Supporting Information

Hwang et al. 10.1073/pnas.1705381114

SI Materials and Methods

Tanimoto Coefficients. Each of the ligands in the 105,646 holo structures used by LBias was converted into 2D fingerprints using the *FP2* feature from the *Open Babel* package. The 2D fingerprint is a 1,024-bit-long string where each bit is either 1 or 0 reflecting whether the ligand does or does not contain a given chemical group. Fingerprints were used to calculate pairwise Tanimoto coefficients that measure molecular similarity. Specifically,

$$\text{Tanimoto coefficient}(X,Y)=\frac{n(X\cap Y)}{n(X)+n(Y)-n(X\cap Y)}, \quad \textbf{[S1]}$$

where $n(X)$ and $n(Y)$ are the number of 1's in the fingerprint of each ligand X and Y, and $n(X\cap Y)$ is the number of 1's that appear at common positions in the fingerprints of X and Y. For a given two molecules, the Tanimoto coefficient ranges from 0 to 1 with larger values reflecting greater similarity.

Precision–Recall Curves. For a given query protein, all of its residues were sorted based on the value of R described above. The list is scanned in order and for each true-positive ligand-binding residue, *r*, encountered, a precision [TP/(TP+FP)] and recall [TP/(TP+FN)] is obtained and plotted. TP and FP are obtained by counting the number of true positives and false positives ranked above *r*, and FN is obtained by counting the number of true positives ranked below *r*. This procedure is carried out until all true positives are accounted for (i.e., the recall reaches 1).

Naive Bayes for LT-Scanner/Seq. We arbitrarily chose 10 and 20 as the total number of bins for $Sim_{\text{LT-scanner}}$ of LT-scanner and −log of BLAST e-value and divided them into equal intervals. Likelihood ratios (LRs) for each bin are calculated as the ratio of known drug–target interactions with a score in a given bin divided by the ratio of drug–target interactions in a negative set using 10-fold cross-validation.

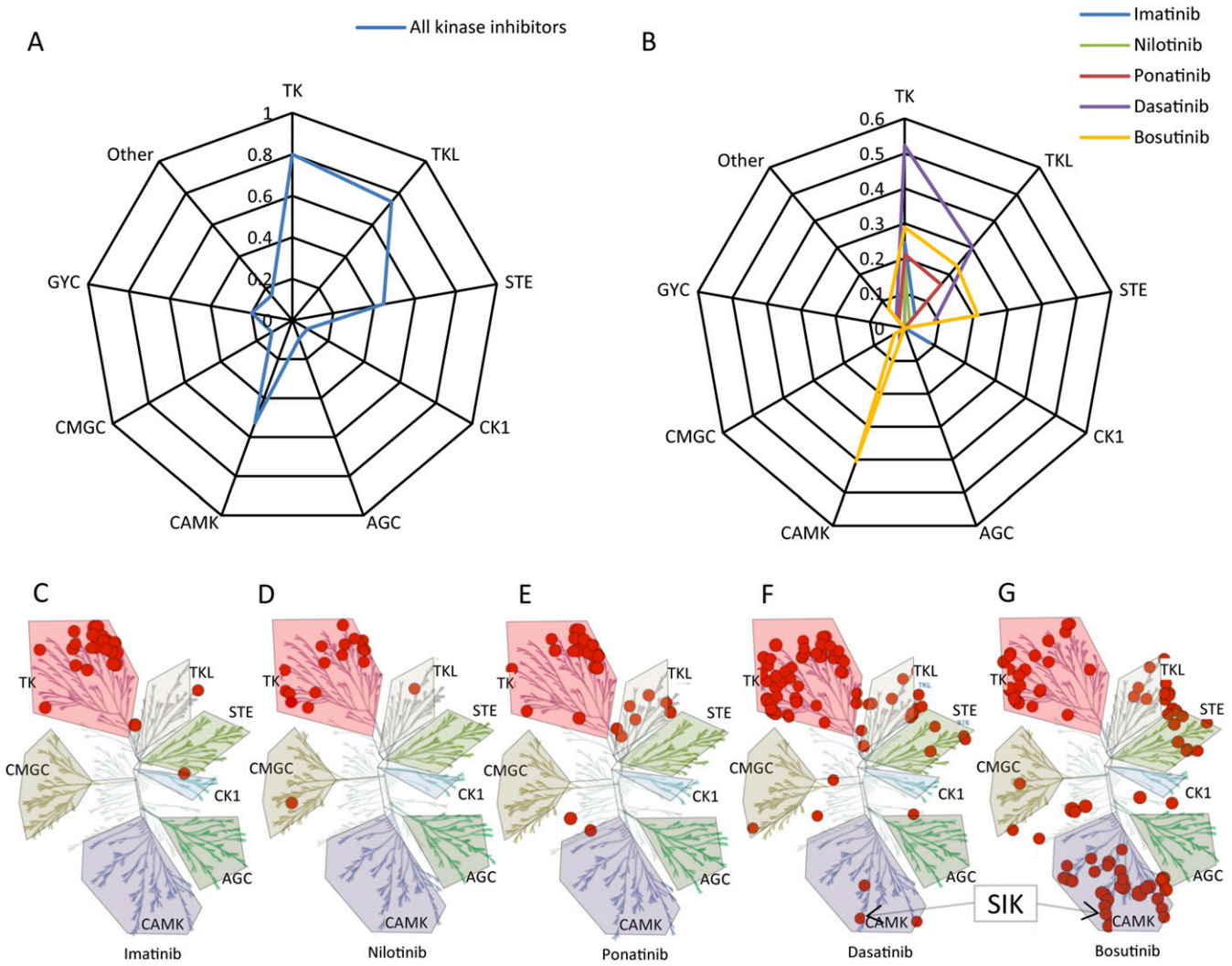

Fig. S1. Distribution of predicted kinase inhibitor targets among the different known kinase families. (*A*) Radar diagram show the distribution, as a fraction, among kinase families of predicted inhibitor targets for 21 known kinase inhibitors. (*B*) Radar diagram of predicted targets of five BCR-ABL tyrosine kinase inhibitors. *C–G* show the distribution of each of the five BCR-ABL tyrosine kinase inhibitors on the human kinome phylogenetic tree. The trees were generated by the Cell Signaling Technology website (https://www.cellsignal.com/). The kinase groups are TK (tyrosine kinase) group, TKL (tyrosine kinase like) group, CaMK (Ca^{2+}/calmodulin-dependent protein kinase) group, STE (homolog-of-sterile kinase) group, CK1 (casein kinase) group, AGC (PKA–PKG–PKC kinase) group, CMGC (CDK, cyclin-dependent kinase; MAP, mitogen-activated kinase; GSK, glycogen synthase kinase; CKD-like kinase) group, and GYC groups, and their relative distances are often shown in the phylogenetic tree. SIK is shown as a predicted target of Dasatinib and Bosutinib.

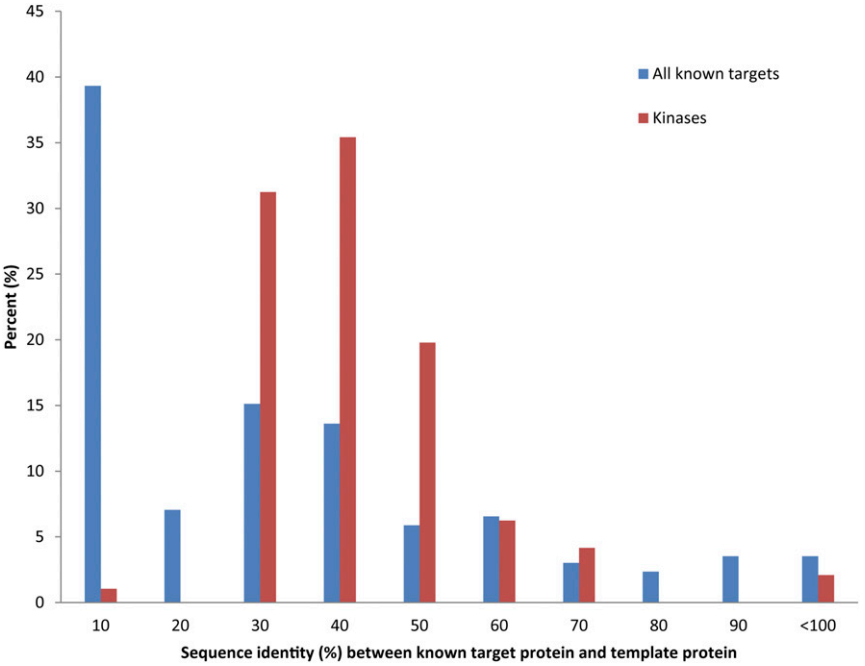

Fig. S2. Distribution of sequence identities (in percentage) between known human target proteins and template proteins that were used for predictions.

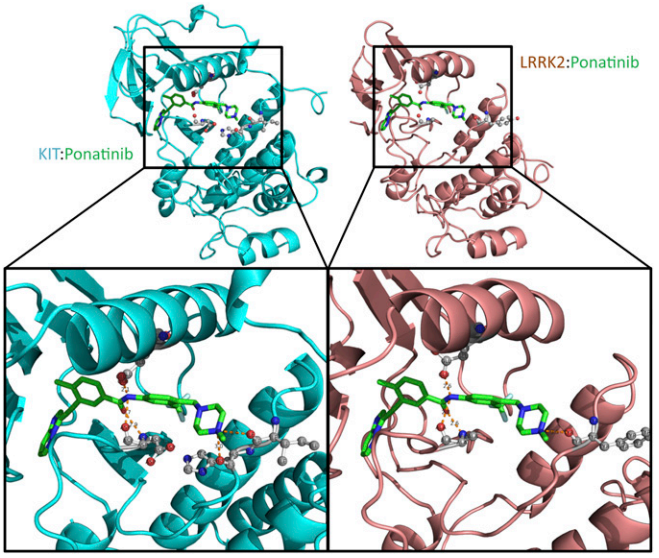

Fig. S3. Hydrogen bond similarity between KIT and LRRK2. The figure shows an example of interaction similarities between KIT (cyan):ponatinib (green) and LRRK2 (salmon):ponatinib (green) using hydrogen bonds. KIT makes five hydrogen bonds (dotted orange lines) with ponatinib. (PDB ID code 4u0i) OE1 atom of KIT GLU640 makes a hydrogen bond with N2 atom of ponatinib and the other four hydrogen bonds are formed through backbone atoms that are O atoms of ILE789, HIS790, and ASP810 and N atom of ASP810 that interact with N4, N4, N2, and O1 atoms of ponatinib, respectively. Based on LT-scanner prediction, LRRK2 makes four (out of the five) similar hydrogen bonds (dotted orange lines) as seen in KIT:ponatinib cocrystal structure. OE2 atom of LRRK2 GLU37 is predicted to form a hydrogen bond with N2 atom of ponatinib. Other three predicted hydrogen bonds are among backbone atoms that are O atoms of TYR101, ALA127, and N atom of ALA127 that interact with N4, N2, and O1 atoms of ponatinib, respectively. It is noteworthy that LT-scanner predicted LRRK2 possibly form similar hydrogen bonds compared with KIT to bind ponatinib through backbone atoms even though their residue types are different.

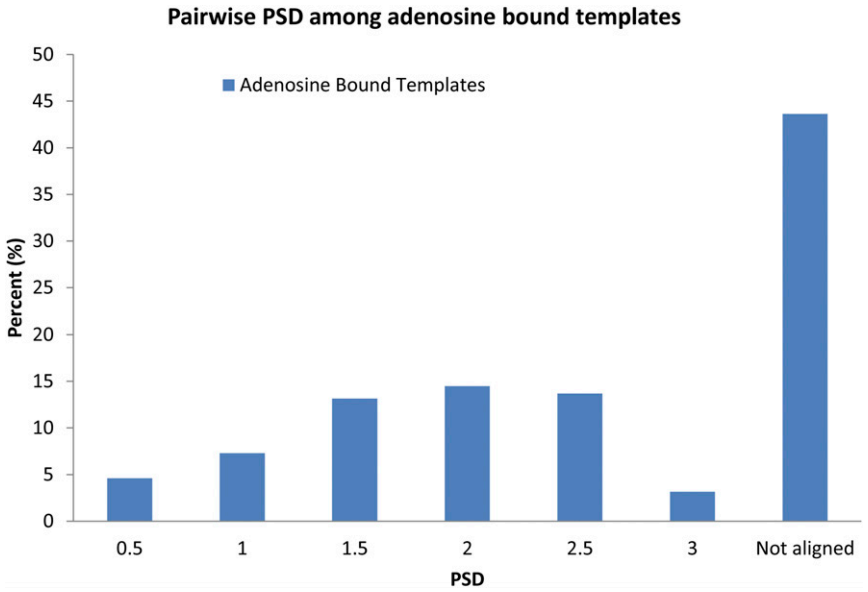

Fig. S4. The figure shows pairwise protein structure distance (PSD) among template proteins for adenosine target prediction. PSD > 1 or "not aligned" means that each protein of a pair most likely belongs to two different protein folds.

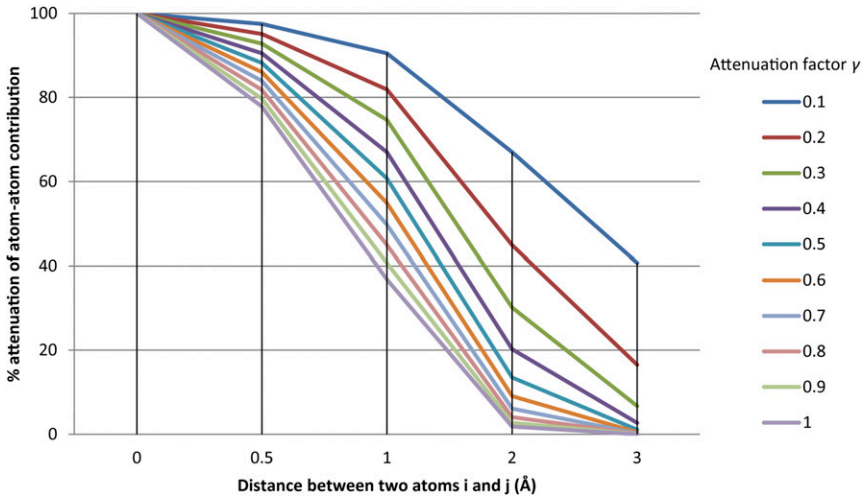

Fig. S5. The figure shows attenuation percentage (*y* axis) of a pair of atoms at difference distance (*x* axis) for different gamma attenuation factors γ (0.1 $\sim$ 1). γ value of 0.7 has 50% attenuation for matching atom pairs at 1 Å away.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)