Modeling Covalent-Modifier Drugs

Ernest Awoonor-Williams, Andrew G. Walsh, and Christopher N. Rowley

Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

Abstract

In this review, we present a summary of how computer modeling has been used in the development of covalent modifier drugs. Covalent modifier drugs bind by forming a chemical bond with their target. This covalent binding can improve the selectivity of the drug for a target with complementary reactivity and result in increased binding affinities due to the strength of the covalent bond formed. In some cases, this results in irreversible inhibition of the target, but some targeted covalent inhibitor (TCI) drugs bind covalently but reversibly. Computer modeling is widely used in drug discovery, but different computational methods must be used to model covalent modifiers because of the chemical bonds formed. Structural and bioinformatic analysis has identified sites of modification that could yield selectivity for a chosen target. Docking methods, which are used to rank binding poses of large sets of inhibitors, have been augmented to support the formation of protein–ligand bonds and are now capable of predicting the binding pose of covalent modifiers accurately. The pK_a's of amino acids can be calculated in order to assess their reactivity towards electrophiles. QM/MM methods have been used to model the reaction mechanisms of covalent modification. The continued development of these tools will allow computation to aid in the development of new covalent modifier drugs.

Keywords: review, covalent modifiers, irreversible inhibition, computer modeling, docking, QM/MM, bioinformatics, pK_a, cysteine, Michael addition, kinase

1. Introduction

The general mechanism for the inhibition of an enzyme or receptor by a small molecule drug is for the drug to bind to the protein, attenuating

its activity. The canonical mode by which a drug will bind to its target is through non-covalent interactions, such as hydrogen bonding, dipole–dipole interactions, and London dispersion interactions. Kollman and coworkers estimated that small molecules that bind to proteins through non-covalent interactions have a maximum binding affinity of 6.3 kJ/mol per non-hydrogen atom [1], so these binding energies are generally sufficiently weak for the binding to be reversible. This establishes a measurable equilibrium between the bound and unbound states.

A sizable class of drugs bind to their targets by an additional mode; a covalent bond is formed between the ligand and its target. These drugs contain a moiety that can undergo a chemical reaction with an amino acid side chain of the target protein, covalently modifying the protein. Dissociating this covalent modifier from the target requires these protein–ligand bonds to be broken. In the cases where the dissociation is strongly exergonic, the equilibrium will lie so far towards the bound state that the inhibition is effectively irreversible. Some covalent protein–ligand reactions are only weakly exergonic, so covalent modification can be reversible in some instances. These ligands also interact with the protein through conventional non-covalent intermolecular forces, so the total binding energy in the covalently-bound state results from both covalent and non-covalent interactions. Refs. [2, 3, 4, 5, 6, 7] are recent reviews on covalent-modifier drugs.

The covalent modification of proteins can involve many types of chemical motifs in the inhibitor and involve a variety of amino acids. Catalytic residues in the active site often have depressed pK_a's, so they are more likely to occupy the deprotonated state that is reactive towards electrophiles. Catalytic serines have been a popular target and two of the most famous covalent modifier drugs (Figure 1) act by acylating a catalytic serine; aspirin targets Ser530 of cyclooxygenase and penicillin targets DD-transpeptidase. Catalytic series, threonine, and cysteine residues have all been targeted by covalent modifiers [8].

In recent years, there have been extensive efforts to develop covalent modifier drugs that undergo reactions with the thiol group of non-catalytic cysteine residues. Cysteines are relatively rare amino acids, comprising only 2.3% of the residues in the human proteome [9]. This limits the number of off-target reactions that are possible. These non-catalytic residues are less likely to be conserved within a family of proteins, which creates opportunities to select for a specific target in a large family of proteins. Although this type of covalent modification does not inactivate the catalytic residues directly, the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

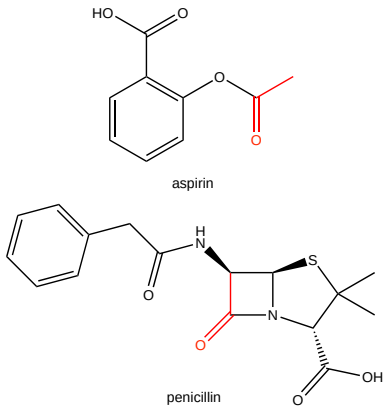

Figure 1: Aspirin and penicillin are early examples of covalent modifier drugs. These drugs bind to their target by acetylating a catalytic serine residue. The reactive moiety is drawn in red.

covalent linker serves to anchor the drug binding site and achieve a stronger binding affinity. Large-scale screens have shown that reactive fragments have unexpectedly high specificity for individual proteins, suggesting that covalent modifiers have a lower risk of promiscuity than had been assumed [10, 11, 12].

These advantages must be balanced against the drawbacks associated with covalent protein–ligand binding. Covalent protein–ligand adducts are believed to trigger immune responses in some cases [13]. Further, the inhibitor must be carefully tuned so that it will only bind irreversibly to its target because irreversible inhibition of an off-target receptor could result in adverse drug reactions. The chemical reactivity of the warhead also creates the potential that the inhibitor will be chemically degraded in an inactive form through metabolism or other types of chemical reactions before it reaches the target. This constrains the reactivity of the warhead.

To develop drugs of practical use that have the advantages of covalent modification but avoid the disadvantages, researchers have developed targeted covalent inhibitors (TCIs). Typically, these compounds have a non-covalently binding framework that is highly selective for the target. For example, covalent modifier ibrutinib (Figure 2) shares the diaminopyrimidine scaffold that has been successfully employed in the development of Bruton's tyrosine kinase-selective non-covalent inhibitors. The reactive warhead is an acrylamide, which is a moderately-reactive electrophile. Thio-Michael additions are generally only weakly exergonic, so these additions are often

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

reversible [14]. This allows the inhibitor to dissociate if it reacts with a thiol of a protein other than its target. High selectivity is achieved by the combination of selective non-covalent interactions and the additional strength of the covalent interaction between the warhead and a complementary reactive amino acid.

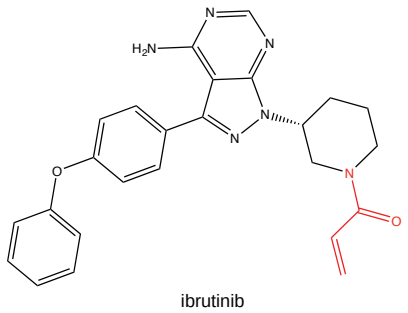

Figure 2: Ibrutinib is an example of a targeted covalent inhibitor. The reactive acrylamide warhead is indicated in red.

1.1. Physical Parameters of Covalent Modification

The strength of the binding of a non-covalent inhibitor to its target can be quantified by the equilibrium constant, K_I, for the association of the ligand (inhibitor, I) and its target (enzyme, E) to form a protein–ligand complex (E·I). This association can also be defined terms of the Gibbs energy of binding through the relation $\Delta G_{non-covalent} = -RT \ln K_I C^\circ$, where C° is the standard state concentration [15]. The binding of a covalent modifier involves additional steps. The protein–ligand complex (E·I) undergoes reaction to form the covalent adduct (I–E). The rate of this process is characterized by its rate constant, $k_{inact.}$. In some cases, this reaction is reversible and the covalent adduct can revert to the non-covalent protein–ligand complex with the rate constant $k_{-inact.}$. If the reaction is strongly exergonic, $k_{inact.}$ will be much larger than $k_{-inact.}$, so the binding will effectively be irreversible. The total binding energy of ligand is results from both the covalent and non-covalent protein–ligand interactions ($\Delta G_{non-covalent}$).

$$E + I \underset{}{\overset{K_I}{\rightleftharpoons}} E{\cdot}I \underset{k_{-inact.}}{\overset{k_{inact.}}{\rightleftharpoons}} E\text{–}I$$

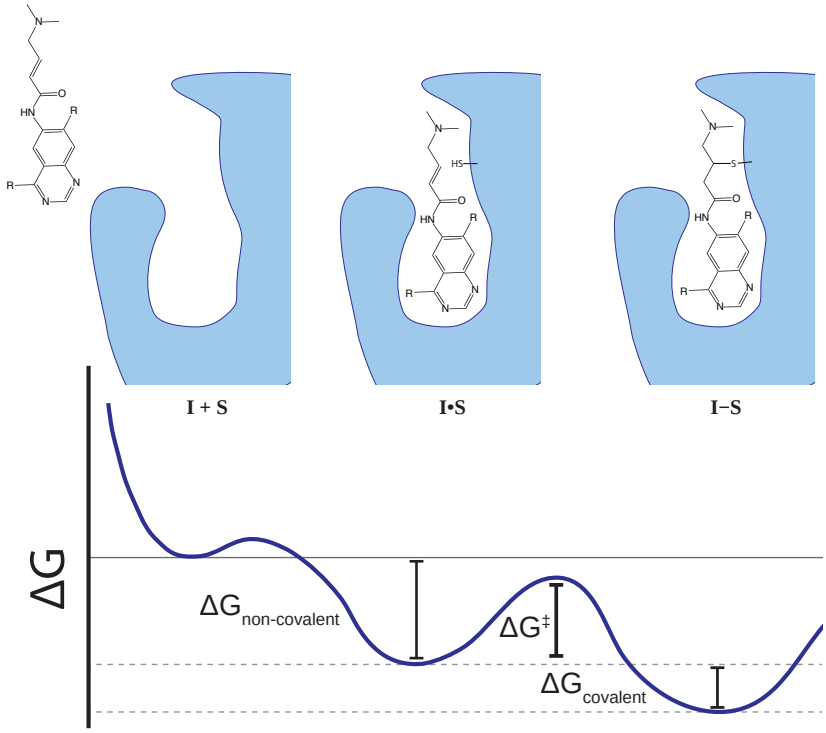

Figure 3: Schematic binding profile for the formation of a covalent protein–ligand complex. In this example, afatinib binds non-covalently to the active site of EGFR (I·S), then undergoes a chemical reaction with the thiol group of Cys-797 to form a covalent thioether adduct (I–S).

The rate at which an inhibitor reacts with the target ($k_{inact.}$) can be calculated using transition state theory. Conventional transition state theory is the simplest and most widely used theory, which relates the rate of reaction to the Gibbs energy profile along the reaction coordinate[1]. Using transition state theory, the rate of reaction can be calculated by the Gibbs energy of activation ($\Delta G^{\ddagger}$ of the rate limiting step [17]),

$$k_{TST} = \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\ddagger}}{k_B T}\right). \quad (1)$$

The mechanism of covalent modification can be complex and involve mul-

[1] A discussion of the limitations of this model in enzymatic reactions is available in Ref. 16.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

tiple reaction steps. For example, the mechanism of covalent modification of a cysteine by a Michael acceptor; deprotonation of the thiol to form a thiolate, formation of an enolate intermediate, and protonation of the enolate to form the thioether product (Figure 4). A comprehensive model for covalent modification would require calculation of $\Delta G_{non-covalent}$, $\Delta G_{covalent}$, as well as the rate-limiting barriers of the chemical reaction ($\Delta G^{\ddagger}$).

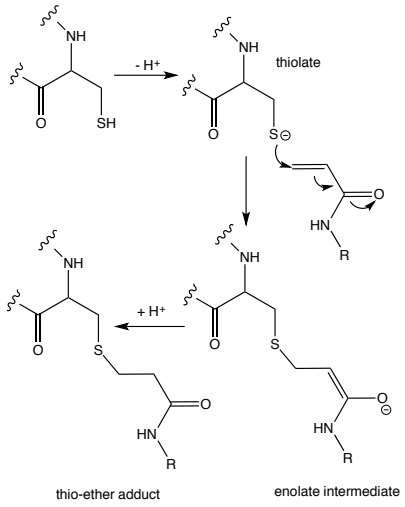

Figure 4: The mechanism of the addition of an acrylamide warhead to a cysteine thiol.

1.2. Computer Modeling in Drug Discovery

Computer modeling is used extensively in the pharmaceutical industry to aid in the development of new drugs. The membrane permeability of a drug can be estimated by empirical computational methods or molecular simulation [18, 19, 20]. Docking algorithms are used to rapidly screen large databases of compounds for their ability to bind a protein or nucleic acid that is targeted for inhibition [21, 22, 23]. Other methods, such as free energy perturbation (FEP), are used to calculate the binding affinities of a drug to a protein ($\Delta G_{non-covalent}$) [24, 25, 26, 27]. These methods are generally based on molecular mechanical force fields or other simplified representations of the protein and ligand, which typically only describe the intermolecular component of protein–ligand binding. Covalent modification inherently involves the making and breaking of chemical bonds, so these methods must be adapted to describe this mode of binding. In this review, we present methods

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

for modeling covalent-modifier drugs. Another recent review is available in Ref. [28].

2. Structural Analysis

The Protein Data Bank (PDB) now contains over 100,000 biological macromolecular structures [29]. These data have been invaluable for the development of covalent-modifier drugs. Cocrystals showing covalent bond formation between covalent-modifier drugs have helped confirm these drugs act as covalent modifiers and unambiguously indicate the site of modification. For example, Figure 5 shows the covalent binding of ibrutinib to *Toxoplasma gondii* calcium-dependent protein kinase 1 (TgCDPK1). The availability of these structures has enabled structure-based design of covalent modifiers through docking and mechanistic modeling.

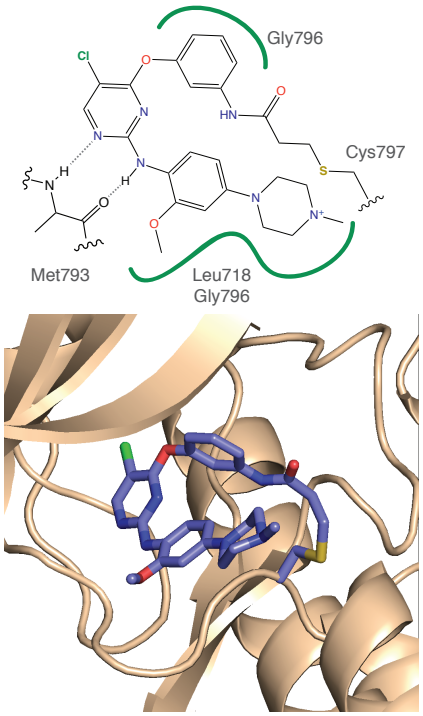

Figure 5: Ibrutinib bound to TgCDPK1 (PDB ID: 4IFG). The acrylamide warhead has undergone a Michael addition to the thiol group of Cys-797, forming a covalent thio-ether adduct between the protein and the ligand. Non-covalent interactions also stabilize the protein–ligand complex.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

This structural data is particularly valuable for designing drugs that are selective for one member of a family of structurally similar targets. The protein kinase family is a prototypical example of this. This family contains hundreds of members with a large degree of structural similarity in their kinase domains. Due to their roles in cell signaling and division, individual members of this family are drug targets, but there is a significant risk of adverse drug reactions due to c [30]. Several kinase-targeting covalent modifiers have been developed, which have the potential for higher affinity and selectivity [31, 32]. Structural analysis of the available kinase structures has been essential for identifying poorly-conserved drugable residues within the active site of the target, which allows a covalent-modifier to be designed with a warhead in a complementary position.

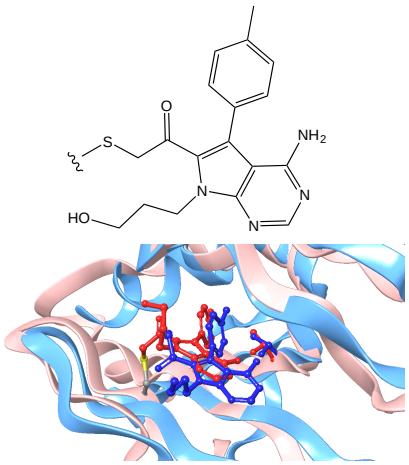

Figure 6: Docked structures of the pyrrolopyrimidine fluoromethylketone ligand (fmk) of Cohen et al. [33] bound to RSK2 (red, PDB ID: 2QR8) and c-Src kinase (blue, PDB ID: 3F6X). Both proteins have a threonine in the gatekeeper position, limiting the selectivity that can be achieved by established structural designs. Fmk binds selectively to RSK2 by forming a covalent bond to with a non-conserved cysteine residue (Cys-436). A valine (i.e. Val-281) holds this position in c-Src.

Cohen et al. reported an early structural study where a covalent-modifier drug was designed to selectively inhibit the RSK1 and RSK2 protein kinases [33]. Proteins in the kinase family have been selectively targeted by non-covalent inhibitors because a residue, known as the gatekeeper, can block binding of the ligand to a hydrophobic pocket in the active site. Inhibitors with a hydrophobic group that binds to this pocket will be selective for

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

these kinases. The RSK1, RSK2, and Src kinase all have a small threonine gatekeeper, so they all bind inhibitors of this class with similar affinity. A non-conserved cysteine is present in the glycine-rich loop on the N-terminal lobe adjacent to the binding site in RSK1 and RSK2, which is absent in Src. A pyrrolopyrimidine ligand with a fluoromethylketone warhead was synthesized, where the non-covalent interactions are optimal for a threonine-gatekeeper binding site and the warhead is situated such that it will react with the loop cysteine residue of RSK1 and RSK2. This molecule was found to selectively inhibit RSK1 and RSK2 by covalent modification, but had a lower affinity for other Src kinase proteins that lacked a cysteine residue at the targeted position. This is illustrated by docked structures in Figure 6.

Zhang et al. performed a more extensive bioinformatic analysis of the human kinase family and found that approximately 200 members of the family had a cysteine residue in the vicinity of the binding site [34]. These proteins were divided into 4 groups based on the general areas of the binding site where the cysteine is located. These residues are not highly conserved across the family, so covalent modification that targets one of these residues will be selective for a small subset of the kinase family.

A structural analysis by Leproult et al. of the active sites of kinase proteins also showed that there was significant variation in the location of active site cysteine residues [35]. The available crystal structures of human kinases were grouped into three conformational classifications: active, inactive C-helix-out, and inactive DFG-out. This analysis showed that there is a large number of targetable cysteines in the kinase family, so the strategy of designing a targeted covalent inhibitor could have wide application for these drug targets. Non-covalent inhibitor imatinib binds to the DFG-out conformation of ABL1, KIT, and PDGFRα kinases, but has the highest affinity for KIT kinase. The authors synthesized variants of imatinib with different electrophilic warheads. These inhibitors had a lower affinity for ABL1, but had a higher affinity for KIT and PDGFRα. The new inhibitors were shown to bind non-covalently to these proteins.

These structural studies of the kinase family highlight the potential of TCIs. The catalytic-domain binding sites of the human protein kinase family has limited structural variation, so it is difficult to design a selective inhibitor. Targeted covalent modifiers can be designed by adding an electrophilic warhead to existing non-covalent drug scaffolds. This warhead can only form a covalent linkage with a cysteine residue at a complementary position. Only a small number of proteins in the kinase family have a cysteine

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

residue at this position, so binding to this site by a covalent addition is selective for a small number of kinases. This design principle may also be effective to achieve selectivity for a target in a family of similar proteins.

Wu et al. generalized this type of analysis by constructing a web-based database of known covalently-bound structures collected from structural and literature databases[36]. At the time of its public release, the database contained 462 protein structures and 1217 inhibitors. This searchable database hyperlinks to entries in the PDB for the protein–ligand complex and to chemical database entries (e.g., UniPROT) of the ligands. For each druggable cysteine, the solvent-accessible surface area (SASA) and conservation within its family are also reported. The calculated pK_a of the residue is also reported, although these calculations were performed using the PROPKA3.1 package, which has poor accuracy for cysteine residues [37]. The database is accessible through the URL: `http://www.cysteinome.org/`.

3. Docking Algorithms

There are many established codes that can screen databases of small molecules for the ability to bind to a protein. For large numbers of compounds to be screened in a tractable period of time, these methods use highly efficient algorithms to estimate if a molecule has the appropriate geometry to bind to the target site. These docking algorithms also assign "scores" to the various binding poses based on estimates of the intermolecular interactions in the bound state. To allow high throughput screens of a large databases of compounds, these models are generally highly simplified, so the calculated interaction scores serve to rank the poses approximately and are not rigorous Gibbs energies of binding.

Conventional docking methods were developed to describe non-covalent protein–ligand interactions, so the stabilization that occurs through the covalent bond formation is not included by these algorithms. Moreover, covalent linkage places the ligand at a short distance from the covalently-modified residue, which is strongly disfavored by the standard steric repulsion terms in non-covalent docking algorithms. The covalent linkage also imposes additional constraints on the pose due to the geometry of warhead-residue bond. To address these issues, new methods have been developed to model the covalent docking of a ligand with the protein (Table 1).

Support for covalent docking has been implemented into the FITTED modeling suite [45, 46, 39]. Compounds containing appropriate reactive war-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

Table 1: Algorithms for predicting covalent-binding poses and the programs they are implemented in.

Algorithm	Program	Reference
MacDOCK	DOCK/MIMIC	38
FITTED	FORECASTER	39
DOCKovalent	DOCK Blaster (web)	40
CovDock	Glide/Prime	41
CovDock-VS	Glide	42
Two-point attractor / flexible side chain	AutoDock	43
DOCKTITE	MOE	44

heads are identified from the ligand database. Poses where the warhead is in close proximity to a reactive amino acid are automatically identified and used to construct a covalently-bound adduct. This approach uniquely allows both covalent and non-covalent binding models to be identified simultaneously and automatically. The binding poses are ranked by the RankScore and MatchScore algorithms. This methodology was to identify novel covalent inhibitors of prolyl oligopeptidase that exhibited high selectivity and affinity to Ser554 [47, 48, 49].

Del Rio et al. developed a computational workflow for evaluating the binding of a covalent modifier to a protein [50]. A database of kinase proteins with non-covalently-bound structures was constructed from the Protein Data Bank. From these structures, inhibitors bound near cysteine residues that had strong non-covalent binding energies were selected. These structures were used as scaffolds for the design of new inhibitors with electrophilic warheads that bind near cysteine residues. The covalently-bound protein–ligand complexes for these compounds were constructed and simulated using molecular dynamics. Inhibitors which required a large distortion in the protein structure in order to form the covalent ligand were rejected.

Ouyang et al. developed CovalentDock by modifying AutoDock5.2 [51]. This method is distinct from some other codes because ligand–protein interaction potential explicitly includes the energy of the covalent linkage. This linkage is described using a Morse potential, where the potential energy minimum forms a covalent bond but the bond can form or dissociate dynamically. This program is also available through a web-based interface, CovalentDock Cloud [52].

London et al. developed the DOCKovalent docking algorithm for covalent

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

docking using the web-based docking server DOCK Blaster [40]. This server uses a library of electrophile-containing ligands, including α, β-unsaturated carbonyls, aldehydes, boronic acids, cyanoacrylamides, alkyl halides, carbamates, α-ketoamides, and epoxides. These ligands were modified to assume the geometry they will hold when covalently bound to the target. A set of likely rotamers is generated for each compound. To search for stable binding poses, protein–ligand complexes are generated where the protein–ligand linkage is constrained to its ideal geometry. The DOCK3.6 scoring algorithm is used to rank the ligands. This method was successfully applied to a novel inhibitor of AmpC β-lactamase where a covalent bond is formed between a boronic acid warhead and the catalytic serine residue. It was also effective at finding drugs capable of binding to kinase proteins RSK2, MSK1, and JAK3 through reaction of ligands containing a cyanoacrylamide warhead with non-catalytic cysteine residues in or near the active site (Cys436 in RSK2, Cys440 in MSK1, and Cys909 in JAK3).

Bianoco et al. recently implemented two covalent docking algorithms into AutoDock Vina, a widely-used docking program [53]. The first algorithm is the two-point attractor method, where an artificial potential is defined that favors overlap between two artificial sites attached to the warhead of covalent modifiers to the C_β–O sites of a target serine or the C_β–S sites of a target cysteine. This allows the drug to move freely in the active site, while favoring conformations where the ligand is in a covalently-bound pose. The second algorithm treats the ligand as a flexible side chain of the protein. The flexible side chain method predicted the RMSD of bound poses in better agreement with the experimental structures than the two site model.

The Glide docking program developed by Schrödinger Inc. has also been modified to support modeling covalent binding ligands [41]. This algorithm mutates the target residue into an alanine residue and performs a conventional docking simulation to generate a library of hundreds of poses. These poses are filtered to select those where the warhead is within 5 Å of the target residue based on a rotamer library. Covalently-bound structures are generated from these bound poses. Further optimization and clustering is performed on these poses using the Prime refinement program [54, 55], which are ranked to yield a final optimal binding pose. Ligands can be screened by calculating an apparent affinity, which is estimated from the average of the scores for the covalently-bound and non-covalently-bound poses. On a test set of 38 covalently-bound protein–ligand complexes, this method was able to predict the binding pose with an RMSD of 1.52 Å from the exper-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

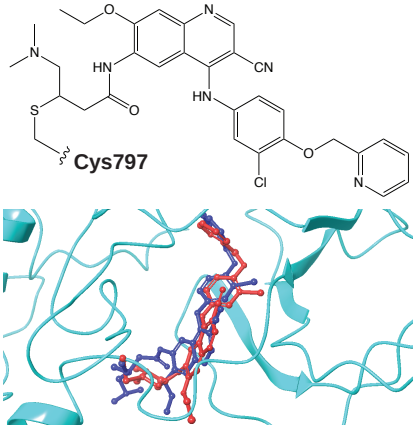

Figure 7: Binding of neratinib (HKI-272) to EGFR modeled using CovalentDock. The RMSD of the non-hydrogen atoms of the ligand are 2.14 Å with respect to the experimental crystallographic structure (PDB ID: 2JIV). The experimental and predicted ligand poses are colored in red and blue, respectively.

imental crystallographic structure. An example of the predicted pose of a covalently-bound inhibitor using this method is presented in Figure 7.

Warshaviak modified the CovDock workflow to enable fast structure-based virtual screening of covalent modifiers [42]. In the CovDock-VS workflow, the target residue is also mutated into an alanine, but a restraint is imposed so that the warhead remains within 5 Å of the target residue during the conformational search. This restricts the initial search so that only poses where covalent bond formation is possible are identified. Covalently-bound structures are generated from these poses, energy-minimized, then clustered to identify unique poses. These poses are directly scored using the GlideScore algorithm, omitting the additional structural refinement stage in CovDock. This simplified workflow had a throughput that was 10–40 times faster than CovDock but the mean RMSD of predicted binding poses for the test set of 21 structures only increased to 1.87 Å from the 1.52 Å RMSD of the original CovDock workflow.

Ai et al. presented a technique steric-clashes alleviating receptor (SCAR) [56], where the covalently-binding residue is mutated into a sterically smaller residue, such as serine or glycine. This allows the ligand to dock in poses similar to the covalently-modified mode without experiencing a steric clash. These poses are ranked according to their non-covalent binding score. This procedure was evaluated by comparing the predicted pose to the experimen-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

tal crystallographic structures of covalently-inhibited AdoMetDC. This technique predicted the binding pose with an RMSD of 3.0 Å, which is comparable to other covalent docking methods. In their complete workflow, the ensemble of docked poses was filtered to select those where the warhead was within 1 Å of the targeted residue. The mean RMSD of the top-ranked structures after imposing this constraint was reduced to 1.9 Å. This procedure was successfully applied to discover novel covalent inhibitors of S-adenosylmethionine decarboxylase.

Scholz et al. presented the implementation of DOCKTITE [44], a covalent docking method into the Molecular Operating Environment (MOE [57]) modeling suite. This method searches a database of potential ligands for molecules possessing one of 21 electrophilic warhead motifs. The structure of the ligand is adjusted to reflect its structure when covalently-bound. Constraints are imposed to force the ligand to occupy a geometry consistent with a covalent linkage and a conformational search is performed to identify the low-energy poses of the ligand in the receptor. Of the 76 covalent-modifier test set developed by Oyang et al., the top-ranked pose predicted by DOCKTITE was 2.4 Å, comparable to other covalent-docking methods. A recent application of DOCKTITE was reported by Schirmeister et al., who found that the relative affinities of covalently-binding dipeptide nitriles inhibitors of rhodesain were correctly predicted by the calculated affinity scores [58].

These implementations have made the docking of covalent-modifiers drugs practical and accessible, although further development of these methods is still needed. Databases containing drugs with warheads must be developed and the codes must be able to identify the potential modes of modification. In some of these codes, the user is required to provide the site of covalent modification, so the screening performed by this code is not yet fully automatic. The energy associated with covalent bond formation (i.e., $\Delta G_{covalent}$) is not immediately available in conventional scoring algorithms, so the absolute strength of binding cannot be realistically estimated by these algorithms either.

4. Calculation of the pKa's of Targeted Residues

Generally, the first step in the mechanism for covalent modification of cysteines, lysines, and serines is their deprotonation to yield their more reactive form (Figure 8). In the case of the modification of a cysteine residue by an electrophile, the thiol group of the amino acid side-chain must be deproto-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

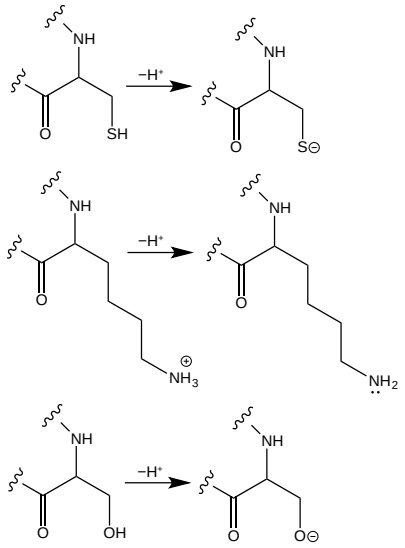

Figure 8: Deprotonation reactions involving cysteine, lysine, and serine. Covalent modification of these residues typically involves reaction in their deprotonated, nucleophilic states.

nated to form the reactive thiolate nucleophile. The stability of the thiolate is thus a significant parameter for the inhibition of a target site by a drug molecule.

The equilibrium between the thiol and thiolate states of a cysteine residue in a protein is defined by its pK_a. Cysteines with low pK_a's are more likely to exist in their reactive thiolate state, so they will be more susceptible to covalent modification by electrophilic inhibitors. The standard pK_a of a cysteine residue is 8.6 [59], but pK_a's of cysteines have been reported to range from 2.9 to 9.8. This broad range results from the intermolecular interactions that the thiol and thiolate states of the cysteine experience inside the protein. Catalytic cysteines in enzymes like cysteine proteases tend to have nearby cationic residues, like histidine or asparagine, which lowers their pK_a's by stabilizing the thiolate state of the cysteine [60]. Conversely, the thiolate state of the cysteine residue will experience repulsive interactions with nearby anionic residues, raising the pK_a. Amino acids buried in hydrophobic pockets of the protein can also have elevated pK_a's because they do not experience stabilizing interactions with water molecules.

Calculating the pK_a of an amino acid side chain in a protein is a long-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

standing challenge in computational biophysics. Traditionally, the pK_a of an amino acid side chain is estimated based on the relative stability of the charged and neutral states. Continuum electrostatic models were among the earliest methods used [61], although the approximations incorporated in their methodology limit their accuracy. Since this time, these models have been continually improved, and some methods that make use of an explicit solvent representation perform well for predicting the pK_a's of aspartic and glutamic acid residues [62, 63].

The methods for the prediction of the pK_a of cysteine residues are less established. In a recent paper, methods for calculating the pK_a's of cysteine residues in proteins were evaluated for a test set of 18 cysteine pK_a's in 12 proteins [37]. Three methods that use an implicit solvent representation were tested, namely: PROPKA, H++, and MCCE. The root-mean-square deviation (RMSD) of the calculated pK_a's with respect to experimental values were large, with some methods having essentially no predictive power. H++ was the most accurate of the three implicit methods, although the RMSD was still 3.4. A method using an all-atom explicit-solvent model with replica exchange molecular dynamics thermodynamic integration (REMD-TI) was more accurate. When used with the CHARMM36 force field, this method was able to predict the pK_a's of cysteine residues in the test set with an RMSD of 2.4. Plots illustrating the correlation between predicted and experimental values for these two methods are presented in Figure 9.

Although REMD-TI gives reasonably accurate cysteine pK_a's, new methods for calculating the pK_a's of cysteines will be needed to allow the reactivity of a cysteine residue to be predicted quantitatively. Improved force fields and the use of algorithms capable of describing variable protonation states of other residues within the protein may improve the accuracy of these methods. Additionally, experimental measurement of more cysteine pK_a's in proteins will allow for a more thorough and comprehensive evaluation of existing pK_a methods. Of particular importance are the pK_a's of noncatalytic residues in protein active sites, which are typically the target of TCIs.

5. Quantum Chemical Methodology

Covalent modification inherently involves the breaking and making of chemical bonds, so researchers have turned to quantum chemistry to model the mechanisms, kinetics, and structures involved in covalent modification. Density functional theory (DFT) is widely used for modeling biological sys-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

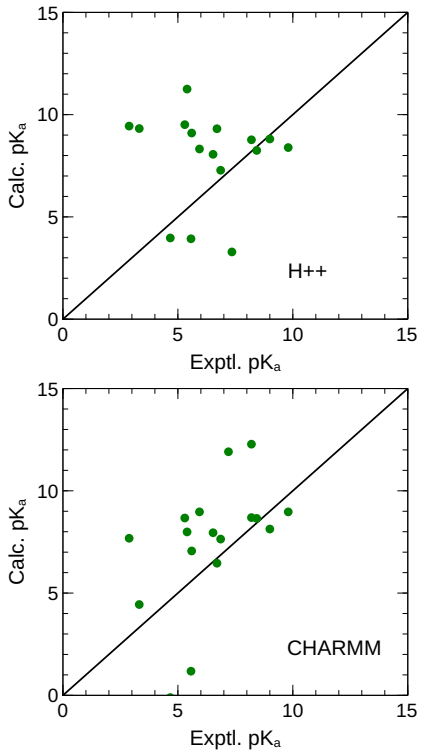

Figure 9: Predictions of cysteine pK$_a$'s using implicit-solvent H++ method and an explicit-solvent replica-exchange molecular dynamics free energy calculation method using the CHARMM force field. Adapted from Ref. [37].

tems because of its ability to describe large chemical systems with quantitatively accurate energies and structures.

Early models of electrophilic thiol additions were unable to identify the enolate/carbanion intermediates that occur in the canonical mechanism for a thio-Michael addition. The failure of conventional DFT methods to describe these reactions stems from an issue in contemporary DFT known as delocalization error [64, 65, 66, 67]. DFT calculates inter-electron repulsion in a way that erroneously includes repulsion between an electron and itself, which must be corrected for in an approximate way through the exchange-correlation functional. The result of this effect is a spurious delocalization of electrons to reduce their self-interaction.

Delocalization error is an issue when DFT is used to model thiol addi-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

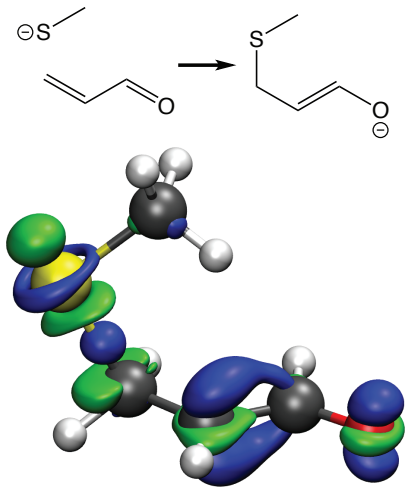

Figure 10: The charge transfer between a thiolate and Michael acceptor calculated using ωB97X-D/aug-cc-pVTZ. Charge is transferred from methylthiolate (top) to the acrolein Michael acceptor (bottom). Areas in blue indicate an increase in charge density while areas in green correspond to a decrease in charge density when the two fragments interact. Charge is lost from the thiolate anion and gained in the space between the S–C_α sigma bond, the π molecular orbital of the C_α–C bond, and the p_z orbital of the O atom, corresponding to an oxygen-centered anion.

tions. The thiolate intermediate features a diffuse, sulfur-centered anion. When some popular DFT functionals are used (e.g., B3LYP or PBE), self-interaction error causes the energy level of the highest occupied molecular orbital (HOMO) to be positive, making the anionic electron formally unbound. When the thiolate is complexed with a Michael acceptor, delocalization error spuriously stabilizes a non-bonded state where electron density is transferred from the HOMO of the thiolate to orbitals of the Michael acceptor. For some electrophiles, this complex is the most stable form and these methods predict that there is no enolate/carbanion intermediate.

One popular method to define the exchange functional in density functional theory calculates the Hartree–Fock exchange energy using the DFT Kohn–Sham orbitals, a technique known as exact exchange. Hybrid DFT functionals have been developed where part of the exchange energy is calculated by exact exchange. These functionals generally outperform "pure" functionals that do include an exact exchange component.

Issues with delocalization error have led to the development of range-separated DFT functionals, where the exchange-correlation functional uses

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

a large component of exact exchange for long-range inter-electron exchange-correlation. Smith et al. showed that range separated DFT functionals such as ωB97X-D predicted a stable thiocarboanion intermediate, while popular methods like B3LYP predicted that this intermediate could not exist as a distinct species (Figure 11). This result was corroborated by highly accurate CCSD(T) calculations [68]. Some hybrid functionals that have a high component of exact exchange globally, such as PBE0 or M06-2X, also predicted a stable carbanion intermediate.

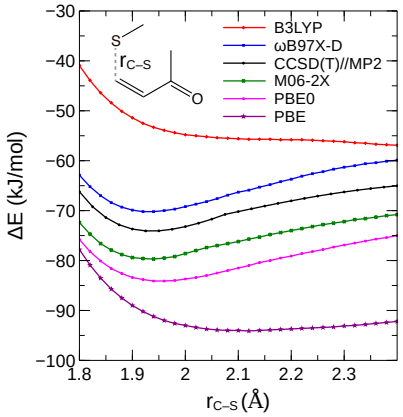

Figure 11: The potential energy surfaces for the addition of methylthiolate to methyl vinyl ketone, calculated using DFT and ab initio methods. The PES for the B3LYP and PBE functionals fail to predict a stable enolate intermediate. High-level ab initio (CCST(T)), range-separated functions (e.g., ωB97X-D) and hybrid functionals (e.g., PBE0) predict a moderately-stable enolate intermediate with a minimum near C_β–S = 1.9 Å.

The reaction energies of thiol additions are also sensitive to the DFT functional used. Krenske et al. studied the addition of methyl thiol to α,β-unsaturated ketones [69]. The calculated reaction energies were sensitive to the quantum chemical method used, but the M06-2X and B2PLYP functionals provided results in close agreement with high-level ab initio results (CBS-QB3). Smith et al. showed that the ωB97X-D and PBE0 functionals also provide accurate thiol-addition reaction energies [68].

6. Warhead Design

An additional factor in the design of covalent modifiers is the selection of an appropriate functional group for reaction with the target residue of

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

the protein. This warhead is typically an electrophilic group. The type of amino acid undergoing modification is the first design criteria. Covalent modifiers of cysteine residues often feature acrylamides or other electron-deficient alkenes, which can undergo Michael additions to the cysteine residues to form thioether adducts.

Quantum chemistry has been used to model the reaction mechanisms of covalent modification, providing information about the reaction kinetics and thermodynamics for various warheads. For Michael additions to cysteines, a model thiol (e.g., methylthiol) has commonly been used to represent the cysteine residue. The transition state and carbanion intermediate stability can theoretically be used to estimate the rates of reaction ($k_{inact.}$). The Gibbs energy for the net reaction determines whether the addition is spontaneous or non-spontaneous and the degree to which it is reversible.

Taunton and coworkers have pursued a line of development of cysteine-targeting covalent inhibitors with acrylonitrile warheads with electron donating aryl or heteroaryl groups at both the α and β positions [70]. The most effective electrophiles formed a carbanion intermediate with a high proton affinity. This warhead was successfully applied to develop a high-affinity 1,2,4-triazole-activated acrylonitrile covalent modifier that was selective for RSK2 kinase.

Smith and Rowley implemented an automated workflow to assess the stability of the carbanion intermediate and thioether product for the addition of a model thiol (methylthiol) to a large set of substituted olefins [71]. All combinations of $-H$, $-CH_3$, $-C(=O)NH(CH_3)$, $-CN$, and $-C(=O)OCH_3$ substituents. The lowest energy conformations of each species were identified by a replica exchange molecular dynamics method [72]. Generally, it was found that substitution of the alkene core has a large effect both on the stability of the intermediate and products, but conventional warheads fell into a narrow range where the addition was weakly spontaneous and went through a moderately stable intermediate (e.g., 120–160 kJ/mol). This is illustrated in Figure 12, with the approximate range of appropriate warheads highlighted in red. Dimethyl fumarate and N-methylacrylamide, which are established covalent warheads, are indicated. The complete set of calculated reaction energies and intermediate stabilities for the full set of warheads are available in the supporting information of Ref. 71.

The calculated reaction energies of warheads used in successful TCIs are weakly exergonic, which makes these reactions reversible. This allows them to dissociate from non-targeted thiols should they react with an off-target thiol

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

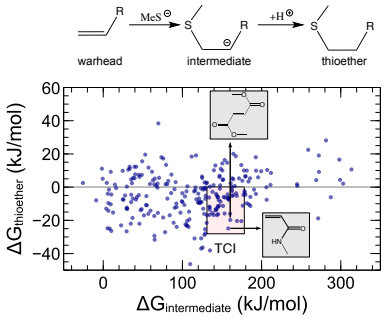

Figure 12: The calculated stability of the carbanion intermediate vs. the stability of the thioether product for the full data set of model thiol additions from Smith and Rowley [71]. The region of potential TCI warheads is highlighted, where the thiol undergoes a weakly exergonic addition ($\Delta G_{thioether} < 0$) through a moderately-stable carbanion intermediate (130 kJ/mol $< \Delta G_{intermediate} <$ 180 kJ/mol).

in the cell prior to reaching their target. This is consistent with an emerging principle of TCI design, which is that the warhead should only be moderately reactive in order to avoid promiscuous modification [73, 74, 75]. Because thiol additions to warheads like acrylamides are reversible, modification of an off-target protein can be reversed. The moderate rate of reaction due to a moderately stable enolate intermediate favors the formation of the covalent bond only after the inhibitor has complexed to its target through selective non-covalent interactions. Of the hundreds of putative warheads evaluated in this study, only a small number have the appropriate thiol-addition kinetics and thermochemistry to serve as a therapeutic TCI.

Krenske et al. studied the addition of methyl thiol to α, β- unsaturated ketones [69]. α-methyl, β-methyl, and α-phenyl ketones were found be more readily reversible than the unsubstituted vinyl ketone. A subsequent study by Krenske et al. used DFT calculations to study the Michael acceptor warheads with aryl groups at the α-position and electron withdrawing groups at the β-position [76]. These calculations were consistent with the experimental observation by Taunton et al., who observed that electrophiles with two electron withdrawing groups at the α-position (e.g., amide and cyano) and an aryl at the β-position yielded a warhead that reacted covalently but reversibly with thiols [75].

7. QM/MM Models of Covalent Modification

Studies of covalent modification using model reactants in the gas phase or using a continuum solvent model do not provide a rigorous description of how the protein environment affects the reaction between the protein and the inhibitor. Paasche et al. found that continuum solvent models provided limited success in describing the cysteine–histidine proton transfer reactions associated with cysteine protease functoin [77]. Describing the full enzyme, inhibitor, and solvent using a quantum mechanical model would be prohibitively computationally demanding, so it is not practical to apply these methods naively to model the covalent modification of a protein.

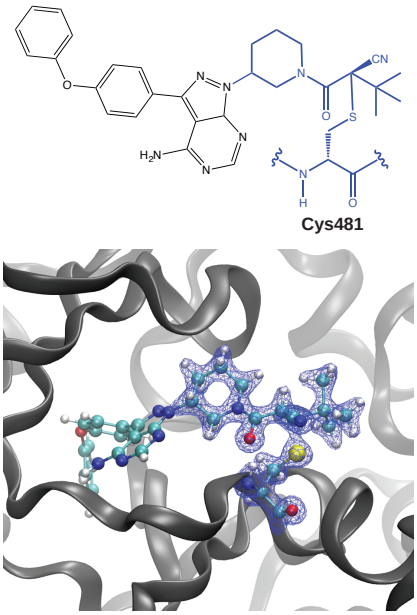

Figure 13: An example QM/MM model of Bruton's tyrosine kinase in complex with a covalent modifier (Ref. 75, PDB ID: 4YHF). The covalently-modified Cys481 residue and the cyanoacrylamide warhead define the QM region. The calculated electron density of the QM region is represented by the blue mesh. The remainder of the protein (gray) and inhibitor comprise the MM region.

Quantum mechanics/molecular mechanics (QM/MM) methods allow for a critical component of a chemical system to be described using a quantum mechanical model, while the rest of the system is represented using a molecular mechanical model (Figure 13). As the size of the QM region is reduced

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

to a relatively small size, the computational expense of these QM/MM calculations is tractable. This is well-suited for modeling chemical reactions involving proteins, like enzymatic reaction mechanisms, where the chemical reaction only directly involves a small number of atoms, but the rest of system provides an essential environment. Analogously, the covalent modification of proteins can also be described using a QM/MM model, where the reactive warhead of the inhibitor and the residue being modified are described using QM, while the balance of the system, such as the solvent and the rest of the protein are described using an MM model. If needed, additional sections of the inhibitor and protein can be included in the QM region. QM/MM methods are now available in codes such as Gaussian [78], CHARMM [79], Amber [80], NAMD [81], and ChemShell [82, 83].

One example of QM/MM modeling of covalent modification probed the mechanisms involved in the covalent modification of Ser530 of cyclooxygenase by aspirin. Tosco and Lazzarato [84] constructed a QM/MM model where aspirin molecule and nearby amino acid residues in the enzyme active site (e.g., Ser530, Tyr348, and Tyr385) were modeled using semiempirical SCC-DFTB QM method and the balance of the system was modeled using the CHARMM force field. The aim was to propose a putative reaction mechanism for the irreversible inactivation of cyclooxygenase by aspirin. The results obtained suggests that acetylation of Ser530 in cyclooxygenase by aspirin occurs under intramolecular general base catalysis conditions, where the vicinal carboxylate group of aspirin abstracts a proton from the hydroxy group of Ser530 in cyclooxygenase, followed by a nucleophilic attack by the O_γ of Ser530 on the acetyl carbonyl carbon of aspirin. The reaction undergoes a tetrahedral intermediate, with Tyr385 stabilizing the anionic intermediate state formed.

Tóth et al. [85] also reported a QM/MM study of the reaction mechanism, transition state, and potential energy surface of the inhibition reaction of cyclooxygenase by aspirin. In their approach, static ONIOM-type QM/MM were used, where the HF, B3LYP, MP2, and B97-D models were used to describe the QM region, which was comprised of the aspirin molecule and surrounding amino acids. Local minimum geometries and transition states were among the few properties determined in their study. The authors concluded that the transesterification reaction of aspirin and cyclooxygenase occurs through a concerted mechanism.

QM/MM methods have also been employed in studying the mechanism of enzyme inhibition in peptide cleaving enzymes, such as cysteine proteases [86, 87]. A wide range of warheads will react readily with the acidic catalytic

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

cysteine, so covalent-modifiers have frequently been used to inhibit these targets. QM/MM studies have examined the inhibition of cysteine proteases by epoxides [88], aziridines [89], peptidyl aldehydes [90], vinyl sulfones [91], and nitroalkene-based inhibitors [92] of cysteine proteases, among others. It should be noted that some of these studies used DFT methods (e.g., B3LYP) that underestimate the stability of the enolate intermediate [71].

Engels and coworkers have published a series of computational studies [87, 89, 93, 94, 95] that explore the factors governing the kinetics, regiospecificity and stereoselectivity of epoxide- and aziridine-based inhibitors of cysteine proteases. QM/MM and molecular dynamics (MD) methods were used to investigate basic mechanistic principles and inhibition processes of these enzymatic reactions. Hydrogen bonds between the inhibitor and other residues in the active site were found to affect the reaction mechanism significantly, demonstrating that the explicit representation of the active site was necessary to describe the reaction mechanism. These studies aided the design and synthesis of new, potent inhibitors of this family of enzymes [89, 96, 97].

Schirmeister et al. [98] developed QM/MM-based workflow for the development of covalent inhibitors. Using an existing protein-inhibitor structure, the substituents on the warhead are systematically varied. A QM/MM model of the protein-inhibitor complex is used to calculate the potential energy surface of the covalent modification, which provided the activation energy and reaction energy. The docking programs FlexX and DOCKTITE were used to identify variants on the drug scaffold that would improve the non-covalent component of the binding energy. The covalent competence of this new warhead were tested by a second round of QM/MM calculations. This workflow was used to develop novel covalent vinyl sulfone-based inhibitors of rhodesian, a parasite protease belonging to the papain family of cysteine proteases. The halogenated warhead designed by this procedure was predicted to be a reversible covalent modifier, which was confirmed experimentally.

Another QM/MM workflow for the design of covalent inhibitors was reported by Fanfrlik et al. [91], which uses a QM-based scoring function based on a hybrid QM/semiempirical QM model to describe the process of covalent binding in protein–ligand complexes. The performance of the algorithm was evaluated on a series of vinyl sulfone-based inhibitors of *S. mansoni* cathepsin B1. The calculated Gibbs energy difference between the non-covalently-bound state and the covalently-bound state was found to correlate to the log of the experimental IC_{50} with a coefficient of determination of 0.69.

QM/MM modeling has the potential play a significant role in understand-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

ing and predicting the mechanisms, kinetics, and thermodynamics of covalent modification. More accurate QM methods, improved algorithms to interface the QM and MM regions, and more extensive configurational sampling will be needed to make these methods quantitatively accurate. In combination with other simulation methods to calculate the non-covalent binding and deprotonation steps, QM/MM methods can be used to calculate the kinetic and thermodynamics terms of covalent modification rigorously. Ultimately, the integration of these methods will make it possible to model the action of covalent modifier drugs in a comprehensive way through the calculation of $\Delta G_{non-covalent}$ and $\Delta G^{\ddagger}$.

8. Conclusions

Computational methods for modeling the covalent modification of proteins have developed rapidly over the last 10 years. Docking programs such as AutoDock, Glide, and MOE now include functionality to find poses of docked covalent modifiers. The design of reactive warheads has been aided by quantum chemical modeling of model reactions. One of the most promising areas of this field is the use of computer modeling to examine the reaction mechanisms of covalent modification; pK_a calculations can be used to determine the reactivity of targeted residues and QM/MM models can be used to elucidate the reaction mechanisms, activation energy, and covalent binding energy. With further developments in computing power and more accurate computational methods, these methods may eventually allow the estimation of rates of inactivation and the covalent component of the binding energy. The maturation of these methods will allow computer modeling to contribute to the development of covalent-modifier drugs to the same degree that they have contributed to the development of non-covalent drugs.

9. Acknowledgements

The authors thank NSERC of Canada for funding through the Discovery Grant program (Application 418505-2012). EAW thanks the School of Graduate Studies at Memorial University for a graduate fellowship. EAW also thanks ACENET for an Advanced Research Computing Fellowship. Computational resources were provided by Compute Canada (RAPI: djk-615-ab) through the Calcul Quebec, Westgrid, and ACENET consortia. We thank Schrödinger Inc. for a temporary license of Maestro/Glide. We thank Professor Erin Johnson for assistance with the charge transfer figure.

10. References

[1] I. D. Kuntz, K. Chen, K. A. Sharp, P. A. Kollman, The maximal affinity of ligands, Proc. Natl. Acad. Sci. USA 96 (1999) 9997–10002.

[2] M. H. Potashman, M. E. Duggan, Covalent Modifiers: An Orthogonal Approach to Drug Design, J. Med. Chem. 52 (2009) 1231–1246.

[3] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, The resurgence of covalent drugs, Nat. Rev. Drug Discov. 10 (2011) 307–317.

[4] A. S. Kalgutkar, D. K. Dalvie, Drug discovery for a new generation of covalent drugs, Expert Opin. Drug Discov. 7 (2012) 561–581.

[5] A. J. Wilson, J. K. Kerns, J. F. Callahan, C. J. Moody, Keap calm, and carry on covalently, J. Med. Chem. 56 (2013) 7463–7476.

[6] R. A. Bauer, Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies, Drug Discov. Today 20 (2015) 1061–1073.

[7] T. A. Baillie, Targeted covalent inhibitors for drug design, Angew. Chem. Int. Ed. 55 (2016) 13408–13421.

[8] J. C. Powers, J. L. Asgian, . D. Ekici, K. E. James, Irreversible inhibitors of serine, cysteine, and threonine proteases, Chemical Reviews 102 (2002) 4639–4750.

[9] A. Miseta, P. Csutora, Relationship between the occurrence of cysteine in proteins and the complexity of organisms, Mol. Biol. Evol. 17 (2000) 1232–1239.

[10] C. Jöst, C. Nitsche, T. Scholz, L. Roux, C. D. Klein, Promiscuity and selectivity in covalent enzyme inhibition: A systematic study of electrophilic fragments, J. Med. Chem. 57 (2014) 7590–7599.

[11] K. M. Backus, B. E. Correia, K. M. Lum, S. Forli, B. D. Horning, G. E. González-Páez, S. Chatterjee, B. R. Lanning, J. R. Teijaro, A. J. Olson, D. W. Wolan, B. F. Cravatt, Proteome-wide covalent ligand discovery in native biological systems, Nature 534 (2016) 570–574. Letter.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[12] C. G. Parker, A. Galmozzi, Y. Wang, B. E. Correia, K. Sasaki, C. M. Joslyn, A. S. Kim, C. L. Cavallaro, R. M. Lawrence, S. R. Johnson, I. Narvaiza, E. Saez, B. F. Cravatt, Ligand and target discovery by fragment-based screening in human cells, Cell 168 (2017) 527–541.e29.

[13] D. S. Johnson, E. Weerapana, B. F. Cravatt, Strategies for discovering and derisking covalent, irreversible enzyme inhibitors, Future Med. Chem. 2 (2010) 949–964.

[14] M. H. Johansson, Reversible michael additions: Covalent inhibitors and prodrugs, Mini-Reviews in Medicinal Chemistry 12 (2012) 1330–1344.

[15] H.-J. Woo, B. Roux, Calculation of absolute protein–ligand binding free energy from computer simulations, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 6825–6830.

[16] Z. D. Nagel, J. P. Klinman, A 21st century revisionist's view at a turning point in enzymology, Nat. Chem. Biol. 5 (2009) 543–550.

[17] M. H. Olsson, J. Mavri, A. Warshel, Transition state theory can be used in studies of enzyme catalysis: lessons from simulations of tunnelling and dynamical effects in lipoxygenase and other systems, Philosophical Transactions of the Royal Society B: Biological Sciences 361 (2006) 1417–1432.

[18] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, Calculation of molecular lipophilicity: State-of-the-art and comparison of logp methods on more than 96,000 compounds, J. Pharm. Sci. 98 (2009) 861–893.

[19] E. Awoonor-Williams, C. N. Rowley, Molecular simulation of nonfacilitated membrane permeation, Biochimica et Biophysica Acta (BBA) - Biomembranes 1858 (2016) 1672–1687.

[20] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, J. C. Gumbart, Simulation-based approaches for determining membrane permeability of small compounds, J. Chem. Inf. Model. 56 (2016) 721–733.

[21] E. Yuriev, M. Agostino, P. A. Ramsland, Challenges and advances in computational docking: 2009 in review, J. Mol. Recogn. 24 (2011) 149–164.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[22] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S. H. Bryant, Structure-based virtual screening for drug discovery: a problem-centric review, The AAPS Journal 14 (2012) 133–141.

[23] G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, Computational methods in drug discovery, Pharmacol. Rev. 66 (2013) 334–395.

[24] J. Wang, Y. Deng, B. Roux, Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials, Biophys J. 91 (2006) 2798–2814.

[25] J. Michel, J. W. Essex, Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations, J. Comput. Aided Mol. Des. 24 (2010) 639–658.

[26] J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, V. S. Pande, Alchemical free energy methods for drug discovery: progress and challenges, Curr. Opin. Struct. Biol. 21 (2011) 150–160.

[27] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, J. Am. Chem. Soc 137 (2015) 2695–2703.

[28] H. M. Kumalo, S. Bhakat, M. E. S. Soliman, Theory and applications of covalent docking in drug discovery: Merits and pitfalls, Molecules 20 (2015) 1984–2000.

[29] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235.

[30] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity, Nat. Biotech. 29 (2011) 1046–1051.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[31] T. Barf, A. Kaptein, Irreversible protein kinase inhibitors: Balancing the benefits and risks, J. Med. Chem. 55 (2012) 6243–6262.

[32] Q. Liu, Y. Sabnis, Z. Zhao, T. Zhang, S. Buhrlage, L. Jones, N. Gray, Developing irreversible inhibitors of the protein kinase cysteinome, Chem. Biol. 20 (2013) 146–159.

[33] M. S. Cohen, C. Zhang, K. M. Shokat, J. Taunton, Structural bioinformatics-based design of selective, irreversible kinase inhibitors, Science 308 (2005) 1318–1321.

[34] J. Zhang, P. L. Yang, N. S. Gray, Targeting cancer with small molecule kinase inhibitors, Nat. Rev. Cancer 9 (2009) 28–39.

[35] E. Leproult, S. Barluenga, D. Moras, J.-M. Wurtz, N. Winssinger, Cysteine mapping in conformationally distinct kinase nucleotide binding sites: Application to the design of selective covalent inhibitors, J. Med. Chem. 54 (2011) 1347–1355.

[36] S. Wu, H. L. (Howard), H. Wang, W. Zhao, Q. Hu, Y. Yang, Cysteinome: The first comprehensive database for proteins with targetable cysteine and their covalent inhibitors, Biochem. Biophys. Res. Commun. 478 (2016) 1268 – 1273.

[37] E. Awoonor-Williams, C. N. Rowley, Evaluation of methods for the calculation of the pKa of cysteine residues in proteins, J. Chem. Theory Comput. 12 (2016) 4662–4673.

[38] X. Fradera, J. Kaur, J. Mestres, Unsupervised guided docking of covalently bound ligands, J. Comput. Aided Mol. Des. 18 (2004) 635–650.

[39] N. Moitessier, J. Pottel, E. Therrien, P. Englebienne, Z. Liu, A. Tomberg, C. R. Corbeil, Medicinal chemistry projects requiring imaginative structure-based drug design methods, Acc. Chem. Res. 49 (2016) 1646–1657.

[40] N. London, R. M. Miller, S. Krishnan, K. Uchida, J. J. Irwin, O. Eidam, L. Gibold, R. Bonnet, P. Cimermančič, B. K. Shoichet, J. Taunton, Covalent docking of large libraries for the discovery of chemical probes, Nat. Chem. Biol. 10 (2014) 1066–1072.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[41] K. Zhu, K. W. Borrelli, J. R. Greenwood, T. Day, R. Abel, R. S. Farid, E. Harder, Docking covalent inhibitors: A parameter free approach to pose prediction and scoring, J. Chem. Inf. Model. 54 (2014) 1932–1940.

[42] D. Toledo Warshaviak, G. Golan, K. W. Borrelli, K. Zhu, O. Kalid, Structure-based virtual screening approach for discovery of covalently bound ligands, J. Chem. Inf. Model. 54 (2014) 1941–1950.

[43] G. Bianco, S. Forli, D. S. Goodsell, A. J. Olson, Covalent docking using autodock: Two-point attractor and flexible side chain methods, Protein Sci. 25 (2016) 295–301.

[44] C. Scholz, S. Knorr, K. Hamacher, B. Schmidt, Docktite – a highly versatile step-by-step workflow for covalent docking and virtual screening in the molecular operating environment, J. Chem. Inf. Model. 55 (2015) 398–406.

[45] C. R. Corbeil, P. Englebienne, N. Moitessier, Docking ligands into flexible and solvated macromolecules. 1. development and validation of FITTED 1.0, J. Chem. Inf. Model. 47 (2007) 435–449.

[46] C. R. Corbeil, N. Moitessier, Docking ligands into flexible and solvated macromolecules. 3. impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs, J. Chem. Inf. Model. 49 (2009) 997–1009.

[47] J. Lawandi, S. Toumieux, V. Seyer, P. Campbell, S. Thielges, L. Juillerat-Jeanneret, N. Moitessier, Constrained peptidomimetics reveal detailed geometric requirements of covalent prolyl oligopeptidase inhibitors, J. Med. Chem. 52 (2009) 6672–6684.

[48] S. De Cesco, S. Deslandes, E. Therrien, D. Levan, M. Cueto, R. Schmidt, L.-D. Cantin, A. Mittermaier, L. Juillerat-Jeanneret, N. Moitessier, Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells, J. Med. Chem. 55 (2012) 6306–6315.

[49] G. Mariaule, S. De Cesco, F. Airaghi, J. Kurian, P. Schiavini, S. Rocheleau, I. Huskić, K. Auclair, A. Mittermaier, N. Moitessier, 3-oxohexahydro-1h-isoindole-4-carboxylic acid as a drug chiral bicyclic scaf-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

776 fold: Structure-based design and preparation of conformationally con-
777 strained covalent and noncovalent prolyl oligopeptidase inhibitors, J.
778 Med. Chem. 59 (2016) 4221–4234.

779 [50] A. Del Rio, M. Sgobba, M. D. Parenti, G. Degliesposti, R. Forestiero,
780 C. Percivalle, P. F. Conte, M. Freccero, G. Rastelli, A computational
781 workflow for the design of irreversible inhibitors of protein kinases, J
782 Comput. Aided Mol. Des. 24 (2010) 183–194.

783 [51] X. Ouyang, S. Zhou, C. T. T. Su, Z. Ge, R. Li, C. K. Kwoh, Co-
784 valentdock: Automated covalent docking with parameterized covalent
785 linkage energy estimation and molecular geometry constraints, J. Com-
786 put. Chem. 34 (2013) 326–336.

787 [52] X. Ouyang, S. Zhou, Z. Ge, R. Li, C. K. Kwoh, Covalentdock cloud:
788 A web server for automated covalent docking, Nucleic Acids Res. 41
789 (2013) W329.

790 [53] G. Bianco, S. Forli, D. S. Goodsell, A. J. Olson, Covalent docking using
791 autodock: Two-point attractor and flexible side chain methods, Protein
792 Sci. 25 (2016) 295–301.

793 [54] M. P. Jacobson, R. A. Friesner, Z. Xiang, B. Honig, On the role of the
794 crystal environment in determining protein side-chain conformations, J.
795 Mol. Biol. 320 (2002) 597–608.

796 [55] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E.
797 Shaw, R. A. Friesner, A hierarchical approach to all-atom protein loop
798 prediction, Proteins: Structure, Function, and Bioinformatics 55 (2004)
799 351–367.

800 [56] Y. Ai, L. Yu, X. Tan, X. Chai, S. Liu, Discovery of covalent ligands via
801 noncovalent docking by dissecting covalent docking based on a "steric-
802 clashes alleviating receptor (SCAR)" strategy, J. Chem. Inf. Model. 56
803 (2016) 1563–1575.

804 [57] Molecular operating environment (MOE) 2013.08 chemical computing
805 group inc. 1010 sherbooke st. west suite #910 montreal q, canada h3a
806 2r7, 2017.

[58] T. Schirmeister, J. Schmitz, S. Jung, T. Schmenger, R. L. Krauth-Siegel, M. Gütschow, Evaluation of dipeptide nitriles as inhibitors of rhodesain, a major cysteine protease of trypanosoma brucei, Bioorg. Med. Chem. Lett. 27 (2017) 45–50.

[59] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, C. N. Pace, pK values of the ionizable groups of proteins, Protein Sci. 15 (2006) 1214–1218.

[60] 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, K. Brocklehurst, A classical enzyme active center motif lacks catalytic competence until modulated electrostatically, Biochemistry 36 (1997) 9968–9982.

[61] D. Bashford, M. Karplus, pKa's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model, Biochemistry 29 (1990) 10219–10225.

[62] T. Simonson, J. Carlsson, D. A. Case, Proton binding to proteins: pka calculations with explicit and implicit solvent models, J. Am. Chem. Soc. 126 (2004) 4167–4180.

[63] E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. M. Olsson, J. K. Shen, J. Warwicker, S. Williams, J. M. Word, Progress in the prediction of pKa values in proteins, Proteins: Struct., Funct., Bioinf. 79 (2011) 3260–3275.

[64] E. R. Johnson, P. Mori-Snchez, A. J. Cohen, W. Yang, Delocalization errors in density functionals and implications for main-group thermochemistry, J. Chem. Phys. 129 (2008) 204112.

[65] A. J. Cohen, P. Mori-Sánchez, W. Yang, Insights into current limitations of density functional theory, Science 321 (2008) 792–794.

[66] J. Autschbach, M. Srebro, Delocalization error and functional tuning in kohnsham calculations of molecular properties, Acc. Chem. Res. 47 (2014) 2592–2602.

[67] A. Wasserman, J. Nafziger, K. Jiang, M.-C. Kim, E. Sim, K. Burke, The importance of being self-consistent, Annu. Rev. Phys. Chem. (2017).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[68] J. M. Smith, Y. Jami Alahmadi, C. N. Rowley, Range-separated dft functionals are necessary to model thio-Michael additions, J. Chem. Theory Comput. 9 (2013) 4860–4865.

[69] E. H. Krenske, R. C. Petter, Z. Zhu, K. N. Houk, Transition states and energetics of nucleophilic additions of thiols to substituted ,-unsaturated ketones: Substituent effects involve enone stabilization, product branching, and solvation, J. Org. Chem. 76 (2011) 5074–5081.

[70] S. Krishnan, R. M. Miller, B. Tian, R. D. Mullins, M. P. Jacobson, J. Taunton, Design of reversible, cysteine-targeted michael acceptors guided by kinetic and computational analysis, J. Am. Chem. Soc. 136 (2014) 12624–12630.

[71] J. M. Smith, C. N. Rowley, Automated computational screening of the thiol reactivity of substituted alkenes, J. Comput.-Aided Mol. Des. 29 (2015) 725–735.

[72] G. K., R. C.N., An explicit-solvent conformation search method using open software., PeerJ 4 (2016) e2088.

[73] I. M. Serafimova, M. A. Pufall, S. Krishnan, K. Duda, M. S. Cohen, R. L. Maglathlin, J. M. McFarland, R. M. Miller, M. Frödin, J. Taunton, Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles, Nat. Chem. Biol. 8 (2012) 471–476.

[74] P. A. Schwartz, P. Kuzmic, J. Solowiej, S. Bergqvist, B. Bolanos, C. Almaden, A. Nagata, K. Ryan, J. Feng, D. Dalvie, J. C. Kath, M. Xu, R. Wani, B. W. Murray, Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance, Proc. Natl. Acad. Sci. USA 111 (2014) 173–178.

[75] 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, J. Taunton, Prolonged and tunable residence time using reversible covalent kinase inhibitors, Nat. Chem. Biol. 11 (2015) 525–531. Article.

[76] E. H. Krenske, R. C. Petter, K. N. Houk, Kinetics and thermodynamics of reversible thiol additions to mono- and diactivated Michael acceptors:

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

Implications for the design of drugs that bind covalently to cysteines, J. Org. Chem. 81 (2016) 11726–11733.

[77] A. Paasche, T. Schirmeister, B. Engels, Benchmark study for the cysteine–histidine proton transfer reaction in a protein environment: Gas phase, COSMO, QM/MM approaches, J. Chem. Theory Comput. 9 (2013) 1765–1777.

[78] 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, X. Li, M. Caricato, 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. A. Montgomery, Jr., 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, D. J. Fox, Gaussian16 Revision A.03, 2016. Gaussian Inc. Wallingford CT.

[79] S. Riahi, C. N. Rowley, The CHARMM–TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties, J. Comput. Chem. 35 (2014) 2076–2086.

[80] R. C. Walker, M. F. Crowley, D. A. Case, The implementation of a fast and accurate QM/MM potential method in Amber, J. Comput. Chem. 29 (2008) 1019–1031.

[81] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with namd, J. Comput. Chem. 26 (2005) 1781–1802.

[82] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

et al., QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis, Comp. Theor. Chem. 632 (2003) 1–28.

[83] S. Metz, J. Kstner, A. A. Sokol, T. W. Keal, P. Sherwood, Chemshell – a modular software package for QM/MM simulations, Wiley Interdiscip. Rev. Comput. Mol. Sci. 4 (2014) 101–110.

[84] P. Tosco, L. Lazzarato, Mechanistic insights into cyclooxygenase irreversible inactivation by aspirin, ChemMedChem 4 (2009) 939–945.

[85] L. Toth, L. Muszbek, I. Komaromi, Mechanism of the irreversible inhibition of human cyclooxygenase-1 by aspirin as predicted by QM/MM calculations, J. Mol. Graph. 40 (2013) 99–109.

[86] S. Ma, L. S. Devi-Kesavan, 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. 129 (2007) 13633–13645.

[87] M. Mladenovic, R. F. Fink, W. Thiel, T. Schirmeister, B. Engels, On the origin of the stabilization of the zwitterionic resting state of cysteine proteases: a theoretical study, J. Am. Chem. Soc. 130 (2008) 8696–8705.

[88] K. Arafet, S. Ferrer, S. Martí, V. Moliner, Quantum mechanics/molecular mechanics studies of the mechanism of falcipain-2 inhibition by the epoxysuccinate E64, Biochemistry 53 (2014) 3336–3346.

[89] V. Buback, M. Mladenovic, B. Engels, T. Schirmeister, Rational design of improved aziridine-based inhibitors of cysteine proteases, J. Phys. Chem. B 113 (2009) 5282–5289.

[90] M. Shokhen, N. Khazanov, A. Albeck, The mechanism of papain inhibition by peptidyl aldehydes, Proteins: Struct., Funct., Bioinf. 79 (2011) 975–985.

[91] J. Fanfrlík, P. S. Brahmkshatriya, J. Rezac, A. Jílkova, M. Horn, M. Mares, P. Hobza, M. Lepsík, Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors, J. Phys. Chem. B 117 (2013) 14973–14982.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017

[92] A. Latorre, T. Schirmeister, J. Kesselring, S. Jung, P. Johé, U. A. Hellmich, A. Heilos, B. Engels, R. L. Krauth-Siegel, N. Dirdjaja, et al., Dipeptidyl nitroalkenes as potent reversible inhibitors of cysteine proteases rhodesain and cruzain, ACS Med. Chem. Lett. 7 (2016) 1073–1076.

[93] M. Mladenovic, T. Schirmeister, S. Thiel, W. Thiel, B. Engels, The importance of the active site histidine for the activity of epoxide- or aziridine-based inhibitors of cysteine proteases, ChemMedChem 2 (2007) 120–128.

[94] M. Mladenovic, K. Junold, R. F. Fink, W. Thiel, T. Schirmeister, B. Engels, Atomistic insights into the inhibition of cysteine proteases: first QM/MM calculations clarifying the regiospecificity and the inhibition potency of epoxide-and aziridine-based inhibitors, J. Phys. Chem. B 112 (2008) 5458–5469.

[95] M. Mladenovic, K. Ansorg, R. F. Fink, W. Thiel, T. Schirmeister, B. Engels, Atomistic insights into the inhibition of cysteine proteases: first QM/MM calculations clarifying the stereoselectivity of epoxide-based inhibitors, J. Phys. Chem. B 112 (2008) 11798–11808.

[96] R. Vicik, H. Helten, T. Schirmeister, B. Engels, Rational design of aziridine-containing cysteine protease inhibitors with improved potency: Studies on inhibition mechanism, ChemMedChem 1 (2006) 1021–1028.

[97] R. Vicik, M. Busemann, C. Gelhaus, N. Stiefl, J. Scheiber, W. Schmitz, F. Schulz, M. Mladenovic, B. Engels, M. Leippe, K. Baumann, T. Schirmeister, Aziridide-based inhibitors of cathepsinl: Synthesis, inhibition activity, and docking studies, ChemMedChem 1 (2006) 1126–1141.

[98] T. Schirmeister, J. Kesselring, S. Jung, T. H. Schneider, A. Weickert, J. Becker, W. Lee, D. Bamberger, P. R. Wich, U. Distler, et al., Quantum chemical-based protocol for the rational design of covalent inhibitors, J. Am. Chem. Soc. 138 (2016) 8332–8335.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2857v1 | CC BY 4.0 Open Access | rec: 9 Mar 2017, publ: 9 Mar 2017