Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2013 J. Phys.: Conf. Ser. 454 012028
(http://iopscience.iop.org/1742-6596/454/1/012028)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 54.191.40.80
This content was downloaded on 01/09/2017 at 13:08

Please note that terms and conditions apply.

You may also be interested in:

Structural relaxation and binding energy calculations of FK506 binding protein complexes using the large-scale DFT code CONQUEST
T Otsuka, N Okimoto, M Taiji et al.

High performance computing for drug development on K computer
Hideaki Fujitani, Keiko Shinoda, Takefumi Yamashita et al.

Quantum chemical study of conformational preferences of intermediates and transition states in the alkaline hydrolysis of dimethyl phosphate
Y Takano, M Kita and H Nakamura

Conformational analysis investigation into the influence of nano-porosity of ultra-permeable ultra-selective polyimides on its diffusivity as potential membranes for use in the "green" separation of natural gases
Tarek M Madkour

An efficient computational method for the implementation of a semi-classical instanton approach using discretized path integrals
T Kawatsu and S Miura

Understanding the structure and dynamic of odorants in the gas phase using a combination of microwave spectroscopy and quantum chemical calculations
Halima Mouhib

Molecular modelling and free-energy calculations of protein—ligand binding
Viktor B. Luzhkov
Evaluations of the conformational search accuracy of CAMDAS using experimental three-dimensional structures of protein-ligand complexes

A Oda¹,²,³, N Yamaotsu¹, S Hirono⁴, Y Takano³, S Fukuyoshi¹, R Nakagaki¹ and O Takahashi²
¹ Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
² Faculty of Pharmaceutical Sciences, Tohoku Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai, Miyagi 981-8558, Japan
³ Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
⁴ School of Pharmacy, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
E-mail: oda@p.kanazawa-u.ac.jp

Abstract. CAMDAS is a conformational search program, through which high temperature molecular dynamics (MD) calculations are carried out. In this study, the conformational search ability of CAMDAS was evaluated using structurally known 281 protein-ligand complexes as a test set. For the test, the influences of initial settings and initial conformations on search results were validated. By using the CAMDAS program, reasonable conformations whose root mean square deviations (RMSDs) in comparison with crystal structures were less than 2.0 Å could be obtained from 96% of the test set even though the worst initial settings were used. The success rate was comparable to those of OMEGA, and the errors of CAMDAS were less than those of OMEGA. Based on the results obtained using CAMDAS, the worst RMSD was around 2.5 Å, although the worst value obtained was around 4.0 Å using OMEGA. The results indicated that CAMDAS is a robust and versatile conformational search method and that it can be used for a wide variety of small molecules. In addition, the accuracy of a conformational search in relation to this study was improved by longer MD calculations and multiple MD simulations.

1. Introduction

In computer-aided drug design (CADD) trials, a technique to construct accurate three-dimensional (3D) molecular structures is important [1]. A wide variety of compounds has to be considered for drug design and development, but their experimental 3D structures, such as X-ray crystallographic structures and NMR structures, are not always obtained. Thus, computational approaches to determine the 3D structures of small molecules are widely used. Although structural optimizations are carried out to procure the most stable structures of isolated molecules, the optimized structures of isolated small molecules are occasionally inappropriate for drug design. In drug design and development trials, the molecular structures of small ligands in biopolymer-ligand complexes play important roles instead of isolated ligands. Because of this, the interactions between drug target biopolymers and drug candidates (small molecule) attracted attention. The structures of small molecules in the complexes are frequently
different from the most stable state structures of isolated molecules [2]. Thus, many types of conformers should be generated for small ligand molecules, and the appropriate conformer is then selected from the generated conformer set. The appropriate conformer which can be used as the ligand structure in biopolymer-ligand complex has to be included in the generated conformer set. Sequentially, the appropriate conformer is selected from the conformer set by using computational methods, such as structure-based drug design (SBDD) and/or ligand-based drug design (LBDD) techniques. The selected conformer is eventually used for the next steps of drug design and development trials. Because the conformational differences cause the differences in physical and chemical properties of molecules [3], the computational method for generating a conformer set should be carefully considered. Several software programs have been developed for conformational searches, such as CAMDAS (Conformational Analyzer with Molecular Dynamics And Sampling) [4] which is the molecular-dynamics-based method and OMEGA [5] which is the knowledge-based method. The conformers obtained are mostly used for several CADD calculations, i.e. computational docking and virtual screening [6] in SBDD or molecular superposition [7] and 3D-quantitative structure-activity relationship (QSAR) study [8] in LBDD. This attests that conformational searching is one of the key steps in CADD.

As mentioned above, the constructions of a wide variety of conformers are required for conformational searches, and several computational methods have been developed for this purpose. Computational techniques such as systematic search, random search, and distance geometry methods have been applied to conformational searches [1]. Moreover, methods using molecular dynamics (MD) calculations are also frequently used. Although some conformational search methods have to include anomalous procedures for particular molecules such as macrocycle compounds, a conformational search method using MD can be carried out without any pre-processing. To generate a wide variety of conformers by using MD calculations, simulations are generally carried out in an extremely high temperature because the energy barriers among conformers need to be climbed over. The effectiveness of high temperature MD in searching conformational space has already been evaluated. For example, replica exchange MD (REMD) simulations which include the high temperature MD of replica and simulated annealing using MD have been applied to several molecular systems [9-10]. Through these procedures, efficient sampling can be performed. In REMD, a temperature above 500 K is frequently used. In fact, quadruple-digit temperatures are sometimes adopted. Because MD programs only solve Newton’s equation of motion, particular treatments such as “conformation dictionary” are unnecessary for any compounds. In addition, an unusual conformational change, e.g. torsional rotations of unsaturated bonds and ring flips, can occur because of high temperature. Thus, high performance conformational searches can be carried out by using high temperature MD calculations.

Previously, we have developed a computational program CAMDAS in which high temperature MD calculations are carried out for the conformational search of molecules [4]. In this software, clustering of conformers generated by high temperature MD is performed, where RMSDs are used as distances, and the representative conformers are outputs. The current version of CAMDAS 2.1 can use several reliable molecular force fields, i.e. MM2 [11], MMFF [12], and AMBER force field [13], and several file formats (PDB, mol2, and AMBER parameter file formats) can be treated. In addition to user-friendliness, improvements for its computational accuracy are implemented. The multi-copy procedure, in which multiple initial conformers are generated from input coordinates by using distance geometry method and multiple conformational search trials are simultaneously carried out, has been implemented. Previously, we have reported the conformational searching ability of CAMDAS by using N-acetylatedamine-N'-methylamide and cyclooctadecane as a test set [4]. Recently, CAMDAS has been used in several studies related to drug design and development. For example, in some studies, CAMDAS was used together with SBDD techniques for the investigations of ligand recognition mechanisms of cytochrome P450 2C9 and 2C19 [14] and for molecular modelling of a human acidic mammalian chitinase in complex with argifin which is a cyclopentapeptide chitinase inhibitor with a flexible ring structure [15]. By using CAMDAS not only with SBDD but also with LBDD procedures,
useful results were obtained for the design and synthesis of novel delta opioid receptor agonists [16], for the estimation of the 3D pharmacophore of ligands for rat multidrug-resistance-associated protein 2 [17], and for the identification of the 3D pharmacophore of k-opioid receptor agonists [18]. CAMDAS is also useful for a wide variety of molecular modelling trials other than drug design, such as the determination of the stereostructure of luminamicin [19]. On the other hand, exhaustive tests using a large test set for evaluating the search ability of CAMDAS were not carried out, and the influences of computational settings on search results were not clearly investigated.

In this study, the ability of CAMDAS for conformational search was evaluated by using structurally known protein-ligand complexes. The results indicated the usefulness of high-temperature MD for the conformational search and revealed the desirable settings for CAMDAS.

2. Methods

2.1. Test set

The test set consisted of 260 systems which were extracted from 281 protein-ligand complexes in a Glide test set [20]. From the Glide test set, only the systems with rotatable bonds of ligand molecules more than or equal to 1 were used as tests. The numbers of rotatable bonds were counted by using OMEGA [5]. The test systems are listed in table 1.

<table>
<thead>
<tr>
<th>121p</th>
<th>1aaq</th>
<th>1acm</th>
<th>iaco</th>
<th>1add</th>
<th>1adf</th>
<th>1ake</th>
<th>lapl</th>
<th>lapu</th>
<th>lapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1apw</td>
<td>1atl</td>
<td>1avd</td>
<td>1azm</td>
<td>1b6j</td>
<td>1fb6k</td>
<td>1b6i</td>
<td>1b6m</td>
<td>1bag</td>
<td>1bbp</td>
</tr>
<tr>
<td>1bkm</td>
<td>1bma</td>
<td>1bra</td>
<td>1byb</td>
<td>1clb</td>
<td>1c3i</td>
<td>1c5p</td>
<td>1c83</td>
<td>1c84</td>
<td>1c86</td>
</tr>
<tr>
<td>1c87</td>
<td>1c8k</td>
<td>1ebh</td>
<td>1ehx</td>
<td>1edc</td>
<td>1edg</td>
<td>1cil</td>
<td>1enx</td>
<td>1com</td>
<td></td>
</tr>
<tr>
<td>1cps</td>
<td>1ctr</td>
<td>1ctt</td>
<td>1d3d</td>
<td>1d3p</td>
<td>1dd7x</td>
<td>1dd8f</td>
<td>1ddb</td>
<td>1dbm</td>
<td>1dd6</td>
</tr>
<tr>
<td>1dd5</td>
<td>1ddf</td>
<td>1did</td>
<td>1die</td>
<td>1dih</td>
<td>1dm2</td>
<td>1dog</td>
<td>1drl</td>
<td>1dwb</td>
<td>1dwc</td>
</tr>
<tr>
<td>1dwd</td>
<td>1e5i</td>
<td>1eap</td>
<td>1ebg</td>
<td>1ecv</td>
<td>1eed</td>
<td>1ejn</td>
<td>1ela</td>
<td>1elb</td>
<td>1ele</td>
</tr>
<tr>
<td>1eld</td>
<td>1ele</td>
<td>1epb</td>
<td>1eta</td>
<td>1etr</td>
<td>1ets</td>
<td>1ett</td>
<td>1etzq</td>
<td>1f0r</td>
<td>1f0s</td>
</tr>
<tr>
<td>1f0t</td>
<td>1f0u</td>
<td>1fen</td>
<td>1f8h</td>
<td>1fh9</td>
<td>1fhd</td>
<td>1fis</td>
<td>1fkg</td>
<td>1g5</td>
<td>1frp</td>
</tr>
<tr>
<td>1fvt</td>
<td>1g45</td>
<td>1g46</td>
<td>1g48</td>
<td>1g4j</td>
<td>1g40</td>
<td>1g52</td>
<td>1g53</td>
<td>1g54</td>
<td>1g5b</td>
</tr>
<tr>
<td>1gLp</td>
<td>1glq</td>
<td>1gsn</td>
<td>1hbc</td>
<td>1hcf</td>
<td>1hfc</td>
<td>1hgg</td>
<td>1hhg</td>
<td>1hgi</td>
<td></td>
</tr>
<tr>
<td>1hj</td>
<td>1hjj</td>
<td>1hjs</td>
<td>1hkp</td>
<td>1hpr</td>
<td>1hri</td>
<td>1hsq</td>
<td>1hst</td>
<td>1hte</td>
<td>1htf</td>
</tr>
<tr>
<td>1hti</td>
<td>1hvr</td>
<td>1htv</td>
<td>1icn</td>
<td>1ida</td>
<td>1igj</td>
<td>1imbv</td>
<td>1ive</td>
<td>1ivd</td>
<td>1ivd</td>
</tr>
<tr>
<td>1ive</td>
<td>1ivf</td>
<td>1lh</td>
<td>1lep</td>
<td>1ldm</td>
<td>1lic</td>
<td>1lmo</td>
<td>1lna</td>
<td>1lst</td>
<td>1lmc</td>
</tr>
<tr>
<td>1mdr</td>
<td>1mfe</td>
<td>1ml</td>
<td>1mpq</td>
<td>1mnc</td>
<td>1mrk</td>
<td>1mpu</td>
<td>1mvp</td>
<td>1mns</td>
<td>1nnb</td>
</tr>
<tr>
<td>1nsd</td>
<td>1ndw</td>
<td>10kl</td>
<td>1phd</td>
<td>1ppg</td>
<td>1pha</td>
<td>1phd</td>
<td>1phf</td>
<td>1phg</td>
<td></td>
</tr>
<tr>
<td>1pec</td>
<td>1ppc</td>
<td>1pph</td>
<td>1ppk</td>
<td>1ppp</td>
<td>1pro</td>
<td>1psd</td>
<td>1rbb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ps</td>
<td>1rhl</td>
<td>1rls</td>
<td>1rne</td>
<td>1rnt</td>
<td>1rob</td>
<td>1sbg</td>
<td>1sht</td>
<td>1snc</td>
<td>1sre</td>
</tr>
<tr>
<td>1srj</td>
<td>1stj</td>
<td>1tdb</td>
<td>1thy</td>
<td>1tka</td>
<td>1tp</td>
<td>1tmn</td>
<td>1tng</td>
<td>1tnh</td>
<td>1tni</td>
</tr>
<tr>
<td>1tnj</td>
<td>1tnk</td>
<td>1ml</td>
<td>1mph</td>
<td>1trk</td>
<td>1tyl</td>
<td>1ukx</td>
<td>1upw</td>
<td>1xid</td>
<td>1xie</td>
</tr>
<tr>
<td>2ack</td>
<td>2ada</td>
<td>2ak3</td>
<td>2cgr</td>
<td>2cht</td>
<td>2cmd</td>
<td>2cte</td>
<td>2dbi</td>
<td>2gbp</td>
<td>2rbb</td>
</tr>
<tr>
<td>2ctx</td>
<td>2mcp</td>
<td>2phh</td>
<td>2pk4</td>
<td>2ply</td>
<td>2r04</td>
<td>2r07</td>
<td>2smb</td>
<td>2sun</td>
<td>2tpi</td>
</tr>
<tr>
<td>2uji</td>
<td>2xis</td>
<td>2yxs</td>
<td>2ypj</td>
<td>3ca</td>
<td>3cpa</td>
<td>3dfk</td>
<td>3hvt</td>
<td>3mth</td>
<td>3ptb</td>
</tr>
<tr>
<td>3pi</td>
<td>4aah</td>
<td>4cts</td>
<td>4dfi</td>
<td>4fab</td>
<td>4fbp</td>
<td>4fnx</td>
<td>4hmg</td>
<td>4hpe</td>
<td>4ttn</td>
</tr>
<tr>
<td>4tin</td>
<td>4tmn</td>
<td>4tpi</td>
<td>4ts1</td>
<td>5abp</td>
<td>5cts</td>
<td>5p2p</td>
<td>5tn</td>
<td>5tmn</td>
<td>6cpa</td>
</tr>
<tr>
<td>6rti</td>
<td>6tim</td>
<td>6tmn</td>
<td>7cna</td>
<td>7tmm</td>
<td>8abp</td>
<td>8atc</td>
<td>8gch</td>
<td>9abp</td>
<td>9hvp</td>
</tr>
</tbody>
</table>

For the CAMDAS calculations, only the ligand structures were extracted from the complex structures registered in the Protein Data Bank (PDB) [21]. Hydrogen atoms were added for the ligand structures by using an SYBYL6.9 [22] software. The complemented ligand structures were optimized
through AM1 method using MOPAC2007 [23]. The addition of hydrogen is a generally required process for a conformational search. In fact, it is indispensable to using both CAMDAS and OMEGA. The procedure of filling the valence can be performed not only by means of SYBYL but also by means of other programs such as Maestro and ChemBio 3D. Meanwhile, the influence of the procedures is removed through structural optimizations with the use AM1 method. The optimized PDB structures are defined as “test set i.”

Because the conformers included in set i were constructed from crystal structures and only structural optimizations were performed, the conformations of set i were similar to the structures of crystal. The ability of conformational search program may be overestimated by using set i as initial structures, because the conformational search tests using initial conformers which are similar to experimental structures may result in undue success. Thus, the preliminary conformational search was carried out before the main tests in order to remove the influences of crystal structures. The generated conformations through the preliminary search were used as initial structures of the main tests. The flow chart of the tests is shown in figure 1. The preliminary and main conformational searches were performed using CAMDAS. The generated conformers using preliminary searches were optimized through AM1 method, and the lowest energy conformers (set ii), the conformers with the smallest RMSDs in comparison with experimental structures (set iii), and the conformers with the largest RMSDs (set iv) in comparison with crystal structures were selected. That is to say that in relation to experimental structures, conformers iii were the nearest and conformers iv were the farthest. Conformer sets ii, iii, and iv were used as initial structures of the main tests as well as set i.

![Flowchart of conformational search tests.](image)

### 2.2. Evaluations for search results

The results of the main tests were compared with the experimental structures, and the accuracy of search results were evaluated using the RMSDs between the generated conformers and the experimental structures. In drug design and development studies, protein-ligand complex structures play important roles. Therefore, conformational search programs are required to reproduce ligand conformers in protein-ligand complexes. In this study, “reasonable” conformers were defined as the conformers whose RMSDs from the experimental structure were smaller than or equal to 2.0 Å. When
at least one reasonable conformer is obtained, the conformational search trial for the test system is defined as successful. The numbers of successful systems were compared among search trials using several initial structures (sets i, ii, iii, and iv) and several settings. Thus, the dependencies of the results of conformational searches on initial structures and settings were investigated. In addition to CAMDAS tests, conformational searches using OMEGA [5], in which the molecules are divided into fragments and conformers are fast constructed by assembling the fragments, were also performed for comparison.

2.3. Computational setting
For both the preliminary and the main conformational searches using CAMDAS, the same settings described below were used. The general AMBER force field (GAFF) [24] was adopted. The assignment of atom types and the atomic charge calculations were carried out by using the antechamber module [25] of AMBER9 [26]. Although the molecular files for CAMDAS calculations can be created by using other software programs such as SYBYL, the antechamber module, which is currently included in the free software of AmberTools, is easier to obtain. For CAMDAS calculations, the dielectric constant was set to 80, and the conformational searches using 100000 steps of MD calculations under 1200 K were performed. 1200 K is sometimes used as the temperature for REMD and annealing [10]. The conformational sampling was conducted every 100 steps during MD trajectory. Meanwhile, the maximum number of sampled conformers was set to 10000. After the conformational sampling, structural optimizations with maximum 1000000 cycles were carried out for the sampled conformers. The clustering of optimized conformers was carried out after conformer generation steps. In the clustering, two conformers were placed in the same cluster when the RMSD between them is less than 1.5 Å. After the clustering, the representative conformers were extracted from the obtained clusters. This procedure (1000000-steps MD and 1.5 Å threshold) was defined as procedure I. The OMEGA calculations for the comparative study were performed using default settings. Because the conformational search results were unaffected by the initial conformers in the default settings of OMEGA calculations, only conformer set ii was used as initial conformers. MOPAC2007 was used for AM1 calculations, and an implicit water model by using COSMO was adopted.

2.4. Additional settings
In addition to procedure I, different settings of CAMDAS calculations were tested. In procedure I, 1,000,000-steps MD was carried out, and clustering threshold was set to RMSD < 1.5 Å. On the other hand, the procedure with 1,000,000-steps MD and a threshold of 0.5 Å (procedure II), the procedure with 5,000,000-steps MD and a threshold of 1.0 Å (procedure III), the procedure with 5,000,000-steps MD and a threshold of 0.5 Å (procedure IV), and the procedure with multicopy MD calculations using the same settings as procedure I (procedure V) were carried out. In the multicopy MD of procedure V, ten types of initial conformations were generated and ten MD calculations were simultaneously carried out. For procedures II to V, only conformer set ii was used. The versions of the conformational search programs were CAMDAS 2.1 and OMEGA 2.2.1.

3. Results and discussion
3.1. Dependencies on initial conformations
In table 2, the results of conformational searches using four types of initial conformers with procedure I are shown. The results of OMEGA calculations are also shown for comparison. The success rate $p$ was calculated as follows:

$$p = \frac{n_{\text{success}}}{n_{\text{all}}}$$

where, $n_{\text{success}}$ is the number of systems for which reasonable conformers were obtained by the conformational search, and $n_{\text{all}}$ is the total number of test systems. “Reasonable conformer” is defined as the conformer whose RMSD against an experimental structure is less than or equal to 2.0 Å. When at least one reasonable conformer is obtained for one test system, the test is defined as “success”. In
the test, the \( n_{\text{all}} \) is equal to 260 as shown in table 1. In addition to success rate, the PDB ID of the test system, where the RMSD of the best conformer is the largest of all 260 test systems, and the best RMSD of the system (the largest, best RMSD) are described in table 2. For example, in the conformational search for 1ake using initial conformer i, the RMSD of the best conformer obtained from the search was 2.310 Å. RMSDs of other conformers were worse than 2.310 Å; nevertheless, “2.310 Å” was the best RMSD for 1ake. For all test systems, the best RMSD was less than or equal to 2.310 Å. In this case, we can say that 1ake was “the system where the RMSD of the best conformer is the largest of all 260 test systems” and that 2.310 Å was the “largest, best RMSD.” In the table, the averages of generated conformers are also shown.

As shown in table 2, CAMDAS can generate reasonable conformers for almost all the test systems regardless of initial conformers. The success rates for initial conformers i and iii were better than those for conformers ii and iv. Conformer i was generated from the crystal structure using simple optimizations only, and conformer iii had the most similar structure to the experimental structure. However, the difference of success rates between them was only around 3% despite the differences of initial conformers. The results indicated that the conformational searches by CAMDAS were robust for the difference of initial structures. Although the structural optimization using MOPAC were carried out to generate initial conformers, the robustness suggested that such pre-processing not always have to be carried out for CAMDAS. In addition, the success rates of CAMDAS were comparable to that of OMEGA, which is one of the most widely used and the most reliable conformational search programs [27-28]. The results suggested that CAMDAS is useful for the conformational searches of small molecules.

As shown in table 2, CAMDAS can generate reasonable conformers for almost all the test systems regardless of initial conformers. The success rates for initial conformers i and iii were better than those for conformers ii and iv. Conformer i was generated from the crystal structure using simple optimizations only, and conformer iii had the most similar structure to the experimental structure. However, the difference of success rates between them was only around 3% despite the differences of initial conformers. The results indicated that the conformational searches by CAMDAS were robust for the difference of initial structures. Although the structural optimization using MOPAC were carried out to generate initial conformers, the robustness suggested that such pre-processing not always have to be carried out for CAMDAS. In addition, the success rates of CAMDAS were comparable to that of OMEGA, which is one of the most widely used and the most reliable conformational search programs [27-28]. The results suggested that CAMDAS is useful for the conformational searches of small molecules.

<table>
<thead>
<tr>
<th>conformer</th>
<th>i</th>
<th>ii</th>
<th>iii</th>
<th>iv</th>
<th>OMEGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_{\text{success}} )</td>
<td>256</td>
<td>248</td>
<td>255</td>
<td>249</td>
<td>252</td>
</tr>
<tr>
<td>success rate</td>
<td>0.985</td>
<td>0.953</td>
<td>0.981</td>
<td>0.958</td>
<td>0.969</td>
</tr>
<tr>
<td>largest best RMSD/Å</td>
<td>2.310</td>
<td>2.474</td>
<td>2.477</td>
<td>2.403</td>
<td>3.862</td>
</tr>
<tr>
<td>(PDB ID)</td>
<td>(1ake)</td>
<td>(1adf)</td>
<td>(5cts)</td>
<td>(5cts)</td>
<td>(1ake)</td>
</tr>
<tr>
<td>num. of conformers</td>
<td>69.67</td>
<td>68.65</td>
<td>68.47</td>
<td>67.74</td>
<td>127.55</td>
</tr>
</tbody>
</table>

Although the success rate of CAMDAS calculations were comparable with that of OMEGA as mentioned above, the results of CAMDAS were better than those of OMEGA in terms of the calculations for the largest best RMSD. The system where the RMSD of the best conformation is the largest of all CAMDAS test was 5cts using conformer iii. The best RMSD of the test of 5cts with initial conformer iii using CAMDAS was only 2.477 Å. On the other hand, the system where the RMSD of the best conformation is the largest in OMEGA test was 1ake, and the best RMSD of 1ake using OMEGA was 3.862 Å. Thus, CAMDAS can generate conformations with an RMSD < 2.5 Å even for the systems where reasonable conformations cannot be generated. Because the numbers of rotatable bonds are 23 and 16 for the ligands of 5cts and 1ake, respectively, their flexibilities are relatively high (in general, the number of rotatable bonds for drug-like molecules are less than or equal to ten [29]). The results indicated that CAMDAS can be effectively used for flexible molecules. The best conformations are shown in figures 2 and 3. Figure 2 is the best conformation generated by CAMDAS for 5cts using initial conformer iii, while figure 3 is the best conformation calculated by OMEGA for 1ake. For comparison, both conformations were superposed onto crystal structures. As seen in figures 2 and 3, the structures colored by elements are generated conformation, and the structures in light blue are crystal structures. The hydrogen atoms were removed from this figure. In figure 2, although the conformation of the flexible chain moiety was different between the generated conformation and the crystal structure, many functional groups such as carboxyl, phosphate, and adenine groups were well-overlapped. Because the interactions between functional groups of the drug target and the ligand molecule are important for drug design trials, the generated conformations that
have functional groups with locations similar to crystal structures seem to be useful for practical use despite an RMSD > 2.0 Å. The conformation shown in figure 2 was the worst case of the “best conformations” in this study, because the RMSD of the best solution for 5cts using conformer iii was 2.477 Å. Moreover, it was the largest RMSD of all “best conformations.” Therefore, CAMDAS can generate conformations where, even in the worst case, the functional groups were located in appropriate positions. On the other hand, for the result of 1ake obtained by using OMEGA which is shown in figure 3, the distance between two adenine moieties in the computational result was shorter than that in the crystal structure. Furthermore, both adenine moieties did not overlap the crystal structure. Because the locations of adenine moieties and phosphate groups were different from the crystal structure, the calculated locations of functional groups which are potentially important for protein-ligand interactions seemed to be inappropriate when using OMEGA. As shown in table 2 and figures 2 and 3, although the success rates were comparable between CAMDAS and OMEGA, the effectiveness for failed systems where reasonable conformations cannot be generated was different. CAMDAS can generate the reasonable positions of conformers in which the functional groups are located. Additionally, CAMDAS seems to be a more robust and fail-safe method for the conformational searches of wide varieties of small molecules. In general, ligand conformations in the protein-ligand complexes are not always the same as the conformations of isolated ligand molecules which are located in the minima of potential surface of isolated ligand [2]. Thus, conformational search methods such as OMEGA sometimes fail to search the ligand conformations in protein-ligand complexes. Even for such cases, CAMDAS can generate conformations close to the ligand structures in the protein-ligand complexes. A knowledge-based approach is adopted by OMEGA. In contrast, CAMDAS uses high temperature MD simulations. Therefore, CAMDAS can search the conformers without any previous knowledge, and it may be the reason behind the robustness of CAMDAS. Although the inaccuracy of force field parameters is considered a cause of the difference between the calculated conformation and the ligand structure in protein-ligand complexes [2], the results indicated that GAFF is appropriate for conformational searches of small ligands.

**Figure 2.** The best conformation with largest RMSD obtained by using CAMDAS. The molecules colored light blue are crystal structures, and the molecules colored by elements are generated conformers.

**Figure 3.** The best conformation with largest RMSD obtained by using OMEGA. The molecules colored light blue are crystal structures, and the molecules colored by elements are generated conformers.
The numbers of generated conformers are also shown in table 2. The number of generated
conformers by using OMEGA is around twice as many as that by CAMDAS. For molecular
superposition studies such as pharmacophore search and 3D-QSAR studies, all conformers have to be
tested, and it is O(n^2) problem. That is, the increase of the number of conformers to be tested is
directly linked to the increase of computational costs. Thus, less number of generated conformations
is better, when the similar success rates can be achieved. Because of this point of view, CAMDAS
seems to be more useful than OMEGA. In OMEGA, the fragment decomposition and the
reconstruction of small molecules are carried out for conformational searches, in contrast to CAMDAS
in which high temperature MD calculations of complete small molecules are performed. Because
CAMDAS generates the conformations by using MD trajectory in accordance with the physical law
without artificial decomposition of molecules, only natural structures are generated. The generations
of natural conformations may be the reason of efficiency of CAMDAS by which high success rates
can be obtained despite less number of generated conformations than by OMEGA. OMEGA gives an
output with a large number of conformers even for small and rigid molecules; however, structural
optimizations and clustering might reduce the number of generated conformations similar to
CAMDAS.

3.2. Results of conformational searches using additional settings
In table 3, the results by using different settings for conformational searches are shown. The detailed
searches, which are frequently required for the ligand-based drug design trials of large flexible
molecules, can be carried out using these settings. For the test calculations of this study, conformer i
was used as a set of initial conformations. As shown in the table, these settings can be used to perform
very accurate conformational searches, even though conformer ii which was constructed without
reference to crystal structures was used as initials. For example, the number of failed systems were
only two using method IV (1hbv and 5cts), and the best RMSDs of the failed systems were 2.125 Å
(1hbv) and 2.013 Å (5cts). Using method V, conformational searches for eight systems (1eed, 1ejn,
1fq5, 1ida, 1ppi, 1psq, 1rme, 9hvp) could not be carried out and no conformations were generated,
because multicopy method could not generate 10 different initial conformers. For other test systems,
the number of failed systems was only three (1ake, 1hvr, 5cts) using method V. The best RMSDs were
2.147 Å, 2.024 Å, and 2.377 Å for 1ake, 1hvr, and 5cts, respectively. The conformational search trials
with no solutions could be easily decided as “failure”, and recalculation could be carried out
immediately. Thus, the search trials with no solutions were “desirable failure” in comparison with the
search trials in which many solutions were obtained but reasonable solutions were not generated. The
result indicated the usefulness of the multicopy method through which multiple MD simulations were
simultaneously carried out.

<table>
<thead>
<tr>
<th>setting</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_success</td>
<td>248</td>
<td>254</td>
<td>256</td>
<td>258</td>
<td>249*</td>
</tr>
<tr>
<td>success rate</td>
<td>0.953</td>
<td>0.977</td>
<td>0.985</td>
<td>0.992</td>
<td>0.988*</td>
</tr>
<tr>
<td>largest best RMSD/Å</td>
<td>2.474, 2.613</td>
<td>2.251, 2.125</td>
<td>2.377</td>
<td>2.777</td>
<td></td>
</tr>
<tr>
<td>(PDB ID)</td>
<td>(1adf)</td>
<td>(1ake)</td>
<td>(1aaq)</td>
<td>(1hbv)</td>
<td>(5cts)</td>
</tr>
<tr>
<td>num. of conformers</td>
<td>68.65</td>
<td>380.04</td>
<td>563.23</td>
<td>1606.63</td>
<td>502.84</td>
</tr>
</tbody>
</table>

* CAMDAS calculations with setting V could not be carried out for eight molecules, and n_all was equal
to 252 for setting V.
3.3. Influences of molecular flexibilities
The relationship between molecular flexibilities and the success rates are shown in figure 4. In the
figure, the results by setting I and OMEGA are illustrated. The molecules in which the numbers of
rotors are less than or equal to 5 are defined as “rigid”, and the molecules with 6 ~ 10 rotors are
defined as “medium”, and the molecules with 11 ~ rotors are “flexible”. In Veber’s drug-likeness
criterion [29], the drug-like molecules have ten or less rotatable bonds. Thus, “rigid” and “medium”
meet the Veber’s criterion in terms of the number of rotatable bonds. As shown in the figure, the
success rates for “flexible” are the worst of all molecular groups. Thus, more detailed procedures such
as settings IV and V may be useful for these flexible molecules. The success rates for “flexible” group
were 0.976 and 0.974 by using setting IV and V, respectively (data not shown). These success rates
were larger than those of setting I and OMEGA. On the other hand, CAMDAS could generate
reasonable conformations for all “rigid” and “medium” molecules, except for the tests using
conformer iv as initials. Because conformer iv was the farthest conformer from the crystal structure,
the results suggested that CAMDAS could generate reasonable conformations for drug-like molecules
unless the unnatural initial conformations are used. On the other hand, OMEGA failed the
conformational searches even for rigid molecules. Because it is difficult to prepare the countermeasure
for unanticipated failures, CAMDAS seems to be more suitable for the conformational searches of
drug-like molecules than OMEGA.

4. Conclusion
The conformational search abilities of CAMDAS were evaluated by using various initial conformers
and settings. The success rates of conformational searches by CAMDAS were comparable to those by
OMEGA, which is one of the most efficient and widely used conformational search methods.
Although the unanticipated failures rarely occurred when OMEGA was used, only “manageable
failures” occurred in the case of using CAMDAS. For example, OMEGA rarely failed for the
conformational searches of rigid molecules, and all the generated conformations were rarely far from
the crystal structures. On the other hand, the reasonable conformations could be obtained for all rigid
molecules through CAMDAS, and the conformations whose RMSDs were around 2.5 Å (or less)
could be obtained even for the failed tests. CAMDAS searched the conformations in accordance with
the physical law, and any artificial treatments were not adopted. The manageability of CAMDAS may
have been caused by the procedure without artificial treatments. The results indicated that CAMDAS
is a useful and robust tool for conformational searches of small molecules.

Acknowledgments
This work was supported by a grant-in-aid for scientific research (23790137) from Japan Society for
the Promotion of Science.
References

[5] Omega 2.2.1 (Boston, MA: OpenEye Scientific Software, Inc)
[22] SYBYL 6.9 (St. Louis: Tripos Inc.)