European Journal of Medicinal Chemistry 44 (2009) 1405–1409

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

MED-3DMC: A new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space

Olivier Sperandio [a,b,*], Marc Souaille [b], François Delfaud [b], Maria A. Miteva [a], Bruno O. Villoutreix [a]

[a] Université Paris Descartes, Inserm U648, 45 rue des Saints-Pères, 75006 Paris, France
[b] Medit SA, 2 rue du Belvédère 91120 Palaiseau, France

ARTICLE INFO

Article history:
Received 16 May 2008
Received in revised form
16 September 2008
Accepted 18 September 2008
Available online 21 October 2008

Keywords:
Virtual ligand screening
Multiconformational ensemble
MED-3DMC

ABSTRACT

Obtaining an efficient sampling of the low to medium energy regions of a ligand conformational space is of primary importance for getting insight into relevant binding modes of drug candidates, or for the screening of rigid molecular entities on the basis of a predefined pharmacophore or for rigid body docking. Here, we report the development of a new computer tool that samples the conformational space by using the Metropolis Monte Carlo algorithm combined with the MMFF94 van der Waals energy term. The performances of the program have been assessed on 86 drug-like molecules that resulted from an ADME/tox profiling applied on cocrystalized small molecules and were compared with the program Omega on the same dataset. Our program has also been assessed on the 85 molecules of the Astex diverse set. Both test sets show convincing performance of our program at sampling the conformational space.

1. Introduction

The generation of relevant conformations for small molecules represents a versatile problem that requires the correct balance between maintaining the list of low energy conformations as short as possible (for combinatorial reason) and within a given energy window, and providing enough conformational diversity such that the Holy grail bioactive conformation can be found among the generated structures. A substantial number of computer tools exists to apply such a task and various approaches, from tools building linkers on the fly and combining them with pre-generated fragment libraries for the ring systems [1,2] to purely stochastic methods [3] among others [4–6].

Several studies [7–12] have been undertaken to get insight into the properties of the bioactive conformation and to know whether or not it resembles the global minimum conformation and what type of energy window and diversity criterion should be used. Although they might disagree on the final answer, they seem to agree that the conformation of the ligand when bound to the protein is equally interesting than the conformation, even crystallized, of the molecule alone in solution. Any computer tool that generates conformational ensemble should provide the bioactive conformation among the many generated conformations. To address the afore-mentioned trade-off between diversity and energetic criterion, we introduce MED-3DMC, a Metropolis Monte Carlo algorithm based on a SMARTS mapping of the rotational bonds and the MMFF94 van der Waals energy term. In order to validate and to assess our program performance we used a list of 86 drug-like ligands derived from the list of molecules described by Kirchmair et al. [9] and compare it to the performance of the commercial program Omega. Our program was able to reproduce the bioactive conformation with an average RMSD equal to 0.67 Å when imposing a maximum of 50 conformations, an energy window of 25 kcal mol^{-1} and a RMSD diversity criterion of 0.8 Å. Doing so, our program performs similarly to Omega on the same test set, with even a better behaviour when applied on certain molecules with low to medium number of rotatable bonds (3–5). Similarly, we used our tool on the Astex diverse set (85 molecules) and showed again convincing performances of our tool.

2. Materials and methods

2.1. MED-3DMC algorithm

The present approach (Fig. 1) is based on a Metropolis Monte Carlo (MMC) algorithm using the van der Waals potential energy term of the MMFF94 force field for the evaluation of the Boltzmann factor, as well as a SMARTS expression recognition procedure to detect the degrees of freedom of the treated molecules (rotatable bonds + non-planar trigonal Nitrogen). The SMARTS expressions

* Corresponding author. Université Paris Descartes, Inserm U648, 45 rue des Saints-Pères, 75006 Paris, France. Tel.: +33 1 57 27 83 90; fax: +33 1 57 27 83 97.
E-mail address: olivier.sperandio@univ-paris5.fr (O. Sperandio).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.09.052

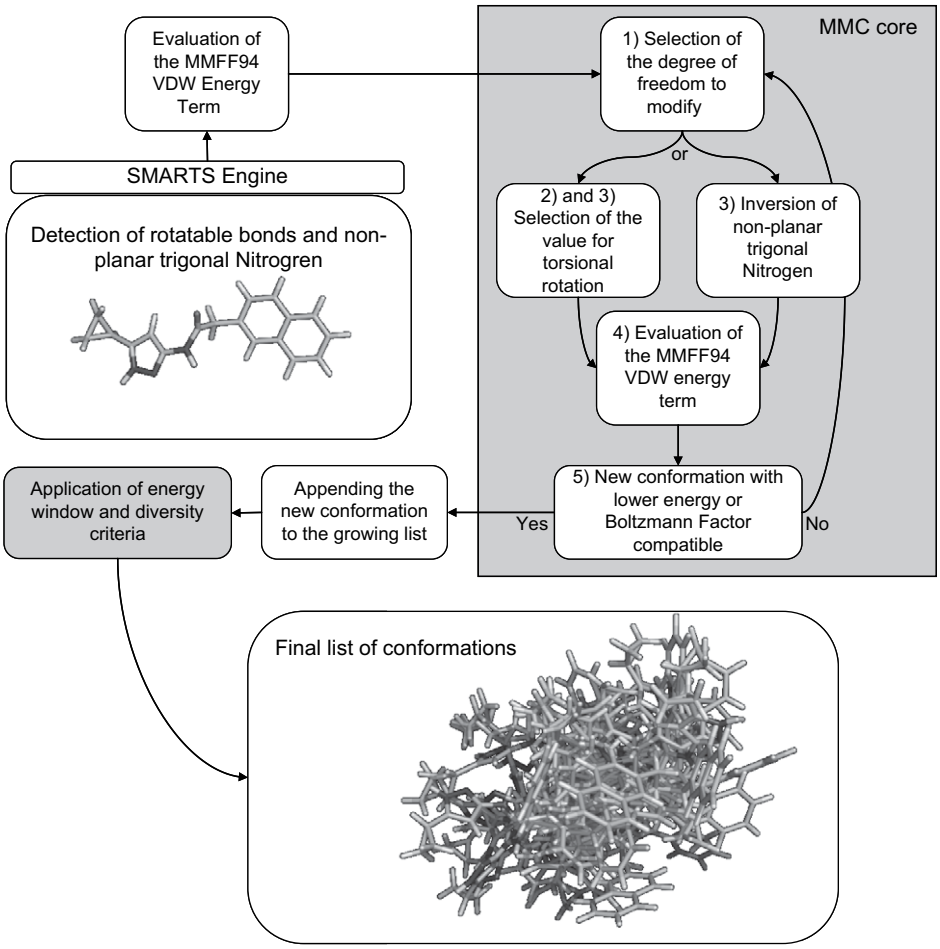

Fig. 1. MED-3DMC procedure.

that can be used for the identification of the rotatable bonds are as follows:

```
[!$(*#*)&!D1&!H3&!$([O;H1])&!$([N;H2])]-
!@[!$(*#*)&!D1&!H3&!$([O;H1])&!$([N;H2])]
[!$(*#*)&!D1&!H3&!$([O;H1])&!$([N;H1]C=O)&!$
(C=(O)[N;H1])]-!@[!$(*#*)&!D1&!H3&!$([O;H1])]
```

The first expression represents any rotatable bond that is in this case any non-terminal atom being neither a methyl, nor a primary amine, nor a hydroxyl, bound to another atom sharing these characteristics. The second expression excludes the peptide bond condition as well. Therefore, the use of a SMARTS expression authorizes a total parameterization of the notion of rotatable bond.

The core of the MMC algorithm (Fig. 1) is processed as follow: (1) a rotatable bond or a trigonal non-planar Nitrogen is chosen randomly; (2) if a rotatable bond has been chosen, then an angle of rotation is selected randomly from a list of dihedral variation, starting from 30 degrees to 180 degrees with intervals of 30 degrees; (3) the geometric transformation is applied, that is, either a dihedral rotation with the selected angle on the selected rotatable bond or the Nitrogen inversion; (4) the van der Waals (VDW) MMFF94 energy term is evaluated and compared to the VDW energy of the previous MMC step; (5) if the energy difference is negative or if the Boltzmann factor (BF) is satisfactory (Metropolis criterion) then the new conformation is accepted. If not, a new conformation is generated from the previously accepted conformation. The global procedure goes on until the number of Monte Carlo cycles is reached.

After the global number of Monte Carlo cycles (usually 10), we designed a procedure that ensures conformational diversity and restrains the generated conformations within a user-defined energy window. The reason why we applied this filter after the Monte Carlo cycles is to prevent any interference with the Markov chain of the MMC. The simulation temperature used for the evaluation of the Boltzmann factor has been set to 1200 K. This has proven to give a higher conformational diversity. The global procedure consists of 10 Monte Carlo cycles of 2000 steps, with each cycle starting with a new seed but from the same initial conformation. Unlike the evaluation of the Boltzmann factor, the potential energy term calculated during the energy window calibration is not restrained to the sole van der Waals term, as we can also evaluate the torsional term (e.g. to eliminate conformations with too high torsional strain energy).

The procedure that ensures conformational diversity is based on a two-step process. First, a pair-wise torsion angle RMSD calculation is applied on all generated conformations, using one referenced dihedral angle for each of the rotatable bonds, with a threshold of 5 degrees. This permits to rapidly discard too close conformations without having to superimpose them (longer calculation) because it uses internal coordinates. Once this first diversity filter is applied, a standard pair-wise Cartesian RMSD based filter is applied to ensure conformational diversity with a user-defined value. A general rule of thumb concerning the CPU time required on Intel Xeon 2.4 GHz with 2Gb memory to generate one conformer ensemble is on average 40–80 s from 1 to 8 rotatable bonds in a compound, and increases quadratically for compounds with more than 8 rotatable bonds.

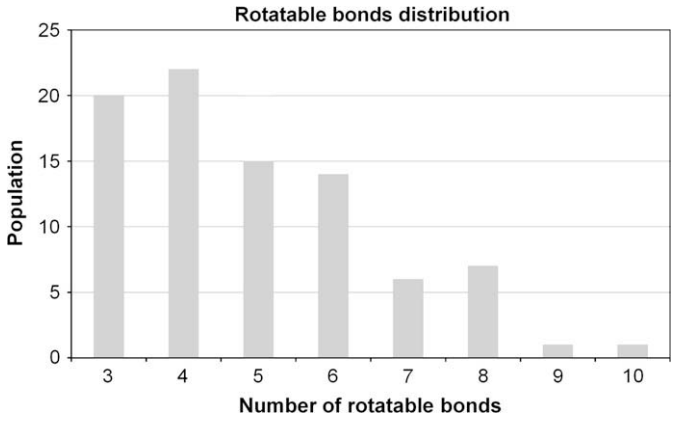

Fig. 2. Rotatable bonds distribution.

3. Validation dataset

3.1. Kirchmair dataset

The 86 molecules of the validation dataset were derived from the 778 molecules described by Kirchmair et al. [9]. An ADME/tox filtering step was performed on these molecules with the program Filter [13], with the default drug-like parameter file and by imposing a minimum number of 3 rotatable bonds for each molecule.

3.2. Astex diverse set

The Astex diverse set [14] was also chosen for evaluating the performance of MED-DMC. This set contains 85 chemically diverse molecules for which the structures and conformations have been determined experimentally from X-ray crystallography and derived from the Protein Data Bank (PDB).

4. Results and discussion

In order to validate and to assess the performance of our tool, we investigated 86 drug-like molecules with different sets of parameters. We further investigated our package on the Astex diverse set which contains 85 molecules (see below). With regard to the Kirchmair's dataset, the 86 molecules were taken from the set of 778 molecules used in a recent study with which Kirchmair et al. [9] assessed the performance of two major programs that construct 3D conformations of small molecules, Omega [1,2] and Catalyst [5]. The reason why the number of molecules has been shrunk so drastically with respect to Kirchmair's validation set (778 molecules) is due to our ADME/tox filtering performed by the program Filter, and the supplementary criterion we used that imposed a minimum of 3 rotatable bonds. In the spectra of this work we intended to focus principally on drug-like compounds that are meant to become orally available drugs in the sense of ADME/tox and Lipinski rules (see Fig. 2 for number of rotatable bonds distribution). The 778 molecules used by Kirchmair et al. not only contained drug-like compounds but also peptide-like molecules, cofactors, or compounds that would be by no mean used as starting point for developing orally available drug.

Table 1
Runs parameters.

Runs	Max confs	RMSD criterion	E_{window} (kcal mol^{-1})
50–0.8	50	0.8	25
50–1.0	50	1.0	25
50–1.2	50	1.2	25
100–0.8	100	0.8	25
500–0.4	500	0.4	25
500–0.6	500	0.6	25

The names of the different runs are defined as X–Y with X equal to the maximum number of conformations generated, and Y the value of the RMSD diversity criterion.

Table 2
Average of the best RMSD for the 6 sets of runs.

Runs	50–0.8	50–1.0	50–1.2	100–0.8	500–0.4	500–0.6
Average of the best RMSDs	0.67	0.71	0.76	0.62	0.50	0.52
Standard deviation	0.4	0.33	0.33	0.33	0.29	0.29

The names of the different runs are defined as X–Y with X equal to the maximum number of conformations generated, and Y the value of the RMSD diversity criterion.

The study of Kirchmair et al. and others [7–11] has shown that among the parameters that play an important role in the generation of molecular conformers are: the energy window with respect to the global minimum; the RMSD value used to impose conformational diversity; and obviously the maximum number of generated conformers. Those have a major influence on the quality of the results, particularly on the RMSD value obtained between the bioactive conformation and the closest generated conformation.

Therefore, we tried several sets of values for the RMSD diversity criterion and also for the maximum number of conformers. Moreover, preliminary tests showed that an energy window of 25 kcal mol^{-1} gives a better conformational diversity. This is in good agreement with the above mentioned study of Kirchmair et al. [9] assessing the program Omega, among others, whose default energy window has also been set to 25 kcal mol^{-1}. Thus we designed several series of validation runs on the 86-molecule dataset, as shown in Table 1.

Those correspond to several strategies that could be employed depending on the virtual screening project. For large chemical libraries (up to several thousands) a maximum number of conformations of 50 or 100 still represent a limit, whereas on focused libraries, a maximum of 500 can easily be used. Regarding the RMSD diversity criterion, we observed that a general rule of thumb is to use 0.8 Å as a standard value when generating for each molecule a maximum number of 50 or 100 conformations, which is too high when this number is equal to 500. The RMSD between the best-generated conformer (closest to the bioactive conformation) has been monitored for each of the 6 sets of runs (see Table 2). We observed that, in all cases, even when generating only 50 conformers, the average of the best RMSDs is always inferior to

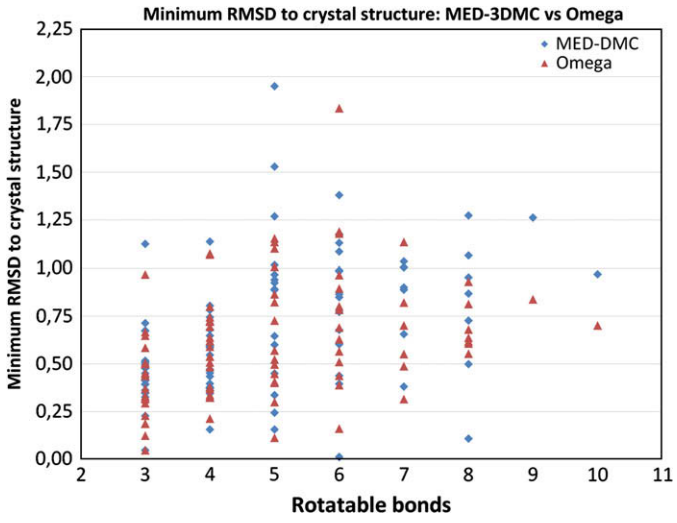

Fig. 3. Best RMSD distribution for the run 50–08 in function of the number of rotatable bonds: MED-3DMC vs. Omega.

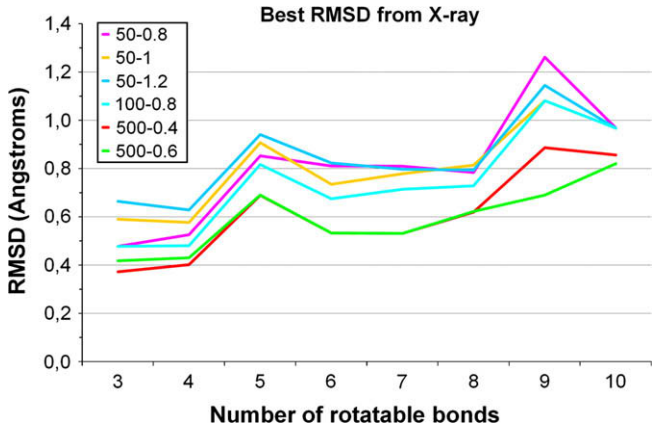

Fig. 4. Variation of the average RMSD between the best-generated conformation and the bioactive conformation, with respect to the number of rotatable bond.

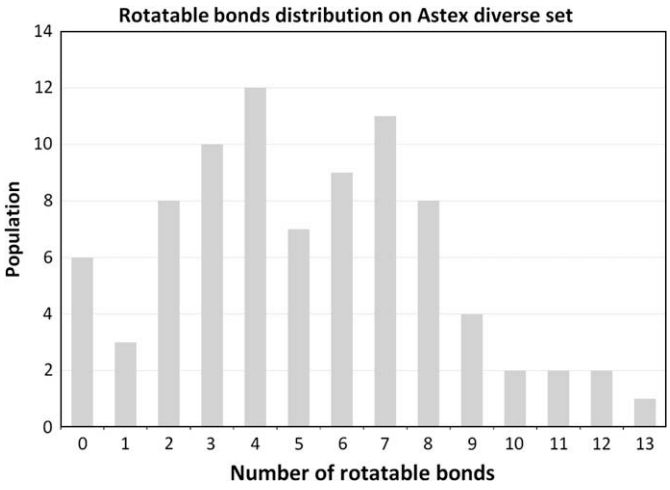

Fig. 6. Rotatable bonds distribution on Astex diverse set.

0.8 Å. This represents a very encouraging result, because it means that MED-3DMC could potentially be used for large chemical libraries for which it is required to generate molecular conformations with a maximum limit of 50 conformers. This indicates that even in this case, the program manages to efficiently sample the conformational space and provides at least one structure very close (low RMSD) to the bioactive conformation.

We compared the performance of our tool with Omega on the exact same validation set (86 drug-like molecules) using identical parameter settings, among which some were suggested by Kirchmair et al. [9], that is, an energy window of 25.0 kcal mol^{-1}, and a RMSD diversity criterion of 0.8 Å, plus an identical maximum of 50 conformations per molecule. Fig. 3 represents a plotting of the minimum RMSD of the generated ensemble with respect to the corresponding bioactive conformation for both MED-3DMC and Omega as a function of the number of rotatable bonds. We note that we obtain similar performance in terms of minimum RMSD to the bioactive conformation and even better results for certain structures, noticeably at lower number of rotatable bonds. Such plotting gives, besides the average and standard deviation shown in Table 2, an estimation of the distribution of the best RMSDs with respect to the number of rotatable bonds in the case of the generation of a conformational ensemble for a large chemical library. For both programs the minimum RMSD with respect to the bioactive conformation does not correlate linearly with the number of rotatable bonds ($R^2_{\text{MED-3DMC}} = 0.166$, and $R^2_{\text{Omega}} = 0.115$) as it can be seen for other programs such as Balloon [3]. It is clear from Fig. 3 that the vast majority of the best RMSDs obtained are below 1 Å,

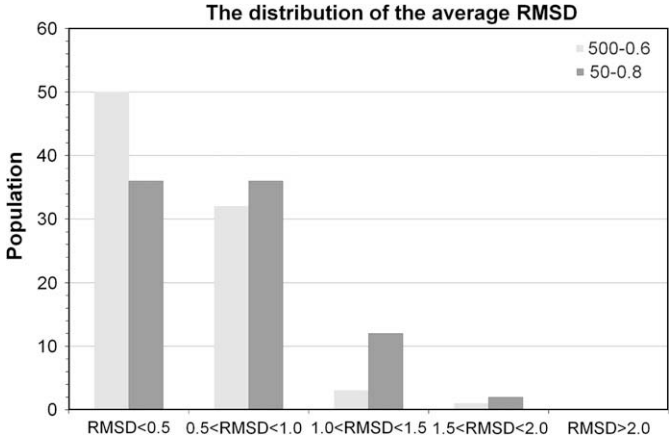

Fig. 5. Distribution of the average RMSD for both the 500–0.6 and the 50–0.8 runs.

and not only for molecules containing a small number of flexible bonds.

In order to address the variation of the best RMSDs observed in Table 2 (see standard deviation values), the best-averaged RMSD has also been plotted as a function of the number of rotatable bonds in order to analyse the dependency of the performance with the degree of flexibility of the molecules (Fig. 4). We note in Fig. 4 that with the runs 500–0.4 and 500–0.6 the best RMSD equal or inferior to 0.6 Å even with 8 rotatable bonds and close to 0.8 Å with 9 or 10 rotatable bonds.

With a maximum number of 50 conformations, we can observed that at a low number of rotatable bonds (3 or 4), a RMSD criterion of 0.8 Å seems to provide better results, but does not seem to influence the quality of the conformation at higher number of rotatable bonds. Interestingly, even at a higher number of rotatable bonds the average RMSD to the bioactive conformation remains at about 0.8 Å, which is reasonable, considering the flexibility of the molecules and the small number of generated conformations.

In Fig. 5 is represented, for the 500–0.6 and 50–0.8 runs the distribution of the best average RMSDs when put into RMSD bins of 0.5 Å. For 500–0.6, the principal observation is that the majority (95% of the molecules) of the best RMSD is inferior to 1 Å. This means that MED-3DMC can generate a pool of 500 conformers per molecule with almost in every case at least one conformation with a RMSD inferior to 1 Å with respect to the bioactive conformation.

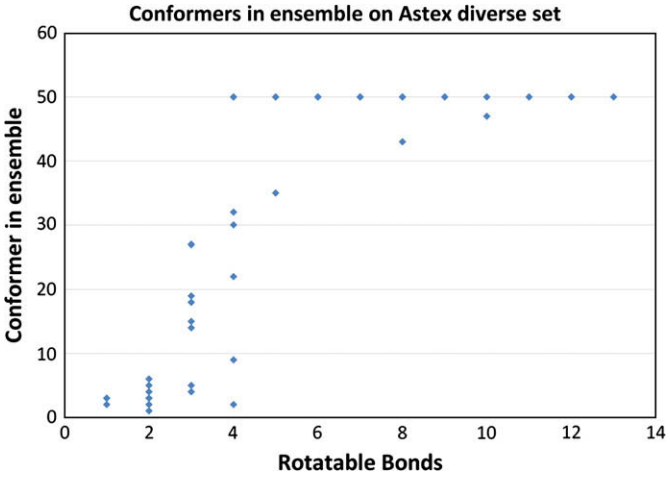

Fig. 7. The number of generated conformers (truncated to 50 conformers) per ensemble as a function of the number of rotatable bonds on Astex diverse set.

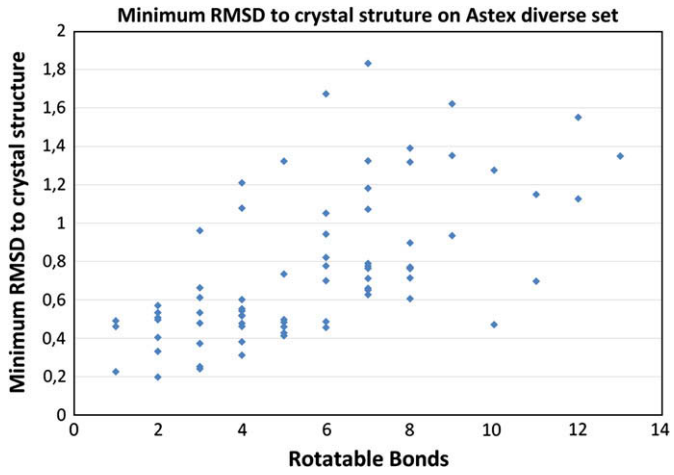

Fig. 8. The minimum RMSD [Å] of optimal superimposition of generated conformers and the experimental protein-bound conformation taken from the Astex diverse set (from PDB) as a function of the number of rotatable bonds.

In the case of the 50–0.8 run, the most part of the molecules (98% of the molecules) have been generated with a RMSD inferior to 1.5 Å and nearly 84% of them have been generated with a RMSD inferior to 1 Å. This provides an interesting estimation of the performance of MED-3DMC to generate conformation ensembles with an efficient sampling of the conformational space.

Finally, we assessed MED-3DMC on the Astex diverse set that contains 85 diverse co-crystallized compounds. Fig. 6 shows the number of rotatable bonds distribution. Fig. 7 shows the number of generated conformers as a function of the number of rotatable bonds. The number of generated conformers with MED-3DMC is not directly correlated to the number of rotatable bonds ($R^2 = 0.231$). It can actually depend on the size of the chemical groups present on each side of the rotatable bonds considered by the program, the bulkier the group is, the lower the number of conformations will be, due to steric clashes along the conformational sampling. For instance, one can cite two interesting examples from the Astex diverse set illustrating this point: two molecules that possess 5 (pdb code 1xm6) and 4 (pdb code 1l2s) rotatable bonds, and that triggered an ensemble of 35 and 122 conformations, respectively (before truncation to 50 conformations), because the steric environment of the peripheral groups are quite different. Another Astex compound from the diverse set (pdb code 1l7f) which has 10 rotatable bonds triggered only an ensemble of 47 conformations (with no truncation), because of the disposition and size of the peripheral chemical groups.

In Fig. 8 is represented the minimum RMSD of optimal superimposition of generated conformers and the bioactive conformation taken from the Astex diverse set. It can be easily seen that the vast majority of the conformer ensembles generated possess at least one conformer within 1.4 Å from the bioactive conformation. This can be compared to a recent study in which a 3D conformation generator program, Balloon, is presented [3] and validated compounds from the Astex diverse set. Some comparison can be made between Balloon and MED-3DMC by focusing on Astex compounds containing less than 14 rotatable bonds (i.e., molecules that were used in both studies) (see Fig. 6).

5. Conclusion

MED-3DMC is a new tool to generate conformation ensembles of small drug-like molecules and can be applied either to design large multiconformational chemical libraries or smaller libraries such as focused libraries. The performances of this tool have been initially assessed on a validation set of 86 drug-like compounds. The results suggest an efficient sampling of the conformational space, with an average RMSD to the bioactive conformation inferior to 0.8 Å, even when generating as low as 50 conformers per compound. Moreover, when aiming at generating focused libraries, that is, when allowing the program to generate 500 conformers per molecule, MED-3DMC displays a RMSD to the bioactive conformation below 1 Å for 95% of the molecules and below 0.5 Å for 58% of the molecules.

The results obtained in the present study by MED-3DMC show similar efficiency at sampling the conformational space of drug-like compounds than Omega on the same dataset of 86 compounds [9]. Besides, when applying our program on Astex diverse set (85 molecules), the results shows that the majority of the generated conformer ensembles have one conformation at least within 1.4 Å from the bioactive conformation, even when imposing a maximum of 50 conformers per compound, which represent similar performance with a recently published study on the same validation set [3].

The total parameterization of the flexibility definition as SMARTS codes allows the user to tune the conformer ensemble generation according to each project. Therefore, we think that MED-3DMC could be used in a variety of virtual screening campaigns, either in structure-based screening, for example rigid body docking of small molecules with tools such as FRED [15], SHEF [16], or in the context of a multistep structure-based protocol [17] or in a ligand-based screening, such as pharmacophore-based screening [18].

6. Software licensing

Commercial information about MED-3DMC is available at www.medit.fr. Questions about MED-3DMC licensing should be addressed to info@medit.fr. Researchers from the Inserm Institute U648 have no financial interests in MEDIT and collaborated with this company only for the present project. Therefore, MEDIT SA has the exclusivity for MED-3DMC sales.

Acknowledgment

We would like to thank the French Ministry of Research for co-funding this work and the INSERM institute for support. We thank OpenEye Scientific Software for providing Filter.

References

[1] T.B. Leite, D. Gomes, M.A. Miteva, J. Chomilier, B.O. Villoutreix, P. Tuffery, Nucleic Acids Res. (2007).
[2] Omega 2.0 Santa Fe, New Mexico (2006).
[3] M.J. Vainio, M.S. Johnson, J. Chem. Inf. Model 47 (2007) 2462–2474.
[4] J. Sadowski, J. Gasteiger, G. Klebe, J. Chem. Inf. Comput. Sci. 34 (1994) 1000.
[5] Catalyst, San Diego, California.
[6] J. Li, T. Ehlers, J. Sutter, S. Varma-O'brien, J. Kirchmair, J. Chem. Inf. Model 47 (2007) 1923–1932.
[7] S. Renner, C.H. Schwab, J. Gasteiger, G. Schneider, J. Chem. Inf. Model 46 (2006) 2324–2332.
[8] E. Perola, P.S. Charifson, J. Med. Chem. 47 (2004) 2499–2510.
[9] J. Kirchmair, G. Wolber, C. Laggner, T. Langer, J. Chem. Inf. Model 46 (2006) 1848–1861.
[10] J. Bostrom, J. Comput. Aided Mol. Des. 15 (2001) 1137–1152.
[11] J. Bostrom, J.R. Greenwood, J. Gottfries, J. Mol. Graph Model 21 (2003) 449–462.
[12] J. Tirado-Rives, W.L. Jorgensen, J. Med. Chem. 49 (2006) 5880–5884.
[13] Filter V. Santa Fe, New Mexico; 2006.
[14] M.J. Hartshorn, M.L. Verdonk, G. Chessari, S.C. Brewerton, W.T. Mooij, P.N. Mortenson, C.W. Murray, J. Med. Chem. 50 (2007) 726–741.
[15] FRED Santa Fe, New Mexico (2005).
[16] W. Cai, J. Xu, X. Shao, V. Leroux, A. Beautrait, B. Maigret, J. Mol. Model 14 (2008) 393–401.
[17] M.A. Miteva, W.H. Lee, M.O. Montes, B.O. Villoutreix, J. Med. Chem. 48 (2005) 6012–6022.
[18] S.A. Khedkar, A.K. Malde, E.C. Coutinho, S. Srivastava, Med. Chem. 3 (2007) 187–197.