Journal of Medicinal Chemistry

ARTICLE

pubs.acs.org/jmc

# Cysteine Mapping in Conformationally Distinct Kinase Nucleotide Binding Sites: Application to the Design of Selective Covalent Inhibitors

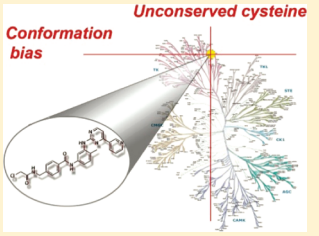
Emeline Leproult,[†] Sofia Barluenga,[‡] Dino Moras,[†] Jean-Marie Wurtz,*[,†] and Nicolas Winssinger*[,‡]

[†]Institut de Génétique et de Biologie Moléculaire et Cellulaire—CNRS (UMR 7104), INSERM (U964), 1 rue Laurent Fries, Université de Strasbourg, 67400 Illkirch, France

[‡]Institut de Science et Ingénierie Supramoléculaire—CNRS (UMR 7006), Université de Strasbourg, 8 allée Gaspard Monge, 67000 Strasbourg, France

Supporting Information

**ABSTRACT:** Kinases have emerged as one of the most prolific therapeutic targets. An important criterion in the therapeutic success of inhibitors targeting the nucleotide binding pocket of kinases is the inhibitor residence time. Recently, covalent kinase inhibitors have attracted attention since they confer terminal inhibition and should thus be more effective than reversible inhibitors with transient inhibition. The most robust approach to design irreversible inhibitors is to capitalize on the nucleophilicity of a cysteine thiol group present in the target protein. Herein, we report a systematic analysis of cysteine residues present in the nucleotide binding site of kinases, which could be harnessed for irreversible inhibition, taking into consideration the different kinase conformations. We demonstrate the predictive power of this analysis with the design and validation of an irreversible inhibitor of KIT/PDGFR kinases. This is the first example of a covalent kinase inhibitor that combines a pharmacophore addressing the DFG-out conformation with a covalent trap.




## INTRODUCTION

The sequencing of the human genome resulted in the identification of 518 protein kinases, which constitutes about 1.7% of all human genes.[1] Most protein kinases belong to a single superfamily containing a eukaryotic protein kinase[1] catalytic domain. This domain is composed of approximately 270 amino acids and promotes the transfer of the ATP $\gamma$ phosphate to serine, threonine, or tyrosine residues of protein substrates (Figure 1A). Kinases are responsible for the phosphorylation of about one-third of the proteome, endowing them with a central role in regulating biochemical pathways. Dysfunction in kinase regulation is responsible for numerous pathologies ranging from inflammation to cancers[2−6] and has been the subject of numerous drug discovery efforts. Thus far, the most successful approaches to designing selective kinase inhibitors targeting the nucleotide binding pocket have been based on the presence of an inactive conformation adopted by some kinases, which provides an enhanced level of discrimination within the kinome. Indeed, "type 1½"[7] and "type 2"[8] inhibitors, such as lapatinib[9] and imatinib,[10] which target the inactive "C-helix-out"[11] (Figure 1B) and "DFG-out"[12] (Figure 1C) conformations, respectively, have been reported to be more specific than inhibitors targeting the active conformation,[13] which is highly conserved throughout the kinome. Alternatively, efforts have focused on the unique position of cysteine residues within the nucleotide binding site to develop irreversible inhibitors (Figure 2). Indeed, the most selective kinase inhibitor reported to date takes advantage of two selectivity filters, a bulky gatekeeper residue in conjunction with a cysteine residue present in only 11 kinases, to achieve inhibition of ribosomal protein S6 kinases (RSKs) only.[14] Irreversible inhibitors have a long history in therapeutic intervention,[15] and several efforts in the kinase area have progressed to clinical trials: HKI-272,[16] CI-1033,[17] EKB-569,[18] and BIBW 2992,[19] which are directed against lung cancer by inhibiting the epidermal growth factor receptor (EGFR) kinase.[20] Encouragingly, these irreversible inhibitors overcame some of the limitations of reversible EGFR inhibitors as they retain activity against the otherwise debilitating T790—M mutation,[16] while maintaining a very good selectivity profile.[21,22]

To date, only four distinct cysteine positions in the ATP binding site have been harnessed for irreversible inhibition (the positions are marked in orange in Figure 3A), and efforts have been restricted to cysteines accessible in the "active conformation". A cysteine, located just after the hinge region and present in 11 kinases, is targeted with covalent EGFR[16,23−26] or Bruton's tyrosine kinase (BTK)[27] inhibitors. A second cysteine, located on the second $\beta$-strand, is targeted by the covalent inhibitor CMK,[14] directed against RSK kinases. Covalent inhibitors exploiting a third cysteine shared by 48 kinases, which is localized before the DFG motif, have been discovered from a natural product family (resorcylic acid lactones such as hypothemycin).[28−31] The most recent covalent inhibition that has been reported targets a

**Received:** October 28, 2010
**Published:** February 15, 2011





dx.doi.org/10.1021/jm101396q | J. Med. Chem. 2011, 54, 1347–1355

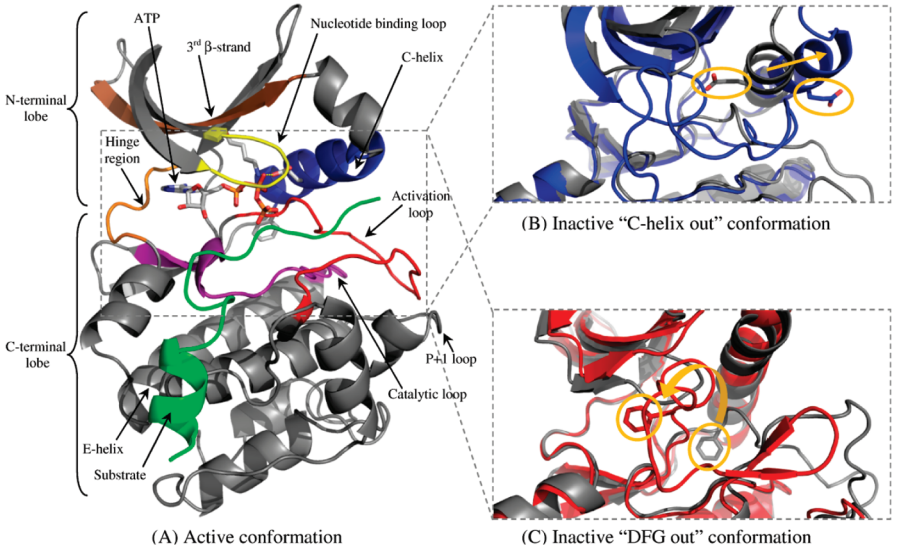



**Figure 1.** Catalytic kinase domain and its structural elements. (A) cAMP-dependent kinase (PDB code 1ATP) in the active conformation. The nucleotide binding loop is in yellow; the third $\beta$-strand of the $\beta$-sheet, bearing the catalytic lysine, is in brown; the C-helix, bearing the catalytic glutamic acid residue, is in blue; the catalytic loop is in purple; the activation loop, starting with the DFG motif, is in red; the peptide substrate is in green; ATP is represented in sticks and located in its binding site. Nucleotide and substrate binding sites are outlined in dashed line. (B) Cell division protein kinase 2 (CDK2): superimposition of the inactive C-helix-out conformation (in blue; PDB code 1PXN) with the active one (in gray; PDB code 3BHV). In the C-helix-out conformation, the glutamic acid residue (shown in stick in the two structures), located on the C-helix, is moved away from the ATP binding site due to the ionic interaction disruption with the catalytic lysine residue, involving a general displacement of the C-helix. Apart from the DFG motif, the activation loop is also highly displaced. (C) KIT kinase: superimposition of the inactive DFG-out conformation (in red; PDB code 1T46) with the active one (in gray; PDB code 1PKG). The phenylalanine of DFG motif is shown in stick in the two structures, to see its displacement in the DFG-out conformation, due to a general movement of the activation loop.

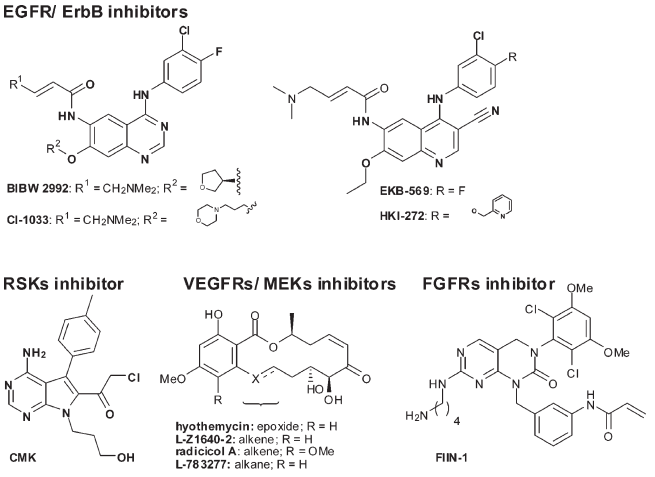



**Figure 2.** Representative irreversible kinase inhibitors reported to date.

cysteine located on the nucleotide binding loop of basic fibroblast growth factor receptor (FGFR) kinases.[32]

In their review, Zhang et al.[8] identified about 200 kinases bearing a cysteine in the nucleotide pocket, thus highlighting the broad applicability of this approach. While several reports have described the use of bioinformatics approaches[8,14,23,28] to identify a targetable cysteine, a systematic analysis taking into account the kinase conformation and accessibility to the thiol has not been reported.

## RESULTS

**Workflow Overview.** The workflow presented here allows access to all residues whose side chain robustly participates in the kinase nucleotide binding site for a large number of protein kinase crystal structures (Figure 3). The beginning of the workflow is tuned to process all available crystal structures of the kinase catalytic domain and classify them into conformations. As a result, the structures were mainly classified into three conformations: active,[13] inactive C-helix-out,[11] and inactive DFG-out.[12] This classification is important for the subsequent analysis since the shape of the ATP binding pocket varies in function of the conformation.[13] The workflow allowed side chains that were not pointing toward the ATP binding site to be discarded. Furthermore, positions with low occurrence in the sequence alignment (non-robust positions) were also not included. At the end of the workflow, a list of residue positions robustly participating with the ATP binding site was obtained for each conformation. Such information was propagated to kinases without structural data associated. A first application consisted in finding out which kinases have accessible cysteines in their ATP binding site, with the objective to design covalent inhibitors.

**Cysteine Residues Inside the ATP Binding Site.** As a result of the workflow, 33 robust residue positions were obtained to describe the ATP binding site of kinases adopting the active conformation. In such a conformation, 211 proteins (43% of eukaryotic protein kinases, Table SI1 in the Supporting Information) have at least one cysteine in the ATP binding site, distributed over 27 positions (Table 1 and Figure 3A). We then turned our attention to the other conformations. To date, 42% of crystallized human kinases adopt inactive conformations (to be submitted, Leproult et al.). The inactive C-helix-out conformation confers a globally larger ATP pocket than in the active conformation, and six additional cysteine positions were identified (Table 1 and Figure 3B) resulting in 66 kinases with a targetable cysteine (Table SI2 in the Supporting Information). For the inactive DFG-out conformation,

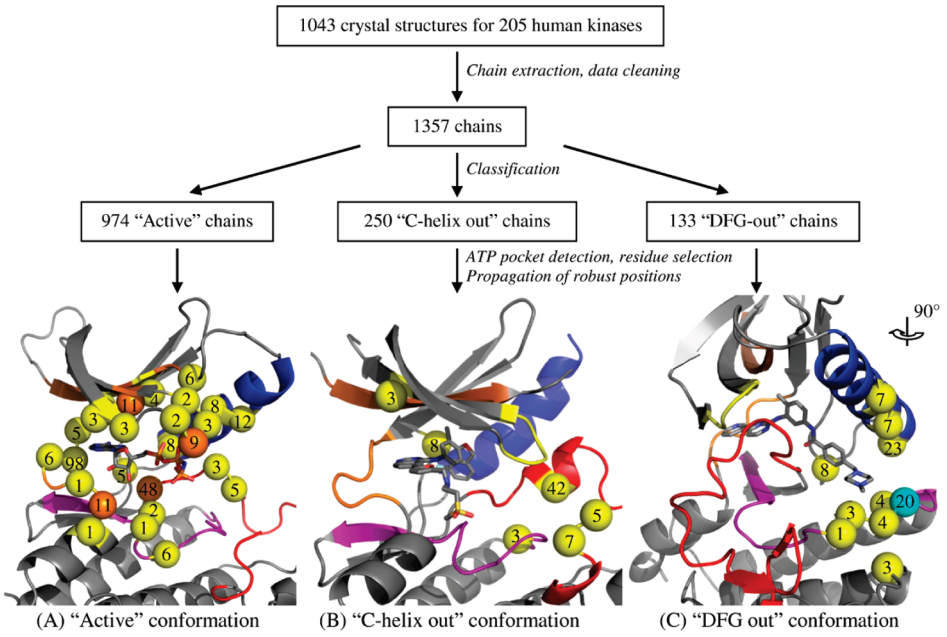



**Figure 3.** Workflow overview and mapping of the cysteine positions with respect to the kinase conformation. The number inside each sphere denotes the number of kinases sharing a given cysteine position. Orange spheres denote previously targeted cysteine positions. (A) Active conformation, 27 cysteine positions; (B) C-helix-out conformation, six additional positions; and (C) DFG-out conformation, 10 additional positions. The position marked with a cyan sphere corresponds to the targeted position in the present work.

**Table 1. Number of Kinases Having Cysteines in the Nucleotide Binding Pocket and Number of Positions Where Cysteines Are Located, Related to the Kinase Conformation**[a]

| | | | Ns | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Conformation | Nc | Nk | AGC | CAMK | CK1 | CMGC | RGC | STE | TK | TKL | OTHER | Nx |
| Active | 27 | 211 (47) | 15 | 36 | 2 | 16 | 5 | 28 | 42 | 13 | 54 | 44 |
| DFG-out | 10* | 127 (7) | 5 | 35 | 1 | 26 | 1 | 19 | 20 | 5 | 15 | 11 |
| C-helix-out | 6* | 66 (2) | 32 | 6 | 1 | 3 | 0 | 10 | 7 | 1 | 6 | 1 |

[a] Nc, number of cysteines positions; *, specific positions of the inactive kinase conformations of which all positions found for the active conformation have to be added; Nk, number of kinases having at least one cysteine (number of kinases having more than one cysteine); Ns, number of kinases per kinase subfamily; and Nx, number of kinases having crystal structures in the desired conformation.

10 additional cysteine positions (Table 1) with respect to the active conformation were found (Figure 3C). These are all located in the extended pocket characteristic of the DFG-out conformation. Globally, 127 kinases have at least one cysteine located in this extension (Table SI3 in the Supporting Information), which is not accessible in the active conformation and provides opportunities that have thus far not been explored.

**Design of a Covalent Kinase Inhibitor Targeting the Inactive DFG-out Conformation.** The analysis of cysteine positioning revealed a unique opportunity to discriminate among kinases targeted by pharmacophores designed for the DFG-out conformation, for example, imatinib[10] (Figure 4B). This conformation has been observed for 16% of the crystallized kinases, and indeed, while imatinib was designed to selectively target Abelson tyrosine kinase (ABL), it is now known also to inhibit two other therapeutic targets:[33,34] the mast/stem cell growth factor receptor (KIT) and the platelet-derived growth factor receptor (PDGFR). Among these three kinases, only KIT and PDGFR kinases bear a cysteine at the beginning of the catalytic loop (position equivalent to Cys788/Cys814 in KIT/PDGFRα kinases, Table SI3 in the Supporting Information), which could be exploited as a second selectivity filter to further refine the selectivity of an inhibitor and provide terminal inhibition. This position is located in the extension of the ATP binding site when the kinase adopts the inactive DFG-out conformation (the position is marked in cyan in Figure 3C).

Two crystal structures of the KIT and ABL kinases exhibiting the DFG-out conformation in complex with imatinib are available in the Protein Data Bank (PDB)[35] (Figure 4A; codes: 1T46 and 2HYY, respectively). The cysteine suggested by our study is solvent exposed and should thus be accessible for covalent bond formation. The striking proximity between this cysteine and the methylpiperazine group of imatinib in the KIT crystal structure suggested that modification of imatinib to include a suitably positioned electrophilic moiety should result in a reaction with cysteine. As shown in Figure 4B, starting from the imatinib heteroaromatic core structure, the aniline was acylated with 4-chloromethyl benzoyl chloride to afford compound **1**. The benzyl chloride was then converted into the benzyl amine (compound **2**) under the action of ammonia and acylated with chloroacetic anhydride, crotyl anhydride, or methylacrylic anhydride to provide compounds **3**, **4**, and **5**, respectively. This

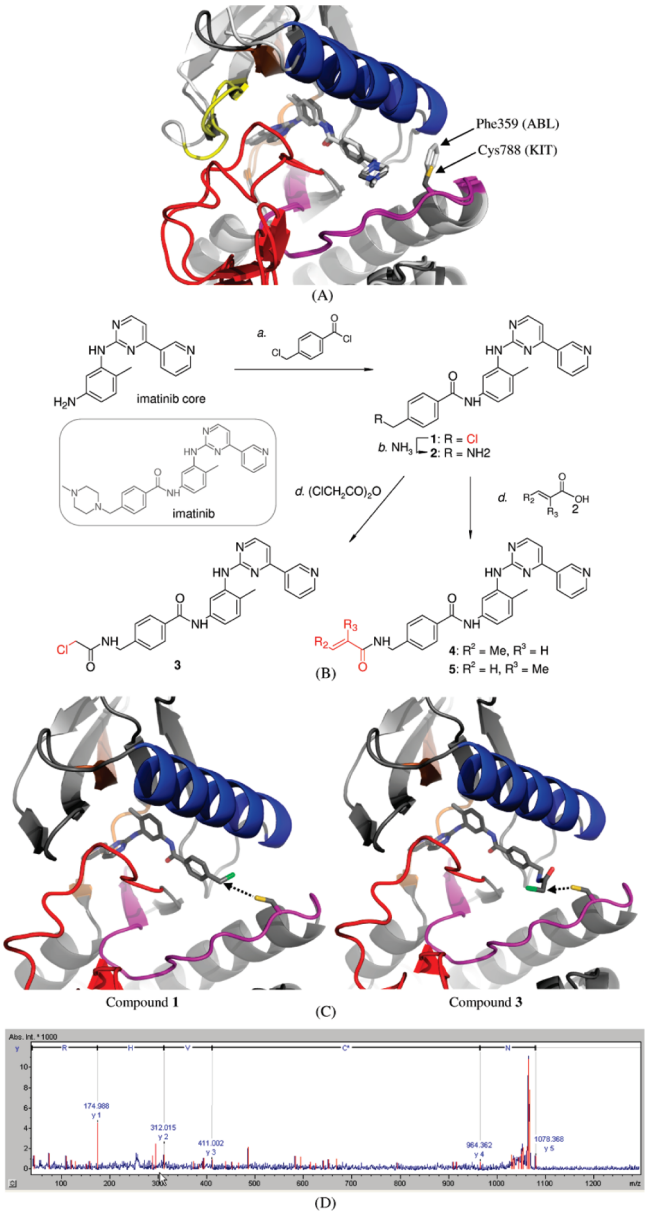



**Figure 4.** Design, synthesis, and characterization of irreversible inhibitors of KIT and PDGFR kinases in the DFG-out conformation. (A) Superimposition of two kinase crystal structures adopting the inactive DFG-out conformation in complex with imatinib: ABL (in light gray; PDB code 2HYY) and KIT (in dark gray; PDB code 1T46). (B) Synthesis of compounds **1**−**5** (electrophilic traps are shown in red). (C) Homology modeling of PDGFR kinase in complex with molecular models of imatinib analogues **1** and **3** bearing different electrophilic traps. (D) Sequencing by MS/MS fragmentation of PDGFR tryptic digest showing the expected shift in molecular weight for the cysteine adduct having reacted with compound **3**.

synthesis thus afforded four compounds with slightly different distances between the heteroaromatic core and the electrophilic trap as well as two different reactivity modes (Michael addition or nucleophilic substitution). While molecular models suggested that compound **3** (Figure 4C) should have the optimal alignment, we reasoned that a certain level of mobility in the side chain may in fact favor other geometries. The inhibition $IC_{50}$ of these five compounds, together with imatinib, was measured for ABL1, KIT, and PDGFRα (Table 2). While this assay measures kinase activity and thus requires the phosphorylated state that favors the active conformation, we reasoned that the comparison between the five compounds and the imatinib would remain relevant and would provide an indication whether the modifications had had a deleterious impact. As the kinase substrate is added simultaneously with the kinase and the assay is allowed to proceed for only 60 min, we did not anticipate a pronounced impact stemming from covalent inhibition. Aside from compound **1**, all other derivatives lost their activity against ABL, presumably due to the loss of ionic interaction between the positively charged methylated nitrogen of the piperazine ring of imatinib[36] and the carbonyl groups of Ile 789 and His 790 (KIT kinase numbering). On the other hand, the compounds containing the different electrophilic traps retained their activity against KIT and PDGFR. To confirm the formation of a covalent adduct between the imatinib analogues and the γ-sulfur of the targeted cysteine residue, tryptic digests of the kinases incubated with the different inhibitors were analyzed by mass spectrometry (MS). Compound **3**, bearing the chloroacetamide group, indeed led to the disappearance of the fragment containing the targeted cysteine residue and to the appearance of a new peak corresponding to the peptide−inhibitor adduct for both KIT and PDGFRα. Further sequencing by MS/MS fragmentation indicated the expected shift in molecular weight (450 amu) for the cysteine adduct having reacted with compound **3** (Figure 4D). Other compounds failed to show similar covalent adducts. This can be rationalized by a less optimal positioning of the electrophile (for example, the angle between compound **1** warhead and the sulfur atom is unfavorable for $S_N2$ substitution; Figure 4C) or a lack of sensitivity in the mass spectrometry assay.

To further assess the specificity of compound **3**, which did form a covalent adduct with KIT and PDGFRα, it was profiled in a competition assay with 440 kinases (376 distinct human kinases and mutants or kinases of pathogenic organisms[33,37]) measuring residual binding of kinases to competitive binders. From this profile, nine kinases [KIT, PDGFRα/β, c-Jun NH2-terminal kinases (JNK) 1−3, epithelial discoidin domain-containing receptor 1 (DDR1), BRAF (V600E), and macrophage colony-stimulating factor 1 receptor (CSF1R)] were identified as having less than 10% residual binding (i.e., 90% binding inhibition; see Table SI4 in the Supporting Information for a complete list). Among the 20 kinases that have a cysteine at the suitable position, 17 were tested, and only four show significant binding to compound **3** (KIT, PDGFRα/β, and CSF1R; Table 3). When comparing the binding of kinases with less than 10% residual binding (i.e., 90% binding inhibition) for either compound **3** or imatinib[33] (Table 4), the modification of the methylpiperazine group for the chloroacetamide affords an improved affinity for the JNK1−3 as well as the mutant of BRAF(V600E) despite the lack of a predictable covalent interaction.

## DISCUSSION

The workflow presented here allowed the analysis of a large number of crystal structures of human protein kinases, which takes into account the conformational diversity within the nucleotide binding pocket. Indeed, depending on the conformation adopted by the kinase catalytic domain, the ATP binding site is more or less extended in different directions, hence providing opportunities to refine inhibitor selectivity.[13] Nevertheless, despite the success of kinase inhibition, there has been a high attrition of kinase inhibitors in clinical development due to off-target toxicity[33] and lack of efficacy.[16,38,39] The preponderance of

**Table 2. $IC_{50}$ Inhibition of Kinase Activity by Imatinib and Five Analogues, of Which Four Have an Electrophilic Moiety**

| | $IC_{50}$ (M) (residue at the position corresponding to the cysteine targeted by electrophilic trap) | | | |
|---|---|---|---|---|
| Compound | ABL1 wt (phenylalanine) | KIT wt (cysteine) | PDGFRα wt (cysteine) | Electrophilic moiety |
| imatinib | $2.7 \times 10^{-6}$ | $3.70 \times 10^{-7}$ | $1.60 \times 10^{-7}$ | no |
| **1** | $6.95 \times 10^{-6}$ | $2.45 \times 10^{-7}$ | $1.39 \times 10^{-7}$ | yes |
| **2** | $>2 \times 10^{-5}$ | $7.88 \times 10^{-7}$ | $1.54 \times 10^{-6}$ | no |
| **3** | $>2 \times 10^{-5}$ | $7.88 \times 10^{-7}$ | $1.08 \times 10^{-6}$ | yes |
| **4** | $>2 \times 10^{-5}$ | $4.71 \times 10^{-6}$ | $4.89 \times 10^{-6}$ | yes |
| **5** | $>2 \times 10^{-5}$ | $2.74 \times 10^{-6}$ | $4.11 \times 10^{-6}$ | yes |

**Table 3. Affinity Profile of Compound 3 for the 20 Kinases Containing the Suitably Positioned Cysteine[a]**

| kinase | compound **3** | imatinib |
|---|---|---|
| AnpA | ND | ND |
| CDKL1 | 100 | ND |
| CDKL4 | ND | ND |
| CSF1R | 5 | 0 |
| Fer | 95 | 91 |
| Fes | 100 | 78 |
| Fgfrl | 100 | 80 |
| Fgfr2 | 92 | 100 |
| Fgfr3 | 100 | 97 |
| Fgfr4 | 100 | 99 |
| Fltl | 86 | 78 |
| Flt3 | 81 | 76 |
| Flt4 | 100 | 85 |
| Jak3 (domain 1) | 100 | 100 |
| Kit | 0 | 0 |
| PDGFRα | 0.7 | 0.15 |
| PDGFRβ | 0 | 0 |
| Pyk2 | 68 | 93 |
| STLK6 | ND | ND |
| Vegfr2 (KDR) | 64 | 57 |

[a] The assay was performed as described[33] using 1 μM compound **3** and compared to the results reported for imatinib (the reported data were obtained using 10 μM imatinib). The data represent residual binding of a competitive binder to the kinases. Thus, the lowest values reflect the strongest inhibition.

**Table 4. Table of Kinases for Which Compound 3 or Imatinib Resulted in Less than 10% Residual Binding of Competitor (i.e., More Than 90% Binding Inhibition)[a]**

| kinase | residue at the position corresponding to the cysteine targeted by electrophilic trap | compound **3** | imatinib |
|---|---|---|---|
| ABL1 | Phe | 67 | **1.4** |
| ABL2 | Phe | 49 | **0.4** |
| BLK | Ser | 69 | **0.8** |
| BRAF (V600E) | Ile | **4.6** | 20 |
| CAMKK1 | Ile | 100 | **1.4** |
| CAMKK2 | Ile | 100 | **2.1** |
| CSF1R | Cys | **5** | **0** |
| DDR1 | Phe | **0.5** | **0.2** |
| DDR2 | Phe | 31 | **7.4** |
| DRAK1 | Val | 100 | **5.2** |
| EPHA8 | Tyr | 100 | **7.2** |
| GAK | Ile | 18 | **5.1** |
| Jnk1 | Ile | **0.45** | 14 |
| Jnk2 | Ile | **4.6** | 75 |
| Jnk3 | Ile | **1.9** | **5.6** |
| Kit | Cys | **0** | **0** |
| Kit (V559D) | Cys | **0** | **0** |
| Kit (V559D, V654A) | Cys | 7 | **0.3** |
| LCK | Tyr | 64 | **1** |
| LYN | Tyr | 76 | **3.2** |
| Mark2 | Ile | 100 | **8.8** |
| Mst2 | Lys | 100 | **4.8** |
| PDGFRα | Cys | **0.7** | **0.15** |
| PDGFRβ | Cys | **0** | **0** |

[a] The assay was performed as described[33] using 1 μM compound **3** and compared to the results reported for imatinib (the reported data were obtained using 10 μM imatinib). The data represent residual binding of a competitive binder to the kinases. Thus, the lowest values reflect the strongest inhibition.

cysteine residues within the nucleotide binding pocket offers the opportunity to rationally design inhibitors capable of terminally inhibiting such kinases. While several inhibitors in clinical development have already highlighted the potential of covalent kinase inhibition, with the most advanced candidates in phase III, reported efforts have been restricted to very few kinases. In addition, a systematic analysis of the cysteine landscape taking into account kinase conformation and solvent accessibility of the cysteine's thiol has not been reported. While the endogenous function of most of these cysteines remains unchartered, at least in the case of the Iκb signaling pathway, one of the natural modulation of the Iκb kinase activity is leveraged on covalent inhibition by a prostaglandin bearing a Michael acceptor.[40]

Clearly, the covalent link between an inhibitor and its target generates long dissociation half-lives,[41] which prolongs efficacy beyond the clearance of the inhibitor, thus potentially allowing a reduction in drug exposure and a decreased risk of off-target effects.[42] Moreover, covalent inhibitors can be efficacious against kinases endowed with high $k_m$ value for ATP,[16,43] by ultimately shifting the equilibrium between the free and the inhibitor-bound forms.[38] Another important advantage of covalent inhibitors is the opportunity to refine their selectivity profile by exploiting unconserved cysteine among kinase family,[15,42] a strategy that is complementary and synergic with other design approaches.

The design of such covalent inhibitors requires the identification of all human kinases having cysteine residues in their ATP binding site. The active kinase conformation gives access to 27 cysteine positions in the ATP binding site (Figure 3A) shared between 211 kinases (Table 1 and Table SI1 in the Supporting Information). The C-helix-out and DFG-out kinase conformations give access to 6 (Figure 3B) and 10 (Figure 3C) additional cysteine positions shared between 66 and 127 kinases, respectively (Table 1 and Tables SI2 and SI3 in the Supporting Information). Kinases bearing equivalent cysteines are not necessarily from the same kinase subfamily. For example, the majority of the 11 kinases sharing the cysteine position equivalent to Cys797 in EGFR kinase (Table SI1 in the Supporting Information) belong to the tyrosine kinase (TK) subfamily, unlike the five kinases sharing the cysteine position equivalent to Leu792 in EGFR kinase (Table SI1 in the Supporting Information), which belong to TK, calcium/calmodulin-dependent kinases (CAMK), AGC, and other subfamilies. Interestingly, the number of kinases having one cysteine in a particular position of the nucleotide binding site is not homogeneous for all positions (Figure 3A−C). While most cysteine positions are populated with more than one kinase, other factors may contribute to the selectivity of inhibitors targeting a particular cysteine residue. For example, broad profiles of the resorcylic acid lactones targeting a cysteine residue present in 48 kinases show a marked selectivity for vascular endothelial growth factor receptors (VEGFRs) and mitogen-activated protein kinase kinases (MEKs). Surprisingly, despite the numerous opportunities offered by cysteines in the nucleotide binding pocket, only a few cysteine positions have thus far been targeted by covalent inhibitors[14,18,29,32] (Figure 2 and Figure 3A).

The results of this structural analysis were exploited to design compounds **1** and **3**−**5** to retune selectivity through covalent trapping. The targeted cysteine appears in the nucleotide binding pocket only when these kinases adopt the inactive DFG-out conformation. Among the kinases that can adopt the required conformation to accommodate the imatinib pharmacophore, only KIT, PDGFRs, and CSF1R have a suitably positioned cysteine residue at the beginning of the catalytic loop. Indeed, compound **3** that forms a covalent adduct with KIT and PDGFR did inhibit their activity but not that of ABL kinase. A broad profile of compound **3**'s affinity against a large panel of kinases reveals that the electrophilic trap did not confer promiscuity for other cysteine-containing kinases (see Figure 5 for graphical representation of selectivity). Indeed, among the kinases bearing the suitably positioned cysteine residue (20 kinases), only a small subset (KIT, PDGFRs, and CSF1R) are known to accommodate the imatinib pharmacophore and were targeted by compound **3**. The substitution of the methylpiperazine group in imatinib for the chloroacetamide in compound **3** led in one case to an improved affinity (JNKs) among the kinases that are known to bind imatinib (over 20), albeit the cellular relevance of such reversible and transient interactions remains to be established. This is the first example of a covalent kinase inhibitor that exploits a cysteine not available in the active conformation in combination with a pharmacophore broadly selective for the DFG-out conformation.

In conclusion, the structural analysis described herein provides a thorough landscape of cysteine positions that can be exploited for irreversible inhibition. Taking into account the different conformations of kinases has revealed previously unrecognized opportunities. The identification of an inhibitor that discriminates between ABL and KIT/PDGFRs illustrates the opportunities to enhance selectivity by harnessing unconserved cysteine residues. Taken together, the results provide a framework to extrapolate this approach to other members of the kinome.

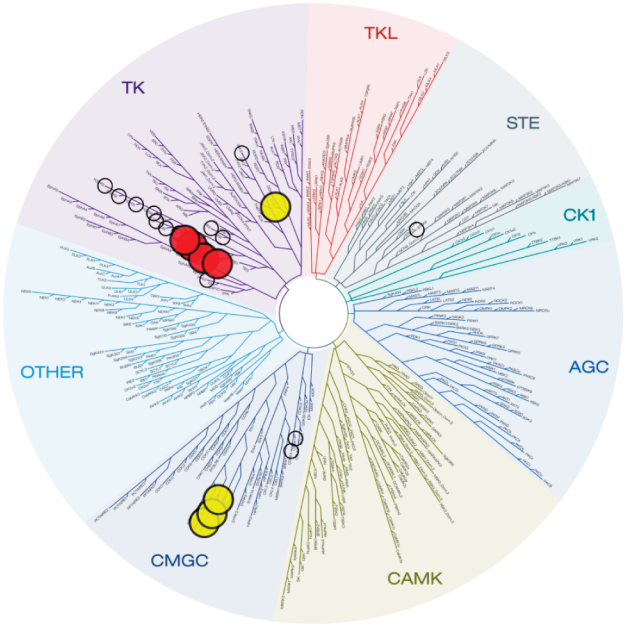



**Figure 5.** Graphical representation of compound **3**'s selectivity. Red- and yellow-filled circles represent covalently and reversibly targeted kinases (less than 10% residual binding of competitor), respectively. Empty circles represent kinases bearing the suitably positioned cysteine but devoid of binding due to pharmacophore specificity.

## EXPERIMENTAL SECTION

**Human Kinase Data Set.** Sequences corresponding to the 491 kinase catalytic domains of the 478 human eukaryotic protein kinases and their sequence alignment file were retrieved from the published data generated by Manning et al.[1] This alignment has been refined for the purpose of this study (available upon request). For each kinase, the corresponding crystal structures were downloaded (July 2010) from the Protein Data Bank[35] via the BIRD system.[44,45] The part of the workflow (Figure 3) explaining the data cleaning of all chains extracted from crystal structures, as well as the classification into conformations, is detailed in Leproult et al. (to be submitted). This led to a total of 1357 chains, of which 974, 250, and 133 chains were classified into active, C-helix-out, and DFG-out conformations, respectively.

**ATP Binding Site Pocket Detection.** Each chain was prepared with the PDB2PQR program[46], which was efficient for automatically removing ligands, solvent, and ions. This was also used to add hydrogens to all atoms in a manner consistent with favorable hydrogen bonding. To automatically detect empty pockets inside the chains, the in-house Pck program[47] was used. This program is dedicated to the geometric-based detection and characterization of pockets. This uses the alpha shape theory[48] to represent the surface of proteins and then implements algorithms specific of some types of pockets. Among these algorithms, CAST[49] was preferentially chosen due to its ability to detect partially buried pockets, as observed for the kinase ATP binding site. Because the classification into conformations required the superimposition of all chains (to be submitted, Leproult et al.), the pocket corresponding to the ATP binding site was located by testing whether it contains a virtually created 3D point placed in the ATP binding site nearby the

hinge region. In the end, only 23 chains were reported as having no detected ATP binding site pocket. Further inspection of these chains revealed that most of them have either a small pocket or a side chain that blocks a part of the ATP binding site, resulting in an autoinhibition phenomenon.

**Amino Acid Selection Inside the ATP Binding Site.** Among the amino acids having atoms participating in the detected ATP pocket, those having at least one atom of the side chain that participates in the pocket are selected for the next part of the workflow. For those that only have main chain atoms that participate in the binding site, the secondary structure information using DSSP program[50] is taken into account. Indeed, if the amino acid is located on a flexible structural element, such as a loop or a turn, then the amino acid is selected. The only peculiar case concerns glycine amino acid when located on a nonflexible structural element. Because a glycine amino acid does not have a side chain, the two hydrogens attached to the $C_\alpha$ atom are investigated. If the one generating an R chirality when mutated to the side chain of a cysteine participates in the detected ATP binding site pocket, then the glycine is selected.

**Robust Amino Acid Positions and Propagation to All Human Kinases.** Each previously selected amino acid is highlighted in the sequence of the corresponding kinase, inside the sequence alignment file.[1] This gives access to an amino acid position. Once the amino acid positions have been obtained for all chains, the ones appearing in more than 30% of all chains, with respect to the conformational class, are considered as robust positions. This step avoids keeping rare positions located in abnormally extended ATP binding sites. Indeed, the detection of some large ATP binding sites by Pck was due to the proximity of the ATP binding site with other domains or the lack of structural elements forming the ATP binding site in the chain. Next, robust amino acid positions are propagated to all human kinases using the sequence alignment. This allows the inference of amino acids having a side chain subject to participation in the ATP binding site for every human kinase in every conformation. The robust cysteine positions are the focus of this study and can be visualized in 3D on the following website: http://lbgi.igbmc.fr/Kinatown.

**Inhibition $IC_{50}$ for ABL1 wt, KIT wt, and PDGFRα wt Kinases.** A radiometric protein kinase assay ($^{33}$PanQinase Activity Assay) was used for measuring the kinase activity of ABL1 wt, KIT wt, and PDGFRα wt kinases (Proqinase, Freiburg, Germany). All kinase assays were performed in 96-well FlashPlates from Perkin-Elmer (Boston, MA) using 50 μL of assay buffer (60 mM HEPES-NaOH, pH 7.5, 3 mM $MgCl_2$, 3 mM $MnCl_2$, 3 μM $Na_3VO_4$, 1.2 mM DTT, 50 μg $mL^{-1}$ PEG2000, and 1 μM [$\gamma$-$^{33}$P]ATP), 20 ng of kinase, and a generic substrate (polyGluTyr for KIT and polyAlaGluLsTyr for ABL and PDGFRα) with 1% DMSO. The test compound concentration ranged from 20 μM to 0.1 nM (semilog dilution). The assays were performed by premixing the ATP solution with the test compound and addition of this solution to the kinase/substrate solution. After 60 min at 30 °C, the reaction was stopped with 50 μL of 2% $H_3PO_4$, plates were aspirated and treated with 200 μL of 0.9% NaCl (2×), and the level of $^{33}$P incorporation was determined with a microplate scintillation counter (MicroBeta, Wallac).

**Mass Spectrometry of KIT Kinase and PDGFRα Kinase Complexes.** A 50 μM solution of each compound was incubated with 2 μg of each kinase in 60 mM Hepes-NaOH, pH 7.5 (3 mM $MgCl_2$ and 3 Mm $MnCl_2$), for 12 h. The proteins were then isolated by one-dimensional sodium dodecyl sulfate−polyacrylamide gel electrophoresis gel, subjected to in-gel trypsin digest, and analyzed by matrix-assisted laser desorption/ionization. As compound **2** cannot form a covalent adduct, this sample was used as a negative control. Only compound **3** showed a new peak corresponding to the expected molecular weight from the peptide adduct (peptide containing Cys 466, MW = 1092.487 for recombinant KIT Kinase, and Cys 491 for recombinant PDGFRα kinase, MW = 1078.472). To further confirm the identity of the peptide sequence of the new peak, MS/MS experiments were performed, which indeed provide the expected degradation product corresponding to the following sequences: KIT = NC(compound **3**)IHR; PDGFRα = NC(compound **3**)VHR.

**Kinase Selectivity Profile.** The selectivity profile was measured by using the KinomeScan technology (http://www.kinomescan.com/) based on active site-dependent competition binding assays.[33,37] The results are expressed as a percentage of signal between a negative (DMSO) and a positive (known binder) control. Residual binding (%) = [(tested compound − positive control)/(tested compound − negative control) × 100].

## ■ ASSOCIATED CONTENT

**Supporting Information.** Tables, figure, synthetic protocols and full characterization of compounds **1**−**5**, peptide sequences, and mass spectrometric characterization of kinase−inhibitor adduct. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

**Corresponding Author**

*(J.-M.W.) Tel: +33 388 655 796. E-mail: wurtz@igbmc.fr. (N.W.) Tel: +33 368 855 112. E-mail: winssinger@unistra.fr.

## ■ ACKNOWLEDGMENT

The Institut Universitaire de France (IUF) and the Institut de Recherche Pierre Fabre (IRPF) are gratefully acknowledged for their support.

## ■ ABBREVIATIONS USED

ABL, Abelson tyrosine kinase; BTK, Bruton's tyrosine kinase; CAMK, calcium/calmodulin-dependent kinases; CDK2, cell division protein kinase 2; CSF1R, macrophage colony-stimulating factor 1 receptor; DFG, aspartate−phenylalanine−glycine; DDR1, epithelial discoidin domain-containing receptor 1; EGFR, epidermal growth factor receptor; FGFR, basic fibroblast growth factor receptor; JNK, c-Jun NH2-terminal kinase; KIT, mast/stem cell growth factor receptor; MEK, mitogen-activated protein kinase kinase; MS, mass spectrometry; PDB, Protein Data Bank; PDGFR, platelet-derived growth factor receptor; RSK, ribosomal protein S6 kinase; TK, tyrosine kinases; VEGFR, vascular endothelial growth factor receptor

## ■ REFERENCES

(1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912–1934.

(2) Bennett, B. L. c-Jun N-terminal kinase-dependent mechanisms in respiratory disease. *Eur. Respir. J.* **2006**, *28*, 651–661.

(3) Dhillon, A. S.; Hagan, S.; Rath, O.; Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **2007**, *26*, 3279–3290.

(4) Wang, Y. Mitogen-Activated Protein Kinases in Heart Development and Diseases. *Circulation* **2007**, *116*, 1413–1423.

(5) Cuenda, A.; Rousseau, S. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochim. Biophys. Acta* **2007**, *1773*, 1358–1375.

(6) Pandya, N.; Santani, D.; Jain, S. Role of mitogen-activated protein (MAP) kinases in cardiovascular diseases. *Cardiovasc. Drug Rev.* **2005**, *23*, 247–254.

(7) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "Gatekeeper Door": Exploiting the Active Kinase Conformation. *J. Med. Chem.* **2010**, *53*, 2681–2694.

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

(9) Scaltriti, M.; Verma, C.; Guzman, M.; Jimenez, J.; Parra, J. L.; Pedersen, K.; Smith, D. J.; Landolfi, S.; Ramon y Cajal, S.; Arribas, J.; Baselga, J. Lapatinib, a HER2 tyrosine kinase inhibitor, induces stabilization and accumulation of HER2 and potentiates trastuzumab-dependent cell cytotoxicity. *Oncogene* **2009**, *28*, 803–814.

(10) Buchdunger, E.; Zimmermann, J.; Mett, H.; Meyer, T.; Müller, M.; Druker, B. J.; Lydon, N. B. Inhibition of the Abl protein-tyrosine kinase in vitro and in vivo by a 2-phenylaminopyrimidine derivative. *Cancer Res.* **1996**, *56*, 100–104.

(11) Andzelm, E. R.; Lew, J.; Taylor, S. Bound to activate: Conformational consequences of cyclin binding to CDK2. *Structure* **1995**, *3*, 1135–1141.

(12) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **2003**, *112*, 859–871.

(13) Liao, J. J. Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *J. Med. Chem.* **2007**, *50*, 409–424.

(14) Cohen, M. S.; Zhang, C.; Shokat, K. M.; Taunton, J. Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science* **2005**, *308*, 1318–1321.

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

(16) Yun, C.; Mengwasser, K. E.; Toms, A. V.; Woo, M. S.; Greulich, H.; Wong, K.; Meyerson, M.; Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2070–2075.

(17) Allen, L. F.; Eiseman, I. A.; Fry, D. W.; Lenehan, P. F. CI-1033, an irreversible pan-erbB receptor inhibitor and its potential application for the treatment of breast cancer. *Semin. Oncol.* **2003**, *30*, 65–78.

(18) Laheru, D.; Croghan, G.; Bukowski, R.; Rudek, M.; Messersmith, W.; Erlichman, C.; Pelley, R.; Jimeno, A.; Donehower, R.; Boni, J.; Abbas, R.; Martins, P.; Zacharchuk, C.; Hidalgo, M. A Phase I Study of EKB-569 in Combination with Capecitabine in Patients with Advanced Colorectal Cancer. *Clin. Cancer Res.* **2008**, *14*, 5602–5609.

(19) Spicer, J.; Rudman, S. EGFR inhibitors in non-small cell lung cancer (NSCLC): The emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. *Targeted Oncol.* **2010**, *5*, 245–255.

(20) For another example (structure undisclosed), see Engelman, J. A.; Zejnullahu, K.; Gale, C.; Lifshits, E.; Gonzales, A. J.; Shimamura, T.; Zhao, F.; Vincent, P. W.; Naumov, G. N.; Bradner, J. E.; Althaus, I. W.; Gandhi, L.; Shapiro, G. I.; Nelson, J. M.; Heymach, J. V.; Meyerson, M.; Wong, K.; Jänne, P. A. PF00299804, an Irreversible Pan-ERBB Inhibitor, Is Effective in Lung Cancer Models with EGFR and ERBB2Mutations that Are Resistant to Gefitinib. *Cancer Res.* **2007**, *67*, 11924–11932.

(21) Cha, M. Y.; Lee, K.; Kim, J. W.; Lee, C. G.; Song, J. Y.; Kim, Y. H.; Lee, G. S.; Park, S. B.; Kim, M. S. Discovery of A Novel Her-1/Her-2 Dual Tyrosine Kinase Inhibitor for the Treatment of Her-1 Selective Inhibitor-Resistant Non-small Cell Lung Cancer. *J. Med. Chem.* **2009**, *52*, 6880–6888.

(22) Rabindran, S. K.; Discafani, C. M.; Rosfjord, E. C.; Baxter, M.; Floyd, M. B.; Golas, J.; Hallett, W. A.; Johnson, B. D.; Nilakantan, R.; Overbeek, E.; Reich, M. F.; Shen, R.; Shi, X.; Tsou, H.; Wang, Y.; Wissner, A. Antitumor Activity of HKI-272, an Orally Active, Irreversible Inhibitor of the HER-2 Tyrosine Kinase. *Cancer Res.* **2004**, *64*, 3958–3965.

(23) Blair, J. A.; Rauh, D.; Kung, C.; Yun, C.; Fan, Q.; Rode, H.; Zhang, C.; Eck, M. J.; Weiss, W. A.; Shokat, K. M. Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **2007**, *3*, 229–238.

(24) Zhou, W.; Ercan, D.; Chen, L.; Yun, C.; Li, D.; Capelletti, M.; Cortot, A. B.; Chirieac, L.; Iacob, R. E.; Padera, R.; Engen, J. R.; Wong, K.; Eck, M. J.; Gray, N. S.; Janne, P. A. Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **2009**, *462*, 1070–1074.

(25) Fry, D. W.; Bridges, A. J.; Denny, W. A.; Doherty, A.; Greis, K. D.; Hicks, J. L.; Hook, K. E.; Keller, P. R.; Leopold, W. R.; Loo, J. A.; McNamara, D. J.; Nelson, J. M.; Sherwood, V.; Smaill, J. B.; Trumpp-Kallmeyer, S.; Dobrusin, E. M. Specific, irreversible inactivation of the epidermal growth factor receptor and erbB2, by a new class of tyrosine kinase inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12022–12027.

(26) Wood, E. R.; Shewchuk, L. M.; Ellis, B.; Brignola, P.; Brashear, R. L.; Caferro, T. R.; Dickerson, S. H.; Dickson, H. D.; Donaldson, K. H.; Gaul, M.; Griffin, R. J.; Hassell, A. M.; Keith, B.; Mullin, R.; Petrov, K. G.; Reno, M. J.; Rusnak, D. W.; Tadepalli, S. M.; Ulrich, J. C.; Wagner, C. D.; Vanderwall, D. E.; Waterson, A. G.; Williams, J. D.; White, W. L.; Uehling, D. E. 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. U.S.A.* **2008**, *105*, 2773–2778.

(27) Pan, Z.; Scheerens, H.; Li, S.; Schultz, B. E.; Sprengeler, P. A.; Burrill, L. C.; Mendonca, R. V.; Sweeney, M. D.; Scott, K. C. K.; Grothaus, P. G.; Jeffery, D. A.; Spoerke, J. M.; Honigberg, L. A.; Young, P. R.; Dalrymple, S. A.; Palmer, J. T. Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem* **2007**, *2*, 58–61.

(28) Ohori, M.; Kinoshita, T.; Yoshimura, S.; Warizaya, M.; Nakajima, H.; Miyake, H. Role of a cysteine residue in the active site of ERK and the MAPKK family. *Biochim. Biophys. Res. Commun.* **2007**, *353*, 633–637.

(29) Jogireddy, R.; Dakas, P.; Valot, G.; Barluenga, S.; Winssinger, N. Synthesis of a Resorcylic Acid Lactone (RAL) Library Using Fluorous-Mixture Synthesis and Profile of its Selectivity Against a Panel of Kinases. *Chem.—Eur. J.* **2009**, *15*, 11498–11506. Dakas, P.-Y.; Barluenga, S.; Totzke, F.; Zirrgiebel, U.; Winssinger, N. Modular Synthesis of Radicicol A and Related Resorcylic Acid Lactones, Potent Kinase Inhibitors. *Angew. Chem., Int. Ed.* **2007**, 6899–6902.

(30) Rastelli, G.; Rosenfeld, R.; Reid, R.; Santi, D. V. Molecular modeling and crystal structure of ERK2-hypothemycin complexes. *J. Struct. Biol.* **2008**, *164*, 18–23.

(31) Schirmer, A.; Kennedy, J.; Murli, S.; Reid, R.; Santi, D. V. Targeted covalent inactivation of protein kinases by resorcylic acid lactone polyketides. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4234–4239.

(32) Zhou, W.; Hur, W.; McDermott, U.; Dutt, A.; Xian, W.; Ficarro, S. B.; Zhang, J.; Sharma, S. V.; Brugge, J.; Meyerson, M.; Settleman, J.; Gray, N. S. A structure-guided approach to creating covalent FGFR inhibitors. *Chem. Biol.* **2010**, *17*, 285–295.

(33) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.

(34) Santarius, T.; Shipley, J.; Brewer, D.; Stratton, M. R.; Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **2010**, *10*, 59–64.

(35) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(36) Aleksandrov, A.; Simonson, T. A molecular mechanics model for imatinib and imatinib:kinase binding. *J. Comput. Chem.* **2010**, *31*, 1550–1560.

(37) Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329–336.

(38) Michalczyk, A.; Klüter, S.; Rode, H. B.; Simard, J. R.; Grütter, C.; Rabiller, M.; Rauh, D. Structural insights into how irreversible

inhibitors can overcome drug resistance in EGFR. *Bioorg. Med. Chem.* **2008**, *16*, 3482–3488.

(39) Blencke, S.; Zech, B.; Engkvist, O.; Greff, Z.; Orfi, L.; Horváth, Z.; Kéri, G.; Ullrich, A.; Daub, H. Characterization of a Conserved Structural Determinant Controlling Protein Kinase Sensitivity to Selective Inhibitors. *Chem. Biol.* **2004**, *11*, 691–701.

(40) Rossi, A.; Kapahi, P.; Natoli, G.; Takahashi, T.; Chen, Y.; Karin, M.; Santoro, M. G. Anti-inflammatory cyclopentenone prostaglandins are direct inhibitors of I[kappa]B kinase. *Nature* **2000**, *403*, 103–118.

(41) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730–739.

(42) Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009**, *52*, 225–233.

(43) Knight, Z. A.; Shokat, K. M. Features of Selective Kinase Inhibitors. *Chem. Biol.* **2005**, *12*, 621–637.

(44) Toursel, T.; Bard, N.; Bolze, R.; Caron, E.; Desprez, F.; Heymann, M.; Friedrich, A.; Moulinier, L.; Nguyen, N. H.; Poch, O. Décrypthon grid - grid resources dedicated to neuromuscular disorders. *Stud. Health Technol. Inform.* **2010**, *159*, 124–133.

(45) BIRD 2010; http://alnitak.u-strasbg.fr/wikili/index.php/BIRD.

(46) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson−Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

(47) Albou, L.; Schwarz, B.; Poch, O.; Wurtz, J. M.; Moras, D. Defining and characterizing protein surface using alpha shapes. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 1–12.

(48) Edelsbrunner, H.; Facello, M.; Liang, J. On the definition and the construction of pockets in macromolecules. *Pac. Symp. Biocomput.* **1996**, 272–287.

(49) Liang, J.; Edelsbrunner, H.; Woodward, C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884–1897.

(50) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.