Improved pose and affinity predictions using different protocols tailored on the basis of data availability

Philip Prathipati*, Chioko Nagao, Shandar Ahmad, Kenji Mizuguchi

National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki City, Osaka- 567-0085

*Corresponding author name and address:
Philip Prathipati
National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan
Email address: philip@nibiohn.go.jp

Keywords

Data-optimal protocol, kinetic stability, thermodynamic stability, docking, scoring, QSAR, GLMNET

Abstract

The D3R 2015 grand drug design challenge provided a set of blinded challenges for evaluating the applicability of our protocols for pose and affinity prediction. In the present study, we report the application of two different strategies for the two protein targets HSP90 and MAP4K4. HSP90 is a well-studied target system with numerous co-crystal structures and SAR data. Furthermore the D3R HSP90 test compounds showed high structural similarity to existing HSP90 inhibitors in BindingDB. Thus, we adopted an integrated docking and scoring approach involving combination of both pharmacophoric and heavy atom similarity alignments, local minimization and quantitative structure activity relationships modeling, resulting in the reasonable prediction of pose (with the root mean square deviation [RMSD] values of 1.75Å for mean pose 1, 1.417Å for the mean best pose and 1.85Å for the mean all poses) and affinity (ROC AUC= 0.702 at 7.5 pIC50 cut-off and R = 0.45 for 180 compounds). The second protein, MAP4K4, represents a novel system with limited SAR and co-crystal structure data and little structural similarity of the D3R MAP4K4 test compounds to known ligands. For this system, we implemented an exhaustive pose and affinity prediction protocol involving docking and scoring using the PLANTS software which considers side chain flexibility together with protein-ligand fingerprints analysis assisting in pose prioritization. This protocol through fares poorly in pose prediction (with the RMSD values of 4.346Å for mean pose 1, 4.69Å for mean best pose and 4.75Å for mean all poses) and produced reasonable affinity prediction (AUC= 0.728 at 7.5 pIC50 cut-off and R = 0.67 for 18 compounds, ranked 1st among 80 submissions).

Introduction

Computer-aided drug design approaches for pose and affinity predictions have traditionally been classified into ligand- and protein structure-based approaches.[1-3] The former is generally applicable when a sufficient number of protein-ligand co-crystal structures and structure-activity relationships (SAR) data (in the form of k_i, k_d or IC_{50}) are available [1,2] and primarily uses quantitative structure activity relationships (QSAR) and pharmacophore models for the prediction of affinity and pose, respectively [1,2]. Whereas

the latter is preferred in their absence or scarcity [1,2] and uses docking and scoring protocols for pose and affinity predictions respectively [1,2]. Due to their heavy reliance on statistical measures, identifying a QSAR or pharmacophore model that reflects the biological reality of the molecular recognition event in question from a pool of models is difficult.[4] Furthermore in view of their ability to better capture the features relevant for explaining the training set data, they often present high predictive performance for internal test set and modest predictive abilities for an external test set depending on the applicability domain of the model. Protein structure-based approaches on the other hand can perform *de novo* pose and affinity predictions[4] and have greater applicability domain and presents reasonable quantitative or qualitative predictive ability. However, this approach continues to suffer from the problems related to scoring functions and pose prioritization[4].

Pharmacophore models based on bioactive conformations derived from protein-ligand co-crystal structures [5-7]could mitigate some of the problems associated with ligand-based approaches for pose prediction. On the other hand, protein structure-based approaches [7] could be enhanced by introducing a protocol for pose prioritization assisted by protein ligand fingerprint analysis of the predicted poses. Such a procedure has been shown to greatly benefit the *de novo* pose and affinity prediction.[8-12]. The D3R grand challenge offered a great opportunity to evaluate these integrated protocols in an unbiased manner through expert evaluation on previously unseen data.

The D3R grand drug design challenge for pose and affinity prediction consisted of two protein targets HSP90 and MAP4K4. The assessments consisted of pose predictions for six HSP90 ligands and 30 MAP4K4 ligands, together with affinity predictions for 180 HSP90 and 18 MA4K4 ligands. While HSP90 [13-16] is a well-known protein target with numerous inhibitors reported in the literature with k_i, k_d and IC_{50} data, MAP4K4[17,18] is a novel target with few co-crystal structures and very few compounds with k_i, k_d or IC_{50} data. Hence for the D3R challenge, we applied two different strategies for pose and affinity predictions of HSP90 and MAP4K4 ligands, using the integrated protocols presented above. The pose prediction protocol for HSP90 inhibitors consisted of a ligand-based pharmacophore alignment approach,

followed by local docking. Predictive QSAR models derived using a large set of HSP90 SAR data, was used for affinity predictions. For MAP4K4, we used a *de novo* docking protocol augmented with protein ligand fingerprint analysis for pose prediction and the free energy measures of these poses were used for affinity predictions.

Materials and methods

HSP90 pose prediction protocol 1 to 4: The ligands and proteins were prepared (converted to 3D coordinates, protonation, lowest energy conformer generation, etc) according to default settings of the software programs MarvinView 16.2.8.0 (2015, ChemAxon, http://www.chemaxon.com), Discovery Studio (Version 3.5.0). and PyRx [Python Prescription 0.8]. The 180 HSP90 inhibitors provided by the D3R organizers together with 15,926 HSP90 subtype (we considered all the five functional human genes encoding Hsp90 protein isoforms, namely HSP90AA1, HSP90AA2, HSP90AB1, HSP90B1 and TRAP1) ligands obtained from bindingDB website, were clustered using WGCNA[4,19] (with parametric settings: "signed" and "module cut-off as 0.45") based on the computed ECFP_12 fingerprint[20,21], which resulted in ~40 clusters (supplementary material 1, data presented in column "moduleColorsAutomatic_ net_12_bicor_t_signed_0.45"). These WGCNA-generated clusters were further grouped manually into three clusters using visual inspection. The HSP90 IDs for the three clusters were as follows: cluster-1: hsp90_1 to hsp90_61, represented by the co-crystal structure of the Protein Data Bank (PDB) entry 4YKR, cluster-2: hsp90_63 to hsp_90_120, represented by the co-crystal structure of 4XDX, cluster-3: hsp90_121 to hsp90_179, represented by the co-crystal structure of 4KYK.

The three co-crystal structure ligands of 4YKR, 4XDX, 4YKY were taken as templates for the three clusters and MIXED (ATOM and pharmacophore) alignments were performed using Open3Dalign.[22] This method represents the pharmacophore aligned docking approach. The docked poses were further minimized and scored within the active site of 4YKR, 4XDX, 4YKY and 4TOZ using SMINA[23] and presented as docking poses numbered 1-4, respectively.

HSP90 scoring protocols 1 to 4: The SAR data in terms of 'k_i' and 'IC_{50}' values of 15,926 ligands for Hsp90 homologs (HSP90AA1, HSP90AA2, HSP90AA2, HSP90B1 and TRAP1) were extracted from BindingDB. The 'IC_{50}' values were converted to Ki values using the variant of Cheng-Prusoff equation.[4] ECFP12 fingerprints were computed for the 15,926 ligands and cross-validated machine-learning models for 1262 training set compounds were derived using Lasso[24], Elastic Net, [24] Ridge regression[24] and random forest. These models regressed 9-log10K_i values (y) on ECFP12 fingerprints (x) (using the jchemmapper[25] package). The predicted k_i values from the models presented reasonable agreement with the experimental ki values both for the training and test sets as presented in results and discussion section. It may be noted that Lasso and Elastic Net perform both parameter shrinkage and variable selection automatically, while Ridge regression and random forest use a different heuristic for dimensionality reduction (without *a priori* feature selection). Since different techniques account for different limitations of modeling high-dimensional data,[26] we decided to evaluate the performance of all the four methods in the present study. Ridge regression and random forest further differ in the polynomials that they model, to account for non-linear dependencies. Elastic Net further differs from Lasso by penalizing correlation among the features. Thus, our scoring scheme incorporates predictions from Lasso, Elastic Net, Ridge regression and Random forest for HSP90 QSAR models (supplementary material 1, data presented in columns "ridge_predicted", "enet_predicted", "lasso_predicted" and "RF_predicted").

HSP90 pharmacophore and local docking protocol for free energy prediction: The HSP90 protocol for pose prediction used a ligand-based MIXED (ATOM and pharmacophore) alignment protocol. However, scoring the aligned pose without optimization may lead to steric clashes and non-optimal orientations of ligand hydrogens with respect to the protein. Thus for measuring the free energies of the binding poses determined by Open3Dalign[22], we decided to use an empirical scoring function, SMINA[23] which optimizes the ligands' co-ordinates (along with hydrogens) within the active site of the following protein crystal structures '4YKR', '4XDX', '4YKY, '3T0Z' and '2JJC' (supplementary material 1, data presented in rows pred_4yky, pred_4YKR, pred_2XDX, pred_2JJC and pred_3TOZ). Although SMINA is

not a force-field based scoring function, it indirectly maps to physical forces such as electrostatic interactions, and are parameterized to reproduce binding affinities or other data.

MAP4K4 pose prediction protocol: The *de novo* pose prediction protocol for MAP4K4 primarily employed the enhanced protocol involving PLANTS1.2[27] for docking and PyPLIF[9] for pose prioritization. Essentially, SPORES[27] was used to split the MAP4K4 PDB file '4U44' into the protein and ligands using the 'splitpdb' module. The protein was recognized, protonated and stored as protein.mol2, while the reference ligand was also recognized, protonated and stored as ligand.mol2. The docking was performed using PLANTS1.2[27] The bind module of PLANTS1.2 was used to identify the binding site (defined as non-hydrogen atoms within 5 Å of any non-hydrogen atom of the reference ligand). The PLANTS1.2 docking procedure produced 50 docking poses for each of the 30 MAP4K4 compounds. Since protein-ligand complexes can be either a thermodynamically or a kinetically stable entity, we prioritized the best binding poses using the following three methods instead of the default free energy approach alone used in many docking protocols. The first method prioritized the best poses using the lowest free energy value. The second used a hybrid score (geometric mean) of the free energy and the Tanimoto similarity to the reference ligand (ligand code 3D9 of the PDB entry 4U44, the most potent MAP4K4 inhibitor) based on the protein-ligand interaction fingerprints of selected evolutionarily conserved and functionally relevant residues implicated in ligand-mediated kinase inactivation (see Results and Discussion for the selected residues). The third protocol was similar to the second, except that the Tanimoto similarity is computed using all the protein-ligand interaction fingerprints. The second and third protocols involved submitting the docking output from PLANTS as input to PyPLIF[9] for the calculation of the protein ligand fingerprints (with selected and all the residues of the binding site) and the Tanimoto similarity of the docked poses with respect to the reference pose(the co-crystal structure of '4U44'.

MAP4K4 scoring protocol: The PLANTS CHEMPLP scores (the default scoring function used in PLANTS) of each of the 18 ligands were based on the poses prioritized using the three methods described above (Supplementary material 4, data presented in columns PLANTS global minimum, PLANTS +

PyPLIF _subset, PLANTS + PyPLIF _all). The free energy scores obtained from SMINA scoring function served as the affinity predictions for stage-2 of the MAP4K4 contest (Supplementary material 4 , data presented in column SMINA_free_energies).

Performance evaluation metrics in terms of RMSD: Detailed procedures for evaluating final pose predictions can be accessed from the D3R website. Unfortunately, some of our predicted poses did not conform to the format requirements of D3R, leading to a lack of cumulative evaluation for all the values (as D3R required error-free computation of RMSDs in *all* poses). To complete the picture for our entire set of predictions, we extended these evaluations by performing RMSD calculations ourselves using Discovery studio visualizer 4.0 which uses a heavy atom based RMSD computation. The three kinds of mean RMSD values ('mean pose 1', 'mean best pose' and 'mean all poses') were computed to enable comparison with other contestants for which the RMSD values were provided by D3R. During the course of our communication we realized that D3R organizers RMSD computation was based on maximum common substructure functionality of the OEChem Python toolkit.

Results and discussion

The D3R grand drug design challenge consisted of two targets HSP90 and MAP4K4. HSP90 is a well-known target, with several SAR data-points and protein-ligand complexes, hence we attempted to verify the utility of ligand- and protein-based 3D pharmacophoric alignments, followed by local docking for pose and free energy prediction. We further employed machine learning-based QSAR models for affinity prediction. The pose and affinity prediction protocol for the MAP4K4 target, was performed using three different *de novo* docking protocols. (See Materials and Methods). Two of three pose prediction protocols involved using protein-ligand interaction fingerprints computed using the PyPLIF software. The free energy values corresponding to the poses selected by the three protocols served as the affinity values. The pose prediction evaluations in the form of RMSDs are presented in table 1 for the two targets and for the 3 types of poses. Table 2 presents the affinity prediction evaluations in the form of Pearson and Kendal tau and

Spearman correlation are presented for the 3 affinity prediction protocols of MAP4K4 and the 4 affinity prediction protocols of HSP90. Figure 1 depicts the comparison of the performance vis-à-vis other participants protocols of our pose and affinity prediction protocols for D3R HSP90 and MAP4K4 datasets. Figure 2 depicts quantitative (scatterplots) and discriminative ability (ROC curve) of the random forest QSAR model against the HSP90 training and test sets. Figure 3 depicts the quantitative (scatterplots) and discriminative ability (ROC curve) of our best affinity prediction protocols for the D3R's 180 HSP90 and 18 MAP4K4 ligands. Figure 4 illustrates the classification ability of all our HSP90 and MP4K4 affinity prediction protocols used for the D3R challenge, using ROC curves.

Table 1: Overview of our Pose prediction performance for the D3R grand challenge vis-à-vis all other submissions, presented in parenthesis.

Performance measures	mean RMSD of pose 1	mean RMSD of best pose	mean RMSD of all poses	# ligands
MAP4K4 (mean, median of all groups submissions)	4.346(5.21,5.12)	4.753 (4.20,4.17)	4.69 (5.34,5.37)	30
HSP90 (mean, median of all groups submissions)	1.751 (3.01,3.14)	1.417 (2.27,1.81)	1.854 (3.41,4.12)	6

Table 2: Correlation between experimental activities and the predicted affinities for the D3R grand challenge vis-à-vis all submissions presented in parenthesis.

Performance measures	R Pearson	R Kendal tau	R spearman	# ligands
MAP4K4, score1 (mean all, median all)	0.672 (0.22,0.28)	0.516 (0.15,0.19)	0.620	18
MAP4K4, score2 (mean all, median all)	0.469 (0.22,0.28)	0.326 (0.15,0.19)	0.400	18
MAP4K4, score3 (mean all, median all)	0.462 (0.22,0.28)	0.326 (0.15,0.19)	0.370	18
HSP90, score 1 (mean all, median all)	0.451 (0.22,0.24)	0.312 (0.15,0.16)	0.420	180
HSP90, score 2 (mean all, median all)	0.405 (0.22,0.24)	0.28 (0.15,0.16)	0.370	180
HSP90, score 3 (mean all, median all)	0.389 (0.22,0.24)	0.255 (0.15,0.16)	0.380	180
HSP90, score 4 (mean all, median all)	0.381 (0.22,0.24)	0.252 (0.15,0.16)	0.390	180

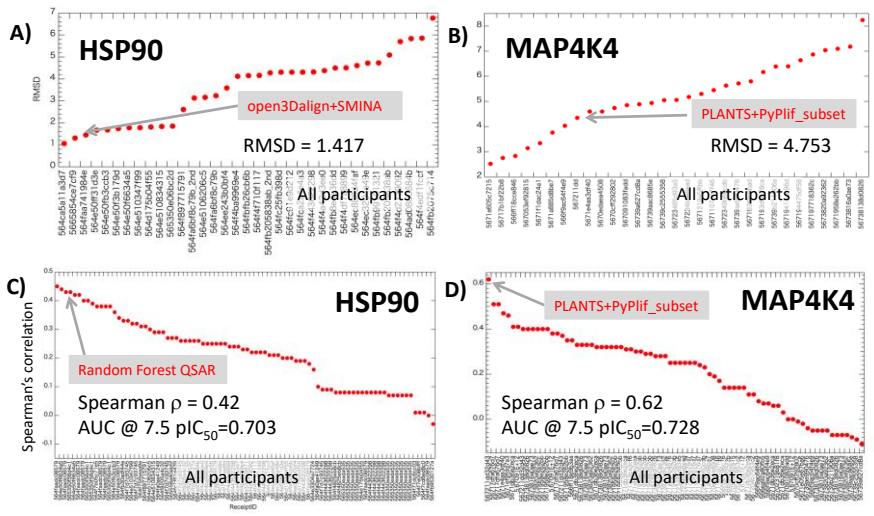

Figure 1: Comparison of the performance of our pose prediction protocols vis-à-vis other participants in D3R; (A) Open3Dalign and SMINA assisted HSP90 mean pose prediction RMSD of 1.417 is among the top ranked,(B) PLANTS docking augmented with PyPLIF analysis presents MAP4K4 mean pose prediction RMSD 0f 4.753 is among the median ranked, (C) with a spearman ρ of 0.42 and ROC AUC of 0.703 the random forest QSAR ranked highly and (D) With a spearman ρ of 0.62 and ROC AUC 0f 0.728 PLANTS+PyPlif protocol is among the top ranked.

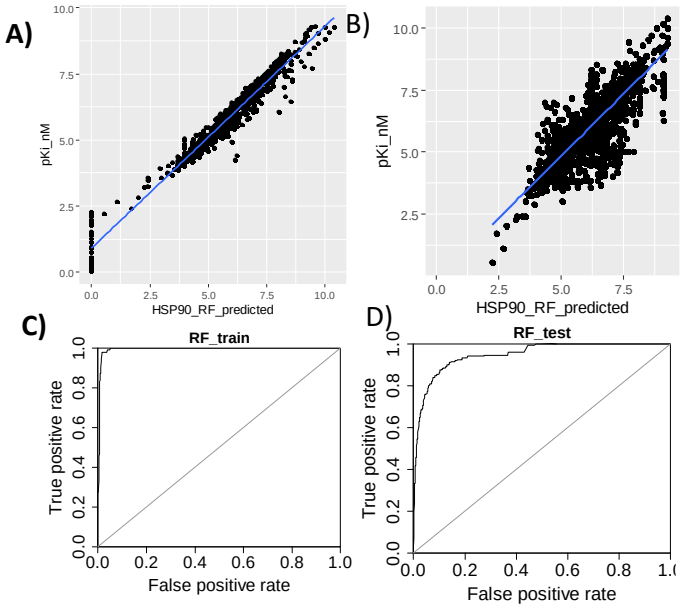

Figure 2: **A)** Scatterplot of the correlation between the random forest QSAR model predicted HSP90 activity scores ($k_{i\ p}$nM values) in the x-axis plotted versus the observed HSP90 values ($k_{i\ p}$nM values) in the y-axis for the training set (n=1262, R2= 0.96), **B)** ROC curve (AUC= 0.995) for the 1262 HPS90 training set predictions based on the random forest QSAR models. The 1262 HSP90 training set compounds were split into two classes using their experimentally pIC50 at a cuf-off of 7.5, **C)** Scatterplot of the correlation between the random forest QSAR model predicted HSP90 activity scores ($k_{i\ p}$nM values) in the x-axis plotted versus the observed HSP90 values ($k_{i\ p}$nM values) in the y-axis for the test set (n=14664,

R2= 0.79) and **D)** ROC curve (AUC= 0.949) for the 14664 HPS90 training set predictions based on the random forest QSAR models. The 1262 HSP90 training set compounds were split into two classes using their experimentally pIC50 at a cuf-off of 7.5.

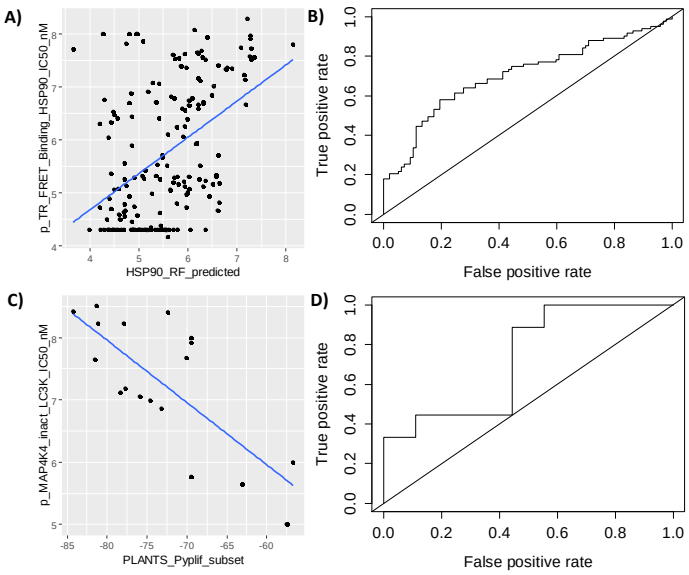

Figure 3: **A)** HSP90 predicted activity scores ($k_{i\ p}$nM values) in the x-axis plotted versus the observed HSP90 values (pIC50 nm values) in the y-axis for D3R HSP90 test set (n=180, R= 0.45), **B)** ROC curve (AUC=0.702) for the 180 HPS90 predictions based on the random forest QSAR models. The 180 HSP90 dataset was split into two classes using their experimentally pIC50 at a cuf-off of 7.5, **C)** MAP4K4 predicted activity scores (PLANTS docking score values)in the x-axis plotted versus the observed MAP4K4 values (pIC50 nm values) in the y-axis for D3R MAP4K4 test set (n=18, R= 0.67), **D)** ROC curve (AUC=0.728) for the 18 MAP4K4 predictions based on the random forest QSAR models. The 18 MAP4K4 compound set was split into two classes using their experimentally pIC50 at a cuf-off of 7.5.

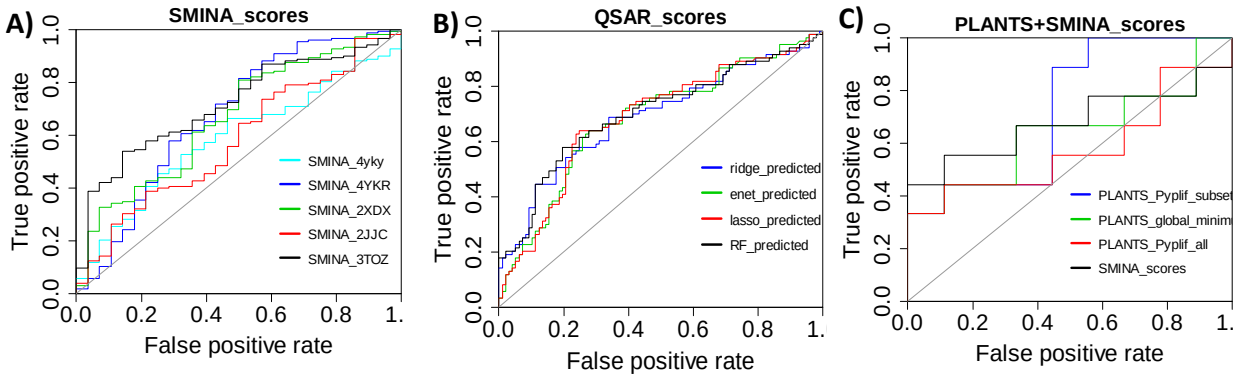

Figure 4: **A)** Receiver operating characteristic (ROC) curves for the D3R HSP90 cross-docking binding affinity predictions against 4yky, 4YKR, 2XDX, 2JJC and 3TOZ with AUCs estimated to be 0.587, 0.675, 0.669, 0.581 and 0.718 respectively. The 180 HSP90 dataset was split into two class using their experimentally pIC50 at a cuf-off of 7.5 The open3Dalignment based docking approach also involved using

SMINA for local docking, minimization and score estimation, **B)** ROC curves for the D3R HSP90 affinity predictions obtained from QSAR models derived from Ridge regression, Elastic Net, Lasso and Random Forest with AUCs estimated to be 0.691, 0.684, 0.687 and 0.702 respectively. The 180 HSP90 dataset was split into two classes using their experimentally pIC50 at a cuf-off of 7.5 and **C)** ROC curves for D3R MAP4K4 affinity predictions obtained using protocols PLANTS+PyPLIF _sub, PLANTS global minimum, PLANTS+PyPLIF _all and SMINA demonstrated AUCs of 0.729, 0.642, 0.580 and 0.679. The 18 D3R MAP4K4 dataset was split into two classes using their experimentally pIC50 at a cuf-off of 7.5.

Utility of ligand pharmacophore based 3D alignments for placing ligands within the active site of proteins (HSP90 pose prediction):

For proteins with crystal structure complexes, ligand alignment-based approaches can greatly aid in the placement of new ligands within the active site of a protein. While a traditional docking approach employs an electrostatic and/or geometric complementarity of the active site of a protein, the ligand alignment-based approach uses either 3D shape similarity, 3D electrostatic similarity, pharmacophoric similarity or heavy atom similarity or a combination of both pharmacophoric and heavy atom similarity. In the present study, we used the MIXED (atom and pharmacophore) alignment protocol using Open3Dalign.[22]. To avoid steric clashes and non-optimal orientations of ligand hydrogens with respect to the protein, we further minimized the binding poses determined by Open3Dalign through a local docking protocol provided in SMINA. This application optimizes the ligand co-ordinates (including hydrogens) within the active site of the protein and further scores the interaction between the minimized ligand and the protein crystal structure (in our case PDB IDs ‘4YKR’, ‘4XDX’, ‘4YKY, ‘3T0Z’ and ‘2JJC’). We submitted the minimized docking poses against ‘4YKR’, ‘4XDX’, ‘4YKY’ and ‘3TOZ’ for the root mean square deviation (RMSD) evaluation. Overall, RMSD values of 1.75Å (mean pose 1), 1.417Å (mean best pose), and 1.85Å (mean all poses) demonstrate the utility of the MIXED ligand alignment for pose prediction, since the average ligand poses were within 2Å RMSD of the crystal poses (Table 1 and Supplementary material 2) The low RMSD values could also be due to the high structural similarity of the HSP90 test compounds to the template co-crystal structures (‘4YKR’, ‘4XDX’, ‘4YKY’). Another group who used similar protocol in the D3R contest presented similar results albeit using a different type of analysis.

Utility of SMINA scoring function for explaining the variation in activities: Empirical scoring functions, incorporating elements of both force-field based and knowledge-based scoring schemes, are typically trained using existing protein-ligand crystal structure.[23] The inability of empirical scoring functions to explain the variation in affinity was widely reported in the literature[28]. Several reasons, ranging from limitations in our understanding and quantification of the molecular recognition process to the need for customized scoring functions for specific protein families, have been implicated. In the present study, free energy scores for both HSP90 (correlation coefficient R = 0.20) and MAP4K (correlation coefficient R = 0.34) obtained with SMINA scoring were less than 0.40, suggesting a requirement for an improved or customized scoring function to explain the variation in activities more accurately. Previous studies have reported better correlation using customized scoring functions [23] and are best suited for classifying actives from inactives. The reasonable classification ability of SMINA scoring function for the 180 D3R HSP90 inhibitory data is depicted in figure 4A. The modest classification ability of SMINA and PLANTS scoring functions for D3R's MAP4K4 datasets is illustrated in figures 4C. These results are consistent with other participants' results[29],[30] in D3R exercise.

Machine learning-based QSAR models to explain the variation in HSP90 activities: Docking-based scoring functions generally do not capture subtle changes in structure, which lead to a large variation in activities (often termed activity cliffs).[23,28,31] Ligand-based approaches are better tuned to handle activity cliffs, since they are often trained to capture molecular features correlated with such a variation. In the present study, ECFP12 fingerprints were computed for the 15,926 HSP90 ligands and cross-validated machine-learning models were derived using Lasso, Elastic Net, Ridge and RF (see Materials and Methods). The predicted k_i values from these models presented reasonable agreement with the experimental k_i values (mean R2 for the training set with Lasso (0.84), Enet(0.80), Ridge(0.82) and RF(0.96) models and R2 for the test set with Lasso(0.66), Enet(0.67), Ridge(0.68) and RF(0.79) models) (Supplementary material 1, data presented in columns ridge_predicted, enet_predicted, lasso_predicted and RF_predicted). We further evaluated the performance of each of the four methods individually using Pearson correlation co-efficient,

Kendall-tau and Spearman correlation coefficients (table 2). Overall, Ridge regression and RF produced better Pearson correlation values (R=0.43, ranked 3rd and R=0.45, ranked 1st, respectively) than elastic net (R=0.38) and Lasso (R=0.38). All these scores were modestly in the D3R evaluation, however in view of ability to explain the variance of ~20% only, additional methodological improvements are required for affinity predictions. As depicted in figure 4B, these models however demonstrate moderate discriminative ability (ranging from 0.684 to 0.701) against the D3R's HSP90 dataset.

Docking protocol combining PLANTS and PyPLIF fingerprints for MAP4K4 pose and affinity predictions: The MAP4K4 target system of the D3R contest presents an interesting problem for pose and affinity prediction. The 30 MAP4K4 test ligands are structurally diverse, with little structural similarity to existing co-crystal structures of this protein. Furthermore, there are very few compounds (~10) with k_i, k_d or IC_{50} data in the bindingDB[32] database. Given this background, for the MAP4K4 target we decided to apply a *de novo* docking protocol using the PLANTS docking software. The PLANTS software (Protein-Ligand ANT System)[27] is based on a class of stochastic optimization algorithms called ant colony optimization (ACO); it considers full ligand flexibility and flexible protein side-chains. The PLANTS suite also includes the pre-processing of the ligands and the protein performed by the program SPORES (Structure PrOtonation and REcognition System).

Given the *de novo* nature of docking studies, pose prioritization protocol is a critically important aspect of pose and affinity prediction. Several approaches have been suggested in the literature for this purpose, including the selection of the pose with the lowest energy or the lowest RMSD to the reference ligand. However, from protein folding and other thermodynamic studies[33], it is widely recognized that the global minimum is not always the preferred solution in biological systems. Given the limited time scale for a biological process, a kinetically stable solution (one of the local minima) is most often the preferred solution over a thermodynamic solution (global minimum). This concept has also been corroborated in several articles evaluating the redocking performance of existing docking software, where the pose with the least RMSD does not often present the lowest interaction energy [34-37]. Given this background, pose

prioritization protocols that can take advantage of known ligand-bound structures are of special interest. In this regard, structure interaction fingerprints (SIFt) were introduced with an objective to represent and analyze 3D protein-ligand interactions by encoding them by a one-dimensional binary string. The construction of SIFt is a two-step process consisting of (1) identification of residues interacting with the ligand and (2) classification of ligand-residue interactions into any of the predetermined types (e.g., hydrogen bond accepter/donor pairs and pi-interacting partners). Among several available methods, PyPLIF is the only available open source software for the calculation of these protein-ligand fingerprints.

In view of the above, the MAP4K4 pose prioritization was performed using three different protocols, combining the free energy value and the Tanimoto similarity to the reference ligand based on protein-ligand interaction fingerprints (see Materials and Methods). Using this approach, the average RMSD values were found to be unacceptable (mean pose 1 = 4.346Å, mean best pose = 4.69Å and mean all poses = 4.75Å) and ranked 8th in the final evaluation (Table 1 and Supplementary material 3). However, the affinity predictions afforded good R values (pose prioritization protocol 1 = 0.67 [ranked 1], pose prioritization protocol 2 = 0.47 [ranked 2] and pose prioritization protocol 3 = 0.46 [ranked 3]] (Table 2 and Supplementary material 4).

Affinity prediction of the poses prioritized using a composite score comprising both fingerprint similarity and free energy performs better than poses prioritized using free energy alone.

Protein-ligand interaction fingerprints are commonly computed using all the residues within the active site (generally within 5A of the reference ligand) of a protein.[9] However, from site-directed mutagenesis, residue conservation analysis and other residue level functional analysis often highlight the preferential role of a subset of residues, which are involved in the ligand-induced modulation of protein function. A kinase inhibitor mediated functional modulation subset typically consists of five residues, which includes two hinge-binding residues, a gate keeper residue and two catalytic site residues (the catalytic lysine and aspartic acid from the DFG motif).

Thus for the MAP4K4 pose and affinity prediction challenge, we assessed the performance of the composite score computed using two different approaches, against the pose prioritized using free energy alone. The first pose prioritization approach used a composite score comprising the geometric mean of the Tanimoto coefficient and the free energy value with the protein-ligand interaction fingerprints of five selected residues only. Poses prioritized by this method achieved an excellent Pearson correlation coefficient (R = 0.67 and ranked 1st among all the contestants). In contrast, the second approach, in which all the residues of the active site (within 5Å of the ligand) were considered for computing the protein-ligand interaction fingerprint, afforded a sub-optimal but still modest performance of the affinity prediction (with an R value of 0.47 and ranked 2nd). Notably, the free energy based approach without these two steps afforded an R value of 0.46 (ranked 3rd).

The local optimization of the pharmacophore aligned compounds and the inability to consider bridging water molecules affects HSP90 pose prediction results.

The local docking approach using the SMINA software optimizes, the docked conformation obtained from 'open3Dalign' alignments, within the active site of various HSP90 protein structures (figures 5 and 6). The success of this docking approach is critically dependent on both the quality of the alignments and also on the minimization problem. Minimization of the docked structures performs a local docking and the minimized solutions in some instances tend to greatly differ for molecules like HSP90_44 (figure 6) where a major part of its substructure accesses solvent exposed region of HSP90. Hence the poor RMSD values in spite of a good chemical similarity in case HSP90 compounds are mostly due to the 'minimization problem'. Since the HSP90 ligands are relatively rigid with extensive delocalization of pi-electron densities that confers certain conformational rigidity to the D3R HSP90 chemical structures. Hence we decided to use a single low energy conformations obtained from ChemAxon's 'cxcalc' suite for performing alignments. However the high RMSD could also due to the alignment problem. The pi-electron density delocalization especially to neighboring substructures like aromatic systems is highly context specific, if sufficient delocalization via various electronic effects has occurred the structures assume flat planar conformations

else they can acquire dihedral angles which are perpendicular(and other angles) to the substructures they are connected to. Hence for more accurate alignments and for chemical structures with large number of rotatable bonds, multiple conformers could be generated using the 'qmd' (quenched molecular dynamics) module of open3Dalign prior to alignments. Furthermore structural water molecules play an important role in the recognition of ligands by the ~~heat shock protein~~ HSP90 as rightly pointed out by the D3R organizers.[38] While four conserved water molecules bridge the interactions between HSP90 with ADP or PU3 (a purine based lead compound) [39], Radicolol on the other hand, displaces one of these water molecules and noticeably shifts two others [40]. However, current methods do not deal adequately with shifting the positions of bridging water molecules and hence we ignored such detailed analysis. This omission may account at least partly for the inability of our method to achieve considerably low RMSD values. In view of our inability to consider bridging water molecules, we decided to use predicted ki values obtained as scores from various QSAR model trained using a dataset of HSP90 inhibitors (Figure2, Figure 3 and Figure 4B).

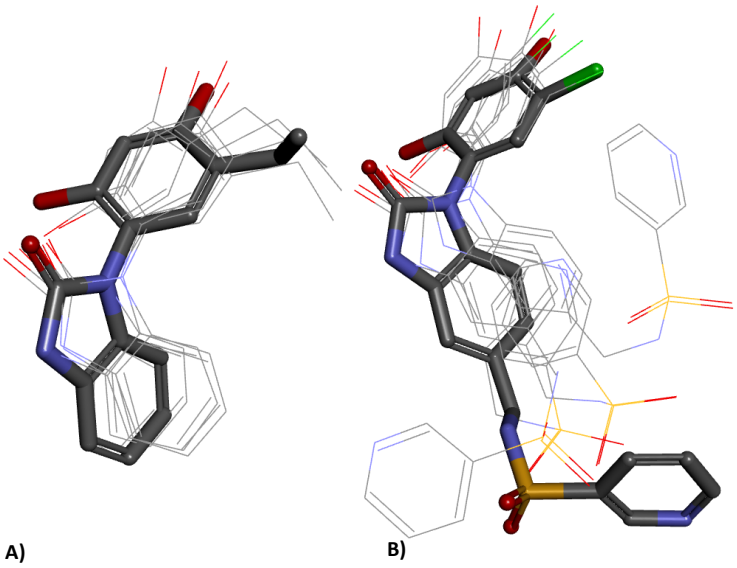

Figure 5: (a) HSP90_40 exemplifies the good docking prediction and low RMSD values for a compound which only accesses buried binding pocket. The co-crystal structure derived binding mode provided by the D3R organizers is depicted as a stick and our predicted poses as lines. The various RMSD values are presented in supplementary material 2 (b) HSP90_44 is the example of a compound which accesses both buried and solvent accessible binding pocket. Since solvent accessible generally do not have anchor

residues ligand pose prediction is challenging and leads to high RMSD values. The co-crystal structure derived binding mode provided by the D3R organizers is depicted as a stick and our predicted poses as lines. The various RMSD values are presented in supplementary material 2

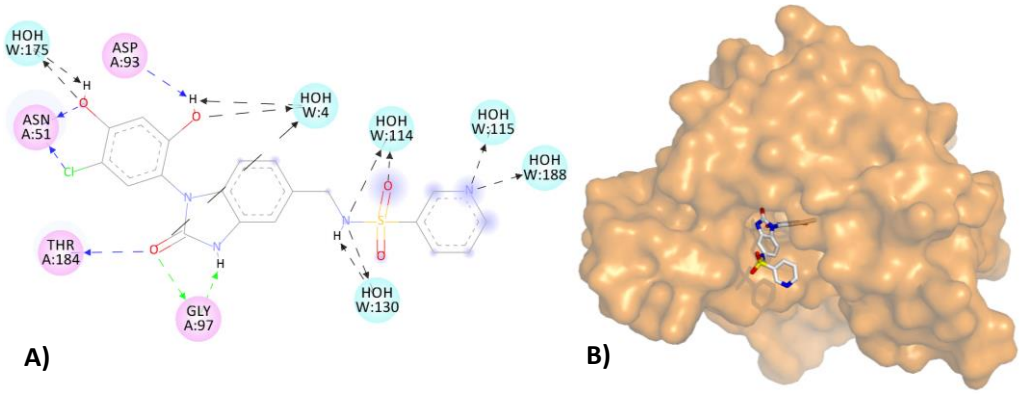

Figure 6: (A) 2D interactions of HSP90_44 with 4YKT, the lack of anchor resides to in teract with HSP90_44's sulphonamide fragment and the importance of bridging water molecules (W:175 and W:4) is also highlighed, (B) The surface depiction of HSP90 showing the sulphonamide fragment of HSP90_44 sticking out into the solvent exposed region.

Solvent exposed ligand-binding site, use of incorrect tautomeric forms of MAP4K4 ligands and PLANTS docking approach are responsible for deviations between predicted and experimental binding modes

A typical kinase ligand-binding site consists of the hinge-binding region, the catalytic site, the gate-keeper region and a solvent exposed region. Docking a ligand to these solvent exposed regions is often challenging, as there are very few polar residues to which ligands can anchor. These solvent exposed regions are not believed to contribute to binding affinities and are often included primarily to improve pharmacokinetic and pharmacodynamic properties of the ligands. Furthermore, docking small molecular fragments also represents a significant challenge.[41] Firstly, fragments are more promiscuous in their binding modes than larger 'drug-like' molecules making their predictions difficult; second, docking scoring functions are inaccurate even for the larger molecules against which they have often been parameterized, and are likely to be still less accurate for fragments[42],[43,44]. Unconstrained docking may thus not be widely accepted as a reliable method to for pose and affinity predictions, and there are few if any studies that compare docking-predicted fragment geometries to subsequent structural results. Kinase inhibitory fragments

typically access only the hinge-binding pocket and anchor via hydrogen bonding interactions with the main chain atoms (cys108 in case of MAP4K4) and also via pi-stacking or hydrophobic interactions with aromatic residue of the hinge loop. In case of our analysis, an incorrect tautomer of MAP32 was docked which further exacerbated the deviation in the predicted pose. We have erroneously used the structures provided by the D3R organizers without tautomer enumeration. For the small fragments hinge binding predictions are critically impacted by tautomer enumeration and prioritization (Figure 7). On the other hand pose prediction of MAP4K4 inhibitors which access both the hinge and catalytic site are predicted relatively better (Figure 8). Other sources of poor pose predictions for MAP4K4 in terms of their RMSD estimates include of the use of PLANTS docking approach which considers side chain flexibility and the use of an approach, different from D3R to calculate the RMSD measures. In all constrained docking approaches, such as PoPSS[45], which predict the binding pose of a query ligand using 3D shape similarity with known crystallographic ligands may be more appropriate for docking kinase fragments which only target the hinge region.

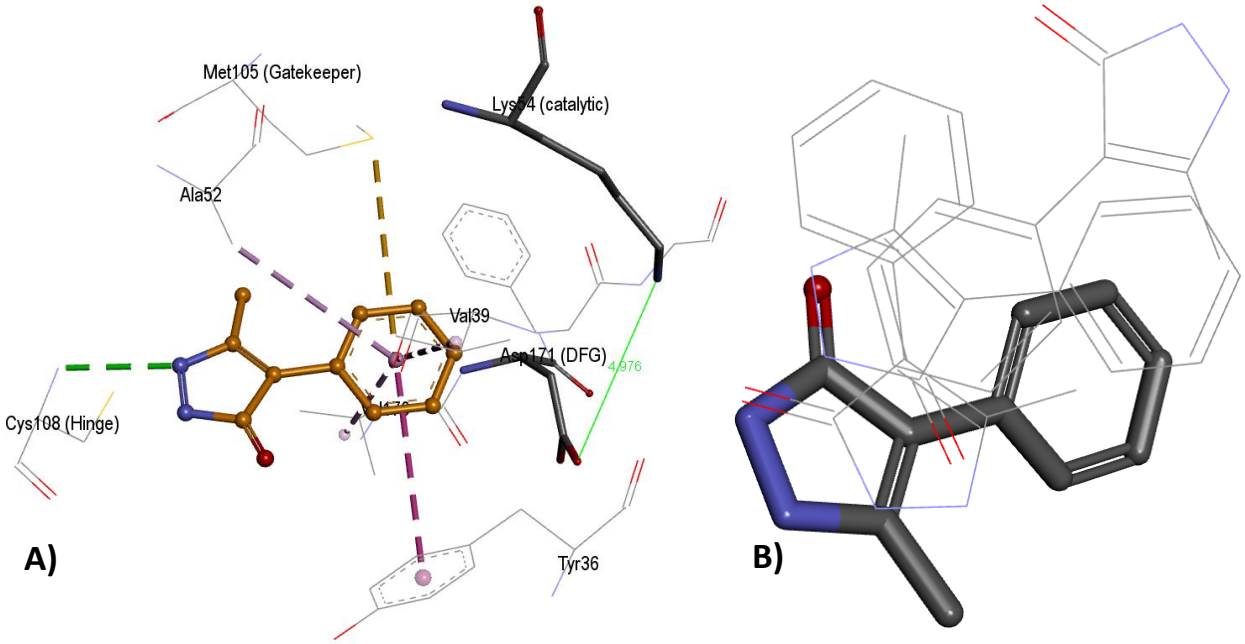

Figure 7: (A) 3D interactions of MAP32 with 4U44 highlight the hinge-binding interactions of MAP32 via the hydrogen bond. The catalytic site comprising of the Lys54 and Asp 171 is not occupied (B) The poor docking prediction for MAP32 (small fragment) can be attributed to the lack of strong anchor residues in

the hinge region (for RMSD values of 'MAP32' pose predictions please kindly refer to supplementary material 3). The co-crystal structure derived binding mode provided by the D3R organizers is depicted as a stick and our predicted poses as lines.

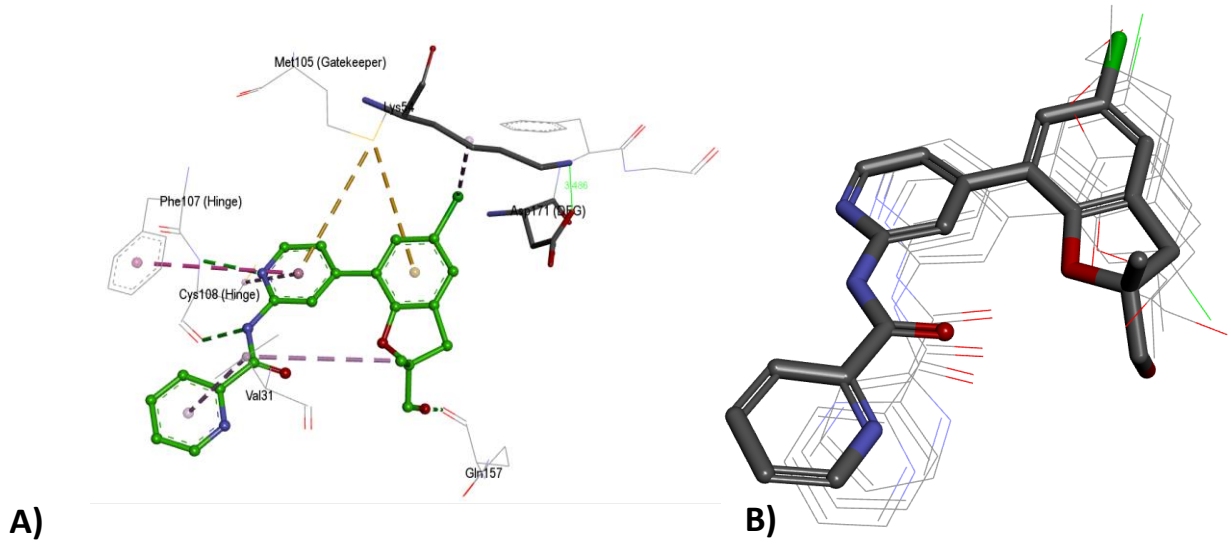

Figure 8: (A) 3D interactions of MAP25 with 4U44 highlight the hinge-binding (hydrogen bonding with Cys108 and pi interactions with Phe107) and catalytic site interactions of MAP25. The catalytic site comprising of the Lys54 and Asp 171 is also occupied (B) The good docking prediction for MAP25 can be attributed to the strong anchor residues in the hinge region (Phe107 and Lys54) and conformational rigidity of the molecule (for RMSD values of 'MAP25' pose predictions please kindly refer to supplementary material 3). The co-crystal structure derived binding mode provided by the D3R organizers is depicted as a stick and our predicted poses as lines.

Better affinity prediction performance of the relatively novel target MAP4K in comparison to the well-established target HSP90 could be a dataset attribute characteristic.

The prior knowledge in terms of activity as well as crystal structures for HSP90 is much greater in comparison to MAP4K4. However the affinity prediction performance however seems counter intuitive. The relatively poorer performance of MAP4K4 could just be an artifact of the dataset attributes such as the size and the spread of activity data. The HSP90 dataset with around 180 compounds is 10-fold larger than the MAP4K4 dataset with 18 compounds. Furthermore a significant portion of HSP90 compounds are inactive and show no variation in activities. Hence we analyzed the receiver operating characteristic AUC for both HSP90 and MAP4K4, where similar AUC values of 0.702 for HSP90 and 0.728 for MAP4K4 were observed (Figure 3).

Counter intuitive performance of MAP4K4 pose versus affinity prediction performance.

The poor MAP4K4 pose prediction results in terms of mean RMSD of 4.346 seems counter-intuitive, in light of a comparatively better affinity prediction performance in terms of ROC AUC value 0.728 at 7.5 pIC50 cut-off and R = 0.67. In addition to small size of this dataset, the following argument could explain the counter-intuitive results. Kinase fragment docking using our protocol, wherein protein-ligand interaction fingerprint analyses were used for automated pose prioritization, needs further validation and revision. Pose selection is often performed manually using visual inspection or automatically selecting the lowest energy pose. In the present studies we tried to address the automatic selection of poses which engage with hinge-, gatekeeper- and the catalytic- site residues. An exhaustive survey of the kinase fragments crystallized with their cognate proteins reveals that chemically similar fragments bind alternatively to either hinge or the catalytic site (consisting of the catalytic lysine and DFG motif). The compounds which failed the pose prediction assessments were predominantly hinge binding fragments. The protein-ligand protocol which employed a composite score comprising both fingerprint similarity and free energy primarily prioritized poses in which the fragment bound with the catalytic site since they are relatively more polar than the hinge pocket. These fragments primarily engaged the catalytic lysine via cation-pi and hydrogen bonding interactions. The plausibility of these alternative poses could explain why good affinity predictions were seen even when the pose prediction seems to far from reasonable. Furthermore the PLANTS docking procedure considers the side chain flexibility of amino acid residues, which could have also impacted the pose prediction accuracy's in few other D3R MAP4K4 molecules. Pose prediction RMSD are known to be significantly impacted by the use of flexible docking procedures.[28,31,46] Finally our RMSD estimation method employed the DS Visualizer (which uses an all heavy atom approach to calculate the degree of overlap) differs from the protocol used by the organizers who used the maximum common substructure functionality of the OEChem Python toolkit.

Conclusion

In the present study, we report the application and results of two different strategies for the two protein systems in the D3R grand drug design challenge. For the well-studied HSP90 target system, our integrated docking and scoring approach performed well in pose and poorly for affinity prediction. For the second protein MAP4K4 with limited data available, we implemented an exhaustive docking and scoring protocol assisted by novel pose prioritization using protein-ligand interaction fingerprints. It resulted in reasonable affinity prediction and poor pose predictions.

ACKNOWLEDGMENTS

This work was in part supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (Grant Numbers 25430186 and 25293079) and from Japan Agency for Medical Research and Development ("The adjuvant database project grant number 16ak0101010h0005") to K.M.

Supplementary material

Supplementary material 1

Supplementary material 2

Supplementary material 3

Supplementary material 4

References

1. Prathipati P, Dixit A, Saxena AK (2007) Current computer aided molecular design 92:29
2. Prathipati P, Mizuguchi K (2016) Curr Top Med Chem 16(9):1009
3. Walters WP, Stahl MT, Murcko MA (1998) Drug Discovery Today 3(4):160
4. Prathipati P, Mizuguchi K (2015) J Chem Inf Model
5. Prathipati P, Pandey G, Saxena AK (2005) J Chem Inf Model 45(1):136
6. Prathipati P, Saxena AK (2005) J Comput Aided Mol Des 19(2):93
7. Prathipati P, Saxena AK (2006) J Chem Inf Model 46(1):39
8. Barillari C, Marcou G, Rognan D (2008) J Chem Inf Model 48(7):1396
9. Radifar M, Yuniarti N, Istyastono EP (2013) Bioinformation 9(6):325
10. Brewerton SC (2008) Curr Opin Drug Discov Devel 11(3):356
11. Da C, Kireev D (2014) J Chem Inf Model 54(9):2555

12. Deng Z, Chuaqui C, Singh J (2004) J Med Chem 47(2):337
13. Roy KK, Singh S, Saxena AK (2011) Mol Divers 15(2):477
14. Saxena AK, Saxena S, Chaudhaery SS (2010) SAR QSAR Environ Res 21(1):1
15. Saxena S, Chaudhaery SS, Varshney K, Saxena AK (2010) SAR QSAR Environ Res 21(5-6):445
16. Verkhivker GM, Dixit A, Morra G, Colombo G (2009) Curr Top Med Chem 9(15):1369
17. Crawford TD, Ndubaku CO, Chen H, Boggs JW, Bravo BJ, Delatorre K, Giannetti AM, Gould SE, Harris SF, Magnuson SR, McNamara E, Murray LJ, Nonomiya J, Sambrone A, Schmidt S, Smyczek T, Stanley M, Vitorino P, Wang L, West K, Wu P, Ye W (2014) J Med Chem 57(8):3484
18. Wang L, Stanley M, Boggs JW, Crawford TD, Bravo BJ, Giannetti AM, Harris SF, Magnuson SR, Nonomiya J, Schmidt S, Wu P, Ye W, Gould SE, Murray LJ, Ndubaku CO, Chen H (2014) Bioorg Med Chem Lett 24(18):4546
19. Langfelder P, Horvath S (2008) BMC Bioinformatics 9(1471-2105 (Electronic)):559
20. Prathipati P, Ma NL, Keller TH (2008) J Chem Inf Model 48(12):2362
21. Rogers D, Hahn M (2010) J Chem Inf Model 50(5):742
22. Tosco P, Balle T, Shiri F (2011) J Comput Aided Mol Des 25(8):777
23. Koes DR, Baumgartner MP, Camacho CJ (2013) J Chem Inf Model 53(8):1893
24. Friedman J, Hastie T, Tibshirani R (2010) J Stat Softw 33(1):1
25. Hinselmann G, Rosenbaum L, Jahn A, Fechner N, Zell A (2011) J Cheminform 3(1):3
26. Boughorbel S, Al-Ali R, Elkum N (2016) PLoS One 11(1):e0146413
27. Korb O, Stutzle T, Exner TE (2009) J Chem Inf Model 49(1):84
28. Smith RD, Dunbar JB, Jr., Ung PM, Esposito EX, Yang CY, Wang S, Carlson HA (2011) J Chem Inf Model 51(9):2115
29. Sunseri J, Ragoza M, Collins J, Koes DR (2016) J Comput Aided Mol Des
30. Ye Z, Baumgartner MP, Wingert BM, Camacho CJ (2016) J Comput Aided Mol Des
31. Damm-Ganamet KL, Smith RD, Dunbar JB, Jr., Stuckey JA, Carlson HA (2013) J Chem Inf Model 53(8):1853
32. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) Nucleic Acids Res 35(Database issue):D198
33. McCammon JA, Gelin BR, Karplus M (1977) Nature 267(5612):585
34. Kontoyianni M, McClellan LM, Sokol GS (2004) J Med Chem 47(3):558
35. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) J Med Chem 49(20):5912
36. Leach AR, Shoichet BK, Peishoff CE (2006) J Med Chem 49(20):5851
37. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) Genome Biol 4(5):P3
38. Yan A, Grant GH, Richards WG (2008) J R Soc Interface 5 Suppl 3:S199
39. Wright L, Barril X, Dymock B, Sheridan L, Surgenor A, Beswick M, Drysdale M, Collier A, Massey A, Davies N, Fink A, Fromont C, Aherne W, Boxall K, Sharp S, Workman P, Hubbard RE (2004) Chem Biol 11(6):775
40. Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH (1999) J Med Chem 42(2):260
41. Chen Y, Shoichet BK (2009) Nat Chem Biol 5(5):358
42. Marcou G, Rognan D (2007) J Chem Inf Model 47(1):195
43. Hubbard RE, Chen I, Davis B (2007) Curr Opin Drug Discov Devel 10(3):289
44. Klebe G (2006) Drug Discov Today 11(13-14):580
45. Kumar A, Zhang KY (2016) J Comput Aided Mol Des
46. Dunbar JB, Jr., Smith RD, Damm-Ganamet KL, Ahmed A, Esposito EX, Delproposto J, Chinnaswamy K, Kang YN, Kubish G, Gestwicki JE, Stuckey JA, Carlson HA (2013) J Chem Inf Model 53(8):1842