Journal of Molecular Graphics and Modelling 27 (2009) 628–636

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

A new method for in-silico drug screening and similarity search using molecular dynamics maximum volume overlap (MD-MVO) method

Yoshifumi Fukunishi [a,b,*], Haruki Nakamura [a,c]

[a] Biomedicinal Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan
[b] Pharmaceutical Innovation Value Chain, BioGrid Center Kansai, 1-4-2 Shinsenri-Higashimachi, Toyonaka, Osaka 560-0082, Japan
[c] Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

Article history:
Received 30 July 2008
Received in revised form 13 October 2008
Accepted 15 October 2008
Available online 1 November 2008

Keywords:
Umbrella sampling
Umbrella potential
Generalized ensemble method
Molecular alignment
Docking score index method
Molecular dynamics simulation

ABSTRACT

We developed a new molecular dynamics simulation method for molecular overlapping (alignment) and ligand-based in-silico drug screening based on molecular similarity. The molecular system consists of the query compound and the other compound(s) selected from a compound library. The newly introduced intermolecular interaction between compounds is proportional to the molecular overlap instead of the van der Waals and Coulomb interactions between atoms of different molecules. This method was able to achieve both conformer generation of molecules and molecular overlapping (alignment) at the same time. After an energy minimization and following short-time MD simulation, the molecules in the system were overlapped with each other and the similarity between compounds was measured by the volume of the overlap. We applied this MD simulation method to ligand-based in-silico drug screening and found that it worked well for several targets.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Ligand-based drug screening is one of the most popular methods of in-silico drug screening, and there have been many reports on it [1–5]. Ligand-based screening is essentially a similarity search based on known active compounds. Recent progress of the similarity search is summarized in the review article [5]. To perform a similarity search, one approach is to use a molecular descriptor like the MACCS key, developed by Molecular Design Limited (MDL) together with a Daylight descriptor (Daylight Chemical Information Systems Inc., Aliso Viejo, CA, USA); another approach is to overlap the known active compound with compounds selected from a compound library. In the latter approach, the two molecules can be overlapped directly, or a pharmacophore can be generated by aligning several known active compounds, and then overlapping the database compounds with this pharmacophore. This method is known as a pharmacophore search, and the comparative molecular field analysis (CoMFA) method is one of the well-known methods of this category [6,7]. These methods could be regarded as a type of quantitative structure–activity relationship (QSAR) or 4D-QSAR method [8–11].

The similarity search by molecular overlapping is a somewhat time-consuming method compared to the similarity search by molecular descriptor, since both the conformer generation of compounds and the overlapping are time-consuming processes. The similarity search by molecular overlapping is suitable for scaffold hopping (lead hopping) [12] and the generated pharmacophore could be helpful in the lead-optimization process.

In the similarity search by molecular descriptor, each compound is projected into high-dimensional space constructed by the molecular descriptors. The MACCS key is one of the most popular descriptors. BCUT and CATS descriptors, which are based on the molecular topology or 3D structure, are also used [13,14]. Usually, this kind of similarity search is computationally very fast, and thus is useful in generating the wide chemical space and classifying the compounds in the library.

In this paper, we experimented with a similarity search by molecular overlapping using molecular dynamics (MD) simulation. Using our method, conformer generation and overlapping could be performed at the same time. Usually, conformer generation of a large-size compound like a peptide is difficult, but MD simulation is suitable for such conformer generation using simple MD simulation at high-temperature or generalized ensemble methods [15–24]. CAMDAS is one of these applications of the MD to the conformer analysis in medicinal chemistry [25]. One of the molecular similarity measures is the overlap volume of

* Corresponding author. Tel.: +81 3 3599 8290; fax: +81 3 3599 8099.
E-mail address: y-fukunishi@aist.go.jp (Y. Fukunishi).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.10.003

molecules [26–28]. The volume overlap could be approximated by the overlap of Gaussian type functions that represent the distributions of electrons [26–28]. We previously developed the maximum volume overlap (MVO) method, which is a molecular overlapping method for prediction of protein–ligand complex structures [29]. The overlap of molecules is represented by the overlap of Gaussian type functions. The MVO method is a scoring method that selects protein–ligand complex structures with higher geometrical accuracy than the top-scoring complex structure of other methods, using the structural information of known protein–ligand complexes. To apply this method, one or more protein–ligand complex structures for the target protein must be known. A number of predicted structures are generated by the protein–compound docking program for a new ligand, and one of these structures, which shows the maximum overlap with the ligand coordinates of the known protein–ligand complex, is selected as the most probable complex structure. The MVO method was shown to be very effective. Namely, the number of correctly predicted structures (RMSD <2 Å) was drastically improved, with 2- and 2.5-fold improvements obtained by the single-reference and two-reference docking methods compared to the conventional docking method.

In this study, we applied MD simulation to the MVO scoring, expecting that both conformer generation and molecular overlapping could be achieved by this method. One of the weak points of the MD simulation is its time-consuming nature. The recent availability of specialist computer hardware for MD simulation can accelerate the MD simulation by about 100 times [30,31]. As a consequence, application of MD simulation is expected to become more popular in future pharmaceutical science.

2. Method

2.1. Molecular dynamics maximum volume overlap (MD-MVO) method

We developed a new similarity search method, consisting of a molecular dynamics simulation method with a restrained potential: namely, the potential used in the MVO method previously developed by us [29], which represents the overlap of two or more molecules. This method consists of several steps. The whole system consists of the template molecule (known active compound), query molecule(s) (compounds in the database), and other environmental molecules (proteins, solvent). At first, the conventional Hamiltonian represents the intramolecule interactions and the intermolecule interactions except the template–query interactions. At second, the overlap between the template and the query represents the template–query interaction. At third, the whole system is treated as a simulated annealing problem by using an MD simulation. At last, the similarity measure between the template and the query is given as the overlap between the template and the query.

Let molecules A, B and C be the template molecule (known active compound), query molecule(s) (compounds in the database), and other molecules, that are the target protein and solvent molecules, respectively. Let $H_{\alpha\alpha}$ and $H_{\alpha\beta}$ (=$H_{\beta\alpha}$) be the Hamiltonian of molecule α, and the interaction energy between molecules α and β, respectively. Here, α and β are A, B and C. The total Hamiltonian (H) of this system is:

$$H = H_{AA} + H_{BB} + H_{CC} + 0.5 \times \lambda_1 (H_{AB} + H_{BA}) + 0.5 \times \lambda_2 (H_{AC} + H_{CA} + H_{BC} + H_{CB}), \quad (1)$$

where λ_1 and λ_2 are parameters. Here H_{AA}, H_{BB} and H_{CC} are the conventional Hamiltonians representing the sum of the usual kinetic energy and the usual potential energy of atomic interactions. H_{AC} (=H_{CA}) and H_{BC} (=H_{CB}) are the conventional potential energies.

H_{AB} (=H_{BA}) is the newly introduced interaction potential. Let $\{x_A^i, y_A^i, z_A^i\}$ be the $\{x, y, z\}$ coordinates of the ith atom of molecule A, and let $\{x_B^j, y_B^j, z_B^j\}$ be the $\{x, y, z\}$ coordinates of the jth atom of molecule B. Let q_i^A and q_j^B be the atomic charge of the ith atom of molecule A and the atomic charge of the jth atom of molecule B. Then,

$$H_{AB} = \alpha \sum_{j=1}^{N_A} \sum_{i=1}^{N_B} w(i,j) \exp(-c((x_A^i - x_B^j)^2 + (y_A^i - y_B^j)^2 + (z_A^i - z_B^j)^2)) \quad (2)$$

and

$$w(i,j) = \begin{cases} -1 & \left|q_i^A - q_j^B\right| < q_{thr} \\ 0 & \left|q_i^A - q_j^B\right| \geq q_{thr} \end{cases} \quad (3)$$

where α, N_A, N_B, c, w, and q_{thr} are the conversion factor, the number of atoms of molecule A, the number of atoms of molecule B, a coefficient, a switching function, and a threshold value, respectively. In the present study, the parameters α, c and q_{thr} were set as 1 kcal/mol, 1.0 Å^{-2} and 0.2 in the atomic unit, respectively, as in our previous study [29]. This method is called the MVO method.

The energy gradient of H_{AB} is:

$$F(x_A^i) = -\frac{\partial H_{AB}}{\partial x_A^i} = \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} -2c(x_A^i - x_B^j) w(i,j) \exp(-c((x_A^i - x_B^j)^2 + (y_A^i - y_B^j)^2 + (z_A^i - z_B^j)^2)) \quad (4)$$

and

$$F(x_B^j) = -\frac{\partial H_{AB}}{\partial x_B^j} = \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} 2c(x_A^i - x_B^j) w(i,j) \exp(-c((x_A^i - x_B^j)^2 + (y_A^i - y_B^j)^2 + (z_A^i - z_B^j)^2)).$$

H_{AB} works to overlap molecules A and B.

The intramolecular interactions of molecules A, B, and C are calculated by the usual Hamiltonians. Namely, these Hamiltonians consist of bond, angle, torsion angle, improper torsion angle, 1–4 and 1–5 van der Waals and Coulomb interactions. The intermolecular interaction between A and B is the MVO potential by H_{AB}. This implies that there is no van der Waals or Coulomb interactions between A and B. The intermolecular interaction between A and C (and between B and C) is the usual Hamiltoian H_{AC} (H_{BC}), which is scaled by the parameter λ_2.

We can apply the MD simulation with a simulated annealing to the system with this Hamiltonian. As the result, A and B are shown to be potentially overlapped by the MVO.

The score S_{AB} is the measure used to evaluate the overlap between molecules A and B. The simplest definition is:

$$S_{AB} = -H_{AB}. \quad (6)$$

The alternative definitions are:

$$S_{AB} = \frac{H_{AB}}{(H_{AA} + H_{BB}) \times 0.5} \quad (7)$$

or

$$S_{AB} = -\frac{H_{AB}}{\sqrt{H_{AA} H_{BB}}} \quad (8)$$

While many kinds of definitions are available, our screening tests showed that the S_{AB} by Eq. (7) is the most useful. After the MD simulation based on the Hamiltonian by Eq. (1), the score of the

final snap-shot structure is evaluated by Eq. (7). Test molecules that show high scores for the query molecule are chosen as the hit compounds.

2.2. Docking score index (DSI) method

We also used a ligand-based drug screening method based on a protein–compound affinity matrix, called the DSI method [32,33]. This is a sort of "affinity fingerprint" approach [34–36]. The molecular descriptor is a set of docking scores of molecules with many proteins, and the DSI method finds a compound that is similar to the template compound (known active compound) based on the descriptor. To reduce the computational error of the docking score, the principal component analysis (PCA) method was applied to the protein–compound interaction matrix, which was given by thorough docking calculations between the set of many protein pockets and chemical compounds. In this method each compound and protein pocket is depicted as a point in the PCA space of compounds and proteins, respectively. Compounds that are close to the known active compounds are selected as candidate hit compounds [32].

Furthermore, we assumed that the protein–compound binding free energy of a compound could be improved by a linear combination of its docking scores with many different proteins [33,37],

$$s_a^{new^i} = \sum_b s_b^i M_a^b, \tag{9}$$

where $s_a^{new^i}$, s_b^i and M_a^b are the modified docking score of the ath protein and the ith compound, the raw docking score of the bth protein and the ith compound, and the constant coefficient, respectively. The problem is how to determine the coefficient M_a^b without any experimental observation of the binding free energy.

The machine-learning docking score index (ML-DSI) method could optimize the docking scores of the protein–compound affinity matrix to maximize the database enrichment of the known active compounds, providing an optimized focused library. The ML-DSI method optimized the M_a^b coefficient by the Monte-Carlo method to maximize the database enrichment. The ML-DSI method requires more than two known active compounds as the template compounds, since this approach is a machine-learning approach. A test calculation showed that about 70% of the active compounds were found within the first 1% of the database by the MS-DSI method.

2.3. Machine-learning score modification multiple target screening (MSM-MTS) method

We also used a structure-based drug screening method based on a protein–compound affinity matrix, called the MSM-MTS method [37]. This is also a sort of "affinity fingerprint" approach. The basic idea of the MTS method is that the potential active compound is a compound, which shows the strongest affinity with the target protein. Thus, based on the protein–compound affinity matrix, the compounds, which show the strongest affinity with the target protein, are selected as the hit compounds. If the active compounds of the target protein were known, the docking score modification by Eq. (9) is applied as same as the ML-DSI method. The MSM-MTS method could optimize the docking scores of the protein–compound affinity matrix to maximize the database enrichment of the known active compounds, providing an optimized focused library. The MSM-MTS method optimized the M_a^b coefficient by the Monte-Carlo method to maximize the database enrichment as same as the ML-DSI method. The MSM-MTS method can be applied, even if only one known active compound is available as the template compound. A test calculation showed that about 40% of the active compounds were found within the first 1% of the database by the MSM-MTS method [37].

3. Result

3.1. Screening procedure by the MD-MVO method without target protein structure

We modified the cosgene MD program of myPresto free software package (http://presto.protein.osaka-u.ac.jp/myPresto4/index_e.html) [38] and applied the modified program to in-silico drug screening test of five well-known targets. The drug screening test is a leave-one-out test. Namely, one known active compound is given to the program as the known active compound, and then the program finds the other active compound from a compound library. In this section, λ_1 in Eq. (1) was set to 1 and λ_2 was set to 0. This means that the system consists of only a template compound (known active compound) and a query compound of the database without protein structure or solvent.

The five targets were selected for the validation test of the MD-MVO method. These five targets were the human immunodeficiency virus protease-1 (HIV), cyclooxygenase-2 (COX2), thermolysin (THR), glutathione S-transferase (GST) and μ opioid receptor (μOR). The compound set consisted of 28 inhibitors of THR, 14 inhibitors of COX2, 19 inhibitors of HIV, 12 inhibitors of GST, 12 agonists of μOR and 11,050 potential-negative compounds of the Coelacanth chemical compound library (Coelacanth Corporation, East Windsor, NJ, USA), which is a random library. The other decoy sets were also prepared to confirm the screening results. Two decoy sets for COX2 and HIV were retrieved from the directory of useful decoy (DUD) database [39]. 1000 compounds were randomly selected from the decoy sets for COX2 and HIV, respectively. Usually only one hit compound is found out of 10^4 randomly selected compounds; thus, we expected that there was no or only a few hit compounds among these 11,050 compounds of the Coelacanth decoy set and the 1000 compounds of the DUD sets. The active compounds of COX2, HIV, GST, THR, and μOR are listed in Appendix A. These compounds and the target proteins were used in our previous studies and the molecular structures were depicted in our previous reports [37]. The 3D coordinates of the 11,050 chemical compounds of the Coelacanth chemical compound library were generated by the Concord program (Tripos, St. Louis, MO) from the 2D Sybyl SD files provided by the Coelacanth Chemical Corporation. The 3D coordinates of the known ligands were generated by the Chem3D program (Cambridge Software, Cambridge, MA, USA). We used the general AMBER force field (GAFF) [40], and the molecular topology files were generated by the tplgeneL/myPresto. The atomic charges were calculated by the Gasteiger method of the Hgene/myPresto [41,42].

The computational procedure of the in-silico screening was as follows. Each system consisted of a template compound, which is a known active compound, and a query compound selected from the compound library. Then, the alignment (overlapping) of the two molecules was performed by the following four steps, and the score was calculated by Eq. (7).

- Step 1: The centers of mass of the two molecules were overlapped. Then, one of the molecules was rotated randomly. Five initial coordinates were prepared by the random rotation.
- Step 2: To the five initial coordinates, the energy minimization was applied. In this step, λ_1 in Eq. (1) was set to 0 and λ_2 was set

to 0. This means that there was no interaction between the two molecules.

- Step 3: The MD was applied to the five systems with the temperature = 700K, time step = 2.0 fs with the rigid model, and harmonic restraint potential. The harmonic potential restrains the centers of mass of the two molecules to the origin of the coordinate. It is 0 Å-radius CAP potential with the force constant of 1.0 kcal/(mol Å^2). Without this restraint potential, these two molecules come off each other at the high temperature. After the 10 ps MD, the temperature set to 100 K and 1 ps MD gave the final coordinates. In this step, λ_1 in Eq. (1) was set to 1 and λ_2 was set to 0.
- Step 4: Energy minimization was applied to the five systems. In this step, λ_1 in Eq. (1) was set to 1 and λ_2 was set to 0. The similarity score was calculated for the energy minimized structure. The similarity score of the given molecule was the best value of the MVO potential values of the five trial systems.

We tried two procedures. In the first method, both the template and query molecules could move without position restraint (method A). In the second method, the coordinates of the template compound were fixed to the binding position of the protein–ligand complex structure through steps 1–4 by the position restraint potential (method B). The force constant for the position restraint is 1 kcal/(mol Å^2). Method A is a ligand-based screening method, and method B is a sort of structure-based screening method. But, the binding pose of template molecule could be determined by conformational analysis by NMR or molecular overlapping of several known active compounds.

In addition to the MD-MVO method, we also applied the DSI method to the five targets, COX2, GST, HIV, THR, and μOR. The number of template compounds was only one; thus, the machine-learning approach was not available. The number of principal components used was 10. The protein set was exactly the same as that used in our previous study [33] and the number of proteins was 180. The compound set was exactly the same as that used for the MD-MVO method.

3.2. Screening results without target protein structure

Let x and $f(x)$ be the numbers of compounds (%) selected from the total compound library and from the database enrichment curve, respectively. The surface area under the database enrichment curve (q) is a measure of the database enrichment.

$$q = \int_0^{100} f(x)\,dx \tag{10}$$

Higher q values correspond to better database enrichment, and $0 < q < 100$. The q value by a random screening is 50. The q value is almost the same as the area under the receiver operating characteristic (ROC) curve (AUC), when the number of active compounds is much smaller than the number of the decoy compounds.

Table 1 shows the q values and the hit ratios at the first 1% of the entries in the database obtained by the MD-MVO and the DSI methods. Fig. 1 shows the database enrichment curves of these targets by the MD-MVO and the DSI methods. The PDB names of the complexes, in which the active compounds are included, are listed in Table 1.

The MD-MVO method was very effective for COX2, GST and HIV, and it was also effective for THR and μOR. The hit ratio and the q value depended on the template compound. The DSI method was also effective for COX2 and HIV, but it was not effective for GST and THR at all. We compared the current results with the results by our previous studies, since we used almost the same compound library and the active compounds were prepared in the same manner. The screening result depends on the compound library used.

Table 1
The q values and the hit ratios by the MD-MVO and DSI methods.

	Method A		DSI		Method B		MSM-MTS	
	q value	Hit ratio	q value	Hit ratio	q value	Hit ratio	q value	Hit ratio
COX2								
1cx2	83.4	35.7	67.7	6.3	81.3	33.3	85.7	6.7
1pxx	98.0	57.1	69.6	6.3	91.2	41.7	79.5	13.3
3pgh	85.7	35.7	71.8	0.0	87.8	25.0	87.6	25.0
4cox	75.6	28.6	72.9	12.5	65.0	16.7	81.2	20.0
GST								
10gs	51.1	33.3	55.3	8.3	76.2	50.0	35.9	9.1
12gs	76.9	75.0	57.5	16.7	79.9	66.7	58.1	9.1
18gs	86.0	66.7	47.1	8.3	90.1	41.7	75.7	18.2
1aqx	81.7	33.3	46.3	8.3	86.3	58.3	64.7	9.1
HIV								
1aid	46.7	0.0	66.4	5.0	54.1	5.0	37.2	5.3
1bv7	89.7	35.0	76.0	15.0	93.9	40.0	77.4	10.5
1hte	62.1	15.0	70.6	10.0	63.8	10.0	37.7	5.3
1mes	82.4	45.0	71.9	15.0	90.2	40.0	77.4	10.5
THR								
1hyt	87.5	31.4	36.7	2.0	85.4	35.7	75.7	12.5
1lna	68.9	2.0	37.2	2.0	79.1	21.4	37.0	4.2
1os0	42.2	19.6	37.3	3.9	66.4	21.4	73.0	4.2
1qf0	47.0	17.6	37.3	3.9	69.9	28.6	47.6	4.2
1qf1	53.0	15.7	37.3	3.9	73.4	14.3	74.8	8.3
1qf2	49.3	11.8	37.3	3.9	66.9	7.1	82.2	4.2
1thl	40.7	9.8	37.3	3.9	58.5	21.4	70.2	8.3
1tlp	42.3	9.8	45.9	15.7	61.4	14.3	59.7	0.0
1tmn	46.0	13.7	46.0	17.6	55.9	21.4	23.3	0.0
1z9g	85.0	19.7	37.3	3.9	82.4	21.4	75.3	37.5
1zdp	85.4	37.3	37.3	3.9	84.6	21.4	79.2	12.5
2tmn	82.3	31.4	46.5	27.5	86.7	28.6	76.4	8.3
5tln	66.2	15.7	37.3	3.9	76.2	14.3	76.0	20.8
5tmn	42.0	9.8	37.3	3.9	61.4	7.1	88.7	8.3
μOR								
End1	35.6	0.0	75.7	13.3	–	–	–	–
End2	34.9	0.0	69.9	13.3	–	–	–	–
End3	75.1	9.1	65.9	6.7	–	–	–	–
End4	90.4	9.1	73.5	6.7	–	–	–	–
Average	66.4	24.1	53.5	8.4	–	–	–	–
Except μOR								
Average	67.6	27.1	50.8	8.1	75.7	27.2	66.8	10.6

The q value is the area under the database enrichment curve. The q value is 50 for random screening, and the maximum value is 100. The hit ratio is the hit ratio at the first 1% of the entries in the database.

The average hit ratio of 27.1% by method A and that of 27.2% by method B in the current study were higher than that of 8.1% by the DSI method, while these values were much lower than the hit ratio of 70% by the machine-learning docking score index (ML-DSI) method [33]. Only one active compound was used as the template compound in the current study; on the contrary, several active compounds were used as the set of template compounds in the previous study. If the number of template compounds is one, the machine-learning approach cannot work in the ML-DSI method. Without the machine-learning, the DSI method gave the hit ratio of 12.4% even if several active compounds were used as the template compounds [33]. Thus, the results by the current method are better than those by the previous method. Of course, the ML-DSI and the DSI methods are several-hundreds times faster than the MD-MVO method.

The MD-MVO method was compared to another shape matching program, ROCS version 2.3.1 (OpenEye Scientific Soft-

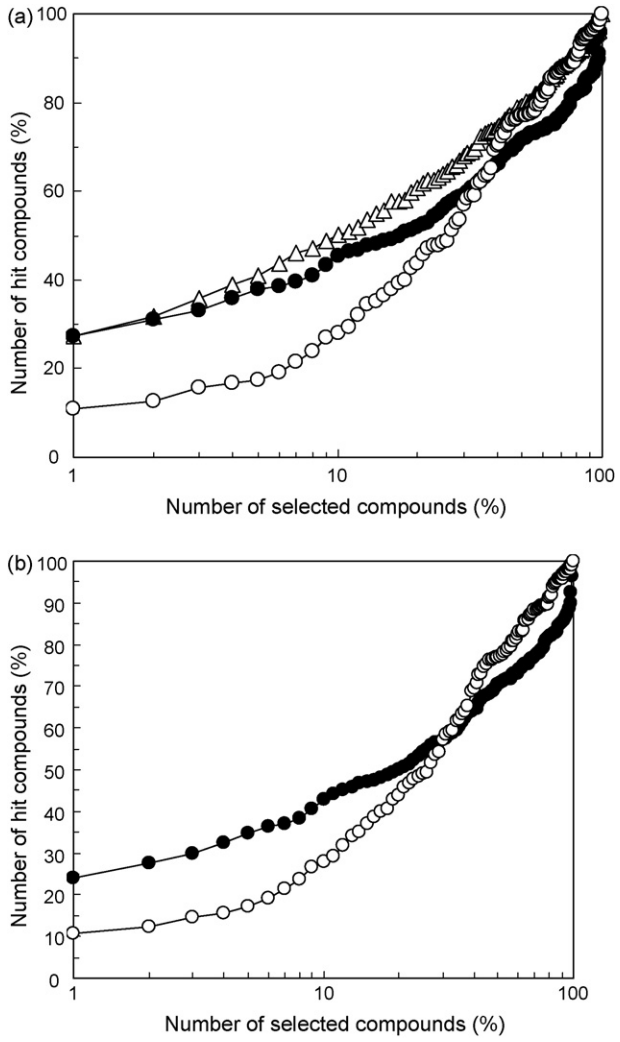

Fig. 1. Database enrichment curves by the MD-MVO and the DSI methods. (a) The averaged database enrichment curves for COX2, GST, HIV, and THR. The filled circles, triangles and the open circles represent the averaged database enrichment curves by method A (MD-MVO), method B (MD-MVO), and the DSI method. In method A, both the query and test compounds could move freely in the MD simulation. In method B, the coordinates of the query compounds were fixed to those of the actual binding positions. (b) The averaged database enrichment curves for COX2, GST, HIV, THR and μOR. The filled circles and the open circles represent the averaged database enrichment curves by the method A (MD-MVO) and the DSI method. Since the protein–ligand complex structure of μOR is unknown, method B could not applied to μOR.

ware, Inc., Santa Fe, NM, USA) [26–28]. The decoy set was the DUD of 1000 compounds. The results were summarized in Table 2. The average hit ratio at the first 1% of the entries in the database was 28.8% by the MD-MVO method (method A), and this value is almost equivalent to the value of 23.9% by the ROCS software. Also, the average q value obtained by the MD-MVO method was 81.9, which was better than the average q value, 61.5, obtained by the ROCS software. These results show that the MD-MVO method is better than or is almost equivalent to a shape matching method, ROCS. However, the ROCS was about 30,000 times faster than the MD-MVO method, because the MD-MVO method requires several thousands MD-simulated annealing steps.

In addition, the MD-MVO method was compared to a graph matching program, ChemFinder version 7.0 (Cambridge Software, Cambridge, MA, USA). The decoy set was the DUD of 1000 compounds. The similarity threshold value was set to 40% for ChemFinder. For COX2, average 79 compounds were selected from the dataset and average 1.5 active compounds were selected. It means that the average hit ratio at the first 7.9% of the entries in the database was 9.8% by ChemFinder. The average hit ratio at the first 8% of the entries in the database was 38.2% by the MD-MVO method (method A). For HIV, average 224 compounds were selected from the dataset and average 10 active compounds were selected. It means that the average hit ratio at the first 22.4% of the entries in the database was 47.5% by ChemFinder. The average hit ratio at the first 22% of the entries in the database was 51.4% by the MD-MVO method (method A). These results show that the MD-MVO method is better than or is almost equivalent to a graph matching method, ChemFinder.

Table 2
Decoy-set dependence of screening results by the MD-MVO method (method A) and the comparison between the MD-MVO method and ROCS.

	MD-MVO (method A)				ROCS	
	Coelacanth		DUD		DUD	
	q value	Hit ratio	q value	Hit ratio	q value	Hit ratio
1cx2	83.4	35.7	85.8	28.6	52.8	23.1
1pxx	98.0	57.1	98.5	50.0	64.5	30.8
3pgh	85.7	35.7	85.3	33.3	64.8	30.8
4cox	75.6	28.6	74.0	33.3	68.5	15.4
1aid	46.7	0.0	59.8	5.0	29.2	4.5
1bv7	89.7	35.0	93.3	40.0	75.8	36.4
1hte	62.1	15.0	70.3	5.0	59.3	13.6
1mes	82.4	45.0	88.2	35.0	77.0	36.4
Average	77.9	31.5	81.9	28.8	61.5	23.9

The q value and the hit ratio (%) at the first 1% of the entries in the database.

For μOR, an intrinsic opioid peptide was selected as the template. The amino acid sequence of the selected intrinsic opioid peptide, endomorphin, is Tyr-Pro-Trp-Phe-NH_2 [43]. Since the first amino acid, Tyr, is common to all μ-opioid peptides, along with the phenyl group–C–C–N, the N-terminal of the peptide should be important. Thus, we prepared four peptides, Tyr-Pro-Trp-Phe-NH_2 (End1), Tyr-Pro-Trp-Phe (End2), Tyr-Pro-Trp (End3), and Tyr-Pro (End4). The four N-terminal amino acid residues of β-endorphin, which is another μ-opioid, are Tyr-Gly-Gly-Phe. Comparing the N-termini of endomorphin and endorphin, Tyr is essential and the 4th residue, Phe, would be important; we are not sure whether End2, End3, and End4 are active or not.

The MD-MVO method showed a good hit ratio for End4, and the hit ratio and the q value by the MD-MVO method were better than those by the DSI method. The longer the template peptide is, the worse the hit ratio is. The hit ratios and the q values for End1, End2, and End3 by the MD-MVO method were worse than those by the DSI method. To find a compound that mimics the peptide activity using the MD-MVO method, the query peptide must be a small fragment of the peptide.

The average q value by method B was better than that by method A. Thus, the consideration of the binding position of the active compound is important in the MD-MVO method. The average database enrichments by methods A and B were good, but methods A and B showed low q values <50 for 10 out of 30 targets and 0 out of 26 targets, respectively. For method B, the lowest hit ratio was 5.0% for $q = 54.1$. This means that the method B worked in many cases. The previous study showed that many in-silico screening programs achieved high hit ratios in almost half cases and they fail to achieve high hit ratio in the other half cases [44]. Comparing to these results, the hit ratio by the MD-MVO method was not so bad.

Method A is a ligand-based screening method. On the contrary, method B is a sort of structure-based screening method, since this method utilizes the binding pose of the known active compound.

Thus, we compared the results by method B to those by the MSM-MTS method, which is a structure-based screening method previously developed by us [37]. The known active compound for the MSM-MTS method is the same compound of the template compound for method B. The results were summarized in Table 1. The hit ratio and the q value obtained by method B was slightly better than those obtained by the MSM-MTS method. The throughput of the MSM-MTS method is much higher than that of method B, but still method B has an advantage to the MSM-MTS method. The binding pose of the template molecule could be determined besides the X-ray structure analysis of protein–ligand complex structure, since the NMR analysis or CoMFA method [6–11] could predict the binding pose of ligand without the protein–ligand complex structure.

The decoy-set dependence of screening result was also examined. The in-silico (virtual) screening results could depend on the used decoy set, so that the hit ratio by a screening method based on a single decoy set could have some artificial effects. Instead of the Coelacanth decoy set, the decoy sets retrieved from the DUD were adopted. Table 2 shows the q values and the hit ratio at the first 1% of the entries in the database obtained by the MD-MVO method (method A). The results based on the DUD were similar to the results based on the Coelacanth decoy set. Thus, the efficiency of the MD-MVO method was confirmed.

We also studied the effect of the computational procedure on the screening result. In step 3, high temperatures were set at 500 and 700 K, and the screening results did not depend on the temperature of the MD simulation. In steps 1 and 4, we tried minimization steps of 100, 500, and 1000 steps, and the screening results did not depend on the number of minimization steps. Table 3 shows the MD-step dependence of hit ratio. The decoy set was the DUD. Four cases were examined; case 1 with 2500 steps at high temperature and 1250 steps at low temperature, case 2 with 5000 steps at high temperature and 2500 steps at low temperature, case 3 with 10,000 steps at high temperature and 5000 steps at low temperature, and case 4 with 20,000 steps at high temperature and 10,000 steps at low temperature. The high temperature was 700 K and the low temperature was 100 K. The hit ratio and the q value did not strongly depend on the number of MD simulation steps. The hit ratio and the q values of case 4 were worse than those of case 4. The initial coordinates depend on the random number and the MD simulation could generate different coordinates, thus, the small differences among these data are within the error range.

The screening results depended on the radius of the CAP constraint. The CAP potential restrains the center of mass of the molecule within a spherical region. The screening results with the CAP radius of 5 Å were worse than the screening results with the CAP radius of 0 Å (the harmonic restraint potential). The screening results did not depend on the integral method or the time step of the MD simulation. The screening results by the Verlet-leap frog method with a SHAKE restraint [45] and time step of 1.5 fs, by the velocity-Verlet method with a rigid model and time step of 2 fs [46], and by the Verlet-leap frog method with an all-atom model and time step of 0.5 fs were almost equivalent to each other.

The screening result strongly depended on the number of initial coordinates in step 1. Table 3 shows the conformer dependence of the hit ratio for the DUD decoy set. The screening results with five initial coordinates were better than those with only one initial coordinate. When the number of initial coordinates was 10 or 20, the hit ratios and the q values were not improved comparing to the results with five initial coordinates, and they were rather worse than those with five conformers. The initial coordinates depend on the random number, thus, the small differences among these data are within the error range. The CPU time is proportional to the number of initial coordinates, so that the screening with five initial coordinates was suitable for this study.

3.3. Screening procedure and results by the MD-MVO method with target protein structure

In this section, both λ_1 and λ_2 in Eq. (1) were set to 1. This means that the system consists of a template compound (known active compound), a query compound of the database with a target protein structure. This procedure could be performed only when the 3D coordinates of the protein–template complex structure are known.

COX2 (PDB: 1cx2) was selected as the target protein of the MD-MVO method. The template molecule is the ligand of the COX2 complex structure (PDB: 1cx2). Two query molecules were used; one (molecule **a**) is the ligand of a COX2 complex structure (PDB: 4cox) and the other (molecule **b**) is a decoy molecule selected from the Coelacanth library. The SMILES representations of molecules **a** and **b** are shown in Appendix B. We used the AMBER99 force field [47] and its atomic charges for the protein molecule. The atomic charges of the template and query molecules were calculated by the Gasteiger method of the Hgene/myPresto [41,42].

The computational procedure of the MD-MVO method was as follows. Each system consisted of a template compound, a query compound, and the target protein. Then, the alignment (overlapping) of the two molecules was performed by the following four steps, and the score was calculated by Eq. (7).

- Step 1: The query compound was docked to the target protein by using the protein–compound flexible docking program, sievgene/myPresto [48]. Top-ranked five coordinates were obtained as the initial coordinates.
- Step 2: To each of the five initial coordinates, the energy minimization was applied, individually. In this step, λ_1 in Eq. (1) was set to 0 and λ_2 was set to 1. The position restraint potential fixed the all main-chain atoms of the target protein. Namely,

Table 3
Performance of the MD-MVO method (method A) under several conditions.

Case[b]	Initial[a]							
	One conformer		Five conformers		10 conformers		20 conformers	
	q value[c]	Hit ratio[c]	q value[c]	Hit ratio[c]	q value[c]	Hit ratio[c]	q value[c]	Hit ratio[c]
Case 1	–	–	71.9	16.2	–	–	–	–
Case 2	–	–	71.4	20	–	–	–	–
Case 3	68.1	11.9	81.9	28.8	73.6	19.6	80.5	21.7
Case 4	–	–	79.2	26.7	–	–	–	–

[a] The number of initial coordinates.
[b] The definitions of the case 1–4 for MD step numbers and temperatures are described in the text.
[c] The average q value and the average hit ratio (%) at the first 1% of the entries in the DUD database. The used templates were listed in Table 2; four COX2 templates and four HIV templates.

Table 4
Results by the MD-MVO method for the "template + query + target protein" system.

Rank[a]	Compound **a**					Compound **b**	
	Score[b]	RMSD[c]	Score[d]	RMSD[e]	MVO score[f]	Score[b]	MVO score[f]
0	–	–	–	–	0.55	–	0.61
1	−3.57	2.25	−3.47	1.06	0.47	−3.53	0.45
2	−3.50	2.40	−3.44	0.26	0.45	−3.48	0.45
3	−3.43	1.97	−3.54	0.99	0.45	−3.46	0.46
4	−3.38	2.59	−3.50	1.13	0.44	−3.40	0.45
5	−3.34	6.78	−2.45	1.56	0.40	−3.38	0.37

[a] The number (1–5) represents the rank number by sievgene docking program. "0" represents the result by MD-MVO method (method A) with $\lambda_1 = 1$ and $\lambda_2 = 0$.
[b] The score represents the sievgene docking score.
[c] The RMSD represents the RMSD error of the heavy atoms of the ligand (compound **a**) obtained by sievgene from the experimental complex structure (PDB ID: 4cox).
[d] The sievgene score of the final snap-shot structure obtained by the MD-MVO method with $\lambda_1 = 1$ and $\lambda_2 = 1$.
[e] The RMSD represents the RMSD error of the heavy atoms of the ligand (compound **a**) obtained by the MD-MVO method with $\lambda_1 = 1$ and $\lambda_2 = 1$ from the experimental complex structure (PDB ID: 4cox).
[f] MD-MVO score of the final snap-shot structure obtained by the MD-MVO method with $\lambda_1 = 1$ and $\lambda_2 = 1$.

there was no interaction between the template and the query molecules, but the interaction energies between the template and the protein, and those between the query and the protein were considered.

- Step 3: The MD was applied to each of the five systems with the temperature = 700K, time step = 2.0 fs with the rigid model, respectively. The position restraint potential fixed the all main-chain atoms of the target protein. After the 10 ps MD, the simulated annealing procedure with the temperature 100 K and 1 ps MD steps gave the final coordinates. In these steps, both λ_1 and λ_2 in Eq. (1) were set to 1.
- Step 4: Energy minimization was applied to each of the five systems, individually. In this step, both λ_1 and λ_2 in Eq. (1) were set to 1. The position restraint potential fixed the all main-chain atoms of the target protein. The similarity score was calculated for the energy minimized structure. During the procedures from step 2 to step 4, 10 Å cutoff length was applied to the van der Waals and the electrostatic interactions. The dielectric constant was set to $4R$, where R is the inter-atomic distance in Å.

The results were summarized in Table 4. The MVO score by method A with $\lambda_1 = 1$ and $\lambda_2 = 0$ for a decoy molecule **b** was better than that of the active compound molecule **a**. It suggests that the method is not very powerful enough to identify the correct active compounds. The sievgene score of molecule **a** was better than that of molecule **b**, where the smaller value of the sievgene score represents the stronger affinity. This result is correct, but the single docking score is not always reliable as well known. The MD-MVO method with the target protein structure with $\lambda_1 = 1$ and $\lambda_2 = 0$ was applied to the five coordinates obtained by the docking program. Then, the best MVO score of molecule **a** was 0.47, which is slightly larger than that of molecule **b**, 0.46, suggesting that molecule **a** can be a better candidate than molecule **b** as the correct answer. In addition, the MD-MVO method improved the predicted complex structure. The RMSD values by the sievgene docking program were drastically improved by the MD-MVO method.

4. Discussion

The screening results did not depend on the computational conditions of the MD simulation but on the number of the initial coordinates. This means that the local minimum search by the energy minimization and the MD simulation was almost completed and the global minimum search was not completed by the current procedure. The side-chain rotation of the amino acid side chain on the protein surface takes about 100 ps at room temperature [49]. The 10 ps MD simulation in step 3 is obviously too short to generate a sufficient number of conformations for molecular overlapping in the current study. The hit ratio could be improved by using an exhaustive simulated annealing or Monte-Carlo conformational search method, increasing the number of the initial coordinates, and using a longer MD simulation than the current MD simulation. The MD-MVO method searched only 300 compounds a day by the current procedure, and the screening speed was 10^3 to 10^4 times slower than the usual similarity search method based on a molecular descriptor. The improvement of the hit ratio requires a large-scale PC cluster.

In the current study, only one active compound was used as the template compound. Theoretically, two or more active compounds could be used as the template compounds as in the previous work [33,37]. In our previous work, the MVO potential was used to select a more precise protein–ligand complex structure among the many complex structures predicted by the software than the best-scored complex structure by the docking software. The accuracy of the selected protein–ligand complex structure was improved by increasing the number of reference compound structures. The reference compound corresponds to the known active compound. Thus, the hit ratio by the MD-MVO method could be improved by using two or more known active compounds.

The MD-MVO method is a kind of QSAR method so that the MD-MVO method can find a compound that matches the pharmacophore indicated by the template compound, but the MD-MVO method cannot find a compound that shows a different binding position from the indicated pharmacophore, just like the COMBINE method [50]. In the MD-MVO method, the force field by the target protein could be considered as molecule C in Eq. (1). The consideration of the target protein structure could improve the limit of the current procedure.

The energy minimization step of the MD-MVO method minimizes the potential energy part of Eq. (1). This means that the MVO potential by Eq. (2) and the score by Eq. (7) are not optimized and the realistic molecular structure should be kept. The measure used in the structural search is different from the measure for the evaluation. The purpose of the MD-MVO method is not generation of a canonical ensemble but estimation of molecular similarity. This is not the unusual approach. In the protein–compound docking study, the Hamiltonian or the total energy, on which the protein–compound complex structure is generated, is usually different from the so-called docking score, which represents the protein–compound binding free energy.

In practical use, the consensus results of method A and the DSI method should be more useful than the results obtained by sole screening method. Also, the consensus results of method B and the MSM-MTS method should be more useful than the results obtained

by sole method. The consensus scoring improved the screening results by the DSI and MTS methods [51], so that the idea of consensus scoring would improve the result by the MD-MVO method.

In the current study, we did not evaluate the conformer search efficiency of the MD-MVO method. The previous study revealed that the conformer search by the MD simulation has an advantage comparing to the usual conformer search method like a random search method [25]. The conformer search by the MD simulation is generally time-consuming for compounds with small mass weights. But, for peptides, the conformer search by the MD simulation could be more effective than the usual conformer search method, since the atomic conflict (van der Waals contact) could occur frequently by the random rotation of the chemical bonds. Thus, the MD-MVO method could be useful when the active compound (template) is a peptide.

The generalized ensemble method could search entire conformer space, but it is very time-consuming, since it constructs the canonical ensemble of the whole system. The generalized ensemble method takes several hours to several weeks to examine the conformer of peptides. On the contrary, the MD-MVO method takes several minutes to examine the conformer of peptides in the current study. This is an advantage of the MD-MVO method against the generalized ensemble method.

The MD-MVO method with a target protein structure could be more precise than the MD-MVO method without a target protein structure. However, at the moment, the MD-MVO method with a target protein structure is not very suitable for practical in-silico screening, since this method is much slower than the MD-MVO method without a target protein structure. The CPU time of the MD-MVO method with a target protein structure for 1 step was 0.25 s and the total CPU time for 1 initial coordinates was 4000 s. On the contrary, the total CPU time for one initial coordinates was 60 s for the MD-MVO method without a target protein structure.

5. Conclusion

We developed a new molecular dynamics simulation method for molecular overlapping (alignment) and ligand-based in-silico drug screening based on molecular similarity. The new method, named the MD-MVO method, performs a molecular overlapping of two or more compounds by using the molecular dynamics simulation with a newly introduced potential function. The newly introduced intermolecular interaction between compounds is proportional to the molecular overlap instead of the van der Waals and Coulomb interactions between atoms of different molecules. This method performs both conformer generation of molecules and molecular overlapping at the same time. After an energy minimization and following brief MD simulation at high temperature, the molecules were overlapped with each other and the similarity between compounds was measured by the overlapped volume.

We applied the MD-MVO method to the ligand-based in-silico drug screenings of five targets. The compound library consisted of several active compounds of each target and about 10^4 decoy compounds. One of the known active compounds was selected for each target as a template compound, and then the other compounds were overlapped to the template compound one by one. The MD-MVO method worked well for most of the targets. The average hit ratio at 1% compounds selected from the compound library was 27%, and the average area under the database enrichment curve (q) was 68–76%. Especially, when the coordinates of the query compounds were fixed to the true binding position of the protein–ligand complex structures, the hit ratio by the MD-MVO method was much higher than the hit ratio by the DSI method. When the coordinates of the all compounds were free, the MD-MVO method worked well. Namely, for the intrinsic μ opioid peptide endomorphin the MD-MVO method could show a q value of 90.4% and an enrichment factor of 9.1. The MD-MVO method could be efficient when only one active compound is known or the active compound is a peptide. In other cases, the other similarity search method could be more efficient than the MD-MVO method.

Acknowledgements

This work was supported by grants from the New Energy and Industrial Technology Development Organization of Japan (NEDO) and the Ministry of Economy, Trade, and Industry (METI) of Japan.

Appendix A

COX2 active compounds: Sc-558 (1-phenylsulfonamide-3-trifluoromethyl-5-parabromophenylpyrazole), diclofenac, indomethacin, arachidonic acid, diflunisal, etodolac, ketoprofen, naproxen, nimesulide, prostaglandin H2, piroxicam, rofecoxib, sulindac, and suprofen.

HIV protease-1 active compounds: Ligand of 1aid, ligand of 1ajv, ligand of 1bv7, ligand of 1dif, ligand of 1hpx, ligand of 1hte, ligand of 1htf, ligand of 1htg, ligand of 1hvi, ligand of 1hvr, ligand of 1hxb, ligand of 1hxw, ligand of 1ivp, ligand of 1jld, ligand of 1mes, ligand of 1odw, ligand of 1pro, ligand of 2upj, ligand of 4phv.

GST active compounds: Ligand of 10gs, ligand of 11gs, ligand of 12gs, ligand of 18gs, ligand of 1aqv, ligand of 1aqx, ligand of 1pgt, ligand of 20gs, ligand of 21gs, ligand of 2gss, ligand of 2pgt, and ligand of 3pgt.

Thermolysin active compounds: Aspartic acid, aspartame, phenyl alanine, ligand of 1hyt, ligand of 1os0, ligand of 1pe5, ligand of 1pe7, ligand of 1pe8, ligand of 1qf0, ligand of 1qf1, ligand of 1qf2, ligand of 1thl, ligand of 1z9g, ligand of 1zdp, ligand of 4tln, ligand of 4tmn, ligand of 5tln, ligand of 5tmn, ligand of 6tmn, ligand of 7tln, ligand of benzyloxycarbonyl-D-Ala, ligand of benzyloxycarbonyl-D-Asp, ligand of benzyloxycarbonyl-D-Glu, ligand of benzyloxycarbonyl-D-Thr, ligand of benzyloxycarbonyl-L-Ala, ligand of benzyloxycarbonyl-L-Asp, ligand of benzyloxycarbonyl-L-Glu, and ligand of benzyloxycarbonyl-L-Thr.

μ-opioids: Endomorphin, alfentanil, fentanyl, hydromorphone, levorphanol, loperamide, methadone, morphine, pethidine, phenazocine, propoxyphene, and sufentanyl.

Appendix B

Molecule **a**: c12n(C(=O)c3ccc(C)cc3)c(C)c(CC(O)=O)c1cc(cc2)OC, molecule **b**: c1(n2c3c(c(NCc4ccccc4)ncn3)c(C)n2)ccccc1.

References

[1] H. van de Waterbeemd, B. Testa, G. Folkers, Computer-Assisted Lead Finding and Optimization—Current Tools for Medicinal Chemistry, Wiley–VCH, Weinheim, 1997.
[2] A.R. Leach, Molecular Modeling—Principles and Applications, second edition, Pearson Education Limited, Edinburgh Gate, England, 2001, pp. 668–687.
[3] W.G. Richards, D.D. Robinson, in: D.G. Truhlar, W.J. Howe, A.J. Hopfinger, J. Blaney, R.A. Dammkoehler (Eds.), Rational Drug Design, Springer-Verlag, New York, 1999, pp. 39–49.
[4] J. Gasteiger, T. Engel, Chemoinformatics: A Textbook, Wiley–VCH, Weinheim, 2003.
[5] H. Eckert, J. Bajorath, Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches, Drug Discov. Today 12 (2007) 225–233.
[6] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.

[7] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on biding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[8] A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, C. Duraiswami, Construction of 3D-QSAR models using the 4D-QSAR analysis formalism, J. Am. Chem. Soc. 119 (1997) 10509–10524.
[9] A.J. Hopfinger, A. Reaka, P. Venkatarangan, J.S. Duca, S. Wang, Construction of a virtual high throughput screen by 4D-QSAR analysis: application to a combinatorial library of glucose inhibitors of glycogen phosphorylase *b*, J. Chem. Inf. Comput. Sci. 39 (1999) 1151–1160.
[10] A.J. Hopfinger, A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis, J. Am. Chem. Soc. 102 (1980) 7196–7206.
[11] A.J. Hopfinger, Inhibition of dihydrofolate reductase: structure–activity correlations of 2,4-diamino-5-benzylpyrimidines based upon molecular shape analysis, J. Med. Chem. 24 (1981) 818–822.
[12] G. Schneider, W. Neidhart, T. Giller, G. Schmid, Scaffold-hopping by topological pharmacophore search: a contribution to virtual screening, Angew. Chem. Int. Ed. 38 (1999) 2894–2896.
[13] R.S. Pearlman, K.M. Smith, Metric validation and the receptor-relevant subspace concept, J. Chem. Inf. Comput. Sci. 39 (1999) 28.
[14] S. Pickett, in: H.J. Boehm, G. Schneider, R. Mannhold, H. Kubinyi, G. Folkers (Eds.), Protein–ligand Interactions from Molecular Recognition to Drug Design—Methods and Principles in Medicinal Chemistry, Wiley–VCH, Weinheim, 2003, pp. 88–91.
[15] D.D. Frantz, D.L. Freeman, J.D. Doll, Reducing quasi-ergodic behavior in Monte Carlo simulations by *J*-walking: applications to atomic clusters, J. Chem. Phys. 93 (1990) 2769–2784.
[16] K. Hukushima, K. Nemoto, Exchange Monte Carlo Method and application to spin glass simulations, J. Phys. Soc. Jpn. 65 (1996) 1604–1608.
[17] J.P. Valleau, Density-scaling Monte Carlo study of subcritical Lennard-Jonesium, J. Chem. Phys. 99 (1993) 4718–4728.
[18] B.A. Berg, T. Neuhaus, Multicanonical algorithms for first order phase transitions, Phys. Lett. B 267 (1991) 249–253.
[19] J. Lee, New Monte Carlo algorithm: entropic sampling, Phys. Rev. Lett. 71 (1993) 211–214.
[20] A.P. Lyubartsev, A.A. Marsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, New approach to Monte Carlo calculation of the free energy: method of expanded ensembles, J. Chem. Phys. 96 (1992) 1776–1783.
[21] F. Wang, D.P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states, Phys. Rev. Lett. 86 (2001) 2050–2053.
[22] B.A. Berg, T. Celik, New approach to spin-glass simulations, Phys. Rev. Lett. 69 (1992) 2292–2295.
[23] N. Nakajima, H. Nakamura, A. Kidera, Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides, J. Phys. Chem. B 101 (1997) 817–824.
[24] J.G. Kim, Y. Fukunishi, H. Nakamura, Multicanonical molecular dynamics algorithm employing adaptive force-biased iteration scheme, Phys. Rev. E 70 (2004) 057103.
[25] H. Tsujishita, S. Hirono, CAMDAS: an automated conformational analysis system using molecular dynamics, J. Comput. Aided Mol. Des. 11 (1997) 305–315.
[26] J.A. Grant, M.A. Gallardo, B.T. Pickup, A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape, J. Comput. Chem. 17 (1996) 1653–1666.
[27] T.S. Rush III, J.A. Grant, L. Mosyak, A. Nicholls, A shape-based 3-D scaffold hopping method and its application to a bacterial protein–protein interaction, J. Med. Chem. 48 (2005) 1489–1495.
[28] P.C.D. Hawkins, A.G. Skillman, A. Nicholls, Comparison of shape-matching and docking as virtual screening tools, J. Med. Chem. 50 (2007) 74–82.
[29] Y. Fukunishi, H. Nakamura, Prediction of protein–ligand complex by docking software guided by other complex structures, J. Mol. Graph. Model. 26 (2008) 1030–1033.
[30] D. Sugimoto, Y. Chikada, J. Makino, T. Ito, T. Ebisuzaki, M. Umemura, A special-purpose computer for gravitational many-body problems, Nature 345 (1990) 33–35.
[31] G. Kikugawa, R. Apostolov, N. Kamiya, M. Taiji, R. Himeno, H. Nakamura, Y. Yonezawa, Application of MDGRAPE-3, a special purpose board for molecular dynamics simulations, to periodic biomolecular systems, J. Comput. Chem., in press.
[32] Y. Fukunishi, Y. Mikami, K. Takedomi, M. Yamanouchi, H. Shima, H. Nakamura, Classification of chemical compounds by protein–compound docking for use in designing a focused library, J. Med. Chem. 49 (2006) 523–533.
[33] Y. Fukunishi, S. Hojo, H. Nakamura, An efficient in silico screening method based on the protein–compound affinity matrix and its application to the design of a focused library for cytochrome P450 (CYP) ligands, J. Chem. Info. Mod. 46 (2006) 2610–2622.
[34] H. Briem, I.D. Kuntz, Molecular similarity based on DOCK-generated fingerprints, J. Med. Chem. 39 (1996) 3401–3408.
[35] U.F. Lessel, H. Briem, Flexsim-X: a method for the detection of molecules with similar biological activity, J. Chem. Inf. Comput. Sci. 40 (2000) 246–253.
[36] H. Briem, U.F. Lessel, In vitro and in silico affinity fingerprints: finding similarities beyond structural classes, Perspect. Drug Discov. Des. 20 (2000) 231–244.
[37] Y. Fukunishi, S. Kubota, H. Nakamura, Noise reduction method for molecular interaction energy: application to in silico drug screening and in silico target protein screening, J. Chem. Info. Mod. 46 (2006) 2071–2084.
[38] Y. Fukunishi, Y. Mikami, H. Nakamura, The filling potential method: a method for estimating the free energy surface for protein–ligand docking, J. Phys. Chem. B 107 (2003) 13201–13210.
[39] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, J. Med. Chem. 49 (2006) 6789–6801.
[40] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[41] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.
[42] J. Gasteiger, M. Marsili, A new model for calculating atomic charges in molecules, Tetrahedron Lett. (1978) 3181–3184.
[43] J.E. Zadina, L. Hackler, L. Ge, A. Kastin, A potent and selective endogenous agonist for the m-opiate receptor, Nature 386 (1997) 499–502.
[44] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring function, J. Med. Chem. 49 (2006) 5912–5931.
[45] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comp. Phys. 23 (1997) 327–341.
[46] T. Terada, A. Kidera, Generalized form of the conserved quantity in constant-temperature molecular dynamics, J. Chem. Phys. 116 (2002) 33–41.
[47] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, P.A. Kollman, AMBER 8, UCSF, 2004.
[48] Y. Fukunishi, Y. Mikami, H. Nakamura, Similarities among receptor pockets and among compounds: analysis and application to *in silico* ligand screening, J. Mol. Graph. Model. 24 (2005) 34–45.
[49] C.L. Brooks III, M. Karplus, B.M. Pettitt, Proteinsm. A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, John Wiley and Sons, New York, 1988.
[50] A.R. Ortiz, M.T. Pisabarro, F. Gago, R.C. Wade, Prediction of drug binding affinities by comparative binding energy analysis, J. Med. Chem. 38 (1995) 2681–2691.
[51] K. Omagari, D. Mitomo, S. Kubota, H. Nakamura, Y. Fukunishi, A method to enhance the hit ratio by a combination of structure-based drug screening and ligand-based screening, Computational Biol. Chem.: Adv. Appl. 1 (2008) 19–28.