Novel Method for the Evaluation of 3D Conformation Generators

Terufumi Takagi,* Michiko Amano, and Masaki Tomimoto

Pharmaceutical Research Division, Takeda Pharmaceutical Co., Ltd., 2-17-85, Juso-honmachi, Yodogawa-ku, Osaka 532-8686, Japan

Received October 23, 2008

Conformation generation is a common and key process of computer-aided drug design. The reliability of the docking simulations, pharmacophore development, and 3D-QSAR analyses depends on the accuracy of conformations of small molecules used as input information for each program. Many conformation generators have been developed with the aim of efficiently generating all the putative bound conformations that small molecules adopt when they interact with macromolecules. Conformation generators have been evaluated by whether they can reproduce the experimentally determined bioactive conformations of bound small molecules. These bioactive conformations are usually obtained from publicly available crystal structures of protein−ligand complexes. However, it is difficult to obtain 2 or more than 2 bioactive conformations of one compound because multiple complex structures of a single molecule with various macromolecules are rarely available. Present methods, therefore, simply check whether a set of generated conformations includes the corresponding bioactive conformation. The overall validity of the entire set of generated conformations against bioactive conformation space has never been checked. In this work we developed a novel method for the evaluation of conformation generators, which makes it possible to measure the performance of a conformation generator based on its ability to reproduce the overall bioactive conformation space. We also determined the optimum parameter sets for OMEGA (OpenEye) based on the coverage of bioactive conformation space and computational efficiency. Our evaluation method elucidated that increasing the number of generated conformations is not necessary to obtain better reproducibility of the overall bioactive conformation space. Our method can be applied to the evaluation of the algorithm and/or design of the conformation generator program itself.

INTRODUCTION

Conformation generation is an important process of computer-aided drug design, such as structure-based drug design and ligand-based drug design.[1] The properly generated conformations of small molecules is a key to success of pharmacophore development, 3D-QSAR analyses, docking simulations, and so on. Therefore, it is essential to have accurate conformation generating algorithms.

The term "accurate conformation generation" generally means the generation of the conformation ensemble containing all the putative bound conformations that a small molecule adopts when it interacts with macromolecules. Perola and Charifson showed in their analysis of the Protein Data Bank (PDB) that experimentally determined bioactive conformations (referred to as "X-ray bioactive conformation" in the following sections) of small molecules with 1 to 3 rotatable bonds are slightly distorted, and the internal energy gap from the corresponding energy minimum conformations is usually 3.3 kcal/mol or less.[2] This observation can be inferred that the energy requirement for the bound conformation of a compound with 1 to 3 rotatable bonds is to be within 3.3 kcal/mol of its global energy minimum. They also showed that the internal energy gap for the X-ray bioactive conformation of a compound with 4 to 6 rotatable bonds is usually 6.9 kcal/mol or less.[2] It can also be inferred that the energy requirement for the bound conformation of a compound with 4 to 6 rotatable bonds is to be within 6.9 kcal/mol of its global energy minimum. Many conformations generated from one compound satisfy the energy requirement for the bound conformation. Thus, the conformation ensemble generated by accurate conformation generation should contain only the putative bound conformations and all the putative bound conformations. The term "conformation ensemble that contains all the putative bound conformations" in the following sections means the conformation ensemble contains only the putative bound conformations and all putative bound conformations.

Several conformation generators, using various algorithms, have been developed for generating more accurate conformation ensemble, such as Catalyst,[3,4] Confort,[5] Corina,[6] Flo99,[7,8] Macromodel,[9] and OMEGA,[10] and they have all been rigorously evaluated to investigate which program is the best for a conformation generator.[11−17] In these studies, conformation generators have been evaluated by their ability to reproduce the X-ray bioactive conformation of a small molecule bound to a macromolecule. A test set of the X-ray bioactive conformations is usually prepared from publicly available crystal structures of protein−ligand complexes in the PDB.[18] However, it is difficult to obtain 2 or more than 2 X-ray bioactive conformations of one compound because multiple complex structures of a single molecule with various macromolecules are rarely available. Present methods, therefore, simply confirm that a set of generated conforma-

* Corresponding author phone: +81-6-6308-9053; fax: +81-6-6308-9010; e-mail: takagi_terufumi@takeda.co.jp.

10.1021/ci800393w CCC: $40.75
Published on Web 05/12/2009

tions includes a X-ray bioactive conformation that is one of many putative bound conformations.[16] The overall validity of generated conformations and the completeness of bioactive conformation space have been ignored.

In this work, we have developed a novel evaluation method for conformation generators based on the coverage of bioactive conformation space and computational efficacy. By using this method, we can evaluate whether conformation generators actually generate conformation ensemble that contains all the putative bound conformations. This method makes it possible not only to optimize calculation conditions for conformation generators but also to compare the accuracy of conformation generators based on their ability to reproduce the overall bioactive conformation space.

EXPERIMENTAL METHODS

1. Bioactive Conformation Space. In this study, "bioactive conformation space" means conformation ensemble that contains all the putative bound conformations. To evaluate conformation generators accurately, it is crucial to have a standard bioactive conformation space. We constructed the standard bioactive conformation space by the stochastic conformation search of MOE, in which all bonds, including ring and side-group bonds, are allowed to rotate randomly and simultaneously to generate energetically allowed conformations.[19] Since stochastic conformation searches generate conformations randomly, some putative bound conformations may be unable to be generated. Thus, it might be argued that stochastic conformation searches are unsuitable for generating conformations exhaustively compared with systematic conformation searches, in which conformations are generated by incrementing dihedral angles systematically. However, all conformations can be generated only when systematic conformation searches are applied with minute dihedral angle increments for all rotatable bonds. From a practical point of view, users must set rougher increments and/or fix dihedral angles to reduce the calculation time while maintaining the quality of the conformation search. As torsion angles are changed systematically by a fixed increment, some conformations are rejected because of high internal energy arising from collisions of atoms, even though very similar conformations are acceptable. Furthermore, the conformations of ring structures, such as piperidine, azepine, and steroids, depend on their initial conformations, because dihedral angles in a ring cannot vary independently, throughout systematic conformation searches. On the other hand, stochastic conformation searches can avoid these issues when they are given the sufficient number of times for the iterative process of stochastically generating conformations. When we generate conformations with a stochastic conformation search, there is a risk that conformations are generated unevenly. However, this risk is reduced with the sufficient number of times for the iterative process. Therefore, we selected the stochastic conformation search to construct the standard bioactive conformation space. Since most conformation generators have been developed based on the concept of systematic conformation searches, it should be preferable that the standard bioactive conformation space for evaluating those conformation generators is constructed by a different method.

We defined the required conditions for the bound conformations based on Perola and Charifson's analysis.[2] They analyzed 150 protein−ligand complex structures in the PDB, elucidated 90% of the compounds with 1 to 3 rotatable bonds, and adopted their bound conformations with the internal energy gap of 3.3 kcal/mol or less from the corresponding energy minimum conformations (the gap of 6.9 kcal/mol or less for compounds with 4 to 6 rotatable bonds). Therefore, for a compound with 1 to 3 rotatable bonds, we generated conformations with 3.3 kcal/mol or less internal energy by using the stochastic conformation search of MOE (ver. 2007.0902)[20] and defined the generated conformation ensemble as its bioactive conformation space. Likewise, for a compound with 4 to 6 rotatable bonds, we generated conformations with 6.9 kcal/mol or less internal energy gap and defined the generated conformation ensemble as its bioactive conformation space. Ideally, these bioactive conformation spaces contain all the putative bound conformations.

Parameters of the stochastic conformation search of MOE assigned nondefault values are listed below.

Forcefield: The *forcefield* setting determines what type of forcefield parameter set is used in the refinement process—that is, initial energy calculation and energy minimization of generated conformations. As we defined the required conditions for the bound conformations based on Perola and Charifson's analysis,[2] we set this parameter to MMFF94[21−23] as in their analysis.

Dielectric constant: The *dielectric constant* setting determines the dielectric constant of the electrostatic energy calculation in the refinement process. This parameter was set to $4 \times r$ (r = interatomic distance) for the same reason described above.

Cartesian Minimization rms Gradient: The *Cartesian minimization rms gradient* setting determines the termination of the energy minimization in the refinement process. When the root-mean-square gradient falls below the given value, the energy minimization is terminated. A bound conformation is not necessarily a local energy minimum conformation, but it should not be highly distorted with very high internal energy. Thus, moderate refinement instead of full refinement (which always leads to a local energy minimum) must be applied. Therefore, this parameter was set to 1.0, which is 1000 times greater than the default value.

Energy Cutoff: The *energy cutoff* setting determines the energy window of the internal energy gap from the energy minimum conformation. When the internal energy gap of a generated conformation does not fall within the energy window, the conformation is rejected. The lower limit of the energy window is set to the internal energy of the lowest energy conformation among all generated conformations. As we defined the required conditions for the bound conformations based on Perola and Charifson's analysis,[2] we set this parameter to 3.3 or 6.9 (kcal/mol) according to the number of rotatable bonds of a compound.

Failure Limit: The *failure limit* setting determines the termination of the conformation search process. When all attempts to generate a new conformation result in conformations already generated continuously more than a given number of times, the conformation search process is terminated. This parameter was set to 1000, which is 50 times greater than the default value to generate conformations thoroughly.

Rms Tolerance: Throughout the conformation search process, two conformations with their heavy-atom root-mean-

square distance (rmsd) value less than a given value are considered duplicates. In the superposition process, molecular symmetry is taken into account (e.g., an unsubstituted phenyl group). The *rms tolerance* was set to 0.3 (Å), because we considered that two conformations with their heavy-atom rmsd values less than 0.3 are generally the same.[24]

Conformational Limit: The *conformational limit* setting determines the limit on the resultant number of conformations. If a value *i* is specified, after the refinement process and the removal of duplicates, only *i* lowest energy conformations are retained and the rest are discarded. This parameter was set to 100,000, which was large enough to allow the number of conformations to be controlled later with a combination of the *energy cutoff*, the *failure limit*, and the *rms tolerance* settings.

Iteration Limit: The *iteration limit* setting determines the maximum number of attempts to generate a new conformation in the overall conformation search process. Although it is desirable to continue the attempts infinitely for constructing the true bioactive conformation space, the conformation search process must be halted to get a result. Considering the balance between calculation time and exhaustiveness, this parameter was set to 1,000,000, which was 100 times greater than the *conformational limit* value described above.

MM Iteration Limit: The *MM iteration limit* setting determines the maximum number of energy minimization steps for each conformation. This parameter was set to 10,000, which was large enough that all conformations were always energy minimized until the rms gradient fell below the specified *Cartesian minimization rms gradient* value described above.

2. The Least-Squares Superposition Procedure. In this study, we considered that a conformation is successfully reproduced by a conformation generator when the heavy-atom rmsd value, between the conformation and the best-fitting conformation in an ensemble generated by the conformation generator, was less than 0.5 Å after superposition, according to the procedure applied by Boström et al.[12] This process was used to check whether the bioactive conformation spaces contained the X-ray bioactive conformations as well. When a heavy-atom rmsd value between the X-ray bioactive conformation and the best-fitting conformations in the bioactive conformation space was less than 0.5 Å, we considered that the X-ray bioactive conformation is contained in the bioactive conformation space.

3. Occupancy Rate against Overall Bioactive Conformation Space. To evaluate a conformation generator accurately, we calculated the occupancy rate of a conformation ensemble generated by the conformation generator against the bioactive conformation space. In this study, when the heavy-atom rmsd value between a putative bound conformation in the bioactive conformation space and the best-fitting conformation in the generated conformation ensemble with a conformation generator was less than 0.5 Å, we considered that the putative bound conformation was successfully reproduced by the conformation generator. Checking whether each putative bound conformation was reproduced or not allowed us to evaluate to what degree the bioactive conformation space was reproduced by the conformation generator. The occupancy (reproducibility) rate was calculated as follows

$$\text{Occupancy (reproducibility) rate (\%)} = 100 \times N_{\text{reproduced}}/N_{\text{overall}}$$

where $N_{\text{reproduced}}$ is the number of putative bound conformations reproduced by the conformation generator, and N_{overall} is the number of putative bound conformations in the bioactive conformation space.

When the occupancy rate of a conformation ensemble generated by a conformation generator against the overall bioactive conformation space is nearly 100%, the conformation generator is regarded to be excellent because it reproduces almost all putative bound conformations. On the other hand, if the occupancy rate is very low, despite having reproduced some X-ray bioactive conformations, a conformation generator should be regarded as being of poor quality.

4. Protein−Ligand Complexes. In this study, we used 36 protein−ligand complex structures in the PDB for optimizing the calculation conditions of OMEGA 1.8.1, one of the most widely used conformation generators, with our novel evaluation method based on the occupancy rate against the overall bioactive conformation space. Although our evaluation method is innovative in that we can evaluate conformation generators without using any X-ray bioactive conformations, we decided to use protein−ligand complex structures for two reasons. The first is to confirm whether the bioactive conformation space constructed by the stochastic conformation search of MOE definitely contains the X-ray bioactive conformations. The second is to check whether OMEGA 1.8.1, optimized based on the coverage of the bioactive conformation space, reproduces the X-ray bioactive conformation.

We selected 36 protein−ligand complex structures same as previously selected by Boström et al.[12] The criteria are as follows:

- Crystal structures should have high resolution (≤2.0 Å).
- The B-factor of the ligands should be small (preferably below 30).
- The ligands should not include rotatable bonds that cannot be determined by protein crystallography (e.g., hydroxyl torsions).
- The ligands should not include unusual moieties.
- The ligands should be reasonably small, flexible, and druglike.

5. OMEGA 1.8.1. OMEGA 1.8.1[10] is one of the most widely used conformation generators in the computer-aided drug design. In this study, we optimized the calculation conditions of OMEGA 1.8.1, based on the occupancy rate against the overall bioactive conformation space, as a practical example of our novel evaluation method for conformation generators.

OMEGA 1.8.1 generates a conformation ensemble in three steps, as follows. The first step involves the generation of a starting conformation from the connection table of the input small molecule. This employs a random-coordinate distance-geometry algorithm[25] to generate starting conformations that are refined using the MMFF[21−23,26,27] as a force field for the following torsion-driving step. In the second step, OMEGA generates multiple conformations from each starting conformation, generated in the first step, by using a depth-first, divide-and-conquer algorithm, where the input structure from the first step is transformed into small fragments that are then joined together to build the molecule's conforma-

tions using torsion and ring libraries for allowed torsion angles and ring conformations defined by OpenEye Science Software. A selectable energy window for conformations, using an OMEGA-specific variant of the Dreiding force field[28] favoring macromolecule-bound conformations, optionally limits which conformations are generated. In the final step, the resultant set of conformations produced in the previous step is filtered using the heavy-atom rmsd values between conformations and the internal energy gap of the conformations. No additional postprocessing, beyond this rmsd filtering, was performed in this study.

Conformation generation with OMEGA 1.8.1 was carried out with the following parameter set. Parameters assigned nondefault values are listed below.

MMFF94s: The *MMFF94s* setting determines whether the original MMFF94 force field[21−23] or the MMFF94s force field[26,27] will be used in the refinement process. Conformation ensembles generated with OMEGA 1.8.1 were compared to putative bound conformations in the bioactive conformation space, in which putative bound conformations were generated with the MMFF94 force field as described above. If our goal was a straightforward comparison between the OMEGA-generated conformation ensemble and the bioactive conformation space, it might be better to use the MMFF94 force field. However, our true goal was an evaluation of conformation generators rather than a comparison. The evaluation method for conformation generators should not be restricted by calculation conditions, and both MMFF94 and MMFF94s force fields are acceptable. We decided to use the MMFF94s force field based on the practical utility of using OMEGA 1.8.1 for the subsequent calculations, such as docking simulations and pharmacophore development. Therefore, this option was set to true.

Itermax: The *itermax* setting determines the maximum number of iterations of energy minimization in the refinement step. This parameter was set to 10,000, which is large enough to refine the starting conformations fully.

Numconfs: The *numconfs* setting determines the number of starting conformations generated with its distance geometry algorithm and the subsequent refinement. This parameter must be larger than *keepconfs* setting (see below). This parameter was set to 10.

Keepconfs: The *keepconfs* setting determines the number of the starting conformations that will move on to the second step — that is, multiple conformation generation with the torsion-driving mechanism. When the *numconfs* setting is 10 and the *keepconfs* setting is 2, OMEGA 1.8.1 generates 10 starting conformations and will pass the two lowest energy nonduplicated conformations on to the torsion-driving step. This parameter directly affects the resultant conformation ensemble, and the values used were 1, 3, or 5.

Ewindow: The *Ewindow* setting determines the energy window of the internal energy gap from the minimum energy conformation. When the internal energy gap of a generated conformation exceeds the energy window, the conformation will be rejected in the torsion-driving step. The lower boundary of the energy window is set to the internal energy of the lowest energy conformation among all generated conformations. The internal energy in this step is calculated based on the Dreiding force field.[28] The generated conformations are filtered later based on their internal energy measured with the MMFF94s force field (see *finalcut* setting described below). Therefore, this parameter was set to 1000, which is large enough to keep most of the generated conformations in the torsion-driving step.

Maxconf: The *maxconf* setting determines the maximum number of generated conformations in the final result for each small molecule. This parameter was set to 10,000, which is large enough to allow control of the number of conformations at a later stage using a combination of the *rms* and *finalcut* settings (see below).

Rms: After the torsion-driving step, two conformations with heavy-atom rmsd values less than the *rms* setting are considered duplicates. The parameter affects the generated conformation ensemble, and the values used were 0.1, 0.2, 0.3, and 0.5.

Finalcut: The *finalcut* setting determines the upper limit of the internal energy gap from the lowest energy conformation. The internal energy was measured by the MMFF94s force field. When the internal energy gap of a generated conformation exceeds the given value, the conformation will be discarded. This parameter affects the generated conformation ensemble, and the values used were 5, 10, 20, 30, 40, and 50.

Finalrms: The *finalrms* setting determines the criterion for the determination of duplicate conformations at the end of the calculation. As we did not apply any energy minimization after the torsion-driving step, this parameter was set equal to the *rms* value described above.

RESULTS

1. Construction of the Bioactive Conformation Space. The 36 protein−ligand complex structures are listed in Table 1, and their ligand structures are shown in Figure 1. This is the same set that Boström et al. used in their evaluation of OMEGA 1.0b.[12] We constructed the bioactive conformation spaces for the subset of compounds with 6 or less rotatable bonds (26 compounds), because the calculation time was too long for the remaining compounds with 7 or more rotatable bonds. The 26 bioactive conformation spaces were constructed according to the approach described in the Experimental Methods section, and the results are shown in Table 2. It is the essential prerequisite that the generated bioactive conformation space is the desired conformation space, although it is difficult to obtain any precise picture of coverage and spread of a conformation space.[29] We therefore confirmed the following two aspects of the bioactive conformation space. One aspect is the confirmation that the bioactive conformation space contains the corresponding X-ray bioactive conformation. Figure 2 illustrates the concept of the bioactive conformation space. The bioactive conformation space (gray circle) should contain all the putative bound conformations (orange triangles — conformations close in energy to the global energy minimum), the global energy minimum (black triangle), and the X-ray bioactive conformation (blue star). To confirm the presence of the X-ray bioactive conformation in the generated bioactive conformation space, heavy-atom rmsd values were calculated for the X-ray bioactive conformation against all conformations in the bioactive conformation space. We observed that for all 26 compounds, the minimum rmsd values were less than 0.5 Å (Table 2), which means that the stochastic conformation search using MOE could successfully repro-

Table 1. 36 Compounds Used in This Study

compound no.	PDB code	HET code	no. of rotatable bonds
1	1a28	STR	1
2	1tng	AMC	1
3	1tnh	FBA	1
4	1qft	HSM	2
5	1ftm	AMQ	3
6	1phg	MYT	3
7	3bto	SSB	3
8	1ia3	TQ5	3
9	1fcy	564	3
10	1d3g	BRE	3
11	1c83	OAI	4
12	1ecv	878	4
13	1fcz	156	4
14	1gr2	KAI	4
15	1ian	D13	4
16	1frb	ZST	5
17	1bju	GP6	5
18	1dyr	TOP	5
19	2izg	BTN	5
20	1cbx	BZS	5
21	5std	UNN	5
22	6std	MS2	5
23	7std	INY	5
24	1cbs	REA	5
25	1dam	DTB	6
26	3std	MQ0	6
27	1ejn	AGB	7
28	1if8	SBS	7
29	1caq	DPS	8
30	1mtv	BX3	8
31	1mtw	DX9	8
32	1pph	TOS-APM-PIP	8
33	1f0u	RPR	11
34	1fkg	SB3	11
35	1fkh	SBX	11
36	1ppc	NAS-GLY-APH-PIP	11

duce the X-ray bioactive conformation for all 26 compounds. The other aspect is the confirmation that the stochastic conformation search generates the conformation ensembles exhaustively. To check the exhaustiveness of the conformation generation by using the stochastic conformation search, we calculated occupancy rates of the bioactive conformation spaces with respect to the conformation ensemble generated by using the systematic conformation search of MOE. The systematic conformation search was applied with default parameter sets other than "on" for energy minimization after systematic search in torsion angles, "0.3" (Å) for rmsd criterion of duplicate conformations, and "3.3" or "6.9" (kcal/mol) for internal energy cutoff according to the number of rotatable bonds. As a result, we observed that occupancy rates of the generated bioactive conformation space with respect to the conformation ensemble generated by using the systematic conformation search were all 100% for the 26 compounds (data not shown). This result meant that the bioactive conformation space constructed by the stochastic conformation search completely covered the conformation ensemble generated by the systematic conformation search. These two confirmations supported the validity of the bioactive conformation space.

2. Occupancy Rate of Conformation Ensemble Generated with OMEGA 1.8.1. We first conducted a pilot study to determine the occupancy rates of the generated conformation ensembles for the 26 compounds using OMEGA 1.8.1, with respect to the bioactive conformation space. Our initial parameter set (*keepconfs* = 1, *final rms* = 0.3, and *finalcut* = 25), which we had previously used successfully, is referred to as the primal parameter set. First, heavy-atom rmsd values were calculated for the X-ray bioactive conformations against all conformations in the generated conformation ensembles. Second, occupancy rates of the generated conformation ensembles with respect to the bioactive conformation space were calculated. The minimum rmsd values and the occupancy rates are shown in Table 3. In the pilot study with the primal parameter set, 3 compounds (**14**, **21**, **26**) showed large rmsd values (>0.5 Å), which shows that we were not able to reproduce these 3 X-ray bioactive conformations. The occupancy rates in Table 3 showed that the conformation ensembles of some compounds (**2**, **3**, **4**, **8**, **11**, **12**) fully covered the bioactive conformation space, while the others did not. The occupancy rate was found to depend on the compound being examined. The occupancy rates of the 3 compounds (**14**, **21**, **26**) that did not reproduce their X-ray bioactive conformations were found to be relatively low. The reason for these low rates is presumably because the primal parameter set was insufficient to reproduce the X-ray bioactive conformations. Some compounds (**6**, **7**, **9**, **10**, **13**, **15**, **19**, **20**, **22**, **25**) gave surprisingly low occupancy rates yet successfully reproduced the X-ray bioactive conformations, which can happen when the overlap between the bioactive and the generated conformation spaces is small, but by chance the X-ray bioactive conformation happens to be found in the region of the overlap. This result supports our initial concept regarding the evaluation of conformation generators. Thus, reproducibility of the X-ray bioactive conformation is not enough to evaluate the overall validity of the generated conformation ensemble, because the X-ray bioactive conformation is just one of multiple putative bound conformations.

3. Optimization of OMEGA 1.8.1 Parameters. The 13 compounds (**6**, **7**, **9**, **10**, **13**, **14**, **15**, **19**, **20**, **21**, **22**, **25**, **26**) that gave low occupancy rates with the primal parameter set (Table 3) were examined in a search for an optimized parameter set to use in OMEGA 1.8.1. The occupancy rates of the conformation ensembles generated for these 13 compounds, with various parameter sets, are shown in Table 4. These results showed that parameter set 66 gave the best performance. In general, a larger number of generated conformations lead to higher occupancy rates and higher probabilities of reproducing the X-ray bioactive conformation. However, it is worthless to simply generate more conformations if it means that the generated conformation ensemble will contain unnecessary conformations, even if it also contains X-ray bioactive conformations. That is to say, the ideal result is to generate a conformation ensemble that contains fewer conformations while maintaining a high occupancy rate. We evaluated the results from various parameter sets, taking into account not only the occupancy rates but also the total number of generated conformations. In this manner, we selected parameter set 66. The detailed selection process was as follows:

- Search for the highest occupancy rate of a compound X.
- Select parameter sets for compound X that give occupancy rates of 80% or greater than the highest occupancy rate for that compound. (The selected parameter sets are indicated with bold letters in Table 4.)

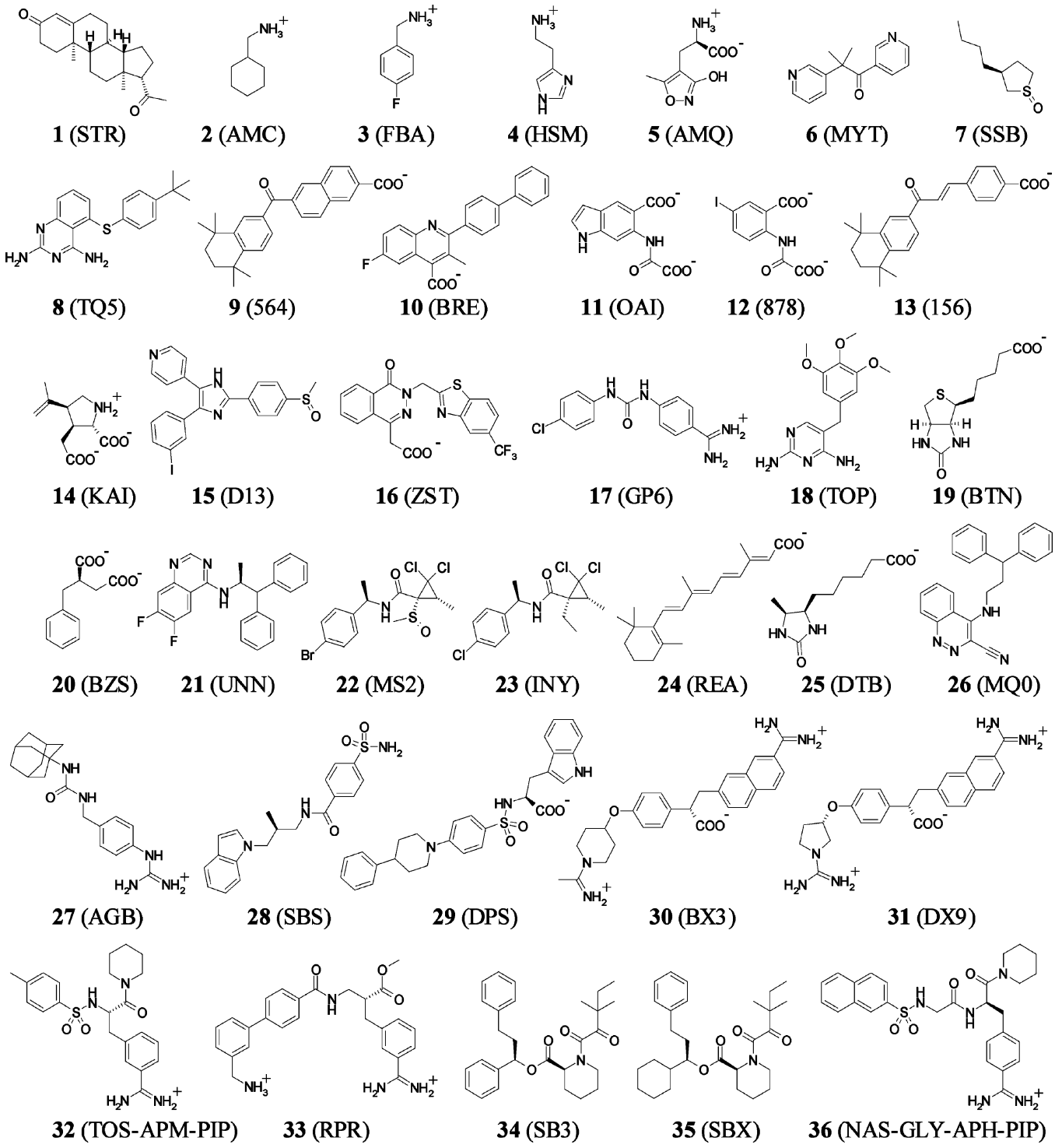

Figure 1. Chemical structures of the 36 compounds examined in this study. HET codes assigned in PDB are shown in parentheses.

- Identify the five parameter sets for compound X that require the smallest number of generated conformations. (The 5 most efficient parameter sets are indicated with underlined letters in Table 4.)
- Apply the above process to all compounds.

The results shown in Table 4 consequently indicated that both parameter sets 65 and 66 were the most efficient parameter set for 4 compounds among the tested 13 compounds. We judged parameter set 66 to provide the best performance among the tested 72 parameter sets, because parameter set 66 showed good occupancy rates for 12 compounds (shown with bold letters in Table 4) while parameter set 65 showed good occupancy rates for 11 compounds out of 13 compounds. We hereinafter refer to it as the optimized parameter set. The comparison between the primal parameter set and the optimized parameter set was shown in Table 5. The *keepconf* setting of the optimized parameter set was modified to 5 from 1 in the primal parameter set. The *finalcut* setting of optimized parameter set was modified to 50 from 25 in the primal parameter set.

In Figure 3, we showed the relationships between the number of generated conformations and the occupancy rates for each of the 13 compounds. In each graph, the 72 results shown in Table 4 were plotted, and the result from the

Table 2. 26 Standard Bioactive Conformation Spaces

compound no.	no. of conformations in the bioactive conformation space	heavy-atom rmsd value (Å)[a]
1	5	0.06
2	5	0.09
3	3	0.09
4	20	0.14
5	86	0.24
6	259	0.19
7	64	0.13
8	161	0.20
9	589	0.15
10	12	0.20
11	5	0.33
12	6	0.33
13	1684	0.19
14	186	0.14
15	1783	0.19
16	9538	0.29
17	21	0.46
18	2071	0.24
19	3242	0.15
20	580	0.25
21	489	0.39
22	428	0.20
23	315	0.18
24	2907	0.27
25	8752	0.18
26	9410	0.18

[a] The heavy-atom rmsd value between the best-fitting conformation and the corresponding X-ray bioactive conformation.

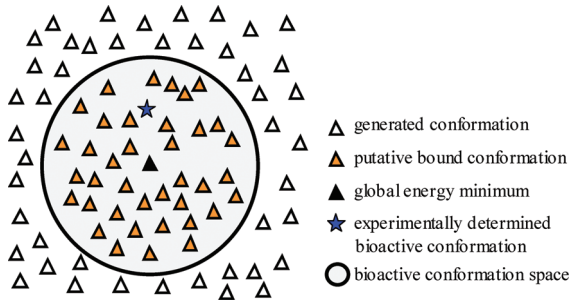

Figure 2. Conceptual diagram of bioactive conformation space. The bioactive conformation space (gray circle) should include all the putative bound conformations (orange triangles), the global energy minimum (black triangle), and experimentally determined bioactive conformation (blue star).

optimized parameter set was indicated by a red cross. For all compounds, the greater number of generated conformations was found to yield higher occupancy rates. However, Figure 3 showed that the occupancy rates did not always reach 100%, no matter how large the generated conformation ensembles were. On the other hand, there were cases where a relatively high occupancy rate could be achieved with a small conformation ensemble. The conformation ensembles generated with the optimized parameter set yielded reasonably high occupancy rates against the bioactive conformation space with much fewer conformations than the parameter set which led to the maximum number of generated conformations in all cases.

4. Reproduction of the X-ray Bioactive Conformation. It was also checked whether the optimized parameter set could reproduce the 36 X-ray bioactive conformations. The conformation ensembles of all 36 compounds were generated using OMEGA 1.8.1 with the optimized parameter set. The average calculation time for each compound was 54.9 min, and 3860 conformations were generated on average. The minimum rmsd values between the best-fitting conformation in the generated conformation ensemble and the corresponding X-ray bioactive conformation (0.38 Å on average) are shown as orange bars in Figure 4. In particular, when we focused on the first 26 compounds (**1**−**26**) with 6 or less rotatable bonds, all of their X-ray bioactive conformations were successfully reproduced (i.e., all of their minimum rmsd values were below 0.5 Å). Furthermore, the average calculation time and average size of the conformation ensembles were just 65 s and 583 conformations, respectively. The minimum rmsd values using the primal parameter set are shown as cyan bars in Figure 4. For almost all of the 36 compounds, rmsd values derived using the optimized parameter set were smaller than those from the primal parameter set. The rmsd values of 3 compounds (**14**, **21**, **26**) were drastically reduced to less than 0.5 Å, which indicated that the X-ray bioactive conformations of these 3 compounds were not reproduced with the primal parameter set but were with the optimized parameter set. The optimized parameter set was derived from the optimization of the occupancy rate against the bioactive conformation space and the size of the generated conformation ensemble. The reproducibility of the X-ray bioactive conformation was not taken into account at all. However, the optimized parameter set provided conformation ensembles with better reproducibility, suggesting that we can generate a conformation ensemble containing conformations closer to the X-ray bioactive conformation using a conformation generator optimized on the basis of the overall validity of the generated conformation ensemble.

Table 3. Conformation Ensembles Generated Using OMEGA 1.8.1 with the Primal Parameter Set

compound no.	rmsd (Å)[a]	occupancy rate (%)[b]
1	0.30	80.0
2	0.12	100.0
3	0.12	100.0
4	0.23	100.0
5	0.36	86.0
6	0.31	34.4
7	0.18	32.8
8	0.17	100.0
9	0.15	43.5
10	0.22	41.7
11	0.36	100.0
12	0.36	100.0
13	0.39	53.9
14	0.58	34.4
15	0.39	23.0
16	0.39	71.1
17	0.34	95.2
18	0.27	81.3
19	0.29	30.3
20	0.25	52.8
21	0.59	53.6
22	0.39	37.9
23	0.35	92.4
24	0.26	97.5
25	0.26	45.8
26	0.79	52.5

[a] The heavy-atom rmsd values between the best-fitting conformation and the corresponding experimentally determined bioactive conformation. [b] Occupancy rate against the bioactive conformation space.

Table 4. Occupancy Rates of Conformation Ensembles Generated by Using OMEGA 1.8.1 with Various Parameter Sets[a]

				compound no.												
parameter set	keepconfs	(final)rms	finalcut	6	7	9	10	13	14	15	19	20	21	22	25	26
1	1	0.1	5	17.0	**26.6**	35.0	41.7	35.4	13.4	12.5	7.9	17.2	15.3	18.2	11.1	18.7
2	1	0.1	10	25.5	**32.8**	**43.6**	41.7	43.5	15.1	22.9	26.5	30.7	31.5	32.5	34.0	38.9
3	1	0.1	20	33.2	**32.8**	**44.1**	41.7	**59.2**	15.1	**25.4**	33.1	**60.2**	52.4	40.7	**49.1**	53.2
4	1	0.1	30	37.5	**32.8**	**45.3**	41.7	**60.0**	73.1	**25.4**	34.9	**65.9**	**58.1**	41.8	**52.9**	57.3
5	1	0.1	40	42.1	**32.8**	**48.0**	41.7	**60.0**	**76.9**	**25.4**	35.4	**66.9**	**60.3**	42.5	**53.0**	60.3
6	1	0.1	50	42.5	**32.8**	**49.1**	**91.7**	**60.0**	**83.9**	**25.4**	41.4	**67.2**	**60.7**	54.0	**55.2**	62.5
7	1	0.2	5	17.0	25.0	35.0	41.7	34.6	12.4	12.5	7.9	14.0	15.3	18.2	11.1	18.7
8	1	0.2	10	25.5	**32.8**	**43.6**	41.7	42.5	14.0	22.9	26.1	27.1	31.5	32.5	33.8	38.9
9	1	0.2	20	33.2	**32.8**	**44.1**	41.7	**58.3**	24.2	**25.4**	32.8	51.7	52.4	40.7	**48.9**	53.1
10	1	0.2	30	37.5	**32.8**	45.3	41.7	**59.1**	**70.4**	**25.4**	34.6	**57.8**	58.1	41.8	**52.7**	57.2
11	1	0.2	40	42.1	**32.8**	**48.0**	41.7	**59.1**	**73.1**	**25.4**	35.1	**60.2**	**60.3**	42.5	**52.8**	60.2
12	1	0.2	50	42.5	**32.8**	**49.1**	**91.7**	**59.1**	**79.6**	**25.4**	41.1	**61.6**	**60.7**	54.0	**55.0**	62.3
13	1	0.3	5	16.6	25.0	35.0	41.7	29.4	11.3	7.9	14.5	14.5	14.7	16.4	15.5	16.9
14	1	0.3	10	23.9	**32.8**	**43.3**	41.7	37.4	12.4	18.5	25.3	23.8	30.9	29.2	32.9	36.2
15	1	0.3	20	32.4	**32.8**	**43.5**	41.7	**53.4**	12.4	23.0	29.3	49.0	48.9	37.4	43.9	50.1
16	1	0.3	30	35.9	**32.8**	**44.7**	41.7	**53.9**	63.4	23.0	30.7	54.3	**56.6**	38.6	45.8	54.0
17	1	0.3	40	40.5	**32.8**	**47.4**	41.7	**53.9**	**70.4**	23.0	31.1	**57.9**	**59.1**	39.3	46.0	56.8
18	1	0.3	50	40.9	**32.8**	**48.2**	**91.7**	**53.9**	**78.5**	23.0	36.4	**59.3**	**59.5**	49.8	**48.2**	58.9
19	1	0.5	5	10.0	21.9	35.0	25.0	21.6	12.9	8.4	9.6	9.7	14.3	12.1	9.5	9.8
20	1	0.5	10	16.6	25.0	35.0	25.0	29.9	12.9	14.8	18.5	17.4	29.7	21.0	18.7	21.3
21	1	0.5	20	20.8	25.0	35.7	25.0	40.3	26.3	14.8	22.4	33.1	42.5	29.0	25.7	33.2
22	1	0.5	30	22.8	25.0	36.3	25.0	40.9	60.2	14.8	23.3	37.1	51.3	30.8	26.9	36.0
23	1	0.5	40	28.2	25.0	37.4	25.0	40.9	65.1	14.8	23.6	40.5	54.2	31.3	27.0	39.8
24	1	0.5	50	28.2	25.0	38.2	58.3	40.9	**69.4**	14.8	27.0	41.0	54.4	40.2	27.0	42.4
25	3	0.1	5	27.4	**31.3**	**44.7**	**91.7**	42.1	13.4	**24.3**	11.5	21.4	26.8	21.7	22.9	19.5
26	3	0.1	10	34.4	**32.8**	**47.2**	**91.7**	44.8	15.1	**27.4**	29.6	40.0	52.6	38.6	41.8	45.7
27	3	0.1	20	42.9	**32.8**	**47.4**	**91.7**	**61.3**	25.3	**28.8**	34.9	**68.4**	**64.8**	45.1	**52.3**	65.3
28	3	0.1	30	46.7	**32.8**	**51.1**	**91.7**	**61.4**	**73.1**	**28.8**	37.4	**69.8**	**66.7**	46.0	**55.0**	**100.0**
29	3	0.1	40	**52.9**	**32.8**	**52.8**	**91.7**	**61.4**	**76.9**	**28.8**	42.3	**71.2**	**66.7**	46.5	**57.5**	**100.0**
30	3	0.1	50	**54.8**	**32.8**	**52.8**	**100.0**	**61.5**	**83.9**	**28.8**	44.4	**71.2**	**67.3**	**59.6**	**58.2**	**100.0**
31	3	0.2	5	27.4	**31.3**	**44.3**	**91.7**	41.4	12.9	**24.3**	11.5	19.8	26.6	20.6	22.7	19.4
32	3	0.2	10	34.4	**32.8**	**44.8**	**91.7**	44.3	14.5	**27.4**	29.3	35.0	52.6	36.4	40.7	44.5
33	3	0.2	20	42.5	**32.8**	**45.0**	**91.7**	**60.5**	26.3	**28.8**	34.6	**63.6**	**64.2**	41.8	**51.0**	63.4
34	3	0.2	30	46.3	**32.8**	**49.2**	**91.7**	**60.5**	**69.9**	**28.8**	37.1	**66.4**	**65.4**	43.0	**53.5**	69.7
35	3	0.2	40	**52.5**	**32.8**	**50.8**	**91.7**	**60.5**	**73.1**	**28.8**	42.1	**67.9**	**65.4**	43.2	**56.1**	**100.0**
36	3	0.2	50	**54.4**	**32.8**	**50.8**	**91.7**	**60.6**	**80.1**	**28.8**	44.2	**68.4**	**66.1**	**56.1**	**56.7**	**100.0**
37	3	0.3	5	27.0	**31.3**	**44.3**	**91.7**	36.2	11.3	17.6	15.6	18.4	26.0	17.5	24.3	16.9
38	3	0.3	10	33.6	**32.8**	**44.5**	**91.7**	38.5	12.4	**24.8**	26.7	29.7	49.7	34.3	36.6	39.3
39	3	0.3	20	40.2	**32.8**	**44.7**	**91.7**	**54.8**	14.0	**26.4**	30.4	**60.0**	**62.8**	40.2	46.4	57.2
40	3	0.3	30	42.9	**32.8**	**49.1**	**91.7**	**54.9**	**68.8**	**26.4**	32.1	**62.9**	**63.4**	40.9	48.4	63.5
41	3	0.3	40	47.9	**32.8**	**50.3**	**91.7**	**54.9**	**74.7**	**26.4**	38.1	**63.3**	**63.4**	41.1	**51.0**	66.5
42	3	0.3	50	49.4	**32.8**	**50.3**	**91.7**	**54.9**	**81.7**	**26.4**	39.1	**63.6**	**64.4**	53.0	**51.3**	67.3
43	3	0.5	5	19.7	**26.6**	36.2	58.3	28.7	8.6	16.5	11.7	15.2	27.4	13.6	16.3	10.1
44	3	0.5	10	25.5	**26.6**	36.2	58.3	30.9	10.2	16.5	22.5	19.5	51.9	27.8	22.6	25.9
45	3	0.5	20	31.7	**26.6**	36.8	58.3	44.8	10.2	16.5	24.8	41.0	**57.1**	34.3	27.8	41.1
46	3	0.5	30	35.5	**26.6**	40.2	58.3	44.9	52.2	16.5	26.5	43.1	**58.1**	35.0	28.8	45.9
47	3	0.5	40	39.8	**26.6**	42.4	58.3	44.9	**70.4**	16.5	30.3	44.7	**58.5**	35.0	29.9	49.5
48	3	0.5	50	42.1	**26.6**	42.4	58.3	44.9	**74.2**	16.5	30.8	46.7	**59.1**	44.6	30.6	50.7
49	5	0.1	5	33.2	**31.3**	**46.0**	**91.7**	45.4	13.4	**24.7**	12.0	21.4	34.6	23.4	23.9	19.5
50	5	0.1	10	39.0	**32.8**	**47.7**	**91.7**	49.2	15.1	**27.2**	36.3	40.0	**59.9**	393	42.1	45.7
51	5	0.1	20	44.4	**32.8**	**48.6**	**91.7**	**65.9**	25.3	**28.7**	**52.4**	**68.4**	**66.7**	45.1	**53.0**	65.3
52	5	0.1	30	**51.7**	**32.8**	**52.8**	**91.7**	**66.0**	**73.1**	**28.7**	**55.9**	**69.8**	**67.3**	46.5	**55.6**	**100.0**
53	5	0.1	40	**58.3**	**32.8**	**53.5**	**91.7**	**66.0**	**76.9**	**28.7**	**62.6**	**71.2**	**67.3**	46.5	**58.3**	**100.0**
54	5	0.1	50	**62.9**	**32.8**	**53.5**	**100.0**	**66.0**	**83.9**	**28.7**	**64.3**	**71.2**	**68.1**	**68.2**	**60.0**	**100.0**
55	5	0.2	5	33.2	**31.3**	**45.3**	**91.7**	43.2	12.9	**24.7**	12.0	19.8	34.4	22.4	23.8	19.4
56	5	0.2	10	38.6	**32.8**	**45.7**	**91.7**	45.1	14.5	**27.2**	35.5	35.0	**59.5**	36.7	41.1	44.5
57	5	0.2	20	43.6	**32.8**	**46.5**	**91.7**	**62.6**	26.3	**28.7**	**51.5**	**63.6**	**66.5**	42.5	**51.6**	63.4
58	5	0.2	30	**51.7**	**32.8**	**50.6**	**91.7**	**62.6**	**69.9**	**28.7**	**54.9**	**66.4**	**66.9**	43.9	**54.0**	69.7
59	5	0.2	40	**57.1**	**32.8**	**50.9**	**91.7**	**62.6**	**73.1**	**28.7**	**61.8**	**67.9**	**66.9**	43.9	**56.7**	**100.0**
60	5	0.2	50	**61.4**	**32.8**	**50.9**	**91.7**	**62.6**	**80.1**	**28.7**	**63.4**	**68.4**	**67.5**	**65.4**	**58.2**	**100.0**
61	5	0.3	5	31.3	**31.3**	**45.3**	**91.7**	37.4	11.3	17.9	14.9	18.4	33.3	19.6	20.4	16.9
62	5	0.3	10	36.7	**32.8**	**45.5**	**91.7**	40.4	12.4	**25.1**	33.3	29.7	**57.3**	34.3	36.5	39.3
63	5	0.3	20	40.5	**32.8**	**46.3**	**91.7**	**57.4**	14.0	**25.9**	47.1	**60.0**	**64.8**	39.5	46.7	57.2
64	5	0.3	30	47.5	**32.8**	**50.4**	**91.7**	**57.4**	**68.8**	**25.9**	50.6	**62.9**	**65.2**	41.1	**48.9**	63.5
65	5	0.3	40	**51.7**	**32.8**	**50.4**	**91.7**	**57.4**	**74.7**	**25.9**	**56.1**	**63.3**	**65.2**	41.4	**51.7**	66.5
66	5	0.3	50	**54.8**	**32.8**	**50.4**	**91.7**	**57.4**	**81.7**	**25.9**	**57.7**	**63.6**	**65.8**	**59.3**	**52.8**	67.3
67	5	0.5	5	23.6	**28.1**	34.8	58.3	29.9	8.6	16.2	12.7	15.2	38.9	14.3	16.8	10.1
68	5	0.5	10	30.1	**28.1**	34.8	58.3	44.7	10.2	16.2	27.6	19.5	**57.1**	27.1	23.1	25.9
69	5	0.5	20	34.4	**28.1**	35.8	58.3	48.1	10.2	16.2	36.8	41.0	**62.2**	33.6	28.1	41.1
70	5	0.5	30	38.2	**28.1**	38.4	58.3	48.5	52.2	16.2	40.0	43.1	**63.0**	34.3	29.1	45.9
71	5	0.5	40	40.5	**28.1**	39.0	58.3	48.5	**70.4**	16.2	44.0	44.7	**63.0**	34.6	30.7	49.5
72	5	0.5	50	42.1	**28.1**	39.0	58.3	48.5	**74.2**	16.2	44.7	46.7	**63.4**	47.2	31.8	50.7

[a] High occupancy rates, which means 80% or greater than the highest occupancy rate for the compound, are indicated with bold letters. Among the high occupancy rates, several occupancy rates that are achieved by the five smallest number of generated conformations are indicated with underlined letters.

DISCUSSION

We described here a novel evaluation method for conformation generators. Previously, conformation generators have been evaluated according to the reproduction of a limited number of X-ray bioactive conformations, usually just one for each compound. However, when a researcher uses a conformation generator, the expected output is a set of conformations that are all valid as X-ray bioactive conforma-

Table 5. Parameters of the Primal and the Optimized Parameter Sets

	primal parameter set[a]	optimized parameter set[a]
-MMFF94s	true	true
-Itermax	10000	10000
-Keepconf	1	5
-Numconf	10	10
-Ewindow	1000	1000
-Maxconf	10000	10000
-Rms	0.3	0.3
-Finalcut	25	50
-Finalrms	0.3	0.3

[a] Parameters assigned default values are not shown in this table.

tions. To meet this expectation, we proposed that the overall validity of the generated conformation ensemble should be the goal of such an algorithm. Our evaluation method compares the generated conformation ensemble with the bioactive conformation space. Therefore, it is necessary to have proper bioactive conformation space. Since our target is druglike compounds for drug discovery research, we generated the bioactive conformation space according to Perola's analysis which had been performed by using crystal structures of protein and druglike small molecule complexes. As an example, we studied the validity of OMEGA-generated conformation ensemble, leading to that in some cases, while OMEGA 1.8.1 reproduced the X-ray bioactive conformations, the occupancy rates of OMEGA-generated conformation ensemble with respect to the bioactive conformation spaces were quite low. For example, the X-ray bioactive conformation of compound **9** could be successfully reproduced using OMEGA 1.8.1. with the primal parameter set (Figure 5A). However, the occupancy rate of the generated conformation ensemble for compound **9** was quite low (43.5%), which gave us concern about the validity of conformation generators based on a limited number of X-ray bioactive conformations. Figure 5B shows the superposition to a conformation of compound **9** in the bioactive conformation space with the closest conformation among the conformation ensemble generated by using OMEGA 1.8.1 with the primal parameter set. The conformation was arbitrarily chosen from the bioactive conformation space and was not the X-ray bioactive conformation but one of the putative bound conformations that satisfied the conditions for the X-ray bioactive conformation outlined by Perola and Charifson.[2] OMEGA 1.8.1 with the primal parameter set could not reproduce the conformation. This is one example in which the conformation generator could reproduce the X-ray bioactive conformation but could not reproduce all the putative bound conformations—leading to a low occupancy rate. The result indicates that an evaluation method based on occupancy rate generated by a conformation generator is more practical than the reproducibility of a single X-ray bioactive conformation.

We optimized the parameter set of OMEGA 1.8.1 as a working example of the proposed evaluation method for conformation generators. We derived the optimized parameter set (parameter set 66 in Table 4) from detailed analysis of the size of generated conformation ensembles and their occupancy rates. Compared to the primal parameter set, the rmsd values of 3 compounds (**14**, **21**, **26**) were drastically reduced to less than 0.5 Å. Thus, using the optimized parameter set we could successfully reproduce the X-ray bioactive conformations of all 26 compounds with 6 or less rotatable bonds. The optimized parameter set could not reproduce the X-ray bioactive conformations of 7 compounds with 7 or more rotatable bonds (**28**, **30**, **32**−**36**). However, rmsd values of some compounds were reduced drastically and all fell below 0.91 Å (Figure 4), which indicates that the optimized parameter set could generate a much closer agreement between the calculated and the X-ray bioactive conformation. We optimized the parameter set based on only the occupancy rate and the size of the generated conformation ensemble, but that reproducibility was not taken into account. Nevertheless, the optimized parameter set provided conformation ensembles with improved reproducibility, which suggests that the occupancy rate against the bioactive conformation space is a more important factor in evaluating conformation generators than the reproducibility of the X-ray bioactive conformation, confirming our original hypothesis.

In the process of optimizing OMEGA 1.8.1 parameter set, we varied three parameter settings, such as *keepconf*, *finalcut*, and *rms* (*finalrms*) settings. While the *keepconf* setting and the *finalcut* setting of the optimized parameter set were modified to 5 from 1 and to 50 from 25, respectively, in the primal parameter set, the *rms* (*finalrms*) setting stayed the same in the primal and the optimized parameter sets. Since the *rms* (*finalrms*) setting determines duplicate conformations, a smaller value of this parameter leads to a larger number of generated conformations in a conformation ensemble. This larger conformation ensemble is expected to achieve a higher occupancy rate; however, it was not observed even when the *rms* (*finalrms*) setting was set to smaller values than 0.3 Å. This means that the *rms* (*finalrms*) setting = 0.3 (Å) was small enough to cover the bioactive conformation space, when the bioactive conformation space was generated with rmsd = 0.3 Å as the criterion of duplicate conformations and the reproduction of putative bound conformations was measured with rmsd = 0.5 Å. On the other hand, a lower occupancy rate was observed when the *rms* (*finalrms*) setting was set to 0.5 (Å). This means that the *rms* (*finalrms*) setting = 0.5 Å was too large to cover the bioactive conformation space. When we focus on 18 parameter sets (13−18, 37−42, and 61−66 in Table 4) all in which the *rms* (*finalrms*) setting was set to 0.3 Å, the optimized parameter set 66 is the one that provided the largest conformation ensemble because both the *keepconf* setting (5) and the *finalcut* setting (50) were set to the largest values among the tested parameter sets. Since the *keepconf* setting determines the number of starting conformations for the torsion-driving step, a larger value of this parameter leads to a larger conformation ensemble. For further investigation of the optimized *keepconf* setting, we checked the relationship between the occupancy rate and the number of generated conformations when the *keepconf* setting was changed from 1 to 10 and observed that larger conformation ensembles actually achieved higher occupancy rates (data not shown). Consequently, the preparation of multiple starting conformations for the torsion-driving step is important to obtain accurate conformation ensembles that show high occupancy rates with respect to the bioactive conformation space. This is likely attributed to the fact that saturated rings usually take multiple energetically stable conformations and that the torsion-driving mechanism rotate only dihedral angles with

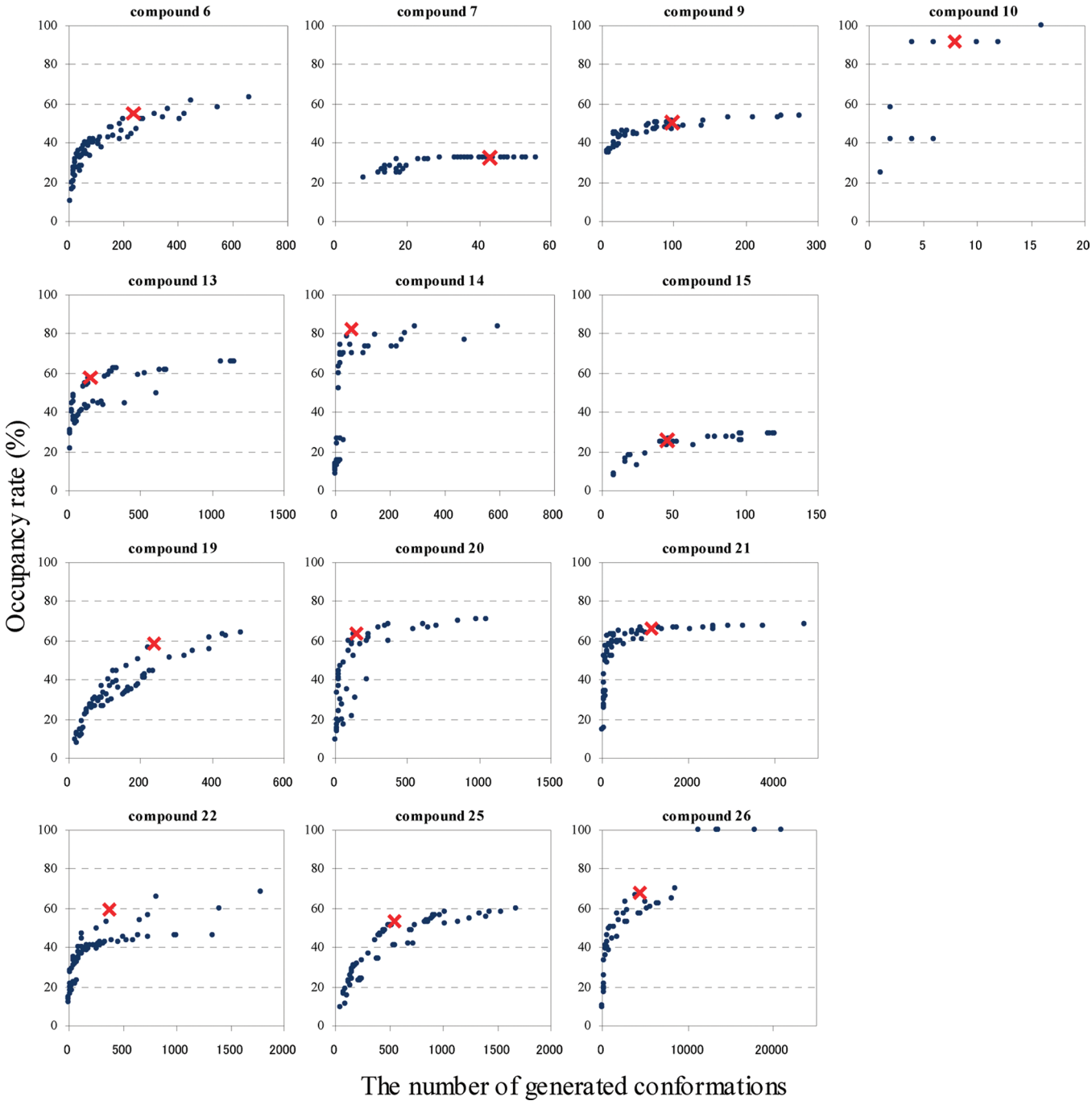

Figure 3. Relationship between the number of generated conformations and the occupancy rate. All the results from 72 parameter sets are plotted as blue dots. The results from the optimized parameter set (parameter set 66) are indicated by red crosses.

keeping both bond lengths and bond angles as they are given from starting conformations. Since the *finalcut* setting determines the upper limit of the internal energy gap of the generated conformations, a larger value of this parameter leads to a larger conformation ensemble. For further investigation of the optimized *finalcut* setting, we checked the relationship between the occupancy rate and the number of generated conformations when the *finalcut* setting was changed from 2 to 6000 (kcal/mol). The relationships of several compounds were shown in Figure 6. To illustrate all cases in the same graph, the number of generated conformations was standardized as the ratio to the number of obtained conformations when the *finalcut* setting was set to 50 in each case. Figure 6 shows the occupancy rates almost or completely plateaued far below 100% when the *finalcut* setting was set to larger values than 50 (i.e., in the area of X > 1.0). This result means that the *finalcut* setting = 50 (kcal/mol) was large enough to cover the bioactive conformation space, when the bioactive conformation space was generated with 3.3 or 6.9 kcal/mol as the criterion of the internal energy gap of the putative bound conformations according to the number of rotatable bonds. The result also means that a larger number of generated conformations are not necessary to give better reproducibility of the overall bioactive conformation space. On the other hand, Figure 6 also shows a smaller number of generated conformations that lead to lower occupancy rates when the *finalcut* setting was set to smaller values than 50 (i.e., in the area of X < 1.0). As a consequence, the *finalcut* setting = 50 (kcal/mol) is optimum in terms of the coverage of bioactive conformation space

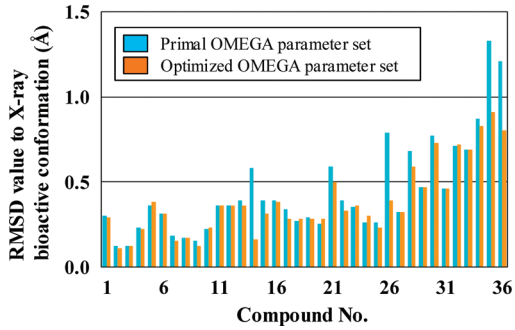

Figure 4. Rmsd values of the best-fitting conformation from each generated conformation ensemble with respect to the experimentally determined bioactive conformation for all 36 compounds. Rmsd values were calculated using OMEGA 1.8.1 with either the primal parameter set (cyan bar) or the optimized parameter set (orange bar).

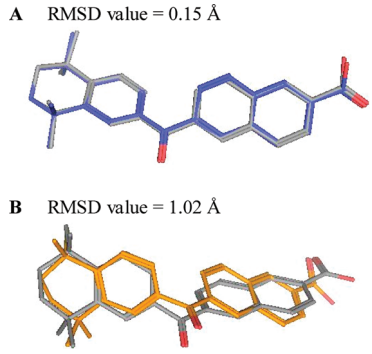

Figure 5. Superposition to compound **9** of X-ray bioactive conformation colored by blue (**A**) and a putative bound conformation colored by orange (**B**) in the bioactive conformation space with the best-fitting conformation (gray) from the conformation ensemble generated using OMEGA 1.8.1 with the primal parameter set.

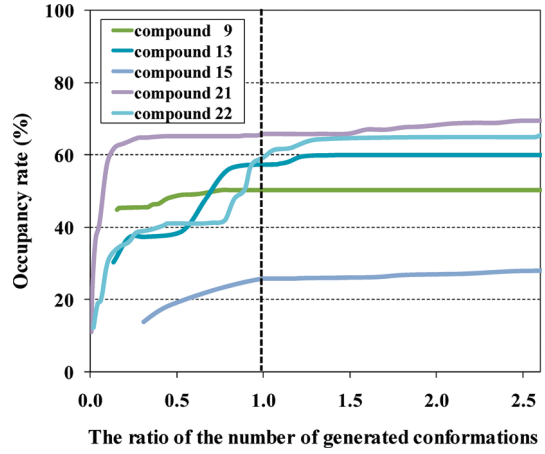

Figure 6. Relationship between the occupancy rate and the number of generated conformations with various values of the *finalcut* setting. The number of generated conformation is presented as the ratio to the number of obtained conformations when the *finalcut* setting is set to 50. The dashed line ($X = 1$) indicates the *finalcut* setting is set to 50.

and computational efficiency, intriguingly across the compounds. It is surprising that the enormous difference between 3.3 or 6.9 kcal/mol as the internal energy criterion of the bioactive conformation space and 50 kcal/mol of the generated conformation ensemble by using OMEGA 1.8.1. The difference is likely attributed to the torsion-driving algorithm. The algorithm can rapidly generate a conformation ensemble from a single starting conformation by rotating its dihedral angles systematically according to the predefined torsion-driving library. For example, dihedral angle increments are set to 120° for a bond between sp^3 carbons and to 30° for a bond between sp^2 carbon and sp^3 carbon. These values were defined from careful studies of the balance between scientific rationale and calculation time. However, the increments can be coarse in generating many conformations from a single starting conformation only by changing its dihedral angles. When the algorithm rotates a dihedral angle of a starting conformation, such as 120°, some intramolecular atomic collisions can occur, although the collision can be avoided with a little less or more rotation of the dihedral angle. Therefore, the *finalcut* setting of the optimized parameter set could be 50 kcal/mol, which was much higher than the required condition for the X-ray bioactive conformation analyzed by Perola and Charifson.[2] It would be quite difficult to find the optimum value without our evaluation method which allowed us to discuss the calculation condition in more detail than the reproduction of the X-ray bioactive conformations.

We expect that our novel evaluation method for conformation generators will allow for making better use of existing conformation generators and may lead to the development of further improved conformation generators.

ACKNOWLEDGMENT

The authors gratefully acknowledge Drs. Kaneyoshi Kato and Hiroyuki Kimura for their suggestions and encouragement and Drs. Toshimasa Tanaka and Matthew Lardy for helpful discussions. The authors thank Drs. David Cork and Douglas Cary for critical reading of the manuscript.

REFERENCES AND NOTES

(1) Schwab, C. H. Conformational analysis and searching. In *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, Germany, 2003; Vol. 1, pp 262−301.
(2) 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.
(3) Sprague, P. W. Automated chemical hypothesis generation and database searching with Catalyst. *Perspect. Drug Discovery Des.* **1995**, *3*, 1–20.
(4) Sprague, P. W.; Hoffman, R. Catalyst pharmacophore models and their utility as queries for searching 3D databases. In *Computer-Assisted Lead Finding and Optimization*: van de Waterbeemd, H., Testa, B., Folker, G., Ed.; Verlag Helvetica Chimica Acta: Basel, 1997; pp 225−240.
(5) *CONFORT, version 7.2*; Tripos Inc.: St. Louis, MO, U.S.A., 2005.
(6) *CORINA, version 3.4*; Molecular Networks: Erlangen, Germany, 2006.
(7) McMartin, C.; Bohacek, R. S. Flexible matching of test ligands to a 3D pharmacophore using a molecular superposition force field: comparison of predicted and experimental conformations of inhibitors of three enzymes. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 237–250.
(8) McMartin, C.; Bohacek, R. S. QXP: powerful rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.
(9) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Macromodel-an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.
(10) *OMEGA, version 1.8.1*; OpenEye Science Software: Santa Fe, NM, U.S.A., 2004.
(11) 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.
(12) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449–462.

(13) Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T. Comparative Analysis of Protain-Bound Ligand Conformations with Respect to Catalyst's Conformational Space Subsampling Algorithms. *J. Chem. Inf. Model.* **2005**, *45*, 422–430.
(14) 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.
(15) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. Fast and efficient in silico 3D screening: toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J. Comput.-Aided Mol. Des.* **2007**, *47*, 2182–2196.
(16) 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.
(17) Borodina, Y. V.; Bolton, E.; Fontaine, F.; Bryant, S. H. Assessment of conformational ensemble sizes necessary for specific resolutions of coverage of conformational space. *J. Chem. Inf. Model.* **2007**, *47*, 1428–1437.
(18) 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.
(19) Ferguson, D. M.; Raber, D. J. A new approach to probing conformational space with molecular mechanics: random incremental pulse search. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378.
(20) *MOE, version 2007.0902*; Chemical Computing Group: Montreal, Quebec, Canada, 2007.
(21) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
(22) Halgren, T. A. Merck molecular force field. II. MMFF van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.
(23) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.
(24) (b) Sadowski, J.; Gasteiger, G.; Klebe, G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
(25) 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.
(26) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
(27) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730–748.
(28) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. Dreiding: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(29) Chen, I.-J.; Foloppe, N. Conformational sampling of druglike molecules with MOE and Catalyst: implications for pharmacophore modeling and screening. *J. Chem. Inf. Model.* **2008**, *48*, 1773–1791.

CI800393W