A Knowledge-Based Weighting Approach to Ligand-Based Virtual Screening

Nikolaus Stiefl* and Andrea Zaliani

Eli Lilly Research Laboratories, Essener Bogen 7, D-22419 Hamburg, Germany

Received August 12, 2005

On the basis of the recently introduced reduced graph concept of ErG (extending reduced graphs), a straightforward weighting approach to include additional (e.g., structural or SAR) knowledge into similarity searching procedures for virtual screening (wErG) is proposed. This simple procedure is exemplified with three data sets, for which interaction patterns available from X-ray structures of native or peptidomimetic ligands with their target protein are used to significantly improve retrieval rates of known actives from the MDL Drug Report database. The results are compared to those of other virtual screening techniques such as Daylight fingerprints, FTrees, UNITY, and various FlexX docking protocols. Here, it is shown that wErG exhibits a very good and stable performance independent of the target structure. On the basis of this (and the fact that ErG retrieves structurally more dissimilar compounds due to its potential to perform scaffold-hopping), the combination of wErG and FlexX is successfully explored. Overall, wErG is not only an easily applicable weighting procedure that efficiently identifies actives in large data sets but it is also straightforward to understand for both medicinal and computational chemists and can, therefore, be driven by several aspects of project-related knowledge (e.g., X-ray, NMR, SAR, and site-directed mutagenesis) in a very early stage of the hit identification process.

INTRODUCTION

The idea that structurally similar compounds are likely to exhibit similar biological properties is a widely accepted concept in pharmaceutical research.[1−5] Application of this concept, in silico similarity searching, has therefore become a standard method in rational drug design.

In practice, a compound active against a given target structure is encoded by a structural descriptor, and then, in-house or third-party databases are searched for structurally similar compounds by means of a similarity metric. Since the actives often stem from high-throughput screens,[5] multiple actives are frequently available as query structures. Hence, specific focus has recently been given to the question of how to make the best use of the latter additional knowledge. Methods incorporating this information comprise, among others, data fusion,[6,7] binary kernel discrimination,[8] naïve Bayes classification,[9−11] and fingerprint scaling.[12,13] In validation studies,[10,14,15] it was shown that the retrieval rates for these methods are significantly enhanced when compared to using a single query structure alone.

Moreover, this concept is not only true for ligand-based virtual screening protocols. It was recently shown that so-called structural interaction fingerprints (SIFt) significantly improve hit rates of structure-based docking protocols by postprocessing the hitlists of virtual-screening runs.[16,17]

In this paper, the reverse of this latter process, that is, the combination of structural target information and ligand-based similarity searching, will be presented for the newly developed *e*xtended *r*educed *g*raph (ErG) descriptor.[18] ErG can be described as a hybrid approach of reduced graphs as introduced by Gillet et al.[19] and Barker et al.[20] and binding property pairs.[21] Here, advantage was taken especially of the binding property component in order to improve on similarity search hit rates.

The paper is organized as follows. First, the workflow and the theoretical background of the ErG approach are briefly reviewed and the details of other applied virtual screening techniques are given. Next, the fundamentals of the knowledge-based weighting procedure of the ErG descriptor are described and applied to three example structures and data sets. The results obtained are then compared to results obtained by similarity searching using Daylight fingerprints (DFPs),[22] UNITY,[23] and the FTrees[24] technique as well as different FlexX[25] docking protocols. Finally, the applicability of wErG as a prefilter for docking is investigated in terms of numerical as well as a structural results. Reasons for the different behavior of all of the methods are discussed.

METHODS

ErG. ErG can be described as a hybrid approach of reduced graphs[19,20] and binding property pairs[21] that combines a chemically intuitive description of the chemical graph with a good visualization ability.

Property Assignment, Graph Reduction, and Descriptor Calculation. For a detailed description of the ErG procedure, the reader is referred to ref 18. Here, only the most important parts of the methodology are reviewed. These can be divided into three main steps (see also Figure 1).

First, the relevant atoms are formally charged. Then, pharmacophoric properties are assigned to all atoms of the original molecular graph with potential H-bonding capabilities (A, acceptor; D, donor) as well as formally charged atoms ("+", positively charged; "−", negatively charged) on the basis of an *in-house* protocol. Atoms that are assigned both H-bond donor and H-bond acceptor properties are

* Corresponding author: Current address: Novartis Pharma AG, Werk Klybeck, Postfach, CH-4002 Basel, Switzerland. E-mail: nikolaus.stiefl@novartis.com. Phone: +41 61 6963068. Fax: +41 61 6964870.

10.1021/ci050324c CCC: $33.50
Published on Web 02/07/2006

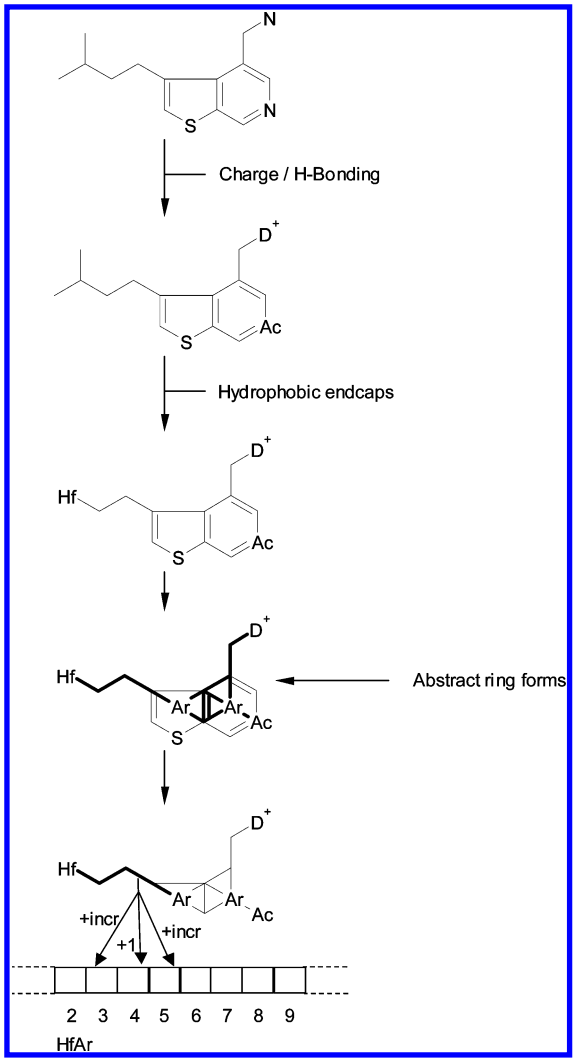


Figure 1. Flow scheme of the ErG procedure exemplified with the help of a sample structure. A fuzzy incrementation with *incr* is applied for the conversion of the reduced graph into the descriptor vector (see text).

identified as "flip-flop" atoms and are handled later in the process in an explicit way.

Next, "endcap" groups that comprise hydrophobic features of three lateral atoms as well as lateral thioethers are assigned a hydrophobic flag (Hf) at the central atom. That way, important size and shape properties of the molecules under study are captured.

In a third step, the ring systems of the molecule are encoded to their reduced form. Here, for each ring of a size less than eight atoms, a ring centroid is generated and assigned an aromatic (Ar for aromatic rings, rings with more than 50% sp^2-hybridized atoms, and aliphatic rings directly attached to aromatic rings) or hydrophobic (Hf for other aliphatic ring systems) flag. These centroids are then connected to all substituted ring atoms as well as bridgehead atoms. All other atoms of the ring system are deleted. That way, intersystem distances of ring and nonring systems are kept at a comparable level. Centroids of highly fused globular ring systems such as adamantane are collapsed into one centroid (see ref 18).

It should be noted that, in ErG, emphasis is placed on the concept that ring systems are handled separately from the H-bonding and charged features. That way, fewer overall property types are used and often similarity can more easily be identified between compounds with different structural skeletons.

After these three steps, atoms of the generated reduced graph are either nonflagged or flagged as charged (positive/negative), H-bonding (acceptor/donor), hydrophobic (endcaps and aliphatic rings), and aromatic. Only the flagged atoms (property points) are subsequently encoded into the numerical ErG descriptor by means of a radial distribution function (RDF).[26] Here, property−property-distance triplets (PPDT) of the type

Property Point 1−topological distance−Property Point 2

are used, where the topological interfeature distance is calculated over the reduced graph. The maximum distance taken into account in ErG is 15, owing to the scarcity of the assigned feature points.[18] With default settings, this results in a descriptor vector **v** of size 315. For each identified PPDT, the corresponding entry in **v** is incremented by 1 and the neighboring distance bins are incremented by *incr* to account for minor distance deviations in the pharmacophoric pattern.[27] On the basis of comparison studies,[18] the default setting of *incr* in the fuzzy incrementation procedure is set to 0.3.

For molecules with assigned flip-flop atoms, multiple vectors are generated by the enumeration of all possible combinations of the H-bonding features. This is done to account for the finding that, in most crystallized ligand−protein complexes, flip-flop atoms seem to interact as either H-bond donor *or* H-bond acceptor but not as both at the same time.[18] For practical applications, the enumeration should be restricted to compounds with five flip-flop atoms to avoid a numerical explosion of the database. However, here, a full enumeration of all of the compounds was performed for reasons of comparability.

Similarity Searching. After generation of the ErG descriptor vector, similarity searching was performed with an algebraic version of the Tanimoto similarity coefficient:[28,29]

$$S_{A,B} = \frac{\sum_{i=1}^{m} n_{A,i} n_{B,i}}{\sum_{i=1}^{m} (n_{A,i})^2 + \sum_{i=1}^{m} (n_{B,i})^2 - \sum_{i=1}^{m} n_{A,i} n_{B,i}} \qquad (1)$$

where m is the size of the ErG vector **v**, $n_{A,i}$ is the entry in vector field i in compound A, and $n_{B,i}$ is the entry in vector field i in compound B.

For each query structure, the database was searched and ranked by decreasing similarity value. Since the actives within the database are known, the results are given as $RT_{1\%}$, that is, the percent of actives found within the first 1% of the ranked database according to

$$RT_{1\%} = \frac{\#\text{actives\_identified}}{\#\text{actives\_overall}} \times 100\% \qquad (2)$$

where #actives_identified is the number of known actives found in the first 1% of the sorted database and #actives_overall is the number of known actives within the whole database.

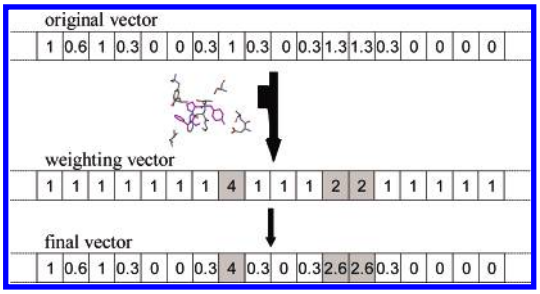


Figure 2. Incorporation of structural knowledge into the similarity search procedure by weighting the original ErG data vector.

Knowledge-Based Weighting. Usually, not all of the assigned pharmacophoric properties of a molecule are part of the pharmacophore important for ligand−receptor interactions. Put differently, when using all possible pharmacophoric points in the ErG descriptor generation, only some are relevant for similarity searching, while others are encoding mainly "noise". This was also identified in our earlier publication.[18] To reduce the amount of noise, different approaches such as normalization procedures or modular fingerprints are viable procedures, depending on the information available to the project.

Owing to the good interpretability of ErG, another method to improve the signal-to-noise ratio is the incorporation of structural knowledge into the search query prior to the actual search procedure. Here, on the basis of specific knowledge derived from, for example, X-ray or NMR data, moieties of the ligand that are known to interact with the receptor are weighted to have a higher influence during the search than other parts of the molecule. This is achieved by the multiplication of the query's ErG vector with a weighting vector **w** to give the search vector **s**.

$$\mathbf{s} = \mathbf{v} \cdot \mathbf{w} \tag{3}$$

where a Hadamard (elementwise) multiplication is applied. **s** is then used for the similarity search procedure of the database. During the search, each database vector is weighted with **w** in the same way as **v** prior to similarity calculation. That way, different weighting vectors **w** (arising from different targets) can easily be applied to a stored database of ErG vectors.

The weighting vector itself is generated as follows. First, the vector **w** is initialized with ones. Next, all pharmacophoric points interacting with the receptor are identified. Then, all possible combinations (PPDTs) of the latter points are generated and the corresponding entries in the weight vector **w** are set to 2 if the PPDT comprises H-bonding properties or hydrophobic points only. Otherwise, that is, if aromatic or charged properties are part of the PPDT, a value of 4 is applied. The latter is done to account for the generally higher interaction energies found for buried charges and aromatic interactions.[30] It is important to note that the full weighting vector is used; that is, all other assigned properties are still taken into account, even though they are given much less weight during the search procedure. That way, the size and shape of the original search query is still included in the search, which would not be true if **w** were initialized with zeros (i.e., variable selection). An example weighting procedure is given in Figure 2. In this study, search procedures with (wErG) and without (ErG) knowledge-based weighting were applied to identify possible enhancements.

Similar in spirit to this weighting approach is the SMARTS searching on reduced graphs proposed by Harper et al.[31] With both methods, structures exhibiting a similar pharmacophoric arrangement can easily be incorporated into the search without the drawbacks of substructure search methods based on the exact molecular structure. That way, additional knowledge (e.g., derived from a crystal structure) can be included in the search procedure. In contrast to the method presented here, compounds identified with SMARTS searches have to exhibit a particular substructure (the SMARTS query). Put differently, if a compound is missing one part of the search query, it will not be identified as interesting. When weighting specific feature combinations, however (as in wErG), compounds with minor deviations from the known search structure still have a good chance to be identified as interesting. Depending on the target of the search, this behavior can be positive or negative. Therefore, it is thinkable that both methods are used in combination (e.g., the more general weighting followed by the more specific SMARTS searching).

Daylight Fingerprints. Since DFPs[22] are used throughout the industry as a standard similarity search method, here, all compounds were also studied by this method. Even though ErG and DFP are rather different in their basic approach, a comparison helps to better understand the performance of the ErG procedure. For DFP similarity searches, results were calculated as for standard ErG but the binary form of the Tanimoto similarity coefficient was used:

$$S_{A,B} = \frac{c}{a + b - c} \tag{4}$$

Here, a is the number of unique fragments in compound A, b is the number of unique fragments in compound B, and c is the number of unique fragments shared by compounds A and B.

FTrees. The second ligand-based search technique applied is FTrees. The FTrees algorithm is a similarity measure based on tree matching, which is described in detail elsewhere.[24] The main difference between FTrees and other molecular descriptors such as ErG or DFP is that FTrees do not work on a linear descriptor vector but directly on a graph representation (the *feature tree*) of the original molecule. The feature tree itself can be described as a type of reduced graph with pharmacophoric properties assigned to each graph node. For similarity searching, subtrees of two feature trees are matched and a similarity measure describing the overall feature tree similarity is calculated on the basis of a weighted average over the similarity of the matches. Hence, FTrees use a "nonstandard" similarity measure as compared to ErG and other linear descriptors. Therefore, and because it was not done before, the comparison of ErG to FTrees is particularly interesting.

Apart from charge assignment, ligand preparation for FTrees was identical to the FlexX protocol. However, since FTrees are sensitive to the assignment of correct formal charges (see FTrees User Guide[32]), two separate runs were performed. One made use of the internal formal charge assigner of the FTrees fcharges.dat file (FTic), and the other used the external in-house charge assignment utility mentioned above (FTch).

FlexX Docking. Prior to virtual screening with FlexX or other docking techniques, different sorts of filters are often applied to reduce the number of structures to be docked and, thus, the computational cost. For this task, pharmacophore search techniques such as UNITY are often employed. However, these methods are still computationally expensive. Hence, one of the main targets of this paper was to investigate the applicability of wErG as a prefilter to docking protocols.

To be able to compare the structures retrieved by the knowledge-based weighting approach to other virtual screening techniques that take structural knowledge into account, a docking protocol, here, FlexX (version 1.13.5), was applied. FlexX is a docking method based on an incremental construction algorithm and is described in detail elsewhere.[25] Prior to docking, all molecules were 2D/3D converted with CORINA[33] and Gasteiger−Hueckel charges were assigned using SYBYL 7.0.[23] A further data set was generated by formally charging all molecules outside FlexX to represent a protonation state at physiological pH on the basis of an *in-house* protocol. This was done to investigate a possible dependence of docking results on the presence of charged moieties. Throughout the paper, the corresponding docking results will be referred to as FX (standard FlexX docking) and FXch (FlexX docking with formally charged molecules), respectively.

Additionally, during this study, a complete version upgrade of FlexX became available (version 2.0). Owing to major changes especially in scoring (access scaling[30]), charge assignment, and atom type recognition, it was interesting to compare the results obtained by this version (FX2) to those obtained by the previous one. It was anticipated that, in particular, access scaling should improve virtual screening results, because of a lower number of molecules docked to the outer rim of the cavity.

Protein preparation was performed with the FlexX module of SYBYL 7.0. First, proteins were loaded and all water molecules were deleted. Then, the binding cavity was defined as all full amino acids with a radius of 9 Å of the cocrystallized ligand. For 1DQ8, a virtual ligand was created (see the Data Sets subsection) and used for cavity definition.

To better take account of other features available in FlexX docking, a third docking approach was employed, in which pharmacophoric constraints were applied for each target active site (FX2ph). For this, the pharmacophore module of FlexX was used. With this module, specific important interactions or combinations thereof can be specified by the user as mandatory for the docking results; that is, a docking solution not fulfilling these constraints is discarded. That way, docking solutions are more restricted to being located within the area of interest, and it is hoped that the rate of false positives will be reduced during the virtual screening run. For this, version 2.0 of FlexX was used because of good results obtained with standard FlexX version 2.0 (see the Results and Discussion section).

For virtual screening purposes, the figure of merit employed is $RT_{1\%}$ (see eq 2). Accordingly, the docked database was ranked on the basis of the scoring value of the native FlexX docking score to give the number of actives in the first 1% of the sorted database.

wErG/FlexX 2.0 Combination. It has previously been observed that, for ErG, the average retrieval rates for the first 5% ($RT_{5\%}$) are generally much larger than those for DFP.[18] The reason for this behavior is the high structural homology of the hits identified by DFP. These actives are mainly enriched toward the top of the DFP-ranked database. From a certain similarity value on, the enrichment in actives is more or less similar to random because of the structural bias of the DFP descriptor. Even though, for ErG, structurally close analogues will also appear at the very top of the ranked database, as a result of the abstract description of the search query, the overall substructural similarity of the actives within the first 1% of the database is often much lower. Accordingly, the cutoff similarity value of "random enrichment" is met at lower levels, resulting in more actives with a broader structural spectrum identified. Hence, it was interesting to see if it would be possible to take advantage of this and use wErG as a prefilter for FlexX to improve retrieval rates (fewer false positives) as well as the computational cost, as compared to standard FlexX docking. That way, 3D information is indirectly included in the wErG approach since compounds with bad steric properties will have low scores within the docking runs. Hence, a similarity search was performed with wErG as above, only that, this time, the top 5% of the sorted database (i.e., 6690 compounds) was used as input for FX2 and FX2ph. These compounds were then ranked by their docking score, and the retrieval rate in the first 1% of the original database size was calculated.

UNITY Searching. To be able to compare the numerical results obtained with wErG to those of a search technique with similar information content (ligand and protein information), a UNITY flexible search (UFS) was performed for all three data sets.[23] UFS employs a so-called directed tweak minimization that is guided by the given pharmacophore query, followed by a selection of reasonable conformations. As a search query, all of the structural features identified for the ErG weighting procedure were used. The respective spatial location tolerances of the pharmacophoric properties were set to 0.5 Å for H-bond donor/acceptor, 1.5 Å for aromatic, 1.8 Å for hydrophobic, and 1 Å for positive/negative charged. Additionally, a spatial constraint was included in the query, to reduce the number of false positives, which can be substantial for more general pharmacophores. Here, a so-called surface volume was used. The latter defines a volume based on the protein surface within the database structure that must completely fit to be considered as a hit. Since the receptor-based surface itself would be a too-strict search criterion, the volume is generated on the basis of a shrunken surface (default amount of shrinkage: 1 Å). The inclusion of spatial constraints into UFS is often difficult, since the right balance between a specific and a general query is often crucial for the number of retrieved compounds. If the query is too specific, only a very low number of compounds is retrieved, whereas if a too-general query is employed, up to 20% of the database compounds are sometimes retrieved. This topic will be briefly discussed for the queries used here. To obtain comparable results, Lipinski filters were not applied prior to the search. Because UFS is often used as a prefilter to docking, the comparison of wErG is particularly interesting.

Data Sets. To evaluate the new approach, three data sets for which structural information about the ligand−target complex was available were identified. Prerequisites for the selection of a data set for evaluation in this study were as

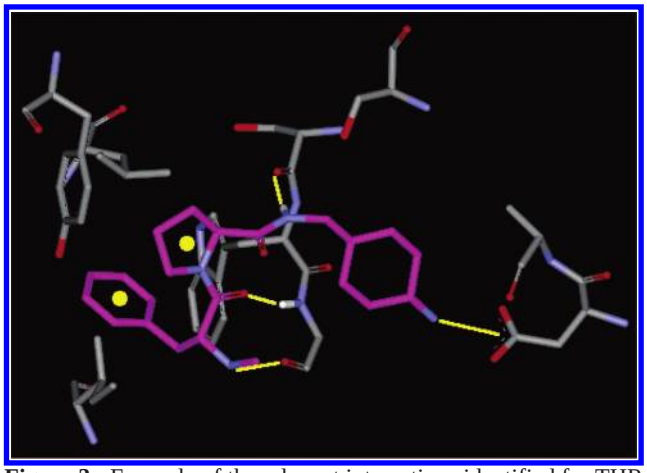


Figure 3. Example of the relevant interactions identified for THR (1TOM; the ligand carbon atoms are displayed in purple for reasons of clarity). Interacting moieties are highlighted in yellow. It should be noted that noninteracting pharmacophoric features such as the aliphatic cyclohexyl ring are not included in the weighting scheme.

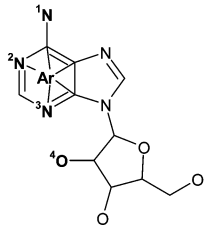
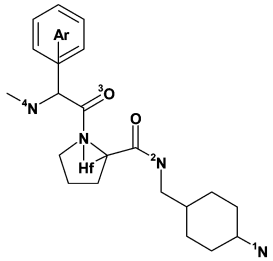
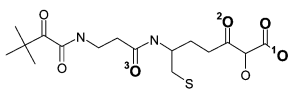
follows. First, since it was found earlier[18] that, especially, peptides as well as peptidimometics have many pharmacophoric points assigned, but only some of them seem to be relevant for ligand−receptor interactions, the cocrystallized ligand should be either a native ligand or a peptide/peptidomimetic. Second, the ligand itself should be of reasonable size. Third, the composition of the data sets should be available and the number of known actives should be representative enough to produce relevant results.

To achieve this, the MDL Drug Data Report database (MDDR)[34] was first "cleaned" by stripping salts as well as removing duplicates, compounds containing atoms usually not appearing in organic compounds (e.g., magnesium), and molecules with a molecular weight of more than 800. This resulted in a set of 133 809 unique molecules. After that, the reduced MDDR was searched to identify activity classes with a minimum of 50 actives and for which at least one crystal structure of the target protein was available in the Brookhaven Protein Databank.[35] That way, a set of adenosine kinase inhibitors (AK; 59 active molecules, PDB code: 1BX4); thrombin inhibitors, the set of which essentially resembles the data set as published in ref 15 (THR; 760 active molecules, PDB code: 1TOM); and HMG-CoA reductase inhibitors (HMG; 1016 active molecules, PDB code: 1DQ8) were selected.

Similar to steps typically taken in medicinal chemistry projects, the ligand−protein interactions of all of the crystal structures were visually inspected and the relevant structural entities identified for both the knowledge-based weighting (see Figure 3) and the pharmacophoric constraints. Here, all H-bond acceptor and donor interactions (including water-mediated contacts) as well as hydrophobic and aromatic contacts between the ligand and protein were identified as important. This kind of analysis could also be performed by an automated approach (e.g., GRID analysis) to reduce the possible bias included in the analysis; however, since it is a common approach to use expert knowledge (i.e., the modeler) for interaction analysis, this approach was followed here too.

The selected structural and pharmacophoric features for both methods are given in Table 1.

Table 1. Structural Features of the Query Structures Identified as Relevant for Ligand−Receptor Interactions[a]

Data	ErG	FX2ph	UFS
AK		ASN 14 (ND2): essential H-Donor	N1: donor atom N2: acceptor atom N3: acceptor atom O4: donor atom Ar
THR		ASP 189 (OD2): optional H-Acceptor, GLY 219: optional H-Acceptor At least 1 constraint has to be satisfied	N1: positive N N2: donor atom O3: acceptor atom N4: donor atom Ar Hf
HMG		LYS 735: essential H-Donor	O1: negative center O2: acceptor atom O3: acceptor atom

[a] For wErG, the relevant weighted atoms are presented in boldface, and for FXpharm, the specific pharmacophoric constraints are given. For UFS, the pharmacophoric properties are assigned to the same boldface atoms as for wErG. For H-bond donor/acceptor atoms, the respective assignment is corresponding to the numerical label. (D, donor atom; A, acceptor atom; Hf, hydrophobic center; Ar, aromatic center.)

Table 2. Numerical Results ($RT_{1\%}$) Obtained for All Three Data Sets and All Employed Techniques[a]

data	DFP	ErG	FTic	FTch	wErG	UFS[b]	FX	FXch	FX2	FX2ph	Rnd
AK	13.6	10.2	*18.6*	*18.6*	**22.0**	8.5	3.4	6.8	5.1	11.9	0.04
THR	18.3	21.1	*25.3*	23.3	**29.3**	5.9	4.9	1.6	11.3	12.2	0.56
HMG	0.9	0.9	0.1	*1.1*	**3.4**	0.8	0.3	0.1	0.0	0.5	0.76

[a] The top performing technique is printed in boldface and the second best in italics. Random retrieval rates are denoted by Rnd. [b] Results obtained with spatial constraints. For UFS, only the compounds fulfilling the whole pharmacophore are obtained; i.e., if the number of retrieved compounds is less than 1%, only those compounds are used.

For HMG, a "virtual needle" was generated (ErG) to account for the interactions of hydroxymethylglutaryl as well as coenzyme A. Here, both structures were connected by two additional carbon atoms to give a single compound. That way, the approximate distances between the two moieties are retained. To reduce the size of this virtual needle, the newly generated molecule was cleaved in such a way that the lateral part of coenzyme A (including the adenosine, sugar, and phosphate substructures) was deleted. The generated needle is displayed in Table 1.

RESULTS AND DISCUSSION

For reasons of clarity, first, general comments on the results are given and then more specific details on the different targets are highlighted. Given that the main interest in this study is the inclusion of structural knowledge in the search procedure, the specific comments are restricted to results that included structural knowledge of the target site (i.e., FlexX docking and wErG).

An overview of the numerical results obtained with all of the different approaches is listed in Table 2.

It can be nicely seen that, in terms of absolute numbers, the standard ErG approach and FTrees perform similarly to DFP (see Table 2), with FT being seemingly the best general approach of the pure similarity search techniques. However, when inspecting the overlay of actives identified by FTrees, it becomes apparent that sometimes hits are identified by "wrong" subgraph matching (e.g, EXTREG:323106), especially if internal charge assignment is applied. Hence, the slightly better performance of FT as compared to the other similarity search techniques has to be put into perspective. Interestingly, the performance of FTch seems to be more robust than FTrees with internal charge assignment (especially for HMG—retrieval rates are above random only for FTch). Although the data sets are not large enough for a comparative study in this case, charge assignment seems to influence FTrees results (see also FTrees User Guide[32]). Hence, for the rest of the paper, only the results obtained with the more robust FTch will be discussed in detail.

A comparison of the techniques including structural knowledge of the target highlights the very good and robust performance of wErG independent of the target. In particular, the good overall performance (wErG is the overall best performing technique for all three data sets) is an interesting feature for a virtual screening technique, since it gives confidence for future applications. The obvious improvement of wErG as compared to standard ErG indicates that the signal-to-noise ratio has increased significantly. In summary, wErG exhibits the best overall performance as compared to all of the other techniques. Therefore, the easy translation of important interaction patterns into the ErG descriptor vector, which permits this straightforward weighting approach, ensures the success of the wErG technique.

It should be noted that, as a result of the spatially constrained queries used, the results given for UFS are hard to compare with those of the other techniques. For the three data sets, the numbers of overall retrieved compounds are 91 (AK), 59 (THR), and 431 (HMG), which is less than the first 1% of the ranked database used by the other methods. One can nicely see that the more specific queries (AK and THR) retrieve much fewer compounds than the more general three-point pharmacophore from HMG. If no surface volume is used, the number of retrieved compounds for the three data sets increases to 3383 (AK), 750 (THR), and 13180 (HMG), which is far beyond the 1% limit used for the other techniques. Ranking of the AK and HMG hitlist within UNITY to obtain $RT_{1\%}$ values on the basis of a torsional minimization procedure is a very time-consuming step, making this procedure a nearly obsolete step if UFS is used as a prefilter to docking. Nevertheless, for AK, THR, and HMG, the retrieval rates are changed to 16.9%, 19.3%, and 1.4%, respectively, which is comparable to the similarity measures without structural knowledge but worse than wErG. Hence, the results obtained with UFS are not discussed in more detail. Of course, one could try to optimize the UNITY queries if more knowledge is available about which interaction patterns are essential and which do not have to be fulfilled (i.e., partial matching). However, this is not within the scope of this article, especially since the focus here is put on an early project phase with little knowledge of potential multiple binding modes.

In contrast to AK and THR, for HMG, the results are close to random (random $RT_{1\%} = 0.76$) for all "standard" methods. Here, only for one technique, for which additional knowledge is included in the screening procedure (wErG), are improvements observed.

Another interesting aspect of the comparison of DFP with standard ErG and FTch is the orthogonality of the methods, that is, the number of compounds identified by all of the methods and those specific to a single technique (see Table 3).

It can be clearly seen that, as a general rule, none of the methods overlaps to a large extent with any of the others. One exception is AK, where DFP and FTch show a very high degree of overlap (the DFP hits are all identified by FTch). Additionally, caution needs to be taken when interpreting the results for HMG, as a result of the low retrieval rates. Nevertheless, all three methods can be described as complementary to each other for the data sets under study. This is in line with findings reported in the FTrees and ErG original publications.[24,18]

However, since structural diversity is of major interest during a similarity search, not only were the mere numbers

Table 3. Overlap and Specificity of the Actives Identified by DFP (D), Standard ErG (E), and FTch (F)[a]

data set	D&E&F[b]	D&E[c]	D&F	E&F	D sp[d]	E sp	F sp	all[e]
AK	3 (6)	3 (6)	8 (8)	3 (6)	0	3	3	59
THR	57 (139)	80 (139)	85 (139)	108 (160)	31	29	41	760
HMG	0 (9)	2 (9)	0 (9)	2 (9)	7	5	9	1016

[a] The number of actives identified by the corresponding worst technique is given in parentheses. [b] "&" denotes the overlap between actives identified by the corresponding techniques; i.e., for D&E&F, the number given describes the number of actives identified by all methods (overlap). [c] Same as for footnote *b*, only that this time, only two techniques are compared. [d] "sp" denotes the number of actives identified only by the corresponding technique and by none of the others (specific). [e] Number of all known actives.

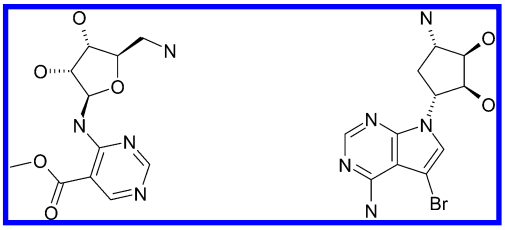


Figure 4. Sample structures of AK actives specifically identified by ErG (left) and FTch (right) only. Each method identifies a certain cluster of actives.

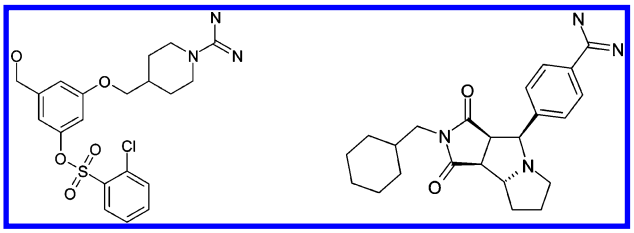


Figure 5. THR-active compounds not exhibiting a "peptide-like" appearance identified by ErG (left, EXTREG:250945) and FTch (right, EXTREG:242337).

of overlapping structures investigated but also a closer look was taken at the corresponding compounds specific for each method. For AK, the main difference between the compounds identified specifically by standard ErG and FTch, respectively, can be described thus: FTch identifies a cluster of compounds with an adenosine analogue two-ring system, whereas ErG is able to identify actives with a single ring system as active (see Figure 4).

For THR, given the high numbers of actives identified, the focus was shifted toward the ability of each technique to retrieve actives deviating from the query's "peptidic" structural appearance. Since the THR data set contains mainly other peptidomimetics, identifying any of the few other structural classes without changes to the search protocol[36] is a challenging task. For DFP, no such structure was identified; however, for ErG and FTch, one compound each was identified as really structurally different (see Figure 5).

The overall performance of the different docking approaches in comparison to those of the similarity search procedures supports the findings of ref 3. Nevertheless, given that docking is a frequently applied technique for virtual screening in industry as well as academia, higher retrieval rates might have been expected. One possible explanation could be the historical growth of the database and, with it, the high similarity of some of the active compounds (i.e., similarity search methods will perform better). Another explanation could be that FlexX is not suited for any of the three different binding cavities. However, given that the pockets' properties are not very similar, the latter reason is not too probable.

Overall, docking with externally applied formal charges did not show any general trend in terms of improvement of the results obtained with standard FlexX docking (FX). Given the low number of different data sets, no general statement on the reasons for this can be given here; however, for the rest of the paper, FXch results will not be discussed further.

In contrast to FXch, FX2 and FX2ph docking improved the results for AK as well as THR as compared to FX. For HMG, however, results stayed at a random level. Hence, at least for the data sets and active sites investigated here, FX2 and FX2ph seem to be the more appropriate methods of FlexX docking for virtual screening, and an improvement between the two versions of FlexX can be observed. It seems that the new access-scaling of FlexX 2.0 in combination with the pharmacophoric constraints seems to positively influence performance and makes significant differences in scoring the respective docking poses. Hence, for the sake of brevity, throughout the rest of the manuscript, general statements holding for all of the docking results will be given; however, detailed discussions of the results will be restricted to FX2ph.

With respect to computational costs, database docking for each target with FX took approximately 2.5 days on eight nodes with dual 3.06 GHz CPUs of a blade server running RedHat Enterprise 3.0. For FX2 and FX2ph, this time was reduced to approximately 1.5 days. For ErG and wErG, similarity searching on a single such processor took approximately 15 s for each query structure.

AK. For the adenosine kinase inhibitor data set, six out of seven actives identified by FX2ph were all also identified by wErg (13 actives), suggesting that, for this data set, the docking methods are not able to select actives highly dissimilar to the cocrystallized ligand (see Figure 6a).

Given the rather deep and small pocket of 1BX4, this is not surprising since, on one hand, larger, more dissimilar ligands (see Figure 6b) are hard to fit into the pocket without any relaxation of the protein or movement of the side chains and, on the other hand if a larger, inactive ligand fits well into the pocket, it is not highly penalized for its size as compared to lower-scoring smaller active ligands. Hence, false positives will reduce the retrieval rates for this data set.

For wErG, the retrieved active compounds are structurally still similar to the needle; however, minor structural changes such as a reduced aromatic ring system as well as different substitution patterns can be identified. One reason for the small variations in the identified actives is the high density of pharmacophoric points given the size of the needle structure. If such a needle is used for searching databases, the reduced graph approach is prone to the retrieval of other compounds with similar pharmacophoric densities.[18] Hence,

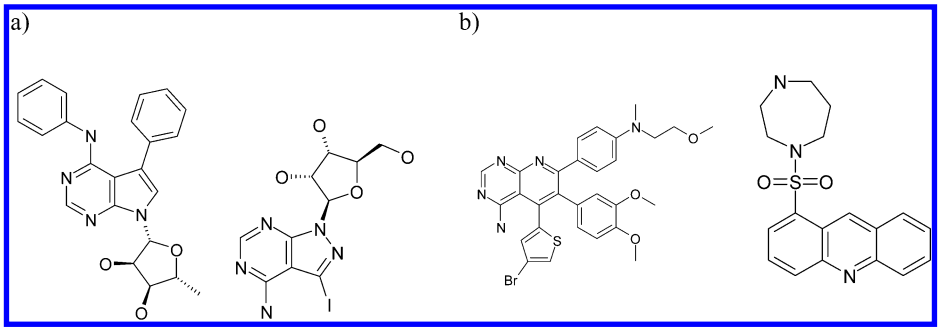


Figure 6. (a) Example structures of AK actives identified by FX2ph (right) and wErG (left). (b) Examples of larger actives and actives with fewer pharmacophoric points that were not identified by any of the methods.

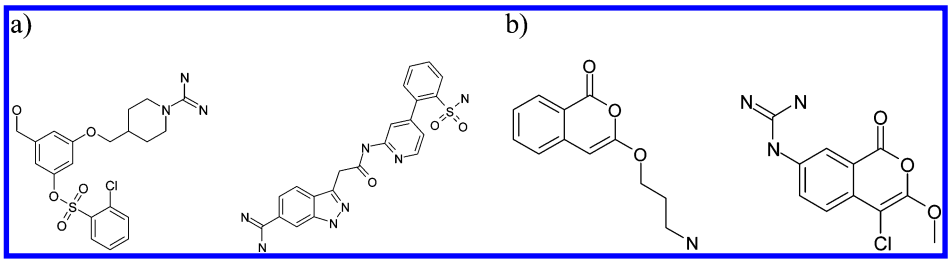


Figure 7. (a) Examples of THR nonpeptidic structures identified by wErG (left) and FX2ph (right). (b) Example structures of small, nonpeptidic THR actives that are not identified by any of the methods.

other actives with fewer pharmacophoric points (e.g., see Figure 6b) are not identified.

THR. For the thrombin inhibitor data set, the actives identified by FX2ph (38 actives) showed an overlap of only 15 with wErG, even though wErG retrieves many more actives than all of the other docking methods (wErG: 223 actives). This suggests that, at least to a certain extent, all of the methods identify a similar set of actives, meaning none of the methods behaves completely orthogonal to the others. In line with this is the finding that, even though all of the methods are capable of identifying actives with nonpeptidic characteristics (see Figure 7a), none of the methods is able to select any of the smaller, nonpeptidic compounds available in the list of active thrombin inhibitors (see Figure 7b).

For wErG, this is not really surprising, because the standard ErG methodology usually retrieves compounds that are not only similar with respect to pharmacophoric properties but also similar in size. However, since docking methods do not depend on a specific needle, smaller molecules that are able to interact well with the protein should be picked up by this method more easily. One possible explanation for not identifying these actives by docking is that, even though they are assigned an active flag, their activity is much lower than that of the peptidic needles and, therefore, they *should* have a much lower docking score. Nevertheless, other possible reasons include such problems as using a static binding pocket and the dependence of the scoring on molecular size. Here, an approach to include negative training data in scoring like that taken, for example, by Pham et al. is a promising step to improve upon the latter issue.[37]

HMG. In contrast to AK and THR, HMG is a particularly difficult data set for both techniques—docking as well as reduced graph similarity searching (see also Table 2). This is due to two different reasons: the binding site and the structure of the known actives. First, the binding cavity found in the crystal structure as well as the ligand is very large. Here, docking without spatial or pharmacophoric constraints will result in mainly large compounds achieving high docking scores, because of the numerous interactions these compounds can pick up anywhere within the active site. Hence, even though smaller ligands will still have reasonably high docking scores, they will usually not show up among the best solutions. Second, another, but no less severe, problem can be identified when inspecting the "virtual" query structure. Within ErG and other reduced graph approaches,[20] amidic substructures are not assigned a specific hydrophobic/aromatic/π-electron feature (except in FTrees). Hence, since most of the known actives exhibit a hydrophobic ring system at the position where the amide bonds of the query structure are located, these compounds are very difficult to identify by ErG. Here, if the user were more interested in more aromatic compounds, the virtual needle could be easily adapted to incorporate at least one such hydrophobic/aromatic group with the corresponding H-bond acceptor/donor properties. Additionally, most of the known actives comprise a lactone ring, even though they interact with the protein in their ring-open form (carboxylate interacting with LYS 735); that is, the structures of these compounds in the MDDR are a prodrug form. The latter is problematic for both similarity searching as well as docking and is reflected by the fact that wErG, the only method with retrieval rates significantly higher than random, is not able to retrieve any of the prodrug forms as active. Moreover, given that the groups interacting with the receptor active site of the ring-open form are geometrically completely differently located from where they would be theoretically in a closed-ring form, it is good to see that none of these molecules is picked up, especially by the docking methods.

Surprisingly, even though the influence of the cavity size is reduced for FX2ph, for HMG no increase in performance between FX and FX2ph is observed. Reasons for this might be the importance of a correct charge assignment and a corresponding proper weighting of the charge influence in the docking score. Although the pharmacophoric constraint improves the placement of the base fragment toward the carboxylate group and, therefore, more known actives will

Table 4. Retrieval Rates ($RT_{1\%}$) Obtained by Docking Only the Top 5% of a Previous wErG Run In Comparison to Docking the Full Database with FX2 and FX2ph

data set	wErG/FX2	FX2	wErG/FX2ph	FX2ph
AK	15.25	5.08	15.25	11.86
THR	30.00	11.32	27.63	12.24
HMG	0.49	0.00	2.46	0.49

be built up in a "correct" pose for final scoring, many other compounds are ranked higher than the known actives. Accordingly, also for HMG, the simple weighting included in wErG results in the best overall performance between the techniques.

wErG/FlexX Docking Combination. For AK, THR, and HMG, roughly 59%, 49%, and 9%, respectively, of all the actives are found within the first 5% of the wErG similarity sorted database. For HMG, the number is comparatively low due to the high number of prodrugs in the set of actives. Nevertheless, the actives within the first 5% of wErG included, for AK, 100% (FX2 and FX2ph); for THR, 69% (FX2) and 66% (FX2ph); and for HMG, 0% (FX2ph) of the actives identified by docking.

As can be clearly seen in Table 4, the combination of wErG with the docking methods results in comparatively high retrieval rates given that only 5% of the whole database is docked. For THR, the number of actives retrieved is the highest found with any of the methods employed.

Again, HMG behaves very differently than the other two data sets. Even though a direct comparison is hard to perform because of the low retrieval rates of both docking techniques, a closer inspection of the actives identified with FX2ph and wErG/FX2ph showed that actives retrieved by docking only comprised one substructure, representing the hydroxymethylglutaryl moiety attached to a flat aromatic central ring system, whereas if wErG filtering was applied, the central ring system is more nonaromatic with multiple substituents. Hence, the overall shape of most of the wErG-filtered actives is much more bulky than that of the actives retrieved by docking only. Overlay, of the wErG-filtered active compounds with the enzyme−ligand complex of the also quite bulky Mevastatin (PDB code: 1HW9[38]) as well as the native ligand−enzyme complex used for docking (1DQ8), highlights a steric clash of Mevastatin with Leu862 of the 1DQ8 crystal. As identified by Istvan et al.,[38] the C-terminal end of HMG-CoA reductase undergoes significant movements upon statine binding to enlarge the binding pocket for the inhibitors, which is not reflected in the binding pocket used for docking. Even though this movement also occurs upon binding of the more flat actives identified by docking only, these molecules can also be fitted into the smaller native pocket, which is not true for the more bulky ligands more similar to the virtual needle.

Overall, in this specific case, the actives identified by the wErG/docking approach are more diverse in terms of overall shape. However, one has to keep in mind that this might also be due to the low actual retrieval rate. Of course, it can be seen that the results obtained by this prefilter depend somewhat on the virtual needle applied. Nevertheless, if more information is available within a specific project, the virtual needle can be tuned to include as much information available as possible. Additionally, because of the short calculation time of wErG, running similarity searches with multiple needles usually still improves on the final computational costs, making the wErG prefilter for docking an interesting tool for virtual screening.

CONCLUSION

In this paper, a knowledge-based weighting approach to similarity searching based on the ErG methodology is presented (wErG). The inclusion of structural and other relevant knowledge (e.g., mutation experiments) in a virtual screening protocol is simple to achieve by this procedure because of the good comprehensibility of the single ErG vector entries; that is, every vector entry encodes a specific property−property-distance combination. That way, it is possible to include the intuition and knowledge of medicinal and computational chemists in a straightforward manner and consequently improve upon the retrieval rates of the virtual screen. Additionally, the signal-to-noise ratio of assigned pharmacophoric properties of the reduced graph and interacting pharmacophoric points is enhanced.

The new procedure is applied to three data sets with different properties, showing that, for all of them, the inclusion of specific structural knowledge (either by wErG or by pharmacophoric constraints of a docking protocol) improves the retrieval rates of known actives. Even the usage of a "virtual" query structure (HMG) generated on the basis of disconnected native ligands resulted in reasonable retrieval rates of actives. Hence, even the inclusion of knowledge about the binding modes of native ligands or derived pharmacophore patterns should help in the identification of novel lead structures.

Additionally, the combination of wErG and FlexX docking exhibits promising results, suggesting that prefiltering with a similarity searching technique able to perform scaffold-hopping still shows enough diversity to be valuable as a standard tool in reducing the computational cost of docking large libraries. This, in combination with the enhancements in speed and retrieval rates identified with the new FlexX 2.0 version, allows the usage of virtual screening of large libraries even on small clusters or single CPU machines.

In summary, with a simple weighting scheme that is easy to understand for both medicinal and computational chemists, the retrieval rates for three targets crystallized with native ligands and a peptidic analogue were significantly improved. Therefore, this knowledge-based weighting of similarity searching techniques is an easy to apply procedure to efficiently identify actives in large data sets. Especially in combination with more generic methods such as reduced graphs, it results in diverse hitlists, which then can then be further pruned by docking to include more 3D and structural knowledge.

ACKNOWLEDGMENT

The authors thank Knut Baumann, Michael Bodkin, Howard Broughton, and Ian Watson for fruitful discussions.

REFERENCES AND NOTES

(1) Cramer, R. D., III.; Redl, G.; Berkoff, C. E. A Novel Approach to the Problem of Drug Design. *J. Med. Chem.* **1974**, *17*, 533−535.

(2) Bajorath, J. Virtual Screening in Drug Discovery: Methods, Expectations and Reality. *Curr. Drug Discovery* **2002**, *3*, 25−28.

(3) Sheridan, R. P.; Kearsley, S. K. Why do we need so many Chemical Similarity Search Methods? *Drug Discovery Today* **2002**, *7*, 903−911.
(4) Schneider, G.; Böhm, H.-J. Virtual Screening and Fast Automated Docking Methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(5) Bender, A.; Glen, R. C. Molecular Similarity: a Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.
(6) Ginn, C. M. R.; Willett, P.; Bradshaw, J. Combination of Molecular Similarity Measures Using Data Fusion. *Perspect. Drug Discovery Des.* **2000**, *20*, 1−16.
(7) Salim, N.; Holliday, J.; Willett, P. Combination of Fingerprint-Based Similarity Coefficients Using Data Fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435−442.
(8) Harper, G.; Bradshaw, J.; Gittins, J. C.; Green, D. V. S.; Leach, A. R. Prediction of Biological Activity for High-Throughput Screening Using Binary Kernel Discrimination. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1295−1300.
(9) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 170−178.
(10) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708−1718.
(11) Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. Molecular Surface Point Environments for Virtual Screening and the Elucidation of Binding Patterns (MOLPRINT 3D). *J. Med. Chem.* **2004**, *47*, 6569−6583.
(12) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Profile Scaling Increases the Similarity Search Performance of Molecular Fingerprints Containing Numerical Descriptors and Structural Keys. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1218−1225.
(13) Xue, L.; Stahura, F. L.; Bajorath, J. Similarity Search Profiling Reveals Effects of Fingerprint Scaling in Virtual Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2032−2039.
(14) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Topological Descriptors for Similarity-Based Virtual Screening Using Multiple Bioactive Reference Structures. *Org. Biomol. Chem.* **2004**, *2*, 3256−3266.
(15) 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.
(16) 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.
(17) 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.
(18) Stiefl, N.; Watson, I. A.; Baumann, K.; Zaliani, A. ErG: 2D-Pharmacophore Descriptions for Scaffold-Hopping. *J. Chem. Inf. Model.* **2006**, *46*, 208−220.
(19) Gillet, V. J.; Willett, P.; Bradshaw, J. Similarity Searching Using Reduced Graphs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 338−345.
(20) Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. Further Development of Reduced Graphs for Identifying Bioactive Compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 346−356.
(21) Kearsley, S. K.; Sallamack, S.; Fluder, E. M.; Andose, J. D.; Mosley, R. T.; Sheridan, R. P. Chemical Similarity Using Physiochemical Property Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 118−127.
(22) DAYLIGHT Inc., Mission Viejo, CA.
(23) *SYBYL*, version 7.0; Tripos Inc.: St. Louis, Missouri, 2004.
(24) Rarey, M.; Dixon, J. S. Feature Trees: A New Molecular Similarity Measure Based on Tree Matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471−490.
(25) 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.
(26) Baumann, K. An Alignment-Independent Versatile Structure Descriptor for QSAR and QSPR Based on the Distribution of Molecular Features. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 26−35.
(27) Stiefl, N.; Baumann, K. Mapping Property Distributions of Molecular Surfaces: Algorithm and Evaluation of a Novel 3D Quantitative Structure−Activity Relationship Technique. *J. Med. Chem.* **2003**, *46*, 1390−1407.
(28) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(29) Chen, X.; Reynolds, C. H. Performance of Similarity Measures in 2D Fragment-Based Similarity Searching: Comparison of Structural Descriptors and Similarity Coefficients. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1407−1414.
(30) Stahl, M.; Böhm, H.-J. Development of Filter Functions for Protein−Ligand Docking. *J. Mol. Graphics Modell.* **1998**, *16*, 121−132.
(31) Harper, G.; Bravi, G. S.; Pickett, S. D.; Hussain, J.; Green, D. V. S. The Reduced Graph Descriptor in Virtual Screening and Data-Driven Clustering of High-Throughput Screening Data. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2145−2156.
(32) The FTrees User Guide can be obtained from BIOSOLVEIT. http://www.biosolveit.de (accessed May 2005).
(33) *CORINA*; Molecular Networks: Erlangen, Germany.
(34) The MDL Drug Data Report database is available from MDL Information Systems Inc., San Leandro, CA.
(35) Brookhaven Protein Databank. http://www.rcsb.org (accessed May 2005).
(36) Sheridan, R. P.; Singh, S. B.; Fluder, E. M.; Kearsley, S. K. Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1395−1406.
(37) Pham, T. A.; Jain, A. N. Parameter Estimation for Scoring Protein−Ligand Interactions Using Negative Training Data. *J. Med. Chem.* [ASAP article] **2005**, DOI: 10.1021/jm050040j.
(38) Istvan, E. S.; Deisenhofer, J. Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Science* **2001**, *292*, 1160−1164.

CI050324C