JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

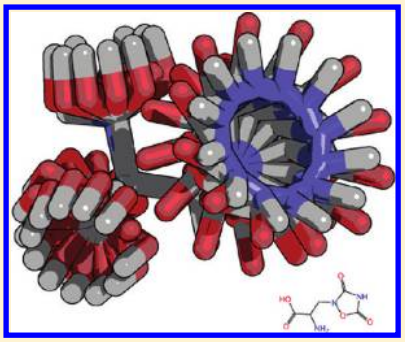
Freely Available Conformer Generation Methods: How Good Are They?

Jean-Paul Ebejer,[†,‡] Garrett M. Morris,[‡] and Charlotte M. Deane*[,†]

[†]Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K., and [‡]InhibOx Limited, Oxford Centre for Innovation, New Road, Oxford, OX1 1BY, U.K.

Supporting Information

ABSTRACT: Conformer generation has important implications in cheminformatics, particularly in computational drug discovery where the quality of conformer generation software may affect the outcome of a virtual screening exercise. We examine the performance of four freely available small molecule conformer generation tools (Balloon, Confab, Frog2, and RDKit) alongside a commercial tool (MOE). The aim of this study is 3-fold: (i) to identify which tools most accurately reproduce experimentally determined structures; (ii) to examine the diversity of the generated conformational set; and (iii) to benchmark the computational time expended. These aspects were tested using a set of 708 drug-like molecules assembled from the OMEGA validation set and the Astex Diverse Set. These molecules have varying physicochemical properties and at least one known X-ray crystal structure. We found that RDKit and Confab are statistically better than other methods at generating low rmsd conformers to the known structure. RDKit is particularly suited for less flexible molecules while Confab, with its systematic approach, is able to generate conformers which are geometrically closer to the experimentally determined structure for molecules with a large number of rotatable bonds (≥10). In our tests RDKit also resulted as the second fastest method after Frog2. In order to enhance the performance of RDKit, we developed a postprocessing algorithm to build a diverse and representative set of conformers which also contains a close conformer to the known structure. Our analysis indicates that, with postprocessing, RDKit is a valid free alternative to commercial, proprietary software.



INTRODUCTION

The vast majority of small molecule drugs work through physical interaction with specific biological macromolecules, usually proteins. The principal determinants of molecular recognition are complementarity of shape and properties between the two molecular entities. Thus the biological function of a drug is intimately related to its three-dimensional structure. Furthermore, most drug molecules are flexible and can adopt a variety of shapes (conformations) in aqueous solution, existing as an ensemble of low-energy conformations in equilibrium with one another. The "biologically active" conformation (that which binds to the target protein) may be similar to one of the solution conformations or it may be a new conformation induced by protein binding.[1] Different proteins may induce different conformations of the same ligand. For example, two arbitrarily selected conformations of the same ligand *cellotriose*, with the Protein Data Bank (PDB)[2] chemical component identifier CTR, bound to two different cellulases from different micro-organisms (PDB entries 2rfz and 2xqo), have a root-mean-square deviation (rmsd) of 3.25 Å from each other. This also serves to highlight that a molecule's "shape" may be more accurately described by an ensemble of low-energy conformations that it may adopt, and thus the diversity of conformations generated by a method is an important consideration.[3]

Therefore, for Computer-Aided Drug Design (CADD) studies, an effective method is required for generating conformational models that captures the bioactive conformation as one of a set of diverse, energetically accessible conformations (because this bioactive conformation will generally not be known in advance). Figure 1 shows a comparison of the crystal structure with the various conformations obtained for each of the methods reviewed here, when only a single conformation is generated. CADD methods that use such models are widespread and include shape-based similarity searches,[4] pharmacophore modeling,[5,6] 3D QSAR,[7] structure-based,[8−10] and ligand-based[11] methods in virtual screening.

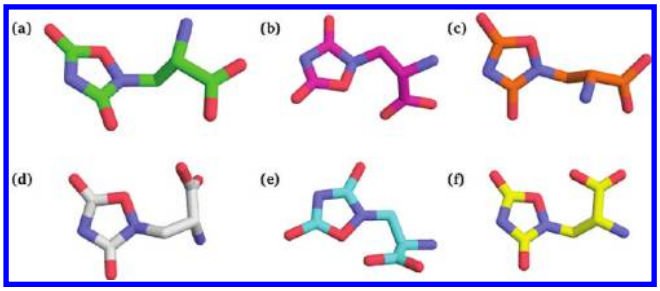


Figure 1. Examples of randomly selected conformations for *quisqualate* (PDB chemical component identifier QUS) with four rotatable bonds generated by the various tools in our review: (a) the X-ray crystal structure (taken from PDB entry 1p1o), (b) Confab, (c) RDKit, (d) Balloon, (e) Frog2, and (f) MOE.

Received: October 2, 2011
Published: April 8, 2012

The aim of the conformer generation process is to build a set of representative conformers that covers the conformational space of a given molecule. The generated conformers should reasonably sample the energy landscape and not produce highly unlikely structures. In applications such as virtual screening, conformational models must be generated for a very large number of compounds, and, as a result, methods that are very computationally intensive are impractical. All the conformer generation methods we consider are therefore based on a molecular mechanics approach.

While reviews of conformer generation software have been published, most of these do not reflect recent developments[12] and/or review only commercial software.[3,13−15] Test set sizes in these publications vary from 32 to 778 molecules.

In this study we benchmark the performance of four freely available conformer generation methods and one implemented in the commercial package MOE. Key metrics are the ability to reproduce the experimentally determined conformation, the coverage of conformational space, and the computational efficiency of the methods.

Conformer Generation. Conformer generation algorithms may be broadly classified as either systematic or stochastic.[16] In the systematic approach, a regular sampling of each of the dimensions of the search space, and hence of all possible conformers, is conducted. This is achieved by incrementing the torsion angles of all rotatable bonds by some predefined amount. Each conformer is therefore enumerated. This is impractical for molecules with a large number of rotatable bonds, due to the combinatorial explosion in the number of rotameric states. In the stochastic approach, the conformational space is randomly sampled using techniques such as Monte Carlo simulated annealing,[17,18] genetic algorithms,[19−21] and distance geometry.[22,23]

Knowledge-based methods[24,25] use predefined libraries for torsion angles and ring conformations.[26,27] These libraries are created by considering known, experimentally determined structures in databases such as the Cambridge Structural Database (CSD)[28] and the PDB. A molecule is decomposed into its constituent fragments and libraries of the possible conformations of these fragments are used to reassemble the whole molecule, in either a stochastic or systematic manner. Energetically favorable conformers are subjected to energy optimization in torsion angle space. The output conformer ensemble is typically generated based on energetic and geometric criteria.

Tools Compared. The conformer generation tools reviewed and compared in this study are Balloon,[20] Confab,[29] Frog2,[30,31] RDKit,[32] and MOE.[33] All of these tools are free and publicly available, with the exception of MOE. Confab, Frog2, and RDKit are open source. Frog2 is also accessible through a Web interface. We give a brief description of each tool and refer the interested reader to the cited literature and the user manuals of these software packages for more details.

RDKit. RDKit uses the distance geometry approach described by Blaney et al.[34] In this approach, a matrix representing the lower and upper bound of all pairwise distances in a molecule is created. This matrix effectively describes the whole structural space for a molecule. The triangle inequality rule is applied to smooth and further refine this matrix (where the three vertices of the triangle are different atoms, so for atoms A, B, and C the distance $AC \leq AB + BC$). To generate multiple conformers, random distance matrices which satisfy the bounds matrix are generated. These conformers are then typically cleaned up using a force field. RDKit does not guarantee that the generated structures are low energy, but it is possible to discriminate and keep only conformers which are a certain rmsd threshold apart. The documentation[35] suggests optimizing the generated conformers using the Universal Force Field (UFF).[36] It also states RDKit's conformer generation is designed to supply 3D structures quickly.

RDKit is an open source cheminformatics toolkit made available under the permissive Berkeley Software Distribution (BSD) license.

Balloon. Balloon uses distance geometry to generate an initial 3D structure for a ligand, followed by a multiobjective genetic algorithm approach which modifies torsion angles, the stereochemistry of double bonds, tetrahedral chiral centers, and ring conformations. The aim is to generate conformers which are near the global energy minimum, while also being diverse and distinct. Generated conformers are optimized using a parametrized force field. Conformers which are closer than an rmsd threshold to a lower energy conformer are discarded, ensuring that the conformers generated are different from each other.

Balloon is free and supports a number of different operating systems (Linux, Mac OS X, and Windows). The source code is not available, and its use is governed by a proprietary license.

Confab. Confab is a knowledge-based conformer generation tool. It uses a systematic approach to generate and test all conformers described by a set of torsion rules. Conformers are generated by varying torsion angles; therefore, no conformers are generated for molecules with zero rotatable bonds. The number of conformers to test may be specified by a cutoff (default 10^6), in which case the conformational space is visited randomly in order to ensure adequate sampling. Only the conformers within a certain energy threshold of the lowest energy conformation are kept. Conformers are also discarded if they are similar in shape to other selected structures (i.e., their rmsd falls within a user-selected value; the default is 0.5 Å). Note that it is not possible to generate a user-specified number of conformers using Confab. Furthermore, and unlike the other tools reviewed here, it only accepts 3D structures as an input. Confab does not explore ring conformations.

Confab is an open source project available under the GNU General Public License (GPL) version 2 license.

Frog2. Given a one- or two-dimensional description of a molecule, Frog2 will break it down into a graph of rings and acyclic elements. This graph is then used to generate conformers. For the ring nodes, a conformation is selected from a library using a knowledge-based approach. The acyclic elements are built using literature-based canonical bond lengths and valence angles. Various combinations of dihedral angles are considered, supplemented by using a Monte Carlo search to vary these angles by a small amount. Conformations are then energy minimized using the AMMOS force field.[37]

Frog2 has a Web interface,[38] and the source code is available under the GNU General Public License (GPL) version 3 license.

MOE. Molecular Operating Environment (MOE) is a fully integrated commercial drug discovery software package.[33] It offers a wealth of functionality covering structure-based design, pharmacophore discovery, protein modeling, molecular simulations, cheminformatics, QSAR, and medicinal chemistry applications. MOE offers three methods for conformer generation: systematic search, stochastic search, and low mode molecular dynamics. The systematic search method generates conformers by rotating the dihedral angles of the molecule by a discrete predefined amount (this type of generation is only suitable for molecules with a small number of rotatable bonds). In the

stochastic search method all the rotatable bonds in the molecule are randomly rotated (including ring bonds), and stereochemistry may also be randomly inverted. The low mode molecular dynamics simulation generates conformers by running a brief molecular dynamics simulation, with velocities initialized to low-frequency vibrational modes. Although this method is efficient for small molecules, chiral centers are rarely inverted. Irrespective of the method chosen the output is subjected to energy minimization (by default, a modified version of the MMFF94[39] force field).

MOE is commercially available from Chemical Computing Group.

Other Tools. There are several other popular, commercial conformer generation software tools not featured in this review.

CORINA is a commercial product available from Molecular Networks[40] which generates a single, high-quality conformer. It takes a rule and data-based approach to generate a low-energy conformer.[41] First, bond lengths and bond angles are set to standard values based on a table. Bond lengths are specific to atom types, hybridization states, and bond order of a particular atom pair. Bond angles depend on the atom type and hybridization state of the central atom. Second, the molecule is fragmented into ring systems and acyclic parts. CORINA can handle small-ring, rigid polymacrocyclic systems and flexible macrocyclic systems in different ways and is able to produce a list of conformations for ring systems. The resulting geometries are optimized using a reduced force field.[42] ROTATE is a complementary program also available from Molecular Networks which generates diverse conformational ensembles by applying a set of rules that resulted from a statistical analysis on the conformational preferences of experimentally determined molecular structures of small molecules.

OMEGA is another commercial conformer generation tool available from OpenEye Scientific Software[43] and uses a systematic, knowledge-based approach. It works by first assembling the initial 3D structure from a library of fragments. Second, it exhaustively enumerates all rotatable torsions using predefined libraries and finally samples this large conformational space using geometric and energy criteria.[44]

For a more detailed and comprehensive review of these (and other) tools we direct the interested reader to a review by Schwab.[5]

Test Set. Our validation of freely available conformer generation methods is based on a test set of 708 distinct small molecules. The selection of these molecules is derived from the work of Hawkins et al.[44] as well as the ligands found in the Astex Diverse Set.[45] These molecules come from high quality X-ray crystal structures in the PDB and CSD. A list of PDB and CSD identifiers and Simplified Molecular Input Line Entry Specification (SMILES)[46] representations of the ligands used in this study can be found in the Supporting Information. For the PDB structures we also supply the reference file used. The molecular weight, heavy atom counts, and rotatable bond distributions for the molecules in the test set are shown in Figure 2. Note that these distributions are similar to the distributions for drug compounds published in the literature[47,48] and comparable to other data sets used for conformer validation.[13,15]

METHODS AND MATERIALS

Conformer Generation Tools. The program installations used in this review are described in Table 1.

Default parameters recommended by the creators of the respective programs have been used throughout when running

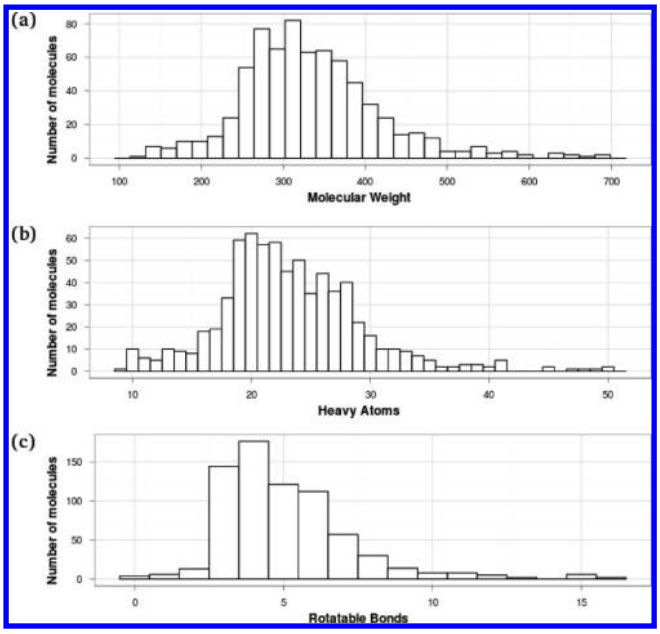


Figure 2. Test set distributions for (a) molecular weight, (b) number of heavy atoms, and (c) number of rotatable bonds.

Table 1. Tools Used for Conformer Generation

tool	version	platform	cost/license	distribution
MOE	2010.10	Linux (Ubuntu 10.10 64-bit)	commercial	binary
BALLOON	1.2.0.915	Linux (Ubuntu 10.10 64-bit)	free/ proprietary	binary
CONFAB	1.0.1	Linux (Ubuntu 10.10 64-bit)	free/GNU GPL	source
FROG2	2.13 (patched from authors)	Linux (Ubuntu 10.10 64-bit)	free/GNU GPL	source
RDKIT	2011.09.1	Linux (Ubuntu 10.10 64-bit)	free/BSD	source

the conformer generation applications, but in order to make a fair comparison we try to replicate the same behavior across all the different tools. For each tool, input and output molecule file parameters were specified as command line arguments. Wherever present the option to generate conformers which are at least 0.5 Å apart was set (before energy minimization). The following points of interest are relevant to each application:

- MOE is very rich in terms of the options for conformer generation. In this work a MOE Scientific Vector Language (SVL) script was written that uses the Conformation Import functionality (i.e., conf_Import) with all filtering options turned off. The conf_Import function breaks a large molecule into smaller fragments. For common fragments it uses conformations from a standard library. If a fragment is not found in the standard library, these fragments are generated using a stochastic conformation search. The number of conformers was specified by setting the outputConformationLimit parameter. Note that the generated structures undergo energy minimization using MOE's MMFF94 modified force field (the outputRefine parameter was set to 1). The output was converted to an MDL SD file to keep the same workflow and input/output requirements as the other tools. The mmSuperposeRMSD parameter which sets the criterion for conformer equality was set to 0.5 Å rmsd.

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146−1158

- BALLOON was run using the settings listed on the usage page (on the Web site[49]), with the MMFF94 force field (option f) and the number of conformers to generate (option nconfs) specified. Note that the latter parameter does not specify the number of final conformations, which may be slightly more or less depending on the flexibility of the molecule, but rather the initial ensemble size. The default value for RMSDtol, which specifies the interconformer rmsd for the final pruning of conformers, is 0.5 Å.
- For FROG2, apart from specifying the work path (option wrkPath) and log file (option logFile), the multi option was also set to 10, 50, or 100, depending on the number of conformers being generated. This setting, together with the gnb option, limits the number of conformations for each compound. We also set the rmsd parameter to generate conformers which are at least 0.5 Å different from each other.
- We generated RDKIT conformers by writing a simple Python script as shown in Section 3.5 in the RDKIT User Manual.[35] As a first step, molecules are loaded from a SMILES file, and hydrogens are added to each using RDKIT. The only difference from the code snippet provided in the manual was that a call to AllChem.EmbedMultipleConfs (instead of AllChem.EmbedMolecule) was made. This function generates a user-defined number of conformers rather than just one. Also, the parameter pruneRmsThresh in this function call is set to 0.5 Å to ensure that the generated conformers are at least this far apart from each other. The script then loops over the conformers to energy minimize them using RDKIT's implementation of the UFF force field, as suggested in the manual.
- The CONFAB conformers were generated using default parameters, which include an rmsd diversity cutoff of 0.5 Å. The conformers outputted from CONFAB were subjected to energy minimization using the MMFF94 force field. This was done using the Obminimize program available in OpenBabel version 2.3.1.[50−52]

Test Set Selection. One of the main challenges was to acquire a diverse test set of high quality structures of small molecules on which to base our conformer tools comparison.

Our starting test set was originally built by combining the 85 bound small molecules from the Astex Diverse Set and a further 677 molecules used in the testing of commercially available conformer generation tool OMEGA.[44]

The Astex Diverse Set is made up of high quality, high resolution crystal structures of complexes containing drug-like ligands from the PDB. It is typically used for protein−ligand docking performance validation. From this set of 85 small molecules, we removed the ligand for PDB entry 1x8x, since this was a duplicate ligand also seen in entry 1of6. This duplication may be an oversight in the Astex Diverse Set as the PDB entries label the ligands as D- and L-isomers of tyrosine for 1of6 and 1x8x respectively, but in fact these structures are both L-tyrosine.

The molecules from the OMEGA test set included the PDB codes of 197 drug-like, high quality structures from well resolved structures in the PDB and 480 molecules from the CSD originating from a previous publication.[27] Unfortunately, there is often more than one ligand associated with each PDB code provided, but the ligand identifier is not supplied. In order to determine which ligand to use for a particular PDB code, the following procedure was applied: (1) apply similar filtering criteria used in the OMEGA validation publication (i.e., $0 \leq$ rotatable bonds ≤ 16, and $8 \leq$ heavy atom count ≤ 50); (2) use the CoFactor database[53] to remove any cofactors present in the structure file; (3) review the literature describing the PDB deposition to determine which was the inhibitor molecule; (4) search the ligands in DrugBank[54] to find a match to one of the ligands listed (and manually select the match).

If after this rigorous manual process there were still multiple potential ligands for a particular structure, the PDB entry was removed from the test set. For each ligand in our test set which has more than one instance in its corresponding PDB entry, we selected an arbitrary structure.

Each identifier for the CSD molecules corresponds to exactly one 3D structure.

The selected molecules from the OMEGA test set were validated using the test set distributions published in the OMEGA paper.

Test Set Preparation. A one-dimensional SMILES representation with stereochemical information was produced from the reference 3D structures of the molecules in the test set using both OPENBABEL 2.3.1 and RDKIT 2011.09.1. From these two equivalent SMILES representations we generated the corresponding InChI keys (with the stereochemistry layer). Where there were mismatches between these keys, the molecule's SMILES representation was inspected manually and, where possible, replaced with the correct SMILES. Otherwise the molecule was removed from the test set. This gives us confidence that the SMILES representations with stereochemistry information (which are available in the Supporting Information) have been correctly generated and are true representations of the 3D reference molecule.

Also, the canonical SMILES and InChI keys were generated for each ligand (also using OPENBABEL 2.3.1) to make sure that there were no duplicates across the whole set. The SMILES representation of the molecule is the starting point of our conformer generation process, ensuring that the conformers are not geometrically biased by the coordinates of the original X-ray structures.

Not all conformer generation tools accept a SMILES string as an input: CONFAB accepts only a 3D structure file. In order to get around this limitation, we generated 3D coordinates from the SMILES representation using OPENBABEL 2.3.1 (with the gen3d option). We then used this as the input file to CONFAB. Once again, we generated InChI keys (with the stereochemistry layer) on the 3D files generated by OPENBABEL and compared them to the InChI keys of the 3D reference molecule file from the test set. If these did not match, we removed the molecule from the test set so as not to negatively bias the results for CONFAB.

The above processing removed 37 out of the initial 197 OMEGA PDB entries, 11 out of the initial 480 OMEGA CSD entries, and 5 out of the 84 Astex Diverse Set entries (duplicate ligand entry 1x8x had already been removed). The remaining 629 molecules from the OMEGA test set were added to the 79 small molecules from the Astex Diverse Set to give a combined set of 708 molecules.

Determining Molecular Descriptors and rmsd between Molecules. The molecular weight, the number of heavy atoms, and the number of rotatable bonds were calculated using the PYBEL API[55] (available when building OPENBABEL 2.3.1 with Python language bindings). The rotatable bonds definition used does not include ring bonds as rotatable.

The rmsd between the experimental structures and conformers generated by each tool were calculated using OBFIT (a program in the OPENBABEL 2.3.1 suite) and considering only heavy atoms.

Number of Conformers Generated. In a small number of cases the software tested failed to generate any conformers (CONFAB: 20 cases, FROG2: 4 cases for the 50 generation run, and BALLOON: 2 cases for the 50 conformer generation run). In some cases this may be attributed to software bugs (e.g., segmentation fault). CONFAB does not generate conformers for molecules with zero rotatable bonds as its approach is based on changing torsion angles of rotatable bonds.

RDKIT was set to generate 10, 50, and 100 conformers for each molecule. Similarly, FROG2 has a command line argument to generate a set number of conformers per stereoisomer. BALLOON has a parameter for the number of conformers to generate, but this is used as an indication only and the number may slightly vary depending on the flexibility of the structure. MOE has an option to limit the number of conformations output. CONFAB does not have any options in this regard so the number of conformers varies greatly e.g. the ligand UN6 in PDB entry 2f70 has 10 rotatable bonds and generated 53,340 conformers. The histogram of the number of conformers generated by CONFAB is shown in the Supporting Information, Figure A. Moreover, sometimes CONFAB generates just one conformer for flexible molecules. This is a known issue because if a molecule has a very large conformer space and a systematic search is carried out within that space only a small fraction of conformers will fall within 50 kcal/mol of the lowest energy conformer. Apart from this, in general and as expected, fewer conformations are generated for less flexible structures (smaller number of rotatable bonds).

For tools which generate more than the required number of conformers (i.e., mostly CONFAB but occasionally also BALLOON) we sample the set randomly.

Statistical Tests. The following tests were performed to find whether the distributions of minimum rmsd values from the theoretical conformer to the experimentally determined structure were statistically significantly different for each method.

First, a Kruskal–Wallis rank sum test was performed for each toolkit's rmsd distribution when generating 50 conformers. This test was selected as there are five groups (tools), and we cannot assume a normal distribution in the underlying minimum crystallographic rmsd values. Second, since the Kruskal–Wallis test only indicates if there is an effect, a *posthoc* test using the pairwise Wilcoxon signed-rank test (paired and with Bonferroni correction) was carried out on toolkit pairs to find which conformer generation methods are statistically different from each other.

RESULTS AND DISCUSSION

We consider the following three criteria in our review: (1) **accuracy**: how close in terms of positional rmsd is a generated conformer to one of the experimentally observed X-ray crystallographic structures? (2) **diversity**: how different or similar are the generated conformations? (3) **speed**: how much computational time is required to generate the conformers?

Our findings show that on average RDKIT and CONFAB are the best conformer generators among the tools we considered. Statistically there is no difference between selecting one of these two tools. However, RDKIT is much faster.

Even if the conformer diversity threshold parameter is set, RDKIT generates many similar conformers after energy minimization (with rmsd <0.5 Å) for molecules with few rotatable bonds. We later describe an algorithm which corrects this.

Quality of Conformers Generated. In order to determine which of the methods most frequently produces conformers closest to the crystallographic conformation, we carried out the following test. Taking each molecule in the test set in turn, we generated 50 conformers using each of the five methods, and for each method we computed the rmsd from each of the 50 conformers to the same arbitrarily selected X-ray structure for that ligand. We then selected the minimum rmsd value, which gives us the closest conformer to the X-ray structure. For tools such as CONFAB that do not have the option of generating a user-specified number of conformers, we sampled randomly from the generated set (for more details see Methods and Materials). The effects of selecting randomly, selecting the minimum energy conformers, and selecting the minimum rmsd structures out of the whole conformer ensemble for CONFAB are shown in Supporting Information, Figure B.

The pairwise comparison of methods can be performed using these minimum crystallographic rmsd values, as shown in Figure 3. In each panel two methods are compared: each point corresponds to a molecule from the test set with its minimum rmsd for each of the two methods providing the *x*- and *y*-coordinates. Note that the points on each graph are smoothed using a kernel density function (i.e., the default function, bkde2D, in R's smoothScatter routine) to avoid overplotting. The diagonal line indicates the theoretical position if both methods performed identically for every molecule in the test set.

Taking the plot of RDKIT (*y*-axis) versus FROG2 (*x*-axis) in Figure 3 as an example, it can be seen that more points fall below the diagonal than above it. In practice, this means that generally the rmsd between a theoretical conformer and its corresponding experimentally determined structure is greater for FROG2 than RDKIT. Thus, RDKIT tends to be more accurate in terms of generating experimentally observed structures of drug-like molecules than FROG2.

The distribution of the minimum crystallographic rmsd values for all 708 small molecules for each method based on 10, 50, and 100 conformer generation runs is shown in Figure 4 (a). Each boxplot shows the minimum rmsd value (from the X-ray crystal structure) for a set of conformers for each molecule in the test set. Thus each boxplot is made up of 708 data points.

Figure 4 (a) shows that RDKIT and CONFAB have better (lower) average minimum rmsd values than the other tools and for more than 75% of the ligands (when generating 50 and 100 conformers) they generate a structure with an rmsd that is less than 0.9 Å from the experimentally observed conformation of the molecule.

Using the OMEGA tool, it is reported that approximately 83% of the PDB structures and 84% of the CSD structures are reproduced within a 1 Å rmsd of the experimental structure (generating a maximum of 200 conformers and using default settings).[44] In order to make our test set comparable to the validation set used in the OMEGA study, we removed the ligands from our test set originating from the Astex set. For the 50 conformer run, RDKIT generates 70.0% of the 160 PDB structures and 90.4% of the 469 CSD structures in our set within a 1 Å rmsd of the experimental structure. All the other tools show inferior performance to RDKIT.

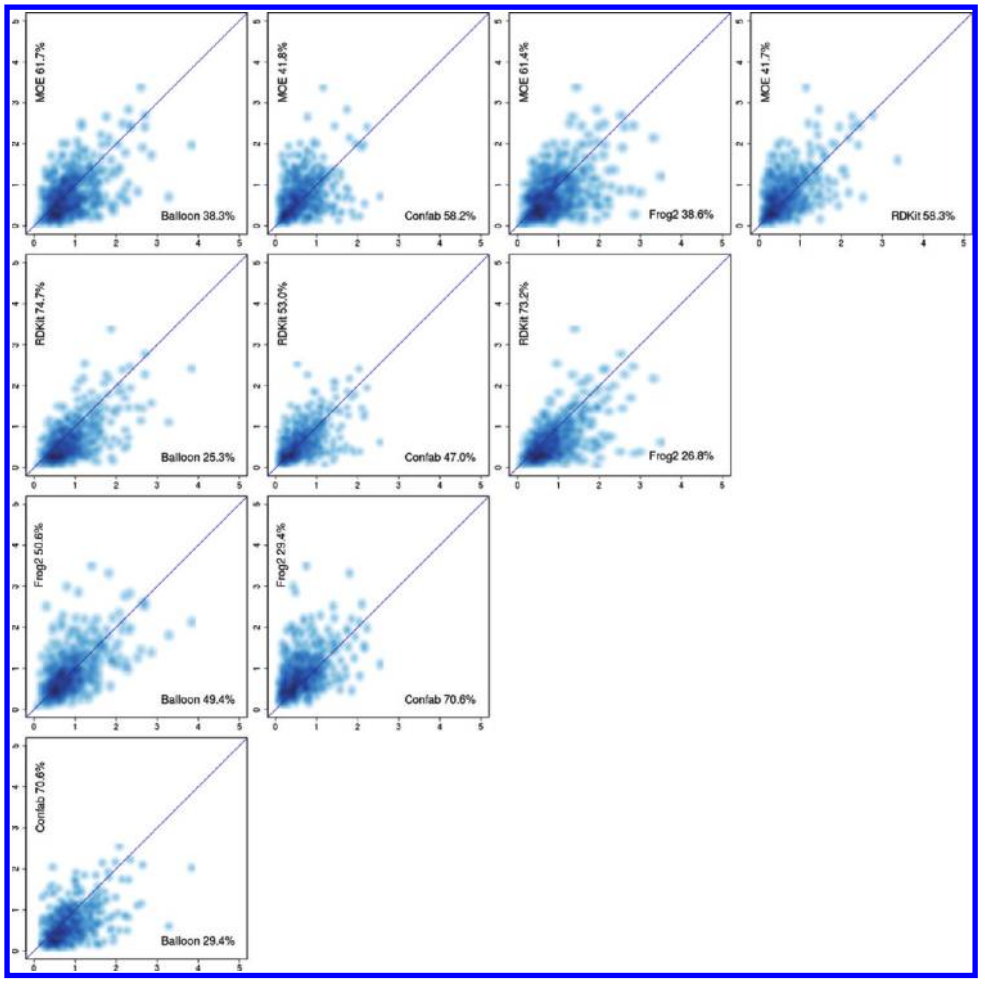


Figure 3. Pairwise comparison of minimum rmsd values between all the conformers generated by the method and the X-ray structure for all toolkits. These plots represent the densities of the 708 molecules in the test set. It can be seen that RDKit performs better than the other toolkits, because it tends to generate conformations with rmsd values from the crystallographic structure that are generally lower than the second method it is compared to, e.g. RDKit versus Frog2, there is a higher density of points below the diagonal than above. The percentages next to the method labels show the proportion of all ligands in the data set (708) for which that particular method is closer to crystallographic structure than the other method plotted on the same panel. All units shown are in Å.

Plots of the average minimum rmsd from the experimentally determined structure (with the standard error indicated by error bars) versus the number of rotatable bonds in the test set for the 50 conformer generation run are shown in Figure 4 (b). Unsurprisingly, as the number of rotatable bonds in the molecule increases it becomes more difficult to generate the crystallographic form. With a couple of exceptions, both RDKit and Confab achieve better results than the other methods, with RDKit doing better or similar to Confab when the number of rotatable bonds is less than ten, and Confab showing better performance for molecules with ten or more rotatable bonds.

When considering all 50 conformers generated by each tool across all 708 molecules (rather than just the best conformer), the percentage of the generated structures with crystallographic rmsd values less than 2 Å is as follows: RDKit 76.7%; Confab 73.2%; Balloon 70.9%; Frog2 70.1%; and MOE 63.4%. These percentages at this frequently used threshold[14,21] give an indication of how plausible the generated conformations tend to be. On the other hand, a lower percentage might indicate more diversity in the conformational model.

The distributions for the 50 conformer runs presented in Figure 4 are all statistically significantly different from each other with the exception of the Balloon-Frog2 and Confab-RDKit pairs ($\alpha = 0.05$). This means that there is a statistical difference in selecting one conformer tool over another.

A Kruskal–Wallis test revealed a significant effect of the toolkit on the minimum rmsd distance to the experimentally determined structures ($\chi^2(4) = 194.961$, $p < 0.01$). The Wilcoxon signed-rank test (paired) with Bonferroni correction showed significant differences between the tools shown in Table 2.

Difficult Cases. As discussed earlier, the larger the number of rotatable bonds the more difficult it is for these tools to generate the X-ray crystallographic structure. The reason for this is that the conformer space grows exponentially with the number of rotatable bonds.

There are a small number of molecules which consistently score badly across all tools. These molecules typically have a central core with flexible, large parts of the molecule stemming from that core as shown in Figure 5 (a). The rmsd calculation is based on the position of these heavy atoms, and even a small torsional rotation of the bonds emanating from the "core" will offset many of the heavy atoms. The presence of macrocycles in the molecule, such as the one in Figure 5 (b), also makes the conformer generation tools perform badly as most of them are unable to sample the ring conformer space correctly. In most cases, the tools tested are unable to reproduce the

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146–1158

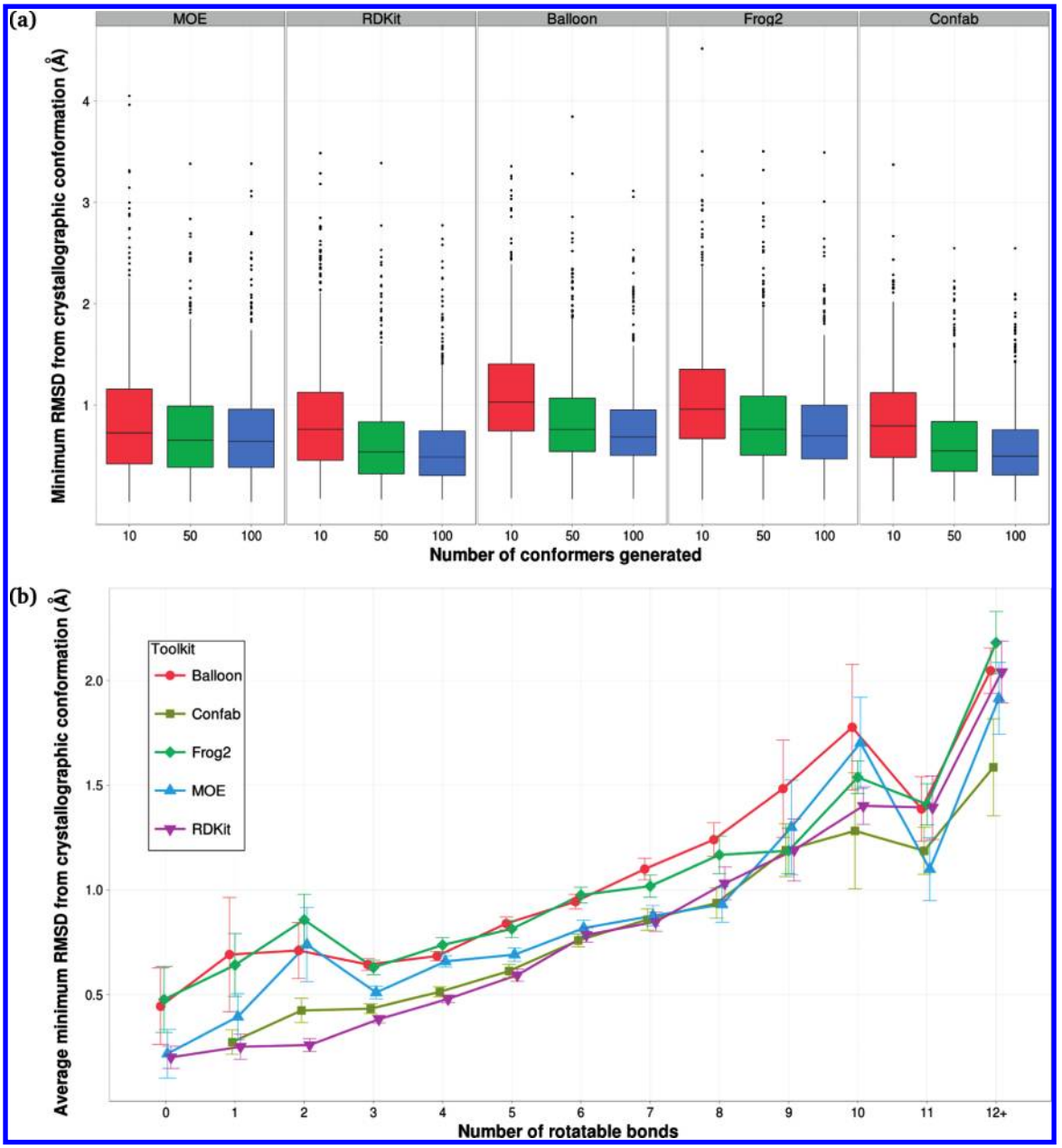


Figure 4. (a) Box plots showing the minimum rmsd from the crystal structure for each of the 708 molecules in the test set, when generating 10, 50, and 100 conformers. In general, as the number of conformers generated increases, the mean rmsd from the crystal structure decreases. (b) The variation in the ability of each method to reproduce crystallographic conformations as the number of rotatable bonds increases. Note that the data points in this graph are jittered horizontally to avoid overplotting. Also, the molecules with 12 or more rotatable bonds are grouped together. It can be seen that in general, among all methods, RDKit and Confab are best at finding the lowest crystallographic rmsd values for the molecules in the test set.

Table 2. Statistical Tests for Generated Conformers Show That There Is a Difference between the Generated Conformer Sets for the below Pairs

toolkits	median (Å)	Z score	p-value
Balloon vs Confab	0.663	12.357	<0.01
Balloon vs MOE	0.697	6.496	<0.01
Balloon vs RDKit	0.646	13.706	<0.01
Confab vs Frog2	0.651	−11.844	<0.01
Confab vs MOE	0.598	−5.357	<0.01
Frog2 vs MOE	0.696	5.730	<0.01
Frog2 vs RDKit	0.637	13.052	<0.01
MOE vs RDKit	0.581	6.003	<0.01

experimentally determined ring conformation for rings with seven or more atoms (refer to Supporting Information, Figure C for more details).

Diversity of the Set of Generated Conformers. Diversity is an important consideration in conformer generation. Running a docking experiment using rigid ligands where these have very similar structures is clearly suboptimal both in terms of the experiment's running time and the space needed to store these conformers (assuming they are stored on the file system or in a database rather than generated on the fly). On the other hand, structures that are very different from the experimentally determined conformation may be unlikely in practice because of energetic and conformational constraints.

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146−1158

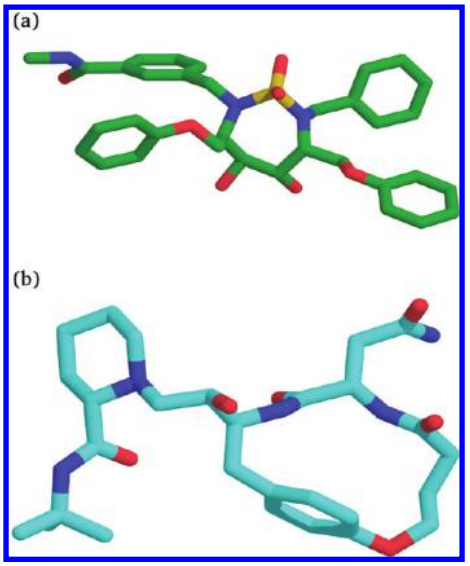


Figure 5. (a) Small molecule with identifier NM1 taken from PDB entry 1g2k in the test set produces conformers across all toolkits which are, on average, 3.459 Å from the experimental structure. (b) Small molecule with identifier PI4 taken from PDB entry 1b6l in the test set produces conformers across all toolkits which are, on average, 2.790 Å from the experimental structure. In this case, the inability to reproduce the experimental structure may be attributed to the macrocycle.

A representative sample that covers all the low energy structures in each molecule's conformational space is therefore required.

For every molecule in the test set and then for each conformer generation method, we have calculated the pairwise rmsd distances between every pair of generated conformers (for the 50 conformer generation run) and partitioned the results by sets of molecules with an odd number of rotatable bonds, and the toolkit used. The resulting graph in Figure 6 shows that as expected, the average of the minimum crystallographic rmsd increases with the number of rotatable bonds. This is because as the flexibility of the molecules increases, the size of the search spaces increases, and this allows for more diverse conformers to be generated. Note that even if the conformer diversity threshold for these tools was set to 0.5 Å, this is applied before the energy minimization step.

RDKit appears to produce many very similar conformations for molecules with low numbers of rotatable bonds (Figure 6). The multimodal peaks can be attributed to the fact that when there are few rotatable bonds, the conformer space is limited and energy minimization will make some of the conformers converge to similar geometries. This highlights the problem with checking for diversity before running the selected energy minimization protocol (as happens with most of these tools).

Conformer Generation Speed. All the conformer generation runs were executed in isolation and benchmarked using the same hardware (using an Intel Core i7−2600K CPU running at 3.40 GHz with 4 GB RAM) and the same operating system (Ubuntu Linux 10.10 64-bit) to get comparable results. Also, the same system tool (i.e., the time command line utility in Linux) was used for benchmarking the execution speed of each conformer generation tool.

We measured the CPU time taken by the conformer generation runs. Our data set consists of 708 separate SMILES files for each molecule in our test set. Each file is run as a separate process. The average CPU time (in seconds) per molecule in the 708 molecule test set for 10, 50, and 100 generated conformers is shown in Figure 7.

Frog2 is an order of magnitude faster than any other tool. Speed is of primary importance for Frog2 considering its Web application nature. MOE shows constant time regardless of the

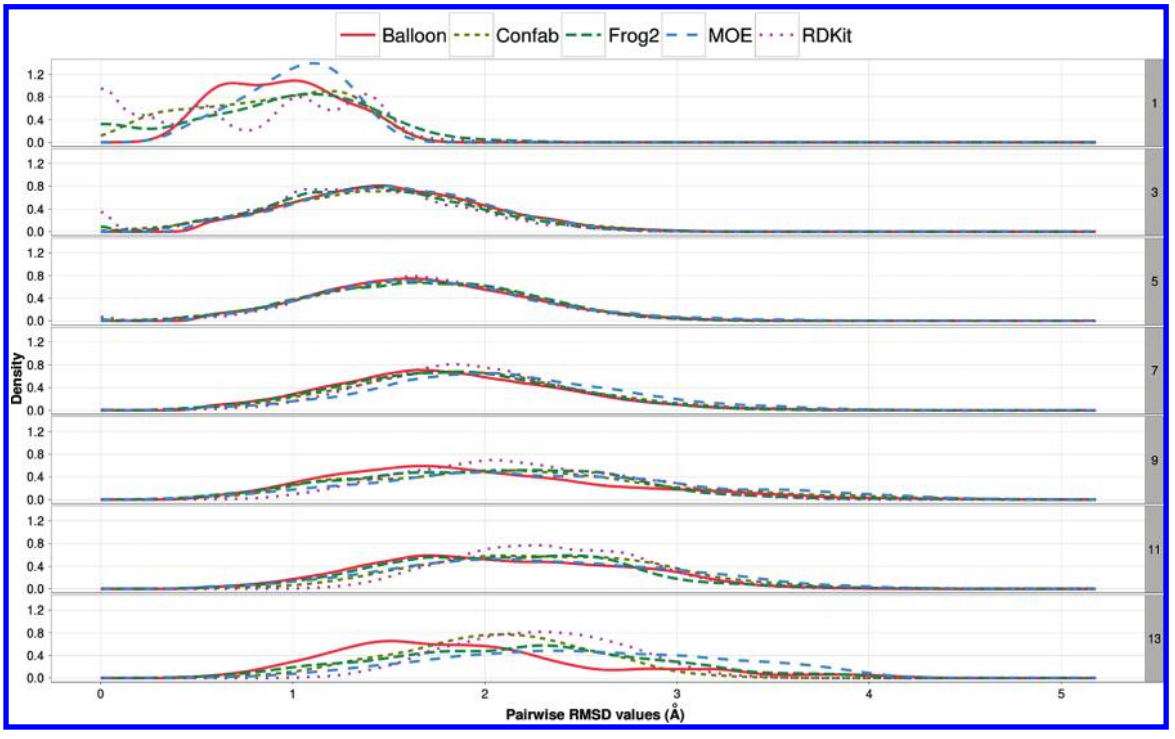


Figure 6. Density-smoothed distribution showing the variation in pairwise rmsd values for conformers per molecule for each toolkit, for sets of molecules with an odd number of rotatable bonds.

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146−1158

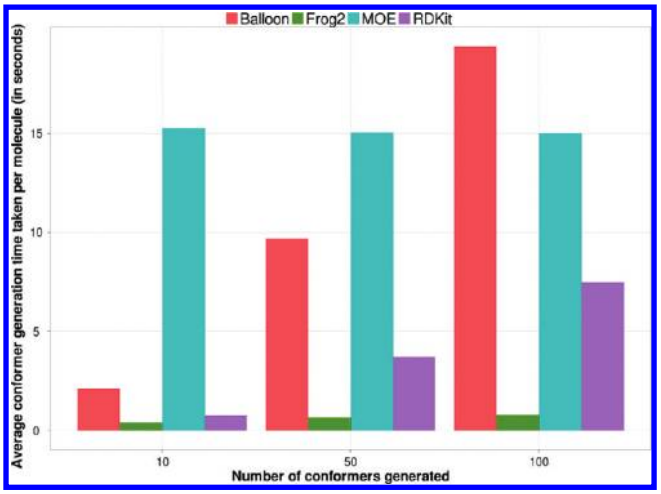


Figure 7. The average time (in seconds) for each molecule for each of 10, 50, and 100 conformer generation runs. CONFAB is not shown due to the inability to generate a specific number of conformers.

number of conformers generated. When running a different experiment, we noticed that MOE was considerably faster when all the molecules were in the same file and a single MOE process launched as opposed to launching a new process for every molecule. The reason for this is 2-fold; first there is a setup cost involved in running a conformer generation job (such as license acquisition or database creation) and, second, the common fragments file is cleared at every single molecule run. RDKIT and BALLOON both show an increase in the time taken with respect to the number of conformers generated, with BALLOON having a much steeper gradient (i.e., slower).

CONFAB is not included in this comparison because it is not possible to generate a specific number of conformers and make a side-by-side comparison to the other conformer generation programs. The CONFAB run (without energy minimization) took 254 min of CPU time for a total of 638,887 conformers generated for the 708 molecules in the test set. The MMFF94 energy minimization for this whole conformer set took approximately 53 h. This figure is biased by the huge number of conformers (>10,000) generated for a few molecules (refer to Supporting Information, Figure A for more details).

A Note on Energy Minimization. While an in-depth comparison of energy minimization on small molecules using a variety of force fields is beyond the scope of this paper, we comment on the effect of energy minimization on the generated conformers. Taking the set of 10 conformers for each molecule in the data set generated by RDKIT, we investigate the effect of the force field on the ability to reproduce the conformation of the experimentally determined structure.

Figure 8 shows that minimization is important to increase the quality of the conformational models. Moreover, energy minimization is a critical step after RDKIT's distance geometry approach as it is required to produce "clean" structures. The MMFF94 force field seems to be better suited for molecules with zero or one rotatable bonds, while UFF does better for the rest of the molecules. The mean rmsd value for all conformers from their respective experimental structures is of 1.113 Å when not using any optimization, 0.878 Å when minimizing using UFF, and 0.945 Å when minimizing using MMFF94.

Another point of interest is the difference between the energy of the experimental structure compared to the energies of the conformers generated (shown in Supporting Information, Figure D). The prevalence of positive values on this histogram shows that the experimental structures have higher energies than the conformers generated. This may indicate an inherent deficiency in these force fields, that the force field used when refining the crystal structure is different, or could be due to crystal packing effects or constraints placed by the bound protein when the ligand crystal structure is taken from a complex.

Generating ten conformers for every molecule in the data set using RDKIT and without any energy minimization took 1 min 46 s

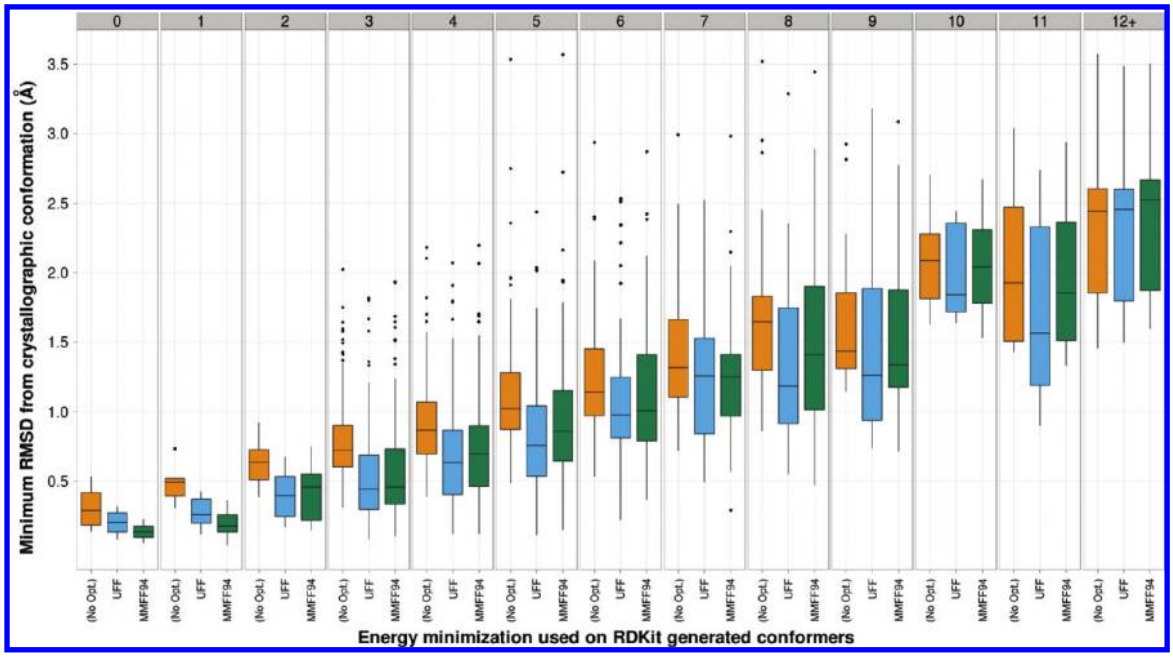


Figure 8. Boxplots showing the minimum rmsd from the experimental structure to the RDKIT set of 10 generated conformers for every molecule in the test set, partitioned by number of rotatable bonds. We show the results for conformers with no optimization (yellow) and with UFF (blue) and MMFF94 minimization (green).

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146–1158

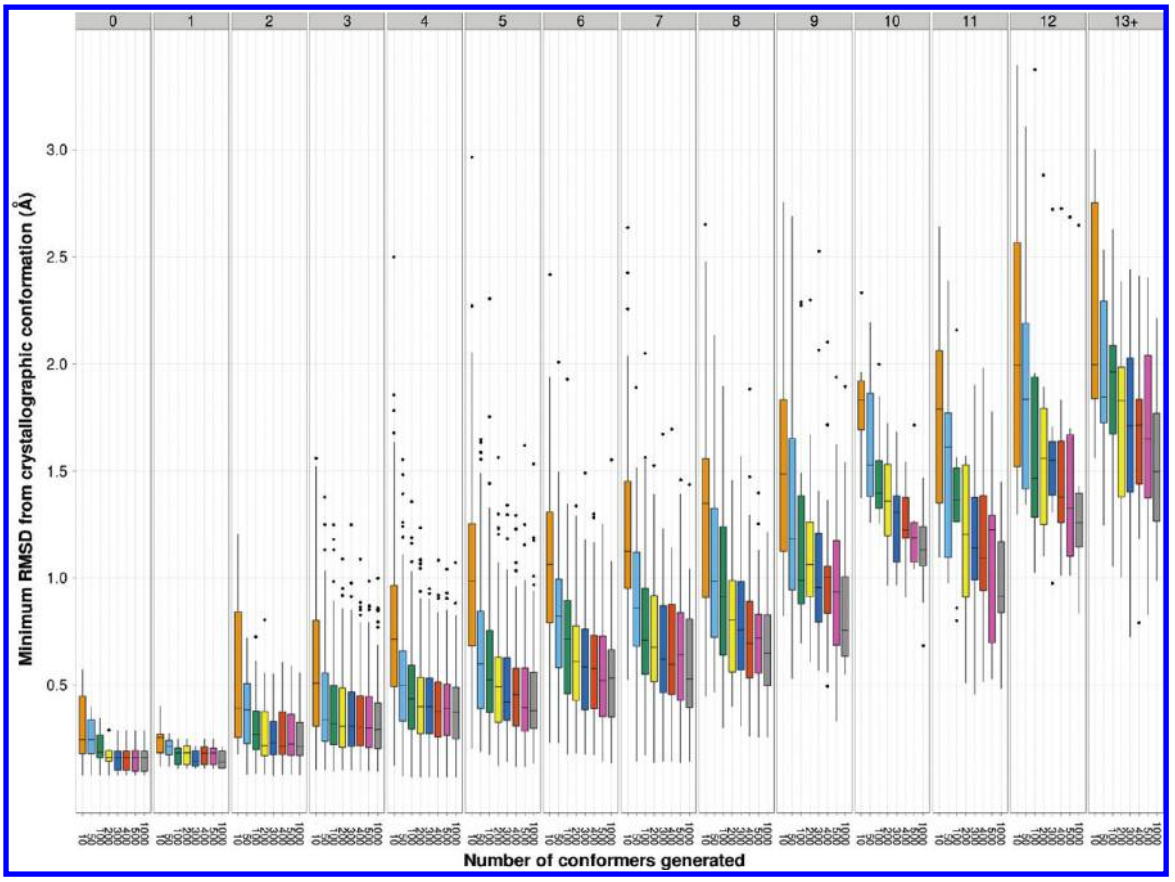


Figure 9. Variation in minimum crystallographic rmsd versus the number of rotatable bonds in the ligand and how these values change with number of conformers generated using RDKit.

on the machine described in the previous section. UFF minimization on these molecules took 8 min 2 s, while MMFF94 minimization took 113 min 10 s. Considering energy minimization makes up the most computationally expensive operation in the conformer generation workflow it may be beneficial to optimize the energy minimization parameters (e.g., maximum iterations or force tolerance).

RDKit - Conformer Generation Postprocessing. The positive results for RDKit in terms of accuracy and speed make it a viable candidate for inclusion in a computational drug discovery project.

One of the main issues with RDKit is that the number of conformers generated must be specified. This means that for less flexible molecules with a small number of (or even no) rotatable bonds, most conformers generated will be similar to each other. This has important repercussions for both the space needed to store the conformers as well as processing time.

RDKit has an option to keep only conformations that are at least a particular rmsd threshold apart from one another (pruneRmsThresh), but this gives distinct conformers before the force field is applied. Theoretically, performing a force field-based energy minimization after the filtering might cause two different conformers close to one another in conformational space to fall into the same local energy minimum and become structurally very similar after energy minimization.

We therefore developed the following alternative approach for filtering the conformers generated by RDKit to resolve this conformational diversity problem: (1) Using RDKit, generate n conformers in set C_{gen}. (2) Energy minimization (using the UFF force field) is performed on every conformer. The conformer list is sorted by increasing energy value and the lowest energy conformer (the first conformer in the list), c_{low}, is recorded. (3) Remove c_{low} from C_{gen} and add it to C_{keep}. (4) For each conformer, c, in C_{gen}, compute the rmsd between c and each conformer in C_{keep}.

(a) If any rmsd value is smaller than a fixed threshold, d_{min}, discard c as we already have a representative of that point in conformational space.
(b) Otherwise add c to C_{keep}

There is ongoing debate as to whether bioactive conformations lie close to an energy minimum[56] or are significantly above it.[1,57] We use the lowest energy conformer as an initial starting point for the sampling of conformational space.

The advantage of sorting the list by increasing energy values is that this will give us the lowest energy conformer in the d_{min} partition.

At the end of the above procedure the set C_{keep} will contain the lowest energy conformer and all its elements will be at least d_{min} Å rmsd apart.

We have performed several experiments to determine the optimal values for the parameters specifying the initial number of conformers to be generated (n) and the minimum threshold distance between each conformer kept (d_{min}). The value of n is a function of the number of rotatable bonds. As expected, more flexible molecules require a larger value of n to cover sufficiently the conformational space.

A good starting value of n is one which still generates a conformer that is similar to the X-ray crystallographically determined structure. Figure 9 shows the minimum rmsd to the

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146−1158

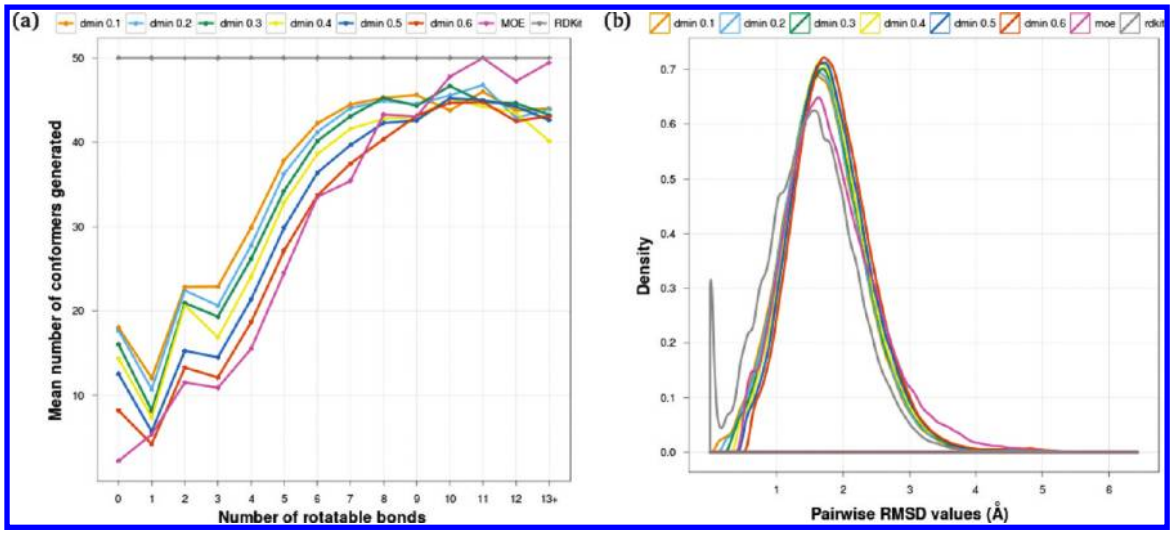


Figure 10. (a) This graph shows the average number of conformers generated for each rotatable bond subset using the RDKit-based algorithm presented here. Note that increasing the d_{min} threshold filters out more conformers (if d_{min} was large enough only one conformer would be left after postprocessing). (b) This graph shows the density distribution of the pairwise rmsd values between conformers of a molecule. Note also that the initial peak of very similar conformers for RDKit is removed using the post processing algorithm.

experimentally determined structure versus multiple RDKit runs that generate 10, 50, 100, 200, 300, 400, 500, and 1000 conformers for each molecule in our test set (partitioned by number of rotatable bonds). Based on Figure 9, plausible values for n could be for $n_{rot} \leq 7$, use $n = 50$; for $8 \leq n_{rot} \leq 12$, use $n = 200$; and for $n_{rot} \geq 13$, use $n = 300$.

We also estimated the value of the clustering threshold, d_{min}, and studied its effects on the number of filtered conformers. Figure 10 shows that as expected the larger the value of d_{min}, the more conformers are filtered.

When applying our postprocessing algorithm on RDKit, using $n = 50$ and $d_{min} = 0.35$ Å we obtain distinct conformers and the initial peak which can be seen in Figure 10(b) for RDKit disappears. This is because similar molecular geometries are filtered out from the generated conformer set. It can also be seen that RDKit with filtering behaves similarly to MOE in terms of number of conformers generated and conformer diversity.

So, using the number of conformers to generate (n) as

$$n = \begin{cases} 50 & \text{if } n_{rot} \leq 7 \\ 200 & \text{if } n_{rot} \geq 8 \text{ and } n_{rot} \leq 12 \\ 300 & \text{otherwise} \end{cases}$$

and a d_{min} value of 0.35 Å, the total CPU running time for the whole test set was 102 min. This is significantly faster than using the one-size-fits-all value of 300 conformers per molecule without compromising the quality of the results. The minimum crystallographic rmsd distributions were similar to those shown for the 50, 200, and 300 conformer generation runs in Figure 9.

CONCLUSIONS

We have reviewed the performance of four free and/or open source conformer generation software packages, Balloon, Confab, Frog2, and RDKit, and compared them to the Conformation Import method implemented in the commercially available package MOE.

We are interested in three measures, specifically the ability of these tools to generate a conformation close to the experimentally observed structure; the coverage of the conformational space of a molecule; and the performance of these tools in terms of speed. These are critical aspects of the computational drug discovery process.

For our benchmarks, we rebuilt the data set used to validate another popular commercially available conformer generation toolkit, Omega, and augmented it with the ligands present in the Astex Diverse Set. The resultant data set consists of 708 molecules from the PDB and CSD with high resolution X-ray crystal structures and which are mostly drug-like in their properties.

When considering the ability to generate conformers which are structurally similar to the experimentally determined structures, we have found that both RDKit and Confab do better than the other toolkits, with the latter performing better with more flexible molecules (i.e., ≥ 10 rotatable bonds). This can be attributed to the systematic exploration of the conformer space as opposed RDKit's stochastic approach.

When analyzing the ability of a method to explore or "cover" conformational space (by measuring the pairwise rmsd between each conformer generated) RDKit tends to generate more similar conformers than the other methods. We presented a postprocessing algorithm we developed to filter out similar structures from the RDKit output, using the lowest energy conformer as the starting point of the conformational space sampling.

In terms of speed, Frog2 was the fastest conformer generator by an order of magnitude and is only slightly affected by the number of conformers generated. After that, RDKit is significantly faster than the other toolkits.

Finally, the choice of a conformer generation tool depends on a number of factors other than the primary ones considered above, e.g. ability to explore the energetic landscape, the ability to integrate with other software (either through source code or in a workflow), licensing model, and pricing. Even in conformer generation, open source tools offer a viable alternative to commercial, closed source, proprietary software.

ASSOCIATED CONTENT

Supporting Information

We provide the PDB entry identifiers, the PDB chemical component three letter codes, and the CSD codes of the 708

dx.doi.org/10.1021/ci2004658 | J. Chem. Inf. Model. 2012, 52, 1146−1158

ligands in our test set (filename ci2004658_si_002.txt). We also supply the SMILES representations of each molecule in the test set (filename ci2004658_si_003.zip). For the PDB molecules we also supply the experimentally determined 3D reference structures used in this study (filename ci2004658_si_004.zip). The supplementary figures referred to in the text may be found in the document ci2004658_si_001.pdf. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: deane@stats.ox.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors declare no affiliations with any of the tools reviewed. This work is supported by European Union Framework Programme 7-funded Marie Curie Initial Training Network STARS, grant agreement PITN-GA-2009-238490. The authors thank Dr. Paul Finn and Dr. Simone Fulle for their helpful suggestions and discussions.

REFERENCES

(1) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.

(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(3) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. Conformational Sampling of Bioactive Molecules: A Comparative Study. *J. Chem. Inf. Model.* **2007**, *47*, 1067−1086.

(4) Hahn, M. Three-Dimensional Shape-Based Searching of Conformationally Flexible Compounds. *J. Chem. Inf. Model.* **1997**, *37*, 80−86.

(5) Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies* **2010**, *7*, e245−e253.

(6) Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D. Comparison of Conformational Analysis Techniques To Generate Pharmacophore Hypotheses Using Catalyst. *J. Chem. Inf. Model.* **2005**, *45*, 461−476.

(7) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in Drug Design - A Review. *Curr. Top Med. Chem.* **2010**, *10*, 95−115(21).

(8) Makino, S.; Kuntz, I. D. Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **1997**, *18*, 1812−1825.

(9) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938−950.

(10) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.

(11) Stahura, F. L.; Bajorath, J. New Methodologies for Ligand-Based Virtual Screening. *Curr. Pharm. Des.* **2005**, *11*, 1189−1202(14).

(12) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Model.* **1994**, *34*, 1000−1008.

(13) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. 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.* **2006**, *46*, 1848−1861.

(14) Boström, J. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137−1152.

(15) Chen, I.-J.; Foloppe, N. Conformational Sampling of Druglike Molecules with MOE and Catalyst: Implications for Pharmacophore Modeling and Virtual Screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773−1791.

(16) Gu, J.; Bourne, P. E. *Structural Bioinformatics*, 2nd ed.; Wiley-Blackwell: Hoboken, New Jersey, USA, 2009; Chapter 27, p 639.

(17) Chang, G.; Guida, W. C.; Still, W. C. An internal-coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc. 111*, 4379-4386.

(18) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. Applications of simulated annealing to the conformational analysis of flexible molecules. *J. Comput. Chem.* **1991**, *12*, 342−349.

(19) Mekenyan, O.; Dimitrov, D.; Nikolova, N.; Karabunarliev, S. Conformational Coverage by a Genetic Algorithm. *J. Chem. Inf. Model.* **1999**, *39*, 997−1016.

(20) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462−2474.

(21) Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H. Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC Bioinf.* **2009**, *10*, 101.

(22) Spellmeyer, D. C.; Wong, A. K.; Bower, M. J.; Blaney, J. M. Conformational analysis using distance geometry methods. *J. Mol. Graphics Modell.* **1997**, *15*, 18−36.

(23) Havel, T.; Kuntz, I.; Crippen, G. The theory and practice of distance geometry. *Bull. Math. Biol.* **1983**, *45*, 665−720.

(24) Klebe, G.; Mietzner, T. A fast and efficient method to generate biologically relevant conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583−606.

(25) Feuston, B. P.; Miller, M. D.; Culberson, J. C.; Nachbar, R. B.; Kearsley, S. K. Comparison of Knowledge-Based and Distance Geometry Approaches for Generation of Molecular Conformations. *J. Chem. Inf. Model.* **2001**, *41*, 754−763.

(26) Sadowski, J.; Boström, J. MIMUMBA Revisited: Torsion Angle Rules for Conformer Generation Derived from X-ray Structures. *J. Chem. Inf. Model.* **2006**, *46*, 2305−2309.

(27) Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M. Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis. *J. Chem. Inf. Model.* **2008**, *48*, 1−24.

(28) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.

(29) O'Boyle, N.; Vandermeersch, T.; Flynn, C.; Maguire, A.; Hutchison, G. Confab - Systematic generation of diverse low-energy conformers. *J. Cheminf.* **2011**, *3*, 8.

(30) Leite, T. B.; Gomes, D.; Miteva, M.; Chomilier, J.; Villoutreix, B.; Tufféry, P. Frog: a FRee Online druG 3D conformation generator. *Nucleic Acids Res.* **2007**, *35*, W568−W572.

(31) Miteva MA., G. F.; Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622−W627.

(32) *RDKit*: Open-source cheminformatics. http://www.rdkit.org (accessed December 12, 2011).

(33) *MOE* (The Molecular Operating Environment) Version 2010.10. http://www.chemcomp.com (accessed December 13, 2011).

(34) Blaney, J. M.; Dixon, J. S. *Distance Geometry in Molecular Modeling*; *Rev. Comp. Chem.*; John Wiley & Sons, Inc.: Hoboken, New Jersey, USA, 2007; Vol. 5, Chapter 6, pp 299−335.

(35) Landrum, G. Getting Started with the RDKit in Python, Version Q1 2011. http://www.rdkit.org/GettingStartedInPython.pdf (accessed December 12, 2011).

(36) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(37) Pencheva, T.; Lagorce, D.; Pajeva, I.; Villoutreix, B.; Miteva, M. AMMOS: Automated Molecular Mechanics Optimization tool for in silico Screening. *BMC Bioinf.* **2008**, *9*, 438.

(38) *Frog2* web interface. http://bioserv.rpbs.univ-paris-diderot.fr/cgi-bin/Frog2 (accessed December 7, 2011).

(39) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(40) Molecular Networks. http://www.molecular-networks.com (accessed December 6, 2011).
(41) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(42) Sadowski, J.; Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(43) OpenEye Scientific Software. http://www.eyesopen.com (accessed December 6, 2011).
(44) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.
(45) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance. *J. Med. Chem.* **2007**, *50*, 726−741.
(46) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28*, 31−36.
(47) Feher, M.; Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J. Chem. Inf. Model.* **2003**, *43*, 218−227.
(48) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251−264.
(49) BALLOON website. http://users.abo.fi/mivainio/balloon (accessed December 7, 2011).
(50) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(51) *The Open Babel Package*, version 2.3.1. http://openbabel.sourceforge.net (accessed December 6, 2011).
(52) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J. K.; Willighagen, E. L. The Blue Obelisk-Interoperability in Chemical Informatics. *J. Chem. Inf. Model.* **2006**, *46*, 991−998.
(53) Fischer, J. D.; Holliday, G. L.; Thornton, J. M. The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics* **2010**, *26*, 2496−2497.
(54) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035−D1041.
(55) O'Boyle, N.; Morley, C.; Hutchison, G. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 5.
(56) Butler, K. T.; Luque, F. J.; Barril, X. Toward accurate relative energy predictions of the bioactive conformation of drugs. *J. Comput. Chem.* **2009**, *30*, 601−610.
(57) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* **1995**, *3*, 411−428.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on April 19, 2012, with incorrect or missing file names in the Supporting Information Section. The correct version was published ASAP on April 26, 2012.