Journal of Computer-Aided Molecular Design, **13:** 35–49, 1999.
KLUWER/ESCOM

Methodological developments and strategies for a fast flexible superposition of drug-size molecules

Gerhard Klebe*, Thomas Mietzner & Frank Weber
BASF AG, Hauptlaboratorium, D-67056 Ludwigshafen, Germany

Received 16 December 1997; Accepted 12 July 1998

Key words: knowledge-based conformational search, molecular similarity, molecular superposition, torsional library, virtual screening

Summary

An alternative to experimental high through-put screening is the virtual screening of compound libraries on the computer. In absence of a detailed structure of the receptor protein, candidate molecules are compared with a known reference by mutually superimposing their skeletons and scoring their similarity. Since molecular shape highly depends on the adopted conformation, an efficient conformational screening is performed using a knowledge-based approach. A comprehensive torsion library has been compiled from crystal data stored in the Cambridge Structural Database. For molecular comparison a strategy is followed considering shape associated physicochemical properties in space such as steric occupancy, electrostatics, lipophilicity and potential hydrogen-bonding. Molecular shape is approximated by a set of Gaussian functions not necessarily located at the atomic positions. The superposition is performed in two steps: first by a global alignment search operating on multiple rigid conformations and then by conformationally relaxing the best scored hits of the global search. A normalized similarity scoring is used to allow for a comparison of molecules with rather different shape and size. The approach has been implemented on a cluster of parallel processors. As a case study, the search for ligands binding to the dopamine receptor is given.

Introduction

Lead identification is the key element in the drug discovery process in pharmaceutical industry. At present in most places a dual approach is followed. With the introduction of molecular test systems, high-throughput screening of compound libraries containing some hundred thousand entries is feasible within several weeks to months [1]. Binding of a low-molecular weight ligand to a particular receptor protein is registered through an easily detectable signal. These random screening techniques are complemented by computer methods to provide ideas about new lead compounds. Meanwhile more than 14 million compounds are described and computer-accessible. In large companies in-house databases with some hundred thousand entries are available. How can these data be used in the lead finding process?

If the 3D structure of the target protein is known, fast docking algorithms indicate possible candidates from such libraries [2]. In the absence of a detailed structure of the target protein methods comparing candidate molecules with a known reference ligand can help in the design of novel compounds [3]. In order to provide the essential prerequisites for receptor binding, a similarity analysis of each database entry with a given reference compound is required. These approaches have to compare the shape and the distribution of physicochemical properties in space [4]. A reliable scoring of the detected similarity between probe and reference molecule is needed. Since molecular shape is dependent on the adopted conformation, an adequate consideration of molecular flexibility is essential for the success of this approach. Accordingly an efficient and representative coverage of the relevant part of conformation space must be provided [5].

*To whom correspondence should be addressed. Present address: Inst. of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany

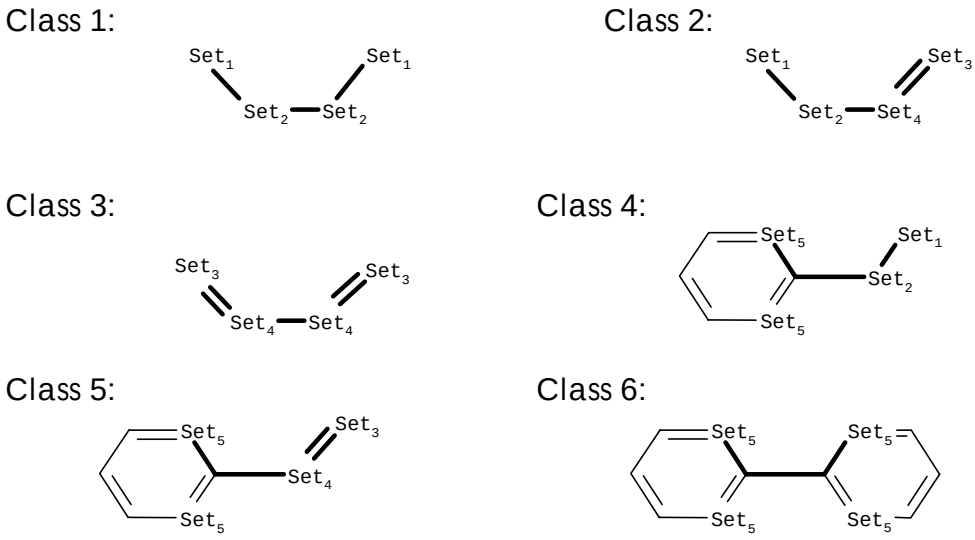


Figure 1. From a list of atom types (cf. five different sets in Table 2), all possible combinations were generated to construct torsion angle fragments. Three different bond types (single, double, aromatic) were allowed. The central bond of a torsional fragment was requested to be a single bond. In total six different fragment classes are distinguished.

In the present paper we describe a series of consecutive algorithms that allow for an automatic flexible alignment of entries from a database with a given reference compound in order to perform 'virtual screening' in the computer. These techniques include a fast knowledge-based conformational analysis exploiting torsional preferences from structural databases to generate a limited set of well distributed biologically relevant conformations. Structural superposition or alignment is performed in two consecutive steps, first by aligning rigid conformers in space followed by a local conformational relaxation of the pre-aligned molecules. None of these steps requires any preconceived matching pairs of centers or assignments of a pharmacophore directly associated with the molecular framework. The entire screening process is fully automated and operates on samples of several thousand candidate molecules from databases.

A well distributed set of biologically relevant conformers

Recently, we described the computer program MIMUMBA [6]. This program is based on empirical knowledge taken from crystal structures of small organic molecules. MIMUMBA generates conformers based on conformational preferences of torsional fragments retrieved from the Cambridge Structural Database (CSD) [7]. We showed [6, 8] that this approach allows for a fast and efficient coverage of the part of conformational space comprising many if not most of the biologically relevant conformations. The major advantage of scanning observed conformational preferences instead of sampling local minima of a molecular-mechanics program is the following. The shape of a particular energy surface defined by a force-field (usually for vacuum conditions) will be strongly modified once a molecule is exposed to a structured molecular environment involving directional interactions and varying dielectric conditions. The width and in particular the relative hight of local minima of a force field will significantly change (or minima even vanish or newly occur). Accordingly, which local minima remain and what is their relative importance? A tempting alternative to circumvent this problem is the use of conformational preferences derived by statistical means from a broad ensemble of different molecules exposed to various crystal packings. They comprise implicitly the (mean) information about the influence of a structured molecular environment with varying dielectric conditions – a situation very similar to that of a ligand exposed to the binding site of a protein.

For conformational analysis, a molecule under consideration is split into rigid fragments and flexible ring portions. For the latter, low-energy conformations are taken from a list of predefined ring templates [9]. In the recent version of MIMUMBA, the ring conformational analysis can optionally be performed by

Table 1. Sample input for a QUEST run to extract

```
COMM
COMM number: 100079
COMM class: C3_SO2
COMM torsion: Cany 1 C3(H)(H) 1 SO2(O2)(O2) 1 Cany
COMM period: 360
COMM symmetry: 180
COMM
STOP 20000
SCREEN 153
T1 *CONN
ELDEF HT=N,O,F,P,S,Cl,Br,I
ELDEF HL=F,Cl,Br,I
ELDEF R=C,HT
AT1 C
AT2 C 2 2 T4
AT3 S 4 0 T4
AT4 C
AT5 O
AT6 O
BO 1 2 1
BO 2 3 1 A
BO 3 4 1
BO 3 5 99
BO 3 6 99
GEOM
DEFINE T1 1 2 3 4
TRANSFORM T1A = ABS T1
HIST T1A
NFRAG −99
END
SAVE 0
QUEST T1
```

Table 2. Composition of the various sets

No.	Set1	Set2	Set3	Set4	Set5
1	Cany	C3(H)(H)	N2	N2	Car(H)
2	Nany	C3(H)(Cany)	C2	C2(H)	Car(R)
3	O3	C3(H)(X)	O2	C2(Cany)	Nar
4	Sany	C3(Cany)(Cany)	S2	C2(X)	
5	P3	C3(Cany)(X)			
6	Halo	C3(X)(X)			
7	F	Nany(H)			
8	OCO2	N3(Cany)			
9		N3(X)			
10		Npl3(Cany)			
11		Npl3(X)			
12		O3			
13		SO(2O2)			
14		SO2(2O2)(2O2)			
15		N4(H)(H)			
16		N4(H)(Cany)			
17		N4(H)(X)			
18		N4(Cany)(Cany)			
19		N4(Cany)(X)			
20		N4(X)(X)			
21		S3			
22		P3(H)			
23		P3(Cany)			
24		P3(X)			
25		P3(Oany)(Oany)			
26		P3(Oany)(Cany)			
27		C3(F)(F)			
28		C2(OCO2)(OCO2)			
29		N2(OCO2)(OCO2)			
30		Ccat(Nany)			

X= heteroatom: N, O, S, ...

the algorithm implemented in the program CORINA [10]. Values for the various dihedral angles in the open-chain portions are taken from a pre-collected internal database. It contains the results from statistical analyses of conformational preferences observed in crystal structures. In a first version, these data were individually collected to cover the most frequently occurring cases. The database compiled by this strategy contained 276 fragments. However, applying the program routinely discovered many missing torsional fragments. In order to cope with such unspecified cases default values were assigned in the early version of the program [6].

To improve on this unsatisfactory situation, we embarked into an automatic retrieval of torsional fragments from the CSD. The retrieval conditions were defined as described previously [6], but we entirely performed the searches with version 5.11 of QUEST. A sample input is given in Table 1.

The query input files for QUEST have been generated by the program MTYP, written by us, in a systematic manner. From a given list of atom and bond types it forms all possible torsion angle combinations. The setting of the individual atom and bond types was selected according to those used in SYBYL [11]. A torsional search fragment is constituted by four members. Each member belongs to a group of atoms corresponding to a different 'set'. In total we distinguish five different sets (Table 2).

The rationale for this grouping in five sets is determined by classifying the bond types in a torsional fragment into single, double and aromatic bonds. The

central bond of a torsional fragment has to be a single bond. The adjacent bonds can be either single, double or aromatic bonds. Using this scheme together with the atom-type grouping we can generate six different classes of fragments (Figure 1). Depending on the type of substitution, the entries in the individual classes possess different local symmetry. As a consequence, torsion angles reproduce the same geometry after different periodicities. The various possibilities considered in each of the six classes are given in Figure 2. All data retrieved from the CSD were symmetrizied according to the rules given in Figure 2. In principle in a particular example taken from the CSD the atoms forming a torsional fragment (e.g. substituents at the central bond) can give rise to molecular asymmetry (local chirality). This was not taken into consideration. Thus, during data retrieval and compilation no control over the absolute configuration of the fragments was accounted for. As a consequence, the subsequently performed symmetrization was accomplished as if the data were obtained from racemic structures.

Considering all possible permutations and excluding chemically unreasonable combinations such as four oxygen atoms in a sequence, a total of 25124 QUEST runs were generated. Most queries fell into class 1 and 2 whereas the others were much less populated (Table 3). Interestingly, most of the queries from class 1 did not retrieve any data or only one single example from the database. In total only 5% of the examples in class 1 are populated with ≥ 2 hits (Table 3). Out of the total 25124 queries only 2002 showed two or more examples and about 1000 had more than 10 hits. The individual statistics are given in Table 4.

Since the obtained histograms have to be translated into pseudopotentials as described previously [6] too sparsely populated distributions cannot be evaluated. Accordingly, we restructured our definition of the sets. All halogen atoms were combined into a general type Halo. The carboxylate and NO_2 group were introduced. The frequently occurring CF_3 and carboxylate groups in set 2 were treated as such, solely considering the combination of set1(7) with set2(27) and set1(8) with set2(28). This definition yielded in 15853 QUEST runs resulting in 2053 distributions populated with two or more examples. Out of these distributions 711 were selected having on average more than 2 entries per 10° within the corresponding symmetry range.

To handle the less frequently occurring fragments in the database, additional searches have been performed. Less detailed definitions of the search fragment have been chosen. All bond and atoms types at the inner central bond were selected as previously described. For the terminal atoms (set1, set3, Figure 1) generalized types R and X have been defined (R denotes any heavy atom, X denotes any heteroatom). Substituents at the inner positions (set2 and set4) of the torsional fragments have been distinguished being either hydrogen or non-hydrogen. Using these generalized atom types the searches revealed 204 and 183 different QUEST runs. Out of these only those having on average more than 2 examples per 10° in the periodicity range were considered for the further analysis.

In order to handle efficiently peptidic portions in ligands, data from well resolved protein structures ($<$ 2 Å) have been used. A total of 59395 tripeptide sequences have been extracted [12] with the program WHATIF [13] from the Brookhaven Protein Databank (PDB) [14]. These data were input for the statistical analysis of 97 individual torsional patterns describing various ϕ, ψ, χ_n angles in the different amino acids. Since data only from natural amino acids have been used, the chirality of the search fragments has been considered. The periodicity for these search fragments corresponds to full 360°.

The distributions from the above described searches have been combined into a new internal database. This selection comprises 1013 independent patterns. After the topological analysis of a molecule in MIMUMBA, the program first scans the part of the database that contains the results from the detailed searches in the PDB and subsequently those from the CSD. For any torsional fragments not to be matched among these entries, the program finally performs an additional search among those entries that were compiled using the more generalized atom types (R, X).

Improvements of the torsion angle selection in the conformer build-up procedure

In the previous version of MIMUMBA, torsion angle distributions in the various histograms have been ranked energetically according to a 30° grid on a relative energy scale between 0 and 10 kJ. Angles used in the conformer build-up procedure were selected according to this 30° raster. Torsion angles at weakly populated grid values, corresponding to high energies were discarded. If more than one torsion angle distrib-

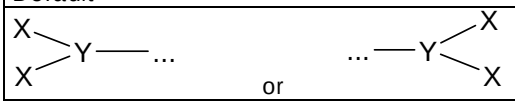
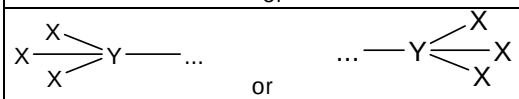
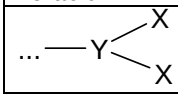
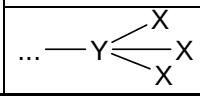
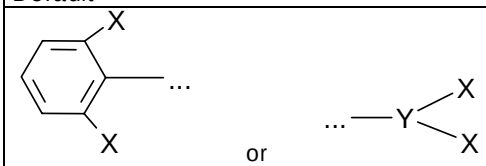
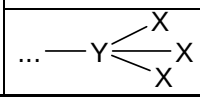
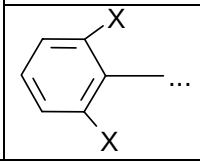
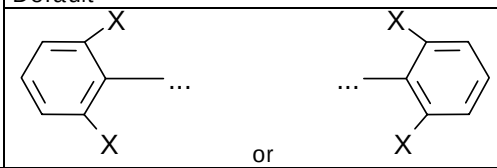
Class	Fragment	Periodic.	Symmetry
1	Default	360	180
		180	90
		120	60
2	Default	360	180
		180	90
		120	60
3	All	360	180
4	Default	360	180
		180	90
		120	60
5	Default	360	180
		180	90
6	Default	360	180
		180	90

Figure 2. For the six different classes of torsional fragments different local symmetry has to be considered. In consequence, the torsion angles reproduce the same angle after different periodicities. The symmetry given in the list corresponds to the angular range after which the histograms obtain a symmetry-related (e.g. mirrored) shape.

Table 3. Statistics for the various CSD searches (Version 5.11)

	Class1	Class2	Class3	Class4	Class5	Class6	All
Sets	$1-2-2-1$	$3=4-2-1$	$3=4-4=3$	$5\sim5-2-1$	$5\sim5-4=3$	$5\sim5-5\sim5$	
Quests	20031	3612	96	1283	90	12	25124
Hits <2	18938	3111	36	992	27	2	23111
Hits ≥ 2	1077	501	60	291	63	10	2002
Hits≥ 2, rel.(%)	5	14	63	23	70	83	8
Hits ≥ 10	474	249	45	177	49	8	1002

Table 4. Statistics for the QUEST runs

Hits	Quests-Entries	
0	13347	0 or 1 hit: 13986 Quests-Entries
1	639	
2	247	2 or more hits: 2044 Quests-Entries
3	172	
4	135	
5	88	5 or more hits: 1490 Quests-Entries
6	79	
7	74	
8	59	
9	56	
10	51	10 or more hits: 1134 Quests-Entries
11–15	150	
16–20	101	
21–25	80	21 or more hits: 883 Quests-Entries
26–30	53	
31–40	91	
41–50	56	51 or more hits: 603 Quests-Entries
51–75	136	
76–100	56	
101–150	71	
151–200	38	
201–300	47	
301–500	56	
501–1000	50	
1001–2000	39	
2001–5000	26	
>5000	6	

ution in the database could be matched to a particular torsional fragment in the probe molecule, only those 30° grid values were kept where all of the considered database entries showed reasonable population. As energy ranking for such an angle the highest value from all matching distributions has been used. During the build-up procedure the energy rankings of the different torsion angles were added. Conformers that obtained an energy ranking beyond a predefined threshold were discarded.

In the present version of MIMUMBA we modified this selection procedure. In contrast to the original version of the database, the individual frequencies are now collected at $10 \times n + 5°$ ($n = 0, 1, \ldots 35$, depending on the symmetry and periodicity of the considered fragment, Figure 2). In a first step all matching histograms are selected and individually translated into pseudopotentials [6]. The potential value at a given

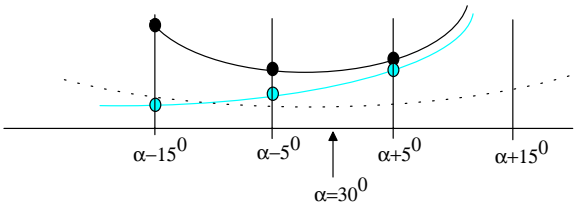


Figure 3. For a probe angle $\alpha = 30°$ the following situation might occur. At first, the potential described by the solid line should be considered. Since the potentials at $\alpha - 15°$ and $\alpha + 5°$ both are larger compared to the one at $\alpha - 5°$ (strict local minimum), $\alpha = 30°$ is accepted as favorable angle. If the situation would correspond to the dotted line, $\alpha = 30°$ would not be considered as favorable. If $\alpha - 5°$ and $\alpha + 5°$ have the same value (dashed line) the potentials at $\alpha \pm 15°$ have to be compared. If both are the same or larger, $\alpha = 30°$ is selected as starting value.

angle is expressed as the negative logarithm of its frequency:

$$\mathrm{Pot}(\tau) = -\mathrm{A} \ln \mathrm{f}(\tau)$$

As in the original version, in those ranges where no experimental data are found (frequency=0), potential values are interpolated by a quadratic polynomial that meets at its maximum an adjustable parameter which was finally set to 30 kJ/mol. Subsequently, the potentials are superimposed and averaged. To select preferred angles, the following strategy has been used. Only in those angular ranges where the potential shows a minimum, start values are collected. The selection is performed in steps of 15°, however, 30° are used as minimal distance between two adjacent favorable angles. For example, an angle at $\alpha = 30° * n$ is only selected, if the potential at $\alpha - 5°$ or $\alpha + 5°$ is a strict local minimum. Figure 3 explains the required condition. For $n = 1$ and $\alpha = 30°$, the potential values at $\alpha - 5°$ and $\alpha + 15°$ or $\alpha - 15°$ and $\alpha + 5°$ must be larger than the values at $\alpha + 5°$ or $\alpha - 5°$ respectively. If the potential has the same value at $\alpha - 5°$ and $\alpha + 5°$, the values at $\alpha - 15°$ and $\alpha + 15°$ are compared. If both show the same or larger values with respect to $\alpha \pm 5°$, α is accepted as a preferred angle. Subsequently, the angles $\alpha = 30° * n + 15°$ are analyzed. These angles are selected if they fall to strict local minima or to a plateau with strictly increasing boundaries no further than $\pm 20°$ apart from α.

The program determines the theoretically possible start structures covering all possible combinations of preferred open-chain angles and ring conformations below a particular energy threshold. If this number exceeds 2×10^6 start structures the energy threshold is gradually reduced and simultaneously the torsion angle selection is performed in a more stringent manner.

Structurally diverse conformers are generated by selecting torsion angles at the periphery of the molecule in larger steps. If required for rather large molecules, the torsional fragments toward terminal groups can be completely frozen.

Compared to the previous version of the program, a more efficient selection of the angles is accomplished. Accordingly, by using less start structures a similar diversity of conformers is generated. Or by keeping the complexity of the search, larger ligands with more rotatable bonds can be analyzed. In our previous paper we described the conformational analysis of methotrexate as an upper-limit case for MIMUMBA. With the old version we generated 445 conformers. The one energetically ranked at position 242 approximated best the experimentally given geometry in the complex 3DFR (PDB code) with an rmsd of 0.76 Å. We reconsidered the search for methotrexate with the new selection algorithm and the extended torsion library. Using roughly the same computing time we now generated 492 conformers within 6 min 22 s on an SGI with R4000 processor (mean 0.78 s per conformer). In this search already the conformer ranked at position 110 (being 5.5 energy units ('kJ/mol') above the global minimum) showed an rmsd of 0.64 Å with respect to the conformation of methotrexate in the complex with the PDB code 3DFR. This is a substantial improvement in search efficiency (rank 110 compared to 242) achieved at the same time for a slightly better rmsd (0.64 compared to 0.76 Å).

In order to assess the reliability and the efficiency of the new angle selection, the following tests were performed. To assess the quality of the potential model, local minima generated by MIMUMBA were compared with the crystallographically determined binding site geometries of 63 protein ligands [15]. The examples in this test suite comprised 3 up to 45 rotatable bonds. Starting with the experimentally determined geometry, bond lengths and angles were relaxed using the MaxiMin force-field [16] in SYBYL [11]. Subsequently, the structures were optimized into the next local minimum in the torsion angle force-field of MIMUMBA. In total the residual rms deviations between crystallographically determined and minimized structures amount on average to 0.07 Å per rotatable bond (the total deviation is obtained by multiplying 0.07 Å with the number of rotatable bonds of the considered ligand). A subset of 28 ligands [17] containing up to 13 rotatable bonds was used to generate multiple conformers with MIMUMBA. Now, the analysis starts with an arbitrary conformation of the ligand. Between 6 and 150 distinct conformers for each of the considered ligands have been generated and subsequently evaluated with respect to its resemblance with the protein-bound conformation. Using the one that approximates most closely the experimentally observed geometry, on average, a deviation of 0.1 Å per rotatable bond is observed (for details including CPU timings cf. Table 5). This figure is close to the one obtained by pure relaxation of the crystallographically observed conformation (see above). The following test has been performed to obtain an objective estimate for the upper limit of rmsd values to be expected in superimposing ligands of the present size: the different conformers were superimposed with random orientation keeping only their centers of mass in common position, and, on the mean, an rms deviation of 0.3 Å per rotatable bond has been obtained [18]. This value has to be compared with the 0.1 Å/bond for MIMUMBA. For a molecule with 15 rotatable bonds this means an expected mean rmsd of 1.5 Å for MIMUMBA compared to 4.5 Å for a random mass-centered superposition.

Structural alignment of conformers considering physicochemical properties

In order to compare the different conformers of a test molecule with a given reference by physicochemical properties we apply the procedure SEAL, described by Kearsley and Smith [19]. In its original version, a simultaneous consideration of steric and electrostatic properties [20] is performed. In a first attempt we included atom-based lipophilicities [21] and calibrated several adjustable parameters by reproducing the superposition of a set of ligand pairs binding to common protein receptors. For these test examples the actual binding mode and accordingly the relative structural alignment is known from protein crystallography. A parameter setting has been finally accepted that approximates on average best the experimentally observed alignments over the entire data set [22].

In order to consider in addition hydrogen-bonding properties, a strategy is followed similar to that in the program DISCO [23]. The various functional groups are classified, e.g. as carboxylate, amide, alcohol or amine. To define putative hydrogen-bonding sites about these groups the results from mapping composite crystal-field environments in small molecule crystal structures have been evaluated [24, 25]. They have been classified according to a small set of generic

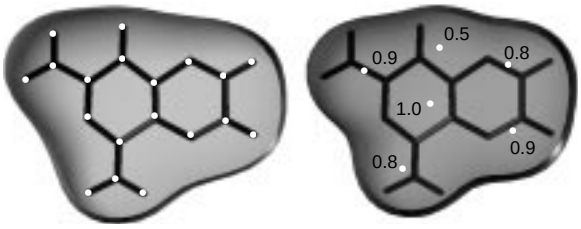


Figure 4. Steric density distribution about the pteridine moiety of methotrexate described by a set of 19 atom-based Gaussians (left, Gaussians localized at each position of the 19 atoms, white dots) and 6 'floating' Gaussians not localized at the atomic positions (right, Gaussians centered at the positions indicated by the white dots). The contour level has been set to an arbitray level in order to approximate the van der Waals volume of the molecule by both representations. Convincing similarity is indicated. In the right part the centers of the six Gaussians are indicated together with a relative weighting (highest Gaussian set to 1).

types. For each such generic type, representative spatial positions are defined at the population maxima of these distributions, e.g. four such centers on a spherical cone about a carboxylate oxygen with two in and one above and below a plane through the functional group (distance 1.9 Å). For bifunctional groups capable to operate as donor and acceptor, for instance, an alcohol group, both acceptor and donor centers are generated with three-fold symmetry (cf. Table 6). Any such center that clashes or falls closer than 1.8 Å with an atom of the probe molecules is discarded from the list. Finally, the different centers are weighted according to the total number of sites attributed to a particular functional group. We could show in another study [26] that bivalent oxygen atoms bound to two non-hydrogen atoms of which at least one is formally assigned to sp^2-type hybridization are no or very weak hydrogen-bond acceptors. Accordingly these oxygen atoms are not considered as potential hydrogen-bonding partners in our approach.

To quantify the spatial similarity of two molecules [27], their shape is approximated by a set of spatial Gaussian-type functions centered at the atomic positions. The use of Gaussians to express molecular properties in similarity matching has been followed in other approaches [28]. Subsequently, for each molecule these functions are associated with a vector of physicochemical properties derived from atom-based descriptors. By selecting a weak distance-dependent attenuation for these Gaussians, the associated steric, electrostatic, and lipophilic properties are considered in a rather delocalized fashion. Similar to the treatment of the molecular shape, the generated hydrogen-bond donor and acceptor sites are associated to Gaussian functions. Regarding the composite crystal-field environments, especially for acceptors, one easily detects rather broad distributions in space. This 'smearing' out in space will be considered in the comparative alignment by applying the same weak distance-dependent attenuation as used for the other properties. In order to reflect the directionality of hydrogen bonding in the comparison a directionality term is introduced into the corresponding associated vectors.

$$f_{\mathrm{mol1}}(x) = \sum_{i=1}^{m} \begin{pmatrix} w_S \cdot \mathrm{steric} \\ w_E \cdot \mathrm{elec} \\ w_L \cdot \mathrm{lipo} \\ w_D \cdot H_{\mathrm{don}} \\ w_A \cdot H_{\mathrm{acc}} \end{pmatrix}_i^{\mathrm{mol1}} e^{-2\alpha \| x - x_i^{\mathrm{mol1}} \|^2},$$

$$f_{\mathrm{mol2}}(x) = \sum_{j=1}^{n} \begin{pmatrix} w_S \cdot \mathrm{steric} \\ w_E \cdot \mathrm{elec} \\ w_L \cdot \mathrm{lipo} \\ w_D \cdot H_{\mathrm{don}} \\ w_A \cdot H_{\mathrm{acc}} \end{pmatrix}_j^{\mathrm{mol2}} e^{-2\alpha \| x - x_i^{\mathrm{mol2}} \|^2},$$

$$A_F = \int \langle f_{\mathrm{mol1}}(x) | f_{mol2}(x) \rangle dx =$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \left\langle \begin{pmatrix} w_S \cdot \mathrm{steric} \\ w_E \cdot \mathrm{elec} \\ w_L \cdot \mathrm{lipo} \\ w_D \cdot H_{\mathrm{don}} \\ w_A \cdot H_{\mathrm{acc}} \end{pmatrix}^{\mathrm{mol1}} \middle| \begin{pmatrix} w_S \cdot \mathrm{steric} \\ w_E \cdot \mathrm{elec} \\ w_L \cdot \mathrm{lipo} \\ w_D \cdot H_{\mathrm{don}} \\ w_A \cdot H_{\mathrm{acc}} \end{pmatrix}_j^{\mathrm{mol2}} \right\rangle$$

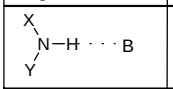
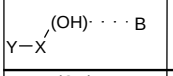
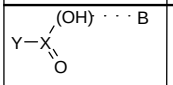
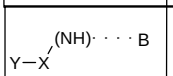
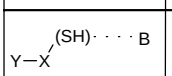
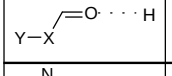
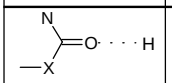
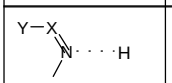
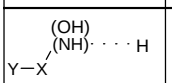
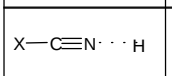
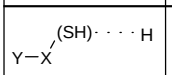
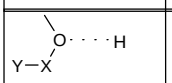
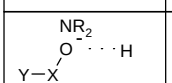
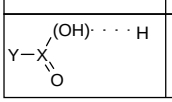
$$\cdot e^{-\alpha \| x_i^{\mathrm{mol1}} - x_j^{mol2} \|^2} \cdot C,$$

$$C = \int e^{-4\alpha \| x - \frac{1}{2}(x_i^{\mathrm{mol1}} + x_j^{\mathrm{mol2}}) \|^2} \mathrm{dx} = \mathrm{const.}$$

To evaluate the similarity of two molecules in space, the integral of the scalar product of the associated functions is computed [19, 22, 27]. The obtained similarity score A_F is maximized by adjusting the alignment of the two molecules. Starting from a set of pre-selected orientations, it is subsequently optimized by minimizing the mutual distances between molecular portions having similar physicochemical properties and hydrogen-bonding features. This method does not require predefined pairs of matching centers associated with the molecular framework. Accordingly, also strongly deviating bonding skeletons can be compared and aligned.

This evaluation of the scalar products also considers deviations in the hydrogen-bond directionalities

Table 5. Generic types of H-bond donors and acceptors. Representative spatial positions have been defined at the population maxima of the corresponding composite crystal-field environments. They are given by the internal coordinates d (in Å), a (in °) and t (in °) based on the sketched labeling. X and Y are neighboring atoms, mostly C, however in some cases also P, S, N, or O are allowed (see below)

fragment	d	a	t
	H···B 1.9	N–H···B 180	X–N–H···B 0
	O···B 2.8	X–O···B 109	Y–X–O···B +/-60, 180
	O···B 2.8	X–O···B 120	Y–X–O···B 0, 180
	N···B 2.9	X–N···B 109	Y–X–N···B +/-60, 180
	S···B 3.4	X–S···B 109	Y–X–S···B +/-60, 180
	O···H 1.8	C=O···H 120	X–C=O···H 0, +/-90, 180
	O···H 1.9	C=O···H 120, 180	X–C=O···H 0, +/-90, 180
	N···H 1.9	X=N···H 120	Y–X=N···H 180
	O···H N···H 1.9	X–O···H X–N···H 109	Y–X–O···H Y–X–N···H +/-60, 180
	N···H 1.8	C≡N···H 180	X–C≡N···H 180
	S···H 2.4	X–S···H 109	Y–X–N···H +/-60, 180
	O···H 1.9	X–O···H 109	Y–X–O···H +/-90
	N···H O···H 1.9	X–N···H X–O···H 109	Y–X–N···H Y–X–O···H +/-60, 180
	O···H 1.9	X–O···H 120	Y–X–O···H 0, +/-90, 180

of different functional groups being compared. However, by analyzing their influence on the obtained alignments we noticed that only the consideration of directionalities toward a donor group really improves the alignment. In consequence, only for these groups the directionality is considered in the approach (see above).

In order to improve on the speed of the algorithm we express the original steric distribution produced by the sum of atomic Gaussian functions loaded with the steric weights in a different way. At first, the maximum of the atom-based steric distribution is determined. Then, at the position of this maximum a new Gaussian function with a steric weight, equivalent to the value of the maximum is produced. The usual attenuation $\alpha = 0.2$ is applied. Subsequently, the difference between the original atom-based distribution and the newly generated Gaussian function is calculated. The origin of the latter function is modified so that the integrated square difference between the original atom-based and the approximated distribution is minimal. At the maximum of the remaining difference density the next Gaussian function is placed, again with a weight corresponding to the maximum at this position. This procedure of determining the difference density and adding new Gaussian functions is followed until the residual difference density falls below a predefined threshold. This threshold was set to a value not larger than 50% of the steric density produced by a Gaussian function attributed to a carbon atom. Finally, after finishing the placement procedure, the weights and positions of all Gaussian functions are modified so that the integrated square difference between the original atom-based and the approximated distribution is minimal. For the optimization steps a quasi-Newton minimizer is used. Figure 4 gives the steric density distribution of the pteridine moiety of methotrexate either by the original set of atom-based Gaussians (top) and the fewer 'floating' Gaussians (bottom). For the upper case in total 19 atom-based Gaussians were used, for the lower one only 6 functions were required. Both densities are drawn at the same contour level and show convincing similarity.

Since the newly generated Gaussian functions are no longer atom-based, electrostatic and lipophilic properties have to be assigned in a different way. In this step, the electrostatic and lipophilic weights attributed to the new Gaussian functions are calculated so that the integrated square difference between the original atom-based and approximated distribution of these properties is a minimum. The putative hydrogen-bonding sites are treated independently. To each such position one Gaussian function is associated.

Table 6. Test set of 28 protein ligands analyzed by MIMUMBA (molecule given by its PDB code [17], rmsd from experimental reference in Å, number of rotatable bonds, total number of considered conformations in the generated set, rank of best corresponding conformer, CPU time (min:s) for the complete analysis)

Molecule	Rmsd	Bonds	Rmsd/bond	Total conform.	Best conform.	CPU time
1cbx	0.42	5	0.09	39	10	0:10.09
1dhf	0.83	9	0.08	150	66	4:33.27
1gpy	0.54	3	0.08	5	3	0:06.74
1stp	0.39	5	0.08	150	12	0:28.21
1thl	1.42	11	0.13	150	100	6:21.69
1tlp	1.51	12	0.13	150	79	6:22.84
1tmn	1.19	13	0.09	150	75	6:15.90
2tmn	0.29	5	0.06	69	5	0:08.82
3cpa	0.50	5	0.10	36	9	0:05.36
3gpb	1.13	3	0.19	36	1	0:08.22
3tapap	0.48	8	0.06	150	21	2:12.83
3tmn	1.07	6	0.18	54	27	0:07.25
4dfr	1.04	9	0.11	150	85	5:08.57
4tapap	0.55	8	0.07	141	11	0:59.54
4tln	0.74	3	0.25	12	7	0:03.54
4tmn	1.00	12	0.09	150	68	8:36.11
5gpb	0.52	3	0.08	58	13	0:11.08
5tln	1.08	8	0.12	150	93	2:47.46
5tmn	1.03	11	0.09	150	32	9:39.09
6cpa	1.48	12	0.12	150	136	8:29.27
7cpa	1.51	13	0.12	150	79	11:14.17
8cpa	1.67	12	0.14	150	148	9:12.73
1dwd	0.43	10	0.04	150	22	6:08.91
cbz	0.63	7	0.09	94	25	0:15.39
1dwc	1.32	10	0.13	150	6	7:04.35
ppp	0.94	7	0.13	38	8	0:07.65
rthior	0.45	6	0.08	97	3	0:15.29
thior	0.52	6	0.09	82	5	0:12.51

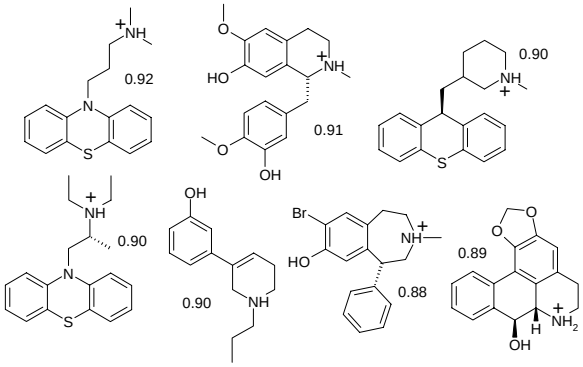
In the new version of the program the steric description of a molecule is expressed by about 50% fewer Gaussian functions compared to the original atom-based assignment. Since during the alignment procedure all pairwise combinations of Gaussian functions of the two molecules have to be integrated, a significant speed-up is achieved.

A further speed-up of the search has been achieved by a more sophisticated selection of appropriate starting positions for the similarity optimization. In the original version of the program these positions were selected by random. The new strategy considers a pre-alignment in terms of lipophilic and H-bonding properties of the two molecules. In a first step, the description by Gaussian functions of the two molecules is generated. Then the program creates a bipartite graph whose vertices correspond to center pairs of Gaussian functions that are both loaded by either lipophilic or putative hydrogen-bond donor or acceptor properties. The first center of each pair originates from the reference molecule, the second one stems from the test molecule. To each vertex a weight is associated:

$$\mathrm{vxwgt(i)} = W_{LH} * (w_R(i) * w_T(i))^{0.5}$$

where W_{LH} is a general factor to distinguish between lipophilic and H-bonding vertices; w_R and w_T denote the individual weights for H-bonding or lipophilicity of the paired centers in the reference and test molecule. Two vertices are connected by an edge if their

Table 7. Best hits from a search for compounds structurally similar to apomorphine (with relative similarity score)



mutual distance in the reference and test molecule are comparable. The edges are sorted according to their differences in distance. Only the best corresponding edges are considered further. As a maximum, 10 edges per vertex are allowed, however not more than 1000 in total. The pre-alignment routine seeks for the completely connected set of vertices (cliques) in this graph. Each clique that consists of at least four vertices gives rise to a potential starting orientation. The latter is produced by a rigid fit of the centers of the Gaussians in the test molecule onto their counterparts in the reference. The obtained orientation is rejected if the centers of mass of the two molecules are significantly apart in space. Furthermore, the centers of corresponding Gaussians in the reference and test molecule must fall close to each other. As a condition, the ratio between the fit deviation (rmsd) and clique weight has to diminish below a predefined threshold. The clique weight is the sum of its vertex weights. Orientations passing this criterion are sorted according to their clique weights. The orientations with the highest clique weights are submitted to a full optimization as described above. We limit the number of starting orientations to:

$$\text{trials} \leq 5 + (T_M - 5) * N_v/1000$$

with N_v denoting the number of vertices in the graph and T_M being the maximal number of trials allowed by the user (default: 25).

The described introduction of the hydrogen-bonding properties and the modified expression by fewer Gaussians required a new calibration of the various adjustable parameters w. Using a data set of 190 ligand pairs mutually binding to the same protein, these parameters were optimized and validated similarly to our previous study [22]. The newly obtained values are $w_{steric} = 0.12$, $w_{electrostatic} = 100$, $w_{lipo} = 10$, $w_{donor} = 30$, and $w_{accept} = 30$. Across this reference set, the observed alignments could be reproduced in 40% of the cases with an rms-deviation below 0.8 Å and 65% were found below 1.6 Å. Considering the inherent accuracy limits of about 0.7 Å for such a superposition of two experimentally determined protein-ligand complexes, the observed agreements are rather convincing. Since the alignment function exhibits several minima, the approach generates multiple superpositions, each with a particular similarity scoring. For 56% of the test cases, the solution with the highest similarity scoring and an rmsd below 1.2 Å also approximates best the experimentally observed alignment. Considering the first and second solution of the investigated cases, this figure increases to 64%. However, since in 36% of the tested examples, the approach cannot come-up with a solution below rmsd < 1.2 Å that is ranked among the best three cases, there is still room to improve, either in the superposition geometry or the similarity scoring.

The different solutions of the approach suggest alternative binding modes. For example, the 'reverse' binding mode of the two antiviral compounds with very similar bonding skeletons are proposed since the best solution corresponds to the 'regular' superposition [29, 30]. However, the second best solution, ranked only 4% less in similarity, approximates correctly the experimentally observed binding mode.

Local conformational optimization of the spatial superposition

The alignment procedure described so far does not consider molecular flexibility. In order to reflect some 'local' flexibility in the superposition process, the alignment function mentioned above has been introduced as additional term into the potential function used in the optimization step of MIMUMBA [6]. In addition to our initially reported version [22] we now included the hydrogen-bonding properties in this simultaneous comparison.

This local optimization step requires an initial orientation and start conformation resulting from the above described conformational analysis and subsequent rigid alignment. The pre-aligned molecules are minimized considering individually their 'MIMUMBA force-fields' and at the same time their similarity in space is optimized. To speed-up the MI-

MUMBA force-field all hydrogen atoms are merged with their neighboring heavy atoms following the formalism of usual united-atom approaches [31]. The associated partial charges and lipophilicities are merged in the same way as described in the previous section.

Generation of candidate molecules

The goal of the present approach is an automatic screening of a large sample of candidate molecules with a given reference structure. The candidate molecules have been collected from (a) examples of peptidomimetics described in literature [32], (b) entries from the CSD [3] being flagged as 'drug molecules' and (c) entries from the Available Chemical Directory (ACD) [33] or the World Drug Index (WDI) [34]. The peptidomimetics were model-built using SYBYL [11]. 3D geometries have been generated for the ACD and WDI entries by CORINA [10]. For the CSD entries the crystallographic coordinates have been used. Atom and bond types were assigned with SYBYL and after adding missing hydrogens an optimization of bond distances and angles has been performed. Partial charges were computed using the AM1 method [20]. In the next step up to 50 conformers were generated using MIMUMBA. For each conformer atom-based lipophilicities [21] and putative hydrogen-bonding sites were computed. Their steric shape was approximated by sets of Gaussian functions and the associated property vectors were calculated. The entire data have been stored in binary format and can be used as input for the subsequent alignment studies.

Consecutive steps toward a flexible alignment

The different algorithms described above have been combined into an automatic procedure. First of all a reference compound with given conformation is required. This conformation has to be available from a different study. It can either result from a detailed QSAR analysis, from a docking experiment, or from a synthetic rigidization program that finally reveals only one (or few) possible binding site conformation(s). In principle, the approach described in the following can operate on different conformations of the reference compound. However, the computational effort increases linearly with the number of considered conformations of the reference molecule.

Similar to the multiple conformers of the candidate molecules, partial charges, atom-based lipophilicities and putative hydrogen-bonding sites are calculated for the reference. In a next step the various conformers of the candidate molecules are compared and aligned with the reference. For each superposition the similarity score is determined and stored. To make the approach available for virtual screening, not only different conformers and superpositions of one single test molecule with a reference have to be compared but a whole variety of test molecules with different sizes will be scored against the reference. To allow for a consistent mutual similarity measure among all candidate molecules, some kind of a 'normalized' scoring index is required. This A_F^{norm} is used as defined previously [22]. All values fall into the interval between zero and one.

$$A_F^{norm} = \frac{A_F}{\left[\left(A_{F_{candidate}}^{self}\right)\left(A_{F_{reference}}^{self}\right)\right]^{1/2}}$$

After all conformers of the candidate molecules have been aligned and scored the obtained scoring indices are sorted. The highest ranked alignments are further evaluated by optimizing simultaneously intramolecular flexibility and intermolecular similarity (s. above). The top ranked alignments yielded by this final step are stored and can be further analyzed by computer graphics.

Implementation of a parallel version on a PC cluster

The procedure to compute the rigid alignment has been ported to a cluster of three LINUX-Pentium PCs (100 MHz) connected via ethernet. The non-parallelized version of the program runs about as fast as on a SGI R4000 Indy. To implement the program in parallel on all processors simultaneously, the parallelization software MPI [35] has been used. The starting orientations within the alignment procedure were distributed evenly among the processors. This yielded a speed-up by a factor of 2.6 compared to the run-time on a single processor.

A case study: Screening for putative dopamine ligands

As a case study for virtual ligand screening a search for putative dopamine ligands has been performed.

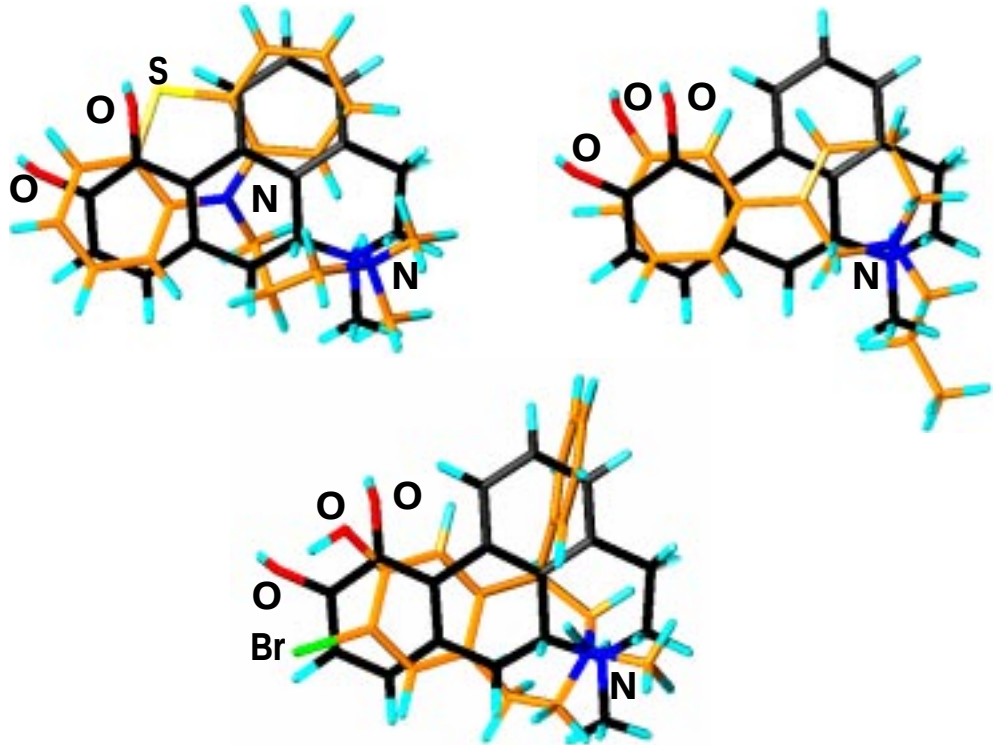


Figure 5. Obtained superposition from a virtual screening for dopaminergic compounds (upper left: phenothiazine, upper right: tetrahydropyridine, bottom: benzazepine, chemical formulae see Table 7) retrieved from the WDI using apomorphine (black) as a structural reference.

Since the search results from a pair-wise fitting of candidate molecules onto a given reference of known 3D structure a potent rigid ligand binding to the dopamine receptor had to be selected. Apomorphine as potent antagonist serves this purpose. In order to check whether the approach is capable to retrieve molecules actually known to bind to the dopamine receptor, 28 examples from a study of Martin et al. [23] were included in the set of candidate molecules. Furthermore, this set was complemented by a random selection of 658 molecules taken from the World Drug Index [34]. For all compounds of the test set the procedure described in the 'generation of candidate molecules' section was applied.

After superimposing and ranking multiple conformers of the various candidate molecules, 50% of the known dopamine ligands were found with a similarity score A_F of more than 0.80 with respect to apomorphine. The first hits were derivatives of apomorphine itself with different substituents at the two aromatic moieties. The subsequent hits from the dopaminergic series were examples with a structurally reduced apomorphine skeleton. Those of the dopaminergic examples ranked beyond 0.6 possess quite a different skeleton compared to apomorphine (e.g. zopiclone). More interesting are hits from WDI, since some putative new leads could be suggested. In Table 7, the chemical formulae of first seven best-scored solutions are sketched. Closer inspections of these structures reveals that some of them are already known for their binding to the dopamine receptor (e.g. three phenothiazine derivatives, a thioxanthene, a benzazepine, a tetrahydro isoquinoline, a substituted tetrahydropyridine or an apomorphine derivative, described as neuroleptics or Antiparkinson agents). Figure 5 shows the actual superpositions onto apomorphine (black) of these best solutions retrieved from the WDI. The typical pattern known for dopaminergic compounds can be detected, perhaps some additional substituents (e.g. OH) might be required in some of the examples. The total screening procedure required 50:27 h on an SGI workstation with an R4000 processor.

Summary and conclusions

In the present paper several methodological improvements have been described to make molecular superposition methods applicable to virtual ligand screening. To consider efficiently molecular flexibility a comprehensive torsional library derived from crystal data has been developed. This library is used in the rule-based conformational search program MIMUMBA. It follows a combinatorial strategy in the angle selection step. For molecules with several rotatable bonds this approach can easily result in a combinatorial explosion. Accordingly a more sophisticated procedure to select the angular settings has be implemented. For the molecular comparison a strategy is followed considering shape-associated physicochemical properties in space. Our former approach using steric, electrostatic and lipophilic properties has been extended by hydrogen-bonding features. To speed up computing, molecular shape is now approximated by fewer Gaussian functions, not necessarily located at the atomic positions but at maxima of the (residual) density. Furthermore, a more sophisticated selection of starting positions for similarity optimization helps to enhance computational efficiency. The comparison is performed in two steps, first by a global alignment search operating on multiple rigid conformers and subsequently by conformationally relaxing the best scored hits from the global search. The method has been implemented on a cluster of parallel processors. To demonstrate the potential of the approach for lead finding a search for putative dopamine receptor ligands has been performed. This example retrieves from a sample set of structures a subset known to bind to this receptor among the best scored hits. In addition some interesting proposals are suggested. The approach is reasonably fast, accordingly a search in data sets of several thousand candidate molecules appears feasible. Further examples showing the potential of the approach have already been published by us [18, 36].

Acknowledgements

Part of the present work has been funded by the BMBF projects RELIWE and ÄBAV. The offered financial support is gratefully acknowledged.

References

1. Brandt, D.W., Drug Discov. Today, 3 (1998) 61.
2. Blaney, J.M. and Dixon, J.S., Perspect. Drug Discov. Design, 1 (1993) 301; Colman, P.M., Curr. Opin. Struct. Biol., 4 (1994) 868; Kuntz, I.D., Science, 257 (1992) 1078; Lengauer, T. and Rarey, M., Curr. Opin. Struct. Biol., 6 (1996) 402; Lybrand, T.P., Curr. Opin. Struct. Biol., 5 (1995) 224.
3. Downs, G.M. and Willet, P., In 'Reviews in Computational Chemistry' Vol. 7, pp. 1–66, Lipkowitz, K. and Boyd, D. (Eds), VCH; Good, A.C. and Mason, J.S., In 'Reviews in Computational Chemistry' Vol. 7, pp. 67–118, Lipkowitz, K. and Boyd, D. (Eds), VCH; Bures, M.G., Practical Application of Computer-Aided Drug Design, pp. 39–72, Marcel Dekker, New York, NY, USA (1997).
4. Hahn, M., J. Chem. Inf. Comput. Sci., 37 (1997) 80; Perry, N.C. and v. Geerestein, V.J., J. Chem. Inf. Comput. Sci., 32 (1992) 607; Parretti, M.F., Kroemer, R.T., Rothmann, J.H. and Richards, W.G., J. Comput. Chem., 18 (1997) 1344; McMartin, C. and Bohacek, R.S., J. Comput.-Aided Mol. Design, 9 (1995) 237; Kato, Y., Inoue, A., Yamada, M., Tomioka, N. and Itai, A., J. Comput.-Aided Mol. Design, 6 (1992) 475.
5. Perkins, T.D.J., Mills, J.E.J. and Dean, P.M., J. Comput.-Aided Mol. Design, 9 (1995) 479; Hurst, T., J. Chem. Inf. Comput. Sci., 34 (1994) 190; Thorner, D.A., Wild, D.J., Willett, P. and Wright, P.M., J. Chem. Inf. Comput. Sci., 36 (1996) 900; Jones, G., Willett, P. and Glen, R.C., J. Comput.-Aided Mol. Design, 9 (1995) 532.
6. Klebe, G. and Mietzner, T., J. Comput.-Aided Mol. Design, 8 (1994) 583.
7. Allen, F.H. and Kennard, O., Acc. Chem. Res., 16 (1983) 146.
8. Klebe, G., in 'Structure Correlation', Eds. Bürgi, H.B. and Dunitz, J.D., VCH (1994) p. 543.
9. DeClerq, P.J., Hoflack, J. and Cauwbergh, S., QCPE Program No. QCMP079.
10. Sadowski, J., Rudolph, C. and Gasteiger, J., Tetrahedron Comput. Methodol., 3 (1990) 537.
11. Program SYBYL from Tripos Ass., Inc, St. Louis, MO, USA.
12. The authors are grateful to Dr. G. Vriend (EMBL, Heidelberg, Germany) for providing the data of tripeptide fragments.
13. Vriend, G., J. Mol. Graph., 8 (1990) 52.
14. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rogers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
15. The following small molecule ligands have been used (labeled by their Refcodes in the PDB):1cbx, 1dhf, 1dwc, 1dwd, 1dwe, 1eed, 1ent, 1gpy, 1hiv, 1hos, 1ivp, 1ivq, 1pph, 1stp, 1thl, 1tlp, 1tmn, 1tpp, 2er7, 2hat, 2mip, 2r04, 2r06, 2r07, 2rm2, 2rr1, 2rs1, 2rs3, 2rs5, 2tmn, 3cpa, 3gpb, 3tmn, 4dfr, 4er1, 4er2, 4gpb, 4hvp, 4phv, 4tapap, 4tln, 4tmn, 5er1, 5er2, 5gpb, 5hvp, 5tln, 5tmn, 6cpa, 7cpa, 7hvp, 8cpa, 8hvp, 9hvp, lysozyme: α-, β- N-acetylglucosamide, thermolysin: β-ppp, cbzphe, thiorphan, retro-thiorphan.
16. Clark, M., Cramer, III, R.D. and v. Opdenbosch, N., J. Comput. Chem., 10 (1989) 982.
17. The following subset of ligands from [15] has been used: 1cbx, 1dhf, 1dwd, 1dwe, 1gpy, 1pph,1stp, 1thl, 1tlp, 1tmn, 2tmn, 3cpa, 3gpb, 3tmn, 4dfr, 4tln, 4tmn, 4-tapap, 5gpb, 5tln, 5tmn, 6cpa, 7cpa, 8cpa, β-ppp, cbzphe, thiorphan, retro-thiorphan.
18. Klebe, G., Perspect. Drug Discov. Design, 3 (1995) 85.
19. Kearsley, S.K. and Smith, G.M., Tetrahed. Comput. Meth., 3 (1990) 615.
20. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.

21. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
22. Klebe, G., Mietzner, T. and Weber, F., J. Comput.-Aided Mol. Design, 8 (1994) 751.
23. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P., J. Comput.-Aided Mol. Design, 7 (1993) 83.
24. Klebe, G., J. Mol. Biol., 237 (1994) 212.
25. Bruno, I.J., Cole, J.C., Lommerse, J.P.M., Rowland, R.S., Taylor, R. and Verdonk, M.L., J. Comput.-Aided Mol. Design, 11 (1997) 525.
26. Böhm, H.J., Brode, S., Hesse, U. and Klebe, G., Chem. Eur. J., 2 (1996) 1509.
27. Good, A.C. and Richards, W.G., J. Chem. Inf. Comput. Sci., 33 (1993) 112.
28. Mestres, J., Rohrer, D.C. and Maggiora, G.M., J. Comput. Chem., 18 (1997) 934; Mestres, J., Rohrer, D.C. and Maggiora, G.M., J. Mol. Graphics, 15 (1997) 114; Leherte, L., Latour, T. and Vercauteren, D., J. Comput.-Aided Mol. Design, 10 (1996) 55; Grant, J.A., Gallardo, M.A. and Pickup, B.T., J. Comput. Chem., 17 (1996) 1653; Lemmen, C. and Lengauer, T., J. Comput.-Aided Mol. Design, 11 (1997) 357; Lemmen, C., Hiller, C. and Lengauer, T., J. Comput.-Aided Mol. Design, 12 (1998) 491; Nissink, J.W.M., Verdonk, M.L., Kroon, J., Mietzner, T. and Klebe, G., J. Comput. Chem., 18 (1997) 638.
29. Badger, J., Minor, I., Kremer, M.J., Oliveira, M.A., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.D., Dutko, F.J., Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. USA, 85 (1988) 3304.
30. Klebe, G., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design, ESCOM, Leiden, 1993, pp. 173–199.
31. Gerber, P.R. and Müller, K., J. Comput.-Aided Mol. Design, 9 (1995) 251.
32. Ujvary, I., BIOSTER, available through Synopsys Scientific Systems Ltd., 175 Woodhouse Lane, Leeds, UK.
33. The Available Chemicals Directory is distributed through Molecular Design Ltd., San Leandros, CA (USA).
34. World Drug Index; Version 2/96, Derwent Information, 1996.
35. MPI: A Message Passing Interface Standard, Message Passing Interface Forum, April 15, 1994, Univ. of Tennessee, Knoxville, TN, USA.
36. Böhm, H.J. and Klebe, G., Angew. Chem., Int. Ed. Engl., 35 (1996) 2588.