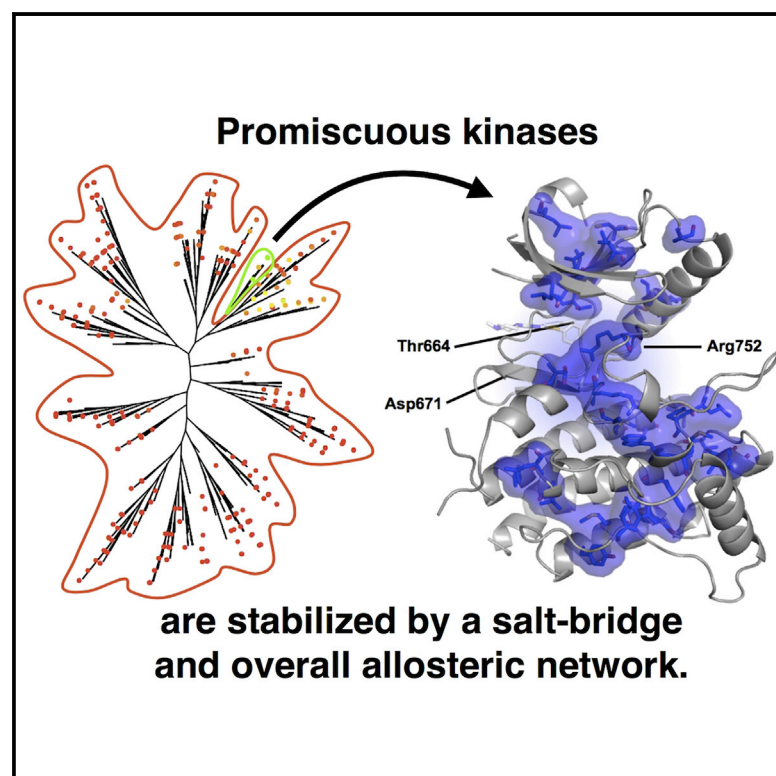# Cell Chemical Biology

## What Makes a Kinase Promiscuous for Inhibitors?

### Graphical Abstract

### Authors

Sonya M. Hanson, George Georghiou,
Manish K. Thakur, W. Todd Miller,
Joshua S. Rest, John D. Chodera,
Markus A. Seeliger

### Correspondence

john.chodera@choderalab.org (J.D.C.),
markus.seeliger@
stonybrook.edu (M.A.S.)

### In Brief

Hanson et al. identify a group of kinases
that bind a surprising number of inhibitors
and reveal the molecular basis for this
promiscuity. Functional clustering of
proteins identifies groups of co-inhibited
kinases. The findings have implications
for the development of specific kinase
inhibitors and drug resistance.

### Highlights

- Clustering of kinome by inhibition phenotype identifies
  promiscuous kinases

- Promiscuous tyrosine kinases are stable in the DFG-Asp-out
  inactive conformation

- Residues define promiscuous kinases and stabilize inactive
  conformation

CellPress

# Article

CellPress

# What Makes a Kinase Promiscuous for Inhibitors?

Sonya M. Hanson,[1,2,5] George Georghiou,[1,5] Manish K. Thakur,[1] W. Todd Miller,[3] Joshua S. Rest,[4] John D. Chodera,[2,*] and Markus A. Seeliger[1,6,*]

[1]Department of Pharmacological Sciences, Stony Brook University, Stony Brook, NY 11794-8651, USA
[2]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065-1115, USA
[3]Department of Physiology and Biophysics, Stony Brook University, Stony Brook, NY 11794-8651, USA
[4]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: john.chodera@choderalab.org (J.D.C.), markus.seeliger@stonybrook.edu (M.A.S.)
https://doi.org/10.1016/j.chembiol.2018.11.005

## SUMMARY

**ATP-competitive kinase inhibitors often bind several kinases due to the high conservation of the ATP binding pocket. Through clustering analysis of a large kinome profiling dataset, we found a cluster of eight promiscuous kinases that on average bind more than five times more kinase inhibitors than the other 398 kinases in the dataset. To understand the structural basis of promiscuous inhibitor binding, we determined the co-crystal structure of the receptor tyrosine kinase DDR1 with the type I inhibitors dasatinib and VX-680. Surprisingly, we find that DDR1 binds these type I inhibitors in an inactive conformation typically reserved for type II inhibitors. Our computational and biochemical studies show that DDR1 is unusually stable in this inactive conformation, giving a mechanistic explanation for inhibitor promiscuity. This phenotypic clustering analysis provides a strategy to obtain functional insights not available by sequence comparison alone.**

## INTRODUCTION

Protein kinases represent one of the largest enzyme families in the human genome and act as signaling mediators in a variety of cellular processes (Manning et al., 2002). Because many diseases are associated with aberrant protein kinase activity, targeted kinase inhibitors are clinically highly successful, such as imatinib in cancer therapy (Druker et al., 2006; Fabbro, 2015; Wu et al., 2016).

ATP-competitive kinase inhibitors can be classified by the conformation of the highly conserved Asp-Phe-Gly (DFG) motif of the kinase upon inhibitor binding (Zhao et al., 2014). Type I inhibitors bind to the active, DFG-Asp-in, conformation. Type II inhibitors bind an inactive conformation in which the aspartate of the DFG motif faces away from the active site into the bulk solvent (DFG-Asp-out) (Wodicka et al., 2010). While both type I and type II kinase inhibitors have been clinically successful, specific kinase inhibition remains challenging due to the high conservation of the ATP binding pocket (Zhang et al., 2009). For example,

the type I inhibitor dasatinib and the type II inhibitor imatinib bind 86 and 19 kinases out of 317, respectively (Karaman et al., 2008). While low inhibitor selectivity seems to be clinically tolerable for treating certain types of leukemia, inhibition of off-target kinases often limits the application of kinase inhibitors against solid tumors (Cohen et al., 2017; Eckstein et al., 2014; Lee and Wang, 2009).

For this reason, the specificity relationship between inhibitor and kinase is typically viewed from the perspective of the inhibitor (i.e., which kinases does a single inhibitor target?). Instead, examining this relationship from the perspective of the kinase (i.e., which inhibitors does an individual kinase bind?) can lead to a kinome-wide understanding of inhibitor binding behavior (Anastassiadis et al., 2011; Huang et al., 2010). To identify relationships among kinases as determined by the similarity of their inhibition phenotype—the profile of kinase inhibitors for which they have measurable affinity—we perform hierarchical clustering on a previously published dataset assessing the inhibition of 406 kinase constructs by 645 inhibitors (Drewry et al., 2017). We identify two groups of kinases with strikingly different promiscuity toward kinase inhibitors. The group of promiscuous kinases consists of eight Ser/Thr and Tyr kinases including established clinical targets (PDGFRA/B, KIT, and CSF1R), as well as kinases that are not prominent clinical targets (DDR1, DDR2, YSK4, and MEK5) (Wu et al., 2015). Importantly, kinases that are the target of many drug development programs such as EGFR, Abl, BRAF, and IGF1R are not part of this promiscuous group.

To determine the structural basis for promiscuity toward kinase inhibitors, we solved the co-crystal structure of DDR1 in complex with two type I inhibitors: the Aurora kinase inhibitor, VX-680, and the pan-tyrosine kinase inhibitor, dasatinib (Harrington et al., 2004; Lombardo et al., 2004). DDR1 is a receptor tyrosine kinase that binds to the extracellular matrix and is characterized by low kinase activity and slow activation kinetics. Surprisingly, our structures show that DDR1 binds both type I inhibitors in the DFG-Asp-out conformation, which is the binding conformation typically reserved for type II inhibitors. This suggests that DDR1 is stable in the DFG-Asp-out inactive conformation. This in itself is unusual since the first structures of kinases in the DFG-Asp-out conformation were considered to be induced by the high-affinity type II inhibitors (Nagar et al., 2002; Schindler et al., 2000). Here, we show that the DFG-Asp-out conformation is not only stable in DDR1 but facilitates promiscuous inhibitor
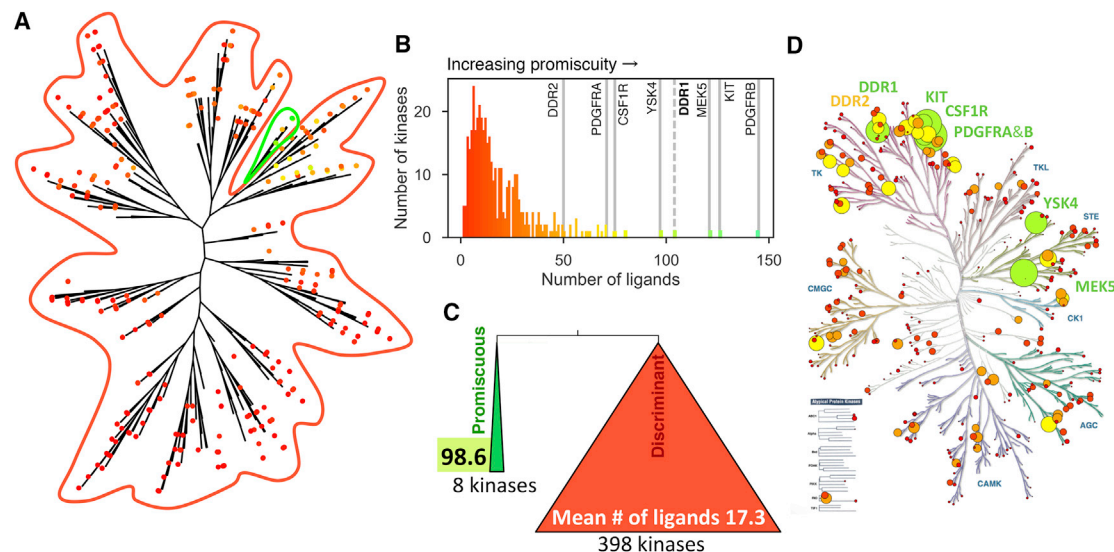
**Figure 1. Hierarchical Clustering of the Human Kinome by Inhibition Phenotype Reveals a Group of Highly Promiscuous Kinases**

(A) Hierarchical clustering analysis of the PKIS2 dataset assessing KinoBead binding inhibition for 406 kinases by 645 ligands at 1 μM concentration. Individual kinases are shown as dots and colored by the number of inhibitors capable of displacing more than 90% of kinase from covalently tethered pan-kinase inhibitors (green being most inhibitors, red being fewest inhibitors). The promiscuous branch is circled in green.

(B) The most basal branch of this dendrogram is a single branch of eight promiscuous kinases that bind a mean number of 98.6 ± 31.9 ligands (green) while the other 398 in the panel bind 17.3 ± 14.0 ligands (red) (mean ± SD).

(C) Number of kinases that are inhibited by given number of inhibitors to more than 90%.

(D) Superposition of the promiscuity of kinases onto the kinase phylogenetic tree. Circle diameter is proportional to kinase promiscuity for emphasis, and colors are as in (C).

See also Figures S1 and S2.

binding. Moreover, we find within the subset of promiscuous kinases a conserved salt bridge that stabilizes the DFG-Asp-out conformation. Disruption of this salt bridge shifts the population of DDR1 toward the active DFG-Asp-in conformation and increases the specific kinase activity 10-fold. The study provides an example of how large functional datasets can be used to group proteins by functional phenotype instead of sequence homology to elucidate a common mechanism. While here we define a single phenotypically different subgroup of kinases and its underlying mechanism, we expect that further analyses of this type will reveal additional phenotype-based kinase subgroups and mechanisms.

## RESULTS

### Phenotypic Clustering Identifies a Group of Unusually Promiscuous Kinases

Several screens of kinase inhibitor collections against large panels of kinases have been conducted in an effort to characterize the selectivity of kinase inhibitors and to determine which fraction of the kinome can be inhibited with existing inhibitors (Drewry et al., 2017; Fabian et al., 2005; Karaman et al., 2008; Klaeger et al., 2017). Hierarchical clustering has been performed on these datasets to identify patterns among ligands (Paricharak et al., 2013). Here, we perform hierarchical cluster analysis on a recently published dataset of 645 small-molecule inhibitors and 406 human kinases (392 wild-type kinases and 14 variants) (Drewry et al., 2017) to identify the relationships among kinases based on their binding phenotype to inhibitors. We obtain a tree

in which eight kinases form a cluster distant from all other kinases (Figures 1A and S1). When we assessed for each kinase the number of ligands that inhibit KinoBead binding by more than 90% at 1 μM inhibitor, the group of eight kinases includes the most promiscuous kinases in this panel (Figure 1B). We find that there is a wide spread of promiscuity ranging from extremely discriminant kinases (five kinases are inhibited by only one inhibitor each) to promiscuous kinases that are inhibited by more than 100 inhibitors. Strikingly, the group of eight promiscuous kinases binds a mean of 98.6 inhibitors (standard deviation [SD] = 31.9) per kinase whereas the group of 398 discriminant kinases binds a mean of 17.3 (SD = 14.0) inhibitors (Wilcoxon p value <0.0001) (Figure 1C).

The eight promiscuous kinases include PDGFRA, PDGFRB, KIT, CSF1R, DDR1, DDR2, MEK5, and YSK4 and are inhibited by 75, 145, 126, 75, 104, 50, 121, and 97 compounds, respectively (Figure 1B). One question is whether the number of bound inhibitors reflects the inhibitor development effort targeted against this kinase rather than promiscuity. The group of eight kinases includes both clinical targets (PDGFRA, PGFDRB, KIT, and CSF1R) and kinases of minor clinical interest (DDR1, DDR2, MEK5, and YSK4) (Wu et al., 2015). This indicates that these eight kinases are not inhibited by many compounds simply because they are the most prominent drug targets. Overlaying kinase promiscuity on the kinome phylogenetic tree clearly shows that our eight most promiscuous kinases do not share most recent common ancestry (Figure 1D).

Next, we set out to determine the structural and energetic basis for promiscuity toward kinase inhibitors. We focused on
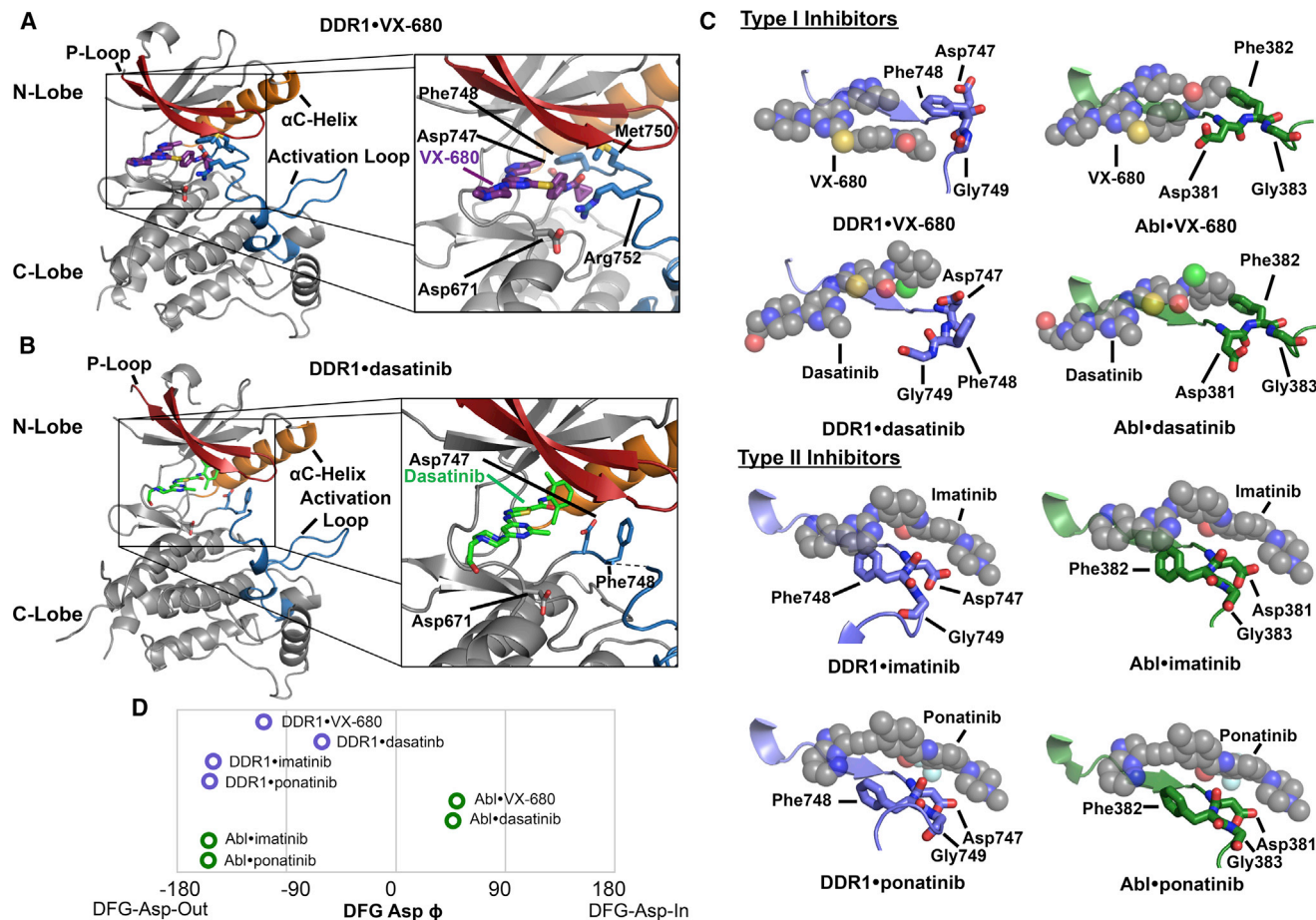
**Figure 2. DDR1 Binds Type I Inhibitors in the DFG-Asp-Out Conformation**

(A) Co-crystal structure of DDR1·VX-680 (PDB: 6BRJ) with activation loop (blue), phosphate binding P loop (red), and helix αC (orange).

(B) The co-crystal structure of DDR1·dasatinib (PDB: 6BSD) with regulatory elements colored as in (A).

(C) Comparison of type I (top, VX-680/dasatinib) and type II (bottom, imatinib/ponatinib) inhibitors binding to Abl (green, right column) and DDR1 kinase (blue, left column). Only the protein surrounding the DFG motif is shown for clarity.

(D) Comparison of Φ dihedral angle for Asp747 in DDR1 (blue) and Abl (green) bound to the type I and type II inhibitors in (C).

See also Figure S3.

DDR1 since it is a promiscuous kinase in various inhibitor screens and binds to the type II inhibitors imatinib and nilotinib, as well as the type I inhibitor VX-680 (Day et al., 2008; Karaman et al., 2008; Manley et al., 2010; Rix et al., 2010). DDR1 is a receptor tyrosine kinase that binds extracellular collagen and exhibits unusually slow activation kinetics and low kinase activity. DDR1 acts as a sensor for anchorage to the extracellular matrix, which is important for cellular adhesion, migration, and invasion. With regard to elucidating its role in disease, chemical biology efforts have resulted in DDR1-specific inhibitors (Canning et al., 2014; Deng et al., 2013; Kothiwale et al., 2015; Murray et al., 2015; Richters et al., 2014). These studies, however, have not addressed why DDR1 is so promiscuous. Here, we set out to understand the structural basis for promiscuity in DDR1.

## DDR1 Binds Both Type I and Type II Inhibitors in the Inactive Conformation

First, we wanted to determine whether DDR1 binds type I and type II kinase inhibitors in similar conformations as less promis-

cuous kinases. We therefore co-crystallized DDR1 with type I inhibitors, since previously published structures of DDR1 were only in complex with type II inhibitors (Canning et al., 2014; Deng et al., 2013; Murray et al., 2015). We expected that DDR1 would bind the Aurora kinase inhibitor VX-680 and the pan-tyrosine kinase inhibitor dasatinib in the DFG-Asp-in conformation observed with other kinases (Tokarski et al., 2006; Young et al., 2006). To our surprise, we found that DDR1 adopts in both complexes the inactive DFG-Asp-out conformation (DDR1 residues Asp747-Phe748-Gly749) typically observed in kinases binding type II inhibitors (Figures 2A and 2B, Table 1).

Since DDR1 binds VX-680 in the DFG-Asp-out conformation, we were curious to see whether the inhibitor interacted differently with DDR1 than with Abl and Aurora kinases. The interactions between VX-680 and DDR1 were similar to interactions seen with Abl and Aurora kinases; however, the hydrophobic cyclopropyl group of VX-680 was shielded by side chains of Phe748 and Met750, and a unique salt bridge formed between Asp671 and Arg752 (Figure 2A). In the equivalent structures of

**Table 1. Data Collection and Refinement Statistics (Molecular Replacement)**

| | DDR1·Dasatinib PDB: 6BSD | DDR1·VX-680 PDB: 6BRJ |
|---|---|---|
| Crystal Parameters | | |
| Space group | P2₁2₁2₁ | P2₁2₁2₁ |
| a, b, c (Å) | 61.7, 72.3, 74.7 | 61.5, 75.4, 77.2 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 |
| Data Collection Statistics | | |
| Wavelength (Å) | 1.075 | 1.075 |
| Resolution (Å) | 47.6–2.61 (2.64–2.61) | 48.1–2.23 (2.26–2.22) |
| $R_{merge}$ | 0.174 (1.00) | 0.115 (0.558) |
| I/σI | 13.3 (2.25) | 24.4 (2.97) |
| Completeness (%) | 99.8 (98.0) | 99.3 (93.04) |
| Redundancy | 13.2 (8.0) | 12.2 (4.5) |
| Refinement Statistics | | |
| Protein atoms | 2,218 | 2,183 |
| B factors (Å²) | 42.2 | 37.3 |
| Ligand/water | 33/61 | 33/104 |
| B factors (Å²) | 41.6/38.5 | 39.3/42.1 |
| $R_{work}/R_{free}$ | 0.1781/0.2552 | 0.1849/0.2225 |
| RMSD | | |
| Bond lengths (Å) | 0.009 | 0.014 |
| Bond angles (°) | 1.16 | 1.04 |
| Ramachandran Plot (% residues) | | |
| Most favored | 265 (98.2) | 264 (98.5) |
| Additional allowed | 5 (1.85) | 4 (1.49) |
| Disallowed regions | 0 | 0 |

Highest-resolution shell is shown in parentheses. RMSD, root-mean-square deviation.

Aurora and Abl kinases bound to VX-680, this hydrophobic moiety is shielded by the phosphate binding loop (P loop) folding over the active site (Adrián et al., 2006; Structural Genomics Consortium et al., 2008; Elkins et al., 2012; Salah et al., 2011).

Similar to our DDR1·VX-680 structure, dasatinib also binds DDR1 in the DFG-Asp-out conformation with similar interactions between the drug and DDR1 as seen in structures of Abl kinase bound to dasatinib (Figures 2B and 2C) (Shah et al., 2004). The DFG motif in DDR1·dasatinib is rotated in comparison with DDR1·VX-680, suggesting that the DFG motif may be in an intermediate state. Furthermore, due to the lack of electron density, we were unable to build in residues following the DFG motif: Met750, Ser751, and Arg752. Some electron density is present where the side chain of Arg752 would fall and, when aligned to the activation loop of DDR1·VX-680, suggests that the salt bridge between Asp671 and Arg752 would be intact.

When bound to type II inhibitors (imatinib and ponatinib), the conformation of the DFG motif of DDR1 resembles the DFG-Asp-out conformation in Abl. However, the DFG-Asp-out conformation in DDR1·VX-680 and DDR1·dasatinib differs from the conformations of Abl bound to either type I or type II inhibitors (Figures 2C and 2D). Similar to other structures of kinases in the DFG-Asp-out conformation, such as Abl kinase bound to imatinib, the salt bridge between the catalytic lysine and the

αC helix is intact in DDR1. Both type I and type II inhibitors form similar interactions with DDR1 as they do in other kinases, including in particular the hinge region of the kinase (the loop connecting the N and C lobes) (Zhang et al., 2009). DDR1 is able to accommodate the larger type II kinase inhibitors by allowing inhibitors to bind in a hydrophobic pocket at the back of the active site that is located under the αC helix and only accessible in the DFG-Asp-out inactive conformation (Canning et al., 2014; Murray et al., 2015; Zhang et al., 2009). However, in the case of type I inhibitors, DDR1 binds them in the DFG-Asp-out conformation and can adjust the position of the DFG motif to accommodate them (Figures 2C and 2D). This curious behavior suggests that the inactive conformation of DDR1 is relatively stable, and that accessing the active DFG-Asp-in conformation may be energetically unfavorable for DDR1, which also helps explain the low cellular kinase activity (Leitinger, 2014).

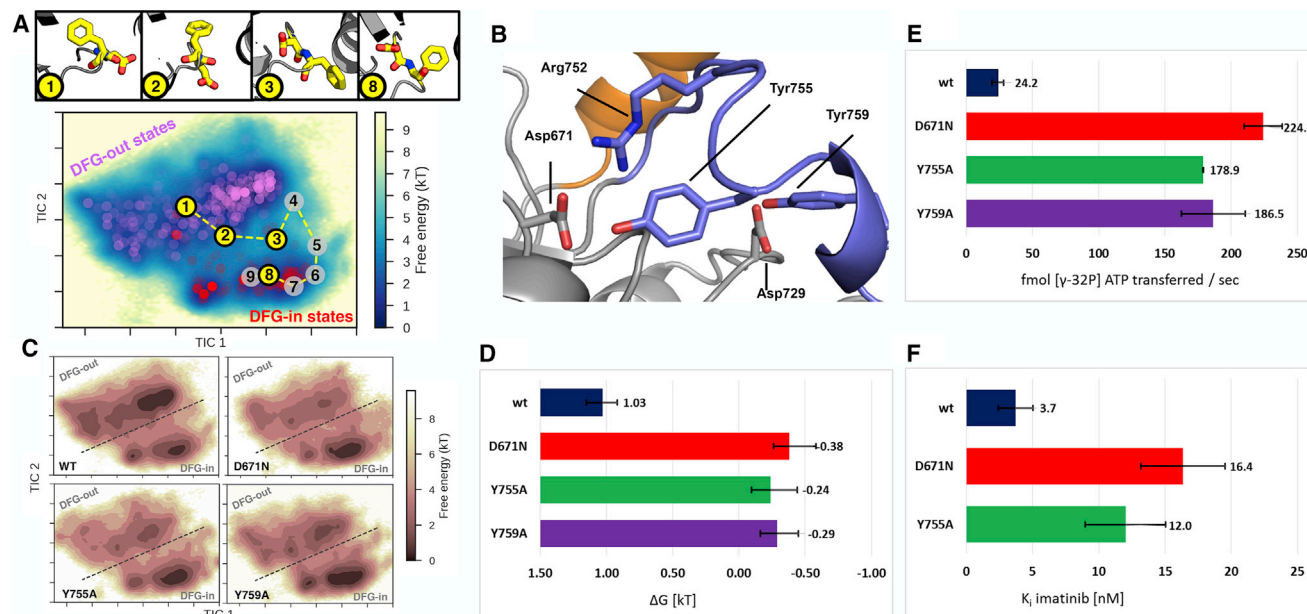### DDR1 Is Stable in the DFG-Asp-Out Conformation

To determine the stability of the DFG-Asp-out conformation in DDR1, we performed massively distributed molecular dynamics (MD) simulations starting from available crystal structures of DDR1, all of which are in the DFG-Asp-out conformation. Simulations were initiated from the VX-680 bound structure reported here, as well as previously published structures of DDR1 in complex with ponatinib and imatinib (PDB: 3ZOS and 4BKJ, respectively) (Canning et al., 2014). In these initial ~450-μs aggregate simulations of DDR1 protonated at Asp747, we observed four independent transitions from the DFG-Asp-out to the DFG-Asp-in conformation (Figure 3A).

To collect a larger number of DFG transitions, we initiated new simulations from snapshots of these four simulations in which the transition was observed. The resulting free energy landscape for DDR1 wild-type clearly shows a preference of ~1 kT for the DFG-Asp-out conformation over the DFG-Asp-in conformation (Figures 3A and 3C).

### Mutations Destabilize the DFG-Asp-Out Conformation and Increase Kinase Activity

Next, we wanted to understand the biophysical basis for the stability of the DFG-Asp-out conformation. We noticed in the structure of DDR1·VX-680 that a network of salt bridges and hydrogen bonds in DDR1 around Asp671, Arg752, Tyr755, and Tyr759 were likely stabilizing the DFG-Asp-out conformation (Figure 3B). We therefore performed MD simulations to determine the free energy landscape for three DDR1 constructs in which the stabilizing interactions were disrupted by mutation (D671N, Y755A, or Y759A) (Figure 3C). We find that all mutants are more stable in the DFG-Asp-in conformation by about ~0.5 kT (Figure 3D).

We hypothesized that these mutations disrupt the inactive (DFG-Asp-out) conformation and increase kinase activity. Indeed, we found that the DDR1 D671N mutant was 9.3-fold more active in biochemical kinase assays compared with wild-type DDR1 (activity increased from 24 fmol s⁻¹ to 224 fmol s⁻¹) (Figure 3E). This increase in activity suggests that the Asp671-Arg752 salt bridge is critical for the stabilization of the inactive conformation and that destabilizing it allows the kinase to more readily adopt the active conformation. The Y755A and Y759A mutants also showed an increase in activity compared with

**Figure 3. Wild-Type DDR1 Is Stable in the DFG-Asp-Out Conformation, and Mutations Destabilize this Inactive Conformation**

(A) Snapshots from one of the trajectories seen to flip from DFG-Asp-out to DFG-Asp-in superimposed on the final free energy landscape for wild-type (WT) DDR1 are highlighted in yellow. hidden Markov Model macrostates representing DFG-in (red) and DFG-out (purple) states are shown, and transparency is proportional to the membership of each k-means cluster center to the macrostate.

(B) The Asp671-Arg752 salt bridge, Asp671-Tyr755 hydrogen bond, and Asp729-Tyr759 interaction thought to stabilize the inactive DFG-Asp-out conformation.

(C) Free energy landscapes for WT, D671N, Y755A, and Y759A simulations superimposed onto the first two (slowest) time-lagged independent components of the WT simulations. Dotted line corresponds to the linear separation between states seen in (A).

(D) Free energy difference between DFG-out and DFG-in states calculated from the free energy landscapes in (C). Positive values indicate stabilization in the DFG-out conformation; negative values indicate stabilization in the DFG-in conformation. Error bars indicate 95% confidence intervals from a Bayesian Markov state model with 1,000 samples.

(E) Kinase activity assays of DDR1. Values are presented as mean values ± SEM.

(F) Binding affinity of DDR1 wt and mutant proteins for the type II inhibitor imatinib. Binding affinity for imatinib is reported as inhibitory constant $K_i$ for competition with a general kinase inhibitor. Values are presented as mean values ± SEM.

See also Figure S4 and Table S1.

wild-type DDR1 (7.5-fold and 7.8-fold increase, respectively), implicating them in the stabilization of the inactive DDR1 conformation (Figure 3E). Interestingly, the fold change in kinase activation upon mutation is consistent with the ~5-fold change in DFG-Asp-in conformation predicted by the simulations (Figure 3D).

If the mutations shift the population of DDR1 from the inactive DFG-Asp-out to the active DFG-Asp-in conformation, we would predict that the mutant proteins bind less tightly to type II inhibitors such as imatinib that require the DFG-Asp-out conformation for binding. Consistent with this model, we find that the tested mutants bind imatinib 3- to 4-fold less tightly (Figure 3F).

## DISCUSSION

In this study we grouped protein kinases by their inhibition phenotype toward a set of 645 kinase inhibitors. We identified eight promiscuous kinases that bind on average five times more inhibitors than the other kinases, and we selected DDR1 as a model kinase in which to explore the potential mechanism of this promiscuous behavior.

The sequence homology-based clustering of the human kinome by Manning et al. (2002) has been extremely useful to organize kinases into families by sequence relationship. Since kinases are such prominent drug targets, we expect that clustering of the kinases by their inhibition phenotype could allow medicinal chemists to quickly visualize what kinases and signaling cascades are often co-inhibited. This classification could reveal interactions relevant to polypharmacology or negative side effects.

Importantly, promiscuity is not only a feature of kinase but may also be affected by the specificity of inhibitors tested to identify promiscuous kinases. The inhibitors studied here were chosen to be selective probes, and their selectivity scores are superior to those of the average clinically approved kinase inhibitor. However, when we repeat this clustering analysis with independent inhibitor sets including clinically approved drugs, we identify essentially the same promiscuous kinases (Figure S2B).

Similarly, clusters of kinases can be analyzed for their shared preference toward certain inhibitor scaffolds. For example, we asked which chemical scaffolds preferentially inhibit the group of eight promiscuous kinases. We found that 40 ligands bound significantly more tightly to the promiscuous kinases than the rest. These inhibitors were enriched for scaffolds containing: 2,4-dianilinopyrimidine, 4-anilinoquinoline, 2-aminobenzimidazole, 6-phenoxy-imidazopyridazine, and 4-anilino or phenoxy pyrrolopyrimidine. Further analysis of these and other ligands that target certain clades of kinases, clustered by phenotype
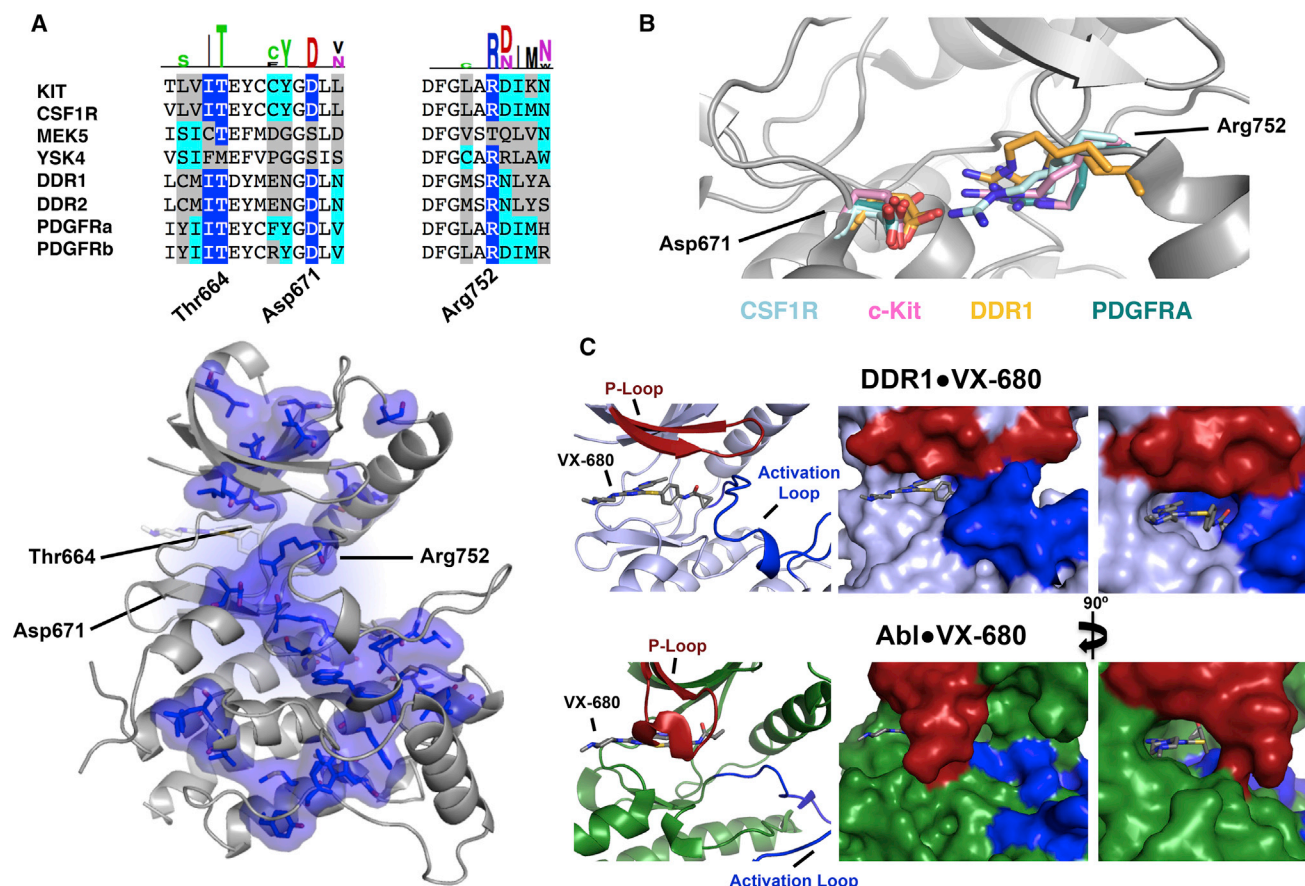
**Figure 4. Structural Hallmarks of DDR1 and Other Promiscuous Kinases**

(A) Sequence alignment of the eight promiscuous kinases around Thr664 (the "gatekeeper" mutant), Asp671, and Arg752 (DDR1 numbering). Dark blue highlights residues enriched in the promiscuous set by 50% or more and cyan indicates residues enriched by 10%–50% according to Two Sequence Logo comparison (using p value cutoff of 0.01) of the promiscuous kinases with the other 483 kinases analyzed. Bottom panel: All residues enriched in our promiscuous set by 50% or more are shown as blue surface on the DDR1-VX680 structure.

(B) Potential salt bridges between residues equivalent to DDR1 Asp671 and Arg752 are shown for representative structures: CSF1R (PDB: 3KRL and 2IOV) in pale cyan, c-Kit (PDB: IT45 and IT46) in pink, DDR1 (PDB: 5FDP and 5BVO) in orange, and PDGFRA (PDB: 5GRN) in teal.

(C) The structures of VX-680 bound to DDR1 (top) and Abl (bottom, PDB: 2F4J) illustrate the hydrophobic shield formed by the activation loop (blue) and the P loop (red), respectively.

See also Figure S5.

and not genotype, will improve our understanding of targeted kinase inhibitor design.

Since the DFG-Asp-out conformation appears to facilitate the promiscuity of DDR1, we wanted to understand whether other promiscuous kinases are also stable in the DFG-Asp-out conformation. Therefore, we compared the kinase domain sequences of the eight promiscuous kinases with the rest of the human kinome using the Two Sample Logo server (Manning et al., 2002; Vacic et al., 2006). This analysis revealed that residues stabilizing the DFG-Asp-out conformation in DDR1 (Asp671 and Arg752) are enriched among the promiscuous tyrosine kinases (Figures 4A and S5), predicting that this salt bridge could stabilize the DFG-Asp-out conformation also in the other promiscuous kinases. In fact, we found that 25 of the 27 available crystal structures of promiscuous kinases in the PDB are in the DFG-Asp-out conformation and form this salt bridge (Figure 4B and Table S2) (Illig et al., 2011; Mol et al., 2004; Murray et al., 2015; Walter et al., 2007; Wang et al., 2016). This includes the apo

structures of CSF1R, KIT, and PGDFRA showing that the DFG-Asp-out conformation is energetically accessible in the absence of ligands (Liang et al., 2016; Mol et al., 2004; Walter et al., 2007). It will be interesting to determine the structural basis for the two promiscuous non-tyrosine kinases MEK5 and YSK4. They both lack the equivalent to the Asp671-Arg752 (DDR1 numbering) salt bridge, and currently no structure is available for these kinases.

While we identify DDR1 here as a promiscuous kinase, DDR1 can still be inhibited with selective inhibitors. The salt bridge formed between the activation loop and the C lobe of DDR1 forms a hydrophobic cage that facilitates the binding of many chemically different inhibitors. However, this unique hydrophobic cage has been exploited by highly selective inhibitors for DDR1, such as DDR1-IN-1 (Canning et al., 2014). Interestingly, DDR1-IN-1 bypasses interactions with the threonine gatekeeper residue in DDR1 and instead forms a hydrogen bond with the αD helix in the C lobe of DDR1. This interaction is only possible due

to anchoring of the activation loop to the C lobe by the Asp671-Arg752 salt bridge in the DFG-Asp-out conformation adapted by DDR1.

In addition to providing insights about promiscuous kinases in general, the results presented here suggest a molecular mechanism for the unusually low kinase activity of DDR1 (Leitinger, 2014). Among promiscuous kinases, Tyr755 is only present in DDR1 and DDR2. This tyrosine is part of DDR1/2's YxxxYY motif, which is an autoinhibitory motif conserved between DDR1/2 and the insulin receptor kinase (IRK) family (Artim et al., 2012; Leitinger, 2014). This autoinhibition is released upon tyrosine phosphorylation at this motif. Unlike the promiscuous tyrosine kinases, the IRK family kinases do not contain the Asp-Arg salt bridge. Consequently, they are catalytically more active but less promiscuous than DDR1. The presence of both the Tyr755 and the Asp-Arg salt bridge explains why, despite other sequence similarities, DDR1 is less active than both other promiscuous kinases and the kinases of the IRK family (Wei et al., 1995).

Consistent with previous simulations on the DFG-Asp-out/-in interconversion of Abl kinase, we only observe the DFG flip with protonated Asp747 (Shan et al., 2009). We showed previously that the $pK_a$ for the DFG-Asp in Abl is elevated at 6.5. Further "constant pH" simulations whereby protonation events can happen freely would be required to predict the $pK_a$ of Asp747 in DDR1 and the absolute populations of DFG-in/-out more precisely (Radak et al., 2017). Importantly, our simulations predict the change in DFG populations (5-fold), in good agreement with the 7- to 10-fold increase in biochemical kinase activity.

Interestingly, we observed that during the DFG-Asp-out/-in interconversion in the DDR1 simulations the αC helix remained in the "in" conformation. This is different from other kinases, e.g., Abl, in which the αC helix transiently moves out of the way for the DFG flip to occur (Shan et al., 2009). We speculate that differences in the mechanism of the DFG flip could lead to differences in the kinetics of the DFG flip. For Abl kinase, the transition of the protonated DFG aspartate through a hydrophobic pocket accessible upon αC helix movement can become the rate-limiting step for the DFG flip (Shan et al., 2009). Since the DFG flip in DDR1 appears to occur without movement of the αC helix, this might indicate the absence of this rate-limiting step.

To examine the relationship of the relative stability of the DFG-Asp-out conformation to the promiscuous property of DDR1 and other kinases, we compared the conformations of DDR1·VX-680 and Abl·VX-680. As stated earlier, Abl binds VX-680 in the active conformation, where the activation loop extends outward, but the P loop of the kinase folds over to shield the inhibitor from the solvent. In DDR1 the kinase is in the inactive conformation, and instead of the P loop the activation loop provides the hydrophobic contact VX-680 needs to bind (Figure 4C). Similarly, in structures of KIT, DDR1, and CSF1R bound to imatinib, the activation loop provides the hydrophobic shield that the kinked P loop of Abl provides.

To test our hypothesis that the relative stability of the DFG-Asp-out conformation is related to promiscuity, we looked at data available in the PKIS2 set for a kinase that was tested in a state stabilized in DFG-Asp-in and the DFG-Asp-out conformation. We found data for activation loop phosphorylated and non-phosphorylated Abl kinase. Phosphorylated Abl is stabilized in the DFG-Asp-in active conformation (Hari et al., 2013). The number of compounds Abl binds with high affinity increases in the non-phosphorylated state, in which the DFG-Asp-out conformation is favored (Figure S2C). This implies that also for Abl kinase, the relative stability of the DFG-Asp-out conformation correlates with promiscuity.

How could disease-related mutations affect the stability of the DFG-Asp-out conformation and the ability of kinases to bind ligands with high affinity? The Two Sample Logo analysis identified residues across the entire kinase domain that are specific to the group of promiscuous kinases. These residues could affect the overall stability of the DFG-Asp-out conformation, and in turn affect the ability of the kinase to bind ligands (Figure 4A). We found that seven of them correspond to sites of clinical mutations that confer imatinib resistance in Abl (Azam et al., 2003) (Figure S5B). Two of these residues (Abl resistance mutants M370T/I and M491I) are distant from the imatinib binding site and the mechanism of their resistance is unclear. Our model suggests that mutations at these sites destabilize the DFG-Asp-out conformation and thereby weaken the affinity for imatinib. In addition, the D681N/Y/G mutations in PDGFRA (Asp671 of the salt bridge in DDR1) are activating mutations that confer resistance to imatinib and sunitinib (COSMIC Study: COSU419, COSU375) (Zehir et al., 2017). This is consistent with our previous finding that a distributed network of residues stabilizes the DFG-Asp-out conformation in kinases (Seeliger et al., 2007). Mutating these residues reduces the stability of the DFG-Asp-out conformation, conferring resistance to ligands such as imatinib that favor this conformation. Similarly, when we compare the sequences of DDR1 and the less promiscuous DDR2, we find that no amino acids within 5 Å of the inhibitors differ between DDR1 and DDR2. This indicates that in fact differences in secondary shell or even more remote residues underlie the difference in promiscuity between DDR1 and DDR2 (Figures S5C–S5E).

Our data indicate that the ability of these promiscuous kinases to bind chemically diverse inhibitors is defined by the hydrophobic pocket formed by the activation loop, which is only accessible in the DFG-Asp-out conformation. Inhibitors do not artificially induce the DFG-Asp-out conformation as was once surmised, but it is a stable, accessible conformation of kinases. Analysis of the available apo structures of the promiscuous kinases PDGFRA, c-Kit, and CSF1R and our MD analysis of DDR1 show that the DFG-Asp-out inactive conformation is the preferred conformation of this unusual set of kinases (Table S2). This analysis of a large chemical genomic dataset identified a phenotypically distinct class of medically important signaling enzymes. Further analysis revealed a shared structural mechanism underlying this characteristic that would not have been obtained by sequence comparison alone. We speculate that future analysis of similar datasets will yield further insight into the function, regulation, and druggability of enzymes.

## SIGNIFICANCE

**In this study we have analyzed a publicly available dataset of protein kinase inhibitors. Rather than analyzing the specificity of kinase inhibitors as is typically done to identify which off-target kinases an inhibitor will affect, we analyzed**

**CellPress**

how kinases differ in their ability to bind a set of fairly specific inhibitors. We were able to group the 400 kinases into families with similar inhibitor binding properties. To our surprise, a group of eight kinases was significantly more promiscuous than the rest. We found that the stability of the DFG-Asp-out inactive conformation underlies the promiscuity of these kinases. As more large-scale functional datasets become available, we expect that grouping of proteins by functional conservation (e.g., with respect to inhibitor binding) will complement the insight from grouping by sequence conservation to reveal how conserved structural features underlie the function of proteins. In the case of kinases, such studies may aid the development of better therapeutics (e.g., by identifying clusters of commonly co-inhibited kinases) and understanding of the mechanism of resistance mutations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Hierarchical Clustering of a Large Kinase Inhibitor Set
  - Cloning and Constructs
  - Protein Expression and Purification
  - DDR1 Crystallization
  - X-ray Diffraction Data and Processing
  - DDR1 Kinase Activity Assays
  - DDR1 Drug Binding Assay
  - Molecular Dynamics Simulation of DDR1 Wild-Type
  - Molecular Dynamics of DDR1 Mutants and Wild-Type Started along the DFG Flip
  - Molecular Simulation Analysis Using Markov Models
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

### SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and two tables and can be found with this article online at https://doi.org/10.1016/j.chembiol.2018.11.005.

### AUTHOR CONTRIBUTIONS

Conceptualization: J.D.C., M.A.S., G.G., S.M.H.; Formal Analysis: J.D.C., M.A.S., G.G., S.M.H., W.T.M., J.S.R.; Resources: M.A.S., J.C., W.T.M., G.G., M.K.T.; Writing – Original Draft: M.A.S., S.M.H., G.G., J.D.C.; Writing – Reviewing & Editing: M.A.S., J.C., S.M.H., G.G.; Investigation: G.G., S.M.H., M.K.T.; Visualization: M.A.S., S.M.H., G.G., M.K.T., J.D.C.; Supervision: M.A.S., J.D.C.; Funding acquisition: J.D.C., M.A.S., W.T.M., J.S.R.

### REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., Mccoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Adrián, F.J., Ding, Q., Sim, T., Velentza, A., Sloan, C., Liu, Y., Zhang, G., Hur, W., Ding, S., Manley, P., et al. (2006). Allosteric inhibitors of Bcr-abl-dependent cell proliferation. Nat. Chem. Biol. *2*, 95–102.

Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H., and Peterson, J.R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nat. Biotechnol. *29*, 1039–1045.

Artim, S.C., Mendrola, J.M., and Lemmon, M.A. (2012). Assessing the range of kinase autoinhibition mechanisms in the insulin receptor family. Biochem. J. *448*, 213–220.

Azam, M., Latek, R.R., and Daley, G.Q. (2003). Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. Cell *112*, 831–843.

Canning, P., Tan, L., Chu, K., Lee, S.W., Gray, N.S., and Bullock, A.N. (2014). Structural mechanisms determining inhibition of the collagen receptor DDR1 by selective and multi-targeted type II kinase inhibitors. J. Mol. Biol. *426*, 2457–2470.

Casnellie, J.E. (1991). Assay of protein kinases using peptides with basic residues for phosphocellulose binding. Methods Enzymol. *200*, 115–120.

Cheng, Y., and Prusoff, W.H. (1973). Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem. Pharmacol. *22*, 3099–3108.

Cohen, N.A., Kim, T.S., and DeMatteo, R.P. (2017). Principles of kinase inhibitor therapy for solid tumors. Ann. Surg. *265*, 311–319.

Day, E., Waters, B., Spiegel, K., Alnadaf, T., Manley, P.W., Buchdunger, E., Walker, C., and Jarai, G. (2008). Inhibition of collagen-induced discoidin domain receptor 1 and 2 activation by imatinib, nilotinib and dasatinib. Eur. J. Pharmacol. *599*, 44–53.

Deng, X., Elkins, J.M., Zhang, J., Yang, Q., Erazo, T., Gomez, N., Choi, H.G., Wang, J., Dzamko, N., Lee, J.-D., et al. (2013). Structural determinants for ERK5 (MAPK7) and leucine rich repeat kinase 2 activities of benzo[e]pyrimido-[5,4-b]diazepine-6(11H)-ones. J. Biotechnol. *70*, 758–767.

Drewry, D.H., Wells, C.I., Andrews, D.M., Angell, R., Al-Ali, H., Axtman, A.D., Capuzzi, S.J., Elkins, J.M., Ettmayer, P., Frederiksen, M., et al. (2017). Progress towards a public chemogenomic set for protein kinases and a call for contributions. PLoS One *12*, e0181585.

Druker, B.J., Guilhot, F., O'Brien, S.G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M.W.N., Silver, R.T., Goldman, J.M., Stone, R.M., et al. (2006). Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N. Engl. J. Med. 355, 2408–2417.

Eastman, P., Friedrichs, M.S., Chodera, J.D., Radmer, R.J., Bruns, C.M., Ku, J.P., Beauchamp, K.A., Lane, T.J., Wang, L.P., Shukla, D., et al. (2013). OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theory Comput. 9, 461–469.

Eastman, P., Swails, J., Chodera, J.D., Mcgibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.-P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol. 13, e1005659.

Eckstein, N., Röper, L., Haas, B., Potthast, H., Hermes, U., Unkrig, C., Naumann-Winter, F., and Enzmann, H. (2014). Clinical pharmacology of tyrosine kinase inhibitors becoming generic drugs: the regulatory perspective. J. Exp. Clin. Cancer Res. 33, 15.

Eid, S., Turk, S., Volkamer, A., Rippmann, F., and Fulle, S. (2017). KinMap: a web-based tool for interactive navigation through human kinome data. BMC Bioinformatics 18, 16.

Elkins, J.M., Santaguida, S., Musacchio, A., and Knapp, S. (2012). Crystal structure of human aurora B in complex with INCENP and VX-680. J. Med. Chem. 55, 7841–7848.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Fabbro, D. (2015). 25 years of small molecular weight kinase inhibitors: potentials and limitations. Mol. Pharmacol. 87, 766–775.

Fabian, M.A., Biggs, W.H., Treiber, D.K., Atteridge, C.E., Azimioara, M.D., Benedetti, M.G., Carter, T.A., Ciceri, P., Edeen, P.T., Floyd, M., et al. (2005). A small molecule-kinase interaction map for clinical kinase inhibitors. Nat. Biotechnol. 23, 329–336.

Fiser, A., and Sali, A. (2003). Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol. 374, 461–491.

Hari, S.B., Perera, B.G., Ranjitkar, P., Seeliger, M.A., and Maly, D.J. (2013). Conformation-selective inhibitors reveal differences in the activation and phosphate-binding loops of the tyrosine kinases Abl and Src. Acs Chem. Biol. 8, 2734–2743.

Harrington, E.A., Bebbington, D., Moore, J., Rasmussen, R.K., Ajose-Adeogun, A.O., Nakayama, T., Graham, J.A., Demur, C., Hercend, T., Diu-Hercend, A., et al. (2004). VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. Nat. Med. 10, 262–267.

Hu, J., Ahuja, L.G., Meharena, H.S., Kannan, N., Kornev, A.P., Taylor, S.S., and Shaw, A.S. (2015). Kinase regulation by hydrophobic spine assembly in cancer. Mol. Cell Biol. 35, 264–276.

Huang, D., Zhou, T., Lafleur, K., Nevado, C., and Caflisch, A. (2010). Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. Bioinformatics 26, 198–204.

Illig, C.R., Manthey, C.L., Wall, M.J., Meegalla, S.K., Chen, J., Wilson, K.J., Ballentine, S.K., Desjarlais, R.L., Schubert, C., Crysler, C.S., et al. (2011). Optimization of a potent class of arylamide colony-stimulating factor-1 receptor inhibitors leading to anti-inflammatory clinical candidate 4-cyano-N-[2-(1-cyclohexen-1-yl)-4-[1-[(dimethylamino)acetyl]-4-piperidinyl]pheny l]-1H-imidazole-2-carboxamide (JNJ-28312141). J. Med. Chem. 54, 7860–7883.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958.

Karaman, M.W., Herrgard, S., Treiber, D.K., Gallant, P., Atteridge, C.E., Campbell, B.T., Chan, K.W., Ciceri, P., Davis, M.I., Edeen, P.T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. Nat. Biotechnol. 26, 127–132.

Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.-A., Reinecke, M., Ruprecht, B., Petzoldt, S., Meng, C., et al. (2017). The target landscape of clinical kinase drugs. Science 358, https://doi.org/10.1126/science.aan4368.

Kothiwale, S., Borza, C.M., Lowe, E.W., Pozzi, A., and Meiler, J. (2015). Discoidin domain receptor 1 (DDR1) kinase as target for structure-based drug discovery. Drug Discov. Today 20, 255–261.

Lee, S.J., and Wang, J.Y.J. (2009). Exploiting the promiscuity of imatinib. J. Biol. 8, 30.

Leitinger, B. (2014). Discoidin domain receptor functions in physiological and pathological conditions (Elsevier), pp. 39–87.

Liang, L., Yan, X.-E., Yin, Y., and Yun, C.-H. (2016). Structural and biochemical studies of the PDGFRA kinase domain. Biochem. Biophys. Res. Commun. 477, 667–672.

Lombardo, L.J., Lee, F.Y., Chen, P., Norris, D., Barrish, J.C., Behnia, K., Castaneda, S., Cornelius, L.A., Das, J., Doweyko, A.M., et al. (2004). Discovery of N-(2-chloro-6-methyl- phenyl)-2-(6-(4-(2-hydroxyethyl)- piperazin-1-yl)-2-methylpyrimidin-4- ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. J. Med. Chem. 47, 6658–6661.

Manley, P.W., Drueckes, P., Fendrich, G., Furet, P., Liebetanz, J., Martiny-Baron, G., Mestan, J., Trappe, J., Wartmann, M., and Fabbro, D. (2010). Extended kinase profile and properties of the protein kinase inhibitor nilotinib. Biochim. Biophys. Acta 1804, 445–453.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science 298, 1912–1934.

Mccoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. 61, 458–464.

Meng, Y., Lin, Y.L., and Roux, B. (2015). Computational study of the "DFG-flip" conformational transition in c-Abl and c-Src tyrosine kinases. J. Phys. Chem. B 119, 1443–1456.

Mobitz, H. (2015). The ABC of protein kinase conformations. Biochim. Biophys. Acta 1854, 1555–1566.

Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.C., and Wilson, K.P. (2004). Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. J. Biol. Chem. 279, 31655–31663.

Murray, C.W., Berdini, V., Buck, I.M., Carr, M.E., Cleasby, A., Coyle, J.E., Curry, J.E., Day, J.E., Day, P.J., Hearn, K., et al. (2015). Fragment-based discovery of potent and selective DDR1/2 inhibitors. ACS Med. Chem. Lett. 6, 798–803.

Nagar, B., Bornmann, W.G., Pellicena, P., Schindler, T., Veach, D.R., Miller, W.T., Clarkson, B., and Kuriyan, J. (2002). Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). Cancer Res. 62, 4236–4243.

Noé, F., Wu, H., Prinz, J.H., and Plattner, N. (2013). Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. J. Chem. Phys. 139, 184114.

Onufriev, A., Bashford, D., and Case, D.A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 55, 383–394.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326.

Paricharak, S., Klenka, T., Augustin, M., Patel, U.A., and Bender, A. (2013). Are phylogenetic trees suitable for chemogenomics analyses of bioactivity data sets: the importance of shared active compounds and choosing a suitable data embedding method, as exemplified on kinases. J. Cheminform. 5, 49.

Parton, D.L., Grinaway, P.B., Hanson, S.M., Beauchamp, K.A., and Chodera, J.D. (2016). Ensembler: enabling high-throughput molecular simulations at the superfamily scale. PLoS Comput. Biol. 12, e1004728.

Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noé, F. (2013). Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. *139*, 015102.

Prinz, J.H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J.D., Schutte, C., and Noe, F. (2011). Markov models of molecular kinetics: generation and validation. J. Chem. Phys. *134*, 174105.

Radak, B.K., Chipot, C., Suh, D., Jo, S., Jiang, W., Phillips, J.C., Schulten, K., and Roux, B. (2017). Constant-pH molecular dynamics simulations for large biomolecular systems. J. Chem. Theory Comput. *13*, 5933–5944.

Richters, A., Nguyen, H.D., Phan, T., Simard, J.R., Grütter, C., Engel, J., and Rauh, D. (2014). Identification of type II and III DDR2 inhibitors. J. Med. Chem. *57*, 4252–4262.

Rix, U., Remsing Rix, L.L., Terker, A.S., Fernbach, N.V., Hantschel, O., Planyavsky, M., Breitwieser, F.P., Herrmann, H., Colinge, J., Bennett, K.L., et al. (2010). A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells. Leukemia *24*, 44–50.

Salah, E., Ugochukwu, E., Barr, A.J., von Delft, F., Knapp, S., and Elkins, J.M. (2011). Crystal structures of ABL-related gene (ABL2) in complex with imatinib, tozasertib (VX-680), and a type I inhibitor of the triazole carbothioamide class. J. Med. Chem. *54*, 2359–2367.

Scherer, M.K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., Wehmeyer, C., Prinz, J.-H., and Noé, F. (2015). PyEMMA 2: a software package for estimation, validation, and analysis of markov models. J. Chem. Theory Comput. *11*, 5525–5542.

Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science *289*, 1938–1942.

Seeliger, M.A., Nagar, B., Frank, F., Cao, X., Henderson, M.N., and Kuriyan, J. (2007). c-Src binds to the cancer drug Imatinib with an inactive Abl/c-kit conformation and a distributed thermodynamic penalty. Structure *15*, 299–311.

Shah, N.P., Tran, C., Lee, F.Y., Chen, P., Norris, D., and Sawyers, C.L. (2004). Overriding Imatinib resistance with a novel ABL kinase inhibitor. Science *305*, 399–401.

Shan, Y., Seeliger, M.A., Eastwood, M.P., Frank, F., Xu, H., Jensen, M.Ø., Dror, R.O., Kuriyan, J., and Shaw, D.E. (2009). A conserved protonation-dependent switch controls drug binding in the Abl kinase. Proc. Natl. Acad. Sci. U S A *106*, 139–144.

Shirts, M., and Pande, V.S. (2000). COMPUTING: screen savers of the world unite! Science *290*, 1903–1904.

Structural Genomics Consortium; China Structural Genomics Consortium; Northeast Structural Genomics Consortium, Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U.,

Arrowsmith, C., Hui, R., et al. (2008). Protein production and purification. Nat. Methods *5*, 135–146.

Tokarski, J.S., Newitt, J.A., Chang, C.Y., Cheng, J.D., Wittekind, M., Kiefer, S.E., Kish, K., Lee, F.Y., Borzillerri, R., Lombardo, L.J., et al. (2006). The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. Cancer Res. *66*, 5790–5797.

Vacic, V., Iakoucheva, L.M., and Radivojac, P. (2006). Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics *22*, 1536–1537.

Walter, M., Lucet, I.S., Patel, O., Broughton, S.E., Bamert, R., Williams, N.K., Fantino, E., Wilks, A.F., and Rossjohn, J. (2007). The 2.7 A crystal structure of the autoinhibited human c-Fms kinase domain. J. Mol. Biol. *367*, 839–847.

Wang, Z., Bian, H., Bartual, S.G., Du, W., Luo, J., Zhao, H., Zhang, S., Mo, C., Zhou, Y., Xu, Y., et al. (2016). Structure-based design of tetrahydroisoquinoline-7-carboxamides as selective discoidin domain receptor 1 (DDR1) inhibitors. J. Med. Chem. *59*, 5911–5916.

Wei, L., Hubbard, S.R., Hendrickson, W.A., and Ellis, L. (1995). Expression, Characterization, and crystallization of the catalytic core of the human insulin receptor protein-tyrosine kinase domain. J. Biol. Chem. *270*, 8122–8130.

Wodicka, L.M., Ciceri, P., Davis, M.I., Hunt, J.P., Floyd, M., Salerno, S., Hua, X.H., Ford, J.M., Armstrong, R.C., Zarrinkar, P.P., et al. (2010). Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. Chem. Biol. *17*, 1241–1249.

Wu, P., Nielsen, T.E., and Clausen, M.H. (2015). FDA-approved small-molecule kinase inhibitors. Trends Pharmacol. Sci. *36*, 422–439.

Wu, P., Nielsen, T.E., and Clausen, M.H. (2016). Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov. Today *21*, 5–10.

Young, M.A., Shah, N.P., Chao, L.H., Seeliger, M., Milanov, Z.V., Biggs, W.H., Treiber, D.K., Patel, H.K., Zarrinkar, P.P., Lockhart, D.J., et al. (2006). Structure of the kinase domain of an imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680. Cancer Res. *66*, 1007–1014.

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med. *23*, 703–713.

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

Zhao, Z., Wu, H., Wang, L., Liu, Y., Knapp, S., Liu, Q., and Gray, N.S. (2014). Exploration of type II binding mode: a privileged approach for kinase inhibitor focused drug discovery? ACS Chem. Biol. *9*, 1230–1241.

**CellPress**

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Chemicals, Peptides, and Recombinant Proteins | | |
| VX-680 (MK-0457, Tozasertib) | Selleck Chemicals | Cat # S1048 |
| Dasatinib | Cayman Chemical | Cat # 11498 |
| Imatinib | Novartis | |
| Axltide | Sigma Aldrich | Cat # 12-516 |
| [$\gamma$-32P] ATP | Perkin Elmer | Cat # NEG002A |
| Trichloroacetic acid | Sigma Aldrich | Cat # T4885-500G |
| TEV Protease | Lab generated | N/A |
| Tris-HCl | Sigma Aldrich | Cat # T3253-5KG |
| PMSF | EMD Millipore | Cat # 52332-25GM |
| NP-40 | Fluka | Cat # 74385 |
| $\beta$-mercaptoehanol | Sigma Aldrich | Cat # M6250 |
| Sodium Chloride | Fisher Scientific | Cat # BP358-10 |
| NiNTA resin | Qiagen | Cat # 30230 |
| Imidazole | Alfa Aesar | Cat # A10221-2500g |
| Glycerol | Fisher Scientific | Cat # G37-20 |
| PEG 3350 | Sigma Aldrich | Cat # P3015-500G |
| Bis-Tris | Sigma Aldrich | Cat # B9754-100G |
| Ethylene glycol | Sigma Aldrich | Cat # 102466 |
| Ammonium Iodide | Sigma Aldrich | Cat # 03101-100G |
| Kinase tracer 178 | Invitrogen | Cat # PV5593 |
| LanthaScreen Eu-anti-His Antibody | Invitrogen | Cat # PV5596 |
| EGTA | Acros Organics | Cat # 428570100 |
| Brij-35 | Acros Organics | Cat # 329581000 |
| MgCl2 | Sigma Aldrich | Cat # M9272 |
| HEPES | Spectrum | Cat # H1084 |
| Experimental Models: Cell Lines | | |
| Sf9 | ATCC | Cat # CRL-1711 CVCL_0549 |
| Software and Algorithms | | |
| HKL-2000 | HKL Research, Inc | http://www.hkl-xray.com/hkl-2000 SCR_015547 |
| CCP4 – Phaser | University of Cambridge; Cambridge; United Kingdom | https://www.phenix-online.org/documentation/reference/phaser.html SCR_014219 |
| Phenix Crystallography Suite | Lawrence Berkeley National Laboratory, University of California at Berkeley; California; USA | https://www.phenix-online.org/SCR_014224 |
| Coot | MRC Laboratory of Molecular Biology | http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/SCR_014222 |
| Kaleidagraph | Synergy software | |
| Other | | |
| Phosphocellulose paper | Milipore | IPVH 00010 |
| Fluorimeter | Horriba Scientific | Fluoromax 4 |
| Microplatereader | Molecular Devices | SpectraMax M5 |

**Continued**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
| --- | --- | --- |
| Deposited Data | | |
| Structure of DDR1·VX680 | This study | PDB 6BRJ |
| Structure of DDR1·Dasatinib | This study | PDB 6BSD |
| Molecular simulations for DDR1 | This study | Open Science Framework at https://osf.io/4r8x2/. |
| Dendrogram presented in Figure 1A | This study | GitHub at https://github.com/choderalab/DDR1_and_kinase_promiscuity_materials |
| Kinase inhibition data | (Drewry et al., 2017) | |
| Structure of Abl·VX680 | https://www.rcsb.org/ | PDB: 2F4J |
| Structure of Abl·Dasatinib | https://www.rcsb.org/ | PDB: 2GQG |
| Structure of DDR1·Imatinib | https://www.rcsb.org/ | PDB: 4BKJ |
| Structure of CSF1R | https://www.rcsb.org/ | PDB: 3KRL |
| Structure of CSF1R | https://www.rcsb.org/ | PDB: 2IOV |
| Structure of c-Kit | https://www.rcsb.org/ | PDB: IT45 |
| Structure of c-Kit | https://www.rcsb.org/ | PDB: IT46 |
| Structure of DDR1 | https://www.rcsb.org/ | PDB: 5FDP |
| Structure of DDR1 | https://www.rcsb.org/ | PDB: 5BVO |
| Structure of PDGFRA | https://www.rcsb.org/ | PDB: 5GRN |
| Structure of DDR1·ponatinib | https://www.rcsb.org/ | PDB: 3ZOS |
| Structure of Abl·ponatinib | https://www.rcsb.org/ | PDB: 3OXZ |
| Structure of Abl·imatinib | https://www.rcsb.org/ | PDB: 2HYY |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, John Chodera (john.chodera@choderalab.org ) and Markus A. Seeliger (markus.seeliger@stonybrook.edu).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sf9 cells were grown in Sf-900-II medium supplemented with 5% fetal bovine serum and 1x antibiotic/antimitotic solution. Cells were grown according to the Bac-to-Bac protocol from Invitrogen.

## METHOD DETAILS

### Hierarchical Clustering of a Large Kinase Inhibitor Set

Pairwise distances were computed among the 645 inhibitors and 406 kinases to generate the Published Kinase Inhibitor Set 2 (PKIS2) in (Drewry et al., 2017). Inhibition refers here to competitive binding of the inhibitor of interest to the kinase and thereby preventing binding of the kinase to immobilized generic, ATP-competitive kinase inhibitors using the commercial DiscoverX KINOMEscan service.

Distance calculations were computed as Euclidean distances and performed using the dist function in R (version 3.4.0). We performed hierarchical clustering analysis of this dataset using the 'complete linkage' method (hclust), in which the similarity of two clusters of kinases is calculated as the similarity of the affinity fingerprint of their most dissimilar members. The resulting dendrogram, in which a cluster of eight kinases is most distant from all other clusters on the tree, i.e., forms the basal cluster when rooted to the midpoint of the tree (Figures 1 and S1). To test the robustness of this result, a consensus tree of 1000 bootstrapped datasets was generated. In this tree, DDR1 and other promiscuous kinases separated into two clades, and not one, distant from other kinases (with YSK4, MEK5, KIT, and PDGFRB in one and RAF1, BRAF, DDR1, DDR2, p38α, and p38β in the other, see Figure S2A for details). Overall, this bootstrap result supports the general trend in which we found our promiscuous kinases distinct from the majority of kinases. We also found that results using the UPGMA or 'average' method, while varying in detail, also followed this trend (Figure S1C).

In addition to construction of this affinity-based tree, we quantified promiscuity by counting the number of inhibitors that bind a single kinase with 90% or greater inhibition. The more inhibitors inhibit a given kinase, the more promiscuous is the kinase. A cutoff of either 90% inhibition or 75% inhibition results in similar patterns (Figure S1B), though throughout this paper we have used the 90% cutoff. Mapping these promiscuity values onto the kinome phylogenetic tree was performed using the kinhub.org KinMap tool (Eid et al., 2017). To define ligand chemotypes that are more predominant in the promiscuous kinases, ligands were sorted by those

with p values < 0.05 when comparing the mean and SD inhibition within promiscuous kinases to the mean and SD of all other kinases using the Bonferroni-corrected Student's t-test.

## Cloning and Constructs

The construct of human DDR1 kinase domain (residues 526-876, DDR1a numbering, UniProtKB accession ID: Q08345-2) was amplified from a pDNR-Dual mammalian expression vector containing full length human DDR1 isoform A (residues 1-876; purchased from the ASU Gene Repository) and subcloned into the pFastBac HTb vector (Invitrogen) using BamHI and XhoI restriction sites located after the His6-tag and TEV protease cleavage site. Individual mutations were introduced into pFastBac HTb DDR1 KD (D671N, Y755A, and Y759A) by site directed mutagenesis and verified by DNA sequencing. Cloning and baculovirus generation for the kinase domain of DDR1a utilized the Bac-to-Bac baculovirus expression system (Invitrogen).

## Protein Expression and Purification

His6-tagged DDR1 kinase domain (KD) was expressed in *Spodoptera frugiperda* (Sf9) insect cells (RRID: CVCL_0549) and purified following the Bac-to-Bac system protocol from Invitrogen. Sf9 cells were grown in Sf-900-II medium supplemented with 5% fetal bovine serum and 1x antibiotic/antimitotic solution. A 1 L culture of Sf9 cells at $0.8 \times 10^6$ cells/mL was infected with 30 mL of P3 virus and incubated for 72 hours. Infected cells were centrifuged for 5 minutes at 3000 g at $4°C$ using the SLC-6000 rotor. Cells were resuspended in lysis buffer (50 mM Tris pH 8.5, 5 mM $\beta$-mercaptoehanol, 100 mM NaCl, 1 mM PMSF, and 1% NP-40) in a ratio of 5 mL lysis buffer per 1 g of cells. Lysates were centrifuged at 20,000 g for 60 minutes at $4°C$. The resulting supernatant was loaded onto 1 mL of NiNTA resin (Qiagen) per 2 L of insect cells on a gravity column in the cold room ($4°C$). The resin was washed with 10 column volumes of low salt wash buffer (20 mM Tris pH 8.5, 20 mM imidazole, 500 mM NaCl, 5 mM $\beta$-mercaptoehanol, and 10% glycerol) and 10 column volumes of high salt wash buffer (20 mM Tris pH 8.5, 1 M NaCl, 5 mM $\beta$-mercaptoehanol and 10% glycerol) followed by 2 column volumes of low salt wash buffer. The protein was eluted and fractionated with 20 mM Tris pH 8.5, 125 mM NaCl, 200 mM Imidazole, 10% glycerol and 5 mM $\beta$-mercaptoehanol. The His6-tag of the protein was then cleaved by TEV protease digest overnight at $4°C$. The cleaved protein was further purified by size exclusion chromatography (GE 16/60 Superdex 200) in 20 mM Tris pH 8.5, 125 mM NaCl, 20 mM imidazole, 10% glycerol and 5 mM $\beta$-mercaptoehanol. The typical yields for DDR1 were approximately 2 mg of protein per liter of Sf9 cells. Protein identity, purity, and verification of no post-translational state was confirmed by LC-MS/MS (Figure S3). No post-translational modifications were detected. For storage, protein was frozen in liquid $N_2$ and stored at $-80°C$.

## DDR1 Crystallization

DDR1 KD was complexed with dasatinib at a concentration of 10 mg/mL protein and 423 $\mu$M inhibitor, in buffer containing 20 mM Tris pH 8.5, 125 mM NaCl, 5 mM $\beta$-mercaptoehanol and 10% glycerol. DDR1·dasatinib crystals were grown using the hanging drop vapor diffusion method and micro seeding in a mother liquor of 18% PEG 3350 and 0.1 M Bis-Tris pH 5.5. Micro seeds were generated from DDR1·dasatinib crystals grown previously in a mother liquor of 22% PEG 3350, and 0.1 M Bis-Tris pH 5.5. Crystals were cryoprotected in mother liquor plus 20% ethylene glycol and stored in liquid nitrogen.

DDR1·VX-680 complex was formed at 10 mg/mL protein and 423 $\mu$M VX-680 in an identical buffer to the DDR1·dasatinib complex. The hanging drop vapor diffusion method and DDR1·dasatinib micro seeds were used to grow crystals in a mother liquor of 18% PEG 3350, 0.05 M NH$_4$I and 0.1 M Bis-Tris pH 5.5. DDR1a·VX-680 crystals were cryoprotected in mother liquor plus 20% glycerol, and stored in liquid nitrogen.

## X-ray Diffraction Data and Processing

X-ray diffraction data were collected at the National Synchrotron Light Source at Brookhaven National Laboratories beamline x29. Data for all protein-drug complexes were collected at 100 K and 1.075 Å wavelength.

DDR1·VX-680 and DDR1·dasatinib crystals diffracted to a resolution of 2.2 Å and 2.6 Å respectively (Table 1). Data for both complexes were processed in space group P2$_1$2$_1$2$_1$ using HKL2000 (Otwinowski and Minor, 1997). The structures were solved by molecular replacement using the kinase domain of DDR1 bound to imatinib (PDB: 4BKJ; residues 599-913) with the $\alpha$C-helix (residues 660-680), activation loop (residues 775-814), and ligand removed as a search model in Phaser (Mccoy et al., 2005). The models were built in Coot (Emsley and Cowtan, 2004) and refined in Phenix (Adams et al., 2002).

## DDR1 Kinase Activity Assays

*In vitro* kinase activity for DDR1 KD was measured by substrate peptide phosphorylation using [$\gamma$-$^{32}$P] ATP in a phosphocellulose paper binding assay (Casnellie, 1991). Reactions included 20 mM Tris pH 7.4, 10 mM MgCl$_2$, 400 $\mu$M ATP, 200 $\mu$M Axltide (KKSRGDYMTMQIG), 0.5 $\mu$M DDR1 KD and 50-100 cpm/pmol [$\gamma$-$^{32}$P] ATP. Reaction mixtures were incubated for 60 minutes at $30°C$ and stopped by the addition of trichloroacetic acid (TCA) and centrifuged to separate the precipitated kinase from the soluble substrate peptide. Supernatant was blotted onto phosphocellulose paper and washed three times with phosphoric acid. The phosphocellulose paper was dried and radioactivity was quantified by scintillation counting.

### DDR1 Drug Binding Assay

Drug binding assays were performed using the Invitrogen LanthaScreen Eu kinase binding assay according to manufacturer's instructions. First we determined the dissociation constant of DDR1 wt and mutant proteins to the Alexa Fluor-647-labeled tracer 178 molecule that binds to the ATP binding site of DDR1. 5 uL of DDR1 at 15 nM concentration was mixed with 5 uL of Eu-labeled anti-His6 antibody at 6 nM concentration and 5 uL of tracer at concentration ranging from 1500 nM to 0 nM in assay buffer (50 mM Hepes pH7.5, 10 mM MgCl2, 1 mM EGTA, 0.01% Brij-35) and 3% DMSO. A series of control experiments contained 30 μM dasatinib in the assay buffer. Samples were mixed in a white 384-well plate and incubated in the dark at room temperature for 60 min. Time-resolved fluorescence energy transfer was measured in a Molecular devices SpectraMax M5 Microplate reader using the following settings: Excitation 332 nm, Emission 620/665 nm, Emission cutoff 550 nm, TRF integration delay 50 μs, Integration time 400 μs, 100 flashes per read, calibration on. The TR-FRET ratio at 665 and 620 nm was calculated for each concentration of tracer. The TR/FRET ratio for the dasatinib control was subtracted from the TR FRET ratio of the dasatinib-free samples at the same tracer concentration. The tracer binding isotherm was fit to a quadratic binding equation to yield the dissociation constant $K_D^{tracer}$ of the tracer to each of the kinase constructs tested.

Binding of imatinib to DDR1 proteins was measured by competition with tracer 178. 5 μL of DDR1 at 15 nM and 6 nM Eu-labeled anti-His body in assay buffer was mixed with 5 μL of tracer 178 at 30 nM in assay buffer and 5 μL of imatinib (30 μM – 0.004 μM) in assay buffer with 3% DMSO. Samples were mixed in a white 384-well microtiter plate and incubated in the dark for 60 min at room temperature. TR FRET ratios were determined as described above. The TR FRET signal at different imatinib concentrations was fit to the following equation using Kaleidagraph:

$$TR\text{-}FRET(imatinib) = ((m1\text{-}m2)/(1+(m3/[conc.\ Imatinib])))+m2$$

where m1 is the minimum TR FRET value, m2 is the maximum TR FRET value and m3 is the concentration at which the amplitude of the TRFRET signal decreases by 50% ($IC_{50}^{imatinib}$).

Using the Cheng-Prusoff relationship, the $IC_{50}^{imatinib}$ values are converted into inhibitory constants $K_i^{imatinib}$ (Cheng and Prusoff, 1973):

$$K_i^{imatinib} = IC_{50}^{imatinib}/(1+([tracer]/K_D^{tracer}))$$

where [tracer] is the concentration of tracer 178 in the binding competition reaction and $K_D^{tracer}$ is the dissociation constant of tracer for the specific kinase or mutant kinase tested.

### Molecular Dynamics Simulation of DDR1 Wild-Type

Simulations of WT DDR1 kinase domain (residues 567-875 according to human DDR1a numbering) were initiated from the structure of DDR1 in complex with VX-680 (this study, PDB ID 6BRJ), as well as complexes with imatinib and ponatinib (PDB IDs 4BKJ and 3ZOS, respectively) (Canning et al., 2014). In all cases, the ligand was removed and only the kinase domain was simulated. Separate sets of simulations were performed with the Asp747 in the DFG motif either protonated or deprotonated, producing a total of six initial structures. Previous simulations of Abl kinase showed that protonation of the Asp in the DFG motif enhances the likelihood of seeing the interconversion between DFG-Asp-in and DFG-Asp-out, the so-called DFG flip (Shan et al., 2009).

Simulations were run using OpenMM 6.3 derived core (core21 v0.0.18) on Folding@home after a short equilibration period (consisting of 100 ps implicit solvent simulation using the OBC GBSA implicit solvent model (Onufriev et al., 2004) followed by 100 ps explicit solvent simulation under isothermal-isobaric (NPT) conditions) (Eastman et al., 2017; Shirts and Pande, 2000; Eastman et al., 2013). Simulations were run in a box of 16,046 explicit TIP3P water molecules (Jorgensen et al., 1983) using the Amber99SB-ILDN force field (Lindorff-Larsen et al., 2010), at 300 K, using the Leapfrog Langevin integrator with a timestep of 2 fs, and a collision rate of 1 / ps. A molecular-scaling Monte Carlo barostat was used with default update interval of 25 steps and pressure of 1 atm. Particle Mesh Ewald (PME) with a cutoff of 9 Å and default parameters was used for long-range electrostatics. For each of the six WT starting conditions, 50 simulations of ∼3 μs each were run totaling ∼150 μs, resulting in ∼0.9 ms in aggregate simulation time for all starting structures.

### Molecular Dynamics of DDR1 Mutants and Wild-Type Started along the DFG Flip

To enhance sampling of the DFG flip, new simulations were adaptively seeded along the DFG flip observed from the unbiased simulations above. From this original set of WT DDR1 simulations, four of the Asp747-protonated simulations interconverted from the DFG-Asp-out conformation to the DFG-Asp-in conformation. From each of these four simulations, nine snapshots were chosen at even intervals along the coordinate originally used to identify trajectories: the distance between the CZ atom of the Phe748 of the DFG motif and CA atom of the Gly716 in the αE-helix, seen to be either comparable or preferable to the Asp747 phi dihedral as an order parameter. This resulted in 36 structures from which new simulations were initiated for eight chemically distinct systems: WT, D671N, Y755A, and Y759A, each with Asp747 either protonated or deprotonated. Simulations from the original three starting structures (DDR1·VX-680, 4BKJ, and 3ZOS) were also started for all eight of these systems. Additionally all eight systems were started from the DDR1·dasatinib structure, with the unresolved loop in the N-lobe of this structure modeled in from the DDR1·VX-680 structure. To ensure uniform preparation for all 320 (8 systems x 40 structures) of these new simulations, Ensembler was used to automate the process (Parton et al., 2016). Ensembler uses Modeller to model in mutants where needed (Fiser and Sali, 2003). The overall simulation protocol was the same as described above for the WT simulations and 20 simulations of 1 μs each were

generated by Folding@home for each of 320 starting conformations, totaling 6.5 ms in aggregate for simulations restarted along the DFG flip, more than previously published for any kinase.

### Molecular Simulation Analysis Using Markov Models

Free energy landscapes of DDR1 simulations were built using the PyEMMA Python library (Scherer et al., 2015). Trajectories were first featurized according to 25 features of interest that were either previously used to investigate the DFG-Asp-in vs. DFG-Asp-out conformation (Mobitz, 2015; Meng et al., 2015; Hu et al., 2015) or were of particular interest to this system, e.g. the Asp671-Arg752 distance (Table S1). The time-lagged independent component analysis (TICA) projection was then calculated from these coordinates to find the most kinetically relevant linear combination (Perez-Hernandez et al., 2013), and cumulative sum of the eigenvalues began to flatten despite the small number coordinates used (Figure S4A). Commute mapping and a lag time of 25 ns were used. K-means clustering of the TICA coordinates was used to generate a 300 microstate model and corresponding transition matrix, which was then used to determine the parameters of a two-state Bayesian hidden Markov Model (HMM) (Noé et al., 2013). The two-state HMM separated the DFG-Asp-in and DFG-Asp-out states for the WT simulation (Figure S4D), and the largest time-scale separation consistently supported a two-state model (Figure S4C), though these two states were different between WT and mutants (Figure S4E). Markov state models (MSM) (Prinz et al., 2011) were then built for all the mutant variants within the TICA space of the WT-protonated simulations. A lag time of 50 ns was used for constructing the MSM. Note that for the WT simulations, all simulations, including initial simulations and re-seeded simulations were analysed, though the nature of the MSM analysis ought to correct for any error introduced by this additional data. Chapman-Kolmogorov tests of the WT HMM and the MSMs of all mutants indicated that these models were valid representations of our real data (Figure S4B). To compare the relative free energies of the states, a line was found using the scikit learn C-support vector classification algorithm with C=1.0 and gamma=0.7 that separates the two HMM states in our WT simulations (Figure S4D). The ratio between the populations of the two states was used to estimate the free energy difference between the states.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Data are presented as mean values ± standard error of the mean (s.e.).

### DATA AND SOFTWARE AVAILABILITY

The structures of DDR1·VX-680 and DDR1·dasatinib were deposited in the PDB under ID codes 6BRJ and 6BSD respectively.

Data generated in this study are available at https://github.com/choderalab/DDR1_and_kinase_promiscuity_materials and https://osf.io/4r8x2/ and upon request to the corresponding authors, John D. Chodera (john.chodera@choderalab.org) and Markus A. Seeliger (markus.seeliger@stonybrook.edu).
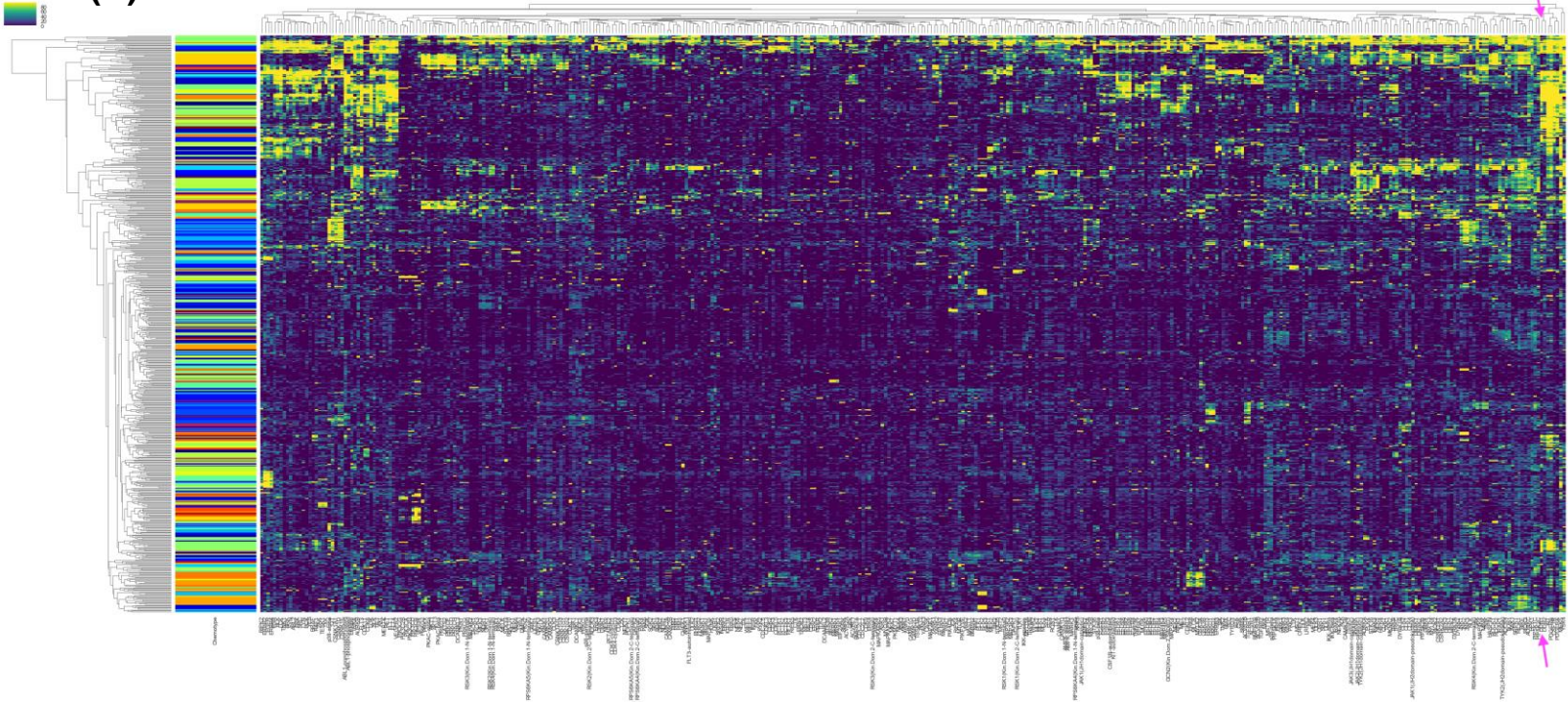
# Supplemental Information

# What Makes a Kinase Promiscuous for Inhibitors?

**Sonya M. Hanson, George Georghiou, Manish K. Thakur, W. Todd Miller, Joshua S. Rest, John D. Chodera, and Markus A. Seeliger**
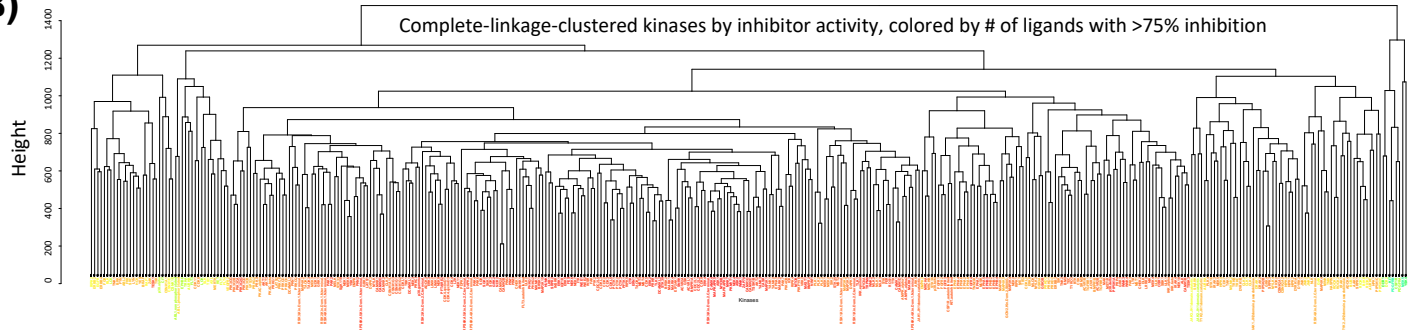
**(A)**



**(B)**



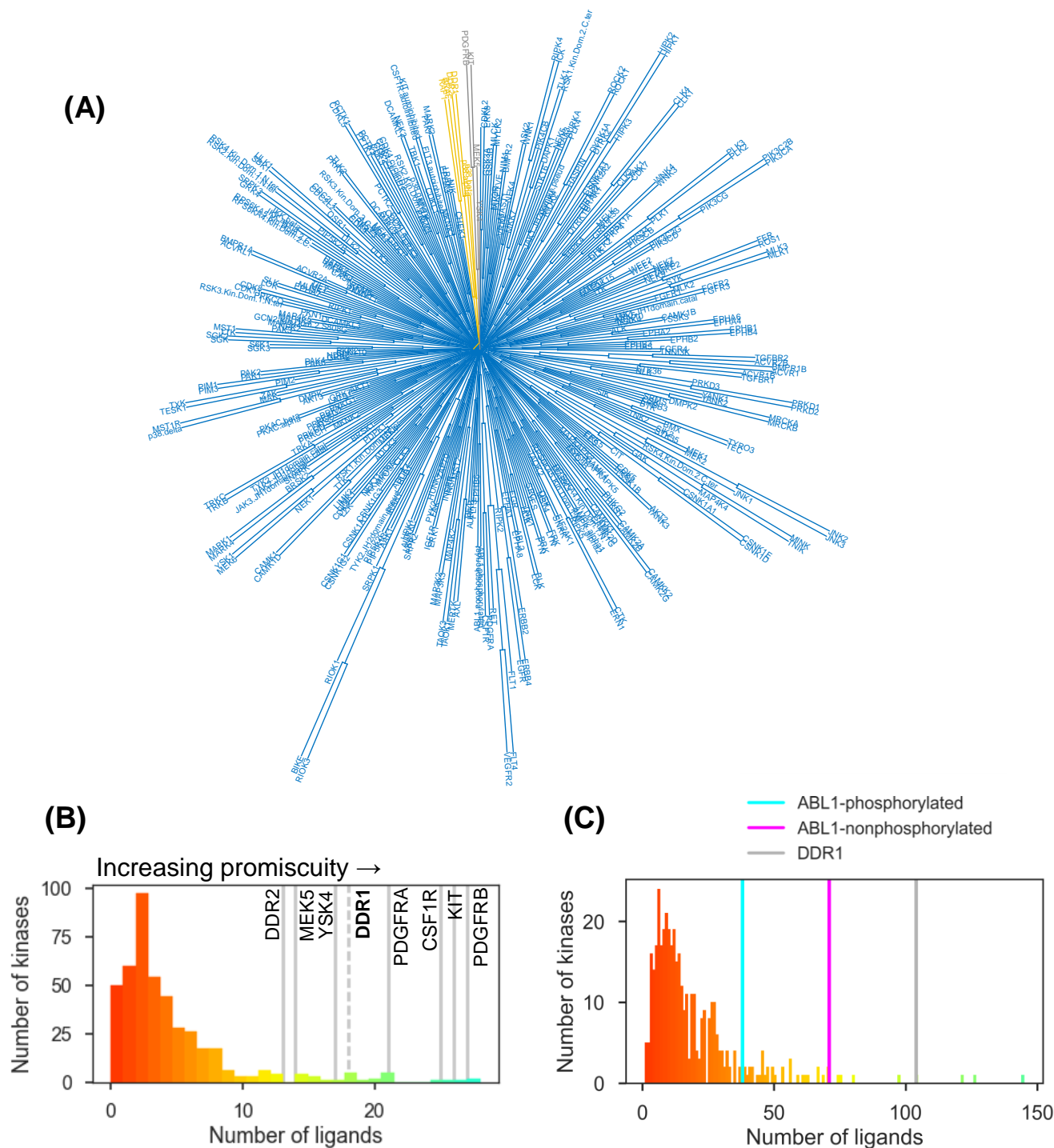Complete-linkage-clustered kinases by inhibitor activity, colored by # of ligands with >75% inhibition

**(C)**



UPGMA-clustered kinases by inhibitor activity, colored by # of ligands with >75% inhibition

**Figure S1: Heatmap and alternative methods for clustering results. Related to Figure 1.** A) The full heatmap resulting from complete linkage hierarchical clustering of inhibition data from Drewry et al. X-axis are kinases. Y-axis are ligands, colored blocks indicate chemotypes. 0% inhibition: dark blue, 100% inhibition: yellow. Pink arrows indicate DDR1. B) Same as A, complete-linkage clustering of kinases by inhibitor activity, here colored by the number of ligands that bind with >75% inhibition. C) UPGMA or 'average' clustering results, colored as in B. Searchable PDFs of all three figures are available via GitHub.

**(A)**

**(B)**

Increasing promiscuity →

*(Panel B: histogram — Number of kinases vs. Number of ligands, with vertical markers labeled DDR2, MEK5, YSK4, DDR1, PDGFRA, CSF1R, KIT, PDGFRB)*

**(C)**

- ABL1-phosphorylated
- ABL1-nonphosphorylated
- DDR1

*(Panel C: histogram — Number of kinases vs. Number of ligands, with vertical lines for ABL1-phosphorylated, ABL1-nonphosphorylated, and DDR1)*

**Figure S2: Bootstrap consensus tree and expanded promiscuity analysis. Related to Figure 1.** A) Consensus tree of 1000 bootstrapped datasets using the complete-linkage method. Grey cluster indicates the 'most distant' group in this consensus tree and consists of the kinases PDGFRB, KIT, MEK5, and YSK4. The yellow cluster indicates the 'second most distant' group in this consensus tree and consists of the kinases DDR1, DDR2, RAF1, BRAF, p38α, and p38β. B) Promiscuity in Davis, 2011 dataset. Analyzing the smaller but insightful dataset of Davis et al 2011 to estimate promiscuity trends results in similar trends to those described in Figure 1 for analysis of the Drewry et al 2017 dataset. Here a cutoff of 50 nM is used to count high affinity ligands. C) Abl and Abl-phosphorylation promiscuity. Counting ligands that bind kinases at 90% Inhibition or higher clearly shows ABL1-nonphosphorylated is more promiscuous than ABL1-phosphorylated.

**(A)**
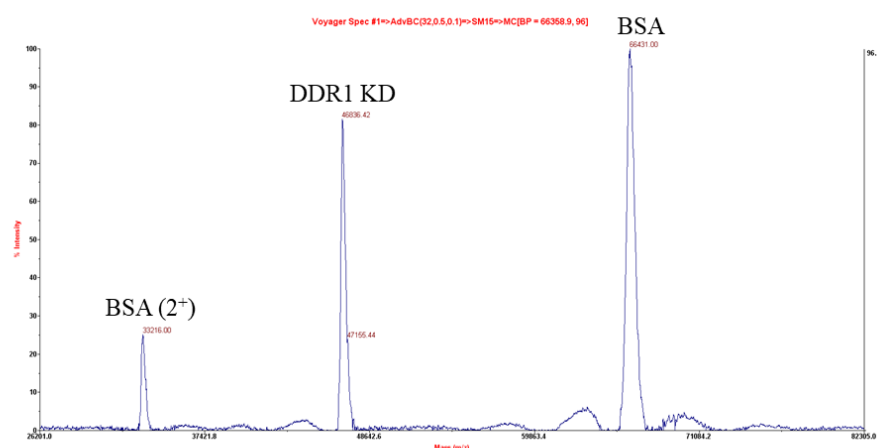


**(B)**
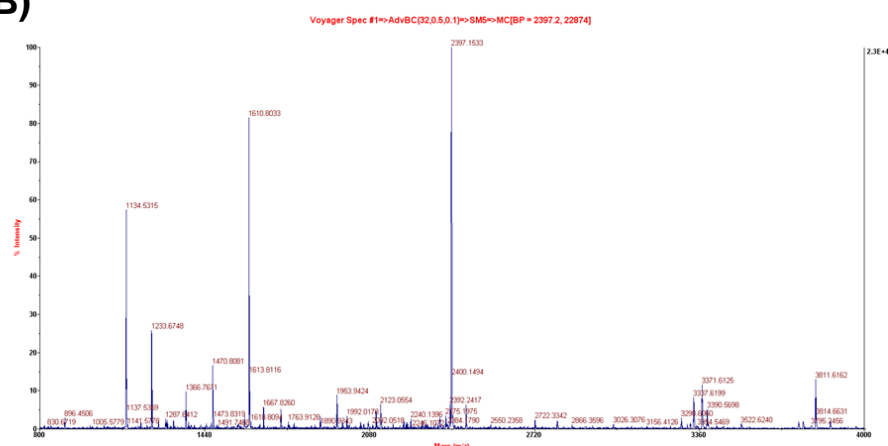


**(C)**

72 peptides, 395 spectra, 68.6% coverage

```
  1 MSYYHHHHHH DYDIPTTENL YFQGAMGQEP RPRGNPPHSA PCVPNGSAAY
 51 SGDYMEPEKP GAPLLPPPPQ NSVPHYAEAD IVTLQGVTGG NTYAVPALPP
101 GAVGDGPPRV DFPRSRLRFK EKLGEGQFGE VHLCEVDSPQ DLVSLDFPLN
151 VRKGHPLLVA VKILRPDATK NARNDFLKEV KIMSRLKDPN IIRLLGVCVQ
201 DDPLCMITDY MENGDLNQFL SAHQLEDKAA EGAPGDGQAA QGPTISYPML
251 LHVAAQIASG MRYLATLNFV HRDLATRNCL VGENFTIKIA DFGMSRNLYA
301 GDYYRVQGRA VLPIRWMAWE CILMGKFTTA SDVWAFGVTL WEVLMLCRAQ
351 PFGQLTDEQV IENAGEFFRD QGRQVYLSRP PACPQGLYEL MLRCWSRESE
401 QRPPFSQLHR FLAEDALNTV
```

**Figure S3: LC MS/MS of DDR1 kinase domain. Related to Figure 2.** A) MADLI-TOF of DDR1 kinase domain compared to bovine serum albumin (BSA) shows a theoretical mass of 46,855 daltons. B) Mass spectrometric analysis of tryptic digest of the purified kinase domain of DDR1. C) Sequence coverage of the tryptic DDR1 peptides detected via MALDI-TOF. N-terminal fragments were too large to detect.
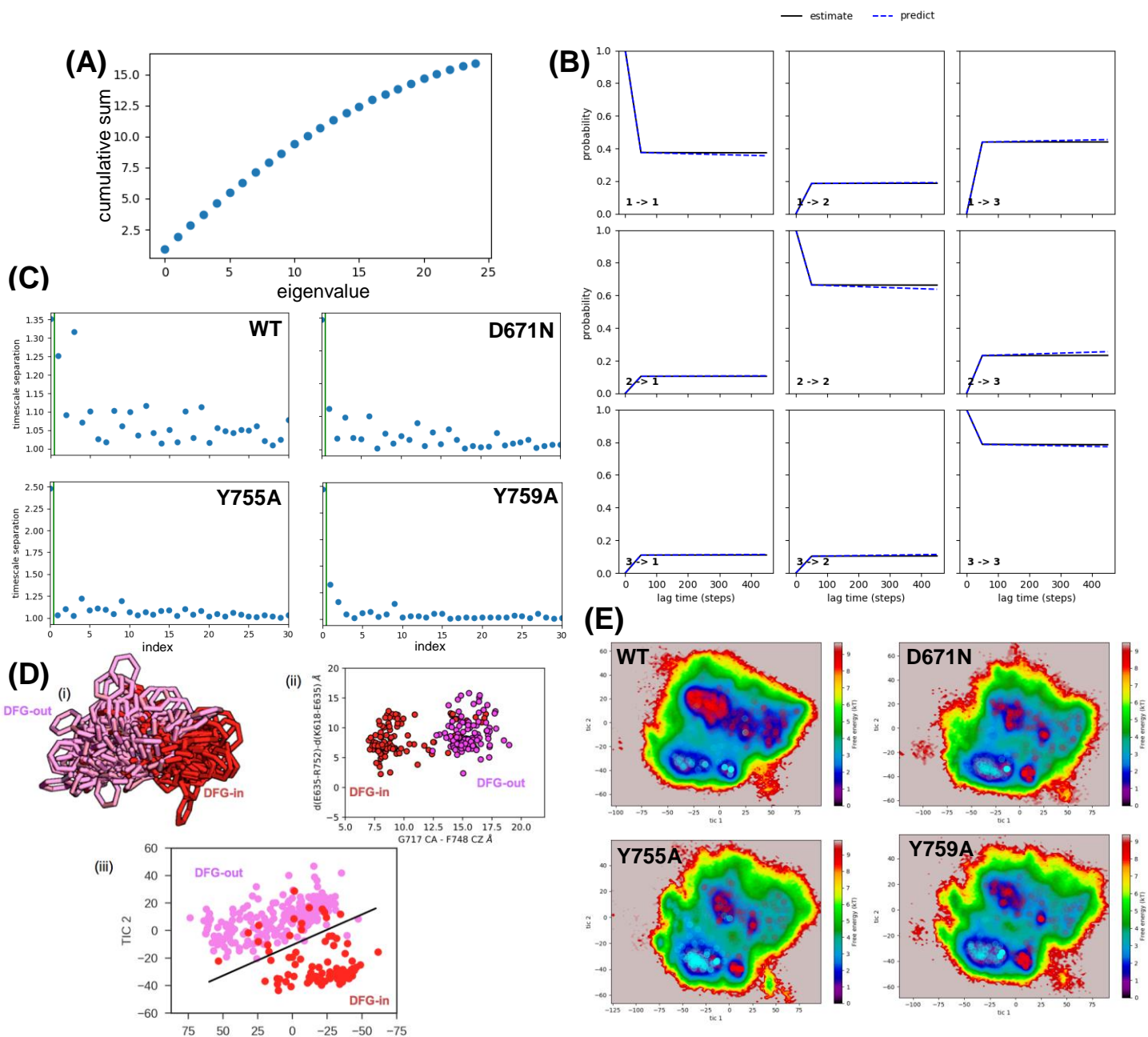
**Figure S4: Further validation of MSM and HMM models. Related to Figure 3**. A) The cumulative sum of the TICA eigenvalues for the WT analysis shows a leveling off. B) Chapman-Kolmogorov (CK) test for WT MSM. CK test results for mutant MSMs on the WT TICA space look similar. C) Timescale separations for MSMs for all mutants in the WT TICA space. Green line indicates the largest timescale separation. D) i) Overlayed phenalanine sidechains of the DFG motif of 100 samples of each state chosen by the probability of being found in each WT HMM macrostate. This shows the heterogeneity of both DFG-out (magenta) and DFG-in (red) states, while still clearly representing distinct states. ii) The same 100 states here plotted onto a y-axis representative of the C-helix rotation (the difference of distances E635-R752 and K618-E635) and an x-axis representative of the DFG-flip (the distance between the CA of G716 in the αE helix and the CZ of F748 in the DFG motif). iii) The line calculated using the scikit learn C-support vector classification algorithm with C=1.0 and gamma=0.7 by an SVM to separate the two DFG-in and DFG-out WT macrostates (this is the dashed lines in Figure 3C). E) Two-state HMM analysis of mutant simulations defines states differently than for the WT. Here K-means cluster centers are colored as DFG-out (red) and DFG-in (cyan).

**(A)**
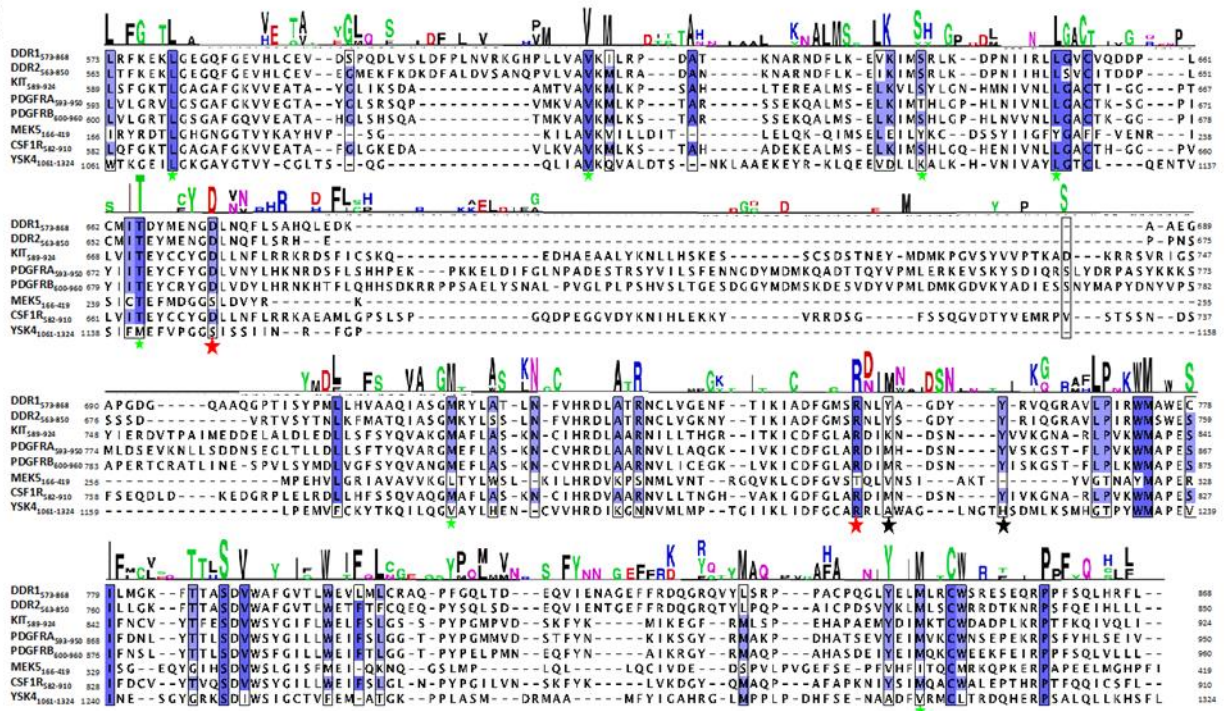
**(B)**
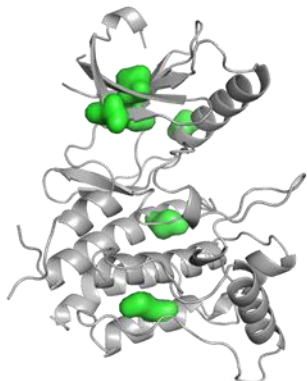
| Abl Imatinib Resistance Mutant | Enriched residues in promiscuous group | Equivalent DDR1 residue |
|---|---|---|
| L267R/V | L | L579 |
| V289A | V | V617 |
| K310E/R | S | S640 |
| L320F | L | L651 |
| T334I/S/G | T | T664 |
| M370T/I | M | M717 |
| M491I | M | M847 |

**(C)**

**(D)**

**(E)**

**Figure S5: Sequence alignments. Related to Figure 4. (A)** WebLogo result from Two Sequence Logo (TSL) analysis (using p-value cutoff 0.01) is presented above the sequence alignment. All 41 residues identified by the TSL analysis as enriched by 50% or more in the promiscuous kinase are highlighted and colored by sequence identity. Residues highlighted by red stars are those involved in the salt bridge proposed to stabilize the DFG-Asp-out conformation. Residues highlighted by black stars are those mutated in mutagenesis experiments of DDR1 (these were not enriched in the TSL analysis, but are highlighted here for clarity). Residues highlighted by green stares are those found enriched in promiscuous kinases that are also found to correlate with imatinib resistance mutants in Abl, illustrated further in Fig S6.

**(B)** Abl resistance mutants correlate to residues enriched in promiscuous kinases. Imatinib resistance mutants found in Abl (Azam et al 2003) that correspond to enriched residues in the eight promiscuous kinases as defined in Two Sample Logo analysis. Seven residues are represented as green surfaces within the DDR1•VX680 structure.

**(C, D)** Sequence differences between DDR1 and DDR2 kinase domains may explain the different levels of promiscuity exhibited by both kinases. Active site of DDR1 bound to VX-680 (C) and dasatinib (D), respectively. All residues within 5Å of VX-680 and dasatinib are highlighted as cyan spheres and in highlighted and colored by sequence identity in the sequence alignment. No significant differences were found in the active site of the two kinases to explain the difference in promiscuity between DDR1 and DDR2.

**(E)** Sequence alignment of DDR1 compared to DDR2 with residues highlighted that were found to be enriched in promiscuous kinases in the TSL analysis (DDR1a numbering in red). There are several differences that span the entire kinase domain that may explain the difference in promiscuity between DDR1 and DDR2.

**Supplemental Tables**

T1. List of 25 features used to build find TICA coordinates for trajectories. Related to Figure 3.

| Feature Type | Residues / Atoms involved | Reference |
|---|---|---|
| Pseudo-dihedral | A746 CB – A746 CA – D747 CA – D747 CG | Meng et al, 2015 |
| Pseudo-dihedral | A746 CB – A746 CA – F748 CA – F748 CG | Meng et al, 2015 |
| Pseudo-dihedral | M750 CB – M750 CA – F748 CA – F748 CG | Meng et al, 2015 |
| Pseudo-dihedral | A746 CA – D747 CA – F748 CA – G749 CA | Mobtiz, 2015 |
| Pseudo-dihedral | D747 CA – F748 CA – G749 CA – M750 CA | Mobitz, 2015 |
| Phi, psi, and chi dihedrals | D747, F748, G749 | N/A |
| Distance | G716 CA – F748 CZ | N/A |
| Closest heavy atom distance | K618 – E635 | Shan et al, 2009 |
| Closest heavy atom distance | E635 – R752 | Shan et al, 2009 |
| Closest heavy atom distance | D671 – R752 | N/A |
| Closest heavy atom distance | M639 – L650 | Hu et al, 2015 |
| Closest heavy atom distance | L650 – H727 | Hu et al, 2015 |
| Closest heavy atom distance | H727 – F748 | Hu et al, 2015 |

T2. List of promiscuous kinase PDBs along with their relevant ligand and conformation. Related to Figure 4.

| Kinase | PDB ID | Inhibitor/Ligand | Type I/II Inhibitor | DFG Conformation |
|--------|--------|------------------|---------------------|------------------|
| PDGFRA | 5GRN | WQ-C-159 | Type II | Out |
| PDGFRA | 5K5X | Apo structure | NA | Out |
| c-Kit | 4U0I | Ponatinib | Type II | Out |
| c-Kit | 1PKG | ADP | NA | In |
| c-Kit | 1T46 | Imatinib | Type II | Out |
| c-Kit | 3G0E | Sunitinib | Type II | Out |
| c-Kit | 1T45 | Apo structure | NA | Out |
| DDR1 | 4CKR | DDR1-IN-1 | Type II | Out |
| DDR1 | 5FDX | D2164 | Type II | Out |
| DDR1 | 5BVN | N-[5-({[[(3-fluorophenyl)carbamoyl]amino}methyl]- 2-methylphenyl]imidazo[1,2-a]pyridine-3-carboxamide | Type II | Out |
| DDR1 | 5FDP | D2099 | Type II | Out |
| DDR1 | 5BVK | 1-(2-chlorophenyl)-3-(pyridin-3-ylmethyl)urea | Type II | Out |
| DDR1 | 4BKJ | Imatinib | Type II | Out |
| DDR1 | 3ZOS | Ponatinib | Type II | Out |
| DDR1 | 5BVW | Dasatinib | Type I | Out |
| DDR1 | 5BVO | N-(5-{(1S)-1-[(5-fluoro-1,3-benzoxazol-2- yl)amino]ethyl}-2-methylphenyl)imidazo[1,2- a]pyridine-3-carboxamide | Type II | Out |
| CSF1R | 3KRL | 5-cyano-N-[4-(4-methylpiperazin-1-yl)-2-piperidin-1-ylphenyl]furan-2-carboxamide | Type II | Out |
| CSF1R | 4R7I | Imatinib | Type II | Out |
| CSF1R | 3LCO | 3-({4-methoxy-5-[(4-methoxybenzyl)oxy]pyridin- 2-yl]methoxy)-5-(1-methyl-1H-pyrazol-4-yl)pyrazin- 2-amine | Type II | Out |
| CSF1R | 3DPK | 8-cyclohexyl-N-methoxy-5-oxo-2-{[4-(2-pyrrolidin- 1-ylethyl)phenyl]amino}-5,8-dihydropyrido[2,3- d]pyrimidine-6-carboxamide | Type II | Out |
| CSF1R | 2I1M | 5-CYANO-FURAN-2-CARBOXYLIC ACID [5-HYDROXYMETHYL- 2-(4-METHYL-PIPERIDIN-1-YL)-PHENYL]-AMIDE | Type II | Out |
| CSF1R | 4R7H | PLX3397 | Type II | Out |
| CSF1R | 2I0Y | 5-CYANO-FURAN-2-CARBOXYLIC ACID [5-HYDROXYMETHYL- 2-(4-METHYL-PIPERIDIN-1-YL)-PHENYL]-AMIDE | Type II | Out |
| CSF1R | 3LCD | N~3~-(2,6-dichlorobenzyl)-5-(4-{[(2R)-2-(pyrrolidin-1-ylmethyl)pyrrolidin-1-yl]carbonyl}phenyl)pyrazine-2,3-diamine | Type I | In |
| CSF1R | 2I0V | 6-CHLORO-3-(3-METHYLISOXAZOL-5-YL)-4-PHENYLQUINOLIN-2(1H)-ONE | Type II | Out |
| CSF1R | 3BEA | 8-(2,3-dihydro-1H-inden-5-yl)-2-({4-[(3R,5S)-3,5-dimethylpiperazin-1-yl]phenyl}amino)-5-oxo-5,8-dihydropyrido[2,3-d]pyrimidine-6-carboxamide | Type II | Out |
| CSF1R | 2OGV | Apo structure | NA | Out |