Reactivities of the Front Pocket N-Cap Cysteines in Human Kinases

Ruibin Liu,[†] Shaoqi Zhan,^{†,‡} Ye Che,[¶] and Jana Shen^{*,†}

[†] Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201

[‡] Joint first author

[¶] Pfizer Worldwide Research, Development and Medical, Groton, CT 06340

Received June 28, 2021; E-mail: jana.shen@rx.umaryland.edu

Abstract: Discovery of targeted covalent inhibitors directed at nucleophilic cysteines is attracting enormous interest. The front pocket (FP) N-cap cysteine has been the most popular site of covalent modification in kinases. Curiously, a long-standing hypothesis associates the N-cap position with cysteine hyper-reactivity; however, traditional computational methods suggest that the FP N-cap cysteines in all human kinases are predominantly unreactive at physiological pH. Here we applied a newly developed GPU-accelerated continuous constant pH molecular dynamics (CpHMD) tool to test the N-cap hypothesis and elucidate the cysteine reactivities. Simulations showed that the N-cap cysteines in BTK/BMX/TEC/ITK/TKX, JAK3, and MKK7 sample the reactive thiolate form to varying degrees at physiological pH; however, those in BLK and EGFR/ERBB2/ERBB4 which contain an Asp at the N-cap+3 position adopt the unreactive thiol form. The latter argues in favor of the base-assisted thiol-Michael addition mechanisms as suggested by the quantum mechanical calculations and experimental structure-function studies of EGFR inhibitors. Analysis revealed that the reactive N-cap cysteines are stabilized by hydrogen bond as well as electrostatic interactions, and in their absence a N-cap cysteine is unreactive due to desolvation. To test a corollary of the N-cap hypothesis, we also examined the reactivities of the FP N-cap+2 cysteines in JNK1/JNK2/JNK3 and CASK. Additionally, our simulations predicted the reactive cysteine and lysine locations in all 15 kinases. Our findings offer a systematic understanding of cysteine reactivities in kinases and demonstrate the predictive power and physical insights CpHMD can provide to guide the rational design of targeted covalent inhibitors.

INTRODUCTION

Protein kinases are enzymes that catalyze protein phosphorylation reactions, and represent one of the largest gene families with over 500 genes discovered so far.¹ Dysregulation of kinase functions plays significant roles in cancer, immunology, and a host of other diseases.¹ In the past decade, the highest number of approvals by the US Food and Drug Administration (FDA) have gone to inhibitors targeting kinases as compared to other protein families.² While traditional kinase inhibitors are small molecules that bind the ATP-binding site via reversible interactions,¹ targeted covalent kinase inhibitors (TCKIs) are emerging at a rapid pace^{3,4} due to the advantages such as prolonged duration of action, enhanced potency, and increased selectivity.^{5,6} Starting from a low affinity reversible inhibitor, a TCKI is typically developed by structure-guided incorporation of an electrophilic warhead that covalently links with a nearby nucleophilic side chain.⁵

Figure 1. Human kinases with a FP N-cap or N-cap+2 Cys. (a) 15 kinases possessing a FP N-cap or N-cap+2 Cys displayed in the human kinome tree (generated with KinMap⁷). (b) Structure of the kinase domain of BMX showing a FP N-cap Cys (orange sphere, PDB ID: 3sxs⁸). The frequently occurring Asn/Asp/Glu at N-cap+3 position is shown as a blue sphere. (c) Names of the kinases possessing a FP N-cap or N-cap+2 Cys, the families they belong to, and the sequences of nearby amino acids. The FP N-cap Cys is colored orange, the N-cap+2 Cys is colored yellow, and nearby residues making a significant contribution to its pK_a shift are colored blue, magenta, green, or brown (same as in Fig. 2).

Owing to the high nucleophilicity of thiolate and the non-catalytic role in kinases, cysteine is the most targeted amino acid in covalent drug discoveries.^{1,5,21} According to a survey from 2018,³ over half of the TCKIs are directed at a Cys located in the linker at the N-terminal end (N-cap) of the αD helix near the front pocket (FP, Fig. 1). Sequence alignment (by us and others^{5,22}) showed that 11 kinases in the human kinome possess such a FP N-cap Cys, including BTK/BMX/TEC/ITK/TKX from the Tec family, BLK from the Src family, JAK3 from the JakA family, EGFR/ERBB2/ERBB4 from the EGFR family, and MKK7 from the STE7 family (Fig. 1), all of which are therapeutic targets.

Table 1. Summary of kinases with a FP N-cap or N-cap+2 Cys and the covalent inhibitors^a directed at the location

Kinase	Family	Group	PDB	Residue	PROPKA ^b	TCKI
FP N-cap Cys						
BTK	Tec	TK	3pj3	C481	9.6	ibrutinib, acalabrutinib, zanubrutinib, tirabrutinib, ⁹ orelabrutinib ¹⁰
BMX	Tec	TK	3sxs	C496	9.9	ibrutinib
TEC	Tec	TK	4hct ^a	C449	9.2	ibrutinib, ritlecitinib ^c
ITK	Tec	TK	3miy	C442	10.7	ibrutinib
TXK	Tec	TK	3t9t ^a	C350	10.4	acalabrutinib
BLK	Src	TK	4mxy ^a	C319	10.6	ibrutinib, acalabrutinib
JAK3	JakA	TK	5lwm	C909	10.3	ritlecitinib ¹¹
EGFR	EGFR	TK	4zjv	C797	11.4	afatinib, osimertinib, dacomitinib, neratinib, olmutinib, ¹² pyrotinib, ¹³ mobocertinib, ¹⁴ ibrutinib, acalabrutinib
ERBB2	EGFR	TK	3pp0	C805	10.8	acalabrutinib
ERBB4	EGFR	TK	2r4b	C803	11.0	ibrutinib
MKK7	STE7	STE	6qfr	C218	9.1	investigational inhibitors ^{15,16}
FP N-cap+2 Cys						
JNK1	MAPK	CMGC	2xrw	C116	10.0	investigational inhibitors ¹⁷
JNK2	MAPK	CMGC	3e7o	C116	10.2	investigational inhibitors ¹⁷
JNK3	MAPK	CMGC	6emh	C154	10.4	investigational inhibitors ¹⁸
CASK	CASK	CAMK	3mfr	C100	13.0	unknown

^a Inhibitors without references are from the FDA approved ones from the FDA Orange Book (<https://www.accessdata.fda.gov/scripts/cder/ob/index.cfm>). Those with references are in the last stage of clinical trial; some of them have gained regulatory approval in foreign countries. ^b PROPKA calculations were performed with PROPKA3.¹⁹

^c Template structure used for homology modeling with SWISS-MODEL.²⁰

BTK/BMX/TEC/ITK/TXK are key components of T-cell receptor signaling and contribute to acting reorganization and cell polarization.^{8,23–25} BLK is an oncogene and a potential therapy target in cutaneous T-cell lymphoma.²⁶ JAK3 drives signaling through cytokine receptors, and is associated with various immune-mediated diseases.^{27,28} EGFR (or ERBB1 or HER1), ERBB2 (or HER2), and ERBB4 (or HER4) play essential roles in regulating cell proliferation, survival, differentiation and migration, and understanding their roles has fueled the development of targeted therapies for cancer in the past decade.^{29–32} MKK7 is associated with multiple myeloma and various neurological diseases.^{15,16,33}

The FP N-cap Cys in all human kinases have been covalently modified. Most notably, the FDA-approved drug ibrutinib covalently modifies the N-cap Cys in BTK/BMX/TEC/ITK, BLK, JAK3, and EGFR/ERBB4 (Table 1).^{22,34,35} The same Cys in EGFR is also targeted by newer covalent drugs such as afatinib, neratinib, acalabrutinib, spebrutinib, mavelertinib, and osimertinib. The popularity of the FP N-cap Cys in TCKI discovery can be explained by its accessibility (near the ATP binding site) and expected reactivity (see below). A N-cap Asp has been known to stabilize a helix, which was attributed to the favorable interaction between the negative aspartate and the positive helix dipole.^{36,37} Similarly, a N-cap Cys was found to be helix stabilizing, which was attributed to the thiolate-helix dipole interaction.^{38,39} It was argued that Cys is the least frequent N-cap residue, because it is in the thiolate form prone to chemical modification.³⁸ Indeed, the N-cap Cys in myoglobin has a measured pK_a of 6.5, which is 1.9 or 1.6 units lower than the (mutant) Cys at N2 or N1 position, and 2 units lower than the pK_a of 8.6 for the model Cys in a peptide.³⁹ Thus, the consensus is that a N-cap Cys is in

the hyper-reactive thiolate form at physiological pH, which we will refer to as the N-cap hypothesis.

Kinetics experiments demonstrated that the observed (apparent) rate constants of thiol reactions with electrophiles are inversely related to the thiol pK_a 's.^{40–43} Thus, it is generally accepted that cysteines with low pK_a 's are more prone to covalent modification due to the increased availability of the nucleophilic thiolate form, and the Cys pK_a can be used as a proxy for reactivity in TCKI design.^{1,5,21,38} Curiously, however, according to the popular empirical pK_a prediction program PROPKA,¹⁹ the N-cap Cys in the aforementioned 11 kinases have pK_a 's of 9.1–11.4 (Table 1), suggesting that they are mainly in the thiol form and consequently unreactive or have low reactivities towards electrophiles. Note, the intrinsic nucleophilicity of a thiolate often increases and sometimes stays constant or decreases with the thiol pK_a value depending on the molecule by molecule basis;^{40,43} this is not a topic of this paper.

Accurate prediction of Cys pK_a 's is challenging using traditional computational methods. For a dataset of 18 proteins, Rowley and coworker showed that the root-mean-square errors (RMSE's) from the experimental pK_a 's are 3.4–4.7 using the static structure based Poisson-Boltzmann (PB) and empirical PROPKA calculations.⁴⁴ They obtained a RMSE of 2.4–3.2 using the molecular dynamics (MD) based thermodynamic integration (TI) method, which is several orders of magnitude slower than the PB or PROPKA calculations. This level of accuracy is however on par with that (RMSE of 2.7) from the null model which assumes the solution (or model) pK_a of 8.6⁴⁵ for all cysteines.

Recently, we developed a GPU-accelerated continuous constant pH MD (CpHMD) method in Amber package⁴⁶ for

accurate and rapid prediction of protein pK_a 's based on independent pH⁴⁷ or pH replica-exchange titration⁴⁸ simulations in the GB-Neck2 implicit solvent.⁴⁹ Using replica-exchange CpHMD, which significantly accelerates the pK_a convergence, we scanned the kinome structure database and found that the catalytic (roof) Lys in a dozen of human kinases can become reactive in the rare DFG-out/ α C-out inactive conformation.⁵⁰ Most recently, we implemented an asynchronous algorithm to allow replica-exchange simulations on a single or any number of GPU cards.⁵¹ Benchmark study based on 24 proteins, including those with large pK_a downshifts examined by Rowley and coworker⁴⁴ and additional ones with large pK_a upshifts relative to the model value, gave a RMSE of 1.2, which is more than two units lower than the PB or empirical calculations. The accuracy of predicting thiolates at physiological pH is about 81% with the replica-exchange CpHMD, in contrast to the accuracy below 50% with the PB or empirical methods.⁵²

Motivated by the importance of the N-cap Cys in TCKI design and intrigued by the contradiction between the N-cap hypothesis and the high pK_a 's predicted by the empirical method, here we employed the asynchronous replica-exchange CpHMD to determine the pK_a 's of all 11 human kinases that possess a FP N-cap Cys. We found that most N-cap Cys have pK_a 's in the range of 7.7–8.5, thus providing nucleophilic thiolates for the direct thiol-Michael addition with electrophiles at physiological pH; however, surprisingly, the N-cap Cys in EGFR/ERBB2/ERBB4 and BLK remain in the thiol form up to pH 10.5. Analysis of the pH-dependent conformational environment surrounding the N-cap Cys provides insights into the varied reactivities and support to the general base assisted mechanism for the Michael addition with the N-cap Cys in EGFR/ERBB2/ERBB4 and BLK. To further explore the N-cap hypothesis, we also determined the pK_a 's of the N-cap+2 Cys in JNK1/JNK2/JNK3 and CASK. Additionally, our study identified reactive Cys and Lys locations in all 15 kinases, offering new opportunities for TCKI discovery. Together, our findings offer a systematic understanding of the Cys structure-reactivity relationship and demonstrate the utility of CpHMD simulations in aiding the rational design of Cys- and Lys-targeted covalent kinase inhibitors.

RESULTS AND DISCUSSION

CpHMD titration showed varied reactivities for the FP N-cap Cys. We performed the GPU-accelerated replica-exchange CpHMD titration simulations^{47,48,51} to determine the Cys and Lys pK_a values for the aforementioned 11 human kinases possessing a FP N-cap Cys, which is located in the linker at the N-terminal end of the α D helix (position 52 of the ATP binding site sequence according to KLIFS database⁵³). Except for TEC, TXK, and BLK, for which homology models were built using SWISS-MODEL,²⁰ simulations started from the X-ray crystal structures with inhibitors, ions, and crystal water molecules removed. A pH range 6.0–10.5 was used, and each replica lasted 30–50 ns until convergence (for more details see Methods). Considering the inverse relationship between the thiol reactivity and pK_a ,^{40,54} and to facilitate discussion, we group the Cys (or later Lys) in four categories (Table 2): unreactive ($pK_a > 9.4$), somewhat reactive (pK_a 8.4–9.4), reactive (pK_a 7.4–8.4), and hyper-reactive $pK_a < 7.4$. Accordingly, the fraction of thiolate (or neutral lysine) state at physiological pH is $< 1\%$ (unreactive),

1–10% (somewhat reactive), 10–50% (reactive), and $> 50\%$ (hyper-reactive).

Table 2. Reactivity definitions used in this work

Cys/Lys pK_a	Deprot. fraction	Reactivity
< 7.4	$> 50\%$	hyper reactive
7.4–8.4	10–50%	reactive
8.5–9.4	1–10%	somewhat reactive
> 9.4	$< 1\%$	unreactive

Our simulations gave a wide pK_a range from 7.6 to above 10.5 for the FP N-cap Cys, suggesting that its reactivity varies despite the same N-cap location. Among all, the N-cap Cys in BTK/BMX of the Tec family and JAK3 of the JakA family are nearly hyper-reactive, with the calculated pK_a 's around 7.7 (Fig. 2 and Table S1). The N-cap Cys in TEC/TXK/ITK of the Tec family are also reactive, with the calculated pK_a 's around 8.4 (Fig. 2 and Table S1), close to the value (8.6) for the model alanine penta-peptide (AACAA) in solution.⁴⁵ The N-cap Cys in MKK7 is somewhat reactive with a calculated pK_a of 9.1 (Fig. 2 and Table S1). Interestingly, the N-cap Cys in EGFR/ERBB2/ERBB4 of the EGFR family and BLK of the Src family are predicted as unreactive with pK_a values near or above 10.5 (Fig. 2 and Table S1), in agreement with our previous work.⁵⁰ Noticeably, our calculated pK_a 's of 7.7 and 8.4 for the FP N-cap Cys in BTK and ITK are in quantitative agreement with the experimental values of 7.7 and 8.5, respectively.⁵⁵ However, curiously, our simulation showed that EGFR's N-cap Cys remained protonated in the pH range 7.5–10.5, in disagreement with the pK_a of 5.5 estimated based on the Bromobimane fluorescence labeling experiment.⁵⁶ We will come back to the discussion.

Figure 2. Predicted pK_a 's and reactivities for the FP N-cap and N-cap+2 Cys in 15 kinases. The reactivity definitions are given in Table 2. The exact pK_a values for the N-cap Cys in EGFR/ERBB2/ERBB4 and BLK are not given, as the thiol form was dominant even at the highest simulation pH 10.5. The kinase names are shown in different colors based on the environment of the FP N-cap Cys: blue (N-cap+3 Asn), magenta (N-cap+3/2 Asp/Glu and N-cap+2/3 Ser/Arg/Lys), and green (N-cap+3 Asp/Glu). The kinases with a FP N-cap+2 Cys are shown in brown.

Sequence alignment of the 11 kinases showed that the FP N-cap Cys is often followed by an Asn or Asp residue at N-cap+3 position (Fig. 1 b and c). A previous experimental study suggested that the N-cap+3 residue may form hydrogen bond (h-bond) or electrostatic interaction with the N-cap residue.³⁸ Our previous computational work^{50,52} demonstrated that the pK_a of a Cys can be significantly shifted

by h-bonding with proximal residues. Experimental work on various proteins also suggested the role of local h-bonding and electrostatics in stabilizing Cys thiolates.^{39,43,57} Thus, to understand the varied reactivities of the FP N-cap Cys, we binned the 11 kinases into three groups according to the nearby residues that provide stabilization or destabilization of the N-cap thiolate: Asn at N-cap+3 (BTK/BMX/TEC/TXK); Asp/Glu at N-cap+3/2 and Ser/Arg/Lys at N-cap+2/3 positions (ITK, JAK3, and MKK7); and Asp/Glu at N-cap+3 without other nearby stabilizing h-bond or electrostatic partners (EGFR/ERBB2/ERBB4 and BLK). Below we present analysis of the pH-dependent h-bond and electrostatic interactions to elucidate the FP N-cap Cys reactivities.

BTK/BMX/TEC/TXK: FP N-cap thiolate stabilized by the N-cap+3 Asn and other h-bond donors. BTK/BMX/TEC/TXK of the Tec family contain an Asn residue at the N-cap+3 position relative to the FP N-cap Cys. The calculated pK_a 's of the N-cap Cys in BTK (C481) and BMX (C496) are 7.7 and 7.8, respectively, indicating that they are nearly hyper-reactive according to our definition (Table 2). Trajectory analysis shows that when deprotonated, Cys481 in BTK accepts h-bonds from both the side chain and backbone amide of Asn484 at N-cap+3 as well as the hydroxyl group and backbone amide of Thr410 (Fig. 3a). The pH-dependent occupancy of the h-bond formation and the pH-dependent deprotonation of Cys481 are perfectly correlated, suggesting that the h-bond interactions contribute to the pK_a downshift of Cys481 (Fig. 3b), consistent with our previous findings that thiolates tend to be stabilized by nearby h-bond donors.^{50,52} Similarly, the thiolate form of Cys496 in BMX is stabilized by h-bonding with the sidechain and backbone of Asn499 at N-cap+3 as well as the salt-bridge interaction with Arg540 (Fig. S1).

Figure 3. The FP N-cap thiolate is stabilized by hydrogen bonding with N-cap+3 Asn and other residues in BTK. (a) A zoomed-in view of the structural environment of Cys481 in BTK with the Asn484–Cys481 and Thr410–Cys481 h-bonds shown. A snapshot from simulation at pH 9.0 is used. (b) Deprotonated fraction of C481 at different pH. Curve represents the best fit to the Henderson-Hasselbalch equation. (c) Occupancy of the h-bond formation between Cys481 thiolate and Asn484 (sidechain and backbone) or Thr410 (sidechain and backbone) at different pH. A h-bond is defined by the heavy-atom donor-acceptor distance of 3.5 Å and the donor-hydrogen-acceptor angle of 150°.

The calculated pK_a 's of the FP N-cap Cys in TEC (Cys449) and TXK (Cys350) are about 8.5, suggesting that they are

somewhat reactive at physiological pH. In TEC, Cys449 thiolate accepts h-bonds from Asn452 at N-cap+3 and Ser378, as well as a salt bridge with Arg493 positioned at N-cap+44. Cys350 in TXK forms h-bonds with Asn353 at N-cap+3 and salt-bridge with Arg394 at N-cap+44. It is interesting to notice that the N-cap+44 Arg which interacts with the N-cap thiolate is the second Arg of the sequence HRDLAARN on the catalytic (C) loop which is highly conserved among protein tyrosine kinases.⁵⁸ It is also noteworthy that the N-cap Cys does not form h-bond or salt-bridge with i-2 or i-3 residues such as Asn, Ser, Glu, and Arg (Fig. 1c).

ITK, JAK3, and MKK7: FP N-cap thiolate destabilized by N-cap+3/2 Asp/Glu but stabilized by N-cap+2/3 Ser/Arg/Lys or N-cap+1 Lys/Thr. Several of the N-cap Cys are in close proximity to an anionic side chain (Asp or Glu), which would increase the Cys pK_a due to electrostatic repulsion. However, h-bonding and salt-bridge formation with the nearby residues can stabilize the thiolate form and thereby decreasing the N-cap Cys pK_a . These opposing effects are experienced by the N-cap Cys in ITK (Cys442), JAK3 (Cys909), and MKK7 (Cys218), resulting in the calculated pK_a 's of 8.4, 7.8, and 9.1, respectively.

Simulations of ITK showed that the deprotonated Cys442 can accept h-bonds from the side chain hydroxyl and backbone amide of Ser444 at N-cap+2 and the backbone amide of Asp445 at N-cap+3, as well as form a salt-bridge with Arg486 positioned at N-cap+44 (Fig. 4a). The excellent correlation between the increasing occupancies of h-bond (or salt bridge) formation and the increasing degree of Cys deprotonation suggests that these interactions stabilize the thiolate form, which compensates for the electrostatic repulsion from the nearby carboxylate group of Asp445 (Fig. 4b and c). As a result, the calculated pK_a of Cys442 is 8.4, indicating that it is somewhat reactive towards electrophiles.

Figure 4. Stabilizing interactions for the FP N-cap Cys thiolate in ITK. (a) Zoomed-in view of the structural environment of Cys442 in ITK taken from the simulation at pH 9.5. (b) pH-dependent deprotonated fraction of the FP Cys442. Curve represents the best fit to the Henderson-Hasselbalch equation. (c) pH-dependent occupancies of the h-bond formation between Cys442 thiolate and the sidechain and backbone of Ser444 (purple), backbone of Asp445 (yellow), and the side chain of Arg486 (light pink).

The compensating interactions are somewhat different for JAK3 and MKK7, in which an attractive electrostatic interaction with a nearby cationic sidechain is also a major contrib-

utor. In JAK3, the N-cap Cys909 thiolate is stabilized by salt-bridge interactions with Arg953 at N-cap+44 and h-bonding with Tyr904 at N-cap-5 at pH below pH 8 (Fig. S2). As a result, despite the repulsion with Asp912 at N-cap+3, the pK_a of Cys909 is downshifted to 7.8, lower than the corresponding N-cap Cys in ITK.

In MKK7, the sidechain-to-sidechain salt-bridge and sidechain-to-backbone h-bond interactions with Lys221 at N-cap+3 as well as the h-bond formation with Ser144 are the stabilizing forces for the thiolate state of the N-cap Cys218 (Fig. S2); these forces compensate for the destabilization due to Glu220 at N-cap+2 position. Overall, the number of stabilizing h-bond/salt bridge interactions with the N-cap Cys in MKK7 is less than those in ITK or JAK3, which may explain why the N-cap Cys pK_a in MKK7 (9.1) is the highest among the three kinases.

EGFR/ERBB2/ERBB4 and BLK: FP N-cap thiolate destabilized by N-cap+3 Asp/Glu and desolvation without compensating interactions.

The N-cap Cys is buried in EGFR/ERBB2/ERBB4 (of the EGFR family) and BLK (of the Src family) to a similar extent as in the aforementioned kinases; however, there is no nearby h-bond donor or attractive electrostatic interaction partner. Furthermore, the N-cap+3 position is occupied by an anionic Asp (or Glu in ERBB4) which can potentially destabilize the thiolate form of the N-cap Cys. Thus, it is not surprising that the calculated pK_a 's are above 10.5 (the highest simulation pH), suggesting that the N-cap Cys in these kinases are unreactive towards electrophiles.

Curiously, the N-cap Cys in all four kinases have been targeted by the FDA-approved covalent drugs. Most notably, Cys797 in EGFR was the target of the first TKI,⁵ and it was subsequently targeted by numerous investigational and FDA-approved inhibitors including ibrutinib, afatinib, and neratinib (Table 1). Analysis of the X-ray structure of EGFR (e.g., PDB ID 4zjv) reveals that the Cys797 thiol may donate a h-bond to the carboxylate group of the N-cap+3 Asp800 (Fig. 5). Similarly, the X-ray structures of ERBB2/ERBB4 and BLK show that the N-cap thiol sulfur is within the h-bond distance to the carboxylate oxygens of N-cap+3 Asp.

Figure 5. The FP N-cap Cys thiol can donate a h-bond to the N-cap+3 aspartate in EGFR. Zoomed-in view of the structural environment of the N-cap Cys797 in EGFR from the X-ray crystal structure (PDB: 4zjv³²). A potential h-bond from Cys797 thiol to the carboxylate of Asp800 is indicated as a dashed line.

The high (calculated) pK_a of the N-cap and the potential formation of a thiol–aspartate h-bond in the holo structure together suggest that the Michael addition reactions with EGFR/ERBB2/ERBB4 and BLK may proceed via a base-assisted mechanism, in which the N-cap+3 Asp serves as a general base to deprotonate the N-cap Cys in the first stage

of the reaction. Indeed, a recent combined quantum mechanical (QM)/MM simulation of EGFR⁵⁹ demonstrated that in the proximity of a covalent inhibitor carrying an acrylamide warhead, the proton on the N-cap Cys797 thiol can be transferred to the carboxylate of the N-cap+3 Asp, resulting in a stable thiolate/Asp-COOH pair, from which the thiolate then attacks the $C\beta$ of acrylamide generating a carbanion before being reprotonated by the N-cap+3 Asp-COOH to result in the thiol adduct.⁵⁹ Our result is consistent with this base-assisted mechanism. We should also note that in our apo simulations of the EGFR/ERBB2/ERBB4 and BLK, the thiol group of the N-cap Cys did not form a h-bond with the N-cap+3 Asp/Glu, supporting the hypothesis that the proton transfer only occurs in the presence of an electrophilic warhead.⁵⁹ Another possibility for the covalent modification of an unreactive Cys is to make use of a base group adjacent to the Michael acceptor warhead. The structure-activity data of EGFR inhibitors suggested that a basic amine, e.g., dialkylamino or pyrrolidine group can abstract a proton from the Cys thiol and thereby catalyzing the reaction.^{30,60,61}

Other reactive Cys and Lys locations. Since in the CpHMD simulations all Cys and Lys were allowed to titrate simultaneously, we were able to determine their pK_a values. Table 3 lists the locations of all reactive Cys and Lys predicted by our simulations. Interestingly, a Cys on $\beta 4$ sheet (i-17/18 with respect to the N-cap Cys) exists in 10 kinases and is found reactive or hyper-reactive in BMX/TEC/TXK of the Tec family and ERBB2/EGFR of the EGFR family. The positions of other reactive Cys are proximal to the ATP binding site and have already been targeted by clinical and investigational TKIs. These locations include the P loop (Cys147 of MKK7), DFG-1 (Cys276 of MKK7), and the A loop (Cys891 of ERBB4).

The local conformational environment of the N-cap Cys is not affected by the conformation of the DFG-motif (in or out) or αC helix (in or out). As such, the calculated pK_a is insensitive to the variation of the kinase conformation. For example, using five X-ray structures (in different DFG/ αC conformations according to KLIFS⁵³), the calculated pK_a of the N-cap Cys481 is 7.8–8.4 (7.8 with PDB 3pj3, 8.3 with PDB 6o8i, 8.4 with PDB 3piy, 8.3 with PDB 5p9j, and 7.9 with PDB 3oct). In contrast, although the roof catalytic Lys has a pK_a above 10.4 in the DFG-in or αC -in states, the pK_a value can be significantly downshifted when the kinase adopts the DFG-out/ αC -out conformation⁵⁰. Our simulations predicted that the roof Lys in BTK (Lys430, PDB 3pj3),²⁵ BMX (Lys445, PDB 3sxs),⁸ and JAK3 (Lys855, PDB 5lwm)⁶² are reactive. Interestingly, although the X-ray structures for these kinases show a DFG out state, αC -Glu in BTK and JAK3 is within a salt-bridge distance to the roof Lys. During the simulations, however, the Glu–Lys salt-bridge disrupted, resulting in the sampling of the αC -Glu out conformations. The facile transition between the αC -Glu in and out positions is consistent with our previous explicit-solvent simulations of the c-Src kinase.⁶³ The Lys at the DFG-3 position has not been targeted before but it is also in the ATP binding site, and the simulations predicted that the DFG-3 Lys in ERBB2 and ERBB4 are reactive, thus providing a new opportunity for covalent targeting. We note that the P-loop and DFG-1 Cys as well as the roof Lys have been previously targeted in other kinases.³

Table 3. Summary of reactive Cys and Lys in the 15 kinases discussed in this work^a

Kinase	Family	Cys	Location	Reactivity	Major contributors	Lys	Location	Reactivity
FP N-cap Cys								
BTK	Tec	C481	front pocket	reactive	N(+3), T(-71)	K430	roof	reactive
BMX	Tec	C479	β 4 sheet	reactive		K445	roof	somewhat
		C496	front pocket	reactive	N(+3), R(+44)			
		C549	β 8 sheet	reactive				
TEC	Tec	C406	α C helix	reactive				
		C432	β 4 sheet	reactive				
		C449	front pocket	reactive	S(-71), N(+3), R(+44)			
TXK	Tec	C333	β 4 sheet	hyper				
		C350	front pocket	somewhat	N(+3), R(+44)			
		C402	β 8 sheet	somewhat				
ITK	Tec	C442	front pocket	reactive	S(+2), D(+3), R(+44)			
JAK3	JakA	C909	front pocket	reactive	Y(-5), R(+2), R(+44)	K855	roof	hyper
		C1028	α G helix	reactive				
MKK7	STE7	C147	P loop	hyper				
		C218	front pocket	reactive	S(-74), K(+3)			
		C276	DFG-1	reactive				
		C341	loop (α F/ α G)	hyper				
EGFR	EGFR	C775	loop (α C/ β 4)	reactive				
		C781	β 4 sheet	reactive				
ERBB2	EGFR	C789	β 4 sheet	hyper		K860	DFG-3	reactive
ERBB4	EGFR	C891	A loop	reactive		K858	DFG-3	reactive
BLK	Src	C460	loop (α G/ α H)	somewhat				
FP N-cap+2 Cys								
JNK1	MAPK	C116	front pocket	reactive	Q(+1), S(+39)			
		C245		reactive	α H helix			
JNK2	MAPK	C116	front pocket	reactive	Q(+1), S(+39)	K153	β 7 sheet	somewhat
JNK3	MAPK	C154	front pocket	hyper	S(-82), Q(+1), S(+39)			
		C283		reactive	loop (α G/ α H)			
CASK	CASK	C100	front pocket	somewhat				

^aThe reactivity levels are defined in Table 2. Residues or locations that have been previously targeted by covalent inhibitors are labeled in bold font. Residues that make significant contributions to the pK_a shift of the N-cap Cys are listed with the sequence positions (relative to N-cap) given in the parenthesis. The calculated pK_a 's of all listed residues are given in Table S1.

Comparison to other pK_a calculations. The CpHMD predicted pK_a 's are in stark contrast to the structure-based calculations based on the empirical PROPKA approach and on the solutions to the PB equation. Most N-cap Cys pK_a 's predicted by the PB calculation (based on the H++ server) are between 9.0 and 11.4, similar to those from PROPKA (Table S2). Most notably, while CpHMD simulations gave the pK_a 's of about 7.7–7.8 for the N-cap Cys in BTK, BMX, and JAK3 (the lowest among the 11 kinases, Table S1), the PROPKA calculated pK_a 's are 9.6, 9.9, 10.3, and 9.0, and those from PB are 9.6, 9.9, 8.8, and 9.0, respectively (Table S2). The large discrepancy, which has been noted in our previous benchmark studies,^{48,50} can be attributed to the h-bonds and electrostatic interactions that are absent in the X-ray structures but emerge from the pH-dependent conformational dynamics captured by CpHMD and not by structure-based calculations.⁶⁴ For example, in BTK, the sidechain hydroxyl and backbone amide of Thr410 donate h-bonds to stabilize the N-cap Cys481 in a pH-dependent manner (Fig. 3); however, these interactions are absent in the X-ray structure and therefore not accounted for by the PROPKA or PB calculations. Similarly, in ITK, the stabilizing interactions with Ser444 and Arg486 (Fig. 4) are absent in the X-ray structure and therefore the pK_a 's from the PROPKA and PB calculations are over 10.4, two units higher than the value of 8.4 from the CpHMD

titration.

The TI simulations and alternative constant pH methods based on the hybrid-solvent Monte-Carlo (MC)/MD and non-equilibrium MD/MC schemes of Rowley and coworker⁶⁵ gave the pK_a 's of 10.4–13.0 for the N-cap Cys in BTK/BMX/ITK, JAK3, and 11.1–13.5 for EGFR (Table S2). Except for EGFR, these results are in disagreement with our CpHMD simulations which suggested that all four N-cap Cys are reactive with the pK_a 's of 7.7–8.4. With regards to the N-cap Cys797 in EGFR, Rowley and coworker calculated its pK_a with either protonated or deprotonated Asp800. Interestingly, the calculated pK_a was above 10 in both cases.⁶⁵

The FP N-cap+2 Cys in JNKs and CASK. JNK1/JNK2/JNK3 of the MAPK family and CASK of the CASK family possess a Cys similar in position as the N-cap Cys discussed so far. This N-cap+2 Cys, Cys116 in JNK1/JNK2, Cys154 in JNK3, or Cys100 in CASK is the second N-terminal residue in the α D helix, and has an Ile in the N-cap+3 position instead of Asn/Asp/Glu. The calculated pK_a 's for Cys116 in JNK1/JNK2 and Cys154 in JNK3 are 7.5, 8.0, and 6.3 respectively, while that for Cys100 in CASK is 9.0 (Table S2). Analysis showed that in JNK1 and JNK2, Cys116 thiolate can accept h-bond formation with Gln117 and Ser155. In JNK3, Cys154 thi-

olate is additionally stabilized by the h-bond formation with Ser72, which may explain the lower pK_a (hyper-reactivity) of Cys154 compared to Cys116 in JNK1/JNK2. As to CASK, the deprotonation of Cys100 is mainly correlated with the salt-bridge formation with Arg302 and to a lesser extent with the h-bond formation with Asn299. The number of h-bonds formed by Cys100 in CASK is lower than that by Cys116 in JNK1/JNK2 or Cys154 in JNK3. As a result, the pK_a upshift due to desolvation is not offset, leading to a higher pK_a value of Cys100 as compared to Cys116 or Cys154 in JNKs.

CONCLUDING REMARKS

N-cap Cys has long been hypothesized as hyper-reactive due to the thiolate-helix dipole interactions.^{38,39} In support of this hypothesis, the N-cap Cys in all human kinases have been covalently targeted. Our data demonstrated that the N-cap position alone may not stabilize the thiolate form making the Cys reactive at physiological pH. CpHMD simulations confirmed the reactivities of the N-cap Cys in BTK/BMX/TEC/TXK/ITK and JAK3, with the calculated pK_a 's ranging from 7.7 to 8.5. The calculated pK_a 's of 7.7 and 8.4 for BTK and ITK are in excellent agreement with the respective experimental values of 7.7 and 8.5.⁵⁵ In comparison, the calculated pK_a is 9.1 for the N-cap Cys in MKK7, suggesting a lower reactivity compared to the aforementioned Tec family kinases and JAK3.

Surprisingly, the calculated pK_a 's are above 10.4 for the N-cap Cys in EGFR/ERBB2/ERBB4 and BLK, suggesting that the thiol form is predominant at physiological pH despite the N-cap position. Curiously, the calculated pK_a for EGFR is 5 units higher than that from the pH-dependent bromobimane-thiol reaction rate. This discrepancy cannot be explained by the lack of polarization of additive force field based CpHMD simulations, which may lead to a small underestimation of the helix dipole⁶⁶ and consequently the pK_a downshift due to the thiolate-dipole interaction. We suggest that the thiol addition reaction of EGFR and ERBB2/ERBB4 as well as BLK may be initiated from a thiol state by following a base-assisted mechanism. EGFR/ERBB2/ERBB4 and BLK all contain an Asp in the N-cap+3 position relative to the N-cap Cys, and in the holo crystal structures, this Asp is within h-bond distance to the Cys. A recent QM/MM study suggested that in the proximity of an inhibitor, a Cys...Asp h-bonded pair may form, from which the thiolate then attacks the acrylamide warhead on the inhibitor.⁵⁹ Thus, it is possible that the Cys797 pK_a determined from the bromobimane-thiol reaction.⁵⁶ refers to the Cys797...Asp800 pair in the presence of bromobimane.

Other base-assisted mechanisms are also possible for thiol-Michael addition reactions. Several experimental studies of the structure-activity relationships of EGFR inhibitors suggested that an amine group proximal to the electrophilic warhead may act as a general base to abstract the proton from the N-cap thiol yielding the nucleophilic thiolate.^{30,60,61} Additionally, a nearby water may in principle serve as a general base. Finally, in the absence of a general base, a direct addition mechanism is conceivable, whereby in the first step the carbonyl oxygen abstracts the proton from the N-cap thiol, as proposed by a most recent QM/MM study of BTK modification by ibrutinib assuming the N-cap Cys is protonated.⁶⁷

A corollary to the N-cap thiolate hypothesis is that a Cys placed further down the helical sequence, e.g., at N2 or N3 position, should have a higher pK_a . The CpHMD simulations of JNK1/JNK2/JNK3 and CASK showed that this needs not be the case. The calculated pK_a 's of the N-cap+2 Cys in

JNKs range from 6.3 to 8.0, consistent with the fact that they have all been covalently modified.^{17,18} The calculated pK_a of N-cap+2 Cys in CASK indicates it is somewhat reactive; no covalent inhibitor targeting this position has been reported yet.

Our analysis showed that while the solvent accessibility of the N-cap or N-cap+2 Cys in various kinases is very similar (all largely buried), the h-bond and electrostatic environment can be very different, which gives rise to the varied Cys pK_a 's. Specifically, interaction with nearby h-bond donors is a major determinant for the pK_a downshift, while solvent exclusion is a major determinant of the pK_a upshift, consistent with our previous findings in the context of other proteins.^{50,52} Notably, the presence of a h-bond donor at the N-cap+3 position (e.g., Asn) to the stabilizes the thiolate form, e.g., in BTK/BMX/TXK. By contrast, the presence of an anionic sidechain (e.g. Asp or Glu) at the N-cap+3 position destabilizes the thiolate form. Although ITK, JAK3, BLK, and EGFR/ERBB2/ERBB4 all have an N-cap+3 Asp next to the N-cap Cys, only ITK and JAK3 have favorable h-bond interactions that stabilize the thiolate form. As a result, the N-cap Cys in ITK and JAK3 are reactive, whereas those in BLK and EGFR/ERBB2/ERBB4 are unreactive.

A unique capability of CpHMD simulations is to capture pH-dependent formation of h-bond and electrostatic interactions that are nascent or absent in the static (X-ray) structure but can nonetheless significantly impact the protonation state. Such interactions with the N-cap or N-cap+2 Cys cannot be accounted for by traditional structure-based PB or empirical calculations; the predicted high pK_a 's are dominated by the large contribution from the desolvation penalty.

The main caveats of the current CpHMD approach include the neglect of polarization, which may lead to a small underestimation of the Cys pK_a downshift due to underestimation of the helix dipole, as well as the neglect of proton transfer between Cys and a nearby general base. QM studies will be conducted in the future to further test the hypothesis that the base-assisted mechanisms underlie the Michael addition reactions with the N-cap Cys in EGFR/ERBB2/ERBB4 and BLK. Taken together, our work offers a systematic understanding of cysteine reactivities in kinases and demonstrates the utility of GPU-accelerated CpHMD simulations in making accurate predictions and rationalization of Cys reactivities to guide the design of targeted covalent inhibitors.

METHODS AND PROTOCOLS

Structure Preparation. Except for TEC, TXK, and BLK, the initial structures were taken from the Protein Data Bank (PDB): BTK (3pj3),²⁵ BMX (3sxs),⁸ ITK (3miy),²⁴ JAK3 (5lwm),⁶² EGFR (4zjv),³² ERBB2 (3pp0),³¹ ERBB4 (2r4b),³⁰ JNK1 (2xrw),⁶⁸ JNK2 (3e7o),⁶⁹ JNK3 (6emh),⁷⁰ and CASK (3mfr).⁷¹ The structures of TEC, TXK, and BLK were built using SWISS-MODEL.²⁰ For each structure, the open source tool pdbfixer from the OpenMM package⁷² was used to remove non-protein heavy atoms and to add missing residues, atoms, acetylated N terminus and amidated C terminus, disulfide bonds (if present), and hydrogen atoms. Lys and Cys had one dummy hydrogen, while His had two dummy hydrogens.

Simulation Protocol. The generalized Born based CpHMD simulations^{47,48} were performed using the pmemd

engine of AMBER18 (AMBER 2018).⁴⁶ The pH replica-exchange titration simulations were performed on the GPUs using our asynchronous replica-exchange implementation.⁵¹ The proteins were represented by the ff14sb protein force field,⁷³ and solvent represented by the GBNeck2 (igb=8) implicit solvent model⁴⁹ with mbondi3 intrinsic Born radii and 0.15 M ionic strength. Modifications to the intrinsic Born radii were made for His and Cys sidechains, as discussed in our previous work.^{47,48,50} Prior to the CpHMD runs, energy minimization was performed in the GB solvent using 500 steps of steepest descent and 500 steps of the conjugate gradient method with a harmonic force constant of 50 kcal/mol/Å² applied to the heavy atoms. Following minimization, four stages of equilibration were performed at 300 K using 2000 steps for each stage. A harmonic force constant applied to the heavy atoms was gradually decreased from 5.0, 2.0, 1.0, to 0 kcal/mol/Å².

One set of pH replica-exchange CpHMD simulations was carried out for each protein. For BTK, ITK, and ERBB4, 8 independent pH titrations were also performed at pH conditions from 6.0 to 9.5, with a 0.5 unit interval. Each pH condition was run for 10 ns. These simulations gave very similar results as the replica-exchange runs. The pH conditions in the replica-exchange simulations were from 7.0 to 10.5, except BTK which had a pH range of 6.0–9.0 and MKK7 which had a pH range of 7.0–11.0. The pH conditions were placed at every 0.5 unit, except for MKK7 which had an extra pH condition of 9.25. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm allowing a 2-fs timestep. Exchanges between adjacent pH replicas were attempted every 1000 steps (2 ps) according to the Metropolis criterion.⁷⁴ Each replica lasted 50 ns (aggregate time of 400 ns) for BTK, BMX, TXK, ITK, JAK3, MKK7, and 30 ns (aggregate time of 240 ns) for the other kinases (these proteins had faster convergence). Each replica underwent Langevin dynamics at 300 K with an effectively infinite cutoff (999 Å) for nonbonded interactions. All side chains of His, Cys, and Lys were allowed to titrate, with the experimental model alanine penta-peptide pK_a's of 6.5, 8.5, and 10.4, respectively.⁴⁵ If not stated, all settings were identical to our previous work.^{47,50,52}

pK_a calculations. The unprotonated fraction (*S*) of a titratable residue at different pH values was calculated using the definitions of the protonated ($\lambda < 0.2$) and deprotonated ($\lambda > 0.8$) states, as in our previous work.^{47,48,52} The pK_a was calculated by fitting the *S* values to the generalized Henderson Hasselbalch (HH) equation, $S = 1/(1 + 10^{n(pK_a - pH)})$, where *n* is the Hill coefficient.

The bootstrap method was used to estimate the errors of the calculated pK_a's.⁴⁸ The *S* values based on the 10-ns simulations at different pH were combined to generate *mⁿ* sets of titration data, where *m* is the number of the 10 ns simulations per pH and *n* is the number of pH conditions. For each set of titration data (*S* vs pH), a pK_a value was calculated. An average and standard deviation of all the calculated pK_a's are reported (Fig. 2 and Table S1).

Supporting Information Available

Supplemental tables and figures are included.

ACKNOWLEDGEMENT. The authors acknowledge National Institutes of Health (R01GM098818 and R01CA256557) for funding.

References

- (1) Ferguson, F. M.; Gray, N. S. Kinase Inhibitors: The Road Ahead. *Nat. Rev. Drug Discov.* **2018**, *17*, 353–377.
- (2) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A Comprehensive Map of Molecular Drug Targets. *Nat. Rev. Drug Discov.* **2017**, *16*, 19–34.
- (3) Zhao, Z.; Bourne, P. E. Progress with Covalent Small-Molecule Kinase Inhibitors. *Drug Discov. Today* **2018**, *23*, 727–735.
- (4) Abdeldayem, A.; Raouf, Y. S.; Constantinescu, S. N.; Morigli, R.; Gunning, P. T. Advances in Covalent Kinase Inhibitors. *Chem. Soc. Rev.* **2020**, *49*, 2617–2687.
- (5) Liu, Q.; Sabnis, Y.; Zhao, Z.; Zhang, T.; Buhrlage, S. J.; Jones, L. H.; Gray, N. S. Developing Irreversible Inhibitors of the Protein Kinase Cysteine. *Chem. Biol.* **2013**, *20*, 146–159.
- (6) Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discovery Today* **2015**, *20*, 1061–1073.
- (7) Eid, S.; Turk, S.; Volkamer, A.; Rippmann, F.; Fulle, S. KinMap: A Web-Based Tool for Interactive Navigation through Human Kinome Data. *BMC Bioinformatics* **2017**, *18*, 16.
- (8) Muckelbauer, J.; Sack, J. S.; Ahmed, N.; Burke, J.; Chang, C. Y.; Gao, M.; Tino, J.; Xie, D.; Tebben, A. J. X-Ray Crystal Structure of Bone Marrow Kinase in the X Chromosome: A Tec Family Kinase. *Chemical Biology & Drug Design* **2011**, *78*, 739–748.
- (9) Dhillon, S. Tirabrutinib: First Approval. *Drugs* **2020**, *80*, 835–840.
- (10) Dhillon, S. Orelabrutinib: First Approval. *Drugs* **2021**, *81*, 503–507.
- (11) Robinson, M. F.; Damjanov, N.; Stamenkovic, B.; Radunovic, G.; Kivitz, A.; Cox, L.; Manukyan, Z.; Banfield, C.; Saunders, M.; Chandra, D.; Vincent, M. S.; Mancuso, J.; Peeva, E.; Beebe, J. S. Efficacy and Safety of PF-06651600 (Ritlecitinib), a Novel JAK3/TEC Inhibitor, in Patients With Moderate-to-Severe Rheumatoid Arthritis and an Inadequate Response to Methotrexate. *Arthritis Rheumatol.* **2020**, *72*, 1621–1631.
- (12) Kim, E. S. Olmutinib: First Global Approval. *Drugs* **2016**, *76*, 1153–1157.
- (13) Blair, H. A. Pyrotinib: First Global Approval. *Drugs* **2018**, *78*, 1751–1755.
- (14) Gonzalez, F. et al. Mobocertinib (TAK-788): A Targeted Inhibitor of EGFR Exon 20 Insertion Mutants in Non-Small Cell Lung Cancer. *Cancer Discov.* **2021**.
- (15) Wolle, P.; Engel, J.; Smith, S.; Goebel, L.; Hennes, E.; Lategahn, J.; Rauh, D. Characterization of Covalent Pyrazolopyrimidine-MKK7 Complexes and a Report on a Unique DFG-in/Leu-in Conformation of Mitogen-Activated Protein Kinase Kinase 7 (MKK7). *J. Med. Chem.* **2019**, *62*, 5541–5546.
- (16) Schröder, M.; Tan, L.; Wang, J.; Liang, Y.; Gray, N. S.; Knapp, S.; Chaikuad, A. Catalytic Domain Plasticity of MKK7 Reveals Structural Mechanisms of Allosteric Activation and Diverse Targeting Opportunities. *Cell Chem. Biol.* **2020**, *27*, 1285–1295.e4.
- (17) Zhang, T. et al. Discovery of Potent and Selective Covalent Inhibitors of JNK. *Chem. Biol.* **2012**, *19*, 140–154.
- (18) Muth, F.; El-Gokha, A.; Ansideri, F.; Eitel, M.; Do, E.; Sievers-Engler, A.; Lange, A.; Boeckler, F. M.; Lam, M.; Laufer, S. A. Tri- and Tetrasubstituted Pyridinylimidazoles as Covalent Inhibitors of C-Jun N-Terminal Kinase 3. *J. Med. Chem.* **2017**, *60*, 594–607.
- (19) Olsson, M. H. M.; Sondergaard, C. R.; Rosikowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (20) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.
- (21) Chaikuad, A.; Koch, P.; Laufer, S. A.; Knapp, S. The Cysteine of Protein Kinases as a Target in Drug Development. *Angewandte Chemie International Edition* **2018**, *57*, 4372–4385.
- (22) Berglöff, A.; Hamasy, A.; Meinke, S.; Palma, M.; Krstic, A.; Månsson, R.; Kimby, E.; Österborg, A.; Smith, C. I. E. Targets for Ibrutinib Beyond B Cell Malignancies. *Scand. J. Immunol.* **2015**, *82*, 208–217.
- (23) Schwartzberg, P. L.; Finkelstein, L. D.; Readinger, J. A. TEC-Family Kinases: Regulators of T-Helper-Cell Differentiation. *Nat. Rev. Immunol.* **2005**, *5*, 284–295.
- (24) Kutach, A. K.; Villaseñor, A. G.; Lam, D.; Belunis, C.; Janson, C.; Lok, S.; Hong, L.-N.; Liu, C.-M.; Deval, J.; Novak, T. J.; Barnett, J. W.; Chu, W.; Shaw, D.; Kuglstatter, A. Crystal Structures of IL-2-Inducible T Cell Kinase Complexed with Inhibitors: Insights into Rational Drug Design and Activity Regulation. *Chemical Biology & Drug Design* **2010**, *76*, 154–163.
- (25) Kuglstatter, A.; Wong, A.; Tsing, S.; Lee, S. W.; Lou, Y.; Villaseñor, A. G.; Bradshaw, J. M.; Shaw, D.; Barnett, J. W.; Browner, M. F. Insights into the Conformational Flexibility of Bruton's Tyrosine Kinase from Multiple Ligand Complex Structures. *Protein Science* **2011**, *20*, 428–436.
- (26) Petersen, D. L. et al. B-Lymphoid Tyrosine Kinase (Blk) Is an Onco-

- gene and a Potential Target for Therapy with Dasatinib in Cutaneous T-Cell Lymphoma (CTCL). *Leukemia* **2014**, *28*, 2109–2112.
- (27) Telliez, J.-B. et al. Discovery of a JAK3-Selective Inhibitor: Functional Differentiation of JAK3-Selective Inhibition over Pan-JAK or JAK1-Selective Inhibition. *ACS Chem. Biol.* **2016**, *11*, 3442–3451.
- (28) Fragoulis, G. E.; McInnes, I. B.; Siebert, S. JAK-Inhibitors. New Players in the Field of Immune-Mediated Diseases, beyond Rheumatoid Arthritis. *Rheumatology* **2019**, *58*, i43–i54.
- (29) Wieduwilt, M. J.; Moasser, M. M. The Epidermal Growth Factor Receptor Family: Biology Driving Targeted Therapeutics. *Cell Mol. Life Sci.* **2008**, *65*, 1566–1584.
- (30) 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.
- (31) Aertgeerts, K.; Skene, R.; Yano, J.; Sang, B.-C.; Zou, H.; Snell, G.; Jennings, A.; Iwamoto, K.; Habuka, N.; Hirokawa, A.; Ishikawa, T.; Tanaka, T.; Miki, H.; Ohta, Y.; Sogabe, S. Structural Analysis of the Mechanism of Inhibition and Allosteric Activation of the Kinase Domain of HER2 Protein. *J. Biol. Chem.* **2011**, *286*, 18756–18765.
- (32) Park, E.; Kim, N.; Ficarro, S. B.; Zhang, Y.; Lee, B. I.; Cho, A.; Kim, K.; Park, A. K. J.; Park, W.-Y.; Murray, B.; Meyerson, M.; Beroukhi, R.; Marto, J. A.; Cho, J. A.; Eck, M. J. Structure and Mechanism of Activity-Based Inhibition of the EGF Receptor by Mig6. *Nat. Struct. Mol. Biol.* **2015**, *22*, 703–711.
- (33) Huang, Y.-W. A.; Zhou, B.; Wernig, M.; Südhof, T. C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and A β Secretion. *Cell* **2017**, *168*, 427–441.e21.
- (34) Pan, Z. et al. Discovery of Selective Irreversible Inhibitors for Bruton's Tyrosine Kinase. *ChemMedChem* **2007**, *2*, 58–61.
- (35) Chen, J.; Kinoshita, T.; Sukbunthorn, J.; Chang, B. Y.; Elias, L. Ibrutinib Inhibits ERBB Receptor Tyrosine Kinases and HER2-Amplified Breast Cancer Cell Growth. *Mol. Cancer Ther.* **2016**, *15*, 2835–2844.
- (36) Hol, W. G. J. Effects of the α -Helix Dipole upon the Functioning and Structure of Proteins and Peptides. *Adv. Biophys.* **1985**, *19*, 133–165.
- (37) Serrano, L.; Fersht, A. R. Capping and A-Helix Stability. *Nature* **1989**, *342*, 296–299.
- (38) Anderson, T. A.; Sauer, R. T. Role of an Ncap Residue in Determining the Stability and Operator-Binding Affinity of Arc Repressor. *Biophys. Chem.* **2002**, *100*, 341–350.
- (39) Miranda, J. L. Position-Dependent Interactions between Cysteine Residues and the Helix Dipole. *Prot. Sci.* **2003**, *12*, 73–81.
- (40) Bulaj, G.; Kortemme, T.; Goldenberg, D. P. Ionization-Reactivity Relationships for Cysteine Thiols in Polypeptides [†]. *Biochemistry* **1998**, *37*, 8965–8972.
- (41) Peskin, A. V.; Winterbourn, C. C. Kinetics of the Reactions of Hypochlorous Acid and Amino Acid Chloramines with Thiols, Methionine, and Ascorbate. *Free Radical Biol. Med.* **2001**, *30*, 572–579.
- (42) Tosatto, S. C. E.; Bosello, V.; Fogolari, F.; Mauri, P.; Roveri, A.; Toppo, S.; Flohé, L.; Ursini, F.; Maiorino, M. The Catalytic Site of Glutathione Peroxidases. *Antioxid. Redox Signal.* **2008**, *10*, 1515–1526.
- (43) Ferrer-Sueta, G.; Manta, B.; Botti, H.; Radi, R.; Trujillo, M.; Denicola, A. Factors Affecting Protein Thiol Reactivity and Specificity in Peroxide Reduction. *Chem. Res. Toxicol.* **2011**, *24*, 434–450.
- (44) Awoonor-Williams, E.; Rowley, C. N. Evaluation of Methods for the Calculation of the pK_a of Cysteine Residues in Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 4662–4673.
- (45) Platzer, G.; Okon, M.; McIntosh, L. P. pH-Dependent Random Coil 1H, 13C, and 15N Chemical Shifts of the Ionizable Amino Acids: A Guide for Protein pK_a Measurements. *J. Biomol. NMR* **2014**, *60*, 109–129.
- (46) Case, D. A. et al. *AMBER 2018*; 2018.
- (47) 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.
- (48) Harris, R. C.; Shen, J. GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: pKa Predictions with Single-pH Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4821–4832.
- (49) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.
- (50) 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.
- (51) Henderson, J. A.; Verma, N.; Harris, R. C.; Liu, R.; Shen, J. Assessment of Proton-Coupled Conformational Dynamics of SARS and MERS Coronavirus Papain-like Proteases: Implication for Designing Broad-Spectrum Antiviral Inhibitors. *J. Chem. Phys.* **2020**, *153*, 115101.
- (52) Harris, R. C.; Liu, R.; Shen, J. Predicting Reactive Cysteines with Implicit-Solvent-Based Continuous Constant pH Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2020**, *16*, 3689–3698.
- (53) van Linden, O. P. J.; Kooistra, A. J.; Leurs, R.; de Esch, I. J. P.; de Graaf, C. KLIFS: A Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space. *J. Med. Chem.* **2014**, *57*, 249–277.
- (54) Winterbourn, C. C.; Metodiewa, D. Reactivity of Biologically Important Thiol Compounds with Superoxide and Hydrogen Peroxide. *Free Radi. Biol. Med.* **1999**, *27*, 322–328.
- (55) Zapf, C. W.; Gerstenberger, B. S.; Xing, L.; Limburg, D. C.; Anderson, D. R.; Caspers, N.; Han, S.; Aulabaugh, A.; Kurumbail, R.; Shakyia, S.; Li, X.; Spaulding, V.; Czerwinski, R. M.; Seth, N.; Medley, Q. G. Covalent Inhibitors of Interleukin-2 Inducible T Cell Kinase (Itk) with Nanomolar Potency in a Whole-Blood Assay. *J. Med. Chem.* **2012**, *55*, 10047–10063.
- (56) Truong, T. H.; Ung, P. M.-U.; Palde, P. B.; Paulsen, C. E.; Schlessinger, A.; Carroll, K. S. Molecular Basis for Redox Activation of Epidermal Growth Factor Receptor Kinase. *Cell Chem. Biol.* **2016**, *23*, 837–848.
- (57) Britto, P.; Knipling, L.; Wolff, J. The Local Electrostatic Environment Determines Cysteine Reactivity of Tubulin. *J. Biol. Chem.* **2002**, *277*, 29018–29027.
- (58) McTigue, M. A.; Wickersham, J. A.; Pinko, C.; Showalter, R. E.; Parast, C. V.; Tempczyk-Russell, A.; Gehring, M. R.; Mroczkowski, B.; Kan, C.-C.; Villafranca, J. E.; Appelt, K. Crystal Structure of the Kinase Domain of Human Vascular Endothelial Growth Factor Receptor 2: A Key Enzyme in Angiogenesis. *Structure* **1999**, *7*, 319–330.
- (59) 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.
- (60) 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.
- (61) 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.
- (62) Forster, M.; Chaikuad, A.; Bauer, S. M.; Holstein, J.; Robers, M. B.; Corona, C. R.; Gehringer, M.; Pfaffenrot, E.; Ghoreschi, K.; Knapp, S.; Laufer, S. A. Selective JAK3 Inhibitors with a Covalent Reversible Binding Mode Targeting a New Induced Fit Binding Pocket. *Cell Chem. Biol.* **2016**, *23*, 1335–1340.
- (63) Tsai, C.-C.; Yue, Z.; Shen, J. How Electrostatic Coupling Enables Conformational Plasticity in a Tyrosine Kinase. *J. Am. Chem. Soc.* **2019**, *141*, 15092–15101.
- (64) Huang, Y.; Yue, Z.; Tsai, C.-C.; Henderson, J. A.; Shen, J. Predicting Catalytic Proton Donors and Nucleophiles in Enzymes: How Adding Dynamics Helps Elucidate the Structure-Function Relationships. *J. Phys. Chem. Lett.* **2018**, *26*.
- (65) Awoonor-Williams, E.; Rowley, C. N. How Reactive Are Druggable Cysteines in Protein Kinases? *J. Chem. Inf. Model.* **2018**, *58*, 1935–1946.
- (66) Huang, J.; MacKerell, A. D. Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (AAQAA)₃ Peptide. *Biophys. J.* **2014**, *107*, 991–997.
- (67) 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.
- (68) Garai, A.; Zeke, A.; Gögl, G.; Törő, I.; Fördős, F.; Blankenburg, H.; Bárkai, T.; Varga, J.; Alexa, A.; Emig, D.; Albrecht, M.; Reményi, A. Specificity of Linear Motifs That Bind to a Common Mitogen-Activated Protein Kinase Docking Groove. *Sci. Signal.* **2012**, *5*, ra74–ra74.
- (69) Shaw, D.; Wang, S. M.; Villaseñor, A. G.; Tsing, S.; Walter, D.; Browner, M. F.; Barnett, J.; Kuglstat, A. The Crystal Structure of JNK2 Reveals Conformational Flexibility in the MAP Kinase Insert and Indicates Its Involvement in the Regulation of Catalytic Activity. *J. Mol. Biol.* **2008**, *383*, 885–893.
- (70) Ansideri, F.; Macedo, J. T.; Eitel, M.; El-Gokha, A.; Zinad, D. S.; Scarpellini, C.; Kudolo, M.; Schollmeyer, D.; Boeckler, F. M.; Blaum, B. S.; Laufer, S. A.; Koch, P. Structural Optimization of a Pyridinylimidazole Scaffold: Shifting the Selectivity from P38 α Mitogen-Activated Protein Kinase to c-Jun N-Terminal Kinase 3. *ACS Omega* **2018**, *3*, 7809–7831.
- (71) Mukherjee, K.; Sharma, M.; Jahn, R.; Wahl, M. C.; Südhof, T. C. Evolution of CASK into a Mg²⁺-Sensitive Kinase. *Sci. Signal.* **2010**, *3*, ra33–ra33.
- (72) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
- (73) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (74) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.