Reactivities of the Front Pocket Ncap Cysteines in Human Kinases

Ruibin Liu†, Shaoqi Zhan†‡, Ye Che¶, Jana Shen†
†Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201
‡Joint first author
¶Discovery Sciences, Pfizer Worldwide Research and Development, Groton, CT 06340

Abstract

The front pocket (FP) Ncap cysteine is the most popular site of covalent modification in kinases. A long-standing hypothesis associates the Ncap position with cysteine hyper-reactivity; however, traditional computational predictions suggest the FP Ncap cysteines as predominantly unreactive. Here we applied the state-of-the-art continuous constant pH molecular dynamics (CpHMD) to test the Ncap hypothesis. Simulations found that the Ncap cysteines of BTK/BMX/TEC/ITK/TXK, JAK3, and MKK7 are reactive to varying degrees; however, those of BLK and EGFR/ERBB2/ERBB4 possessing a Ncap+3 aspartate are unreactive. Analysis suggested that hydrogen bonding as well as electrostatic interactions drive the reactivity, and their absence renders a Ncap cysteine unreactive. To further test the Ncap hypothesis, we examined the FP Ncap+2 cysteines in JNK1/JNK2/JNK3 and CASK. Our work offers a systematic understanding of the cysteine structure-reactivity relationship and illustrates the use of CpHMD to differentiate cysteines towards the design of targeted covalent inhibitors with reduced chemical reactivities.

Graphical Abstract

jana.shen@rx.umaryland.edu.

Supporting Information

Table S1 contains a summary of the CpHMD calculated pKa’s; Table S2 contains the pKa’s calculated with traditional computational methods; Table S3 contains the distances between the Ncap Cys and Ncap+3 Asp in the crystal structures; Fig. S1-S6 contain the titration curves and analysis of the pH-dependent hydrogen bond and salt-bridge interactions.

The findings of this work may be of interest to ComputChem LLC, of which Dr. Jana Shen is a founder and scientific advisor.
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.\(^5\)

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 front pocket (FP) Cys located in the linker as the N-terminal cap (Ncap) of the \(\alpha\)D helix (Fig. 1a, 1b). Sequence alignment (by us and others\(^5,22\)) showed that 11 kinases in the human kinome possess such a FP Ncap Cys, including BTK/BMX/TEC/ITK/TXK 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. 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 Ncap Cys in all human kinases have been covalently modified. Most notably, the FDA-approved drug ibrutinib covalently modifies the Ncap Cys in BTK/BMX/TEC/ITK, BLK, JAK3, and EGFR/ERBB4 (Table 1).\textsuperscript{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 FP Ncap Cys in BTK was found to react with a range of chloroacetamide and acrylamide electrophilic fragments.\textsuperscript{36} The popularity of the FP Ncap Cys in TCKI discovery can be explained by its accessibility (near the ATP binding site) and expected reactivity (see below). A Ncap Asp has been known to stabilize a helix, which was attributed to the favorable interaction between the negative aspartate and the positive helix dipole.\textsuperscript{37,38} Similarly, a Ncap Cys was found to be helix stabilizing, which was attributed to the thiolate-helix dipole interaction.\textsuperscript{39,40} It was argued that Cys is the least frequent Ncap residue, because it is in the thiolate form prone to chemical modification.\textsuperscript{39} Indeed, the Ncap Cys in myoglobin has a measured pK\textsubscript{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\textsubscript{a} of 8.6 for the model Cys in a peptide.\textsuperscript{40} Thus, the consensus is that a Ncap Cys is in the hyper-reactive thiolate form at physiological pH, which we will refer to as the Ncap hypothesis.

Kinetics experiments demonstrated that the observed (apparent) rate constants of thiol reactions with electrophiles are inversely related to the thiol pK\textsubscript{a}’s.\textsuperscript{41-44} Thus, it is generally accepted that cysteines with low pK\textsubscript{a}’s are more prone to covalent modification due to the increased availability of the nucleophilic thiolate form, and the Cys pK\textsubscript{a} can be used as a proxy for reactivity in TCKI design.\textsuperscript{1,5,21,39} Curiously, however, according to the popular empirical pK\textsubscript{a} prediction program PROPKA,\textsuperscript{19} the Ncap Cys in the aforementioned 11 kinases have pK\textsubscript{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, despite the fact that the observed thiol nucleophilicity increases with lower pK\textsubscript{a}, the intrinsic nucleophilicity may decrease or stay constant depending on the specific molecule.\textsuperscript{41,44}

Accurate prediction of Cys pK\textsubscript{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\textsubscript{a}’s are 3.4–4.7 using the static structure based Poisson-Boltzmann (PB) and empirical PROPKA calculations.\textsuperscript{45} 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 smaller than the PB or PROKA 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\textsubscript{a} of 8.6\textsuperscript{46} for all cysteines.

Recently, we developed a GPU-accelerated continuous constant pH MD (CpHMD) method in Amber package\textsuperscript{47} for accurate and rapid prediction of protein pK\textsubscript{a}’s based on independent pH\textsuperscript{48} or pH replica-exchange titration\textsuperscript{49} simulations in the GB-Neck2 implicit solvent.\textsuperscript{50} Using replica-exchange CpHMD, which significantly accelerates the pK\textsubscript{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/αC-out inactive conformation.\textsuperscript{51} Most recently, we implemented an asynchronous algorithm to allow replica-exchange simulations on a single or any number of GPU cards.\textsuperscript{52} Benchmark study based on 24 proteins, including those with large pK\textsubscript{a} downshifts examined by Rowley and coworker\textsuperscript{45}.
and additional ones with large pKₐ 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.

Motivated by the importance of the Ncap Cys in TCKI design and intrigued by the contradiction between the Ncap hypothesis and the high pKₐ’s predicted by the empirical method, here we applied the asynchronous replica-exchange CpHMD to determine the pKₐ’s of all 11 human kinases that possess a FP Ncap Cys. We found that most Ncap Cys have pKₐ’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 Ncap 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 Ncap Cys provides insights into the varied reactivities and support to the general base assisted mechanism for the Michael addition with the Ncap Cys in EGFR/ERBB2/ERBB4 and BLK. To further explore the Ncap hypothesis, we also determined the pKₐ’s of the Ncap+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 Ncap Cys.

We performed the GPU-accelerated replica-exchange CpHMD titration simulations to determine the Cys and Lys pKₐ values for the aforementioned 11 human kinases possessing a FP Ncap Cys, which is located in the linker as the N-terminal cap of the αD helix (position 52 of the ATP binding site sequence according to KLIFS database). Except for TEC, TXK, and BLK, for which homology models were built using SWISS-MODEL, 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 pKₐ and to facilitate discussion, we group the Cys (or later Lys) in four categories (Table 2): unreactive (pKₐ > 9.4), somewhat reactive (pKₐ 8.4–9.4), reactive (pKₐ 7.4–8.4), and hyper-reactive pKₐ < 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).

Our simulations gave a wide pKₐ range from 7.6 to above 10.5 for the FP Ncap Cys, suggesting that its reactivity varies despite the same Ncap location. Among all, the Ncap Cys in BTK/BMX of the Tec family and JAK3 of the JakA family are nearly hyper-reactive, with the calculated pKₐ’s around 7.7 (Fig. 2 and Table S1). The Ncap Cys in TEC/TKX/ITK of the Tec family are also reactive, with the calculated pKₐ’s around 8.4 (Fig. 2 and Table S1), close to the value (8.6) for the model alanine penta-peptide (AACAA) in solution. The Ncap Cys in MKK7 is somewhat reactive with a calculated pKₐ of 9.1 (Fig. 2 and Table S1).
Interestingly, the Ncap 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), consistent with our previous work.\textsuperscript{51} Noticeably, our calculated $pK_a$'s of 7.7 and 8.4 for the FP Ncap Cys in BTK and ITK are in agreement with the respective values of 7.7 and 8.5 noted by Zapf and coworkers in the covalent inhibition study of ITK although the source of these $pK_a$'s was not mentioned.\textsuperscript{56} However, curiously, our simulation showed that EGFR’s Ncap 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.\textsuperscript{57} We will come back to the discussion of this discrepancy.

Sequence alignment of the 11 kinases showed that the FP Ncap Cys is often followed by an Asn or Asp residue at Ncap+3 position (Fig. 1 b and c). A previous experimental study suggested that the Ncap+3 residue may form hydrogen bond (h-bond) or electrostatic interaction with the Ncap residue.\textsuperscript{39} Our previous computational work\textsuperscript{51,53} 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.\textsuperscript{40,44,58} Thus, to understand the varied reactivities of the FP Ncap Cys, we binned the 11 kinases into three groups according to the nearby residues that provide stabilization or destabilization of the Ncap thiolate: Asn at Ncap+3 (BTK/BMX/TEC/TXK); Asp/Glu at Ncap+3/2 and Ser/Arg/Lys at Ncap+2/3 positions (ITK, JAK3, and MKK7); and Asp/Glu at Ncap+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 Ncap Cys reactivities.

**BTK/BMX/TEC/TXK: FP Ncap thiolate stabilized by the Ncap+3 Asn and other h-bond donors.**

BTK/BMX/TEC/TXK of the Tec family contain an Asn residue at the Ncap+3 position relative to the FP Ncap Cys. The calculated $pK_a$’s of the Ncap 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 Ncap+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.\textsuperscript{51,53} Similarly, the thiolate form of Cys496 in BMX is stabilized by h-bonding with the sidechain and backbone of Asn499 at Ncap+3 as well as the salt-bridge interaction with Arg540 (Fig. S1).

The calculated $pK_a$’s of the FP Ncap 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 Ncap+3 and Ser378, as well as a salt bridge with Arg493 positioned at Ncap+44. Cys350 in TXK forms h-bonds with Asn353 at Ncap+3 and salt-bridge with Arg394 at Ncap+44. It is interesting to notice that the Ncap+44 Arg which interacts with the Ncap thiolate is the second Arg of the highly conserved sequence
HRDLAARN on the catalytic loop.\textsuperscript{59} It is also noteworthy that the Ncap Cys in TEC and TXK does not form h-bond or salt-bridge with Ncap-2 or Ncap-3 residues such as Asn, Ser, Glu, and Arg (Fig. S1).

**ITK, JAK3, and MKK7: FP Ncap thiolate destabilized by Ncap+3/2 Asp/Glu but stabilized by Ncap+2/3 Ser/Arg/Lys or Ncap+1 Lys/Thr.**

Several of the Ncap Cys are in close proximity to an anionic side chain (Asp or Glu), which would increase the Cys p\(K_a\) due to electrostatic repulsion. However, h-bonding and salt-bridge formation with the nearby residues can stabilize the thiolate form and thereby decreasing the Ncap Cys p\(K_a\). These opposing effects are experienced by the Ncap Cys in ITK (Cys442), JAK3 (Cys909), and MKK7 (Cys218), resulting in the calculated p\(K_a\)’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 Ncap+2 and the backbone amide of Asp445 at Ncap+3, as well as form a salt-bridge with Arg486 positioned at Ncap+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 p\(K_a\) of Cys442 is 8.4, indicating that it is somewhat reactive towards electrophiles.

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

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

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

The Ncap 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 Ncap+3 position is occupied by an anionic Asp (or Glu in ERBB4) which can potentially destabilize the thiolate form of the Ncap Cys. Thus, it is not surprising that the calculated p\(K_a\)’s are above 10.5 (the highest simulation pH), suggesting that the Ncap Cys in these kinases are unreactive towards electrophiles.
Curiously, the Ncap 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 TCKI, and it was subsequently targeted by numerous investigational and FDA-approved inhibitors including ibrutinib, afatinib, and neratinib (Table 1). Analysis of the X-ray structures of EGFR (e.g., PDB ID 4zjv) reveals that the Cys797 thiol may donate a h-bond to the carboxylate group of the Ncap+3 Asp800 (Fig. 5). Similarly, the X-ray structures of ERBB2/ERBB4 show that the Ncap thiol sulfur can be within 3.5 Å to the nearest carboxylate oxygen of the Ncap+3 Asp (Table S3).

The high (calculated) pKᵦ of the Ncap 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 Ncap+3 Asp serves as a general base to deprotonate the Ncap Cys in the first stage of the reaction. Indeed, a recent combined quantum mechanical (QM)/MM simulation of EGFR demonstrated that in the proximity of a covalent inhibitor carrying an acrylamide warhead, the proton on the Ncap Cys797 thiol can be transferred to the carboxylate of the Ncap+3 Asp, resulting in a stable thiolate/Asp-COOH pair, from which the thiolate then attacks the Cβ of acrylamide generating a carbanion before being reprotonated by the Ncap+3 Asp-COOH to result in the thiol adduct. Our result is consistent with this base-assisted mechanism, although the proton transfer energetics may be underestimated in the QM/MM study based on the SCC-DFTB method. We should also note that in our apo simulations of EGFR/ERBB2/ERBB4 and BLK, the Ncap thiol rarely formed a h-bond with the Ncap+3 Asp/Glu, supporting the hypothesis that a proton transfer may only occur in the presence of inhibitor due to the restricted motion of the Cys and Asp sidechains. 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.

**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 the pKᵦ values. Table 3 lists the locations of all reactive Cys and Lys predicted by our simulations. Interestingly, a Cys on β4 sheet (Ncap-17/18) 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 TCKIs. 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 Ncap Cys is not affected by the conformation of the DFG-motif (in or out) or αC helix (in or out). As such, the calculated pKᵦ is insensitive to the variation of the kinase conformation. For example, using five X-ray structures (in different DFG/αC conformations according to KLIFS), the calculated pKᵦ of the Ncap Cys481 in BTK 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 pKᵦ above 10.4 in the DFG-in or αC-in states, the pKᵦ value can be significantly downshifted when the kinase adopts the DFG-out/αC-out
conformation\textsuperscript{51}). Our simulations predicted that the roof Lys in BTK (Lys430, PDB 3pj3),\textsuperscript{25} BMX (Lys445, PDB 3sxs),\textsuperscript{8} and JAK3 (Lys855, PDB 5lwm)\textsuperscript{67} are reactive. Interestingly, although the X-ray structures for these kinases show a DFG out state, $\alpha$C-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 $\alpha$C-Glu out conformations. The facile transition between the $\alpha$C-Glu in and out positions is consistent with our previous explicit-solvent simulations of the c-Src kinase.\textsuperscript{68} 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.\textsuperscript{3}

Comparison to other pK\textsubscript{a} calculations.

The CpHMD predicted pK\textsubscript{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 Ncap Cys pK\textsubscript{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\textsubscript{a}’s of about 7.7–7.8 for the Ncap Cys in BTK, BMX, and JAK3 (the lowest among the 11 kinases, Table S1), the PROPKA calculated pK\textsubscript{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,\textsuperscript{49,51} 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.\textsuperscript{69} For example, in BTK, the sidechain hydroxyl and backbone amide of Thr410 donate h-bonds to stabilize the Ncap 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\textsubscript{a}’s from the PROKA 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 coworker\textsuperscript{70} gave the pK\textsubscript{a}’s of 10.4–13.0 for the Ncap 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 Ncap Cys are reactive with the pK\textsubscript{a}’s of 7.7–8.4. With regards to the Ncap Cys797 in EGFR, Rowley and coworker calculated its pK\textsubscript{a} with either protonated or deprotonated Asp800. Interestingly, the calculated pK\textsubscript{a} was above 10 in both cases.\textsuperscript{70}

The FP Ncap+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 Ncap Cys discussed so far. This Ncap+2 Cys, Cys116 in JNK1/ JNK2, Cys154 in JNK3, or Cys100 in CASK is the second N-terminal residue in the $\alpha$D helix, and has an Ile in the Ncap+3 position instead of Asn/Asp/Glu. The calculated pK\textsubscript{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 thiolate is additionally stabilized by the h-bond formation with Ser72, which may explain the lower pK\textsubscript{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\textsubscript{a} upshift due to desolvation is not offset, leading to a higher pK\textsubscript{a} value of Cys100 as compared to Cys116 or Cys154 in JNKs.

CONCLUDING DISCUSSION

Ncap Cys has long been hypothesized as hyper-reactive due to the thiolate-helix dipole interactions.\textsuperscript{39,40} In support of this hypothesis, the Ncap Cys in all human kinases have been covalently targeted. Our data demonstrated that the Ncap position alone may not stabilize the thiolate form making the Cys reactive at physiological pH. CpHMD simulations confirmed the reactivities of the Ncap Cys in BTK/BMX/TEC/TXK/ITK and JAK3, with the calculated pK\textsubscript{a}’s ranging from 7.7 to 8.5. The calculated pK\textsubscript{a}’s of BTK and ITK are very similar to those noted in an experimental study of ITK\textsuperscript{56} (personal communication with the Pfizer team that performed the study).\textsuperscript{71} In comparison, the calculated pK\textsubscript{a} is 9.1 for the Ncap Cys in MKK7, suggesting a lower reactivity compared to the aforementioned Tec family kinases and JAK3.

Surprisingly, the calculated pK\textsubscript{a}’s are above 10.4 for the Ncap Cys in EGFR/ERBB2/ERBB4 and BLK, suggesting that the thiol form is predominant at physiological pH despite the Ncap position. Curiously, the calculated pK\textsubscript{a} for EGFR is 5 units higher than that from the pH-dependent bromobimane-thiol reaction rate.\textsuperscript{57} 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 dipole\textsuperscript{72} and consequently the pK\textsubscript{a} downshift due to the thiolate-dipole interaction. Base-catalyzed mechanism is common for thiol-Michael addition reactions in the absence of a deprotonated thiol.\textsuperscript{73,74} Considering that EGFR and ERBB2/ERBB4 have a Ncap+3 Asp and the carboxyl group is within the h-bond distance to the Ncap thiol in some crystal structures, we speculate that the Ncap+3 Asp may act as a general base. Similarly, a 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.\textsuperscript{60} Therefore, the Cys797 pK\textsubscript{a} determined through the bromobimane-thiol reaction\textsuperscript{57} might reflect the Cys797… Asp800 pair in the presence of bromobimane. However, we note that the proton transfer energy in the aforementioned QM/MM study\textsuperscript{60} may be underestimated by the SCC-DFTB method.\textsuperscript{61}

Other base-assisted mechanisms may also explain the 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 Ncap thiol yielding the nucleophilic thiolate.\textsuperscript{30,62,63} 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 Ncap thiol, as proposed by a most recent QM/MM study of BTK modification by ibrutinib assuming the Ncap Cys is protonated.\textsuperscript{75}

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

Our analysis showed that solvent exposure of the Ncap or Ncap+2 Cys is very low, corroborating our previous finding that all reactive cysteines are largely buried\textsuperscript{53} and consistent with an early bioinformatics survey which suggests that cysteine is the least exposed residue in proteins.\textsuperscript{76} However, the h-bond and electrostatic environment of the Ncap and Ncap+2 Cys can be very different, which gives rise to the varied Cys $pK_a$’s. Specifically, interaction with the 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.\textsuperscript{51,53} Notably, the presence of a h-bond donor at the Ncap+3 position (e.g., Asn) provides stabilization to the thiolate form in BTK/BMX/TXK. By contrast, the presence of an anionic sidechain (e.g. Asp or Glu) at the Ncap+3 position destabilizes the thiolate form. Although ITK, JAK3, BLK, and EGFR/ERBB2/ERBB4 all have an Ncap+3 Asp next to the Ncap Cys, only ITK and JAK3 have favorable h-bond interactions that stabilize the thiolate form. As a result, the Ncap 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 the pH-dependent formation of h-bond and electrostatic interactions that are either nascent or completely absent in the static (X-ray) structure but can nonetheless significantly impact the protonation state. Such interactions with the Ncap or Ncap+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 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 the underestimation of helix dipole as well as the neglect of a possible proton transfer between Cys and a nearby general base, which can be located on the protein (e.g. Ncap+3 Asp) or inhibitor. CpHMD simulations with the holo structures and QM/MM studies will be conducted in the future to test the hypothesis that a base-assisted mechanism underlies the Michael addition reactions with the Ncap Cys in EGFR/ERBB2/ERBB4 and BLK. Taken together, our work offers a systematic understanding of the cysteine structure-reactivity relationship and demonstrates the utility of CpHMD simulations in differentiating cysteine sites to assist in the design of targeted covalent inhibitors with reduced chemical reactivities. These capabilities would also allow the CpHMD tool to complement the chemical proteomic platform\textsuperscript{36,77} in the discovery of new druggable sites in the proteome.
Experimental

Structure Preparation.—Except for TEC, TXK, and BLK, the initial structures were taken from the Protein Data Bank (PDB): BTK (3pj3),8 BMX (3xsx),8 ITK (3miy),24 JAK3 (5lwm),67 EGFR (4zjv),32 ERBB4 (2r4b),30 JNK1 (2xrw),78 JNK2 (3e7o),79 JNK3 (6emh),80 and CASK (3mfr).81 The structures of TEC, TXK, and BLK were built using SWISS-MODEL.20 For each structure, the open source tool pdbfixer from the OpenMM package82 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 (GB) based CpHMD simulations48,49 were performed using the pmdm engine of AMBER18 (AMBER 2018).47 The pH replica-exchange titration simulations were performed on the GPUs using our asynchronous replica-exchange implementation.52 The proteins were represented by the ff14sb protein force field,83 and solvent represented by the GBNek2 (igb=8) implicit solvent model50 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.48,49,51 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 500 steps of 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 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.84 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 pKₐ’s of 6.5, 8.5, and 10.4, respectively.46 If not stated, all settings were identical to our previous work.48,51,53

pKₐ calculations.—The unprotonated fraction (S) of a titratable residue at different pH values was calculated using the definitions of the protonated (λ < 0.2) and deprotonated (λ > 0.8) states, as in our previous work.48,49,53 The pKₐ 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. The $S$ values based on the 10-ns simulations at different pH were combined to generate $m^n$ 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).

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**ACKNOWLEDGEMENTS**

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

**ABBREVIATIONS**

| Abbreviation | Description |
|--------------|-------------|
| CpHMD        | continuous constant pH molecular dynamics |
| FDA          | U.S. Food and Drug Administration |
| FP           | front pocket |
| GB           | Generalized Born |
| h-bond       | hydrogen bond |
| HH           | Henderson-Hasselbalch |
| MD           | molecular dynamics |
| MC           | Monte-Carlo |
| MM           | molecular mechanics |
| Ncap         | N-terminal cap |
| PB           | Poisson-Boltzmann |
| RMSE         | root-mean-square error |
| TCKI         | targeted covalent kinase inhibitor |
| TI           | thermodynamic integration |
| QM           | quantum mechanics |

**References**

(1). Ferguson FM; Gray NS Kinase Inhibitors: The Road Ahead. Nat. Rev. Drug Discov 2018, 17, 353–377. [PubMed: 29545548]
(2). Santos R; Ursu O; Gaulton A; Bento AP; Donadi RS; Bologa CG; Karlsson A; Al-Lazikani B; Hersey A; Oprea TI; Overington JP A Comprehensive Map of Molecular Drug Targets. Nat. Rev. Drug Discov 2017, 16, 19–34. [PubMed: 27910877]

(3). Zhao Z; Bourne PE Progress with Covalent Small-Molecule Kinase Inhibitors. Drug Discov. Today 2018, 23, 727–735. [PubMed: 29337202]

(4). Abdeldayem A; Raouf YS; Constantinescu SN; Moriggl R; Gunting PT Advances in Covalent Kinase Inhibitors. Chem. Soc. Rev 2020, 49, 2617–2687. [PubMed: 3227030]

(5). Liu Q; Sabnis Y; Zhao Z; Zhang T; Buhrlage SJ; Jones LH; Gray NS Developing Irreversible Inhibitors of the Protein Kinase Cysteine. Chem. Biol 2013, 20, 146–159. [PubMed: 23437444]

(6). Bauer RA Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. Drug Discovery Today 2015, 20, 1061–1073. [PubMed: 26002380]

(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. [PubMed: 28056780]

(8). Muckelbauer J; Sak S; Ahmed N; Burke J; Chang CY; Gao M; Tino J; Xie D; Tebben AJ X-Ray Crystal Structure of Bone Marrow Kinase in the X Chromosome: A Tec Family Kinase. Chemical Biology & Drug Design 2011, 78, 739–748. [PubMed: 21883956]

(9). Dhillon S Tirabrutinib: First Approval. Drugs 2020, 80, 835–840. [PubMed: 32382949]

(10). Dhillon S Orelabrutinib: First Approval. Drugs 2018, 78, 103–507. [PubMed: 33704654]

(11). Robinson MF; Damjanov N; Stamenkovic B; Radunovic G; Kivitz A; Cox L; Manukyan Z; Banfield C; Saunders M; Chandra D; Vincent MS; Mancuso J; Peeva E; Beebe JS 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. [PubMed: 32419304]

(12). Kim ES Olmutinib: First Global Approval. Drugs 2016, 76, 1153–1157. [PubMed: 27357069]

(13). Blair HA Pyrotinib: First Global Approval. Drugs 2018, 78, 1751–1755. [PubMed: 30341682]

(14). Gonzalvez F; Vincent S; Baker TE; Gould AE; Li S; Wardwell SD; Nadworny S; Ning Y; Zhang S; Huang W-S; Hu Y; Li F; Greenfield MT; Zech SG; Das B; Narasimhan NI; Clackson T; Dalgarno D; Shakespeare WC; Fitzgerald M; Chouitar J; Griffin RJ; Liu S; Wong K-K; Zhu X; Rivera VM Mobocertinib (TAK-788): A Targeted Inhibitor of EGFR Exon 20 Insertion Mutants in Non-Small Cell Lung Cancer. Cancer Discov. 2021, 11, 1673–1687.

(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. [PubMed: 31083997]

(16). Schröder M; Tan L; Wang J; Liang Y; Gray NS; 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. [PubMed: 32783966]

(17). Zhang T; Inesta-Vaquera F; Niepel M; Zhang J; Ficarro SB; Machleidt T; Xie T; Marto JA; Kim N; Sim T; Laughlin JD; Park H; LoGrasso PV; Patricelli M; Nomanbhoy TK; Sorger PK; Alessi DR; Gray NS Discovery of Potent and Selective Covalent Inhibitors of JNK. Chem. Biol 2012, 19, 140–154. [PubMed: 22284361]

(18). Muth F; El-Gokha A; Asnider F; Eitel M; Do E; Sievers-Engler A; Lange A; Boeckler FM; Lam M; Laufler SA Tri- and Tetrasubstituted Pyridinylimidazoles as Covalent Inhibitors of C-Jun N-Terminal Kinase 3. J. Med. Chem. 2017, 60, 594–607. [PubMed: 27977190]

(19). Olsson M; Søndergaard CR; Rostkowski M; Jensen JH PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions. J. Chem. Theory Comput 2011, 7, 525–537. [PubMed: 26596171]

(20). Waterhouse A; Bertoni M; Bienert S; Studer G; Tauriello G; Gumieny R; Heer FT.de Beer TAP;Rempfer C; Bordoli L; Lepore R; Schwede T SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. Nucleic Acids Research 2018, 46, W296–W303. [PubMed: 29788355]
(35). Chen J; Kinoshita T; Sukbuntherng J; Chang BY; Elias L Ibrutinib Inhibits ERBB Receptor Tyrosine Kinases and HER2-Amplified Breast Cancer Cell Growth. Mol. Cancer Thera 2016, 15, 2835–2844.

(36). Backus KM; Correia BE; Lum KM; Forli S; Horning BD; González-Páez GE; Chatterjee S; Lanning BR; Teijaro JR; Olson AJ; Wolan DW; Cravatt BF Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. Nature 2016, 534, 570–574. [PubMed: 27309814]

(37). Hol WGJ Effects of the α-Helix Dipole upon the Functioning and Structure of Proteins and Peptides. Adv. Biophys 1985, 19, 133–165. [PubMed: 2424281]

(38). Serrano L; Fersht AR Capping and A-Helix Stability. Nature 1989, 342, 296–299. [PubMed: 2812029]

(39). Anderson TA; Sauer RT Role of an Ncap Residue in Determining the Stability and Operator-Binding Affinity of Arc Repressor. Biophys. Chem 2002, 100, 341–350.

(40). Miranda JL Position-Dependent Interactions between Cysteine Residues and the Helix Dipole. Prot. Sci 2003, 12, 73–81.

(41). Bulaj G; Kortemme T; Goldenberg DP Ionization-Reactivity Relationships for Cysteine Thiols in Polypeptides. Biochemistry 1998, 37, 8965–8972. [PubMed: 9636038]

(42). Peskin AV; Winterbourn CC Kinetics of the Reactions of Hypochlorous Acid and Amino Acid Chloramines with Thiols, Methionine, and Ascorbate. Free Radical Biol. Med 2001, 30, 572–579. [PubMed: 1182528]

(43). Tosatto SCE; 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. [PubMed: 18500926]

(44). 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. [PubMed: 21391663]

(45). Awoonor-Williams E; Rowley CN Evaluation of Methods for the Calculation of the pKₐ of Cysteine Residues in Proteins. J. Chem. Theory Comput 2016, 12, 4662–4673. [PubMed: 27541839]

(46). Thurlkill RL; Grimsley GR; Scholtz JM; Pace CN pK Values of the Ionizable Groups of Proteins. Protein Sci. 2006, 15, 1214–1218. [PubMed: 16597822]

(47). Case DA; Ben-Shalom IY; Brozell SR; Cerutti DS; Cheatham T III; Cruzeiro VWD; Darden TA; Duke RE; Ghoreishi D; Gilson MK; Gohlke H; Goetz AW; Greene D; Harris R; Homeyer N; Huang Y; Izadi S; Kovalenko A; Kurtzman T; Lee TS; LeGrand S; Li P; Lin C; Liu J; Luchko T; Luo R; Mermelstein DJ; Merz KM; Miao Y; Monard G; Nguyen C; Nguyen H; Omelyan I; Onufriev A; Pan F; Qi R; Roe DR; Roitberg A; Sagui C; Schott-Verdugo S; Shen J; Simmerling CL; Smith J; Salomon-Ferrer R; Swails J; Walker RC; Wang J; Wei H; Wolf RM; Wu X; Xiao L; York DM; Kollman PA AMBER 2018: 2018.

(48). Huang Y; Harris RC; Shen J Generalized Born Based Continuous Constant pH Molecular Dynamics in Amber: Implementation, Benchmarking and Analysis. J. Chem. Inf. Model 2018, 58, 1372–1383. [PubMed: 29949356]

(49). Harris RC; 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. [PubMed: 31661616]

(50). Nguyen H; Roe DR; Simmerling C Improved Generalized Born Solvent Model Parameters for Protein Simulations. J. Chem. Theory Comput 2013, 9, 2020–2034. [PubMed: 25788871]

(51). 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. [PubMed: 30945531]

(52). Henderson JA; Verma N; Harris RC; 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. [PubMed: 32962355]
(53). Harris RC; 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. [PubMed: 32330035]

(54). van Linden OJP; Kooistra AJ; Leurs R; de Esch IJP; de Graaf C KLIFS: A Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space. J. Med. Chem 2014, 57, 249–277. [PubMed: 23941661]

(55). Winterbourn CC; Metodiewa D Reactivity of Biologically Important Thiol Compounds with Superoxide and Hydrogen Peroxide. Free Radi. Biol. Med 1999, 27, 322–328.

(56). Zapf CW; Gerstenberger BS; Xing L; Limburg DC; Anderson DR; Caspers N; Han S; Aulabaugh A; Kurumbail R; Shaky SA; Li X; Spaulding V; Czerwinski RM; Seth N; Medley QG 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. [PubMed: 23098091]

(57). Truong TH; Ung PM-U; Palde PB; Paulsen CE; Schlessinger A; Carroll KS Molecular Basis for Redox Activation of Epidermal Growth Factor Receptor Kinase. Cell Chem. Biol 2016, 23, 837–848. [PubMed: 27427230]

(58). Britto P; Knipling L; Wolff J The Local Electrostatic Environment Determines Cysteine Reactivity of Tubulin. J. Biol. Chem 2002, 277, 29018–29027. [PubMed: 12023292]

(59). McTigue MA; Wickersham JA; Pinko C; Showalter RE; Parast CV; Tempczyk-Russell A; Gehring MR; Mroczkowski B; Kan C-C; Villafranca JE; 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. [PubMed: 10368301]

(60). 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. [PubMed: 25658136]

(61). Awoonor-Williams E; Isley WC; Dale SG; Johnson ER; Yu H; Becke AD; Roux B; Rowlie CN Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols. J. Comput. Chem 2020, 41, 427–438. [PubMed: 31512279]

(62). Tsou H-R; Manuwa N; Johnson BD; Reich MF; Gruber BC; Ye F; Nilakantan R; Shen R; Discafani C; DeBlanc R; Davis R; Koehn FE; Greenberger LM; 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. [PubMed: 11495584]

(63). Wissner A; Overbeek E; Reich MF; Floyd MB; Johnson BD; Manuwa N; Rosfjord EC; Discafani C; Davis R; Shi X; Rabindran SK; Gruber BC; Ye F; Hallett WA; Nilakantan R; Shen R; Wang Y-F; Greenberger LM; Tsou H-R 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. [PubMed: 12502359]

(64). Henise JC; Taunton J Irreversible Nek2 Kinase Inhibitors with Cellular Activity. J. Med. Chem 2011, 54, 4133–4146. [PubMed: 21627121]

(65). Zhao A; Lee SH; Mojena M; Jenkins RG; Patrick DR; Huber HE; Goetz MA; Hensons OD; Zink DL; Vilella D; Dombrowski AW; Lingham RB; Huang L Resorcylic Acid Lactones: Naturally Occurring Potent and Selective Inhibitors of MEK. J. Antibiot 1999, 52, 1086–1094.

(66). Ahmad R; Raina D; Meyer C; Kharbanda S; Kufe D Triterpenoid CDDO-Me Blocks the NF-kappaB Pathway by Direct Inhibition of IKKbeta on Cys-179. J. Biol. Chem 2006, 281, 35764–35769. [PubMed: 16998237]

(67). Forster M; Chaikuad A; Bauer SM; Holstein J; Robers MB; Corona CR; Gehringer M; Pfaffnenrot E; Ghoreschi K; Knapp S; Laufer SA Selective JAK3 Inhibitors with a Covalent Reversible Binding Mode Targeting a New Induced Fit Binding Pocket. Cell Chem. Biol 2016, 23, 1335–1340. [PubMed: 27840070]

(68). 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. [PubMed: 31476863]

J Med Chem. Author manuscript; available in PMC 2022 February 03.
(69). Huang Y; Yue Z; Tsai C-C; Henderson JA; Shen J Predicting Catalytic Proton Donors and Nucleophiles in Enzymes: How Adding Dynamics Helps Elucidate the Structure-Function Relationships. J. Phys. Chem. Lett. 2018, 9, 1179–1184. [PubMed: 29461836]

(70). Awoonor-Williams E; Rowley CN How Reactive Are Druggable Cysteines in Protein Kinases? J. Chem. Inf. Model 2018, 58, 1935–1946. [PubMed: 30118220]

(71). The Cys pKa’s were determined from the pH dependence of inactivation kinetics via Kitz-Wilson analysis (Kitz and Wilson, J. Biol. Chem. 1962, 237, 3235–3239). The pKa values can somewhat vary depending on a number of factors, e.g. ATP concentration and protein construction (phosphorylated or not). For example, the Ncap Cys pKa of ITK was determined as 8.2±0.1 without ATP and 8.5±0.1 with 0.5 mM ATP. For example, C443 pKa of ITK was determined to be around 8.2 ± 0.1 with no ATP in presence, or 8.5 ± 0.1 under 0.5 mM ATP.

(72). Huang J; MacKerell AD Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (AAQAA)3 Peptide. Biophys. J. 2014, 107, 991–997. [PubMed: 25140435]

(73). Krenske EH; Petter RC; Zhu Z; Houk KN Transition States and Energetics of Nucleophilic Additions of Thiols to Substituted α,β-Unsaturated Ketones: Substituent Effects Involve Enone Stabilization, Product Branching, and Solvation. J. Org. Chem. 2011, 76, 5074–5081. [PubMed: 21574592]

(74). Nair DP; Podgórski M; Chatani S; Gong T; Xi W; Fenoli CR; Bowman CN The Thiol-Michael Addition Click Reaction: A Powerful and Widely Used Tool in Materials Chemistry. Chem. Mater. 2014, 26, 724–744.

(75). Voice AT; Tresadern G; Twidale RM; van Vlijmen H; Mulholland AJ Mechanism of Covalent Binding of Ibrutinib to Bruton’s Tyrosine Kinase Revealed by QM/MM Calculations. Chem. Sci. 2021, 12, 5511–5516. [PubMed: 33995994]

(76). Marino SM Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. J. Mol. Biol. 2010, 404, 902–916. [PubMed: 20950627]

(77). Weerapana E; Wang C; Simon GM; Richter F; Khare S; Dillon MB; Bachovchin DA; Mowen K; Baker D; Cravatt BF Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. Nature 2010, 468, 790–795. [PubMed: 21085121]

(78). Shaw D; Wang SM; Villasenor AG; Tsing S; Walter D; Browner MF; Barnett J; Kuglstatter 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. [PubMed: 18801372]

(79). Eastman P; Swails J; Chodera JD; McGibbon RT; Zhao Y; Beauchamp KA; Wang L-P; Simmonett AC; Harrigan MP; Stern CD; Wiewiora RP; Brooks BR; Pande VS OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. PLoS Comput. Biol. 2017, 13, e1005659. [PubMed: 28746339]

J Med Chem. Author manuscript; available in PMC 2022 February 03.
Figure 1. Human kinases with a FP Ncap or Ncap+2 Cys.
(a) The 15 kinases possessing a FP Ncap or Ncap+2 Cys displayed in the human kinome tree as orange or yellow circles, respectively. The kinome tree was generated with KinMap.7
(b) Structure of the BMX kinase domain showing a FP Ncap Cys (orange sphere, PDB ID: 3sx8). (c) Names of the kinases possessing a FP Ncap (orange) or Ncap+2 Cys (yellow), the families and groups they belong to, and the nearby amino acid sequences. The nearby residues making a significant contribution to its pKₐ shift are colored blue, magenta, green, or brown (same as in Fig. 2). For JAK3, the second domain which has a crystal structure is studied here.
Figure 2. Predicted pK$_a$'s and reactivities for the FP Ncap and Ncap+2 Cys in 15 kinases. The reactivity definitions are given in Table 2. The exact pK$_a$ values for the Ncap 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 Ncap Cys: blue (Ncap+3 Asn), magenta (Ncap+3/2 Asp/Glu and Ncap+2/3 Ser/Arg/Lys), and green (Ncap+3 Asp/Glu). The kinases with a FP Ncap+2 Cys are shown in brown.
Figure 3. The FP Ncap thiolate is stabilized by hydrogen bonding with Ncap+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°.
Figure 4. Stabilizing interactions for the FP Ncap 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).
Figure 5. The FP Ncap Cys thiol may donate a h-bond to the Ncap+3 aspartate in EGFR. Zoomed-in view of the structural environment of the Ncap Cys797 in EGFR from the apo X-ray crystal structure (PDB: 4zjv). The distance between the thiol sulfur and the nearest carboxyl oxygen of Asp800 is 3.3 Å. The potential h-bond is indicated as a dashed line.
Table 1.

Summary of kinases with a FP Ncap or Ncap+2 Cys and the covalent inhibitors\(^\text{a}\) directed at the location

| Kinase | Family | Group | PDB  | Residue | PROPKA\(^b\) | TCKI |
|--------|--------|-------|------|---------|-------------|------|
| **FP Ncap 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    | 4hec\(^a\) | C449    | 9.2         | ibrutinib, ritlecinitib\(^c\) |
| ITK    | Tec    | TK    | 3miy | C442    | 10.7        | ibrutinib |
| TXK    | Tec    | TK    | 3t9g\(^a\) | C350    | 10.4        | acalabrutinib |
| BLK    | Src    | TK    | 4mxy\(^a\) | C319    | 10.6        | ibrutinib, acalabrutinib |
| JAK3   | JakA   | TK    | 5bwm | C909    | 10.3        | ritlecinitib\(^11\) |
| EGFR   | EGFR   | TK    | 4zjv | C797    | 11.4        | afatinib, osimertinib, dacomitinib, neratinib, olimtinib,\(^12\) pyro tinib,\(^13\) mobocertinib,\(^14\) ibrutinib, acalabrutinib |
| ERBB2  | EGFR   | TK    | 3pp0 | C805    | 10.8        | aca labrutinib |
| ERBB4  | EGFR   | TK    | 2r4b | C803    | 11.0        | ibrutinib |
| MKK7   | STE7   | STE   | 6qfr | C218    | 9.1         | investigational inhibitors\(^15,16\) |
| **FP Ncap+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.\(^20\)
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     |
Table 3.

Summary of the predicted reactive Cys and Lys in the 15 kinases discussed in this work.\(^a\)

| Kinase Family | Cys Location | Reactivity | Major contributors | Lys Location | Reactivity |
|---------------|--------------|------------|--------------------|--------------|------------|
| **FP Ncap Cys** | | | | | |
| BTK Tec | C481 front pocket | reactive | N(+3), T(−71) | K430 roof | reactive |
| BMX Tec | C479 β4 sheet | reactive | N(+3), R(+44) | K445 roof | somewhat |
| C496 front pocket | reactive | | | | |
| TEC Tec | C406 αC helix | reactive | | | |
| C432 β4 sheet | reactive | | | | |
| C449 front pocket | reactive | | | | |
| TXK Tec | C350 front pocket | somewhat | N(+3), R(+44) | | |
| C333 β4 sheet | hyper | | | | |
| ITK Tec | C422 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 αG helix | reactive | | | | |
| MKK7 STE7 | C147 P loop | hyper | | | |
| C218 front pocket | reactive | | S(−74), K(+3) | | |
| C276 DFG-1 | reactive | | | | |
| C341 loop (αF/αG) | hyper | | | | |
| EGFR | C775 loop (αC/β1) | reactive | | | |
| C781 β4 sheet | reactive | | | | |
| ERBB2 EGFR | C789 β4 sheet | hyper | | K860 DFG-3 | reactive |
| ERBB4 EGFR | C891 A loop | reactive | | K858 DFG-3 | reactive |
| BLK Src | C460 loop (αG/αH) | somewhat | | | |
| **FP Ncap+2 Cys** | | | | | |
| JNK1 MAPK | C116 front pocket | reactive | Q(+1), S(+39) | | |
| C245 | reactive | αH helix | | | |
| JNK2 MAPK | C116 front pocket | reactive | Q(+1), S(+39) | K153 β7 sheet | somewhat |
| JNK3 MAPK | C154 front pocket | hyper | S(−82), Q(+1), S(+39) | | |
The 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 Ncap Cys are listed with the sequence positions (relative to Ncap) given in the parenthesis. The calculated $pK_a$'s of all listed residues are given in Table S1.

| Kinase | Family | Cys | Location   | Reactivity | Major contributors | Lys | Location | Reactivity |
|--------|--------|-----|------------|------------|--------------------|-----|----------|------------|
| CASK   | CASK   | C283| reactive loop ($\alpha$G/\alpha$H$) |            |                    |     |          |            |
| CASK   | C100   |     | front pocket | somewhat   |                    |     |          |            |