JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

ALFA: Automatic Ligand Flexibility Assignment

Javier Klett,[†] Álvaro Cortés-Cabrera,[†,‡] Rubén Gil-Redondo,[†,§] Federico Gago,[‡] and Antonio Morreale*[,†,∥]

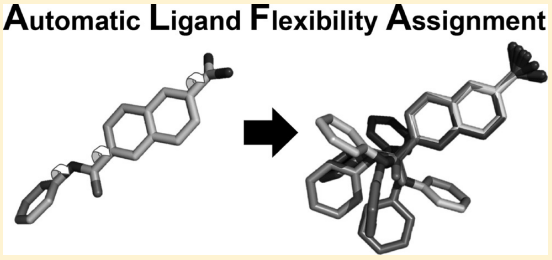
[†]Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Campus de Cantoblanco UAM, E-28049 Madrid, Spain

[‡]Área de Farmacología, Departamento de Ciencias Biomédicas, Unidad Asociada de I+D+I al IQM-CSIC, Universidad de Alcalá, Alcalá de Henares, E-28871 Madrid, Spain

[§]SmartLigs Bioinformática S.L., Fundación Parque Científico de Madrid, c/Faraday, 7. Campus de Cantoblanco UAM, E-28049 Madrid, Spain

Supporting Information

ABSTRACT: ALFA is a fast computational tool for the conformational analysis of small molecules that uses a custom-made iterative algorithm to provide a set of representative conformers in an attempt to reproduce the diversity of states in which small molecules can exist, either isolated in solution or bound to a target. The results shown in this work prove that ALFA is fast enough to be integrated into massive high-throughput virtual screening protocols with the aim of incorporating ligand flexibility and also that ALFA reproduces crystallographic X-ray structures of bound ligands with great accuracy. Furthermore, the application includes a graphical user interface that allows its use through the popular molecular graphics program PyMOL to make it accessible to nonexpert users. ALFA is distributed free of charge upon request from the authors.



INTRODUCTION

Molecules are known to exist in a range of environmentally dependent conformations. The selection of a diverse set of representative conformers for a small molecule is a challenging task in modern theoretical approaches to drug discovery. Different tools have been developed to overcome this problem[1−6] and they perform, in general, reasonably well according to the discussion by Ebejer et al.[6] insofar as they successfully provide ligand conformations that are closely similar to those found in experimentally determined target-bound complexes (almost exclusively from X-ray crystallography) for a high percentage of the molecules comprised in different validation sets.[7,8]

Validating the performance of these tools can be done from a dual perspective: on one hand, by assessing the quality of the conformers in terms of how plausible they are from a structural and/or energetic point of view; on the other hand, by estimating whether or not the set of generated conformers is diverse enough to appropriately populate the phase space of the molecule under study. As the diversity of the set increases, the probability of finding structures more similar to the experimental ones also increases. Different strategies have been developed to achieve both goals, and they usually rely on energy and/or shape criteria. Energy evaluation has been used in many different ways as in ConFab,[3] which is based on a force-field energy function, or in more sophisticated methods[9] that rely on first-principles calculations. Similarity can be measured using shape descriptors[10] or other metrics such as the root-mean-squared deviation (RMSD), Tanimoto combo scores,[10] etc.

Having a similarity measure ready at hand, the next step is to classify the solutions and reduce the set by selecting only those conformers that add diversity to the ensemble. This is done by means of clustering algorithms such as *k*-means, hierarchical clustering, etc. For these methods to work, the all vs all distance matrices need to be precalculated and stored. This turns out to be a problem when dealing with very flexible molecules because the number of possible conformations grows exponentially with the number of rotatable bonds and so do the computational requirements. There are different approaches to select conformations on the basis of results from clustering algorithms. Some of them apply clustering as a mere refining tool or to classify the solutions while they are being produced[3] (i.e., *on the fly*); whereas, others rely, for example, on different energy criteria or Monte Carlo methods.[10]

It is well-known that when force field-based potential energies are used as a criterion to select conformers, a certain bias exists toward those conformations favored by the parameters implemented in the force field and certain areas of the phase space can be left unexplored. This limitation is not so crucial if one considers only molecules in solution but may be of importance when the molecule is bound to its macromolecular target, as the binding site environment is thought to be ultimately responsible for the real pose and may select a conformation that does not exactly correspond to a minimum on the potential energy surface.

Received: July 30, 2013
Published: January 6, 2014

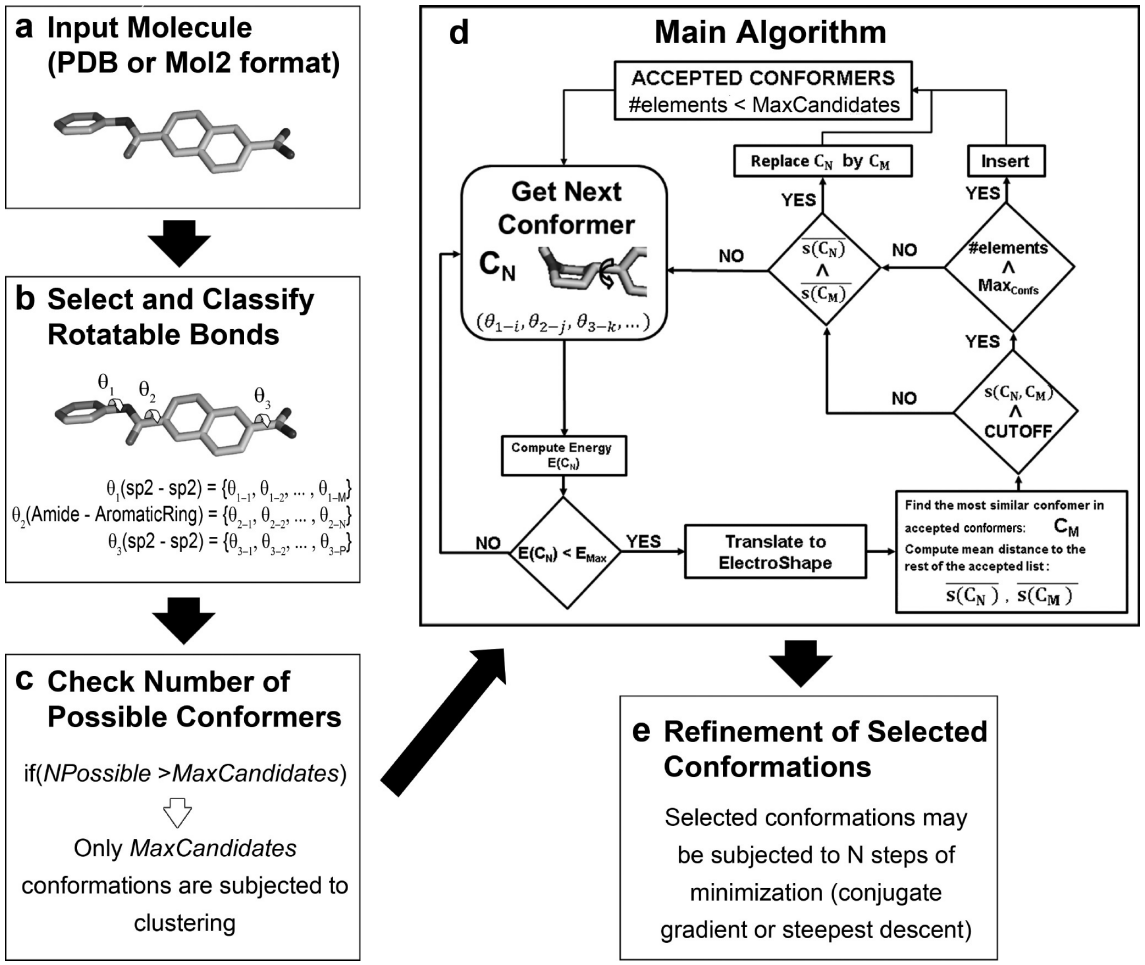


Figure 1. Schematic representation of the different steps performed by ALFA to generate a set of conformers starting from a molecule in PDB or MOL2 format.

Finally, experimental structures are more likely to contain gross errors in atomic coordinates than model-built conformers, e.g. those generated by CORINA,[11] which contain well-defined theoretical values. It is not surprising, therefore, that in some cases current RMSD fitting tools fail to produce a numerical value and simply provide an error code or message. This really happens more often than one would initially think, between 20 and 40% of cases in our hands, depending on the program used (RDKit,[12] OpenBabel RMSD,[13] and OpenEye RMSD[14]). At the core of this problem is the fact that the atom and bond types assigned on the basis of those faulty geometries do not agree with those present in the theoretical models.

In this study we present ALFA, a fast and accurate *on-the-fly* automatic ligand flexibility assignment tool, and discuss the results obtained using it as a conformational ensemble generator for small molecules. ALFA tackles some of the problems outlined above and provides a set of diverse ligand structures representative of both target-bound and "free" solvated states. The program implements: (a) a knowledge-based set of preferred torsion angles according to the atom types that define each rotatable bond; (b) ElectroShape (ES) descriptors[15] to represent each single conformer based on electrostatic and shape properties which are then compared using the Manhattan distance; (c) a selection algorithm designed to achieve enough diversity while covering as much conformational space as possible; (d) an energy minimization routine to refine selected conformations; and (e) a friendly graphical user interface (GUI) integrated within the popular molecular graphics program PyMOL.[16] ALFA is able to reproduce fairly well the experimentally found structures of small molecules isolated in solution and also in complex with their macromolecular targets.

METHODS

Rule-Based Generation of Conformers. As do other methods of this kind[3,10] ALFA relies on a broad and detailed knowledge-based set of preferred torsion angles called rules (Supporting Information Excel file and supplementary Figure 1), which depend on the atom types defining the rotatable bond (aromatic, aliphatic, amide, etc.). This definition applies to any single bond connected to two rigid fragments, and its belonging to a predetermined class depends on the atom types to which it is covalently connected. Rules are stored as SMARTS (simplified molecular-input line-entry system [SMILES] arbitrary target specification) patterns and are sorted in a hierarchical way according to the ambiguity of the patterns, that is, the more general cases are assigned first. These are later overridden by a more specific rule, if it exists, which ensures that the presence of other structural factors such as certain substituents or the electronic properties of the neighboring atoms is taken into account. Each rule has preassigned the most probable discrete values for the rotatable bond it defines. These values have been gathered from textbooks or manuals for initial simple rules such

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

as bonds between sp^3-sp^3 or sp^2-sp^3 atoms. Values for more complex rules were obtained by performing a dihedral angle analysis using the torsional and nonbonded terms from the Generalized AMBER force field (GAFF)[17] on prototype structures representing each type of rotatable bond. Up to 13 different local minima are selected in these rules to better describe all the possible angle values. Discretization of the dihedral angle space significantly reduces the number of possible conformations compared to more exhaustive methods, but still its computational cost increases dramatically with the number of rotatable bonds found in the input molecule.

Starting from the 3D structure of a molecule in PDB or MOL2 format (Figure 1a), ALFA performs two main tasks: rotatable bond assignment and conformer enumeration (Figure 1b). First, each pair of connected fragments and the corresponding rotatable bond are assigned to a rule. For example, a system with two saturated carbon rings linked by a carbon atom will first match the sp^3-sp^3 rule (*general*) but, due mainly to the steric effects between both rings, the default values for sp^3-sp^3 dihedral angles will be modified by application of a more *specific* rule. Second, once a bond involved in a torsion has been detected and assigned, the different values associated with the rule are used to define the structure of each conformer in the enumeration process.

As shown in Figure 1c, the number of possible conformations (N_{Possible}) is checked at the beginning of the process. To avoid the combinatorial explosion, if N_{Possible} > MaxCandidates, where MaxCandidates is a predefined cutoff value (300 000 by default), then ALFA will generate up to MaxCandidates conformers in a random fashion from the whole set of conformers. The detailed list of rules with their preferred dihedral angle values is stored in the parameter file TorsionRules.h (see the Supporting Information).

Measuring Similarity with ElectroShape. We use ES descriptors[15] as a fast and reliable means of measuring the similarity between conformers. ES defines the electrostatic and shape properties of a given conformer in terms of the distributions of Euclidian distances from the atoms to five different centroids. Atoms are considered as discrete points belonging to $\mathbb{R}^4$ where the first three coordinates define the position in the space and the fourth is the atomic charge. Centroids are defined as follows: (a) C_1, the geometric center of all atoms, (b) C_2, the atom whose position is the furthest from C_1, (c) C_3, the atom most distant from C_2, (d) $C_4 = C_1 + \vec{c} + (0, 0, 0, \mu q_{\max})$; and (e) $C_5 = C_1 + \vec{c} + (0, 0, 0, \mu q_{\min})$, where $\vec{c} = (\|\vec{a}\|/2\|\vec{a} \times \vec{b}\|)\vec{a} \times \vec{b}$, $\vec{a} = C_3 - C_1$, $\vec{b} = C_2 - C_1$, μ is a scaling factor (25), and $q_{\max}$ and $q_{\min}$ are the maximum and minimum atomic charges, respectively.

Then, descriptors are built using the mean, standard deviation, and third moment of the distribution of distances from each atom to each centroid, summing up to a total of fifteen scalar values that define the ES descriptor for a given conformation. Descriptors for a pair of conformers (Struct_i, Struct_j) can be compared by defining the similarity measure as a function of the Manhattan distance (d_{man}) between them:

$$s(\text{Struct}_i, \text{Struct}_j) = \frac{1}{1 + d_{\text{man}(\text{Struct}_i, \text{Struct}_j)}}$$

It has been noted that different charge models can have a critical influence on the performance of the method.[15] Accordingly, ALFA allows the user to provide his/her own charges in place of the Gasteiger–Hückel charges that are automatically assigned using OpenBabel.[13]

Classification Algorithm. Each conformation is subjected to the following protocol (Figure 1d):

(1) A pseudoenergy score is assigned that only considers the repulsive term of the commonly used AMBER force field 12–6 Lennard-Jones potential. The conformer is rejected if the score is higher than a predefined cutoff value; otherwise it is accepted. All those conformers with steric clashes are thus eliminated. If the conformer is accepted, it is added to the accepted set, and the process continues until the list is complete (MaxCandidates = 200 by default). It should be noted that if the set of accepted conformers is empty, either we are dealing with the first conformer of the list or none of the conformers tested so far satisfied the energy cutoff value. Finally, if no conformer succeeds in reaching this cutoff the program outputs just the starting structure.

(2) The accepted conformer is translated into ES descriptors and checked for selection. The selection criterion depends on whether or not MaxConfs has been reached. In the former case (|S| = MaxConfs, where MaxConfs represents the maximum number of conformers that will be selected by the algorithm) two possibilities are considered: (i) the current conformation is discarded and the next one is tried out (the new conformation is so similar to one already selected that it does not increase diversity) or (ii) the most similar conformer already selected is replaced (the new conformation is similar to one already selected but it adds more diversity to the set). In the latter case (|S| < MaxConfs), it remains to be decided whether (i) to discard the current conformation and try out the next one, (ii) to insert it into the selected list (the new conformation increases diversity), or (iii) to replace another conformer from the list, that is, the most similar conformer already selected.

Decisions to be made in 2 above depend on a similarity cutoff (s_{cutoff}) to guarantee the maximum diversity and on a series of comparisons. Suppose there is, at least, one selected conformation in the list S ($S = \{C_{S1}, ..., C_{Sk}\}$). Then, (a) the similarity between the new candidate (C_N) and the set of selected conformers ($s(C_N, C_{S_k})$) is calculated, (b) the most similar conformer to C_N in the list ($C_{\max}$) is identified by calculating $s_{\max}(C_N) = \max(s(C_N, C_{S_k}))$, and (c) the mean similarity value between C_N and the rest of selected conformers excluding $C_{\max}$ ($\overline{s(C_N)} = \sum_k s(C_N, C_{S_k})/(|S| - 1)$) is computed. Then, if the list is already full and $s_{\max}(C_N) > s_{\text{cutoff}}$, C_N is rejected because diversity would not be increased. However, if $s_{\max}(C_N) < s_{\text{cutoff}}$, C_N will be selected only if it adds more diversity to the list, that is, if $\overline{s(C_N)} < \overline{s(C_{\max})}$. On the other hand, if the list is not yet completed, any C_N will be automatically selected unless $s_{\max}(C_N) > s_{\text{cutoff}}$. If C_N is selected, the algorithm checks for the possibility that the list's diversity could be enriched by replacing a conformer from the list with C_N. The algorithm iterates, on the fly, over each new conformer until MaxCandidates is reached or all possible conformations have been checked in cases where N_{Possible} < MaxCandidates.

Finally, the selected set of conformers can be subjected to a steepest descent or conjugate gradient energy minimization process[13] (Figure 1e) to relax the structures because it must be borne in mind that the rules consist of discrete numerical values

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314–323

for the angles. Although minimizing the potential energy of the structures increases the computation time, the conformers thus relaxed correspond to energy minima within the context of the force field employed.

The algorithm requires only one input parameter from the user, which is MaxConfs. It is worth remarking that in many cases the value of MaxConfs is not even reached since the algorithm tries to provide the minimum number of conformations while covering the maximum diversity (Figure 4, Results section).

An adequate choice of s_{cutoff} is of main concern because it is critical to achieve a good performance in identifying at least one conformer from the list close enough to one of the experimental structures. Taking the set of ligands from the ASTEX diverse test set[8] as a benchmark, we computed their conformations over a range of s_{cutoff} values, namely 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, and 0.95. The performance of each set was assessed by measuring the RMSD between the coordinates of each conformer in a given set and those from its experimental counterpart (X-ray structure). The best s_{cutoff}, on average, for a given number of rotatable bonds was selected (Table 1).

Table 1. Different Values for s_{cutoff} Depending on the Number of Rotatable Bonds[a]

number of rotatable bonds	s_{cutoff}
$N < 7$	0.75
$N = 7$	0.80
$8 \leq N < 10$	0.85
$10 \leq N < 12$	0.90
$N \geq 12$	0.95

[a]Ligands were taken from the ASTEX test set.[8]

Ligand structures were prepared according to common protocols described elsewhere.[18,19]

ALFA Input/Output Formats. ALFA accepts PDB and MOL2 formatted files as input for ligand structures and produces multi-PDB and multi-MOL2 files containing the set of selected conformations.

Programming Details. The algorithm has been implemented in C++ programming language using the OpenBabel C++ toolkit, from which we have used different subroutines for reading and writing PDB and MOL2 formatted files, SMARTS patterns matching, partial charge assignment, rotational searches, and energy refinement.

The ES methodology was implemented in a small and simple tool written in C++ that was used to measure the similarity between each conformer and its corresponding experimental structure, as well as the similarities between all the conformers generated for each molecule in the ASTEX and validation sets (see below).

RMSD Fitting Tool. A new RMSD fitting tool has been developed (the VF2-based algorithm[20]) which first builds an equivalency matrix between two graphs and then iteratively reduces the level of requirements to consider two atoms as equivalent whenever the algorithm is unable to find a suitable match with the initial coordinates, most likely due to a poor template geometry. This means that bond order is only taken into account when there is, at least, an equivalent atom in one molecule for each atom in the second molecule. To avoid considering the excessive number of possibilities that arise from these matrices the program applies a shuffling algorithm to find the equivalency between the two molecules with the lowest RMSD.

ALFA Performance. To validate ALFA we used the test set of 667 ligands published by Watts et al.[7] The complexes were downloaded from the PDB database and the ligand coordinates were saved as the reference structures; in the case of multiple ligand entries within the same file, only the first set was considered. The 3D coordinates were then transformed into SMILES strings using the OpenBabel toolkit (except for the complex in PDB entry 1QBV, which contains an incomplete set of coordinates for the ligand). These strings were then used as input for CORINA[11] to generate a set of idealized 3D geometries for each ligand (including distinct ring conformations when flexible cycles were present) to which MMFF94 atomic charges were assigned using OpenBabel. Finally, we used ALFA to generate conformational diversity using default parameters as explained above.

As the accuracy of the method depends on the MaxConfs parameter, ALFA calculations were performed with different values for MaxConfs, namely 20, 50, 100, and 200. Moreover, since ALFA employs a random algorithm for conformer selection, each calculation was repeated five times taking the average value for comparisons using MaxCandidates = 300 000.

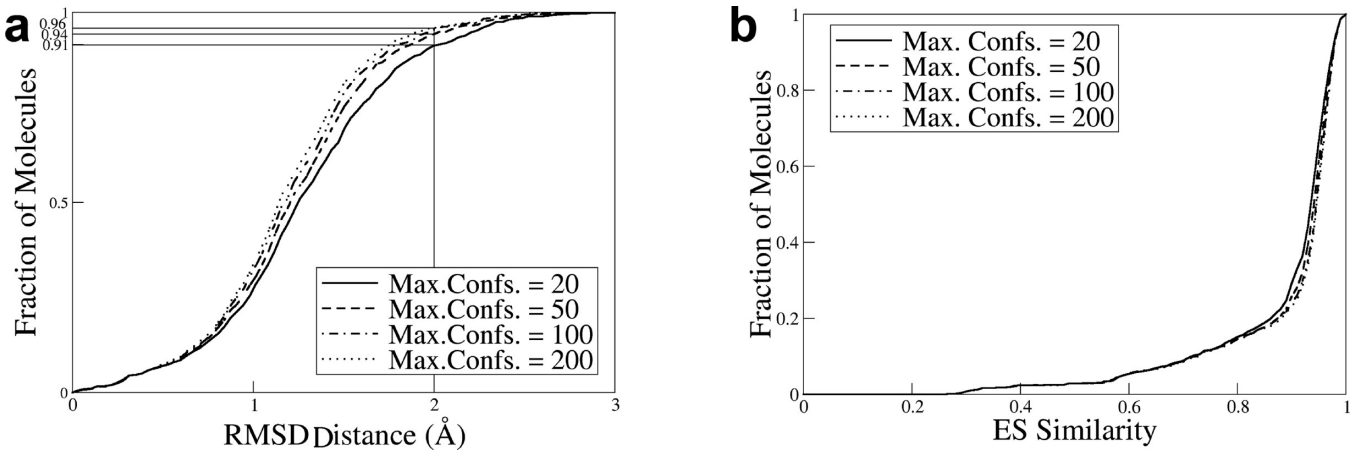


Figure 2. Cumulative density plots for (a) minimum RMSD and (b) maximum ES similarity measures depending on the number of conformers accepted for the test set.[7]

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

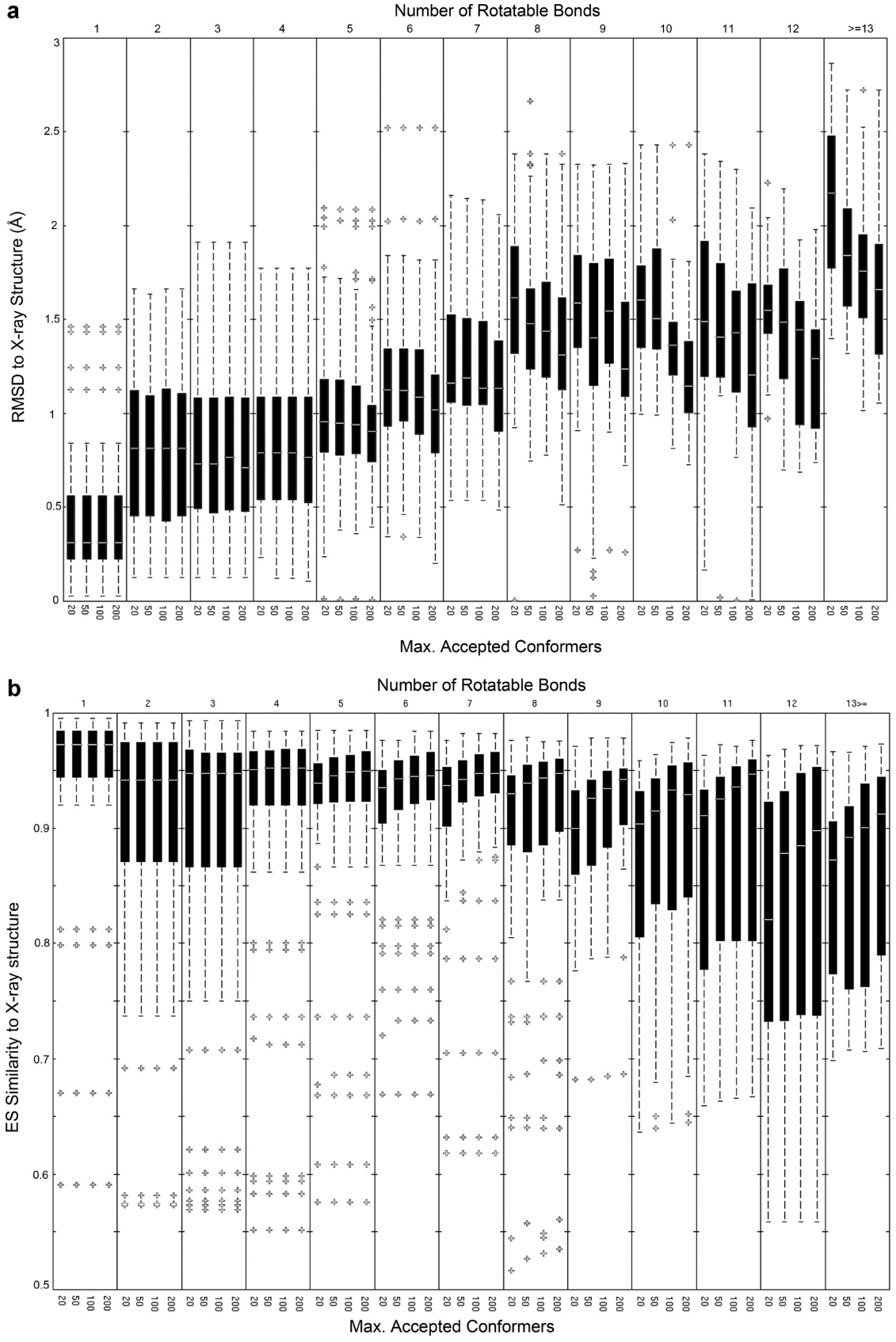


Figure 3. Boxplots showing ALFA accuracy dependence on the number of rotatable bonds and conformers accepted for (a) minimum RMSD and (b) maximum ES similarity measures using the test set.[7] In both figures outliers are marked as crosses.

ALFA Graphical User Interface. All of the functionalities available in the ALFA code have been implemented in a GUI written in the Python programming language. This allows its facile use as a plugin to the popular molecular visualization

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

program PyMOL. The GUI front end uses PyMOL software version 1.2 or higher and has been prepared to be executed on Linux operating systems. Only negligible amounts of main memory or storage capabilities are required to run and hold the outcomes of the calculations.

RESULTS

Accuracy. Program performance was defined by its capability to reproduce the experimental structure of the molecules in the test sets. To this end, RMSD and ES similarity between each conformer and its corresponding experimental structure were measured, and those conformers with the minimum RMSD and the maximum ES similarity, respectively, were selected. Figure 2 shows the cumulative density plots for both metrics in each set of conformers (MaxConfs = 20, 50, 100, and 200).

In terms of the RMSD (Figure 2a) and depending on MaxConfs (20, 50, 100, and 200), it was found, as expected, that as the number of selected candidates increases so does the accuracy (92%, 94%, 96%, and 96%, respectively). On the other hand, no significant differences were observed when ES similarity was used (Figure 2b) as the accuracy of ALFA was almost the same irrespective of MaxConfs values.

The dependence of the accuracy on the number of rotatable bonds was also tested for different values of the MaxConfs parameter (20, 50, 100, and 200; Figure 3).

It is worth remarking that all of the generated conformers have an RMSD value lower than 2 Å for molecules containing up to six rotatable bonds (Figure 3a) whereas ES similarity is equally good (similarity above 0.85) up to nine rotatable bonds (Figure 3b). Moreover, we also noted that in most cases MaxConfs was not reached (Figure 4).

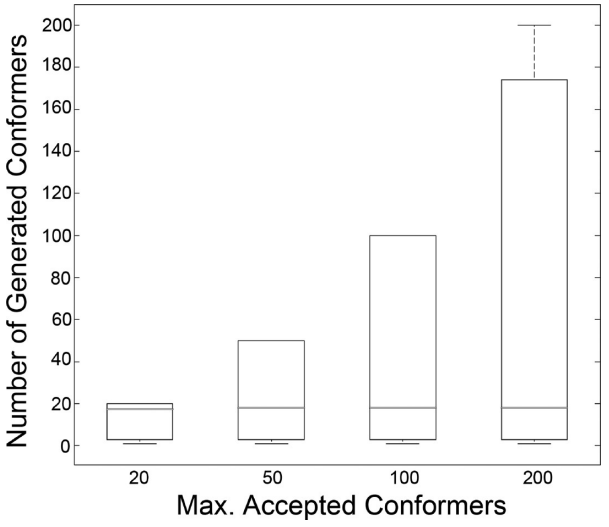


Figure 4. Boxplots depicting the relationship between the number of generated conformers and different values of MaxConfs.

Computational Performance. Runtimes for the execution of ALFA (in a logarithmic scale) per number of rotatable bonds using different values for MaxConfs (20, 50, 100, and 200) are shown in Figure 5. Calculations for small molecules (up to seven rotatable bonds) can take ∼1 s; whereas, for more complex ligands runtimes increase to ∼10 s. These tests were performed on a 2007 Intel Xeon at 3.0 GHz with a cache size of 512 Kb and only negligible amounts of main memory or disk space were required to run and store the outcomes of the calculations, respectively.

Diversity. Obtaining a diverse set of conformers can be critical for different purposes, for example in virtual screening or in pharmacophore modeling. Figure 6a and b shows boxplots for the different all-versus-all RMSD and ES similarity values, respectively, depending on the number of rotatable bonds and for different values of MaxConfs.

ALFA Graphical User Interface. Following the current trend of producing easy-to-use software for researchers who do not have a strong theoretical chemistry background, we have designed and developed a simple GUI that is integrated as a plugin within the popular PyMOL molecular graphics program.[16]

Once the module is executed, a window with two tabs is shown. The first tab (Figure 7a), **Configure**, is used to name the ALFA executable path and the working directory that will store the results and the temporary files created when a new conformational analysis is performed. The second tab (Figure 7b) contains four different configurable sections where the user can select: (i) the names for the input, output, and reference files (the third one optional, as it is used only if RMSD calculations are required); (ii) the maximum number of conformers to be generated and the maximum number of conformers to be written out; (iii) the method to group the conformers (either maximum dissimilarity according to the ES method or a force field-based energy cutoff); and (iv) force field-based energy minimization of selected conformers (either with the steepest descent or conjugate gradient method) and definition of the ending criterion (maximum number of steps or an energy convergence cutoff). Finally, on the right-hand side, the plugin shows the output from ALFA and gives the user the possibility of running/stopping a job or loading the results of the last job into the PyMOL graphics screen.

DISCUSSION

We describe a new computational tool named ALFA that provides a plausible set of diverse conformations for a given small molecule in a very short time. This set contains some structures that are close to the target-bound experimental conformation of the ligand and others representing some of the conformations that the isolated molecule can adopt in solution or bound to other targets. A typical workflow consists of three distinct steps: (1) each rotatable bond is typified and a set of angular arrangements is assigned, (2) a selection algorithm is applied with the aims of achieving enough diversity while covering as much conformational space as possible, and (3) energy minimization of selected conformations is (optionally) performed. A similar approximation is implemented in the program OMEGA,[21] which has been recently reviewed.[22]

Using ALFA as the conformer generator and the CRDOCK program[18] as the docking engine, we have shown that even small deviations in critical parts of the molecule between the best selected conformers and the experimental X-ray structure can have a notable negative influence on docking performance. At the core of this problem is the discretization of the dihedral angle values that are stored within the rules, which in some cases can be very different from the experimental ones.

It has been argued that evaluating docking performance on the sole basis of comparisons between experimental (X-ray) and modeled conformations using the RMSD may introduce a certain bias and have a detrimental impact on the results. That is, some conformers could be discarded, especially those with high

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

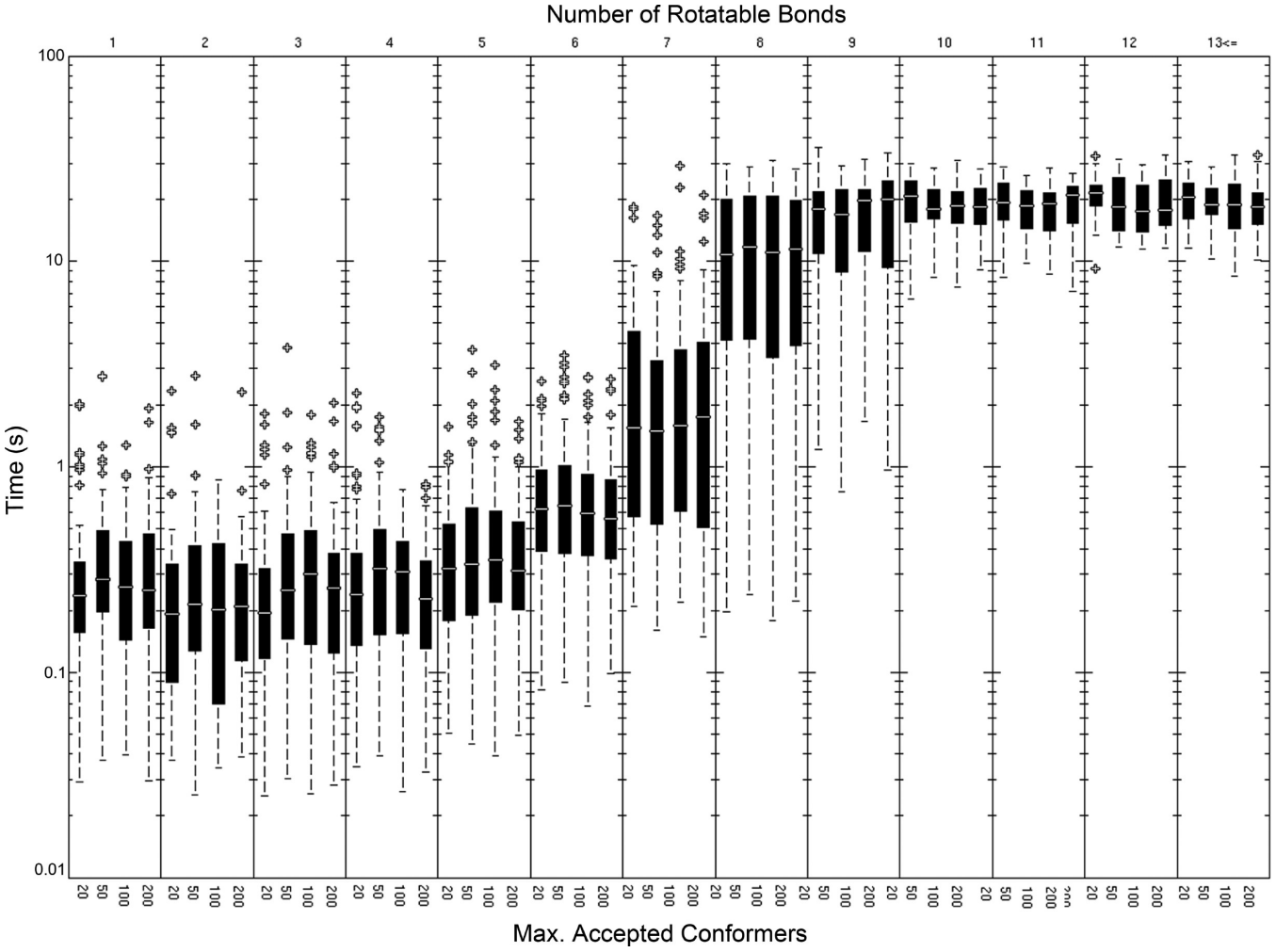


Figure 5. Boxplots showing ALFA execution times (logarithmic scale) dependence on the number of rotatable bonds and on different values of MaxConfs. Outliers are marked as crosses. Tests were performed on a 2007 Intel Xeon at 3.0 GHz with a cache size of 512 Kb.

RMSD values, because the major differences arise from the spatial location of some distal part of the molecule that is not really essential for complex stabilization despite the fact that the main interacting core is in agreement with the experimental pose. This work, through the use of ES descriptors,[15] represents a new approach to try and overcome this limitation. Although we have experienced some of these RMSD-related problems while evaluating our results with ALFA, we realize that this is the most commonly used evaluation criterion, and so it is worth to present our results in the same way.

Let us point out that the performance shown in Figure 2a in terms of RMSD values is very similar to that reported by Ebejer et al.[6] From their *minimum RMSD* results, we can see how ALFA's average performance is similar to those of Balloon, Confab, Frog2, MOE, or RDKit. All of these programs, as well as ALFA, can generate sets of conformers with at least one displaying an RMSD value below 1 Å with respect to the X-ray structure for those molecules with up to six rotatable bonds and higher when the number of rotatable bonds increases. It is important to point out that RDKit and MOE predict the X-ray structures more accurately than do the other tools including ALFA. However, in our experience, this does not guaranty an improvement in performance when these conformations are used within a pool of others for rigid docking, as the location and classification of solutions are highly dependent on the scoring function employed. Likewise, as already published,[18] using conformations generated by ALFA for docking purposes gives good results that are comparable to those obtained with other common docking tools. Finally, it is worth remarking that the results obtained by evaluating RMSD values for small molecules is not without dangers, especially in view of the fact that some widely used programs only succeed in obtaining low RMSD values for the ligands in about 50% of the cases. We believe that the RMSD tool shown in this work, although more CPU-intensive, provides a reasonable way of measuring these values when crystallographic conformations are available, especially in cases of poor experimental ligand geometries.

In terms of computation times, ALFA reaches a significant speed-up for ligands containing up to seven rotatable bonds as it is able to generate them in less than half a second. For more complex molecules it may need up to ten seconds. Although it is not the fastest program, it is able to maintain constant computation times in terms of the number of conformers generated, irrespective of the number of rotatable bonds, especially for molecules with more than eight such bonds (Figure 5). Other tools have average times per molecule ranging from 1 to up to 18 s.[6]

Finally, when comparing our program with other similar tools, a key difference is that ALFA was designed bearing in mind the maximization of the diversity within the set of conformers, where

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

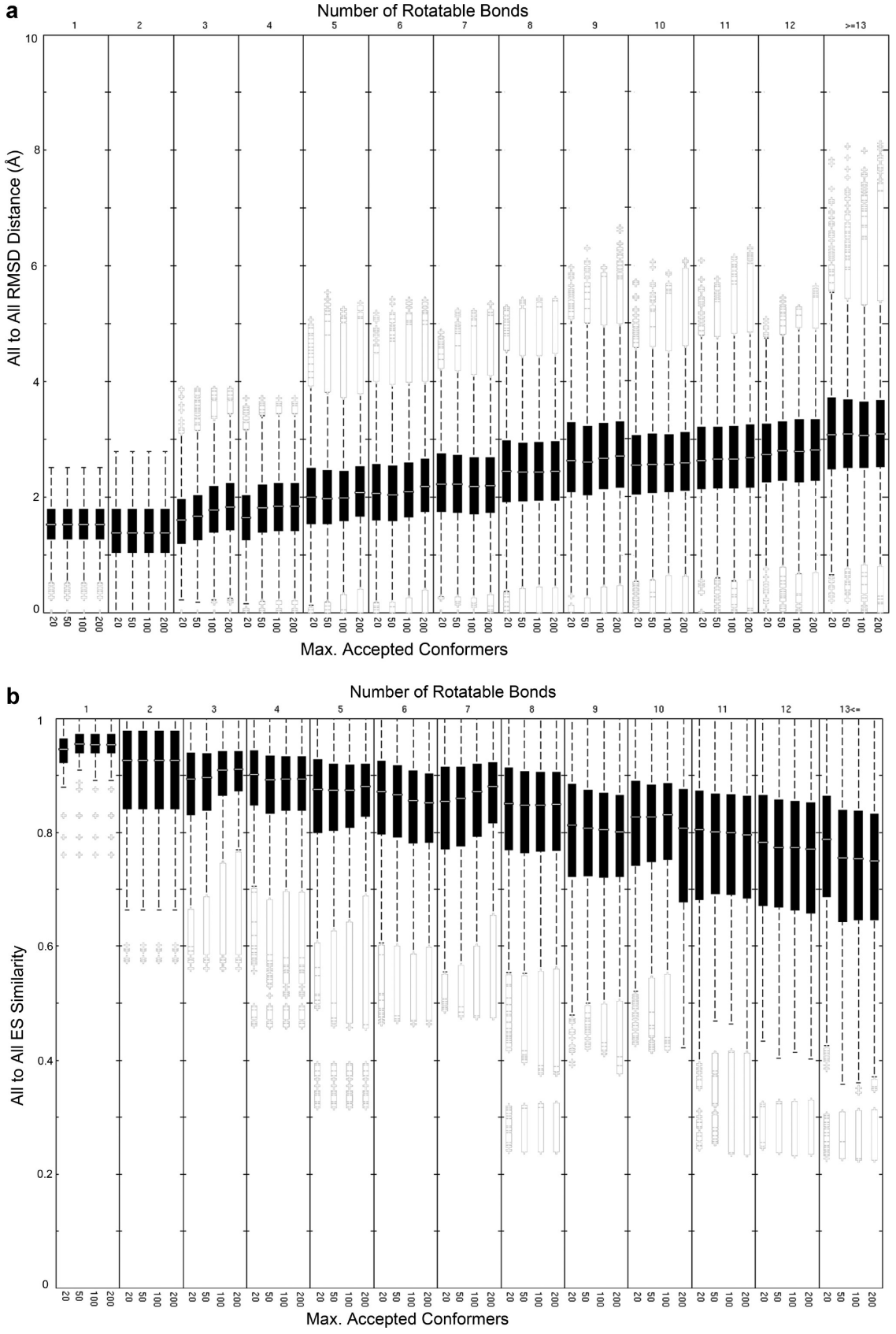


Figure 6. Conformational diversity achieved by ALFA. The all-versus-all RMSD (a) and ES similarity values (b) are displayed as a function of the number of rotatable bonds and for different values of MaxConfs.

it clearly shows a better performance, as can be inferred from Figure 7 in the work of Ebejer et al. This picture shows the distribution of the pairwise distances between conformers for molecules with 1, 3, 5, 7, 9, 11, and 13 rotatable bonds. This

dx.doi.org/10.1021/ci400453n | J. Chem. Inf. Model. 2014, 54, 314−323

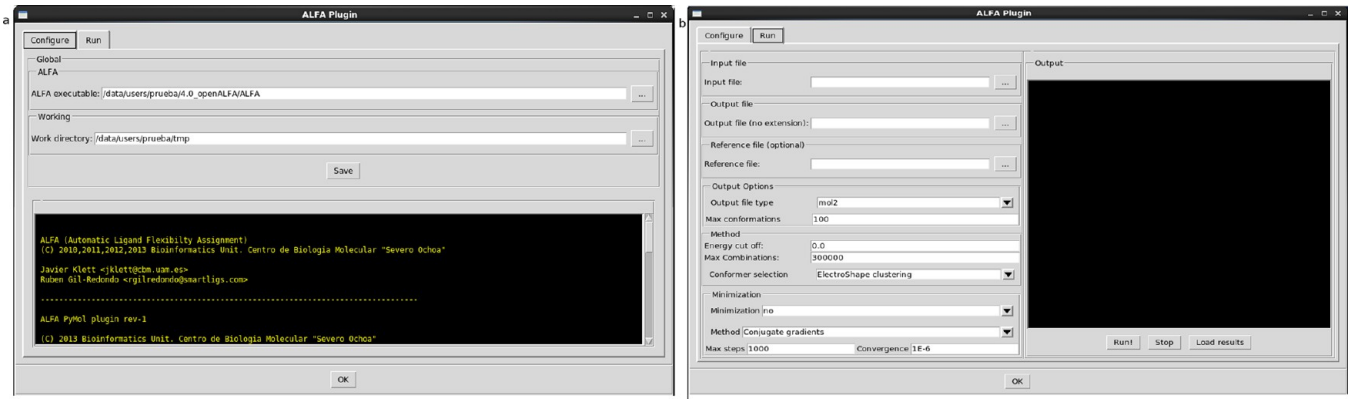


Figure 7. PyMOL plugin.

metric can be used as a diversity indicator: a smaller average distance implies closer, more structurally similar conformers and, therefore, a reduced diversity. In all cases, we observe a significant improvement when ALFA is used. For small molecules (up to 6 rotatable bonds) ALFA increases by ∼0.5 Å the average of the pairwise distance distribution, whereas for more complex molecules (above 6 rotatable bonds) this increase reaches 1 Å. This suggests that ALFA generates a set of conformations which, on average, is more diverse than those produced by other tools, while keeping similar accuracy and time scales.

The PyMol plugin we have created for ALFA represents a novelty and also possibly an advantage for the scientific community at large as it makes the calculation easily accessible to nonexpert users (e.g., chemists without a formal training in theoretical chemistry and/or modeling) who desire a trustworthy conformational analysis of small molecules without knowing the intricate details of the methods. Default parameters are meant to work well in general cases, while highly complicated molecular structures (not commonly found in drug-like molecules) will require the fine-tuning of certain parameters. Although it is true that other tools, e.g. Autodock Vina,[23] provide a docking engine that considers ligand flexibility within the target binding site they do not offer, unlike ALFA, a simple conformer generator that can be used in the absence of the receptor. Therefore, an additional benefit from the application we present here is the possibility of generating a set of diverse ligand conformations to be used as different starting points for molecular dynamics (MD) simulations to improve the sampling of phase space. This kind of studies can also be conducted when the experimental pose of a ligand within the binding site of a protein is unknown. A set of different docking solutions can always be selected and each of the corresponding complexes can be subjected to MD simulations to study their stability and analyze ligand−receptor interactions; because of its deterministic nature, MD will yield different trajectories depending on initial coordinates that can be later assessed for convergence, etc. This extended sampling procedure can also be applied to cases where the X-ray structure is unable to locate a side chain in the receptor or a functional group in the ligand.

CONCLUSIONS

ALFA represents a new approach to the problem of generating conformational variation for typical ligands using a dissimilarity criterion that is different from those used by other programs reported so far. Comparisons, which take into account not only shape but also electrostatics, are performed in a very fast way. This helps to achieve diversity within the set of generated conformers, not just in terms of molecular shapes but also in terms of charge distributions around the molecules, which can be a crucial aspect for success in rigid docking procedures. This new tool has been implemented as a user-friendly plugin for the popular molecular graphics PyMOL and is made freely available upon request for the benefit of the scientific community at large.

ASSOCIATED CONTENT

Supporting Information

Computer readable file (ci400453n_si_001.xlsx) that contains a detailed list of rules and preferred values for dihedral angles together with a supplementary figure showing the prototype structures used for rule derivation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: antonio.morreale@repsol.com. Telephone number: +34 917 536 489.

Present Address

[‖]Repsol Technology Center,Ctra. de Extremadura, A-5, km 18, Móstoles, E-28923 Madrid, Spain.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the Spanish CICYT (SAF2009-13914-C02-02 and SAF2012-39760-C02-02 to F.G.) and Comunidad Autónoma de Madrid (S-BIO-0214-2006 [BIPEDD] and S2010-BMD-2457 [BIPEDD2] to A.M. and F.G.). A.C.C. is grateful to the Spanish Ministry of Education for a research fellowship (FPU 2009-0203). J.K. was supported in part by a grant from the Spanish Ministerio de Economía y Competitividad (BFU2011-24595). A.M. gratefully acknowledges financial support from Fundación Severo Ochoa through the AMAROUTO program.

REFERENCES

(1) 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.

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

(3) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab-Systematic generation of diverse low-energy conformers. *J. Cheminform.* **2011**, *3*, 1−9.
(4) Andronico, A.; Randall, A.; Benz, R. W.; Baldi, P. Data-driven high-throughput prediction of the 3-D structure of small molecules: review and progress. *J. Chem. Inf. Model.* **2011**, *51*, 760−776.
(5) Pérez-Nueno, V. I.; Ritchie, D. W. Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model.* **2011**, *51*, 1233−1248.
(6) Ebejer, J. P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52*, 1146−1158.
(7) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534−546.
(8) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins: Struct., Funct., Bioinf.* **2002**, *49*, 457−471.
(9) Fox, S.; Wallnoefer, H. G.; Fox, T.; Tautermann, C. S.; Skylaris, C. K. First principles-based calculations of free energy of binding: application to ligand binding in a self-assembling superstructure. *J. Chem. Theory Comput.* **2011**, *7*, 1102−1108.
(10) 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.
(11) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(12) Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, 115−119.
(13) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 1−14.
(14) *OEChem*, version 1.7.2.4; OpenEye Scientific Software: Santa Fe, 2012.
(15) Armstrong, M. S.; Morris, G. M.; Finn, P. W.; Sharma, R.; Moretti, L.; Cooper, R. I.; Richards, W. G. ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 789−801.
(16) *The PyMOL Molecular Graphics System*, version 1.5.0.4; Schrödinger, LLC.
(17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(18) Cortes Cabrera, A.; Klett, J.; Dos Santos, H. G.; Perona, A.; Gil-Redondo, R.; Francis, S. M.; Priego, E. M.; Gago, F.; Morreale, A. CRDOCK: an ultrafast multipurpose protein-ligand docking tool. *J. Chem. Inf. Model.* **2012**, *52*, 2300−2309.
(19) Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, Á.; Perona, A.; Gil-Redondo, R.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: An Ultrafast and Accurate Scoring Function for Protein−Protein Docking. *J. Chem. Theory Comput.* **2012**, *8*, 3395−3408.
(20) Cordella, L. P.; Foggia, P.; Sansone, C.; Vento, M. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1367−1372.
(21) Hawkins, P. C.; 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.
(22) Hawkins, P. C.; Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.
(23) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.