

Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors

Dennis M. Krüger^[a] and Andreas Evers^{*[b]}

Structure- and ligand-based virtual-screening methods (docking, 2D- and 3D-similarity searching) were analysed for their effectiveness in virtual screening against four different targets: angiotensin-converting enzyme (ACE), cyclooxygenase 2 (COX-2), thrombin and human immunodeficiency virus 1 (HIV-1) protease. The relative performance of the tools was compared by examining their ability to recognise known active compounds from a set of actives and nonactives. Furthermore, we investigated whether the application of different virtual-screening methods in parallel provides complementary or redundant hit lists. Docking was performed with GOLD, Glide, FlexX and Surflex. The obtained docking poses were rescored by using nine different scoring functions in addition to the scoring functions

implemented as objective functions in the docking algorithms. Ligand-based virtual screening was done with ROCS (3D-similarity searching), Feature Trees and Scitegic Functional Fingerprints (2D-similarity searching). The results show that structure- and ligand-based virtual-screening methods provide comparable enrichments in detecting active compounds. Interestingly, the hit lists that are obtained from different virtual-screening methods are generally highly complementary. These results suggest that a parallel application of different structure- and ligand-based virtual-screening methods increases the chance of identifying more (and more diverse) active compounds from a virtual-screening campaign.

Introduction

In the last few years, virtual screening (VS) has become a major part of modern drug discovery for the fast and effective identification of novel bioactive ligands from large compound databases.^[1,2] Numerous structure- and ligand-based approaches have been reported.^[3–6]

Ligand-based virtual-screening methods differ in the descriptors of the molecular structures and properties and the metric used to describe the similarity between molecules. Several approaches allow molecules to be combined into a model that can itself be used as reference for ligand-based virtual screening.^[3,7–9] Different methodologies can make use of either 2D or 3D descriptors. Given 3D conformations of one or more active ligands derived from structure determination methods or from molecular modelling, 3D-similarity^[5] or pharmacophore searches^[10] represent an option for the virtual screening of compound libraries.

Finally, once the structure of the target protein is known, ligands can be subjected to molecular docking and scoring to provide potential candidates for experimental testing. Due to the high dimensionality of the configuration space and the complexity of the energetics governing the protein–ligand interactions, the docking approach is generally the computationally most demanding procedure. Therefore, different virtual-screening methods are often applied sequentially, with hierarchical filters applied due to the different complexities of the different screening steps with respect to their computational requirements. Several success stories based on this cascaded virtual-screening approach have been reported in literature.^[11–19]

The recent literature covers several retrospective comparisons of different structure- or ligand-based virtual-screening tools.^[11,12,20–28] One study was performed in our group. We compared the relative performance of structure- and ligand-based virtual-screening methods for four biogenic amine-binding GPCRs.^[11] In that study, ligand-based virtual screening was superior to docking. However, several aspects of our study were not “fair”: 1) The protein structures were not derived by crystallography but by homology modelling (based on the crystal structure of bovine rhodopsin). 2) For the generation of ligand-based (pharmacophore, Feature Tree and 2D PLS) models, multiple ligands were used as training sets, whereas for docking, we considered only one (rigid) protein structure for each target. With the present study, we would like to compare the performance when docking is performed into **one**

[a] D. M. Krüger
Institut für pharmazeutische und medizinische Chemie
Heinrich-Heine-Universität Düsseldorf
Universitätsstrasse 1, 40225 Düsseldorf (Germany)

[b] Dr. A. Evers
Research & Development, Drug Design, Sanofi–Aventis Deutschland GmbH
65926 Frankfurt am Main (Germany)
Fax: (+49) 69-331399
E-mail: Andreas.Evers@sanofi-aventis.com

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cmdc.200900314>: Eight tables with enrichment factors derived from all scoring functions applied to each of the docking algorithms and one table with enrichment factors derived from all scoring functions applied to ROCS. Furthermore, the MDDR entry numbers of active and decoy compounds are provided.

rigid protein structure and ligand-based searches are based (only) on **one** reference ligand. Another aspect that has not been addressed either in our study or in others is the question in how far the different virtual-screening approaches provide complementary or redundant hit lists. The question to ask when assessing different tools for a virtual-screening campaign might indeed not only be which tool to prefer over the other but which tools should be used in parallel or how different methods should be combined with each other.

This study compares the performance of structure- and ligand-based virtual-screening tools for four targets (ACE, COX-2, HIV and thrombin) whose protein structure is known from crystallography. In addition, we shall investigate the complementarity of the hit lists generated by the different approaches.

For the structure-based screening of the virtual compound libraries, we used the docking programs Glide,^[29] GOLD,^[30] Surflex^[31] and FlexX.^[32] The obtained docking poses were rescored and ranked with different scoring functions (see the “Experimental” Section). For ligand-based virtual screening, we applied 3D-similarity searches (using ROCS^[7]), and 2D-similarity searches using Feature Trees^[5] or Scitegic Functional Fingerprints (FCFP4).

Computational Methods

Virtual screening targets

Protein structures, reference ligands, screening set: In order to allow for a fair comparison between the different approaches, we used one (rigid) protein crystal structure for the docking approach (neglecting protein flexibility), and one ligand as reference for the ligand-based virtual-screening approaches (neglecting the existence of further active ligands with different scaffolds). The protein crystal structures were obtained from the Protein Data Bank (PDB).^[33] The query molecules used as reference for ligand-based virtual screening were extracted from these crystal structures (Figure 1).

To compare the performance of the different virtual screening protocols, we compiled diverse screening data sets of 50 “active” and 950 inactive compounds for each target, extracted from the MDL Drug Data Report (MDDR). Since the inactive data set might still contain actives simply because activity annotation might be missing in the MDDR, we will refer to these as decoys. Further details are described in the “Experimental” Section.

Model generation and retrospective virtual screening

Docking and scoring: Ligands from each screening data set were docked into the associated protein cocrystal structures by using GOLD, Glide (SP and XP), FlexX and Surflex. In all cases, standard parameters were used. All docking poses were rescored and ranked by using nine different scoring functions in addition to the scoring functions implemented in the docking algorithms. To investigate whether considering knowledge about ligand binding further improves the docking results, we imposed “interaction constraints” on the docking procedure for the best docking program (Glide SP; see the Experimental Section).

3D-similarity searching (ROCS): ROCS was used to perform shape-based overlays of conformers, generated with OMEGA,^[34] onto the reference ligand extracted from the appropriate PDB file. The reference ligand was provided as 1) its bioactive conformation, as ex-

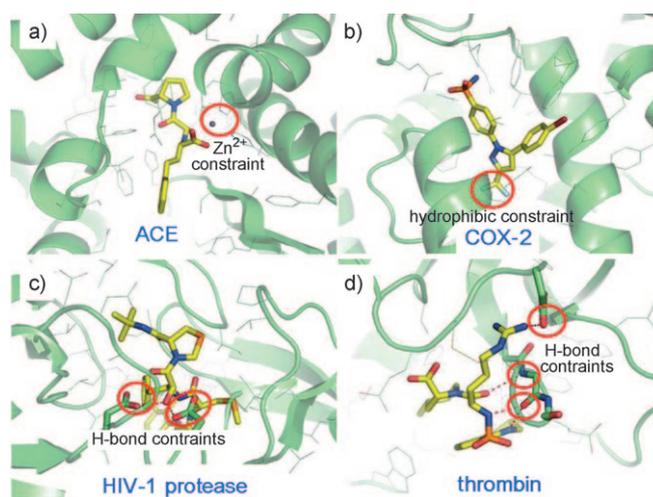


Figure 1. Protein structures and their complexed ligands that were used as references for virtual screening. The docking constraints are ringed in red. a) ACE crystal structure (PDB: 1uze^[49]) complexed with enalapril. As docking constraint, we imposed a hydrogen-bond acceptor able to interact with the Zn²⁺ ion. b) COX-2 crystal structure (PDB: 1cx2^[50]) complexed with S58. A hydrophobic feature that is common to all crystallized COX-2 inhibitors was used as constraint for docking. c) HIV-1 protease (PDB: 1hpx^[52]) complexed with KNI-272. Hydrogen-bonds to the catalytic aspartates were imposed as docking constraints. d) Thrombin (PDB: 1dwc^[53]) complexed with MD805. Three hydrogen-bond contacts were defined as constraints for docking.

tracted from the PDB file, 2) the Omega conformer closest to this cocrystallised conformation or 3) the global-energy-minimum conformer obtained after a MacroModel^[35] conformational search. The compounds of the screening data base were ranked according to “ComboScore”, “ColorScore” and “ShapeTanimoto”. A detailed description of both the conformer generation with Omega and the parameterization of ROCS is provided in the Experimental Section.

Feature Tree: The reference molecules and the molecules of the screening sets were converted into feature tree models. To compare reference and screening compounds, we applied the split-search, match-search and dynamic match-search algorithms.^[5,36] In the virtual screening, all candidate ligands were ranked according to their resulting similarity to the respective reference ligand.

2D-similarity searching: 2D-similarity searches were performed by using Scitegic’s functional class fingerprints (FCFP), the Tanimoto index was employed as similarity coefficient for arithmetic superposition and similarity ranking of the screening sets.

Results

In the following, we will compare the relative performance of the different methods in retrieving known active compounds. Subsequently, we will analyse the potential complementarity of the different virtual-screening methods by providing a comparison of the hit lists obtained from the different approaches. We will show selected examples where active ligands have been retrieved among the top-scorers of one virtual-screening method compared to another. This might help to understand the general strengths and weaknesses of the different virtual-screening methodologies.

Docking and scoring

We first applied “free docking”, that is, we did not define any constraints in the docking procedure based on knowledge about ligand binding. All docking poses obtained from GOLD, GlideSP, GlideXP, Surflex and Flex were ranked with the scoring functions implemented as objective functions in the docking algorithms and, furthermore, with nine other different scoring functions. For reasons of clarity, we will only show the enrichments obtained from the combination of the best docking program and scoring function. Details are listed in the Supporting Information.

Analysis of these results (see Figure 2 and Table S1) revealed that GlideSP on average outperforms the other docking programs. In a next step, we wanted to investigate whether—or to show that—considering knowledge about ligand binding in terms of docking constraints improves the enrichments. These docking constraints (shown in Figure 1) were defined after careful visual inspection of all available crystal structures of the four investigated targets.

For ACE (Figure 1a), GlideSP provided the best enrichment for the top 10 scorers, identifying nine actives. Considering the top-ranked 50 compounds, all docking programs provided comparable enrichments. Consideration of the Zn²⁺ constraint considerably improved the docking results for ACE. Indeed, from the set of 50 active compounds and 950 decoys, only 23 actives and 19 decoys could be accommodated in the ACE

binding site. Thus, this stringent constraint identifies nearly 50% of the actives, whereas only 2% of the decoys are retrieved as false positives.

For docking into the COX-2 crystal structure (Figure 1b), GlideSP again outperforms the other docking programs. Here, consideration of a constraint did not (substantially) improve the virtual screening results. This is probably due to the fact that the COX-2 ligands bind in diverse orientations and that there are no dominant (polar) interactions used by a large fraction of active binders.

It must be stated that HIV-1 protease (Figure 1c) represents a challenge for structure-based virtual screening, since it has a large, flexible binding site. Analysis of superimposed HIV-1 protease crystal structures revealed that different ligands induce different binding-site conformations. Due to the large size of the binding site, known HIV-1 protease inhibitors generally have a large molecular weight and a partial peptidic or peptidomimetic character. Since this represents a handicap for both ligand- and structure-based virtual screening and since the goal of this study was not to show that virtual screening works in “easy” cases, but rather to provide a fair comparison between ligand- and structure-based virtual screening, we decided to include this pharmaceutically relevant target in our study.

For HIV-1 protease, the best enrichment for the ten top scorers is obtained by Surflex. However, visual inspection revealed that the docking poses were unrealistic. The top scorers re-

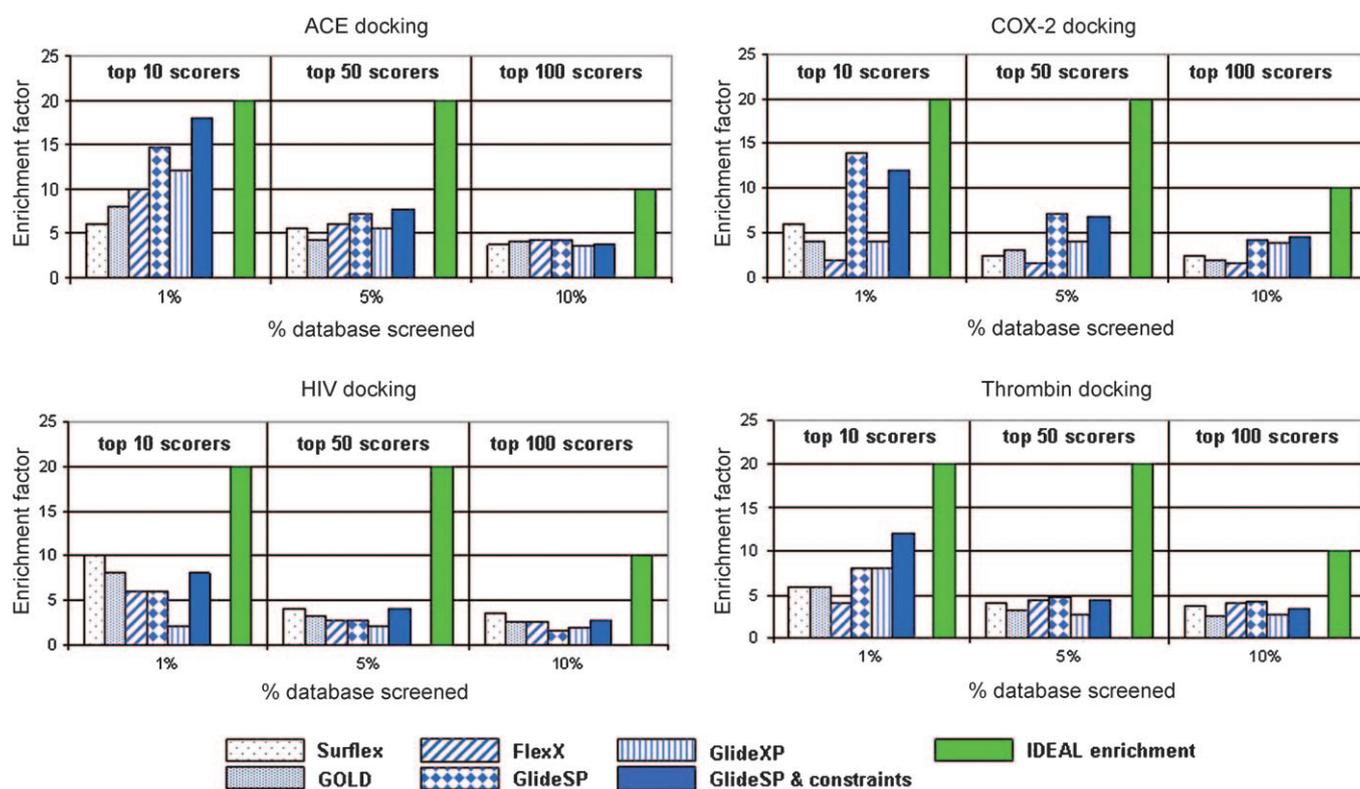


Figure 2. Comparison of docking tools in virtual screening. Enrichments at 1, 5 and 10% are shown. For each docking program, the scoring function providing the best enrichment is listed.

trieved by Surflex were indeed the large, peptide-like ligands; however, they showed collapsed conformations in the binding site and did not establish directed interactions with the protein. This example points out the importance of visual inspection of docking poses. The application of Surflex to the identification of HIV ligands in virtual screening would possibly provide large peptidic ligands. Since the polar interactions observed in the HIV crystal structures were not reproduced in this example by Surflex, it is unlikely that such virtual screening would be successful. The docking results obtained by GlideSP could, again, be improved by including constraints. By defining the essential hydrogen bonds to the catalytic aspartates, more time is left to search the remaining conformational and configurational space for optimising the interactions with the protein binding site. Visual inspection of docking poses and comparison with crystal structures of similar ligands revealed that the poses generated by GlideSP are reasonable.

Similar to HIV-1 protease, **thrombin** (Figure 1 d) is a protease with a large fraction of known ligands having large molecular weight and peptidic or peptidomimetic character. As for the other proteases, the best enrichment among the ten top scorers was obtained when docking with GlideSP. Consideration of constraints (see Figure 1 d) considerably improved the results. To be fair, we have to admit that finding the three constraints required considerable “manual” input and optimisation. Different parameter settings and possible interaction constraints were varied, and the constraints shown in Figure 1 d represent

the combination of constraints that provided the best final enrichment.

On average, GlideSP was the best docking program. Interestingly, GlideSP substantially outperformed GlideXP, in contrast to a study reported by Zhou et al.^[20] It was reported that the Glide extra precision (XP) methodology substantially enhanced the ability to pick out known active compounds from a random data base. Our interpretation is that GlideXP is highly sensitive to an accurate receptor conformation and provides very exact docking solutions for ligands that fit into the provided receptor conformation. We assume that these ligands are indeed retrieved as top scorers. For those compounds that induce slight rearrangements of the binding site, GlideXP seems to be less forgiving than GlideSP. These compounds are consequently not retrieved among the top scorers by GlideXP.

3D-similarity searches (ROCS)

Enrichment plots for all ROCS searches are given in Figure 3, enrichment factors are provided in Table S2 in the Supporting Information. Not surprisingly, for all targets, the best enrichments are obtained when the reference conformation corresponds or is close to the bioactive conformation instead of the lowest-energy conformer retrieved from a conformational search. For ACE and COX-2, good enrichment factors are also obtained when using the lowest-energy conformer as template for similarity searches. For COX-2, this observation can be ex-

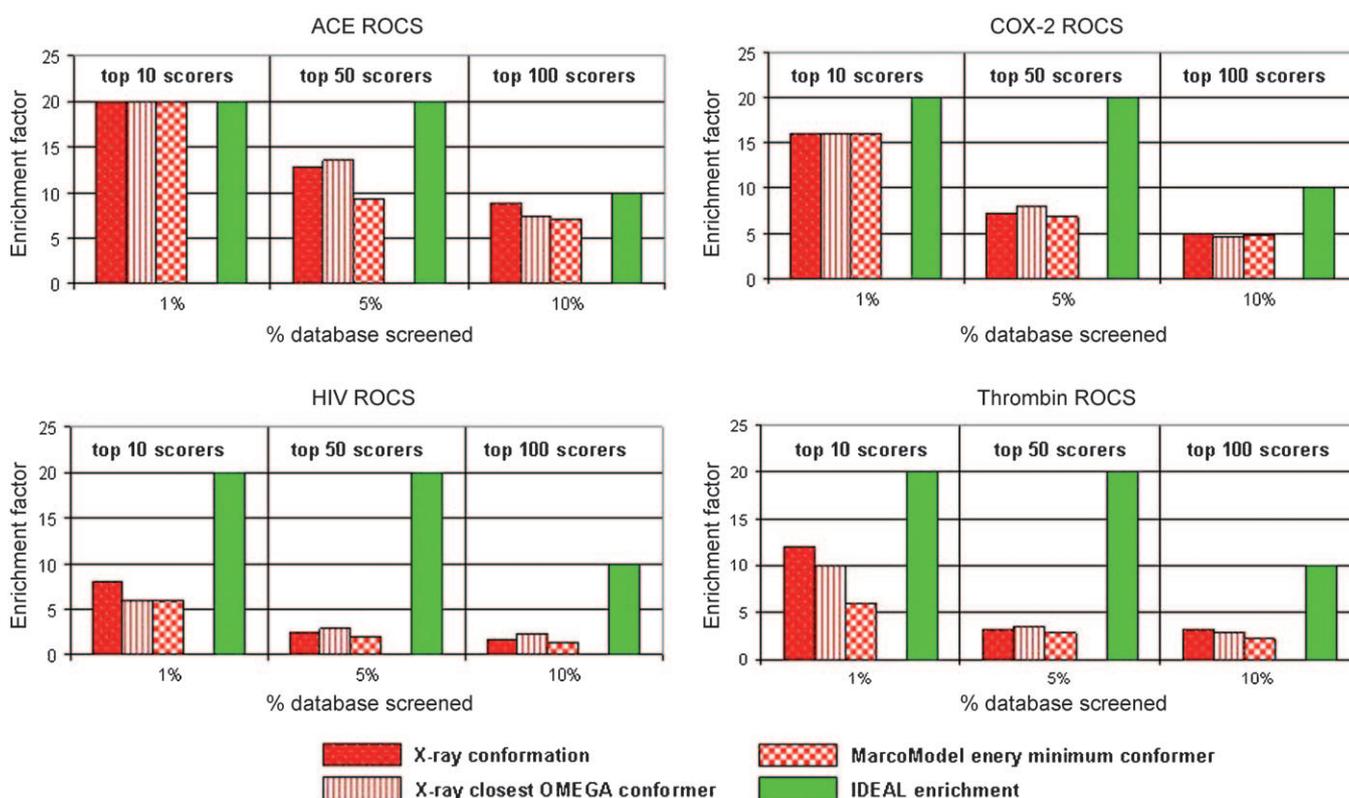


Figure 3. Performance of ROCS in virtual screening. Enrichments at 1, 5 and 10% are shown. For each ROCS screen, the objective function (ShapeTanimoto, ColorScore or ComboScore) providing the best enrichment is plotted (the enrichments obtained for the different objective functions are shown in Table S2). The best objective functions are ColorScore for ACE and HIV-1, ShapeTanimoto for COX-2 and ComboScore for thrombin.

plained by the fact that the lowest-energy conformer of the ligand S58 is relatively close to the bioactive conformation (rmsd = 1.13 Å).

The enalapril (ACE) lowest-energy conformer deviates from the bioactive conformation (rmsd = 2.23 Å). However, since a large fraction of the ACE actives have a similar topology of rotatable bonds to enalapril, these compounds can also be mapped onto the lowest-energy conformer of enalapril. For HIV-1 (rmsd = 4.42 Å) and thrombin (rmsd = 3.92 Å), however, the lowest-energy conformers deviate substantially from the bioactive conformation, and this explains the poor enrichments from these searches.

That the enrichments obtained from HIV-1 and thrombin inhibitors are poorer than those obtained from ACE is probably due to the fact that the HIV-1 and thrombin data sets are more diverse than the ACE data set; this is reflected by the observation that excellent enrichments are obtained for ACE with similarity searches based on simple (FCFP4) 2D descriptors. As indicated in Table S2, the performance of ROCS depends on the applied scoring function implemented into ROCS (ShapeTanimoto, ColorScore, ComboScore). The results vary from target to target.

Feature Trees

The results obtained from virtual screening with Feature Trees are plotted in Figure 4. Enrichment factors are provided in Table S2. As mentioned before, Feature Tree searches were per-

formed by applying the match-search, split-search and dynamic match-search algorithms. These different search algorithms perform differently among the four different targets. There is no one search algorithm that performs consistently better or worse than others.

2D-similarity searches with Scitegic Fingerprints (FCFP4)

Scitegic fingerprint searches have been performed as a benchmark for other virtual-screening tools to address the question of how much better a virtual-screening tool is than a "simple 2D descriptor". As described in the Experimental Section, the compilation of active and decoy data sets for virtual-screening evaluation was based on a diversity selection using this FCFP4 descriptor. Thus, the enrichments obtained from similarity searches based on this descriptor reflect 1) how diverse the active sets are compared to the decoy set and 2) how different the reference compounds are from the respective active sets. The enrichments obtained from FCFP4 similarity searches are plotted in Figure 5, and the corresponding enrichments are listed in Table S2.

Comparison

The relative performance of the different virtual-screening tools is provided in Figure 5. For each tool, the best results (obtained with the optimal parameter setting and scoring function) are listed. An important goal of this study was to in-

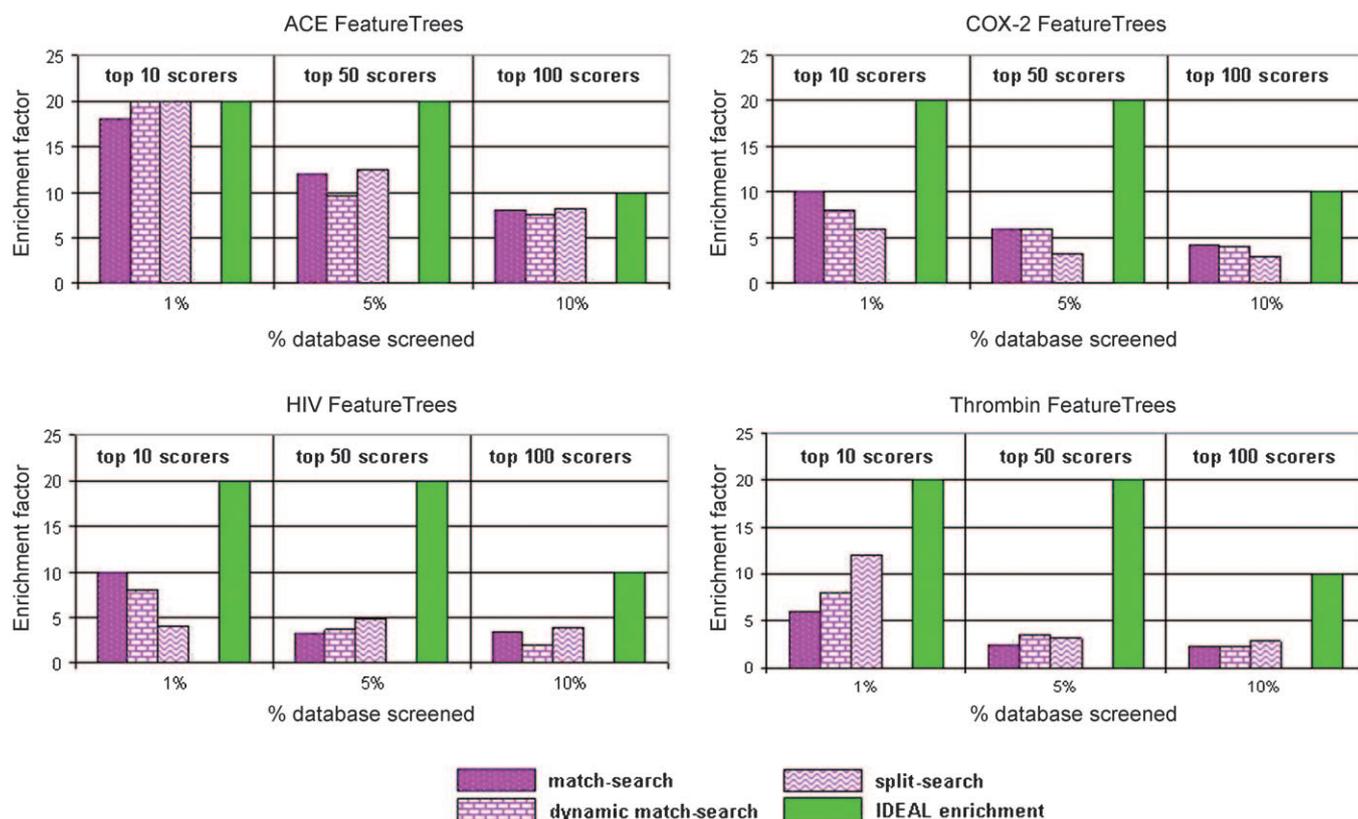


Figure 4. Performance of Feature Trees in virtual screening. Enrichments at 1, 5 and 10% are shown.

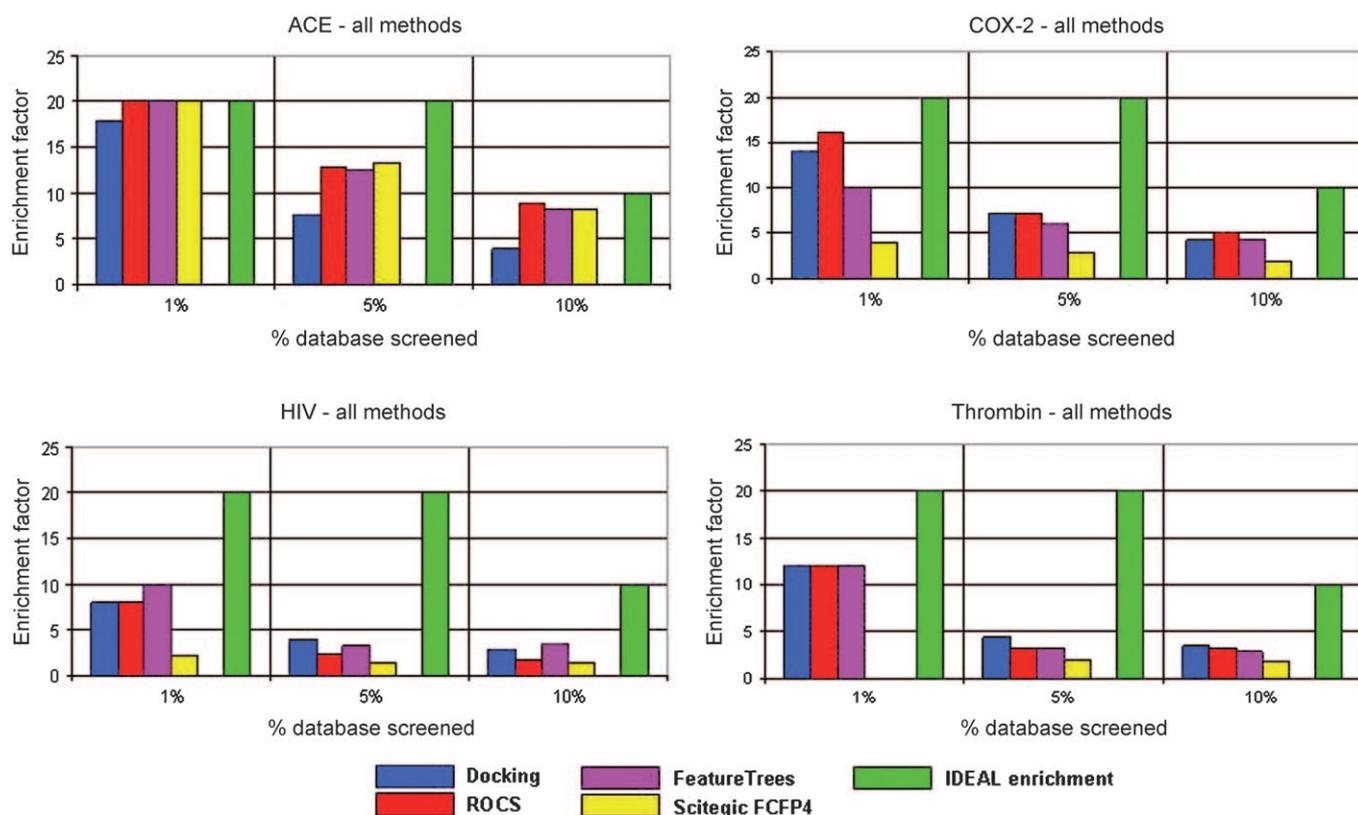


Figure 5. Comparative performance of docking, ROCS, Feature Trees and Scitegic FPs in virtual screening. Enrichments at 1, 5 and 10% are shown. For each virtual-screening method, the results obtained from the best parameter settings (i.e., scoring function, constraint setup for docking, reference ligand conformation and scoring function for ROCS and search algorithm for Feature Trees) are considered.

investigate whether different virtual-screening methods provide us with similar or different hit lists of (active) compounds. Therefore, for each target, we analysed the top 50 scorers and

merged the resulting hit lists. The redundancy and complementarity of the different virtual-screening tools observed for the four different targets is provided in Figure 6.

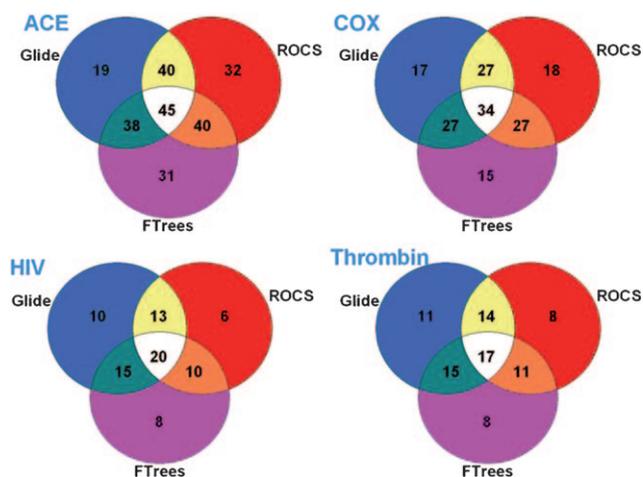


Figure 6. Redundancy and complementarity analysis for three different virtual-screening tools. For each virtual-screening method, the number of actives retrieved among the 50 top scorers is provided in the corresponding circles (e.g., for COX, ROCS retrieved 18 compounds and FTrees 15). The intersection planes between two circles reveal how many unique compounds have been retrieved by the two corresponding virtual-screening tools if the top 50 hit lists are merged (e.g., for COX, ROCS + FTrees retrieved 27 unique compounds).

ACE

For ACE, all virtual-screening tools perform remarkably well. As mentioned before, the docking results could be substantially improved by including an interaction constraint with the Zn^{2+} ion. Nevertheless, docking is outperformed by the ligand-based approaches. Most strikingly, the 2D-similarity searches based on FCFP4 fingerprints perform best. Although the selection of actives was based on a maximum diversity of ACE inhibitors available in the MDDR using these FCFP4 fingerprints, the ACE active set was evidently not diverse enough compared to the set of 950 decoys. In the context of this observation, it is not amazing that the ligand-based virtual-screening approaches perform so well. This observation points out the importance of data-set design for virtual-screening-validation studies. Analysing the complementarity of the hit lists obtained from the different virtual-screening tools (see Figure 6a) reveals that the different tools are obviously able to retrieve compounds with different characteristics. When merging the top 50 hit lists obtained from Glide (19 actives), ROCS (32 actives) and Feature Trees (31 actives), 45 (i.e. 90%) of all 50 ACE inhibitors are retrieved; this demonstrates that each method is

able to find compounds that are not identified by the other methods.

For us, an important aspect of a virtual-screening tool is the interpretability of the results. We want to visually analyze and, thus, understand *how* a compound matches to a model and *why* a compound is identified by one virtual-screening method but not another. For docking, intuitive interpretability is provided by visualizing the docking mode in the protein structure and analysing the interactions between protein and ligand. In the case of ROCS, the 3D superimposition of a screening compound onto the reference can be inspected. If the reference compound was extracted from a cocrystal structure, visualization is even possible in the context of the protein environment. Thus, visual analysis as a postfilter of virtual screenings results obtained from docking or ROCS can be performed with a standard molecular viewer.

An interesting example of a compound being retrieved by ROCS, Feature Tree and Scitegic fingerprints, but not by docking, is shown in Figure 7. Utibapril (Figure 7b) was retrieved on

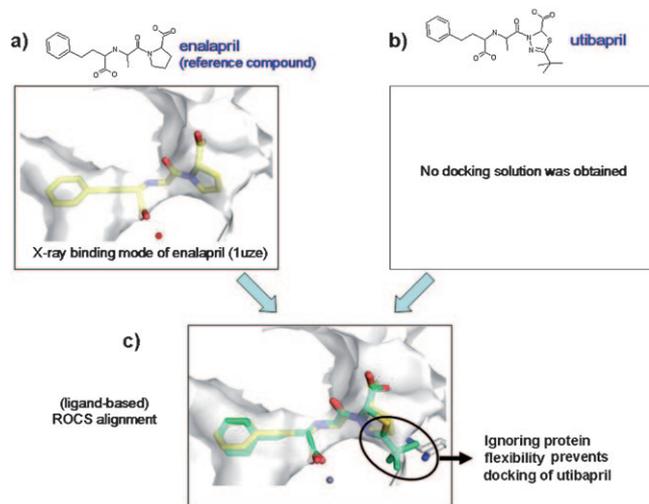


Figure 7. The ACE inhibitor utibapril was identified by ROCS but not by docking into the ACE crystal structure. a) The reference compound enalapril in its co-crystallised conformation. b) 2D structure of utibapril. c) 3D alignment of utibapril onto enalapril generated with ROCS. The protein environment of ACE is shown in surface representation.

rank 3 by ROCS (Feature Tree: rank 5, FCFP4: rank 8) when mapped onto the reference compound enalapril (shown in Figure 7a). No docking mode for utibapril was obtained with GlideSP. This can be rationalized by analysing the ligand-based (ROCS) alignment of utibapril onto the reference compound enalapril in the context of the ACE protein environment. Obviously, the *tert*-butyl-group of utibapril, which is not present in enalapril, would clash with the protein when docking into the protein conformation of 1uze, as indicated by the surface representation in Figure 7c. We suppose that ignorance of protein flexibility is the reason for missing utibapril when docking with GlideSP.

COX-2

For COX-2, the Scitegic fingerprints are outperformed by the other virtual-screening tools, thus indicating that all these methods are able to retrieve compounds with scaffolds different from that of the reference structure S58. Interestingly, as shown in Figure 6, each method is able to find active compounds (among the 50 top scorers) not found by the other methods: for example, 18 COX-2 inhibitors are retrieved by ROCS and 15 actives by FTrees to provide a total number of 33 actives. Of these, 27 compounds are unique; this shows that the hit lists obtained from FTrees and ROCS are highly complementary. Additional redundancy is observed when adding the compounds from the docking approach to the virtual screening hit list. Figure 8b shows a compound (1) that was retrieved

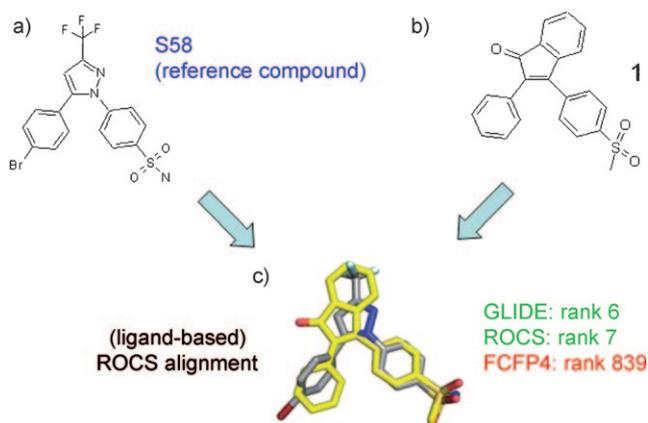


Figure 8. ROCS-generated 3D alignment of COX-2 inhibitor 1 onto the reference compound S58. Compound 1 was found among the top scorers by Glide and ROCS, but is highly dissimilar to S58 based on Scitegic FCFP4 fingerprints.

among the top scorers by ROCS (rank 7) and Glide (rank 6), but was only found at rank 839 based on Scitegic FCFP4 fingerprints. This example demonstrates that these virtual-screening approaches are potentially able to retrieve compounds that are different with respect to their chemical scaffolds but which share 3D similarities and/or are able to establish similar interactions with the target protein.

HIV-1 protease

It was mentioned in the docking section that HIV-1 protease represents a challenge for both ligand- and structure-based virtual-screening approaches. Indeed, the enrichments obtained from virtual-screening approaches for this target are lower than those for the other targets. As shown in Figure 5 and Tables S1 and S2, the best enrichment is obtained with Feature Trees. Encouragingly, Figure 6 demonstrates that the different virtual-screening tools provide complementary hit lists. Each method is able to retrieve active binders that are not found by the other methods. Figure 9 demonstrates that, also for HIV-1 protease, the different methods provide compounds with dif-

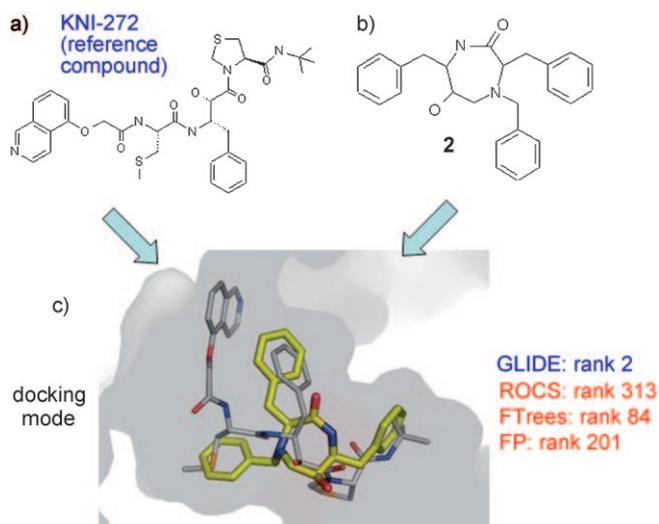


Figure 9. Docking mode of compound 2 in the HIV-1 crystal structure. Compound 2 was retrieved on rank 2 by GlideSP but it was not identified among the top 50 scorers by the ligand-based approaches.

ferent characteristics. In Figure 9b, a compound (2) is shown that was retrieved by docking (rank 2), but not by the ligand-based approaches. This compound (a cyclic urea) has a totally different scaffold from that of the reference compound KNI-272 (Figure 9a) and binds with a different topology into the HIV-1 binding site. Thus, it is not surprising that compound 2 was not identified as active by the ligand-based virtual-screening approaches. Visual analysis of further HIV-1 protease crystal structures revealed that the docking mode obtained for compound 2 is realistic.

Thrombin

For thrombin, Glide, ROCS and Feature Trees perform equally well (see Figure 5 and Tables S1 and S2). Again, the enrichments are considerably better than those obtained from 2D-similarity searches with Scitegic FCFP4 fingerprints, thus indicating that the identified virtual screening hits are considerably different from the reference compound MD805. We must note, however, that considerable manual input was necessary to define docking constraints to provide an enrichment as good as those obtained from FTrees or ROCS. Again, comparison of the top 50 hit lists revealed complementarity between the different virtual-screening tools.

Discussion and Conclusion

Complementarity of hit lists

This study shows that different virtual-screening tools generally provide complementary hit lists. It was shown that the different virtual-screening approaches are obviously able to retrieve actives with different characteristics. Although the test data set was relatively small (1000 compounds for each target), we observed only a small overlap of the top 50 scorers. These top 50

scorers represent 5% of the test data set. In realistic prospective virtual-screening scenarios, we might extract 1000 compounds from a virtual collection of 1 million compounds, that is, 0.1%. Thus, it is likely that in this situation the overlap of hit lists obtained from different virtual-screening approaches will be extremely low. Therefore, in order to increase the likelihood of identifying actives with different characteristics, we recommend applying different (structure- and ligand-based) virtual-screening tools in parallel and merging the resulting hit lists. Of course, the number of targets considered in this study is too low for a conclusive statement, but we assume that similar trends regarding the complementarity of the screening methods will be found for other targets as well. A similar conclusion was drawn from a study performed by Bajorath et al.,^[28] proposing that a parallel selection of candidate compounds from individual rankings is generally superior to rank fusion. An explanation for this observation is provided by Sheridan and Kearsley:^[37] different methods select different actives for the same biological activity, and the same method might work better on some activities than others in a way that is difficult to predict, since receptors are diverse, and chemical groups that appear equivalent to one descriptor (or one receptor) might not be equivalent to another.

Relative performance of different virtual-screening approaches

The results presented here show that structure- and ligand-based virtual-screening methods provide comparable enrichments in detecting active compounds. Furthermore, it seems generally (but not always) beneficial to combine ligand- and structure-based approaches if possible, that is, by considering knowledge about ligand binding as docking constraints or using a target-bound ligand conformation when performing 3D-similarity searches.

Application of virtual-screening hypotheses derived from multiple ligands or protein structures

Since it was not a scope of this study, we did not investigate in how far the selection of different or multiple ligands influences the success of a ligand-based virtual screening campaign. The selection of the reference ligand will change the goodness of a ligand-based virtual screen drastically. The application of multiple (diverse) ligands, either by combination into a model or by performing multiple searches for each reference ligand, will generally improve the enrichment. On the other hand, it was shown by Sheridan et al.^[38] that virtual-screening results from docking are also extremely sensitive to exactly which crystal structure is used for a particular target. Thus, when several crystal structures with different binding-site conformations are available for one target, the application of ensemble docking into different crystal structures and subsequent fusion of the individual hit lists might be a further way to improve the enrichment.

Cascaded or parallel virtual screening?

Often, virtual screening is performed in a cascaded manner, by using hierarchical filters of increasing computational complexity. After the elimination of chemical structures with unwanted properties, 2D-similarity searches of known ligands are often performed. 3D-similarity or 3D-pharmacophore searches are options for further reducing the number of compounds for molecular docking. Numerous successful applications of this strategy have been reported in the literature.^[15–19] In this study, we have shown that different virtual-screening methods generally seem to provide complementary hit lists of actives. Moreover, it is known that ligand-based virtual screenings based on different reference ligands or docking into different protein structures^[46] provide different sets of actives. We wonder whether cascaded virtual-screening is the best strategy for identifying a maximum number of actives with diverse chemical structures. Depending on the computational resources available for virtual screening, we suggest applying different virtual-screening methods and reference protein and ligand structures in parallel and finally combining the resulting hit lists whenever possible.

Novelty of hits

Another important aspect of virtual screening is the novelty of hits. Good et al. suggest counting additional hits only when the chemotype of a molecule is not equal to a template chemotype or any other chemotype that already exists in the hit list.^[39] This approach results in a chemotype enrichment that highlights a measure of ligands with totally different chemotypical properties. We tried to assess the novelty of hits by comparing the hit rates with a “simple 2D descriptor” as suggested by Jain et al.^[40] by taking advantage of the fact that 2D fingerprints are only able to find compounds similar to the reference structure. The observation that docking, ROCS and Feature Trees outperform the FCFP4 fingerprints for three of the four investigated targets shows that these methods are principally able to retrieve compounds with novel scaffolds (i.e., scaffolds different from a reference compound). Considering the high degree of substructure encoding, it is not surprising that FCFP4 fingerprints performed worst at finding novel chemotypes. On the other hand, it should be mentioned that even the identification of a close analogue by virtual screening can be beneficial, because an extended SAR or a small modification might guide the way into patent-free chemistry space.

Data-set design

We are aware that the results from a virtual screening study are sensitive to the design of the test data base.^[22,47,48] In our opinion, it is hardly possible to design the one and only “perfect” evaluation data set. A pragmatic solution is to compile a “decoy” set from that data base which is considered for the next “real” virtual screening campaign, for example, a representative set of the company’s screening collection or a database of compounds from commercial vendors. Such a set will not

perfectly represent the whole of chemical space, but it will show which methods are best able to identify known actives from the virtual-screening set under consideration.

The importance of visual inspection

Analysis of the hit rates and enrichment factors should not be the only criterion when evaluating the performance of a virtual screen. In our opinion, it is important to understand how a compound matches onto a model, that is, how a ligand fits into a protein binding site, is mapped onto a pharmacophore or aligns with a template ligand. Therefore, we recommend visual inspection of virtual-screening hits. Although this step is slow and lacking in objectivity, it is in our opinion one of the most crucial steps, since we believe that the virtual-screening tools presently available are not sufficiently reliable and discriminating. In addition, aspects like novelty and physicochemical properties or, as shown in this study for HIV-1 protease inhibitors, the detection of artificial binding poses and other factors that might be relevant for a particular virtual screening, are difficult and time consuming to implement into a computer algorithm.

Experimental Section

Protein preparation: For each target, all available PDB crystal structures were retrieved and visualized with Relibase.^[41–43] Considering resolution and the activity of the cocrystallised ligand, we selected the crystal structures 1uze^[44,49] (ACE), 1cx2^[45,50] (COX-2), 1hpx^[51,52] (HIV) and 1dwc^[53,54] (Thrombin). The protein setup was performed by using Maestro’s “Protein Preparation Wizard” utility and adding hydrogens and assigning the correct bond orders, protonation states and OH torsions. The protein was saved in a maestro-file format and exported into mol2- and pdb-file formats.

Screening set: To compare the performance of different virtual-screening methods, we compiled databases of “active” compounds and decoys from the MDL Drug Data Report (MDDR) database (MDL Information Systems Inc.). To generate the “active” sets, we extracted all compounds with stated activity against the respective targets; this resulted in 570 ACE inhibitors, 980 COX-2 inhibitors, 1008 HIV-1 protease inhibitors and 1269 thrombin inhibitors. After application of molecular property filters ($M_w < 500$, number of H-bond donors or acceptors < 8 , number of rotatable bonds < 16 , number of rings < 8 , $-6 \geq AP \log P \leq 9$) and removal of prodrugs, we extracted the set of 50 “most diverse compounds” for each set using Scitegic’s FCFP4 descriptor.

The same filter criteria were used for the compilation of the decoy data set to extract a diverse set of 1200 compounds without stated activity on the four reference targets. Further removal of compounds after inspection of molecular property filters was performed to ensure a similar property distribution between active and decoy compounds in order to avoid trivial enrichments based on ranking by, for example, molecular weight. This procedure reduced the decoy data set to 950 compounds.

Quantitative description of hit lists: The effectiveness of the screening methods was evaluated by assessing the enrichment of known “actives” within the top-scored compounds, compared to random selection. The enrichments are reported in graphical and tabular form. The enrichment factor is represented by:

$$EF = \frac{\text{Hits}_{\text{sampled}}/N_{\text{sampled}}}{\text{Hits}_{\text{total}}/N_{\text{total}}}$$

EF: enrichment factor, $\text{Hits}_{\text{sampled}}$: number of true hits in the hit list, N_{sampled} : number of compounds in the hit list, $\text{Hits}_{\text{total}}$: number of hits in the full data base, N_{total} : number of compounds in the full data base.

The enrichment factor was calculated based on the assumption that all compounds with MDDR stated activity are active (true actives) and compounds with no stated activity against this target are inactive. Although compounds with potential affinity against the investigated proteins were eliminated from the decoy sets after filtering, that some of the inactives identified among the top-scored compounds by virtual screening reveal actual activity on that target cannot be excluded. The hit rate and enrichment factor would thus be higher.

Docking and scoring

Ligand preparation: Prior to docking, the ligand structures were processed by using the LigPrep 2.0 utility from Schrodinger (Schrodinger, L.L.C., New York) to create tautomers, ring conformations, stereoisomers and protomers.

Active-site definition: The active site was defined as given by the default parameters for each docking tools, that is, 7 Å around the ligand for FlexX and Surflex, residues within 5 Å of the ligand for GOLD and 10 Å around the ligand for Glide.

GOLD 3.1.1 docking: For each of the GA runs, a maximum number of 100000 operations were performed on a population of 100 individuals. Operator weights for crossover, mutation and migration were set to 95, 95 and 10, respectively, which are the standard default settings recommended by the authors for careful work. The distance for hydrogen bonding was set to 4 Å, and the cut-off value for van der Waals was 2.5. For each ligand, 10 poses were saved.

Surflex 2.0 docking: All default parameters, as implemented in the 6.72 release of Sybyl, were used. Cscore calculations were performed for ranking, and for each ligand the ten best poses were saved.

FlexX 2.0.2 docking: Default parameters were used. Cscore calculations were performed for ranking, and for each ligand, the 10 best poses were saved.

Glide 4.0 docking: Distances from a grid point to the receptor surface were compared to distances from the ligand centre to the ligand surface. Good matches were kept, followed by a clash test, subset scoring, greedy scoring and final refinement of 5000 initial poses in the *x/y/z* directions. The resultant 400 refined poses were kept, and then minimized with a distance-dependent dielectric constant and 100 conjugate gradient steps. Final poses were scored with GlideScore with an inclusion of an energy score.

For each ligand, ten poses were saved. Subsequently, we defined a set of constraints for each target to consider interactions known to be essential for the binding of the ligand in its cocrystallised conformation. A detailed description of these constraints is provided in the legend of Figure 1.

Scoring: All docking poses were rescored by using nine different scoring functions (D_Score, G_Score, ChemScore, PMF as implemented in the Cscore^[55] module of Sybyl7.0, DrugScore^[56] (PDB and CSD), Xscore^[57] (HP, HM and HS), Chemscore^[58], SFC_Score^[59] and PLP_Score^[60]). If available, the scoring functions implemented

as objective function in the docking algorithms were included into the set of scoring functions.

Feature Trees: For all molecules of the training and the screening sets, Feature Tree^[36] descriptors were calculated. For each target, the cocrystallised ligand (Figure 1) was automatically converted into a Feature Tree model (FTree).^[5] In the virtual screening, all candidate ligands of the active and decoy sets were ranked according to their similarity to the reference Feature Tree used for the screening. For the similarity searches, we used the split-search, match-search and dynamic match-search algorithm implemented into the Feature Tree program.

3D-similarity searching with ROCS (version 2.1.1): Ligand structures were preprocessed by using the LigPrep 2.0 utility from Schrodinger (Schrodinger, L.L.C., New York) to create tautomers, ring conformations, stereoisomers and protomers. Conformer generation was performed with Omega (version 1.8.1). The maximum number of conformers and ring conformations was set to one million, the value for the energy window was set to -10, and the RMS threshold for determining duplicate conformations was set to 0.8 so as to provide a maximum of conformers from each ligand. ROCS ("rapid overlay of chemical structures") is based on the description of molecules by physical characterization of the shape and electrostatics of a reference and a query molecule. For each target, the reference ligand was used either 1) in its cocrystallised conformation, 2) utilising the Omega conformer closest to the cocrystallised conformation or 3) the global-energy-minimum conformer obtained from a conformational search performed with MacroModel. The compounds of the screening data sets were mapped to the respective reference ligands with ROCS and ranked according to the "ComboScore", "ColorScore" and "ShapeTanimoto".

Keywords: active compounds · docking · drug design · molecular similarity · virtual screening

- [1] K. H. Bleicher, H. J. Bohm, K. Muller, A. I. Alanine, *Nat. Rev. Drug Discovery* **2003**, *2*, 369.
- [2] T. I. Oprea, H. Matter, *Curr. Opin. Chem. Biol.* **2004**, *8*, 349.
- [3] Q. Zhang, I. Muegge, *J. Med. Chem.* **2006**, *49*, 1536.
- [4] G. Schneider, H. J. Böhm, *Drug Discovery Today* **2002**, *7*, 64.
- [5] G. Hessler, M. Zimmermann, H. Matter, A. Evers, T. Naumann, T. Lengauer, M. Rarey, *J. Med. Chem.* **2005**, *48*, 6575.
- [6] M. L. Verdonk, V. Berdini, M. J. Hartshorn, W. T. Mooij, C. W. Murray, R. D. Taylor, P. Watson, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793.
- [7] T. S. Rush 3rd, J. A. Grant, L. Mosyak, A. Nicholls, *J. Med. Chem.* **2005**, *48*, 1489.
- [8] M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, *J. Med. Chem.* **2000**, *43*, 3233.
- [9] T. Kalliokoski, T. Ronkko, A. Poso, *J. Chem. Inf. Model.* **2008**, *48*, 1131.
- [10] O. Guner, O. Clement, Y. Kurogi, *Curr. Med. Chem.* **2004**, *11*, 2991.
- [11] A. Evers, G. Hessler, H. Matter, T. Klabunde, *J. Med. Chem.* **2005**, *48*, 5448.
- [12] M. Rella, C. A. Rushworth, J. L. Guy, A. J. Turner, T. Langer, R. M. Jackson, *J. Chem. Inf. Model.* **2006**, *46*, 708.
- [13] R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson, P. G. Wyatt, *ChemMedChem* **2008**, *3*, 435.
- [14] D. A. Grueneberg, L. Pablo, K. Q. Hu, P. August, Z. Weng, J. Papkoff, *Mol. Cell. Biol.* **2003**, *23*, 3936.
- [15] R. Brenk, L. Naerum, U. Gradler, H. D. Gerber, G. A. Garcia, K. Reuter, M. T. Stubbs, G. Klebe, *J. Med. Chem.* **2003**, *46*, 1133.
- [16] H. Yang, Y. Shen, J. Chen, Q. Jiang, Y. Leng, J. Shen, *Eur. J. Med. Chem.* **2009**, *44*, 1167.
- [17] S. Wu, M. Bottini, R. C. Rickert, T. Mustelin, L. Tautz, *ChemMedChem* **2009**, *4*, 440.
- [18] S. Turk, A. Kovač, A. Boniface, J. M. Bostock, I. Chopra, D. Blant, S. Gobec, *Bioorg. Med. Chem.* **2009**, *17*, 1884.

- [19] S. Grüneberg, M. T. Stubbs, G. Klebe, *J. Med. Chem.* **2002**, *45*, 3588.
- [20] Z. Zhou, A. K. Felts, R. A. Friesner, R. M. Levy, *J. Chem. Inf. Model.* **2007**, *47*, 1599.
- [21] M. Kontoyianni, L. M. McClellan, G. S. Sokol, *J. Med. Chem.* **2004**, *47*, 558.
- [22] G. B. McGaughey, R. P. Sheridan, C. I. Bayly, J. C. Culberson, C. Kretsoulas, S. Lindsley, V. Maiorov, J. F. Truchon, W. D. Cornell, *J. Chem. Inf. Model.* **2007**, *47*, 1504.
- [23] M. Stahl, M. Rarey, *J. Med. Chem.* **2001**, *44*, 1035.
- [24] P. C. Hawkins, A. G. Skillman, A. Nicholls, *J. Med. Chem.* **2007**, *50*, 74.
- [25] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, *Proteins Struct. Funct. Bioinf.* **2004**, *57*, 225.
- [26] K. Moffat, V. J. Gillet, M. Whittle, G. Bravi, A. R. Leach, *J. Chem. Inf. Model.* **2008**, *48*, 719.
- [27] V. I. Pérez-Nueno, D. W. Ritchie, O. Rabal, R. Pascual, J. I. Borrell, J. Teixidó, *J. Chem. Inf. Model.* **2008**, *48*, 509.
- [28] L. Tan, H. Geppert, M. T. Sisay, M. Gutschow, J. Bajorath, *ChemMedChem* **2008**, *3*, 1566.
- [29] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739.
- [30] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727.
- [31] A. N. Jain, *J. Med. Chem.* **2003**, *46*, 499.
- [32] B. Kramer, M. Rarey, T. Lengauer, *Proteins Struct. Funct. Bioinf.* **1999**, *37*, 228.
- [33] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, C. Zardecki, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58*, 899.
- [34] J. Bostrom, J. R. Greenwood, J. Gottfries, *J. Mol. Catal. B* **2003**, *21*, 449.
- [35] MacroModel, Schrödinger, New York, NY, **2005**.
- [36] M. Rarey, J. S. Dixon, *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471.
- [37] R. P. Sheridan, S. K. Kearsley, *Drug Discovery Today* **2002**, *7*, 903.
- [38] R. P. Sheridan, *J. Chem. Inf. Model.* **2008**, *48*, 426.
- [39] A. C. Good, M. A. Hermsmeier, S. A. Hindle, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 529.
- [40] A. N. Jain, A. Nicholls, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133.
- [41] M. Hendlich, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1998**, *54*, 1178.
- [42] M. Hendlich, A. Bergner, J. Gunther, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 607.
- [43] J. Günther, A. Bergner, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 621.
- [44] R. Natesh, S. L. Schwager, E. D. Sturrock, K. R. Acharya, *Nature* **2003**, *421*, 551.
- [45] K. V. Sai Ram, G. Rambabu, J. A. Sarma, G. R. Desiraju, *J. Chem. Inf. Model.* **2006**, *46*, 1784.
- [46] R. P. Sheridan, G. B. McGaughey, W. D. Cornell, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 257.
- [47] N. Huang, B. K. Shoichet, J. J. Irwin, *J. Med. Chem.* **2006**, *49*, 6789.
- [48] A. N. Jain, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 201.
- [49] R. Natesh, S. L. Schwager, H. R. Evans, E. D. Sturrock, K. R. Acharya, *Biochemistry* **2004**, *43*, 8718.
- [50] R. G. Kurumbail, A. M. Stevens, J. K. Gierse, J. J. McDonald, R. A. Stegeman, J. Y. Pak, D. Gildehaus, J. M. Miyashiro, T. D. Penning, K. Seibert, P. C. Isakson, W. C. Stallings, *Nature* **1996**, *384*, 644.
- [51] T. N. Bhat, E. T. Baldwin, B. Liu, Y. S. Cheng, J. W. Erickson, *Nat. Struct. Biol.* **1994**, *1*, 552.
- [52] E. T. Baldwin, T. N. Bhat, S. Gulnik, B. Liu, I. A. Topol, Y. Kiso, T. Mimoto, H. Mitsuya, J. W. Erickson, *Structure* **1995**, *3*, 581.
- [53] D. W. Banner, P. Hadvary, *J. Biol. Chem.* **1991**, *266*, 20085.
- [54] D. W. Banner, P. Hadvary, *Adv. Exp. Med. Biol.* **1993**, *340*, 27.
- [55] R. D. Clark, A. Strizhev, J. M. Leonard, J. F. Blake, J. B. Matthew, *J. Mol. Catal. B* **2002**, *20*, 281.
- [56] H. Gohlke, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2000**, *295*, 337.
- [57] R. Wang, L. Lai, S. Wang, *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11.
- [58] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee, *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425.
- [59] C. A. Sotriffer, P. Sanschagrin, H. Matter, G. Klebe, *Proteins Struct. Funct. Bioinf.* **2008**, *73*, 395.
- [60] D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, S. T. Freer, *Chem. Biol.* **1995**, *2*, 317.

Received: July 29, 2009

Published online on November 11, 2009