Journal of Computer-Aided Molecular Design, 11 (1997) 53–60
ESCOM

J-CAMD 378

A hybrid approach for addressing ring flexibility in 3D database searching

Jens Sadowski*

Bayer AG, PH-R Structural Research, Aprather Weg, D-42096 Wuppertal, Germany

Received 28 May 1996
Accepted 27 August 1996

Keywords: Flexible 3D searching; Directed tweak; Ring flexibility; UNITY; CORINA; NCI database

Summary

A hybrid approach for flexible 3D database searching is presented that addresses the problem of ring flexibility. It combines the explicit storage of up to 25 multiple conformations of rings, with up to eight atoms, generated by the 3D structure generator CORINA with the power of a torsional fitting technique implemented in the 3D database system UNITY. A comparison with the original UNITY approach, using a database with about 130 000 entries and five different pharmacophore queries, was performed. The hybrid approach scored, on an average, 10–20% more hits than the reference run. Moreover, specific problems with unrealistic hit geometries produced by the original approach can be excluded. In addition, the influence of the maximum number of ring conformations per molecule was investigated. An optimal number of 10 conformations per molecule is recommended.

Introduction

In the pharmaceutical industry, searching large structural databases for structures matching precise 3D pharmacophore properties has become a standard method in the process of drug design [1] and a number of new lead structures found by this technique have already been reported [2–4].

The state of the art in 3D database searching is an adequate way of addressing the flexibility of molecules. Flexibility can be introduced in three principal ways: (i) by storing explicitly multiple conformations in the 3D database; (ii) by applying 3D screens based on multiple conformations [5,6]; and/or (iii) by a torsional fitting technique which 'on the fly' generates a conformation compatible with the query. The first two approaches suffer from a dilemma: a conformation analysis that is too fine requires immense computer resources (CPU time and, for the first approach, also disk space) and a search grid that is too coarse may result in a loss of possible bioactive conformations. Thus, the torsional fitting technique seems to be the more elegant and efficient approach.

Amongst the torsional fitting techniques, the 'directed tweak' method [7,8] has become the most popular and seems to be the most efficient approach [9]. Directed tweak optimizes the torsion angles at rotatable bonds with respect to a penalty function describing the distance of the actual conformation of a molecule from a given pharmacophore. In principle, this technique is also applicable to ring systems. This is desirable since there is evidence that binding to a receptor may induce a conformational change also in rather rigid ring systems of substrate molecules [10]. Thus, without addressing the flexibility of rings some potential hits might be missed. But the application of the tweak technique to rings runs into problems. Firstly, ring systems have only limited numbers of significantly different conformations compared to the flexibility of chain fragments. Thus, a continuous fit technique will not be as good as for chain fragments. Secondly, the handling of ring closure by opening one bond per ring and introducing additional constraints into the penalty function makes the hypersurface of the optimization function much more complicated and is a potential source of numerical problems. Convergence problems and questionable hit geometries may be the result. In general, the

*Present address: BASF AG, ZHV/W-A30, D-67056 Ludwigshafen, Germany.

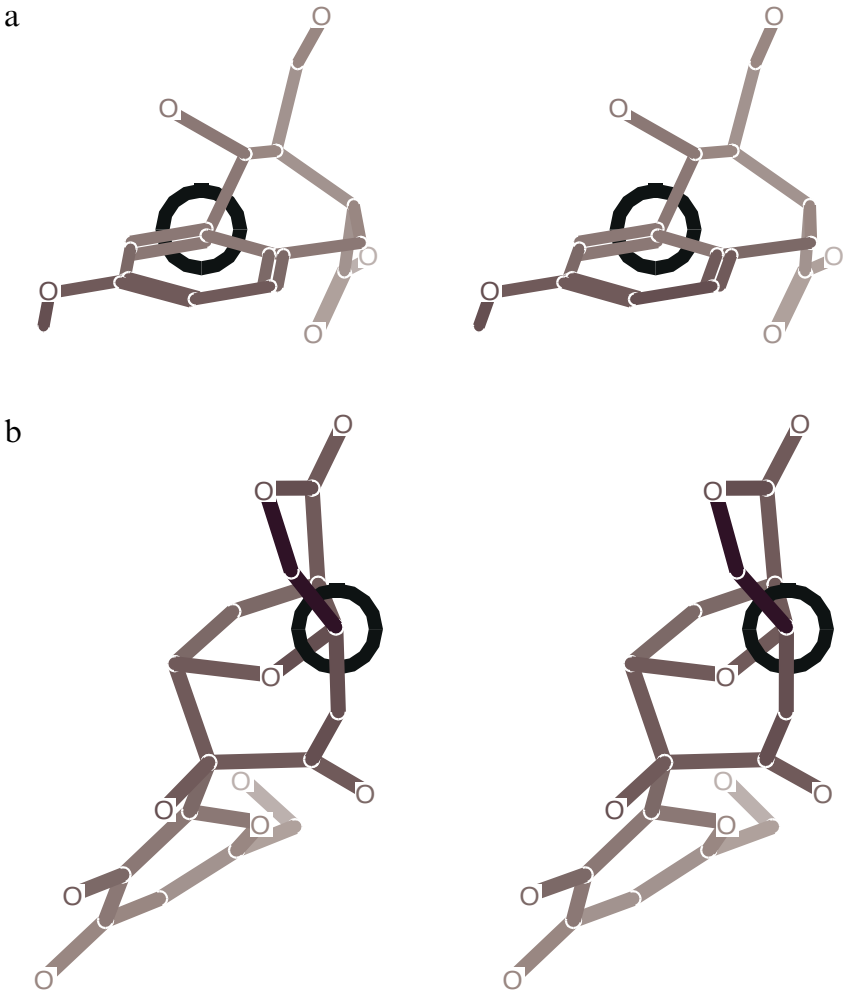

Fig. 1. Stereoplots of two hits from a UNITY ring search with questionable geometries at ring atoms (black circles): (a) nonplanar sp^2 atom; (b) violated tetrahedral stereocenter.

geometries of the hit structures are often rather distorted. Figure 1 shows two symptomatic examples taken from a hit list of a UNITY [11] search with ring flexibility. Common problems are violated sp^2 atoms or stereocenters and atom clashes at the ring closure bonds even when searching with an explicit van der Waals check. The marked atom of the aromatic ring in Fig. 1a is in an extremely unrealistic pyramidal configuration. The marked bridgehead atom of the bicyclo[1.2.3]octane skeleton in Fig. 1b has all four substituents on one side of a plane through itself, and the bridging moiety is inverted and positioned into the ring plane – another extremely unrealistic situation. But this is only the tip of the iceberg since other geometric details like bond lengths, bond angles, and dihedrals are also rather arbitrarily distorted and a long way from a realistic energy level for a receptor-bound conformation.

It is clear that one cannot expect geometries generated by a torsional fitting technique to exactly match an energy minimum of the isolated molecule since the change in internal strain during the fitting process is only rudimentarily taken into account. However, serious violations of the ideal geometry are indicative of a problematic hit.

The specific performance and quality of a flexible ring search has not been studied as comprehensively as the flexible search techniques in general. In this report, an attempt to evaluate this part of the UNITY system [11] will be undertaken. In the Results and Discussion section, it will be shown that a significant proportion of the hits found by a search with ring flexibility are incorrect and disappear when relaxed into the next energy minimum of a force field. These examples indicate that the directed tweak method performs less efficiently and accurately for ring systems. In the following, a simple hybrid approach is presented which addresses the solution of these problems.

Methods

Hybrid approach

The following ideas lead to a hybrid approach that overcomes the above problems of the directed tweak

method. Firstly, the directed tweak performs very efficiently for chain portions of molecules but runs into problems when applied to rings. Secondly, the explicit storage of multiple conformations for addressing flexibility suffers both from the immense requirements for computer resources and from possible losses of bioactive conformations due to a too coarse search grid. But ring systems consisting of small rings (e.g., with up to eight atoms) show only a small number of conformations that represent a coarse grid. Thus, a hybrid approach combining the two techniques can increase the search efficiency. The proposed technique distinguishes between addressing the flexibility of ring portions and chain portions of the molecules:

(1) For chains, the directed tweak method is used.

(2) For rings, multiple conformations are stored in a 3D database.

This method was implemented by combining two commercially available programs: UNITY [11] for the flexible search (directed tweak) and database management, and CORINA [12] for the 2D-to-3D conversion and the generation of multiple ring conformations.

Conformation handling by CORINA

Ring systems represent a special challenge for a 3D structure generator due to the ring closure constraint. CORINA processes rings up to a size of eight atoms by using a table of single ring conformations which implicitly ensure ring closure. In the case of fused or bridged systems, a backtracking procedure finds a contradiction-free set of conformations for each single ring following some geometric and energetic restrictions. Since this strategy works on the torsion angle representations of the ring conformations and uses only logical operations and integer arithmetic, it is extremely fast. The ring conformations are then translated into 3D coordinates, further refined by a simplified force field, and ordered by an energy function that takes into account the torsional energy of the individual rings as well as the Pitzer strain energy caused by exocyclic substituents and the additional strain in fused or bridged ring systems [12a,b]. The list of conformations is completed by additional conformers gained by generating both configurations of 'flappy' ring nitrogens (i.e., uncharged sp^3 nitrogens having one exocyclic substituent). Thus, CORINA performs a partial

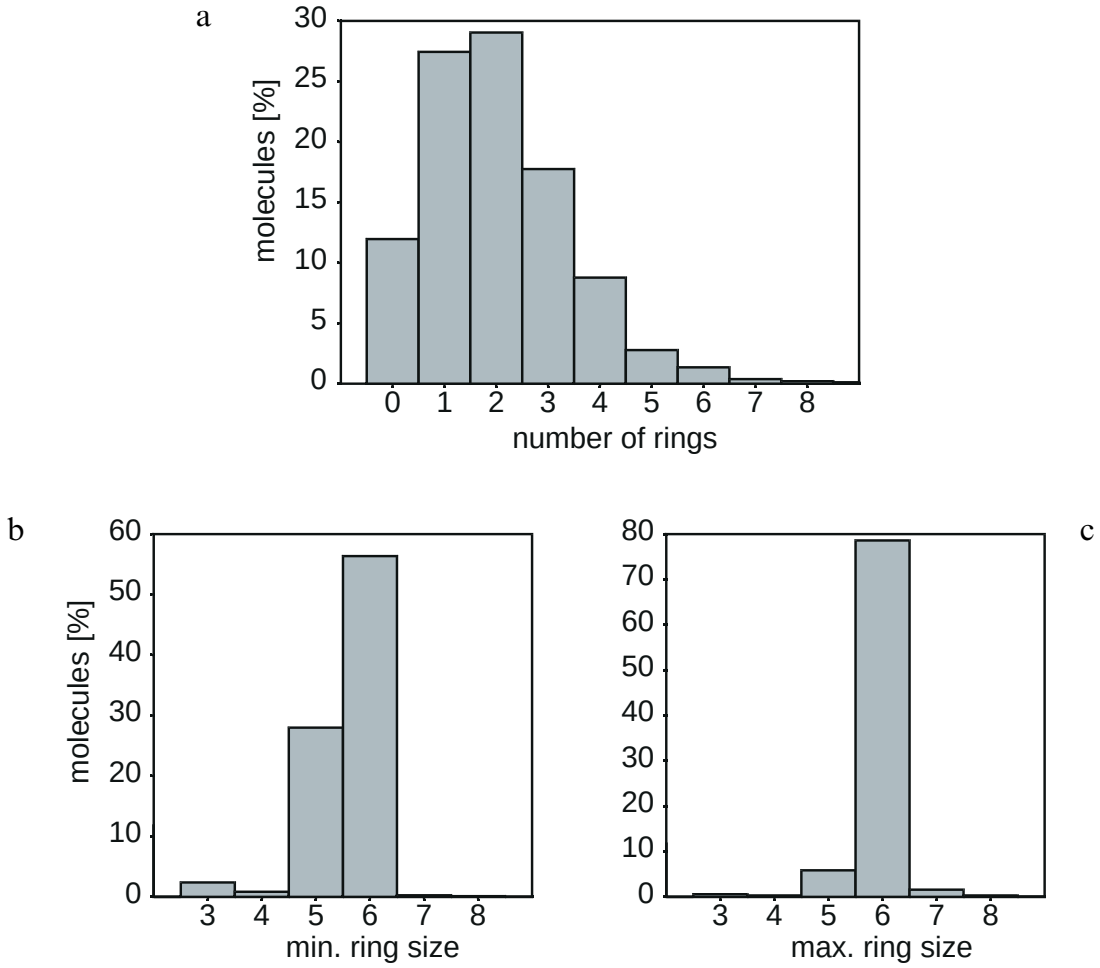

Fig. 2. Ring characteristics of the NCI database: (a) distribution of the number of rings per molecule; (b) distribution of the minimum ring size; (c) distribution of the maximum ring size.

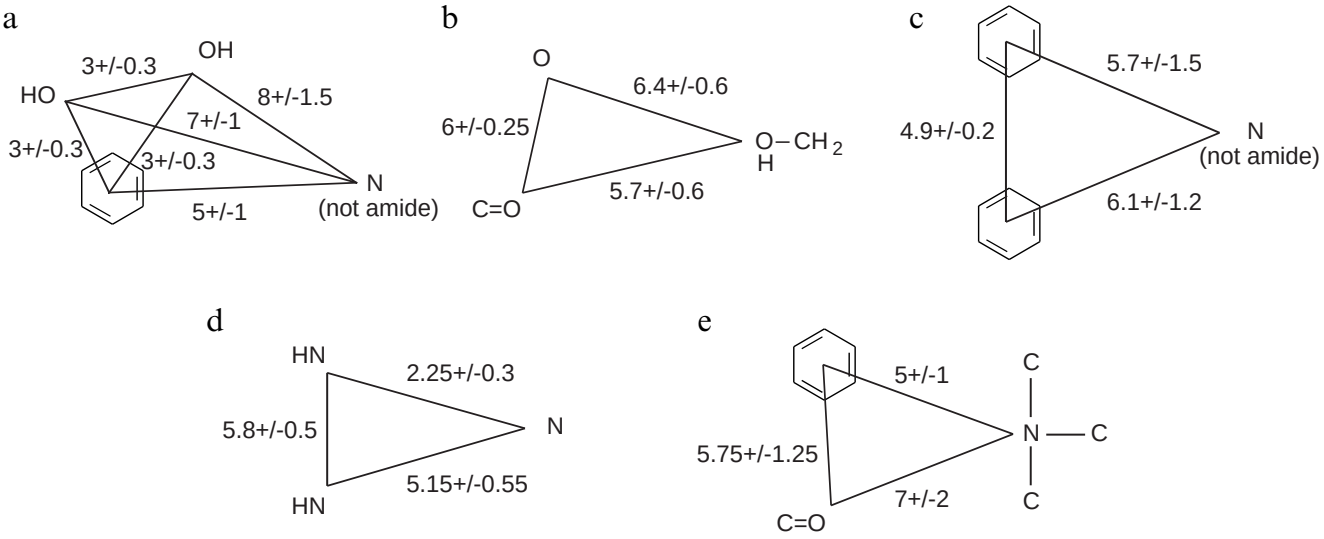

Fig. 3. Definition of queries: (a) dopamine agonist (derived from the dopamine structure); (b) kinase C agonist [2]; (c, d) histamine agonists 1 and 2 [6]; (e) antiarrhythmic agent [8].

conformational analysis restricted to ring systems. For each chain portion, only one suitable conformation is generated. This strategy of the program is comparable to that of other available rule-based 3D structure generators or conformation analysis programs such as, for example, CONCORD [13], COBRA [14], or SCA [15] (for a review see Ref. 12b). It cannot simply be compared to the more general approaches of systematic or stochastic conformation analysis methods (for a general review of conformation analysis methods see Ref. 16). Although the performance of CORINA and a number of other 3D structure generators was evaluated using 639 X-ray structures [12c], the ring conformer generation module of CORINA was not compared specifically to that of other programs. However, the author is not aware of any other available program for the rapid 2D-to-3D conversion of large databases that offers this type of additional feature for handling ring systems.

For molecules having more than one ring system (the term 'ring system' is used for an ensemble of one or more rings having at least one atom in common), which are connected by chains, CORINA offers two techniques:

(1) *The exhaustive method:* All combinations of all conformations of the individual ring systems are generated and ordered according to energy. A possible combinatorial explosion is inherent in this method.

(2) *The compact method:* All ring systems simultaneously change their conformations from the low- to the high-energy levels. All conformations of a particular ring system will be seen, but not all combinations of them. A possible loss of bioactive conformations is the price paid for a significantly smaller number of conformations generated.

The output generated by both techniques can be tailored by restricting the total number of conformations per molecule. These methods are illustrated by using 1-cyclohexyl-2-cyclohexenylethane as an example. For the cyclohexane and cyclohexene rings each having one exocyclic substituent, CORINA generates 14 and four conformations, respectively. Thus, the exhaustive method results in $14 \times 4 = 56$ conformations. The compact method yields $\mathrm{MAX}(14,4) = 14$ conformations.

Of course, these techniques are rather simple. Although the ranking of the conformations by energy is good for a first guess, conformer sampling by geometrical diversity is what really would be needed for 3D database searching. Some authors [17,18] suggest such methods.

Evaluation method

In order to compare the performance of the hybrid technique with alternative approaches, five reference queries were searched in a public domain database. The public part of the NCI database [19] containing 126 705 molecular structures was used as a database for the evaluation. The distribution of rings within the database can be characterized as shown in Fig. 2. Figure 2a gives the distribution of the number of rings per molecule. The maximum is between one and two rings. Twelve percent of the structures contain no rings. Figures 2b and c illustrate the distribution of the maximum and the minimum ring sizes found in a molecule. The distributions are rather narrowly centered around the six-membered rings. Thus, the vast majority of molecules contain small rings having up to seven atoms. This range is covered by CORINA. In addition, the number of flexible ring systems per molecule was investigated. Only 29% of the molecules have at least one flexible ring system, whereas the number of molecules with more than one flexible ring system is even smaller (3%).

Test queries for dopamine agonists (derived from the dopamine structure), kinase C agonists [2], histamine agonists 1 and 2 [6], and antiarrhythmic agents [8] were taken from the literature as shown in Fig. 3.

Two 3D databases were constructed using CORINA version 1.7 and imported into UNITY version 2.4 for the search. One database (NCI) contained one conformation per molecule and the other database (NCI_FLEX) contained multiple ring conformations. The number of conformations per molecule in the NCI_FLEX database was restricted to 25 and the compact method (vide supra) was applied to molecules having more than one flexible ring system. From records consisting of more than one molecule (e.g., salts), all fragments except the largest were removed.

Figure 4 illustrates the evaluation scheme. The five reference queries were searched in both databases and stored in separate hit list files. The one-conformer database was searched with the directed tweak technique with and without ring flexibility. The multi-conformer database was searched without ring flexibility (hybrid approach in the gray frame). All three search runs were performed with an explicit van der Waals check. The hits were subsequently relaxed in a published robust force field [20] into the next local minimum in order to distinguish between unreasonably distorted ring geometries (see the examples in Fig. 1 in the Introduction section) and hits matching the query based on a new ring conformation. The refined hit geometries were stored in separate databases and searched again without ring flexibility, thus revealing the true hits. The results of the one-conformer search without ring flexibility were finally used to figure out the additional amount of hits gained by the alternative approaches for addressing ring flexibility.

Results and Discussion

Comparison of alternative techniques

The above evaluation scheme (Fig. 4) was applied. First, the two 3D databases had to be generated. Table 1 summarizes the results of the two CORINA runs. The increased requirements for computer resources in terms of CPU time and disk space are fairly moderate and do not seem to be problematic.

The flexibility of the rings contained in the database is illustrated by the distribution of the number of conformations per molecule as given in Fig. 5. The majority of molecules do not contain a flexible ring. Only 29% of the molecules in the database have rings with two or more low-energy conformations. Thus, the effect of ring flexi-

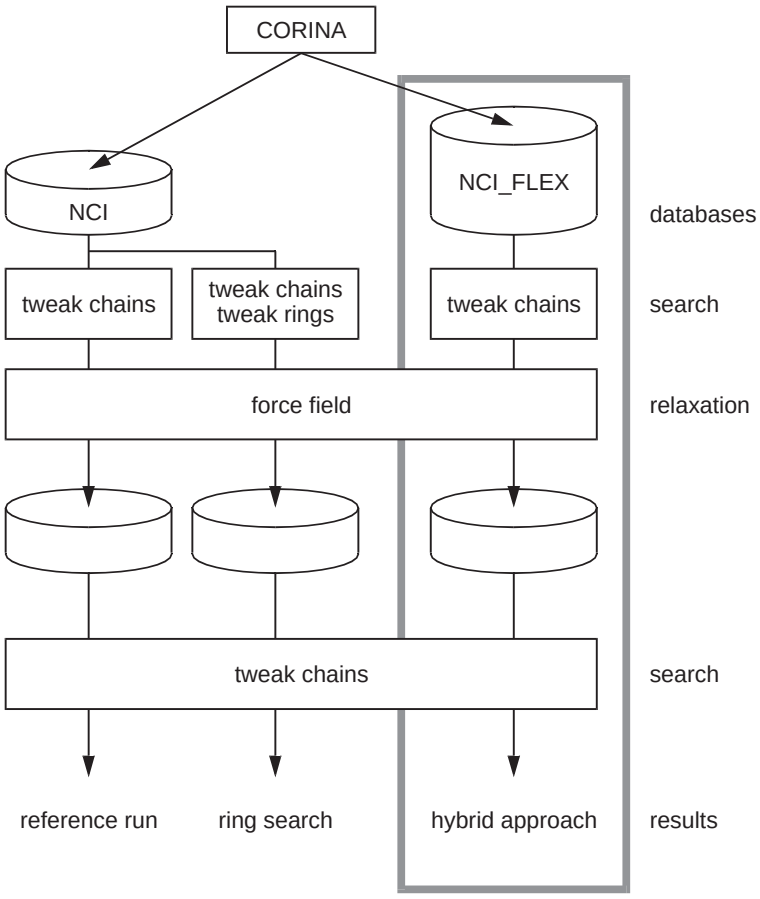

Fig. 4. Evaluation scheme (hybrid approach in the gray frame).

TABLE 1
RESULTS OF THE TWO CORINA RUNS

	NCI	NCI_FLEX
Maximum conformations	1	25
CPU time (s)[a]	28 368	112 962
Conversion rate (%)	99.6	99.6
Number of 3D structures	126 147	372 836
Conformations per molecule	1.0	3.0
UNITY database size (MB)	155	291

[a] On an SGI R8000 under IRIX 6.1.

bility on the hit rates of a flexible search can be expected to be in the same order of magnitude.

Secondly, the five reference queries (Fig. 3) were searched. The results of the three different search runs are given in Table 2. Clearly, the hybrid approach (NCI_FLEX, tweak chains) results in a significant additional number of hits (6–23%) compared to the reference run without ring flexibility. This correlates with the proportion of 29% of molecules in the database having two or more conformations (Fig. 5). The application of the directed tweak method to rings yielded a significantly smaller additional number of hits. Moreover, in one case (histamine 2) the tweaking of rings decreased the hit number by 2%. This indicates that a number of 'stable' hit structures found without ring flexibility in the reference run may also be hidden by the flexible ring search due to an 'instable' hit structure falling in a local minimum during relaxation and being unable to fulfill the query in the subsequent search run without ring flexibility.

Influence of force field relaxation on the hit rates

In order to distinguish between artifacts introduced by the force field relaxation of the hit geometries and artifacts introduced by the flexible search, the hit numbers of all three search runs before and after relaxation were compared (Table 3). Clearly, the hit numbers of the first and the third search runs (i.e., the reference run and the hybrid approach run) were comparably stable with an average loss of hits below 5%. This is the bias brought into the procedure by the force field since the force field of the structure generator and the one used for the optimization are not identical and some of the hit molecules moved slightly out of the query tolerances. On the other hand, the search run with ring flexibility lost a significant proportion of 7–29% of the hits – far higher than the loss found in the reference run. This is a clear indication that the hits found by tweaking the rings are of questionable quality and contain a significant proportion of unrealistic geometries.

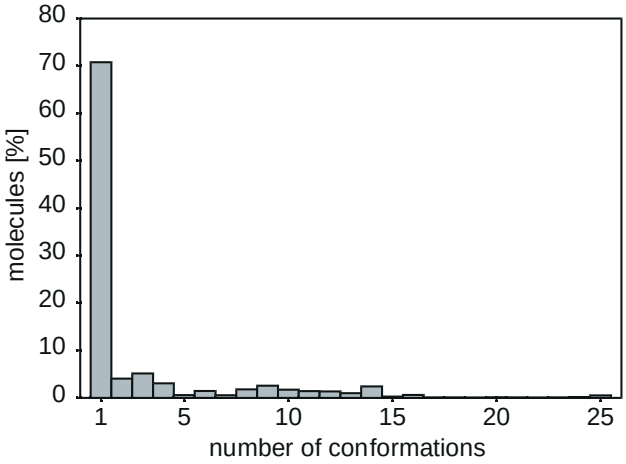

Fig. 5. Distribution of the number of conformations per molecule.

In the histamine 2 case, the number of hits found with a ring search dropped even below the hit count of the reference run. This might be an indication that the ring search also hides true hits found without ring flexibility as discussed above. Moreover, there is a significant number – up to 14% – of hits found without ring flexibility which do not appear in the hit list of the ring search even before the force field relaxation (data not shown). This was not seen for the hit lists of the hybrid approach. This observation was not investigated in detail but it might be another indication of the uncertainty brought in by tweaking the rings.

Investigating the optimal number of conformations

In an additional test sequence, the influence of the maximal number of conformations with respect to performance criteria such as hit rate, database size, and search time was investigated. Different hybrid search runs were performed on databases with a maximum of 1, 5, 10, 15, 20, and 25 conformations per molecule. Figure 6 shows plots of the hit rate, the database size, and the

TABLE 2
HIT NUMBERS OF THE THREE SEARCH RUNS

Query	NCI tweak chains	NCI tweak chains tweak rings	NCI_FLEX tweak chains (hybrid approach)
Dopamine	117	122 (+4%)	144 (+23%)
Kinase C	490	553 (+13%)	601 (+23%)
Histamine 1	3736	3999 (+7%)	4247 (+14%)
Histamine 2	1932	1885 (–2%)	2050 (+6%)
Antiarrhythmic	1180	1197 (+1%)	1441 (+22%)

TABLE 3
HIT NUMBERS OF THE THREE SEARCH RUNS BEFORE AND AFTER THE FORCE FIELD RELAXATION

Query	NCI tweak chains		NCI tweak chains tweak rings		NCI_FLEX tweak chains (hybrid approach)	
	Before	After	Before	After	Before	After
Dopamine	118	117	141	122	149	144
Kinase C	511	490	776	553	630	601
Histamine 1	3796	3736	4308	3999	4338	4247
Histamine 2	1985	1932	2064	1885	2123	2050
Antiarrhythmic	1207	1180	1329	1197	1497	1441

search time versus the maximum number of conformations per molecule for the five reference queries (in percent relative to the one-conformer database). As expected, the three plots correlate strongly. In most cases, the hit rates converged at a value of 10 conformations. At that point, the database size reached about 170% and the search times increased by a factor of 3–6 depending on the query.

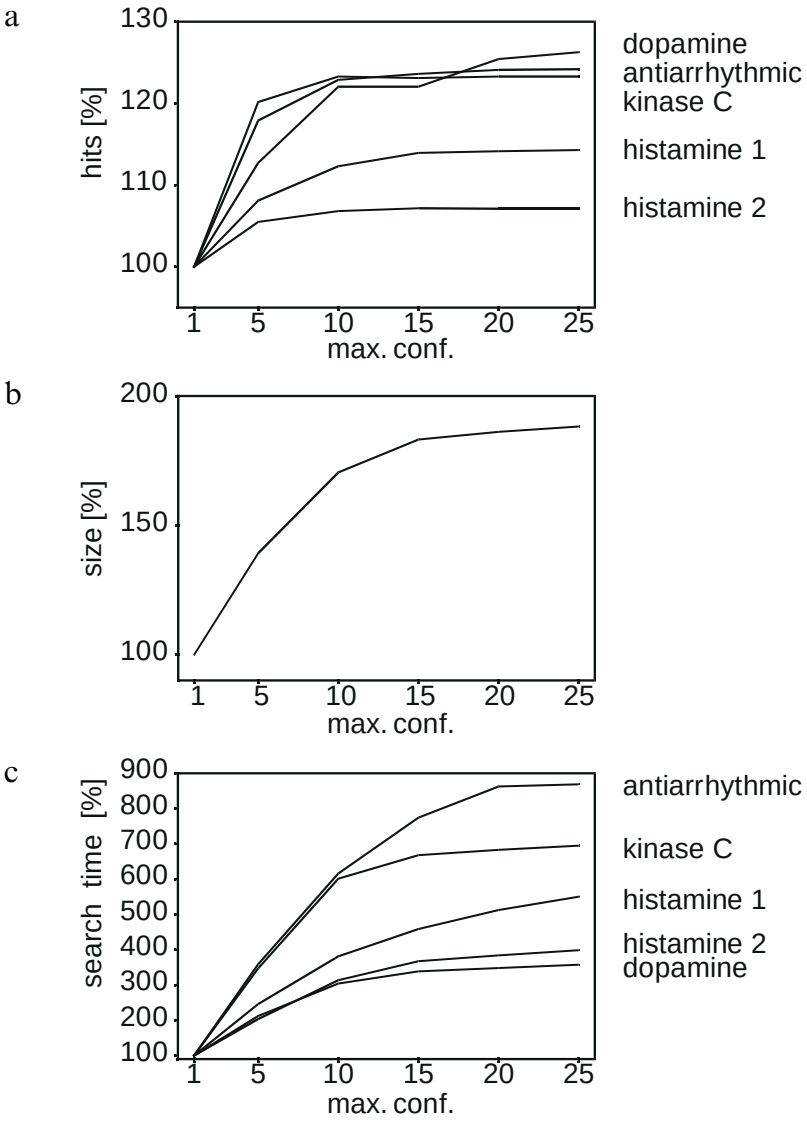

Fig. 6. Investigation of the role of the maximum number of conformations per molecule: (a) hit rate versus maximum conformations; (b) database size versus maximum conformations; (c) search time versus maximum conformations.

Conclusions

A new hybrid approach for flexible 3D database searching has been presented. It combines the strengths of two available techniques for 3D database handling: CORINA [12] for the 2D-to-3D conversion and the generation of multiple ring conformations, and UNITY [11] for the flexible fitting of chain torsions. An evaluation study that was performed by searching five reference queries in a large public domain database with 126 705 molecules showed that

(1) the 'directed tweak' technique is less suitable for ring searches due to a loss of hits and due to unrealistic hit geometries, and

(2) the hybrid approach (CORINA + UNITY) increases the hit rates of flexible 3D database searches significantly (on an average by 10–20%).

A reasonable compromise between hit rate and computer resources can be recommended with 10 ring conformations per molecule. In this case, the hit rates came close to their maximum. The database size increased by a factor of 1.7 and the increase of the search times varied between 3 and 6 times depending on the query compared with the reference run without ring flexibility. Compared to the run with explicit tweaking of rings, the search times increased by factors of 1.5–3.6 (data not shown). For a particular query, the user should decide whether the additional hits found by the hybrid approach justify this additional amount of computer resources.

As a consequence of this study, the structural database of Bayer Pharma Research has been converted into a UNITY database with up to 10 ring conformations per molecule for future pharmacophore search purposes in order to enhance the strength of this drug design method.

Acknowledgements

The assistance of the Munich office of Tripos Associates in questions concerning the use of UNITY is gratefully acknowledged.

References

1 a. Martin, Y.C., Bures, M.G. and Willett, P., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 1, VCH, New York, NY, U.S.A., 1990, pp. 213–263.
b. Good, A.C. and Mason, J.S., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 7, VCH, New York, NY, U.S.A., 1996, pp. 67–117.
2 Wang, S., Zaharevitz, D.W., Sharma, R., Marquez, V.E., Lewin, N.E., Du, L., Blumberg, P.M. and Milne, G.W.A., J. Med. Chem., 37 (1994) 4479.
3 Lam, P.Y., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C.-H., Weber, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S.E., Science, 263 (1994) 380.
4 Kiyama, R., Honma, T., Hayashi, K., Ogawa, M., Hara, M., Fujimoto, M. and Fujishita, T., J. Med. Chem., 38 (1995) 2728.
5 Davies, K. and Upton, R., Tetrahedron Comput. Methodol., 3 (1990) 665.
6 Good, A.C. and Kuntz, I.D., J. Comput.-Aided Mol. Design, 9 (1995) 373.
7 Hurst, T., J. Chem. Inf. Comput. Sci., 34 (1994) 190.
8 Mook, T.E., Henry, D.R., Ozkabak, A.G. and Alamgir, M., J. Chem. Inf. Comput. Sci., 34 (1994) 184.
9 Clark, D.E., Jones, G. and Willett, P., J. Chem. Inf. Comput. Sci., 34 (1994) 197.
10 Furet, P., Caravatti, G., Lydon, N., Priestle, J.P., Sowadski, J.M., Trinks, U. and Traxler, P., J. Comput.-Aided Mol. Design, 9 (1995) 465.
11 UNITY (1995), developed and distributed by Tripos Associates, St. Louis, MO, U.S.A.
12 a. Gasteiger, J., Rudolph, C. and Sadowski, J., Tetrahedron Comput. Methodol., 3 (1990) 537.
b. Sadowski, J. and Gasteiger, J., Chem. Rev., 93 (1993) 2567.
c. Sadowski, J., Gasteiger, J. and Klebe, G., J. Chem. Inf. Comput. Sci., 34 (1994) 1000.
d. CORINA is available from Prof. Dr. J. Gasteiger, Computer-Chemie-Centrum, University of Erlangen-Nürnberg, Nägelsbachstrasse 25, D-91052 Erlangen, Germany.
13 CONCORD (1995) is available from Tripos Associates, St. Louis, MO, U.S.A.
14 COBRA (1995) is available from Oxford Molecular Ltd., Oxford, U.K.
15 De Clerq, P.J., Hoflack, J. and Cauwberghs, S., QCPE Program No. QCMP079, Bloomington, IN, U.S.A.
16 Leach, A., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 2, VCH, New York, NY, U.S.A., 1991, pp. 1–55.
17 Leach, A., J. Chem. Inf. Comput. Sci., 34 (1994) 661.
18 a. Smellie, A., Kahn, S.D. and Teig, S.L., J. Chem. Inf. Comput. Sci., 34 (1995) 285.
b. Smellie, A., Teig, S.L. and Towbin, P., J. Comput. Chem., 16 (1995) 171.
19 Milne, G.W.A., Nicklaus, M.C., Driscoll, J.S., Wang, S. and Zaharevitz, D.W., J. Chem. Inf. Comput. Sci., 34 (1994) 1219.
20 Davies, E.K. and Murrall, N.W., Comput. Chem., 13 (1989) 149.