This is an open access article published under a Creative Commons Non-Commercial No Derivative Works (CC-BY-NC-ND) Attribution License, which permits copying and redistribution of the article, and creation of adaptations, all for non-commercial purposes.

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

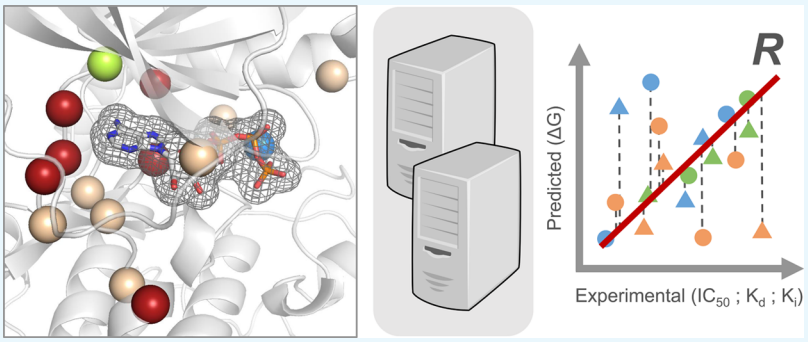
How Far Are We from the Rapid Prediction of Drug Resistance Arising Due to Kinase Mutations?

Mehmet Erguven, Tülay Karakulak, M. Kasim Diril, and Ezgi Karaca*

Cite This: *ACS Omega* 2021, 6, 1254−1265

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: In all living organisms, protein kinases regulate various cell signaling events through phosphorylation. The phosphorylation occurs upon transferring an ATP's terminal phosphate to a target residue. Because of the central role of protein kinases in several proliferative pathways, point mutations occurring within the kinase's ATP-binding site can lead to a constitutively active enzyme, and ultimately, to cancer. A select set of these point mutations can also make the enzyme drug resistant toward the available kinase inhibitors. Because of technical and economical limitations, rapid experimental exploration of the impact of these mutations remains to be a challenge. This underscores the importance of kinase−ligand binding affinity prediction tools that are poised to measure the efficacy of inhibitors in the presence of kinase mutations. To this end, here, we compare the performances of six web-based scoring tools (DSX-ONLINE, KDEEP, HADDOCK2.2, PDBePISA, Pose&Rank, and PRODIGY-LIG) in assessing the impact of kinase mutations on their interactions with their inhibitors. This assessment is carried out on a new structure-based BINDKIN benchmark we compiled. BINDKIN contains wild-type and mutant structure pairs of kinase−inhibitor complexes, together with their corresponding experimental binding affinities (in the form of IC_{50}, K_d, and K_i). The performance of various web servers over BINDKIN shows that they cannot predict the binding affinities (ΔGs) of wild-type and mutant cases directly. Still, they could catch whether a mutation improves or worsens the ligand binding ($\Delta\Delta G$s) where the highest Pearson's R correlation coefficient is reached by DSX-ONLINE over the K_i dataset. When homology models are used instead of K_i-associated crystal structures, DSX-ONLINE loses its predictive capacity. These results highlight that there is room to improve the available scoring functions to estimate the impact of protein kinase point mutations on inhibitor binding. The BINDKIN benchmark with all related results is freely accessible online (https://github.com/CSB-KaracaLab/BINDKIN).

Downloaded via DOKUZ EYLUL UNIV on September 2, 2021 at 11:16:38 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

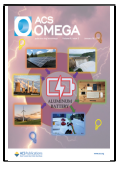
1. INTRODUCTION

In all living organisms, protein kinases are at the heart of numerous signaling pathways.[1] They regulate diverse cell signaling events through phosphorylation. The phosphorylation reaction involves the transfer of an ATP molecule's terminal phosphate to a target residue (serine/threonine/tyrosine/histidine). This transfer adds negative charge to the phosphorylated residue and alters the physicochemical properties of the substrate protein. This phosphorylation is used as a signal to mediate most of the critical cellular processes, such as apoptosis, cell division, cell migration, and transcriptional regulation. Because of the essential role of kinases in cellular homeostasis, deregulation in the protein kinase activity often results in malignancies. Such deregulations might stem from expression level changes, as well as from point mutations at or around the ATP binding pocket of the kinase.

1.1. Fundamentals of Protein Kinase-Mediated Phosphorylation.

Protein kinases mediate catalysis through their interdomain interactions (between the globular N-terminal and C-terminal domains) (Figure 1). This architecture grants modularity to the enzyme to leverage the binding of

Received: September 23, 2020
Accepted: December 11, 2020
Published: January 4, 2021



© 2021 The Authors. Published by American Chemical Society

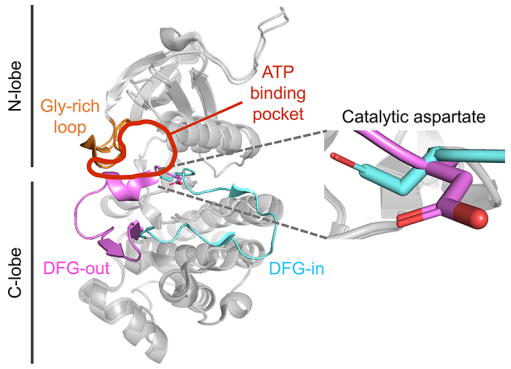


Figure 1. Representative protein kinase structure in DFG-in/-out states. DFG-in and DFG-out conformations of ABL kinase were superimposed (PDB entries, 3KF4 and 3KFA, respectively).[19] In the DFG-in conformation (cyan), the activation loop exposes the ATP-binding pocket. It also orients the catalytic aspartate toward the ATP-binding pocket. In the DFG-out conformation (pink), the activation loop occludes the ATP-binding pocket and orients the catalytic aspartate away from the ATP-binding pocket. The glycine-rich loop is represented in orange.

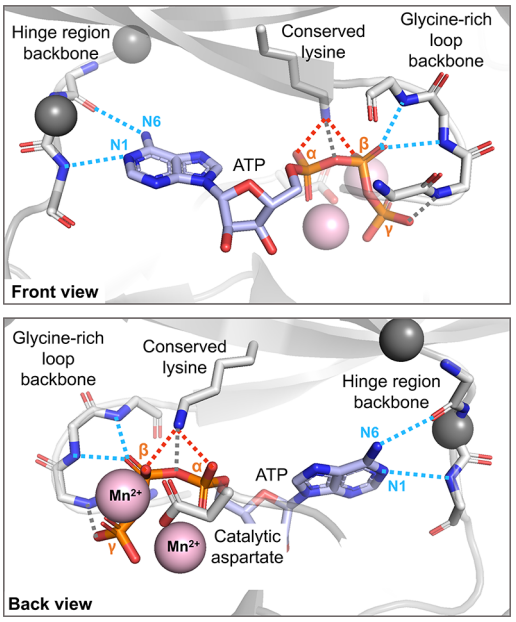


Figure 2. Catalytic subunit of cAMP-dependent protein kinase in complex with an inhibitory peptide (which is not covered by the figure pane) and ATP (PDB entry, 1ATP).[20] The ATP, hinge region, and glycine-rich loop backbone, conserved lysine, and catalytic aspartate are shown in sticks. Manganese ions are depicted as pink spheres. The positions of the two common gatekeeper residues are shown as smaller gray spheres. The blue-, red-, and orange-colored atoms correspond to nitrogen, oxygen, and phosphorus, respectively. The possible dipole−dipole (hydrogen bond), ion−ion (salt bridge), and ion−dipole interactions between the kinase and ATP are depicted in blue-, red-, and gray-colored dashed lines, respectively.

ATP, cofactors, and other protein partners, during the enzyme's catalytic cycle.[2] The smaller N-terminal lobe is mostly composed of antiparallel β-sheets, whereas the larger C-terminal lobe is enriched in α-helices.[3−5] All protein kinases share several catalytically important regions, that is, the gatekeeper residue, the activation loop, the DFG motif, and the glycine-rich loop. The kinase gatekeeper residue resides within the ATP-binding pocket (Figure 2). In the inactive state, the 20−30 residue-long activation loop is found in the DFG-out conformation. In this state, the catalytic aspartate (D of **D**FG), responsible for transferring the phosphate from ATP to the substrate, blocks the catalytic cleft for substrate entry (Figure 1). In the active DFG-in conformation, the ATP-binding pocket is accessible to ATP, while the side chain of the catalytic aspartate faces the ATP-binding pocket (Figure 1).[6,7] The transition from DFG-out to DFG-in states is induced upon phosphorylation of the activation loop or by binding of ATP-competitive kinase inhibitors.[8−10] In its active form, the catalytic aspartate chelates either with magnesium or manganese ions, the essential divalent cofactor cations (Figure 2).[11] Together with these cations, the ATP molecule is coordinated by the glycine-rich loop and a conserved lysine of the ATP-binding pocket. In this coordination, N1 and N6 nitrogen atoms of the adenine ring form specific hydrogen bonds with the backbone of the interdomain hinge region (Figure 2).[12] The partially conserved nonpolar aliphatic residues (leucine, valine, phenylalanine, alanine, and methionine) of the ATP-binding pocket provide van der Waals contacts with ATP's purine moiety.[13] The details on the reaction mechanism and kinetics are provided in the Supporting Information.[4,5,14−18]

1.2. Protein Kinase Deregulation and Drug Resistance due to Point Mutations. Point mutations, within the vicinity of the protein kinase's functionally important regions, can impact cellular processes by altering the specificity of the enzyme or creating a constitutively active protein kinase. These changes can lead to severe clinical outcomes. For example, naturally occurring bulkier side chain substitutions at the gatekeeper position cause drug resistance, as they sterically block the entry of drug molecules into the ATP-binding pocket. Two such gatekeeper mutations are T315I in ABL kinase and T790M in EGFR. These mutations sever the critical hydrogen bond between the gatekeeper threonine hydroxyl and the ligand. They also destabilize the inactive conformation of the enzyme, leading to a constitutively active state. Constitutive activity of these mutants is attributed to the enhanced van der Waals interactions between the phenylalanine of the DFG motif and the mutant gatekeeper residue.[21−24] The hydrogen bond network between the ATP-binding pocket and the adenine ring does not involve the gatekeeper residue side chain. Therefore, the gatekeeper residue mutations do not hinder the binding of ATP.[25−27] Instead, the adenine ring engages in a conserved hydrogen bonding network, involving the hinge region (backbone) of the enzyme (Figure 2). Also, the increased space occupation by the larger gatekeeper residue does not impact the accommodation of ATP because of the large spatial gap between the ATP molecule and the gatekeeper residue.

Frequently, mutations in cancer can also be found within the N-lobe side of the activation loop.[28] For example, V600 of BRAF is a commonly mutated residue in cancer. The V600D/E/G/K/L/M/R mutations of BRAF result in a constitutively active enzyme. This mutation position corresponds to activating D1228V/N/H in MET and D816E/H/V/N/F/Y/I in KIT. Here, we also see that physiochemically opposing mutations can yield to hyperactivation: while mutation of valine to aspartate cause overactivation in BRAF, mutation of aspartate to valine leads to overactivation in MET and KIT kinases. Other such prominent examples are L858R (activation loop) and G719S (proximal to glycine-rich loop) mutants of

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

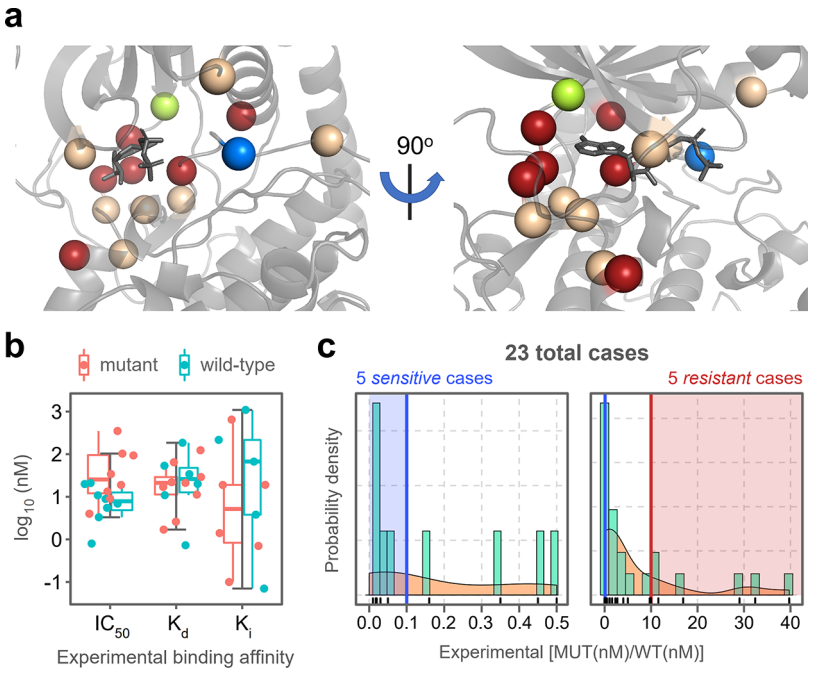


Figure 3. BINDKIN benchmark characteristics. (a) The corresponding positions of the mutated residues are mapped on a representative ATP-bound protein kinase (cAMP-dependent protein kinase with the PDB entry 1ATP).[20] The green sphere is the position where the majority of BINDKIN mutations are located. The blue sphere indicates the location of the *sensitizing* mutations, where the red spheres indicate the mutations conferring *resistance*. The light brown ones correspond to the *neutral* mutations. (b) The distribution of the binding kinetics data for 42 individual cases. The data was split into K_d, K_i, and IC_{50} subsets and plotted as box-and-whisker plots for the corresponding wild-type and mutant cases. (c) The probability distribution of the kinetic data for wild-type and mutant pairs. The blue and red areas indicate the *drug-sensitive* and *drug-resistant* cases, respectively. The KDE (kernel density estimate) is represented by the light-brown area. The distribution of all cases covering the range of 0−40 nM were given on the right. The subset of cases within the range of 0−0.5 nM were explicitly highlighted on the left.

EGFR, which, similar to the gatekeeper residue mutations, destabilize the inactive conformation of the kinase.[29] These observations highlight the complex nature of mutation-induced physiochemical changes in protein kinases.

1.3. Utilization of Kinase Mutations for Functional Studies. Drug discovery and chemical genetic fields frequently make use of ATP-binding pocket mutations to study the kinase function. In the case of drug discovery, the protein kinase gatekeeper and other critical residues are artificially mutated in model systems to study the mechanism of drug resistance. This approach allows scientists to tailor novel compounds for the treatment of drug-resistant tumors.[30] In chemical genetics, naturally occurring mutations are used as a tool to elucidate the biological roles of protein kinases.[31] For this, the kinase gatekeeper residue is mutated to glycine or alanine through gene editing, which enables the binding of a bulky inhibitor. As a result of this reverse engineering, protein kinases become sensitive to inhibitors, which can selectively turn the enzyme off without interfering with other kinases.[26,32] Such engineered kinases include yeast v-Src (I338G), c-Abl (T315A), Cdk2 (F80G),[33] Mps1 (M516G),[34] and human Cdk12 (F813G).[35] Besides their numerous advantages, exploiting the ATP-binding pocket mutations also has potential handicaps, such as incompatibility of the mutation−drug combinations or reduced enzyme activity after mutation. To minimize these risks, prescreening of mutation−drug combinations with binding affinity prediction tools stands out as a promising strategy.

1.4. How Far Are We from Predicting the Impact of Mutations on Protein Kinase−Ligand Binding? Several methods have been proposed to predict protein−ligand binding affinities. However, only recently, the accuracy of these methods has been evaluated for predicting the impact of protein kinase mutations on their ligand binding. In 2018, Hauser et al. compiled a benchmark set, encompassing 144 ligand-binding affinity data of clinically relevant Abl mutations.[36] A year later, Aldeghi et al. assessed the capacity of statistical mechanics, mixed physics- and knowledge-based potentials, and machine-learning approaches in predicting the impact of 31 Abl mutations on their drug interactions.[37] Both of these approaches used a hybrid structure-based Abl benchmark, composed of wild-type co-crystal/docked structures and modeled mutant cases. Moreover, both papers assessed the use of sophisticated tools, the use of which would be too complicated for many experimental biologists. Next to these approaches, there are also web-based protein−ligand binding affinity predictors, which can be easily used by nonexperts to plan/guide their experiments. However, a systematic assessment of these tools within the context of protein kinase mutations has never been performed. To address this need, here, we have benchmarked four web-based protein−ligand scoring functions, DSX-ONLINE, KDEEP, Pose&Rank, and PRODIGY-LIG, as well as two general scoring functions, PDBePISA and HADDOCK Score,[38−45] for predicting the impact of protein kinase mutations on their ligand binding. We have selected these predictors based on their user-friendliness and widespread use. The benchmarking has been carried out on our structure-based **BINDKIN** (effect of point mutations on the **BIND**ing affinity of protein **KIN**ase−ligand complexes) data set.

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

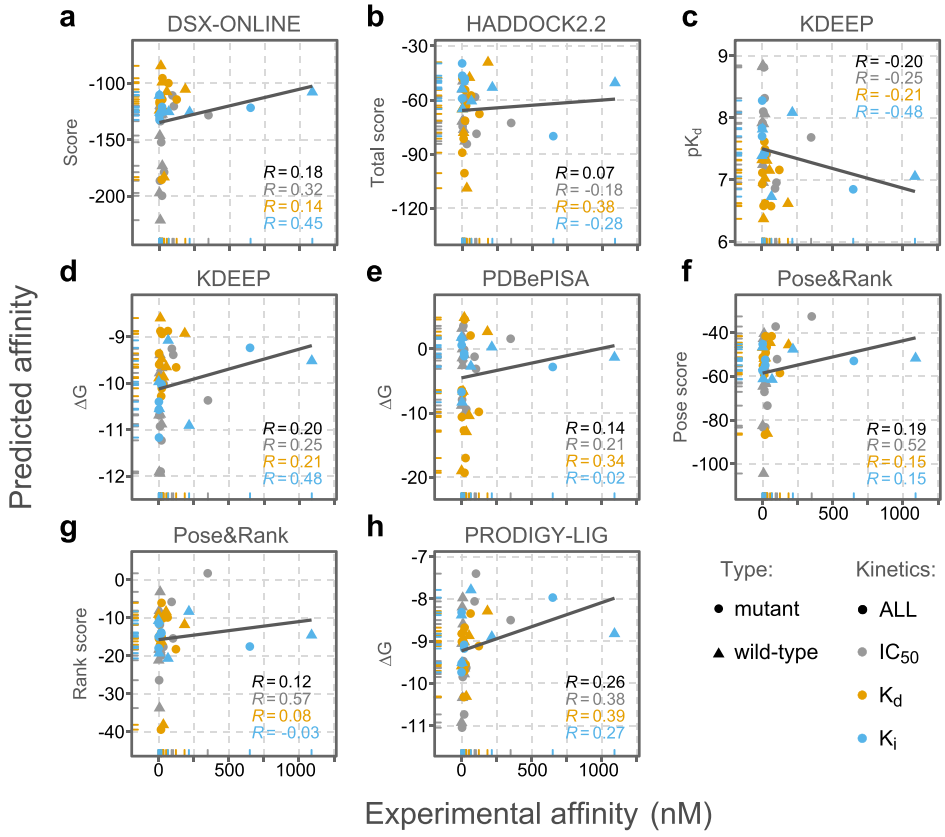


Figure 4. Performance of various scoring functions according to *direct assessment*. Because of the nature of prediction scores, a negative correlation for pK_d (KDEEP) and a positive correlation for the other scoring terms are expected. (a) DSX-ONLINE, (b) HADDOCK2.2 (*HADDOCK score*), (c,d) KDEEP (pK_d and ΔG), (e) PDBePISA (ΔG), (f,g) Pose&Rank (*PoseScore* and *RankScore*), and (h) PRODIGY-LIG (ΔG) performances. The sample sizes for IC_{50}, K_d, and K_i data are n = 16, n = 15, and n = 11, respectively.

2. RESULTS AND DISCUSSION

2.1. Characteristics of the BINDKIN Benchmark. The BINDKIN benchmark covers nine unique protein kinases (seven EGFR, three Abl, three Mps1, three Src, two Cdk2, one ALK, one FGFR, one Kit, and one PKA). In the benchmark, each wild-type crystal structure has a mutant counterpart, where both of the cases are bound to the same drug. While compiling the BINDKIN set, we have made sure that wild-type and mutant structures and protein kinase−drug binding affinities come from the same study, so that the experimental values are comparable. These strict criteria resulted in the collection of 23 protein kinase−drug pairs (with 8, 9, and 6; IC_{50}-, K_d-, and K_i-cases, respectively) and 42 individual cases (with 16, 15, and 11; IC_{50}-, K_d-, and K_i-cases, respectively) (Table S1). Despite being ideal to convert different types of kinetic data to a common metric, that is, Gibbs-free energy (ΔG), we could not perform this conversion due to the unavailability of substrate concentrations.

The 23 mutant cases in the BINDKIN benchmark include 17 single, three double, two triple, and one quintuple point mutants. A total of 36 mutations are distributed across 15 unique positions, most of which are located within or at the boundaries of functionally crucial segments: six positions are at the hinge region; two positions at the activation loop; and one position at the glycine-rich loop (Figure 3a). In their wild-type states, these mutation positions exert diverse biophysical characteristics (charged, polar, and hydrophobic), whereas in their mutant forms, they predominantly turn into hydrophobic and charged amino acids. Approximately, half of the mutations lead to bulkier residues. 8 out of 42 structures are in DFG-out, and the remaining structures are in DFG-in conformation. Except two case pairs, the wild-type and mutant forms of all proteins share a common DFG state. Among these two pairs, the wild-type 4WA9 and 5AP1 are in the DFG-out state, while the mutant 4TWP and 5AP4 are in the DFG-in state. The DFG state change induced by these mutants may cause the production of constitutively active protein kinases.

The 18 ligands presented in BINDKIN are all ATP-competing reversible protein kinase inhibitors. These inhibitors are usually composed of five or six-membered, mostly heterocyclic rings, which enrich the nitrogen content of the ligands (Table S1, Figure S1). The rings are generally linked by secondary amine, ether, or ketone groups. Because of the abundance of unprotonated nitrogen atoms in the rings and unprotonated oxygen atoms in the linker ether and ketone groups, the hydrogen bond acceptors are more abundant than the donors. Regarding the conformational degrees of freedom of inhibitors, BINDKIN contains both rigid and flexible ligands, as the number of rotatable bonds in the ligands range from 0 to 11. We have looked for a relationship between the ligand properties and the reported binding affinities, but could not observe a correlation between them (Figure S2).[46]

BINDKIN spans the binding affinity ranges of 0.80−350, 0.73−185, and 0.07−1090 nM for IC_{50}, K_d, and K_i subsets, respectively (Table S2).[47−62] Therefore, the K_i subset covers the broadest binding affinity range (Figure 3b). We classified a mutation as drug-sensitive if the binding affinity of the mutant complex is at least 10-folds higher (lower IC_{50}, K_d, or K_i) than

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

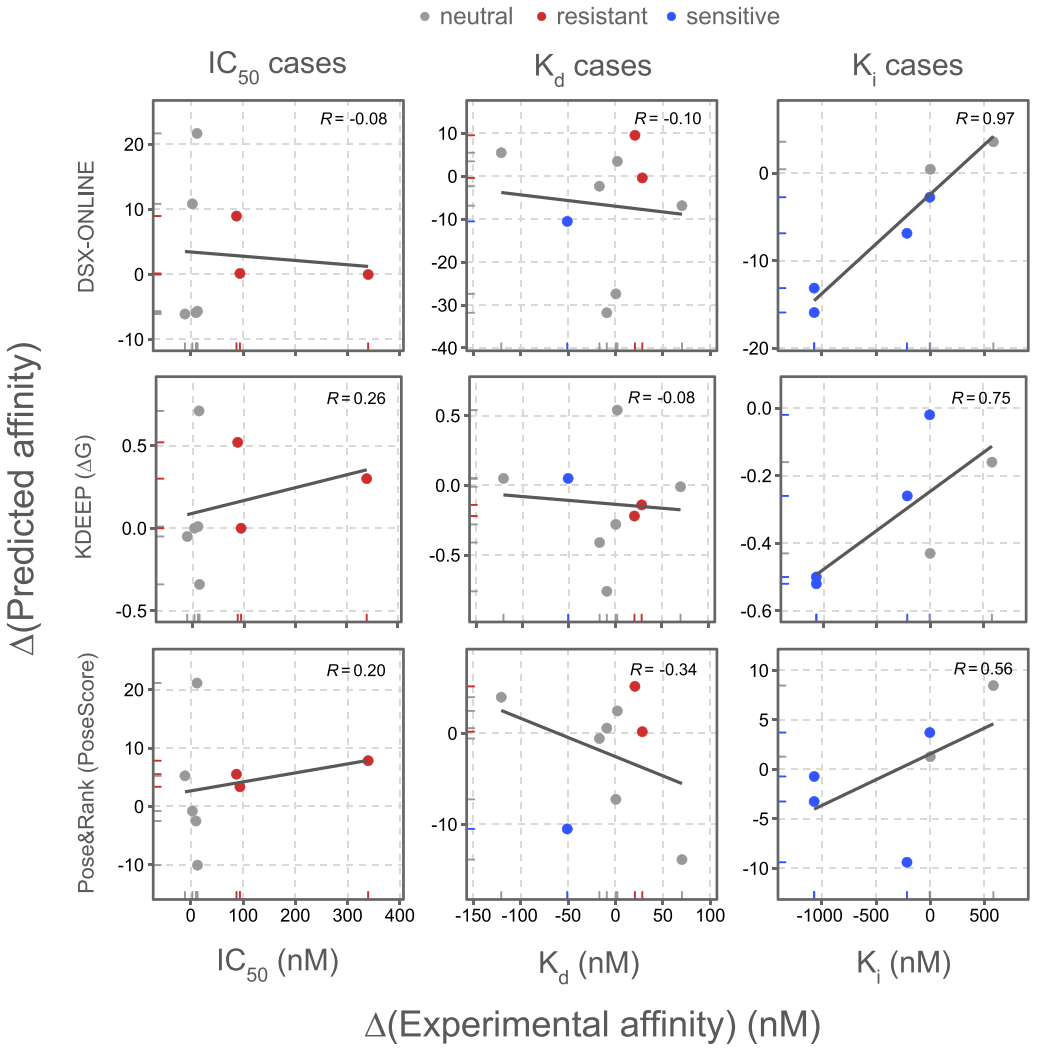


Figure 5. Performance of various scoring functions according to the *delta assessment*. The *delta assessment* of DSX-ONLINE, KDEEP (ΔG), and Pose&Rank (*PoseScore*) is characterized based on the drug resistance/sensitivity of the cases. The data are evaluated as subsets based on the kinetic metrics. The neutral (gray), drug-resistant (red), and drug-sensitive (blue) cases are indicated.

that of its wild-type state.[36,37,63] When the binding affinity of the mutant complex is lower (higher IC_{50}, K_d, or K_i) than one-tenth of its wild-type state, then this case is considered as *drug-resistant*. The remaining cases are classified as *neutral*. According to these criteria, BINDKIN contains 5 *sensitive* (one K_d and four K_i), 5 *resistant* (three IC_{50}, and two K_d), and 13 *neutral* (five IC_{50}, six K_d, and two K_i) cases (Figure 3c). For more details on our BINDKIN benchmark, please see the Methods and https://github.com/CSB-KaracaLab/BINDKIN.

2.2. Majority of Web-Based Scoring Functions can Predict the ΔΔ*G*-Related Changes Caused by Kinase Mutations. Advanced computational approaches have been demonstrated to accurately predict protein kinase−ligand binding affinities.[37,63] Considering that these approaches do not have a broad reach among the experimental biology community, we have focused on user-friendly web servers' performance in predicting the impact of kinase mutations. To this end, we have tested DSX-ONLINE, HADDOCK2.2 (refinement interface), KDEEP, PDBePISA, Pose&Rank, and PRODIGY-LIG servers on the BINDKIN benchmark. To assess the performance of each tool, we have pooled the affinity predictions (n = 42) and analyzed their association with the experimental data. This *direct assessment* shows that the presented scoring functions could not predict the affinities of BINDKIN cases (Figure 4). Here, the highest R-value is produced by PRODIGY-LIG (R = 0.26). When we divide the same set according to their kinetic metrics (IC_{50}, K_d, K_i), the highest correlation is obtained by Pose&Rank over IC_{50} cases (R = ∼0.50). This performance is followed by KDEEP and DSX-ONLINE ($0.40 < R < 0.50$, IC_{50} cases). Assessing wild-type and mutant states separately did not improve the presented correlations (data not shown).

Next to the *direct assessment*, we have used two other metrics, that is, *delta* and *binary assessments* (Figure S3, see the Methods, Section 4.4). The *delta assessment* measures the absolute change in the affinity upon mutation (correlated to $\Delta\Delta G$), improving the R values significantly. This is especially the case for the R-performances of DSX-ONLINE, KDEEP, and *PoseScore* over K_i-cases, which become 0.97, 0.76, and 0.56, respectively. When the performance of these three predictors is analyzed from the perspective of drug resistance/sensitivity, it is observed that for each scoring function, different cases deviate from the linear behavior (Figure 5). This means that distinct cases pose different prediction difficulties for the presented scoring functions. In general, significant binding affinity changes, leading to a gain of drug

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

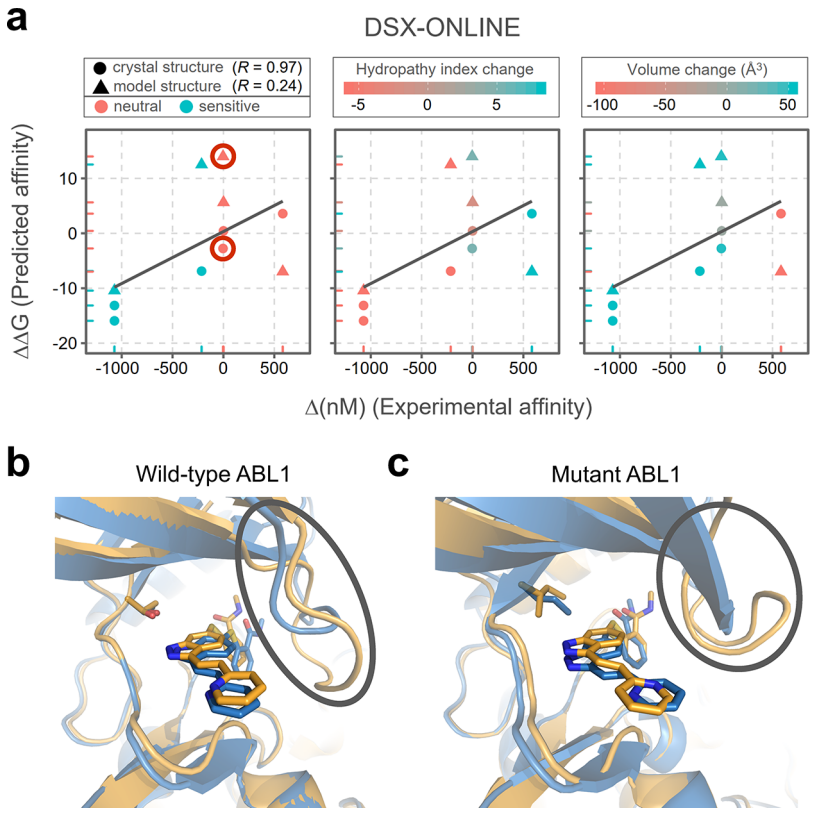


Figure 6. The default scoring function of DSX-ONLINE is tested on the K_i-associated crystal and homology model structures using *delta assessment*. (a) Pink color corresponds to the neutral cases (left) or a decrease in the hydropathy index (middle) or residue volume (right). Turquoise color corresponds to the sensitive cases (left) or an increase in the hydropathy index (middle) or residue volume (right). (b) The crystal structure (PDB entry, 4WA9, orange) and the model structure (blue) of wild-type human ABL1 kinase were aligned (RMSD = 2.9 Å). (c) The crystal structure (PDB entry 4TWP, orange) and the model structure (blue) of T315I mutant human ABL1 kinase were aligned (RMSD = 2.7 Å). The ATP-binding pockets of the superposed structures are shown for both alignments. The ligand, the residue to be mutated, and the mutant residue are shown in sticks. The glycine-rich loops are encircled with the gray ellipses.

sensitivity or resistance, are easier to predict. The K_i-subset is rich in drug-sensitive cases. Consequently, having the K_i-subset dominated by cases with significant binding affinity changes could be the reason for the good prediction performances over the K_i-cases. Unfortunately, none of the presented scoring functions could predict the neutral behavior of IC_{50}- and K_d-cases.

In the *binary assessment*, the accuracy of predictors was measured upon analyzing the agreement between the direction of change in the experimental and predicted binding affinities. According to this metric, DSX-ONLINE accurately predicts the affinity change of all K_i-cases. The success rate of DSX-ONLINE over K_i-cases is followed by *PoseScore* and PRODIGY-LIG, both leading to 83% accuracy. As another important highlight, KDEEP (pK_d) reaches 88% success rate in IC_{50} cases. When the neutral cases were excluded from the analysis, it is observed that DSX-ONLINE, PRODIGY-LIG, and *PoseScore* can predict the drug-sensitive and drug-resistant cases, with 100, 90, and 90% accuracies, respectively. For the neutral cases, the best performance is achieved by DSX-ONLINE and PRODIGY-LIG, with 62 and 54% accuracies. Therefore, both *delta* and *binary assessments* are in agreement with the fact that the mutations causing the most dramatic binding affinity changes (resistance or sensitivity) are easier to predict. However, the amount of neutral cases (n = 13) is almost 3 times more than the drug-sensitive (n = 5) or drug-resistant (n = 5) ones. Therefore, we cannot rule out the fact that small and different sample sizes could have an impact on the observed prediction accuracy.

As an across-metric comparison, over K_i-cases, DSX-ONLINE leads to R values of 0.45 and 0.97 for *direct* and *delta assessments*, respectively, and 100% accuracy for the *binary assessment*. This accuracy improvement is explained by the fact that while strong binders (complexes with lower K_i values) show an orthogonal spread in the direct assessment, they show a linear spread for the *delta assessment* (Figure S4). While *delta* and *binary assessments* agree in this particular case, in some other cases, *delta* and *binary assessments* contradict. For example, PRODIGY-LIG achieves an R-value of 0.29 in the *delta assessment* and 83% accuracy in the *binary assessment* over the K_i-cases. This means that a scoring function may achieve great success on the *binary assessment* because it can predict the direction of affinity change. However, the same scoring function may not be able to predict the amount of change ($\Delta\Delta G$), which would lead to a lower R-value. Within the context of the biological meaning of the assessment approaches, we suggest the concurrent use of *delta* and *binary assessments*. After testing the web servers with crystal structures, we have scored the HADDOCK-refined complexes with the probed web servers too. These efforts have revealed that structural refinement of crystal structures did not improve the prediction accuracy (data not shown).

Finally, among the three different kinetic data types, according to the *direct assessment*, IC_{50} turns out to be the

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

best descriptor for most of the scoring functions. When the complexes are considered as mutant/wild-type pairs (*delta* and *binary assessments*), K_i comes out as the best descriptor (Figure S3). Interestingly, none of the methods could lead to a reliable prediction over the K_d set. K_d is a universal measure of binding, while K_i is defined for enzymes only, which could be the reason for the high R values reported for this set.[64] However, in this case too, we cannot rule out the fact that success rates of the kinetic descriptors may depend on the number of cases available for each descriptor (eight pairs for IC_{50}, nine pairs for K_d, and six pairs for K_i).

2.3. Structure Quality has a Substantial Impact on the Prediction Accuracy. In a real-life scenario, the experimental structure of the kinase under study might not be available. Taking this into consideration, we have homology modeled the K_i-cases and probed these structures with DSX-ONLINE (as this combination leads to the best performance for the *delta and binary assessments*). The mutation-specific features such as hydropathy index change,[65] residue volume change,[66] and drug resistance or sensitivity were chosen for the characterization of the affinity data Figure 6a). As a result of this exercise, we show that the performance of DSX-ONLINE is dramatically worsened by the worsened quality of the input structure (an R-drop from 0.97 to 0.24). We also observe a decrease in the residue volume upon mutation accompanied by an increase of polarity and drug sensitivity. To understand how modeling has impacted the conformations, we have performed a minimal structure examination. The case pair with the visibly highest predicted affinity difference was chosen for this exercise (Figure 6a, red circled data points). In this regard, the human ABL1 kinase−AXI complex is the choice. The wild-type and mutant kinase structures of this case pair (4WA9, wild-type; 4TWP, mutant) has the highest root-mean-square deviation (rmsd) between their backbone heavy atoms, when compared with the other K_i-cases (Figure 6b,c). The structural comparison of ABL1 and ABL T315I mutant shows that the conformation of the glycine-rich loop is distinctly altered upon modeling, misleading the DSX-ONLINE prediction. This glycine-rich loop is an integral element of the ATP-binding pocket, which is in direct contact with ABL's ligand AXI. This example highlights the importance of proper mutation modeling in the prediction of mutant protein kinase−drug binding affinity changes.

3. CONCLUSIONS

Our work investigates whether the web-based scoring tools can predict the impact of protein kinase mutations on their ligand-binding affinities. To test this, we have compiled BINDKIN, the first structure-based affinity benchmark of wild-type and mutant protein kinase−inhibitor complexes. Based on the K_i-associated cases of BINDKIN, we have generated a homology model benchmark too. Using the crystal structure sets and their available experimental binding affinity data, we have assessed the prediction accuracy of six different web-based scoring tools, DSX-ONLINE, HADDOCK2.2, KDEEP, PDBePISA, Pose&Rank, and PRODIGY-LIG. Overall, ligand-specific scoring functions (used by DSX-ONLINE, KDEEP, Pose&Rank, and PRODIGY-LIG) performed better than the general scoring functions (used by PDBePISA and HADDOCK2.2). When the ligand-specific modified HADDOCK[67] scoring function was used, the prediction accuracy did not improve significantly (data not shown).

When a deeper analysis is performed, we have observed that (i) K_i is a sensitive descriptor that can capture the effect of mutations on the binding affinity, (ii) DSX-ONLINE is the best tool in predicting the mutation-induced K_i changes, (iii) the quality of the input structure significantly impacts the prediction accuracy. Therefore, despite the recent advances presented by machine-learning tools in the field of computational structural biology, in our study, a classical knowledge-based approach (DSX-ONLINE) still performed better than a machine-learning approach (KDEEP).

All in all, as an answer to our title question, our results show that we are currently far from rapidly and accurately predicting the impact of protein kinase mutations, especially when the kinase structure is not at hand. A more comprehensive benchmark than BINDKIN is necessary to train and improve the available predictors in a reliable manner. In this regard, drug companies working on mutant oncogenic protein kinases can facilitate a consistent and a more comprehensive data set by providing their related experimental data. From our end, to assist the kinase-research field, we are sharing our files associated with this work online at https://github.com/CSB-KaracaLab/BINDKIN.

4. METHODS

4.1. Collection of the BINDKIN Benchmark. To construct the **BINDKIN** (effect of point mutations on the **BIND**ing affinity of protein **KIN**ase−ligand complexes) benchmark, we performed a thorough search in the Protein Data Bank (PDB)[68] (https://www.rcsb.org/). We obtained the final list of wild-type and mutant kinase−ligand pairs according to the following criteria:

- For each mutant protein kinase−ligand complex, there has to be a wild-type complex, containing the same protein and the ligand.
- Structures of wild-type and mutant complexes should originate from the same research article.
- For each complex pair, there has to be experimentally determined binding affinity data, originating from the same research. The experimental binding affinity data were acquired from PDBbind (http://www.pdbbind-cn.org/index.php),[69−71] Binding DB (https://www.bindingdb.org/bind/index.jsp),[72] and Binding MOAD (http://bindingmoad.org/Search/advancesearch)[73,74] databases.
- The ligand has to be an ATP-competing reversible (noncovalent) inhibitor.

These criteria left us with 23 pairs of wild-type and mutant complexes, making up the BINDKIN benchmark. In the benchmark, three kinases are found in the dimeric state (PDB IDs: 2IW8/2IW9, 3G0F, and 4EOR/4EOK). However, in these cases, the protein−protein dimerization interface is not involved in the ligand interactions. Therefore, these cases have been treated as monomeric kinase complexes. The BINDKIN benchmark includes seven EGFR, three Abl, three Mps1, three Src, two Cdk2, one ALK, one FGFR, one Kit, and one PKA proteins. These complexes present 36 point mutations distributed across 15 unique positions within or in the vicinity of the ATP-binding pocket (Figure 3a).

BINDKIN covers binding modes of 18 different inhibitors (Figure S1). The 2D images of the ligands were generated from their respective unique SMILES strings, using the Smi2Depict tool of ChemDB Chemoinformatics Portal

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

Table 1. List of Benchmarked Servers

tool name	brief explanation	input format	link
DSX-ONLINE[41]	a knowledge-based scoring function composed of distance-dependent pair potentials, torsion-angle potentials, and solvent-accessible surface-dependent potentials	takes the protein and ligand coordinate files in pdb and mol2 formats	http://pc1664.pharmazie.uni-marburg.de/drugscore/index.php
KDEEP[39]	a deep convolutional neural network approach. The web server represents the binding site as a 24 Å voxel for pharmacophore featurization, considering eight pharmacophoric-like features. It reports pK_i and ΔG scores	takes the protein and the ligand coordinate files in pdb and sdf formats, respectively	https://www.playmolecule.org/Kdeep/
Pose & Rank[38]	*PoseScore* and *RankScore* are atomistic distance-dependent statistical scoring functions. The *PoseScore* is optimal for distinguishing the native binding geometries of compounds from other poses, whereas *RankScore* is optimal for distinguishing ligands from nonbinding small molecules	takes the protein and ligand coordinate files in pdb and mol2 formats, respectively	https://modbase.compbio.ucsf.edu/poseandrank/
PRODIGY-LIG[40]	an atomistic contact-based protein−small molecule binding affinity predictor	takes the protein and ligand coordinates in pdb format	https://bianca.science.uu.nl/prodigy/
HADDOCK2.2[42,91]	HADDOCK2.2 is a molecular docking program, using both force field-based and empirically-derived scoring terms. In this study, we used the refinement interface of HADDOCK2.2	Takes the protein and the ligand coordinates in pdb format	https://milou.science.uu.nl/services/HADDOCK2.2/haddockserver-refinement.html
PDBePISA[43−45]	scores according to the average free energy contributions of the hydrogen bonds and salt bridges between the subunits and the solvation free energy of folding	takes the protein and the ligand coordinates in pdb format	https://www.ebi.ac.uk/pdbe/pisa/

(http://cdb.ics.uci.edu/cgibin/Smi2DepictWeb.py).[75] The pharmacophoric characteristics of ligands were evaluated with the ALOGPS web server of the Virtual Computational Chemistry Laboratory (http://www.virtuallaboratory.org/web/alogps/)[76−85] and the pkCSM web server (http://structure.bioc.cam.ac.uk/pkcsm).[86] The available ligand-related data were obtained from RCSB PDB and PubChem (https://pubchem.ncbi.nlm.nih.gov/). All the benchmark-related features are deposited in https://github.com/CSB-KaracaLab/BINDKIN. Further details about the benchmark characteristics are described in the Results and Discussion section.

4.2. BINDKIN-Homology Model Benchmark. When the structure of the protein kinase under study is not at hand, the researchers often resort to homology modeling to predict the relevant structure. Therefore, to simulate a realistic scenario, we established a homology model benchmark for the K_i-associated cases. Here, we chose the K_i-associated cases, as their binding affinities were predicted with the overall highest accuracy based on the available BINDKIN crystal structures.

The wild-type and mutant sequences were structurally modeled with the default settings of the I-TASSER web server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/).[87−89] During modeling, the original coordinates of the wild-type or mutant structures were excluded from the template list. After obtaining the homology models, their corresponding ligands were placed at their catalytic-binding pocket using the original crystal structure as a template. The fitting was performed with PyMOL.[90] In these crude models, steric clashes were observed. To optimize the protein−ligand interface, we applied water refinement on each model using the HADDOCK2.2 refinement interface (https://milou.science.uu.nl/services/HADDOCK2.2/haddockserver-refinement.html).[42] The HADDOCK refinement performs very short molecular dynamics simulation cycles in explicit water.

4.3. Benchmarking of the Web Servers. The molecules present other than the protein and the ligand either led to an error in the web servers or were disregarded by the web servers. We should note here that DSX-ONLINE has an option to include water molecules in the calculations. However, to measure the performance on a standardized set of crystal structures, we decided to use the same input in all servers. Therefore, before benchmarking, the crystallization buffer additives, crystal water molecules, and ions were removed from the co-crystal structures. For each occurrence of multiple conformations, the conformer with the highest occupancy was retained, and the other conformers were removed. In one of the EGFR cases (5EM7), the ligand 5Q4 is present in two distinct conformations. Both conformations were taken into consideration.

The ligand coordinate files were converted to "mol2" and "sdf" formats based on the different file format requirements of the web servers. For mol2 conversion, Open BABEL (v2.4.1), and for sdf conversion, Online SMILES Translator and Structure File Generator (National Cancer Institute) (https://cactus.nci.nih.gov/translate/) were used. For the sdf conversion, the "Aromatic" SMILES representation option (prints the unique SMILES string into the "sdf" file) and the "3D" coordinate option were chosen. The sdf files were protonated by default. The conformations and the 3D coordinates of the ligands were retained during conversion to both mol2 and sdf formats.

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

Six different web-based scoring functions were used to predict the binding affinities listed within BINDKIN. A short description of each web server is provided in Table 1. Among the web servers, DSX-ONLINE became inaccessible shortly after the completion of its benchmarking.

4.4. Assessment of the Prediction Results. Three different approaches were used to assess the prediction results. These are *direct*, *delta*, and *binary assessments*. The *direct assessment* is a classical approach, where a linear correlation between all scores and experimental data (wild-type and mutant) were sought. For the *delta assessment*, a direct relationship between the mutation-induced binding affinity (Δba_{exp}, eq 1) and the mutation-induced score (Δba_{pred}, eq 2) changes was sought.

$$\Delta ba_{exp} = ba_{exp_mutant} - ba_{exp_wild\text{-}type} \quad (1)$$

$$\Delta ba_{pred} = ba_{pred_mutant} - ba_{pred_wild\text{-}type} \quad (2)$$

In the case of *binary assessment*, an agreement in the direction of mutation-induced change was required. For example, if a mutation leads to a worsening of the experimental binding affinity, and it is predicted as such by the computational prediction, it was considered as a successful prediction. If a worsening in the predicted affinity accompanies a mutation-induced improvement in the experimental affinity, then this case was counted as a misprediction. In the *binary assessment*, the percentage of correct predictions was used as the overall success rate.

For linear regression analyses, the associated R and p values and the plots were generated with R (v3.6.1),[92] invoked by RStudio (v1.2.5001).[93] The following R packages were used to generate scatterplots, barplots, or histograms: ggpubr,[94] ggplot2,[95] magrittr,[96] ggrepel,[97] grid,[92] and gridExtra.[98] The heatmap was produced with pheatmap[99] and RColorBrewer[100] R packages.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c04672.

> Proposed catalytic mechanism and kinetics for protein kinases; 2D structure of the ligands present in BINDKIN; physicochemical characteristics of the ligands; performance of the various scoring functions; assessment of DSX-ONLINE over the crystal structures of the K_i cases; content and the basic characteristics of BINDKIN (PDF)

Accession Codes

PDB IDs: 3KF4, 3KFA, 1ATP, 4WA9, 4TWP.

■ AUTHOR INFORMATION

Corresponding Author

Ezgi Karaca − *Izmir Biomedicine and Genome Center, 35330 Izmir, Turkey; Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, 35340 Izmir, Turkey;* orcid.org/0000-0002-4926-7991; Email: ezgi.karaca@ibg.edu.tr

Authors

Mehmet Erguven − *Izmir Biomedicine and Genome Center, 35330 Izmir, Turkey; Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, 35340 Izmir, Turkey*

Tülay Karakulak − *Izmir Biomedicine and Genome Center, 35330 Izmir, Turkey; Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, 35340 Izmir, Turkey*

M. Kasim Diril − *Izmir Biomedicine and Genome Center, 35330 Izmir, Turkey; Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, 35340 Izmir, Turkey*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c04672

Author Contributions

This research was devised and performed by M.E., T.K., and E.K. The paper was written by M.E. and E.K. All authors contributed to careful revising of the manuscript.

Funding

This work was supported by the EMBO Installation Grant (project # 4421). The first author was also supported by a TUBITAK 1002 stipend (project # 217Z248) during this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Gerard Martinez (Acellera Ltd., London, United Kingdom) for his help and guidance during benchmarking of the KDEEP web server. We thank Dr. Daniel Fisher (The Institute of Molecular Genetics of Montpellier (IGMM), Montpellier, France) for helping us clarify the wrong binding kinetic data records found in the PDBbind database (for the PDB entries 4EOR and 4EOK). We also thank Deniz Dogan and Prof. Dr. Alexandre M.J.J. Bonvin for the critical reading of the manuscript.

■ ABBREVIATIONS

BSA, buried surface area (Å^2); MUT, mutant; PDB, the Protein Data Bank; RMSD, root-mean-square deviation; WT, wild-type

■ REFERENCES

(1) Manning, G.; Hunter, T. Eukaryotic Kinomes. *Handbook of Cell Signaling*; Elsevier, 2010; pp 393−397.

(2) Knight, J. D. R.; Qian, B.; Baker, D.; Kothary, R. Conservation, Variability and the Modeling of Active Protein Kinases. *PLoS One* **2007**, *2*, No. e982.

(3) Bossemeyer, D. Protein Kinases - Structure and Function. *FEBS Lett.* **1995**, *369*, 57−61.

(4) Kornev, A. P.; Taylor, S. S. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **2015**, *40*, 628−647.

(5) Weber, T. J.; Qian, W.; States, U. Protein Kinases. *Comprehensive Toxicology*, 3rd ed.; Elsevier, 2018; Vol. *8*, pp 264−285.

(6) Cruzalegui, F. Protein Kinases: From Targets to Anti-Cancer Drugs. *Ann. Pharm. Fr.* **2010**, *68*, 254−259.

(7) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109*, 275−282.

(8) Adams, J. A. Activation Loop Phosphorylation and Catalysis in Protein Kinases: Is There Functional Evidence for the Autoinhibitor Model? *Biochemistry* **2003**, *42*, 601−607.

(9) Martin, M. P.; Zhu, J.-Y.; Lawrence, H. R.; Pireddu, R.; Luo, Y.; Alam, R.; Ozcan, S.; Sebti, S. M.; Lawrence, N. J.; Schönbrunn, E. A Novel Mechanism by Which Small Molecule Inhibitors Induce the DFG Flip in Aurora A. *ACS Chem. Biol.* **2012**, *7*, 698−706.

(10) Vijayan, R. S. K.; He, P.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. Conformational Analysis of the DFG-Out Kinase Motif and Biochemical Profiling of Structurally Validated Type II Inhibitors. *J. Med. Chem.* **2015**, *58*, 466−479.
(11) Modi, V.; Dunbrack, R. L. Defining a New Nomenclature for the Structures of Active and Inactive Kinases. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 6818−6827.
(12) Izarzugaza, J. M.; Hopcroft, L. E.; Baresic, A.; Orengo, C. A.; Martin, A. C.; Valencia, A. Characterization of Pathogenic Germline Mutations in Human Protein Kinases. *BMC Bioinf.* **2011**, *12*, S1.
(13) Endicott, J. A.; Noble, M. E. M.; Johnson, L. N. The Structural Basis for Control of Eukaryotic Protein Kinases. *Annu. Rev. Biochem.* **2012**, *81*, 587−613.
(14) LoGrasso, P. V.; Frantz, B.; Rolando, A. M.; O'Keefe, S. J.; Hermes, J. D.; O'Neill, E. A. Kinetic Mechanism for P38 MAP Kinase. *Biochemistry* **1997**, *36*, 10422−10427.
(15) Maheshwari, S.; Miller, M. S.; O'Meally, R.; Cole, R. N.; Amzel, L. M.; Gabelli, S. B. Kinetic and Structural Analyses Reveal Residues in Phosphoinositide 3-Kinase α That Are Critical for Catalysis and Substrate Recognition. *J. Biol. Chem.* **2017**, *292*, 13541−13550.
(16) Weber, T. J.. Protein Kinases. *Comprehensive Toxicology*, 2nd ed.; Elsevier, 2010; Vol. *2*, pp 473−493.
(17) Adams, J. A. Kinetic and Catalytic Mechanisms of Protein Kinases. *Chem. Rev.* **2001**, *101*, 2271−2290.
(18) Szafranska, A. E.; Dalby, K. N. Kinetic Mechanism for P38 MAP Kinase Alpha. A Partial Rapid-Equilibrium Random-Order Ternary-Complex Mechanism for the Phosphorylation of a Protein Substrate. *FEBS J.* **2005**, *272*, 4631−4645.
(19) Zhou, T.; Commodore, L.; Huang, W.-S.; Wang, Y.; Sawyer, T. K.; Shakespeare, W. C.; Clackson, T.; Zhu, X.; Dalgarno, D. C. Structural Analysis of DFG-in and DFG-out Dual Src-Abl Inhibitors Sharing a Common Vinyl Purine Template. *Chem. Biol. Drug Des.* **2010**, *75*, 18−28.
(20) Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. 2.2 \AA Refined Crystal Structure of the Catalytic Subunit of CAMP-Dependent Protein Kinase Complexed with MnATP and a Peptide Inhibitor. *Acta Crystallogr., Sect. D: Struct. Biol.* **1993**, *49*, 362−365.
(21) Brognard, J.; Hunter, T. Protein Kinase Signaling Networks in Cancer. *Curr. Opin. Genet. Dev.* **2011**, *21*, 4−11.
(22) Mortlock, A.; Foote, K.; Kettle, J.; Aquila, B. Kinase Inhibitors in Cancer. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2014; p B9780124095472110000.
(23) Mortlock, A. A.; Wilson, D. M.; Kettle, J. G.; Goldberg, F. W.; Foote, K. M. Selective Kinase Inhibitors in Cancer. *Comprehensive Medicinal Chemistry III*; Elsevier, 2017; pp 39−75.
(24) Orlando, E.; Aebersold, D. M.; Medová, M.; Zimmer, Y. Oncogene Addiction as a Foundation of Targeted Cancer Therapy: The Paradigm of the MET Receptor Tyrosine Kinase. *Cancer Lett.* **2019**, *443*, 189−202.
(25) Daub, H.; Specht, K.; Ullrich, A. Strategies to Overcome Resistance to Targeted Protein Kinase Inhibitors. *Nat. Rev. Drug Discovery* **2004**, *3*, 1001−1010.
(26) Barouch-Bentov, R.; Sauer, K. Mechanisms of Drug Resistance in Kinases. *Expert Opin. Invest. Drugs* **2011**, *20*, 153−208.
(27) Bello, M.; Guadarrama-García, C.; Rodriguez-Fonseca, R. A. Dissecting the Molecular Recognition of Dual Lapatinib Derivatives for EGFR/HER2. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 293−303.
(28) Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflich, K. P. Targeting Cancer with Kinase Inhibitors. *J. Clin. Invest.* **2015**, *125*, 1780−1789.
(29) Torkamani, A.; Verkhivker, G.; Schork, N. J.N. J. Cancer Driver Mutations in Protein Kinase Genes. *Cancer Lett.* **2009**, *281*, 117−127.
(30) Carter, T. A.; Wodicka, L. M.; Shah, N. P.; Velasco, A. M.; Fabian, M. A.; Treiber, D. K.; Milanov, Z. V.; Atteridge, C. E.; Biggs, W. H.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Mehta, S. A.; Patel, H. K.; Pao, W.; Sawyers, C. L.; Varmus, H.; Zarrinkar, P. P.; Lockhart, D. J. Inhibition of Drug-Resistant Mutants of ABL, KIT, and EGF Receptor Kinases. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11011−11016.
(31) Bishop, A. C.; Buzko, O.; Shokat, K. M. Magic Bullets for Protein Kinases. *Trends Cell Biol.* **2001**, *11*, 167−172.
(32) Kissova, M.; Maga, G.; Crespan, E. The Human Tyrosine Kinase Kit and Its Gatekeeper Mutant T670I, Show Different Kinetic Properties: Implications for Drug Design. *Bioorg. Med. Chem.* **2016**, *24*, 4555−4562.
(33) Bishop, A. C.; Ubersax, J. A.; Petsch, D. T.; Matheos, D. P.; Gray, N. S.; Blethrow, J.; Shimizu, E.; Tsien, J. Z.; Schultz, P. G.; Rose, M. D.; Wood, J. L.; Morgan, D. O.; Shokat, K. M. A Chemical Switch for Inhibitor-Sensitive Alleles of Any Protein Kinase. *Nature* **2000**, *407*, 395−401.
(34) Jones, M. H.; Huneycutt, B. J.; Pearson, C. G.; Zhang, C.; Morgan, G.; Shokat, K.; Bloom, K.; Winey, M. Chemical Genetics Reveals a Role for Mps1 Kinase in Kinetochore Attachment during Mitosis. *Curr. Biol.* **2005**, *15*, 160−165.
(35) Bartkowiak, B.; Yan, C.; Greenleaf, A. L. Engineering an Analog-Sensitive CDK12 Cell Line Using CRISPR/Cas. *Biochim. Biophys. Acta, Gene Regul. Mech.* **2015**, *1849*, 1179−1187.
(36) Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L. Predicting Resistance of Clinical Abl Mutations to Targeted Kinase Inhibitors Using Alchemical Free-Energy Calculations. *Commun. Biol.* **2018**, *1*, 70.
(37) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Predicting Kinase Inhibitor Resistance: Physics-Based and Data-Driven Approaches. *ACS Cent. Sci.* **2019**, *5*, 1468−1474.
(38) Fan, H.; Schneidman-Duhovny, D.; Irwin, J. J.; Dong, G.; Shoichet, B. K.; Sali, A. Statistical Potential for Modeling and Ranking of Protein−Ligand Interactions. *J. Chem. Inf. Model.* **2011**, *51*, 3078−3092.
(39) Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G. KDEEP: Protein−Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. *J. Chem. Inf. Model.* **2018**, *58*, 287−296.
(40) Kurkcuoglu, Z.; Koukos, P. I.; Citro, N.; Trellet, M. E.; Rodrigues, J. P. G. L. M.; Moreira, I. S.; Roel-Touris, J.; Melquiond, A. S. J.; Geng, C.; Schaarschmidt, J.; Xue, L. C.; Vangone, A.; Bonvin, A. M. J. J. Performance of HADDOCK and a Simple Contact-Based Protein−Ligand Binding Affinity Predictor in the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 175−185.
(41) Neudert, G.; Klebe, G. D. S. X: A Knowledge-Based Scoring Function for the Assessment of Protein−Ligand Complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2731−2745.
(42) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720−725.
(43) Krissinel, E.; Henrick, K. Detection of Protein Assemblies in Crystals. In *Computational Life Science*; R. Berthold, M.; Glen, R. C.; Diederichs, K.; Kohlbacher, O.; Fischer, I., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 163−174.
(44) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372*, 774−797.
(45) Krissinel, E. Crystal Contacts as Nature's Docking Solutions. *J. Comput. Chem.* **2010**, *31*, 133−143.
(46) Haupt, V. J.; Daminelli, S.; Schroeder, M. Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key. *PLoS One* **2013**, *8*, No. e65894.
(47) Zhou, T.; Parillon, L.; Li, F.; Wang, Y.; Keats, J.; Lamore, S.; Xu, Q.; Shakespeare, W.; Dalgarno, D.; Zhu, X. Crystal Structure of the T315I Mutant of Abl Kinase. *Chem. Biol. Drug Des.* **2007**, *70*, 171−181.
(48) Pratt, D. J.; Bentley, J.; Jewsbury, P.; Boyle, F. T.; Endicott, J. A.; Noble, M. E. M. Dissecting the Determinants of Cyclin-Dependent Kinase 2 and Cyclin-Dependent Kinase 4 Inhibitor Selectivity. *J. Med. Chem.* **2006**, *49*, 5470−5477.

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

(49) Getlik, M.; Grütter, C.; Simard, J. R.; Klüter, S.; Rabiller, M.; Rode, H. B.; Robubi, A.; Rauh, D. Hybrid Compound Design To Overcome the Gatekeeper T338M Mutation in cSrc#. *J. Med. Chem.* **2009**, *52*, 3915−3926.
(50) Chan, W. W.; Wise, S. C.; Kaufman, M. D.; Ahn, Y. M.; Ensinger, C. L.; Haack, T.; Hood, M. M.; Jones, J.; Lord, J. W.; Lu, W. P.; Miller, D.; Patt, W. C.; Smith, B. D.; Petillo, P. A.; Rutkoski, T. J.; Telikepalli, H.; Vogeti, L.; Yao, T.; Chun, L.; Clark, R.; Evangelista, P.; Gavrilescu, L. C.; Lazarides, K.; Zaleskas, V. M.; Stewart, L. J.; Van Etten, R. A.; Flynn, D. L. Conformational Control Inhibition of the BCR-ABL1 Tyrosine Kinase, Including the Gatekeeper T315I Mutant, by the Switch-Control Inhibitor DCC-2036. *Cancer Cell* **2011**, *19*, 556−568.
(51) Sogabe, S.; Kawakita, Y.; Igaki, S.; Iwata, H.; Miki, H.; Cary, D. R.; Takagi, T.; Takagi, S.; Ohta, Y.; Ishikawa, T. Structure-Based Approach for the Discovery of Pyrrolo[3,2-d]Pyrimidine-Based EGFR T790M/L858R Mutant Inhibitors. *ACS Med. Chem. Lett.* **2013**, *4*, 201−205.
(52) Gurden, M. D.; Westwood, I. M.; Faisal, A.; Naud, S.; Cheung, K.-M. J.; McAndrew, C.; Wood, A.; Schmitt, J.; Boxall, K.; Mak, G.; Workman, P.; Burke, R.; Hoelder, S.; Blagg, J.; Van Montfort, R. L. M.; Linardopoulos, S. Naturally Occurring Mutations in the MPS1 Gene Predispose Cells to Kinase Inhibitor Drug Resistance. *Cancer Res.* **2015**, *75*, 3340.
(53) Breitenlechner, C. B.; Friebe, W.-G.; Brunet, E.; Werner, G.; Graul, K.; Thomas, U.; Künkele, K.-P.; Schäfer, W.; Gassel, M.; Bossemeyer, D.; Huber, R.; Engh, R. A.; Masjost, B. Design and Crystal Structures of Protein Kinase B-Selective Inhibitors in Complex with Protein Kinase A and Mutants. *J. Med. Chem.* **2005**, *48*, 163−170.
(54) Yun, C.-H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **2007**, *11*, 217−227.
(55) Gajiwala, K. S.; Wu, J. C.; Christensen, J.; Deshmukh, G. D.; Diehl, W.; DiNitto, J. P.; English, J. M.; Greig, M. J.; He, Y.-A.; Jacques, S. L.; Lunney, E. A.; McTigue, M.; Molina, D.; Quenzer, T.; Wells, P. A.; Yu, X.; Zhang, Y.; Zou, A.; Emmett, M. R.; Marshall, A. G.; Zhang, H.-M.; Demetri, G. D. KIT Kinase Mutants Show Unique Mechanisms of Drug Resistance to Imatinib and Sunitinib in Gastrointestinal Stromal Tumor Patients. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1542.
(56) Levinson, N. M.; Boxer, S. G. A Conserved Water-Mediated Hydrogen Bond Network Defines Bosutinib's Kinase Selectivity. *Nat. Chem. Biol.* **2014**, *10*, 127−132.
(57) Bunney, T. D.; Wan, S.; Thiyagarajan, N.; Sutto, L.; Williams, S. V.; Ashford, P.; Koss, H.; Knowles, M. A.; Gervasio, F. L.; Coveney, P. V.; Katan, M. The Effect of Mutations on Drug Sensitivity and Kinase Activity of Fibroblast Growth Factor Receptors: A Combined Experimental and Theoretical Study. *EBioMedicine* **2015**, *2*, 194−204.
(58) Johnson, T. W.; Richardson, P. F.; Bailey, S.; Brooun, A.; Burke, B. J.; Collins, M. R.; Cui, J. J.; Deal, J. G.; Deng, Y.-L.; Dinh, D.; Engstrom, L. D.; He, M.; Hoffman, J.; Hoffman, R. L.; Huang, Q.; Kania, R. S.; Kath, J. C.; Lam, H.; Lam, J. L.; Le, P. T.; Lingardo, L.; Liu, W.; McTigue, M.; Palmer, C. L.; Sach, N. W.; Smeal, T.; Smith, G. L.; Stewart, A. E.; Timofeevski, S.; Zhu, H.; Zhu, J.; Zou, H. Y.; Edwards, M. P. Discovery of (10R)-7-Amino-12-Fluoro-2,10,16-Trimethyl-15-Oxo-10,15,16,17-Tetrahydro-2H-8,4-(Metheno)-Pyrazolo[4,3-h][2,5,11]-Benzoxadiazacyclotetradecine-3-Carbonitrile (PF-06463922), a Macrocyclic Inhibitor of Anaplastic Lymphoma Kinase (ALK) and c-Ros Oncogene 1 (ROS1) with Preclinical Brain Exposure and Broad-Spectrum Potency against ALK-Resistant Mutations. *J. Med. Chem.* **2014**, *57*, 4720−4744.
(59) Echalier, A.; Cot, E.; Camasses, A.; Hodimont, E.; Hoh, F.; Jay, P.; Sheinerman, F.; Krasinska, L.; Fisher, D. An Integrated Chemical Biology Approach Provides Insight into Cdk2 Functional Redundancy and Inhibitor Sensitivity. *Chem. Biol.* **2012**, *19*, 1028−1040.
(60) Pemovska, T.; Johnson, E.; Kontro, M.; Repasky, G. A.; Chen, J.; Wells, P.; Cronin, C. N.; McTigue, M.; Kallioniemi, O.; Porkka, K.; Murray, B. W.; Wennerberg, K. Axitinib Effectively Inhibits BCR-ABL1(T315I) with a Distinct Binding Conformation. *Nature* **2015**, *519*, 102−105.
(61) Heald, R.; Bowman, K. K.; Bryan, M. C.; Burdick, D.; Chan, B.; Chan, E.; Chen, Y.; Clausen, S.; Dominguez-Fernandez, B.; Eigenbrot, C.; Elliott, R.; Hanan, E. J.; Jackson, P.; Knight, J.; La, H.; Lainchbury, M.; Malek, S.; Mann, S.; Merchant, M.; Mortara, K.; Purkey, H.; Schaefer, G.; Schmidt, S.; Seward, E.; Sideris, S.; Shao, L.; Wang, S.; Yeap, K.; Yen, I.; Yu, C.; Heffron, T. P. Noncovalent Mutant Selective Epidermal Growth Factor Receptor Inhibitors: A Lead Optimization Case Study. *J. Med. Chem.* **2015**, *58*, 8877−8895.
(62) Bryan, M. C.; Burdick, D. J.; Chan, B. K.; Chen, Y.; Clausen, S.; Dotson, J.; Eigenbrot, C.; Elliott, R.; Hanan, E. J.; Heald, R.; Jackson, P.; La, H.; Lainchbury, M.; Malek, S.; Mann, S. E.; Purkey, H. E.; Schaefer, G.; Schmidt, S.; Seward, E.; Sideris, S.; Wang, S.; Yen, I.; Yu, C.; Heffron, T. P. Pyridones as Highly Selective, Noncovalent Inhibitors of T790M Double Mutants of EGFR. *ACS Med. Chem. Lett.* **2016**, *7*, 100−104.
(63) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Cent. Sci.* **2018**, *4*, 1708−1718.
(64) Rezaei, M.; Li, Y.; Li, X.; Li, C. Improving the Accuracy of Protein-Ligand Binding Affinity Prediction by Deep Learning Models: Benchmark and Model; ChemRxiv; **2019**. DOI: 10.26434/chemrxiv.9866912.v1.
(65) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105−132.
(66) Zamyatnin, A. A. Protein Volume in Solution. *Prog. Biophys. Mol. Biol.* **1972**, *24*, 107−123.
(67) Koukos, P. I.; Xue, L. C.; Bonvin, A. M. J. J. Protein−Ligand Pose and Affinity Prediction: Lessons from D3R Grand Challenge 3. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 83−91.
(68) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(69) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on a Diverse Test Set. *J. Chem. Inf. Model.* **2009**, *49*, 1079−1093.
(70) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.
(71) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(72) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045−D1053.
(73) Ahmed, A.; Smith, R. D.; Clark, J. J.; Dunbar, J. B.; Carlson, H. A. Recent Improvements to Binding MOAD: A Resource for Protein−Ligand Binding Affinities and Structures. *Nucleic Acids Res.* **2015**, *43*, D465−D469.
(74) Smith, R. D.; Clark, J. J.; Ahmed, A.; Orban, Z. J.; Dunbar, J. B.; Carlson, H. A. Updates to Binding MOAD (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *J. Mol. Biol.* **2019**, *431*, 2423−2433.
(75) Chen, J. H.; Linstead, E.; Swamidass, S. J.; Wang, D.; Baldi, P. ChemDB Update Full-Text Search and Virtual Chemical Space. *Bioinformatics* **2007**, *23*, 2348−2351.
(76) Tetko, I. V.; Poda, G. I. Application of ALOGPS 2.1 to Predict Log D Distribution Coefficient for Pfizer Proprietary Compounds. *J. Med. Chem.* **2004**, *47*, 5601−5604.
(77) Tetko, I. V.; Bruneau, P. Application of ALOGPS to Predict 1-octanol/Water Distribution Coefficients, LogP, and LogD, of AstraZeneca In-house Database. *J. Pharm. Sci.* **2004**, *93*, 3103−3110.

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265

(78) Tetko, I. V.; Tanchuk, V. Y. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1136−1145.
(79) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.
(80) Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488−1493.
(81) Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. P. Prediction of N-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407−1421.
(82) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127−130.
(83) Poda, G. I.; Tetko, I. V.; Rohrer, D. C. Towards Predictive ADME Profiling of Drug Candidates: Lipophilicity and Solubility. *229th National Meeting & Expo*; American Chemical Society: San Diego, CA, 2005, p MEDI 514 article.
(84) Tetko, I. V. Computing Chemistry on the Web. *Drug Discovery Today* **2005**, *10*, 1497−1500.
(85) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual Computational Chemistry Laboratory − Design and Description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453−463.
(86) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58*, 4066−4072.
(87) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5*, 725−738.
(88) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12*, 7−8.
(89) Yang, J.; Zhang, Y. I-TASSER Server: New Development for Protein Structure and Function Predictions. *Nucleic Acids Res.* **2015**, *43*, W174−W181.
(90) Schrödinger, LLC. *The PyMOL Molecular Graphics System*, version 1.8, 2015.
(91) Vangone, A.; Rodrigues, J. P. G. L. M.; Xue, L. C.; van Zundert, G. C. P.; Geng, C.; Kurkcuoglu, Z.; Nellen, M.; Narasimhan, S.; Karaca, E.; van Dijk, M.; Melquiond, A. S. J.; Visscher, K. M.; Trellet, M.; Kastritis, P. L.; Bonvin, A. M. J. J. Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. *Proteins: Struct., Funct., Bioinf.* **2017**, *85*, 417−423.
(92) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
(93) RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, Inc.: Boston, MA, 2015.
(94) Kassambara, A. *Ggpubr: "ggplot2" Based Publication Ready Plots*, 2019.
(95) Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag: New York, 2016.
(96) Bache, S. M.; Wickham, H. *Magrittr: A Forward-Pipe Operator for R*, 2014.
(97) Slowikowski, K. *Ggrepel: Automatically Position Non-Overlapping Text Labels with "Ggplot2"*, 2019.
(98) Auguie, B. *GridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017.
(99) Kolde, R. *Pheatmap*; Pretty Heatmaps, 2019.
(100) Neuwirth, E. *RColorBrewer: ColorBrewer Palettes*, 2014.

https://dx.doi.org/10.1021/acsomega.0c04672
ACS Omega 2021, 6, 1254−1265