently bound BTK/ibrutinib adduct to the non-covalently bound reactant state described by Mechanism 2 did not result in a productive reaction pathway. Only a pathway along the approximate diagonal of the FE surface resulted in the expected reaction pathway. When the remainder of the FE surface was sampled with QM/MM umbrella sampling, several spurious proton transfers occurred. Some of these additional proton transfers resulted in the formation of the enol product, where a proton transfer from the developing hydronium ion to the carbonyl oxygen of the inhibitor was observed. Simulations with harmonic restraints added to the system to stop these unwanted proton transfers from occurring still resulted in the formation of the enol product or led to fragmentation of the inhibitor. The additional proton transfers are not explicitly sampled in the generation of the FE

surface, and so the energetics of these additional processes are not captured, and the full FE surface has therefore not been shown. The approximate FE barrier can be estimated from the approximate diagonal pathway that assumes a purely concerted reaction path for this reaction and is 39.0 kcal mol<sup>-1</sup> (Figure 5.6).

The difficulty in producing a productive reaction pathway could be the result of the limitations of using a simple linear combination of multiple distances for such a complex reaction coordinate that involves the cleavage of two bonds and the formation of two bonds. A significant finding from the investigation into Mechanism 2 is that there appears to be a strong preference for the reaction to proceed via an enol intermediate. The developing negative charge that forms in the acrylamide group as the S-C bond forms is more likely to be centered on the carbonyl oxygen than the  $\alpha$ -carbon of the inhibitor, owing to the greater electronegativity of the oxygen atom. This has the effect of attracting the proton of the water molecule to the carbonyl oxygen instead of the  $\alpha$ -carbon. Furthermore, visual inspection of the geometry of the TS (Figure 5.6) shows that the distorted 6-membered ring that forms in the TS is strained, likely resulting in the high barrier for the reaction.

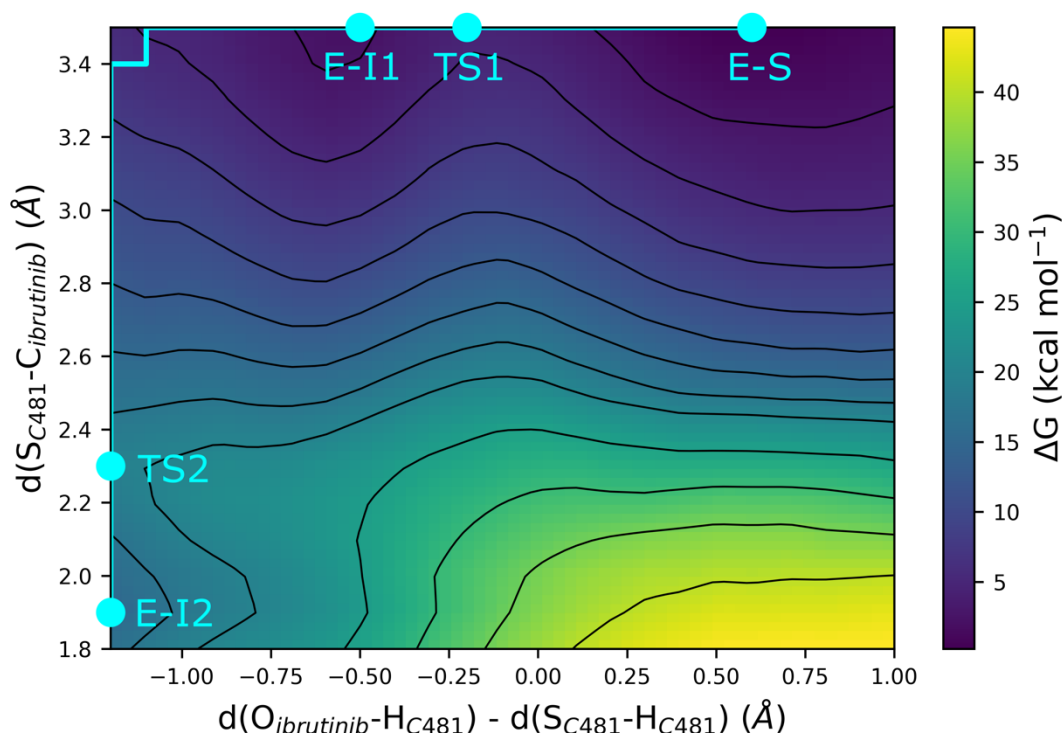
A stepwise variant of the pathway described by Mechanism 2 is possible, where the thiol proton transfers from the sulfur atom of C481 to a nearby water molecule to form H<sub>3</sub>O<sup>+</sup> and a negatively charged thiolate anion. This would result in a hydronium/thiolate ion pair analogous to cysteine thiolate/His<sup>+</sup> ion pair that make up the catalytic dyad in cysteine proteases.<sup>263,264</sup> Although a proton transfer of the thiol proton to a nearby water molecule was attempted, the instability of the H<sub>3</sub>O<sup>+</sup> resulted in the transfer of a proton from the H<sub>3</sub>O<sup>+</sup> to the carbonyl oxygen atom of the inhibitor. The apparent lack of a stable ion pair is not that surprising for BTK. Although stable Cys-S-/His-NH<sup>+</sup> ion pairs are routinely observed in reactions catalyzed by cysteine proteases, these enzymes have evolved to promote reactivity of the cysteine residue based on their neighboring sidechains. This is of course a different

scenario to BTK reactivity, where an exogenous drug molecule targets a non-catalytic cysteine residue. In a more similar scenario, although not an ion pair, a stable Cys-S<sup>-</sup>/Asp-COOH pair was discovered for EGFR kinase that was predicted to be more stable than the Cys-SH/Asp-COO<sup>-</sup> pair. The likely difference between the EGFR system and BTK is the increased stability of the neutral carboxylate compared with the hydronium ion in the BTK system, and the geometric and entropic restraint of a proximal sidechain compared to solvent. Inspection of the reaction center and its immediate microenvironment suggests that there is no opportunity for stabilisation of the thiolate/hydronium ion pair. The edge of the ATP binding pocket where C481 is situated in BTK is reasonably solvent exposed. Whilst the i+3 N484 residue is close enough to provide some stabilisation of the ion pair, the amide side chain of N484 cannot get close enough to the thiolate as it is hindered by the  $\beta$ -carbon of the inhibitor and thus does not provide any stabilisation of the TS in Mechanism 2 (Figure 5.6). The results of the investigation of Mechanism 2 indicate that solvent assisted thiol addition and keto product formation is an unlikely pathway for covalent alkylation of C481. This is indicated by the high reaction barrier of 39.0 kcal mol<sup>-1</sup> for the approximate minimum energy pathway and the apparent preference for the reaction to proceed via a more stable pathway resulting in an enol intermediate.

### 5.3.3 Mechanism 3: direct thiol addition and enol product formation

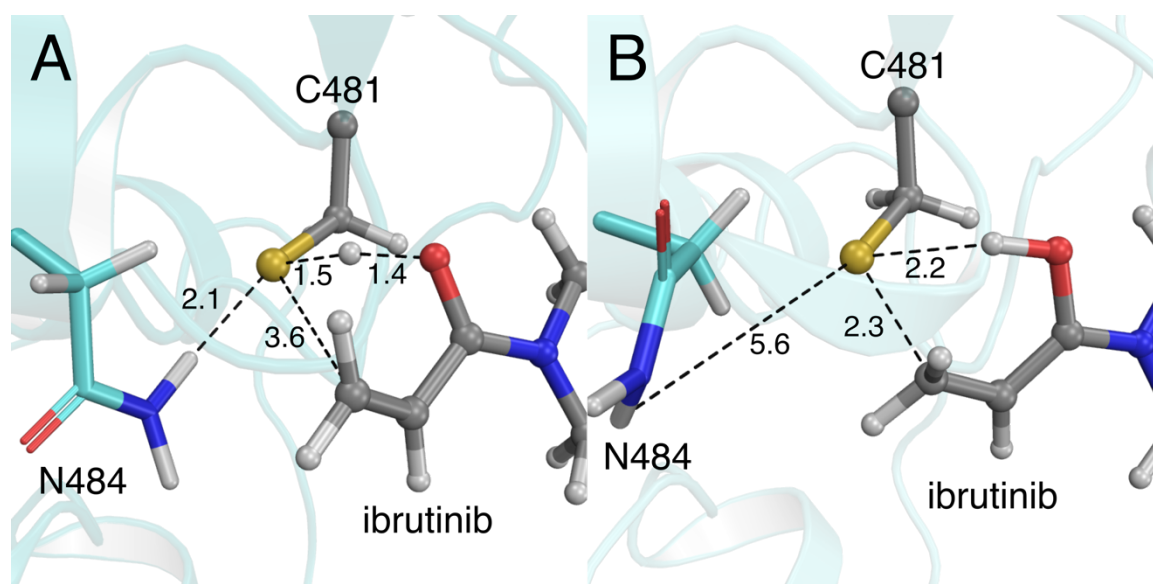
The possibility of the reaction mechanism of BTK inhibition by ibrutinib occurring via a direct proton transfer from the C481 thiol group to the carbonyl oxygen atom in ibrutinib to result in an enol intermediate (Figure 5.3, Mechanism 3) was explored with a 2D FE surface. The first reaction coordinate defined the proton transfer from the thiol side chain of C481 directly to the carbonyl oxygen atom in ibrutinib ( $d[\text{H}_{\text{C481-O}_{\text{ibrutinib}}}] - d[\text{S}_{\text{ibrutinib-H}_{\text{ibrutinib}}}]$ ), and the second reaction coordinate defined the nucleophilic attack step ( $d[\text{S}_{\text{C481-C}\alpha_{\text{ibrutinib}}}]$ ). The FE surface (Figure 5.7) for this reaction indicates that the carbon-sulfur bond formation and proton transfer steps proceed in a stepwise fashion, with the reaction proceeding via an intermediate that

corresponds to stable ion pair formed between the negatively charged cysteine thiolate and the protonated oxygen in the acrylamide warhead (EI1, Figure 5.3). The barrier to the initial PT step and formation of the EI1 reaction intermediate is 4.6 kcal mol<sup>-1</sup>, and EI1 lies 2.0 kcal mol<sup>-1</sup> higher in energy than the reactant complex. Unrestrained classical MD of EI1 confirms that it does not collapse and remains stable after removal of the umbrella sampling restraints. The overall reaction barrier is 17.6 kcal mol<sup>-1</sup>, corresponding to a product-like TS that is formed during carbon-sulfur bond formation, and after the initial PT has occurred. This is consistent with other studies that have predicted the rate limiting step in thiol addition to correspond to S-C formation.<sup>74,120</sup> The enol product lies 15.0 kcal mol<sup>-1</sup> higher in energy than the non-covalent BTK/ibrutinib reactant complex (ES, Figure 5.7).



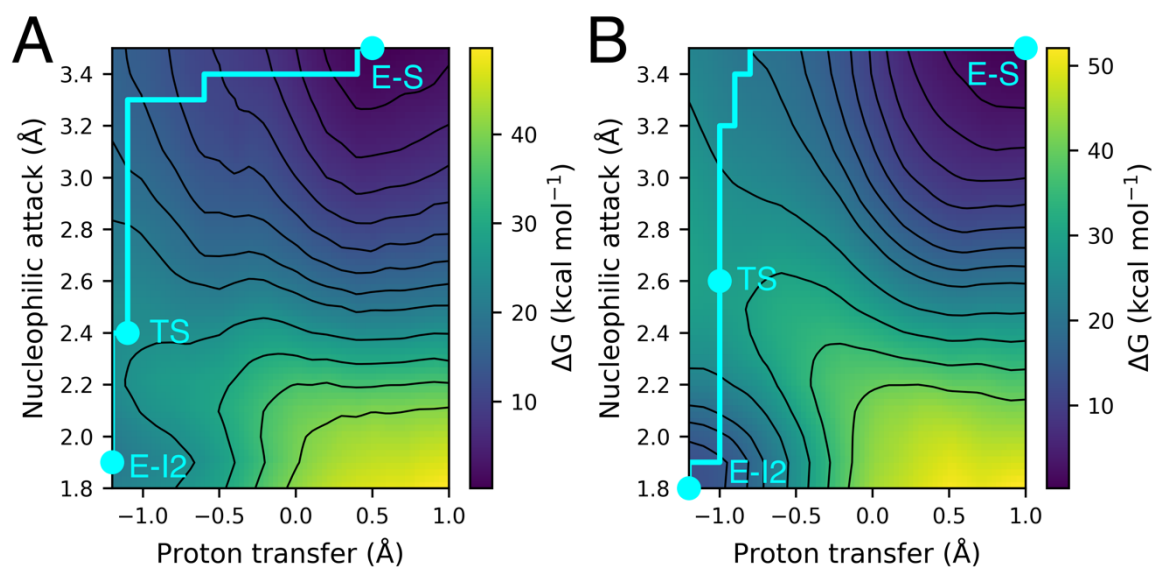
**Figure 5.7** FE surface of the S-C bond formation step and the solvent assisted tautomerisation step at the DFTB3/MM level, with 10 ps of sampling in each umbrella sampling window. The reaction proceeds in a stepwise fashion, through a reaction intermediate EI1. The barrier for the proton transfer step (TS1) is 4.6 kcal mol<sup>-1</sup> and the barrier to S-C bond formation (TS2) is 17.6 kcal mol<sup>-1</sup>.

Gas phase modelling of the reaction between methyl thiolate and 1-(piperidin-1-yl)prop-2-en-1-one (Chapter 4, Figure 4.8) indicates a synchronous reaction pathway between the proton transfer and S-C bond formation steps, with no stable minimum corresponding to a protonated carbonyl oxygen. However, this charged intermediate is unlikely to be stable in the gas phase, as a crucial component of its stability is a hydrogen bonding interaction that forms between the negatively charged C481 thiolate and the amide side chain of the i+3 asparagine residue, N484. Even when the reaction is carried out in BTK, a stable intermediate corresponding to EI1 is not observed if the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> does not form. (Figures 5.8 and 5.9). The formation of the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interaction is dependent on the orientation of the N484 residue. When QM/MM reaction simulations starting from the reactant state are performed for Mechanism 3, the N484 residue is oriented away from C481 and the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interaction does not form and no stable intermediate is predicted as a result. However, when the reaction is investigated in the reverse direction, the N484 residue is oriented towards the C481 residue, thus allowing an interaction to form with the thiolate of C481 (Figure 5.8).



**Figure 5.8** (A) Approximate TS from DFTB3/MM umbrella sampling simulations showing the stabilizing interaction between the amide side chain of N484 and the cysteine thiolate of C481 that results in a stable intermediate (EI1). (B) Approximate TS from DFTB3/MM umbrella sampling simulations showing the absence of the stabilizing N484 interaction which does not lead to a stable EI1 intermediate.

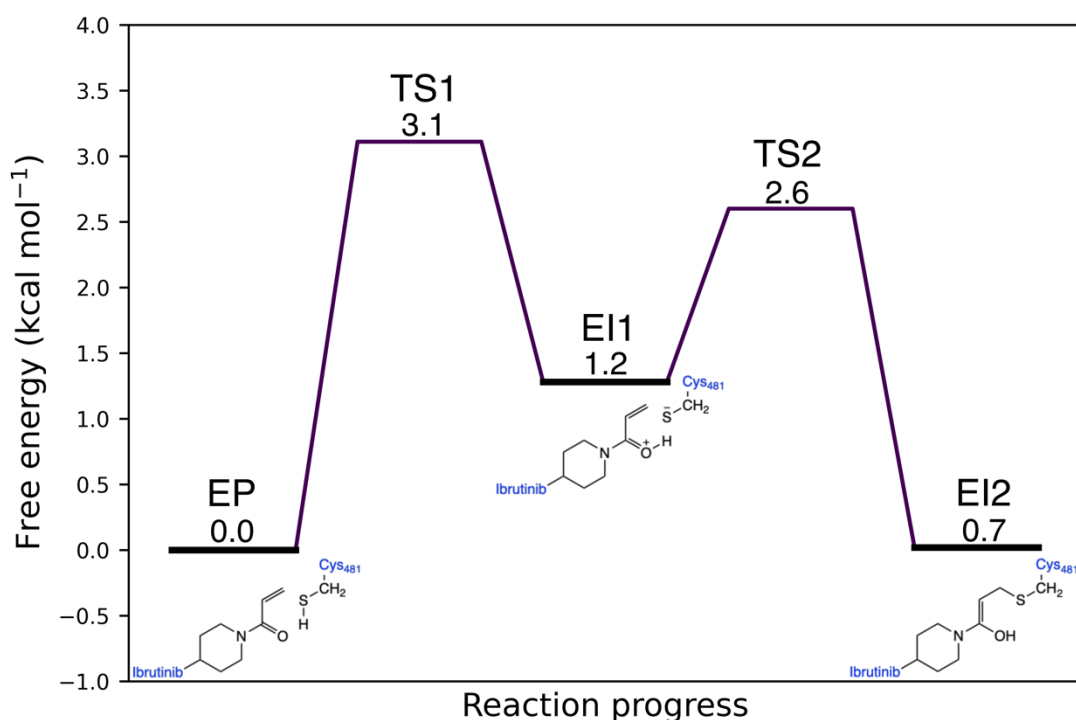
In order to confirm that the formation of the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interaction is dependent on the orientation of the amide side chain of N484 rather than the QM method used, FE surfaces were produced at the DFTB3/MM and  $\omega$ B97X-D/6-31G(d)/MM levels of theory. The starting points for both simulations were a single snapshot where the N484 was oriented away from C481 and the reaction was modelled in the forwards direction (from ES to EI2). Comparison of the FESs produced at the DFTB3/MM and  $\omega$ B97X-D/6-31G(d)/MM levels of theory reveals that no stable EI1 intermediate is predicted by either method (Figure 5.9) and no C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interactions were observed on examination of the reaction trajectories. The barrier to enol formation is 23.9 and 29.2 kcal mol<sup>-1</sup> for DFTB and  $\omega$ B97X-D respectively. The reaction barriers are in good agreement with the same reaction carried out in the reverse direction (EI2 to ES,  $\Delta G^\ddagger=17.6$  kcal mol<sup>-1</sup>), when the N484 interaction with C482 results in a reactive intermediate, EI1. The presence of the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interaction clearly provides a small amount of transition state stabilization and results in a lower barrier.



**Figure 5.9** FE surfaces for the forwards reaction between C481 and the acrylamide warhead of covalent drug ibrutinib at the DFTB3/MM level (A) and  $\omega$ B97X-D/6-31G(d)/MM (B) with 2 ps of sampling in each reaction coordinate window. The barrier heights for the reaction are 23.9 kcal mol<sup>-1</sup> and 29.2 kcal mol<sup>-1</sup> for the DFTB/MM and  $\omega$ B97X-D/6-31G(d) surfaces respectively. DFTB3 underestimates the reaction barrier by approximately 5 kcal mol<sup>-1</sup> compared to  $\omega$ B97X-D/6-31G(d).



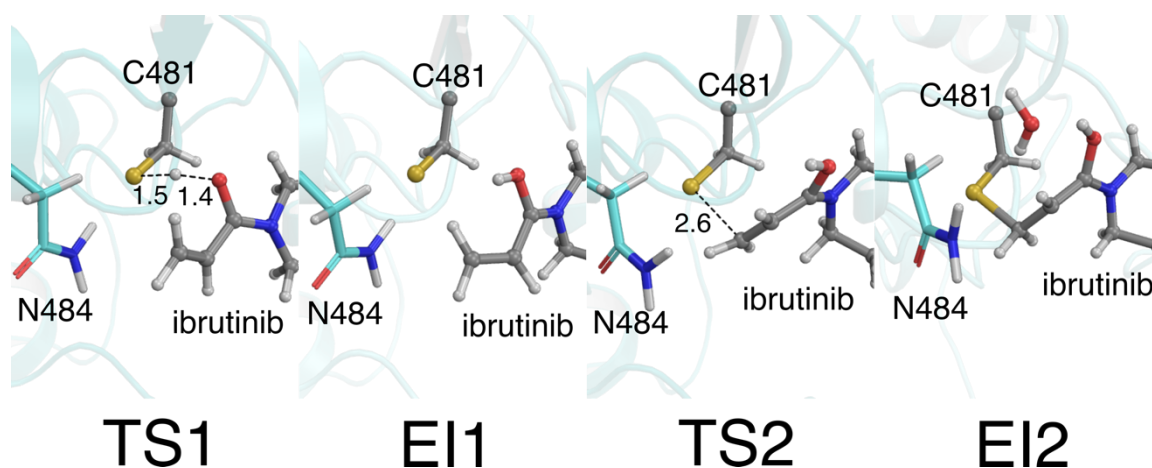
The FE surfaces presented in Figure 5.9 also highlight how similar the reaction pathways are at the DFTB3/MM and  $\omega$ B97X-D/6-31G(d)/MM levels of theory, consistent with the benchmarking calculations performed in Chapter 4 of this thesis and confirm the suitability for investigating mechanistic pathways in BTK. Furthermore, it shows that DFTB3 underestimates the reaction barrier by approximately 5 kcal mol<sup>-1</sup> with respect to  $\omega$ B97X-D/6-31G(d).



**Figure 5.10** Summary of the FE of intermediates and TSs along the pathway of Mechanism 3, when calculated as a series of 1-D stepwise reaction coordinates at the DFTB3/MM level from umbrella sampling simulations where 30 ps of sampling was carried out in each umbrella sampling window. Reaction progress describes the covalent reaction from the non-covalently bound BTK/ibrutinib complex (E-S) to the covalently bound enol intermediate (EI2).

The FE surfaces for Mechanism 3 predict a reasonably high energy barrier to reaction, and a high energy content of the enol intermediate EI2 for the reaction with respect to ES. This is the case regardless of whether the C481-S<sup>-</sup>/Asn-NH<sub>2</sub> interaction forms or not. When the reaction is modelled in a stepwise manner, the energies of the transition states and intermediates along the pathway are reduced significantly (Figure

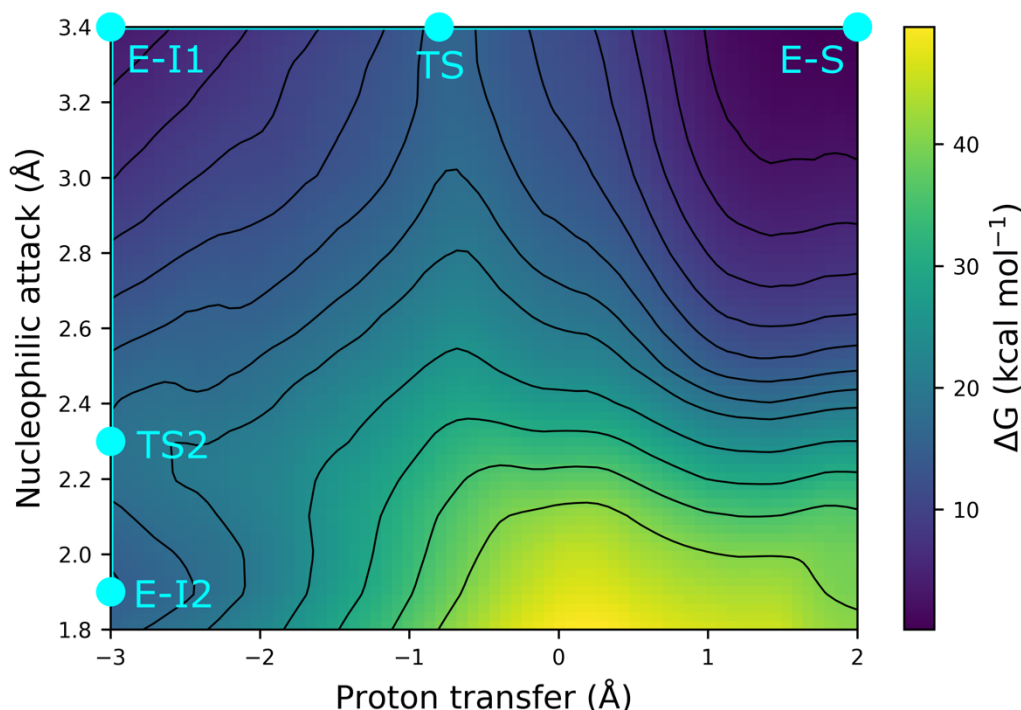
5.10). Removal of the umbrella sampling restraints for the S-C step when the proton transfer from the thiol to the carbonyl oxygen of the inhibitor takes place affords a modest reduction of 1.5 kcal mol<sup>-1</sup> in the FE penalty required to form EI1 from ES. However, a significant reduction in the barrier to S-C formation is observed when the umbrella sampling restraints for the proton transfer are released. This allows for greater flexibility in the acrylamide group during S-C formation, including a rotation around the acrylamide C<sub>carbonyl</sub>-C<sub>α</sub> bond as S-C formation occurs that results in an *E* configuration in the newly formed enolic double bond in EI2 (Figure 5.11). This is in contrast to the concerted mechanism that results in a higher energy *Z* configuration around the enolic double bond in the TS corresponding to S-C formation (Figure 5.8(B)) and the resulting EI2 intermediate. The more favorable geometries adopted by ibrutinib and the thiol side chain of C481 in the stepwise pathway result in the barrier of the S-C formation step being dramatically reduced from 17.6 to 2.6 kcal mol<sup>-1</sup>, and the energy of the enol intermediate EI2 being reduced from 15.0 to 0.7 kcal mol<sup>-1</sup>. This makes Mechanism 3 the lowest energy pathway to enol product formation, and the most feasible reaction mechanism considered so far for the covalent modification of C481 by ibrutinib.



**Figure 5.11** Approximate transition state and intermediate structures taken from DFTB3/MM MD simulations along the reaction path of Mechanism 3, when modelled with a stepwise reaction coordinate.

### 5.3.4 Mechanism 4: solvent assisted thiol addition and enol product formation

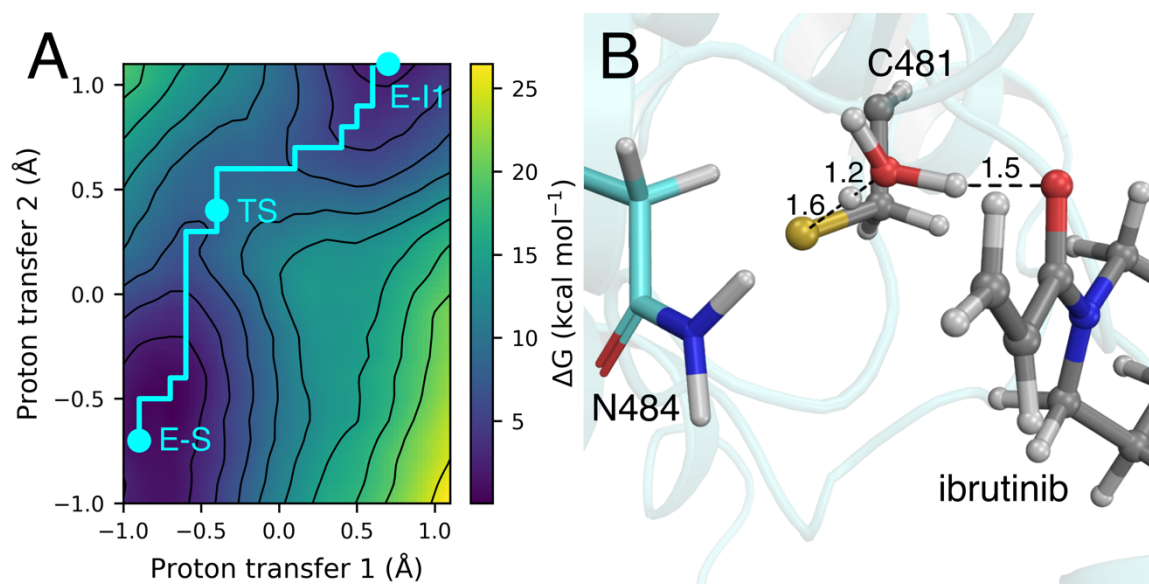
Initially, a FE surface for Mechanism 4 (thiol addition where the thiol proton transfers to the acrylamide carbonyl via a water molecule and S-C formation to form an enol product (Figure 5.3, Mechanism 4)) was modelled on a 2D FE surface. The first reaction coordinate consisted of varying the S-C distance that describes nucleophilic attack ( $d[\text{S}_{\text{C481}}-\text{C}\alpha_{\text{ibrutinib}}]$ ), and the second reaction coordinate defined the simultaneous proton transfer from the thiol to a water, and a second proton transfer from the water to the carbonyl carbon ( $((d[\text{O}_{\text{ibrutinib}}-\text{H1}_{\text{wat}}] - d[\text{H1}_{\text{wat}}-\text{O}_{\text{wat}}] - d[\text{S}_{\text{C481}}-\text{H2}_{\text{wat}}] + d[\text{O}_{\text{wat}}-\text{H2}_{\text{wat}}]))$ ). This FE surface (Figure 5.12) suggested that the reaction was stepwise due to the presence of an intermediate (EI1) corresponding to protonation of the carbonyl oxygen atom of the inhibitor prior to S-C formation. This is consistent with the stepwise pathway modelled for Mechanism 3 that also found the same EI1 intermediate.



**Figure 5.12** FE surface to show thiol addition to C481 for Mechanism 4, with 10 ps of sampling performed in each umbrella sampling window along the MEP. RC1 is solvent assisted proton transfer from thiol to carbonyl oxygen in the acrylamide group of ibrutinib and RC2 is

S-C formation. Presence of a shallow energy minimum indicates a two-step reaction. The overall reaction barrier corresponds to proton transfer and is 16.0 kcal mol<sup>-1</sup>.

The barrier corresponding to the first step (proton transfer and formation of the charged intermediate) was calculated to be 16.0 kcal mol<sup>-1</sup>, and the barrier to the second step (S-C bond formation) was calculated to be 14.8 kcal mol<sup>-1</sup> at the DFTB3/MM level of theory. After discovering that the stepwise pathway for Mechanism 3 led to a lower energy reaction pathway, a similar procedure was used to investigate the solvent assisted proton transfer. The starting structure selected was the final MD snapshot corresponding to EI1 from Mechanism 3. This consisted of generating a 2D surface investigating each proton transfer step (Figure 5.13). PT1 corresponds to the proton transfer of the proton from the thiol group of C481 to a nearby water molecule situated in the cleft between the thiol group and the carbonyl group of the inhibitor and PT2 corresponds to the transfer of a proton from the same water molecule to the carbonyl oxygen atom of ibrutinib.



**Figure 5.13** FE surface for the solvent assisted proton transfer from C481 in BTK to the carbonyl oxygen atom of the covalent inhibitor ibrutinib at the DFTB3/MM level. In each reaction coordinate window 30 ps of sampling was performed along the MEP, and the barrier to the proton transfer was 8.3 kcal mol<sup>-1</sup>.

The MEP for the solvent assisted proton transfer process is concerted, with a FE barrier of 8.3 kcal mol<sup>-1</sup> (Figure 5.13). This is significantly lower than the barrier reported for this step when modelled in combination with S-C bond formation ( $\Delta G^\ddagger=16.0$  kcal mol<sup>-1</sup>, Figure 5.9). This barrier is consistent with a previous study that calculated the barrier of solvent assisted deprotonation of a thiol to be 10.6 kcal mol<sup>-1</sup> in 1,3-dihydroimidazole-2-thione at the MP2/6-311++G(d,p)//M06-2X/6-31+G(d) level.<sup>286</sup> The slightly higher barrier to proton transfer in Figure 5.12 compared with modelling the reaction as a stepwise process (Figure 5.13) is the result of an approximate diagonal pathway being followed between both of the PT reaction coordinates, rather than the true MEP for the reaction.

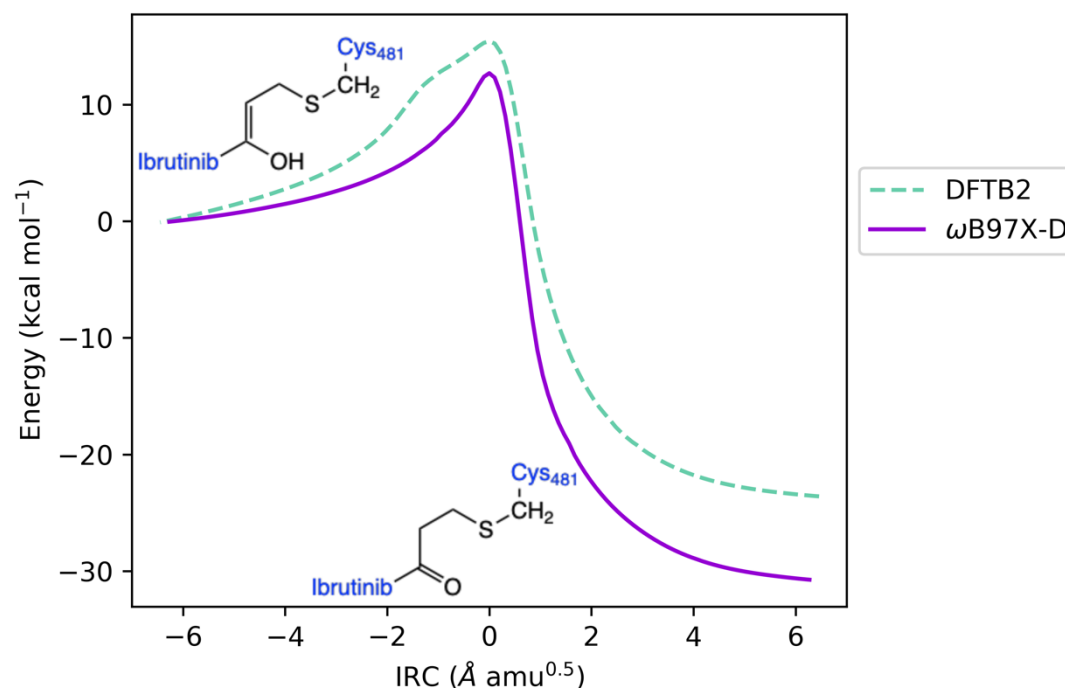
The FE barrier to the solvent assisted proton transfer step is slightly higher than the direct pathway (Mechanism 3) when water is not involved in the reaction. This is due to the high entropic cost of forming the transition state that requires the optimal positioning of a water molecule, the thiol side chain of C481 and the acrylamide warhead of ibrutinib. This makes the water assisted proton transfer pathway less likely to occur along the reaction pathway to enol formation than the direct pathway investigated for this step in Mechanism 3. Solvent assisted mechanisms have been observed in previous reactivity studies involving sulfur.<sup>287</sup> A solvent assisted reaction pathway resulting in enol formation was found to be the most likely mechanism of covalent bond formation between a cysteine residue in a protein phosphatase enzyme and an the  $\alpha,\beta$ -unsaturated moiety of a microcystin.<sup>201</sup> Although several pathways were considered, the pathway involving water was found to be lower in energy and resulted in thermodynamically more stable products when potential energy scans, IRC calculations and transition state optimisations were used to investigate reactivity using the B3LYP density functional in combination with the 6-31G(d) basis set. The solvent assisted enol formation was found to occur in a single step with a barrier of 21.9 kcal mol<sup>-1</sup>, rather than the two-step mechanism modelled in this work. This is likely to be the result of the cluster model used in the microcystin study that contained just 35

atoms comprising the thiol side chain of a cysteine residue and the reactive part of the microcystin. Any potential additional stabilisation from the enzyme environment was therefore neglected. For the reaction between BTK and ibrutinib, stabilisation of the negatively charged thiolate ion is achieved through hydrogen bonding interactions with the amide side chain of the i+3 Asn residue. Although the Asn residue does not directly participate as a proton acceptor in the reaction, it clearly has an essential role in providing stabilisation to the TS structures and intermediates along the reaction pathway in BTK.

### 5.3.5 Solvent assisted tautomerisation

Out of the four pathways considered for covalent modification of C481 in BTK by the acrylamide containing inhibitor ibrutinib, the most viable pathways are the two that proceed via an enol intermediate. A tautomerisation step must occur to complete this pathway and form the keto product, which is the thermodynamic driving force for the reaction. Tautomerisation steps (also referred to as ketonisation steps in the literature) are ubiquitous in many chemical reactions, especially in biochemical systems and have been extensively studied with computational chemistry techniques.<sup>288–291</sup> In general, intramolecular keto-enol tautomerisation reactions are higher in energy than the equivalent reaction when assisted by a water molecule.<sup>284,292</sup> A previous investigation by Paasche *et al.* found that keto-enol tautomerisation in the context of thio-Michael addition is a high energy process and therefore is unlikely to be a viable mechanistic pathway in thiol addition reactions.<sup>199</sup> Unfortunately this study has led other investigations of thiol reactivity to discount tautomerisation pathways as a feasible possibility.<sup>74</sup> However, Paasche *et al.* only considered the keto-enol tautomerisation step to proceed through a high energy enol cation intermediate or enolate intermediate, and neglected the possibility of assistance from solvent. The authors also note that a ketonisation step is possible if catalysed by an efficient water bridge, but they conclude that this is unlikely in an enzyme active site due to water inaccessibility.<sup>199</sup> However, it is feasible for a solvent assisted tautomerisation to occur

in BTK, due to the solvent exposed nature of the edge of the ATP binding pocket in BTK where the reactive C481 resides.



**Figure 5.14** Gas phase IRC calculations of the solvent assisted proton transfer in a model BTK/ibrutinib system consisting of 35 atoms that comprise of methyl thiol, a water molecule and the acrylamide warhead and linker group of ibrutinib. The IRC calculations suggest that there is a low barrier to reaction (9.7 kcal mol<sup>-1</sup> at the  $\omega$ B97X-D/6-31G(d) level) and that DFTB3 overestimates the reaction barrier by 5.8 kcal mol<sup>-1</sup>.

The feasibility of a solvent assisted keto-enol tautomerisation step occurring for ibrutinib covalently bound to C481 in BTK was initially investigated using transition state optimisations followed by IRC calculations in Gaussian16<sup>228</sup> on a model system consisting of 35 atoms. The model comprised of 2-(methylthio)-1-(piperidin-1-yl)ethan-1-one and a single explicit water molecule to mimic ibrutinib bound to a cysteine residue in BTK. A guess TS geometry was built and optimised with a water molecule situated in an optimal position between the carbonyl oxygen atom and the hydrogen attached to the  $\alpha$ -carbon of the acrylamide group in the gas phase (Figure 5.14). The resulting structure was checked for a single imaginary frequency corresponding to the reaction coordinate, and then an IRC calculation was performed to follow the reaction

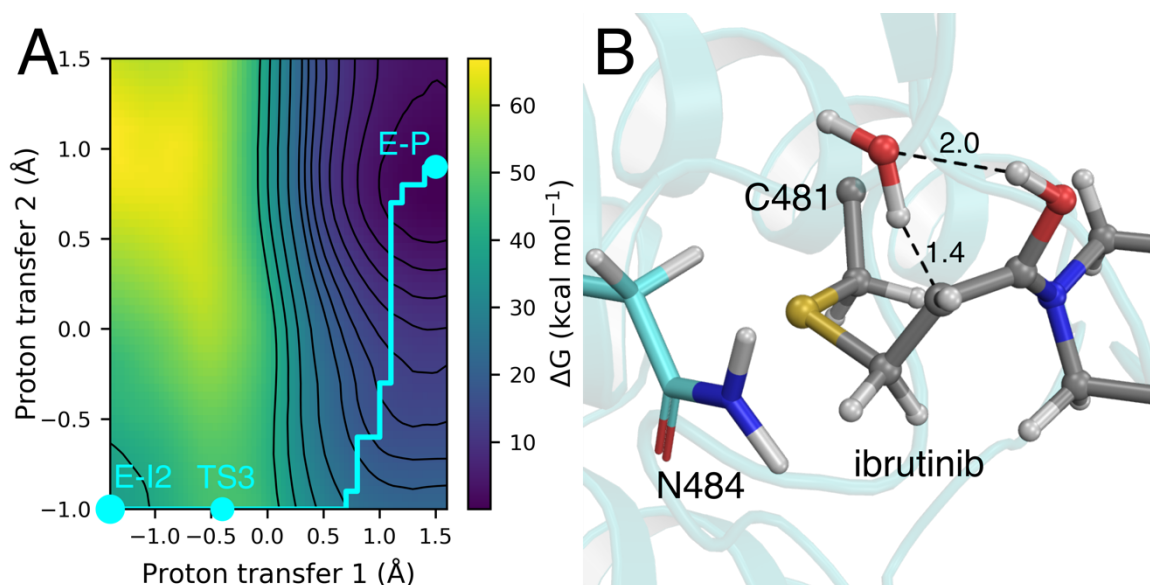
path in both directions to the enol and keto structures. Semi-empirical DFTB2 and the  $\omega$ B97X-D/6-31G(d) density functional were compared as benchmarking calculations performed in Chapter 4 of this thesis suggested that both of these methods performed well for modelling sulfur reactivity.

The results suggest that a solvent assisted tautomerisation is energetically favorable for the formation of the keto product from the enol product at both levels of theory in the gas phase. The keto product is  $-23.5 \text{ kcal mol}^{-1}$  and  $-36.5 \text{ kcal mol}^{-1}$  lower in energy than the enol product for DFTB2 and  $\omega$ B97X-D respectively, highlighting that the formation of the covalently bound keto product is a strong thermodynamic driving force for the reaction. The barrier to keto-enol tautomerisation in this model system is  $15.5 \text{ kcal mol}^{-1}$  and  $9.7 \text{ kcal mol}^{-1}$  for DFTB2 and  $\omega$ B97X-D/6-31G(d), showing that this is clearly a feasible reaction. There is reasonable agreement between the semi-empirical and DFT levels of theory, with DFTB2 actually overestimating the barrier compared with  $\omega$ B97X-D/6-31G(d) by  $5.8 \text{ kcal mol}^{-1}$ . Inspection of the reaction pathway reveals that there are no major discrepancies in geometry between both levels of theory and so the higher barrier predicted by DFTB3 is simply due to error in the DFTB2 energies.<sup>244</sup> The barrier heights and relative stabilities of the enol and keto states are in good agreement with a study that has investigated ketonisation pathways in 4-(m-Chlorobenzylamino)-3-Phenyl-4,5-Dihydro-1H-1,2,4-Triazol-5-One at high levels of QM theory (MP2/6-311++G(d,p)).<sup>284</sup> These results are encouraging, but do not help to answer the question of whether this tautomerisation step can occur in the kinase domain of BTK as geometric constraints and the electrostatic environment in the active site could alter the likelihood of this step occurring.

QM/MM umbrella sampling simulations were used to assess the likelihood of a solvent assisted keto-enol tautomerisation pathway taking place in BTK. The initial geometries were taken from classical molecular dynamics simulations of ibrutinib covalently bound to C481 in BTK. The crystal structure (PDB ID: 5P9J)<sup>99</sup> contains a



water molecule positioned just above the carbonyl oxygen atom ( $d[\text{O}_{\text{ibrutinib}}-\text{O}_{\text{wat}}] = 2.9 \text{ \AA}$ ) and the  $\alpha$ -carbon ( $d[\text{O}_{\text{wat}}-\text{C}\alpha_{\text{ibrutinib}}] = 3.6 \text{ \AA}$ ) of the acrylamide group in ibrutinib. Analysis of a 500ns MD trajectory of ibrutinib covalently bound to BTK to find optimally positioned water molecules for the tautomerisation step showed that a water bridge between the carbonyl oxygen and  $\text{C}\alpha$  in ibrutinib was ideally placed in 7 % of frames. Optimal water placement was defined by simple  $d[\text{O}_{\text{ibrutinib}}-\text{O}_{\text{wat}}]$  and  $d[\text{O}_{\text{wat}}-\text{C}\alpha_{\text{ibrutinib}}]$  distances below  $3.5 \text{ \AA}$ . A suitable snapshot was selected, and two reaction coordinates were chosen to describe the transformation from the keto product to the enol intermediate. These were the transfer of a proton from the  $\alpha$ -carbon of the inhibitor to a water molecule (PT1:  $d[\text{O}_{\text{wat}}-\text{H}_{\text{ibrutinib}}] - d[\text{C}\alpha_{\text{ibrutinib}}-\text{H}_{\text{ibrutinib}}]$ ) and the transfer of a proton from a water molecule to the carbonyl oxygen atom of the inhibitor and formation (PT2:  $d[\text{O}_{\text{ibrutinib}}-\text{H}_{\text{wat}}] - d[\text{O}_{\text{wat}}-\text{H}_{\text{wat}}]$ ).



**Figure 5.15** (A) FE surface of the solvent assisted keto-enol tautomerisation step in BTK with ibrutinib covalently bound to C481. The barrier to this process is  $10.5 \text{ kcal mol}^{-1}$  at the DFTB3 level when 30 ps of sampling is carried out in each umbrella sampling window along the MEP. (B) Approximate transition state of keto-enol tautomerisation (proton transfer 1 =  $-0.6 \text{ \AA}$ , proton transfer 2 =  $-1.0 \text{ \AA}$ ).

The FE surface (Figure 5.15) produced at the DFTB3/MM level for this keto enol tautomerisation step has a barrier of  $10.8 \text{ kcal mol}^{-1}$  to from the covalently bound

keto product (EP, Figure 5.15). The covalent keto adduct is 35.1 kcal mol<sup>-1</sup> lower in energy than the enol adduct. This is in good agreement with the energetics calculated for this step at the same level of theory and at the  $\omega$ B97X-D/6-31G(d) level for the model BTK/ibrutinib system (Figure 5.14). The TS for the solvent assisted keto-enol tautomerisation step (Figure 5.15(B)) close in structure to EI2 and corresponds to the transfer of a proton from water to the C $\alpha$  carbon of ibrutinib.

The results indicate that a solvent assisted keto-enol tautomerisation step is a feasible event in BTK, when ibrutinib is bound in the ATP binding pocket. The reasonably solvent exposed nature of C481 and the region of space around the acrylamide warhead of ibrutinib mean that there is ample space for a water molecule to sit that can then assist in a keto-enol tautomerisation step. A water molecule is only ideally placed to perform this role in a relatively small 7 % of frames from the 500 ns MD trajectory analysed. This is when a water molecule is less than 3.5 Å from the carbonyl oxygen and  $\alpha$ -carbon of the inhibitor based on to  $d[\text{O}_{\text{ibrutinib}}-\text{O}_{\text{wat}}]$  and  $d[\text{O}_{\text{wat}}-\text{C}\alpha_{\text{ibrutinib}}]$  distances. However, this of course corresponds to the keto structure and carrying out the reaction in the reverse direction. The carbonyl oxygen is well solvated throughout the MD trajectory and has a water molecule in close proximity (less than or equal to 3.0 Å) in 73 % of frames. The enol is likely to be just as well solvated, making a solvent assisted proton transfer to form the keto product a facile process given the high degree of solvent interaction and low reaction barrier. A water molecule is observed in an optimal position to allow solvent assisted tautomerisation, as shown in EI2 of Figure 5.11.

## 5.4 Comparisons with experimental kinetic measurements

Studies of the inhibition kinetics of ibrutinib and BTK exist that estimate for the inactivation rate ( $k_{\text{inact}}$ ) of BTK by ibrutinib to be 0.0116 s<sup>-1</sup> and 0.0266 s<sup>-1</sup>.<sup>90,293</sup> Using transition state theory, these rates correspond to  $\Delta G^\ddagger$  values of 20.0 kcal mol<sup>-1</sup>

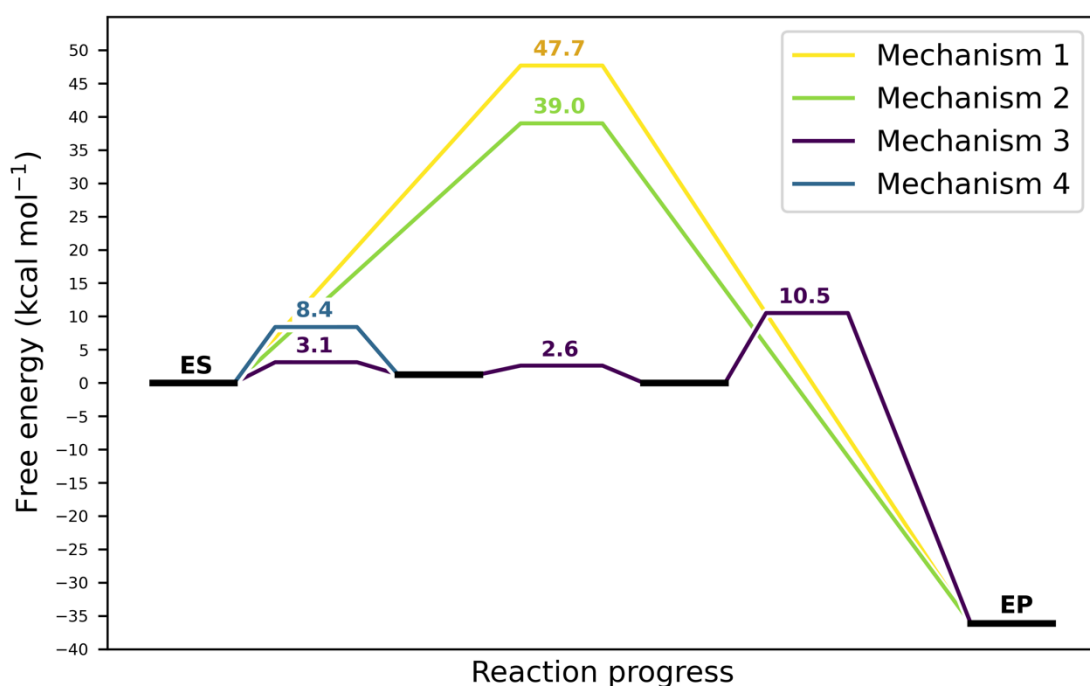
and 19.6 kcal mol<sup>-1</sup> respectively. The barrier to the keto enol tautomerisation step of 10.8 kcal mol<sup>-1</sup> makes this the rate limiting step in the covalent inhibition of BTK by ibrutinib. Although this is clearly an underestimate of the experimental barrier of 19.6 kcal mol<sup>-1</sup>, the comparison of full FE surfaces for thiol addition in BTK at the DFTB3 and  $\omega$ B97X-D/6-31G(d) levels of theory show that DFTB3 underestimates the barrier by approximately 5 kcal mol<sup>-1</sup> compared to  $\omega$ B97X-D. This is consistent with other studies that have found DFTB3 to underestimate reaction barriers.<sup>244,294,295</sup> Furthermore, the reaction simulations were started from MD snapshots where the C481 thiol and acrylamide group in ibrutinib were optimally positioned for the reaction to occur. The reaction barriers reported are therefore likely to underestimate the experimental kinetics as they neglect any conformational FE penalty required for the system to adopt a reactive conformation. Although reactive conformations were observed to be rare events in classical molecular dynamics simulations, they do form occasionally and so the barrier to forming a reactive conformation is likely to be low. Nevertheless, the extensive benchmarking of the DFTB3 method performed in Chapter 4 show that DFTB3 is able to reproduce reaction pathways for thiol addition that closely resemble those predicted from density functional methods ( $\omega$ B97X-D) and post Hartree-Fock methods (MP2). DFTB3 therefore represents an excellent choice for comparing reaction pathways for thiol addition as it provides a good balance between accuracy and speed.

Experimental inactivation kinetics also provide some support of the direct proton transfer mechanism (Mechanism 3) presented in this work. There are ten kinases that contain a cysteine residue in the same position as BTK.<sup>296</sup> Of these, TEC kinase is one example that contains an i+3 asparagine residue analogous to BTK. Ibrutinib exhibits very similar inactivation kinetics with TEC and BTK.<sup>90</sup> Comparison of the experimental kinetics of ibrutinib against kinases that contain an i+3 aspartate residue, such as EGFR and ITK, shows a very different kinetic profile. Experimental inactivation kinetic measurements ( $k_{\text{inact}}/K_i$ ) for these kinases show an increase of two-

to-three orders of magnitude in the selectivity towards ibrutinib.<sup>297</sup> This indicates that an alternative mechanism of action could be occurring in kinases with different i+3 residues. Covalent binding of ibrutinib to kinases that contain i+3 aspartate residues are likely to proceed via a base catalysed mechanism, where proton transfer occurs via the aspartate residue. In contrast, kinases containing an i+3 asparagine result in much higher  $k_{\text{inact}}/K_i$  values<sup>297</sup> and could support the mechanistic pathway presented for BTK where a direct proton transfer to the inhibitor occurs. Although S-C formation is the rate-limiting step in kinases with an i+3 aspartate,<sup>74</sup> the presence of an asparagine in BTK changes the environment around the cysteine and results in a different mechanistic pathway where S-C bond formation is no longer the rate limiting step.

## 5.5 Conclusions

The four mechanistic pathways explored are summarized in Figure 5.16 and indicate that the most probable mechanism for covalent modification of C481 by inhibitors containing an acrylamide warhead is a direct proton transfer from the thiol group to the carbonyl oxygen atom of the acrylamide group followed by a solvent assistant tautomerisation to form the covalently bound keto product of the reaction. This pathway was found to be lower in energy than the equivalent pathway that does involve assistance by a water molecule to form the enol intermediate, and pathways where the reaction occurs without going via an enol intermediate but instead proceeding directly to the keto product (Figure 5.14). The preferred pathway occurs in three steps. In the first step, the thiol proton transfers to the carbonyl oxygen atom via a water molecule with a barrier of 8.3 kcal mol<sup>-1</sup>. In the second step, nucleophilic attack occurs from the newly formed thiolate anion on to the  $\beta$ -carbon of the acrylamide group to form an enol intermediate with a barrier of 5.2 kcal mol<sup>-1</sup>. The enol intermediate then tautomerises to the keto product, where the barrier is reduced to 10.8 kcal mol<sup>-1</sup> by the assistance of a nearby water molecule.



**Figure 5.16** Comparison of the four mechanistic pathways examined in this work, showing the FE of the transition states and minimum energy structures along each pathway relative to the reactant, ES. The lowest energy pathway consists of a solvent assisted proton transfer step, followed by S-C formation and then a final solvent assisted keto-enol tautomerisation step to form the covalently bound keto product formed from the reaction between ibrutinib and the thiol side chain of C481 in BTK.

The rate limiting step in the most likely mechanistic pathway is the solvent assisted keto-enol tautomerisation step with a FE barrier of 10.5 kcal mol<sup>-1</sup>. Although the experimental barrier is underestimated by 9.1 kcal mol<sup>-1</sup>, this is expected, as the use of the semi-empirical DFTB3 Hamiltonian was demonstrated to underestimate the barrier to thiol addition by approximately 5 kcal mol<sup>-1</sup> (Figure 5.9). In spite of this, the reaction pathways generated by DFTB3 are robust and are in close agreement to pathways predicted at the  $\omega$ B97X-D/6-31G(d) level of theory (Figure 5.9). The QM/MM reaction simulations also highlight the importance of performing reactivity studies of this type in a protein environment, owing to the crucial interaction formed between the cysteine thiolate and amide side chain of the neighboring i+3 N484 residue, that stabilises the transition states and reaction intermediates along the reaction pathway. The identification of a suitable proton acceptor of the thiol proton

is also a significant result, given that the amino acid residues surrounding C481 are not appropriate candidates. This is particularly important given the upshifted  $pK_a$  of C481 in BTK. This upshifted  $pK_a$  has been calculated as 10.4 using thermodynamic integration,<sup>65</sup> and as 13.2 from constant pH MD simulations in Chapter 3 of this thesis. The upshift in  $pK_a$  means that the C481 thiol group will not be ionized at physiological pH, and a suitable proton acceptor is therefore required for reaction with a covalent inhibitor.

Understanding the precise mechanism by which C481 in BTK is covalently modified by acrylamide warheads will help in the design of safer, and more selective covalent drugs. There are currently no other studies that have investigated the covalent mechanism of action of ibrutinib against its target BTK at the atomistic level, and insights from this work can help rationally tune the covalent reactivity of acrylamide (and potentially other) covalent inhibitors of BTK. The mechanistic pathways explored in this work highlight the importance of inhibitor conformation, thiol conformation and  $pK_a$ , and the hydration of the binding site. Water plays an essential role in assisting the keto-enol tautomerisation step in the reaction so will also influence the thiol  $pK_a$  and therefore its intrinsic reactivity. The use of substituted acrylamides with hydrophobic or hydrophilic groups could be employed as a strategy to attenuate or enhance reactivity. Substituted acrylamides with appropriate electron withdrawing or donating substituents could be used to subtly tune the electrophilicity of the  $\beta$ -carbon of the acrylamide to enhance its interaction with the sulfur. The conformation of the warhead and its position relative to the thiol group is also another avenue for altering reactivity. The linker group that is used to attach the warhead to the main drug scaffold could be changed by adding or reducing flexibility. This could allow the warhead to adopt a more reactive conformation by allowing it to get closer to the acrylamide prior to reaction. However, of the little kinetic data available for covalent BTK inhibitors, changes in the linker and even warhead show that these have little effect on the reported inactivation rates for five covalent BTK inhibitors (Table

1). For example, tirabrutinib contains a butynamide warhead and different linker group to ibrutinib, but the FE of inactivation differs by only 0.6 kcal mol<sup>-1</sup>. This raises the possibility that reactivity is largely dominated by the protein itself, in particular the orientation and p*K*<sub>a</sub> of the cysteine residue. Consideration of both cysteine reactivity and warhead reactivity will therefore be important for the design of covalent inhibitors.

**Table 5.1** Five covalent BTK inhibitors and corresponding inactivation rates.<sup>293</sup> The FE of activation  $\Delta G^\ddagger$  values are calculated using transition state theory and are shown for comparison.

Inhibitor	BTK inactivation rate, $k_{\text{inact}}$ (s <sup>-1</sup> )	FE of inactivation, $\Delta G^\ddagger$ (kcal mol <sup>-1</sup> )
Ibrutinib	2.66 x 10 <sup>-2</sup>	19.6
Acalabrutinib	5.59 x 10 <sup>-3</sup>	20.5
Zanabrutinib	3.33 x 10 <sup>-2</sup>	19.5
Spebrutinib	1.36 x 10 <sup>-2</sup>	20.0
Tirabrutinib	9.72 x 10 <sup>-2</sup>	20.2

## 5.6 Suggestions for further work

The four mechanisms investigated for the covalent inhibition of BTK by ibrutinib were investigated using the semi-empirical QM method DFTB3. The suitability of using this QM method was explored extensively in Chapter 4 of this thesis, and some additional comparisons in this chapter have demonstrated that the reaction pathways produced by DFTB3 are in excellent agreement with the  $\omega$ B97X-D density functional. However, the energies predicted by DFTB3 are likely to underestimate the barrier to thiol addition (Figure 5.9), and this presents an opportunity to further improve upon the work in this chapter. The umbrella sampling simulations for each step of the lowest energy pathway (Mechanism 3) should be repeated using a higher level of QM theory for thiol addition such as  $\omega$ B97X-D/6-31G(d). As a result of the computational cost of performing QM/MM MD at this higher level of theory, the amount of sampling

performed in each reaction coordinate window will be limited compared to DFTB3. See Chapter 4, Section 4.7 for a discussion on how this can be improved. The aim of this would be to obtain more accurate barrier heights for the reaction pathway that agree with experimental reaction kinetics. This would enable alternative inhibitors to be modelled with QM/MM methods and the barrier heights for the rate determining step to be calculated. These could then be compared amongst a set of covalent inhibitors and allow their reactivity towards C481 in BTK to be ranked. The effect that modifications to the drug scaffold, particularly around the reactive warhead, have on the rate of reactivity could also be investigated. The ability to rank and predict the reactivity of covalently binding inhibitors would be an incredibly useful tool in computer aided drug design.<sup>72</sup>

Another opportunity for future work is the use of an alternative reaction path finding method. Several problems were encountered with the QM/MM umbrella sampling protocol used in this work, on occasions where accurate MEPs could not be obtained for the mechanistic pathways investigated for BTK inhibition by ibrutinib. Some of the problems encountered were in the high energy regions of the FE surface, where side reactions were often observed that conflicted with the sampled reaction coordinates. In addition, the use of a simple linear combination of distances to describe complex reaction coordinates involving the formation and cleavage of several bonds can be problematic.<sup>298</sup> To overcome both of these issues, a combination of the string method<sup>299,300</sup> and a path collective variable<sup>301</sup> can be used. The advantage of the string method is that it first optimizes the MEP between the reactant and product state, and then umbrella sampling can be performed along the pre-defined path.<sup>302</sup> This means a full FE surface does not need to be produced with umbrella sampling, which can be costly as the number of reaction coordinates increases. The use of a path collective variable to describe the transformation from reactants to products is advantageous as a combination of multiple distances, angles or dihedrals are able to more accurately describe the MEP.<sup>303</sup> Furthermore, the string method means that multiple reaction



paths can be considered on the same FE surface, further reducing the computational cost.<sup>302</sup> Employing the adaptive string method<sup>304</sup> to investigate the reaction of ibrutinib in BTK would complement the current work by providing another estimate of the MEP, which would hopefully be in agreement with the current path. It would also provide a more accurate path and transition state structures for further analysis.

Finally, the covalent inactivation by ibrutinib of other kinases that contain an i+3 Asp residue such as EGFR could be modelled. It would be interesting to test if the direct mechanism presented in this chapter was the preferred pathway in EGFR given the change in i+3 residue from Asn to Asp, or whether a base catalyzed mechanism<sup>74</sup> would operate instead. This could then be compared to experimental inactivation kinetic data that suggest a different mechanism of action could be taking place in kinases with i+3 Asn residues compared to kinases with i+3 Asp residues due to differences in  $k_{\text{inact}}/K_i$  measurements.<sup>297</sup> An experimental kinetic study of BTK mutants could also support the mechanism presented in this chapter. For example, if a library of N484 mutants was generated and the rate of inactivation of each mutant measured, this could support the role of Asn in the mechanism.

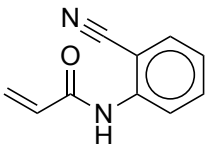
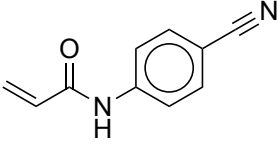
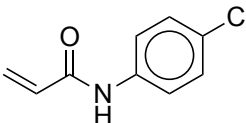
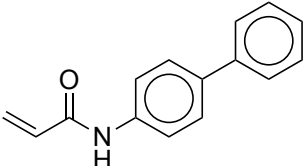
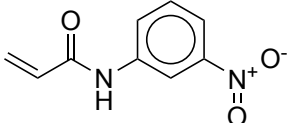
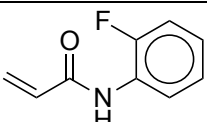
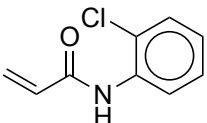
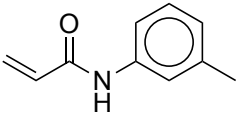
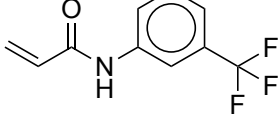
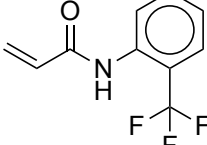
## Appendix 1

The following is the supporting information from the publication in Chapter 2.

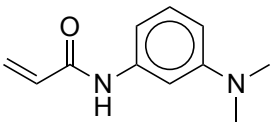
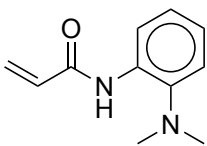
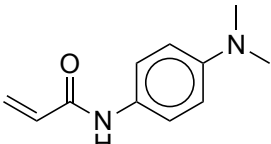
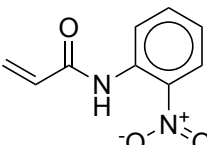
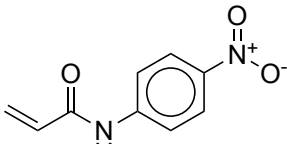
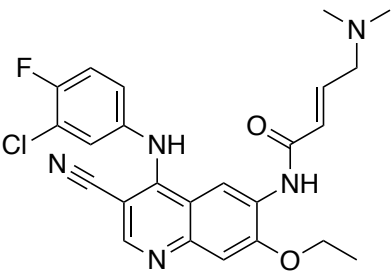
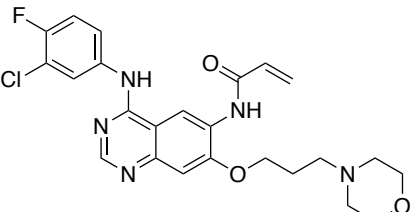
**Table A1.** Structures, PA ( $\Delta\Delta G$ ),  $\Delta E$  and  $\log(k_{\text{GSH}})$  values for all compounds studied in Chapter 2.

Compound ID	SMILES	Structure	PA, $\Delta\Delta G$ (kcal mol <sup>-1</sup> )	Reaction energy, $\Delta E$ (kcal mol <sup>-1</sup> )	$\log(k_{\text{GSH}})$
<b>Data set 1</b>					
6	<chem>O=C(C=C)Nc1ccccc1</chem>		0.00	2.75	-1.05
7	<chem>N#Cc1ccc(NC(C=C)O)c1</chem>		-0.55	4.64	-1.59
8	<chem>O[C@@H]1CN(C(C=C)O)C1=O</chem>		-1.10	5.11	-1.82
26	<chem>Clc1ccc(NC(C=C)O)c1</chem>		-2.69	4.67	-1.89
27	<chem>O=C(C=C)Nc1ccccc1</chem>		-2.32	9.02	-2.14
28	<chem>COc1ccc(NC(C=C)O)c1</chem>		-6.33	7.76	-2.48
29	<chem>CN(C(C=C)O)c1ccccc1</chem>		-2.96	1.01	-2.54

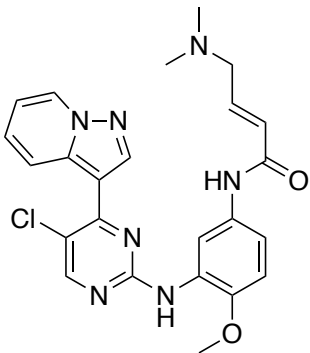
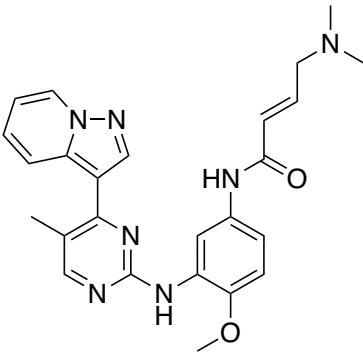
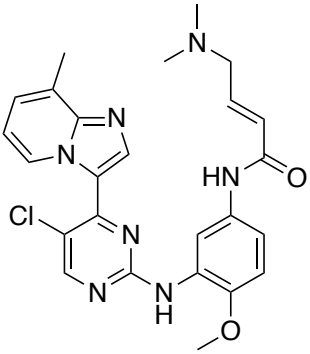
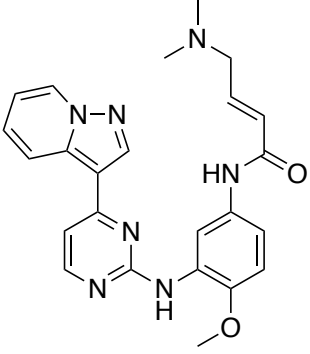
30	<chem>CNC(/C=C/C(F)(F)F)=O</chem>		-5.37	9.02	-2.59
31	<chem>COc1ccc(NC(C=C)O)c1</chem>		-10.29	7.65	-2.85
32	<chem>O=C(C=C)NCc1ccccc1</chem>		-7.56	7.82	-3.10
33	<chem>CNC(C=C)=O</chem>		-10.09	-3.05	-3.15
34	<chem>COc1cc(CN(C(C=C)O)ccc1</chem>		-8.61	11.94	-3.40
35	<chem>O=C(C=C)N1CCCCC1</chem>		-7.24	7.78	-2.40
Data set 2					
9	<chem>C=CC(Nc1ccc(N)cc1)=O</chem>		-5.99	-20.07	-2.51
10	<chem>C=CC(Nc1ccc(F)cc1)=O</chem>		-4.41	-20.32	-2.11
11	<chem>C=CC(Nc1ccc(C)cc1)=O</chem>		-5.81	-20.52	-2.23
36	<chem>C=CC(Nc1ccc(F)cc1)=O</chem>		0.00	-20.62	-1.17
37	<chem>C=CC(Nc1ccccc1C)=O</chem>		-2.83	-20.40	-1.76

38	<chem>C=CC(Nc1cccc1C#N)=O</chem>		-5.23	-20.19	-2.16
39	<chem>C=CC(Nc1ccc(C#N)cc1)=O</chem>		-1.33	-20.62	-1.12
Data set 3					
12	<chem>C=CC(Nc1ccc(Cl)cc1)=O</chem>		-5.99	-20.58	-1.6
13	<chem>C=CC(Nc1ccc(cc1)-c2ccccc2)=O</chem>		-5.33	-20.43	-1.71
14	<chem>C=CC(Nc1ccc([N+](=O)[O-])cc1)=O</chem>		-4.47	-20.79	-1.04
40	<chem>C=CC(Nc1ccccc1F)=O</chem>		-3.77	-20.45	-1.68
41	<chem>C=CC(Nc1ccccc1Cl)=O</chem>		-5.10	-20.79	-1.32
42	<chem>C=CC(Nc1ccc(C)cc1)=O</chem>		-8.39	-20.32	-1.97
43	<chem>C=CC(Nc1ccc(C(F)(F)F)cc1)=O</chem>		-5.77	-20.37	-1.43
44	<chem>C=CC(Nc1ccccc1C(F)(F)F)=O</chem>		-3.36	-21.09	-1.44

45	<chem>C=CC(Nc1ccc(C(F)(F)F)cc1)=O</chem>		-4.11	-20.50	-1.23
46	<chem>C=CC(Nc1ccc(cc1)-c2ccccc2)=O</chem>		-7.04	-20.56	-1.65
47	<chem>C=CC(Nc1ccccc1-c2ccccc2)=O</chem>		-6.77	-20.62	-1.98
48	<chem>C=CC(Nc1ccc(OC(=O)C)cc1)=O</chem>		-6.28	-20.44	-1.7
49	<chem>C=CC(Nc1ccccc1C(=O)OC)=O</chem>		-4.95	-20.58	-1.11
50	<chem>C=CC(Nc1ccc(OC(=O)C)cc1)=O</chem>		-4.37	-20.58	-1
51	<chem>C=CC(Nc1ccccc1OC)=O</chem>		-7.41	-20.40	-1.92
52	<chem>C=CC(Nc1ccc(SC)cc1)=O</chem>		-5.75	-20.46	-1.58
53	<chem>C=CC(Nc1ccccc1S)=O</chem>		-5.66	-20.54	-1.5
54	<chem>C=CC(Nc1ccc(SC)cc1)=O</chem>		-5.32	-20.23	-1.77

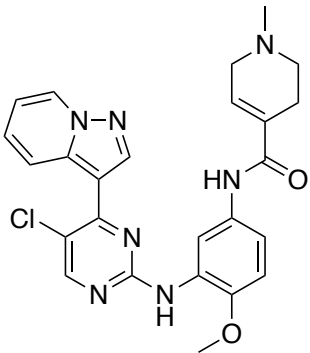
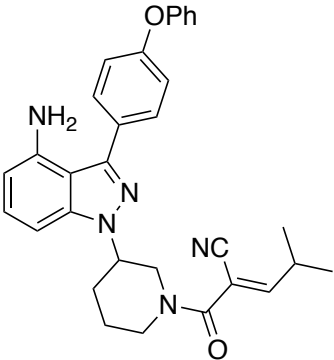
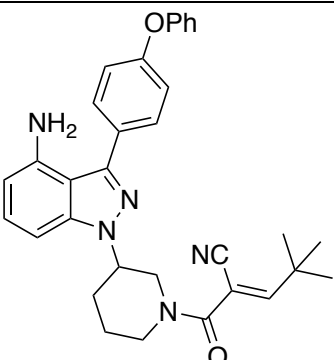
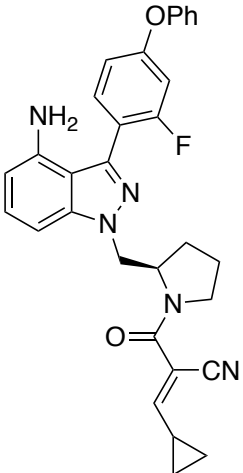
55	<chem>C=CC(Nc1cc(N(C)C)ccc1)=O</chem>		-5.28	-20.29	-1.82
56	<chem>C=CC(Nc1ccccc1)=O</chem>		-7.59	-20.45	-1.55
57	<chem>C=CC(Nc1ccc(N(C)C)cc1)=O</chem>		-9.01	-20.18	-2.03
58	<chem>C=CC(Nc1cc([N+](=O)[O-])ccc1)=O</chem>		0.00	-20.65	-0.416
59	<chem>C=CC(Nc1ccc([N+](=O)[O-])cc1)=O</chem>		-1.22	-20.80	-0.534
Data set 4					
15	<chem>O=C(/C=C/CN(C)C)Nc1cc2c(C1=CC2=C(NC3=CC=C(F)C(C1)=C3)C(C#N)=CN=C2C=C1OCC</chem>		-4.33	-19.88	-1.09
18	<chem>O=C(C=C)Nc1cc2c(C1=CC2=C(NC3=CC=C(F)C(C1)=C3)N=CN=C2C=C1OCCCN4CCOCC4</chem>		0.00	-24.43	-0.96

19	<chem>COC1=CC2=NC=NC(NC3=CC=C(F)C(Cl)=C3)=C2C=C1NC(/C=C/CN4CCCCC4)=O</chem>		-5.41	-20.00	-1.29
20	<chem>O=C(/C=C/CN(C)C)N(C1=CC2=C(NC3=CC=C(F)C(Cl)=C3)N=CN=C2C=C1OC4COCC4</chem>		-3.26	-22.50	-1
21	<chem>O=C(/C=C/CN(C)C)N(C1=CC2=C(NC3=CC=C(OCC4=NCC=C4)C(Cl)=C3)C(C#N)=CN=C2C=C1OCC</chem>		-10.03	-21.64	-1.21
22	<chem>O=C(C=C)NC1=CC2=C(C(NC3=CC(C)=CC=C3)=CC=C3)N=CN=C2C=C1</chem>		-3.31	-22.53	-1.49
Data set 5					
16	<chem>COC(C=CC(NC(/C=C/CN(C)C)=O)=C1)C1NC2=NC=C(Cl)C(C3=CNC4=CC=C(C=C43)=N2</chem>		-13.41	-17.05	-2.25

23	<chem>COC(C=CC</chem> <chem>(NC(/C=C/</chem> <chem>CN(C)C)=O</chem> <chem>)=C1)=C1N</chem> <chem>C2=NC=C(</chem> <chem>Cl)C(C3=C4</chem> <chem>N(C=CC=C</chem> <chem>4)N=C3)=N</chem> <chem>2</chem>		-2.92	-20.19	-2.15
24	<chem>CC(C=N1)</chem> <chem>=C(C2=C3</chem> <chem>N(C=CC=C</chem> <chem>3)N=C2)N=</chem> <chem>C1NC4=CC</chem> <chem>(NC(/C=C/</chem> <chem>CN(C)C)=O</chem> <chem>)=CC=C4O</chem> <chem>C</chem>		0.00	-16.35	-1.84
25	<chem>COC(C=CC</chem> <chem>(NC(/C=C/</chem> <chem>CN(C)C)=O</chem> <chem>)=C1)=C1N</chem> <chem>C2=NC=C(</chem> <chem>Cl)C(C3=C</chem> <chem>N=C4N3C=</chem> <chem>CC=C4C')=</chem> <chem>N2</chem>		-2.60	-18.84	-2.8
60	<chem>COC(C=CC</chem> <chem>(NC(/C=C/</chem> <chem>CN(C)C)=O</chem> <chem>)=C1)=C1N</chem> <chem>C2=NC=CC</chem> <chem>(C3=C4N(C</chem> <chem>=CC=C4)N</chem> <chem>=C3)=N2</chem>		-1.33	-18.43	-2.24



61	<chem>COC(C=CC(NC(/C=C/CN(C)C)=O)=C1)=C1NC2=NC=C(C)C(C3=C4N(C=CC=C4)N=C3)=N2</chem>		-2.72	-20.17	-2.28
62	<chem>COC(C=CC(NC(C=C)O)=C1)=C1NC2=NC=C(Cl)C(C3=CN=C4N3C=CC=C4)=N2</chem>		-1.95	-19.67	-2.16
63	<chem>COC(C=C(CN(C)C)C(NC(C=C)O)=C1)=C1NC2=NC=C(Cl)C(C3=CN=C4N3C=CC=C4)=N2</chem>		-5.68	-25.67	-1.42
64	<chem>COC(C=CC(NC(/C=C/CN(C)C)=O)=C1)=C1NC2=NC=C(C#N)C(C3=C4N(C=C(C=C4)N=C3)=N2</chem>		-0.79	-19.32	-2.22
65	<chem>ClC1=CN=C(C(NC2=NN(C)C(NC(C=O)=C2)N=C1C3=C4N(C=CC=C4)N=C3</chem>		-0.25	-22.43	-1.11

66	<chem>COC(C=CC</chem> <chem>(NC(C1=CC</chem> <chem>N(C)CC1=</chem> <chem>O)=C2)=C2</chem> <chem>NC3=NC=C</chem> <chem>(Cl)C(C4=C</chem> <chem>5N(C=CC=</chem> <chem>C5)N=C4)=</chem> <chem>N3</chem>		-0.82	-8.65	-2.9
Data set 6					
17	<chem>O=C(/C(C</chem> <chem>#N)=C/C)</chem> <chem>N(C1)CCCC</chem> <chem>1N2C3=CC</chem> <chem>=CC(N)=C</chem> <chem>3C(C4=CC</chem> <chem>=C(OC5=C</chem> <chem>C=CC=C5)</chem> <chem>C=C4)=N2</chem>		-12.81	-16.82	-1.96
67	<chem>O=C(/C(C</chem> <chem>#N)=C/C(</chem> <chem>C)C)N(C1)C</chem> <chem>CCC1N2C3</chem> <chem>=CC=CC(N</chem> <chem>)=C3C(C4=</chem> <chem>CC=C(OC5</chem> <chem>=CC=CC=</chem> <chem>C5)C=C4)=</chem> <chem>N2</chem>		-9.28	-14.16	-2.30
68	<chem>O=C(/C(C</chem> <chem>#N)=C/C(</chem> <chem>C)(C)C)N(C</chem> <chem>1)CCCC1N2</chem> <chem>C3=CC=CC</chem> <chem>(N)=C3C(C</chem> <chem>4=CC=C(O</chem> <chem>C5=CC=CC</chem> <chem>=C5)C=C4)</chem> <chem>=N2</chem>		0.00	-15.34	-1.34

69	<chem>O=C(/C(C#N)=C/C1CC1)N(CCC2)[C@H]2CN3C4=CC=C(C(N)=C4C(C5=C(F)C=C(OC6=CC=CC=C6)C=C5)=N3</chem>		-18.58	-21.09	-2.30
70	<chem>O=C(C=C)N(CCC1)[C@H]1CN2C3=CC=CC(N)=C3C(C4=C(F)C=C(OC5=CC=CC=C5)C=C4)=N2</chem>		-18.30	-14.59	-1.53
71	<chem>O=C(/C(C#N)=C/C1CC1)N(CCC2)[C@@H]2CN3C4=CC=CC(N)=C4C(C5=C(F)C=C(OC6=CC=CC=C6)C=C5)=N3</chem>		-8.90	-9.35	-1.92

72	<chem>O=C(/C(C#N)=C/C(N1CCOCC1)(C)C)N(CC2)[C@@H]2CN3C4=CC=CC(N)=C4C(C5=C(F)C=C(OC6=CC=CC=C6)C=C5)=N3</chem>		-10.38	-9.66	-2.22
73	<chem>O=C(/C(C#N)=C/C(N1COC1)C)(C)C)N(CC2)[C@@H]2CN3C4=CC=CC(N)=C4C(C5=C(F)C=C(OC6=CC=CC=C6)C=C5)=N3</chem>		-0.85	-12.37	-0.39

### A1.1 Boltzmann averaging of $\Delta\Delta G_{PA}$ values from multiple low energy conformers

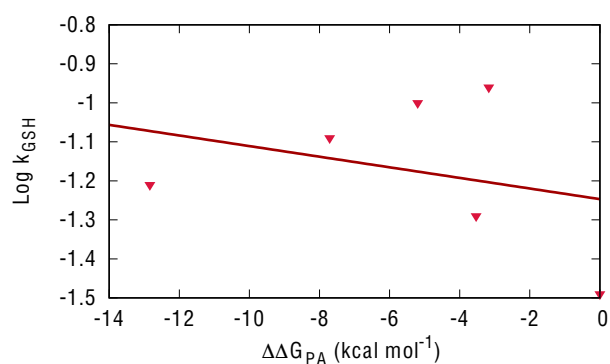
We have calculated a Boltzmann averaged  $\Delta\Delta G_{PA}$  from 4 low energy conformers of each compound in data set 4 according to the following equation.<sup>305</sup>

$$\Delta G_{PA,ave} = -RT \ln \left[ \frac{1}{n} \sum_{i=1}^n \exp \left( \frac{\Delta G_{PA,i}}{RT} \right) \right]$$

**Table A2.**  $\Delta G_{PA}$  values and Boltzmann weighted average for compounds 15, 18, 19, 20, 21 and 22 from data set 4.

	Compound $\Delta G_{PA}$ kcal mol <sup>-1</sup>					
	15	18	19	20	21	22
<b>Conformer</b>						
<b>1</b>	-317.69	-311.12	-312.42	-314.93	-316.26	-301.09
<b>Conformer</b>						
<b>2</b>	-318.96	-313.17	-312.42	-311.84	-317.42	-300.10
<b>Conformer</b>						
<b>3</b>	-316.12	-313.25	-314.58	-316.64	-316.26	-311.07
<b>Conformer</b>						
<b>4</b>	-318.34	-314.52	-314.58	-303.44	-324.32	-311.07
<b>Boltzmann weighted average</b>	-318.36	-313.82	-314.18	-315.85	-323.49	-310.65

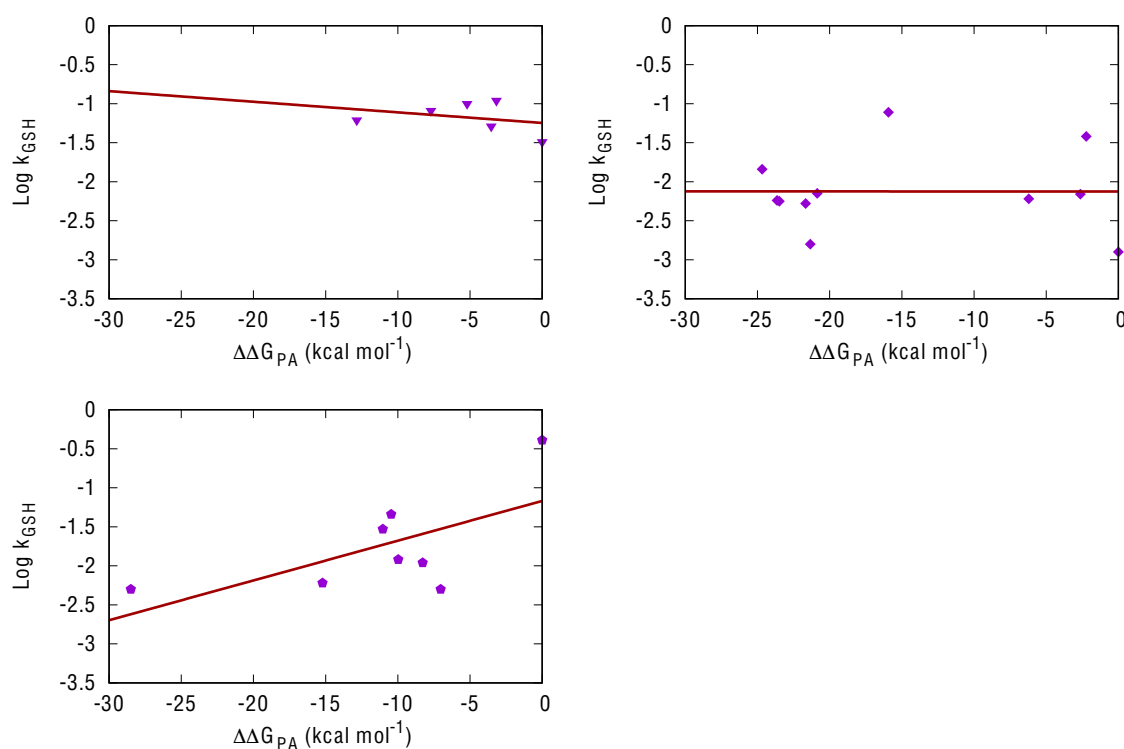
We observe no improvement in correlation between  $\Delta\Delta G_{PA}$  and  $\log k_{GSH}$  when using Boltzmann averaged  $\Delta\Delta G_{PA}$  values for data set 4 compounds. Interestingly, a different compound is predicted to be the most acidic compound compared to the data presented in Figure 2.5, highlighting the strong conformational dependence of PA.



**Figure A1.** Plot of Boltzmann averaged  $\Delta\Delta G_{PA}$  for data set 4 compounds vs  $\log k_{GSH}$ .  $R^2=0.09$

## A1.2 Effects of using alternative optimisation strategies

We investigated the effect of using different optimisation strategies on alternative low energy conformers due to the poor correlation observed between  $\Delta\Delta G_{PA}$  and  $\log k_{GSH}$  for compounds from datasets 4, 5 and 6. We used a procedure nearly identical to the methodology reported in the methods section, but used Gaussian09<sup>306</sup> with tighter convergence criteria, and in some cases the UltraFine (pruned 99,590) integration grid if imaginary frequencies were observed.

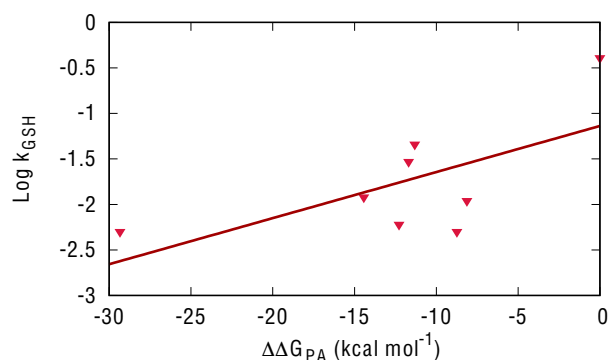


**Figure A2.** Plots of GSH reactivity vs relative proton affinity ( $\Delta\Delta G_{PA}$ ) for (A) data set 4,  $r^2=0.09$ , (B) data set 5,  $r^2=0.00$  (C) data set 6,  $r^2=0.41$ .

The correlation between  $\Delta\Delta G_{PA}$  and  $\log k_{GSH}$  is very similar to the correlation reported in Figure 2.5 for data set 4 and 5, giving no correlation for two of the three datasets. A slightly improved correlation of  $R^2=0.41$  is observed for data set 6, but it

still does not have any predictive utility in drug design, and there is clearly no good prediction of the rank order whatsoever, shown by the low Spearman correlation of 0.399. Although a more rigorous optimisation protocol was used, we still conclude that ligand only methods are not sufficient to accurately predict covalent reactivity ‘in situ’.

### A1.3 Effect of using an alternative DFT functional for PA calculation



**Figure A3.** Plot of  $\Delta\Delta G_{PA}$  for data set 6 compounds vs log  $k_{GSH}$  using the M06-2X density functional.  $R^2=0.41$

Recent reports have recommended that the M06-2X density functional should be used when modelling sulfur reactivity.<sup>231</sup> In particular, the use of the B3LYP density functional has been found to be problematic, as it fails to accurately predict long range behaviour due to errors in the exchange part of the density functional.<sup>127</sup> To ensure that this didn’t affect our protocol for calculating  $\Delta\Delta G_{PA}$ , we took all the compounds in data set 6 and calculated  $\Delta\Delta G_{PA}$  at the M06-2X-D3/6-311+G(d,p)//M06-2X-D3/6-31G(d) level in Gaussian09, revision D.01<sup>306</sup> (analogous to our original procedure, see methods, but using the M06-2X functional in place of B3LYP and using Gaussian instead of Jaguar). We find that the correlation between  $\Delta\Delta G_{PA}$  and experimental reactivity to be identical when comparing B3LYP and M06-2X ( $R^2=0.41$  and  $R^2=0.41$  respectively). We conclude that the choice of density functional has very little impact on calculated PA, and so we have chosen to use the B3LYP density functional for our calculations as they have shown to work well for PA calculations.<sup>125</sup>

## A1.4 Quasi Harmonic Oscillator approximation

For the large and flexible drug-like molecules in data sets 4, 5 and 6, we have used the quasi harmonic oscillator approximation<sup>136</sup> to correct the calculated  $\Delta G_{PA}$  values. For these larger molecules, the contribution of low energy vibrational modes to the free energy can be increasingly important.<sup>135</sup> Using the program GoodVibes<sup>137</sup>, we have calculated the corrected free energy using the Truhlar quasi harmonic oscillator approximation<sup>136</sup> with a frequency cut off of 100  $\text{cm}^{-1}$ . We find that QHO free energy corrections do not deviate significantly from the uncorrected values (Table S2), with average % changes of 0.46, 0.41 and 0.17 for data sets 4, 5 and 6 respectively.

**Table A3.** Uncorrected, and QHO corrected  $\Delta G_{PA}$  values for data set 4, 5 and 6.

	<b>Compound</b>	<b><math>\Delta G_{PA}</math> (kcal mol<sup>-1</sup>)</b>	<b>QHO corrected <math>\Delta G_{PA}</math> (kcal mol<sup>-1</sup>)</b>	<b>Percentage change (%)</b>
<b>Data set 4</b>	15	-317.6889	-319.5106	0.57
	18	-311.1235	-310.6240	0.16
	19	-312.4187	-313.9925	0.50
	20	-314.9329	-318.4469	1.12
	21	-316.2598	-317.3812	0.35
	22	-301.0936	-300.9662	0.04
<b>Data set 5</b>	16	-320.7416	-321.1275	0.12
	23	-318.1141	-320.5526	0.77
	24	-321.9279	-323.3718	0.45
	25	-318.5897	-321.1763	0.81
	60	-320.8936	-323.2831	0.74
	61	-318.9190	-321.1466	0.70
	62	-299.8828	-300.8579	0.33
	63	-299.4848	-300.2768	0.26
	64	-303.4588	-305.0338	0.52
	65	-313.1765	-311.5826	0.51



	66	-297.2508	-298.2297	0.33
<b>Data set 6</b>	17	-296.52912	-296.52912	0.06
	67	-295.28391	-295.28391	0.01
	68	-298.71231	-298.71231	0.44
	69	-316.74201	-316.74201	0.11
	70	-299.29751	-299.29751	0.33
	71	-298.22322	-298.22322	0.03
	72	-303.47687	-303.47687	0.27
	73	-288.25334	-288.25334	0.11

## References

- 1 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, The resurgence of covalent drugs, *Nat. Rev. Drug Discov.*, 2011, **10**, 307–317.
- 2 M. Mann and O. N. Jensen, Proteomic analysis of post-translational modifications, *Nat. Biotechnol.*, 2003, **21**, 255–261.
- 3 F. J. Schopfer, C. Cipollina and B. A. Freeman, Formation and signaling actions of electrophilic lipids, *Chem. Rev.*, 2011, **111**, 5997–6021.
- 4 T. A. Baillie, Targeted Covalent Inhibitors for Drug Design, *Angew. Chemie Int. Ed.*, 2016, **55**, 13408–13421.
- 5 P. Giron, L. Dayon and J. C. Sanchez, Cysteine tagging for MS-based proteomics, *Mass Spectrom. Rev.*, 2011, **30**, 366–395.
- 6 J. S. Martin, C. J. MacKenzie, D. Fletcher and I. H. Gilbert, Characterising covalent warhead reactivity, *Bioorganic Med. Chem.*, 2019, **27**, 2066–2074.
- 7 D. A. Shannon and E. Weerapana, Covalent protein modification: The current landscape of residue-specific electrophiles, *Curr. Opin. Chem. Biol.*, 2015, **24**, 18–26.
- 8 L. Petri, A. Egyed, D. Bajusz, T. Imre, A. Hetényi, T. Martinek, P. Ábrányi-Balogh and G. M. Keserü, An electrophilic warhead library for mapping the reactivity and accessibility of tractable cysteines in protein kinases, *Eur. J. Med. Chem.*, 2020, **207**, 112836.
- 9 M. Gehringer and S. A. Laufer, Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology, *J. Med. Chem.*, 2019, **62**, 5673–5724.
- 10 K. McAulay, E. A. Hoyt, M. Thomas, M. Schimpl, M. S. Bodnarchuk, H. J. Lewis, D. Barratt, D. Bhavsar, D. M. Robinson, M. J. Deery, D. J. Ogg, G. J. L. Bernardes, G. J. L. Bernardes, R. A. Ward, M. J. Waring and J. G. Kettle, Alkynyl Benzoxazines and Dihydroquinazolines as Cysteine Targeting Covalent Warheads and Their Application in Identification of Selective Irreversible Kinase Inhibitors, *J. Am. Chem. Soc.*, 2020, **142**, 10358–10372.
- 11 F. Sutanto, M. Konstantinidou and A. Dömling, Covalent inhibitors: A rational approach to drug discovery, *RSC Med. Chem.*, 2020, **11**, 876–884.
- 12 J. R. Vane and R. M. Botting, The mechanism of action of aspirin, in *Thrombosis Research*, Elsevier Ltd, 2003, vol. 110, pp. 255–258.
- 13 D. J. Waxman and J. L. Strominger, Penicillin-Binding Proteins and the Mechanism of Action of Beta-Lactam Antibiotics, *Annu. Rev. Biochem.*, 1983, **52**, 825–869.
- 14 K. Munson, R. Garcia and G. Sachs, Inhibitor and ion binding sites on the gastric H,K-ATPase, *Biochemistry*, 2005, **44**, 5267–5284.
- 15 R. A. Bauer, Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies, *Drug Discov. Today*, 2015, **20**, 1061–1073.
- 16 I. Kola and J. Landis, Can the pharmaceutical industry reduce attrition rates?, *Nat. Rev. Drug Discov.*, 2004, **3**, 711–715.
- 17 A. Abdeldayem, Y. S. Raouf, S. N. Constantinescu, R. Moriggl and P. T. Gunning, Advances in covalent kinase inhibitors, *Chem. Soc. Rev.*, 2020, **49**, 2617–2687.
- 18 I. D. Kuntz, K. Chen, K. A. Sharp and P. A. Kollman, The maximal affinity of ligands, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9997–10002.
- 19 A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees and C. H. Reynolds, The role of ligand efficiency metrics in drug discovery, *Nat. Rev. Drug Discov.*, 2014, **13**, 105–

- 121.
- 20 A. J. T. Smith, X. Zhang, A. G. Leach and K. N. Houk, Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins, *J. Med. Chem.*, 2009, **52**, 225–233.
- 21 R. A. Copeland, The drug–target residence time model: a 10-year retrospective, *Nat. Rev. Drug Discov.*, 2016, **15**, 87–95.
- 22 J. M. Bradshaw, J. M. McFarland, V. O. Paavilainen, A. Bisconte, D. Tam, V. T. Phan, S. Romanov, D. Finkle, J. Shu, V. Patel, T. Ton, X. Li, D. G. Loughhead, P. A. Nunn, D. E. Karr, M. E. Gerritsen, J. O. Funk, T. D. Owens, E. Verner, K. A. Brameld, R. J. Hill, D. M. Goldstein and J. Taunton, Prolonged and tunable residence time using reversible covalent kinase inhibitors, *Nat. Chem. Biol.*, 2015, **11**, 525–531.
- 23 M. Cully, Rational drug design: Tuning kinase inhibitor residence time, *Nat. Rev. Drug Discov.*, 2015, **14**, 457.
- 24 D. S. Johnson, E. Weerapana and B. F. Cravatt, Strategies for discovering and derisking covalent, irreversible enzyme inhibitors., *Future Med. Chem.*, 2010, **2**, 949–64.
- 25 T. A. Halgren, Identifying and characterizing binding sites and assessing druggability, *J. Chem. Inf. Model.*, 2009, **49**, 377–389.
- 26 A. Leonetti, S. Sharma, R. Minari, P. Perego, E. Giovannetti and M. Tiseo, Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer, *Br. J. Cancer*, 2019, **121**, 725–737.
- 27 S. Cheng, A. Guo, P. Lu, J. Ma, M. Coleman and Y. L. Wang, Functional characterization of BTK C481S mutation that confers ibrutinib resistance: Exploration of alternative kinase inhibitors, *Leukemia*, 2015, **29**, 895–900.
- 28 R. Lagoutte, R. Patouret and N. Winssinger, Covalent inhibitors: an opportunity for rational target selectivity, *Curr. Opin. Chem. Biol.*, 2017, **39**, 54–63.
- 29 A. J. Wilson, J. K. Kerns, J. F. Callahan and C. J. Moody, Keap calm, and carry on covalently, *J. Med. Chem.*, 2013, **56**, 7463–7476.
- 30 J. C. L. Erve, Chemical toxicology: Reactive intermediates and their role in pharmacology and toxicology, *Expert Opin. Drug Metab. Toxicol.*, 2006, **2**, 923–946.
- 31 J. Uetrecht, Idiosyncratic drug reactions: Past, present, and future, *Chem. Res. Toxicol.*, 2008, **21**, 84–92.
- 32 B. K. Park, A. Boobis, S. Clarke, C. E. P. Goldring, D. Jones, J. G. Kenna, C. Lambert, H. G. Lavery, D. J. Naisbitt, S. Nelson, D. A. Nicoll-Griffith, R. S. Obach, P. Routledge, D. A. Smith, D. J. Tweedie, N. Vermeulen, D. P. Williams, I. D. Wilson and T. A. Baillie, Managing the challenge of chemically reactive metabolites in drug development, *Nat. Rev. Drug Discov.*, 2011, **10**, 292–306.
- 33 M. H. Potashman and M. E. Duggan, Covalent modifiers: An orthogonal approach to drug design, *J. Med. Chem.*, 2009, **52**, 1231–1246.
- 34 P. J. Tonge, Drug-Target Kinetics in Drug Discovery, *ACS Chem. Neurosci.*, 2018, **9**, 29–39.
- 35 R. Lonsdale and R. A. Ward, Structure-based design of targeted covalent inhibitors, *Chem. Soc. Rev.*, 2018, **47**, 3816–3830.
- 36 J. M. Strelow, A Perspective on the Kinetics of Covalent and Irreversible Inhibition, *SLAS Discov. Adv. Sci. Drug Discov.*, 2017, **22**, 3–20.
- 37 T. Barf and A. Kaptein, Irreversible Protein Kinase Inhibitors: Balancing the Benefits and Risks, *J. Med. Chem.*, 2012, **55**, 6243–6262.
- 38 Y. Shibata and M. Chiba, The role of extrahepatic metabolism in the pharmacokinetics of the targeted covalent inhibitors afatinib, ibrutinib, and neratinib, *Drug Metab.*

- Dispos.*, 2015, **43**, 375–384.
- 39 A. L. Garske, U. Peters, A. T. Cortesi, J. L. Perez and K. M. Shokat, Chemical genetic strategy for targeting protein kinases based on covalent complementarity, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 15046–15052.
  - 40 S. Wang, Y. Tian, M. Wang, M. Wang, G. Sun and X. Sun, Advanced Activity-Based Protein Profiling Application Strategies for Drug Development, *Front. Pharmacol.*, 2018, **9**, 353.
  - 41 H. Deng, Q. Lei, Y. Wu, Y. He and W. Li, Activity-based protein profiling: Recent advances in medicinal chemistry, *Eur. J. Med. Chem.*, 2020, **191**, 112151.
  - 42 G. M. Simon and B. F. Cravatt, Activity-based proteomics of enzyme superfamilies: Serine hydrolases as a case study, *J. Biol. Chem.*, 2010, **285**, 11051–11055.
  - 43 D. Kato, K. M. Boatright, A. B. Berger, T. Nazif, G. Blum, C. Ryan, K. A. H. Chehade, G. S. Salvesen and M. Bogoy, Activity-Based Probes that Target Diverse Cysteine Protease Families, *Nat. Chem. Biol.*, 2005, **1**, 33–38.
  - 44 M. S. Cohen, C. Zhang, K. M. Shokat and J. Taunton, Biochemistry: Structural bioinformatics-based design of selective, irreversible kinase inhibitors, *Science (80-. )*, 2005, **308**, 1318–1321.
  - 45 G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, Computational methods in drug discovery, *Pharmacol. Rev.*, 2014, **66**, 334–395.
  - 46 E. Awoonor-Williams, A. G. Walsh and C. N. Rowley, Modeling covalent-modifier drugs, *Biochim. Biophys. Acta - Proteins Proteomics*, 2017, **1865**, 1664–1675.
  - 47 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
  - 48 K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid and E. Harder, Docking covalent inhibitors: A parameter free approach to pose prediction and scoring, *J. Chem. Inf. Model.*, 2014, **54**, 1932–1940.
  - 49 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.*, 1997, **267**, 727–748.
  - 50 R. Abagyan, M. Totrov and D. Kuznetsov, ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation, *J. Comput. Chem.*, 1994, **15**, 488–506.
  - 51 A. Scarpino, G. G. Ferenczy and G. M. Keserü, Comparative Evaluation of Covalent Docking Tools, *J. Chem. Inf. Model.*, 2018, **58**, 1441–1458.
  - 52 C. Wen, X. Yan, Q. Gu, J. Du, D. Wu, Y. Lu, H. Zhou and J. Xu, Systematic studies on the protocol and criteria for selecting a covalent docking tool, *Molecules*, 2019, **24**, 2183.
  - 53 H. Zhang, W. Jiang, P. Chatterjee and Y. Luo, Ranking Reversible Covalent Drugs: From Free Energy Perturbation to Fragment Docking, *J. Chem. Inf. Model.*, 2019, **59**, 2093–2102.
  - 54 B. Kuhn, M. Tichý, L. Wang, S. Robinson, R. E. Martin, A. Kuglstatter, J. Benz, M. Giroud, T. Schirmeister, R. Abel, F. Diederich and J. Hert, Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors, *J. Med. Chem.*, 2017, **60**, 2485–2497.
  - 55 A. Böhme, D. Thaens, A. Paschke and G. Schüürmann, Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles - Application to  $\alpha,\beta$ -unsaturated ketones, acrylates, and propiolates, *Chem. Res. Toxicol.*, 2009, **22**, 742–750.
  - 56 M. E. Flanagan, J. A. Abramite, D. P. Anderson, A. Aulabaugh, U. P. Dahal, A. M.

- Gilbert, C. Li, J. Montgomery, S. R. Oppenheimer, T. Ryder, B. P. Schuff, D. P. Uccello, G. S. Walker, Y. Wu, M. F. Brown, J. M. Chen, M. M. Hayward, M. C. Noe, R. S. Obach, L. Philippe, V. Shanmugasundaram, M. J. Shapiro, J. Starr, J. Stroh and Y. Che, Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors, *J. Med. Chem.*, 2014, **57**, 10072–10079.
- 57 J. A. H. Schwöbel, J. C. Madden and M. T. D. Cronin, Examination of Michael addition reactivity towards glutathione by transition-state calculations, *SAR QSAR Environ. Res.*, 2010, **21**, 693–710.
- 58 D. Mulliner, D. Wondrousch and G. Schüürmann, Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations, *Org. Biomol. Chem.*, 2011, **9**, 8400–8412.
- 59 L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls and R. A. Friesner, Automated Transition State Search and Its Application to Diverse Types of Organic Reactions, *J. Chem. Theory Comput.*, 2017, **13**, 5780–5797.
- 60 R. Lonsdale, J. Burgess, N. Colclough, N. L. Davies, E. M. Lenz, A. L. Orton and R. A. Ward, Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity, *J. Chem. Inf. Model.*, 2017, **57**, 3124–3137.
- 61 S. Krishnan, R. M. Miller, B. Tian, R. D. Mullins, M. P. Jacobson and J. Taunton, Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis, *J. Am. Chem. Soc.*, 2014, **136**, 12624–12630.
- 62 A. Voice, G. Tresadern, H. van Vlijmen and A. Mulholland, Limitations of Ligand-Only Approaches for Predicting the Reactivity of Covalent Inhibitors, *J. Chem. Inf. Model.*, 2019, **59**, 4220–4227.
- 63 F. Palazzesi, M. A. Grundl, A. Pautsch, A. Weber and C. S. Tautermann, A Fast Ab Initio Predictor Tool for Covalent Reactivity Estimation of Acrylamides, *J. Chem. Inf. Model.*, 2019, **59**, 3565–3571.
- 64 F. Palazzesi, M. R. Hermann, M. A. Grundl, A. Pautsch, D. Seeliger, C. S. Tautermann and A. Weber, BIreactive: A Machine-Learning Model to Estimate Covalent Warhead Reactivity, *J. Chem. Inf. Model.*, 2020, **60**, 2915–2923.
- 65 E. Awoonor-Williams and C. N. Rowley, How Reactive are Druggable Cysteines in Protein Kinases?, *J. Chem. Inf. Model.*, 2018, **58**, 1935–1946.
- 66 E. Awoonor-Williams and C. N. Rowley, Evaluation of Methods for the Calculation of the pKa of Cysteine Residues in Proteins, *J. Chem. Theory Comput.*, 2016, **12**, 4662–4673.
- 67 F. Hofer, J. Kraml, U. Kahler, A. S. Kamenik and K. R. Liedl, Catalytic Site pKa Values of Aspartic, Cysteine, and Serine Proteases: Constant pH MD Simulations, *J. Chem. Inf. Model.*, 2020, **60**, 3030–3042.
- 68 R. C. Harris, R. Liu and J. Shen, Predicting Reactive Cysteines with Implicit-Solvent-Based Continuous Constant pH Molecular Dynamics in Amber, *J. Chem. Theory Comput.*, 2020, **16**, 3689–3698.
- 69 W. Zhang, J. Pei and L. Lai, Statistical Analysis and Prediction of Covalent Ligand Targeted Cysteine Residues, *J. Chem. Inf. Model.*, 2017, **57**, 1453–1460.
- 70 H. Wang, X. Chen, C. Li, Y. Liu, F. Yang and C. Wang, Sequence-Based Prediction of Cysteine Reactivity Using Machine Learning, *Biochemistry*, 2018, **57**, 451–460.
- 71 E. Awoonor-Williams and C. N. Rowley, The hydration structure of methylthiolate from QM/MM molecular dynamics, *J. Chem. Phys.*, 2018, **149**, 045103.

- 72 C. H. S. Da Costa, V. Bonatto, A. M. Dos Santos, J. Lameira, A. Leitão and C. A. Montanari, Evaluating QM/MM Free Energy Surfaces for Ranking Cysteine Protease Covalent Inhibitors, *J. Chem. Inf. Model.*, 2020, **60**, 880–889.
- 73 J. Ainsley, A. Lodola, A. J. Mulholland, C. Z. Christov and T. G. Karabenchewa-Christova, Combined Quantum Mechanics and Molecular Mechanics Studies of Enzymatic Reaction Mechanisms, in *Advances in Protein Chemistry and Structural Biology*, Academic Press Inc., 2018, vol. 113, pp. 1–32.
- 74 L. Capoferri, A. Lodola, S. Rivara and M. Mor, Quantum Mechanics/Molecular Mechanics Modeling of Covalent Addition between EGFR–Cysteine 797 and *N*-(4-Anilinoquinazolin-6-yl) Acrylamide, *J. Chem. Inf. Model.*, 2015, **55**, 589–599.
- 75 A. Lodola, L. Capoferri, S. Rivara, G. Tarzia, D. Piomelli, A. Mulholland and M. Mor, Quantum Mechanics/Molecular Mechanics Modeling of Fatty Acid Amide Hydrolase Reactivation Distinguishes Substrate from Irreversible Covalent Inhibitors, *J. Med. Chem.*, 2013, **56**, 2500–2512.
- 76 K. Świderek and V. Moliner, Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods, *Chem. Sci.*, 2020, **11**, 10626–10630.
- 77 M. W. van der Kamp and A. J. Mulholland, Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology, *Biochemistry*, 2013, **52**, 2708–2728.
- 78 A. J. Mulholland, Modelling enzyme reaction mechanisms, specificity and catalysis, *Drug Discov. Today*, 2005, **10**, 1393–1402.
- 79 K. E. Ranaghan and A. J. Mulholland, Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods, *Int. Rev. Phys. Chem.*, 2010, **29**, 65–133.
- 80 R. Lonsdale, J. N. Harvey and A. J. Mulholland, A practical guide to modelling enzyme-catalysed reactions, *Chem. Soc. Rev.*, 2012, **41**, 3025.
- 81 L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors, *Science (80-. )*, 2020, **368**, 409–412.
- 82 K. Arafet, F. V. González and V. Moliner, Quantum Mechanics/Molecular Mechanics Studies of the Mechanism of Cysteine Proteases Inhibition by Dipeptidyl Nitroalkenes, *Chem. – A Eur. J.*, 2020, **26**, 2002–2012.
- 83 T. Schirmeister, J. Kesselring, S. Jung, T. H. Schneider, A. Weickert, J. Becker, W. Lee, D. Bamberger, P. R. Wich, U. Distler, S. Tenzer, P. Johé, U. A. Hellmich and B. Engels, Quantum Chemical-Based Protocol for the Rational Design of Covalent Inhibitors, *J. Am. Chem. Soc.*, 2016, **138**, 8332–8335.
- 84 G. Milano, J. P. Spano and B. Leyland-Jones, EGFR-targeting drugs in combination with cytotoxic agents: From bench to bedside, a contrasted reality, *Br. J. Cancer*, 2008, **99**, 1–5.
- 85 C. S. Tan, N. B. Kumarakulasinghe, Y. Q. Huang, Y. L. E. Ang, J. R. E. Choo, B. C. Goh and R. A. Soo, Third generation EGFR TKIs: Current data and future directions, *Mol. Cancer*, 2018, **17**, 29.
- 86 E. Awoonor-Williams and C. Rowley, Calculating the Full Free Energy Profile for Covalent Modification of a Druggable Cysteine in Bruton’s Tyrosine Kinase, *ChemRxiv*, 2020, 1–11.
- 87 G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, The protein kinase complement of the human genome, *Science*, 2002, **298**, 1912–1934.

- 88 J. Cicens, E. Zalyte, A. Bairoch and P. Gaudet, Kinases and cancer, *Cancers (Basel)*., 2018, **10**, 63.
- 89 K. S. Bhullar, N. O. Lagarón, E. M. McGowan, I. Parmar, A. Jha, B. P. Hubbard and H. P. V. Rupasinghe, Kinase-targeted cancer therapies: Progress, challenges and future directions, *Mol. Cancer*, 2018, **17**, 1–20.
- 90 M. Hopper, T. Gururaja, T. Kinoshita, J. P. Dean, R. J. Hill and A. Mongan, Relative Selectivity of Covalent Inhibitors Requires Assessment of Inactivation Kinetics and Cellular Occupancy: A Case Study of Ibrutinib and Acalabrutinib, *J. Pharmacol. Exp. Ther.*, 2020, **372**, 331–338.
- 91 W. Zhou, D. Ercan, L. Chen, C. H. Yun, D. Li, M. Capelletti, A. B. Cortot, L. Chirieac, R. E. Iacob, R. Padera, J. R. Engen, K. K. Wong, M. J. Eck, N. S. Gray and P. A. Jänne, Novel mutant-selective EGFR kinase inhibitors against EGFR T790M, *Nature*, 2009, **462**, 1070–1074.
- 92 C. Sheridan, Companies in rapid pursuit of BTK immunokinase, *Nat. Biotechnol.*, 2012, **30**, 199–200.
- 93 R. A. Ward, M. J. Anderton, S. Ashton, P. A. Bethel, M. Box, S. Butterworth, N. Colclough, C. G. Chorley, C. Chuaqui, D. A. E. Cross, L. A. Dakin, J. É. Debreczeni, C. Eberlein, M. R. V. Finlay, G. B. Hill, M. Grist, T. C. M. Klinowska, C. Lane, S. Martin, J. P. Orme, P. Smith, F. Wang and M. J. Waring, Structure- and Reactivity-Based Development of Covalent Inhibitors of the Activating and Gatekeeper Mutant Forms of the Epidermal Growth Factor Receptor (EGFR), *J. Med. Chem.*, 2013, **56**, 7025–7048.
- 94 F. B. Akher, A. Farrokhzadeh, N. Ravenscroft and M. M. Kuttel, A Mechanistic Study of a Potent and Selective Epidermal Growth Factor Receptor Inhibitor against the L858R/T790M Resistance Mutation, *Biochemistry*, 2019, **58**, 4246–4259.
- 95 R. Castelli, N. Bozza, A. Cavazzoni, M. Bonelli, F. Vacondio, F. Ferlenghi, D. Callegari, C. Silva, S. Rivara, A. Lodola, G. Digiacomo, C. Fumarola, R. Alfieri, P. G. Petronini and M. Mor, Balancing reactivity and antitumor activity: heteroarylthioacetamide derivatives as potent and time-dependent inhibitors of EGFR, *Eur. J. Med. Chem.*, 2019, **162**, 507–524.
- 96 P. K. Singh and O. Silakari, Molecular dynamics guided development of indole based dual inhibitors of EGFR (T790M) and c-MET, *Bioorg. Chem.*, 2018, **79**, 163–170.
- 97 J. M. Bradshaw, The Src, Syk, and Tec family kinases: Distinct types of molecular switches, *Cell. Signal.*, 2010, **22**, 1175–1184.
- 98 S. Pal Singh, F. Dammeijer and R. W. Hendriks, Role of Bruton’s tyrosine kinase in B cells and malignancies, *Mol. Cancer*, 2018, **17**, 57.
- 99 A. T. Bender, A. Gardberg, A. Pereira, T. Johnson, Y. Wu, R. Grenningloh, J. Head, F. Morandi, P. Haselmayer and L. Liu-Bujalski, Ability of Bruton’s tyrosine kinase inhibitors to sequester Y551 and prevent phosphorylation determines potency for inhibition of Fc receptor but not B-cell receptor signaling, *Mol. Pharmacol.*, 2017, **91**, 208–219.
- 100 M. M. Sultan, R. A. Denny, R. Unwalla, F. Lovering and V. S. Pande, Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain, *Sci. Rep.*, 2017, **7**, 15604.
- 101 M. S. Davids and J. R. Brown, Ibrutinib: A first in class covalent inhibitor of Bruton’s tyrosine kinase, *Futur. Oncol.*, 2014, **10**, 957–967.
- 102 C. Liang, D. Tian, X. Ren, S. Ding, M. Jia, M. Xin and S. Thareja, The development of Bruton’s tyrosine kinase (BTK) inhibitors from 2012 to 2017: A mini-review, *Eur. J.*

- Med. Chem.*, 2018, **151**, 315–326.
- 103 K. Kriegsmann, M. Kriegsmann and M. Witzens-Harig, Acalabrutinib, A second-generation bruton's tyrosine kinase inhibitor, in *Recent Results in Cancer Research*, Springer New York LLC, 2018, vol. 212, pp. 285–294.
  - 104 C. S. Tam, J. Trotman, S. Opat, J. A. Burger, G. Cull, D. Gottlieb, R. Harrup, P. B. Johnston, P. Marlton, J. Munoz, J. F. Seymour, D. Simpson, A. Tedeschi, R. Elstrom, Y. Yu, Z. Tang, L. Han, J. Huang, W. Novotny, L. Wang and A. W. Roberts, Phase 1 study of the selective BTK inhibitor zanubrutinib in B-cell malignancies and safety and efficacy evaluation in CLL, *Blood*, 2019, **134**, 851–859.
  - 105 D. P. Duarte, A. J. Lamontanara, G. La Sala, S. Jeong, Y. K. Sohn, A. Panjkovich, S. Georgeon, T. Kükenshöner, M. J. Marcaida, F. Pojer, M. De Vivo, D. Svergun, H. S. Kim, M. Dal Peraro and O. Hantschel, Btk SH2-kinase interface is critical for allosteric kinase activation and its targeting inhibits B-cell neoplasms, *Nat. Commun.*, 2020, **11**, 2319.
  - 106 S. D. Reiff, E. M. Muhowski, D. Guinn, A. Lehman, C. A. Fabian, C. Cheney, R. Mantel, L. Smith, A. J. Johnson, W. B. Young, A. R. Johnson, L. Liu, J. C. Byrd and J. A. Woyach, Noncovalent inhibition of C481S Bruton tyrosine kinase by GDC-0853: A new treatment strategy for ibrutinib-resistant CLL, *Blood*, 2018, **132**, 1039–1049.
  - 107 P. K. Balasubramanian, A. Balupuri and S. J. Cho, Molecular modeling studies on series of Btk inhibitors using docking, structure-based 3D-QSAR and molecular dynamics simulation: A combined approach, *Arch. Pharm. Res.*, 2016, **39**, 328–339.
  - 108 S. Sakthivel and S. K. M. Habeeb, Combined pharmacophore, virtual screening and molecular dynamics studies to identify Bruton's tyrosine kinase inhibitors, *J. Biomol. Struct. Dyn.*, 2018, **36**, 4320–4337.
  - 109 A. Wang, X. E. Yan, H. Wu, W. Wang, C. Hu, C. Chen, Z. Zhao, P. Zhao, X. Li, L. Wang, B. Wang, Z. Ye, J. Wang, C. Wang, W. Zhang, N. S. Gray, E. L. Weisberg, L. Chen, J. Liu, C. H. Yun and Q. Liu, Ibrutinib targets mutant-EGFR kinase with a distinct binding conformation, *Oncotarget*, 2016, **7**, 69760–69769.
  - 110 C. W. Zapf, B. S. Gerstenberger, L. Xing, D. C. Limburg, D. R. Anderson, N. Caspers, S. Han, A. Aulabaugh, R. Kurumbail, S. Shakya, X. Li, V. Spaulding, R. M. Czerwinski, N. Seth and Q. G. Medley, Covalent inhibitors of interleukin-2 inducible T cell kinase (ItK) with nanomolar potency in a whole-blood assay, *J. Med. Chem.*, 2012, **55**, 10047–10063.
  - 111 A. K. Ghosh, I. Samanta, A. Mondal and W. R. Liu, Covalent Inhibition in Drug Discovery, *ChemMedChem*, 2019, **14**, 889–906.
  - 112 J. Zhang, P. L. Yang and N. S. Gray, Targeting cancer with small molecule kinase inhibitors, *Nat. Rev. Cancer*, 2009, **9**, 28–39.
  - 113 Z. Zhao and P. E. Bourne, Progress with covalent small-molecule kinase inhibitors, *Drug Discov. Today*, 2018, **23**, 727–735.
  - 114 P. A. MacFaul, A. D. Morley and J. J. Crawford, A simple in vitro assay for assessing the reactivity of nitrile containing compounds, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1136–1138.
  - 115 T. A. Baillie, Metabolism and toxicity of drugs. Two decades of progress in industrial drug metabolism, *Chem. Res. Toxicol.*, 2008, **21**, 129–137.
  - 116 D. Callegari, K. E. Ranaghan, C. J. Woods, R. Minari, M. Tiseo, M. Mor, A. J. Mulholland and A. Lodola, L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib, *Chem. Sci.*, 2018, **9**, 2740–2749.



- 117 A. Lodola, L. Capoferri, S. Rivara, G. Tarzia, D. Piomelli, A. Mulholland and M. Mor, Quantum Mechanics/Molecular Mechanics Modeling of Fatty Acid Amide Hydrolase Reactivation Distinguishes Substrate from Irreversible Covalent Inhibitors, *J. Med. Chem.*, 2013, **56**, 2500–2512.
- 118 T. C. Schmidt, A. Welker, M. Rieger, P. K. Sahu, C. A. Sotriffer, T. Schirmeister and B. Engels, Protocol for Rational Design of Covalently Interacting Inhibitors, *ChemPhysChem*, 2014, **15**, 3226–3235.
- 119 R. Lonsdale, J. Burgess, N. Colclough, N. L. Davies, E. M. Lenz, A. L. Orton and R. A. Ward, Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity, *J. Chem. Inf. Model.*, 2017, **57**, 3124–3137.
- 120 E. H. Krenske, R. C. Petter and K. N. Houk, Kinetics and Thermodynamics of Reversible Thiol Additions to Mono- and Diactivated Michael Acceptors: Implications for the Design of Drugs That Bind Covalently to Cysteines, *J. Org. Chem.*, 2016, **81**, 11726–11733.
- 121 S. Kolboe, Proton affinity calculations with high level methods, *J. Chem. Theory Comput.*, 2014, **10**, 3123–3128.
- 122 L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls and R. A. Friesner, Automated Transition State Search and Its Application to Diverse Types of Organic Reactions, *J. Chem. Theory Comput.*, 2017, **13**, 5780–5797.
- 123 Z. Liu, C. Patel, J. N. Harvey and R. B. Sunoj, Mechanism and reactivity in the Morita–Baylis–Hillman reaction: the challenge of accurate computations, *Phys. Chem. Chem. Phys.*, 2017, **19**, 30647–30657.
- 124 B. Thapa and H. B. Schlegel, Density Functional Theory Calculation of pKa’s of Thiols in Aqueous Solution Using Explicit Water Molecules and the Polarizable Continuum Model, *J. Phys. Chem. A*, 2016, **120**, 5726–5735.
- 125 Y. Fu, L. Liu, R. Q. Li, R. Liu and Q. X. Guo, First-Principle Predictions of Absolute pKa’s of Organic Acids in Dimethyl Sulfoxide Solution, *J. Am. Chem. Soc.*, 2004, **126**, 814–822.
- 126 A. Fekete and I. Komáromi, Modeling the archetype cysteine protease reaction using dispersion corrected density functional methods in ONIOM-type hybrid QM/MM calculations; The proteolytic reaction of papain, *Phys. Chem. Chem. Phys.*, 2016, **18**, 32847–32861.
- 127 J. M. Smith, Y. Jami Alahmadi and C. N. Rowley, Range-Separated DFT Functionals are Necessary to Model Thio-Michael Additions, *J. Chem. Theory Comput.*, 2013, **9**, 4860–4865.
- 128 S. J. Bennie, M. W. van der Kamp, R. C. R. Pennifold, M. Stella, F. R. Manby and A. J. Mulholland, A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase, *J. Chem. Theory Comput.*, 2016, **12**, 2689–2697.
- 129 D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, B. Honig, M. Ringnalda and W. A. Goddard, Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory, *J. Am. Chem. Soc.*, 1994, **116**, 11875–11882.
- 130 B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff and B. Honig, New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects, *J. Phys. Chem.*, 1996, **100**, 11775–11788.

- 131 H. M. Senn, T. Ziegler, P. A. Nielsen, J. W. Jaroszewski, P.-O. Norrby, T. Liljefors, C. J. Cramer, D. G. Truhlar, M. Orozco, F. J. Luque, J. Tomasi, B. Mennucci, R. Cammi, J. R. Pliego and J. M. Riveros, Selected publications for further reading on continuum solvation models, *J. Am. Chem. Soc.*, 1996, **100**, 597.
- 132 A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang and R. A. Friesner, Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences, *Int. J. Quantum Chem.*, 2013, **113**, 2110–2142.
- 133 2018. Schrödinger Release 2018-3: Maestro, Schrödinger, LLC, New York, NY, Maestro | Schrödinger, .
- 134 V. J. Cee, L. P. Volak, Y. Chen, M. D. Bartberger, C. Tegley, T. Arvedson, J. McCarter, A. S. Tasker and C. Fotsch, Systematic Study of the Glutathione (GSH) Reactivity of *N*-Arylacrylamides: 1. Effects of Aryl Substitution, *J. Med. Chem.*, 2015, **58**, 9171–9178.
- 135 N. L. Haworth, Q. Wang and M. L. Coote, Modeling Flexible Molecules in Solution: A pKa Case Study, *J. Phys. Chem. A*, 2017, **121**, 5217–5225.
- 136 R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation, *J. Phys. Chem. B*, 2011, **115**, 14556–14562.
- 137 I. Funes-Ardoiz and R. S. Paton, GoodVibes: GoodVibes v1.0.1, 2016.
- 138 N. Foloppe and I.-J. Chen, Towards understanding the unbound state of drug compounds: Implications for the intramolecular reorganization energy upon binding, *Bioorg. Med. Chem.*, 2016, **24**, 2159–2189.
- 139 D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, S. Khalid, A. Kuzmanic, C. A. Laughton, J. Michel, A. J. Mulholland, E. Rosta, M. S. P. Sansom and M. W. van der Kamp, Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2019, **9**, 1393.
- 140 R. E. Amaro and A. J. Mulholland, Multiscale methods in drug design bridge chemical and biological complexity in the search for cures, *Nat. Rev. Chem.*, 2018, **2**, 0148.
- 141 I. M. Serafimova, M. A. Pufall, S. Krishnan, K. Duda, M. S. Cohen, R. L. Maglathlin, J. M. McFarland, R. M. Miller, M. Frödin and J. Taunton, Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles, *Nat. Chem. Biol.*, 2012, **8**, 471–476.
- 142 J. Van Der Reest, S. Lilla, L. Zheng, S. Zanivan and E. Gottlieb, Proteome-wide analysis of cysteine oxidation reveals metabolic sensitivity to redox stress, *Nat. Commun.*, 2018, **9**, 1–16.
- 143 L. B. Poole, The basics of thiols and cysteines in redox biology and chemistry., *Free Radic. Biol. Med.*, 2015, **80**, 148–57.
- 144 G. Ferrer-Sueta, B. Manta, H. Botti, R. Radi, M. Trujillo and A. Denicola, Factors affecting protein thiol reactivity and specificity in peroxide reduction, *Chem. Res. Toxicol.*, 2011, **24**, 434–450.
- 145 R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz and C. N. Pace, pK values of the ionizable groups of proteins, *Protein Sci.*, 2006, **15**, 1214–1218.
- 146 G. Bulaj, T. Kortemme and D. P. Goldenberg, Ionization-reactivity relationships for cysteine thiols in polypeptides, *Biochemistry*, 1998, **37**, 8965–8972.
- 147 S. Pinitglang, A. B. Watts, M. Patel, J. D. Reid, M. A. Noble, S. Gul, A. Bokth, A. Naeem, H. Patel, E. W. Thomas, S. K. Sreedharan, C. Verma and K. Brocklehurst, A

- classical enzyme active center motif lacks catalytic competence until modulated electrostatically, *Biochemistry*, 1997, **36**, 9968–9982.
- 148 J. Reijenga, A. van Hoof, A. van Loon and B. Teunissen, Development of methods for the determination of pKa values, *Anal. Chem. Insights*, 2013, **8**, 53–71.
- 149 C. L. Stanton and K. N. Houk, Benchmarking pK prediction methods for residues in proteins, *J. Chem. Theory Comput.*, 2008, **4**, 951–966.
- 150 J. C. Gordon, J. B. Myers, T. Foltá, V. Shoja, L. S. Heath and A. Onufriev, H<sup>++</sup>: A server for estimating pKas and adding missing hydrogens to macromolecules, *Nucleic Acids Res.*, 2005, **33**, W368–W371.
- 151 R. E. Georgescu, E. G. Alexov and M. R. Gunner, Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins, *Biophys. J.*, 2002, **83**, 1731–1748.
- 152 M. H. M. Olsson, C. R. SØndergaard, M. Rostkowski and J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical p K a predictions, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
- 153 C. R. SØndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values, *J. Chem. Theory Comput.*, 2011, **7**, 2284–2295.
- 154 C. I. Bayly, K. M. Merz, D. M. Ferguson, W. D. Cornell, T. Fox, J. W. Caldwell, P. A. Kollman, P. Cieplak, I. R. Gould and D. C. Spellmeyer, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 155 J. Harvey, *Computational Chemistry, Computational Chemistry*, Oxford University Press, Oxford, UK, 2018.
- 156 S. A. Hollingsworth and R. O. Dror, Molecular Dynamics Simulation for All, *Neuron*, 2018, **99**, 1129–1143.
- 157 E. J. M. Lang, L. C. Heyes, G. B. Jameson and E. J. Parker, Calculated pKa Variations Expose Dynamic Allosteric Communication Networks, *J. Am. Chem. Soc.*, 2016, **138**, 2036–2045.
- 158 M. S. Lee, F. R. Salsbury and C. L. Brooks, Constant-pH molecular dynamics using continuous titration coordinates, *Proteins Struct. Funct. Genet.*, 2004, **56**, 738–752.
- 159 J. Khandogin and C. L. Brooks, Constant pH molecular dynamics with proton tautomerism, *Biophys. J.*, 2005, **89**, 141–157.
- 160 J. Mongan, D. A. Case and J. A. McCammon, Constant pH molecular dynamics in generalized Born implicit solvent, *J. Comput. Chem.*, 2004, **25**, 2038–2048.
- 161 Y. Chen and B. Roux, Constant-pH Hybrid Nonequilibrium Molecular Dynamics-Monte Carlo Simulation Method, *J. Chem. Theory Comput.*, 2015, **11**, 3919–3931.
- 162 J. M. Swails, D. M. York and A. E. Roitberg, Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation, *J. Chem. Theory Comput.*, 2014, **10**, 1341–1352.
- 163 B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten and B. Roux, Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems, *J. Chem. Theory Comput.*, 2017, **13**, 5933–5944.
- 164 T. H. Truong, P. M.-U. Ung, P. B. Palde, C. E. Paulsen, A. Schlessinger and K. S. Carroll, Molecular Basis for Redox Activation of Epidermal Growth Factor Receptor Kinase, *Cell Chem. Biol.*, 2016, **23**, 837–848.
- 165 Y. Huang, R. C. Harris and J. Shen, Generalized Born Based Continuous Constant pH Molecular Dynamics in Amber: Implementation, Benchmarking and Analysis, *J. Chem.*

- Inf. Model.*, 2018, **58**, 1372–1383.
- 166 H. Nguyen, D. R. Roe and C. Simmerling, Improved generalized born solvent model parameters for protein simulations, *J. Chem. Theory Comput.*, 2013, **9**, 2020–2034.
- 167 R. Liu, Z. Yue, C. C. Tsai and J. Shen, Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors, *J. Am. Chem. Soc.*, 2019, **141**, 6553–6560.
- 168 A. Onufriev, D. Bashford and D. A. Case, Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model, *Proteins Struct. Funct. Genet.*, 2004, **55**, 383–394.
- 169 E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak and F. Y. Dupradeau, R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments, *Nucleic Acids Res.*, 2011, **39**, W511–W517.
- 170 D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, I. T.E. Cheatham, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, J. L. P. Li, C. Lin, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, S. P. F.Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman, AMBER 2018, University of California, San Francisco.
- 171 G. G. Rhys, C. W. Wood, E. J. M. Lang, A. J. Mulholland, R. L. Brady, A. R. Thomson and D. N. Woolfson, Maintaining and breaking symmetry in homomeric coiled-coil assemblies, *Nat. Commun.*, 2018, **9**, 4132–4132.
- 172 S. Ma, L. S. Devi-Kesavan and J. Gao, Molecular dynamics simulations of the catalytic pathway of a cysteine protease: A combined QM/MM study of human cathepsin K, *J. Am. Chem. Soc.*, 2007, **129**, 13633–13645.
- 173 S. R. Malwal, J. Gao, X. Hu, Y. Yang, W. Liu, J. W. Huang, T. P. Ko, L. Li, C. C. Chen, B. O’Dowd, R. L. Khade, Y. Zhang, Y. Zhang, E. Oldfield and R. T. Guo, Catalytic Role of Conserved Asparagine, Glutamine, Serine, and Tyrosine Residues in Isoprenoid Biosynthesis Enzymes, *ACS Catal.*, 2018, **8**, 4299–4312.
- 174 S. B. Syed, M. Shahbaaz, S. H. Khan, S. Srivastava, A. Islam, F. Ahmad and M. I. Hassan, Estimation of pH effect on the structure and stability of kinase domain of human integrin-linked kinase, *J. Biomol. Struct. Dyn.*, 2019, **37**, 156–165.
- 175 T. K. Harris and G. J. Turner, Structural basis of perturbed pKa values of catalytic groups in enzyme active sites, *IUBMB Life*, 2002, **53**, 85–98.
- 176 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general Amber force field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 177 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 178 P. Florová, P. Sklenovský, P. Banáš and M. Otyepka, Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact, *J. Chem. Theory Comput.*, 2010, **6**, 3569–3579.
- 179 W. L. Jorgensen and J. Tirado-Rives, Potential energy functions for atomic-level simulations of water and organic and biomolecular systems, *Proc. Natl. Acad. Sci. U.*

- S. A., 2005, **102**, 6665–6670.
- 180 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 181 M. W. Mahoney and W. L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *J. Chem. Phys.*, 2000, **112**, 8910–8922.
- 182 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, The missing term in effective pair potentials, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
- 183 R. Bürigi, P. A. Kollman and W. F. Van Gunsteren, Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation, *Proteins Struct. Funct. Genet.*, 2002, **47**, 469–480.
- 184 A. V. Yeager, J. M. Swails and B. R. Miller, Improved Accuracy for Constant pH-REMD Simulations through Modification of Carboxylate Effective Radii, *J. Chem. Theory Comput.*, 2017, **13**, 4624–4635.
- 185 A. Onufriev, D. A. Case and D. Bashford, Effective Born radii in the generalized Born approximation: The importance of being perfect, *J. Comput. Chem.*, 2002, **23**, 1297–1304.
- 186 J. Wang, P. Cieplak, J. Li, Q. Cai, M. J. Hsieh, R. Luo and Y. Duan, Development of polarizable models for molecular mechanical calculations. 4. van der waals parametrization, *J. Phys. Chem. B*, 2012, **116**, 7088–7101.
- 187 A. Voice, G. Tresdern, R. M. Twidale, H. Van Vlijmen and A. J. Mulholland, Mechanism of covalent binding of ibrutinib to Bruton’s tyrosine kinase revealed by QM/MM calculations, *Chem. Sci.*, 2021, 1–7.
- 188 P. A. Jackson, J. C. Widen, D. A. Harki and K. M. Brummond, Covalent Modifiers: A Chemical Perspective on the Reactivity of  $\alpha,\beta$ -Unsaturated Carbonyls with Thiols via Hetero-Michael Addition Reactions, *J. Med. Chem.*, 2017, **60**, 839–885.
- 189 C. Yin, F. Huo, J. Zhang, R. Martínez-Mañez, Y. Yang, H. Lv and S. Li, Thiol-addition reactions and their applications in thiol recognition, *Chem. Soc. Rev.*, 2013, **42**, 6032–6059.
- 190 D. P. Nair, M. Podgórski, S. Chatani, T. Gong, W. Xi, C. R. Fenoli and C. N. Bowman, The Thiol-Michael addition click reaction: A powerful and widely used tool in materials chemistry, *Chem. Mater.*, 2014, **26**, 724–744.
- 191 Y. Sun, H. Liu, L. Cheng, S. Zhu, C. Cai, T. Yang, L. Yang and P. Ding, Thiol Michael addition reaction: a facile tool for introducing peptides into polymer-based gene delivery systems, *Polym. Int.*, 2018, **67**, 25–31.
- 192 J. S. Brown, A. W. Ruttinger, A. J. Vaidya, C. A. Alabi and P. Clancy, Decomplexation as a rate limitation in the thiol-Michael addition of: N -acrylamides, *Org. Biomol. Chem.*, 2020, **18**, 6364–6377.
- 193 R. G. Pearson and J. Songstad, Application of the Principle of Hard and Soft Acids and Bases to Organic Chemistry, *J. Am. Chem. Soc.*, 1967, **89**, 1827–1836.
- 194 J. A. H. Schwöbel, Y. K. Koleva, S. J. Enoch, F. Bajot, M. Hewitt, J. C. Madden, D. W. Roberts, T. W. Schultz and M. T. D. Cronin, Measurement and estimation of electrophilic reactivity for predictive toxicology, *Chem. Rev.*, 2011, **111**, 2562–2596.
- 195 J. Clayden, N. Greeves, S. Warren and P. Wothers, *Organic Chemistry (1st ed.)*, *Organic Chemistry (1st ed.)*, Oxford University Press, 1st edn., 2001.
- 196 C. J. Rosenker, E. H. Krenske, K. N. Houk and P. Wipf, Influence of base and structure in the reversible covalent conjugate addition of thiol to polycyclic enone scaffolds, *Org.*

- Lett.*, 2013, **15**, 1076–1079.
- 197 J. Chen, X. Jiang, S. L. Carroll, J. Huang and J. Wang, Theoretical and Experimental Investigation of Thermodynamics and Kinetics of Thiol-Michael Addition Reactions: A Case Study of Reversible Fluorescent Probes for Glutathione Imaging in Single Cells, *Org. Lett.*, 2015, **17**, 5978–5981.
  - 198 B. E. Thomas and P. A. Kollman, An ab Initio Molecular Orbital Study of the First Step of the Catalytic Mechanism of Thymidylate Synthase: The Michael Addition of Sulfur and Oxygen Nucleophiles, *J. Org. Chem.*, 1995, **60**, 8375–8381.
  - 199 A. Paasche, M. Schiller, T. Schirmeister and B. Engels, Mechanistic Study of the Reaction of Thiol-Containing Enzymes with  $\alpha,\beta$ -Unsaturated Carbonyl Substrates by Computation and Chemoassays, *ChemMedChem*, 2010, **5**, 869–880.
  - 200 P. Carlqvist, M. Svedendahl, C. Branneby, K. Hult, T. Brinck and P. Berglund, Exploring the Active-Site of a Rationally Redesigned Lipase for Catalysis of Michael-Type Additions, *ChemBioChem*, 2005, **6**, 331–336.
  - 201 S. R. Pereira, V. M. Vasconcelos and A. Antunes, Computational study of the covalent bonding of microcystins to cysteine residues - a reaction involved in the inhibition of the PPP family of protein phosphatases, *FEBS J.*, 2013, **280**, 674–680.
  - 202 F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert and D. Baker, Computational design of catalytic dyads and oxyanion holes for ester hydrolysis, *J. Am. Chem. Soc.*, 2012, **134**, 16197–16206.
  - 203 A. J. Cohen, P. Mori-Sánchez and W. Yang, Insights into current limitations of density functional theory., *Science*, 2008, **321**, 792–4.
  - 204 Y. Zhao, N. González-Garda and D. G. Truhlar, Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.
  - 205 A. Dreuw and M. Head-Gordon, Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes, *J. Am. Chem. Soc.*, 2004, **126**, 4007–4016.
  - 206 J. Gräfenstein, E. Kraka and D. Cremer, The impact of the self-interaction error on the density functional theory description of dissociating radical cations: Ionic and covalent dissociation limits, *J. Chem. Phys.*, 2004, **120**, 524–539.
  - 207 P. Mori-Sánchez, A. J. Cohen and W. Yang, Many-electron self-interaction error in approximate density functionals, *J. Chem. Phys.*, 2006, **125**, 201102.
  - 208 D. Hait and M. Head-Gordon, Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number, *J. Phys. Chem. Lett.*, 2018, **9**, 6280–6288.
  - 209 J. R. Brown, Ibrutinib (PCI-32765), the First BTK (Bruton’s Tyrosine Kinase) Inhibitor in Clinical Trials, *Curr. Hematol. Malig. Rep.*, 2013, **8**, 1–6.
  - 210 Z. Zhao and P. E. Bourne, Progress with covalent small-molecule kinase inhibitors, *Drug Discov. Today*, 2018, **23**, 727–735.
  - 211 D. Callegari, K. E. Ranaghan, C. J. Woods, R. Minari, M. Tiseo, M. Mor, A. J. Mulholland and A. Lodola, L718Q mutant EGFR escapes covalent inhibition by stabilizing a non-reactive conformation of the lung cancer drug osimertinib, *Chem. Sci.*, 2018, **9**, 2740–2749.

- 212 S. Martí, V. Moliner and K. Świderek, Examination of the performance of semiempirical  
methods in QM/MM studies of the SN2-like reaction of an adenylyl group transfer  
catalysed by ANT4', *Theor. Chem. Acc.*, 2019, **138**, 1–12.
- 213 S. H. Vosko, L. Wilk and M. Nusair, Accurate spin-dependent electron liquid correlation  
energies for local spin density calculations: a critical analysis, *Can. J. Phys.*, 1980, **58**,  
1200–1211.
- 214 C. Lee, W. Yang and R. G. Parr, Development of the Colle-Salvetti correlation-energy  
formula into a functional of the electron density, *Phys. Rev. B*, 1988, **37**, 785–789.
- 215 A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J.*  
*Chem. Phys.*, 1993, **98**, 5648–5652.
- 216 T. Yanai, D. P. Tew and N. C. Handy, A new hybrid exchange-correlation functional  
using the Coulomb-attenuating method (CAM-B3LYP), *Chem. Phys. Lett.*, 2004, **393**,  
51–57.
- 217 Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group  
thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and  
transition elements: Two new functionals and systematic testing of four M06-class  
functionals and 12 other functionals, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 218 J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with  
damped atom–atom dispersion corrections, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.
- 219 O. A. Vydrov, J. Heyd, A. V. Krukau and G. E. Scuseria, Importance of short-range  
versus long-range Hartree-Fock exchange for the performance of hybrid density  
functionals, *J. Chem. Phys.*, 2006, **125**, 074106.
- 220 O. A. Vydrov and G. E. Scuseria, Assessment of a long-range corrected hybrid  
functional, *J. Chem. Phys.*, 2006, **125**, 234109.
- 221 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *AMI: A New General*  
*Purpose Quantum Mechanical Molecular Model1*, *AMI: A New General Purpose*  
*Quantum Mechanical Molecular Model1*, 1985, vol. 107.
- 222 J. J. P. Stewart, Optimization of parameters for semiempirical methods I. Method, *J.*  
*Comput. Chem.*, 1989, **10**, 209–220.
- 223 J. J. P. Stewart, Optimization of parameters for semiempirical methods V: Modification  
of NDDO approximations and application to 70 elements, *J. Mol. Model.*, 2007, **13**,  
1173–1213.
- 224 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk and T. Frauenheim, Self-  
consistent-charge density-functional tight-binding method for simulations of complex  
materials properties, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1998, **58**, 7260–  
7268.
- 225 M. Gaus, Q. Cui and M. Elstner, DFTB3: Extension of the self-consistent-charge  
density-functional tight-binding method (SCC-DFTB), *J. Chem. Theory Comput.*, 2011,  
**7**, 931–948.
- 226 T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The  
atoms boron through neon and hydrogen, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- 227 D. E. Woon and T. H. Dunning, Gaussian basis sets for use in correlated molecular  
calculations. III. The atoms aluminum through argon, *J. Chem. Phys.*, 1993, **98**, 1358–  
1371.
- 228 Gaussian 16, Revision A.03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,  
M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji,  
M. C. X. Li, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H.  
P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F.

- Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- 229 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
- 230 A. Grossfield, WHAM: the weighted histogram analysis method, WHAM: the weighted histogram analysis method, <http://membrane.urmc.rochester.edu/content/wham/>, (accessed 15 July 2020).
- 231 E. Awoonor-Williams, W. C. Isley, S. G. Dale, E. R. Johnson, H. Yu, A. D. Becke, B. Roux and C. N. Rowley, Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols, *J. Comput. Chem.*, 2020, **41**, 427–438.
- 232 F. Jensen, *Introduction to Computational Chemistry, 3rd Edition, Introduction to Computational Chemistry, 3rd Edition*, Wiley, 2017.
- 233 P. Mori-Sánchez, A. J. Cohen and W. Yang, Localization and delocalization errors in density functional theory and implications for band-gap prediction, *Phys. Rev. Lett.*, 2008, **100**, 146401.
- 234 Y. Zhang and W. Yang, Perspective on ‘Density-functional theory for fractional particle number: Derivative discontinuities of the energy’, *Theor. Chem. Acc.*, 2000, **103**, 346–348.
- 235 J. P. Perdew, R. G. Parr, M. Levy and J. L. Balduz, Density-functional theory for fractional particle number: Derivative discontinuities of the energy, *Phys. Rev. Lett.*, 1982, **49**, 1691–1694.
- 236 T. M. Henderson, B. G. Janesko and G. E. Scuseria, Range separation and local Hybridization in density functional theory, *J. Phys. Chem. A*, 2008, **112**, 12530–12542.
- 237 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, A long-range correction scheme for generalized-gradient-approximation exchange functionals, *J. Chem. Phys.*, 2001, **115**, 3540–3544.
- 238 J. Toulouse, F. Colonna and A. Savin, Long-range - Short-range separation of the electron-electron interaction in density-functional theory, *Phys. Rev. A - At. Mol. Opt. Phys.*, 2004, **70**, 062505.
- 239 É. Brémond, Á. J. Pérez-Jiménez, J. C. Sancho-García and C. Adamo, Range-separated hybrid density functionals made simple, *J. Chem. Phys.*, 2019, **150**, 201102.
- 240 C. Adamo and V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 241 C. J. Cramer, *Essentials of Computational Chemistry Theories and Models Second Edition, Essentials of Computational Chemistry Theories and Models Second Edition*, Wiley, West Sussex, England, 2nd edn., 2004.
- 242 R. Pariser and R. G. Parr, A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I, *J. Chem. Phys.*, 1953, **21**, 466–471.
- 243 A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, Semiempirical Quantum



- Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications, *Chem. Rev.*, 2016, **116**, 5301–5337.
- 244 M. Gaus, X. Lu, M. Elstner and Q. Cui, Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications, *J. Chem. Theory Comput.*, 2014, **10**, 1518–1537.
- 245 Y.-S. Wang, Y.-D. Lin and S. Der Chao, Hydrogen-bonding Structures and Energetics of Acrylamide Isomers, Tautomers, and Dimers: An *ab initio* Study and Spectral Analysis, *J. Chinese Chem. Soc.*, 2016, **63**, 968–976.
- 246 J. R. A. Silva, L. Cianni, D. Araujo, P. H. J. Batista, D. De Vita, F. Rosini, A. Leitão, J. Lameira and C. A. Montanari, Assessment of the Cruzain Cysteine Protease Reversible and Irreversible Covalent Inhibition Mechanism, *J. Chem. Inf. Model.*, 2020, **60**, 1666–1677.
- 247 X. Wang, G. M. Bakanina Kissanga, E. Li, Q. Li and J. Yao, The catalytic mechanism of: S -acyltransferases: Acylation is triggered on by a loose transition state and deacylation is turned off by a tight transition state, *Phys. Chem. Chem. Phys.*, 2019, **21**, 12163–12172.
- 248 R. B. Woodward and R. Hoffmann, Stereochemistry of Electrocyclic Reactions, *J. Am. Chem. Soc.*, 1965, **87**, 395–397.
- 249 V. Mlýnský, P. Banáš, J. Šponer, M. W. Van Der Kamp, A. J. Mulholland and M. Otyepka, Comparison of *ab Initio*, DFT, and semiempirical QM/MM approaches for description of catalytic mechanism of hairpin ribozyme, *J. Chem. Theory Comput.*, 2014, **10**, 1608–1622.
- 250 M. P. Lutolf and J. A. Hubbell, Synthesis and physicochemical characterization of end-linked poly(ethylene glycol)-co-peptide hydrogels formed by Michael-type addition, *Biomacromolecules*, 2003, **4**, 713–722.
- 251 M. P. Lutolf, N. Tirelli, S. Cerritelli, L. Cavalli and J. A. Hubbell, Systematic modulation of Michael-type reactivity of thiols through the use of charged amino acids, *Bioconjug. Chem.*, 2001, **12**, 1051–1056.
- 252 K. Arafet, S. Ferrer and V. Moliner, Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease, *ACS Catal.*, 2017, **7**, 1207–1215.
- 253 A. M. Dos Santos, L. Cianni, D. De Vita, F. Rosini, A. Leitão, C. A. Laughton, J. Lameira and C. A. Montanari, Experimental study and computational modelling of cruzain cysteine protease inhibition by dipeptidyl nitriles, *Phys. Chem. Chem. Phys.*, 2018, **20**, 24317–24328.
- 254 G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 255 J. Kästner, Umbrella sampling, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 932–942.
- 256 J. Kästner and W. Thiel, Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: ‘umbrella integration’, *J. Chem. Phys.*, 2005, **123**, 144104.
- 257 J. Kästner and W. Thiel, Analysis of the statistical error in umbrella sampling simulations by umbrella integration, *J. Chem. Phys.*, 2006, **124**, 234106.
- 258 M. J. Byrne, N. R. Lees, L. C. Han, M. W. Van Der Kamp, A. J. Mulholland, J. E. M. Stach, C. L. Willis and P. R. Race, The Catalytic Mechanism of a Natural Diels-Alderase Revealed in Molecular Detail, *J. Am. Chem. Soc.*, 2016, **138**, 6095–6098.
- 259 S. Re, H. Oshima, K. Kasahara, M. Kamiya and Y. Sugita, Encounter complexes and hidden poses of kinase inhibitor binding on the free-energy landscape, *Proc. Natl. Acad.*

- Sci. U. S. A.*, 2019, **116**, 18404–18409.
- 260 H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan and T. Hou, Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations, *Sci. Rep.*, 2015, **5**, 1–8.
- 261 K. E. Ranaghan and A. J. Mulholland, Conformational effects in enzyme catalysis: QM/MM free energy calculation of the ‘NAC’ contribution in chorismate mutase, *Chem. Commun.*, 2004, **10**, 1238–1239.
- 262 A. Lodola, M. Mor, J. Zurek, G. Tarzia, D. Piomelli, J. N. Harvey and A. J. Mulholland, Conformational effects in enzyme catalysis: Reaction via a high energy conformation in fatty acid amide hydrolase, *Biophys. J.*, 2007, **92**, 20.
- 263 K. Arafet, K. Świderek and V. Moliner, Computational Study of the Michaelis Complex Formation and the Effect on the Reaction Mechanism of Cruzain Cysteine Protease, *ACS Omega*, 2018, **3**, 18613–18622.
- 264 K. Arafet, S. Ferrer, F. V. González and V. Moliner, Quantum mechanics/molecular mechanics studies of the mechanism of cysteine protease inhibition by peptidyl-2,3-epoxyketones, *Phys. Chem. Chem. Phys.*, 2017, **19**, 12740–12748.
- 265 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods, *ACS Catal.*, 2020, **10**, 12544–12554.
- 266 F. Zhu and G. Hummer, Convergence and error estimation in free energy calculations using the weighted histogram analysis method, *J. Comput. Chem.*, 2012, **33**, 453–465.
- 267 M. Souaille and B. Roux, Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations, *Comput. Phys. Commun.*, 2001, **135**, 40–57.
- 268 S. Grimme, C. Bannwarth and P. Shushkov, A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
- 269 J. J. Wilke, M. C. Lind, H. F. Schaefer, A. G. Császár and W. D. Allen, Conformers of gaseous cysteine, *J. Chem. Theory Comput.*, 2009, **5**, 1511–1523.
- 270 C. Bannwarth, S. Ehlert and S. Grimme, GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 271 R. Mera-Adasme, M. Domínguez and O. Denis-Alpizar, A benchmark for the size of the QM system required for accurate hybrid QM/MM calculations on the metal site of the protein copper, zinc superoxide dismutase, *J. Mol. Model.*, 2019, **25**, 1–7.
- 272 P. Pokorná, H. Kruse, M. Krepl and J. Šponer, QM/MM Calculations on Protein-RNA Complexes: Understanding Limitations of Classical MD Simulations and Search for Reliable Cost-Effective QM Methods, *J. Chem. Theory Comput.*, 2018, **14**, 5419–5433.
- 273 E. Rosta and G. Hummer, Free energies from dynamic weighted histogram analysis using unbiased Markov state model, *J. Chem. Theory Comput.*, 2015, **11**, 276–285.
- 274 L. S. Stelzl, A. Kells, E. Rosta and G. Hummer, Dynamic Histogram Analysis To Determine Free Energies and Rates from Biased Simulations, *J. Chem. Theory Comput.*, 2017, **13**, 6328–6342.
- 275 S. Ito, Y. Wang, Y. Okamoto and S. Irle, Quantum chemical replica-exchange umbrella sampling molecular dynamics simulations reveal the formation mechanism of iron phthalocyanine from iron and phthalonitrile, *J. Chem. Phys.*, 2018, **149**, 72332.

- 276 S. Ito, S. Irle and Y. Okamoto, Implementation of replica-exchange umbrella sampling in the DFTB+ semiempirical quantum chemistry package, *Comput. Phys. Commun.*, 2016, **204**, 1–10.
- 277 K. E. Ranaghan, D. Shchepanovska, S. J. Bennie, N. Lawan, S. J. Macrae, J. Zurek, F. R. Manby and A. J. Mulholland, Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution, *J. Chem. Inf. Model.*, 2019, **59**, 2063–2078.
- 278 M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby and T. F. Miller, Embedded mean-field theory, *J. Chem. Theory Comput.*, 2015, **11**, 568–580.
- 279 F. Ding, F. R. Manby and T. F. Miller, Embedded Mean-Field Theory with Block-Orthogonalized Partitioning, *J. Chem. Theory Comput.*, 2017, **13**, 1605–1615.
- 280 Z. Pan, Bruton’s tyrosine kinase as a drug discovery target, *Drug News Perspect.*, 2008, **21**, 357–362.
- 281 Z. Zhang, D. Zhang, Y. Liu, D. Yang, F. Ran, M. L. Wang and G. Zhao, Targeting Bruton’s tyrosine kinase for the treatment of B cell associated malignancies and autoimmune diseases: Preclinical and clinical developments of small molecule inhibitors, *Arch. Pharm. (Weinheim)*, 2018, **351**, 1700369.
- 282 C. Owen, N. L. Berinstein, A. Christofides and L. H. Sehn, Review of Bruton tyrosine kinase inhibitors for the treatment of relapsed or refractory mantle cell lymphoma., *Curr. Oncol.*, 2019, **26**, e233–e240.
- 283 G. Roos, N. Foloppe and J. Messens, Understanding the pKa of redox cysteines: The key role of hydrogen bonding, *Antioxidants Redox Signal.*, 2013, **18**, 94–127.
- 284 N. B. Arslan and N. Özdemir, Direct and solvent-assisted keto–enol tautomerism and hydrogen-bonding interactions in 4-(m-chlorobenzylamino)-3-phenyl-4,5-dihydro-1H-1,2,4-triazol-5-one: a quantum-chemical study, *J. Mol. Model.*, 2015, **21**, 1–10.
- 285 J. J. Vollmer and K. L. Servis, Woodward-Hoffmann rules: Electrocyclic reactions, *J. Chem. Educ.*, 1968, **45**, 214–220.
- 286 E. G. Jayasree and S. Sreedevi, A DFT study on protic solvent assisted tautomerization of heterocyclic thiocarbonyls, *Chem. Phys.*, 2020, **530**, 110650.
- 287 C. A. Bayse, Transition states for cysteine redox processes modeled by DFT and solvent-assisted proton exchange, *Org. Biomol. Chem.*, 2011, **9**, 4748–4751.
- 288 M. J. Ajitha and K. W. Huang, Role of keto-enol tautomerization in a chiral phosphoric acid catalyzed asymmetric thiocarboxylsis of meso-epoxide: A DFT study, *Org. Biomol. Chem.*, 2015, **13**, 10981–10985.
- 289 C. Trujillo, O. Mó and M. Yáñez, A theoretical study of hydration effects on the prototropic tautomerism of selenouracils, *Org. Biomol. Chem.*, 2007, **5**, 3092–3099.
- 290 G. Alagona and C. Ghio, Keto-enol tautomerism in linear and cyclic  $\beta$ -diketones: A DFT study in vacuo and in solution, *Int. J. Quantum Chem.*, 2008, **108**, 1840–1855.
- 291 R. Marrero-Carballo, F. Tun-Rosado, G. J. Mena-Rejón, D. Cáceres-Castillo, J. Barroso, F. Murillo, G. Merino and R. F. Quijano-Quñones, The base-catalyzed keto-enol tautomerism of chrysophanol anthrone. A DFT investigation of the base-catalyzed reaction, *Mol. Simul.*, 2019, **45**, 716–723.
- 292 K. Jana and B. Ganguly, DFT Study to Explore the Importance of Ring Size and Effect of Solvents on the Keto-Enol Tautomerization Process of  $\alpha$ - And  $\beta$ -Cyclodiones, *ACS Omega*, 2018, **3**, 8429–8439.
- 293 A. Kaptein, G. de Bruin, M. Emmelot-van Hoek, B. van de Kar, A. de Jong, M. Gulrajani, D. Demont, T. Covey, D. Mittag and T. Barf, Potency and Selectivity of BTK Inhibitors in Clinical Development for B-Cell Malignancies, *Blood*, 2018, **132**,

- 1871–1871.
- 294 M. Gruden, L. Andjeklović, A. K. Jissy, S. Stepanović, M. Zlatar, Q. Cui and M. Elstner, Benchmarking density functional tight binding models for barrier heights and reaction energetics of organic molecules, *J. Comput. Chem.*, 2017, **38**, 2171–2185.
- 295 J. I. Mujika, X. Lopez and A. J. Mulholland, Mechanism of C-terminal intein cleavage in protein splicing from QM/MM molecular dynamics simulations, *Org. Biomol. Chem.*, 2012, **10**, 1207–1218.
- 296 S.-T. Bong, L. N.-S. Law and J. W.-F. Law, A review on the potential of Bruton’s tyrosine kinase (Btk) inhibitor – Ibrutinib for treatment of Multiple Myeloma (MM), *Prog. Microbes Mol. Biol.*, 2019, **2**, 28.
- 297 A. Licican, L. Serafini, W. Xing, G. Czerwieniec, B. Steiner, T. Wang, K. M. Brendza, J. D. Lutz, K. S. Keegan, A. S. Ray, B. E. Schultz, R. Sakowicz and J. Y. Feng, Biochemical characterization of tirabrutinib and other irreversible inhibitors of Bruton’s tyrosine kinase reveals differences in on - and off - target inhibition, *Biochim. Biophys. Acta - Gen. Subj.*, 2020, **1864**, 129531.
- 298 K. Zinovjev and I. Tuñón, Reaction coordinates and transition states in enzymatic catalysis, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, e1329.
- 299 L. Maragliano and E. Vanden-Eijnden, On-the-fly string method for minimum free energy paths calculation, *Chem. Phys. Lett.*, 2007, **446**, 182–190.
- 300 L. Maragliano, A. Fischer, E. Vanden-Eijnden and G. Ciccotti, String method in collective variables: Minimum free energy paths and isocommittor surfaces, *J. Chem. Phys.*, 2006, **125**, 024106.
- 301 D. Branduardi, F. L. Gervasio and M. Parrinello, From A to B in free energy space, *J. Chem. Phys.*, 2007, **126**, 054103.
- 302 K. Zinovjev, J. J. Ruiz-Pernía and I. Tuñón, Toward an automatic determination of enzymatic reaction mechanisms and their activation free energies, *J. Chem. Theory Comput.*, 2013, **9**, 3740–3749.
- 303 K. Zinovjev, S. Martí and I. Tuñón, A collective coordinate to obtain free energy profiles for complex reactions in condensed phases, *J. Chem. Theory Comput.*, 2012, **8**, 1795–1801.
- 304 K. Zinovjev and I. Tuñón, Adaptive Finite Temperature String Method in Collective Variables, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.
- 305 R. Lonsdale, K. T. Houghton, J. Żurek, C. M. Bathelt, N. Foloppe, M. J. de Groot, J. N. Harvey and A. J. Mulholland, Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9, *J. Am. Chem. Soc.*, 2013, **135**, 8001–8015.
- 306 Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, M. C. X. Li, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian, Inc.,

Wallingford CT, 2016.