—————**ARTICLES**—————

FieldChopper, A New Tool for Automatic Model Generation and Virtual Screening Based on Molecular Fields

Tuomo Kalliokoski,* Toni Ronkko, and Antti Poso

Department of Pharmaceutical Chemistry, University of Kuopio, P.O. Box 1627, 70211 Kuopio, Finland

Received June 21, 2007

Algorithms were developed for ligand-based virtual screening of molecular databases. FieldChopper (FC) is based on the discretization of the electrostatic and van der Waals field into three classes. A model is built from a set of superimposed active molecules. The similarity of the compounds in the database to the model is then calculated using matrices that define scores for comparing field values of different categories. The method was validated using 12 publicly available data sets by comparing the method to the electrostatic similarity comparison program EON. The results suggest that FC is competitive with more complex descriptors and could be used as a molecular sieve in virtual screening experiments when multiple active ligands are known.

1. INTRODUCTION

Virtual screening can be defined as the ranking of compounds based on their bioactivity with knowledge about the active compounds or information about the structure of a target protein.[1] It is analogous to High-Throughput Screening (HTS), where a large number of compounds are screened in vitro. However, in vitro screening of large chemical libraries is expensive, and virtual screening is widely used to reduce the number of compounds that need to be tested.[2]

Protein-based virtual screening is usually the preferred method if the crystal structure of the drug target is known.[3] A comprehensive review on protein-based virtual screening has been published.[4] In spite of several successful applications, there are still major methodological weaknesses associated with protein-based virtual screening.[5]

Ligand-based virtual screening is based on the similarity between a query and the compounds in the database. There are a plethora of 2D- and 3D-based molecular descriptors available.[6] 2D-based molecular fingerprint searching is perhaps the most widely used method and has been proven to be an efficient approach.[1] Several 3D-methods have also been proposed. One way to approach the problem is to superimpose a compound selected from a database into a template molecule and calculate the similarity of superimposed conformations. Several algorithms for this purpose have been described in the literature.[7–12]

When several active ligands are known, different kinds of methods can be employed to improve virtual screening performance. One of the most efficient methodologies is to build a pharmacophore with active molecules and use this as a database query.[13,14] Crystallized protein−ligand complexes can also be used as a basis of pharmacophores.[15–17]

Molecular fields describe the properties of a compound by the potentials around the molecule. Molecular fields have also been used for automatic pharmacophore and model generation. The seminal work in this area is Comparative Molecular-Field Analysis (CoMFA), which is the most used 3D-QSAR method.[18] Putta et al. described a shape-feature based method where molecules are represented in a binary descriptor.[19] From this descriptor, the subset that is most relevant to bioactivity is identified and used to score virtual libraries. Jain developed Surflex-Sim, which uses a surface-based morphological similarity function while minimizing the overall molecular volume of the aligned structures to form a hypothesis on binding.[20] Kotani and Higashura described Comparative Molecular Active Site Analysis (CoMASA).[21] CoMASA extracts 3D maps from a set of active compounds, which then could be used as 3D database queries.

In this paper, a novel model generation and virtual screening method (FieldChopper, FC) based on the van der Waals and electrostatic molecular fields is presented. FC contains two differences to previously published methods. The method uses multiple fields types simultaneously, and it can be used with different superimpositioning algorithms. The algorithms are described in detail, and its virtual screening performance is evaluated using 12 drug targets. FC is available free of charge to academic researchers for noncommercial use (please contact the corresponding author for details).

2. METHODS

2.1. Overview. An overview of FC is shown in Figure 1. First, a template molecule for superimposition is selected which is used in both model generation and scoring algorithms. A model is built by superimposing a set of bioactive molecules onto the template molecule and running the model generation algorithm. Then, a 3D multiconfor-

* Corresponding author e-mail: tuomo.kalliokoski@uku.fi.

10.1021/ci700216u CCC: $40.75 © 2008 American Chemical Society
Published on Web 05/20/2008

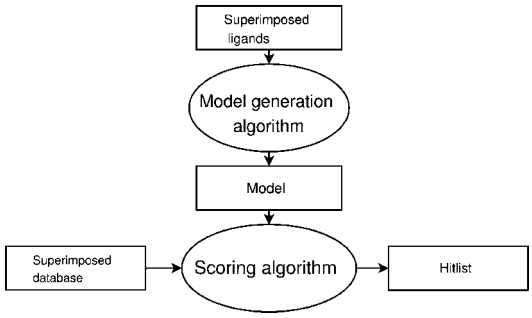

Figure 1. Overall view of FC.

mation database is screened by superimposing the compounds on to the same template and scoring them against the model. Finally, the scores are saved in a hit-list that can be used in the selection of compounds for in vitro testing.

There are many automatic superimpositioning methods proposed in the literature.[14] For practical purposes, the superimpositioning method has to be fast enough to handle thousands of molecules within a reasonable time. For example, ROCS[22] and BRUTUS[23] are suitable algorithms.

2.2. Algorithms. The most important FC algorithms are those constructed for model generation and scoring. FC uses the electrostatic potentials and van der Waals volumes to describe molecules. One common way to represent these fields is to use a rectilinear 3D-lattice that is equally spaced and to calculate the interaction between the molecule and a probe atom in each grid point. This principle is applied in the widely used CoMFA.[18] FC uses grid-spacing of 1 Å by default and an sp3 carbon with a charge of +1.0 as a probe atom. The size of the grid is determined according to the molecules used in the model.

The van der Waals volume is approximated as

$$E = \frac{v^6}{r^6} \tag{1}$$

where v is the van der Waals radius for the atom; and r is the distance between the grid point and the atom.

The electrostatic potential can be calculated with the following equation based on Coulomb's law[24]

$$E = \frac{kQ}{r^2} \tag{2}$$

where k is a conversion factor; Q is the partial charge on the atom; and r is the distance between grid point and the atom.

All programming was done with C++ (GNU C++ compiler version 4.1.1) except for the graphical user interface, which was written in Java. This makes FC readily transferable to various platforms. X86/X64 Linux-clusters were used for the software development and calculations.

2.2.1. Model Generation Algorithm. The model generation algorithm detects similar grid points between active compounds. Each grid point is analyzed one at a time. The values from active molecules are classified into three bins (Table 1) for each point. This results in a three-class histogram from which the peaks are detected. If the class frequency is over one-third of the sum of classes, then it is classified as a peak. The amount of the peaks can vary from zero to two peaks within one histogram.

Table 1. Classification Limits for the van der Waals and Electrostatic Potentials

interaction	bin 1	bin 2	bin 3
van der Waals	inside (>=1)	near (>=0.001, <=1.0)	outside (<0.001)
electrostatic	negative (<−0.1)	near zero (>=−0.1, <0.1)	positive (>=0.1)

Table 2. van der Waals Scoring Matrix

model	molecule		
peak(s)	inside	near	outside
Inside	1	−5	−5
Inside&Near	1	1	−5
Inside&Outside	1	−5	1
Near	−10	1	−5
Near&Outside	−10	1	1
Outside	−10	−5	1
None	1	1	1

Table 3. Electrostatic Scoring Matrix

model	molecule		
peak(s)	negative	zero	positive
Negative	2	0	−2
Negative&Zero	1	1	0
Negative&Positive	1	0	1
Zero	0	1	0
Zero&Positive	0	1	1
Positive	−2	0	2
None	0	0	0

The peaks are used in the scoring algorithm. Since most of the electrostatic grid is empty, important grid points for the activity are detected using the van der Waals histograms. The grid points having a peak in their "Near Molecule" bin are taken into the electrostatic scoring, and all other points are excluded. For the van der Waals volumes, all grid points are used in the scoring, since a van der Waals volume describes the overall shape of the binding site. Classification limits for the van der Waals histograms are selected so that compounds larger than those used in the model are punished in the scoring algorithm.

2.2.2. Scoring Algorithm. The scoring algorithm requires a previously generated model and a superimposed 3D-molecule as input data. The van der Waals volume and the electrostatic potential are generated, and each grid point is scored. First, the grid point is classified into one of three classes described in the model generation algorithm. Then, this class is compared with the peaks in the model and scored using the scoring matrices (Tables 2 and 3). The score for a field is simply the sum of values from the scoring matrices. The total score is defined as

$$S = W_V P_V + W_E P_E \tag{3}$$

where S is the total score; W_V is the weight for the van der Waals score; P_V is the van der Waals score; W_E is the weight for the electrostatic score; and P_E is the electrostatic score.

Since there are fewer points in the electrostatic score than in the van der Waals score, the latter score needs to be scaled down. In this study, arbitrary values of 0.2 for W_V and 1.0 for W_E were selected. It should be noted that these values

Table 4. Selected Targets for Retrospective Virtual Screening and Number of Ligands and Decoys[a]

target	class	ligands	decoys
androgen receptor (AR)	nuclear hormone receptors	59	2628
estrogen receptor (ER_agonist)	nuclear hormone receptors	52	2355
glucocorticoid receptor (GR)	nuclear hormone receptors	63	2797
epidermal growth factor receptor kinase (EGFr)	kinases	429	14894
fibroblast growth factor receptor 1 kinase (FGFr1)	kinases	103	4205
P38 mitogen activated protein kinase (P38)	kinases	241	8387
tyrosine kinase SRC (SRC)	kinases	140	5793
factor Xa (FXa)	serine proteases	127	5095
dihydrofolate reductase (DHFR)	folate enzymes	186	7145
acetylcholineesterase (AChE)	other enzymes	90	3714
cyclooxygenase-2 (COX2)	other enzymes	333	12464
enoyl ACP reductase (InhA)	other enzymes	70	3035

[a] Ligands used in the models (15 per target) are not included.

are probably not optimal, and the weights should be modified according to the nature of the target.

2.3. Retrospective Virtual Screening. *2.3.1. Selection of Targets.* Recently, a diverse test set for benchmarking molecular docking programs called Database of Useful Decoys (DUD) was compiled by Huang and co-workers.[25] DUD decoys have two advantages over other test sets. They are publicly available in electronic form which makes it possible to benchmark different methods. More importantly, DUD decoys are matched to ligands via their the physical properties, which makes them more demanding compared to randomly selected decoys.

DUD version 2 was downloaded from the Web site http://dud.docking.org (accessed Oct 26, 2006). Twelve targets from DUD's 40 targets were selected for this study based on the number of ligands (Table 4). Those classes that had at least 50 active ligands were used in this study. All other protein classes are included except for the metalloenzymes, which were excluded because there were not enough ligands in DUD for any of the metalloenzymes.

2.3.2. Decoy Sets. As suggested in the original DUD paper,[25] two decoy sets were created. The "own decoys" set contains the DUD decoys selected for the target, and the "combined decoys" set consists of all the decoys used in the selected 12 targets combined together. This "combined decoys" consists of about 60% of the DUD decoys, so direct comparisons cannot be made with the "amalgamated test set" in the original DUD paper. The "combined decoys" set is used to simulate a virtual screening scenario, where there is a larger number of heterogeneous compounds available.

2.3.3. Conformation Generation and Partial Charges. Since all the methods used in this study consider molecules as rigid structures, conformations had to be pregenerated using OMEGA with the MMFF94s force field.[26] The number of conformations was limited to ten for the virtual screening sets. Partial charges were assigned using the MMFF94 method implemented in MolCharge.[27]

2.3.4. Molecule Superimpositioning. Both BRUTUS and ROCS were used to produce superimpositions for FC. Both methods produced three different superimpositionings for each conformation in the database. The main difference between these two methods is the way that molecular energy fields are represented. ROCS uses a set of analytic Gaussian functions, and BRUTUS is a grid-based method.[12]

For ROCS, ComboScore with default ImplicitMillsDean force field was used to score the superimposed structures. The ComboScore is simply the sum of the shape Tanimoto and scaled color score. The shape Tanimoto measures the overall complementary of two shapes

$$T_{AB} = \frac{O_{A,B}}{I_A + I_B - O_{A,B}} \tag{4}$$

where $O_{A,B}$ is the optimized overlap between molecules A and B, I_A is the self-overlap for molecule A, and I_B is the self-overlap for molecule B .

The color score is calculated by combining all the color atoms in the query and summing the best color interaction with the hit molecule. The color atoms and interactions are defined by 2D-SMARTS based rules. A more detailed description of ROCS can be found elsewhere.[7,22,28]

BRUTUS version 0.8.9 evaluates the similarities of molecules through molecular volume and electrostatic energy fields.[12] The similarity of the molecular volumes and electrostatic energy fields is computed separately with the Hodgkin index[29]

$$H_{AB} = \frac{2\int_v \rho A \rho B dV}{\int_v \rho_A^2 dV + \int_v \rho_B^2 dV} \tag{5}$$

where ρ_A and ρ_B are density functions of energy fields A and B. The volumic similarity S_v and electrostatic similarity S_e thus computed are then combined into the total score T_s

$$T_s = wS_v + (1-w)S_e \tag{6}$$

where w is a weighting factor. In this study, w was set to 0.5.

Template molecules for BRUTUS, ROCS, and EON were taken in their bioactive conformation from the Protein Databank using the same crystal structures as in the original DUD paper.[25,30] Atom and bond types were corrected with SYBYL.[31] Hydrogen atoms and partial charges with MMFF94 were added using the same software. The template molecule selected for superimpositioning might not have been ideal in some of the cases studied here. This might have a negative effect on the models, and probably they could be improved significantly. However, if there had been any major human intervention, then the results would have been difficult to compare.

2.3.5. Other Virtual Screening Methods. Since FC requires information on several active compounds instead of a single ligand, it should outperform similarity metrics which rely only on a single active conformation. EON[32] calculates the electrostatic Tanimoto similarity between a query molecule and a set of superimposed molecules from ROCS. EON is fast enough to handle large databases. Recently, a novel

melanin-concentrating hormone receptor 1 antagonist was discovered from a combinatorial library of over 3 million compounds with EON.[33]

The similarity of two electrostatic fields A and B can be expressed as

$$E_{AB} = \frac{\int A(\vec{r}) * B(\vec{r})}{\int A(\vec{r}) * A(\vec{r}) + \int B(\vec{r}) * B(\vec{r}) - \int A(\vec{r}) * B(\vec{r})} \quad (7)$$

where $\int A(\vec{r})$ is the electrostatic field of A and $\int B(\vec{r})$ is the electrostatic field of B.

In this study, compounds were ranked using the ET_Combo score, which is simply the sum of the shape similarity T_{AB} and the electrostatic similarity E_{AB}. The molecules used for superimpositioning were used in similarity calculations.

As one of the reviewers pointed out, one could calculate the similarities with all 15 molecules used in the models. This does indeed yield superior results compared to FC and reflects the diversity of active molecules (data included in the Supporting Information). However, this strategy becomes troublesome very quickly as the number of active compounds and the size of the database increases.

2.3.6. Model Building. The FC model should have a diverse selection of active compounds with different chemical scaffolds. They should also be superimposable into the model. Even though these factors affect the model and its performance, they are somewhat subjective and complicate comparisons with simpler methods.

To ensure objective selection of compounds for the FC models, the following protocol was applied. Since the activities for DUD ligands were not readily available, compounds were only considered to be either active or inactive. To maximize the chemical diversity of the model compounds, GRIND-descriptors for the ligands were calculated using three probes (DRY, O carbonyl, and N amide).[34] Then, two-component principal component analysis (PCA) was performed with ALMOND.[35] From these analyses, 15 compounds were selected for each target using the Kennard-Stone uniform subset selection algorithm implemented by Daszykowski and co-workers with GNU Octave.[36,37] The compounds used in the FC models were removed from all virtual screening data sets.

BRUTUS was used to superimpose the compounds used in the model. Three possible solutions were generated for each molecule. From these superimpositions, the best one for the model was selected by visual inspection.

2.3.7. Evaluating VS Results. The best ranked superimpositioning was used as a score for the compounds. Virtual screening is often evaluated by using enrichment factors. The enrichment factor is defined as[38]

$$F = \frac{AC}{BD} \quad (8)$$

where A is the number of ligands retrieved in a subset; B is the total number of ligands in the database; C is the total number of molecules in database; and D is the total number of molecules in a subset.

However the enrichment factor is a problematic measure, since it relies on cutoff fraction f made at a certain point, and this can be sensitive to small changes in ranking.[39] Another issue is that the maximum enrichment factor is limited by C/B or $1/f$, whichever is smaller.[40] The Receiver Operating Characteristic (ROC) curve has been suggested as a better measure of VS,[41] and the Area under ROC curve (ROC AUC) can be used as a measure of overall enrichment. If the ROC AUC is 0.6 or lower, then the performance is poor. The highest possible value is 1.0. The details of the ROC method have been described by Triballeau and co-workers.[41] ROC and ROC AUCs were calculated with the ROCR package.[42] The Wilcoxon signed ranks test was used to assess the statistical significance of the results ($\alpha = 0.05$).[43,44]

However, ROC AUC suffers from the "early recognition" problem as recently pointed out by Truchon and Bayly.[45] This means that one should weigh the top of the hit-list more than the bottom, since molecular databases are huge and only the very top of the hit-list can be tested in vitro. To evaluate early enrichment in this study, enrichment factors at the top 1% of the hit-list (EF_1) were calculated. Since the top 1% of the "own decoys" set is smaller than the number of active compounds, only the enrichment factors for the "combined decoys" set are presented.

Perhaps more important than simple enrichment, virtual screening should produce unique chemical structures for lead discovery ("scaffold hopping"). However, it is very difficult to assess scaffold hopping objectively with this data, since active compounds are not classified in any way.

Table 5. van der Waals Peak Distributions in Models (1 Å Resolution)

target	points	Inside	Inside& Near	Near	Near& Outside	Outside
AR	16675	240	200	3582	850	11803
ER_agonist	14283	235	191	3468	791	9598
GR	16875	292	247	4029	969	11338
EGFr	18975	259	248	4035	1196	13237
FGFr1	22599	256	301	4199	1535	16308
P38	22707	213	354	3825	1941	16374
SRC	25839	220	316	3920	1656	19727
FXa	27869	292	313	4442	1374	21448
DHFR	15525	210	270	3586	1286	10173
AChE	24025	272	272	4307	1191	17983
COX-2	15341	243	238	3631	854	10375
InhA	19251	287	239	4084	1217	13424

3. RESULTS AND DISCUSSION

3.1. Model Building and Scoring Algorithms. To obtain an overview of the FC models, an analysis of peak distributions was performed. The numbers of different peaks are presented in Tables 5 and 6. The models displayed a very similar distribution of the van der Waals peaks. The reason for this phenomenon is that template molecules used for superimpositioning are roughly of the same size. The differences in "Outside" peaks in the van der Waals scoring are attributable to larger grid boxes for certain targets. No histograms with "Inside&Outside" or "None" peak cases were found in van der Waals material. With respect to the electrostatics, the nuclear hormone targets (AR,ER_agonist, and GR) had distributions different from the other targets. This reflects their partial charges, which are close to zero in most cases. Positive electrostatics seem to be dominant, which is due to the positively charged nitrogens in the original DUD data.

Since there were positively charged compounds in the model, active and inactive compounds were mostly dif-

Table 6. Electrostatic Peak Distributions in Models (1 Å Resolution)[a]

target	points	Neg.	Neg.& Zero	Neg.& Pos.	Zero	&Pos.	Pos.	None
AR	16675	180	628	38	2123	857	211	21
ER_agonist	14283	95	333	79	2513	729	145	19
GR	16875	61	531	90	2833	700	265	20
EGFr	18975	20	28	231	27	2132	2238	9
FGFr1	22599	12	29	840	30	1149	3024	18
P38	22707	1	21	153	17	2633	2156	9
SRC	25839	2	32	175	55	2349	2229	16
FXa	27869	2	2	105	16	501	4604	3
DHFR	15525	59	188	406	276	2175	1111	149
AChE	24025	0	2	31	4	211	4791	1
COX-2	15341	262	321	287	1025	1313	920	42
InhA	19251	128	135	272	106	3024	1096	8

[a] Only those peaks are shown that are included in the scoring process (points near the surface).

Table 7. Effect of Grid Spacing on Selected Models: ROC AUCs for "Own Decoys" Sets

	grid spacing			
target	0.5 Å	1.0 Å	2.0 Å	5.0 Å
AR	0.802	0.803	0.807	0.640
EGFr	0.805	0.798	0.814	0.691
FXa	0.914	0.915	0.915	0.907
DHFR	0.830	0.830	0.835	0.733
COX-2	0.901	0.896	0.897	0.868
mean	0.851	0.849	0.853	0.768
median	0.830	0.830	0.835	0.733

Table 8. Different Classification Limits: ROC AUCs for "Own Decoys" Set

target	0.5 * limits	normal limits	2.0 * limits
AR	0.809	0.803	0.762
EGFr	0.738	0.798	0.810
FXa	0.911	0.915	0.919
DHFR	0.797	0.830	0.833
COX-2	0.879	0.896	0.903
mean	0.827	0.849	0.845
median	0.809	0.830	0.833

ferentiated by fitting into "positive peak. This is especially true for FXa, where FC exhibits a dramatic difference with EON (see the next section).

The effect of grid spacing was studied on one target from each protein family that had an ROC AUC value approximately equal to 0.8 or higher at 1 Å resolution (Table 7). It seems that the spacing of 1 Å or 2 Å is optimal for most cases. Surprisingly, grid spacing of 5 Å still yields high ROC AUC on COX-2 and FXa. This kind of crude spacing will however lead to several ties in the hit-list and thus complicate the selection of the top compounds.

The classification limits can also be adjusted. The default classification limits were used in retrospective screening, since they produced the highest average ROC AUC (Table 8). However, the differences are insignificant. It is possible that different limits should be used for different kinds of targets, but that is is beyond of the scope of this paper.

The orientation of molecules in the model is a critical step. This is illustrated in Figure 2, which shows two different FC models for AR. Both had the same crystal structure as a starting point. In one of the models, the coordinates of crystal

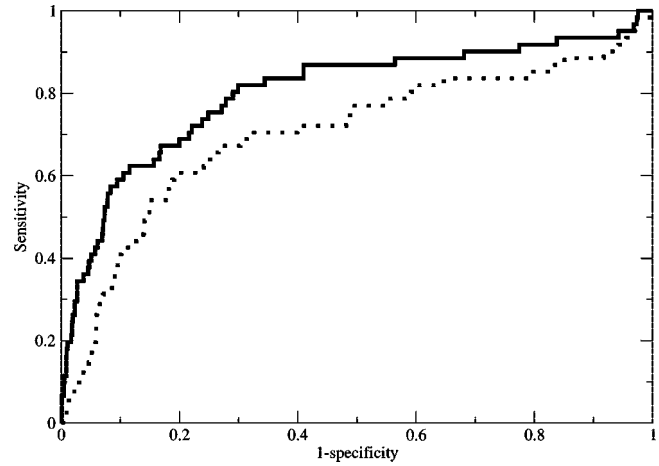

Figure 2. The effect of orientation of the molecules in the model. ROC curves for two different FC models (AR, own decoys set). The original model (solid line) outperforms the new model (dotted line).

Table 9. ROC AUCs with Superimpositionings from BRUTUS and ROCS[a]

	combined decoys set		own decoys set	
target	FC/BRUTUS	FC/ROCS	FC/BRUTUS	FC/ROCS
AR	0.930	0.898	0.803	0.810
ER_agonist	0.933	0.915	0.775	0.767
GR	0.912	0.627	0.814	0.594
EGFr	0.822	0.732	0.798	0.777
FGFr1	0.620	0.472	0.585	0.515
P38	0.777	0.740	0.735	0.801
SRC	0.640	0.513	0.702	0.644
FXa	0.928	0.868	0.915	0.906
DHFR	0.860	0.939	0.830	0.957
AChE	0.874	0.863	0.516	0.567
COX-2	0.909	0.867	0.896	0.891
InhA	0.804	0.767	0.832	0.826
mean	0.834	0.767	0.767	0.755
median	0.867	0.815	0.801	0.789

[a] Wilcoxon signed pairs test: combined decoys set $P = 0.01 < 0.05$, own decoys set $P = 0.5186 > 0.05$.

structure were transferred to another position. The two models have different performances, which is probably due the differences in superimpositionings. The authors acknowledge this as a major problem with FC. However, a similar problem exists with other grid-based methods like CoMFA.[46]

3.2. Retrospective Virtual Screening. FC ROC AUC averages were quite similar with both BRUTUS and ROCS superimpositions, even though there is a statistically significant difference in the combined decoys set (Table 9). It seems that FC could be used with both methods. This is not surprising, since both methods have been shown to produce reasonable superimpositionings.[7,12]

The early enrichment measured by enrichment factors at 1% (EF_1) of a ranked database for FC and EON are shown in Table 10. Both methods displayed a similar overall performance when one examines average and median. FC outperformed EON on nuclear hormone targets (AR, ER_agonist, and GR), whereas EON exhibited higher enrichment on AChE and InhA. There was high enrichment on COX-2, which was also reported in the previous study on EON by Nicholls and co-workers.[8]

The "own decoys" set seems to be more demanding than the larger set, since ROC AUCs are lower for both methods

Table 10. Enrichment Factors at 1% of Ranked Database[a]

	combined decoys set	
target	FC	EON
AR	66.19	64.47
ER_agonist	61.62	28.88
GR	30.20	6.36
EGFr	9.10	10.73
FGFr1	0.00	0.00
P38	0.42	3.32
SRC	0.72	4.29
FXa	4.73	0.79
DHFR	1.08	3.76
AChE	37.79	64.47
COX-2	54.41	45.69
InhA	11.43	28.57
mean	23.14	18.10
median	10.26	8.55

[a] Both methods have similar overall performance. FC outperforms EON on nuclear hormone targets (AR, ER_agonist, and GR), while EON has a higher enrichment factor on AChE and InhA. The maximum enrichment factor is 100.

Table 11. ROC AUCs for All Targets[a]

	combined decoys set		own decoys set	
target	FC	EON	FC	EON
AR	0.930	0.677	0.803	0.676
ER_agonist	0.933	0.739	0.775	0.619
GR	0.912	0.593	0.814	0.725
EGFr	0.822	0.713	0.798	0.702
FGFr1	0.620	0.298	0.585	0.456
P38	0.777	0.572	0.735	0.596
SRC	0.640	0.338	0.702	0.292
FXa	0.928	0.249	0.915	0.399
DHFR	0.860	0.767	0.830	0.769
AChE	0.874	0.910	0.516	0.778
COX-2	0.909	0.878	0.896	0.884
InhA	0.804	0.715	0.832	0.758
mean	0.834	0.621	0.767	0.638
median	0.867	0.695	0.801	0.689

[a] FC outperforms EON in both data sets (Wilcoxon signed rank test: combined decoys $P = 0.001465 < 0.05$, own decoys $P = 0.02100 < 0.05$).

with the "own decoys" test set (Table 11). This was also noted in the original DUD article. However, since the main purpose of FC is to process larger databases, the differences in larger data sets are more important.

The median of ROC AUC for FC is over 0.8 for both decoy sets, which is significantly higher than the ROC AUC value for EON. This suggests that FC could be used as a crude molecular sieve. FC can be calculated quickly and therefore is suitable for large databases.

The DUD decoys are selected such that their 2D fingerprints are dissimilar to the active molecules, so it is possible that good enrichment in this study is simply related to 2D differences. However, this is unlikely because FC showed improved performance compared to another 3D-method, EON.

3.3. Conclusions and Future Remarks. In this paper, a novel method for ligand-based virtual screening was described. Even though there was a statistically significant difference between the ROC AUC values of FC and EON, it is obvious that this does not necessarily represent a measure of the actual usefulness of the method. This study serves only as a proof of concept on how to use 3D models for high-throughput virtual screening. The models used in this study were crudely built by a semiautomatic process. It is most likely that the results could be improved by more careful selection of compounds and by adjusting the various parameters utilized in the method.

There has been an increasing interest in the early prediction of ADMET (Absorption Distribution, Metabolism, Excrection and Toxicity) properties in silico, and a large number of 3D-QSAR models have been published for prediction of such properties.[47] The prospective usefulness of most of these models is however debatable, since normally only small training and test sets have been used instead of a full virtual screening scenario.[46] Due to the problems related with 3D-QSAR, the most commonly used ADMET models are therefore still based on simple 2D-properties.[48] However, rapid 3D-based methods have also been shown to be efficient (for example, VolSurf).[49] It is our next task to investigate if FC is suitable for such predictions.

ACKNOWLEDGMENT

We thank the Finnish Funding Agency for Technology and Innovation for financial support and CSC-Scientific Computing Ltd. for computing resources. The anonymous reviewers are thanked for their constructive comments on the manuscript.

Supporting Information Available: The lists of compounds used in the FC models, the hit-lists from FC and EON, and the EON results with 15 query molecules. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11*, 1046–1053.
(2) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
(3) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **2005**, *26*, 11–22.
(4) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
(5) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. Critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(6) Todeschini, R.; Consonni, V. Handbook to molecular descriptors. In *Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kybinyi, H., Folkers, G., Eds.; Wiley-VCH: Weinheim, Germany, 2000; Vol. 11.
(7) Grant, J. A.; Gallard, M. A.; Pickup, B. T. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.
(8) Nicholls, A.; MacGuish, N. E.; MacGuish, J. D. Variable selection and model validation of 2D and 3D molecular descriptors. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 451–474.
(9) Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. Field interaction and geometrical overlap: A new simplex and experimental design based computational procedure for superposing small ligand molecules. *J. Med. Chem.* **2003**, *46*, 1359–1371.
(10) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular field extrema as descriptors of biological activity: definition and validation. *J. Chem. Inf. Model.* **2006**, *46*, 665–676.
(11) Wolber, G.; Dornhofer, A. A.; Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 773–788.

(12) Ronkko, T.; Tervo, A. J.; Parkkinen, J.; Poso, A. BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition. II. Description and characterization. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 227–236.
(13) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 647–671.
(14) Langer, T.; Hoffman, R. D. Pharmacophores and Pharmacophore Searches. In *Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Folkers, G., Eds.; Wiley-VCH: Weinheim, Germany, 2006; Vol. 32.
(15) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2004**, *45*, 160–169.
(16) Chen, J.; Lai, L. Pocket v.2: Further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.* **2006**, *46*, 2684–2691.
(17) Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics* **2006**, *22*, 1449–1455.
(18) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(19) Putta, S.; Lemmen, C.; Beroza, P.; Greene, J. A novel shape-feature based approach to virtual screening. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1230–1240.
(20) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947–961.
(21) Kotani, T.; Higashiura, K. Comparative Molecular Active Site Analysis (CoMASA). 1. An approach to rapid evaluation of 3D QSAR. *J. Med. Chem.* **2004**, *47*, 2732–2742.
(22) *ROCS, version 2.3*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(23) Tervo, A. J.; Ronkko, T.; Nyronen, T. H.; Poso, A. BRUTUS: optimization of a grid-based similarity function for rigid-body molecular superposition. 1. Alignment and virtual screening applications. *J. Med. Chem.* **2005**, *48*, 4076–4086.
(24) Dill, K. A.; Bromberg, S. *Molecular driving forces*, 1st ed.; Garland Science: New York, NY, 2002; pp 370−371.
(25) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(26) *OMEGA, version 2.1*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(27) *MolCharge, version 1.3.1*; OpeneEye Scientific Software, Inc.: Santa Fe, NM.
(28) Jennings, A.; Tennant, M. Selection of molecules based on shape and electrostatic similarity: proof of concept of "Electroforms". *J. Chem. Inf. Model.* **2007**, *47*, 1829–1838.
(29) Hodgkin, E. E.; Richards, W. G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105–110.
(30) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(31) *SYBYL, version 7.1*; Tripos, Inc.: St. Louis, MO.
(32) *EON, version 2.0.1*; OpenEye Scientific Software, Inc.: Santa Fe, NM.
(33) Muchmore, S. W.; Souers, A. J.; Akritopoulou-Zanze, I. The use of three-dimensional shape and electrostatic similarity searching in the identification of a melanin-concentrating hormone receptor 1 antagonist. *Chem. Biol. Drug Des.* **2006**, *67*, 174–176.
(34) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.
(35) *ALMOND, version 3.3*; Molecular Discovery, Ltd.: Perugia, Italy.
(36) Daszykowski, M.; Walczak, B.; Massart, D. L. Representative subset selection. *Anal. Chim. Acta* **2002**, *468*, 91–103.
(37) *GNU Octave, version 2.1.73*. http://www.gnu.org/software/octave (accessed Oct 16, 2007).
(38) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **2006**, *46*, 401–415.
(39) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74–82.
(40) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.
(41) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual screening workflow development guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2537.
(42) Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **2006**, *21*, 3940–3941.
(43) Demsar, J. Statistical comparisons on classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
(44) *R, version 2.6.0*. http://www.r-project.org (accessed Nov 8, 2007).
(45) Truchon, J.-F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "Early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
(46) Doweyko, A. M. 3D-QSAR illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587–596.
(47) Norinder, U.; Bergstrom, C. A. S. Prediction of ADMET properties. *ChemMedChem* **2006**, *1*, 920–937.
(48) Lobell, M.; Hendrix, M.; Hinzen, B.; Keldenich, J.; Meier, H.; Schmeck, C.; Schohe-Loop, R.; Wunberg, T.; Hillisch, A. In silico ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* **2006**, *1*, 1229–1236.
(49) Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204–2216.

CI700216U