Drug Discovery Today: Technologies
Vol. 7, No. 4 2010

3D Pharmacophore Elucidation and Virtual Screening

Conformations and 3D pharmacophore searching

Christof H. Schwab*

Molecular Networks GmbH, Henkestr. 91, 91052 Erlangen, Germany

Several methods have been developed and published over the past years to generate sets of diverse and pharmacologically relevant conformations which can be used within 3D pharmacophore search protocols to increase the number of meaningful hits of such experiments. This review gives some insights into the general challenges and problems in the area of 3D structure and conformation generation and focuses on some available and recent software technologies and approaches applicable for this task. The methods, algorithms and philosophies behind the approaches are briefly described and discussed and some examples on the performance and results obtained with the different tools are given.

Section editor:
Gerhard Wolber – Freie Universität Berlin, Germany

Introduction

Most molecules of pharmacological interest can adopt more than one conformation of nearly equal energy by rotation around single bonds. 3D pharmacophores are very sensitive toward the data they are searched against, that is the three-dimensional structures of the database molecules. Therefore, the success of a 3D pharmacophore search experiment heavily relies not only on the quality and accuracy, but also on the conformational diversity of the 3D structures stored in the database. In addition, a single 3D geometry of a molecule may miss a pharmacophore, although, it is able to exhibit the appropriate conformation. To avoid such false negative hits and to ensure that all molecules that are searched are present in their *bioactive* conformation, that is, the preferred conformation in the receptor-bound state, most pharmacophore methods employ sets of molecular conformations – conformational ensembles – as well as a conformation generation tool that calculates the different molecular geometries either before or on-the-fly during the search. By contrast, too many conformations will not only increase search and computation times, but, and even more importantly, may dramatically increase the number of false positive hits.

Brief historical background and classification of methods

The major goal of any conformation generation tool that is employed in drug design is to generate and identify the bioactive conformation of a molecule within a reasonable amount of time. Doubtlessly, this is not possible by generating a single 3D structure. Therefore, conformational ensembles that are ideally biased toward the conformational space that is considered to contain the bioactive conformation or at least geometries that are similar to it have to be calculated. Clearly, the less conformations have to be generated, the faster the pharmacophore search is the better the ratio of true to false positive hits is.

A large number of different approaches have been developed over the past decades that generate 3D structures and conformational ensembles starting already in the late eighties of the past century. Reviews about these developments until the early 2000s can be found at [1–4].

*Corresponding author:. C.H. Schwab (schwab@molecular-networks.com)

1740-6749/$ DOI: 10.1016/j.ddtec.2010.10.003

In general, conformer generators can be classified into rule- and data-based systems, systematic (or grid) searches, numerical methods, random approaches and genetic algorithms (not discussed here). However, most of the more recent technologies can be regarded as hybrid approaches, as they employ combinations of the methods named above.

Description of problem

What is so special about the bioactive conformation? The major issue is that during binding to the biological receptor the molecule undergoes a transition from the unbound, 'free' status in aqueous solution into a bound status exposed to directed electrostatic and steric forces by the amino acids of the binding site. Enthalpic and entropic contributions, such as the loss of water molecules, may stabilize the bound structure in a different geometry that the ligand exhibits in solution, in solid state, or in gas phase. Studies have shown that ligands often bind in more extended conformations to maximize the interaction surface with the receptor [5]. Discussions if receptor-bound ligands have energy contents significantly above the global energy minima [6,7] or, as a recent study proposes, if bioactive conformations are closer to energy minima [8] are matter of an ongoing scientific debate and a thorough discussion here would break the limitations of this article.

Another challenge is the fact of molecules that bind to more than one biological target with acceptable binding affinities (sometimes also called 'promiscuous ligands'). The ligand-bound conformations at the various targets may significantly differ, but each pharmacophore toward the molecule that is active should produce a hit. Therefore, a thorough sampling of the conformational space including all biologically relevant geometries is a crucial step for successful pharmacophore-based virtual screening.

There are several additional issues related to conformational sampling that are frequently discussed in the literature, but only two should be mentioned here. One question is which methods and means are adequate for testing and evaluating the performance of conformer generation tools (including reproduction of the bioactive conformations as well as coverage of the conformational space) and another topic is the preparation of useful test datasets for such test runs (see, e.g. [9,10]).

General workflow of a conformational search

The general workflow of a conformational search procedure can be described as following.

Identification of the rotatable bonds in the molecule

Usually, only single bonds between heavy atoms are considered as rotatable and often single bonds to terminal groups (e.g. a methyl group) are excluded. Owing to the ring closure restriction, cyclic systems are orders of magnitudes more rigid than the acyclic (open-chain) parts and have to be treated separately, for example, by pre-calculated, allowed ring conformations. Most program systems determine a set of rotatable bonds and flexible rings purely on the basis of the molecular graph.

Generation of conformations with the implemented algorithm

At this step, the molecule is expanded into conformational space starting from the connection table information either with or without 3D coordinates (starting geometry). Bond lengths and angles are usually kept fixed and only the tetrahedral (or torsion) angles of the rotatable bonds are varied. Especially systematic searches have to deal with the challenge of the combinatorial explosion at this step. The total number of possible conformations N that can be generated in systematic searches increases exponentially with the number of rotatable bonds n supposed that for each rotatable bond a number k of torsion angles (increments) is applied around a full rotation of 360° (see Eqn (1)).

$$N = \left(\frac{360}{k}\right)^n \tag{1}$$

Methods and algorithms that identify and prune pathways in the search tree that lead to unfavorable conformations (e.g. highly strained ring conformations or atom clashes) at an early stage speed up the search dramatically. In addition, intermediate energy or symmetry consideration can reduce the number of conformations that have to be generated and further processed. Optionally, the generated 'raw' conformations have to be energy-minimized and geometry-optimized at this step (e.g. by a force field).

Checking for duplicates and very similar geometries

Redundant information, such as identical conformation or those that are very similar have to be identified and removed. This can be done by the calculation of inter-conformational distances either in the Cartesian space (comparison of the 3D coordinates of the heavy atoms of two conformations) or by comparing internal coordinates such as the torsion angle values. The usually applied metric for this comparison is the RMSD value (root mean square deviation, see Eqn (2)) which can also be used to calculate the similarity between a generated conformation and an experimental structure, for example, the bioactive conformation (Fig. 1).

$$\mathrm{RMSD} = \sqrt{\frac{\sum_i^N (\Delta d)^2}{N}} \tag{2}$$

where N is the number of heavy atoms (non-hydrogen atoms) and Δd is the distance between the ith corresponding atom pair of the conformations. Δd can be calculated either in Cartesian space that is, the relative positions of the 3D coordinates between corresponding atom pairs are used, or

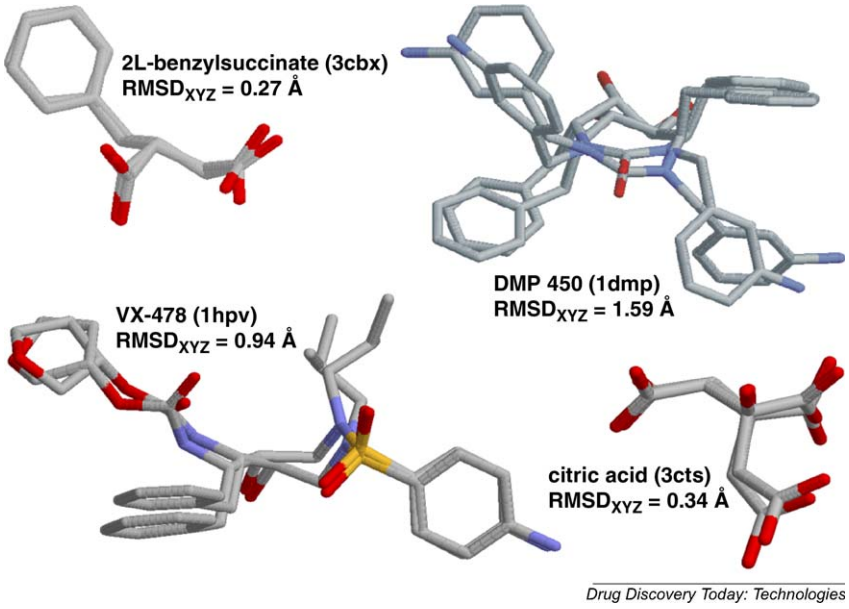

Figure 1. Four examples of superimpositions of the experimentally determined bioactive conformation with the best-matching conformation obtained from an automatic conformational search. The RMSD$_{XYZ}$ (root mean square deviation in Cartesian space in [Å]) as well as the PDB code of the receptor–ligand complex is shown.

by comparing internal coordinates, such as the torsion angle values of corresponding bonds.

Selection of a set of representative conformations biased toward area of application

This step can also be embedded and performed in one of the steps above. The major goal of an exhaustive search in conformational space is to retrieve all possible conformations of a molecule, for example, to identify the global minimum. This exhaustive but computationally demanding search should also retrieve the bioactive conformation. However, in pharmacophore-based screening the goal is to rapidly generate a relatively small conformational ensemble around the ligand-bound conformation to be able to screen databases of millions of compounds in reasonable amounts of time. Therefore, a subsampling of the entire accessible conformational space is usually performed.

Available technologies

CatConf, or **ConFirm**, from Accelrys (http://accelrys.com) that is part of the Discovery Studio pharmacophore modeling protocols and tools (formerly available in the Catalyst pharmacophore modeling platform) provides two different search modes, *fast* and *best*, which significantly differ in their methods and implementations [11,12]. The *fast* mode applies a modified systematic search (quasi-exhaustive search) with a fuzzy grid, if atom clashes occur, for the open chain portions of a molecule whereas for the ring systems a set of predefined ring conformations is used (ring templates). In the next step, the conformations are relaxed in a restricted CHARMm force field [13] that has been modified to prevent the generation of duplicate geometries and only conformations within a (user-) defined energy window compared to the lowest energy conformation are submitted to the subsequent step. Finally, heuristics are applied to maximize the conformational diversity of the subset of conformations and the user can optionally limit the maximum number of output conformations. The *best* mode of ConFirm uses a distance geometry approach, a relatively time-consuming procedure that can be regarded as a numerical method. Distance geometry uses a matrix of internal coordinates of a molecule (bond lengths and higher distances) with lower and upper bounds. After procedures that are called *triangle smoothing* and (*energy*) *embedding*, the method is able to generate a set of conformations with 'raw' 3D coordinates; however, ring and acyclic portions of the molecule can be treated together in one step. For refinement, again a modified version of the CHARMm force field is employed. A method called poling [14] is applied that biases the sampling of conformations toward geometries that are far from a local minimum but energetically near to each other by placing huge, artificial barriers (poles) on the energy surface. This technique enables the exploration of the low-energy regions of the conformational space and produces conformations that do not correspond to a local energy minimum. The *best* mode of ConFirm is able to well reproduce the ligand-bound conformation of molecules while computation times are significantly higher than in *fast* mode. Some performance results of ConFirm in study [15] are summarized in Table 1.

Table 1. Performance of conformer generators

Program	ConFirm/ Catalyst	CEASAR	MOE	CORINA/ROTATE	OMEGA	MacroModel	ConfGen	CONFORT
Version	4.11	(Discovery Studio 1.7)	2006.08	1.6	2.3	7.0	2.0	3.9
Mode	Fast/best	–	Conformation import	Automatic	Default	AMBER, GB/SA	Very fast-comprehensive	Intermediate energy filter
Dataset	778 drug-like molecules from PDB	918 drug-like molecules from PDB	256 drug-like molecules from PDB	778 drug-like molecules from PDB	197 drug-like molecules from PDB	32 drug-like molecules from PDB	667 drug-like molecules from PDB	32 drug-like molecules from PDB
Mean RMSD to bioactive conformation (Å)	1.18/0.95	0.95	–	0.97	0.67	0.50	–	0.86
Reproduced bioactive conformations (RMSD $\leq$1.0 Å) (%)	29.0/38.7	61.0	68.0-87.0	62.5	83	91.0	52.0–71.0	66.0
Average number of generated conformers per molecule	35/165	100 (maximum)	46-1175	125.5	123.3	154	14.3–146.4	23
CPU times (s/mol)	1.5/155	0.2	6.0	6.5	2.1	1129	0.5–8.0	369
Hardware platform, operating system	x86 Linux PC, 2.8 GHz	x86 Linux PC, 3.4 GHz	PC, 2.2 GHz	x86 Linux PC, 2.8 GHz	x86 Linux PC, 2.8 GHz	SGI IRIX 6.5, R10000, 250 MHz	x86 Linux PC, 2.4 GHz	SGI IRIX 6.5, R10000, 250 MHz
Reference	[15]	[16]	[18]	[21]	[9]	[25]	[26]	[25]

Comment: The statistics on conformation search studies carried out with the different program systems discussed in this review and listed in this table are collected from the individual publications (see References above). Please note that most of the studies have been performed with different datasets, with different analysis protocols and on different hardware platforms. This table is intended to provide the reader with some general information about the performance of conformation generation tools and not for a direct comparison of the performance of the different program systems listed in this table. It is highly recommended to read the original publications for a complete overview and further details of the individual studies.

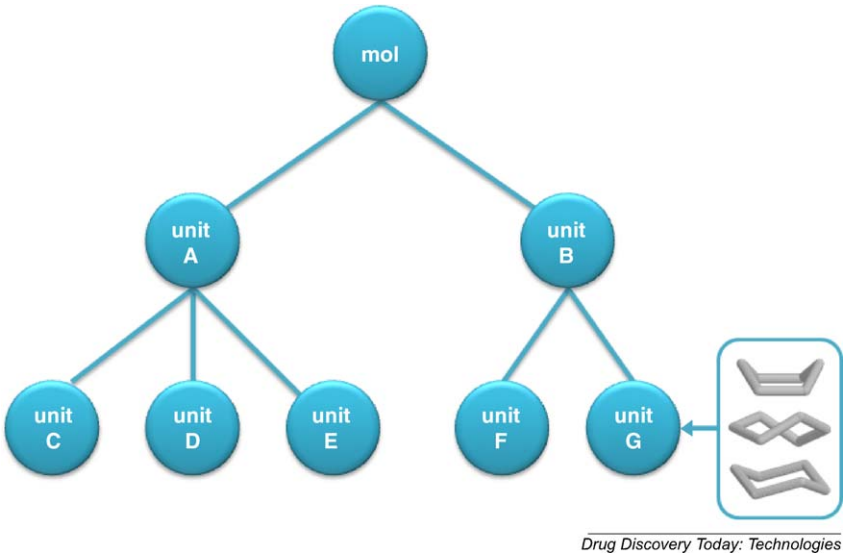

Figure 2. Tree representation of a hypothetical molecule as used in CAESAR. The root node 'mol' represents the entire molecule that is partitioned into conformational units 'A' and 'B' (nodes) of approximately equal structural complexity that are connected by rotatable bonds (edges). Then, units 'A' and 'B' are further dissected into the smallest possible units 'C' to 'G'. Conformational templates are assigned to each node to build up conformational ensemble (e.g. the terminal unit 'G' is a single ring system and different ring conformations are assigned to it).

CEASAR (Conformer Algorithm based on Energy Screening and Recursive Buildup) [16] that is available for conformation generation in the Accelrys Discovery Studio product can be regarded as a systematic search. Internally, the program represents a molecule as a tree where the nodes are the smallest conformational units the molecule can be fragmented into (such as individual ring systems or small rigid fragments, for example, a methylene or ethylene fragment) and the edges are rotatable bonds linking the units (Fig. 2). This representation is based purely on the molecular graph (no 3D information yet). To expand this tree representation into 3D space and to generate conformations, lists of conformational templates are assigned to the nodes (conformational units). For ring systems, the template library method from Catalyst is used (*vide supra*). The small rigid fragments can also be used to generate stereoisomers. Then, the conformations are assembled recursively according to the tree representation starting from the smallest fragment. The fragments (nodes) are linked (edges) by applying a grid of possible torsion angles values (between six and 12 values). During this stepwise build-up procedure, precursors of the final molecular conformations are already checked if a certain energy threshold or a certain number of intermediate conformations is exceeded to control the generation process. Energy calculations are performed by using the parameters of the Catalyst force field. Furthermore, a new method that takes into account symmetries in the molecule prevents the generation of duplicate conformations at early stages of the build-up process (symmetry unique torsions). CEASAR performs up to 20 times faster than the *fast* mode of ConFirm while performing slightly better in reproducing bioactive conformations. Some results of conformational searches with CEASAR from the original paper are summarized in Table 1 [16].

The modeling suite **Molecular Operating Environment**, **MOE**, from Chemical Computing Group (http://www.chemcomp.com) contains modules for systematic and stochastic searches [17] as well as a protocol called Conformational Import. The systematic search applies a grid search by rotating each rotatable bond by a fixed angle increment (15° for cyclic, 60° or 120° for acyclic bonds). After checking for atom clashes and removal of structures with too close van-der-Waals contacts, the conformations can optionally be energy-minimized. In a final step, only conformations within a certain energy window (with respect to the lowest energy conformation) are retained and duplicate geometries (based on RMSD) are rejected. Systematic searches are usually slow; however, perform well if a complete sampling of the conformational space is required. The stochastic search generates a set of conformations by repetitively and randomly changing the torsion angles of the rotatable bonds (cyclic and acyclic portions, perturbation of around 30°) and the algorithm is similar to the RIPS (*R*andom *I*ncremental *P*ulse *S*earch) method [17]. After an energy minimization step, only conformations within a certain energy window and which are not too similar are kept. The entire search process stops if a certain number of cycles to generate conformations are reached or if a certain number of conformations have been generated. These types of stop criterion are typical for stochastic (or random) methods. The stochastic method in MOE is fast, but the default parameters have to be adjusted by the user to sample the pharmacologically relevant space according to [18]. The Conformational Import protocol is a

combination of a rule- and data-based system and a stochastic search and is recommended for generating multi-conformations for large datasets (e.g. for 3D pharmacophore searching). Optionally, the molecules can be first submitted to some filters (e.g. consistent protonation state, removal of molecules with reactive groups, and physicochemical properties). Then, each molecule is fragmented according to some rules and matching fragment conformations from a pre-calculated 3D template library are used to build up different geometries. If a certain fragment cannot be found in the template library, conformations are generated during run time by the stochastic search described above. By joining the different 3D templates, a set of conformations is obtained for the molecule under investigation. Furthermore, the conformations can be optionally relaxed in a force field (refinement option). The Conformation Import protocol is well suited for processing large amounts of molecules. Interestingly, the reproduction of bioactive conformation performs better without the final refinement step in study [18], the results of which are briefly summarized in Table 1.

The program systems **CORINA** [19] and **ROTATE** [20] available from Molecular Networks (http://www.molecular-networks.com) are rule- and data-based approaches. CORINA is widely used for generating high-quality, low-energy, single 3D structures (e.g. of corporate databases or huge compound collections). To identify a single low-energy conformation, CORINA internally generates a set of different ring geometries of the cyclic parts of a molecule based on a library of generic ring templates including strain energies (Fig. 3) which are finally relaxed in a reduced, but extremely fast, force field. These sets of low-energy ring conformations can optionally be output. The open-chain portions are extended as much as possible by setting the torsion angles to *anti* or *trans* configurations, unless a *cis* double bond is specified which effectively minimizes non-bonding interactions. In a following step, ROTATE can vary the torsion angles of the acyclic portions. The program uses a set of rules that has been derived by a statistical analysis of the conformational preferences of small molecule crystal structure taken from the Cambridge Structural Database. Finally, similar conformations (in Cartesian or torsion angle space) are grouped to families and each family is represented by a single conformation to reduce the amount of output of conformations. ROTATE offers a fully automatic mode, but all parameters can also be adjusted by the user. Both programs perform reliably and robust and are able to well reproduce the bioactive conformations; however, the user has to work with two program systems to cover the cyclic and acyclic portions of a molecule. Results of the performance of CORINA and ROTATE are given in Table 1 [21].

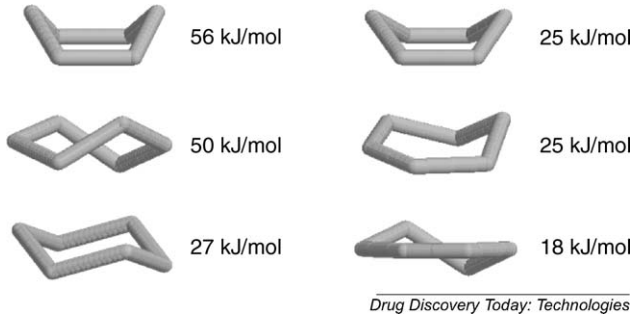

Figure 3. Pre-defined generic ring conformations as used in CORINA for a saturated six-membered ring system (left) and a six-membered ring system with one double bond (right) sorted by their symbolic strain energy. The structures on the left side represent the well-known chair, twisted and boat conformation of six-membered rings (from bottom up).

OMEGA from OpenEye Scientific Software (http://www.eyesopen.com) is a fast systematic, knowledge-based method [9]. The program mainly consists of five steps of which the first two have to be performed only once unless the knowledge-base has to be changed. The first step (*fragment database preparation*) fragments a very large collection of commercially available compounds into ring systems and small open-chain linkers and for each of these fragments at least one conformation is generated using an approach similar to distance geometry (embedding) with a subsequent minimization in a reduced Merck molecular force field (MMFF94) [22]. In the second step (*torsion sampling dictionary*), a pre-defined hierarchical set of torsion angle patterns is populated with different torsion angle values (for each pattern) derived from experimental or calculated 3D structures. In the third step (*3D structure generation*) the fragments from the fragment database calculated in the first step that match the query molecule are assembled using geometric and chemical rules. Thus, for each input molecule at least a single conformation can be generated. In the forth step (*torsion driving*), the torsion angle rules calculated in step two that match to the conformations generated in step three are applied to generate larger conformational ensembles for each molecule. Conformations with severe atom clashes and duplicates are removed. In the final step (*sampling*), a subset of all generated conformations is sampled using a scoring and ranking scheme until a fixed number of final conformations is reached or only conformations below a certain threshold of the score are remaining. The scoring is based on atom clashes in the conformations that are assessed using force field calculations (without a minimization step). OMEGA performs fast and samples the conformational space well around bioactive conformations. Table 1 briefly summarizes the results of a study carried out with OMEGA version 2.3 [9].

MacroModel from Schrödinger (http://www.schrodinger.com) is a very well-known and widely used force field-based molecular modeling package and offers a variety of methods for conformation generation [23]. Besides systematic and stochastic searches, the so-called low-mode conformational

search (LMCS) and large-scale low mode (LLMOD) are available in MacroModel [24]. These effective and accurate algorithms explore the conformational space by applying a (vibrational) mode-following or eigenvector-following technique. In addition, the program includes hydration and solvent accessible surface area models. Geometry optimization can be carried out with a variety of different force field packages, such as MM2, MM3, MMFFs AMBER, AMBER94 and many more. MacroModel can accurately reproduce bioactive conformations, especially if solvation models are applied during the calculations. However, computation times are too high to employ it in virtual screening experiments with large compound collections. In Table 1 some results of a conformational search study with MacroModel are given [25].

More recently, scientists of Schrödinger presented a new method called **ConfGen** [26]. The core algorithm of ConfGen can be regarded as rule- and data-based and was originally developed to generate diverse conformations for the docking program Glide [27]. In principle, the program performs three steps to generate ensembles of conformations. In the first step (*variable feature identification*), rotatable bonds in acyclic portions of the molecule and ring systems are determined. For ring systems, ConfGen uses a set of over 1200 precalculated templates including their relative energies that have been generated by conformational searches using MacroModel. Several criteria (e.g. an upper limit of the relative energy of a ring template within the molecule) limit the total number of generated ring conformations. In the second step (*conformer generation*), acyclic bonds that are considered to be rotatable are treated as individual torsion angle patterns. For each of these patterns a tabulated potential is calculated using a truncated version of OPLS_2001 including some corrections due to molecular symmetries or too few minima. The minima of the resulting potentials are then used to generate the conformations, whereas for the rotatable bonds in the inner part of the molecule (molecular core) all combinations of minima are used and only the lowest energy geometry is applied to the peripheral groups (i.e. the terminal groups including one rotatable bond). After an energy and a steric clash filter, the so-called extension score (ES) is calculated to rank the generated core conformations according to their extendedness. This procedure biases the set of final conformations toward more extended conformations following the finding that ligands tend to bind to their biological receptor in a more stretched geometry (*vide supra*). Finally, the peripheral groups are sampled again (either one at a time, *rapid mode*, or all combinations, *thorough mode*). The final step (*conformer selection and refinement*) includes some filters to remove conformations with unwanted electrostatic properties, polar contacts or conformations that have a high local concentration of heavy atoms. A user-defined number of conformations is sampled from the ES-ranked list of conformations generated so far which can then be optimized

Table 2. Comparison summary table

Technology type	Systematic/ numerical	Systematic	Systematic/stochastic/ rule- and data-based	Rule- and data-based	Systematic, knowledge-based	Systematic/stochastic/ numerical	Rule- and data-based	–
Specific technologies, companies and company web sites	ConFirm (Catalyst), Accelrys, http://www.accelrys.com	CEASAR, Accelrys, http://www.accelrys.com	MOE, Chemical Computing Group, http://www.chemcomp.com	CORINA/ROTATE, Molecular Networks, http://www.molecular-networks.com	OMEGA, OpenEye Scientific Software, http://www.eyesopen.com	MacroModel, Schrödinger, http://www.schroedinger.com	ConfGen, Schrödinger, http://www.schroedinger.com	CONFORT, Tripos, http://www.tripos.com
Reference	[11,12,14]	[16]	[18]	[19,20]	[9]	[23]	[26]	[29]

in an OPLS_2005 force field [28]. Duplicates or conformations that are too similar (RMSD-based) are rejected. Some statistics of a study performed with ConfGen are summarized in Table 1 [26].

CONFORT was developed by Prof. Pearlman, University of Texas, Austin and is distributed by Tripos (http://www.tripos.com). CONFORT performs an exhaustive conformational analysis of a molecule [29]. Two different search modes either generate a user defined number of conformations or output a set of maximally diverse conformations. The diversity metric is based on inter-conformational distances which circumvents the generation of duplicate structures. The conformations are relaxed and optimized by applying only internal coordinates and analytic gradients as well as by the Tripos force field package (see Table 1 for some results from [25]).

The technologies described above are summarized in Table 2. Of course, this list of technologies is by far not comprehensive and there are many more tools existing, even commercially available or developed by industry or academic groups for in-house use. The package **QXP** for structure-based drug design relies on a random search technique to generate conformations that are finally relaxed in a force field [30]. The program **MIMUMBA** by Klebe and Mietzner from BASF company is a rule and data-based system using crystallographic data (similar to ROTATE, *vide supra*) for exploring the conformational space of small molecules [31,32]. The group of Agrafiotis at Johnson & Johnson, Pharmaceutical R&D developed a conformation generator based on stochastic proximity embedding (SPE) and self-organizing superimposition (SOS) to generate 'raw' 3D coordinates of conformational ensembles that can be finally relaxed in a force field [33]. Griewel *et al.* at the Center of Bioinformatics, University of Hamburg in the group of Prof. Rarey recently presented the **TCG** algorithm (TrixX Conformer Generator) that fragments a molecule into components and starts building up the conformation step-wise from the most central component using pre-computed 3D fragments (components) and dihedral angles including intermediate energy filtering at each step [34].

Furthermore, some open source packages, such as **Open Babel** (http://openbabel.org) or **RDKit** (http://www.rdkit.org), provide means to generate sets of molecular conformations.

Conclusions

A lot of different methods and approaches for the automatic generation of conformations have been developed over the past decades. Although, it is an old discipline with a long tradition, the ongoing activities in the community and discussions in the literature show that this field still is an exciting, important and controversially received topic of interest and research. Looking back into history, conformer generators were originally designed and are still used for the identification of the global minimum of a molecule. However, the rapidly growing availability of computer resources, the explosion of available biological and chemical data and the development of virtual screening techniques, such as pharmacophore searches, led to a paradigm shift, or an extension of the paradigm in this area of research. Relevant conformations that is, a small conformational ensemble comprising a maximum of a couple of dozens of structures around the bioactive one, have to be generated in a minimum amount of time to make such virtual screening experiments meaningful, feasible and successful. Most of the technologies described above perform reasonable well for this task and there are no clear favorites or significantly superior techniques. Probably, the methods implemented in MacroModel and CONFORT should be used preferentially if high-quality, accurate and an exhaustive sampling of the conformational space is required (e.g. if the global minima of a small series of molecules is investigated) due to higher computation times. Also, molecular dynamics calculations that are not discussed in this brief review are well suitable for these problems [35]. Differences in performance as outlined in Table 1 are also due to different test datasets, different protocols to analyze the results and to calculate the statistics of the runs as well as different hardware platforms and CPU speed. A combined study with either a unified evaluation protocol (test dataset, analysis tools and CPU power) or each protocol with all available conformer generators would gain a better and deeper insight into the differences in the performance of the methods and accomplishing such a study is encouraged at this point. As in most scientific areas and disciplines, there is still a room not only for the improvement of the existing technologies but also for new developments. An accurate prediction of the bioactive conformation of a ligand by first principles and by generating a single 3D structure is still not possible. Further increasing computer power will probably allow for the generation and screening of larger conformational ensembles in reasonable amounts of time in the future. Entirely novel future concepts and ideas, adapted especially from the rapidly growing area of information technology and theory, will probably further leverage successes in hunting for the bioactive conformation.

References

1 Leach, A. (2001) *Molecular Modelling: Principles and Applications* (2nd edn), Pearson Education EMA
2 Sadowski, J. *et al.* (2004) 3D structure generation and conformational searching. In *Computational Medicinal Chemistry and Drug Discovery* (Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J.P., eds), pp. 151–212, Dekker Inc
3 Sadowski, J. (2003) 3D structure generation. In *Handbook of Chemoinformatics – From Data to Knowledge* (Gasteiger, J and Engel, Th., eds), pp. 231–261, Wiley–VCH
4 Schwab, C.H. (2003) Conformational analysis and searching. In *Handbook of Chemoinformatics – From Data to Knowledge* (Gasteiger, J. and Engel, Th., eds), pp. 262–301, Wiley–VCH

5 Perola, E. and Charifson, P.S. (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* 47, 2499–2510
6 Nicklaus, M.C.W.S. *et al.* (1995) Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* 3, 411–442
7 Mobley, D.L. and Dill, K.A. (2009) Binding of small-molecule ligands to proteins: 'what you see' is not always 'what you get'. *Structure* 17, 489–498
8 Keith, T. *et al.* (2009) Toward accurate relative energy predictions of the bioactive conformation of drugs. *J. Comput. Chem.* 30, 601–610
9 Hawkins, P. *et al.* (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* 50, 572–584
10 Loferer, M.J. *et al.* (2007) Analyzing the performance of conformational search programs on compound databases. *J. Comput. Aid. Mol. Des.* 25, 700–710
11 Smellie, A. *et al.* (1995) Analysis of conformational coverage. 1. Validation and estimation of coverage. *J. Chem. Inf. Comput. Sci.* 35, 285–294
12 Smellie, A. *et al.* (1995) Analysis of conformational coverage. 2. Application of conformational models. *J. Chem. Inf. Comput. Sci.* 35, 295–304
13 Brooks, B.R. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217
14 Smellie, A. *et al.* (1995) Poling: promoting conformational variation. *J. Comput. Chem.* 16, 171–187
15 Kirchmair, J. *et al.* (2006) Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* 46, 1848–1861
16 Li, J. *et al.* (2007) CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J. Chem. Inf. Model.* 47, 1923–1932
17 Ferguson, D.M. and Raber, D.J. (1989) A new approach to probing conformational space with molecular mechanics: random incremental pulse search. *J. Am. Chem. Soc.* 111, 4371–4378
18 Chen, I-J. and Foloppe, N. (2008) Conformational sampling of drug-like molecules with MOE and Catalyst: implications for pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* 48, 1773–1791
19 Sadowski, J. *et al.* (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Comput. Sci.* 34, 1000–1008
20 Renner, S. *et al.* (2006) Impact of conformational flexibility on three-dimensional similarity searching using correlation vectors. *J. Comp. Inf. Model.* 46, 2324–2332
21 Wolber, G. (2009) Pers. commun., University of Innsbruck, Austria (the 778 compounds of Ref. [15] were processed with CORINA and ROTATE in July 2009)
22 Halgren, T.A. (1996) The Merck molecular force field. I. Basis, form, scope, parameterisation, and performance of MMFF94. *J. Comput. Chem.* 17, 490–499
23 Mohamadi, F. *et al.* (1990) MacroModel – an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* 11, 440–467
24 Kolossváry, I. and Guida, W.C. (1996) Low mode search. An efficient automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* 118, 5011–5019
25 Boström, J. (2002) Reproducing the conformation of protein-bound ligands: a critical evaluation of several popular conformational searching tools. *J. Comput. Aid. Mol. Des.* 15, 1137–1152
26 Watts, K.S. *et al.* (2010) ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Comp. Inf. Model.* 50, 534–546
27 Friesner, R.A. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749
28 Banks, J.L. *et al.* (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* 26, 1752–1780
29 Pearlman, R.S., Balducci, R. (1998) Confort: a novel algorithm for conformational analysis. National Meeting of the American Chemical Society, New Orleans, LA, USA.
30 McMartin, C. and Bohacek, R.S. (1997) QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput. Aid. Mol. Des.* 11, 333–344
31 Klebe, G. and Mietzner, T. (1994) A fast and efficient method to generate biologically relevant conformations. *J. Comput. Aid. Mol. Des.* 8, 583–606
32 Klebe, G. *et al.* (1999) Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput. Aid. Mol. Des.* 13, 35–49
33 Tresadern, G. and Agrafiotis, D.K. (2009) Conformational sampling with stochastic proximity embedding and self-organizing superimposition: establishing reasonable parameters for their practical use. *J. Chem. Inf. Model.* 49, 2786–2800
34 Griewel, A. *et al.* (2009) Conformational sampling for large-scale virtual screening: accuracy versus ensemble size. *J. Chem. Inf. Model.* 49, 2303–2311
35 Lybrand, T.P. (1990) Computer simulations of biomolecular systems using molecular dynamics and free energy perturbation methods. In *Reviews in Computational Chemistry – Vol. 1* (Lipkowitz, K.B. and Boyd, D.B., eds), pp. 295–320, VCH Verlagsgesellschaft mbH