Assessment of Scaffold Hopping Efficiency by Use of Molecular Interaction Fingerprints

Jennifer Venhorst,[†] Sara Núñez,*[,†] Jan Willem Terpstra, and Chris G. Kruse

Solvay Pharmaceuticals, Research Laboratories, C. J. van Houtenlaan 36, 1381 CP Weesp, The Netherlands

Received February 1, 2008

A novel scoring algorithm based on molecular interaction fingerprints (IFPs) was comparatively evaluated in its scaffold hopping efficiency against four virtual screening standards (GlideXP, Gold, ROCS, and a Bayesian classifier). Decoy databases for the two targets under examination, adenosine deaminase and retinoid X receptor alpha, were obtained from the Directory of Useful Decoys and were further enriched with approximately 5% of active ligands. Structure and ligand-based methods were used to generate the ligand poses, and a Tanimoto metric was chosen for the calculation of the similarity interaction fingerprint between the reference ligand and the screening database. Database enrichments were found to strongly depend on the pose generator algorithm. In spite of these dependencies, enrichments using molecular IFPs were comparable to those obtained with GlideXP, Gold, ROCS, and the Bayesian classifier. More interestingly, the molecular IFP scoring algorithm outperformed these methods at scaffold hopping enrichment, regardless of the pose generator algorithm.

1. Introduction

Identifying active compounds from large databases by applying structural knowledge through in silico approaches has become an important part of current drug discovery. Much of the drive in making use of virtual screening (VS) methods has arisen from the increased pressure to reduce the costs involved in experimental high-throughput screening.[1] The ambition of VS methods is thus to accelerate the lead finding and lead optimization discovery stages, ultimately resulting in more candidates in the developmental pipeline in shorter time frames.

A number of VS approaches have been comparatively evaluated by several groups with respect to their ability to provide substantial ligand enrichment among top-ranking hits.[2–6] The result is a series of elegant studies that challenged the ability of various VS methods to assign low scores to decoys while assigning high scores to annotated active compounds in the database. These studies also critically characterize the common strengths and pitfalls of VS methods with the following conclusions. First, database enrichments bear a high dependence on the target studied, that is, the quality of the crystal structure (or homology model), topology of the binding pocket, cavity size, and magnitude of induced-fit motions upon ligand binding. Second, the physical and structural correlation between the decoys and the ligands is critical in assessing enrichment factors in virtual screens. This concern is illustrated in the works of Verdonk et al.,[7] who concluded that if there are significant differences in physical properties (e.g., size) between ligands and decoys, docking enrichments can appear artificially good. Third, the enrichment of actives from among a large database of decoys can also vary depending on the virtual screening algorithm, the various constraints imposed for pose generation, and the software/hardware characteristics.

Recently, the competence of VS approaches at scaffold hopping has also generated much attention and research in the drug design community.[8–13] In an attempt to improve the poor performance of the energy-based and 2D fingerprint similarity scoring methods in scaffold hopping,[14] several new descriptors have been recently introduced. Among these, the GRID molecular interaction field descriptor,[12] the SHED descriptor,[8] and the CATS descriptor[15] have been fruitfully used in different virtual screening strategies.

In this study, we present a comparative evaluation of a novel scoring method based on a molecular IFP algorithm[6] against the popular VS programs GlideXP,[16] Gold,[17] ROCS,[18] and a Bayesian classifier as to their scaffold hopping efficiency. Our results show that (i) the molecular IFP scoring algorithm is equally suited to quantifying virtual screening success as the popular internal scoring functions of GlideXP, Gold, or ROCS, and (ii) the IFP algorithm is overall a more efficient scoring method at scaffold hopping regardless of the pose generator algorithm.

2. Methodology

a. Protein Target Selection and Preparation. In this study, two proteins were used as test cases. The first one was adenosine deaminase (ADA), an enzyme involved in purine metabolism which irreversibly deaminates adenosine. Many crystal structures of ADA in complex with assorted inhibitors exemplify its rather polar active site (Figure 1). The second protein target was retinoid X receptor alpha (RXRα), a nuclear receptor that activates transcription upon binding of 9-*cis*-retinoic acid or other RXR agonists. Abundant crystal structures of RXRα in complex with different ligands reveal this nuclear receptor binding site as considerably hydrophobic (Figure 1). The choice of very disparate targets for the underlying analysis was intended to identify possible biases that may exist in the virtual screening algorithms with regard to the nature of the ligand database and/or protein−ligand interactions.

The crystal structures of ADA and RXRα were obtained from the ZINC database[19] and the RCSB Protein Data Bank,[20] respectively, in PDB format. Water molecules and ions were removed from the original source. Hydrogen atoms were added, and the positions of hydrogen atoms involved in polar interactions were subsequently minimized to optimize their hydrogen bond interactions using the Protein Preparation module in Maestro.[16] Protein residues were inspected, and the tautomeric states of histidines, hydroxyl group orientations, and protonation states of titratable residues were adjusted accordingly. The resulting receptor models were saved to MAE and MOL2 file formats, compatible with GlideXP and Gold, respectively.

* To whom correspondence should be addressed. Fax: +31-294-477138. Phone: +31-294-479868. E-mail: sara.nunez@solvay.com.

[†] Both authors have contributed equally to the studies presented.

10.1021/jm8001058 CCC: $40.75
Published on Web 05/01/2008

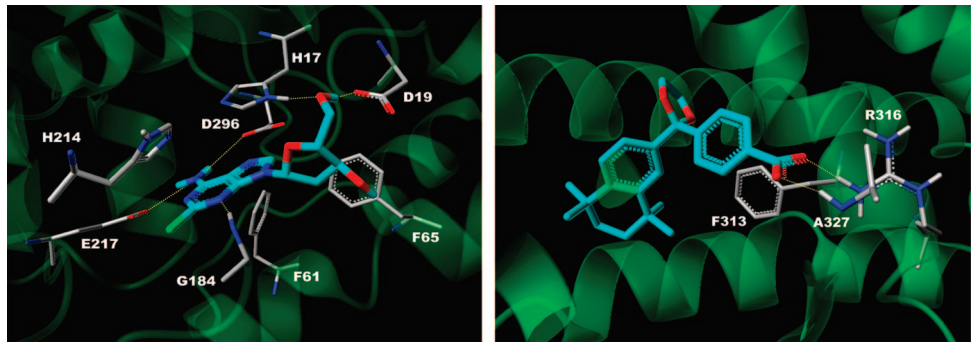


Figure 1. Binding sites of adenosine deaminase (left) and retinoid X receptor alpha (right). The reference ligands are colored in blue and protein residues that are actively involved in ligand binding are annotated accordingly.

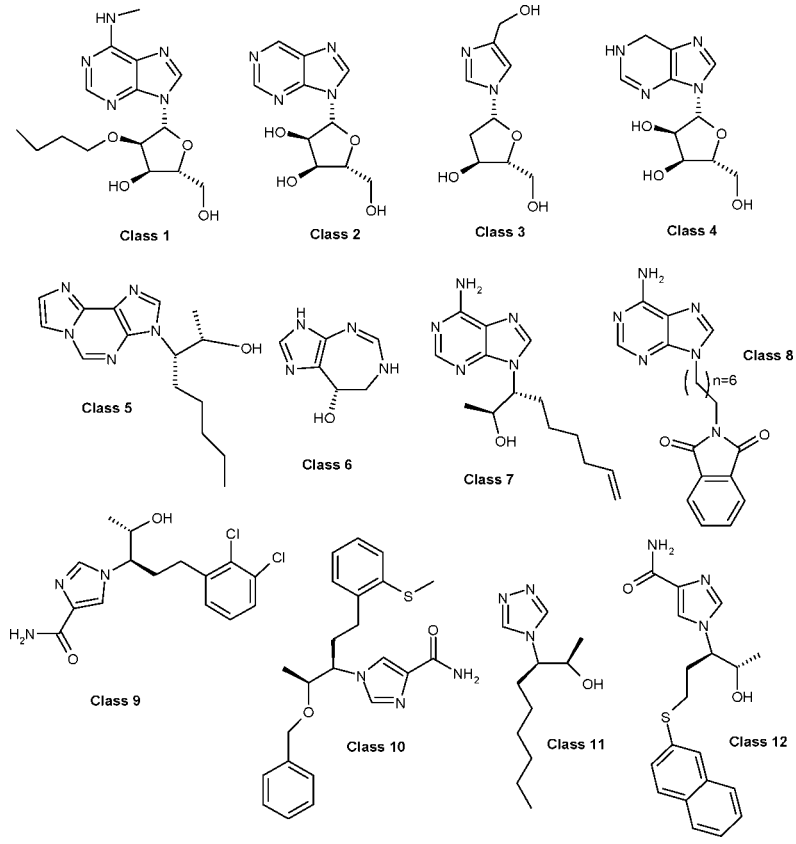


Figure 2. Representative molecules from the 12 structural classes in the active ligand set of ADA.

b. Database Selection and Preparation. In the interest of improving and validating existing and upcoming VS algorithms, it is desirable that the modeling community uses benchmarked data sets in their investigations so that biases are minimized, thus facilitating the assessment and comparison of the VS studies. A recent publication describes an unprecedented benchmarking of sets for molecular docking.[21] The result of this generous effort is a publicly available database of physically matched decoys and ligands, the Directory of Useful Decoys (DUD), which provides a more stringent decoy criterion with which to evaluate virtual screening performance. DUD is currently drawn from the ZINC database,[19] a database of commercially available compounds for virtual screening.

Relevant decoy libraries for the two targets under investigation in this study were used and further enriched with active ligands obtained from both the ZINC and Integrity Prous databases.[22] The databases were prepared in the following manner. Molecules were converted from 2D SDF to 3D MOL2 format using SYBYL,[23] and OMEGA[24] was subsequently used to obtain the lowest-energy conformer for each database molecule. The screening database of ADA consisted of 821 decoys and 48 active ligands, and that of RXRα consisted of 706 decoys and 38 active ligands. We used Classpharmer[25] to cluster the active ligands into distinct structural classes. In addition, some classes were further broken down in subclasses in order to maintain an even average of compounds in each class. Classpharmer was used to generate the subclasses, and the criterion that compounds belonging to each subclass have dissimilar molecular IFPs was imposed. The result was a balanced clustering of ADA and RXRα active ligands among 12 and 10 classes, respectively (Figures 2 and 3).

c. Molecular Interaction Fingerprint. Recently, studies by Rognan[6] and Deng[26] introduced two novel topological scoring methods based on molecular interaction fingerprints (IFPs) as a better descriptor to quantify docking success. IFPs are simple bit strings that convert 3D information on protein−ligand interactions into 1D bit vector representations that can be quickly compared to a reference pose by the use of traditional metrics, e.g., Tanimoto similarity coefficient.[27]

Specifically, the molecular IFP algorithm was used as a simple but efficient postdocking processing method for prioritizing the most

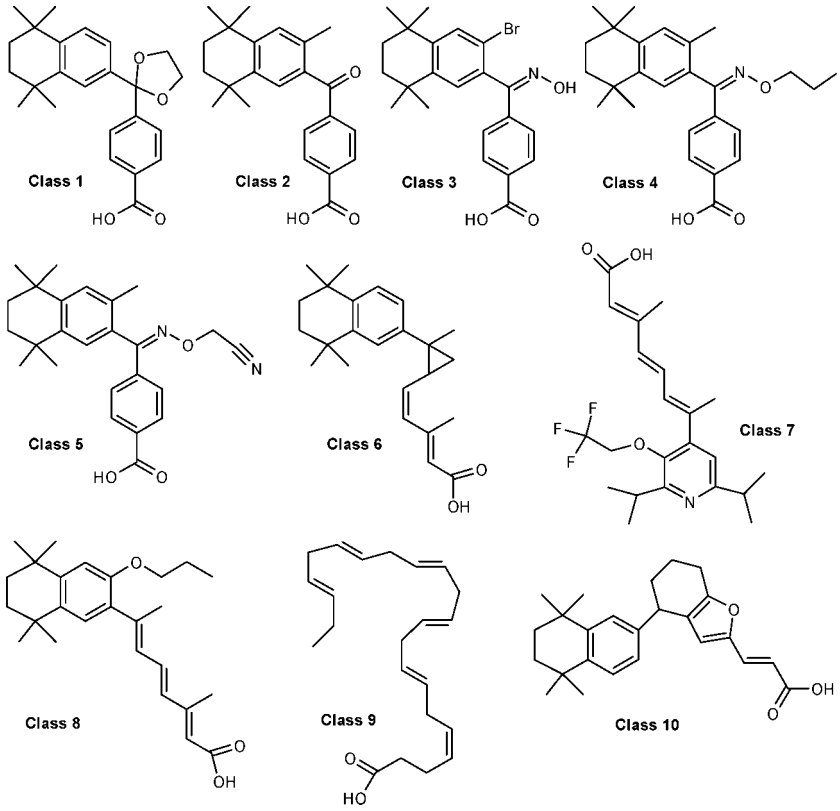


Figure 3. Representative molecules from the 10 structural classes in the active ligand set of RXRα.

Table 1. Seven-Bit Molecular Interaction Fingerprint Used for Every Residue in the Protein

bit vector position	protein atom type	ligand atom type	interaction
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

relevant poses of both fragments and druglike compounds.[6] A comparable structural molecular interaction fingerprint (SIFt) was reported that incorporates the binding interactions of various fragments in a combinatorial library and translates desirable ligand−target binding interactions into library filtering constraints.[26] The authors demonstrated that their new algorithm, coupled with classification models, is a valuable structure-based tool for the design of novel combinatorial chemical libraries.

Our current work evaluates Rognan's IFP algorithm in its ability to enrich libraries with novel bioactive scaffolds. The underlying assumption in the use of topological IFPs as a scoring method is that pose generators like GlideXP and Gold are capable of sampling the binding site cavity for generating reliable poses; however, they are simply not accurate enough at correctly scoring them. Our approach makes use of these correct poses and applies the molecular IFP as an enhanced scoring algorithm by comparing the fingerprint of the binding pose of a cocrystallized reference ligand to the IFPs of the screening database ligands.

For the computation of the protein−ligand interactions and subsequent scoring with the Tanimoto metric, a 7-bit fingerprint was used (Table 1).

For the purpose of the similarity analysis, two reference ligands were taken for each target. These ligands represent the binding modes of crystallized ADA and RXRα ligands, and their IFPs are shown in Table 2. Each binding site residue is associated with a 7-bit fingerprint, and the bit is turned on if the interaction between the residue and the ligand is present. It becomes evident from the comparison of both molecular interaction fingerprints that the binding site of ADA has a more polar character than that of RXRα, which in fact mirrors the nature of their ligands.

Given the fact that the molecular interaction fingerprint vector shown in Table 2 does not consider repulsive van der Waals and electrostatic interactions between a given ligand and the protein, a postprocessing analysis was required prior to the database enrichment computation with the IFP scoring method. Hence those ligand poses which presented abnormal electrostatic (elec) or van der Waals (vdW) energies (i.e., E_{elec} and $E_{vdW} > 0$) were discarded before the IFP similarity comparison and subsequent database ranking.

Finally, because different software packages were used, each with a proprietary file format with occasional discrepancies in atom type assignments, standardization of the database ligand poses was performed with UNITY[23] prior to the generation of the molecular interaction fingerprint.

d. Structure-Based Algorithms. The GlideXP (Maestro version 8.0) and Gold (version 3.0) docking algorithms were used in this study. With regard to the virtual screening settings of GlideXP, the crystallized ligand was chosen as the center of mass for generation of the grid for the docking calculation. Rigid protein docking was carried out using the "extra precision" (XP) mode which combines a powerful conformational sampling protocol with a custom scoring function that is specifically designed to eliminate false positives. The remaining options were kept as default and no constraints were used. The GlideXP Score was used to select the 50 best poses for each ligand in the database.

With regards to Gold, the binding pocket of each target was defined from the crystallographic coordinates of the ligand, taking into account any residue within 15 Å of any atom of the ligand. Docking runs were performed under the standard default settings mode keeping the protein rigid. No imposed geometrical constraints were used, and the GoldScore function was used to obtain a maximum of 50 conformations of each ligand. Subsequently, the 50 best poses for each ligand were rescored with the ChemScore energy function.

Table 2. Molecular Interaction Fingerprint of the Reference Ligands of ADA (top) and RXRα (bottom)[a]

ligand	HIS17	ASP19	LEU58	PHE61	LEU62	PHE65	TYR102	SER103	LEU106	CYS153	MET155	ALA183	GLY184	HIS214	GLU217	ASP296
113529	1101000	1000100	1000000	1010000	1000000	1010000	1000000	1000000	1000000	1000000	1000000	1000000	1001000	1010000	1000100	1000000
ZINC01614355	1101000	1000100	0000000	1010000	1000000	1010000	1000000	1000000	1000000	1000000	1000000	1000000	1001000	0000000	0000000	1000100
ligand	VAL265	ILE268	ALA271	ALA272	TRP305	ASN306	LEU309	ILE310	PHE313	ARG316	ILE324	LEU326	ALA327	VAL342	ILE345	VAL349
338166	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	0001010	1000000	1000000	0000000	1000000	1000000	1000000
ZINC01540632	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1000000	1010000	0001010	1000000	1000000	0001000	1000000	1000000	1000000

[a] Only the most important binding residues are shown.

e. Ligand-Based Algorithm. Because of its speed and good performance in previously reported VS studies,[28] ROCS was chosen as the ligand-based VS method. ROCS ranks molecules on the basis of their similarity to a known active molecule (i.e., reference ligand) in 3D shape space, using atom-centered Gaussian functions to allow rapid maximization of molecular overlap (volume and atomic features), so-called ComboScore. Prior to the overlap calculation, OMEGA (version 2.2.1) was used to create the 500 lowest-energy conformations of each ligand in the database, and subsequently, ROCS (version 2.2) was used to obtain the 50 best overlap poses of each ligand in the database with respect to two reference ligands (see Section 2c for more details on the reference ligands).

In addition, the scoring efficiency of ROCS was compared to that of the IFP method using ROCS as the pose generator. For this purpose, the 50 best overlap poses of each ligand in the database to the crystal-like reference ligand poses were collected and embedded in the protein, preserving the same orientation within the binding cavity.

f. Naïve Bayesian Classifiers. Learned Bayesian models were generated for each target with Pipeline Pilot[29] using two reference ligands ("actives"), and a random subset of 20 decoys ("inactives"). Two-dimensional structural information was used as input data, and the extended connectivity fingerprint (ECFP-12), *A*log*P*, molecular weight, number of hydrogen bond acceptors, number of hydrogen bonds donors, number of rotational bonds, and molecular polar surface area descriptors were used for model building. The Bayesian model was then used to rank the remaining compounds in the screening database derived from the normalized probability.

Additionally, a second pair of Bayesian classifiers was learned using the molecular interaction fingerprints using several nonredundant IFPs of the two reference ligands ("actives"), and similarly for a small, random subset of 20 decoys ("inactives"). Subsequently, 50 IFPs were then generated for each ligand in the database using Gold, and those poses which presented unacceptable van der Waals or electrostatic interaction energies were discarded. Thereafter, the model was used to rank the database ligands according to the normalized probability.

3. Results and Discussion

a. Overall Database Enrichments. In this study, we investigated the ability of different VS methods at enriching benchmarked decoy libraries spiked with approximately 5% of true ligands. Moreover, we assessed the internal scoring functions of GlideXP, Gold, ROCS, and a naı¨ve Bayesian classifier at scaffold hopping efficiency and compared them with the novel topological IFP scoring method based on protein−ligand interaction fingerprints.

At this point, we would like to make the distinction between pose generator and scoring algorithm efficiencies. One must ensure that the pose generator and scoring algorithm are capable of characterizing the true ligands, reproducing crystal-like poses and ranking them well. Both play an important role in discriminating ligands from decoys in a screening database, and any VS study must ensure the prior assessment of docking algorithms and scoring functions using benchmarked screening ligand-decoy libraries.

Accordingly, we tested GlideXP and Gold in their ability to reproduce accurate poses of two reference ligands for which crystallographic data were available. Our analysis reveals that the rmsd of docked pose versus crystallographic pose is lower than 1.5 Å for ADA and RXRα, which falls within the acceptable limits of accuracy and reproducibility. These results therefore demonstrate that for the two proteins used in this evaluation, ADA and RXRα, docking into a single protein structure by multiple compound classes was able to reproduce the observed binding mode even when the protein is held rigid.

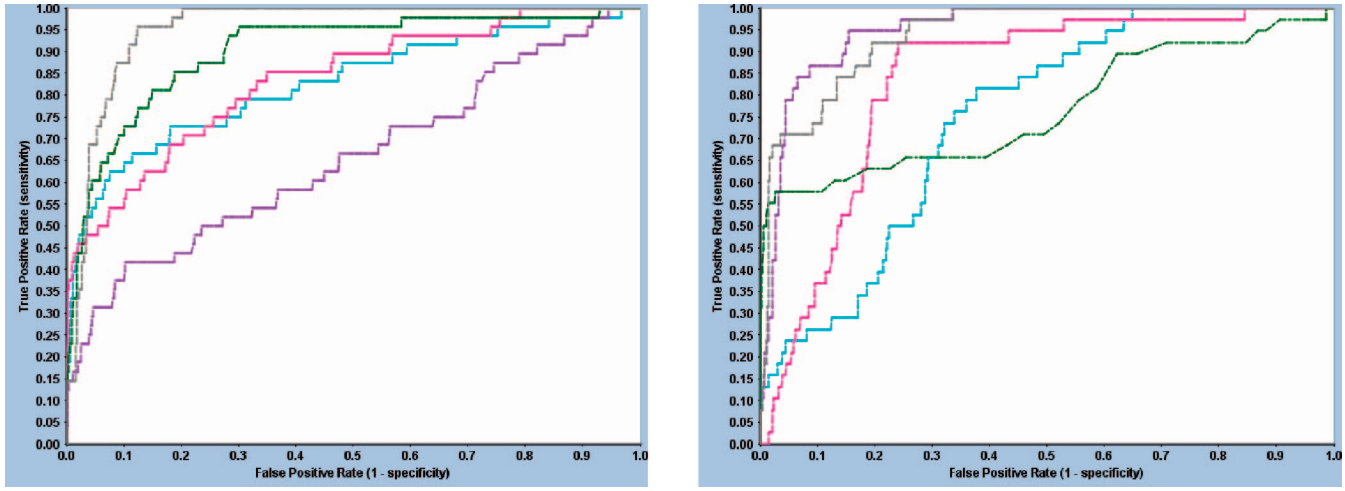


Figure 4. Database enrichments of ADA (left) and RXRα (right). The coloring scheme is as follows: light blue (GoldScore), pink (GlideXP), violet (ChemScore), green (ROCS), and grey (Bayesian).

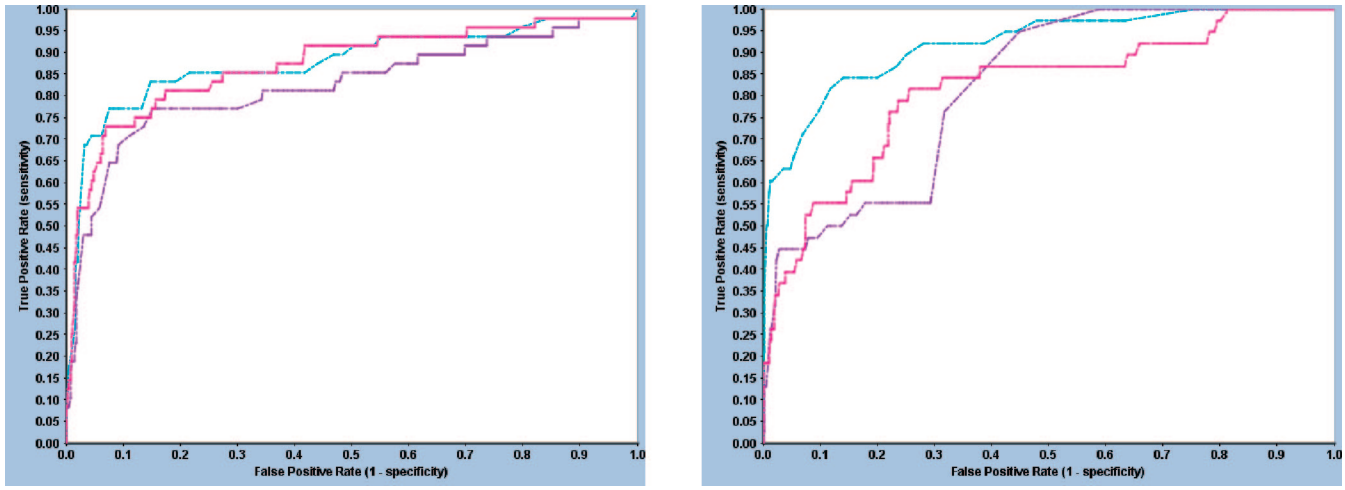


Figure 5. Database enrichments of ADA (left) and RXRα (right) calculated with the IFP scoring method. The coloring scheme is as follows: light blue (GoldScore for ADA and ChemScore for RXRα), pink (Bayesian), and violet (ROCS).

Table 3. Scaffold Hopping Enrichment at 5% (left) and 10% (right) of the Database Screened for ADA[a]

Method/Class	1	2	3	4	5	6	7	8	9	10	11	12
GoldScore												
ChemScore												
Gold IFP												
ROCS												
ROCS IFP												
Bayes												
Bayes IFP												

Method/Class	1	2	3	4	5	6	7	8	9	10	11	12
GoldScore												
ChemScore												
Gold IFP												
ROCS												
ROCS IFP												
Bayes												
Bayes IFP												

[a] Color-coding is as follows. Red: not a single active ligand has been retrieved for that particular class; green: at least one ligand has been retrieved for that particular class.

Database enrichments for ADA and RXRα were then calculated with Pipeline Pilot, and the results are summarized in Figure 4 with Receiver Operating Characteristic (ROC) plots. The areas under the ROC curves are a robust method for measuring performance. These plots represent the proportion of all actives recovered versus the proportion of all inactives recovered as one proceeds from the top to the bottom of the ranked list. An immediate observation drafted from Figure 4 is that the database enrichments of ADA and RXRα are strongly dependent on the algorithm used. We observed that the various rescoring options in Gold perform differently, with ChemScore and GoldScore performing better for hydrophobic and polar ligands, respectively. With regards to GlideXP, the number of database ligands that did not find any pose within the binding pockets of ADA and RXRα was substantially high, this phenomenon being more severe for RXRα. These observations are not unusual as the GlideXP scoring method encompasses a less flexible criterion in its energy function in order to discriminate ligands from decoys. For the docking runs, protein flexibility was not taken into account, and this can be a problem in virtual screening. Typically, the target is represented by a single conformation because allowing conformational changes

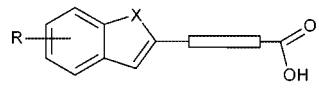
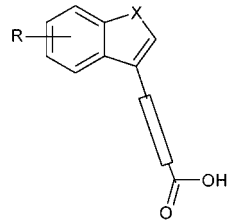
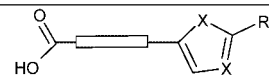
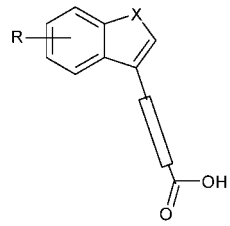
Table 4. Scaffold Hopping Enrichment at 5% (left) and 10% (right) of the Database Screened for RXRα[a]

Method/Class	1	2	3	4	5	6	7	8	9	10
GoldScore										
ChemScore										
Gold IFP										
ROCS										
ROCS IFP										
Bayes										
Bayes IFP										

Method/Class	1	2	3	4	5	6	7	8	9	10
GoldScore										
ChemScore										
Gold IFP										
ROCS										
ROCS IFP										
Bayes										
Bayes IFP										

[a] Color-coding is as follows. Red: not a single active ligand has been retrieved for that particular class; green: at least one ligand has been retrieved for that particular class.

Table 5. Schematic Representation and Physical Chemical Descriptors of the Seven Active Compounds Representing Five New Lead Classes[a]

Name	Schematic diagram	AlogP	NBA	NBD	MPSA
Class 1		6	5	1	94
Class 2		6	2	1	42
Class 3	Proprietary scaffold	6	2	1	42
Class 4		6	5	1	124
Class 5		6	4	1	85

[a] AlogP, NBA (number of hydrogen bond acceptors), NBD (number of hydrogen bond donors), and MPSA (molecular polar surface area, Å^2). X = C, N, O, S. The Tanimoto similarity matrix for each class with respect to the reference ligands can be found in the Supporting Information (Table 3).

during the docking run scales exponentially in the degrees of freedom and CPU time, which poses a problem in VS campaigns due to the large screened databases. A simple way to account for protein flexibility may be to soften the criterion for the steric fit between ligand and receptor by attenuating the repulsive term in the Lennard-Jones potential, thus allowing for a closer approach between ligand and protein.[30] In the present study, we did not soften the potential for the sole reason that a "softer" potential, although better at identifying known true ligands, can lead to many false positives.

Database enrichments of ADA were superior to those of RXRα with ROCS. One plausible reason is that the ligands of RXRα are more hydrophobic than those of ADA, making the feature overlap component of the scoring function undistinguishable for ligands and decoys, increasing the false positive rate. The naı¨ve Bayesian models showed comparable or higher library enrichments than Gold, GlideXP, and ROCS.

It is likely that some of the virtual screening enrichments obtained in this study are lower than those reported in the literature using related targets and VS methods, raising the question of to what extent virtual screening results depend on the database and VS settings. This further illustrates the need of the modeling community adopting the use of benchmarked data sets in order to minimize biases, better facilitating the evaluation and comparison of the VS algorithms.

The ability of the internal scoring functions of Gold and ROCS were thereafter compared with the IFP topological scoring method at database ranking efficiency. Additionally, the Bayesian model trained with IFPs was used to classify the database ligands. The database ranking and scaffold hopping efficiency evaluation of GlideXP with the IFP method was discontinued due to the poor database enrichments obtained with this algorithm for both ADA and RXRα.

Figure 5 shows the database enrichments obtained with the IFP scoring algorithm using Gold and ROCS as pose generators, as well as the Bayesian classifier trained with IFPs. From the comparison of Figures 4 and 5, we can detect that the molecular IFP descriptor is equally suited to quantifying virtual screening success as the well-accepted scoring algorithms of Gold, ROCS, and the naïve Bayesian classifier in the top 5% and 10% of the database screened for RXRα and ADA, respectively.

From the database enrichment results of this study, we propose a Tanimoto coefficient threshold of 0.8 for discriminating true ligands from decoys. A previous study suggested a threshold of 0.6,[6] however, only future applications of this new metric on additional benchmarks will help refine this threshold.

It should be emphasized that our database enrichment approach with IFP scoring was slightly different from previously published constrained VS investigations. In the present case, pose generation was totally unconstrained, and poses were simply postprocessed to select the ones in agreement with the reference molecular interaction fingerprint. Our protocol was also different in spirit from template matching, 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 offer the advantage of reconciling structure-based design with ligand-based data mining. Molecular IFPs are thus a fuzzy but very promising method of selecting virtual poses/hits that satisfy user-defined prerequisites (e.g., the crystallized pose of a ligand in a complex). Furthermore, the IFP method is well-suited for in silico fragment screening, thus providing a promising link between structure-based screening and fragment-based drug discovery.

b. Overall Scaffold Hopping Enrichment. Next, the competence of Gold, ROCS, and a naı¨ve Bayesian classifier in scaffold hopping enrichment was thereafter challenged and compared to those computed with the novel scoring method based on molecular interaction fingerprints.

Scaffold hopping success was used as a measure for the ability of recovering actives from as many structural classes as possible. The scaffold enrichment results for ADA and RXRα are summarized in Tables 3 and 4. The color coding characterizes those classes which are populated with at least one true active (green) or no actives (red). The scaffold hopping examination was done both at 5% and 10% of the database screened. Additionally, the scaffold enrichment results for ADA and RXRα in percentage of actives retrieved of each structural class at 5% and 10% of the database screened can be found in the Supporting Information (Tables 1 and 2).

Subsequently, we performed an analogous scaffold hopping analysis using the IFP topological scoring method using Gold and ROCS as pose generators and a Bayesian classifier learned with IFPs. The classes which encompass the reference ligands are class 1 and 4 for ADA, and class 1 and 10 for RXRα.

From Tables 3 and 4, it can be appreciated that the scaffold efficiencies calculated with the IFP method for ADA and RXRα are appreciably different. A plausible explanation lies again in the differing natures of the ligands of each target. While the ligands of ADA are rather polar, giving rise to rich molecular IFPs, the IFPs of the RXRα ligands are highly aspecific. In this respect, the default IFP rules are partly accountable for the lack of specificity as they do not consider a long conjugated system as aromatic. In essence, a nonspecific molecular IFP leads to a higher number of false positives, thus decreasing the database enrichments and scaffold hopping efficiency.

Regarding the performance of the novel IFP scoring method for a specific target, Tables 3 and 4 show that the scaffold hopping efficiency using the molecular interaction fingerprint in combination with Gold, ROCS, or a naı¨ve Bayesian classifier was higher than the internal scoring algorithms of the VS methods themselves both at 5% and 10% of the screened database. In view of the remarkable performance of the molecular interaction fingerprint as a topological scoring method, a VS campaign of our database was carried out for RXRα in the search of novel lead scaffolds in the next section.

c. Virtual Screening of RXRα. In this section, the VS campaign for RXRα using the molecular interaction fingerprint as a topological scoring method in the search of novel scaffolds is described. Our in-house compound stock was the screening database, and Gold was used as the pose generator. We made a selection of our compound stock of those compounds, which have proven to be active in related nuclear receptors using Pipeline Pilot. Thus, a proprietary collection of 935 compounds was selected for virtual screening with Gold with the options described in the Methodology Section.

All compounds with an IFP Tanimoto coefficient greater or equal to 0.8, with acceptable van der Waals and electrostatic interactions, were filtered (31 compounds). Thereafter, only those compounds with a polar bit turned on with ARG316[31] were selected (nine compounds). Experimental RXRα induced transactivation dose response curves for these nine compounds were determined. The induction of transactivation was measured at the concentration range, 1 nM to 10 μM. Out of the selection of nine compounds, we found that 7 out of 9 were active experimentally, with only two false positives retrieved. The true hits displayed EC_{50}s values between 46 and 477 nM (inactive compounds are defined as those compounds not showing any significant transactivation activity at a concentration of 1 μM). This thus underlines the value of the molecular IFP as an effective, alternative scoring method. More interestingly, 70% of the true positives represent new structural classes for RXRα, affirming the competence of the molecular IFP at scaffold hopping (Table 5).

4. Conclusions

The current study presents a comparative evaluation of a novel scoring method based on molecular interaction fingerprints against the popular virtual screening programs GlideXP, Gold, ROCS, and a Bayesian classifier as to their scaffold hopping efficiency.

We have demonstrated the overall validity of these virtual screening methods for identifying new leads from a pool of ligands with similar physicochemical properties. Moreover, we compared the IFP topological scoring method to the scoring functions of GlideXP, Gold, and ROCS and found that the IFP method exhibits comparable database enrichments and superior scaffold hopping performance. The use of molecular IFPs as an attractive, low-dimensional representation of protein−ligand interactions with promising rank-ordering predictions and with applicability for the virtual identification and profiling of novel scaffolds at the different stages of drug discovery projects is thus recommended.

Acknowledgment. We acknowledge the help of Dr. Antoniou and Dr. McCormack with the proof-reading of the manuscript, and Laboratoire Fournier for providing the transactivation data of RXRα.

Supporting Information Available: Tables 1 and 2 show the scaffold enrichment results for ADA and RXRα in percentage of actives retrieved of each structural class at 5% and 10% of the database screened. Table 3 shows the Tanimoto similarity matrix

of the five novel classes found in the virtual screening of RXRα and its reference ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kraemer, O.; Hazemann, I.; Podjarny, A. D.; Klebe, G. Virtual screening for inhibitors of human aldose reductase. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 814–823.

(2) Jain, A. N. Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403.

(3) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.

(4) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

(5) Seifert, M. H.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discovery Devel.* **2007**, *10*, 298–307.

(6) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195–207.

(7) 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.

(8) Gregori-Puigjane, E.; Mestres, J. SHED: Shannon entropy descriptors from topological feature distributions. *J. Chem. Inf. Model.* **2006**, *46*, 1615–1622.

(9) Zhang, Q.; Muegge, I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J. Med. Chem.* **2006**, *49*, 1536–1548.

(10) Barker, E. J.; Buttar, D.; Cosgrove, D. A.; Gardiner, E. J.; Kitts, P; Willett, P.; Gillet, V. J. Scaffold hopping using clique detection applied to reduced graphs. *J. Chem. Inf. Model.* **2006**, *46*, 503–511.

(11) 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.

(12) Ahlstrom, M. M.; Ridderstrom, M.; Luthman, K.; Zamora, I. Virtual screening and scaffold hopping based on GRID molecular interaction fields. *J. Chem. Inf. Model.* **2005**, *45*, 1313–1323.

(13) Zhao, H. Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective. *Drug Discovery Today* **2007**, *12*, 149–155.

(14) Rush, T. S., III; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.

(15) Naerum, L.; Norskov-Lauritsen, L.; Olesen, P. H. Scaffold hopping and optimization towards libraries of glycogen synthase kinase-3 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1525–1528.

(16) Schrödinger, L. L. C., New York.

(17) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(18) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.

(19) Irwin, J. J.; Shoichet, B. K. ZINC: A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

(20) RCSB Protein Data Bank, http://www.pdb.org.

(21) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.

(22) Prous Science Integrity, http://www.prous.com/integrity.

(23) SYBYL, Tripos International, 1699 Sout Hanley Rd., St. Louis, Missouri, 63144.

(24) OMEGA, OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, New Mexico, 2001.

(25) Bioreason, Inc., 3900 Pase del Sol, Santa Fe, New Mexico 87507 (version 3.5).

(26) 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.

(27) Tanimoto, T. T. *IBM Internal Report*; Nov. 17, 1957.

(28) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.

(29) Scitegic, 9665 Chesapeake Drive, Suite 401, San Diego, California 92123-1365.

(30) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076–5084.

(31) Egea, P. F.; Mitschler, A.; Rochel, N.; Ruff, M.; Chambon, P.; Moras, D. Crystal structure of the human RXRα ligand-binding domain bound to its natural ligand: 9-*cis*-retinoic acid. *EMBO J.* **2000**, *19*, 2592–2601.

JM8001058