

# Comparison of Ligand- and Structure-Based Virtual Screening on the DUD Data Set

Modest von Korff,\* Joel Freyss, and Thomas Sander

Department of Research Informatics, Actelion Ltd., Gewerbestrasse 16, CH-4123 Allschwil, Switzerland

Received August 29, 2008

Several in-house developed descriptors and our in-house docking tool ActDock were compared with virtual screening on the data set of useful decoys (DUD). The results were compared with the chemical fingerprint descriptor from ChemAxon and with the docking results of the original DUD publication. The DUD is the first published data set providing active molecules, decoys, and references for crystal structures of ligand-target complexes. The DUD was designed for the purpose of evaluating docking programs. It contains 2950 active compounds against a total of 40 target proteins. Furthermore, for every ligand the data set contains 36 structurally dissimilar decoy compounds with similar physicochemical properties. We extracted the ligands from the target proteins to extend the applicability of the data set to include ligand based virtual screening. From the 40 target proteins, 37 contained ligands that we used as query molecules for virtual screening evaluation. With this data set a large comparison was done between four different chemical fingerprints, a topological pharmacophore descriptor, the Flexophore descriptor, and ActDock. The Actelion docking tool relies on a MM2 forcefield and a pharmacophore point interaction statistic for scoring; the details are described in this publication. In terms of enrichment rates the chemical fingerprint descriptors performed better than the Flexophore and the docking tool. After removing molecules chemically similar to the query molecules the Flexophore descriptor outperformed the chemical descriptors and the topological pharmacophore descriptors. With the similarity matrix calculations used in this study it was shown that the Flexophore is well suited to find new chemical entities via “scaffold hopping”. The Flexophore descriptor can be explored with a Java applet at <http://www.cheminformatics.ch> in the submenu Tools→Flexophore. Its usage is free of charge and does not require registration.

## INTRODUCTION

Ligand-based virtual screening is established as a standard procedure in drug discovery. The descriptors used to describe the molecules are an important factor for success in ligand-based virtual screening. There exists a broad variety, from molecular bulk descriptors over 2D chemical fingerprints to descriptors taking pharmacophoric features and molecule conformations into account.<sup>1</sup> Theoretically, any of these descriptors can be used to find new bioactive molecules. The challenge is to find bioactive molecules that open a door to an unexplored area of the chemical space. A successful virtual screening results in new scaffolds, which can be explored by medicinal chemistry. The experience of recent years has shown that there is no gold standard, which can be used to search for new molecules.<sup>2–4</sup> This impression is only a qualitative one, because it is almost impossible to compare different descriptors from the literature. In many publications proprietary data sets are used; the descriptors under consideration are only applied to a limited number of targets, or only one descriptor is used. No standard data sets exist for the unbiased validation of virtual screening procedures. The procedure comprises the data set, a similarity metric, the query, and the description of the molecules. Similarity metrics range from Tanimoto coefficients for ligand based virtual screening to docking scores for structure based methods. The query can be a protein crystal structure

or the descriptor of a bioactive molecule. During the last years the need for validation data sets was realized, which resulted in several publications.<sup>5–7</sup> A drawback in most of these publications was that only the bioactive molecules were specified and the inactive molecules, termed decoys, were not. However, the nature of the decoy compounds used has a strong impact on the success of the virtual screening.<sup>4,6</sup> One of the most recent approaches to create a validation data set is the data set of useful decoys (DUD).<sup>8</sup> In this study, the DUD was used for its broad variety of targets and ligands and because crystal structures are provided for all targets along with the structures of further active molecules and the decoys. The aim of this study was to explore the chemical space from the bioactivity point of view. Seven different candidates are compared: Actelion's in-house developed docking tool, four chemical fingerprints, a topological pharmacophore descriptor, and the recently introduced Flexophore descriptor.<sup>9</sup> Which of the candidates would succeed best in the detection of bioactive molecules? Which one would be able to detect new chemical entities? Are there real advantages in using a quite complex descriptor like the Flexophore?

## METHODS

**Descriptors.** Four of the six molecular descriptors used in this study were chemical fingerprints. Except for the ChemAxCFp<sup>10</sup> descriptor, they were all developed in-house. All four chemical fingerprint descriptors are topological descriptors and are therefore devoid of 3D information. A

\* Corresponding author phone: +41 61 565 63 23; fax: +41 61 565 65 00; e-mail: [modest.korff@actelion.com](mailto:modest.korff@actelion.com).

**Table 1.** Distance Table from the Protein Data Bank for Primary Amides to a Alcoholic Hydroxyl Group

		Primary Amide to Alcohol															
<i>r</i>	...	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	4.75	...	6.75
<i>N(r)</i>	...	0	0	3	58	322	1204	1186	832	870	904	877	940	1237	1437	...	2691
<i>g(r)</i>	...	0	0	0.06	0.86	3.91	12.18	10.2	6.11	5.54	5.03	4.31	4.11	4.84	5.06	...	4.76
$\Delta W(r)$	...	16.86	9.56	3.93	0.21	-1.61	-2.10	-1.84	-1.20	-0.53	-0.07	0.16	0.23	0.22	0.19	...	0.03

comprehensive introduction to molecular fingerprints was given by Xue and Bajorath.<sup>11,12</sup>

**Actelion Fragment Fingerprint (FragFp).** The FragFp is a dictionary based descriptor with a length of 512 bits. Each bit represents a substructure fragment of which some include wildcard atoms.<sup>13</sup> The dictionary of 512 substructures was created by a computational procedure, which had been optimized to balance two goals: (1) anyone of these fragments should occur frequently in organic molecule structures and (2) all fragments should be linear independent concerning their substructure-match-pattern in diverse organic compound sets. To generate a descriptor vector the molecular structure is searched for any of the substructures in the dictionary. For any match, the corresponding bit of the vector is set to 1. Any molecular structure is represented by a binary vector of length 512. The FragFp descriptor belongs to the same class as the "MDL structure keys"<sup>14</sup> which have recently been shown to perform better in virtual screening than 3D descriptors.<sup>15</sup>

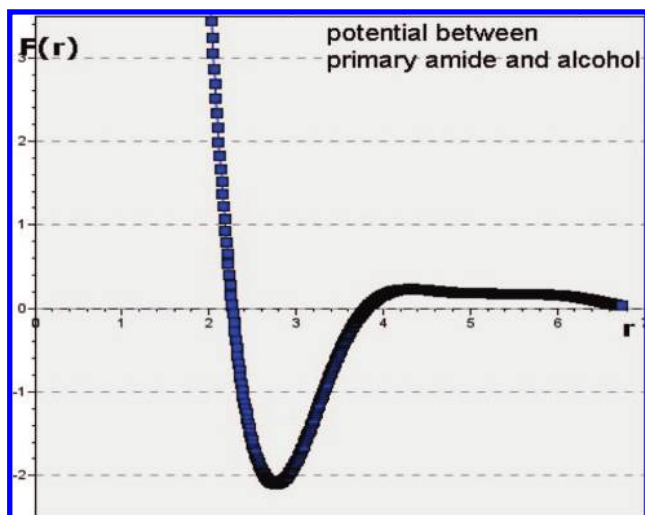
**Actelion Path Fingerprint (PathFp).** The PathFp is a molecular graph path walking fingerprint descriptor. All distinguishable paths with up to 7 atoms are hashed into a descriptor vector of 512 bits. This descriptor is conceptually similar to ChemAxCFp, the chemical fingerprints from ChemAxon,<sup>10</sup> and to the Daylight<sup>16</sup> fingerprints.

**Actelion Sphere Fingerprint (SphereFp).** The SphereFp locates three circular layers with increasing bond distance around each atom. This yields four fragments starting with the naked central atom, adding one layer at a time. Every fragment is encoded as a canonical string (id-code), similar to the generation of canonical SMILES.<sup>17,18</sup> The string is then assigned to one of the unique 1024 bits. For this the hash value of the id-code is calculated, and the corresponding bit in the vector is set. The Hashlittle algorithm from Jenkins<sup>19</sup> is used as binning function, which takes a text

string as input and returns an integer value between 0 (inclusive) and 1024 (exclusive). This hash function showed in preliminary experiments a good uniform distribution of the generated hash values. A calculation of  $e = 4$  spheres is done for all atoms  $g$  in the molecule serving once as center atom, which results in  $b = eg$  number of id-codes. The number of bits set in the descriptor vector is  $\leq b$ , depending on the number of unique id-codes and hash collisions.

**Actelion Topological Pharmacophore Histogram (TopPPHist).** The TopPPHist descriptor is an in-house development quite similar to a descriptor developed together with ChemAxon.<sup>20</sup> The TopPPHist descriptor is closely related to the atom pair descriptor<sup>21</sup> and to the binding property pair's descriptor.<sup>22</sup> For the detection of the pharmacophore points (PPs) a combination of the defined substructures and logic expressions was used. The substructures for the pharmacophore point definitions were chosen according to Greene et al.<sup>23</sup> From preliminary experiments, it was decided to use the following pharmacophore types: hydrogen bond donor (d), hydrogen bond acceptor (a), hydrophobic (h), positive charged (-), negative charged (-), and aromatic (r). The PPs are put into a relationship by the histograms of the topological distance counts. The topological distances between each pair of atoms belonging to a certain pharmacophore type are determined. The maximum topological distance was set to 12 bond lengths. A histogram is generated for every possible combination of pharmacophore types that contains one entry for any two atoms that match the particular pharmacophore types. The number of bonds between these atoms determines the bin of the histogram, which is increased by one. The counts are added to the histogram for the pharmacophore point pair combination. All resulting histograms from one molecule are written into a descriptor vector.

**Flexophore.** The generation of the Flexophore descriptor and its similarity metric were recently described in detail in this journal.<sup>9</sup> The Flexophore descriptor represents the molecule by a complete graph. This is a simple graph where each pair of vertices is connected by an edge. The vertices are labeled with enhanced MM2 atoms types.<sup>24</sup> A vertex can contain several labels. The edges are histograms of the vertex distances resulting from diverse molecule conformations. The similarity function introduces the target point of view into the descriptor comparison. Interaction statistics for the enhanced MM2 atom types were derived from the crystallographic data in the Protein Data Bank (PDB).<sup>25</sup> A similarity matrix of these atom types, concerning their interaction behavior, was created. The generated table contains the similarity between any two vertices. Edge similarities are calculated from the fraction of overlap between two distance histograms. To calculate the similarity of two Flexophore descriptors, the maximum common substructure (MCS) of two descriptor graphs is determined, and an overall similarity score is calculated for the best match of all vertices

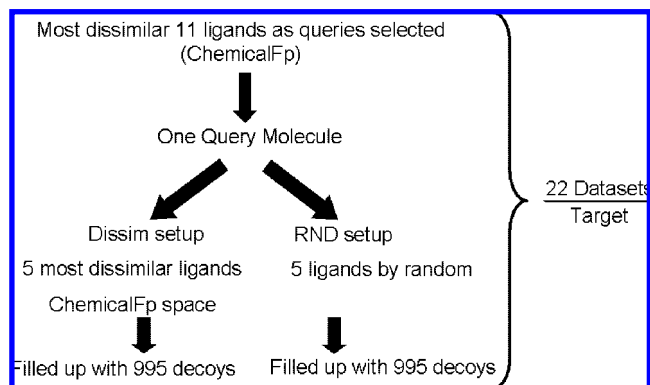


**Figure 1.** Smoothed energy potential between primary amide and alcohol.

**Table 2.** Results of the Docking Set Tested on FlexScreen, Glide, Gold, Flexx, and ActDock<sup>a</sup>

	FlexScreen	Glide	Gold	FlexX	ActDock
1a28	0.33				0.31
1a4q	0.74				4.49
1a6w	0.88				0.6
1abe	0.38	0.17	0.86	1.16	0.33
1abf	0.58	0.2		1.27	0.25
1aoe	0.51				
1apt	1.04	0.58	1.62	1.89	1.35
1apu	0.73				1.11
1aqw	7.76				5
1atl	3.93	0.94		2.06	0.85
1b58	0.96				
1bma	1.03	9.31		13.41	
1byb	0.79	10.49		1.62	
1c1e	5.14				0.64
1c5c	0.43				4.29
1c5x	1.88				0.42
1c83	0.48				0.48
1cbs	0.38	1.96		1.68	0.51
1cil	1.63	3.82		3.85	0.41
1coy	0.49	0.28	0.86	1.06	0.81
1d0l	1.11				
1d3h	0.47				0.34
1ejn	0.4				1.12
1eta	8.43	2.92	11.21	8.46	
1f3d	0.68				0.46
1fen	0.56	0.66		1.39	0.86
1flr	0.58				0.53
1glp	0.33	0.34		0.47	0.72
1glq	1.28	0.29	1.35	6.43	4.98
1hfc	2.49	2.24		2.51	0.46
1hfv	1.04				
1hsb	0.41				1.87
1hsl	0.79	1.31	0.97	0.59	1.37
1hvr	0.64	1.5		3.35	
1hyt	1.08	0.28	1.1	1.62	0.7
1ida	1.14	11.88	12.12	11.95	
1jap	1.4				0.83
1kel	6.4				0.92
1lcp	0.56	1.98		1.65	3.94
1lic	1.08	4.87	10.78	5.07	3.37
1lna	2.49	0.95		5.4	1.93
1lst	0.6	0.14	0.87	0.71	0.27
1mld	0.63	0.32		1.45	1.29
1mmq	0.66	0.92		0.52	
1mrg	0.5	0.3		0.81	0.42
1mrk	0.86	1.2	1.01	3.55	3.04
1mts	0.57				
1nco	0.34	6.99		5.85	3.61
1phd	0.95	1.22	0.85	0.65	
1phg	0.28	4.32	1.35	4.74	
1ppc	1.38	7.92		3.05	
1pph	2	4.31		4.91	3.45
1qbr	0.46				
1qbu	0.5				
1rds	0.59	3.75	4.78	4.89	
1rnt	1.02	0.72		1.9	3.48
1rob	1.16	1.85	3.75	7.7	0.51
1slt	0.87	0.51	0.78	1.63	1.29
1snc	6.19	1.91		7.48	0.9
1srj	7.23	0.58	0.42	2.36	0.42
1tmn	0.83	2.8	1.68	0.86	
1tng	0.36	0.19		1.93	0.61
1tnh	0.82	0.33		0.56	0.34
1tni	2.8	2.18		2.71	1.93
1tnl	2.73	0.23		0.71	3.96
1tyl	7.37	1.06		2.34	1.96
1ukz	6.21	0.37		0.94	0.7
1wap	0.26	0.12		0.57	0.54
1xid	3.22	4.3	0.92	2.01	3.94
2ak3	0.45	0.71	5.08	0.91	0.49
2cmd	0.55	0.65		3.75	0.44
2cpp	0.37	0.17		2.94	2.65
2ctc	1.64	1.61	0.32	1.97	0.36
2fox	0.84				
2gbp	0.66	0.15		0.92	0.18
2qwk	0.97				0.27
2tmn	1.22	0.58		5.16	0.67
2tsc	1.52				
3ert	0.58				0.87
3tpi	0.38	0.49	0.8	1.07	0.3
4dfr	0.94	1.12	1.44	1.4	
5abp	0.43	0.21		1.17	0.34
6rnt	5.83	2.22	1.2	4.79	0.39
7tim	1.25	0.14	0.78	1.49	1.02
rmsd < 1	59.52%	52.63%	44.00%	22.81%	61.90%
rmsd < 2	80.95%	71.93%	76.00%	54.39%	79.37%

<sup>a</sup> The Docking results for FlexScreen, Glide, Gold, and FlexX were taken from ref 28.

**Figure 2.** Scheme to create a Sub<sub>random</sub> data set and a Sub<sub>dissim</sub> data set for Setup 2.

The comparison of two descriptors is computationally expensive. Because the graphs are complete, all pairs of nodes have to be compared. This means a full permutation of the nodes of the smaller graph over all nodes of the larger graph. However, the algorithm for a single substructure match is shortened by breaking the comparison after detecting a nonmatching node pair by using the node similarity threshold. The similarities for all following permutations containing this nonmatching node pair do not have to be calculated. A Java applet to explore the Flexophore descriptor is available at <http://www.cheminformatics.ch> in the submenu Tools→Flexophore.

**ActDock.** All our descriptors were also compared to our in-house docking tool ActDock. ActDock uses a rigid protein-flexible ligand approach. The optimal ligand pose is defined by the minimum of the structure energy and the binding energy. The structure energy is calculated from the MM2 force field, which was implemented in-house as a Java package. To calculate the binding energy, the same interaction statistics was used as for the calculation of the Flexophore descriptors. The development of the algorithm was inspired from the Drugscore scoring function as described by Gohlke et al.<sup>26</sup> Twenty-six thousand files from the PDB database containing protein–ligand complexes were extracted and filtered. Accepted ligands had to be druglike, not overly stretched and at least 60% buried in the cavity. For the description of the ligand-protein interaction 130 atom types were defined. The definition for an atom type included the MM2 atom type plus additional information about the atom valence and its ring conformation. The distance statistics were calculated from the number of occurrences  $N_{i,j}(r)$  between the protein atom  $i$  and the ligand atom  $j$  at a distance between  $r$  and  $r + 0.25$  Å ( $r < 7$  Å).  $N_{i,j}(r)$  was normalized according to distance ( $g_{i,j}(r)$ ) and compared to the reference normalized distribution ( $g(r)$ ) to obtain a distance dependent pair potential  $\Delta W_{i,j}$  that is unbiased from the natural frequency of some atom-types in the PDB database, i.e. N-primary amines to O-ether only have 1300 occurrences in the PDB database but are preferred to O-carbonyl to O-ether with 24000 occurrences.

and edges. A threshold for the node similarity is used to distinguish between the matching and not-matching nodes.



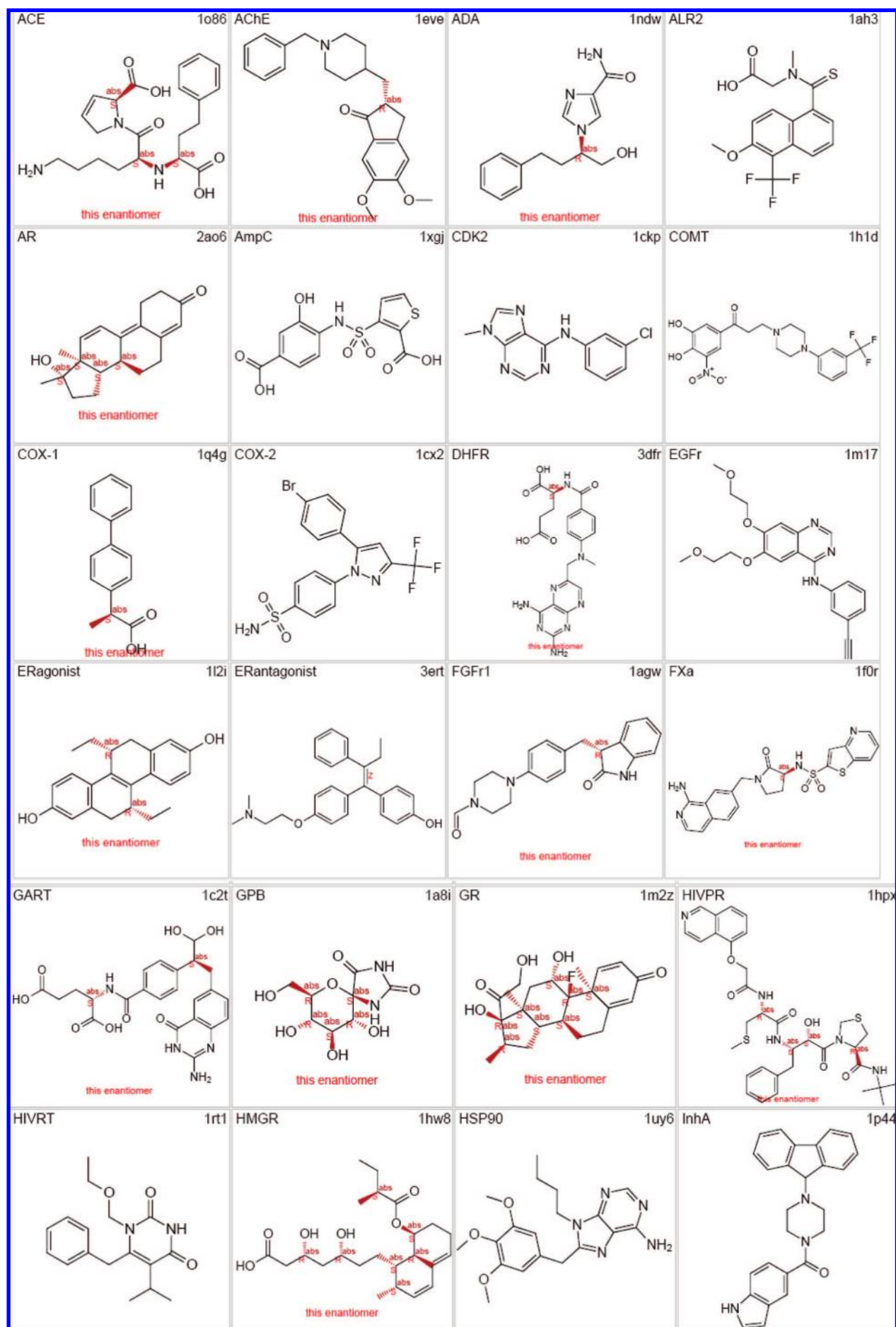
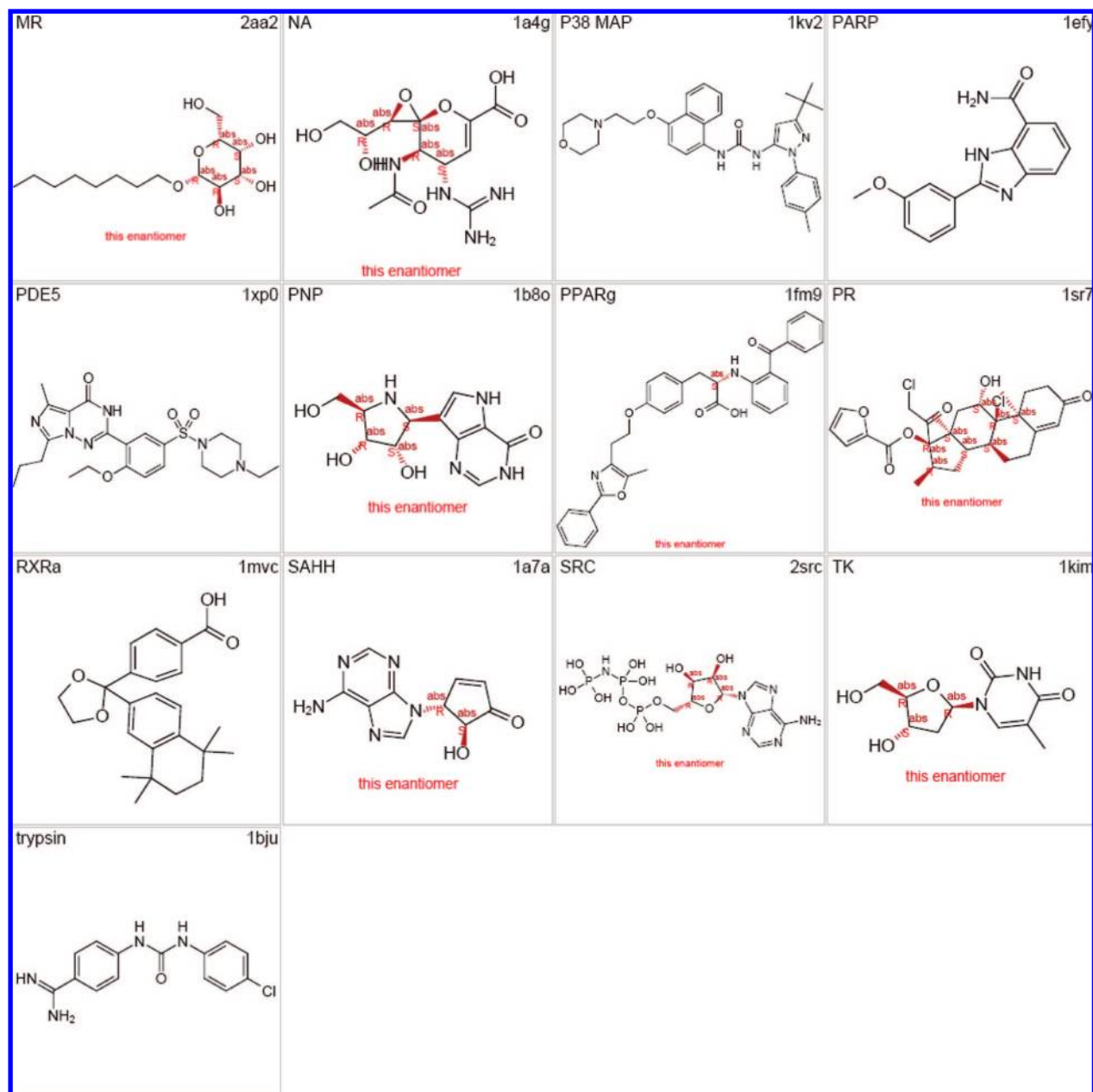
**Chart 1.** Chemical Structures Extracted from the Protein Data Bank Crystals and Used As Query Molecules for Setup<sub>crystal</sub>

Chart 1. Continued

Table 3. Simulated Example for the Rank Based Selection for Set Sub<sub>Dissim</sub> Subsets<sup>a</sup>

ligand	similarity to query			rank			minimum
	FragFp	PathFp	ChemAxCFp	FragFP	PathFp	ChemAxCFp	
1	<b>0.484</b>	<b>0.816</b>	<b>0.698</b>	<b>7</b>	<b>9</b>	<b>9</b>	<b>7</b>
2	0.262	0.256	0.059	3	3	2	2
3	0.143	0.569	0.374	2	5	4	2
4	<b>0.322</b>	<b>0.743</b>	<b>0.898</b>	<b>5</b>	<b>7</b>	<b>12</b>	<b>5</b>
5	0.334	0.101	0.003	6	2	1	1
6	0.969	0.882	0.310	12	11	3	3
7	0.321	0.029	0.839	4	1	10	1
8	<b>0.848</b>	<b>0.791</b>	<b>0.666</b>	<b>11</b>	<b>8</b>	<b>7</b>	<b>7</b>
9	<b>0.617</b>	<b>0.920</b>	<b>0.840</b>	<b>8</b>	<b>12</b>	<b>11</b>	<b>8</b>
10	0.644	0.495	0.440	10	4	6	4
11	0.002	0.877	0.383	1	10	5	1
12	<b>0.642</b>	<b>0.740</b>	<b>0.674</b>	<b>9</b>	<b>6</b>	<b>8</b>	<b>6</b>

<sup>a</sup> Column 1 contains the ligand index. Columns 2–4 contain the similarity between the query and the ligands for the three descriptors. Columns 5–7 contain the rank derived from the similarity values. The lowest rank for each row (ligand) is shown in the last column. The rows with the five highest ranks are in bold; these would be the ligands to be used for the test set.

$$g_{ij}(r) = \frac{N_{ij}(r)/4\pi r^2}{\sum_r (N_{ij}(r)/4\pi r^2)} \quad (1)$$

$$g(r) = \frac{\sum_{ij} g_{ij}(r)}{i * j} \quad (2)$$

$$\Delta W_{ij}(r) = -\ln(g_{ij}(r)/g(r)) \quad (3)$$

To have a smooth and continuous potential without artifacts, we applied smoothing cubic splines to  $\Delta W(r)$ . In Table 1 the  $N$ ,  $g$ , and  $W$  values for the interaction of primary-amines with alcohol groups are given. Application of the cubic splines to the data resulted in the smoothed curve  $F_{ij}(r)$  in Figure 1. The binding energy is the sum of all the individual contributions over each protein–ligand atom pair.

$$\text{binding Energy} = \sum_{a1 \in \text{protein}} \sum_{a2 \in \text{ligand}} F_{\text{atomType}(a1), \text{atomType}(a2)}(\text{dist}(a1, a2)) \quad (4)$$

The structure energy is the energy coming from the ligand intramolecular forces. The MM2 force field that we used to compute this term is coming from the Tinker<sup>27</sup> code originally written in Fortran but rewritten in Java to be totally integrated with our application. It is the sum of all the bonds, angles, stretch–bend, torsion, and VDW (Lennard Jones) terms.

$$\text{structure Energy} = \sum F_{\text{bonds}} + \sum F_{\text{angles}} + \sum F_{\text{stretch\_bend}} + \sum F_{\text{torsions}} + \sum F_{\text{VDW\_LennardJones}} \quad (5)$$

The ActDock's program uses two scoring functions, one for finding the optimal ligand pose and one for estimating the ligand's activity.

$$\text{position Score} = \text{binding Energy} + \text{structure Energy}$$

$$\text{activity Score} = \text{binding Energy}$$

A docking procedure starts by selecting the cavity where the docking should be performed. If a ligand is present, it is used to define the cavity. The cavity is filled with probes to estimate the shape and size of the docking site. The protein is cropped around these probes to keep only the atoms of interest for the interactions (i.e., all atoms that are more than 7 Å away from a probe are deleted). Working with a cropped part of the protein makes the whole process faster and helps the genetic algorithm to converge. Apart from the cavity

selection, there is no preparation to do; the program does not need to know the atom protonation state, the dielectric constant is always set to 10, the bonding information is automatically guessed from the PDB file, and water molecules are ignored. During the docking process, 10 reasonable ligand conformations are generated to start with a random population of 500 positions. After a local optimization the 200 most diverse positions are selected and undergo mutations and crossovers. During the optimization phase the bonds and angles are considered to be rigid to reduce the degrees of freedom; only  $6 + n$  degrees of freedom are considered for the rotation, translation, and the torsions for the  $n$  rotatable bonds. A local optimization is performed after each generation; positions that are either too high in energy, outside of the protein, or too similar are removed. When the genetic algorithm cannot improve the leading positions, a more precise MM2 optimization (including the bond lengths, stretch–bend angles, and angles) is performed on the 10 top positions in order to have a refined score. The final position is the one that minimizes the *positionScore*. ActDock can be used for two different purposes: finding the binding mode of a ligand to a protein or to enrich a library. The *positionScore* is the scoring function used to find the optimal binding mode as it tries to minimize the sum of the binding energy coming from the interactions and the structural energy coming from the internal ligand stress. However, the structure energy (or free energy) cannot be reliably used to compare two different classes of ligands, so the *activityScore* depends only on the binding Energy.

The binding pose generated by ActDock was validated on diverse test sets from literature. While it is nearly impossible to mimic the exact same docking procedure from literature, it is important to have consistent results across all test sets and that the docking program is not biased toward a specific test set, where shape matching, lipophilicity, or hydrophobicity would be enough to explain the results.

ActDock was tested on the data set from Fischer for the FlexScreen docking program, which was also tested on Glide, Dock, and FlexX.<sup>28</sup> The ActDock results show that it performs better (pose within 2 rmsd) for 7 complexes and worse for 8 complexes. The pose was found correctly 79% of the test cases within 2 rmsd, compared to 81% with FlexScreen, 71% with Glide, and 76% with Gold (Table 2). The DUD data set was used to estimate the enrichment rate that can be obtained for different classes of proteins. Each molecule from the data set was docked on the Actelion computing grid. The docking procedure was run in fully automated mode. The natural ligand's cavity was used as the starting position, which means that docking was impossible for ADA and HIVRT because they did not contain a natural ligand. The docking on the Actelion Grid took 20 to 120 s per complex per PC. The molecules were ranked according to their *activityScore*.

**Data.** The DUD was originally developed to assess the performance of structure based virtual screening.<sup>8</sup> It contains 40 different target proteins and almost 3000 ligands. For each ligand 36 decoy molecules were selected. Each of the 36 decoys selected had the same “druglike” features as the ligand. Molecules that were topologically similar to the ligand were not taken as decoys to avoid false positive classification results in the virtual screening. This data preparation gives an advantage to the ligand based methods,

**Table 4.** Calculation Time in Milliseconds Needed for Descriptor Generation and Comparison<sup>a</sup>

method	generation time/ molecule [ms]	no. of comparisons	comp. time [ms]	factor
FragFp	2.77	1.00E+06	438	1
PathFp	0.814	1.00E+06	438	1
SphereFp	3.927	1.00E+06	970	2
TopPPHist	25442	1.00E+06	2331	5
Flexophore	4310	10000	6039	1378
ActDock	----	1	70000	1.6E+08

<sup>a</sup> This was measured on an Intel Dual Core 2.13 GHz processor with 2 GB of RAM. Column 2: descriptor generation time per molecule. Column 3: number of descriptor comparisons. Column 4: time needed for the comparisons. Column 5: Relative time needed for descriptor comparisons. Time needed for comparison of two FragFp descriptors was set to 1.

**Table 5.** Relative Enrichment Rates in Percent for the Comparison of Ligand-Based and Structure-Based Virtual Screening with Setup<sub>crystal</sub><sup>a</sup>

ClassName	DOCK	ActDock	ChemAxCFp	PathFp	SphereFp	TopPPHist	FragFp	Flexophore
ACE	16.25	27.09	59.59	65.01	43.34	65.01	81.26	75.84
ACHe	10.01	7.51	55.04	57.54	82.56	57.54	65.05	55.04
ADA	31.55	0	42.06	52.58	31.55	0	84.12	0
ALR2	39.18	9.79	19.59	9.79	19.59	9.79	9.79	9.79
AR	40.98	3.42	47.81	40.98	61.48	44.4	44.4	27.32
AmpC	0	0	99.13	99.13	99.13	74.35	74.35	86.74
CDK2	18.81	0	37.63	37.63	42.33	23.52	9.41	23.52
COMT	0	0	0	0	0	20.88	0	20.88
COX-1	21.39	0	32.09	42.78	53.48	21.39	42.78	64.17
COX-2	46.19	44.73	66.73	60.86	64.53	69.66	85.79	96.79
DHFR	60.68	10.5	80.52	89.86	75.85	96.86	10.5	91.03
EGFr	9.12	4.87	72.99	90.02	95.49	70.56	9.12	73.6
ERagonist	15.22	26.64	57.08	57.08	68.49	45.66	15.22	41.86
ERantagonist	26.9	13.45	73.97	53.8	67.25	47.07	67.25	26.9
FGFr1	0	2.15	0	0	6.46	0	8.61	12.92
FXa	13.59	25.48	8.49	10.19	16.99	5.1	6.79	23.78
GART	10.99	21.98	10.99	10.99	21.98	10.99	10.99	10.99
GPB	9.12	0	59.31	72.99	18.25	72.99	68.43	36.5
GR	43.02	16.55	26.47	23.16	29.78	26.47	26.47	26.47
HIVPR	0	52.61	0	0	4.78	4.78	57.39	4.78
HIVRT	12.83	0	51.31	44.9	51.31	38.49	38.49	38.49
HMGR	79.26	72.66	99.08	99.08	99.08	66.05	99.08	99.08
HSP90	19.92	9.96	99.6	99.6	99.6	99.6	89.64	69.72
InhA	0	8.95	98.45	98.45	98.45	86.52	98.45	95.47
MR	66.67	33.33	0	0	0	0	0	0
NA	26	10.4	41.6	52	67.6	31.2	83.2	20.8
P38 MAP	3.18	2.12	27.57	31.81	27.57	0	26.51	22.27
PARP	14.43	50.51	7.22	7.22	21.65	7.22	14.43	14.43
PDE5	9.74	24.35	92.55	63.32	92.55	58.45	58.45	58.45
PNP	46.9	75.05	93.81	93.81	93.81	28.14	46.9	93.81
PPARg	0	84.16	56.11	87.28	74.81	87.28	96.63	96.63
PR	0	0	0	0	9.45	0	0	9.45
RXRa	51.95	77.92	38.96	51.95	51.95	12.99	38.96	77.92
SAHH	21.75	36.26	87.02	65.26	65.26	58.01	21.75	36.26
SRC	0	1.96	0	0	5.88	1.96	0	0
TK	0	0	54.76	65.72	32.86	32.86	43.81	54.76
trypsin	0	5.95	0	0	0	5.95	0	5.95
<b>median</b>	14.43	9.96	47.81	52	51.31	31.2	38.96	36.26
<b>mean</b>	20.69	20.55	45.87919	46.89	48.52	37.34	41.46	43.30838
<b>sdv</b>	20.73	24.36	33.26	33.59154	32.66	30.56	32.76	32.37754
<b>ranking median</b>	7	8	3	1	2	6	4	5
<b>ranking mean</b>	7	8	3	2	1	6	5	4

<sup>a</sup> The last five rows in the table summarize the results with the median, arithmetic mean, and the standard deviation of the mean. The last two rows show the ranking for the median and for the mean. The last column contains the results for the docking procedure.

compared to the structure based virtual screening with docking. The resulting bias has to be considered when this data set is used to compare ligand-based and structure-based methods. The 40 protein targets given in the publication<sup>8</sup> were taken from PDB.<sup>25</sup> Three corrections were necessary. According to information in the PDB, the file with the crystal structure 1xq2 was replaced with 2ao6 (AR). According to information on the DUD Web site,<sup>29</sup> the structure first was replaced with the structure 1ndw (ADA), and structure 1p4g was replaced with structure 1q4g (COX-1). The target PDGFRb kinase was not used in this study, because it was a modeled structure with no crystal structure available. From the files with the remaining 39 crystal structures the ligands were extracted. All ligands with a molecular weight below 150 or above 800 were discarded; 50 molecules remained. More than one ligand was obtained for nine targets (number of ligands in parentheses): COMT (2), COX-1 (3), DHFR (2), GART (2), GR (4), HMGR (3), InhA (2), MR (2), and SAHH (2). With the DUD three different data set setups were implemented:

**Setup<sub>crystal</sub>.** This setup enabled the direct comparison of ligand- and structure-based virtual screening. For the ligand based virtual screening the ligand extracted from the protein was taken as query. For 37 targets one ligand was selected; docking for two targets was not feasible (Chart 1). A test data set was all the ligands for the target as given by the DUD together with the designated decoys. In the docking approach, the test molecules were docked into the target protein and sorted according their docking score. For the ligand based virtual screening, the descriptors for the query molecule and the molecules in the DUD data set were calculated. The similarity values between the query descriptor and the test data descriptors were calculated, and the test molecules were sorted according to the similarity values. For all setups, the relative enrichment in the first 1% of the sorted test data set was used as figure of merit. The details are described in the following paragraphs.

**Setup<sub>random</sub>** and **Setup<sub>dissim</sub>** were used to explore the capability of the different descriptors (Figure 2). For each target 22 data subsets were created, with 11 subsets for



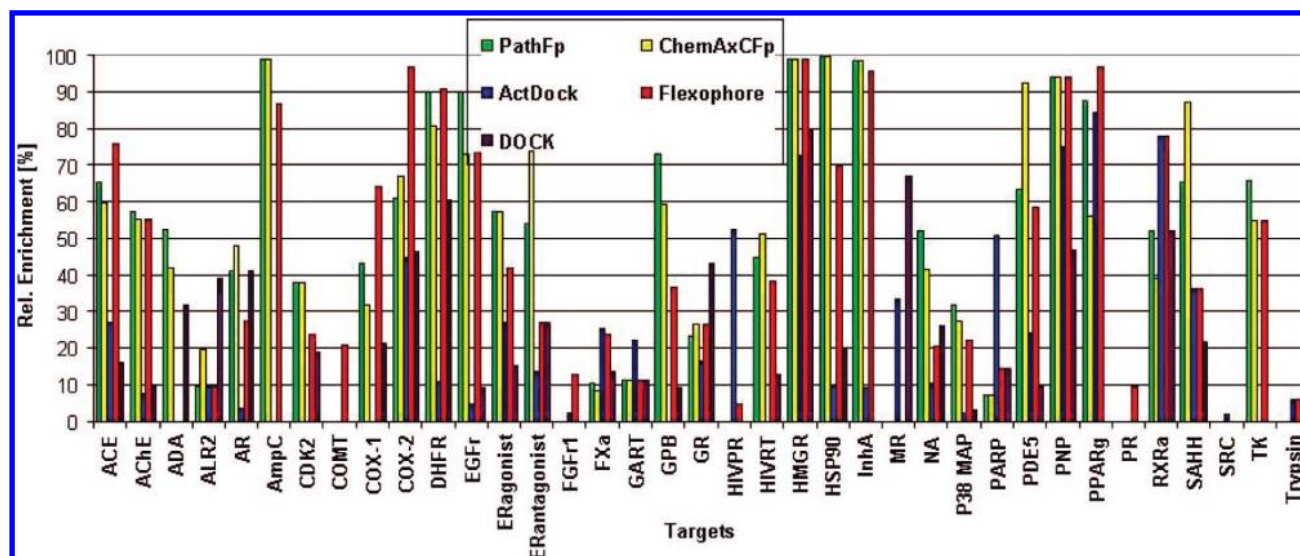
**Table 6.** Number of Clusters Found for a Certain Target and Virtual Screening Method<sup>a</sup>

target	DOCK	Act Dock	Chem AxCFp	PathFp	SphereFp	TopPP Hist	Frag Fp	Flexophore	no. of clusters
ACE	3	4	7	6	5	8	9	8	18
AChE	0	2	3	4	7	4	3	3	18
ADA	2	0	3	3	2	0	3	0	8
ALR2	2	1	1	1	1	1	1	1	14
AmpC	0	0	1	1	1	2	1	1	6
AR	3	0	1	1	1	1	1	1	10
CDK2	4	0	4	4	5	4	2	3	32
COMT	0	0	0	0	0	1	0	1	2
COX-1	2	0	2	2	3	1	3	4	11
COX-2	6	6	5	4	5	9	3	7	44
DHFR	4	14	6	7	8	4	2	9	14
EGFr	5	3	16	17	20	2	3	12	40
ERago	2	2	4	4	4	2	1	5	10
ERanta	4	0	4	3	4	2	3	2	8
FGFr1	0	0	0	0	2	0	0	0	12
FXa	4	3	1	2	3	1	1	4	19
GART	1	0	0	0	1	0	0	0	5
GPB	2	0	4	4	2	3	5	3	10
GR	4	1	3	2	4	1	3	3	9
HIVPR	0	0	0	0	0	0	1	0	3
HIVRT	1	0	3	3	3	2	2	3	17
HMGR	2	4	2	2	2	1	2	2	4
HSP90	1	0	2	2	2	2	2	2	4
InhA	0	0	5	6	4	5	4	6	23
MR	1	1	0	0	0	0	0	0	2
NA	1	1	3	2	1	1	3	1	7
P38MAP	0	1	1	1	1	0	1	1	20
PARP	1	2	1	1	2	1	2	2	7
PDE5	1	0	2	1	2	2	2	1	22
PNP	2	3	2	2	3	2	3	4	4
PPARg	0	0	1	0	1	0	0	0	6
PR	0	0	0	0	1	0	0	1	4
RXRa	1	3	1	1	1	1	1	1	3
SAHH	1	1	1	1	1	1	1	1	2
SRC	0	1	0	0	1	1	0	0	21
TK	0	0	1	1	2	1	1	2	7
trypsin	0	0	0	0	0	0	0	0	7
sum column	60	53	90	88	105	66	69	94	453

<sup>a</sup> The last row is the sum of all clusters in the column. The last column is the total number of clusters in the ligand set.

Setup<sub>random</sub> and for Setup<sub>dissim</sub>, respectively. As shown in Figure 2, a pair of Setup<sub>random</sub> and Setup<sub>dissim</sub> was always created from the same query molecule. The difference in the data set generation was the selection of test ligands. For

the random approach no bias was introduced for the selection of the ligands. For the dissim approach the ligands with the maximum distance in chemical space to the query molecule were taken as test molecules. The complex problem of

**Figure 3.** Results for Setup<sub>crystal</sub>.



**Table 7.** Number of Unique Clusters Found by Pairwise Method Combination

method	ActDock	ChemAxCFp	DOCK	Flexophore	FragFp	PathFp	SphereFp	TopPPHist
ActDock	0	117	95	116	103	116	130	99
ChemAxCFp	117	0	124	121	101	103	120	112
DOCK	95	124	0	123	105	126	139	103
Flexophore	116	121	123	0	111	119	133	108
FragFp	103	101	105	111	0	104	121	91
PathFp	116	103	126	119	104	0	121	109
SphereFp	130	120	139	133	121	121	0	127
TopPPHist	20	44	23	52	44	45	44	0

**Table 8.** Number of Overlapping Clusters Found by Pairwise Method Combination

method	ActDock	ChemAxCFp	DOCK	Flexophore	FragFp	PathFp	SphereFp	TopPPHist
ActDock	0	26	18	31	19	25	28	20
ChemAxCFp	26	0	26	63	58	75	75	44
DOCK	18	26	0	31	24	22	26	23
Flexophore	31	63	31	0	52	63	66	52
FragFp	19	58	24	52	0	53	53	44
PathFp	25	75	22	63	53	0	72	45
SphereFp	28	75	26	66	53	72	0	44
TopPPHist	20	44	23	52	44	45	44	0

determining chemical similarity was tackled by using three different chemical fingerprint descriptors. The dissim approach has the advantage that it avoids overoptimistic results in virtual screening. Ligands for a target are often developed as a series of chemical analogs to a certain hit/lead molecule. If a molecule from such a series became a query molecule it would easily enrich its analogs. The rest of this paragraph describes the algorithm for Setup<sub>random</sub> and Setup<sub>dissim</sub> in detail. The 11 most dissimilar ligands were selected as queries. A dissimilarity criterion served the Tanimoto coefficient of the ActelionFp descriptor, which is described in one of the following paragraphs. For each query molecule two data subsets were generated, Setup<sub>random,i</sub> and Setup<sub>dissim,i</sub>,  $i = 1$  to 11. To create a Setup<sub>random,i</sub> subset, five molecules were taken by random to form the ligand data set, except for the query molecule. The algorithm for the generation of the Setup<sub>dissim,i</sub> selected molecules from the ligand data set which were chemically dissimilar to the query molecule. The chemical fingerprint descriptors FragFP, PathFp, and ChemAxCFp were calculated for the ligands and query molecules. The similarity matrix between the query descriptors and the ligand descriptors was calculated with the Tanimoto coefficient. The similarity values were replaced by the rank in the matrix, whereas for each descriptor type the rank was determined separately. For each ligand the minimum rank was taken as final rank value  $R_{Fin}$ . The five ligands with the highest  $R_{Fin}$  values were taken as test subset. The selection scheme is explained with an example in Table 3. The Setup<sub>random,i</sub> and the Setup<sub>dissim,i</sub> data set were filled with up to 1000 molecules with randomly selected decoys. The total number of data sets for Setup<sub>random</sub> and Setup<sub>dissim</sub> was 858.

**Enrichment.** A relative enrichment factor  $e_{rel}$  at 1.0% of the total data set ( $frac = 0.01$ ) was calculated in order to compare enrichment values of data sets with different numbers of active molecules. Higher fractions are often applied for calculating enrichment values. However, the authors considered that values from 0.001 to 0.01 reflect the reality of high-throughput screening, where rates of true-positives are frequently below 0.001. The relative enrichment

$e_{rel}$  is the achieved enrichment  $e_{abs}$  normalized by the maximum possible enrichment  $e_{abs,max}$ .

$$e_{abs} = n_{hits, fraction} / (n_{active} * frac) \quad (6)$$

$$n_{max \text{ hits, fraction}} = \text{minimum}(n * frac, n_{active}) \quad (7)$$

$$e_{abs,max} = n_{max \text{ hits, fraction}} / (n_{active} * frac) \quad (8)$$

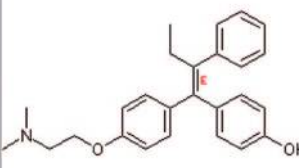
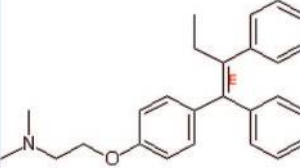
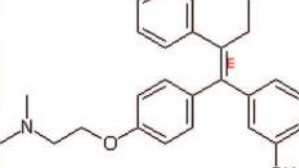
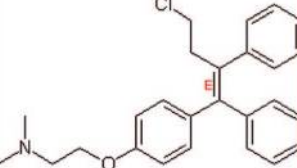
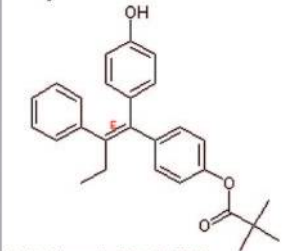
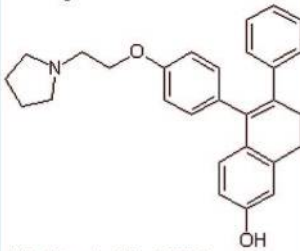
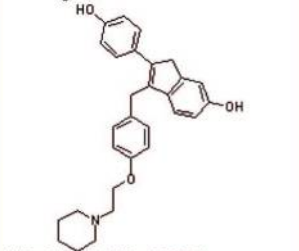
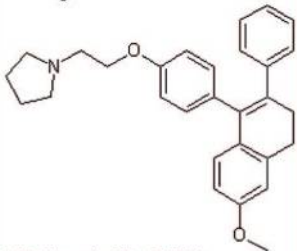
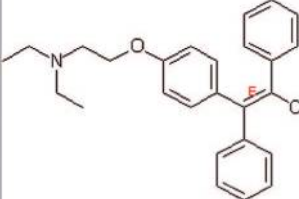
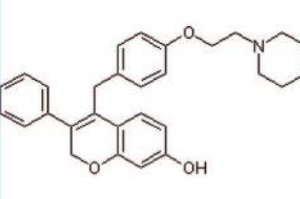
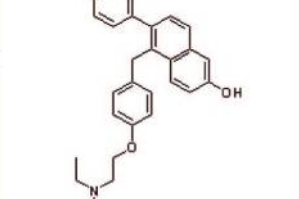
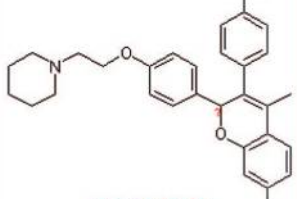
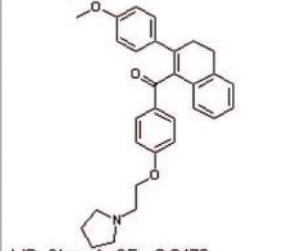
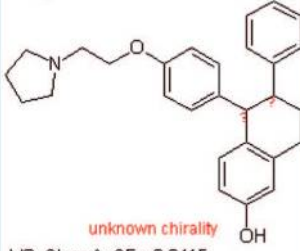
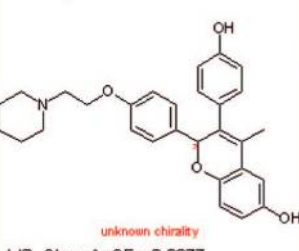
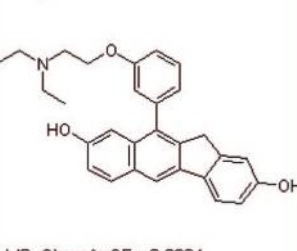
$$e_{rel} = 100 * e_{abs} / e_{abs,max} = 100 * n_{hits, fraction} / n_{max \text{ hits, fraction}} \quad (9)$$

To compare two descriptors, the binary Tanimoto coefficient was used for the chemical fingerprints PathFp, ChemAxCFp, FragFp, and SphereFp. The Tanimoto coefficient was used for the integer vector based TopPPHist descriptor. The more sophisticated metric for the Flexophore descriptor resulted also in values between 0 for dissimilar descriptor pairs and 1 for similar ones. The docking energies from ActDock and from the original DUD publication were treated like the similarity values from the descriptor comparisons. The energy values were sorted in ascending order, and the relative enrichment rates were calculated.

**Orthogonality of Descriptors.** A correlation analysis was done to analyze the relationships between the descriptors. A subset of 300 molecules was selected by random from all ligands in the DUD data set. For this subset the similarity matrices for all six descriptors were calculated. A similarity matrix (SimMatrixDUDSub) for a descriptor is calculated with the pairwise similarities for all ligand combinations. The calculation of one triangle of the matrix is sufficient because the calculation of the similarity values is symmetric. With  $l$  descriptors and  $n$  ligands are the total number of similarity calculations:  $l(n^2-n)/2$ .

The maximum similarity value for all six descriptor metrics is 1.0, which means that the descriptors are identical in the metric space. The minimum similarity is 0.0. The upper triangle of a similarity matrix is written into one column. The columns calculated from the similarity matrices for different descriptors were used to calculate the Pearson's correlation coefficient. An additional similarity matrix (SimMatrixGPCRSub) was calculated from an in-house data set of diverse GPCR ligands.<sup>9</sup> The reason for using a similarity

**Chart 2.** Results of Virtual Screening with Setup<sub>Crystal</sub> Sorted According to Scores for ChemAxCFp

 VS_ChemAxCFp: 0.9999 VS_Energy: 0.21140 VS_TopPPHist: 0.9999 VS_Flexophore: 0.9639	 VS_ChemAxCFp: 0.9999 VS_Energy: 0.18260 VS_TopPPHist: 0.9750 VS_Flexophore: 0.9262	 VS_ChemAxCFp: 0.9548 VS_Energy: 0.45000 VS_TopPPHist: 0.9921 VS_Flexophore: 0.5010	 VS_ChemAxCFp: 0.8758 VS_Energy: 0.45000 VS_TopPPHist: 0.9748 VS_Flexophore: 0.8633
 VS_ChemAxCFp: 0.7808 VS_Energy: 0.19270 VS_TopPPHist: 0.8644 VS_Flexophore: 0.8081	 VS_ChemAxCFp: 0.7696 VS_Energy: 0.40230 VS_TopPPHist: 0.9996 VS_Flexophore: 0.6462	 VS_ChemAxCFp: 0.7621 VS_Energy: 0.39900 VS_TopPPHist: 0.8995 VS_Flexophore: 0.2503	 VS_ChemAxCFp: 0.7604 VS_Energy: 0.23600 VS_TopPPHist: 0.9674 VS_Flexophore: 0.4336
 VS_ChemAxCFp: 0.7337 VS_Energy: -0.00770 VS_TopPPHist: 0.9747 VS_Flexophore: 0.5674	 VS_ChemAxCFp: 0.6756 VS_Energy: 0.43620 VS_TopPPHist: 0.9189 VS_Flexophore: 0.2867	 VS_ChemAxCFp: 0.6713 VS_Energy: 0.12850 VS_TopPPHist: 0.8447 VS_Flexophore: 0.2234	 unknown chirality VS_ChemAxCFp: 0.6509 VS_Energy: 0.22950 VS_TopPPHist: 0.9258 VS_Flexophore: 0.1596
 VS_ChemAxCFp: 0.6476 VS_Energy: 0.21230 VS_TopPPHist: 0.8915 VS_Flexophore: -0.0001	 unknown chirality VS_ChemAxCFp: 0.6415 VS_Energy: 0.45800 VS_TopPPHist: 0.9997 VS_Flexophore: 0.5994	 unknown chirality VS_ChemAxCFp: 0.6377 VS_Energy: 0.23370 VS_TopPPHist: 0.9267 VS_Flexophore: 0.1371	 VS_ChemAxCFp: 0.6364 VS_Energy: 0.06980 VS_TopPPHist: 0.8037 VS_Flexophore: 0.0000

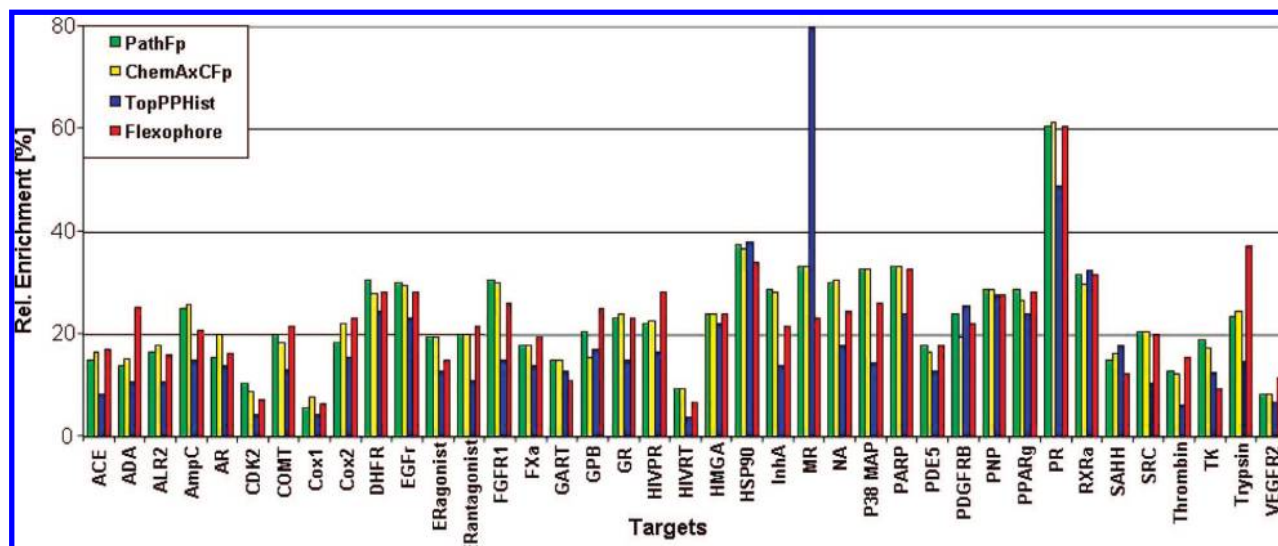
matrix of GPCR ligands was the higher acceptance of GPCR's for chemically diverse ligands than the targets represented in the DUD.

**Computational Details.** With the exception of the implementation of the PathFp and the ChemAxonFp descriptor, all descriptors and their similarity metrics were implemented by the authors. Java 5 was used for this implementation. The processing of 858 test data sets from Setup<sub>random</sub> and Setup<sub>dissim</sub> took approximately 41 min on a workstation with an Intel Dual Core 2.13 GHz processor, equivalent to 350 similarity calculations per second. While the comparison of Flexophore descriptors is a relatively swift process, its generation is rather time-consuming due to the necessity of generating a diverse set of conformers. This takes up to ten

seconds per compound. For this reason, the descriptor generation was distributed to multiple computers on Actelion's computing grid of about 200 computers. For searching commercial libraries of screening compounds with 3 million structures, the similarity calculations are also distributed onto the grid. In Table 4 the calculation times for descriptor generation and the time needed for descriptor comparison and docking are given.

## RESULTS

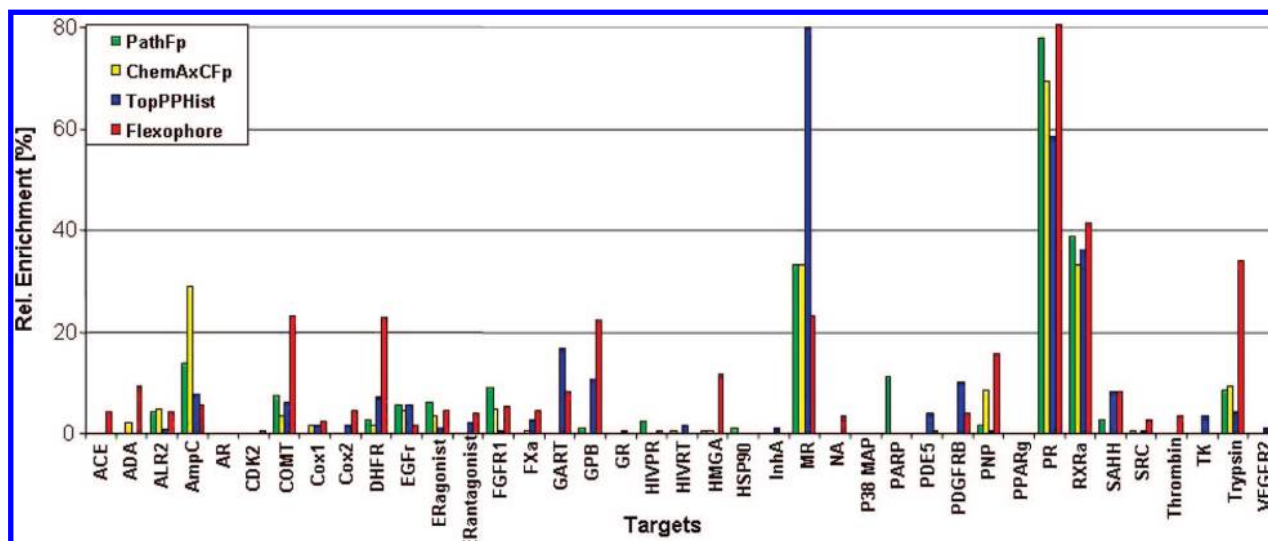
Table 5 and Figure 3 show the results of the virtual screening as relative enrichment rates for Setup<sub>crystal</sub>. The docking results from the original DUD publication were downloaded as energies<sup>29</sup> and recalculated as relative enrich-

Figure 4. Results for Setup<sub>random</sub>.Table 9. Relative Enrichment Rates in Percent for the Descriptor Comparison with Setup<sub>random</sub>

target	TopPPHist	SphereFp	FragFp	Flexophore	PathFp	ChemAxCFp
ACE	8.33	16.67	19.44	17.22	15.00	16.67
ADA	10.87	23.91	21.74	25.36	13.77	15.22
ALR2	10.67	12.67	14.67	16.00	16.67	18.00
AmpC	15.00	22.50	20.83	20.83	25.00	25.83
AR	13.89	19.44	18.89	16.11	15.56	20.00
CDK2	4.44	6.11	9.44	7.22	10.56	8.89
COMT	13.33	20.00	15.00	21.67	20.00	18.33
Cox1	4.35	7.97	8.70	6.52	5.80	7.97
Cox2	15.56	23.89	14.44	23.33	18.33	22.22
DHFR	24.44	22.78	19.44	28.33	30.56	27.78
EGFR	23.33	32.22	21.67	28.33	30.00	29.44
ERagonist	12.78	21.11	17.78	15.00	19.44	19.44
ERantagonist	11.11	26.67	23.89	21.67	20.00	20.00
FGFR1	15.00	26.11	29.44	26.11	30.56	30.00
FXa	13.89	21.11	18.33	19.44	17.78	17.78
GART	12.78	16.11	11.67	11.11	15.00	15.00
GPB	17.22	25.00	22.22	25.00	20.56	15.56
GR	15.00	23.33	20.56	23.33	23.33	23.89
HIVPR	16.67	26.67	30.00	28.33	22.22	22.78
HIVRT	3.89	7.78	5.56	6.67	9.44	9.44
HMGA	22.22	24.44	22.78	23.89	23.89	23.89
HSP90	38.19	34.03	34.03	34.03	37.50	36.81
InhA	13.89	27.22	22.78	21.67	28.89	28.33
MR	80.00	33.33	36.67	23.33	33.33	33.33
NA	17.78	30.00	30.56	24.44	30.00	30.56
P38 MAP	14.44	32.78	17.78	26.11	32.78	32.78
PARP	23.89	33.33	33.33	32.78	33.33	33.33
PDE5	12.78	17.78	9.44	17.78	17.78	16.67
PDGFRB	25.56	15.56	17.22	22.22	23.89	19.44
PNP	27.59	29.31	25.29	27.59	28.74	28.74
PPARg	23.89	31.67	30.56	28.33	28.89	26.67
PR	48.96	60.42	60.42	60.42	60.42	61.46
RXRa	32.46	34.21	30.70	31.58	31.58	29.82
SAHH	17.78	16.67	19.44	12.22	15.00	16.11
SRC	10.56	21.11	17.22	20.00	20.56	20.56
thrombin	6.11	13.33	10.56	15.56	12.78	12.22
TK	12.70	15.87	12.70	9.52	19.05	17.46
trypsin	14.71	27.45	15.69	37.25	23.53	24.51
VEGFR2	6.67	6.11	8.89	11.67	8.33	8.33
<b>median</b>	<b>14.71</b>	<b>23.33</b>	<b>19.44</b>	<b>22.22</b>	<b>20.56</b>	<b>20.56</b>
<b>mean</b>	<b>18.19</b>	<b>23.25</b>	<b>20.98</b>	<b>22.26</b>	<b>22.76</b>	<b>22.65</b>
<b>sdv</b>	<b>13.67</b>	<b>10.01</b>	<b>10.09</b>	<b>9.87</b>	<b>9.97</b>	<b>9.84</b>
<b>rank median</b>	<b>6</b>	<b>1</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>3</b>
<b>rank mean</b>	<b>6</b>	<b>1</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>3</b>

ment rates like the other values in this publication. The referring column in Table 5 is headed with "DOCK". In

Table 5 the last rows contain a summary of the individual values. The median and the arithmetic mean are given; both

Figure 5. Results for Setup<sub>Dissim</sub>.Table 10. Relative Enrichment Rates in Percent for the Descriptor Comparison with Setup 2, Setup<sub>dissim</sub> Data Sets

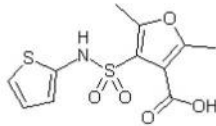
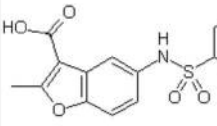
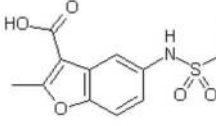
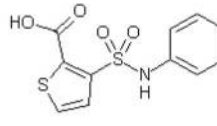
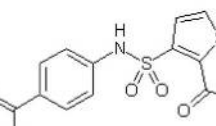
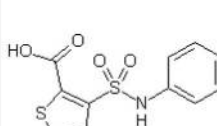
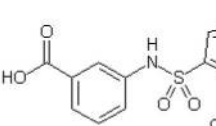
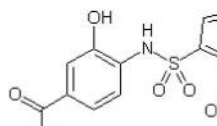
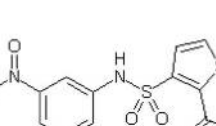
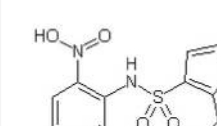
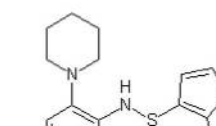
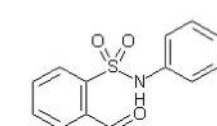
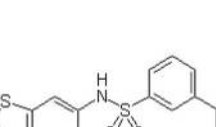
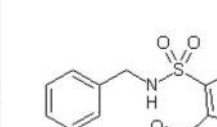
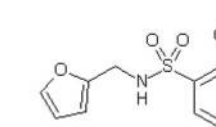
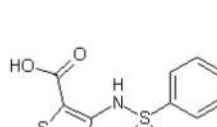
target	TopPPHist	SphereFp	FragFp	Flexophore	PathFp	ChemAxCFp
ACE	0.00	0.00	0.00	4.17	0.00	0.00
ADA	0.00	6.66	5.62	9.41	0.00	2.08
ALR2	0.88	0.00	2.38	4.22	4.05	5.01
AmpC	7.78	8.33	0.00	5.56	13.89	28.89
AR	0.00	0.00	0.00	0.00	0.00	0.00
CDK2	0.00	1.11	0.00	0.56	0.00	0.00
COMT	6.11	12.30	0.00	23.17	7.54	3.33
Cox1	1.55	4.65	3.72	2.27	0.00	1.59
Cox2	1.67	0.56	0.00	4.44	0.00	0.00
DHFR	7.22	5.00	0.00	22.78	2.78	1.67
EGFr	5.56	5.00	1.67	1.67	5.56	4.44
ERagonist	1.11	5.06	2.85	4.46	6.19	3.37
ERantagonist	2.07	4.11	0.67	3.94	0.00	0.00
FGFR1	0.56	3.11	0.00	5.22	9.00	4.67
FXa	2.78	11.67	0.00	4.44	0.00	0.56
GART	16.67	13.89	5.56	8.33	0.00	0.00
GPB	10.73	12.88	0.00	22.26	1.11	0.00
GR	0.56	1.94	0.00	0.00	0.00	0.00
HIVPR	0.00	0.00	0.64	0.64	2.38	0.00
HIVRT	1.67	0.56	0.00	0.00	0.56	0.00
HMGA	0.00	8.89	0.00	11.67	0.56	0.56
HSP90	0.00	1.11	0.00	0.00	1.11	0.00
InhA	1.11	1.11	0.00	0.00	0.00	0.00
MR	80.00	33.33	36.67	23.33	33.33	33.33
NA	0.00	0.00	0.00	3.33	0.00	0.00
P38 MAP	0.00	0.00	0.00	0.00	0.00	0.00
PARP	0.00	2.78	0.00	0.00	11.11	0.00
PDE5	3.89	1.67	0.00	0.56	0.00	0.00
PDGFRB	10.00	0.00	0.00	3.89	0.00	0.00
PNP	0.60	0.60	0.00	15.84	1.85	8.40
PPARg	0.00	2.08	0.00	0.00	0.00	0.00
PR	58.33	66.67	80.56	80.56	77.78	69.44
RXRa	36.11	52.78	35.19	41.67	38.89	33.33
SAHH	8.33	2.78	2.78	8.33	2.78	0.00
SRC	0.56	0.00	0.00	2.78	0.56	0.00
thrombin	0.00	2.22	0.00	3.33	0.00	0.00
TK	3.33	0.00	0.00	0.00	0.00	0.00
trypsin	4.24	12.53	1.96	34.11	8.55	9.53
VEGFR2	1.11	0.56	0.00	0.56	0.00	0.00
median	1.11	2.22	0.00	3.94	0.56	0.00
mean	6.89	7.20	4.51	9.04	5.75	5.26
sdv	16.21	13.96	14.82	15.43	14.48	13.58
rank median	3	2	5	1	4	5
rank mean	3	2	6	1	4	5

values measure the relative enrichment rate at 1% of the data set. The next row shows the variance of the mean, and the

last two rows give the ranking of the median and the mean. According to the median the PathFp was the best descriptor



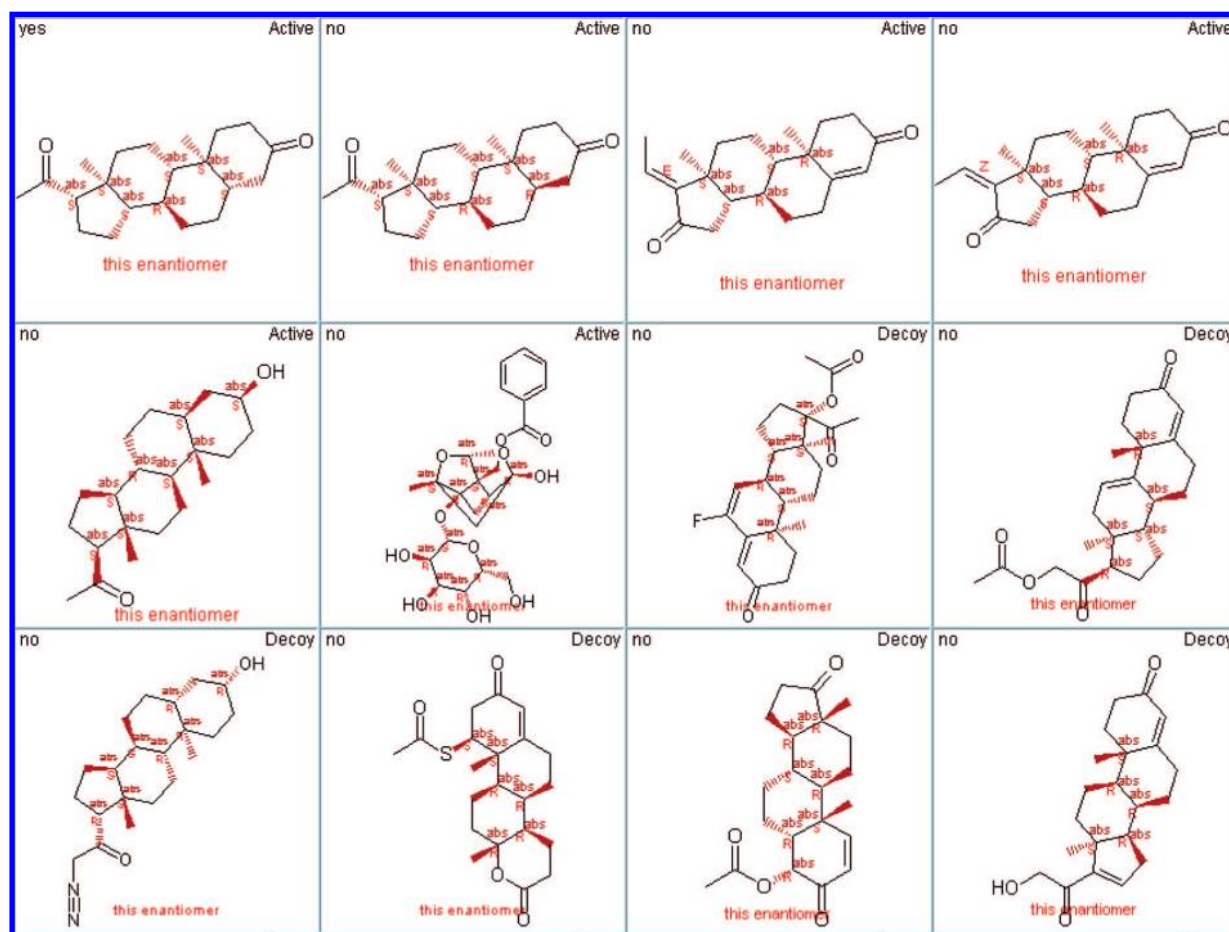
**Chart 3.** Query (First Molecule), Ligands, and Some Decoys from the AmpC Data Set for Setup<sub>Dissim</sub>

 <p>Active</p> <p>ChemAxCFpActive0: 7S_PP3DMM2Active0: VS_FragFpActive0:</p>	 <p>Decoy</p> <p>ChemAxCFpActive0: 0.6444 7S_PP3DMM2Active0: 0.5978 VS_FragFpActive0: 0.3881</p>	 <p>Decoy</p> <p>ChemAxCFpActive0: 0.6364 7S_PP3DMM2Active0: 0.6199 VS_FragFpActive0: 0.3920</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.6276 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4827</p>
 <p>Active</p> <p>ChemAxCFpActive0: 0.6100 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4969</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.6075 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4515</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.6070 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.5059</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.6062 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4397</p>
 <p>Active</p> <p>ChemAxCFpActive0: 0.5925 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.5152</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.5919 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4969</p>	 <p>Decoy</p> <p>ChemAxCFpActive0: 0.5893 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.4576</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.5605 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.3106</p>
 <p>Decoy</p> <p>ChemAxCFpActive0: 0.5589 7S_PP3DMM2Active0: 0.0000 VS_FragFpActive0: 0.2872</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.5558 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.5761</p>	 <p>Decoy</p> <p>ChemAxCFpActive0: 0.5502 7S_PP3DMM2Active0: 0.0000 VS_FragFpActive0: 0.5053</p>	 <p>Active</p> <p>ChemAxCFpActive0: 0.5450 7S_PP3DMM2Active0: -0.0001 VS_FragFpActive0: 0.5029</p>

and reached an average relative enrichment rate of 46.9%, with the median at 51%. Taking into account that the median is a robust estimator, we have the following ranking: PathFp, SphereFp, ChemAxCFp, FragFp, Flexophore, TopPPHist, DOCK, and ActDock. The small inner-descriptor differences between the ranks of the median and of the mean indicate that the result is robust. The difference between the medians of the descriptors and the docking procedures is striking but not surprising. A design principle of the DUD data set was to avoid decoys being topologically similar to the active molecules, while many of the actives are structurally quite similar to the query. The decoys were selected according to similar physicochemical properties to the active molecules. This biased setup prefers chemical fingerprints in the virtual screening. For 12 targets all eight methods gave reasonable results ( $\geq 10\%$  relative enrichment): ACE, COX-2, DHFR, ERagonist, ERantagonist, GART, GR, HMGR, NA, PNP,

RXRa, and SAHH. With the median as figure of merit the PathFp performed the best, closely followed by the SphereFp. The performance of the Flexophore descriptor is below the performance of the chemical fingerprints and better than the performance of the docking procedure. From its design the Flexophore is not free of implicit topological information, the conformers depend on the topology of the molecule, and the PPs are based on the atom types. The performance of the topological pharmacophore descriptor TopPPHist is between that of the chemical fingerprints and the docking procedure.

To allow a more detailed analysis of the virtual screening results in the cluster analysis by Good<sup>4,30</sup> and Oprea were used to determine the cluster memberships of the enriched ligands (Table 6). The published cluster members were taken (\*\_cluster.sdf) together with the cluster identifiers, and the fractions of enriched ligands were searched for these

**Chart 4.** Query (First Molecule), Ligands, and Some Decoys from the MR Data Set for Setup<sub>Dissim</sub>


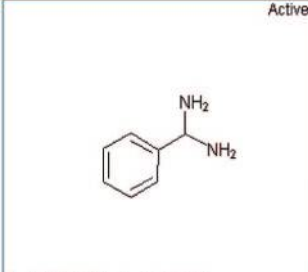
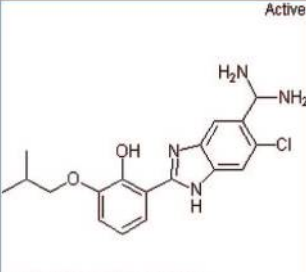
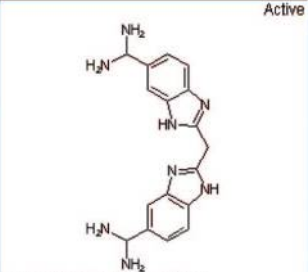
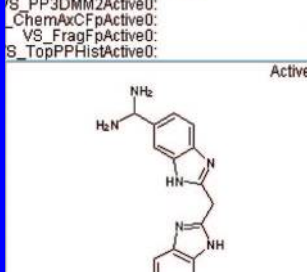
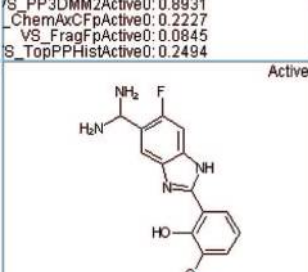
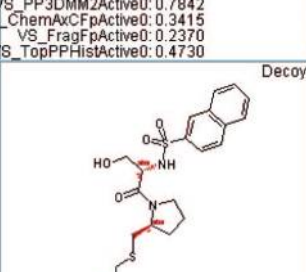
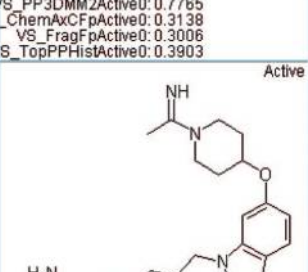
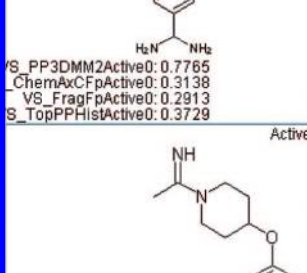
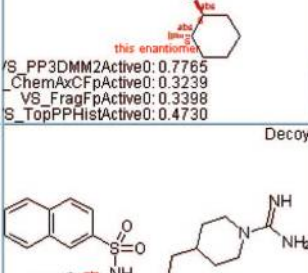
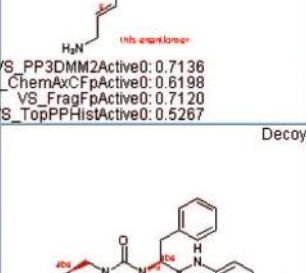
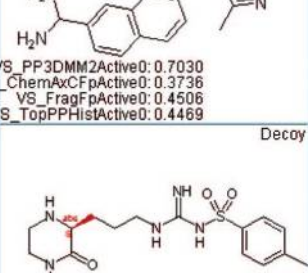
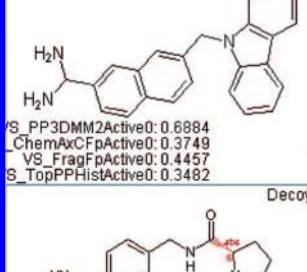
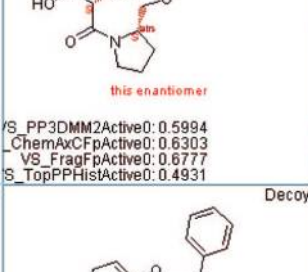
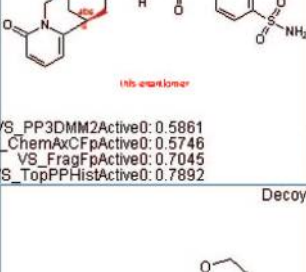
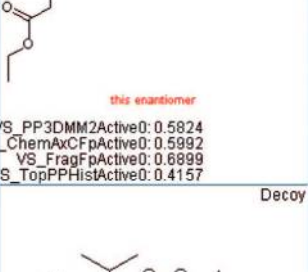
molecules. If at least one ligand was found with a cluster molecule this cluster was counted as present in the enriched fraction. This method is not exhaustive, because the cluster molecules are the lead-like filtered ligands of the DUD data set and do not represent all structures, but it produces a good overview of the diversity of the enriched molecules. Taking the total sum of enriched clusters (Table 6, last row), the best performing descriptor was the SphereFp in enriching 105 clusters. The ranking for the other descriptors in decreasing order of clusters found is as follows: Flexophore, ChemAxCFp, PathFp, FragFp, TopPPHist, DOCK, and ActDock. Remarkably, the number of enriched clusters is clearly lower for FragFp, TopPPHist, ActDock, and DOCK than for the other methods. Tables 7 and 8 contain the results for the pairwise method combinations. It is shown in Table 7 that the maximum number of unique clusters was reached for the combination of the methods DOCK and SphereFp. The second best combination was the combination of Flexophore and SphereFp. The largest overlap in cluster enrichment gave the combinations of ChemAxCFp and PathFp and the former one with SphereFp (Table 8). This result was expected for the combination of ChemAxCFp with PathFp; both descriptors are topological path based fingerprints. More surprising was that the same number of overlapping clusters was found by the combination of ChemAxCFp with SphereFp. Both descriptors rely on different algorithms, and no such correlation was observed in other data sets.

**Detailed Results for Six Representative Systems.** The same targets as in Huang et al. were chosen for comparison.

**Estrogen Receptor (ER<sub>antagonist</sub>).** All eight virtual screening methods delivered good enrichment results. The best performance showed the ChemAxCFp with a relative enrichment rate of 75%; the worst was the ActDock program with a 14% enrichment rate. In Chart 2 the first sixteen enriched molecules are shown. They are sorted according to the best performing ChemAxCFp descriptor. The molecules in row two to four demonstrate that this topological path based descriptor is able to enrich molecules, which are only partially chemically similar. The cluster analysis showed that the ChemAxCFp found four clusters, as many as DOCK and SphereFp did. All three methods found exactly the same clusters: 2, 6, 7, 8, and none of the other methods were able to detect another cluster.

**Thymidine Kinase (TK).** Both docking approaches failed in enrichment in the first one percent of the test data set. The ligand based methods showed good results between 66% relative enrichment for the PathFp and 33% for both the TopPPHist and the SphereFp. While TK is known to be a difficult target for docking, it is an easy one for ligand based methods. The query structure (Chart 1) is small and not very flexible. It contains multiple chemical substructures like alcohol, ether, and amide, which results in multiple entries in the chemical fingerprint vectors. The pharmacophore based descriptors TopPPHist and Flexophore also profit from these substructures. A small rigid molecule with multiple chemical substructures generates a well defined descriptor. The active ligands in the DUD database all belong to the same chemical class and have similar well-defined descriptors. The result is a

**Chart 5.** Query (First Molecule), Ligands, and Some Decoys from the TR Data Set for Setup<sub>Dissim</sub><sup>a</sup>

 <p>Active</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.7765 _ChemAxCFpActive0: 0.2227 _VS_FragFpActive0: 0.0845 S_TopPPHistActive0:</p>	 <p>Active</p> <p>/S_PP3DMM2Active0: 0.8931 _ChemAxCFpActive0: 0.2227 _VS_FragFpActive0: 0.0845 S_TopPPHistActive0: 0.2494</p>	 <p>Active</p> <p>/S_PP3DMM2Active0: 0.7842 _ChemAxCFpActive0: 0.3415 _VS_FragFpActive0: 0.2370 S_TopPPHistActive0: 0.4730</p>	 <p>Active</p> <p>/S_PP3DMM2Active0: 0.7765 _ChemAxCFpActive0: 0.3138 _VS_FragFpActive0: 0.3006 S_TopPPHistActive0: 0.3903</p>
 <p>Active</p> <p>/S_PP3DMM2Active0: 0.7765 _ChemAxCFpActive0: 0.3138 _VS_FragFpActive0: 0.2913 S_TopPPHistActive0: 0.3729</p>	 <p>Active</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.7765 _ChemAxCFpActive0: 0.3239 _VS_FragFpActive0: 0.3398 S_TopPPHistActive0: 0.4730</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.7136 _ChemAxCFpActive0: 0.6198 _VS_FragFpActive0: 0.7120 S_TopPPHistActive0: 0.5267</p>	 <p>Active</p> <p>/S_PP3DMM2Active0: 0.7030 _ChemAxCFpActive0: 0.3736 _VS_FragFpActive0: 0.4506 S_TopPPHistActive0: 0.4469</p>
 <p>Active</p> <p>/S_PP3DMM2Active0: 0.6884 _ChemAxCFpActive0: 0.3749 _VS_FragFpActive0: 0.4457 S_TopPPHistActive0: 0.3482</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5994 _ChemAxCFpActive0: 0.6303 _VS_FragFpActive0: 0.6777 S_TopPPHistActive0: 0.4931</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5861 _ChemAxCFpActive0: 0.5746 _VS_FragFpActive0: 0.7045 S_TopPPHistActive0: 0.7892</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5824 _ChemAxCFpActive0: 0.5992 _VS_FragFpActive0: 0.6899 S_TopPPHistActive0: 0.4157</p>
 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5797 _ChemAxCFpActive0: 0.5642 _VS_FragFpActive0: 0.6707 S_TopPPHistActive0: 0.6638</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5544 _ChemAxCFpActive0: 0.3887 _VS_FragFpActive0: 0.6467 S_TopPPHistActive0: 0.6879</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5281 _ChemAxCFpActive0: 0.4919 _VS_FragFpActive0: 0.6720 S_TopPPHistActive0: 0.3650</p>	 <p>Decoy</p> <p><i>this enantiomer</i></p> <p>/S_PP3DMM2Active0: 0.5041 _ChemAxCFpActive0: 0.6498 _VS_FragFpActive0: 0.5686 S_TopPPHistActive0: 0.4845</p>

<sup>a</sup> The molecules are sorted according their similarity score for the Flexophore (PP3DMM2) descriptor.

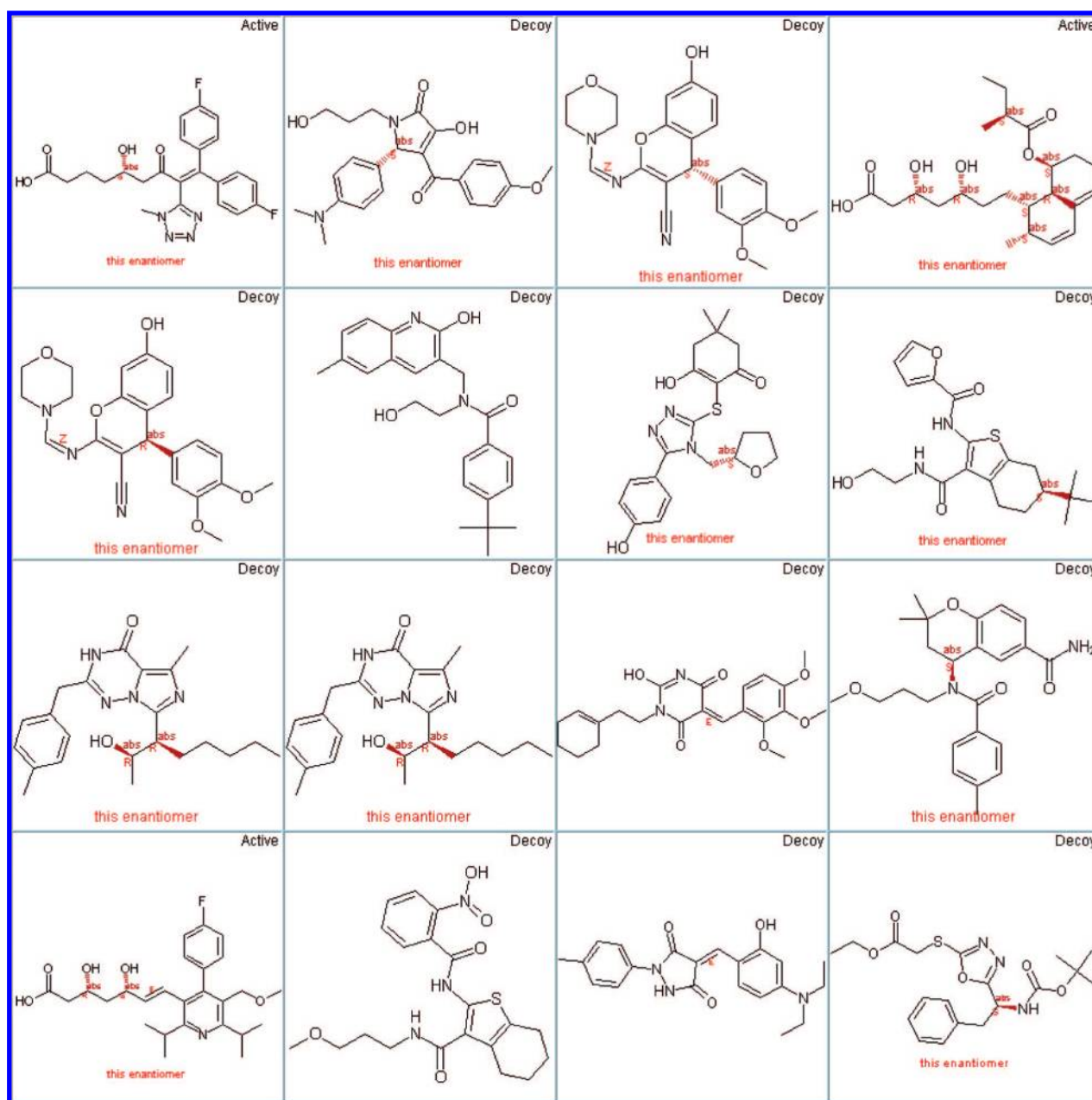
good enrichment in a data set of less similar decoys. As stated by Huang et al. the decoys are chemically dissimilar to the TK ligands.<sup>8</sup> Despite the good virtual screening results, only one of seven clusters was found by the PathFp. The Flexophore and the SphereFp performed better with detecting two clusters, despite their lower enrichment rate.

**P38 Mitogen Activated Protein Kinase (P38 MAP).** The poor results for the docking approaches and for the failed virtual screening with the TopPPHist descriptor showed that P38 MAP is a difficult target. The query molecule is large and flexible, nevertheless the majority of the ligand based methods succeeded. Their relative enrichment rates are located in a narrow range, between 32% for the PathFp and

22% for Flexophore. All descriptors enrich only ligands from the same structural class (N-pyrazole urea) as the query molecule. Other urea analogs (i.e., with an attached isoxazole) were not found by the descriptors. The cluster 2 was only detected by ActDock, ChemAxCFp, PathFp, SphereFp, FragFp, and Flexophore.

**Adenosine Deaminase (ADA).** ActDock, TopPPHist, and Flexophore failed on this metalloenzyme. The docking procedure applied by Huang et al. also failed, and it was only after manual refinement of the docking protocol that the enrichment was observed.<sup>8</sup> All four chemical fingerprints succeeded with the FragFp as best performing descriptor. An enrichment rate of 84% showed that this descriptor detected a pattern, which described the active principle. The



**Chart 6.** Query (First Molecule), Ligands, and Some Decoys from the HMGR Data Set for Setup<sub>Dissim</sub><sup>a</sup>

<sup>a</sup> The molecules are sorted according their similarity score for the Flexophore descriptor.

descriptors FragFp, ChemAxCFp, and PathFp detected three of the eight clusters. The high enrichment rate and the low number of detected clusters indicated a structural bias for the ligands in this data set.

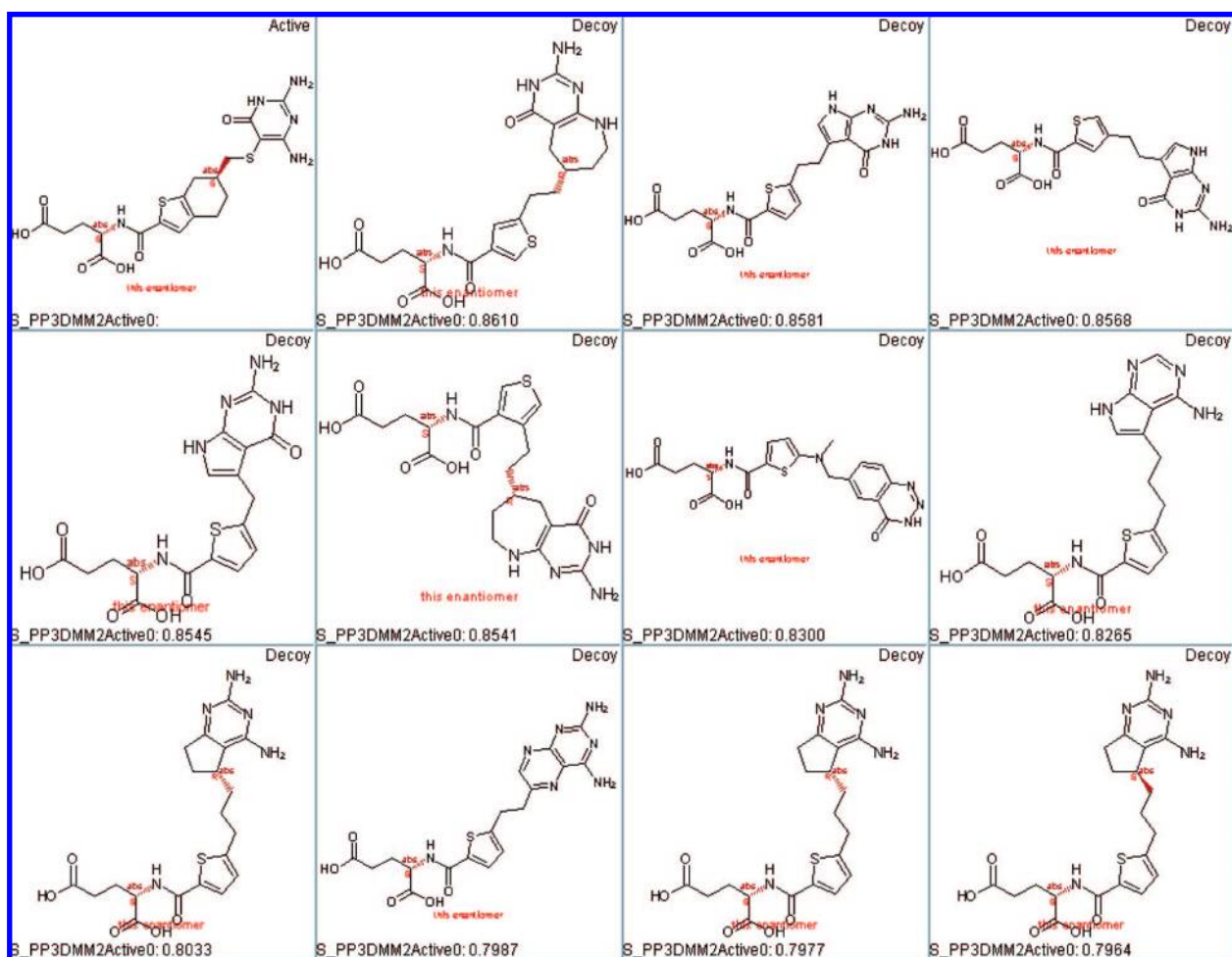
**Aldose Reductase (ALR2).** All virtual screening methods succeeded for ALR2. It is one of the three cases where the DOCK approach performed best with a relative enrichment of 39%. The enrichment fraction of ligands contained two clusters. ChemAxCFp and SphereFp performed second best with 20% enrichment. ActDock and the remaining descriptors reached a relative enrichment rate of 10%. For the descriptors this can be considered a failure, because the enriched structure was the query molecule. The most similar molecule in the complete data set, decoys and ligands, had a similarity of less than 0.6 for the three chemical fingerprints.

**Enoyl ACP Reductase (InhA).** For Huang et al. the worst virtual screening scenario was given by InhA.<sup>8</sup> In contrast, all methods employed in this publication succeeded, while

the Actelion docking tool ActDock performed worst with only 9% enrichment, in close analogy to DOCK. The ligand-based methods performed almost perfectly with a maximum relative enrichment of 98% for ChemAxCFp, PathFp, SphereFp, and FragFp. They were closely followed by Flexophore with a 96% relative enrichment rate and TopP-PHist with 87%. None of the decoys contained the indole-amide, which was a relevant substructural pattern for all descriptors. With six clusters in the fraction of enriched ligands Flexophore was the most successful method in the diversity ranking.

For Setup<sub>random</sub> and Setup<sub>dissim</sub> 11 molecules were chosen from each data set as query, as described in the “Data” section. A comparison of the ligand based and docking methods was not possible because of an insufficient number of crystal structures for the target proteins. For each query molecule a virtual screening was performed on the corresponding test data set with 1000 molecules, spiked with five



**Chart 7.** Query (First Molecule) and Decoys from the GART Data Set for Setup<sub>Dissim</sub><sup>a</sup>

<sup>a</sup> The molecules are sorted according to their similarity score for the Flexophore descriptor.

**Table 11.** Descriptor Correlation Matrix for All Mutual Compound Similarity Values of a Subset of 300 Bioactive Molecules<sup>a</sup>

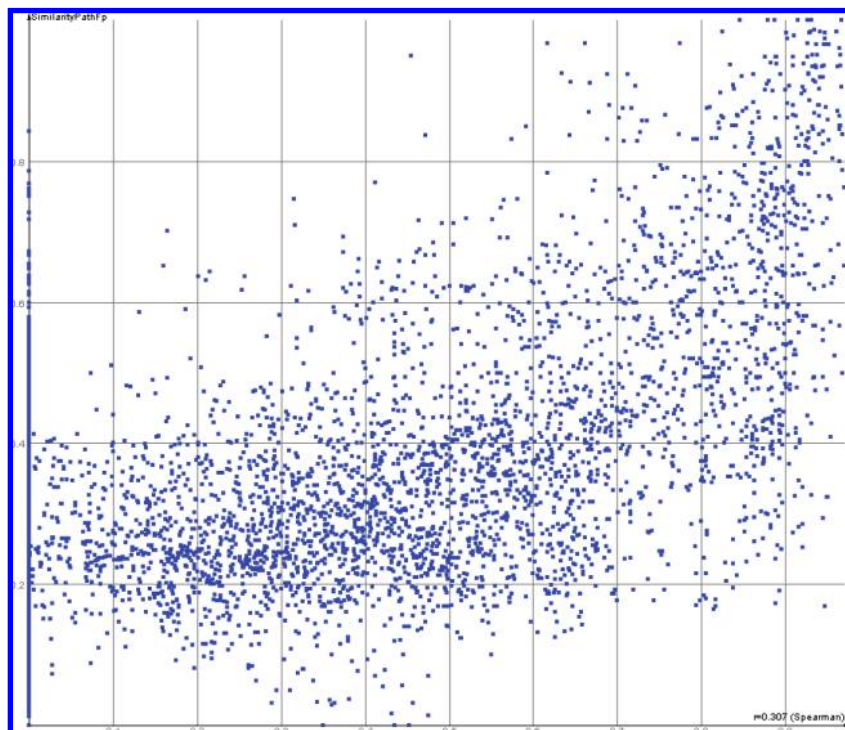
descriptor	ChemAxCFp	FragFp	SphereFp	PathFp	TopPPHist	PP3DMM2
ChemAxCFp		0.649	0.645	<b>0.885</b>	0.567	0.472
FragFp	0.649		0.509	0.591	0.375	0.353
SphereFp	0.645	0.509		0.688	0.41	0.62
PathFp	<b>0.885</b>	0.591	0.688		0.542	0.555
TopPPHist	0.567	0.375	0.41	0.542		0.287
PP3DMM2	0.472	0.353	0.62	0.555	0.287	

<sup>a</sup> Correlation coefficients were calculated according to Pearson; all elements on the diagonal are 1.0 and not shown.

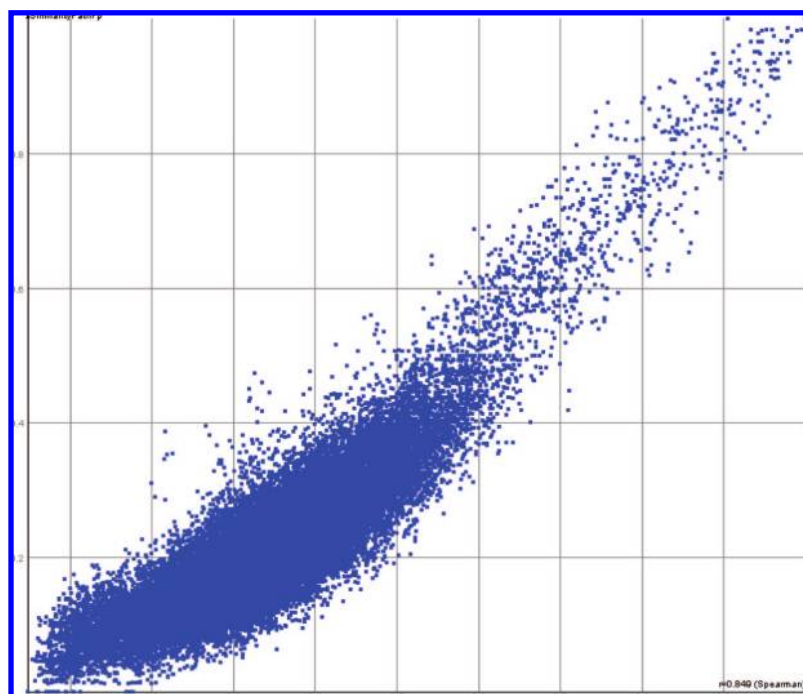
ligands. Figure of merit was the relative enrichment rate, calculated from the average enrichment of the 11 single-query virtual screenings. For Setup<sub>random</sub> the actives in the test sets were selected by random; the results are given in Figure 4 and Table 9. For Setup<sub>dissim</sub> the actives in the test sets were topologically the most dissimilar molecules compared to the query molecule; the results are presented in Figure 5 and Table 10. The variance in performance between the descriptors for Setup<sub>random</sub> is lower than for Setup<sub>crystal</sub>. According to the rank median as a robust estimator, a chemical fingerprint again performed best, but now the Flexophore is the second best descriptor. Only small differences between the four chemical fingerprints and the Flexophore can be observed. The PathFp and the ChemAxCFp performed equally well. Only the TopPPHist descriptor shows a significantly lower performance ranking for the relative enrichment. No conclusion is possible for this setup

whether there are real differences in the descriptors describing the bioactive chemical space. The picture changes for Setup<sub>dissim</sub> (Figure 5 and Table 10); all descriptors performed worse after removing chemically similar ligands from the data set. According to the median ranking the Flexophore descriptor performed best, while the ChemAxCFp and the FragFp lost enrichment capability. The variance in enrichment is much higher than in Setup<sub>random</sub>. For the targets MR, PR, and RXRa all descriptors still showed a good relative enrichment rate. In Setup<sub>random</sub> and Setup<sub>dissim</sub> modeled the TopPPHist descriptor the target MR significantly better than the other methods. For the targets COMT, DHFR, GPB, HMGA, PNP, and VEGFR2 the Flexophore descriptor demonstrated its capability to enrich molecules in data sets where the other descriptors failed.

**Detailed Analysis for Results with Setup<sub>Dissim</sub>.** Because of the larger variance in the enrichment results for Setup<sub>Dissim</sub>



**Figure 6.** Plot of similarity values calculated with the Flexophore descriptor versus (x-axis) similarity values calculated with the PathFp descriptor (y-axis).



**Figure 7.** Plot of similarity values calculated with the ChemAxCFp descriptor versus (x-axis) similarity values calculated with the PathFp descriptor (y-axis).

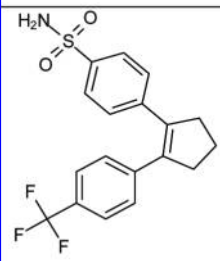
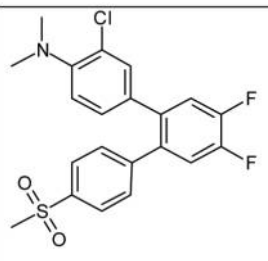
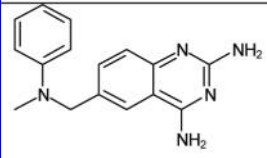
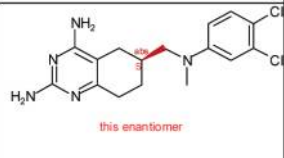
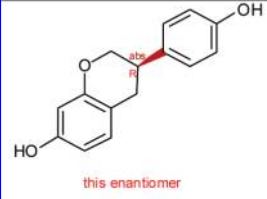
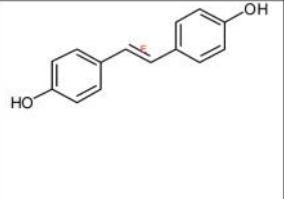
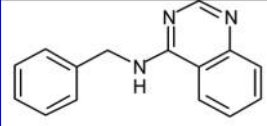
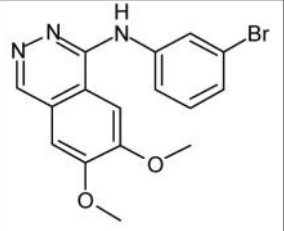
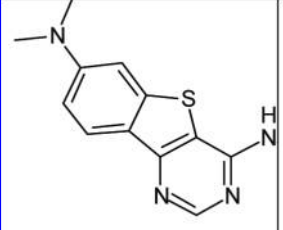
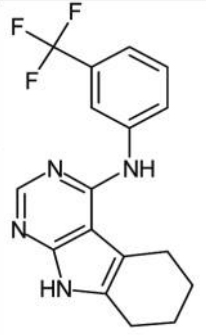
than for Setup<sub>Random</sub>, the four targets in Setup<sub>Dissim</sub> are analyzed in more detail.

**AmpC  $\beta$ -Lactamase (AmpC).** In the Setup<sub>Dissim</sub> the ChemAxCFp descriptor outperformed the other descriptors. In Chart 3 the first enriched molecules for one of the 11 test data sets are given. The ChemAxCFp descriptor was at the detection limit; the similarity for the first ligand found was 0.6. In the authors' experience, the similarity range for the ChemAxCFp descriptor should be between 1.0 and 0.75 to be meaningful. If the highest similarity values are below 0.65,

then the data set has no real relation to the query. Any enrichment of ligands is only due to the fact that the decoys are totally dissimilar to the query molecule. For Flexophore the differences in the number and quality of PPs are too large to detect any similarity within the ligands. Similarly, the substructural patterns for FragFp are too different to reach any enrichment.

**Mineralocorticoid Receptor (MR).** The result of the TopPPHist descriptor in Setup<sub>Dissim</sub> is striking. A look into the query molecule and the enriched ligands allows the

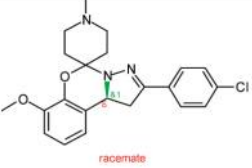
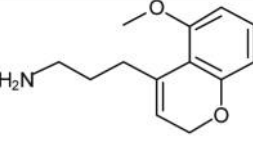
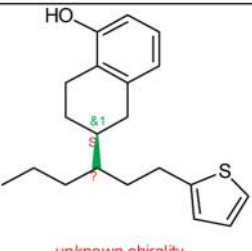
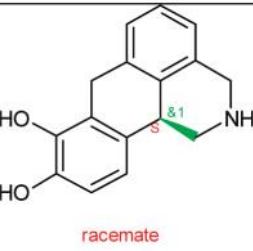
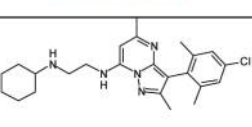
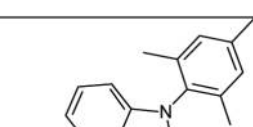
**Chart 8.** Subset from the Similarity Matrix SimMatrixDUDSub

Ligand 1	Ligand 2	Target	FragFp	Sphere Fp	PathFp	Flexo phore
		COX2	0.29	0.18	0.27	0.96
		DHFR	0.29	0.28	0.33	0.96
		ER ago	0.13	0.26	0.22	0.95
		EGFR	0.31	0.18	0.23	0.94
		EGFR	0.35	0.28	0.35	0.92

conclusion that this data set “was made for this descriptor”. Chart 4 shows the query molecule in the upper left corner, and the ligands are shown in the following fields. The scaffold of the query molecule is rigid and contains no heteroatoms. Two oxygen atoms at opposite sides of the scaffold determine the TopPPHist descriptor with one relevant pair of PPs. The resulting descriptor vector is very sparsely occupied and allows a sharp separation between the ligands and the decoys. This sharp separation is necessary, because several decoys similar to the query molecule are also present in the data set. The other descriptors detected additional PPs or substructural features. The relevant information is diluted, and the descriptor performance decreased.

**Trypsin (TR).** For the trypsin data set Flexophore outperformed the other descriptors in Setup<sub>Dissim</sub>. The first sixteen enriched molecules for one of the Setup<sub>Dissim</sub> data sets are shown in Chart 5. The first ligand (second molecule in row one) was found in a perfect substructure match. The analysis of the Flexophore descriptor matches showed that this substructure was also the sole reason for the enrichment of the other matching ligands. The three matching PPs, two for the amines and one for the phenyl, are the only reasons for the enrichment of the diverse chemical structures. This means that the virtual screening could have been replaced with a simple substructure search. The matching substructure is the anchor point for the similarity search; the rest of the

**Chart 9.** Subset from the Similarity Matrix SimMatrixGPCRSub, Calculated from GPCR Ligands

Ligand 1	Ligand 2	Target	FragFp	Sphere Fp	PathFp	Flexophore
		5-HT2B	0.31	0.28	0.31	0.91
		D2	0.30	0.23	0.29	0.93
		NPY1	0.28	0.22	0.31	0.90

**Chart 10.** Comparison of Two Molecules with the Flexophore Descriptors<sup>a</sup>

<sup>a</sup> Molecules are taken from row 3 in similarity matrix SimMatrixGPCRSub (Chart 9); both molecules are active on NPY1. The Flexophore pharmacophore points are visualized with balls within the molecules. Equally colored pharmacophore points on the left and right sides indicate a mapping pharmacophore point. Nonmapping pharmacophore points are indicated in grey.

molecules is only decoration for Flexophore and decreased the similarity. The same observation was made for the targets COMT, DHFR, thrombin, and GPB.

**Hydroxymethylglutaryl-CoA Reductase (HMGR).** The enriched molecules for Flexophore are shown in Chart 6. For HMGR, Flexophore was able to enrich molecules where



the other descriptors failed. Even though the ligands are highly flexible, Flexophore found the necessary PPs to determine a sufficient similarity between the query and the ligands in rows 1 and 4. Flexophore also enriched decoys with a hydroxyl group instead of the carboxylic acid. According to the PDB interaction statistics, both groups have a certain probability to be found close to the same amino acids in a target protein. This broadness of acceptance lowers the enrichment rate but reflects the experience that sometimes a hydroxyl may be sufficient to replace the carboxyl.

#### **Glycinamide Ribonucleotide Transformylase (GART).**

In Chart 7 the molecules are sorted according to their similarity to the query, as calculated with Flexophore descriptor. None of the enriched molecules belong to the set of ligands, but they all show a high similarity in descriptor space. The same effect was observed in the AR data set. According to the authors' experience, the enriched decoys would be chosen to undergo biological screening. This demonstrates one of the dilemmas for virtual screening experiments: the whole test data set should be tested in biological screening on the target proteins. To test the 100,000 molecules of the DUD data set in 40 biological assays for IC<sub>50</sub> values would be an enormous effort. The dilution of biologically tested molecules with nontested decoys might result in a too limited assessment of the descriptor performance.

The results of the descriptor orthogonality calculations are summarized in Table 11 and Figures 6 and 7. Table 11 indicates that there is no correlation between the descriptors, except between PathFp and ChemAxonFp. A correlation of 0.885 was observed for this pair of descriptors. In Figure 6, the mutual compound similarity values of PathFp are plotted versus the ones of Flexophore as an example of a pair of uncorrelated descriptors. The plots of the correlation curves of the other uncorrelated descriptor pairs looked similar. At high similarity values a slight correlation between the two descriptors can be observed. This is due to the fact that descriptor similarity values of very similar molecules are close to 1.0, independent of the descriptor type, but for the majority of similarity values no correlation exists. The similarity correlation plot of the PathFp versus the ChemAxonFp in Figure 7 shows a clear correlation between these two descriptors. This is the only significant correlation among all pairwise descriptor combinations. PathFp and ChemAxonFp are topological path descriptors, and their algorithmic principles are similar enough to describe the chemical space in a related way. This correlation is in line with the similar performance of the two descriptors in the virtual screening experiments. The noncorrelation between the other descriptors explains why these descriptors selected different molecules in the virtual screening experiments. Despite their similar enrichment rates in the virtual screening experiments with Setup<sub>crystal</sub> and Setup<sub>random</sub>, the descriptors perceive the chemical space of biological active molecules in an orthogonal way. Some structural examples underlining the orthogonality between the Flexophore descriptor and the chemical fingerprint descriptors are shown in Charts 89. The charts represent sections of the similarity matrices SimMatrixDUDSub and SimMatrixGPCRSub. Each line shows a pair of molecules being active on the same target along with similarity

values for the descriptors FragFp, SphereFp, PathFp, and Flexophore. In all cases, the Tanimoto values of the three chemical fingerprints are below 0.5, while the Flexophore similarity is always about 0.9. In comparison of any molecule pair shows a clear similarity from the medicinal chemist's perspective. Therefore screening with any of these molecules using a chemical fingerprint descriptor would miss its counterpart as an obvious candidate for biological testing. For the Flexophore descriptor the differences are perceived as being much smaller; it would choose these molecules in a virtual screening experiment as hits. In Chart 9 the pairs of molecules differ much more in terms of chemical structure than in the previous example. The three examples in Chart 9 have chemical fingerprint similarities below 0.5, for good reasons. In all three cases the Flexophore descriptor still indicates high similarities. This is the part of the bioactivity space where the Flexophore unleashes its full potential. In Chart 10 the mapping PPs are visualized for the last example of Chart 9. The PPs are projected into a single conformation of the structure to visualize the relation between PPs and related structural features. The structure itself is not part of the Flexophore descriptor. In the Flexophore descriptor, the PPs are all connected with distance histograms generated by the conformation generator. Equally colored PPs on the left and right sides indicate a mapping pair. Pharmacophore points without counterpart are indicated in gray. In total there are eight mapping PPs, which contribute to the high similarity score of the Flexophore descriptor comparison.

## CONCLUSIONS

A clear target dependency can be seen in all three setups and in the seven virtual screening methods. The high standard deviation for the virtual screening methods shows the necessity to test virtual screening methods on manifold diverse targets. There are large differences in the success of virtual screening caused by the target proteins and the structural classes of the ligands. This finding was only possible because of the diverse target selection done in the creation of the DUD. The diverse targets and the resulting huge variance in the virtual screening success had consequences for the result summary. To get a correct impression for the performance of the different methods it is absolutely necessary to use robust statistics. The virtual screening experiments with the original DUD showed Setup<sub>crystal</sub> superiority for chemical fingerprints, which was expected because of the bias of the data set. The two docking algorithms performed comparably, although the underlying principles are different. The average enrichment rate for both docking programs was almost equal after removing the results for ADA and HIVRT, where no automatic docking was possible because of missing ligands.

With the enrichment rate as a figure of merit the SphereFp performed almost equally well as the PathFp, but a cluster analysis demonstrated that the SphereFp is superior in enriching diverse molecules. This demonstrates that chemical diversity in virtual screening can be reached with chemical fingerprints. The detailed analysis of Setup<sub>crystal</sub> showed that docking- and ligand-based screening performed independently. Targets with a huge binding cavity like TK or

unknown challenges like InhA gave the docking algorithms a tough task. For the ligand-based methods only the binding ligand is important. An appropriate ligand for virtual screening must have chemical features to generate a meaningful descriptor. This was demonstrated with the MR data set for Setup<sub>Dissim</sub> and is discussed later.

The broader virtual screening experiments with Setup<sub>Random</sub> resulted in a good performance for all descriptors. After removing the chemically similar ligands in Setup<sub>dissim</sub>, the Flexophore descriptor was the optimum choice to find active molecules beyond chemical similarity. While enrichment on some targets was extraordinary, it failed on others entirely to detect actives. The results from the detailed analysis may give some hints for successful enrichment. On the MR data set for Setup<sub>Dissim</sub> it was shown that the necessary substructure's features in the query molecule for successful enrichment differ between the descriptors. The query molecule had a sufficient substructure combination to reach an excellent enrichment with the TopPPHist descriptor, though the other descriptors were not able to generate meaningful descriptors.

For all ligand-based methods a strong substructure dependency was observed. This was obvious for the chemical fingerprints ChemAxCFp, FragFp, PathFp, and SphereFp, which rely only on topological graph features. Surprisingly, the Flexophore descriptor also showed this dependency when it outperformed the chemical fingerprints. In the detailed analysis for COMT, GPB, DHFR, thrombin, and trypsin with Setup<sub>Dissim</sub>, Flexophore detected a relevant pharmacophore pattern from one substructure. This pattern was used by Flexophore to enrich the ligands, whereas the rest of the molecule resulted in PPs, which only decreased the similarity score. For the chemical fingerprints this substructure did not result in descriptors, which could be used to get meaningful similarity scores. The appropriate combination of descriptor query molecule, ligands, and decoys determines the success of virtual screening experiments. A successful descriptor is able to extract the structural features from a query molecule, which discriminate between ligands and decoys. Furthermore, the descriptor should have enough generalization power to enrich different structural classes of ligands. In conclusion, the authors recommend the use of a SphereFp-like descriptor choice for virtual screening. The implementation is straightforward and can be done with any canonical molecule description. Additionally, the SphereFp and DOCK together was the best method combination. The combination of SphereFp and Flexophore performed comparably but had the advantage that comparisons with the Flexophore descriptor are considerably faster than docking with the DOCK algorithm.

A detailed similarity analysis for the Flexophore descriptor was done by calculating two similarity matrices. Examples of the first similarity matrix demonstrated that the Flexophore descriptor is able to describe molecules as similar where chemical fingerprints fail. Examples of the second similarity matrix showed that its similarity perception is indeed uncoupled to the topology of the underlying structures. Encouraged from the good results of the test data sets the Flexophore descriptor was applied in several in-house virtual screening projects. The descriptor performed well in suggesting several molecules for

biological testing, which were later confirmed as hits. These new hits were chemically dissimilar to the active molecules already known and encouraged us to regularly apply the Flexophore descriptor in virtual screening in our drug discovery department.

## ACKNOWLEDGMENT

Christian Rufener implemented Actelion's computing grid and the PathFp. We thank Christopher Snyder for editorial assistance.

**Supporting Information Available:** MDL SD files with the structures for Setup<sub>crystal</sub>, Setup<sub>random</sub>, and Setup<sub>dissim</sub> and a tab separated table clusterVS\_DUD.txt with the detected cluster identifiers. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
- (2) Sheridan, R.; Kearsley, S. Why do we need so many chemical similarity search methods. *Drug Discovery Today* **2002**, 7, 903–911.
- (3) Sheridan, R. P. Chemical similarity searches: when is complexity justified. *Expert Opin. Drug Discovery* **2007**, 2, 423–430.
- (4) Good, A. C.; Oprea, T. I. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection. *J. Comput.-Aided Mol. Des.* **2008**, 22, 169–178.
- (5) Yeap, S. K.; Walley, R. J.; Snarey, M.; Hoorn, W. P. v.; Mason, J. S. Designing Compound Subsets: Comparison of Random and Rational Approaches Using Statistical Stimulation. *J. Chem. Inf. Model.* **2007**, 47, 2149–2158.
- (6) Rohrer, S. G.; Baumann, K. Impact of Benchmark Data Set Topology on the Validation of Virtual Screening Methods: Exploration and Quantification by Spatial Statistics. *J. Chem. Inf. Comput. Sci.* **2008**, 48, 704–718.
- (7) Good, A. C.; Hermsmeider, M. A.; Hindle, S. A. Measuring CAMD technique performance: A virtual screening case study in the design of validation experiments. *J. Comput.-Aided Mol. Des.* **2004**, 18, 529–536.
- (8) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, 49, 6789–6801.
- (9) von Korff, M.; Freyss, J.; Sander, T. Flexophore, a new versatile 3D pharmacophore descriptor that considers molecular flexibility. *J. Chem. Inf. Model.* **2008**, 48, 797–810.
- (10) ChemAxon. GenerateMD. <http://www.chemaxon.com/jchem/doc/user/fingerprint.html> (accessed Nov 8, 2007).
- (11) Xue, L.; Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb. Chem. High Throughput Screening* **2000**, 3, 363–372.
- (12) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 233–245.
- (13) Sander, T. ActelionFp; Department of Research Informatics, Actelion: Allschwil, Switzerland, 2002.
- (14) McGregor, M. J.; Pallai, P. V. Clustering of Large Databases of Compounds: Using the MDL "Keys" as Structural Descriptors. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 443–448.
- (15) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. *J. Med. Chem.* **2006**, 49, 6802–6810.
- (16) Daylight Fingerprints. <http://www.daylight.com/meetings/summer-school98/course/basics/fp.html> (accessed Jul 30, 2008).
- (17) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31–36.
- (18) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 97–101.
- (19) Jenkins, R. Hash Functions and Block Ciphers. <http://burtleburtle.net/bob/c/lookup3.c> (accessed Apr 11, 2008).
- (20) Korff, M. v.; Steger, M. GPCR-Tailored Pharmacophore Pattern Recognition of Small Molecular Ligands. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 1137–1147.

- (21) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 64–73.
- (22) 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.
- (23) Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1297–1308.
- (24) Allinger, N. L.; Zhou, X.; Bergsma, J. Molecular Mechanics Parameters. *J. Mol. Struct.* **1994**, 312, 69–83.
- (25) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235–242.
- (26) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, 295, 337–56.
- (27) Ponder, J. W.; Richards, F. M. An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, 8, 1016–1024.
- (28) Fischer, B. K. *High-throughput simulation methods for protein-ligand docking*; Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft: Karlsruhe, Germany, 2007.
- (29) Irwin, J. J. DUD Release 2. <http://dud.docking.org/r2/> (accessed Jul 9, 2008).
- (30) Good, A. DUD Clustering. <http://dud.docking.org/clusters/> (accessed Nov 25, 2008).

CI800303K