Evolutionary divergence in the conformational landscapes of tyrosine vs serine/threonine kinases

Joan Gizzio¹,²†, Abhishek Thakur¹,²†, Allan Haldane¹,³, Ronald M Levy¹,²*

¹Center for Biophysics and Computational Biology, Temple University, Philadelphia, United States; ²Department of Chemistry, Temple University, Philadelphia, United States; ³Department of Physics, Temple University, Philadelphia, United States

Abstract Inactive conformations of protein kinase catalytic domains where the DFG motif has a “DFG-out” orientation and the activation loop is folded present a druggable binding pocket that is targeted by FDA-approved ‘type-II inhibitors’ in the treatment of cancers. Tyrosine kinases (TKs) typically show strong binding affinity with a wide spectrum of type-II inhibitors while serine/threonine kinases (STKs) usually bind more weakly which we suggest here is due to differences in the folded to extended conformational equilibrium of the activation loop between TKs vs. STKs. To investigate this, we use sequence covariation analysis with a Potts Hamiltonian statistical energy model to guide absolute binding free-energy molecular dynamics simulations of 74 protein-ligand complexes. Using the calculated binding free energies together with experimental values, we estimated free-energy costs for the large-scale (~17–20 Å) conformational change of the activation loop by an indirect approach, circumventing the very challenging problem of simulating the conformational change directly. We also used the Potts statistical potential to thread large sequence ensembles over active and inactive kinase states. The structure-based and sequence-based analyses are consistent; together they suggest TKs evolved to have free-energy penalties for the classical ‘folded activation loop’ DFG-out conformation relative to the active conformation, that is, on average, 4–6 kcal/mol smaller than the corresponding values for STKs. Potts statistical energy analysis suggests a molecular basis for this observation, wherein the activation loops of TKs are more weakly ‘anchored’ against the catalytic loop motif in the active conformation and form more stable substrate-mimicking interactions in the inactive conformation. These results provide insights into the molecular basis for the divergent functional properties of TKs and STKs, and have pharmacological implications for the target selectivity of type-II inhibitors.

Editor’s evaluation

This important paper provides a convincing mechanism for the relative binding specificity of Type II inhibitors to kinases. The combination of a sequence-derived Potts model with experimental dissociation constants and calculated free energies of binding to the DFG-out state is highly compelling and goes beyond the current state-of-the-art. Given the importance of kinases in pathophysiological processes, the results will be of interest to a broad audience and, in addition, the combination of computational methods can be applicable to a wide variety of other biophysical processes that involve conformational rearrangements.
Introduction

The human genome contains approximately 500 eukaryotic protein kinases which coordinate signaling networks in cells by catalyzing the transfer of a phosphate group from ATP to serine, threonine, or tyrosine residues (Manning, 1995; Modi and Dunbrack, 2019a). The GO (gene ontology) database identifies 351 (~70%) of these enzymes as serine/threonine kinases (STKs) and 90 (~18%) as tyrosine kinases (TKs). STKs are an ancient class of protein kinases that predate the divergence of the three domains of life (bacteria, archaea, eukaryote) (Stancik et al., 2018), whereas TKs are a more recent evolutionary innovation, having diverged from STKs about 600 million years ago during early metazoan evolution (Miller, 2012; Sebé-Pedrós et al., 2016). Kinases are important therapeutic targets in a large number of human pathologies and cancers. Both TKs and STKs share a striking degree of structural similarity in their catalytic domains, owing to evolutionary selective pressures that preserve their catalytic function; in particular, the location and structure of the ATP binding site are highly conserved which raises significant challenges for the design of small-molecule ATP-competitive inhibitors that are both potent for their intended target(s) and have low off-target activity for unintended kinase targets. The latter is referred to as the ‘selectivity’ of an inhibitor, a property which is difficult to predict and control but is nonetheless very important for developing drugs with minimal harmful side effects.

A particular class of ATP-competitive kinase inhibitors which were proposed to have a high potential for selectivity are called ‘type-II inhibitors’ which only bind when the kinase adopts an inactive ‘DFG-out’ conformation. ‘DFG’ (Asp-Phe-Gly) refers to a conserved catalytic motif located at the N-terminus of an ~20 residue-long ‘activation loop’ that is highly flexible and controls the activation state of the kinase and the structure of the substrate binding surface. The precise arrangement of catalytic residues and the structural organization of large regulatory elements, such as the activation loop and nearby ‘αC-helix’, are strongly coupled to the conformation of the DFG motif and the DFG-1 residue preceding it, which is well described by regions on the Ramachandran map occupied by the Asp, Phe, and DFG-1 residues (beta-turn, right-handed alpha-helix, left-handed alpha-helix) and the $\chi_1$ rotamer state of the DFG-Phe sidechain (trans, gauche-minus, gauche-plus). Recently, Dunbrack and co-workers identified eight major conformational states in the Protein Data Bank (PDB) based on these metrics (Modi and Dunbrack, 2019b). The most common state, which is evolutionarily conserved in all kinases, corresponds to the active ‘DFG-in’ conformation. In this conformation all structural requirements for catalysis are typically met, e.g., a complete hydrophobic spine, a salt bridge between the conserved β3-Lys and αC-Glu residues, and an extended activation loop which forms the substrate binding surface. Inactive kinases in the PDB are most frequently seen in an ‘Src-like inactive’ conformation where the DFG is ‘in’, but the αC-helix is swung outward, breaking the β3-Lys → αC-Glu salt bridge and disassembling the hydrophobic spine. Disassembly of the hydrophobic spine caused by αC-helix rotation increases the cavity volume around the DFG-Phe residue, allowing it to pass through the Src-like inactive conformation and completely ‘flip’ from DFG-in to DFG-out (Levinson et al., 2006; Shan et al., 2013). The classical DFG-out conformation, targeted by type-II inhibitors, displays a highly reorganized activation loop that is folded away from the αC-helix, projecting toward solvent or forming stable secondary structure and substrate-mimicking interactions. We refer to these states of the activation loop collectively as ‘folded’, to describe its ~17 Å reorganization relative to the active ‘extended’ conformation, wherein the substrate binding surface has been ‘folded up’ toward the kinase N-terminal lobe and away from the αC-helix.

In both TKs and STKs, the activation loop undergoes this large-scale conformational change when the DFG motif flips from the active ‘DFG-in’ conformation to the classical DFG-out conformation. The DFG flip swaps the positions of DFG-Phe and DFG-Asp, opening a hydrophobic ‘back pocket’ that is connected to the conserved ATP binding site through the ‘gatekeeper’ residue. Type-II inhibitors typically have a long chemical fragment that allows them to bind across the gatekeeper and form interactions with residues in the back pocket. In contrast, type-I inhibitors (the majority of kinase drugs) occupy the ATP pocket but not the back pocket and can bind to either DFG-in or DFG-out. For these reasons, it has been proposed that type-II inhibition holds greater potential for the design of highly selective drugs (Vijayan et al., 2015; Davis et al., 2011; Anastassiadis et al., 2011); it has been shown that different kinase sequences have different propensities to adopt DFG-out in the absence of inhibitor (Haldane et al., 2016; Hari et al., 2013), and the DFG-out back pocket has been suggested to have a lesser degree of sequence and structural homology between kinases (Liu and Gray, 2006). However, the notion that type-II inhibitors developed to-date are more selective than type-I inhibitors...
has been brought into question (Zhao et al., 2014; Klaeger et al., 2017), suggesting that further consideration of the energetic contributions described above is required.

In order to fully exploit the target-selective potential of type-II inhibitors it is necessary to understand the underlying sequence-dependent principles that control the conformational preferences of their kinase targets, and the extent to which this has been diversified by evolution. This can, in principle, be directly approached using free-energy simulations to estimate the reorganization free-energy required for different kinases to adopt DFG-out, although this is computationally very expensive and of uncertain reliability for conformational changes involving large-scale loop reorganizations, such as the ~17 Å ‘folding’ of the activation loop that accompanies the transition from active DFG-in to the inactive, classical DFG-out state. To accommodate this limitation, we employ modern sequence-based computational methods to characterize the conformational selection process over the entire kinome and combine the sequence-based results with structure-based free energy simulations with the goal of identifying evolutionarily divergent features of the energy landscape that control the preference of individual kinases for the active (DFG-in) vs inactive (DFG-out ‘folded activation loop’) states. To this end, we report evidence that TK catalytic domains have a molecular evolutionary bias that shifts their conformational equilibrium toward the inactive ‘folded activation loop’ DFG-out state in the absence of activation signals. In contrast, STKs as a class have a more stable active conformation which is favored over the DFG-out state due to sequence constraints in the absence of other signals.

As described below, our analysis of a previously published kinome-wide assay suggests that TKs have properties which privilege the binding of type-II inhibitors in comparison to STKs, which leads us to hypothesize an evolutionary divergence in their conformational energy landscapes. To investigate this, we used a Potts Hamiltonian statistical energy model derived from residue-residue covariation in a multiple sequence alignment (MSA) of protein kinase sequences to probe the active DFG-in ↔ classical DFG-out conformational equilibrium as previously described (Haldane et al., 2016). Using an approach that involves ‘threading’ a large number of kinase sequences onto ensembles of active DFG-in and classical DFG-out structures from the PDB and scoring them using the Potts Hamiltonian, we are able to view the evolutionary divergence in TK and STK conformational landscapes. This calculation only probes the free-energy difference between the active DFG-in and classical DFG-out conformations, and by construction does not consider alternative conformations (e.g. ‘Src-like inactive’) that might be important for analyzing the type-II binding pathway. As discussed below, the Potts calculations from this two-state model correlate well with the free-energy cost to adopt the classical DFG-out conformation.

To validate our results, we used the Potts statistical energy threading calculations to guide target selection for a set of more computationally intensive free-energy simulations. These simulations use type-II inhibitors as tools to probe kinase targets that have already reorganized to DFG-out, allowing us to estimate the free-energy of reorganization ($\Delta G_{\text{reorg}}$) as the excess between the absolute binding free-energy (ABFE) calculated from simulations and the standard binding free-energy measured experimentally in vitro, which already includes the cost to reorganize. Although our methods avoid sampling the conformational change directly, we show how important structural determinants of the conformational change can be identified by analyzing residue-pair contributions to the Potts threading calculations, enabling us to reason about the molecular evolutionary basis for the differences in conformational behavior observed for TKs and STKs.

## Results

### Insights into the sequence-dependent conformational free-energy landscape

The binding of type-II inhibitors is achieved once a protein kinase has reorganized to the DFG-out with activation loop folded conformation (classical DFG-out). We sought initial insight into the conformational equilibrium from type-II binding data available publicly in the form of literature-reported dissociation constants ($K_d$). From the binding assay reported by Davis et al., 2011, we report a ‘hit’ where an inhibitor binds to a kinase with $K_d \leq 10$ μM. Using this criterion, a type-II inhibitor hit rate was calculated for each kinase. Analysis of the type-II hit rate distributions for STKs and TKs from the Davis assay (Figure 1A) indicates that STKs, on average, have an unfavorable contribution to the binding of type-II inhibitors relative to TKs.
Figure 1. Viewing the conformational landscape of the human kinome. (A) Hit rate distributions from kinome-wide experimental binding assays with type-II inhibitors for human serine/threonine kinases (STKs; blue, top) and tyrosine kinases (TKs; orange, bottom) with small gatekeepers (solid bars; sidechain volume <110 Å³) and large gatekeepers (hatched bars; sidechain volume >110 Å³). (B) PyMol (pymol, 2015) visualization of two conformational ensembles populated by Abl kinase from recent solution NMR (Nuclear Magnetic Resonance) experiments (Xie et al., 2020). The active DFG-in conformation where the activation loop is ‘extended’ (red, the Protein Data Bank [PDB]: 6XR6) and the inactive classical DFG-out conformation where the activation loop is ‘folded’ (blue, PDB: 6XRG) both exist in the absence of ligands, but there is a free-energy cost to transition between them (Xie et al., 2020). Type-II inhibitors preferentially bind to this folded DFG-out state. (C) Correlation between Potts DFG-out penalty (ΔE Potts) and hit rates for kinases with small gatekeepers only, to control for gatekeeper effects (Pearson correlation of –0.59, p<0.001). (D) Potts DFG-out penalties calculated for the human kinome and plotted using CORAL (Metz et al., 2018); the TK branch appears to have lower penalties relative to the rest of the kinome, which represent STKs. See Figure 1—source data 1 for values of the calculated type-II hit rates and Potts threaded-energy scores over the human kinome.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Figure supplement 1. Conformational landscape of the activation loop derived from the Protein Data Bank (PDB).

Figure supplement 2. Contact frequency differences between the active DFG-in (BLAminus) and classical DFG-out (BBAminus) conformations.

Figure 1 continued on next page
The size of the gatekeeper residue is important for type-II binding as it controls access to a hydrophobic pocket adjacent to the ATP binding site that is traversed by type-II inhibitors (Zuccotto et al., 2010; Bosc et al., 2015; Liu et al., 1998; Ghose et al., 2008; van Linden et al., 2014), and the size of the gatekeeper residue is thought to negatively affect type-II binding (Zuccotto et al., 2010; Azam et al., 2008; Lovera et al., 2015; Yun et al., 2008). Because TKs tend to have small gatekeepers in comparison to STKs (Zuccotto et al., 2010; Taylor and Kornev, 2011), we considered this as a possible explanation behind the bias for TKs to have larger type-II hit rates. By plotting the hit rate distributions for STKs and TKs where the gatekeeper is either small or large (Figure 1A) we confirm that gatekeeper size has an important influence on type-II binding for both STKs and TKs (Azam et al., 2008). However, the hit rate distribution for TKs appears more sensitive to gatekeeper size than STKs. Even with small gatekeepers, there is a significant fraction of STKs that have hit rates of zero compared with TKs, suggesting the difference in hit rates between TKs and STKs cannot be accounted for primarily by the size of the gatekeeper residue.

Recent solution NMR experiments with Abl kinase revealed two DFG-out conformational states (Xie et al., 2020) one where the DFG motif has flipped ‘DFG-in to DFG-out’ but the activation loop remains in a ‘minimally perturbed’ active-like conformation, and the other state is a classical ‘folded’ DFG-out conformation where the activation loop has moved ~17 Å away from the active conformation (Figure 1B), and the DFG motif is in a ‘classical DFG-out’ (Vijayan et al., 2015) or ‘BBAminus’ (Modi and Dunbrack, 2019b) state. Type-II inhibitors were shown to preferentially bind to this folded DFG-out state, confirming observations that Abl is almost always co-crystallized with type-II inhibitors in this conformation. This binding behavior is also exhibited by other kinases, suggested by the large number of activation loop folded DFG-out states seen in type-II bound co-crystal structures (Figure 1—figure supplement 1). Hence, the importance of large-scale activation loop conformational changes in type-II binding and the large number of residue-residue contact changes involved in this transition (Figure 1—figure supplement 2) suggests the sequence variation of the activation loop and the catalytic loop with which it interacts, might contour the conformational landscape differently for TKs compared with STKs. To investigate this, we used a Potts statistical energy model of sequence covariation to estimate the energetic cost of the active DFG-in (activation loop extended) → inactive DFG-out (activation loop folded) transition for human TKs and STKs (see Methods).

Patterns of coevolution of amino acids at different positions in an MSA are thought to largely reflect fitness constraints for fold stability and function between residues close in 3D space (Lapedes et al., 2012; Hopf et al., 2015; Hopf et al., 2017; Marcos et al., 2014), and these coevolutionary interactions can be successfully modeled by a Potts Hamiltonian (Weigt et al., 2009; Lunt et al., 2010) which we inferred using Mi3-GPU, an algorithm designed to solve ‘Inverse Ising’ problems for protein sequences with high accuracy (Haldane and Levy, 2021). The pairwise interactions from the Potts model can be used as a simple threaded energy function to estimate energetic differences between two conformations, based on changes in residue-residue contacts in the PDB (Haldane et al., 2016). We have calculated the threading penalty for all kinases in the human kinome. Our calculations show the Potts predicted DFG-out penalty (\(\Delta E_{\text{Potts}}\)), which is dominated by large-scale reorganization of the activation loop to the folded DFG-out state, is correlated with type-II hit rates (Figure 1C) when controlling for gatekeeper size. From this, we determine that sequence variation of the activation loop and the contacts broken/formed by its large-scale conformational change (Figure 1—figure supplement 2) makes an important contribution to the binding affinity of type-II inhibitors.

Notably, our calculations over the entire human kinome show that the large majority of kinases with large \(\Delta E_{\text{Potts}}\) (unfavorable conformational penalties) are STKs, and the large majority of low-penalty kinases are TKs (Figure 1D). To validate this finding, we next perform an independent and more computationally intensive prediction of the conformational reorganization energy of TKs and

Source data 1. Type-II inhibitor hit rates and Potts threaded-energy penalties for DFG-out calculated for tyrosine kinases and serine/threonine kinases from the human kinome.
STKs for select kinase targets, chosen based on the kinome calculations of \( \Delta E_{Potts} \), and type-II hit rates shown in Figure 1, in which we use type-II inhibitors as probes in ABFE simulations as described in the following section. By comparing the conformational penalties predicted from these structure-based molecular dynamics (MD) free-energy simulations with the Potts conformational penalty scores, we also identify the scale of \( \Delta E_{Potts} \) in physical free-energy units. This allows us to predict physical conformational free energies based on Potts calculations which can be carried out at scale on large numbers of sequences, to evaluate the evolutionary divergence of the conformational penalty between STKs and TKs generally.

**Structure-based free-energy simulations guided by the sequence-based Potts model**

Relative binding free-energy simulations are now widely employed to screen potent inhibitors in large-scale drug discovery studies (Wang et al., 2015). These methods are used to determine the relative free energy of binding between ligands that differ by small substitutions, which permit one to simulate along an alchemical pathway that mutates one ligand to another. By leaving the common core scaffold unperturbed, the cost and difficulty of sampling the transition between unbound (apo) and bound (holo) states of the system are avoided (Wang et al., 2015; Hayes et al., 2022; Guest et al., 2022). Alternatively, alchemical methods to determine ABFEs, such as the ‘double decoupling’ method employed in this work, sample the apo — holo transition along a pathway that decouples the entire ligand from its environment. While more computationally expensive, the advantage of ABFE is that the computed \( \Delta G_{bind} \) can be directly compared with experimental binding affinities, and successful convergence does not rely on the structural similarity of compounds being simulated (Cournia et al., 2020; Li et al., 2020; Heinzelmann and Gilson, 2021; Lee et al., 2020; Gilson et al., 1997; Qian et al., 2019; Sun et al., 2022).

Our alchemical ABFE simulations of type-II inhibitors binding to TKs and STKs simulate the apo and holo states of the kinase domains in the classical DFG-out conformation with the activation loop folded, starting from the experimentally determined co-crystal structure of the holo state. The apo state remains DFG-out with the activation loop folded throughout the simulations, and therefore the calculated ABFE (\( \Delta G_{bind}^{ABFE} \)) excludes the cost to reorganize from DFG-in (\( \Delta G_{reorg} \)). On the other hand, standard binding free-energies (\( \Delta G_{exp}^{o} \)) determined experimentally from inhibition or dissociation constants (Equation 1) implicitly include the free-energy cost to reorganize. Therefore given the experimentally determined total binding free energy, \( \Delta G_{exp}^{o} \), ABFE simulations can be used to separate the free energy of ligand-receptor association in the inactive state (\( \Delta G_{bind}^{ABFE} \)) from the cost to reorganize from the active to inactive state, \( \Delta G_{reorg} \) (Equation 3; Deng et al., 2011; Lin et al., 2014; Lin et al., 2013).

We calculated \( \Delta G_{exp}^{o} \) (Equation 1) from literature reported IC\(_{50}\) or \( K_d \) values, where the standard concentration \( C_0 \) is set to 1 M

\[
\Delta G_{exp}^{o} = k_B T \ln \left( K_d / C_0 \right) \quad (1)
\]

\( \Delta G_{exp}^{o} \) can be expressed as the sum of the free-energy change to reorganize from the active to inactive state, \( \Delta G_{reorg} \) plus the free energy to bind to the inactive state \( \Delta G_{bind}^{ABFE} \) (Equation 2). \( \Delta G_{reorg} \) is therefore the excess free-energy difference between \( \Delta G_{exp}^{o} \) and \( \Delta G_{bind}^{ABFE} \) (Equation 3).

\[
\begin{align*}
\Delta G_{exp}^{o} &= \Delta G_{reorg} + \Delta G_{bind}^{ABFE} \\
\Delta G_{reorg} &= \Delta G_{exp}^{o} - \Delta G_{bind}^{ABFE}
\end{align*}
\]

Type-II inhibitors generally bind when the activation loop is in a folded DFG-out conformation (Figure 1B), which presents major challenges for direct simulations to determine the free energy cost of the conformational change in contrast to the method employed here (Equation 3).

Because the type-II inhibitor imatinib is co-crystallized in a type-II binding mode with MAPK14 (p38α), an STK, and several other TKs (e.g. ABL1, DDR1, LCK, CSF1R, KIT, and PDGFRA), we chose this inhibitor as an initial probe of our hypothesis that TKs evolved to have lower \( \Delta G_{reorg} \) than STKs (Figure 2). In this example we note that TKs bind strongly to imatinib (‘STI’ in Figure 2) with an average \( \Delta G_{exp}^{o} \) of –9.3 kcal/mol, in contrast to the STK MAPK14 which binds this drug very weakly
\( \Delta G_{\text{exp}} = -6.1 \text{ kcal/mol} \). At face value this appears consistent with our analysis from Figure 1D, where we calculated a large Potts DFG-out penalty for MAPK14 (\( \Delta E_{\text{Potts}} = 5.2 \)) and low penalties for TKs, suggesting that the weak binding of imatinib to MAPK14 is due at least partially to large \( \Delta G_{\text{reorg}} \). To confirm this, we used ABFE simulations with the imatinib: MAPK14 complex to evaluate Equation 3, confirming that MAPK14 incurs a large penalty to adopt the DFG-out conformation with the activation loop folded (\( \Delta G_{\text{reorg}} = 5 \text{ kcal/mol} \)) (Figure 2).

Despite the large \( \Delta G_{\text{reorg}} \) predicted for MAPK14 by both the Potts model and the simulations with imatinib described above, highly potent type-II inhibitors have been successfully developed for this kinase. For example, BIRB-796 (Pargellis et al., 2002) binds to MAPK14 about 7 kcal/mol more strongly than imatinib. This stronger binding of BIRB-796 is captured by \( \Delta G_{\text{ABFE}}^{\text{bind}} \) from our simulations (Figure 2), and the calculated value of \( \Delta G_{\text{reorg}} \) for this complex (\( \Delta G_{\text{reorg}} \approx 4 \text{ kcal/mol} \)) is very close to the corresponding estimate of \( \Delta G_{\text{reorg}} \) based on simulations with imatinib (Figure 2). Importantly, this result suggests that STKs can be potently inhibited by type-II inhibitors despite their large \( \Delta G_{\text{reorg}} \).

To support this, we performed additional ABFE simulations with BIRB-796 and calculated \( \Delta G_{\text{reorg}} \) for two additional STKs predicted to have large reorganization penalties (MAPK9 and BRAF, \( \Delta E_{\text{Potts}} < 1 \)). We calculated \( \Delta G_{\text{reorg}} > 8 \text{ kcal/mol} \) for MAPK9 and BRAF, which is consistent with predictions from the Potts model, and comparison of \( \Delta G_{\text{exp}} \) and \( \Delta G_{\text{ABFE}}^{\text{bind}} \) in Figure 2 confirms that BIRB-796 is able to overcome the large \( \Delta G_{\text{reorg}} \) of certain kinases to attain high experimental potencies (e.g. MAPK14 and MAPK9). To further validate this result, we calculated \( \Delta G_{\text{reorg}} \) via ABFE simulations of BIRB-796 binding to a TK predicted by the Potts model to have a low penalty (PTK2B, \( \Delta E_{\text{Potts}} < 1 \)), which again shows consistency with our Potts prediction of the conformational landscape (Figure 2). The relatively weak value of \( \Delta G_{\text{ABFE}}^{\text{bind}} \) for this kinase compared with MAPK14 is also consistent with observations.

Figure 2. Overview of the conformational landscapes between serine/threonine kinases (STKs) and tyrosine kinases (TKs) from absolute binding free-energy simulations, where we compare \( \Delta G_{\text{bind}} \) (hatched bars) from binding free-energy simulations with \( \Delta G_{\text{exp}} \) (solid bars) for the type-II inhibitors imatinib (the Protein Data Bank [PDB] code: STI) and BIRB-976 (PDB code: B96) vs several TKs (orange) and STKs (blue).

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Figure supplement 1. Absolute binding free-energy (ABFE) benchmarking results.

Source data 1. Absolute binding free-energy results for kinases bound to imatinib and BIRB-796.
of the BIRB-796: PTK2B co-crystal structure (PDB: 3FZS), where the binding mode in PTK2B is more weakly associated with the ATP pocket in comparison with MAPK14 (Han et al., 2009).

The analysis above provides initial support for our hypothesis about the evolutionarily divergent STK and TK conformational landscapes. To further develop this approach, we identified five STKs and five TKs which are predicted by the Potts threading calculations to have large and small $\Delta G_{\text{reorg}}$, respectively, and for which there are sufficient experimental structural and inhibitory data (co-crystal structures and binding constants) to calculate an average $\Delta G_{\text{reorg}}$ via Equation 3 for each target using multiple type-II inhibitor probes. For each of the TK and STK targets, these sets of calculations can be visualized as a linear regression of $\Delta G_{\text{exp}}$ vs $\Delta G_{\text{ABFE}}^{\text{bind}}$ where the slope is constrained to one, consistent with Equation 2 (see Methods for details). We employed this workflow for the set of five TK and five STK targets by simulating 22 and 23 type-II inhibitor complexes, respectively.

Figure 3. Using type-II inhibitors as tools to probe the conformational landscape of TKs and STKs. (A) The average $\Delta G_{\text{reorg}}$ calculated via absolute binding free-energy (ABFE) simulations with 23 type-I (stars) and 23 type-II inhibitors (circles) complexes in the active DFG-in and inactive DFG-out state, respectively, computed from five serine/threonine kinase (STK) targets (Table 1) and (B) computed with 22 type-II inhibitors vs five tyrosine kinase (TK) targets in the DFG-out state (Table 1). (C) Kinome plot created with CORAL (Metz et al., 2018) illustrating the selection of five TKs and five STKs which are detailed in Table 1.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Absolute binding free-energy results for five tyrosine kinases and five serine/threonine kinases bound to type-II and type-I inhibitors.

Figure supplement 1. Experimental binding free-energy vs calculated absolute binding free-energy for type-I inhibitors vs serine/threonine kinases (STKs; column A), type-II inhibitors vs STKs (column B), and type-II inhibitors vs tyrosine kinases (TKs; column C).
The result of this workflow for the set of five TKs and their type-II complexes revealed a low average $\Delta G_{\text{reorg}}$ of <1 kcal/mol (Figure 3B), consistent with our initial predictions from Potts $\Delta E_s$ and type-II hit rates (Table 1). On the other hand, the binding free-energy simulations for the set of five STKs and their type-II complexes show an average of ~6 kcal/mol of $\Delta G_{\text{reorg}}$ is required for these kinases to adopt DFG-out conformation, which is also consistent with our initial predictions from the Potts model (Table 1). To verify that the large $\Delta G_{\text{reorg}}$ identified for STKs is a property of conformational selection for DFG-out rather than systematic overestimation of $\Delta G_{\text{ABFE}}^{\text{bind}}$ for these kinases, we performed ABFE simulations of type-I inhibitors binding to the same set of STKs (an additional 23 complexes). For the binding of type-I inhibitors, we expect there to be no reorganization penalty due to the lack of DFG-out conformational selection in their binding mechanism. As anticipated, the calculated values of $\Delta G_{\text{ABFE}}^{\text{bind}}$ for type-I inhibitors are very close to their experimental binding affinities ($\Delta G_{\text{exp}}^{\text{bind}}$) (Figure 3A).

We find that the set of type-II inhibitors complexed with STKs in this dataset tends to have more favorable binding free energies to the reorganized receptor ($\Delta G_{\text{ABFE}}^{\text{bind}}$) than type-II inhibitors complexed with TKs, as shown by their distribution along the horizontal axes in Figure 3. The reason for this can be understood as a consequence of selection bias. Our selection of STK complexes for this study usually involved lead compounds from the literature, which were designed for high on-target experimental potency and published for their pharmaceutical potential, similar to BIRB-796: MAPK14 which is a tightly bound complex with high experimental affinity despite the large $\Delta G_{\text{reorg}}$ incurred by this kinase (Figure 2). This tight binding is reflected by the favorability of the $\Delta G_{\text{ABFE}}^{\text{bind}}$ term which must be implicitly tuned by medicinal chemists to overcome the large $\Delta G_{\text{reorg}}$ found in STKs. Meanwhile, the chemical space of type-II inhibitors studied against TKs appears to be privileged by their low $\Delta G_{\text{reorg}}$, judging by the comparably weak $\Delta G_{\text{ABFE}}^{\text{bind}}$ for these complexes. This ultimately gives rise to similar experimental potencies for the binding of type II inhibitors to TKs and STKs plotted in Figure 3.

The results of the MD binding free-energy simulations when combined with experimental binding affinities, reveal significant differences in the conformational free-energy landscapes between STKs and TKs. The DFG-in (activation loop extended) to DFG-out (activation loop folded)

Table 1. Calculation of reorganization free-energy for five TKs and five STKs. Type-II hit rates from Davis et al. and Potts threaded energy penalties from Figure 1 were used to guide the selection of five serine/threonine kinase (STK) and five tyrosine kinase (TK) targets for absolute binding free-energy simulations. For some kinases, the hit rate binary classifier captures a set of relatively weak hits with average binding which, in context with large Potts penalty (see Figure 1D), might be explained by a large $\Delta G_{\text{reorg}}$ incurred for the folded DFG-out state (Figure 1B). See Table 3—source data 1 for detailed data and references for experimental binding affinities.

| Kinase | Class | Hit rate* (Kd <10 µM) | Potts penalty† ($\Delta E_{\text{rot}}$) | Calculated $\Delta G_{\text{reorg}}$‡ |
|--------|-------|----------------------|--------------------------------------|-------------------------------|
| MELK   | STK   | 3                    | 5.9                                  | 5.6±0.2                        |
| MAPK9  | STK   | 5                    | 4.7                                  | 6.9±0.3                        |
| CDK2   | STK   | 2                    | 5.3                                  | 7.7±0.2                        |
| IRAK4  | STK   | 0                    | 2.5                                  | 5.4±0.2                        |
| BRAF   | STK   | 7                    | 4.0                                  | 6.5±0.1                        |
| ABL1   | TK    | 10                   | −1.0                                 | 1.3±0.3                        |
| LCK    | TK    | 11                   | 0.5                                  | 1.0±0.3                        |
| TIE2   | TK    | 6                    | 1.1                                  | −0.3±0.2                       |
| NTRK2  | TK    | 6                    | −1.0                                 | 1.7±0.1 §                      |
| DDR1   | TK    | 11                   | 0.4                                  | 0.3±0.3                        |

*Type-II inhibitors only, data from Davis et al., 2011.
†Calculations from Figure 1d.
‡$\Delta G_{\text{reorg}}$ was calculated from Equation 3. Reported standard deviations were calculated by propagating error from the simulations used in the calculation of average $\Delta G_{\text{reorg}}$ in units of kcal/mol (see Figure 3—source data 1 for statistics from individual simulations).
§Average and standard deviation calculated from two simulated complexes.
reorganization penalties are strongly correlated with corresponding $\Delta E_s$ calculated from the Potts model ($R^2 = 0.75, P \approx 10^{-3}$) emphasizing the connection between coevolutionary statistical energies in sequence space and physical free-energies in protein conformational space (Figure 4). From this relationship, we can approximate a scale for the Potts $\Delta E$ scores in physical free-energy units which describe the conformational landscapes of folded proteins in a similar manner to that of an earlier study of protein folding landscapes (Morcos et al., 2014); we find that a Potts statistical energy difference $\Delta E$ of one unit corresponds approximately to 1.3 kcal/mol of $\Delta G_{\text{reorg}}$.

**Structural and evolutionary basis for the divergent TK and STK conformational landscapes**

The consistency of the predictions between $\Delta G_{\text{reorg}}$ and $\Delta E_{\text{Potts}}$ identified from Potts-guided free-energy simulations (Figure 4) led us to investigate whether the observed difference in the free energy to reorganize from the active to inactive state is a more general feature that distinguishes TKs from STKs. To this end, we extracted ~200,000 STKs and ~10,000 TKs from the large MSA of Pfam sequences used in the construction of our Potts model, based on patterns of sequence conservation that clearly distinguish the two classes (see Methods). For each sequence, we calculated $\Delta E_{\text{Potts}}$ threaded over the structural database (a total of 4268 active DFG-in and 510 classical DFG-out PDB structures) and plotted the distributions for TKs and STKs, revealing a bias for STKs toward larger Potts conformational penalties (Figure 5). The average difference between these distributions, $\Delta \Delta E = 3.2$, is extremely unlikely to be observed by chance ($p \leq 10^{-15}$, see Methods) and supports the hypothesis that TKs are evolutionarily biased toward a lower free-energy cost to adopt the classical ‘folded activation loop’ DFG-out conformation ($\Delta G_{\text{reorg}}$) compared to STKs. We estimate that $\Delta \Delta E = 3.2$ corresponds to ~4.3 kcal/mol based on the analysis summarized in Figure 4.

To gain insight into the molecular basis for this effect which distinguishes the conformational landscape of TKs from STKs, we examined the residue-residue interactions that make the most significant contributions to the observed $\Delta \Delta E$. The difference between average Potts threaded-energy penalties, $\Delta \Delta E = \langle \Delta E \rangle_{\text{STKs}} - \langle \Delta E \rangle_{\text{TKs}}$, can also be written as a sum over pairs of alignment positions $i$ and $j$ along length $L$ of the aligned kinase domains, $\sum_{|i-j|>4} \Delta \Delta E_{ij}$ (see Methods for details). We find that ~75% of the total contribution to $\Delta \Delta E$ (approximately 3 kcal/mol) can be traced back to a
A small number (10) of residue-residue interactions involving the activation loop, suggesting that mutations within the activation loop are largely responsible for the evolutionary divergence between the conformational free-energy landscapes of TKs and STKs. These interactions occur between important structural motifs responsible for controlling the stability of the active ‘extended’ conformation of the activation loop (Figure 6A), especially the activation loop N-terminal and C-terminal ‘anchors’ (Nolen et al., 2004), and the regulatory ‘RD-pocket’ formed by the HRD motif of the catalytic loop which functions to stabilize or destabilize this conformation depending on the activation loop’s phosphorylation state (Nolen et al., 2004; Johnson et al., 1996). The remaining top $\Delta \Delta E_{ij}$ correspond to contacts that stabilize the DFG-out ‘folded’ conformation of the activation loop for TKs, wherein the kinase’s substrate binding site recognizes its own activation loop tyrosine (Figure 6C, right; Hubbard et al., 1994). We describe these interactions below, focusing on the strongest effects involving these structural motifs that lead to differences in the conformational free-energy landscapes of TKs and STKs. The residue nomenclature we use in our descriptions follows the format $\text{MSA}^{\text{CDK6}}_{\text{ABL1}}$, which is the unique residue numbering in our MSA followed by Abl1 (TK) numbering in the subscript (active PDB: 2GQG, inactive PDB: 1IEP) and CDK6 (STK) numbering in the superscript (active PDB: 1XO2, inactive PDB: 1G3N), corresponding to the original PDB files used to generate Figure 6B and Figure 6C.

The ‘RD-pocket’ (Nolen et al., 2004) is a conserved pocket formed by the Arg and Asp residues of the HRD motif (Johnson et al., 1996) (R123$^{\text{ABL1}}$ and D124$^{\text{ABL1}}$) and a positively charged Lys or Arg that is often present in the N-terminal anchor of the activation loop (R147$^{\text{ABL1}}$, L144$^{\text{ABL1}}$ in the HRD motif and Lys/Arg 147$^{\text{CDK6}}$ in the N-terminal anchor form an unfavorable like-charge interaction when the activation loop is in the active, extended conformation (Nolen et al., 2004). Kinase activation is typically a complex process involving many layers of regulation from other protein domains, cofactors, and phosphorylation events (Endicott et al., 2012) however, a general activation mechanism that applies to the majority of kinases involves quenching the net-charge of the RD-pocket by addition of a negatively charged phosphate group to a nearby residue in the activation loop, stabilizing the active conformation. The conservation of this regulatory mechanism in most protein kinases, particularly those bearing the HRD-Arg residue (termed ‘RD-kinases’ Johnson et al., 1996), explains why Lys or Arg is frequently observed at position 147$^{\text{ABL1}}$, which the Potts model suggests has a greater destabilizing effect on the conformational free-energy landscape.
Figure 6. Molecular basis for the evolutionary divergence between tyrosine kinases (TKs) and serine/threonine kinases (STKs). Visualizing the top interaction pairs that contribute to the result in Figure 5, which is discussed in the main text. (A) Diagrams depicting general features of the active "extended" conformation of the activation loop (left) and the primary structure of these motifs in our multiple sequence alignment (MSA; right) with the HRD (pink), DFG (red), and APE (teal) motifs color-coded for reference. (B–C) Structural examples of a representative STK (CDK6) and TK (Abl) in the active DFG-in conformation (left) and the folded DFG-out conformation (right). Residues are labeled according to their position in our MSA, and colored according to A. The inset (center) displays the $\Delta E_{\text{Potts}}$ of the reference kinase derived from Figure 5 as well as a cartoon depicting their location in the distributions. The diagrams of CDK6 in the active DFG-in conformation (PDB: 1XO2, chain B), CDK6 in the folded DFG-out conformation (PDB: 1G3N, chain A), Abl in the active DFG-in conformation (PDB: 2G2I, chain A), and Abl in the folded DFG-out conformation (PDB: 1IEP, chain A) were generated with PyMol (pymol, 2015). All ligands and some backbone atoms were hidden for clarity.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Interactions belonging to the top five contributions to $\Delta\Delta E$.

Figure supplement 2. Cumulative sum of position pair contributions to $\Delta\Delta E$. 

Gizzio, Thakur et al. eLife 2022;11:e83368. DOI: https://doi.org/10.7554/eLife.83368
effect on the active conformation than Lys (9%) due to interactions with HRD-Arg. The activation loop Arg also forms part of an electrostatic interaction network that stabilizes the ‘Src-like inactive’ conformation in TKs (Ozkirimli and Post, 2006; Wu et al., 2020), a conformation with a ‘partially’ folded activation loop (Figure 1—figure supplement 1) that is suggested to be an intermediate state along the transition to DFG-out (Levinson et al., 2006; Shan et al., 2013). On the other hand, RD-STKs display K147 more frequently (26%), which the Potts model suggests interacts more favorably with the HRD-Arg independently of activation loop phosphorylation, thus contributing to greater stabilization of the active DFG-in conformation for RD-STKs in comparison with TKs. Additionally, the Potts statistical energy analysis suggests that packing interactions between HRD-Arg and V160(161) located near the activation loop C-terminal also contributes to phosphorylation-independent stabilization of the active conformation in RD-STKs (Figure 6B left), appearing in 28% of RD-STKs and only 2% of RD-TKs. In RD-TKs, the residue at this position is usually Arg or Lys which appears to be repelled from the RD-pocket (Figure 6C left) and is suggested by the Potts model to result in a less stable active conformation.

The largest $\Delta \Delta E_{ij}$ term, which contributes 16% of the total difference in average Potts conformational penalties between STKs and TKs, comes from an interaction pair in the C-terminal of the activation loop (162183) and the C-terminal of the catalytic loop (126147). These residues form part of the ‘C-terminal anchor’ (Nolen et al., 2004) which is important for creating a suitable binding site for the substrate peptide. The C-terminal anchor residue 162183 is Pro in TKs and typically Ser or Thr in STKs (Taylor et al., 1995). In STKs, the sidechain hydroxyl of this residue forms a hydrogen bond with K126147 in the catalytic loop, creating a stable binding site for substrate phosphoacceptor residues and stabilizing the C-terminal anchor (Taylor et al., 1995). K126 is also directly involved in catalysis by interacting with and stabilizing the gamma phosphate of ATP (Zheng et al., 1993); hence, it is often referred to as the ‘catalytic lysine’. The hydrogen bond between K126 and S or T162 is almost always formed in the active DFG-in conformation, and we observe breakage of this hydrogen bond in many STKs crystallized in the DFG-out/activation loop folded conformation (e.g. CDK6, Figure 6B), suggesting that deformation of the C-terminal anchor contributes an energetic penalty for the active $\rightarrow$ inactive conformational change. In TKs, however, the catalytic lysine is almost always replaced with Ala, with the exception of a few TKs (e.g. c-Src) which have instead adopted Arg at this position. The C-terminal anchor of TKs containing A126 and P162 is less stable in comparison to STKs containing (K126, T162) or (K126, S162) for which the Potts coupling is very favorable, consistent with the structural observation that (A126, P162) forms weak interactions (Figure 6C). Our analysis suggests the interaction pair (A126, P162) weakens the C-terminal anchor, leading to a less stable active conformation in TKs as compared with STKs. Another significant contribution to the stabilization of the C-terminal anchor in the active DFG-in conformation for STKs comes from interactions between the residue pair (161181, 166186) which are both located within the activation loop. We observe (G161, M166) at this position pair in 33% of STKs, but never in TKs (Figure 6—figure supplement 1D). The Potts coupling between these residues is highly favorable. In contrast, we observe (L161, M166) in 40% of TKs but never in STKs (Figure 6—figure supplement 1H), which have weaker coupling. The bulky sidechains of (L161, M166) observed in TKs cause the activation loop to ‘bulge’ in this C-terminal region which has been previously identified as a feature of TKs that helps shape the substrate binding site to accommodate Tyr residues (Nolen et al., 2004). In addition to this paradigm, our analysis suggests that the C-terminal bulge results in weaker structural constraints on the active conformation relative to STKs.

In summary, TKs are suggested by the Potts statistical energy model which is based on sequence covariation, to have on average, weaker N-terminal anchor, RD-pocket, and C-terminal anchor interactions than STKs. This mechanism of shifting the TK conformational equilibrium away from the active DFG-in/extended activation loop conformation can explain 7 of the top 10 $\Delta \Delta E_{ij}$, accounting for ~80% of these contributions to the divergence between the STK and TK conformational landscapes. The remaining ~20% of the top contributions can be attributed to residue-residue interactions that occur within the folded DFG-out conformation wherein the activation loop of TKs binds to the kinase’s own active site as though it was engaging a peptide substrate in trans (Figure 6C right) (Nolen et al., 2004). STKs, however, rarely adopt a folded DFG-out conformation with this property, and instead the activation loop is typically found to be unresolved and/or projecting outward toward solvent (Figure 6B right). The Potts model suggests that this substrate mimicry of the folded
DFG-out activation loop observed in TKs is highly dependent on the presence of a Tyr phosphorylation site at position 154. The active conformation, the anionic pY154 stabilizes the active conformation by binding to the basic RD-pocket (Figure 6C left, phosphate not shown). However, in the (unphosphorylated) folded DFG-out conformation, this Y154 mimics a substrate by stacking against the TK-conserved P162 residue (Figure 6C right) (Nolen et al., 2004). The substrate mimicking nature of this binding mode is demonstrated by the autophosphorylation dimer structure of TK FGFR3 (PDB: 6PNX) solved recently (Figure 6—figure supplement 1K; Chen et al., 2020).

The striking connection between the ability of TKs to phosphorylate tyrosine substrates and their enhanced access to the DFG-out conformation via substrate-competitive contacts from their own activation loop described above suggests an evolutionary model for the TK conformational behavior characterized in this work. In this model, the coevolution of residues that form substrate-competitive contacts in folded DFG-out appears to be a byproduct of the evolutionary pressure for TKs to phosphorylate other TKs on their activation loop tyrosine residues. STKs, on the other hand, have optimized the binding of Ser and Thr substrates via a different binding mode (Endicott et al., 2012) which does not have the same energetic feedback with the stability of the folded DFG-out conformation. Additionally, the catalytic domains of TKs appear to have a less energetically stable 'extended' activation loop conformation than STKs, which may have encouraged the evolution of more complex mechanisms of allosteric regulation and autophosphorylation which are highly important regulatory mechanisms in TKs (Chen et al., 2020; Beenstock et al., 2016; Lemmon and Schlessinger, 2010). The combined effect of these two TK phenotypes, the former favoring the stabilization of folded DFG-out and the latter favoring destabilization of active DFG-in, may explain their low free-energy cost for the DFG-in → DFG-out conformational change compared with STKs.

Discussion
In this work, we have combined sequence and structure-based approaches to analyze the conformational free energy difference between active DFG-in and inactive DFG-out kinase states. Using a Potts statistical energy model derived from residue-residue covariance in a kinase family multiple sequence alignment, we first threaded all human STKs and TKs onto large ensembles of active DFG-in and classical DFG-out structures from the PDB. We found distinctly different distributions of threading scores for SSTKs compared with TKs, with STKs having a significant conformational reorganization penalty compared with TKs. The molecular basis for the evolutionary divergence in the conformational landscapes was analyzed; a substantial contribution to the difference is associated with sequence position pairs that couple the N and C terminal anchor residues of the activation loop to N and C terminal residues in the catalytic loop, according to the Potts statistical energy analysis. We then used the Potts statistical energy model to guide the selection for structure-based MD binding free energy simulations of 68 protein-ligand complexes; using the calculated binding free-energy estimates together with experimental values, we were able to estimate free-energy costs for the large-scale (~17–20 Å) conformational change of the activation loop by an indirect approach. The structure-based estimates of the reorganization free-energy penalties are consistent with the sequence-based estimates. Additionally, the strong correlation between ∆G_georg and ∆E_potts identified in this study reveals that the conformational landscape has a strong sequence dependence with STKs having an ~4 kcal/mol conformational free energy bias favoring the active state over the inactive state relative to TKs (Figure 4). We note that the most potent type-II inhibitors from the literature which target STKs bind with nanomolar K_s, similar to that for TKs, despite the substantial additional reorganization penalty that STKs must overcome. This suggests that medicinal chemists have implicitly been able to exploit particularly favorable characteristics of the type-II binding pocket to design inhibitors with extremely strong affinities to the DFG-out (activation loop folded) receptor conformations of STKs, and that further analysis of the molecular basis for this tight binding could provide a basis for designing more selective inhibitors.
Materials and methods

Multiple sequence alignment (MSA) and classification of serine/threonine vs tyrosine kinases

An MSA of 236,572 protein kinase catalytic domains with 259 columns was constructed as previously described (McGee et al., 2021). STKs and TKs were classified based on patterns of sequence conservation previously identified by Taylor and co-workers (Taylor et al., 1995); characteristic sequence features of TKs and STKs which form their respective phosphoacceptor binding pockets are found at the HRD +2 (Ala or Arg in TKs, Lys in STKs) and HRD +4 (Arg in TKs, variable in STKs) in the catalytic loop, as well as the APE-2 residue (Trp in TKs, variable in STKs) in the activation loop. These residues correspond to positions 126, 128, and 165 in our MSA, respectively. Kinases which satisfy the conditions for TKs at all three positions were classified as TKs (10,345 raw sequences, 1069 effective sequences), and those that satisfy the condition for STKs at position 126 and are non-overlapping with the TK condition at position 128 were classified as STKs (210,862 raw sequences, 22,893 effective sequences). The effective number of sequences in each class was calculated by summing over sequence weights, where each sequence was assigned a weight defined as the fraction of the number of sequences in the same class that are within 40% identity. In this way, we correct for the effects of phylogenetics in the calculation of sample size as well as other quantities (see below).

Human kinome dataset

497 human kinase catalytic domain sequences were acquired from Modi and Dunbrack, 2019b (excluding atypical kinases). These sequences were aligned using a Hidden Markov Model (HMM) that contains 259 columns (L=259), which was constructed from the same MSA used to derive our Potts model. 447 human kinases remained after filtering-out sequences with 32 or more gaps.

Classification of gatekeeper size

The designation of gatekeeper residues as ‘large’ or ‘small’ was based on sidechain van der Waals volumes (Miller et al., 1987; Figure 1—figure supplement 3), where small gatekeepers have a volume of <110 Å³ (Gly, Ala, Ser, Pro, Thr, Cys, Val), and large gatekeepers have a volume of >110 Å³ (Asn, His, Ile, Leu, Met, Lys, Phe, Glu, Tyr, Gin, Trp, Arg).

PDB dataset and conformational states

X-ray crystal structures of tyrosine, serine/threonine, and dual-specificity eukaryotic protein kinases in the PDB were collected from http://rcsb.org on July 30, 2020. The protein sequences of 7919 chains were extracted from 6805 PDB files by parsing the SEQRES record and aligned to the MSA used to construct our Potts model, using an HMM.

Contact frequency differences between ensembles of active/DFG-in and inactive/DFG-out (classical DFG-out) PDB structures (Figure 1—figure supplement 2) are incorporated into the calculation of Potts threaded energies, which are central to this work. Our classification of the active DFG-in and classical DFG-out conformational states is based on ref 6, which we describe in further detail here:

Active DFG-in (BLAminus)

The BLAminus state of the DFG motif (active DFG-in) is the only active conformation of the activation loop, in contrast to several other inactive DFG-in states (ABAminus, BLAplus, BLBplus, BLBminus, BLBtrans). In the active conformation, all structural requirements for catalytic activity are typically met, e.g., a complete hydrophobic spine, a salt bridge between β3-Lys → αC-Glu, and an extended activation loop that ensures unobstructed substrate-binding, all of which have high correspondence with the BLAminus state (Modi and Dunbrack, 2019b). Our analysis using the software provided in Modi and Dunbrack, 2019b identifies 3643 structures in this conformation belonging to STKs and 625 structures belonging to TKs (4268 structures in total).

Classical DFG-out (BBAminus)

The DFG-out state is characterized by a ‘flip’ of the conserved Phe to occupy the ATP binding pocket which is otherwise occupied by the conserved Asp in the active conformation. The classical DFG-out conformation is associated with the binding of type-II inhibitors which occupy the back pocket region
opened up by the DFG-flip (Vijayan et al., 2015). This conformation corresponds to the BBaminus rotamer state of these residues, and it is the dominant DFG-out conformation associated with the binding mode of type-II inhibitors. This classical DFG-out or BBaminus conformation is correlated with a larger-scale conformational change of the activation loop that involves an \(-17\) Å ‘folding’ transition with respect to the active conformation. This conformational change is usually accompanied by the formation of secondary structure that obstructs the typical substrate binding surface (e.g. PDB ID: 2HIW, Figure 1—figure supplement 1B). Our conformational analysis via Modi and Dunbrack, 2019b identifies 224 structures in this conformational state belonging to STKs and 286 structures belonging to TKs (510 structures in total).

Contact frequency differences

Each of the clustered PDB structures were converted to an adjacency matrix of binary contacts (1 for ‘in-contact’, 0 otherwise). A contact between residues \(i\) and \(j\) in structure \(n\) was assigned when their nearest sidechain atoms (excluding hydrogen) were detected within a distance \(r_{ij}(n) < 6\) Å. The contact frequency \(c_{ij}^A\) in cluster \(A\) (e.g. active DFG-in) was calculated for each residue pair \((i, j)\) by taking a weighted average over all instances of a contact in that cluster –

\[
c_{ij}^A = \frac{1}{\sum w_{n}^{AB}} \sum_{n} w_{n}^{AB} \delta_{ij}(n)
\]

(4)

\[
\delta_{ij}(n) = \begin{cases} 1, & r_{ij}(n) < 6\text{ Å} \\ 0, & r_{ij}(n) \geq 6\text{ Å} \end{cases}
\]

(5)

Where \(N_A\) is the number of PDB chains in cluster \(A\), and weights were calculated with \(w_{n}^{AB} = \frac{1}{2\pi}\), where \(\Delta\) is the number of times the UniProt ID of structure \(n\) is found within either cluster \(A\) (i.e. active) or cluster \(B\) (i.e. DFG-out). In this way, we have down-weighted contributions to the contact differences \(\Delta c_{ij}^{AB} = c_{ij}^{A} - c_{ij}^{B}\) that are due to overrepresentation of specific kinases in the PDB clusters, with the goal of using contact differences to represent conserved features of the conformational transition across many different kinases. Alignment gaps and unresolved residues were accounted for by excluding these counts in the summations. Only \(|i - j| > 4\) were included in the calculation. The PDB clusters used to calculate these contact differences are described above. The contact frequency differences for both STKs and TKs were plotted on a contact map for visualization (Figure 1—figure supplement 2).

Potts model and threaded-energy calculation

Our Potts Hamiltonian was constructed from an MSA of protein kinase catalytic domains as previously described (McGee et al., 2021). The Potts Hamiltonian \(H(S)\) takes the form –

\[
H(S) = \sum_{i<j} J_{S,S_i}^i + \sum_{i} h_{S_i}^i
\]

(6)

where \(L\) is the number of columns in the MSA (\(L = 259\)), \(h\) is a matrix of self-interactions or ‘fields’, and \(J\) is the coupling matrix which has the interpretation of co-evolutionary interactions between residues.

The Potts threaded-energy penalty \(\Delta E(S)\) for sequence \(S\) to undergo the conformational transition \(A \rightarrow B\) is calculated using contact frequency differences between the two conformational ensembles (Haldane et al., 2016) –

\[
\Delta E(S \in X) = -\sum_{i<j} J_{S,S_i}^i \Delta c_{ij}^{AB}(X)
\]

(7)

where \(X\) represents a class or family of sequences for which sequence \(S\) has membership, and \(\Delta c_{ij}^{AB}(X)\) represents the contact frequency difference between conformations \(A\) and \(B\) observed only for other sequences belonging to class \(X\) (e.g. \(X\) \(\equiv\) TKs or \(X\) \(\equiv\) STKs); upper and lower triangle of Figure 1—figure supplement 2, respectively). As described previously (Haldane et al., 2016) the couplings \((J_{S,S_i}^i)\) and fields \((h_{S_i}^i)\) were transformed to the ‘zero-gauge’ prior to calculating \(\Delta E(S)\).
Contributions to average shift in $\Delta E_{\text{Potts}}$ between STKs and TKs ($\Delta \Delta E$). Where $\Delta E$ ($S$) is the Potts conformational penalty for sequence $S$ to undergo the conformational change $A \rightarrow B$, we define $\Delta \Delta E$ as the difference in average $\Delta E$ between two groups of sequences $X$ and $Y$:

$$\Delta \Delta E = \langle \Delta E \rangle_X - \langle \Delta E \rangle_Y .$$  \hspace{1cm} (8)

To help interpret $\Delta \Delta E$ in a structural and coevolutionary context, we can write $\Delta \Delta E$ as a sum over position pairs ($i,j$)

$$\Delta \Delta E = \sum_{i<j} \Delta \Delta E_{ij} .$$  \hspace{1cm} (9)

To evaluate this, we note that average $\Delta E$ can be expressed as a sum over position pairs

$$\langle \Delta E \rangle_X = \sum_{i<j} \langle \Delta E_{ij} \rangle_X ,$$  \hspace{1cm} (10)

where $\langle \Delta E_{ij} \rangle_X$ is calculated as follows

$$\langle \Delta E_{ij} \rangle_X = -\sum_{\alpha} \sum_{\beta} f_{ij}^{\alpha\beta}(X) \Delta c_{i,j}^{\alpha\beta}(X) .$$  \hspace{1cm} (11)

$f_{ij}^{\alpha\beta}(X)$ is the frequency (bivariate marginal) of residues $\alpha$ and $\beta$ at positions $i$ and $j$ for sequences in the MSA which belong to group $X$, which we calculate after applying the MSA-derived phylogenetic weights described above. Finally, by substituting Equation 10 back into Equation 8, we show how $\Delta \Delta E$ can be decomposed into contributions from individual residue pairs

$$\Delta \Delta E_{ij} = -\sum_{\alpha} \sum_{\beta} f_{ij}^{\alpha\beta}(X) \Delta c_{i,j}^{\alpha\beta}(X) \Delta c_{i,j}^{\alpha\beta}(Y).$$  \hspace{1cm} (12)

By viewing the largest (most positive) $\Delta \Delta E_{ij}$ terms, where $X \equiv \text{STKs}$ and $Y \equiv \text{TKs}$ in Equation 12, we are identifying position pairs that cause STKs to have higher penalties than TK in our Potts threading calculations for the active DFG-in to DFG-out conformational change (Figure 6—figure supplement 2).

**Calculation of p-value for $\Delta \Delta E$**

The quantity $\Delta \Delta E$ is a difference between two averages, $\langle \Delta E \rangle_{\text{STK}} - \langle \Delta E \rangle_{\text{TK}}$. Hypothesis testing to determine the statistical significance of this quantity was performed with respect to a null model where the populations of $\Delta E$s for STKs and TKs, from which our samples were drawn, are indistinguishable. To this end, a p-value was calculated for a t-statistic derived from Welch’s t-test (Welch, 1947), where $s$ is the standard error of average $\Delta E$

$$t = \frac{\langle \Delta E \rangle_{\text{STK}} - \langle \Delta E \rangle_{\text{TK}}}{\sqrt{s_{\text{STK}}^2 + s_{\text{TK}}^2}}$$  \hspace{1cm} (13)

where the averages and standard errors are calculated after down-weighting each sequence as described above. This was done to lessen the effects of phylogenetic sampling bias from our MSA and ensure that $\Delta \Delta E$ captures general differences between TKs and STKs, rather than specific TK or STK families.

The one-tailed p-value was calculated using the cumulative t-distribution function generated in python using the SciPy package (Virtanen et al., 2020),

$$p = 1 - t_{\text{cdf}}(t, \nu)$$  \hspace{1cm} (14)

where the degrees of freedom for the t-distribution describing the combined population, $\nu$, was estimated via the Welch-Satterthwaite equation (Welch, 1947) from the degrees of freedom of the two samples $\nu_{\text{STK}}$ and $\nu_{\text{TK}}$. 

---

Gizzo, Thakur et al. eLife 2022;11:e83368. DOI: https://doi.org/10.7554/eLife.83368

17 of 25
Computational and Systems Biology | Structural Biology and Molecular Biophysics

\[ \nu = \left( \frac{N_{STK}^{N_{STK}} + N_{TK}^{N_{TK}}}{N_{STK}^{N_{STK}} + N_{TK}^{N_{TK}}} \right)^2. \]  

(15)

where \( N \) represents the effective number of STKs or TKs, which is an unbiased count of sequences in each dataset that can be obtained by summing the sample weights (\( N_{STK} = 22,893, N_{TK} = 1069 \)). From the calculation of \( \Delta \Delta E = \langle \Delta E \rangle_{STK} - \langle \Delta E \rangle_{TK} = 3.2 \), we determine the corresponding p-value to be less than \( 10^{-15} \), meaning it is highly unlikely for this large of a difference to be observed if the \( \Delta \Delta E \)s for TKs and STKs were randomly drawn from the same distribution rather than distinct distributions.

**Enumeration of absolute binding free-energy simulations**

In this work, we have performed all-atom MD simulations in explicit solvent for a total of 94 different kinase-inhibitor complexes to calculate ABFEs via the alchemical DDM. 74 of these free-energy calculations were guided by insights from the Potts model, specifically the patterns of Potts conformational penalties plotted in Figure 1D. 68 of these complexes correspond to the ABFEs plotted in Figure 3, of which 23 type-II inhibitors and 23 type-I inhibitors ABFEs are plotted for STKs in Figure 3A, and 22 type-II inhibitors ABFEs for TKs are plotted in Figure 3B. Six additional Potts-guided ABFEs corresponding to CSF1R, KIT, PDGFRA, MAPK14, and PTK2B are included in Figure 2. An additional 20 type-I and type-I ½ ABFEs were calculated as part of our benchmarking procedure described in the Methods section.

**Double decoupling method setup**

The double decoupling method (DDM), also known as an ‘alchemical’ method, was applied to compute ABFE (\( \Delta G_{\text{bind}}^{o} \)), as shown in Equation 16 (Deng et al., 2018; Sakae et al., 2020). This method computes the free energies of decoupling the inhibitor from the bulk solvent in the presence and absence of a receptor via a nonphysical thermodynamic cycle where the two end states are connected via the alchemical pathway. The starting holo-structures for ABFE calculations were taken from the available holo crystal structure. The absence of crystal structure prompted us to model the structure of the ligand into the active site of the kinase by superimposing over the binding pose of the available holo crystal structure.

\[ \Delta G_{\text{bind}}^{o} = -\Delta G_{\text{restrain}} + \Delta G_{\text{decouple}} + \Delta G_{\text{gas}} + \Delta G_{\text{bulk}} \]  

(16)

Decoupling of the ligand was achieved by first turning off the coulombic intermolecular interactions followed by Lennard-Jones intermolecular interactions from both the legs. This allows DDM to estimate the free energy, i.e., in the presence of protein (\( \Delta G_{\text{decouple}} \)) and absence of protein, i.e., in the bulk solvent, (\( \Delta G_{\text{bulk}} \)) as shown in Equations 17; 18.

\[ \Delta G_{\text{decouple}} = \Delta G_{\text{decouple}} - \text{Coulomb} + \Delta G_{\text{decouple}} - \text{LJ} \]  

(17)

\[ \Delta G_{\text{bulk}} = \Delta G_{\text{bulk}} - \text{Coulomb} + \Delta G_{\text{bulk}} - \text{LJ} \]  

(18)

Substituting Equations 17; 18 into Equation 19 yields the estimated (\( \Delta G_{\text{bind}}^{o} \)) from DDM

\[ \Delta G_{\text{bind}}^{o} = -\Delta G_{\text{restrain}} + \Delta G_{\text{Coulomb}} + \Delta G_{\text{LJ}} + \Delta G_{\text{gas}} \]  

(19)

where \( \Delta G_{\text{Coulomb}} \) is the electrostatic energy contribution toward the total ABFE, and \( \Delta G_{\text{LJ}} \) is the non-polar energy contribution.

In this study, depending on the system’s convergence, either 20 or 31 total \( \dot{i} \)s were used for decoupling the ligand from bulk solvent. For instance, 5 \( \dot{i} \)s with \( \Delta i=0.5 \) or 11 \( \dot{i} \)s with \( \Delta i=0.1 \) were used for coulombic decoupling and 15 \( \dot{i} \)s with \( \Delta i=0.1 \) or 20 \( \dot{i} \)s with \( \Delta i=0.05 \) were used for decoupling Lennard-Jones interactions in the bulk solvent.

Similarly, depending on the convergence, either 30 or 42 total \( \dot{i} \)s were used for decoupling ligand bound to protein. For instance, 11 or 12 non-uniformly distributed \( \dot{i} \)s were used to restrain the ligand. Decoupling the coulombic interactions between ligand and protein was achieved by either using 4 \( \dot{i} \)s with \( \Delta i=0.25 \) or 10 \( \dot{i} \)s with \( \Delta i=0.1 \), whereas a large number of \( \dot{i} \)s were used for decoupling Lennard-Jones interactions, i.e., 15 \( \dot{i} \)s with \( \Delta i=0.1 \) or 20 \( \dot{i} \)s with \( \Delta i=0.05 \) were used. The correction term developed by Rocklin and coworkers for treating charged ligands during DDM simulations was
adopted (Rocklin et al., 2013). In this regard, it is well documented that the use of a finite-sized periodic solvent box during DDM simulations can lead to non-negligible electrostatic energy contribution toward the calculated total ABFE. Thus, calculated (ΔGααbind) for charged ligand after addition of electrostatic correction term can be expressed as:

\[
\Delta G_{\text{bind}}^{\alpha} = -\Delta G_{\text{restrain}}^{\text{bound}} + \Delta \Delta G_{\text{Coulomb}} + \Delta \Delta G_{LJ} + \Delta \Delta G_{\text{electrostatic\;correction}} + \Delta G^{\text{gas\;restrain}}
\]  

(20)

For a proper convergence during DDM simulations, the application of restrains is crucial. Herein, we have used six relative orthogonal restrains with harmonic potentials that include one distance, two angles, and three dihedral angles restrain between the ligand and the protein with a force constant of 10 kcal mol\(^{-1}\) Å\(^{-2}\) [deg\(^{-2}\)]. At each \(\lambda\), 10–30 ns of decoupling simulation via replica-exchange (Affentranger et al., 2006) were obtained to compute the ΔGbind over a well-converged trajectory.

**Molecular dynamics setup**

In this study, MD simulations were applied to compute the binding free-energy simulations via DDM. GROMACS-2018.8 (Abraham et al., 2015) was used as an MD engine for all simulations. The leap module of AMBER16 was used to add the missing hydrogen atoms to the kinase enzymes. The system was solvated explicitly using TIP3P water boxes (Jorgensen et al., 1983) that extended at least 10 Å from the center of the system in each direction. The topology file for the kinase enzyme was created using the amber forcefield ff14SB (Essmann et al., 1995) and general amber force field 2 (GAFF2) (Wang et al., 2004) were employed to parametrize different inhibitors used in this study. The overall charge of the system was maintained by adding a suitable number of counterions in each system. During the simulations, electrostatic interactions were computed using the particle mesh Ewald method (Essmann et al., 1995) with a cutoff and grid spacing of 10.0 and 1.0 Å, respectively. The NPT (constant Number of particles, Pressure, and Temperature) ensembles with a time step of 2 fs was used in the simulations.

**Benchmarking calculations for absolute binding free-energy simulations**

Accurate prediction of the ABFE difference between the Apo and Holo state of a protein is extremely important to achieve from force field-based MD simulations (Lin et al., 2013; Lovera et al., 2012). In this study, we used type-II inhibitors as probes to estimate ΔGreorg via ABFE simulations, which is the excess free-energy between experimentally determined binding affinity and ΔGbind calculated from ABFE. Target kinases, i.e., MAPK14, CDK2, and JNK1 bound with type-I and I 1/2 inhibitors in the active conformation states (where ΔGreorg is expected to be close to zero) have been regularly used by the computational community as benchmark systems for absolute or relative binding free energy calculations (Wang et al., 2015; Lee et al., 2020; Goel et al., 2021; Kuhn et al., 2020; Gapsys et al., 2019; Khalak et al., 2021). Benchmarking calculations over multiple protein-ligand complexes show close agreement between calculated (ΔGbind) and experimental (ΔGexp) terms (Table 2, Figure 2—figure supplement 1). In the later part of benchmarking studies, we have included ABL1 bound to type-II inhibitors in the DFG-out/folded activation loop state. NMR studies have shown experimentally that ABL1 has a ΔGreorg of 1.2 kcal/mol (Xie et al., 2020), which is consistent with our range of estimates of ΔGreorg for this kinase (Table 2).

### Table 2

| Kinase | # Compounds | MUE | RMSE | \(\langle \Delta G_{\text{exp}} - \Delta G_{\text{ABFE}}^{\text{bind}} \rangle\) |
|--------|-------------|-----|------|--------------------------------------------------|
| CDK2   | 6 (type-I)  | 0.71| 1.03 | 1.89                                             |
| JNK1   | 5 (type-I)  | 0.37| 0.44 | 1.29                                             |
| MAPK14 | 9 (type-I/2) | 0.47| 0.64 | 0.2                                              |
| ABL1   | 6 (type-II) | 1.47| 1.57 | 1.26                                             |

Gizio, Thakur et al. eLife 2022;11:e83368. DOI: https://doi.org/10.7554/eLife.83368
Potts-guided target selection for absolute binding free-energy simulations

Our Potts threaded-energy calculations were used alongside experimental type-II binding data from the large-scale assay by Davis et al., 2011 to identify kinase targets that are likely to have very large or very small \( \Delta G_{\text{reorg}} \). As described in the main text, all-atom MD simulations to calculate ABFEs of type-II inhibitors can be used alongside experimental binding affinities to calculate the free-energy cost for a kinase to reorganize to the DFG-out/folded activation loop conformation. This relation, \( \Delta G_{\text{reorg}} = \Delta G_{\text{exp}} - \Delta G_{\text{bind}}(\text{ABFE}) \), gives the free-energy cost to reorganize in physical energy units (kcal/mol) and can be used to approximate a scale for the Potts statistical energy differences provided one sample a sufficient range of \( \Delta G_{\text{reorg}} \) and \( \Delta E_{\text{Potts}} \). However, ABFE simulations are much more computationally demanding than the Potts threading calculation, which we sought to mitigate by choosing kinase simulation targets which are likely to provide a strong signal, guided by the Potts model. We direct the reader to Table 1, which contains the Potts penalties and type-II hit rates for the targets of interest. For comparison, Figure 1A and Figure 5 provide the overall distributions of hit rates and Potts penalties for TKs and STKs.

A significant challenge for our target selection was the limited availability of type-II inhibitors co-crystallized against STKs which have (a) Potts penalties and type-II hit rates that predict very high \( \Delta G_{\text{reorg}} \), (b) experimental binding affinities available in the literature in the form of IC50, Ki, or Kd, (c) availability of protein-ligand co-crystallized structure(s), and (d) type-II inhibitor complex systems where the activation loop appears to have undergone a large-scale ‘folding’ conformational change relative to the active ‘extended’ conformation. STK complexes that satisfy all four criteria appear to be sparse, which is consistent with the notion that kinases with large reorganization penalties are more difficult to crystallize in the classical DFG-out conformation (Haldane et al., 2016). However, for some STKs with very high Potts threaded-energy penalties (e.g., MELK) there has been significant medicinal chemistry efforts to design potent type-II inhibitors and structurally characterize their complexes using x-ray crystallography. Using type-II co-crystal structures that cover five different STKs with high Potts penalties (Table 1) and five different TKs with low Potts penalties, we were able to sample a wide range of \( \Delta G_{\text{reorg}} \) from a total of 45 ABFE simulations covering 45 type-II inhibitor complexes and 10 different kinase targets. These simulations and subsequent calculations of \( \Delta G_{\text{reorg}} \) for each kinase resulted in a strong correlation with the Potts threaded energy scores (Figure 5), allowing us to establish a scale for the Potts energies in kcal/mol. We have provided detailed results from the ABFE simulations of these 10 kinase targets in the form of supplementary figures (Figure 3—figure supplement 1) where the average \( \Delta G_{\text{reorg}} \) for each kinase is visualized as the y-intercept of a linear regression with the slope constrained to one.

Data availability

Values of \( \Delta G_{\text{ABFE}} \) from all ABFE simulations described in this work, including benchmarking calculations (94 simulations in total), are provided in the form of supplementary tables (Figure 2—source data 1 and Figure 3—source data 1). Potts \( \Delta E_s \), type-II hit rates computed from Shan et al., 2013, the identity of gatekeeper residues and corresponding van der Waals volumes in Å (Stancîk et al., 2018), and the classification of human kinases as TKs or STKs were provided in a separate supplementary table (Figure 1—source data 1).

Code availability

The Mi3-GPU (Haldane and Levy, 2021) source code required to reproduce the Potts model employed in this manuscript can be found at the following link: https://github.com/ahaldane/Mi3-GPU, (copy archived at URL to be added) (v1.1).

Acknowledgements

This research was supported by National Institutes of Health grant number R35-GM132090, and by NIH Computer Equipment Grant (OD020095). Gratitude is also expressed to the OWLSNEST high performance cluster at Temple University for its computing support in this project. We thank Shima Arasteh for helpful discussions related to kinase conformational states and alchemical free-energy simulations.
Additional information

Funding

| Funder                        | Grant reference number | Author                      |
|-------------------------------|------------------------|----------------------------|
| National Institutes of Health | R35-GM132090           | Joan Gizzio, Abhishek Thakur, Allan Haldane, Ronald M Levy |
| National Institutes of Health | OD020095               | Joan Gizzio, Abhishek Thakur, Allan Haldane, Ronald M Levy |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Joan Gizzio, Abhishek Thakur, Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Writing - original draft, Writing - review and editing; Allan Haldane, Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Investigation, Methodology, Writing - review and editing; Ronald M Levy, Conceptualization, Resources, Supervision, Funding acquisition, Investigation, Methodology, Project administration, Writing - review and editing

Author ORCIDs

Joan Gizzio [http://orcid.org/0000-0002-4127-2294](http://orcid.org/0000-0002-4127-2294)
Abhishek Thakur [http://orcid.org/0000-0002-4827-7602](http://orcid.org/0000-0002-4827-7602)
Allan Haldane [http://orcid.org/0000-0002-8343-1994](http://orcid.org/0000-0002-8343-1994)
Ronald M Levy [http://orcid.org/0000-0001-8696-5177](http://orcid.org/0000-0001-8696-5177)

Decision letter and Author response

Decision letter [https://doi.org/10.7554/eLife.83368.sa1](https://doi.org/10.7554/eLife.83368.sa1)
Author response [https://doi.org/10.7554/eLife.83368.sa2](https://doi.org/10.7554/eLife.83368.sa2)

Additional files

Supplementary files
- MDAR checklist
- Supplementary file 1. Figure 1 - Source Data 1
- Supplementary file 2. Figure 2 - Source Data 1
- Supplementary file 3. Figure 3 - Source Data 1

Data availability

Our computational study makes use of experimental data from the literature, which we extracted and curated manually rather than relying on any specific database. Any experimental data used can be found in our supporting information in the form of tables alongside appropriate citations. A large set of experimental "hit rates" were derived from binding affinities available from Davis et al., Nature Biotechnology (2011). The data used to generate various plots in the main text can be found in tables throughout the supporting information, as well as a distinct "supplementary table" which we provide. The Potts model code is made available as a github link, provided in the main text "Code Availability" section.
The following previously published dataset was used:

| Author(s)          | Year | Dataset title                              | Dataset URL                        | Database and Identifier |
|--------------------|------|--------------------------------------------|------------------------------------|-------------------------|
| Davis MI, Hunt JP, | 2011 | Comprehensive analysis of kinase inhibitor selectivity | https://doi.org/10.1038/nbt.1990 | BindingDB, 10.1038/nbt.1990 |
| Herrgard S, Ciceri P, Wodicka LM, Pallares, G, Hocker M, Treiber DK, Zarrinkar PP |

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. DOI: https://doi.org/10.1016/j.softx.2015.06.001

Affentranger R, Tavernelli I, Di Iorio EE. 2006. A novel hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. Journal of Chemical Theory and Computation 2:217–228. DOI: https://doi.org/10.1021/ct050250b, PMID: 16626508

Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. 2011. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nature Biotechnology 29:1039–1045. DOI: https://doi.org/10.1038/nbt.2017, PMID: 22037377

Apperley JF. 2007. Part I: mechanisms of resistance to imatinib in chronic myeloid leukaemia. The Lancet. Oncology 8:1018–1029. DOI: https://doi.org/10.1016/S1470-2045(07)70342-X, PMID: 17976612

Azam M, Seeliger MA, Gray NS, Kuriyan J, Daley GQ. 2008. Activation of tyrosine kinases by mutation of the gatekeeper threonine. Nature Structural & Molecular Biology 15:1109–1118. DOI: https://doi.org/10.1038/nsmb.1486, PMID: 18794843

Barafis M, Lewandowski K, Gniot M, Iwoła M, Lewandowska M, Komarnicki M. 2008. Dasatinib treatment can overcome imatinib and nilotinib resistance in CML patient carrying F359I mutation of bcr-abl oncogene. Journal of Applied Genetics 49:201–203. DOI: https://doi.org/10.1007/BF03195613, PMID: 18439944

Beenstock J, Mooshayef N, Engelberg D. 2016. How do protein kinases take a selfie (autophosphorylate)? Trends in Biochemical Sciences 41:938–953. DOI: https://doi.org/10.1016/j.tibs.2016.08.006, PMID: 27594179

Bosc N, Wroblowski B, Ací-Séche S, Meyer C, Bonnet P. 2015. A proteomic analysis of human kinome: insight into discriminant conformation-dependent residues. ACS Chemical Biology 10:2827–2840. DOI: https://doi.org/10.1021/acschembio.5b00555, PMID: 26411811

Chen L, Marsiglia WM, Chen H, Katiqbak J, Erdjument-Bromage H, Kemble DJ, Fu L, Ma J, Sun G, Zhang Y, Liang G, Neubert TA, Li X, Traseeth NJ, Hoffmann M. 2020. Molecular basis for receptor tyrosine kinase A-loop tyrosine transphosphorylation. Nature Chemical Biology 16:267–277. DOI: https://doi.org/10.1038/s41589-019-0455-7, PMID: 31959966

Cournia Z, Allen BK, Beuming T, Pearlman DA, Radak BK, Sherman W. 2020. Rigorous free energy simulations in virtual screening. Journal of Chemical Information and Modeling 60:4153–4169. DOI: https://doi.org/10.1021/acs.jcim.0c00116, PMID: 32539386

Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. 2011. Comprehensive analysis of kinase inhibitor selectivity. Nature Biotechnology 29:1046–1051. DOI: https://doi.org/10.1038/nbt.1990, PMID: 22037378

Deng N, Zhang P, Cieplak P, Lai L. 2011. Elucidating the energetics of entropically driven protein-ligand association: calculations of absolute binding free energy and entropy. The Journal of Physical Chemistry B 115:11902–11910. DOI: https://doi.org/10.1021/jp204047b, PMID: 21899337

Deng N, Cui D, Zhang BW, Xia J, Cruz J, Levy R. 2018. Comparing alchemical and physical pathway methods for computing the absolute binding free energy of charged ligands. Physical Chemistry Chemical Physics 20:17081–17092. DOI: https://doi.org/10.1039/c8cp01524d, PMID: 29896599

Endicott JA, Noble MEM, Johnson LN. 2012. The structural basis for control of eukaryotic protein kinases. Annual Review of Biochemistry 81:587–613. DOI: https://doi.org/10.1146/annurev-biochem-052410-090317, PMID: 22482904

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh ewald method. J Chem Phys 103:8577–8599. DOI: https://doi.org/10.1063/1.470117

Gapys S, Pérez-Benito L, Aldeghi M, Seelig D, van Vlijmen H, Tresadern G, de Groot BL. 2019. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. Chemical Science 11:1140–1152. DOI: https://doi.org/10.1039/c9sc03754c, PMID: 34084371

Ghose AK, Herbertz T, Pippin DA, Salvino JM, Mallamo JP. 2008. Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. Journal of Medicinal Chemistry 51:5149–5171. DOI: https://doi.org/10.1021/jm800475y, PMID: 18710211

Gilson MK, Given JA, Bush BL, McCammon JA. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophysical Journal 72:1047–1069. DOI: https://doi.org/10.1016/S0006-3495(97)78756-3, PMID: 9138555
Goel H, Hazel A, Ustach VD, Jo S, Yu W, MacKerell AD. 2021. Rapid and accurate estimation of protein-ligand relative binding affinities using site-identification by ligand competitive saturation. *Chemical Science* **12**:8844–8858. DOI: https://doi.org/10.1039/d1sc01781k, PMID: 34257885

Guest EE, Cervantes LF, Pickett SD, Brooks CL, Hirst JD. 2022. Alchemical free energy methods applied to complexes of the first bromodomain of BRD4. *Journal of Chemical Information and Modeling* **62**:1458–1470. DOI: https://doi.org/10.1021/acs.jcim.1c01229, PMID: 35258972

Haldane A, Flynn WF, He P, Vijayan RS, Ley MV. 2016. Structural propensities of kinase family proteins from a potts model of residue co-variation. *Protein Science* **25**:1378–1384. DOI: https://doi.org/10.1002/proc.2954, PMID: 27241634

Haldane A, Ley MV. 2021. Mi3-GPU: MCMC-based inverse Ising inference on GPUs for protein covariation analysis. *Computer Physics Communications* **260**:107312. DOI: https://doi.org/10.1016/j.cpc.2020.107312, PMID: 33716309

Han S, Mistry A, Chang JS, Cunningham D, Griffee M, Bonnette PC, Wang H, Chrunky BA, Apnes GE, Walker DP, Brosius AD, Buckdinder L. 2009. Structural characterization of proline-rich tyrosine kinase 2 (PYK2) reveals a unique (DFG-out) conformation and enables inhibitor design. *Journal of Biological Chemistry* **284**:13193–13201. DOI: https://doi.org/10.1074/jbc.M809383200, PMID: 19244237

Hari SB, Merritt EA, Maly DJ. 2013. Sequence determinants of a specific inactive protein kinase conformation. *Chemistry & Biology* **20**:806–815. DOI: https://doi.org/10.1016/j.chembiol.2013.05.005, PMID: 23790491

Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, Wang L. 2018. Predicting resistance of clinical abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications Biology* **1**:70. DOI: https://doi.org/10.1038/s42003-018-0075-x, PMID: 30159405

Hayes RL, Vilseck JZ, Brooks CL. 2022. Addressing intersite coupling unlocks large combinatorial chemical spaces for alchemical free energy methods. *Journal of Chemical Theory and Computation* **18**:2114–2123. DOI: https://doi.org/10.1021/acs.jctc.1c00948, PMID: 35255214

Heinzelmann G, Gilson MK. 2021. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific Reports* **11**:1116. DOI: https://doi.org/10.1038/s41598-020-80769-1, PMID: 33441879

Hopf TA, Ingraham JB, Poelwijk FJ, Springer M, Sander C, Marks DS. 2015. Quantification of the Effect of Mutations Using a Global Probability Model of Natural Sequence Variation. arXiv. https://arxiv.org/abs/1510.04612

Hopf TA, Ingrahm JB, Poelwijk FJ, Schärfe CPI, Springer M, Sander C, Marks DS. 2017. Mutation effects predicted from sequence co-variation. *Nature Biotechnology* **35**:128–135. DOI: https://doi.org/10.1038/nbt.3769, PMID: 28092658

Hubbard SR, Wei L, Hendrickson WA. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**:746–754. DOI: https://doi.org/10.1038/372746a0

Jakalian A, Jack DB, Basy CL. 2002. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *Journal of Computational Chemistry* **23**:1623–1641. DOI: https://doi.org/10.1002/jcc.10128, PMID: 12395429

Johnson LN, Noble MEM, Owen DJ. 1996. Active and inactive protein kinases: structural basis for regulation. *Cell* **85**:149–158. DOI: https://doi.org/10.1016/s0036-0270(96)00759-8, PMID: 8612268

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**:926–935. DOI: https://doi.org/10.1063/1.445869

Khalak Y, Tresadern G, Aldeghi M, Baumann HM, Mobley DL, de Groot BL, Gapsys V. 2021. Alchemical absolute free energy methods applied to receptor complexes of the first bromodomain of BRD4. *Journal of Chemical Information and Modeling* **62**:3120–3130. DOI: https://doi.org/10.1021/acs.jcim.0c00165, PMID: 32437145

Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig PA, Reinecke M, Ruprecht B, Petzoldt S, Meng C, Zech A, Reiter K, Qiao H, Helm D, Koch H, Schoof M, Canevari G, Casale E, De Paoli SJ, Feuchtling A, et al. 2017. The target landscape of clinical kinase drugs. *Science* **358**:eaan4368. DOI: https://doi.org/10.1126/science.aan4368, PMID: 29191878

Kuhn M, Firth-Clark S, Tosco P, Meye A, Mackey M, Michel J. 2020. Assessment of binding affinity via alchemical free-energy calculations. *Journal of Chemical Information and Modeling* **60**:3120–3130. DOI: https://doi.org/10.1021/acs.jcim.0c00165, PMID: 32437145

Lapedes A, Giraud B, Jarzynski C. 2012. Using Sequence Alignments to Predict Protein Structure and Stability With High Accuracy. arXiv. https://arxiv.org/abs/1207.2484

Lee TS, Allen BK, Giese TJ, Guo Z, Li P, Lin C, McGee TD, Pearlman DA, Radak BK, Tao Y, Tsai HC, Xu H, Sherman W, York DM. 2020. Alchemical binding free energy calculations in AMBER20: advances and best practices for drug discovery. *Journal of Chemical Information and Modeling* **60**:5595–5623. DOI: https://doi.org/10.1021/acs.jcim.1000163, PMID: 32936637

Leommon MA, Schlessinger J. 2010. Cell signaling by receptor tyrosine kinases. *Cell* **141**:1117–1134. DOI: https://doi.org/10.1016/j.cell.2010.06.011, PMID: 20602996

Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M, Karplus M, Cole PA, Kuriyan J. 2006. A src-like inactive conformation in the abl tyrosine kinase domain. *PLOS Biology* **4**:e144. DOI: https://doi.org/10.1371/journal.pbio.0040144, PMID: 16640460

Li Z, Li X, Huang Y, Wu Y, Liu R, Zhou L, Lin Y, Wu D, Zhang L, Liu H, Xu X, Yu K, Zhang Y, Cui J, Zhan CG, Wang X, Luo Hb. 2020. Identify potent SARS-cov-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *PNAS* **117**:27381–27387. DOI: https://doi.org/10.1073/pnas.2010470117, PMID: 33051297

Gizio, Thakur et al. eLife 2022;11:e83368. DOI: https://doi.org/10.7554/eLife.83368
linyl, meng y, jiang w, roux b. 2013. explaining why gleevec is a specific and potent inhibitor of abl kinase. pnas 110:1664–1669. doi: https://doi.org/10.1073/pnas.1214330110, pmid: 23319661

linyl, meng y, huang l, roux b. 2014. computational study of gleevec and g6g reveals molecular determinants of kinase inhibitor selectivity. journal of the american chemical society 136:14753–14762. doi: https://doi.org/10.1021/ja504146x, pmid: 25243930

liu y, shah k, yang f, witucki l, shokat km. 1998. a molecular gate which controls unnatural atp analogue recognition by the tyrosine kinase v-src. bioorganic & medicinal chemistry 6:1219–1226. doi: https://doi.org/10.1016/s0968-0896(98)00099-6, pmid: 9784862

liu y, gray ns. 2006. rational design of inhibitors that bind to inactive kinase conformations. nature chemical biology 2:358–364. doi: https://doi.org/10.1038/nchembio799, pmid: 16783341

lovera s, sutto l, boubeva r, schapozza l, dolker n, gervasio fl. 2012. the different flexibility of c-src and c-abl kinases regulates the accessibility of a druggable inactive conformation. journal of the american chemical society 134:2496–2499. doi: https://doi.org/10.1021/ja210751l, pmid: 22280319

lovera s, morando m, pucheta-martinez e, martinez-torrecaudrada jl, saladino g, gervasio fl. 2015. towards a molecular understanding of the link between imatinib resistance and kinase conformational dynamics. plos computational biology 11:e1004578. doi: https://doi.org/10.1371/journal.pcbi.1004578, pmid: 26606374

lunt b, szurmant h, procaccini a, hoch ja, hwa t, weigt m. 2010. inference of direct residue contacts in two-component signaling. 1st ed. elsevier inc.

maier ja, martinez c, casavajhala k, wickstrom l, hauser ke, simmerling c. 2015. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. journal of chemical theory and computation 11:3696–3713. doi: https://doi.org/10.1021/acs.jctc.5b00255, pmid: 26574453

manning g. 1995. the protein kinase complement of the human genome - supplemental information. bulletin of the world health organization 73:7–14.

mccgee f, hauti s, novingier q, vucetic s, levy rm, carnevalle v, haldane a. 2021. the generative capacity of probabilistic protein sequence models. nature communications 12:6302. doi: https://doi.org/10.1038/s41467-021-26529-9, pmid: 34728624

metz ks, deoudes em, berginski me, jimenez-ruiz i, aksoy ba, hammerbacher j, gomez sm, phanstiel dh. 2018. coral: clear and customizable visualization of human kinome data. cell systems 7:347–350. doi: https://doi.org/10.1016/j.cels.2018.07.001, pmid: 30172842

miller s, janin j, lesk am, chothia c. 1987. interior and surface of monomeric proteins. journal of molecular biology 196:641–656. doi: https://doi.org/10.1016/0022-2836(87)90038-6, pmid: 3681970

miller wt. 2012. tyrosine kinase signaling and the emergence of multicellularity. biochimica et biophysica acta 1823:1053–1057. doi: https://doi.org/10.1016/j.bbamcr.2012.03.009, pmid: 22480439

modi v, dunbrack rl. 2019a. a structurally-validated multiple sequence alignment of 497 human protein kinase domains. scientific reports 9:19790. doi: https://doi.org/10.1038/s41598-019-56499-4, pmid: 31875044

modi v, dunbrack rl. 2019b. defining a new nomenclature for the structures of active and inactive kinases. pnas 116:6818–6827. doi: https://doi.org/10.1073/pnas.1814279116, pmid: 30867294

morcos f, schafer np, cheng rr, onuchic jn, wolynes pg. 2014. coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. pnas 111:12408–12413. doi: https://doi.org/10.1073/pnas.1413575111, pmid: 25114242

nolen b, taylor s, ghosh g. 2004. regulation of protein kinases; controlling activity through activation segment conformation. molecular cell 15:661–675. doi: https://doi.org/10.1016/j.molcel.2004.08.024, pmid: 15350212

okzirimli e, post cb. 2006. src kinase activation: a switched electrostatic network. protein science 15:1053–1062. doi: https://doi.org/10.1110/ps.051999206, pmid: 16597828

pargellis c, tong l, churchill l, cirillo pf, gilmore t, graham ag, grob pm, hickey er, moss n, puv s, regan j. 2002. inhibition of p38 map kinase by utilizing a novel allosteric binding site. nature structural biology 9:268–272. doi: https://doi.org/10.1038/nsb770, pmid: 11896401

pymolf. 2015. the pymol molecular graphics system, version 2.0. schrödinger, llc.

qian y, cabeza de vaca i, vilseck jz, cole dj, tirado-rites j, jorgensen wl. 2019. absolute free energy of binding calculations for macrophage migration inhibitory factor in complex with a druglike inhibitor. the journal of physical chemistry. b 123:8675–8685. doi: https://doi.org/10.1021/acs.jpcb.9b07588, pmid: 31553604

rocklin gj, mobley dl, dill ka, hünenberger ph. 2013. calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. j chem phys 139:184103. doi: https://doi.org/10.1063/1.4826261, pmid: 24320250

sakae y, zhang bw, levy rm, deng n. 2020. absolute protein binding free energy simulations for ligands with multiple poses, a thermodynamic path that avoids exhaustive enumeration of the poses. journal of computational chemistry 41:56–68. doi: https://doi.org/10.1002/jcc.26078, pmid: 31621932

sebepedrós a, peña mi, capella-gutiérrez s, antó m, gabaldón t, ruiz-trillo i, sabidó e. 2016. high-throughput proteomics reveals the unicellular roots of animal phosphosignaling and cell differentiation. developmental cell 39:186–197. doi: https://doi.org/10.1016/j.devcel.2016.09.019, pmid: 27746046

shan y, arkhipov a, kim et, pan ac, shaw de. 2013. transitions to catalytically inactive conformations in egfr kinase. pnas 110:7270–7275. doi: https://doi.org/10.1073/pnas.1220843110, pmid: 23576739

stancilik ia, šestak ms, ji b, axelsson-fisk m, franjevic d, jers c, domazet-lošo t, mjakovic i. 2018. serine/threonine protein kinases from bacteria, archaea and eukarya share a common evolutionary origin deeply
rooted in the tree of life. *Journal of Molecular Biology* **430**:27–32. DOI: https://doi.org/10.1016/j.jmb.2017.11.004, PMID: 29138003

Sun Q, Biswas A, Vijayan RSK, Craveur P, Forli S, Olson AJ, Castaner AE, Kirby KA, Sarafianos SG, Deng N, Levy R. 2022. Structure-based virtual screening workflow to identify antivirals targeting HIV-1 capsid. *Journal of Computer-Aided Molecular Design* **36**:193–203. DOI: https://doi.org/10.1007/s10822-022-00446-5, PMID: 35262811

Taylor SS, Radzio-Andzelm E, Hunter T. 1995. How do protein kinases discriminate between serine/threonine and tyrosine? structural insights from the insulin receptor protein-tyrosine kinase. *The FASEB Journal* **9**:1255–1266. DOI: https://doi.org/10.1096/fasebj.9.13.7557015, PMID: 7557015

Taylor SS, Kornev AP. 2011. Protein kinases: evolution of dynamic regulatory proteins. Trends in *Biochemical Sciences* **36**:65–77. DOI: https://doi.org/10.1016/j.tibs.2010.09.006, PMID: 20971646

Tokarski JS, Newitt JA, Chang CYJ, Cheng JD, Wittekind M, Kiefer SE, Kish K, Lee FYF, Borzilleri R, Lombardo LJ, Xie D, Zhang Y, Klei HE. 2006. The structure of dasatinib (BMS-354825) bound to activated abl kinase domain elucidates its inhibitory activity against imatinib-resistant abl mutants. *Cancer Research* **66**:5790–5797. DOI: https://doi.org/10.1158/0008-5472.CAN-05-4187, PMID: 16740718

van Linden OPJ, Kooistra AJ, Leurs R, de Esch JP, de Graaf C. 2014. KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *Journal of Medicinal Chemistry* **57**:249–277. DOI: https://doi.org/10.1021/jm400378w, PMID: 23941661

Vijayan RSK, He P, Modi V, Duong-Ly KC, Ma H, Peterson JR, Dunbrack RL, Levy RM. 2015. Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *Journal of Medicinal Chemistry* **58**:466–479. DOI: https://doi.org/10.1021/jm501603h, PMID: 25478866

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods* **17**:261–272. DOI: https://doi.org/10.1038/s41592-019-0686-2, PMID: 32015543

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**:1157–1174. DOI: https://doi.org/10.1002/jcc.20035, PMID: 15116359

Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyanskiy D, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, et al. 2015. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society* **137**:2695–2703. DOI: https://doi.org/10.1021/ja512751q, PMID: 25625324

Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. 2009. Identification of direct residue contacts in protein-protein interaction by message passing. *PNAS* **106**:67–72. DOI: https://doi.org/10.1073/pnas.0805923106, PMID: 19116270

Welch ABL. 1947. The generalization of _student_'s problem when several different population variances are involved published by: *Biometrika* **34**:28–35. DOI: https://doi.org/10.1016/j.jmb.2017.11.004, PMID: 33167630

Xie T, Saleh T, Rossi P, Kalliomaki C. 2022. Structure of the catalytic subunit of camp-dependent protein kinase complexed with mgatp and peptide inhibitor. *Biochemistry* **53**:2154–2161. DOI: https://doi.org/10.1021/acs.biochem.2b00358, PMID: 3443157

Zhao Z, Wu H, Wang L, Liu Y, Knapp S, Liu Q, Gray NS. 2014. Exploration of type II binding mode: a privileged conformation. *Journal of Medicinal Chemistry* **53**:2681–2694. DOI: https://doi.org/10.1021/jm401443h, PMID: 20000735
Editors evaluation

This important paper provides a convincing mechanism for the relative binding specificity of Type II inhibitors to kinases. The combination of a sequence-derived Potts model with experimental dissociation constants and calculated free energies of binding to the DFG-out state is highly compelling and goes beyond the current state-of-the-art. Given the importance of kinases in pathophysiological processes, the results will be of interest to a broad audience and, in addition, the combination of computational methods can be applicable to a wide variety of other biophysical processes that involve conformational rearrangements.

Preamble

Our editorial process produces two outputs: i) public reviews designed to be posted alongside the preprint for the benefit of readers; ii) feedback on the manuscript for the authors, including requests for revisions, shown below. We also include an acceptance summary that explains what the editors found interesting or important about the work.

Decision letter

Decision letter after peer review:

Thank you for submitting your article "Evolutionary Divergence in the Conformational Landscapes of Tyrosine vs Serine/Threonine Kinases" for consideration by eLife. Your article has been reviewed by 3 peer reviewers, and the evaluation has been overseen by a Reviewing Editor and Amy Andreotti as the Senior Editor. The following individual involved in review of your submission has agreed to reveal their identity: Robert Best (Reviewer #1).

The reviewers have discussed their reviews with one another, and the Reviewing Editor has drafted this to help you prepare a revised submission.

Essential revisions:

A description of the limitations of the work should be more prominent. You indeed should explicitly state your assumptions about the energy landscape containing two main states, DFG-in and DFG-out. In that context, the conformation of the C-helix should be discussed, even if it has no bearings on the final results.

Reviewer #1 (Recommendations for the authors):

– I was interested in whether the role of the gatekeeper residue would be also captured by the Potts models and what its relative contribution to overall free energy difference would be. For example if one computed an average $\delta \delta E$ for mutating large to small gatekeepers, does this follow the expected trend of small gatekeepers favoring DFG-out, and what is the magnitude of the effect? And whether the effect depends on the context (STK vs TK).
– Showing the location of the 'gatekeeper' residue in the structure in Figure 1 or elsewhere would help the reader to visualize the location of this frequently discussed residue. Also, there is not much discussion of the gatekeeper after Figure 1, even though it is clear that it plays a role in inhibitor binding.

– Are there data for Type I inhibitors for TK's to show in Figure 3? If dG is ~0 between DFG-in and DFG-out, those would also lie close to the y=x line.

Reviewer #2 (Recommendations for the authors):

This paper focuses on an important topic. It explores how the activation loop conformations affect the type II inhibitor binding to Tyr and Ser/Thr kinases. The computational results agree with the available experimental data. It is a remarkably comprehensive, high quality paper. Some minor comments are as follows:

1. The work focuses on the classical "folded" DFG-out conformation of kinases. From the available PDB structures, most of DFG-out kinases have no well-defined "folded" activation loop structures. In many of them, the activation loops after "DFG" motif have no complete density, which indicates that they are flexible and can be disordered. Given that "folded" for protein structures sometimes means that they form α-helix or β strand, the authors may consider changing "folded" into other descriptors or removing it.

2. Smaller gatekeepers and the sequence of the activation loop contribute to the binding of type II inhibitors in the pocket. The gatekeepers have direct contacts with the inhibitors, while the activation loop does not. Can the authors comment on which one would be more important for determining the type II inhibitor binding?

3. The transition energies were calculated from the active DFG-in conformation to the inactive DFG-out conformation. Please note that the DFG-in conformation has two states as well, i.e., the inactive state (with aC-out) and the active state (with aC-in) (PMID: 35003591, PMID: 33971270). Some kinases barely show the DFG-out conformation (PMID: 32961846). The binding of the type II inhibitors in the pockets does not necessarily always go through the active DFG-in conformation (ac-in). It can also go through the inactive DFG-in conformation (ac-out) to DFG-out. Does the calculated energy data consider the alternative structural transitions?

4. In Figure 5, please provide more details about the collected 10,345 tyrosine kinases and 210,862 serine/threonine kinases.

5. How the activation loop sequences are aligned, and are they structurally validated?

Reviewer #3 (Recommendations for the authors):

I find the paper to be well-written and don't have any major critiques. However, my minor concern is about the presentation of the subject in the Introduction. For a reader who is not familiar with the topic, it would sound that the "DFG-out" is the major inactive conformation. And although the authors do talk about "Src-like" confirmation and present distribution of inactive structures, it is in the Methods section or in Supplementary material. The method has its limits and it is better to describe them
directly in the Introduction. A lot of information that would be helpful for a reader to understand the scope and the importance of the paper is moved to the Methods section which is almost 14 pages long, compared to 4 pages of Introduction. I would advise the authors to reconsider their Methods and move some important concepts into the Introduction.

Author response

Essential revisions:

A description of the limitations of the work should be more prominent. You indeed should explicitly state your assumptions about the energy landscape containing two main states, DFG-in and DFG-out. In that context, the conformation of the C-helix should be discussed, even if it has no bearings on the final results.

We thank the editors and reviewers for their useful feedback. As discussed below in our responses to the reviewers, we have moved the discussion of the αC-helix and its role in the conformational landscape (αC-in ↔ αC-out) to the introduction. We have also made it clear in the introduction that our Potts threading calculations are done with a two-state model (active DFG-in αC-in and inactive classical DFG-out). Even though the αC-helix plays a role in the transition pathway between active DFG-in and inactive DFG-out, its conformation is largely the same in the initial and final states. Therefore, the different conformations that can be adopted by the αC-helix do not significantly affect the Potts threaded energies between the initial and final states, which were calculated for the change in threaded energy between the active DFG-in and inactive DFG-out basins only.

Reviewer #1 (Recommendations for the authors):

- I was interested in whether the role of the gatekeeper residue would be also captured by the Potts models and what its relative contribution to overall free energy difference would be. For example if one computed an average δδE for mutating large to small gatekeepers, does this follow the expected trend of small gatekeepers favoring DFG-out, and what is the magnitude of the effect? And whether the effect depends on the context (STK vs TK).

The reviewer brings up an excellent point regarding the effect of the gatekeeper on the conformational landscape, and it is interesting to ask whether our Potts calculations would capture this effect in the context of type-II binding. When comparing the two conformational ensembles or free energy basins studied in our paper (active DFG-in αC-in and “classical” DFG-out) we find that the gatekeeper’s environment, measured by residue-residue contacts, is not significantly different in the two ensembles or basins. Therefore, our calculation of ΔE from the Potts model suggests that gatekeeper size does not directly contribute to the free energy difference between the two basins. Despite this, in Figure 1a we show that the gatekeeper size likely does effect type-II binding. To rationalize this, we note that the Potts calculations focus only on the conformational reorganization component of type-II binding. Meanwhile, the free-energy of binding to the already-reorganized receptor will include the effects of gatekeeper size due to direct interactions between the type-II inhibitors and the
gatekeeper residue. As the reviewer mentions there is a “…trend of small gatekeepers favoring classical DFG-out” in the sense that most kinase structures crystallized in classical DFG-out in the PDB have small gatekeepers. But the very large majority of those structures are co-crystallized with a type-II inhibitor. Our results suggest this trend is related to the more favorable interactions of type-II inhibitors with smaller gatekeeper residues in the reorganized DFG-out binding pocket, rather than the effect of gatekeeper size on the reorganization free energy in the absence of bound type-II inhibitors.

– Showing the location of the ‘gatekeeper’ residue in the structure in Figure 1 or elsewhere would help the reader to visualize the location of this frequently discussed residue. Also, there is not much discussion of the gatekeeper after Figure 1, even though it is clear that it plays a role in inhibitor binding.

We thank the reviewer for the suggestion to modify Figure 1. We have made changes in the Figure by showing the position of the gatekeeper residue to make it easier for readers to visualize. We would also like to address the reviewer’s concern regarding not much discussion of the gatekeeper residue in the manuscript after Figure 1. Our results using the Potts model to estimate the reorganization free energy from the active DFG-in aC-in basin to the classical DFG-out inactive basin are most strongly influenced by the sequence variation of the activation loop and the motifs that it directly interacts with, which does not include the gatekeeper residue. However, as we pointed out in Figure 1a, the type-II hit rates of kinases (especially TKs) with large vs small gatekeepers are significantly different, consistent with current understanding in the literature regarding the role of the gatekeeper in type-II binding. Dissecting the effect of the direct interactions between the gatekeeper residue and type-II inhibitors on binding affinities is outside the scope of our manuscript but will be an interesting subject of future research.

– Are there data for Type I inhibitors for TK’s to show in Figure 3? If dG is ~0 between DFG-in and DFG-out, those would also lie close to the y=x line.

It is true that Type-I inhibitors vs TKs would also lie close to the y=x line. Our reason for plotting type-I inhibitors on the same axis as type-II inhibitors for STKs is because of the large difference between ΔGexp and ΔGbind for type-II inhibitors, which we hypothesize is related to the large ΔGreorg between DFG-in and DFG-out for STKs. This required us to rule out other sources of unexpected systematic error unrelated to the DFG-in ↔ DFG-out equilibrium for STKs, hence this is why we have shown the plot of ΔGreorg ~ 0 for type-I inhibitors vs STKs but not TKs.

Reviewer #2 (Recommendations for the authors):

This paper focuses on an important topic. It explores how the activation loop conformations affect the type II inhibitor binding to Tyr and Ser/Thr kinases. The computational results agree with the available experimental data. It is a remarkably comprehensive, high quality paper. Some minor comments are as follows:

1. The work focuses on the classical “folded” DFG-out conformation of kinases. From the available PDB structures, most of DFG-out kinases have no well-defined "folded" activation
loop structures. In many of them, the activation loops after "DFG" motif have no complete density, which indicates that they are flexible and can be disordered. Given that "folded" for protein structures sometimes means that they form α-helix or β strand, the authors may consider changing "folded" into other descriptors or removing it.

We thank the reviewer for pointing out this potential confusion. We have added clarification for this nomenclature in the Introduction, stating that “folded activation loop” in our nomenclature collectively refers to activation loops which have undergone a large conformational change of ~17Å away from the active/extended state. This nomenclature has been used previously in the literature to describe the activation loop conformation in the classical DFG-out state.

2. Smaller gatekeepers and the sequence of the activation loop contribute to the binding of type II inhibitors in the pocket. The gatekeepers have direct contacts with the inhibitors, while the activation loop does not. Can the authors comment on which one would be more important for determining the type II inhibitor binding?

Indeed, gatekeeper size has a strong effect on type-II binding via direct interactions with the inhibitor; for example, it has been shown by other authors that mutating from small to large gatekeepers (e.g. T315I in Abl) increases the binding free-energy to the DFG-out kinase by >3 kcal/mol. Our observation in Figure 1a that gatekeeper size is correlated with type-II inhibitor hit appears consistent with this. However, we also show in Figure 1a that differences in gatekeeper size do not explain the large shift in type-II hit rates between TKs and STKs. From the results of our Potts model calculations, we proposed that residue substitutions in the activation loop rather than gatekeeper are largely responsible for this shift which is associated with the free energy cost to reorganize from the active DFG-in aC-in conformational basin to the inactive classical DFG-out basin.

3. The transition energies were calculated from the active DFG-in conformation to the inactive DFG-out conformation. Please note that the DFG-in conformation has two states as well, i.e., the inactive state (with aC-out) and the active state (with aC-in) (PMID: 35003591, PMID: 33971270). Some kinases barely show the DFG-out conformation (PMID: 32918948). The binding of the type II inhibitors in the pockets does not necessarily always go through the active DFG-in conformation (ac-in). It can also go through the inactive DFG-in conformation (aC-out) to DFG-out. Does the calculated energy data consider the alternative structural transitions?

The calculated Potts threaded energies consider two conformational basins only (active DFG-in aC-in, and inactive classical DFG-out). We have clarified this in the introduction and point out that the good correlation between ΔGreorg and Potts energies calculated in this way suggests the two-state model is sufficient for distinguishing the contribution of the reorganization free energy from the active DFG-in aC -in basin to the inactive classical DFG-out basin between TKs and STKs. As suggested in the literature, the inactive DFG-in aC-out conformation(also called Src-like inactive) is likely an intermediate on the pathway from DFG-in active to classical DFG-out inactive. This intermediate state may well affect the kinetics of the DFG-in active to classical DFG-out inactive transition, but not the thermodynamics of the transition between the initial DFG-in active and final classical DFG-out inactive free energy basins.
4. In Figure 5, please provide more details about the collected 10,345 tyrosine kinases and 210,862 serine/threonine kinases.

We thank the reviewer for pointing this out. In the revised version of the manuscript, we have added details about the sequence diversity and phylogenetics of these kinases in the Figure 5 caption.

5. How the activation loop sequences are aligned, and are they structurally validated?

This is an excellent question given that sequence alignments of flexible loop regions can be difficult to construct. The kinase domain sequences including activation loops were aligned starting from a PFAM seed, which was used with HHblits to align evolutionarily related kinases from the UniProt database. The aligned activation loops are 20 residues long with gaps and insertions assigned accordingly – typically this occurs near the center of the activation loop where there is significant flexibility and low homology. In general the activation loop alignments are most accurate near the N-terminal and C-terminal residues which are more evolutionarily conserved and not as flexible as the center-region of the activation loop. Our results with the Potts model are derived primarily from accurately aligned regions of the activation loop, which we have confirmed via structural superposition of important TKs and STKs.

Reviewer #3 (Recommendations for the authors):

I find the paper to be well-written and don’t have any major critiques. However, my minor concern is about the presentation of the subject in the Introduction. For a reader who is not familiar with the topic, it would sound that the "DFG-out" is the major inactive conformation. And although the authors do talk about “Src-like” confirmation and present distribution of inactive structures, it is in the Methods section or in Supplementary material. The method has its limits and it is better to describe them directly in the Introduction. A lot of information that would be helpful for a reader to understand the scope and the importance of the paper is moved to the Methods section which is almost 14 pages long, compared to 4 pages of Introduction. I would advise the authors to reconsider their Methods and move some important concepts into the Introduction.

As the reviewer suggests, we have moved the introduction of the Src-like inactive conformation and a description of other features of the kinase conformational landscape from the Methods section to the introduction. We also note that while the major inactive conformation in the absence of inhibitors may differ from kinase to kinase, the only inactive conformation compatible with the type-II binding mode is classical DFG-out, which is why we have placed significant focus on this free-energy basin. We hope that our revised introduction will give a more holistic view of the conformational landscape in the absence of inhibitors, but we emphasize that the classical DFG-out conformational free energy basin is the one which is populated when type-II inhibitors are bound.
Figures and figure supplements

Evolutionary divergence in the conformational landscapes of tyrosine vs serine/threonine kinases

Joan Gizzio and Abhishek Thakur et al.
Figure 1. Viewing the conformational landscape of the human kinome. (A) Hit rate distributions from kinome-wide experimental binding assays with type-II inhibitors for human serine/threonine kinases (STKs; blue, top) and tyrosine kinases (TKs; orange, bottom) with small gatekeepers (solid bars; sidechain volume <110 Å³) and large gatekeepers (hatched bars; sidechain volume >110 Å³). (B) PyMol (pymol, 2015) visualization of two conformational ensembles populated by Abl kinase from recent solution NMR (Nuclear Magnetic Resonance) experiments (Xie et al., 2020). The active DFG-in conformation where the activation loop is ‘extended’ (red, the Protein Data Bank [PDB]: 6XR6) and the inactive classical DFG-out conformation where the activation loop is ‘folded’ (blue, PDB: 6XRG) both exist in the absence of ligands, but there is a free-energy cost to transition between them (Xie et al., 2020). Type-II inhibitors preferentially bind to this folded DFG-out state. (C) Correlation between Potts DFG-out penalty (ΔEPotts) and hit rates for kinases with small gatekeepers only, to control for gatekeeper effects (Pearson correlation of −0.59, p<0.001). (D) Potts DFG-out penalties calculated for the human kinome and plotted using CORAL (Metz et al., 2018); the TK branch appears to have lower penalties relative to the rest of the kinome, which represent STKs. See Figure 1—source data 1 for values of the calculated type-II hit rates and Potts threaded-energy scores over the human kinome.
Figure 1—figure supplement 1. Conformational landscape of the activation loop derived from the Protein Data Bank (PDB). (B) Backbone torsions of the first three residues of the activation loop (XDF of the XDFG motif) broadly control folding of the activation loop into autoinhibitory conformations. The XDF dihedral classification (BLAminus, BLBplus, BBminus, etc.), which informs the positioning of catalytic residues at the activation loop N-terminus, correlates with the spatial trajectory of the first five residues of the activation loop (DFGXXXXX) measured by RMSD (Root Mean Square Deviation) with respect to catalytically primed Aurora A (PDB: 5DNR_A). Plotting the RMSD of the Asp and Phe residues of the DFG motif on the y-axis vs. RMSD of the DFGXXXXX residues on the x-axis captures three major conformational groups – an extended active-like conformation corresponding (circled orange) which is largely represented by the catalytically active BLAminus conformation, a partially folded DFG-in inactive conformation (circled red) which is mostly populated by BLBplus and commonly referred to as Src-like inactive, and the classical, folded DFG-out inactive conformation (circled pink) which is represented by the BBminus state. (B) Cartoon representation of the three major conformational clusters created in PyMol.
Figure 1—figure supplement 2. Contact frequency differences between the active DFG-in (BLAminus) and classical DFG-out (BBAminus) conformations. The contact frequency differences are explicitly involved in the calculation of ΔE Potts (see Methods). The upper triangle depicts the contact frequency difference matrix for kinases annotated in the Protein Data Bank as ‘tyrosine kinases (TKs)’, meanwhile the lower triangle was calculated from kinases annotated as ‘serine/threonine kinases (STKs)’. 
Figure 1—figure supplement 3. Gatekeeper analysis. Thousands of serine/threonine kinase (STK) and tyrosine kinase (TK) sequences (see legend) were analyzed to determine the identities of the gatekeeper residues. Distribution of gatekeeper residues (top) with bins ordered (left to right) according to increasing sidechain volume shows that STKs tend to have Met as a gatekeeper, whereas TKs tend to have Thr. Alternatively, one can look at the distribution of gatekeepers in terms of sidechain volume (bottom) which shows TKs have a greater tendency toward small gatekeepers in comparison with STKs.
Figure 2. Overview of the conformational landscapes between serine/threonine kinases (STKs) and tyrosine kinases (TKs) from absolute binding free-energy simulations, where we compare $\Delta G_{\text{bind}}$ (hatched bars) from binding free-energy simulations with $\Delta G_{\text{exp}}$ (solid bars) for the type-II inhibitors imatinib (the Protein Data Bank [PDB] code: STI) and BIRB-976 (PDB code: B96) vs several TKs (orange) and STKs (blue).
**Figure 2—figure supplement 1.** Absolute binding free-energy (ABFE) benchmarking results. Benchmarking the ABFE results over 26 different co-crystallized inhibitors bound to CDK2, JNK1, MAPK14 in the active state, and ABL1 in the inactive state. Mean unsigned error (MUE) and root mean square error (RMSE) were calculated with respect to the linear model where the intercept of the best-fit regression corresponds to the average difference between ABFE and experiment, and the slope is constrained to one.
Figure 3. Using type-II inhibitors as tools to probe the conformational landscape of TKs and STKs. (A) The average $\Delta G_{\text{reorg}}$ calculated via absolute binding free-energy (ABFE) simulations with 23 type-I (stars) and 23 type-II inhibitors (circles) complexes in the active DFG-in and inactive DFG-out state, respectively, computed from five serine/threonine kinase (STK) targets (Table 1) and (B) computed with 22 type-II inhibitors vs five tyrosine kinase (TK) targets in the DFG-out state (Table 1). (C) Kinome plot created with CORAL (Metz et al., 2018) illustrating the selection of five TKs and five STKs which are detailed in Table 1.
Figure 3—figure supplement 1. Experimental binding free-energy vs calculated absolute binding free-energy for type-I inhibitors vs serine/threonine kinases (STKs; column A), type-II inhibitors vs STKs (column B), and type-II inhibitors vs tyrosine kinases (TKs; column C). The linear regression of the data with the slope constrained to one is plotted, which gives the reorganization free energy as the intercept of this line (see legend). Mean unsigned error (MUE) and root mean square error (RMSE) were calculated with respect to this linear model.
Figure 4. Correlation between $\Delta E_{\text{Potts}}$ and averaged calculations of $\Delta G_{\text{reorg}}$ for five tyrosine kinases (TKs) and five serine/threonine kinases (STKs; Table 1).
Figure 5. The distributions of Potts conformational penalties for (orange) 10,345 tyrosine kinases (TKs) from 471 different species and (blue) 210,862 serine/threonine kinases (STKs) from 2713 different species, showing that TKs tend to have smaller energetic penalties on average. The difference in averages between these distributions is shown ($\Delta \Delta E = 3.2$), which we estimate to be $\approx 4.3$ kcal/mol based on the analysis in Figure 4. The probability density was plotted after down-weighting each sequence by the number of times another sequence in the same class (e.g. TKs) is observed within 40% identity (see Methods for details). The effective number of TKs ($N_{eff} = 1,096$) and STKs ($N_{eff} = 22,893$) is the sum totals of their down-weights, which is an unbiased measure of the sequence diversity in each probability distribution.
Figure 6. Molecular basis for the evolutionary divergence between tyrosine kinases (TKs) and serine/threonine kinases (STKs). Visualizing the top interaction pairs that contribute to the result in Figure 5, which is discussed in the main text. (A) Diagrams depicting general features of the active "extended" conformation of the activation loop (left) and the primary structure of these motifs in our multiple sequence alignment (MSA; right) with the HRD (pink), DFG (red), and APE (teal) motifs color-coded for reference. (B–C) Structural examples of a representative STK (CDK6) and TK (Abl) in the active DFG-in conformation (left) and the folded DFG-out conformation (right). Residues are labeled according to their position in our MSA, and colored according to A. The inset (center) displays the $\Delta E_{\text{potts}}$ of the reference kinase derived from Figure 5 as well as a cartoon depicting their location in the distributions. The diagrams of CDK6 in the active DFG-in conformation (PDB: 1XO2, chain B), CDK6 in the folded DFG-out conformation (PDB: 1G3N, chain A), Abl in the active DFG-in conformation (PDB: 2G2I, chain A), and Abl in the folded DFG-out conformation (PDB: 1IEP, chain A) were generated with PyMol (pymol, 2015). All ligands and some backbone atoms were hidden for clarity.
Figure 6—figure supplement 1. Interactions belonging to the top five contributions to ΔΔE. ΔΔE was decomposed to identify the strongest contributing pairs of positions (see Methods), with the idea that these pairs form interactions in the 3D structure that are important for controlling the conformational equilibrium. The residue pairs shown in this figure belong to the top five ΔΔEij (except for [128, 154], which is the 19th top contributing pair). For each position pair (a–h), the residue identities which make the largest contributions are shown. Location of each position along the kinase primary structure with consistent coloring is shown in (i) for reference. (j–k) The substrate mimicking behavior of the TK DFG-out activation loop, with the trans-autophosphorylating FGFR3 structure for reference (substrate activation loop colored green). (l) The interaction between positions (161, 166) forms part of the activation loop C-terminal bulge in TKs which, according to the Potts model, decreases the energetic cost for DFG-out relative to STKs (m) where formation of this interaction pair is constrained to the active conformation, and transition to DFG-out is penalized by breakage of this contact. PDBs: 1XO2 chain B (active CDK6), 2G2I chain A (active ABL), 4MNE chain B (active BRAF), 5HID chain A (inactive BRAF), 3LCD chain A (active CSF1R), 4R7I chain A (inactive CSF1R), 6PNX (trans-autophosphorylating FGFR3 dimer).
Figure 6—figure supplement 2. Cumulative sum of position pair contributions to ΔΔE. ΔΔE contributions (see Methods) were ordered from largest to smallest magnitude. The top pairs which are discussed in the main text and illustrated in Figure 6 are annotated. When the activation loop is in an active ‘extended’ conformation, the N-terminal anchor (Nolen et al., 2004) is formed by a β6 strand in the N-terminal of the activation loop (146385 to 148387) and a β9 strand near the N-terminal of the catalytic loop (119358 to 121142). In the inactive ‘folded’ conformation of the activation loop, the N-terminal anchor is completely broken as the β6 and β9 strands are peeled apart (Figure 6B and Figure 6C). The Potts model suggests this conformational change is energetically penalized in STKs due to favorable contacts between position 120359 and 146385 in the β6 strand and 120359 in the β9 strand; in 20% of serine/threonine kinases (STKs) we observe V120 and A146 at these positions (Figure 6B). These residues have a significantly more favorable Potts interaction energy than the TK-prevalent pair (F120, S146) which appears in 17% of tyrosine kinases (TKs) and only 1% of STKs (Figure 6C). As reported previously, position 120359 in TKs is a frequent site of drug resistance mutations to type-II inhibitors (Apperley, 2007; Hauser et al., 2018), particularly the F120I mutation (F359I in Abl numbering). This mutation in Abl changes the interaction pair from (F120, S146), seen in only 1% of STKs, to (I120, S146), which is seen in 15% of STKs. The mutant ‘STK-like’ interaction pair (I120, S146) is energetically more favorable according to the Potts model, suggesting that part of the Abl type-II inhibitor resistance mechanism of F120I (F359I) involves increasing the free-energy penalty for the active DFG-in/extended activation loop → classical DFG-out/folded activation loop transition. Consistent with this view, it was found that it is possible to treat patients with this drug-resistance mutation using dasatinib (Barańska et al., 2008; Tokarski et al., 2006): a type-I inhibitor that does not require access to the DFG-out conformation.