Base-Assisted Mechanism of Acrylamide Reactions With Cysteines: QM/MM Simulations of Afatinib Reaction with EGFR

Shuhua Ma,[*,†] Heeral Patel,[†] Craig A. Peeples,[‡] and Jana Shen[*,‡]

† *Department of Chemistry, Jess and Mildred Fisher College of Science and Mathematics, Towson University, Towson, MD 21252*
‡ *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201*

E-mail: sma@towson.edu; jana.shen@rx.umaryland.edu

Abstract

Targeted covalent inhibitor (TCI) design has become increasingly popular. Acrylamides are most commonly used warheads for targeting reactive cysteines; however, the reaction mechanisms of acrylamides in proteins remain controversial, particularly for those involving unreactive cysteines. Using combined semiempirical quantum mechanical (QM)/molecular mechanical (MM) simulations, we investigated the reaction between afatinib, the first TCI drug for cancer treatment, and the front-pocket cysteine Cys797 in EGFR, which has been suggested as unreactive in a recent study. Two base-assisted mechanisms were examined, in which either 1) the β-dimethylaminomethyl substitution or 2) the nearby Asp800 serves as a general base/acid for the Michael addition between afatinib and Cys797 of EGFR. The calculated reaction free energy profiles suggested that while both mechanisms are stepwise, Mechanism 1 is more likely due to a lower free energy barrier for the rate-limiting nucleophilic attack, in close agreement with experimental measurement. Importantly, the simulations demonstrated that Asp800 plays an assisting role by stabilizing the thiolate-afatinib ion-pair reactant state as well as lowering the barrier of the nucleophilic attack. Our work elucidates an important structure-activity relationship of acrylamides that is generally applicable for cysteine-directed TCI designs.

INTRODUCTION

Kinases are enzymes that catalyze the phosphoryl transfer reactions; dysregulation of kinases has been linked to a variety of human diseases especially cancer.[1] In 2013 the first targeted covalent inhibitor (TCI), afatinib, which inhibits a receptor kinase called epidermal growth factor receptor (EGFR, Figure 1a and 1b), obtained the FDA approval for treatment of non-small-cell lung (NSCL) cancer.[2] Over the last decade, FDA approvals of TCIs of various kinases have been steadily increasing.[3] A TCI is comprised of a reversible binding motif and an electrophilic functional group (also called the warhead) that forms a chemical bond with a nucleophilic amino acid (typically cysteine) in the target protein. Understanding of the chemical reaction mechanism of TCIs in the protein environment is important for the rational design of novel inhibitors and improvement of existing ones.

So far, all FDA-approved TCIs contain an acrylamide warhead, which is a type of ac-

tivated olefin capable of forming a covalent bond with cysteine's thiol group through Michael addition. The rate-limiting step of a thiol-Michael addition in solution is commonly believed to be the nucleophilic attack of a deprotonated thiolate on the electron deficient β-carbon of the olefin, resulting in a carbanion intermediate which reprotonates to yield the adduct (Figure 1c).[4–7] In the absence of a deprotonated thiolate, a Michael reaction can be initiated, in which a nearby base abstracts the proton from the thiol group, and the same proton is later transferred to protonate the carbanion intermediate (Figure 1c).[4–6,8] Our recent studies of various kinases[9–11] based on the continuous constant pH molecular dynamics (CpHMD) titration simulations found that many cysteines targeted by the TCIs approved by the FDA or in clinical stages are indeed reactive, as defined by having a deprotonated population greater than 10% at physiological pH 7.4.

Of particular interest is the so-called front-pocket cysteine located at the N-terminal end of the αD helix in 11 kinases including EGFR. The front pocket cysteines of all these kinases have been targeted by the TCIs, making it the most popular site of covalent modification for kinases.[10] The CpHMD simulations suggested that the front-pocket cysteine is reactive in the apo state of 7 kinases, e.g., Cys481 in BTK; however, it is unreactive in EGFR (Cys797) and the related ERBB2 and ERBB4 as well as BLK, due to a negatively charged Asp or Glu sidechain at the Cys+3 position (Asp800 in EGFR).[10] Considering the proximity of the carboxylate to the cysteine thiol group in the cocrystal structures, the Cys+3 Asp or Glu may serve as a general base to abstract the proton from the front-pocket cysteine and thereby initiating the Michael reaction.[10]

The reaction of two TCIs with the front-pocket Cys797 in EGFR has been studied by the combined quantum mechanics/molecular mechanics (QM/MM) simulations which started from the protonated Cys797. Capoferri et al.[13] investigated the reaction mechanism of 13-JAB (Figure 1b) with EGFR using the semiempirical DFTB/MM potential . Using the same method, Callegari et al.[14] examined the effect of EGFR mutations on the reaction barrier with osimertinib (Figure 1b). Surprisingly, both studies[13,14] suggested that the nucleophilic attack of the deprotonated Cys797 is concerted with the proton transfer from the protonated Asp800 without a stable intermediate. This direct (i.e., non-Michael) mechanism is in contradiction with the solution chemistry experiments[4–6,8] and QM calculations of the thiol additions of various activated olefins.[7,15–19] More recently, two QM/MM studies[20,21] examined the reaction mechanism of the acrylamide TCIs with the front-pocket cysteine Cys481 in BTK, which was found to be deprotonated in the CpHMD simulations.[10] The QM/MM simulations employing the density functional theory (DFT) QM description of a truncated α-cyano acrylamide inhibitor and starting from a deprotonated Cys797 confirmed the Michael reaction mechanism involving an intermediate.[21] In contrast, the QM/MM simulations using the DFTB QM description of BTK and protonated Cys481 suggested an alternative mechanism which involves first a proton transfer from Cys481 to the carbonyl group.[20] The contradictory results from the above studies suggest that the reaction mechanism of acrylamides with cysteines in the protein environment remains poorly understood.

In the present work, we use afatinib's reaction with the front-pocket Cys797 in EGFR as a model system to understand the reaction mechanism of acrylamides with unreactive cysteines in proteins. Afatinib is an ATP-competitive anilinoquinazoline derivative; its acrylamide warhead has a dimethylaminomethyl (DMAM) substitution on the β-carbon (Figure 1b). Experimental studies found that the β-DMAM substitution enhances reactivity (i.e., rate constant) of acrylamide compounds and their inhibitory activities against EGFR.[22,23] A more recent experiment[24] demonstrated that the glutathione half-life of the β-methylamino substituted acrylamides decreases with the increasing pK_{a} value (from 6 to 9) of the amino sub-

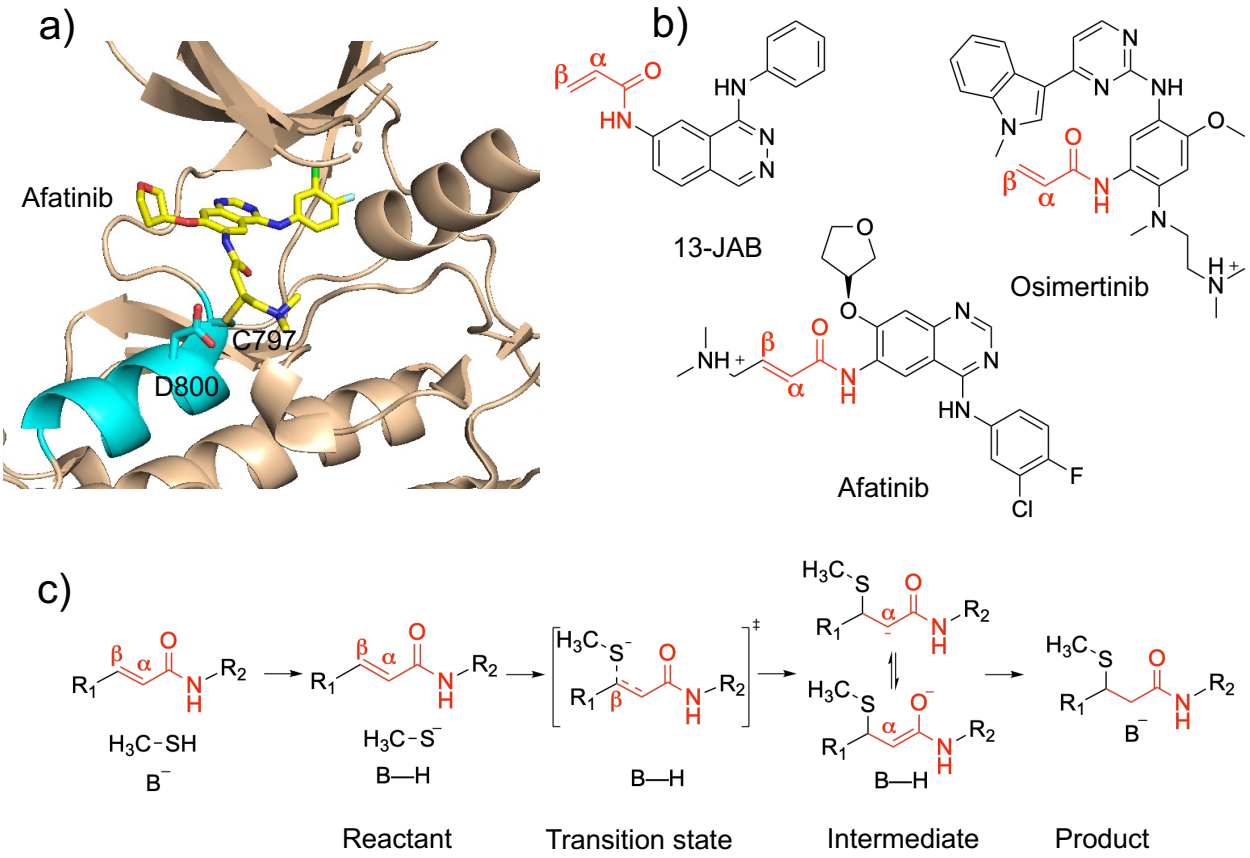


Figure 1: **The TCIs of EGFR studied by QM/MM simulations and the thiol-Michael addition reaction mechanism in solution. a)** The X-ray cocrystal structure of EGFR in covalent complex with afatinib (PDB ID: 4G5J).[12] The αD helix is highlighted in cyan. The covalently modified Cys797 and the nearby Asp800 which can be a potential general base are labeled. **b)** Chemical structures of afatinib and two previously studied[13,14] TCIs of EGFR. **c)** The base-assisted mechanism of Michael addition between acrylamide and thiol (shown here as a methanethiol) in solution. B-H represents a general base. The tautomerization of carbanion/enolate intermediate is shown.

stituent, implying that a positively charged β substitution increases the Michael reaction rate. To explain this, our recent QM calculations of model reactions in solution (implicit solvent model)[19] found that a charged β-DMAM group lowers the reaction barrier of nucleophilic attack. However, in the protein environment, the reaction mechanism of β-DMAM substituted inhibitors may be different from that in solution, especially in the presence of a protonated cysteine. Indeed, while studying the structure-activity relationship of acetylenic thienopyrimidine compounds, Wood et al.[25] hypothesized that the amino substituent at the Cβ position is the general base that deprotonates Cys797 before nucleophilic attack.

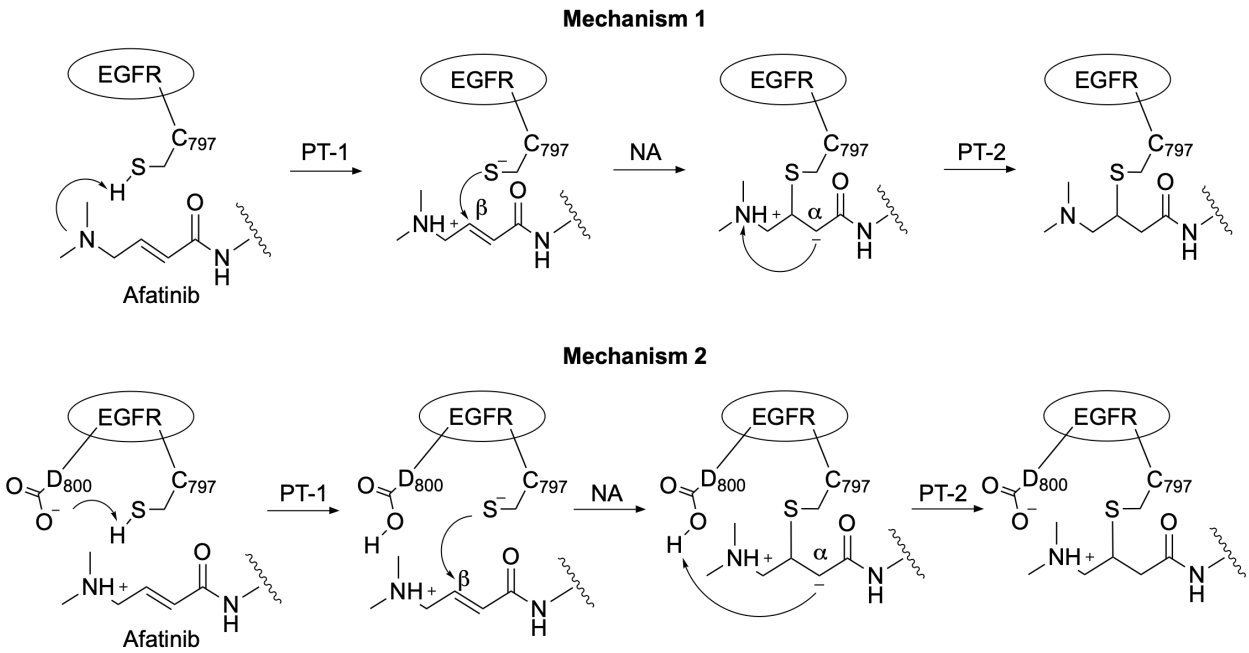
Inspired by the aforementioned work and hypotheses, here we conducted the QM/MM free energy simulations to investigate the reaction mechanism of afatinib with Cys797 in EGFR. We focus on the two most important questions: 1) what's the identity of the general base? and 2) Is there an enolate or carbanion intermediate? We investigated two possible mechanisms, which mainly differ in the identity of the general base (Scheme 1). The free energy simulations demonstrated stepwise proton transfer and nucleophilic attack. The simulations also found that Mechanism 1, in which the β-DMAM group serves as a general base, is more likely due to a lower free energy barrier in the rate-limiting step. We discussed the reaction path and the nature of the intermediate states as well as their stabilization.

RESULTS and DISCUSSION

Summary of the PDDG-PM3/MM simulations. In this study, we investigated two possible reaction mechanisms of afatinib with Cys797 in EGFR (Scheme 1). Mechanism 1 begins with the first proton-transfer step (PT-1), in which Cys797 thiol transfers a proton to the β-DMAM nitrogen of the acrylamide, initiating the nucleophilic attack (NA) of the thiolate on the β-carbon followed by the second proton-transfer step (PT-2), whereby the proton is transferred from the β-DMAM nitrogen to the α-carbon, resulting in the thioether adduct product (Scheme 1, top). In contrast, Mechanism 2 begins with a proton transfer from the cysteine thiol to Asp800 in EGFR (PT-1), following the NA step, the proton on Asp800 is transferred to the α-carbon (PT-2, Scheme 1, bottom). PT-1 and NA, and similarly also NA and PT-2, may be stepwise or concerted, which can be distinguished through QM/MM simulations.

The simulations were initiated from the cocrystal structure of EGFR in complex with the covalently bound afatinib (PDB ID: 4G5J).[12] For Mechanism 1, the QM region includes the sidechain of Cys797 and the entire inhibitor (afatinib), while the Cα atom of Cys797 is the QM/MM boundary atom (Figure 2a and b). For comparison, the simulations of the reaction steps in solution were also performed, the QM region in solution includes methanethiol and afatinib. For Mechanism 2, the QM region includes the the entire inhibitor (afatinib) and the sidechains of Cys797 and Asp800, and the Cα atoms of Cys797 and Asp800 are the QM/MM boundary atoms. Due to the large number of atoms in the QM region (65 atoms for Mechanism 1 and 73 atoms for Mechanism 2), the semiempirical Pairwise Distance Directed Gaussian modified PM3 method (PDDG-PM3)[26,27] was used (details see Methods).

PT-1/NA steps of Mechanism 1 are stepwise in EGFR but concerted in solution. The later has a much higher barrier. To investigate the multi-step reaction Mechanism 1, we first performed two-dimensional QM/MM free energy simulations along the reaction coordinates that represent the proton transfer from the cysteine thiol to the nitrogen of the β-DMAM group on afatinib (PT-1) and the nucleophilic attack (NA) from the cysteine sulfur to the Cβ of afatinib (Figure 2c and d). For PT-1, the difference between the thiol sulfur-to-hydrogen and β-DMAM hydrogen-to-nitrogen distances was used as the reaction



Scheme 1: **Two possible reaction mechanisms of afatinib covalent addition with Cys797 of EGFR investigated in this work.** The mechanism 1 and 2 differ in that either the β-DMAM group of the acrylamide warhead or the nearby Asp800 acts as a general base to abstract the proton from the otherwise unreactive Cys797 (PT-1) before nucleophilic attack (NA) of thiolate on Cβ followed by the reprotonation of Cα atom (PT-2). For clarity, the enolate-carbanion tautomerization is not shown.

coordinate. For NA, the difference between the Cα-to-Cβ and sulfur-to-Cβ distances was used as the reaction coordinate. The resulting free energy surface (Figure 2c) shows that the ion-pair state (R1) with a deprotonated thiolate and charged β-DMAM group is more stable than the neutral state (R0), which means that Cys797 exists as thiolate and β-DMAM group is charged in EGFR upon afatinib binding. Thus, the covalent bond formation process starts from the R1 state, from which the nucleophilic attack occurs and has a free energy barrier of about 20 kcal/mol. The nucleophilic attack results in an intermediate state (INT1), which then underwent a conformational change to form a slightly more stable INT2, the two intermediate states are separated by a small free energy barrier, and INT2 state serves as the reactant state (R2) for the PT-2 step.

As a comparison and reference, we also conducted the QM/MM free energy simulations of the same PT-1 and NA reactions of Mechanism 1 in solution. From the free energy surface, it is clear that methanethiol and afatinib exist in the neutral state, as the ion-pair state (R1) has a much higher free energy (close to 16 kcal/mol than the neutral state (R0, Figure 2d). This result is consistent with the high pK_{a} of 10.33 for methanethiol in water.[28] The free energy surface shows that the nucleophilic attack and proton transfer from methanethiol and β-DMAM is concerted with a much higher free energy barrier than that in EGFR (Figure 2d). The resultant carbanion or enolate intermediate (INT1) has a much higher free energy relative to the reactant state R0.

PT-2 step of Mechanism 1 in EGFR has a higher barrier but is less exergonic compared to that in solution. We next performed the one-dimensional free energy simulations to investigate proton transfer (PT-2) from the β-DMAM nitrogen to the Cα atom of afatinib in both EGFR and solution (Figure 2e and f). For PT-2, the difference between the β-DMAM nitrogen-to-hydrogen and hydrogen-to-Cα distances was used as a reaction coordinate. The free energy of activation in EGFR

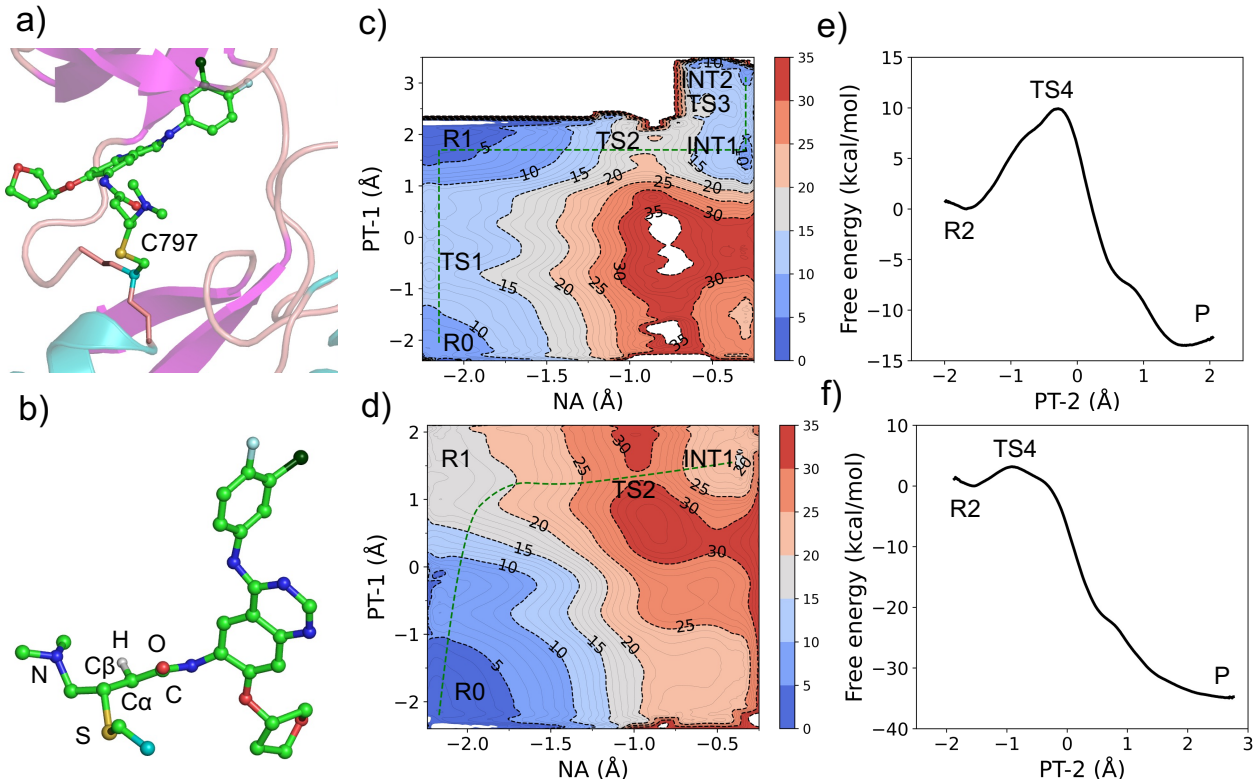


Figure 2: **QM/MM free energy simulations of the covalent bond formation between afatinib and Cys797 in EGFR and the model reaction in solution. a)** Partition of the QM (stick model) and MM (cartoon model) regions in the combined QM/MM simulations. Water (hidden) is in the MM region. **b)** A zoomed-in view of the QM region. Atoms involved in the chemical steps are labeled. Except for the reactive one, hydrogens are hidden for clarity. **c, d)** Calculated free energy surface along the reaction coordinates of the NA ($R_{\mathrm{C}\alpha-\mathrm{C}\beta} - R_{\mathrm{S}-\mathrm{C}\beta}$) and PT-1 ($R_{\mathrm{S}-\mathrm{H}} - R_{\mathrm{H}-\mathrm{N}}$) steps in EGFR (c) and model reaction in solution (d) according to Mechanism 1. R is the distance of the breaking or forming bond. **e, f)** Calculated free energy change along the reaction coordinate of PT-2 ($R_{\mathrm{N}-\mathrm{H}} - R_{\mathrm{H}-\mathrm{C}\alpha}$) in EGFR (e) and model reaction in solution (f) according to Mechanism 1. The state labels are explained in the main text and correspond to those in Scheme 2.

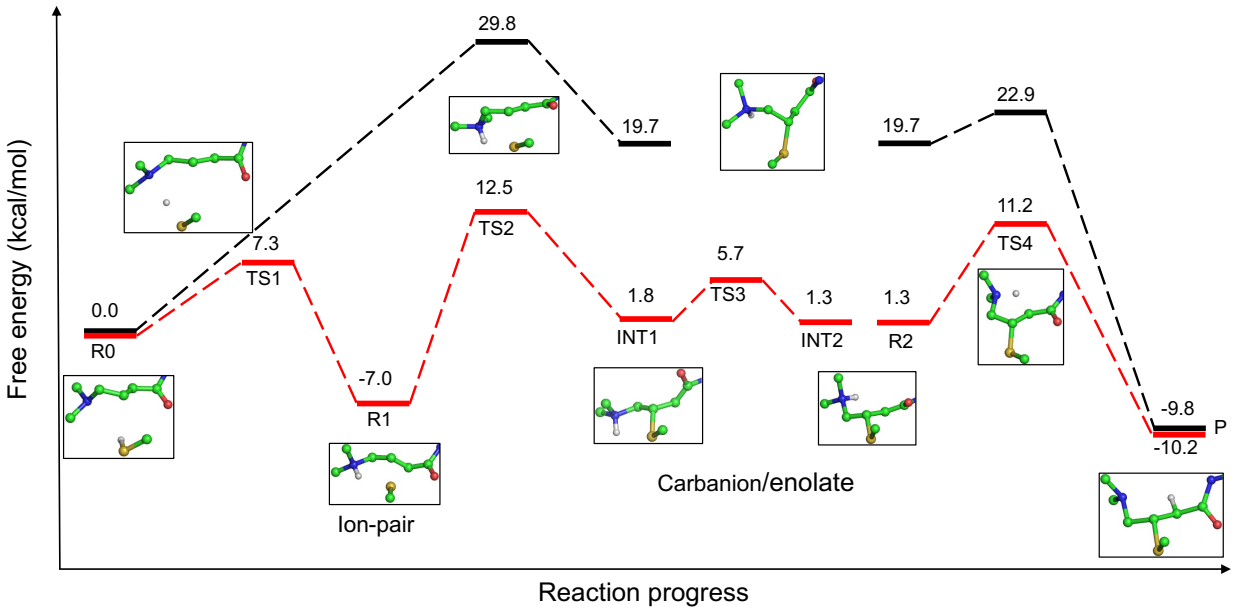
is significantly higher than in the model reaction solution. However, the reaction is significantly less exergonic in EGFR as compared to solution.

Analysis of the reaction profiles of Mechanism 1 and comparison to the less likely Mechanism 2. The free energy profile of Mechanism 1 in EGFR (red) and solution (black) is summarized in Scheme 2. Note, a DFT-based correction is applied (Supplemental methods). In EGFR, the covalent bond formation between afatinib and Cys797 involves three steps, PT-1, NA, and PT-2. The activation barriers for PT-1 (from R0 and R1), NA (from R1 to INT1), and PT-2 (from R2 to P) steps are 7.3 kcal/mol, 19.5 kcal/mol, and 9.9 kcal/mol, respectively. From the INT1 intermediate a conformational rearrangement of the β-DMAM group occurs with a free energy barrier of 3.9 kcal/mol, leading to the slightly more stable INT2 intermediate, which is 8.3 kcal/mol higher in free energy than the reactant R1 state. By contrast, the afatinib reaction with methylthiol in solution proceeds from the R0 state through concerted PT-1 and NA steps, however the NA step is rate-limiting with a free energy barrier of 29.8 kcal/mol, which is about 10 kcal/mol higher than the NA step in EGFR. However, the PT-2 step in water has a much lower free energy barrier of 3.2 kcal/mol, as compared to 9.9 kcal/mol in EGFR. These data show that in both EGFR and solution, the rate-limiting step is the NA step. Note, the former (19.5 kcal/mol in EGFR) is in good agreement with the free energy of activation of 21.6 kcal/mol converted from the measured rate constant of $9 \times 10^{-4} s^{-1}$ for the covalent inhibition of EGFR by afatinib.[29] The reaction free energy in EGFR (from R1 to P) is -3.2 kcal/mol, which is much smaller in magnitude than the corresponding reaction free energy (from R0 to P) of -9.8 kcal/mol in solution. The reaction free energy of the PT-2 step (from R2 to P) in EGFR is -11.5 kcal/mol, which is also much smaller in magnitude than the corresponding reaction free energy of -29.5 kcal/mol in solution.

Now we turn to the afatinib reaction in EGFR according to Mechanism 2, whereby the nearby Asp800 acts as a general base to deprotonate Cys797 (PT-1), followed by NA and reprotonation of Cα on afatinib (PT-2). The individual PT-1, NA, and PT-2 steps have free energy barriers of 11.7, 23.3, and 7.1 kcal/mol, respectively (Figure S8). The carbanion/enolate intermediate state is 15.4 kcal/mol higher in free energy than the reactant state (R2, thiolate^{-}/Asp800^{H}). Overall, the rate-liming NA step has a free energy barrier of 31.3 kcal/mol, which is much higher than that in Mechanism 1, we conclude that Mechanism 1 is a more likely reaction path.

The free energy profiles of the afatinib reaction in both EGFR and solution demonstrate a classic Michael addition in which the NA step is rate-limiting and a stable intermediate is formed before reprotonation to the thiol adduct state. This is consistent with Rowley and coworker's QM/MM simulations of ibrutinib reaction with BTK,[21] although their simulations started from a deprotonated thiolate state.

Formation of a hybrid carbanion/enolate intermediate. To further understand the reaction mechanism (1), we examined the structure changes near the reaction center along the reaction path using the configurations from the free energy simulations (details see Supplemental Methods). The selected distances and Mulliken atomic partial charges of the reaction center along the reaction path are given in Table S6 and S7. Of particular interest is the nature of the intermediate state INT2. In the literature, it is often stated that the Michael addition of acrylamides give an enolate without discussion of the carbanion tautomer.[7,21,30] Our QM/MM simulations in both EGFR and solution demonstrated that the intermediate resulting from the NA step is a hybrid carbanion/enolate. In the INT2 state of EGFR (Figure 3), the S$-$Cβ bond is formed (distance of 1.85 Å as compared to 3.36 Å in the R1 state), and the proton re-

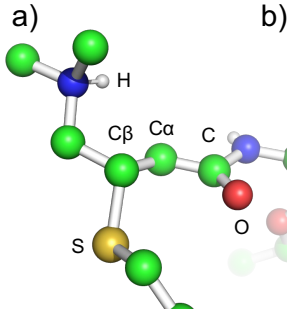


Scheme 2: **The calculated free energy profile along the reaction path in EGFR (red) and in aqueous solution (black) according to Mechanism 1.** The free energies are corrected (see Supplemental methods). The states correspond to those in Figure 2c and d. R0: reactant state with neutral Cys797 (or methanethiol in solution) and β-DMAM substituted acrylamide. R1: reactant state (ion-pair) with the charged Cys797 thiolate (or methylthiolate in solution) and β-DMAM$^+$ substituted acrylamide. TS1: transition state for PT-1 (proton transfer from the thiol to the β-DMAM group). TS2: transition state for the NA step. INT1 and INT2: carbanion/enolate intermediates (differs by a conformational rearrangement of the β-DMAM group). INT2 is also the reactant state (R2) for PT-2 (proton transfer from β-DMAM$^+$ to Cα). P: thioether adduct.

mains bonded to the β-DMAM nitrogen (N–H distance of 1.01 Å). The Cβ−Cα distance is elongated to 1.47 Å from the C=C bond distance of 1.34 Å in the R1 state, although still slightly shorter than the C−C bond distance of 1.54 Å in the product state.

An ideal enolate has a Cα=C bond and a C−O bond, whereas an ideal carbanion has a Cα−C bond and a C=O bond. The INT2 state shows that the Cα−C and C−O distances are 1.39 and 1.27 Å, respectively, which are shorter than the respective C−C and C−O single bond lengths of 1.54 and 1.43 Å, but longer than the respective C=C and C=O double bond lengths of 1.33 and 1.23 Å. In solution, the aforementioned distances are identical to the values in EGFR. Thus, our QM/MM simulations suggest that the intermediate of the NA step is a hybrid carbanion/enolate state. Such a hybrid carbanion/enolate state is also demonstrated by the Mulliken charges (Supplemental Table S7). In INT2 of EGFR, the negative charge is distributed to both Cα (-0.7) and O (-0.55) atoms, with more negative charge on Cα. In solution, the negative charge is distributed slightly more onto the Cα atom, with a charge of -0.73 as compared to -0.50 on the O atom.

a)



Reaction center

b)

	R1	INT2	P
Cβ	-0.14	-0.11	-0.26
Cα	-0.24	-0.70	-0.22
C	0.29	0.27	0.24
O	-0.33	-0.55	-0.40
S--Cβ	3.36	1.85	1.85
Cβ--Cα	1.34	1.47	1.54
Cα--C	1.49	1.39	1.52
C--O	1.25	1.27	1.24

Figure 3: **Changes in the bond lengths and atomic partial charges in the reaction center between the reactant, intermediate, and product state.** a) The geometry of the reaction center in the INT2 state. Nonpolar hydrogens are hidden. Both afatinib and cysteine are cropped, and the protein (EGFR) as well as water are hidden. Relevant atoms are labeled. b) Relevant bond lengths (Å) and Mulliken partial charges (a.u.) in R1 (before NA), INT2 (carbanion/enolate), and P states.

Asp800 stabilizes the ion-pair reactant state. To identify critical residues in EGFR that contribute to the formation of the ion-pair state R1 and the final adduct state, we performed the interaction energy decomposition analysis.[31,32] For each selected residue, we calculated its QM/MM interaction energy with the QM region, which is defined as the difference in the QM energy before and after annihilation of the MM charges on the residue. This calculation was performed for all the trajectory snapshots representing the R0, R1, and P states, and for each state, the interaction energies were averaged for each residue (Supplemental methods for more details). Thus, the interaction energy difference between two states represents the residue-specific contribution to the stabilization or destabilization of the state. Figure 4 shows the residue-based interaction energies for stabilizing (or destabilizing) the R1 state (relative to R0 state) or the P state (relative to R1 state). Three residues make greater than 1 kcal/mol contributions to the stabilization or destabilization of the R1 state relative to the R0 state. Arg803 and Asp855 destabilize the R1 state by 1.2 kcal/mol and 2.5 kcal/mol, respectively, whereas Asp800 stabilizes the R1 state by 8.3 kcal/mol, making it a major contributor to the PT-1 step. Since Asp800 is absent in solution, this may explain why R1 does not occur in solution, as it has much higher free energy (Figure 2d). Examination of the trajectory snapshots revealed that the stabilizing interaction of Asp800 with afatinib arises from the electrostatic interaction (a weak hydrogen bond) between the carboxylate oxygen of Asp800 and the β-DMAM nitrogen. The distance between the reactive hydrogen of β-DMAM and the nearest carboxylate oxygen is $\sim$2.9 Å in the R1 state and it reduces to $\sim$2.6 Å in the TS2 state. Thus, Asp800 not only stabilizes the ion-pair complex but also lower the free energy barrier of the NA step. The latter effect is missing in solution, leading to a much higher barrier. In the product state, however, Asp800 contributes over 8 kcal/mol destabilization energy with respect to the R1 state, which may explain the

less favorable reaction free energy in EGFR as compared to solution (Figure 2).

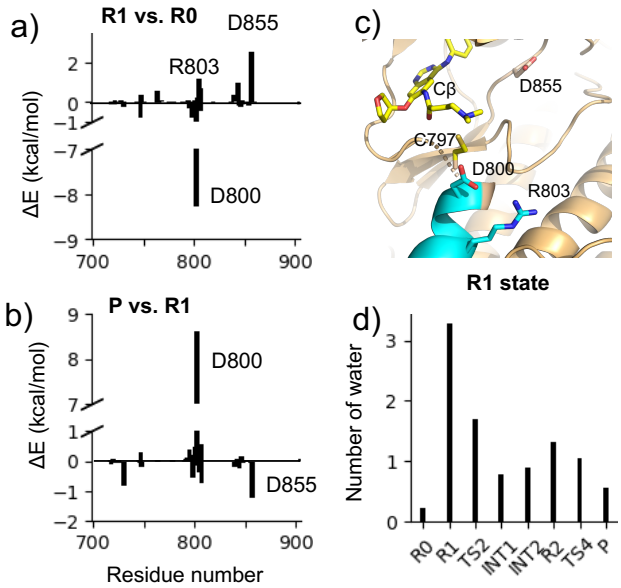


Figure 4: **Interaction energy decomposition and hydration analysis. a,b)** Residue-based electrostatic contributions to the QM/MM energy change of the R1 vs. R0 state (a) and the P vs. R1 state (b). Residues with contributions greater than 1.0 kcal/mol are labeled. **c)** A zoomed-in view of the locations of the three residues making the largest electrostatic energy contributions shown in a) and b). **d)** The average number of water within 3.5 Å of the Cys797 sulfur in different states along the reaction path.

Hydration of the various states along the reaction path. We analyzed the number of water within 3.5 Å of the Cys797 sulfur atom at different states along the reaction path using the aforementioned configurations extracted from the free energy simulations (Figure 4d). As expected, the charged Cys797 thiolate in the R1 state is well hydrated, with an average of 3.3 water molecules surrounding the sulfur atom; by contrast, the cysteine is desolvated in the neutral R0 state, with nearly no water. From the R1 to TS2 and INT1/INT2 states, the average hydration number decreases from 3.3 to 1.7 and 0.8/0.82, implying that the nucleophilic attack of Cys797 thiolate on the Cβ atom is coupled with the desolvation of the thiolate, which is consistent with a previous QM/MM study of the reaction between the covalent inhibitor 13-JAB with EGFR using the semiempirical DFTB/MM potential.[13] The low hydration number remains from the INT1/INT2 states to the product state since the S-C covalent bond is formed.

Conclusions

The covalent bond formation mechanism between afatinib and Cys797 of the EGFR kinase was investigated using the combined PDDG-PM3/MM free energy simulations. Specifically, we evaluated two possible mechanisms, in which either the β-DMAM group on the acrylamide warhead or a nearby Asp800 acts as a general base to initiate the nucleophilic attack of the Cys797 thiolate. The calculated free energy profiles showed that in either mechanism, the reaction follows a classical Michael addition, which involves a stable intermediate and the nucleophilic attack as the rate-limiting step. The Michael addition mechanism of afatinib in EGFR is consistent with the solution chemistry experiments[4,6,8] and QM calculations[7,15,19] as well as the recent DFT QM/MM simulations of the reaction of ibrutinib with the BTK kinase.[21] However, the Michael addition mechanism of afatinib does not support the concerted mechanism concluded from the earlier semiempirical DFTB QM/MM simulations of the reactions of 13-JAB and osimertinib with EGFR.[13,14]

Our QM/MM simulations suggest that Mechanism 1 (β-DMAM is the general base/acid) is more likely than Mechanism 2 (Asp800 is the general base/acid) for afatinib's reaction with EGFR, as the former has a much lower barrier (19.5 kcal/mol) for the rate-limiting nucleophilic attack step, which is in close agreement with the barrier height of 21.6 kcal/mol estimated from the experimental rate constant of afatinib with EGFR.[29] The interaction energy analysis for Mechanism 1 showed that Asp800 is the dominant contributor to the stabilization of the ion-pair (thiolate$^-$/β-DMAM$^+$) R1 state. Thus, our data support a mechanism by which the β-DMAM group is the gen-

eral base while Asp800 plays a supporting role in stabilizing the reactant and the transition state. These findings offer atomic details in support of an early hypothesis regarding the role of β-methylamino substitution on acrylamides and the nearby Asp800.[25] The role of Asp800 is also congruent with the solution simulation results, which showed that in the absence of Asp800 and the protein environment, the PT-1 step is concerted with the NA step and the rate-limiting NA step has a much (10.3 kcal/mol) higher barrier.

Based on our QM/MM simulations, we suggest that the reactions of EGFR inhibitors without a β-methylamino substitution, e.g., osimertinib and 13-JAB, follow a classical Michael addition mechanism with Asp800 as a general base, which is inconsistent with the previous semiempirical DFTB QM/MM simulations.[13,14] The feasibility of this reaction mechanism is supported by the stepwise reaction profile for Mechanism 2 of afatinib, which has a rate-limiting barrier height of 31.3 kcal/mol. One aspect that warrants future investigation is the potential role of water. For afatinib, there might be a third possible mechanism, i.e., a nearby water acts as the general base/acid for the Michael addition while β-DMAM and Asp800 remain charged. The latter aspect can be examined by the recently developed GPU-accelerated all-atom continuous constant pH molecular dynamics method.[33] Notwithstanding, our work elucidates an important structure-activity relationship of acrylamides that is generally applicable for cysteine-directed TCI designs.

Computational Details

Validation of the PDDG-PM3 QM potential. For the QM/MM simulations of afatinib reaction with EGFR, the QM region included the entire inhibitor and the side chain of Cys797 with the Cα as the boundary atom (a total of 65 atoms, Figure 2b). For the solution model reaction simulations, the QM region included the entire inhibitor and methanethiol. To make such large QM regions computationally feasible, the semiempirical Pairwise Distance Directed Gaussian modified PM3 method (PDDG-PM3)[26,27] was employed. The PDDG-PM3 calculated heats of formation of model compounds methanethiol, methylthiolate, and acrylamide ($CH_2{=}CHCONH_2$) are in good agreement with experimental data (Table S5). We also compared the model reaction enthalpies calculated by the PDDG-PM3 method and the long-range corrected meta-GGA DFT method ωB97X-D3BJ,[34,35] which was used in our previous work on the solution reactivities of model acrylamide warheads[19] (Table S2). PDDG-PM3 calculations were performed using CHARMM;[36] ωB97X-D3BJ calculations were performed using the ORCA package (version 5.1.1).[37]

The QM/MM boundary atoms (Cα of Cys797 in Mechanism 1 and 2, and Cα of Asp800 in Mechanism 2) were treated by the generalized hybrid orbital (GHO) method.[38] The GHO parameters for carbon atom in the combined PDDG-PM3/MM potential was obtained following the protocol used in Ref[39] (See more details in SI).

System preparation and MD simulations. For Mechanism 1, EGFR (except for the sidechain of Cys797), ions, and water comprised the MM region. The CHARMM22[40] force field was used for the protein and ions, while the TIP3P model was used for water.[41] The simulations of the reaction system was based on the X-ray co-crystal structure of EGFR in covalent complex with afatinib (PDB ID: 4G5J).[12] The coordinates of the missing residues 721-724 and 747-756 were built based on the crystal structure of EGFR (PDB ID: 5U8L).[42] The protonation states of titratable sites (Asp, Glu, His, Cys, and Lys) in EGFR were adopted from a previous study[9] using the generalized-Born continuous constant pH MD titration simulations[43] based on the crystal structure (PDB ID: 5U8L, ligand removed).[42]

The CHARMM package (version c45b2)[36] was used for the simulations. The system (EGFR-afatinib complex and crystal wa-

ter) was first minimized for 100 steps using the adopted-basis Newton Raphson (ABNR) method and then placed in a rectangular water box with at least 10 Å cushion between the protein heavy atoms and the edges of the box. Water molecules within 2.5 Å of any protein or inhibitor heavy atoms were deleted, and the crystal water molecules were kept. Three water in the bulk that were about 6 Å away from negatively charged residues on protein surface were selected and replaced with sodium ions to neutralize the system. In total, there were 26 crystal water and 12071 bulk water in the system. The solvated system was energy minimized using the PDDG-PM3/MM potential and the ABNR algorithm, first for the protein and QM atoms (40 steps), then water and ions (90 steps), finally the entire system (40 steps). Next, the system was heated to 298 K for 15 ps followed by equilibration for 50 ps with periodic boundary conditions and the constant NPT ensemble at 298 K and 1 atm. The Nośe-Hoover thermostat[44,45] was used. The particle mesh Ewald method[46] was employed for long-range electrostatic calculations with a real-space cutoff of 12 Å and a grid spacing of 1 Å. The nonbonded interactions were smoothly switched off between 12 and 13 Å. The nonbonded pair list was updated every 25 steps. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm.[47]

For QM/MM simulations of the model reaction in solution, the structures of afatinib and methanethiol (built from the side chain of Cys797) were taken from the X-ray co-crystal structure (PDB ID: 4G5J).[12] The system was solvated in a cubic water box (36x36x36 $Å^3$), and partitioned into the QM (afatinib and methanethiol) and MM (water) regions. The solvated system underwent energy minimization for 40 steps for the QM atoms, 60 steps for the MM water, and another 40 steps for the entire system. The heating and equilibration protocols were the same as for the protein system.

A similar system preparation and MD protocol was applied for Mechanism 2, except that the QM part includes 73 atoms (afatinib and the sidechains of Cys797 and Asp800) with two boundary atoms (Cα of Cys797 and Asp800), and there were 12063 bulk water molecules and two sodium ions.

Umbrella sampling free energy simulations. Following the MD equilibration, the free energy simulations were initiated from the EGFR-afatinib adduct state (or the methanethiol-afatinib adduct state for the model reaction) via the two possible mechanisms (Scheme 1). Free energy changes of the chemical steps were calculated in the reverse order of that shown in Scheme 1.

The adaptive umbrella sampling technique[48,49] (module RXNCOR in CHARMM) was used. For the PT-2 step, 19 windows (23 for model reaction) in the reaction coordinate range of -1.8–1.8 Å (-1.6–2.4 Å for the model reaction). The NA and PT-1 steps were simulated through a two-dimensional free energy simulation with 327 windows (341 for the model reaction) in the reaction coordinate range of -2 to 3.2 Å for PT-1 and -2.2 to -0.3 Å for NA. For each umbrella sampling window, 100 ps equilibration was performed followed by 100 ps production run for trajectory collection. For the PT-2 step, a biasing potential and a harmonic restraining potential with the force constant of 30–100 kcal/mol/$Å^2$ were placed on the reaction coordinates. While for the two-dimensional simulation of NA and PT-1 steps, only a harmonic restraining potential with the force constant of 30–100 kcal/mol/$Å^2$ were placed on the reaction coordinates. 500 structures from each simulation window were saved for analysis. The weighted histogram analysis method (WHAM)[50] was used to obtain the unbiased probability densities and free energy changes.

QM/MM interaction energy decomposition analysis. A well-established energy decomposition method[31,32] was applied to calculate the residue-based energy contributions to the stabilization or destabilization of the states along the reaction path. For the saved snapshots of a state, the QM/MM interaction ener-

gies of all residues in EGFR were calculated by sequentially zeroing the MM charges on a residue. The interaction energy between the residue and the QM region is defined as the difference in the QM electrostatic energy before and after the charge annihilation. The results were averaged over all the snapshots of each state.

Acknowledgements

Financial support from the National Institutes of Health (R01CA256557) is acknowledged.

Supporting Information Available

Protocol for the validation of the PDDG-PM3-GHO parameters. Protocol for the corrections of the PDDG-PM3 calculated energies based on the ωB97X-D3BJ DFT calculations. Protocol for selecting trajectory snapshots for the various states. Table S1 compares the calculated geometries of propane using the MM, PDDG-PM3/MM, PDDG-PM3, and HF/6-31G(d) methods. Table S2 contains the Mulliken atomic charges for atoms 1-5 of propane calculated by the PDDG-PM3 and PDDG-PM3/MM methods. Table S3 contains the modified parameters for the GHO carbon atom in combined PDDG- PM3/MM potential. Table S4 contains the calculated proton affinities for the blocked aspartic acid and cysteine. Table S5 compares the gas-phase heats of formation of model compounds from the PDDG-PM3 calculations and experiment. Table S6 and S7 contain the calculated reaction center distances and Mulliken atomic charges. Fig. S1 explains the QM/MM partition of propane. Fig. S2 and S3 compare the gas-phase model reaction enthalpies for Mechanism 1 and 2 from the PDDG-PM3 and DFT calculations. Fig. S4–S5 contain the calculated free energy profiles of the afatinib and model reactions according to Mechanism 1 without the DFT corrections. Fig. S6, S7 contain the calculated free energy profiles of the afatinib reaction in EGFR without (Fig. S6) and with (Fig. S7) the DFT corrections.

Data Availability

The CHARMM MD program is freely available for academic use at `https://academiccharmm.org/news/free-charmm`. The ORCA QM package is freely available for academic use at `https://orcaforum.kofo.mpg.de`. Cartesian coordinates of the model compounds are given in the downloadable PDF file (see Supporting Information). All simulation input files are available for download at `https://github.com/JanaShenLab/Afatinib_reaction_mechanism`.

Abbreviations

density functional theory, DFT; DMAM, dimethylaminomethyl; human epidermal growth factor receptor 1, EGFR; molecular mechanics, MM; molecular dynamics, MD; targeted covalent inhibitor, TCI; quantum mechanical, QM

References

(1) Ferguson, F. M.; Gray, N. S. Kinase Inhibitors: The Road Ahead. *Nat. Rev. Drug Discov.* **2018**, *17*, 353–377.

(2) Boike, L.; Henning, N. J.; Nomura, D. K. Advances in Covalent Drug Discovery. *Nat. Rev. Drug Discov.* **2022**, *21*, 881–898.

(3) Roskoski, R. Orally Effective FDA-approved Protein Kinase Targeted Covalent Inhibitors (TCIs). *Pharmacol. Res.* **2021**, *165*, 105422.

(4) Dmuchovsky, Basil.; Vineyard, B. D.; Zienty, F. B. The Mechanism of the Base-Catalyzed Addition of Thiols to Maleic Anhydride. *J. Am. Chem. Soc.* **1964**, *86*, 2874–2877.

(5) Fishbein, J. C.; Jencks, W. P. Elimination Reactions of .Beta.-Cyano Thioethers: Evidence for a Carbanion Intermediate and a Change in Rate-Limiting Step. *J. Am. Chem. Soc.* **1988**, *110*, 5075–5086.

(6) Heo, C. K. M.; Bunting, J. W. Rate-Determining Steps in Michael-type Additions and E1cb Reactions in Aqueous Solution. *J. Org. Chem.* **1992**, *57*, 3570–3578.

(7) Roseli, R. B.; Keto, A. B.; Krenske, E. H. Mechanistic Aspects of Thiol Additions to Michael Acceptors: Insights from Computations. *WIREs Comput. Mol. Sci.* **2023**, *13*.

(8) Böhme, A.; Thaens, D.; Paschke, A.; Schüürmann, G. Kinetic Glutathione Chemoassay To Quantify Thiol Reactivity of Organic Electrophiles—Application to α,β-Unsaturated Ketones, Acrylates, and Propiolates. *Chem. Res. Toxicol.* **2009**, *22*, 742–750.

(9) Liu, R.; Yue, Z.; Tsai, C.-C.; Shen, J. Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors. *J. Am. Chem. Soc.* **2019**, *141*, 6553–6560.

(10) Liu, R.; Zhan, S.; Che, Y.; Shen, J. Reactivities of the Front Pocket N-Terminal Cap Cysteines in Human Kinases. *J. Med. Chem.* **2022**, *65*, 1525–1535.

(11) Liu, R.; Verma, N.; Henderson, J. A.; Zhan, S.; Shen, J. Profiling MAP Kinase Cysteines for Targeted Covalent Inhibitor Design. *RSC Med. Chem.* **2022**, *13*, 54–63.

(12) Solca, F.; Dahl, G.; Zoephel, A.; Bader, G.; Sanderson, M.; Klein, C.; Kraemer, O.; Himmelsbach, F.; Haaksma, E.; Adolf, G. R. Target Binding Properties and Cellular Activity of Afatinib (BIBW 2992), an Irreversible ErbB Family Blocker. *J. Pharmacol. Exp. Ther.* **2012**, *343*, 342–350.

(13) Capoferri, L.; Lodola, A.; Rivara, S.; Mor, M. 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.

(14) Callegari, D.; Ranaghan, K. E.; Woods, C. J.; Minari, R.; Tiseo, M.; Mor, M.; Mulholland, A. J.; Lodola, A. L718Q Mutant EGFR Escapes Covalent Inhibition by Stabilizing a Non-Reactive Conformation of the Lung Cancer Drug Osimertinib. *Chem. Sci.* **2018**, *9*, 2740–2749.

(15) Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. Mechanistic Study of the Reaction of Thiol-Containing Enzymes with α,β-Unsaturated Carbonyl Substrates by Computation and Chemoassays. *ChemMedChem* **2010**, *5*, 869–880.

(16) Smith, J. M.; Alahmadi, Y. J.; Rowley, C. N. Range-Separated DFT Functionals Are Necessary to Model Thio-Michael Additions. *J. Chem. Theory Comput.* **2013**, 6.

(17) Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P.; Taunton, J. Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *J. Am. Chem. Soc.* **2014**, *136*, 12624–12630.

(18) Krenske, E. H.; Petter, R. C.; Houk, K. N. 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.

(19) Liu, R.; Vazquez-Montelongo, E.; Ma, S.; Shen, J. Quantum Descriptors for Predicting and Understanding the Structure-Activity Relationships of Michael Accep-

tor Warheads. *J. Chem. Inf. Model.* **2023**, *63*, 4912–4923.

(20) Voice, A. T.; Tresadern, G.; Twidale, R. M.; van Vlijmen, H.; Mulholland, A. J. Mechanism of Covalent Binding of Ibrutinib to Bruton's Tyrosine Kinase Revealed by QM/MM Calculations. *Chem. Sci.* **2021**, *12*, 5511–5516.

(21) Awoonor-Williams, E.; Rowley, C. N. Modeling the Binding and Conformational Energetics of a Targeted Covalent Inhibitor to Bruton's Tyrosine Kinase. *J. Chem. Inf. Model.* **2021**, *61*, 5234–5242.

(22) Tsou, H.-R.; Mamuya, N.; Johnson, B. D.; Reich, M. F.; Gruber, B. C.; Ye, F.; Nilakantan, R.; Shen, R.; Discafani, C.; DeBlanc, R.; Davis, R.; Koehn, F. E.; Greenberger, L. M.; Wang, Y.-F.; Wissner, A. 6-Substituted-4-(3-Bromophenylamino)Quinazolines as Putative Irreversible Inhibitors of the Epidermal Growth Factor Receptor (EGFR) and Human Epidermal Growth Factor Receptor (HER-2) Tyrosine Kinases with Enhanced Antitumor Activity. *J. Med. Chem.* **2001**, *44*, 2719–2734.

(23) Wissner, A. et al. Synthesis and Structure-Activity Relationships of 6,7-Disubstituted 4-Anilinoquinoline-3-carbonitriles. The Design of an Orally Active, Irreversible Inhibitor of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor (EGFR) and the Human Epidermal Growth Factor Receptor-2 (HER-2). *J. Med. Chem.* **2003**, *46*, 49–63.

(24) Birkholz, A.; Kopecky, D. J.; Volak, L. P.; Bartberger, M. D.; Chen, Y.; Tegley, C. M.; Arvedson, T.; McCarter, J. D.; Fotsch, C.; Cee, V. J. Systematic Study of the Glutathione Reactivity of *N* -Phenylacrylamides: 2. Effects of Acrylamide Substitution. *J. Med. Chem.* **2020**, *63*, 11602–11614.

(25) Wood, E. R. et al. 6-Ethynylthieno[3,2-d]- and 6-Ethynylthieno[2,3-d]Pyrimidin-4-Anilines as Tunable Covalent Modifiers of ErbB Kinases. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2773–2778.

(26) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods. *J. Comput. Chem.* **2002**, *23*, 1601–1622.

(27) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. Extension of the PDDG/PM3 Semiempirical Molecular Orbital Method to Sulfur, Silicon, and Phosphorus. *J. Chem. Theory Comput.* **2005**, *1*, 817–823.

(28) of Pure and Applied Chemistry, I. U., Serjeant, E. P., Dempsey, B., of Pure and Applied Chemistry, I. U., Eds. *Ionisation constants of organic acids in aqueous solution*; IUPAC chemical data series no. 23; Pergamon Press: Oxford ; New York, 1979.

(29) Schwartz, P. A.; Kuzmic, P.; Solowiej, J.; Bergqvist, S.; Bolanos, B.; Almaden, C.; Nagata, A.; Ryan, K.; Feng, J.; Dalvie, D.; Kath, J. C.; Xu, M.; Wani, R.; Murray, B. W. Covalent EGFR Inhibitor Analysis Reveals Importance of Reversible Interactions to Potency and Mechanisms of Drug Resistance. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 173–178.

(30) Awoonor-Williams, E.; Isley, W. C.; Dale, S. G.; Johnson, E. R.; Yu, H.; Becke, A. D.; Roux, B.; Rowley, C. N. Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols. *J. Comput. Chem.* **2020**, *41*, 427–438.

(31) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. Computer Simulation and Analysis of the Reaction Pathway of

Triosephosphate Isomerase. *Biochemistry* **1991**, *30*, 5826–5832.

(32) Ma, S.; Devi-Kesavan, L. S.; Gao, J. 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.

(33) Harris, J. A.; Liu, R.; Martins de Oliveira, V.; Vázquez-Montelongo, E. A.; Henderson, J. A.; Shen, J. GPU-Accelerated All-Atom Particle-Mesh Ewald Continuous Constant pH Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2022**, *18*, 7510–7527.

(34) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(35) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(36) Brooks, B. et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(37) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152*, 224108.

(38) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

(39) Garcia-Viloca, M.; Gao, J. Generalized Hybrid Orbital for the Treatment of Boundary Atoms in Combined Quantum Mechanical and Molecular Mechanical Calculations Using the Semiempirical Parameterized Model 3 Method. *Theor. Chem. Acc.* **2004**, *111*, 280–286.

(40) MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(42) Zhao, Q.; Ouyang, X.; Wan, X.; Gajiwala, K. S.; Kath, J. C.; Jones, L. H.; Burlingame, A. L.; Taunton, J. Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes. *J. Am. Chem. Soc.* **2017**, *139*, 680–685.

(43) Huang, Y.; Harris, R. C.; Shen, J. Generalized Born Based Continuous Constant pH Molecular Dynamics in Amber: Implementation, Benchmarking and Analysis. *J. Chem. Inf. Model.* **2018**, *58*, 1372–1383.

(44) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemblet. *Mol. Phys.* **1984**, *52*, 255–268.

(45) Hoover, W. G. Canonical Dynamics: Equilibration Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An W Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 4.

(47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(48) Kottalam, J.; Case, D. A. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.* **1988**, *110*, 7690–7697.

(49) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J. Chem. Phys.* **1992**, *97*, 6690–6694.

(50) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.