JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Predicting Ligand Binding Modes from Neural Networks Trained on Protein−Ligand Interaction Fingerprints

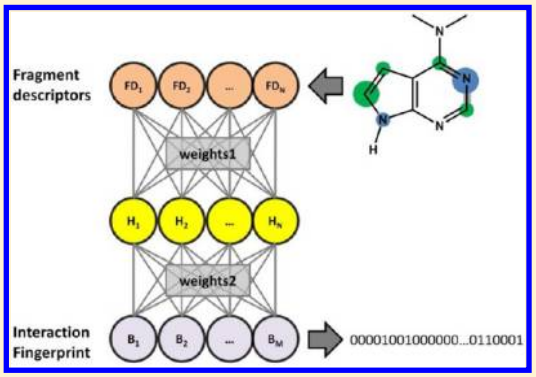
Vladimir Chupakhin,[†,‡] Gilles Marcou,[†] Igor Baskin,[†,§] Alexandre Varnek,[†] and Didier Rognan*[,‡]

[†]Laboratory of Chemoinformatics, UMR 7177 University of Strasbourg/CNRS, 4 rue B. Pascal, 67000 Strasbourg, France

[‡]Laboratory of Therapeutic Innovation, UMR 7200 University of Strasbourg/CNRS, MEDALIS Drug Discovery Center, 74 route du Rhin, F-67400 Illkirch-Graffenstaden, France

Supporting Information

ABSTRACT: We herewith present a novel approach to predict protein−ligand binding modes from the single two-dimensional structure of the ligand. Known protein−ligand X-ray structures were converted into binary bit strings encoding protein−ligand interactions. An artificial neural network was then set up to first learn and then predict protein−ligand interaction fingerprints from simple ligand descriptors. Specific models were constructed for three targets (CDK2, p38-α, HSP90-α) and 146 ligands for which protein−ligand X-ray structures are available. These models were able to predict protein−ligand interaction fingerprints and to discriminate important features from minor interactions. Predicted interaction fingerprints were successfully used as descriptors to discriminate true ligands from decoys by virtual screening. In some but not all cases, the predicted interaction fingerprints furthermore enable to efficiently rerank cross-docking poses and prioritize the best possible docking solutions.



INTRODUCTION

Knowing the target-bound conformation of a ligand as well as the molecular interactions with its binding site (pose) is crucial information in drug design, both at the hit-to-lead and lead optimization steps.[1] Among the computational methods of choice for predicting protein−ligand interactions is molecular docking.[2,3] Despite noticeable breakthroughs in pose prediction,[4] notably for low molecular-weight ligands[5] and flexible binding sites,[6] docking is still hampered by the relative inaccuracy of fast scoring functions to discriminate correct from wrong poses and to rank ligands by decreasing binding free energy.[7,8] Since scoring functions undoubtedly undergo real but still modest improvements,[9] an alternative strategy to rigorous prediction of free energies is to postprocess docking poses to prioritize the most reliable ones irrespective of energetic criteria. Among the many possible approaches, the usage of protein−ligand interaction fingerprints (IFP)[10] has gained considerable popularity.[11] An IFP is a simple bit string registering, from atomic three-dimensional (3D) coordinates, the presence or absence of well-defined interactions (apolar, aromatic, hydrogen bonds, salt bridges, metal coordination) between a ligand and a fixed set of amino acids (or protein atoms) lining its binding site. Interactions are calculated on-the-fly according to a set of rules (atom types) and geometric relationships (distances, angles) between interacting atoms. A bit is then switched on or off as whether an interaction with a residue (residue-based IFP)[10] or an atom (atom-based IFP)[12] from the binding site is verified.

Numerous validations and applications of IFPs have been reported in the last 5 years for (i) mining and clustering docking poses,[13−15] (ii) predicting ligand poses,[16] (iii) prioritizing scaffold posing[15] and hopping,[17] (iv) selecting most valuable virtual hits,[18,19] and (v) designing target-focused libraries.[14] 3D coordinates of protein−ligand complexes obtained either experimentally (X-ray diffraction) or computationally (docking) are mandatory to generate IFPs. *We herewith investigate whether the IFP of a peculiar ligand to a single target could be inferred from the sole two-dimensional (2D) structure of the ligand.* In other words, could a binding mode to a particular target be predicted from simple ligand descriptors and a repository of known binding modes (IFPs)? QSAR/QSPR methods are usually trained to output a single number (bit, biological activity, physicochemical property). We here present an example of multiple output learning based on an artificial neural network reading ligand descriptors and writing IFPs. The current study aims at demonstrating the proof-of-concept for three human targets of interest to the drug design community (cyclin-dependent kinase 2 or CDK2, mitogen-activated protein kinase 14 or p38-α, and heat shock protein 90-α or HSP90-α) and for which numerous protein−ligand high resolution X-ray structures are available.[20] Despite real shortcomings of the current method, we unambiguously evidence that IFPs (in other words binding modes) can be

Received: April 23, 2012

   dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

accurately predicted for most ligands from their simple 2D structures.

METHODS

Protein−Ligand X-ray Structures. X-ray structures of protein−ligand complexes were retrieved in MOL2 file format from the sc-PDB data set (2010 release) of druggable protein−ligand binding sites.[20] Three human targets of pharmaceutical interest (CDK2, UniProt AC: P24941; p38-α, UniProt AC: Q16539; HSP 90-α, UniProt AC: P07900) were chosen for their abundance in the sc-PDB. For every target, active site mutants were discarded. The amino acid side chain was renumbered whenever necessary according to the UniProt numbering and Cα binding site atoms aligned with FuzCav.[21] IFPs were calculated for every protein−ligand complex using the FingerPrintLib program[15] based on OEChem TK.[22] A seven bit-string was written for every active site residue registering seven interaction types (apolar, aromatic face to face, aromatic edge to face, H-bond donated by the protein, H-bond donated by the ligand, ionic bond with protein cation and anion, respectively) verifying standard geometric rules.[15] An active site residue was considered as any amino acid for which at least one of the above-described seven bits was switched on. Since the number of active site residues is dependent on the ligand, a fixed-length IFP was generated for every target by selecting the most frequently interacting amino acids. A frequency threshold of 90% was applied to all targets, excepted for CDK2 for which a threshold of 80% was applied. These thresholds were manually selected in order to obtain consensus sites large enough to represent the different binding modes observed experimentally. This procedure led to the definition of 15, 15, and 11 consensus active site residues for CDK2 (72 entries), p38-α (39 entries), and HSP90-α (35 entries), respectively (Table 1). Final IFPs were generated from these

Table 1. Targets and Ligands Used for the Building the IFP Models and Their Application

	set	CDK2	p38-α	HSP90-α
consensus residues		15	15	11
IFP length		105	105	77
sc-PDB ligand	training	72	39	35
DUD decoys	training	72	39	35
ChEMBL actives	validation	572	1552	71
DUD actives	validation	52	269	23
DUD decoys	validation	1662	7784	784

restricted sets of amino acids. The complete lists of sc-PDB entries and active site residues selected for IFP generation are available as Supporting Information. The final data sets associate chemical structures with IFPs for each complex.

Preparation of the Training Sets and Building the IFP Models. For each of the three targets, we chose an equivalent number of target-specific positive instances (sc-PDB ligands) and of target-specific negative instances (decoys) randomly chosen from the DUD data set.[23] All ligands in SD file format were cleaned with the ChemAxon Standardizer[24] software. Structures were aromatized and neutralized, and explicit hydrogen atoms were removed. Two types of molecular fragment descriptors were calculated with the ISIDA/Fragmentor[25] software: sequences of atoms and atoms and their neighbors with a path length up to 6 atoms. The Stuttgart Neural Network Simulator (SNNS) software[26] was used to build IFP models using a supervised feed-forward multilayer neural network with a resilient propagation algorithm. Every neural network was composed of three layers: input, hidden, and output (Figure 1). The size of input and hidden layer was

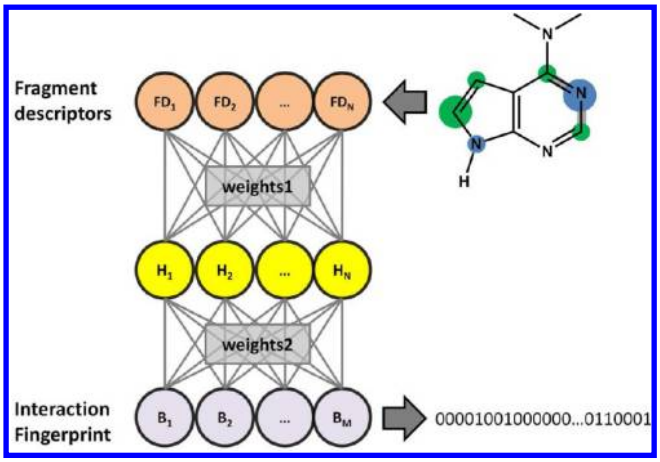


Figure 1. Scheme of the artificial neural network used for IFP prediction (FD: fragment descriptor value, H: hidden layer of the artificial neural network, B: bit in the IFP; weights: weights used for building the models with the neural network).

equal to the number of calculated molecular descriptors (937 for CDK2, 682 for p38-α, and 786 for HSP90), whereas the target-specific output layer had a size equal to the number of bits in the IFP (105 for CDK2 and p38-α and 77 for HSP90-α). As illustrated in Figure 2, each data set was randomly split five

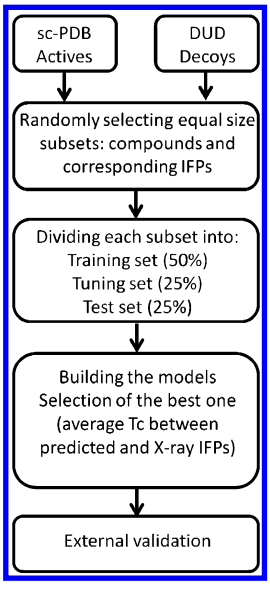


Figure 2. Workflow for IFP modeling.

times into a training set (50% of the total number of entries) to train the model, a tuning set (25%) for early stopping the model to prevent overfitting, and a test set (25%) for external validation of the model. The neural network was trained on the sc-PDB IFPs for positive instances and on empty bit strings (all bits set to 0) for negative instances. The predictive performance of every IFP model was assessed for the test set as the average pairwise Tanimoto coefficient (Tc) between predicted and X-

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

ray IFPs. The best model for every target was chosen on the basis of the highest average Tc for the test set.

Virtual Screening Sets. The relevance of the predicted IFPs was first assessed by a similarity based virtual screening experiment. True actives were either taken from the ChEMBL[27] or the DUD[23] databases. ChEMBL active compounds were selected from binding assays according to their affinity to our targets ($IC_{50} \leq 20\ \mu M$) with a confidence score of 8 or 9. Decoys were selected only from the DUD data set, unless they were part of the training set. Therefore a target-specific set of actives and decoys was assembled. Ligand structures were retrieved in SD file format and prepared as previously described. For every compound, an IFP to the intended target was predicted and compared to each X-ray IFP of the training sc-PDB entries by means of a Tc coefficient. Compounds were ranked by decreasing Tc values. Therefore, for each X-ray IFP of the training set, a specific receiver operating characteristic (ROC)[28] plot was obtained.

External Validation Set. For docking validation purposes, we also collected an external validation set (EV set) consisting of new ligands and their corresponding IFPs for complexes of the studied targets not yet included in the sc-PDB data set (deposited in the PDB after June 2011). Complexes extracted from the RCSC PDB[29] were prepared according to the same procedure as for sc-PDB entries.[30] Ligands of the external validation set were extracted and standardized as described above, and the best models were used to predict the IFP for each of the ligands. The list of PDB identifiers in the EV set is given in the Supporting Information.

Docking. Ligands from the EV set were docked to all protein structures from the training set. Docking was made using default parameters of the GOLD 5.0 software,[31] with a search sphere of 12.5 Å around the center of mass of the cocrystallized ligand and an early termination criterion if the top 3 ranked poses were within 1.5 Å root-mean-square (rmsd) deviation. For all docked poses, IFPs were calculated with FingerprintLib[15] as previously described. For postdocking analysis, all binding sites were aligned using standard settings of the FuzCav software.[21]

RESULTS AND DISCUSSION

Assessment of Predicted IFP Quality. The IFP models output the probability of each interaction from 0 to 1. The probabilities were converted to a bit string, using a threshold above which the corresponding predicted interaction was translated to 1 or 0 otherwise. Several thresholds were tested, and the resulting IFPs were compared to their corresponding X-ray IFPs. However the choice of the threshold appeared to have a relatively weak influence on the resulting IFP, if chosen between 0.2 and 0.8. Therefore we decided to use a threshold of 0.5 in the following. With respect to the (internal) test set, IFP models were able to correctly predict about two-thirds of all interactions on average (Figure 3). The average pairwise Tanimoto between predicted and X-ray IFPs ($Tc_{pred\_vs\_X\text{-}ray}$) is identical for two targets (0.65 for CDK2, p38-α) and higher for HSP90-α (0.79) but always above the acceptable threshold of 0.6 that we previously defined for many independent data sets.[15] It should be mentioned that the better predictions on the HSP90-α set might be due to a lower chemical diversity on the training ligands (Supporting Information).

For each target, models were built and validated on scrambled data, each IFP being randomly associated with a training set compound. The process was repeated 100 times

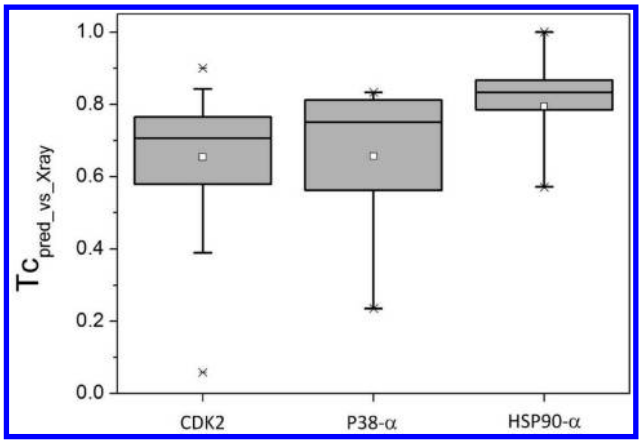


Figure 3. Box-and-whisker plot of the Tanimoto coefficient between X-ray and predicted IFPs for complexes of the training set (CDK2: n = 72; p38-α: n = 39; HSP90-α: n = 35). The box delimits the 25th and 75th percentiles; the whiskers delimit the 5th and 95th percentiles. Median and mean values are indicated by a horizontal line and an empty square. Crosses delimit the 1st and 99th percentiles, respectively.

each. The 95th percentile relates to values of 0.58 for CDK2, 0.59 for p38-α, and 0.53 for HSP90-α. In other words, the above selected threshold of 0.6 insures that the obtained models are chance correlation-free.

Next, IFPs were predicted for complexes from the external test set (EV set). The prediction was considered successful for all cases where $Tc_{pred\_vs\_X\text{-}ray} > 0.6$. Using this cutoff the success rate was 69% (9/13) for CDK2, 71% (10/14) for p38-α, and 73% (16/22) for HSP90-α. Additional tests were performed to compare the distribution of the average bit frequency for predicted IFPs (ChEMBL true actives) and X-ray IFPs (sc-PDB ligands; Figure 4). As to be expected, both distributions for the three targets of interest are indeed correlated, with a correlation coefficient of 0.89 (CDK2), 0.89 (p38-α), and 0.88 (HSP90-α). However, significant differences between bit frequencies of sc-PDB and ChEMBL ligands occur notably for the most frequent hydrophobic interactions and for aromatic−aromatic interactions whose definition (edge-to-face, face-to-face) is highly directional. Interestingly, the distribution of the average frequency of bits describing hydrogen bonds was similar between sc-PDB and ChEMBL sets. Finally, the models generally ignored rare interactions (frequency < 0.1).

A typical limit in predicting IFPs from machine learning models is illustrated Figure 5. Predicted probabilities of on-bit switching for every active site residue have been mapped on a 2D sketch of observed interactions between the 1GK ligand and p38-α in the 3mpt X-ray structure. Predicting the corresponding IFP can be considered successful ($Tc_{pred\_vs\_X\text{-}ray}$ = 0.63) although a real discrepancy occurs on His107 which accepts an H-bond from the cocrystallized ligand but is predicted to engage hydrophobic contacts by the model. This misprediction is due to the following: (i) a lack of registered H-bonds between His107 and p38-α ligands in the sc-PDB training set (the corresponding bit frequency for this interaction is 0.1) and (ii) the absence of any pyrrole-forming hydrogen bond among training sc-PDB ligands. This failure illustrates two problems in our application: the difficulty to learn rarely occurring interactions and the absence of a relevant applicability domain.

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

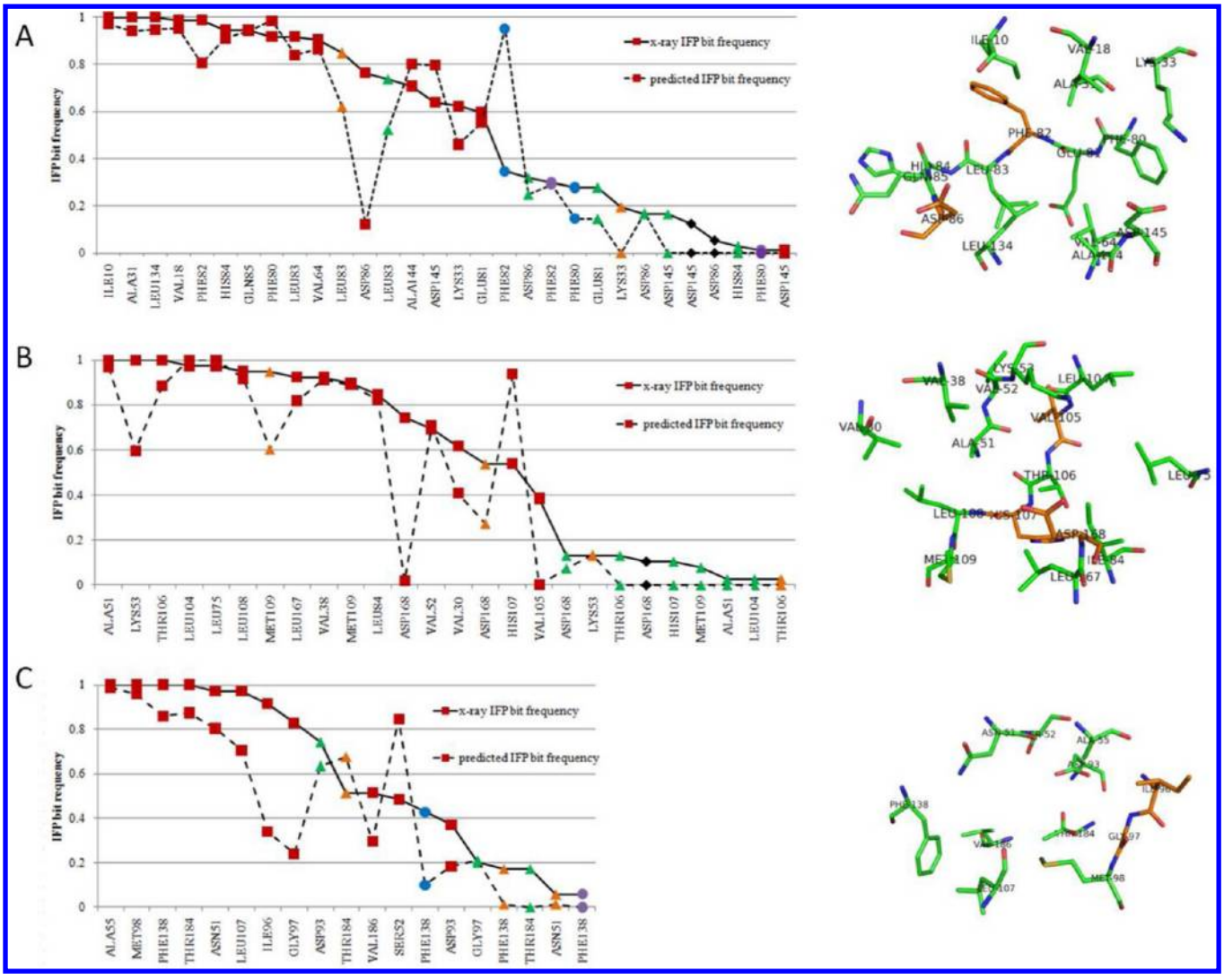


Figure 4. Comparison of the average bit frequency for X-ray IFP (continuous line) and predicted IFP (dotted line) for CDK2 (A), p38-α (B), and HSP90-α ligands (C). Interactions are labeled as follows: red square: hydrophobic interaction; orange triangle: hydrogen bond donated by the protein; green triangle: hydrogen bond donated by the ligand; blue circle: face-to-face aromatic interaction; purple circle: edge-to-face aromatic interaction; dark rhomb: ionic interaction. Consensus binding site residues (right panels) are colored in orange if their predicted interaction frequencies deviate from experiment by more than 0.2.

Virtual Screening with Predicted IFPs. IFPs have been widely used to improve hit rate retrieval from docking poses.[18,32] Along this line, we challenged the suitability of predicted IFPs for similarity-based virtual screening. For the three targets of interest, IFPs for compounds from DUD (DUD actives, DUD decoys) and ChEMBL (ChEMBL actives, DUD decoys) sets were predicted. Then, Tanimoto coefficients between predicted and each of the target-specific X-ray IFPs (sc-PDB set) were iteratively calculated and used to discriminate between active and inactive compounds. As many classifications as reference X-ray structures (CDK2:72, p38-α: 39, HSP90-α: 35) were computed and the corresponding ROC curves inspected. For the three targets and the two data sets (DUD, ChEMBL), predicted IFPs appear to well discriminate true actives from decoys with areas under the ROC curve ranging from 0.76 to 1.00 (Figure 6). Distribution of area under the ROC curves indicates few variations of the ROC values for the two test sets and the three targets (Figure 7), notably for screening CDK2 ligands which provide very comparable ROC curves from IFP similarity measure to each of the 72 reference ligands.

In addition, we computed the ability of the IFP prediction models in the early enrichment in true actives (enrichment in true positives at a constant false positive rate of 1%). Obtained enrichments (Figure 7) ranged from good (CDK2 DUD set, p38-α DUD and CHEMBL sets, HSP90-α ChEMBL set) to excellent (HSP90-α DUD set, CDK2 ChEMBL set) and could not be explained by any bias either in the chemical diversity of training ligands nor in some unexpected high pairwise similarity between ligand sets (DUD actives vs X-ray, ChEMBL actives vs X-ray; Supporting Information).

For the purpose of comparison, classification neural networks models have been directly built using the same ISIDA descriptors as for IFP modeling. Instead of predicting the IFP, the model here directly predicts activity in a binary manner but from exactly the same ISIDA descriptors as previously. The later models were quantitatively of similar accuracy to that based on IFP-based similarity search (Table 2) thereby suggesting that predicted IFPs are suitable descriptors for

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

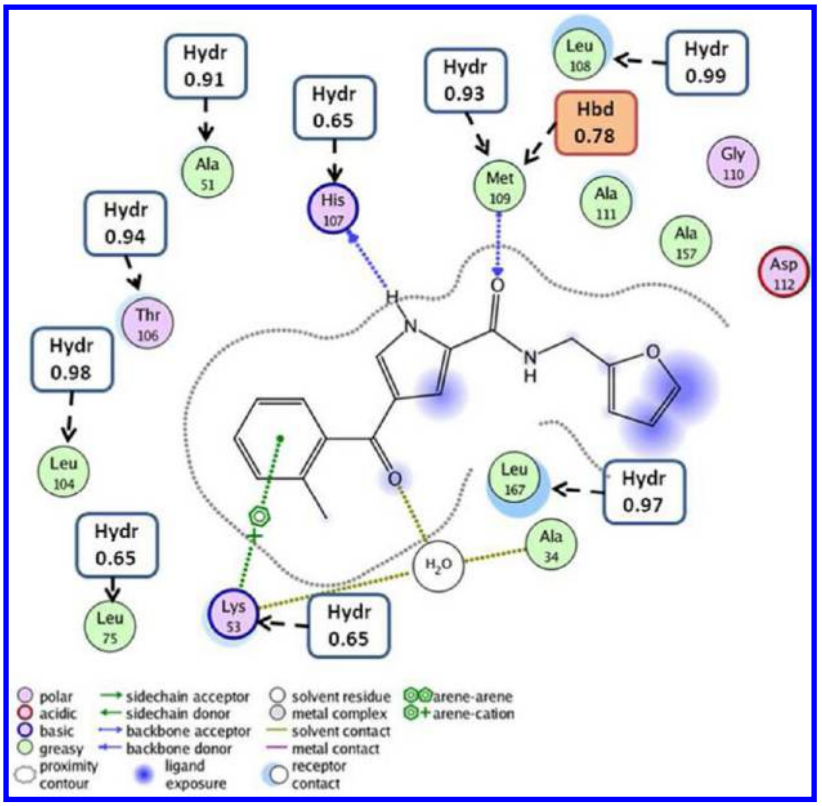


Figure 5. Mapping predicted interactions for ligand 1GK with the p38-α binding site (pdb entry 3MPT). Amino acids without contact probability scores and labels (Hydr, hydrophobic contacts; Hbd, Hydrogen bond) were not included in the consensus binding site. The 2D sketch was rendered by MOE.[37]

virtual screening. It is fair to acknowledge that, at least on the herein presented three data sets, predicting IFPs does not provide a significant added value with respect to the standard ISIDA fragment descriptors. It is however reassuring that predicted IFPs appear to encode all necessary information to really discriminate actives from inactives in virtual screening exercises.

Prospective IFP-Guided Docking. Predicted IFPs were challenged to retrieve near-native poses of ligands with known complex structure (EV set, not used in the modeling stage), out of pools of poses proposed by cross-docking. Several parameters were computed for every docked conformation of the ligands: rmsd of the docked conformation to the X-ray pose and Tanimoto coefficient between docked and predicted IFP ($Tc_{docked\_vs\_pred}$) and between docked and X-ray IFP ($Tc_{docked\_vs\_X\text{-}ray}$).

In cross-docking, every ligand from the EV set has been docked to the binding sites of all complexes of the training set. Conformations were prioritized according to the similarity of the docked to the predicted IFP of that molecule. In case of identical similarity values, the pose with the higher Gold Score was selected (TC_GS strategy). In parallel, a second strategy based on selecting poses according to GoldScore alone was followed (GoldScore strategy).

Overall, TC_GS and GoldScore pose ranking seem to have comparable performances. The success rate (count of cases for which rmsd <2.0 Å) is summarized in Table 3. Only for CDK2, the TC_GS strategy was better than the GoldScore ranking. Yet, the rmsd to the X-ray pose may be misleading since it reflects a global measurement which may suffer from a single but large mismatch.[15] Therefore, the IFP of the docked poses was also compared to the IFP from the X-ray pose (Table 4). According to this criterion, TC_GS is a better scoring function than GoldScore for cross-docking to CDK2 and HSP90.

Figure 8 gives a more detailed view of these results. For each ligand, either rmsd or $Tc_{docked\_vs\_X\text{-}ray}$ of selected poses are plotted as a function of the strategy used. In addition, one pose was selected *a posteriori* as the closest to the crystallographic one according to the rmsd. As expected, the IFP of this pose was also in general − but not always − very similar to the IFP of the X-ray. Detailed analysis of cross-docking results shows that in most of the cases the TC_GS strategy based on predicted IFP yields reasonable poses (Figure 8). This achieved to demonstrate that the predicted IFP convincingly described the binding mode of ligands. However, we could identify different causes of failure which are discussed below and illustrated in Figure 9.

Failure Due to Multiple Binding Site Conformations. In this case, the retrieved ligand forms the same or a very similar pattern of interactions but has a rmsd to X-ray ≥ 2.0 Å, due to conformational differences in the binding sites used for docking. For example, the ligand 3HUB (Figure 9A) may bind into the 1ZYJ X-ray structure of the p38-α site and form the same interaction pattern as with its "own" 3HUB site (Tc = 0.80). Nevertheless, alignment of the 3HUB and 1ZYJ binding sites evidence clear differences in the loop geometries, which makes the 1ZYJ-pose of the 3HUB ligand less than perfect in terms of rmsd (2.76 Å). This situation was observed on several other examples (3PY0, 3HUC, 2XHR). Note that, in all these cases, low rmsd poses do exist. For 3HUC, however, the IFP from the lowest rmsd pose poorly matches the native X-ray IFP

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

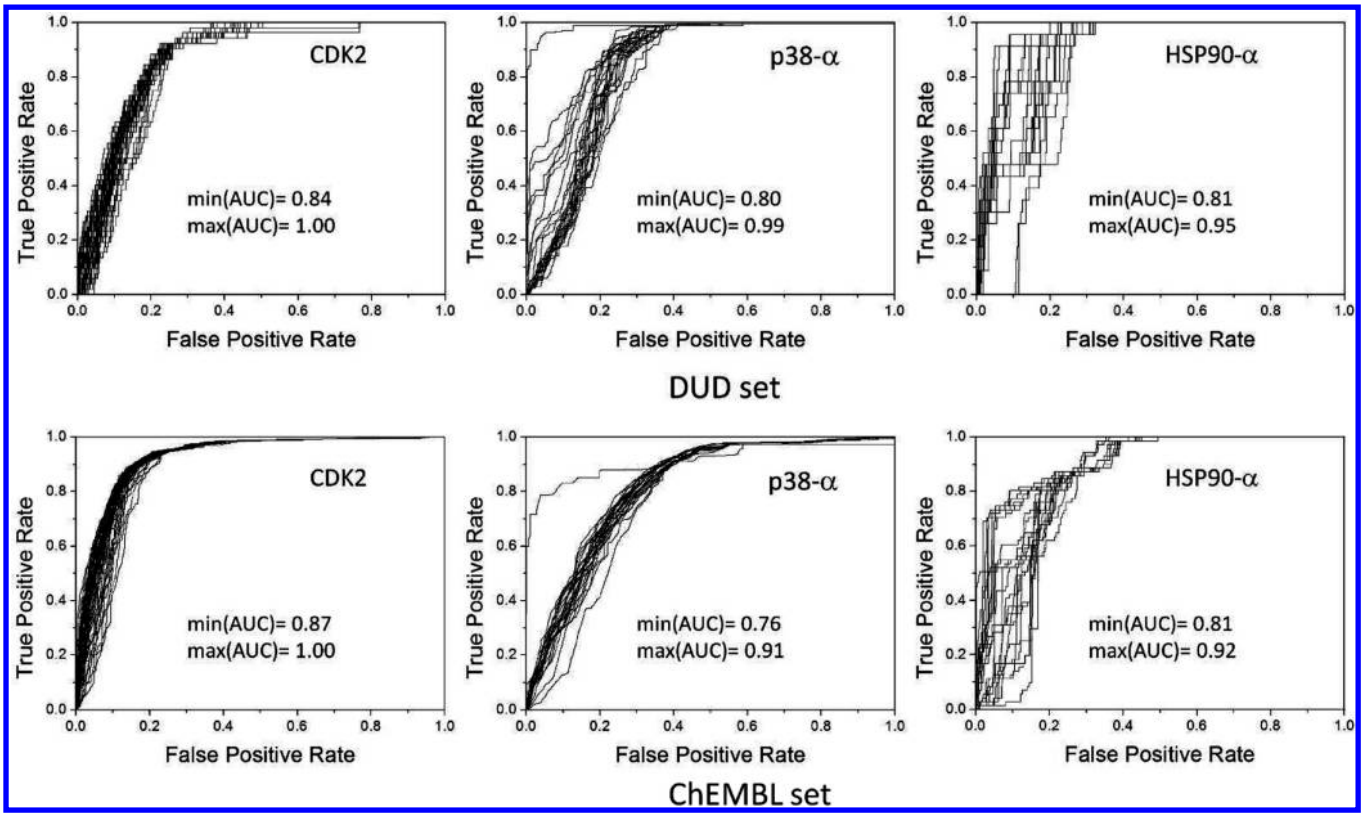


Figure 6. ROC curves for similarity-based virtual screening of DUD (DUD actives, DUD decoys) and ChEMBL (ChEMBL actives, DUD decoys) predicted IFPs against reference X-ray IFPs of the training sc-PDB set (CDK2, 72 complexes; p38-α, 39 complexes; HSP90-α, 35 complexes). For each reference, external IFPs are ranked by decreasing similarity to the reference IFP and a ROC curve for segregating actives from inactives is generated. Minimum and maximum values of ROC AUCs are given as min(AUC) and max(AUC), respectively.

(ligand poses may overlap, but interactions diverge because of moving loops).

The case of HSP90-α ligand from 2XHR is particularly interesting (Figure 9B). The docked IFP is similar to the X-ray IFP (Tc = 0.84) despite a corresponding rmsd value of 5.15 Å. The binding site (2VCJ) in which the docking-based IFP matched most of the native X-ray features has a drastically different conformation of the LGTIAK loop. In spite of the different positioning, the ligand switches the correct interaction bits on but in another active site conformer. This illustrates a possible advantage of predicted IFPs, because they might help to identify the correct binding interactions without the need to highlight the correct pose, all in bypassing the huge challenges linked to flexible active site modeling in docking simulations.

Failure Due to Pharmacophoric Degeneracy. In this scenario, the retrieved pose highlights the same pattern of interactions, but the position of the ligand in the binding site is different from that in the native X-ray structure. The pharmacophore patterns of those ligands are 'degenerate' (usually symmetric) in the sense that, for example, the ligand might display an alternative head-to-tail binding mode, in which all the native contacts are preserved (being fulfilled by alternative/symmetric groups of the ligand). These cases raise the legitimate question whether such alternative binding modes may exist but have been overlooked when fitting the ligand into the electron density map. For instance, the docked and X-ray poses of ligand 3ITZ to p38-α are shown in the Figure 9C. These two conformations form similar interactions (Tc = 0.70), but the orientations in the binding site are totally different (rmsd = 10.05 Å). The pseudosymmetric interaction patterns of p38-α might allow for alternative binding modes or blur the docking. As a matter of fact, the crystal structure 3ITZ is very well-defined, and there are no doubts about the actual position of the ligand. Therefore, it is likely that the method just failed in this case.

Pharmacophoric degeneracy is the more likely, the more unspecific (hydrophobe-dominated) the interaction pattern is. For example, both poses of ligand 3IW6 to p38-α (Figure 9D) form very similar hydrophobic interactions in spite of a large rmsd (9.64 Å). Here again, the 3IW6 structure is refined enough to suspect that this unspecific hydrophobic pattern is misleading the docking. Further examples of such cases are rare upon docking to CDK2 (3LE6) and p38-α (3ITZ, 3IW6) since the corresponding binding sites are polar enough but very frequent when docking to the hydrophobic HSP90-α binding site (2XHT, 2XHX, 2XJJ, 3HEK, 3HYZ, 3K97, 3K98, 3K99, 3MNR; Figure 8 E,F).

Intrinsic Structural Overdetermination According to the rmsd Criterion. In this situation, the interactions generated from protein-buried ligand atoms are well predicted and the rmsd of this substructure to the X-ray pose is low. However, the extreme difficulty in correctly positioning the solvent accessible part of the ligand yields to an overall bad rmsd to the X-ray solution. For example, the rmsd of the docked pose for ligand 2XJX to HSP90-α (Figure 9E) is high (rmsd = 4.39 Å), because the solvent-accessible piperazinyl fragment does not contribute much to binding while the isoindoline group contribute by hydrophobic and nondirectional contacts only —thus different binding modes of these fragments are difficult to discriminate based on IFPs. The buried resorcinol group is nevertheless very

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

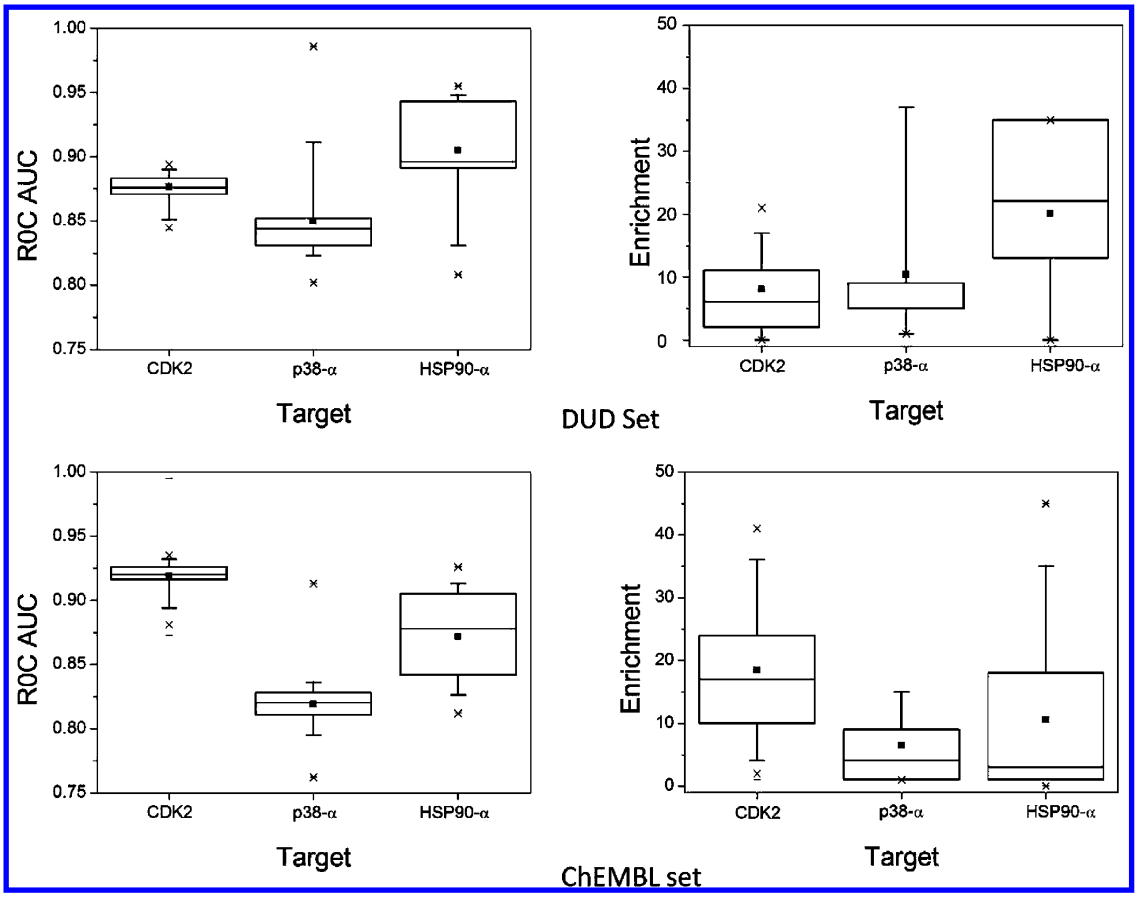


Figure 7. Box-and-whisker plot of the distribution of ROC area under the curve values (left panels) and of the enrichment in true positives at a 1% false positive rate (right panels) for similarity-based virtual screening of DUD (DUD actives, DUD decoys) and ChEMBL (ChEMBL actives, DUD decoys) predicted IFPs against reference X-ray IFPs of the training sc-PDB set (CDK2, 72 complexes; p38-α, 39 complexes; HSP90-α, 35 complexes). The box delimits the 25th and 75th percentiles; the whiskers delimit the 5th and 95th percentiles. Median and mean values are indicated by a horizontal line and an empty square. Crosses delimit the 1st and 99th percentiles, respectively.

Table 2. Comparative Ability (Averaged Area under the ROC Curve) of Predicted IFPs and ISIDA Ligand Descriptors in Discriminating Actives from Decoys in Virtual Screening

	target					
	CDK2		p38-α		HSP90-α	
descriptor	IFP	ISIDA	IFP	ISIDA	IFP	ISIDA
DUD set[a]	0.87	0.85	0.85	0.93	0.90	0.87
ChEMBL set[b]	0.91	0.85	0.82	0.96	0.87	0.90

[a]DUD actives, DUD decoys. [b]ChEMBL actives, DUD decoys.

Table 3. Success Rate of Cross-Docking Experiments According to the rmsd Metric[a]

	GoldScore	TC_GS
CDK2	38% (5/13)	54% (7/13)
p38-α	50% (7/14)	36% (5/14)
HSP90-α	27% (6/22)	13% (3/22)

[a]A success means that the rmsd of the selected docking pose to the X-ray solution is lower than 2.0 Å. Docking poses are selected according the highest GoldScore (GS) or the highest similarity to the predicted IFP (TC_GS).

Table 4. Success Rate of Cross-Docking Experiments According to the IFP Similarity Metric[a]

	GS	TC_GS
CDK2	62% (8/13)	85% (11/13)
p38-α	71% (10/14)	71% (10/14)
HSP90-α	50% (11/22)	68% (15/22)

[a]A success means that the Tanimoto coefficient between the selected pose and the X-ray solution is higher than 0.6. Docking poses are selected according the highest GoldScore(GS) or highest similarity to the predicted IFP (TC_GS).

well positioned, indicating a good reliability of the predicted IFP to guide docking for that substructure. Since that part strongly interacts with the binding site, docked and X-ray IFPs remain globally similar (Tc = 0.92). Docking partially buried ligands to open cavities like that in HSP90-α is a notoriously difficult exercise (see also docking of ligands 2XDK, 3HZ5, 2XAB to HSP90-α; Figure 8E,F), and the failure cannot reasonably be attributed to the IFP prediction model.

Wrong IFP Predictions. We previously demonstrated that rarely occurring data (interaction between a ligand fragment and a binding site residue; recall Figure 5) cannot be accurately predicted by any machine learning model. Therefore, if the predicted IFP for a particular protein−ligand complex is wrong (Tc <0.6), it is logical that IFP-guide docking will also fail in delivering reliable poses. In our current external set, this situation occurs for several complexes (CDK2: 2WPA, 3IGG,

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

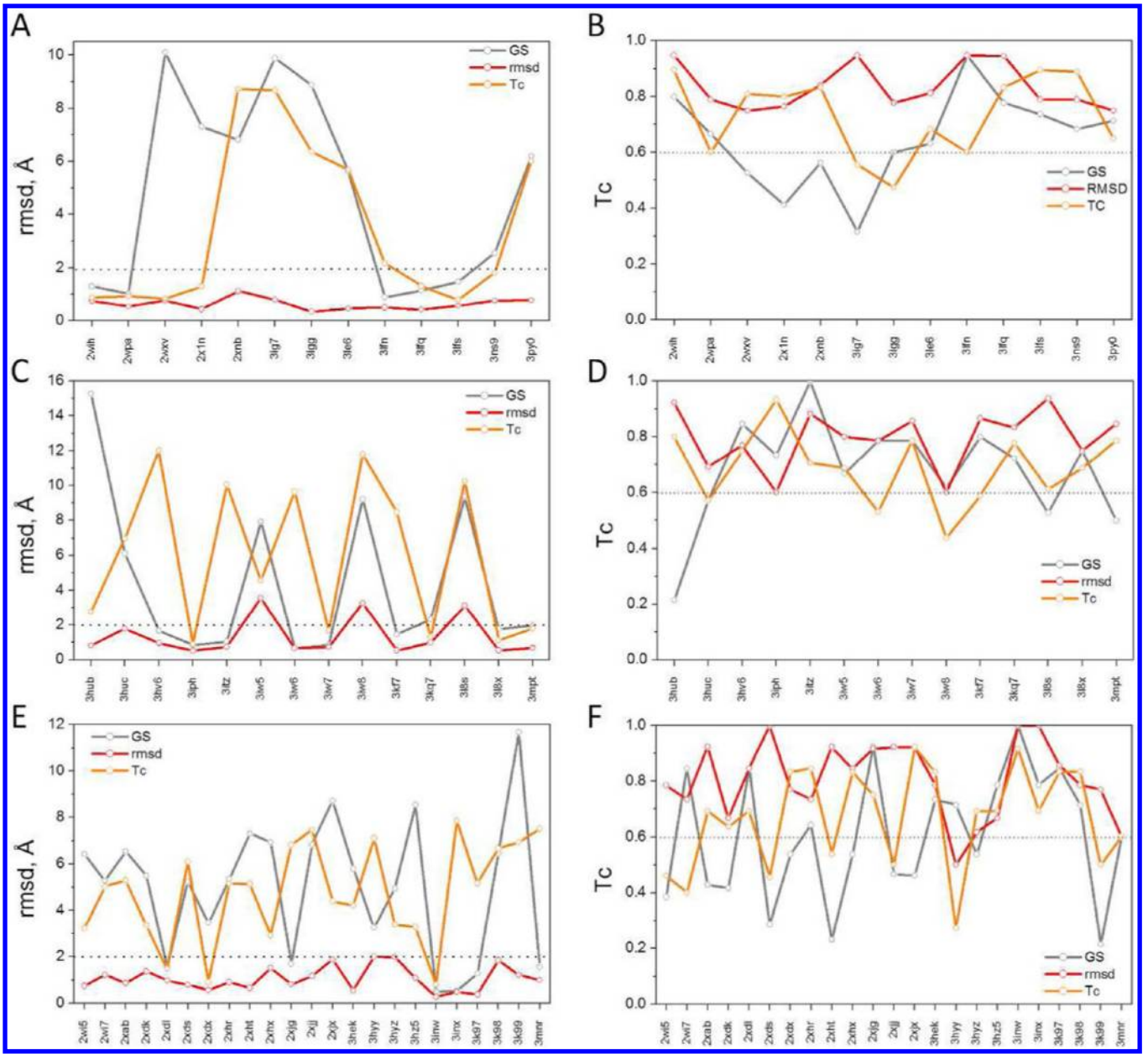


Figure 8. Root Mean Square Deviation (rmsd; panels A, C, E) and IFP similarity (docked vs X-ray pose; B, D, F) of docking poses for ligands of the external validation set for CDK2 (A, B), p38-α (C, D), and HSP90-α (E, F). A predicted pose is selected according to three different ranking strategies: (i) highest similarity to predicted IFP (TC, orange line), (ii) highest GoldScore (GS, gray line), and (iii) lowest rmsd (rmsd; red line). PDB identifiers refer to that of the crossdocked ligands.

3PY0; p38-α: 3IW7, 3IW8, 3L8X; HSP90-α: 2WI7, 2XDS, 3HYY).

Last, not considering protein-bound water molecules in the learning phase will generate IFPs with crucial missing information and erroneous docking. As an example, we highlight here the difficulty in docking ligand 2WI5 to HSP90-α (Figure 9F). The IFP-guided docking suggests a solution with low Tc (0.46) and high rmsd (3.23) to the X-ray pose because three out of the six hydrogen-bonds engaged by the ligand to the target are mediated by a tightly connected network of water molecules.[33]

CONCLUSION

This contribution presents the proof of concept that a machine learning algorithm (here an artificial neural network) can successfully be applied to predict protein−ligand binding modes without explicitly modeling 3D interactions between both partners. The machine learning model is trained on protein−ligand complexes of known 3D structure for which ligands are encoded by fragment descriptors and protein−ligand interactions by binary interaction fingerprints. It is able to recapitulate and next predict the most important interactions in a predicted interaction fingerprint between trained targets and ligands not present in the training set but sharing the same binding site with trained ligands. The method was validated by virtual screening experiments (distinguishing true actives from decoys) and postprocessing docking poses according to predicted IFPs. The models were not yet associated with a relevant applicability domain, taking into account the chemistry and the interactions observed at the modeling stage.

dx.doi.org/10.1021/ci300200r | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

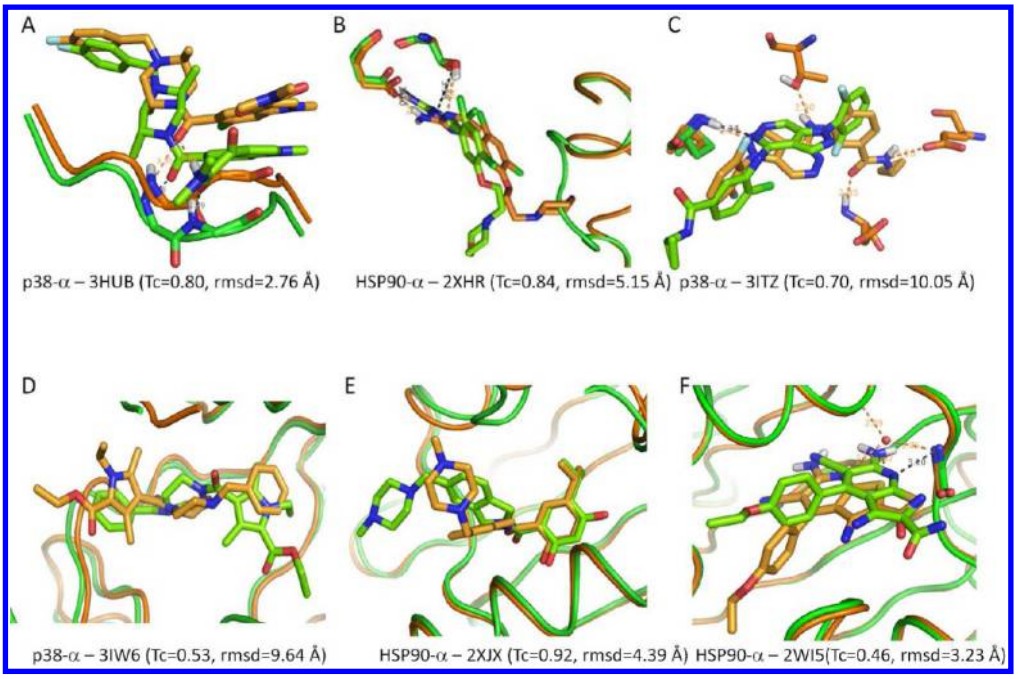


Figure 9. Predicted IFP-guided docking examples (green) matched to X-ray solutions (orange). Ligand nitrogen and oxygen atoms are colored in blue and red, respectively. Carbons atoms are colored in green (predicted pose) and orange (X-ray pose). Proteins are displayed by tubes (green, selected conformer upon cross-docking; orange, native X-ray structure). Important hydrogen bonds are represented by dotted lines (green, predicted pose: orange, X-ray pose) along with the corresponding heavy atom distance. PDB identifiers refer to that of the crossdocked ligands Tc values (docked to X-ray IFPs), and rmsd (docked to X-ray) are given in commas.

A current limitation of the method is the necessity to output a target-dependent fixed-length IFP which forces to define a consensus binding site (unique definition of cavity-lining residues) that may not capture ligand-specific molecular interactions. Target-independent fixed-length interaction fingerprints[34,35] could be an interesting workaround to enable a much easier setup for training and a larger applicability domain. The herein developed IFP prediction method is not limited to neural networks. A simple alternative would be, for a new compound, to get a consensus of the IFP of the k nearest neighbors in the training set. Kernel methods[36] or logistic regressions with PLS might also be applied. We herewith demonstrated the proof-of-concept that predicting molecular interactions from simple ligand descriptors is feasible indeed. To be valuable in a practical drug discovery process, the method has to be applicable to any potential protein of known 3D structure. The tedious part of the method is the manual preparation of consensus binding sites enabling the comparison of fixed-length IFPs for a set of ligands. Current work is ongoing to remove this constraint by defining target-independent interaction fingerprints.[35] Predicting IFPs to guide docking present many advantages: simplicity, speed, and robustness toward conformational and energy indeterminations. It should help to improve the reliability of conventional docking similarly to pharmacophores. Predicted IFPs would also be useful in order to early select compounds able to build key interactions with a target of interest.

■ ASSOCIATED CONTENT

Supporting Information

List of the PDB IDs used for modeling of the IFPs; list of the PDB IDs used for validation of the predicted IFP; list of consensus amino acids used for IFP generation; classification of sc-PDB training ligands by chemotypes; distribution of compounds by class for the three training sets; diversity analysis of the compounds used for building and testing IFP models; self-similarity plot of sc-PDB training compounds; chemical similarity between DUD actives and sc-PDB training compounds; chemical similarity between ChEMBL actives and sc-PDB training compounds; chemical similarity between DUD actives and DUD decoys; chemical similarity between ChEMBL actives and DUD decoys. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rognan@unistra.fr.

Present Address

[§]Faculty of Physics, M. V. Lomonosov Moscow State University, Moscow, 119991, Russia.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to the International Center for Frontier Research in Chemistry (ci-FRC of Strasbourg) for a postdoctoral grant to W.C. CC-IN2P3 (Villeurbanne) and GENCI (Project x2012075024) are acknowledged for providing computational resources.

■ REFERENCES

(1) Kalyaanamoorthy, S.; Chen, Y. P. Structure-based drug design to augment hit discovery. *Drug Discovery Today* **2011**, *16*, 831−839.

(2) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.

(3) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate

docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7−S26.
(4) Rognan, D. Docking methods for virtual screening: Principles and recent advances. In *Virtual screening: Pinciples, challenges and practical guidelines*; Sotriffer, C., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2011; pp 153−176.
(5) Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W. Docking performance of fragments and druglike compounds. *J. Med. Chem.* **2011**, *54*, 5422−5431.
(6) B-Rao, C.; Subramanian, J.; Sharma, S. D. Managing protein flexibility in docking and its applications. *Drug Discovery Today* **2009**, *14*, 394−400.
(7) Smith, R. D.; Dunbar, J. B., Jr.; Ung, P. M.; Esposito, E. X.; Yang, C. Y.; Wang, S.; Carlson, H. A. CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.* **2011**, *51*, 2115−2131.
(8) Novikov, F. N.; Zeifman, A. A.; Stroganov, O. V.; Stroylov, V. S.; Kulkov, V.; Chilov, G. G. CSAR scoring challenge reveals the need for new concepts in estimating protein-ligand binding affinity. *J. Chem. Inf. Model.* **2011**, *51*, 2090−2096.
(9) Ballester, P. J.; Mitchell, J. B. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* **2010**, *26*, 1169−1175.
(10) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337−344.
(11) Brewerton, S. C. The use of protein-ligand interaction fingerprints in docking. *Curr. Opin. Drug Discovery Dev.* **2008**, *11*, 356−364.
(12) Mpamhanga, C. P.; Chen, B.; McLay, I. M.; Willett, P. Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. *J. Chem. Inf. Model.* **2006**, *46*, 686−698.
(13) Chuaqui, C.; Deng, Z.; Singh, J. Interaction profiles of protein kinase-inhibitor complexes and their application to virtual screening. *J. Med. Chem.* **2005**, *48*, 121−133.
(14) Deng, Z.; Chuaqui, C.; Singh, J. Knowledge-based design of target-focused libraries using protein-ligand interaction constraints. *J. Med. Chem.* **2006**, *49*, 490−500.
(15) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195−207.
(16) Chalopin, M.; Tesse, A.; Martinez, M. C.; Rognan, D.; Arnal, J. F.; Andriantsitohaina, R. Estrogen receptor α as a key target of red wine polyphenols action on the endothelium. *PLoS One* **2010**, *5*, e8554.
(17) Venhorst, J.; Nunez, S.; Terpstra, J. W.; Kruse, C. G. Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. *J. Med. Chem.* **2008**, *51*, 3222−3229.
(18) de Graaf, C.; Rein, C.; Piwnica, D.; Giordanetto, F.; Rognan, D. Structure-based discovery of allosteric modulators of two related class B G-protein-coupled receptors. *ChemMedChem* **2011**, *6*, 2159−2169.
(19) de Graaf, C.; Rognan, D. Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J. Med. Chem.* **2008**, *51*, 4978−4985.
(20) Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: a database for identifying variations and multiplicity of 'druggable' binding sites in proteins. *Bioinformatics* **2011**, *27*, 1324−1326.
(21) Weill, N.; Rognan, D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites. *J. Chem. Inf. Model.* **2010**, *50*, 123−135.
(22) *OE Chem TK*, version 1.7.7; OpenEye Scientific Software: Santa Fe, NM, 2011.
(23) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−801.
(24) *ChemAxon Standardizer*, version 5.4.4.1; ChemAxon: Budapest, Hungary, 2010.
(25) Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V. P. Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 693−703.
(26) Stuttgart Neural Network Simulator. http://www.ra.cs.uni-tuebingen.de/SNNS (accessed March 2013).
(27) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(28) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−2547.
(29) 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.
(30) Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J. Chem. Inf. Model.* **2006**, *46*, 717−727.
(31) Jones, G.; Willett, P.; Glen, R. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.
(32) Shahid, M.; Kasam, V.; Hofmann-Apitius, M. An improved weighted-residue profile based method of using protein−ligand interaction information in increasing hits selection from virtual screening: a study on virtual screening of human GPCR A2A receptor antagonists. *Mol. Inf.* **2010**, *29*, 781−791.
(33) Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer, J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **2009**, *52*, 4794−4809.
(34) Perez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245−1260.
(35) Despahy, J.; Ducrot, P.; Raimbaud, E.; Rognan, D. Encoding protein-ligand interaction patterns in fingerprints and graphs. *J. Chem. Inf. Model.*, DOI: http://dx.doi.org/10.1021/ci300566n.
(36) Evgeniou, T.; Micchelli, C. A.; Pontil, M. Learning multiple tasks with kernel methods. *J. Mach. Learn. Res.* **2005**, *6*, 615−637.
(37) *MOE*, version 2011.10; Chemical Computing Group: Montreal, Canada, 2010.