Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 61–70
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Multiple target screening method for robust and accurate in silico ligand screening

Yoshifumi Fukunishi [a,*], Yoshiaki Mikami [b], Satoru Kubota [b], Haruki Nakamura [a,c]

[a] *Biological Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan*

[b] *Japan Biological Information Research Center (JBIRC), Japan Biological Informatics Consortium (JBIC), 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan*

[c] *Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*

Received 18 July 2005; received in revised form 18 November 2005; accepted 18 November 2005
Available online 22 December 2005

Abstract

We developed a new in silico multiple target screening (MTS) method, based on a multi-receptor versus multi-ligand docking affinity matrixes, and examined its robustness against changes in the scoring system. According to this method, compounds in a database are docked to multiple proteins. The compounds among these proteins that are likely bind to the target protein are selected as the members of the candidate-hit compound group. Then, the compounds in the group are sorted into descending order using the docking score: the first (*n*-th) compound is expected to be the most (*n*-th) probable hit compound. This method was applied to the analysis of a set of 142 receptors and 142 compounds using a receptor-ligand docking program, Sievgene [Y. Fukunishi, Y. Mikami, H. Nakamura, Similarities among receptor pockets and among compounds: analysis and application to in silico ligand screening, J. Mol. Graphics Modelling, 24 (2005) 34–45], and the results demonstrated that this method achieves a high hit ratio compared to uniform sampling. We prepared two new scores: the ΔG score, designed to reproduce the protein-ligand binding free energy, and the hit-optimized score, designed to maximize the hit ratio of in silico screening. Using the Sievgene docking score, ΔG score and hit-optimized score, the MTS method is more robust than the multiple active-site correction scoring method [G.P.A. Vigers, J.P. Rizzi, Multiple active site corrections for docking and virtual screening, J. Med. Chem., 47 (2004) 80–89].

Keywords: Receptor-ligand docking; Flexible docking; Docking score; Binding free energy; Scoring system; Database enrichment

1. Introduction

Protein-ligand docking is a key technology for in silico screening, and many protein-ligand docking programs have been reported [1,3–13]. One of the serious problems of in silico screening is the low accuracy of the estimation of the protein-compound binding free energy. This problem is so difficult to solve that in many cases, several scoring functions have been developed independently. One is a docking score to optimize the protein-ligand complexed structure, and another is a score to evaluate the complex structure. One of the later scores is a Gibbs free energy difference (ΔG) score, to estimate the protein-compound binding free energy after the docking simulation using the docking score. Usually, ΔG scores have been used to reproduce the binding free energy of the complexes of proteins and their true binders [7,13]. The other scoring function is the knowledge-based energy function [14,15]. In this approach, the potential of mean force is estimated from the spatial distributions of atom pairs or functional-group pairs of the many complex structures. The knowledge-based energy function can give rather accurate binding free energy than the physical energy functions. Since the ΔG score is not designed to reproduce the ΔGs of inactive compounds but rather, only those of active compounds, the ΔG score does not guarantee a higher hit ratio in the in silico screening than the docking score.

The other approach to solving the problem of the low accuracy of binding free energy is to use the protein-compound affinity matrix. The multiple active site correction (MASC) score S_{ik}' for the *i*-th pocket and the *k*-th compound has been reported by Vigers and Rizzi as follows [2]:

$$S'_{ik} = \frac{S_{ik} - \mu_k}{\sigma_k},$$

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

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.11.006

where S_{ik} is the raw docking score for the i-th pocket and the k-th compound, and μ_k and σ_k are the average and standard deviation of the raw docking scores across all pockets for the k-th compound, respectively. In this method, S_{ik}' is used for screening instead of S_{ik}. The MASC scoring method achieved high database enrichment using DOCK [3], FlexX [4] and GOLD [5] software. We developed a modified version of the MASC scoring method, the receptor selection (RS) method [1]. This study revealed that the screening result strongly depends on the choice of the proteins and the desirable choice of the protein is to choose the similar proteins to the target protein. In the RS method, the similarity of the proteins to the target protein is taken into account, and the database enrichment by the RS method is more robust and better than the original MASC scoring method. However, a critical problem in those methods is that even if the docking score corresponds exactly to the experimental binding free energy, the methods do not always succeed in finding the true binder. The reason is that the true binder has the strongest S_{ik}, but it does not necessarily have the strongest S_{ik}', depending on the choice of the proteins.

In this study, we proposed a new screening method, named the Multiple target screening (MTS) method, based on the protein-compound affinity matrix, and this method can find the true binder if the binding free energy was precisely estimated. Then, we tested the robustness of our screening method using three scores: the docking score, the score to estimate the binding free energy, and the hit-optimized score, which is optimized to maximize the hit ratio of the screening.

2. Methods

2.1. Multiple target screening (MTS) method

We prepare a set of protein pockets $P = \{p_1, p_2, p_3, \cdots p_M\}$, where p_i represents the i-th pocket. The total number of pockets is M. We also prepare a set of compounds $X = \{x^1, x^2, \cdots x^N\}$, where x^k represents the k-th compound. The total number of compounds is N. For each pocket p_i, all compounds of the set X are docked to the pocket p_i with score s_i^k between the i-th pocket and the k-th compound. Here, s_i^k corresponds to the binding free energy, and a lower s_i^k means a higher affinity between the i-th pocket and the k-th compound.

For the k-th compound, $\{s_i^k;\ i = 1, \cdots M\}$ are sorted in descending order, and the order n_i^k is assigned to each i-th pocket depending on its value s_i^k. For example, when $n_i^k = 1$, the i-th pocket binds the k-th compound with the strongest affinity. When $n_i^k = M$, the i-th pocket binds with the weakest affinity. This procedure is repeated until the orders $\{n_i^k;\ i = 1, \cdots, M \mid k = 1, \cdots, N\}$ are determined for all the compounds.

Next, we focused on the target k-th pocket. The compounds having the order $n_i^k = 1$ are assigned as members in the compound group-1. Among the group-1 members, the compound with the lowest s_i^k should be the most probable hit compound. If there is no member in group-1, the compounds having $n_i^k = 2$ are assigned as the members in compound group-2. Among the group-2 members, the compound with the lowest s_i^k should be the most probable hit compound. This procedure is repeated until the most probable hit compound is found.

In the conventional screening method with one target protein, the scores s_i^{k0} with a constant $k0$ are simply sorted in descending order. However, the conventional scores tend to be lower when the interacting interfaces are wider, even though the actual affinities are not very high; this is particularly true for larger compounds. On the contrary, in the current MTS method, we first determine which protein pocket is the strongest binder for each compound, and the compounds that are likely bind the target protein are selected. This procedure assures the identification of probable hit compounds with fewer errors than the conventional method, regardless of the accuracy of the estimated scores.

In the MASC scoring method where the scores are normalized by various compounds and proteins, the true binder would be missed in some cases. In fact, when the binding free energies of an inactive compound (the k-th compound) for many proteins were early zero, the MASC score for the target protein i, S_{ik}' could be a large number when the deviation of the scores is a small number. When the deviation of the scores of the true binder is large, the MASC score of the true binder could be a small number. For these cases, the true binder should be found in the MTS method.

2.2. Scoring function

All docking simulations were performed by the in-house docking software Sievgene. The program used the docking algorithm in combination with rigid-protein and flexible-ligand techniques. Details of the docking algorithm and the computation is described in detail elsewhere [1]. This docking program was developed with a performance yielding about 50% of the reconstructed complexes at a distance of less than 2 Å RMSD for the 132 complexed receptors with the compounds in PDB.

The receptor-ligand interactions accounted for by Sievgene include van der Waals (vdW), Coulomb, hydrogen bond, and hydrophobic interactions. A grid potential represents these interactions, assuming that the receptor structure is rigid. The details of these energy terms were reported by our previous study.

The space is divided into two region, one region is the inside region of the protein and the other is the outside region of the protein. The two regions are divided by the accessible surface of the protein. In the outside region of the protein, the energy terms are calculated as follows. The van der Waals energy is calculated by the AMBER force field. The Coulomb interaction is calculated by the usual manner with the distance dependent dielectric constant, where $\varepsilon = 4R$ (R is the inter atomic distance). The hydrogen bond interaction is estimated by a Gaussian type attractive potential. The hydrophobic interaction is estimated by the accessible surface area method with the atomic solvation parameter = 10 cal/mol/Å^2, which is close to the values 5.4 cal/mol/Å^2 [16] and 8 cal/mol/Å^2 [17]. In the inside region of the protein, the energy surfaces of the van der Waals and the Coulomb interactions are exchanged by smooth

functions, which are functions of the distance to the nearest point of the accessible surface.

We introduced two new scores: the ΔG score, designed to reproduce the protein-ligand binding free energy, and the hit-optimized score, designed to maximize the hit ratio of in silico screening. The docking simulation was performed by using the Sievgene docking score; then the new scores were calculated by fixing the protein-compound complexed structure. The new score to estimate the protein-ligand binding free energy was determined as

$$S_{\Delta G} = c_{\mathrm{rot}}N_{\mathrm{rot}} + c_{\mathrm{AV}}(E_{\mathrm{ASA}} + E_{\mathrm{vdW}}) + c_{\mathrm{ele}}E_{\mathrm{ele}} + c_{\mathrm{hyd}}E_{\mathrm{hyd}} + c_{\mathrm{intra-vdW}}E_{\mathrm{intra-vdW}},$$

where N_{rot}, E_{ASA}, E_{vdW}, E_{ele}, E_{hyd} and $E_{\mathrm{intra\text{-}vdW}}$ represent the number of rotatable bonds, the hydrophobic energy due to the accessible surface area, the vdW energy, the protein-ligand Coulomb potential, the hydrogen bond energy and the intra molecule vdW energy of the ligand. Also, c_{rot}, c_{AV}, c_{ele}, c_{hyd} and $c_{\mathrm{intra\text{-}vdW}}$ are the optimized coefficients for each energy term. For each atom type, the sum of E_{ASA} and E_{vdW} gives one grid potential, and both energy terms are always simultaneously calculated. Thus, these two terms share the same coefficient, c_{AV}. Only the differences among the original Sievgene docking score, the ΔG score and the hit-optimized score are the differences of the coefficients c_{rot}, c_{AV}, c_{ele}, c_{hyd} and $c_{\mathrm{intra\text{-}vdW}}$.

Sievgene utilizes the grid potential to calculate each energy term except the intra-molecule interaction. In this study, the mesh sizes of $100 \times 100 \times 100$ and $60 \times 60 \times 60$ were adopted.

3. Preparation of materials

To determine the coefficients for the ΔG score, we performed a protein-ligand docking simulation based on the known complex structures registered in the Protein Data Bank. Here, 50 complexes accompanied with the experimental binding free energy values were selected from the database that was used in the determination of ΔG score of the PRO_LEADS [7]. The PDB identifiers are summarized in Appendix A. Furthermore, these 50 proteins were divided into two groups, protein sets A and B. The whole 50 proteins were put in alphabetical order, and then the protein set A consists of odd numbered proteins and the protein set B consists of even numbered proteins. Both protein sets A and B consist of 25 proteins. The PDB identifiers of protein sets A and B are summarized in the Appendix A. All water molecules were removed from the proteins and all missing hydrogen atoms were added to form all-atom models of the proteins. The initial conformations of all ligands were set to the crystal structures, and the atomic coordinates of the ligands were optimized by the steepest descent method on the potential surface by the Sievgene docking score.

To evaluate our docking method, we performed a protein-ligand docking simulation based on the known complex structures registered in the Protein Data Bank. Here, 142 complexes were selected from the database used in the evaluation of the GOLD and FlexX [18]. This data set contains a rich variety of proteins and compounds whose structures were all determined by high quality experiments with a resolution of less than 2.5 Å. Almost all the atom coordinates are supplied, and the all-atomic structures around the ligand pockets are quite reliable. Thus, this data set was used in the docking study and in silico screening. We removed from the original data set those complexes containing a covalent bond between the protein and ligand. The PDB identifiers are summarized in the Appendix B. All water molecules and cofactors were removed from the proteins, and all missing hydrogen atoms were added to form the all-atom models of proteins. The conformations of all ligands were randomized before the docking simulation. Furthermore, these 142 proteins were divided into two groups, protein sets C and D. The whole 142 proteins were put in alphabetical order, and then the protein set C consists of odd numbered proteins and the protein set D consists of even numbered proteins. Both protein sets C and D consist of 71 proteins. The PDB identifiers of protein sets C and D are summarized in the Appendix B.

The size distribution of ligands was as follows: 1–9 atoms, 3.6%; 10–19 atoms, 15.2%; 20–29 atoms, 30.9%; 30–39 atoms, 15.4%; 40–49 atoms, 15.8%; 50–59 atoms, 10.6%; and more than 60 atoms, 16.1%. The average ligand size was 37.1 atoms.

For all protein-ligand complexes summarized in the Appendices A and B, the atomic charges of each ligand were determined by the restricted electrostatic point charge (RESP) procedure using HF/6-31G*-level quantum chemical calculations [19]. We used GAMESS and Gaussian98 to perform the quantum chemical calculations [20,21]. The atomic charges of the proteins were the same as the atomic charges in AMBER parm99 [19,22].

4. Results

4.1. Comparison of MTS, MASC scoring and raw scoring screening results when the Sievgene docking score is adopted

Fig. 1 shows database enrichments with the MTS, MASC scoring and raw docking scoring methods applied to 142 receptors versus 142 compounds. The mesh size of the grid potential was set to $100 \times 100 \times 100$. Usually, the in silico screening method is used to select only a small number of compounds from a large number of compounds in a database, so that the number of hits obtained among the first small number of compounds of the database is crucial. Thus, in this study we have addressed only the number of hits found among the first 5% and 10% of the entries in the database. As shown in Fig. 1, the conventional raw scoring method yielded a slight enrichment, with 16.2% and 26.8% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method yielded a 3.5–5.2-fold enrichment, with 25.8% and 34.8% of the ligands found among the first 5% and 10% of the database, respectively. Here the enrichment is the ratio of percent correct hits and the top 5% or 10% of the best scoring compounds. The MASC scoring method gave the best

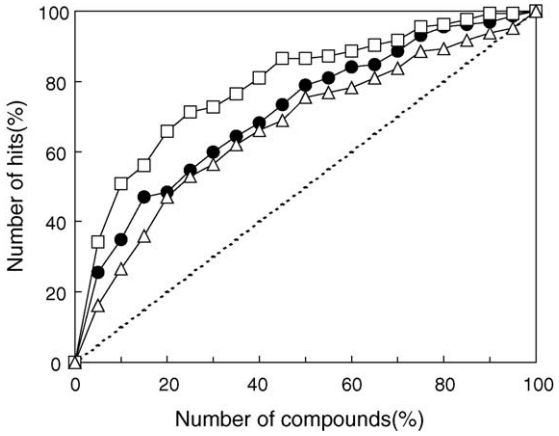

Fig. 1. Database enrichment results with the Sievgene docking score using the potential grid of $100 \times 100 \times 100$. The dashed line, open triangles, open squares and filled circles represent the results obtained with uniform sampling and the conventional raw scoring, MASC scoring, and MTS methods, respectively. The numbers of compounds and hits were scaled to %.

enrichment among these three methods, yielding a 5.1–6.8-fold enrichment, with 34.1% and 50.8% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1.

In order to examine the robustness of the three methods: the MTS, the MASC scoring and the raw scoring methods, we further prepared two docking affinity matrixes. One matrix is the 71 proteins of the protein set C versus the whole 142 compounds, and another matrix is the 71 proteins of the protein set D versus the whole 142 compounds. The database enrichments by the MTS, the MASC scoring and the raw scoring methods were calculated for the both matrixes. For the protein set C, the conventional raw scoring method yielded a slight enrichment, with 19.7% and 28.2% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method gave the better enrichment than the raw scoring method, with 27.5% and 33.8% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method yielded a best enrichment among these methods, with 30.3% and 40.1% of the ligands found among the first 5% and 10% of the database, respectively. For the protein set D, the conventional raw scoring method yielded a slight enrichment, with 19.7% and 28.2% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method gave the better enrichment than the raw scoring method, with 26.8% and 33.8% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method yielded a best enrichment among these methods, with 31.0% and 40.1% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1.

When the mesh size of the grid potential is set to $60 \times 60 \times 60$, the results change drastically as shown in Fig. 2. The conventional raw scoring method yielded a slight enrichment, with 19.7% and 35.9% of the ligands found among the first 5% and 10% of the database, respectively. The result of

Table 1
Database enrichments by various scoring functions and methods

Scoring function	Method	Grid size	Database enrichment at 5% compound (%)	Database enrichment at 10% compound (%)
Sievgene score	MTS	100	25.80	34.80
	MASC	100	34.10	50.80
	Raw	100	16.20	26.80
	MTS + MASC	100	56.34	65.49
Sievgene score	MTS	60	28.90	41.50
	MASC	60	25.40	30.30
	Raw	60	19.70	35.90
	MTS + MASC	60	53.52	61.27
ΔG score	MTS	100	26.10	32.40
	MASC	100	20.40	28.20
	Raw	100	15.50	23.20
Hit-optimized score α	MTS	100	23.90	38.70
	MASC	100	26.70	31.00
	Raw	100	21.80	33.10
Hit-optimized score β	MTS	100	28.20	40.10
	MASC	100	26.80	33.10
	Raw	100	19.00	35.90
Sievgene score[a]	MTS	100	30.28	40.14
	MASC	100	27.46	33.80
	Raw	100	19.72	28.17
Sievgene score[b]	MTS	100	30.99	40.14
	MASC	100	26.76	33.80
	Raw	100	19.71	28.17

[a] The original Sievgene score was applied to the protein data set C.
[b] The original Sievgene score was applied to the protein data set D.

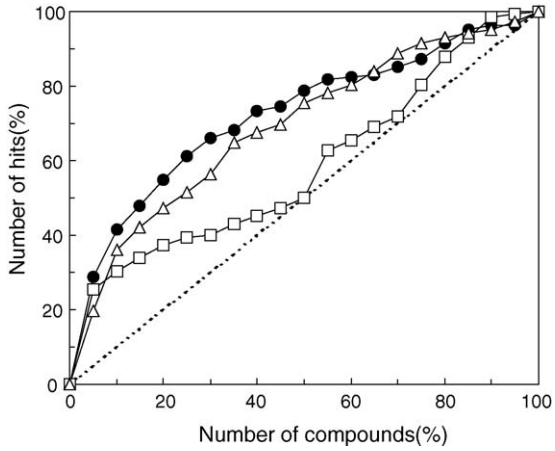

Fig. 2. Database enrichment results with the Sievgene docking score using the potential grid of 60 × 60 × 60. The dashed line, open triangles, open squares and filled circles represent the results obtained with uniform sampling and the conventional raw scoring, MASC scoring, and the MTS methods, respectively. The numbers of compounds and hits were scaled to %.

the MASC scoring method is drastically decreased compared to the result in Fig. 1. The MASC scoring method yielded a 3.0–5.1-fold enrichment, with 25.4% and 30.3% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method gave the best enrichment among these three methods, yielding a 4.2–5.8-fold enrichment, with 28.9% and 41.5% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1.

The enrichment with the grid of 100 × 100 × 100 is better than that with the grid of 60 × 60 × 60, since the larger the dimension of the grid is, the more precise the estimation of the potential is. The enrichments by the MTS and the raw scoring methods do not depend on the grid size so much while that by the MASC scoring method depends heavily on grid size.

These results do not mean that the MTS method is more useful than the MASC scoring method in any case. We can use the both methods to get better result than that by the MTS or the MASC scoring method. We selected 7 compounds (=5% of the total 142 compounds) as follows; 4 compounds by the MTS method and 3 compounds by the MASC scoring method for each protein, and then the database enrichment was calculated. The united method yielded 10.7–11.3-fold enrichment with 53.5% and 56.3% ligand found among the first 5% compounds of the database with 60 × 60 × 60 and 100 × 100 × grid, respectively. When we selected 14 compounds (=10% of the total 142 compounds) for each protein, namely the half was selected by the MTS and the other half was selected by the MASC scoring method, the united method yielded 6.1–6.6-fold enrichment with 61.3% and 65.5% ligand found among the first 10% compounds of the database with 60 × 60 × 60 and 100 × 100 × 100 grid, respectively.

4.2. ΔG score

Starting from the 50 complexed structures, determined by X-ray crystallography, of Appendix A, each ligand structure was optimized using the Sievgene docking score by the steepest descent energy minimization procedure. The maximum value, the minimum value and the average value of the RMSD were 1.01, 0.17 and 0.45 Å, respectively. The RMAD was calculated for all atoms of the ligand and the protein structure was fixed to the X-ray crystallography structure. Then, each energy term was optimized to reproduce the experimental protein-ligand binding free energy by the least square method. The resulting coefficients are summarized in Table 2. The coefficient for intra-molecule vdW interaction was set to zero, since $c_{\text{intra-vdW}}$ became negative when all coefficients were optimized. If $c_{\text{intra-vdW}}$ became negative, the vdW conflict became preferable, and such interaction is not reasonable at all.

Table 2
Coefficients of various scoring functions

	c_{rot}	c_{AV}	c_{ele}	c_{hyd}	$c_{\text{intra-vdW}}$
Sievgene docking score	0.0000	0.0100	0.0100	0.0100	0.0100
ΔG score	0.1351	0.0260	0.0249	0.0409	0.0000
ΔG score[a]	0.2033	0.0319	0.0327	0.0342	0.0000
ΔG score[b]	0.1501	0.0229	0.0116	0.0462	0.0000
Screening score A	0.1351	0.3538	1.8917	0.0038	0.5163
Screening score B	0.1351	0.5516	1.4777	0.1267	0.4106
Screening score[c]	0.1351	0.4790	0.9849	0.0239	0.5032
Screening score[d]	0.1351	0.7600	1.5771	0.0259	0.5737

[a] The ΔG score derived from the protein set A.
[b] The ΔG score derived from the protein set B.
[c] The hit-optimized score derived from the protein set C.
[d] The hit-optimized score derived from the protein set D.

The fitting results are shown in Fig. 3. The average error is 2.47 kcal/mol and the correlation coefficient between the experimental data and the estimated values is 0.66. The ΔG score underestimates the ΔG values of the carboxypeptidase A-phosphonate complex (PDB id: 7CPA) and biotin-streptavidin complex (PDB id: 1STP) by 4 and 10 kcal/mol, respectively. The reasons for this underestimation are unclear. The average error and the correlation coefficient of PRO_LEADS were 2.5 kcal/mol and 0.62, respectively [7], and the correlation coefficient of LigScore was 0.75, making the accuracy of our

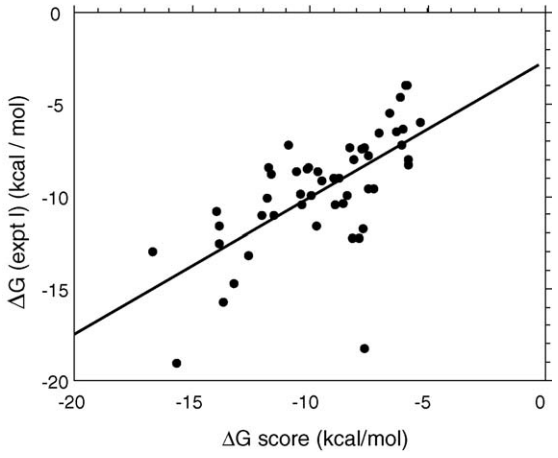

Fig. 3. Correlation between the calculated binding free energy by the ΔG score and the experimental binding free energy. The solid line represents the linear fit by the least square error method. The experimental errors are unknown.

ΔG score comparable to those in previously reported methods [13].

To evaluate the reliability of the ΔG score, we determined the ΔG score for a fraction of the protein set and the ΔG score was applied to the other fraction of the protein set. The ΔG score was determined by using the protein set A. The resulting coefficients are summarized in Table 2. The average error and the correlation coefficient were 2.78 kcal/mol and 0.55, respectively. This ΔG score was applied to the protein set B. The average error and the correlation coefficient were 2.64 kcal/mol and 0.88, respectively. Also, the ΔG score was determined by using the protein set B. The resulting coefficients are summarized in Table 2. The average error and the correlation coefficient were 1.37 kcal/mol and 0.90, respectively. This ΔG score was applied to the protein set A. The average error and the correlation coefficient were 3.39 kcal/mol and 0.47, respectively.

This test suggested that the ΔG score can be applied to the general proteins, while the number of data is only 25 for the protein sets A and B.

4.3. Comparison of MTS, MASC scoring and raw scoring screening results when the ΔG score is adopted

Fig. 4 shows database enrichments with the MTS, MASC scoring and raw scoring methods applied to 142 receptors versus 142 compounds using the ΔG score. The raw scoring method is the conventional single target screening method. Each method was applied as in Section 4.1, and the ΔG score replaced the docking score. Each protein-compound complex was calculated using the Sievgene docking score; the ΔG score was then applied to the protein-compound complex by fixing the atomic coordinates. The mesh size of the grid potential was set to $100 \times 100 \times 100$. The raw scoring method yielded a slight enrichment, with 15.5% and 23.2% of the ligands found among the first 5% and 10% of the database, respectively. The result of the MASC scoring method for the ΔG score was drastically decreased compared to the result for the docking score shown in Fig. 1. The MASC scoring method yielded a 2.8–4.1-fold enrichment, with 20.4% and 28.2% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method for the ΔG score gave the best enrichment among these three methods, yielding a 3.2–5.2-fold enrichment, with 26.1% and 32.4% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1.

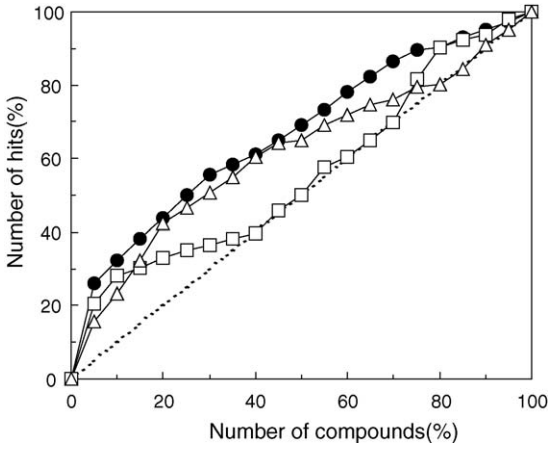

Fig. 4. Database enrichment results with the ΔG score using the potential grid of $100 \times 100 \times 100$. The dashed line, open triangles, open squares and filled circles represent the results obtained with uniform sampling and the conventional raw scoring, MASC scoring, and MTS methods, respectively. The numbers of compounds and hits were scaled to %.

4.4. Hit-optimized score

The ΔG score was introduced to estimate the binding free energy of protein-ligand complexes. The ΔG score works well when the compound is true binder of the protein, but if the compound is not the true binder, the accuracy of the estimated binding free energy is not guaranteed. The most important result of the in silico screening is the hit ratio (database enrichment). Thus, we introduced a hit-optimized score, which maximizes the hit ratio of the in silico screening. In this study, using the X-ray crystal structures of the protein-ligand complexes, the ligand in the complex structure was regarded as a hit compound for each protein.

Using the 142 proteins and their 142 ligands, the database enrichment was calculated by the raw scoring method, which is the conventional screening method. Then, each energy coefficient was optimized to maximize the hit ratio. The parameter optimization was performed by a Monte Carlo method. Each parameter was changed randomly, and then the hit ratio was calculated by the new parameter set. When the hit ratio was improved, the new parameter set was adopted. This procedure was iterated until the hit ratio converged. The hit ratio depends on how many compounds are selected from the database. In this study, 5% or 10% compounds out of the whole 142 compounds were selected to calculate the hit ratio. One parameter set optimizes the hit ratio at the top 5% compounds out of the whole 142 compounds selected by the score with the new parameter set. We call this score the hit-optimized score α. The other parameter set optimizes the hit ratio at the top 10% compounds out of the whole 142 compounds selected by the score with the new parameter set. We call this score the hit-optimized score β. The resulting coefficients were summarized in Table 2. The c_{rot} was set to the same value of the ΔG score, since one coefficient must be fixed to determine the other coefficients. The reason is clear: when the all scores are scaled by the same number, the database enrichment does not change.

Fig. 5 shows database enrichments with the raw docking score, the hit-optimized score α and the hit-optimized score β. The database enrichment with the hit-optimized score is slightly improved at the first 5% and 10% compounds of the database. The raw docking score yielded a slight enrichment, with 16.2% and 26.8% of the ligands found among the first 5% and 10% of the database, respectively. The hit-optimized score α yielded a 3.3–4.4-fold enrichment, with 21.8% and 33.1% of the ligands found among the first 5% and 10% of the database, respectively. The hit-optimized score β yielded a 3.6–3.8-fold

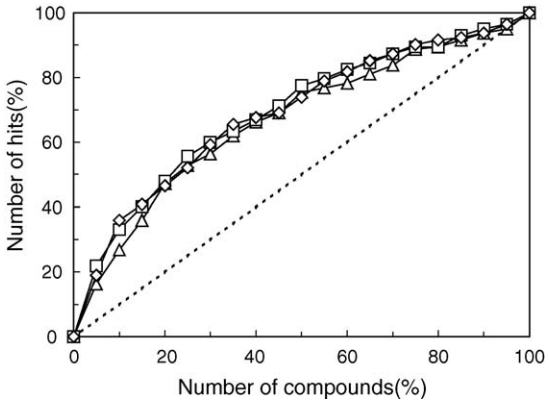

Fig. 5. Database enrichment results with the docking score, the hit-optimized score optimized at 5% compounds, and the hit-optimized score optimized at 10% compounds. The dashed line, open triangles, open diamonds and open squares represent the results obtained with uniform sampling and the conventional raw docking score method with the hit-optimized score set at 5% and 10% compounds, respectively. The numbers of compounds and hits were scaled to %.

enrichment, with 19.1% and 35.9% of the ligands found among the first 5% and 10% of the database, respectively.

The hit-optimized score improved the database enrichment by 5.6–9.2% until 10% of the compounds of the database were selected. When more than 40% of the database was selected, the database enrichments of these three scores were almost equivalent.

4.5. Comparison of MTS, MASC scoring and raw scoring screening results when the hit-optimized score is adopted

Fig. 6 shows database enrichments with the MTS, MASC scoring and conventional raw scoring methods applied to 142 receptors versus 142 compounds. The mesh size of the grid potential was set to 100 × 100 × 100, and the hit-optimized score α was applied. The raw scoring method yielded a slight enrichment, with 21.8% and 33.1% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method yielded a 3.1–5.3-fold enrichment, with 26.7% and 31.0% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method yielded a 3.9–4.8-fold enrichment, with 23.9% and 38.7% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1. The MASC scoring method gave the best enrichment until 5% compounds of the database were selected; beyond that, the MTS method gave the best enrichment among the three methods.

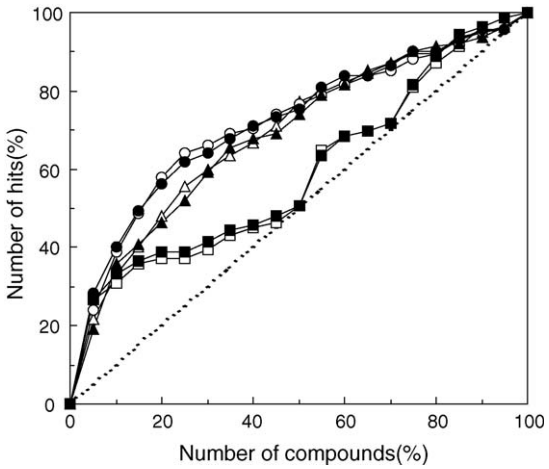

Fig. 6. Database enrichment results with the hit-optimized score using the potential grid of 100 × 100 × 100. The dashed line, open triangles, open squares and open circles represent the results obtained with uniform sampling and the conventional raw scoring, MASC scoring, and MTS methods with the hit-optimized score set at 5% compounds, respectively. The filled triangles, filled squares and filled circles represent the results obtained with uniform sampling and the conventional raw scoring, MASC scoring, and MTS methods with the hit-optimized score set at 10% compounds, respectively. The numbers of compounds and hits were scaled to %.

When the mesh size of the grid potential was set to 100 × 100 × 100, and the hit-optimized score β was applied, the results did not change quantitatively. The raw scoring method yielded a slight enrichment, with 19.0% and 35.9% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method yielded a 3.3–5.4-fold enrichment, with 26.8% and 33.1% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method gave the best enrichment among these three methods, yielding a 4.0–5.6-fold enrichment, with 28.2% and 40.1% of the ligands found among the first 5% and 10% of the database, respectively. These results were summarized in Table 1.

To evaluate the reliability of the hit-optimized score, the hit-optimized score is calculated for a fraction of the protein sets and then the hit-optimized score is applied to the other fraction of the protein sets. The parameters for the hit-optimized score were optimized for the docking affinity matrix of the 71 proteins of the protein data set C versus the whole 142 compounds. The database enrichment at 5% compounds was optimized. The parameters were summarized in Table 2. The raw scoring method with rhe Sievgene docking score yielded a slight enrichment, with 16.2% and 26.1% of the ligands found among the first 5% and 10% of the database, respectively. The raw scoring method with the hit-optimized score yielded a better enrichment than the Sievgene docking score, with 21.1% and 29.6% of the ligands found among the first 5% and 10% of the database, respectively.

The hit-optimized score with this parameter set derived from the protein set C was applied to the docking affinity matrix of the 71 proteins of the protein set D versus the whole 142 compounds. The raw scoring method with the Sievgene docking score showed a database enrichment with 16.2% and 26.1% of the ligands found among the first 5% and 10% of the database, respectively. The raw scoring method with the hit-optimized score showed a better database enrichment than the Sievgene docking score with 19.7% and 28.2% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method showed a database enrichment with 25.4% and 31.6% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method showed the best database enrichment among these three methods with 27.5% and 38.0% of the ligands found among the first 5% and 10% of the database, respectively.

Also, the parameters of the hit optimized score for the docking affinity matrix of the 71 proteins of the protein data set D versus the whole 142 compounds were optimized. The database enrichment at 5% compounds was optimized. The parameters were summarized in Table 2. The raw scoring method with the Sievgene docking score yielded a slight enrichment, with 16.9% and 27.5% of the ligands found among the first 5% and 10% of the database, respectively. The raw scoring method with the hit-optimized score yielded a better enrichment than the Sievgene docking score, with 21.1% and 28.2% of the ligands found among the first 5% and 10% of the database, respectively.

The hit-optimized score with this parameter set derived from the protein set D was applied to the docking affinity matrix of the 71 proteins of the protein set C versus the whole 142 compounds. The raw scoring method with the Sievgene docking score showed a database enrichment with 16.9% and 27.4% of the ligands found among the first 5% and 10% of the database, respectively. The raw scoring method with the hit-optimized score showed a better database enrichment than the Sievgene docking score with 21.1% and 29.6% of the ligands found among the first 5% and 10% of the database, respectively. The MASC scoring method showed a database enrichment with 26.8% and 31.7% of the ligands found among the first 5% and 10% of the database, respectively. The MTS method showed the best database enrichment among these three methods with 28.9% and 39.4% of the ligands found among the first 5% and 10% of the database, respectively.

This test revealed that the hit-optimized score improves the database enrichment by the traditional raw scoring screening method and in these cases the result by the MTS method is better than that by the MASC scoring and the raw scoring method.

5. Discussion

In Figs. 1, 2, 4 and 6, the MTS method is more robust than the MASC scoring method. Even if the score values change using the docking score, ΔG score and hit-optimized score, the database enrichments of the MTS and the conventional raw scoring methods do not change if the order of the scores does not change. On the contrary, the database enrichment of the MASC changes depending on the change of the scores, even if the order of the score does not change. That is why the MTS method is more robust than the MASC scoring method. In some cases, the MASC scoring method can show better database enrichment than the MTS method; for instance, it can be used when many true binders are available, and we can check the database enrichment before the actual large-scale in silico screening.

In this paper, it is shown that the results of the MTS method are better than those by other conventional methods in many cases. In the conventional raw scoring method for one target protein, finding a true compound binder among many compounds directly depends on the precision of the scores. In principle, this becomes very difficult because the error of ΔG estimation is almost as large as the value of ΔG itself.

For a particular protein pocket, there are large variety of interaction schemes with many different compounds, since those compounds consist of a variety of functional groups. On the contrary, for a particular compound, the number of variations of interactions with proteins would be more limited because the varieties of the protein pockets are, at most, the numbers of their functional domains in the protein families or several times of them. Namely, the numbers should be less than 10^5 [23,24]. Thus, finding a true protein binder among many proteins is somewhat easier than finding a true compound binder among many proteins, particularly when the estimation of binding scores has some error. This is why the MTS method is much more robust than the previous methods and can achieve a high enrichment.

In the MTS method, the selective true binder preferentially binds the target protein among the set of proteins. Also, the selective true binder shows the lowest ΔG among the compounds, i.e., it preferentially binds the same target protein. Thus, the MTS method always selects the selective true binder as the top-ranked compound among the compounds of the database, when ΔG values are correctly estimated. Thus, in addition to the robustness, the MTS method could provide better results than those given by the MASC scoring method.

The combination of the MTS and the MASC scoring methods showed a good result. When the 5% compounds out of the whole database were selected for a target protein, the enrichment is higher than that by the 10% compounds. These two methods selected different compounds, thus the union of the results by these two methods can achieve high database enrichment.

The newly developed ΔG score did not improve the hit ratio of in silico screening, since our ΔG score was developed based on the conventional fitting method using the ΔG values of only active compounds. The hit-optimized score did not improve the hit ratio of the MTS and MASC scoring methods, but improved the hit ratio of the conventional raw scoring method. The hit-optimized score succeeded in emphasizing the differences between the active and inactive compounds compared to the ΔG score, but the results of the low hit ratios suggest the poor correlations between the ΔG and the hit-optimized scores.

6. Conclusion

We have developed a new screening method, called the Multiple target screening (MTS) method, based on the protein-compound affinity matrix, and we evaluated the hit ratio (database enrichment) of the raw scoring, MTS, and MASC scoring methods with three scoring functions: the Sievgene docking score, the ΔG score and the hit-optimized score. Also, two kinds of the grid potential were examined, with mesh sizes of $100 \times 100 \times 100$ and $60 \times 60 \times 60$.

The database enrichment by the MTS method was comparable to or slightly better than that by the MASC scoring method. When the mesh size was $100 \times 100 \times 100$ with the Sievgene docking scores, the MASC scoring method was superior to other methods. Although our Sievgene docking score was not optimized to reproduce the binding free energy,

it worked well for screening. In the other cases, the MTS method was the best among those three methods, showing the best database enrichment with the Sievgene docking scores. The MTS method is robust against changes in the scoring function. On the contrary, the MASC scoring method is sensitive to changes in the scoring function. Still, both methods require the same large number of docking simulations for many different proteins. Also, the union of the results by the MTS and the MASC scoring method could show high database enrichment.

Our ΔG score succeeded in reproducing the experimental binding free energy with an average error of 2.47 kcal/mol for 50 protein-ligand complexes. Also, the hit-optimized score improved the database enrichment of the raw scoring method; the database enrichment of the MTS method was somewhat lower than that with the Sievgene docking score.

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

The following PDB identifier list of complexes was used: 1abe, 1abf, 1apu, 1cbx, 1dbb, 1dbj, 1dog, 1dwb, 1ebg, 1epo, 1etr, 1ets, 1ett, 1hpv, 1hsl, 1htf, 1hvr, 1mnc, 1nsd, 1pgp, 1phf, 1phg, 1ppc, 1pph, 1rbp, 1stp, 1tmn, 1tng, 1tnh, 1ulb, 2cgr, 2cpp, 2gbp, 2ifb, 2phh, 2r04, 2tmn, 2tsc, 2ypi, 3ptb, 4dfr, 5abp, 5cpp, 5tln, 6cpa, 7cpa. For 1abe, 1abf, 5abp, and 1htf, two receptor pockets were prepared, since these proteins both bind two ligands each.

The following PDB identifier list is the protein data set A: 1abe, 1abf, 1apu, 1dbb, 1dog, 1ebg, 1etr, 1ett, 1hsl, 1htf, 1mnc, 1pgp, 1phg, 1pph, 1stp, 1tng, 1ulb, 2cpp, 2ifb, 2r04, 2tsc, 3ptb, 5abp, 5cpp and 6cpa.

The following PDB identifier list is the protein data set B: 1abe, 1abf, 1cbx, 1dbj, 1dwb, 1epo, 1ets, 1hpv, 1htf, 1hvr, 1nsd, 1phf, 1ppc, 1rbp, 1tmn, 1tnh, 2cgr, 2gbp, 2phh, 2tmn, 2ypi, 4dfr, 5abp, 5tln and 7cpa.

Appendix B

The following PDB identifier list of complexes was used: 1a28, 1a42, 1a4g, 1a4q, 1abe, 1abf, 1aco, 1ai5, 1aoe, 1apt, 1apu, 1aqw, 1atl, 1b58, 1b9v, 1bkc, 1bma, 1bqq, 1byb, 1byg, 1c1e, 1c5c, 1c83, 1cbs, 1cbx, 1cdg, 1ckp, 1cle, 1com, 1coy, 1cps, 1cvu, 1d0l, 1d3h, 1dd7, 1dg5, 1dhf, 1dog, 1dr1, 1ebg, 1eed, 1ejn, 1epb, 1epo, 1ets, 1f0r, 1f0s, 1f3d, 1fen, 1fkg, 1fki, 1fl3, 1gc7, 1glp, 1hdc, 1hfc, 1hos, 1hpv, 1hsb, 1hsl, 1htf, 1hyt, 1ida, 1ivb, 1jap, 1kj0, 1lah, 1lcp, 1ldm, 1lic, 1lna, 1lst, 1mbi, 1mdr, 1mld, 1mmq, 1mrg, 1mts, 1mup, 1nco, 1ngp, 1nis, 1okl, 1pbd, 1pdz, 1phd, 1phg, 1poc, 1ppc, 1pph, 1pso, 1qbr, 1qbu, 1qpq, 1r55, 1rds, 1rne, 1rnt, 1rob, 1snc, 1srj, 1tlp, 1tmn, 1tng, 1tnh, 1tni, 1tnl, 1tyl, 1xid, 1xie, 1yee, 2aad, 2ack, 2ada, 2cht, 2cmd, 2cpp, 2ctc, 2fox, 2gbp, 2ifb, 2pk4, 2qwk, 2tmn, 3cla, 3cpa, 3erd, 3ert, 3tpi, 4est, 4lbd, 4phv, 5abp, 5cpp, 5er1, 6rnt, and 7tim. For 1abe, 1abf, 5abp, and 1htf, two receptor pockets were prepared, since these proteins both bind two ligands each.

The following PDB identifier list is the protein data set C: 1a28, 1a4g, 1abe, 1abf, 1aco, 1aoe, 1apu, 1atl, 1b9v, 1bqq, 1byg, 1c5c, 1cbs, 1cdg, 1cle, 1coy, 1cvu, 1d3h, 1dg5, 1dog, 1ebg, 1ejn, 1epo, 1f0r, 1f3d, 1fkg, 1fl3, 1hdc, 1hos, 1hsb, 1htf, 1hyt, 1ivb, 1lah, 1ldm, 1lna, 1mbi, 1gcz, 1mmq, 1mts, 1nco, 1nis, 1pbd, 1phd, 1poc, 1pph, 1qbr, 1qpq, 1rds, 1rnt, 1snc, 1bkc, 1tlp, 1tng, 1tni, 1tyl, 1xie, 2aad, 2ada, 2cmd, 2ctc, 2gbp, 2pk4, 2tmn, 3cpa, 3ert, 4aah, 4lbd, 5abp, 5cpp and 6rnt.

The following PDB identifier list is the protein data set D: 1a42, 1a4q, 1abe, 1abf, 1ai5, 1apt, 1aqw, 1b58, 1bma, 1byb, 1c1e, 1c83, 1cbx, 1ckp, 1com, 1cps, 1d0l, 1dd7, 1dhf, 1dr1, 1eed, 1epb, 1ets, 1f0s, 1fen, 1fki, 1glp, 1hfc, 1hpv, 1hsl, 1htf, 1ida, 1jap, 1lcp, 1lic, 1lst, 1mdr, 1mld, 1mrg, 1mup, 1ngp, 1okl, 1pdz, 1phg, 1ppc, 1pso, 1qbu, 1r55, 1rne, 1rob, 1srj, 1kj0, 1tmn, 1tnh, 1tnl, 1xid, 1yee, 2ack, 2cht, 2cpp, 2fox, 2ifb, 2qwk, 3cla, 3erd, 3tpi, 4est, 4phv, 5abp, 5er1 and 7tim.

References

[1] Y. Fukunishi, Y. Mikami, H. Nakamura, Similarities among receptor pockets and among compounds: analysis and application to in silico ligand screening, J. Mol. Graphics Modelling 24 (2005) 34–45.
[2] G.P.A. Vigers, J.P. Rizzi, Multiple active site corrections for docking and virtual screening, J. Med. Chem. 47 (2004) 80–89.
[3] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A Geometric approach to macromolecule-ligand interactions, J. Mol. Biol. 161 (1982) 269–288.
[4] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 261 (1996) 470–489.
[5] G. Jones, P. Willet, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[6] N. Paul, D. Rognan, ConsDock:, A new program for the consensus analysis of protein-ligand interactions, Proteins Struct. Funct. Genet. 47 (2002) 521–533.
[7] C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, Flexible docking using tabu search and an empirical estimate of binding affinity, Proteins Struct. Funct. Genet. 33 (1998) 367–382.
[8] M.R. McGann, H.R. Almond, A. Nicholls, J.A. Grant, F.K. Brown, Gaussian docking functions, Biopolymers 68 (2003) 76–90.
[9] D.S. Goodsell, A.J. Olson, Automated docking of substrates to proteins by simulated annealing, Proteins Struct. Funct. Genet. 8 (1990) 195–202.
[10] J.S. Taylor, R.M. Burnett, DARWIN: a program for docking flexible molecules, Proteins Struct. Funct. Genet. 41 (2000) 173–191.
[11] R. Abagyan, M. Totrov, D. Kuznetsov, ICM: a new method for structure modeling and design: application to docking and structure prediction from the disordered native conformation, J. Comput. Chem. 15 (1994) 488–506.
[12] P.M. Colman, Structure-based drug design, Curr. Opin. Struct. Biol. 4 (1994) 868–874.
[13] A. Krammer, P.D. Kirchhoff1, X. Jiang, C.M. Venkatachalam, M. Waldman, LigScore: a novel scoring function for predicting binding affinities, J. Mol. Graphics Modelling 23 (2005) 395–407.
[14] C. Zhang, S. Liu, Q. Zhu, Y. Zhou, A knowledge-based energy function for protein-ligand, protein–protein, and protein-DNA complexes, J. Med. Chem. 48 (2005) 2325–2335.
[15] I. Muegge, Y.C. Martin, A general and fast scoring function for protein-ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[16] D.G. Hawkins, J.C. Cramer, G.D. Truhlar, Parametrized Models of aqueous free energies of solvation based on pairwise descreening of

solute atomic charges from a dielectric medium, J. Phys. Chem. 100 (1996) 19824–19839.
[17] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptide, Proc. Natl. Acad. Sci. USA 84 (1987) 3086–3090.
[18] J.W.M. Nissink, C. Murray, M. Hartshorn, M.L. Verdonk, J.C. Cole, R. Taylor, A new test set for validating predictions of protein-ligand interaction, Proteins Struct. Funct. Genet. 49 (2002) 457–471.
[19] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074.
[20] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, The general atomic and molecular electronic structure system, J. Comput. Chem. 14 (1993) 1347–1363.
[21] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann Jr., J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A. 9, Gaussian, Inc., Pittsburgh PA, 1998.
[22] 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, University of California, San Francisco, 2004.
[23] C.A. Orengo, D.T. Jones, J.M. Thornton, Protein superfamilies and domain superfolds, Nature 372 (1994) 631–634.
[24] F. Corpet, J. Gouzy, D. Kahn, Recent improvements of the ProDom database of protein domain families, Nucleic. Acids Res. 27 (1999) 263–267.