Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints

Gilles Marcou and Didier Rognan*

Bioinformatics of the Drug, UMR 7175 CNRS-ULP (Université Louis Pasteur- Strasbourg I), 74 route du Rhin, B.P. 24, F-67400 Illkirch, France

Received August 7, 2006

Protein−ligand interaction fingerprints have been used to postprocess docking poses of three ligand data sets: a set of 40 low-molecular-weight compounds from the Protein Data Bank, a collection of 40 scaffolds from pharmaceutically relevant protein ligands, and a database of 19 scaffolds extracted from true cdk2 inhibitors seeded in 2230 scaffold decoys. Four popular docking tools (FlexX, Glide, Gold, and Surflex) were used to generate poses for ligands of the three data sets. In all cases, scoring by the similarity of interaction fingerprints to a given reference was statistically superior to conventional scoring functions in posing low-molecular-weight fragments, predicting protein-bound scaffold coordinates according to the known binding mode of related ligands, and screening a scaffold library to enrich a hit list in true cdk2-targeted scaffolds.

INTRODUCTION

Fragonomics[1,2] is a rapidly evolving drug discovery area enabling the identification of efficient low-molecular-weight hits[3] by either X-ray diffraction[4] or nuclear magnetic resonance spectroscopy.[5] A true advantage of this screening method is that both hit identification and binding mode may be derived simultaneously. By either growing or linking selected fragments, nanomolar leads can then be designed at an incomparable pace.[6] Structural biology screening however requires the setup of a fragment library (compendium of low-molecular-weight compounds) by applying knowledge-based[7] and medicinal chemistry rules.[8] Up to now, these fragment libraries have no particular focus on peculiar targets. Applying a computational protocol to select the most suitable fragments for a given target would thus be highly desirable to hasten the screening process and to open this technology to academia. Molecular docking[9] has now been successfully used for a decade to predict the binding mode of druglike compounds and screen large libraries. However, the well-identified weakness of fast energy-based scoring functions[10] currently limits the systematic use of docking to low-molecular-weight compounds. Very often, one faces the situation where a continuum of poses of comparable binding energy describes the possible interaction of a fragment to its binding site, the scoring function being unable to discriminate near-native from irrelevant poses (Figure 1). Docking under pharmacophoric constraints may slightly improve the situation regarding in silico fragment screening,[11] however still at relatively low accuracy.

To better address this problem, a much stronger emphasis has been put recently on postdocking strategies than on more accurate scoring functions. Docking poses can be challenged for the consistency of topological descriptors (e.g., accessibility of the ligand, holes along the protein−ligand interface, and interaction mismatches)[12] by consensus approaches involving either multiple scoring functions,[13,14] docking tools,[15] or alternative protein coordinates[16] or by combining docking to additional computational tools (e.g., Bayesian statistics,[17] molecular mechanics,[18] MM-PB GB/SA methods,[19,20] or quantum mechanics[21]). As an indicator of the importance of postdocking issues, several standalone programs (e.g., Silver,[22] Post-Dock,[23] and Viscana[24]) are now dedicated to this task. Efficient postprocessing however requires a considerable amount of pre-existing knowledge.

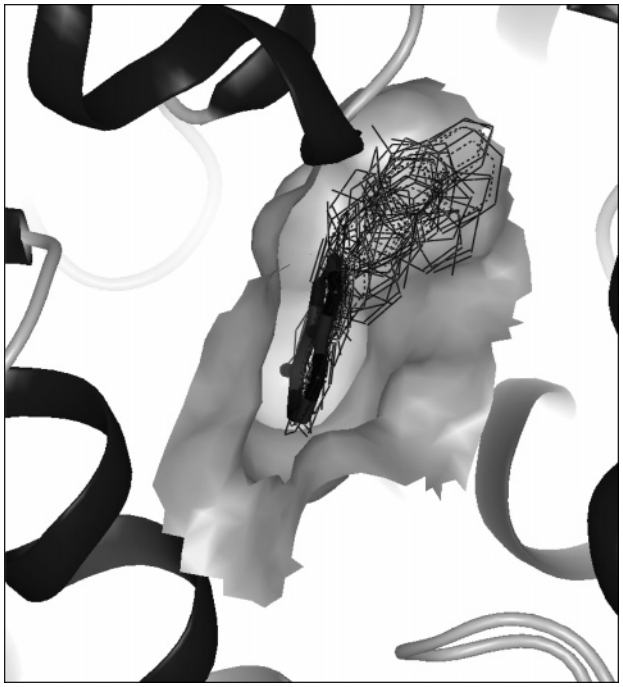

Figure 1. X-ray (sticks) and Glide-predicted poses (gray lines) of a small molecular weight fragment (adenine) to phosphodiesterase 4D. The fragment pocket is delimited by a white surface and X-ray fragment coordinates extracted from adenosine monophosphate (PDB entry 1tb7).

* Corresponding author e-mail: didier.rognan@pharma.u-strasbg.fr.

10.1021/ci600342e CCC: $37.00
Published on Web 12/09/2006

Many postdocking strategies are thus usually target- and/or project-dependent.

In order to rely on simple, robust but efficient postdocking processing, we herewith propose the use of molecular interaction fingerprints (IFP) for prioritizing the most relevant poses for either low-molecular-weight fragments or molecular scaffolds. Practically, IFPs are simple bit strings that convert 3-D information about protein−ligand interactions into simple 1-D bit vector representations that can be quickly compared by the use of traditional metrics (e.g., Tanimoto coefficient or Euclidian distance). Usage of IFPs has been recently pioneered by Biogen Idec.[25] with a particular focus on protein kinases and has previously shown in either residue-based[25] or atom-based implementations[26,27] to outperform conventional scoring functions in predicting correct poses for druglike compounds. Their utility in fragment docking is however still unknown. To assess their utility as postdocking filters, three sets of low-molecular-weight fragments and scaffolds have been docked to their cognate Protein data Bank (PDB) targets using four popular docking tools[28] (FlexX,[29] Gold,[30] Glide,[18] and Surflex[31]). Using a Tanimoto metric (Tc-IFP) measuring the similarity of predicted to X-ray IFP, we unambiguously show that (i) this descriptor is better suited to quantify docking success than the widely used root-mean-square deviation (RMSD) criterion and (ii) scoring poses by decreasing Tc-IFP outperforms conventional scoring functions in discriminating near-native poses from decoys and enriching a hit list in "true active" scaffolds when screening a scaffold library for cdk2 inhibitors.

RESULTS

The first benchmark on 42 low-molecular-weight PDB ligands (Chart 1, Supporting Information) was set up to answer a simple question: *are current docking tools suitable to pose low-molecular-weight fragments?*

All 42 fragments were docked to their cognate proteins using four docking tools (FlexX, Glide, Gold, and Surflex). Corresponding compounds were prefiltered from the sc-PDB database[32] according to "fragmentlike" filters (see the Computational Methods). As expected and previously reported for "druglike" compound data sets,[14] no relationships could be found between the docking score and the RMSD to the true X-ray pose (data not shown).

When the RMSD to the X-ray pose is compared to the herein introduced Tc-IFP value (see the mathematical definition in the Computational Methods section), better correlations are observed (Figure 2A). As expected, Tc-IFP increases when RMSD decreases. Plotting average RMSD values for poses within a defined Tc-IFP interval suggests that a Tc-IFP threshold of 0.60 corresponds to the widely used RMSD threshold of 2.0 Å for quantifying docking success.[29,30] However, the RMSD to Tc-IFP dependency varies across docking tools (Figure 2B). FlexX, Gold, and Surflex poses exhibit moderate relationships between RMSD and Tc-IFP ($0.70 < |r| < 0.77$ and $420 < n < 1260$), which is not the case of Glide, for which the relationship is poor ($|r| = 0.57$ and $n = 974$). The higher tendency of FlexX to generate poses with Tc-IFP values of 0 is noticeable and corresponds to the well-documented tendency of this docking tool to generate poses at the periphery of the binding site.[28] When the above-defined thresholds were used, poses could be classified in four categories assuming that the attribute "*positive*" describes poses with a Tc-IFP value above 0.6 and that the attribute "*true*" describes poses for which RMSD and Tc-IFP agree: true positives (RMSD ≤ 2 Å and Tc-IFP ≥ 0.6), true negatives (RMSD >2 Å and Tc-IFP < 0.6), false positives (RMSD > 2 Å and Tc-IFP ≥ 0.6), and false negatives (RMSD ≤ 2 Å and Tc-IFP < 0.6). Whereas true positives and negatives describe the majority of poses, there is still a significant number of false negatives and, to a much lesser extent, some false positives from all docking programs (Figure 2B and Table 1). The percentages of false positives (7.5% on average) and false negatives (ca. 12%) are relatively stable across the various docking tools used herein (Table 1). From a pure statistical point of view, Gold and Surflex generate the best set of poses (highest proportion of true positives) and consequently the lowest number of true negatives (Table 1). Whereas true positive and true negative poses are straightforward to analyze, false poses require more attention. An exhaustive analysis of the latter poses indicate that false positives are usually planar molecules with symmetry or pseudo-symmetry axes exhibiting high RMSD values whereas most interactions to the target are conserved (e.g., Figure 3A). Conversely, a slight translation of a predicted pose with respect to the true X-ray pose can result in false positives where a partial or complete loss of protein−ligand interactions (e.g., Figure 3B) does not much affect the RMSD value. From here on, the success of docking will thus be quantified by using only the Tc-IFP metric that faithfully reproduce observed molecular interactions in the reference X-ray structures. For the present data set, we considered a docking to be successful if at least 60% of the intermolecular interactions observed in the X-ray structure (Tc-IFP > 0.6) are reproduced by the docking program. As previously noticed (Figure 2A), this threshold roughly corresponds to a 2.0 Å RMSD cutoff.

Whatever the docking engine, using the docking score as a criterion to rank low-molecular-weight fragments is not satisfactory because only modest success rates can be reported (from 40% for FlexX to 70% for Glide; Figure 4). This did not imply that satisfactory solutions were not generated in the set of predicted poses. Hence, when the Tc-IFP scoring function (in the present case, the Tc-IFP of each pose is computed with respect to the X-ray pose) is used, much higher success rates are observed (from 70% for FlexX up to 95% for Glide; Figure 4). As expected, picking the lowest RMSD solution always performs worse than selecting the best Tc-IFP pose, confirming that RMSD is not the best possible criterion for selecting the best pose available. Although the scope of the study is not to compare docking tools, we can notice that Glide is remarkably well-suited to dock small fragments despite the moderate utility of its internal Glidescore scoring function.

The second benchmark was aimed at answering the following question: *given the known binding mode of a molecular scaffold, is it possible to successfully dock bioisosteric scaffolds sharing the same target and binding site*?

We wanted to investigate scaffold docking in a scaffold-hopping context, which means predicting the binding mode of a given scaffold A from the known binding mode of a bioisosteric scaffold B. Therefore, we needed several targets

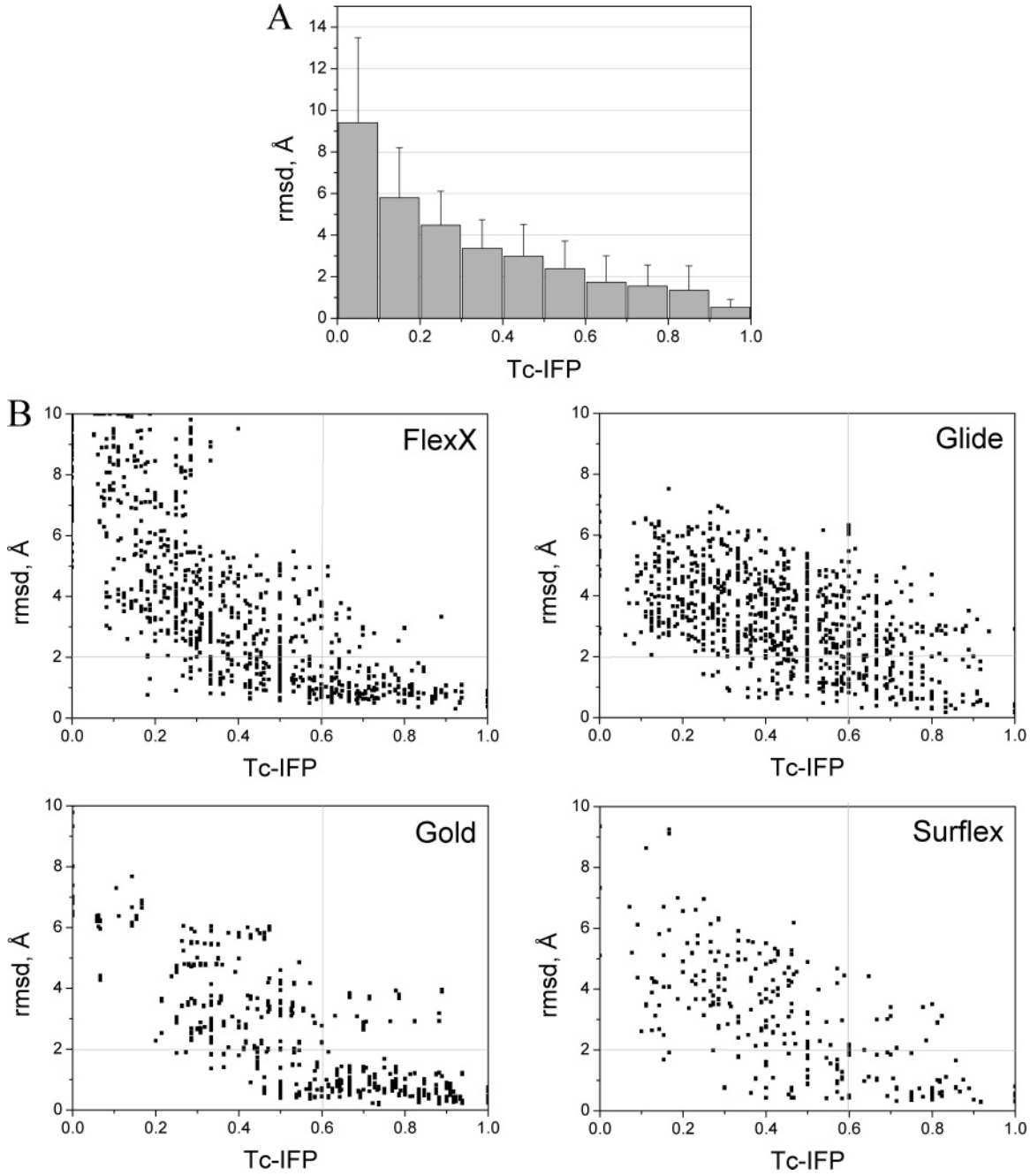

Figure 2. Plotting RMS deviations from the X-ray pose versus similarity of interaction fingerprints Tc-IFP (expressed by a Tanimoto coefficient calculated from IFPs generated by the X-ray and the predicted pose) for a set of 42 low-molecular PDB ligands (data set 1). (A) Average RMSD and standard deviation values (error bars) for increasing Tc-IFP intervals (data collected for a total of 3924 poses from four docking tools). (B) Docking program dependency of RMSD to Tc-IFP. Poses located in the upper-right part are considered as false positives (high RMSD, high Tc-IFP), and poses located in the lower-left part are considered as false negatives (low RMSD, low Tc-IFP). Poses in the upper-left part are true negatives (high RMSD, low Tc-IFP), whereas poses in the lower-right part are true positives (low RMSD, high Tc-IFP).

of pharmaceutical interest for which there is enough scaffold diversity among cocrystallized PDB ligands. Looking at the in-house sc-PDB database,[32] which is a subset of the PDB for which only proteins with druggable binding sites are stored, we could find 10 targets satisfying these conditions (neuraminidase, estrogen receptor α, protein tyrosine phosphatase 1B, phosphodiesterase 4D, tryptophan synthase, ionotropic glutamate receptor type 2, cell division protein kinase 2, carbonic anhydrase type 2, thrombin, and cyclooxygenase type 2). Different scaffolds were considered for each of 10 therapeutically relevant human targets from known sc-PDB ligands.[32] For each target, heterocyclic scaffolds representing roughly 50% of the corresponding full compounds, sharing the same location in the binding site, and exhibiting a wide array of intermolecular interactions were manually selected. In order to design a data set of similar size to the previous database of low-molecular-weight fragments, four different scaffolds were thus chosen for each target to yield a final data set of 40 scaffolds (Chart 2, Supporting Information).

To closely mimic real-life situations, each pose for a particular scaffold was fingerprinted and cross-scored with

Table 1. Classification of Fragment Poses for Four Docking Programs

	FlexX	Glide	Gold	Surflex
TP[a]	16	13	40	30
FP[b]	2	12	9	7
FN[c]	14	9	11	14
TN[d]	68	66	40	49

[a] True positives (RMSD ≤ 2 Å and Tc-IFP ≥ 0.6). [b] False positives (RMSD > 2 Å and Tc-IFP ≥ 0.6). [c] False negatives (RMSD ≤ 2 Å and Tc-IFP < 0.6). [d] True negatives (RMSD > 2 Å and Tc-IFP < 0.6).

respect to the fingerprints generated from the three other scaffolds describing ligands of the same target. The best of these three scores was then saved for each entry and used to assess docking success (Tc-IFP ≥ 0.6) or a lack thereof.

Again, scoring by interaction fingerprint similarity always outperforms energy-based scoring functions for all docking programs used herein (Figure 5). Although some differences might be observed among the different docking tools, the general trend is conserved. The success rate rises from ca. 60% for conventional scoring functions to ca. 85% when using Tc-IFP as a scoring metric. Furthermore, scoring by similarity of interaction fingerprints selects significantly better poses than a simple topological RMSD criterion would do (Figure 5). Despite the low number of chemotypes investigated for each target and the limited value of statistics derived thereof, a decomposition of success rate by target (Figure 6) clearly shows that postprocessing docking poses by Tc-IFP has a considerable value for difficult targets (e.g., ionotropic glutamate receptor type 2 for FlexX, Glide, and Surflex; human carbonic anhydrase, neuraminidase, and thrombin for Glide; COX-2 for Gold; and neuraminidase and phosphodiesterase 4D for Surflex). For easier test cases (e.g., cdk2, estrogen receptor α), conventional scoring by energy performs relatively well and the added-value of IFPs is marginal.

The third benchmark was even more stringent and set up to check *whether scaffolds could be efficiently prioritized by a structure-based virtual screening approach*. A library of 2249 scaffolds (data set 3, see Computational Methods) containing 19 true actives (scaffolds manually extracted from true cdk2 inhibitors; Supporting Information Chart 3) was set up and docked to one X-ray structure of human cdk2 (PDB entry 1dm2) using the Gold3.1 program. Various strategies were used to rank scaffolds. Two of them (Goldscore and Chemscore) are energy-based scoring functions typically used with Gold; the third one is the Tc-IFP metric (using as reference a single IFP derived from each of the 19 true scaffolds); the fourth one is a Laplacian-modified Bayesian classifier[33] using multiple references (all 19 IFPs from true actives, see Computational Methods). The area under the curve (AUC) of receiver-operating characteristic (ROC) plots[34] derived from the four scoring lists was used as an indicator of both the sensitivity (how many true positives are recovered) and the specificity (how many false positives are recovered) of the scoring method (Figure 7). Both energy-based scoring functions (Goldscore and Chemscore) performed worse than random picking (ROC AUC < 0.5). Interestingly, Tc-IFP scoring dramatically improves the in silico screening accuracy whatever IFP is used as a reference (Figure 7). However, the AUC value considerably varies with the reference IFP from 0.5 (e.g., 1di8 and 1jvp) to 0.78 (1e1x). Among the 19 IFP references, five can be considered as good (AUC > 0.7) and 11 as fair (0.6 > AUC > 0.7). Last, the two-class Bayesian categorization model offers the best possible performance with an AUC value of 0.78. When the corresponding ROC plots (Figure 8) are compared, the Bayesian model is by far the most adequate screening strategy. It allows the recovery of 14 out of the 19 true scaffolds (74%) while selecting only 25% of the random decoys (557 scaffolds).

DISCUSSION

The contribution of chemoinformatics to fragonomics has up to now been limited to the design of appropriate fragment libraries using various knowledge-based principles.[7,8] Prioritization of the most interesting fragments using a structure-based approach has been hampered by the limited accuracy of fast empirical scoring functions which are usually unable to discriminate near-native fragment poses from decoys (Figure 1). We therefore propose an alternative nonenergetical approach consisting first in converting protein−ligand interactions in a bit vector called an interaction fingerprint and second in comparing the predicted IFP to a given reference. The interaction fingerprint concept (SIFt: structural interaction fingerprint) has recently been exemplified at Biogen Idec.[35] in a series of elegant studies focused on protein kinases and has shown several promising features: (i) enhancing the quality of pose prediction in docking experiments,[25] (ii) clustering protein−ligand interactions for a panel of related inhibitors according to the diversity of their interaction with a target subfamily,[36] and (iii) assisting target-biased library design.[37] Herewith, we applied IFPs to fragment docking with the hope of bypassing classical problems associated with energy-based scoring functions. IFPs are computed on the fly from 3-D coordinates of both the protein and the ligand using an in-house-developed algorithm utilizing OpenEye's OEChem library[38] and a set of topological rules (Tables 2−4) for defining a panel of eight possible interactions (H-bond, weak H-bond, ionic, hydrophobic, face-to-face aromatic, edge-to-face aromatic, π-cation, and metal complexation). Our IFP implementation varies significantly from that of Biogen because we do not distinguish main-chain from side-chain protein atoms in generating the residue-based fingerprint (this feature was specifically designed at Biogen for protein-kinase inhibitors because interaction with the main-chain atoms of the hinge region is a crucial determinant of ATP-competitive inhibitors[39]). Furthermore, we allow more interaction types than the original SIFt approach (Tables 2 and 3).

Three different data sets have been used to benchmark the use of IFP in low-molecular-weight compound docking. The first one (data set 1) is a compendium of 42 fragments of known protein-bound X-ray coordinates from the sc-PDB database,[32] which comprises structural information about protein, active-site, and cognate PDB ligands which overcome a series of intensive filters (resolution limit, druggability of the binding site, druglikeness of the ligand, etc.)

Because we did not want to bias our approach toward a particular docking tool, four among the best docking programs (FlexX, Glide, Gold, and Surflex)[28,40,41] have been used to generate a set of poses for each ligand. A first aspect

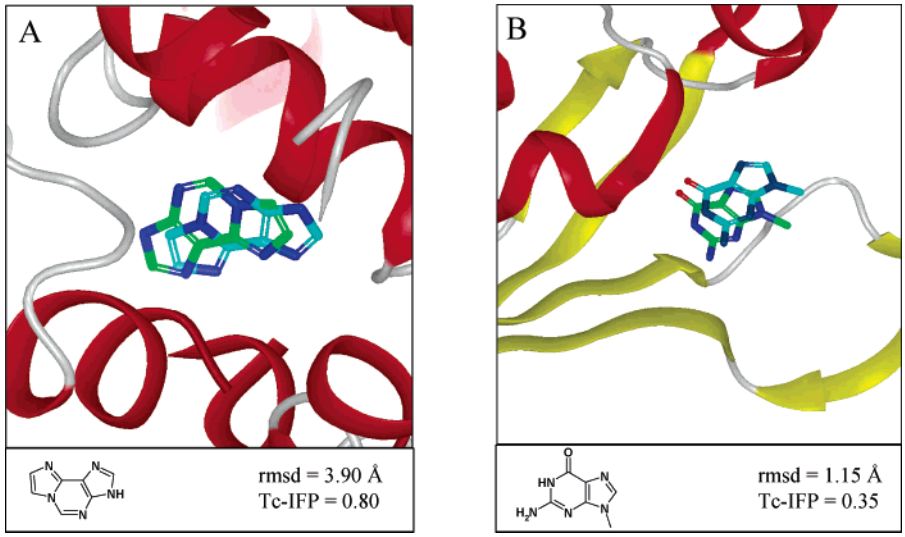

Figure 3. Prototypical examples of mismatches between RMSD and Tc-IFP, for two low-molecular-weight ligands (A: 3H-imidazo[2,1-I]purine to 3-methyladenine DNA glycosylase, PDB entry 1pu8; B: 9-methylguanine to dihydroneopterin aldolase, PDB entry 1rrw). The bound ligand is shown as sticks (cyan carbon, X-ray pose; green carbon, predicted pose).

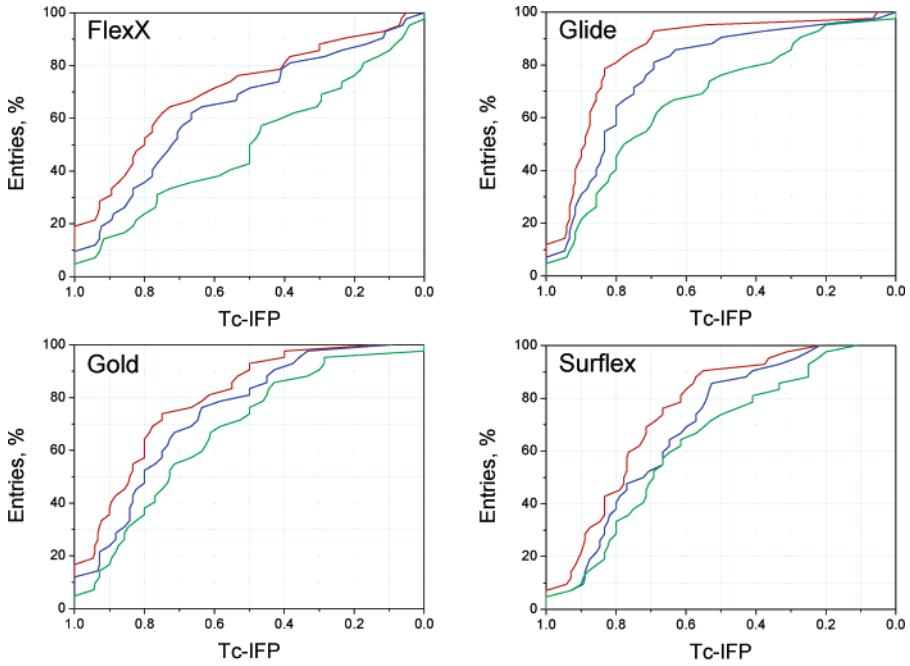

Figure 4. Posing accuracy (expressed by Tc-IFP) of 42 low-molecular-weight ligands (data set 1) obtained with FlexX, Glide, Gold, and Surflex. For each complex, three poses are stored: the best energy pose (top-ranked by the scoring function, green lines), the best Tc-IFP pose (highest IFP similarity to the X-ray pose, red lines), and the best RMSD pose (closest RMSD to the X-ray pose, blue lines).

of the present work was to use a proper metric to estimate docking success. Usually, a RMSD threshold of 2 Å has been used as an indicator of pose accuracy. We felt that this global metric was not really suited to the basic aim of our study because it focuses on ligand coordinates only and thus loses information about the kind of intermolecular interactions which have been reproduced or not (whether a particular residue-based bit string is recovered or not). Using the set of 3924 poses generated for data set 1, we wished first to look at relationships between computed RMSD and the similarity of IFPs (Tc-IFP) using in both cases the X-ray pose as a reference. As expected, there is an overall relationship between RMSD and Tc-IFP (Figure 2A,B). In many cases, a low RMSD correlates with a high Tc-IFP and vice versa. However, there is a significant number of cases for which the RMSD value is misleading. It is either much too high (false positives, Figure 3A) or much too low (false negatives, Figure 3B) with regard to the key protein−ligand interaction which is really reproduced. Whereas the Tc-IFP metric is clearly better than the RMSD criterion for handling false positives (7.5% of all poses), the advantage of the herein introduced metric is more questionable for false negatives (12% on average) where a translation of the predicted pose with respect to the X-ray pose is usually seen (Figure 3B). A total of 7% of the false negatives show a RMSD lower than 0.5 Å and cannot be considered as misdocked because energy refinement of the latter poses or rescoring using other scoring functions based on different physicochemical principles may easily transform false negatives into true positives. It is less likely that refining poses with a RMSD higher than

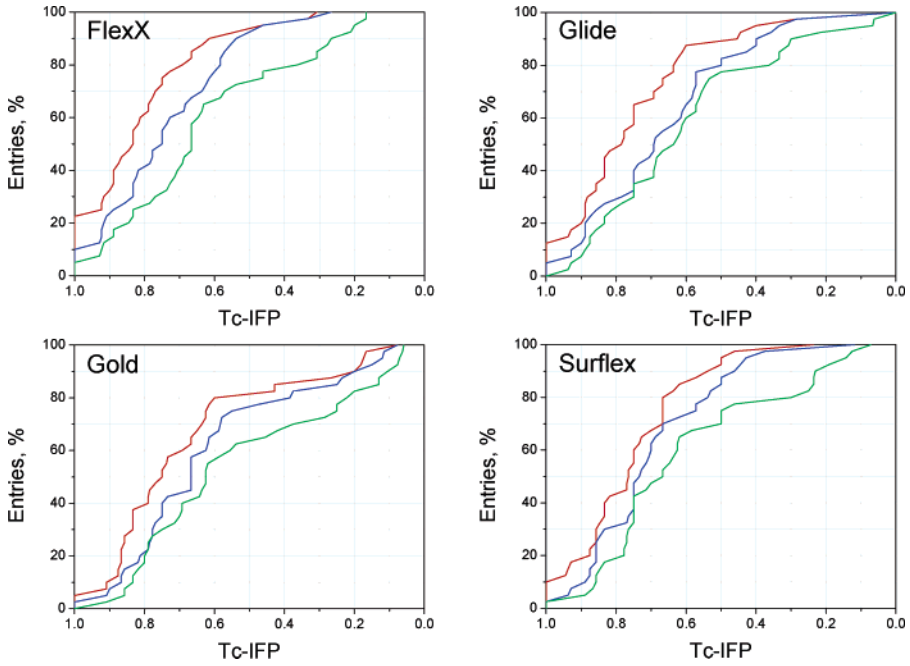

Figure 5. Posing accuracy (expressed by the highest Tc-IFP to any of the three other reference scaffolds) for 40 scaffolds (data set 2) obtained with FlexX, Glide, Gold, and Surflex. For each complex, three poses are stored: the best energy pose (green lines), the best Tc pose (red lines), and the best RMSD pose (blue lines).

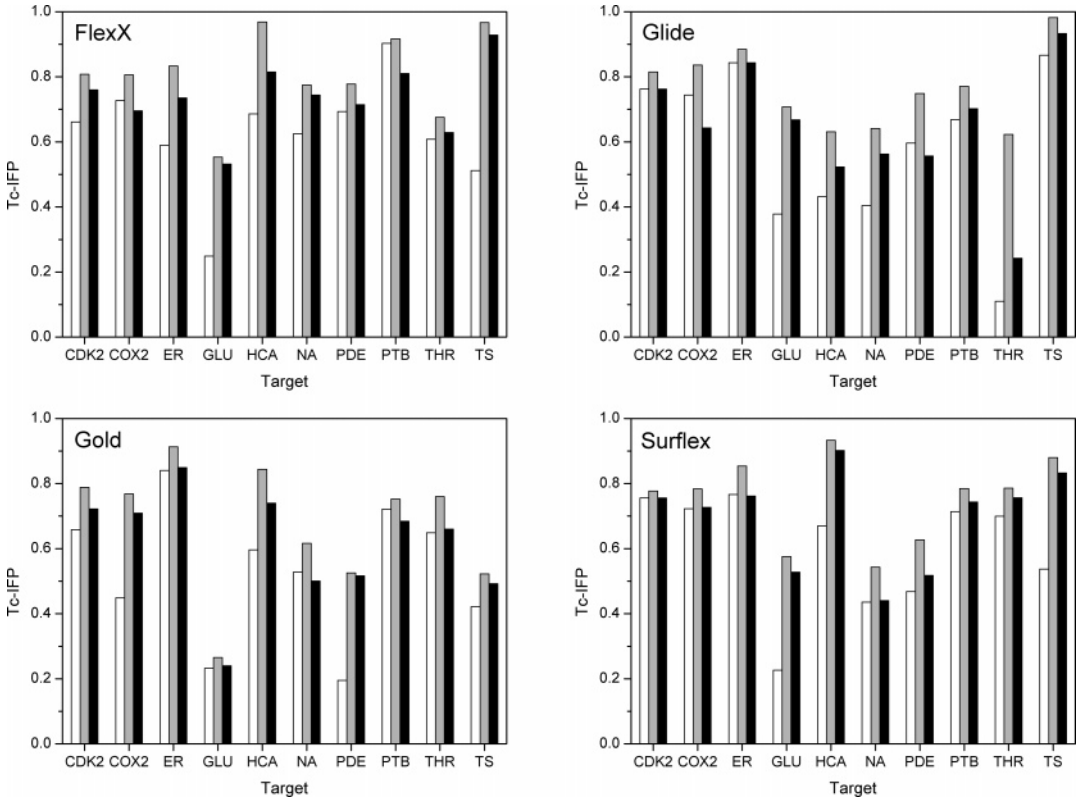

Figure 6. Decomposition of the docking accuracy of 40 scaffolds for 10 pharmaceutical targets. For each complex, three poses are stored: the best energy pose (white bars), the best Tc-IFP pose (gray bars), and the best RMSD pose (back bars).

1 Å would lead to a situation where most of the missing interactions could be recovered. In that case, we believe that the Tc-IFP is more faithful than the RMSD value. This represents the majority (68%) of observed false negatives. For the remaining poses (RMSD between 0.5 and 1 Å), a fuzzier definition of rules used to detect intermolecular interactions could be implemented (e.g., tolerating an H-bond distance up to 4 Å). However, our experience suggests that broadening these rules brings much more noise than the true signal and therefore has not been implemented in our IFPs. Altogether, only 146 poses out of 3924 (3.7%) sharing a RMSD above 1 Å and a Tc-IFP below 0.6 are better described using RMSD than using the Tc-IFP criterion. Altogether, it is therefore strongly advisable to omit the RMSD descriptor to quantify docking success. From here on, the Tc-IFP value will be used only using a 0.6 threshold

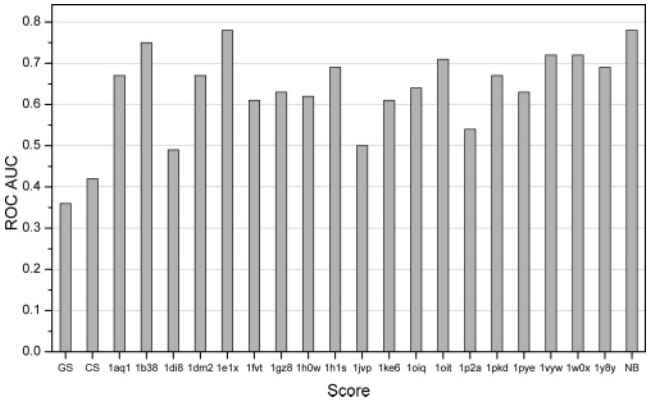

Figure 7. Accuracy of receiver-operating characteristic (ROC) plots expressed by the area under the curve (AUC) for various virtual screening protocols using either energy-based scoring functions (GS, Goldscore; CS, Chemscore) or the Tc-IFP metric with a single IFP (1aq1 to 1y8y) generated from the corresponding PDB structures and a user-defined scaffold definition (Supporting Information Chart 3) or multiple references (NB: Naïve Bayesian model trained on 19 reference IFPs and 2291 decoys).

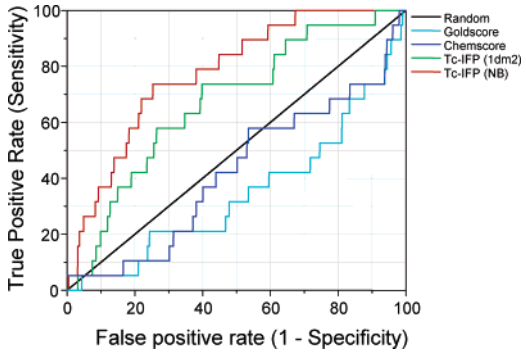

Figure 8. Receiver-operating characteristic (ROC) plot for GOLD docking of 22 249 scaffolds (data set 3) to the X-ray structure of human cdk2 (PDB entry 1dm2) using Tc-IFP similarity score (1dm2, single reference; NB, Bayesian model) and two scoring functions (Goldscore and Chemscore) to discriminate the 19 true active scaffolds from 22 230 decoys.

for quantifying success. It can be seen as a quantitative descriptor for what Kroemer et al. called an interaction-based analysis of contacts in a recent study,[42] coming to the same conclusion.

There is a clear advantage in using Tc-IFP for ranking fragment poses as illustrated by results obtained for data set 1 (Figure 4). Whatever the docking tool, scoring by Tc-IFP is largely superior to energy-based pose ranking. Docking success rises from ca. 40−70% to 70−95% if a Tc-IFP threshold of 0.6 is applied (60% of protein−ligand interactions are recovered in the predicted pose). Although the individual performance of docking tools varies according to the data set and probably to the individual program settings (no particular emphasis was put in fine-tuning program-specific settings for this particular data set), the benefit is independent of the docking program used. This is a very important feature because predicting which docking tool is going to perform well is context-dependent and very difficult.[28] Using a robust docking tool, the user can thus be sure that the Tc-IFP metric will rank docking poses better than a conventional scoring function. Furthermore, comparing RMSD and Tc-IFP rankings clearly shows that the best RMSD poses are definitely not the best to select among the set of proposed solutions. There are still some cases totally unsuitable for fragment docking (Tc-IFP < 0.3) whatever the scoring function. These difficult cases correspond to three main situations (Figure 9): (i) the binding site is highly polar, and different anchors are equally good at positioning the fragment (e.g., 1aj0); (ii) the binding site is almost hydrophobic, and very few polar anchors are available to position the fragment (e.g., 1qy2); (iii) the binding site is largely open to solvent and the fragment partially buried into the protein (e.g., 1rrw, 1tuv, 1yvm, and 3pce).

This first computational experiment suggests that the main problem in fragment docking is usually not sampling the active site space for generating reliable poses (e.g., 95% of Glide poses are correct; Figure 4B) but simply ranking them. Clearly, one cannot exclude misscoring issues for the ca. 20% of wrong answers using FlexX, Gold, or Surflex (Figure 4) where constrained docking using the Tc-IFP metric as the internal scoring function (currently under investigation in our group) is likely to improve the docking accuracy.

When a given set of available poses is used, postprocessing docking outputs by computing IFPs and comparing them to a reference significantly enhance the quality of the first-ranked pose. Of course, it is obvious that self-scoring (scoring a protein−ligand IFP against itself) will outperform scoring by energy as the computational experiment focuses on Tc-IFP values. However, we here wanted to ascertain that, among all predicted poses, some are really close to the experimentally observed solution and that there is no need to constrain the docking algorithm.

Table 2. Atom Flags Used in Determining IFPs[a]

flag	SMARTS definition
donor	[O,N,S][H]
acceptor	[O,N, *−;!+]
cation	[*+]
anion	[*−]
aromatic	[a;R]
hydrophobe	[C,S,F,Cl,Br,I]
weak acceptor	[a:a,A=A,A#A,S+0]
weak donor	[c,CX3,CX2][H]
metal	[Ca,Cd,Co,Cu, Fe,Mg,Mn,Ni,Zn]

[a] (Weak) H-bond donors and acceptors have been determined according to Steiner.[65]

Table 3. Interaction Fingerprint Definition

bit vector position	protein atom flag[a]	ligand atom flag	interaction[b]
1	hydrophobe	hydrophobe	hydrophobic
2	aromatic	aromatic	face-to-face
3	aromatic	aromatic	edge-to-face
4	donor	acceptor	H-bond
5	acceptor	donor	H-bond
6	cation	anion	ionic
7	anion	cation	ionic
8	weak donor	acceptor	weak H-bond
	weak donor	weak acceptor	
	donor	weak acceptor	
9	acceptor	weak donor	weak H-bond
	weak acceptor	weak donor	
	weak acceptor	donor	
10	cation or aromatic	aromatic or cation	π-cation
11	metal	acceptor	metal complexation

[a] See Table 1 for description of atom flags. [b] See Table 3 for description of rules used to identify protein−ligand interactions.

Table 4. Geometric Rules Used in Determining IFPs

interaction	rule 1[a]	rule 2[b]
H-bond	$\overrightarrow{\|DA\|} \leq 3.5\,\AA$	$\left\langle \overrightarrow{DH}, \overrightarrow{HA} \right\rangle \in \left[\frac{-\pi}{4}, \frac{\pi}{4} \right]$
Weak H-bond	$\overrightarrow{\|DA\|} \leq 2.8\,\AA$	$\left\langle \overrightarrow{DH}, \overrightarrow{HA} \right\rangle \in \left[\frac{-\pi}{6}, \frac{\pi}{6} \right]$
Ionic	$\overrightarrow{\|+-\|} \leq 4.0\,\AA$	
Hydrophobe	$\overrightarrow{\|Y_1Y_2\|} \leq 4.5\,\AA$	
Aromatic (Face to face)	$\overrightarrow{\|ac_1ac_2\|} \leq 4.0\,\AA$	$\left\langle \overrightarrow{n_1}, \overrightarrow{n_2} \right\rangle \in \left[\frac{-\pi}{6}, \frac{\pi}{6} \right]$
Aromatic (Edge to face)	$\overrightarrow{\|ac_1ac_2\|} \leq 4.0\,\AA$	$\left\langle \overrightarrow{n_1}, \overrightarrow{n_2} \right\rangle \in \left[\frac{\pi}{6}, \frac{5\pi}{6} \right]$
π-cation	$\overrightarrow{\|ac+\|} \leq 4.0\,\AA$	$\left\langle \vec{n}, \overrightarrow{ac+} \right\rangle \in \left[\frac{-\pi}{6}, \frac{\pi}{6} \right]$
Metal	$\overrightarrow{\|MA\|} \leq 2.8\,\AA$	

[a] *D*, H-bond donor; *A*, H-bond acceptor; +, cation; −, anion; *Y*, hydrophobe; *ac*, geometric center of an aromatic ring; *M*, metal. [b] *H*, hydrogen; *n*, normal to the aromatic ring.

In a second set of experiments, cross-scoring was addressed by selecting 40 scaffolds from pharmaceutically relevant target ligands (data set 2). For each of the 10 selected targets, four different scaffolds were chosen and each scaffold docked into its cognate protein but scored according to the IFP generated by the three others. This computational experiment is therefore closer to a more realistic situation where one wants to prioritize scaffold poses on the basis of the knowledge of bioequivalent scaffolds. We therefore chose scaffolds sharing the same binding site for a particular target. Again, scoring by Tc-IFP outperforms scoring by conventional scoring functions (please note that the standard scoring function of each docking program was used and that no rescoring attempt with external functions was tested). Still, using a 0.6 similarity threshold, docking success increases from ca. 60% to 85% when using the Tc-IFP scoring metric (Figure 5). As for the previous data set of fragments, scoring by Tc-IFP also outperforms scoring by RMSD, which justifies the use of the similarity metric for selecting the appropriate poses. Whereas FlexX was superior in generating appropriate poses for scaffolds than for fragments, the opposite observation is seen for Glide and Gold, Surflex showing very similar accuracy for both data sets (Figures 4 and 5). The noticeable differences in the individual performance of docking tools across both test sets are probably data-set-dependent. More importantly, the same trend is observed: selecting poses ranked by the Tc-IFP metric is the most reliable strategy. This assumption is particularly justified for difficult targets (e.g., ionotropic glutamate receptor type 2 and phosphodiesterase 4D) where energy-based scoring functions usually fail. The binding cleft of the ionotropic glutamate receptor type 2 is lined by polar residues plus two important water molecules, which renders the docking of the corresponding scaffolds very difficult because many orientations can fulfill H-bonding requirements. For phosphodiesterase ligands, docking algorithms tend to locate polar groups of the scaffolds toward the metal binding site and not the Gln369 pocket. It is therefore logical that significant failures occur for the latter active sites.

In a last series of computational experiments, a structure-based virtual screening was undertaken to retrieve scaffolds likely to bind to the human cdk2 target. This enzyme was chosen as the test set for several reasons: (i) it seems to be an adequate target for scaffold docking (recall Figure 6); (ii) it has previously been used for studying the virtual screening accuracy of low-molecular-weight inhibitors;[11] (iii) it has been cocrystallized with various druglike ATP-competitive inhibitors sharing the same binding site,[43] thus offering the possibility to study the influence of the reference IFP in the Tc-IFP scoring. Gold was used as the docking tool for this virtual screening experiment to compare our results with the previously published virtual screening data on low-molecular-weight cdk2 inhibitors which was also realized using Gold.[11] As expected from this latter study,[11] energetic scoring functions are not suited to distinguish "true cdk2" scaffolds from decoys (Figures 7 and 8). The performance of Chemscore was close to that of random picking, whereas that of Goldscore was even worse. This did not mean that good poses have not been generated because scoring them by similarity of their IFP to that of a given reference gave much better results (Figures 7 and 8). However, using the Tc-IFP metric requires first the selection of an IFP reference which might not be an easy task. Hence, human cdk2 has been cocrystallized with various inhibitors all binding to the ATP binding site and exhibiting different chemotypes.[43] One may thus use several X-ray structures for generating the reference

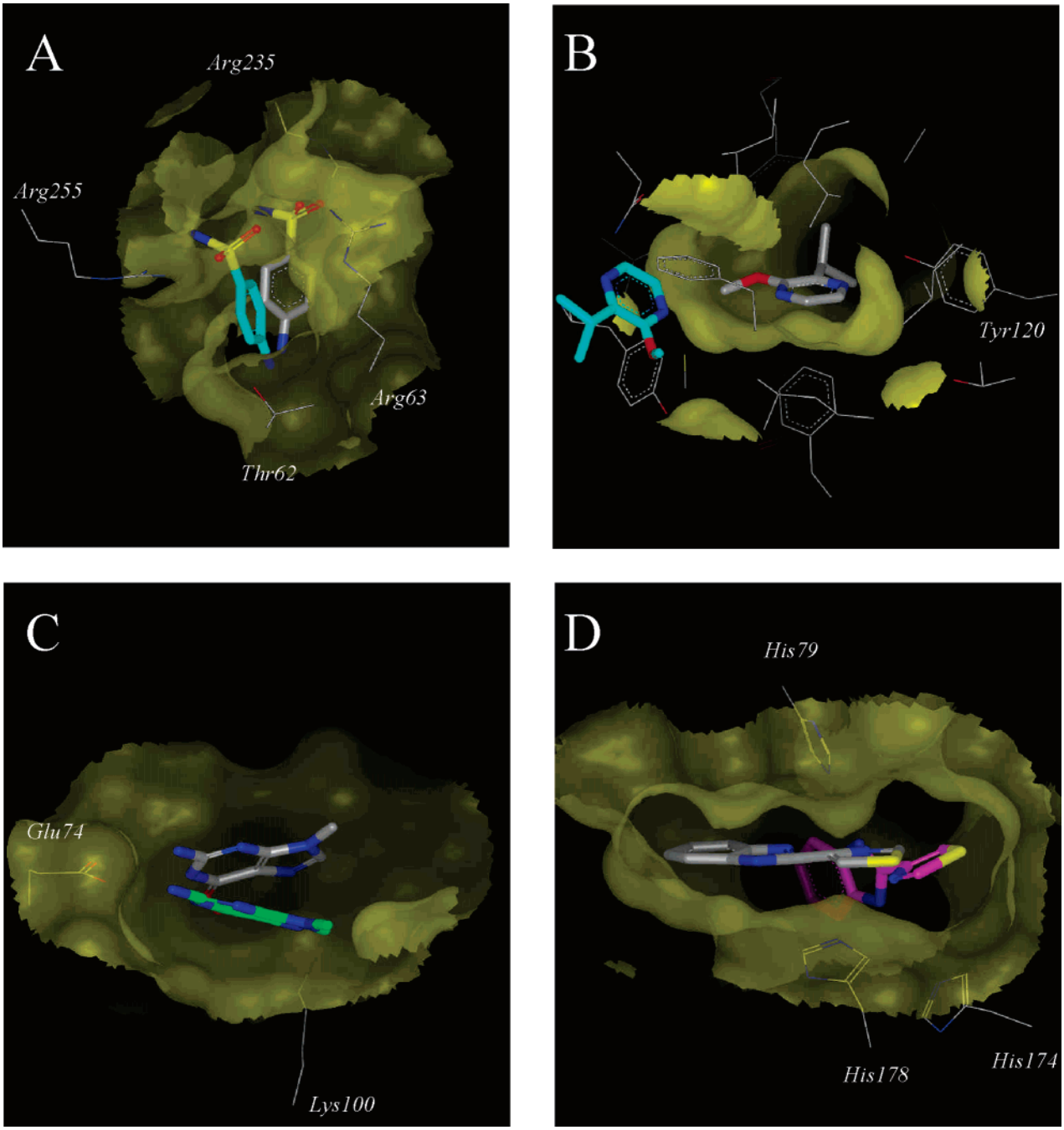

Figure 9. Difficult test cases for fragment docking. The X-ray pose (white-carbon sticks) is shown along with predicted poses (FlexX, cyan-carbons sticks; Gold, green-carbon sticks; Surflex, magenta-carbon sticks). The active site is represented by a solid surface and important protein residues by wire frames. Besides carbon atoms, the atom color coding is the following: oxygen, red; nitrogen, blue; sulfur, yellow. (A) 1aj0: The aniline moiety of the ligand is well-positioned, but the interactions of the sulfonamide group with Arg63 and Arg235 are not reproduced by FlexX, which prefers to anchor the sulfonamide to Arg63 and Arg255. (B) 1qy2: The ligand is embedded in a very apolar subsite with a single H bond to the pyrazine fragment, docked outside the binding site by FlexX. (C) 1rrw: The interaction of the 2-aminouracyl ring with Glu74 is reproduced but not the location of the partially buried imidazole ring, which is forced by Gold to H-bond to Lys100. (D) 1yvm: The thiazole ring is H-bonded to His79 and His178 in the X-ray structure with a benzimidazole ring partially buried upon binding. Surflex places the benzimidazole ring deeper in the active site and predicts H-bonding of the benzimidazole moiety to His79 and His174.

IFP. For human cdk2, the problem is complicated by the fact that the binding site may exist in either a "closed" or "opened" form according to the conformation of lining residues (e.g., Lys33 and Asp145). Because we aimed to dock very low molecular weight ligands (scaffolds) into the protein structure, a closed form (PDB entry 1dm2) previously shown to be acceptable for docking many cdk2 ligands[43] was used here.

Scaffold ranking using energy-based scoring functions (Goldscore and Chemscore) was shown to be even worse than random picking, whereas the usage of IFPs significantly enhances the in silico screening accuracy (Figures 7 and 8). Interestingly, most of the single IFP references available allowed a significant enrichment of virtual hits in true cdk2-targeted scaffolds by using the Tc-IFP scoring function (Figure 7). Although the 1dm2 protein coordinates have been

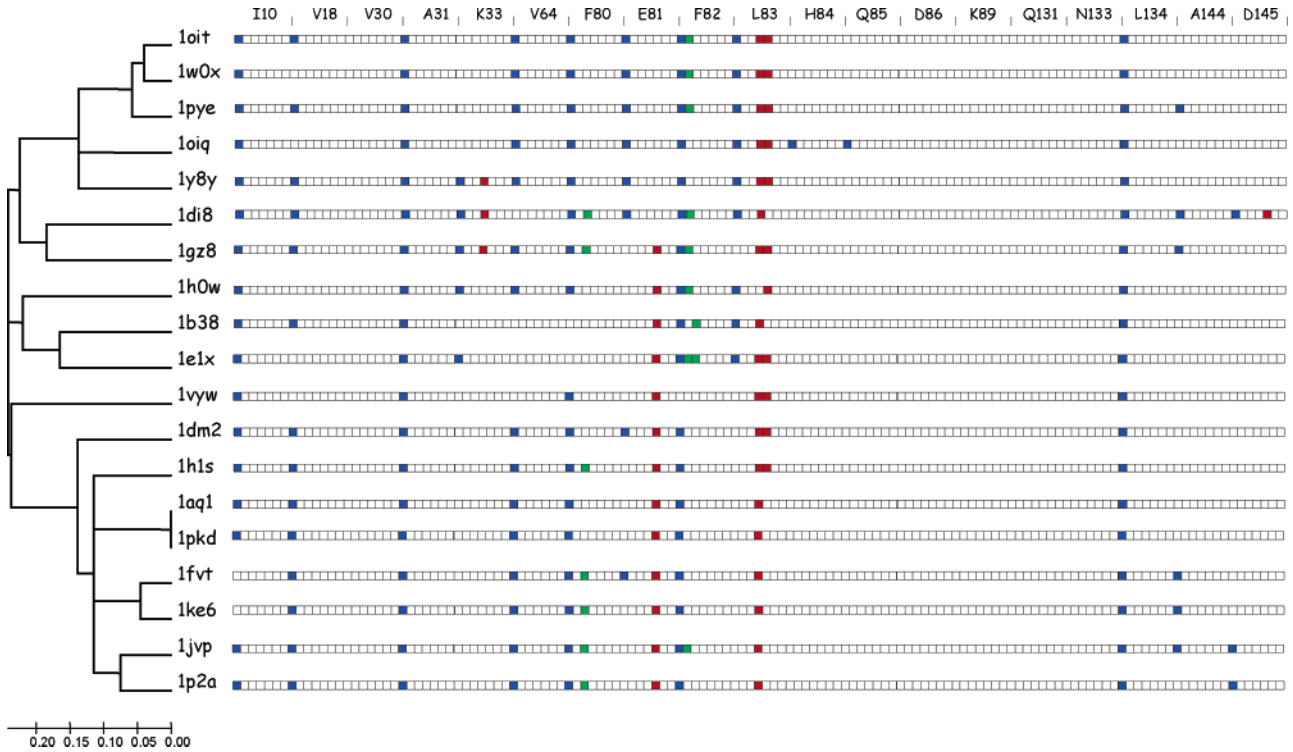

Figure 10. Phylogenic tree obtained by neighbor-joining clustering of 19 reference IFPs (Supporting Information Chart 3) generated from known PDB structures. A bit-to-bit comparison was applied to the 133-bit string to measure distance between two IFPs. On-bits are displayed for each of the 19 residues of the human cdk2 active site with the following color coding: blue, hydrophobic interactions; green, aromatic interactions; red, hydrogen bonds. Bit vector positions are analogous to those indicated in Table 3. Branch lengths are proportional to the distance (scale below the tree) between bit strings.

used, the 1dm2 scaffold was not the best possible choice for generating the most valuable reference IFP (Figure 8). Only three reference IFPs (1di8, 1jvp, and 1p2a) were unsuitable to retrieve true active scaffolds. Clustering all reference IFPs indicates that two of the poor IFPs are close together (1jvp and 1p2a) at the bottom of the IFP tree (Figure 10). Interestingly, the corresponding three scaffolds interact with the Asp145 side chain via either hydrophobic interactions (1jvp and 1p2a) or hydrogen bonding (1di8), two features which are absent in other reference IFPs (Figure 10).

When a Bayesian model trained with known binding modes of a few actives and predicted poses of presumed inactives is used, significant enrichments in true active scaffolds can be obtained by a standard structure-based in silico screening approach without having the need to explicitly define one reference (Figure 7). Assuming that a proper scoring metric has been found to prioritize very low molecular weight compounds, the true advantages of docking scaffolds instead of full compounds are double: (i) less compounds have to be docked to cover a defined chemical space, and therefore a larger proportion of virtual hits can be selected and experimentally tested; (ii) it is a very straightforward computational approach to scaffold hopping[44] and library design. Our results on cdk2-targeted scaffold docking compares very favorably with a previous pharmacophore constrained docking of low-molecular-weight cdk2 inhibitors,[11] although caution should be given to such comparisons because of the different settings in both computational experiments.

CONCLUSIONS

We herewith propose the usage of IFPs instead of the widely used RMSD metric to estimate docking success. IFPs focus on protein−ligand interactions and not on ligand coordinates and represent, therefore, a logical description of docking poses. From the first two benchmark data sets used in this study, we propose a Tc-IFP threshold of 0.6 for discriminating acceptable from bad solutions. However, only future applications of this new metric on additional benchmarks will help to refine the proposed threshold. Our docking/scoring approach is slightly different from several constrained docking protocols published so far.[30,45,46] In the present case, docking is totally unconstrained and poses are simply postprocessed to select the ones in agreement with some reference interaction fingerprint(s) which need not to be derived from the compound to dock. Our protocol is also different in spirit from template matching,[46] which attempts to optimally fit one compound to a reference set of coordinates without any guarantee that protein−ligand interactions are going to be conserved.

IFPs present the noticeable advantage of reconciling structure-based design with ligand-based data mining. Hence, 1-D fingerprints have heavily been investigated in the field of ligand-based virtual screening,[47] notably the usage of single[48] versus multiple baits,[49] fingerprint scaling,[50] and activity-centered fingerprints.[51] Fingerprints are also well-suited for applying machine learning methods[52] (neural networks, Bayesian statistics, support vector machines, recursive partitioning, and binary kernel discrimination) and taking the maximal advantage of multiple references.[53] Using IFPs is thus a fuzzy but very promising way of selecting

virtual poses/hits satisfying user-defined prerequisites which appear especially well-suited for fragment and scaffold in silico screening and thus better link structure-based screening to fragment-based drug discovery.

COMPUTATIONAL METHODS

Ligand Setup. Three ligand databases have been prepared as follows:

Data Set 1. Starting from 2721 "druglike" compounds extracted from the sc-PDB,[32] a database of druggable binding sites (http://bioinfo-pharma.u-strasbg.fr/scPDB/) from the Protein Data Bank,[54] a first topological filtering was established using an automated PipelinePilot[55] workflow to retain compounds according to the following criteria: (1) molecular weight between 150 and 250, (2) H-bond acceptor count below six, H-bond donor count below three, number of rotatable bonds below three, number of rings below three, topological polar surface area below 120 Å,[2] and AlogP below 3. This filtering step afforded a total of 80 ligands, which was further downsized by removing any compound whose similarity (expressed by a Tanimoto coefficient on MDL public keys)[56] was higher than 0.75 to any other ligand when reading data set ligands by increasing molecular weight. A total of 42 ligands (Supporting Information Chart 1) were finally selected as sd files, ionized at physiological pH using Filter2,[38] converted into 3-D mol2 files[57] with Corina,[58] and refined with 100 steps of Powel energy minimization using the TRIPOS force field.[57]

Data Set 2. Four ligands have been chosen for each of 10 selected pharmaceutically relevant targets (neuraminidase, estrogen receptor α, protein tyrosine phosphatase 1B, phosphodiesterase 4D, tryptophan synthase, ionotropic glutamate receptor type 2, cyclin-dependent kinase 2, human carbonic anhydrase II, thrombin, and cyclooxygenase-2) out of the sc-PDB database. The 40 ligands were prepared for docking as previously described for data set 1. The molecular scaffold of each of the 40 ligands was manually selected (Supporting Information Chart 2) and processed as for data set 1 and final 3-D coordinates saved in mol2 format.

Data Set 3. A total of 52 known cdk2 inhibitors were retrieved from the sc-PDB database by a simple keyword-based search (SwissProt id = P24491) and further clustered with the ClassPharmer 3.5 program[59] using medium homogeneity and no redundancy settings (Figure 11). Exact ring closure and exact atom match parameters were chosen to define a total of 19 maximum common substructures (MCS) representing "true active" cdk2-targeted scaffolds (Supporting Information Chart 3). Decoys were selected from the recently described SBI library[60] totaling 21 393 MCS computed from ca. 900 000 druglike commercially available compounds. To avoid decoy and active sets describing non-overlapping chemical spaces, a first filtering step was applied to the full decoy data set to keep 3306 scaffolds with global properties (molecular weight below 250, at least one H-bond donor and acceptor, and at least one aromatic ring) close to "true active" scaffolds. However, to ascertain whether decoy scaffolds have a low probability to recognize the cdk2 ATP binding site and thus bias screening results, the pairwise similarity of all remaining 3306 decoys to 47 reference scaffolds was calculated using circular FCFP_4 fingerprints.[55] Reference scaffolds were generated by ClassPharmer as previously described from a public set[61] of 309 known cdk2 inhibitors exhibiting an IC_{50} value lower than 25 μM. Any decoy whose similarity expressed by a Tanimoto coefficient was higher than 0.5 to any reference scaffold was discarded from the data set to afford a total of 2230 decoys, which were then merged with the 19 true active scaffolds in data set 3 (Figure 11). Further processing (ionization, 3-D coordinates generation, and energy-refinement) was realized as described above for ligand data sets 1 and 2.

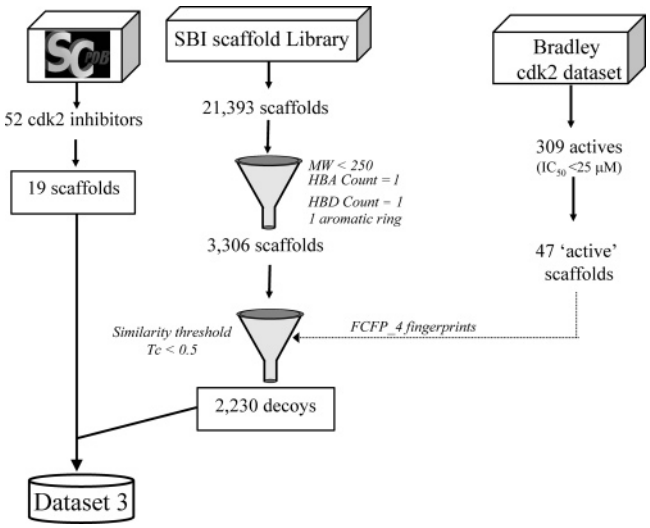

Figure 11. Workflow for the preparation of data set 3 consisting in 19 true actives (scaffolds extracted from cdk2 inhibitors of known protein-bound PDB structure) and 2230 decoys.

Protein Setup. Protein and active site coordinates were directly extracted from the sc-PDB database[32] in readable PDB or MOL2 format for which ionization and tautomeric states were previously manually checked and corrected if necessary. For known PDB complexes, coordinates of polar hydrogen atoms were manually set in order to optimize protein−ligand H bonds first and protein−protein H bonds in a second step.

FlexX 1.13 Docking. Standard parameters of FlexX[29] as implemented in the 7.0 release of the SYBYL package[57] were used for the iterative growing and subsequent scoring of FlexX poses. Active site atoms were defined as previously described in the sc-PDB database.[32] A receptor description file was automatically defined form the PDB coordinates of the hydrogen-free protein/active site coordinates and further checked for consistency of polar hydrogen positions and the tautomeric state of histidine side chains in FlexV. Formal charges were assigned to ligand atoms. The top 30 solutions as scored by FlexXScore were retained and further stored in a single mol2 file.

Glide 3.5 Docking. Glide SP calculations[18] were performed with Impact version 3.5.[62] The grid generation step requires MAE input files of both the ligand and active site including hydrogen atoms. The protein charged groups that were not located in the ligand binding pocket nor involved in salt bridges were neutralized using the Schrödinger pprep script. Important metal ions and cofactors were included in binding sites. The center of the grid enclosing box was defined by the center of the bound ligand as described in the sc-PDB entry. The enclosing box and bounding box dimensions were fixed to 14 and 10 Å, respectively. No further modifications were applied to the default settings. A

total of 30 poses were saved for each ligand.

Gold 3.1 Docking. The active site was defined to encompass any protein atom included in a 10-Å-radius sphere centered on the center of mass of the bound ligand as described in the original sc-PDB entry. Default settings of the slower parameter set ("default 1") and the Goldscore scoring function were used to retain 30 poses for each ligand.

Surflex 1.2 Docking. Surflex[31] employs an idealized active site called a protomol,[63] built from the hydrogen-containing protein mol2 file and based on protein residues that line the active site (see the above definition) using standard parameters. Docking of the ligand was run using default settings. The program returned up to 10 poses for each ligand.

cdk2-Targeted Virtual Screening. Compounds in data set 3 were docked to the X-ray structure of human cdk2 (PDB code: 1dm2) using the Gold v3.1 program as described above (Gold docking). Both Goldscore and Chemscore scoring functions were used to generate two independent sets of 30 poses per ligand.

Interaction Fingerprint (IFP) Computation. IFPs are generated from a list of docked poses and active site coordinates using a C++ library and executables freely available upon request to the authors. On the basis of a list of atom flags (Table 2) inferred from OpenEye's OEChem1.3 library, positions of a bit vector are switched either on or off depending on whether or not predefined intermolecular interactions (Table 3) agree with user-defined rules (Table 4). All geometric rules may be edited and saved in a configuration file without a need to recompile the code. By default, only the first seven bits which correspond to the most frequent protein−ligand interactions are calculated. However, the user may choose to compute less frequent interactions (e.g., weak H bonds, π-cation interactions, and metal complexation) and thus fill the remaining four positions in the bit vector. Using a freely available C clustering library[64] linked to our routine, IFPs may be clustered according to various methods and then converted in dendograms for visualizing the diversity of generated poses. In the present work, only simple seven-bit-long strings were generated. Measuring the distance between two IFPs was realized using a simple Tanimoto metric (Tc-IFP) as follows:

$$\text{Tc-IFP} = \frac{|A \cap B|}{|A \cup B|}$$

where $A \cap B$ is the number of switched-on bits common to IFPs A and B and $A \cup B$ is the sum of switched-on bits in IFPs A and B.

Bayesian Classification of IFPs. A Laplacian-modified Bayesian classifier[33] was first trained by pooling 19 IFPs generated from known X-ray poses of the 19 "true cdk2" scaffolds embedded in data set 3 and 2291 Gold-predicted poses for a set of 200 scaffold decoys randomly chosen from the recently described SBI scaffold library[60] but presenting a molecular weight below 250. Those decoys were docked to the 1dm2 coordinates as previously described. The Bayesian learning model was derived by tagging the 19 X-ray-derived IFPs as "GOOD" and the 2291 predicted IFP decoys as "BAD" using standard settings of the Bayesian classifier in Pipeline Pilot 5.1.[55] The model was then used to classify IFPs computed from Gold poses obtained from the virtual screening of the test set (data set 3) whose decoys show no overlap with those of the training set.

ACKNOWLEDGMENT

This work was supported by the LifeSciHealth (Grant LSHB-CT-2003-503337) programme of the European Community. We sincerely thank Marcel Verdonk and Suzanne Brewerton (Astex Technologies, U. K.) for critical reading of the manuscript.

Supporting Information Available: Chart 1 (data set of 42 low-molecular-weight ligands from the Protein Data Bank), Chart 2 (data set of 40 molecular scaffolds covering 10 pharmacology relevant targets), and Chart 3 (data set of 19 cdk2 inhibitors of known PDB structure). This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Zartler, E. R.; Shapiro, M. J. Fragonomics: Fragment-Based Drug Discovery. *Curr. Opin. Chem. Biol.* **2005**, *9*, 366−370.
(2) Verdonk, M. L.; Hartshorn, M. J. Structure-Guided Fragment Screening for Lead Discovery. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 404−410.
(3) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead Selection. *Drug Discovery Today* **2004**, *9*, 430−431.
(4) Blundell, T. L.; Patel, S. High-Throughput X-ray Crystallography for Drug Discovery. *Curr. Opin. Pharmacol.* **2004**, *4*, 490−496.
(5) Schade, M. NMR Fragment Screening: Advantages and Applications. *IDrugs* **2006**, *9*, 110−113.
(6) Howard, N.; Abell, C.; Blakemore, W.; Chessari, G.; Congreve, M.; Howard, S.; Jhoti, H.; Murray, C. W.; Seavers, L. C.; van Montfort, R. L. Application of Fragment Screening and Fragment Linking to the Discovery of Novel Thrombin Inhibitors. *J. Med. Chem.* **2006**, *49*, 1346−1355.
(7) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for Fragment-Based Lead Discovery. *Drug Discovery Today* **2003**, *8*, 876−877.
(8) Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. Library Design for Fragment Based Screening. *Curr. Top. Med. Chem.* **2005**, *5*, 751−762.
(9) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.
(10) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.
(11) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual Screening Using Protein−Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(12) Stahl, M.; Bohm, H. J. Development of Filter Functions for Protein−Ligand Docking. *J. Mol. Graphics Modell.* **1998**, *16*, 121−132.
(13) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(14) Bissantz, C.; Folkers, G.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(15) Paul, N.; Rognan, D. ConsDock: A New Program for the Consensus Analysis of Protein−Ligand Interactions. *Proteins* **2002**, *47*, 521−533.
(16) Vigers, G. P.; Rizzi, J. P. Multiple Active Site Corrections for Docking and Virtual Screening. *J. Med. Chem.* **2004**, *47*, 80−89.
(17) Klon, A. E.; Glick, M.; Davies, J. W. Combination of a Naive Bayes Classifier with Consensus Scoring Improves Enrichment of High-Throughput Docking Results. *J. Med. Chem.* **2004**, *47*, 4356−4359.
(18) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(19) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. FDS: Flexible Ligand and Receptor Docking with a Continuum Solvent Model and Soft-Core Energy Function. *J. Comput. Chem.* **2003**, *24*, 1637−1656.

(20) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040−4048.
(21) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J. Med. Chem.* **2005**, *26*, 915−931.
(22) Cambridge Cristallographic Data Centre, Cambridge, U. K.
(23) Springer, C.; Adalsteinsson, H.; Young, M. M.; Kegelmeyer, P. W.; Roe, D. C. PostDOCK: A Structural, Empirical Approach to Scoring Protein Ligand Complexes. *J. Med. Chem.* **2005**, *48*, 6821−6831.
(24) Amari, S.; Aizawa, M.; Zhang, J.; Fukuzawa, K.; Mochizuki, Y.; Iwasawa, Y.; Nakata, K.; Chuman, H.; Nakano, T. VISCANA: Visualized Cluster Analysis of Protein−Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening. *J. Chem. Inf. Model.* **2006**, *46*, 221−230.
(25) 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.
(26) Kelly, M. D.; Mancera, R. L. Expanded Interaction Fingerprint Method for Analyzing Ligand Binding Modes in Docking and Structure-Based Drug Design. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1942−1951.
(27) 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.
(28) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. *Proteins* **2004**, *57*, 225−242.
(29) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(30) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein−Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609−623.
(31) Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *J. Med. Chem.* **2003**, *46*, 499−511.
(32) 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.
(33) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of Kinase Inhibitors Using a Bayesian Model. *J. Med. Chem.* **2004**, *47*, 4463−4470.
(34) 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.
(35) Singh, J.; Deng, Z.; Narale, G.; Chuaqui, C. Structural Interaction Fingerprints: A New Approach to Organizing, Mining, Analyzing, and Designing Protein−Small Molecule Complexes. *Chem. Biol. Drug Des.* **2006**, *67*, 5−12.
(36) 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.
(37) 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.
(38) OpenEye Scientific Software, Santa Fe, NM 87507.
(39) Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking. *J. Med. Chem.* **2003**, *46*, 2656−2662.
(40) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *J. Chem. Inf. Model.* **2006**, *46*, 401−415.
(41) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(42) Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlen, M.; Stouten, P. F. Assessment of Docking Poses: Interactions-Based Accuracy Classification (IBAC) versus Crystal Structure Deviations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871−881.
(43) Thomas, M. P.; McInnes, C.; Fischer, P. M. Protein Structures in Virtual Screening: A Case Study with CDK2. *J. Med. Chem.* **2006**, *49*, 92−104.
(44) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.
(45) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. Flexible Docking under Pharmacophore Type Constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.
(46) Cross, S. S. Improved FlexX Docking Using FlexS-Determined Base Fragment Placement. *J. Chem. Inf. Model.* **2005**, *45*, 993−1001.
(47) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.
(48) Godden, J. W.; Stahura, F. L.; Bajorath, J. Anatomy of Fingerprint Search Calculations on Structurally Diverse Sets of Active Compounds. *J. Chem. Inf. Model.* **2005**, *45*, 1812−1819.
(49) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.
(50) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Similarity Search Profiles as a Diagnostic Tool for the Analysis of Virtual Screening Calculations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1275−1281.
(51) Godden, J. W.; Bajorath, J. A Distance Function for Retrieval of Active Molecules from Complex Chemical Space Representations. *J. Chem. Inf. Model.* **2006**, *46*, 1094−1097.
(52) Arimoto, R.; Prasad, M. A.; Gifford, E. M. Development of CYP3A4 Inhibition Models: Comparisons of Machine-Learning Techniques and Molecular Descriptors. *J. Biomol. Screening* **2005**, *10*, 197−205.
(53) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. New Methods for Ligand-Based Virtual Screening: Use of Data Fusion and Machine Learning to Enhance the Effectiveness of Similarity Searching. *J. Chem. Inf. Model.* **2006**, *46*, 462−470.
(54) 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.
(55) SciTegic Inc., San Diego, CA 92123−1365.
(56) Elsevier MDL, San Leandro, CA 94577.
(57) TRIPOS, Assoc., Inc., St. Louis, MO.
(58) Molecular Networks GmbH, D-91052 Erlangen, Germany.
(59) Bioreason Inc., Santa Fe, New Mexico 87507.
(60) Krier, M.; Bret, G.; Rognan, D. Assessing the Scaffold Diversity of Screening Libraries. *J. Chem. Inf. Model.* **2006**, *46*, 512−524.
(61) Bradley, E. K.; Miller, J. L.; Saiah, E.; Grootenhuis, P. D. Informative Library Design as an Efficient Strategy to Identify and Optimize Leads: Application to Cyclin-Dependent Kinase 2 Antagonists. *J. Med. Chem.* **2003**, *46*, 4360−4364.
(62) Schrödinger, Inc., New York, NY 10036.
(63) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites. *Chem. Biol.* **1996**, *3*, 449−462.
(64) http://bonsai.ims.u-tokyo.ac.jp/∼mdehoon/software/cluster/software-.htm (accessed Nov 2006).
(65) Steiner, T. The Hydrogen Bond in the Solid State. *Angew. Chem., Int. Ed.* **2002**, *41*, 48−76.

CI600342E