WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A statistical rescoring scheme for protein–ligand docking: Consideration of entropic effect

Juyong Lee and Chaok Seok*

Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-747, Republic of Korea

ABSTRACT

Computational prediction of protein–ligand binding modes provides useful information on the relationship between structure and activity needed for drug design. A statistical rescoring method that incorporates entropic effect is proposed to improve the accuracy of binding mode prediction. A probability function for two sampled conformations to belong to the same broad basin in the potential energy surface is introduced to estimate the contribution of the state represented by a sampled conformation to the configurational integral. The rescoring function is reduced to the colony energy introduced by Xiang et al. (Proc Natl Acad Sci USA 2002;99:7432–7437) when a particular functional form for the probability function is used. The scheme is applied to rescore protein–ligand complex conformations generated by AutoDock. It is demonstrated that this simple rescoring improves prediction accuracy substantially when tested on 163 protein–ligand complexes with known experimental structures. For example, the percentage of complexes for which predicted ligand conformations are within 1 Å root-mean-square deviation from the native conformations is doubled from about 20% to more than 40%. Rescoring with 11 different scoring functions including AutoDock scoring functions were also tested using the ensemble of conformations generated by Wang et al. (J Med Chem 2003;46:2287–2303). Comparison with other methods that use clustering and estimation of conformational entropy is provided. Examination of the docked poses reveals that the rescoring corrects the predictions in which ligands are tightly fit into the binding pockets and have low energies, but have too little room for conformational freedom and thus have low entropy.

Proteins 2008; 70:1074–1083.

Key words: protein–ligand binding mode; AutoDock; entropy; rescoring; scoring function.

INTRODUCTION

Precise prediction of binding modes and binding affinities of protein–ligand complexes has drawn great interest because of its impact on biological sciences and pharmaceutical industry. Molecular dynamics simulations with proper account of solvation effects are promising because both protein and ligand are allowed fully flexible, and the binding free energy can be calculated in principle from the sampling methods based on statistical mechanics. In practice, computational costs required for the rigorous computation of the free energy is immense, and more practical methods such as MMPB/SA,[1–3] LIE,[4–8] and free energy perturbation methods[9–12] have been developed.

Faster protein–ligand docking methods are therefore being frequently used for structure-based drug design. Proteins are often considered rigid, enabling fast scoring with the use of energy grids. This constraint of frozen protein can be relaxed by considering side-chain rotamers[13,14] or by utilizing an ensemble of backbone conformations.[15–17] However, incorporating protein flexibilities especially in the cases involving large induced conformational change is still a challenge. Scoring function is another important factor determining the accuracy of docking. Especially when conformational search is close to complete with grid-based searches of relatively small ligands, scoring function determines the accuracy.

Various implementations of scoring functions for fast docking are available, and some of them are complementary with others, thus being improved by consensus scoring methods.[18–20] The scoring functions may be roughly classified into energy-based scoring functions[21–25] that resemble energy terms in force fields, empirical scoring functions[26–28] whose parameters are more heavily dependent on empirical data, and knowledge-based scoring functions[29,30] that are derived from the database. Improvements in scoring functions have been reported by more accurate account of partial charges,[31] desolvation effects,[32] or balance of different

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsors: Seoul R&BD Program; MarineBio21, Ministry of Maritime Affairs and Fisheries, Korea.

*Correspondence to: Chaok Seok, Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-747, Republic of Korea. E-mail: chaok@snu.ac.kr

Received 5 June 2007; Revised 19 September 2007; Accepted 26 September 2007
Published online 12 December 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21844

score terms as in consensus scoring functions,[18–20] for example.

Contribution of entropy has been perceived to be important when predicting binding modes. Clustering methods are often employed to effectively account for entropic effect by taking a representative structure in the most populated cluster. Recently it is reported that considering a few high ranking binding modes together improves the success rate of binding mode prediction.[33,34] This fact implies that useful information can be extracted from the ensemble of conformations sampled by the existing docking programs. The sampled structures have information on the energy surface of the system, although the exact details would depend on the specific sampling method. Entropic effect may be extracted from the set of conformations corresponding to each binding mode. Here, a rescoring scheme that analyzes the populations of binding modes in the context of statistical mechanics is proposed. Ruvinsky *et al.* proposed methods similar in spirit,[35–37] in which the configuration integral is approximated from the ranges of coordinates that are covered by sampled conformations or is estimated from the populations of conformation clusters. A related method to account for entropy when predicting binding affinity was also proposed by Salaniwal *et al.*[38]

Our statistical rescoring function is reduced to the "colony energy," proposed by Xiang *et al.*[39] in the context of protein loop modeling, when a particular functional form is used for the probability function introduced here. An explicit connection of the colony energy to the free energy is provided using the probability function for a pair of conformations to be in the same state. This method can be easily combined with preexisting docking scoring functions, and requires very little extra computational costs because no energy minimizations, dynamics simulations, or clustering is needed. When compared with binding poses predicted with the original AutoDock scoring function for 163 complexes with known experimental structures, our rescoring produced improved results in 99 cases with the average RMSD improvement of 1.18 Å, the same predictions in 39 cases, and worse results only in 25 cases by average RMSD of 0.41 Å. The average RMSD of predicted poses decreased from 2.28 to 1.63 Å with the simple rescoring.

METHODS

Rescoring scheme and the colony energy

Suppose that there are N protein–ligand complex conformations sampled using a docking program with an energy-based scoring function. Conformational sampling is usually very sparse in docking simulations because of the need to deal with a large number of complexes, for example, for virtual screening. However, the sampling is supposed to be dense enough that at least several local minima are sampled for those important states with substantial contribution to thermodynamic properties. A state here is defined as an ensemble of local energy minima that are close enough in the conformational space and also in potential energy to be interconvertible at the temperature of interest. In other words, the conformations that belong to the same state are considered to be in the same broad potential energy basin, and different states are separated by large energy barriers compared to thermal energy.

The configuration space is partitioned into n parts corresponding to n different states, and the partition function is written as

$$Z = \int e^{-\beta E(\mathbf{r})} d\mathbf{r} = \sum_{I=1}^{n} Z_I, \qquad (1)$$

where the coordinates are represented collectively as $\mathbf{r}$, $E(\mathbf{r})$ is the potential energy function, and β is $1/(k_B T)$. The contribution due to the Ith state, Z_I, is the integral over the configuration space volume V_I encompassed by the state as follows:

$$Z_I \equiv \int_{V_I} e^{-\beta E(\mathbf{r})} d\mathbf{r}. \qquad (2)$$

Let the jth conformation sampled by a docking program has energy score E_j, where $j = 1, 2, \ldots, N$. Z_I is approximated in terms of the sampled local minima as

$$Z_I = \sum_{j=1}^{N} \rho_{Ij} \int_{v_j} e^{-\beta E(\mathbf{r})}\, d\mathbf{r} \cong \sum_{j=1}^{N} \rho_{Ij}\, e^{-\beta E_j}\, v_j, \qquad (3)$$

where $\rho_{Ij} = 1$ if the jth conformation belongs to the I th state, and 0 otherwise, and v_j is the configuration space volume represented by the jth conformation which also includes the space not explicitly sampled. When N is doubled, v_j is expected be halved approximately. Further simplifying approximation of slowly varying energy within the volume v_j is introduced to obtain the last equality.

For each state I, a reference conformation $i_{\mathrm{ref}}(I)$ is selected, and Z_I is now expressed as

$$Z_I = \sum_{j=1}^{N} \alpha(i_{\mathrm{ref}}(I), j) e^{-\beta E_j}\, v_j, \qquad (4)$$

where $\alpha(i, j) = 1$ if the two conformations i and j belong to the same state, and 0 otherwise. However, it is not known a priori whether the given conformations belong to the same state or not, and there is always an ambiguity in assigning conformations into states. Therefore, a more fuzzy form for $\alpha(i, j)$ is employed, which is interpreted as the "probability" that i and j belong to the same state.

The free energy of the protein–ligand complex can be written as

$$\Delta G = -k_B T \ln\left[\sum_I Z_I\right], \quad (5)$$

where the contribution from the configuration integrals of free ligand and free protein and other constant terms are omitted. Note that only the relative stabilities of different bound states are considered here. Assuming that the contribution from the maximum Z_I dominates, as in the case of a single strong binding mode, Eq. (5) is reduced to

$$\begin{aligned}\Delta G &= -k_B T \ln\left[\max_I Z_I\right] \\ &= -k_B T \ln\left[\sum_{j=1}^{N} \alpha(i_{\mathrm{ref}}(I_{\min}), j) e^{-\beta E_j} v_j\right], \end{aligned} \quad (6)$$

where $I_{\min}$ is the state I that results in the lowest colony energy $\Delta\mathrm{CE}(I)$ defined as

$$\Delta\mathrm{CE}(I) = -k_B T \ln\left[\sum_{j=1}^{N} \alpha(i_{\mathrm{ref}}(I), j) e^{-\beta E_j} v_j\right]. \quad (7)$$

The volume v_j is difficult to estimate, depending both on the potential energy surface and the sampling method. Here it is simply taken to be constant assuming that the sampling is reasonably uniform, leading to

$$\Delta\mathrm{CE}(I) = -k_B T \ln\left[\sum_{j=1}^{N} \alpha(i_{\mathrm{ref}}(I), j) e^{-\beta E_j}\right]. \quad (8)$$

The approximation of constant v_j would be most accurately applied when sampling is close to uniform, for example, when conformations are obtained by independent sampling from multiple random initial conformations.

It is a nontrivial task to explicitly find $i_{\mathrm{ref}}(I)$ and $I_{\min}$, especially with the sparsely sampled conformations. One approach is to cluster the conformations to obtain the reference conformations. In the formalism described here, the reference states $i_{\mathrm{ref}}(I)$ are not tried to be found explicitly. Instead, it is noted that the value of $\alpha(i_{\mathrm{ref}}(I),j)$ does not change very much if $i_{\mathrm{ref}}(I)$ is replaced with any other conformation in the same state I. For convenience, the approximation

$$\sum_{j=1}^{N} \alpha(i_{\mathrm{ref}}(I), j) e^{-\beta E_j} \cong \max_{i \in I} \sum_{j=1}^{N} \alpha(i, j) e^{-\beta E_j} \quad (9)$$

is made. The task of the binding mode prediction is now reduced to finding the conformation $i_{\min}$ that has the minimum value of the colony energy, now assigned to each conformation as

$$\Delta\mathrm{ce}(i) \equiv -k_B T \ln\left[\sum_{j=1}^{N} \alpha(i, j) e^{-\beta E_j}\right]. \quad (10)$$

The state I that contains the conformation $i_{\min}$ corresponds to $I_{\min}$. Since the conformations belonging to the same states are similar by definition, this conformation $i_{\min}$ is taken as the conformation representing the most stable state $I_{\min}$. The colony energy $\Delta\mathrm{ce}(i)$ is equivalent to the colony energy of Xiang *et al.*,[39] proposed for the case of protein loop modeling, when combined with a particular functional form for the probability function $\alpha(i, j)$.

The probability function

The probability function $\alpha(i, j)$ must be 1 when $i = j$, that is, when i and j are in the same energy basin, and must approach 0 when i and j conformations are very different, that is, when i and j are in different basins. An additional approximation $\alpha(i, j) = \alpha(\mathrm{rmsd}_{ij})$ is introduced, where rmsd_{ij} is the root-mean-square distance between the conformations i and j. The following four functional forms have been tested:

1. Step function ("Step"): $\alpha(x) = \begin{cases} 1, & \text{if } x \le x_c \\ 0, & \text{if } x > x_c \end{cases}$
2. Exponential function of x ("Exp1") : $\alpha(x) = \exp(-x/\gamma_1)$
3. Gaussian function of x ("Exp2"): $\alpha(x) = \exp(-x^2/\gamma_2)$
4. Exponential function of x^3 ("Exp3"): $\alpha(x) = \exp(-x^3/\gamma_3)$

As the power of x increases, exponential function approaches the step function. When examining the effect of parameters, a characteristic length is defined for each function: x_c for the step function, the inflection points for "Exp2" and "Exp3," and the point of half probability value for "Exp1."

Protein–ligand complex sets

The number of the protein–ligand complexes tested in this study is 163. The complexes deposited in the protein–ligand binding affinity database[40] were selected with the following two criteria to limit the size of the set: (1) molecular weight of ligand less than 500 Da and (2) no hetero atoms within 8 Å from ligand. The pdb codes are listed in Supplementary Table I. The atomic coordinates for the complexes were extracted from the RCSB PDB database,[41] and processed with the SYBYL7.2 program[42]: ligand and protein were separated, hydrogen atoms were attached to ligands, Gasteiger–Hückel charges were assigned to ligand atoms, polar hydrogen atoms were

DOI 10.1002/prot

Table I
Accumulated Percentages of the Predictions Within Given RMSD and the Average RMSD for the Original AutoDock Score and the Rescoring Method Used With Four Different Probability Functions, Step, Exp1, Exp2, and Exp3

RMSD from native (Å)	AutoDock	Step	Exp1	Exp2	Exp3
≤1.0	22.1	45.4	42.3	43.6	44.2
≤1.5	49.7	75.5	70.6	69.9	71.8
≤2.0	63.8	79.8	77.9	78.5	78.5
≤2.5	71.8	82.8	81.0	81.6	82.2
≤3.0	77.3	85.9	84.7	84.7	85.3
Average RMSD (standard deviation) (Å)	2.28 (1.90)	1.61 (1.47)	1.64 (1.48)	1.66 (1.53)	1.63 (1.51)

attached to protein, and Kollman united charges were assigned to protein atoms with the biopolymer module.

Docking calculation

The AutoDock3.0.5 program[21] was used to sample ligand conformations. The Lamarckian genetic algorithm (GA)[21] was used for conformational sampling. The default parameters for AutoDock runs were maintained in general, except for the number of conformations (*ga_run*) and the maximum number of energy evaluations (*ga_num_evals*). The number of conformations was varied to test the effect of different degrees of the conformation space coverage, and the maximum number of energy evaluations was varied to examine the effect of depth of the energy search. This results in four sets of conformations: (*ga_run*, *ga_num_evals*) = (500, 100,000), (500, 250,000), (250, 100,000), and (250, 250,000), referred to as "Set1," "Set2," "Set3," and "Set4," respectively. Default parameter values were used for other running parameters such as *ga_num_generations* = 27,000, *ga_pop_size* = 50, *ga_mutation_rate* = 0.02, *ga_elitism* = 1, and *ga_crossover_rate* = 0.8.

Comparison with other rescoring methods

In addition to the original AutoDock scoring function, two different rescoring methods were also compared with the new rescoring method. The two methods first cluster the sampled conformations, and then score the clusters by: (1) the size, N_I, of the clusters ("MaxN") and (2) both the occupancy, N_I, and the minimum energy, E_I, of the clusters ("MaxP"). In the MaxN method, the lowest energy conformation in the largest cluster is selected. In the MaxP method, following Ref. 36, the lowest energy conformation in the cluster I of the lowest free energy $\Delta\tilde{G}_I = E_I - k_B T \ln N_I$ is selected. The RMSD criterion for clustering in AutoDock was set to 2 Å, following Refs. 35 and 36, and no effort for optimizing this parameter and others were made.

Application to 11 scoring functions

The ensemble of conformations for 100 protein–ligand complexes generated by Wang *et al.*[43] using AutoDock were adopted, and the rescoring was performed for the 11 scoring functions tested in Ref. 43 (LigScore, PLP, PMF, LUDI, *F*-Score, *G*-Score, *D*-Score, ChemScore, AutoDock, DrugScore, and *X*-Score). Some of the scoring functions (AutoDock, LUDI, ChemScore, and *X*-Score) have an empirical torsional entropy term to account for entropy loss upon binding, but the term depends only on the type of the ligand but not on the conformation of the ligand. Because different conformations of the same ligand are compared in the current binding mode prediction, the torsional entropy term contributes just a constant. Therefore, torsional entropy is not double counted in the present rescoring scheme.

The number of conformations for each complex is 101. The authors of Ref. 43 used the GA of the AutoDock program for conformational sampling, as in "Docking calculation" section. The generation method is a little different from that described in the "Docking calculation" section for the 163 complexes, however. In Ref. 43, the focus was to generate "diverse" conformations, but for the current rescoring scheme to be successful, "degeneracy" in conformation is desired to be collected because the degree of degeneracy reflects the size of the basin of attraction in the binding energy surface. In Ref. 43, the mutation rate and elitism parameters were set to 0.2 and 0.1, respectively, to obtain diverse sampling. In the "Docking calculation" section, the default values of 0.02 and 1 were used to obtain reasonable local convergence. In Ref. 43, the length of the GA runs was varied to obtain the diversity in RMSD to a desirable level. In the "Docking calculation" section, it is determined by the preset value of the maximum number of energy evaluations, and the effect of this parameter was tested. The length of the GA runs turns out to be similar: 50–200 in Ref. 43 and 50–150 in the "Docking calculation" section. The number of independent GA runs, which becomes the number of conformations sampled, was 100 in Ref. 43, but 250 or 500 were tested in the "Docking calculation" section to facilitate probing of the binding energy surface. Therefore, the use of the ensembles of conformations provided by Wang *et al.*[43] is less optimal for the present rescoring method compared to the ensemble generated in the "Docking calculation" section, but application of our rescoring method to the 11 scoring functions

Table II
Changes in Prediction After Rescoring

	Step	Exp1	Exp2	Exp3
No. cases of decreased RMSD	104	98	98	99
Average RMSD change (standard deviation) (Å)	−1.16 (1.78)	−1.14 (1.74)	−1.14 (1.77)	−1.18 (1.817)
No. cases of increased RMSD	30	17	23	25
Average RMSD change (standard deviation) (Å)	0.39 (0.63)	0.39 (0.60)	0.41 (0.71)	0.41 (0.69)
No. cases of equal RMSD	29	48	42	39

illustrates that the method can be still useful in this less optimal application.

RESULTS AND DISCUSSION

Improvement of docking accuracy

The accuracies of the predicted structures with the scoring function of AutoDock and with the rescoring function were examined in terms of RMSDs of the best scored conformations from the known experimental structures. The average of RMSD of the best scored conformations over the 163 complexes and the RMSD distribution are summarized in Table I. Results for the rescoring function with the four different probability functions introduced in Methods are presented together. Only the results for Set1 (see Methods) are shown in Table I because the results for the other three sets are similar: the average RMSDs of the predicted structures with AutoDock score and with the rescoring function with the Exp3 probability function are (AutoDock, Exp3) = (2.28, 1.63), (2.47, 1.67), (2.32, 1.70), (2.23, 1.62) for Set1, Set2, Set3, and Set4, respectively.

There is substantial improvement in the prediction with rescoring. For example, the percentage of the predictions within 1 Å RMSD from native increased from 22.1 to 44.2% when the probability function "Exp3" is used. The degree of improvement is not sensitive to the choice of the functional form of the probability function, as can be seen from Table I. The choice of parameters of the probability function does not influence the results substantially, either, except for the case of step function, as discussed later. The data in Table I were obtained for the characteristic length of 1.5 Å, for which the step function results in the best prediction.

Table II presents the data in a different manner. The number of cases in which the RMSD of the best scored conformation decreased after rescoring is 99 out of 163, for example, when "Exp3" probability function is used, whereas the number of increased cases is only 25. The conformations selected as the best remained the same in 39 cases. The average change in RMSD is −1.18 Å when RMSD decreases and 0.42 Å when RMSD increases.

Two examples are demonstrated in Figures 1 through 4: Figures 1 and 2 are for the complex pdb code 4cla, and Figures 3 and 4 are for 1ivb. In Figures 1 and 3,

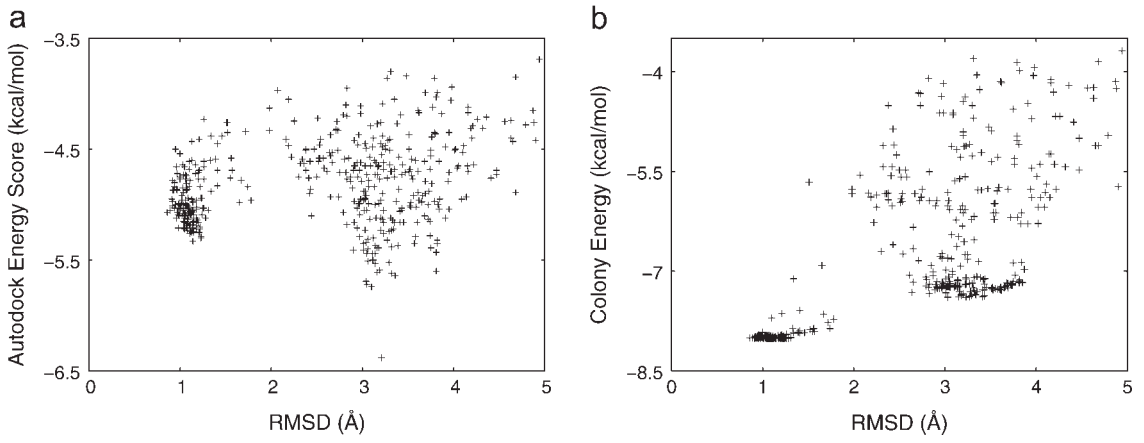

Figure 1
*(**a**) AutoDock score versus RMSD from the crystal ligand pose for ligand conformations generated for the complex 4cla using AutoDock. (**b**) Colony energy versus RMSD after rescoring of the ligand conformations generated for the complex 4cla. Exp3 for the probability function, inflection point of 1.5 Å, and T = 300 K were used. The results are not very sensitive to these parameters, as discussed in the text.*

DOI 10.1002/prot

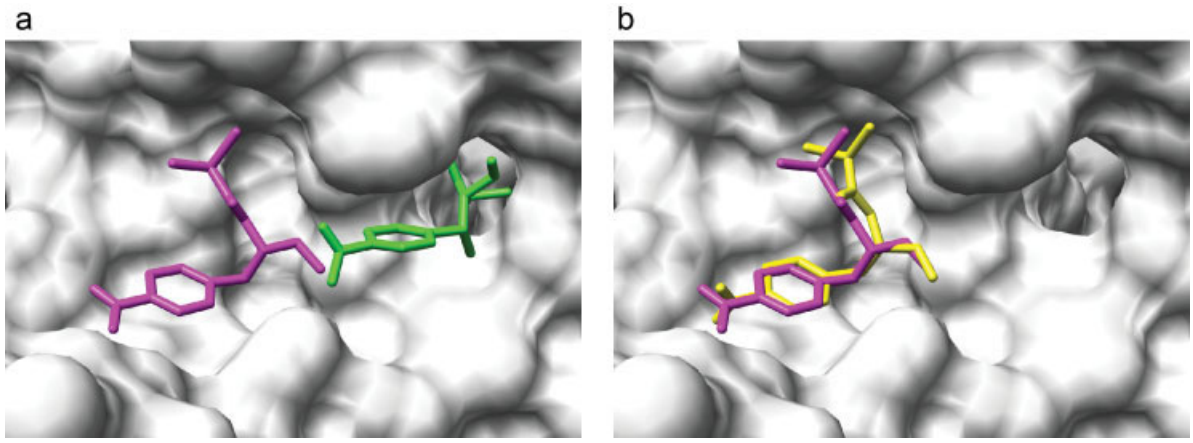

Figure 2

*(**a**) Binding pocket of the complex 4cla. The crystal ligand pose is shown in magenta, the conformation with the lowest energy in green, and the protein surface in gray. UCSF chimera[44,45] was used to prepare this figure and similar figures. (**b**) Binding pocket of the complex 4cla. The crystal ligand conformation is in magenta, the conformation with the lowest colony energy in yellow, and the protein surface in gray. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

scores are plotted as functions of RMSD for the sampled conformations. In Figure 1, the conformations can be roughly grouped into two clusters, one around 3 Å from the native (referred to as "A"), and the other around 1 Å (referred to as "B"). In Figure 1(a), the conformations in the A cluster have lower energy scores than those in the B cluster with the AutoDock scoring function, but more conformations were sampled in the B cluster, implying greater conformational entropy for this state. This entropic effect is reflected to the colony energy score, as in Figure 1(b), where the conformations in the B cluster have lower colony energies. Larger conformational entropy can be interpreted as being originated from the larger room for ligand in the binding pocket for the B state, as illustrated in Figure 2(a,b). In the A state, one end of ligand is packed into a small hole, leaving not much room for conformational flexibility. Such a tightly bound ligand conformation can be more favorable energetically, but unfavorable in terms of entropy, resulting in less stability.

In Figure 3(a), a single conformation about 2.8 Å away from native has the lowest energy. As can be seen from Figure 4(a), this conformation has very good contact with the receptor protein, but shows similar problem as

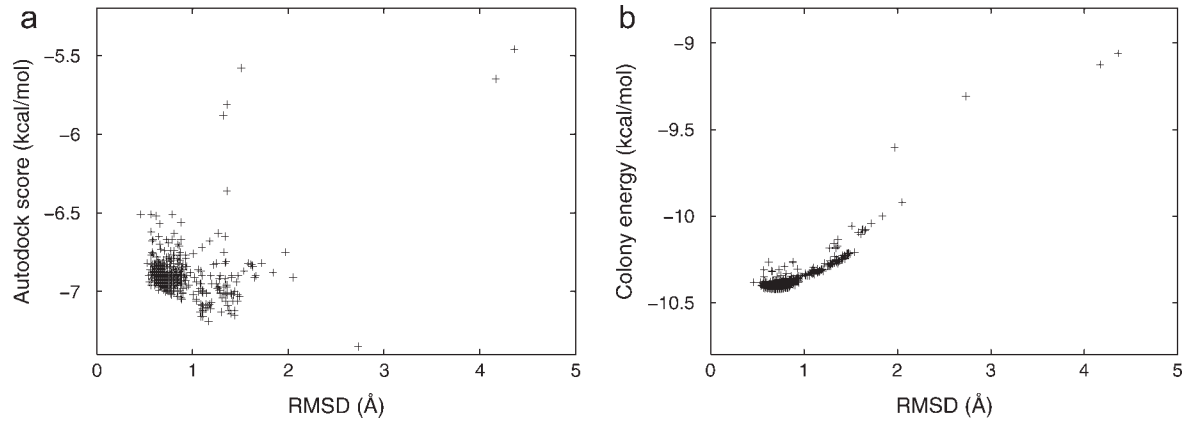

Figure 3

*(**a**) AutoDock score versus RMSD from the crystal ligand pose for ligand conformations generated for the complex 1ivb using AutoDock. (**b**) Colony energy versus RMSD after rescoring of the ligand conformations generated for the complex 1ivd. The same parameters were used as in Figure 1(b).*

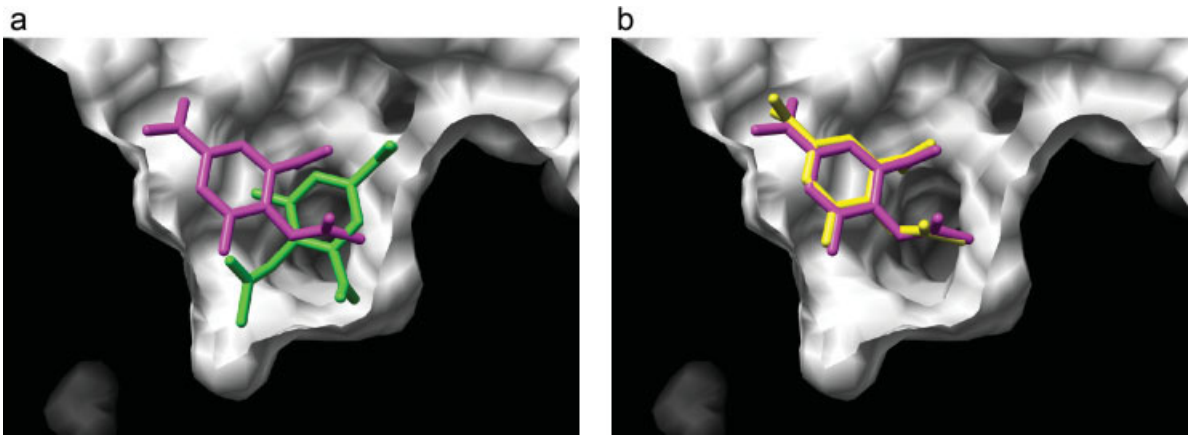

Figure 4

*(**a**) Binding pocket of the complex 1ivb. The crystal ligand pose is shown in magenta, and the conformation with the lowest energy is in green. (**b**) Binding pocket of the complex 1ivb. The crystal ligand conformation is in magenta, and the conformation with the lowest colony energy is in yellow. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

above in that there is not much room for conformational freedom. When rescored, a state much closer to the native pose is predicted to be more stable, as in Figures 3(b) and 4(b).

Dependence on parameters

The central part of the rescoring scheme described in Methods is the probability function. In an accurate formulation, the probability function must be replaced by an indicator function, assigning a particular energy basin to each conformation. Although the probability function is an approximation to the indicator function, correcting the energy score approximately results in improvement in accuracy when states with very different entropies occur, as demonstrated earlier. Four different functional forms have been tried, and the sensitivity of prediction accuracy to the parameters (characteristic lengths) has been tested. Results for different functional forms of the probability function are summarized in Table I, and the dependences of the results on the characteristic lengths are plotted in Figure 5. The prediction accuracy does not sensitively depend on the choice of the probability function except when the step function is used. A cutoff value of 1.5 Å was found to be the best in the case of the step function.

The rescoring scheme does not use clustering and therefore clustering criterion is not needed. The characteristic length in the current method corresponds to clustering criterion, such as the RMSD criterion in AutoDock. When clustering methods are used, prediction accuracy may sensitively depend on the clustering criterion.[37] Figure 5 shows that the result of the current rescoring method is not sensitive to the characteristic length when a smooth function is used for the probability function. This robustness upon change in parameter is one of the merits of our method.

In addition to the parameters of the probability function, dependence on temperature T was also examined. The energy score of a docking program is scaled to reproduce experimental binding affinities, and the energy score thus includes the entropic effect effectively in the scaling factors. In the rescoring scheme introduced here, the most proper energy scale may therefore be different from the energy scale used in the docking program. Proper rescaling of the energy can be performed by introducing "effective" temperature, that is, by treating

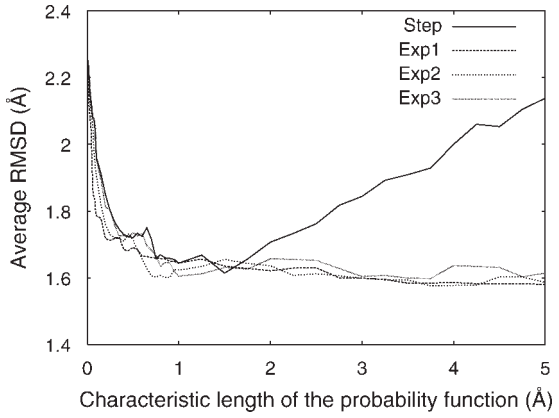

Figure 5

Average RMSD of the best scored ligand as a function of the characteristic length of the probability function. Average RMSD is almost constant for the characteristic length >1 Å, except for the step function.

DOI 10.1002/prot

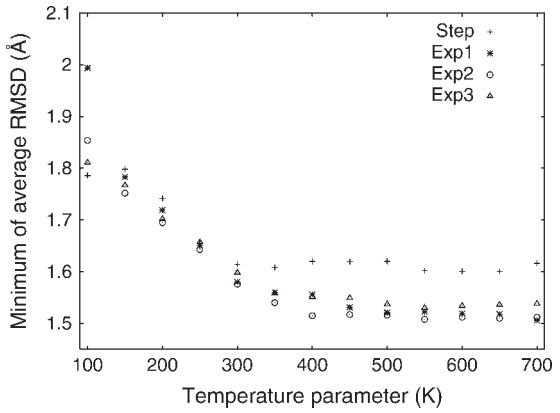

Figure 6

Dependence of the prediction accuracy, as represented by the minimum of average RMSD (over the characteristic length) on the temperature parameter.

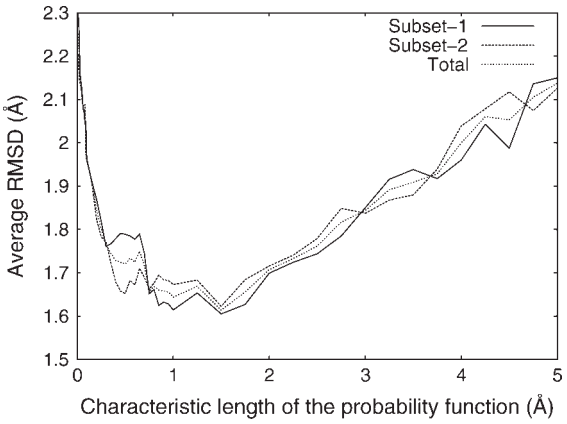

Figure 7

Average RMSD as a function of the characteristic length of the Step probability function for the total set and the two subsets.

temperature as a parameter. The dependence of the prediction accuracy on temperature was therefore examined. Figure 6 shows that the results are not very sensitive for temperature greater than 300 K for the step function and 400 K for other smooth functions. $T = 300$ K was used here unless specified otherwise.

Dependence on training set

To test whether the results are transferable, the total set of 163 protein–ligand complexes were randomly divided into two sets of 82 and 81 complexes, and the parameter and function space for the probability function were searched. The members of the subsets are listed in Supplementary Table II. Dependence on the functional form and the parameter with the subsets is almost the same as that with the total set. The best parameters and the trends of the parameter dependence shown in Figure 7 confirm that essentially the same degree of improvement can be obtained with smaller training sets.

Comparison with other rescoring methods

Table III compares rescoring results of the colony energy (with the Exp3 probability function) with those of the MaxN and MaxP methods described in the "Comparison with other rescoring methods" section. It is seen that average RMSD is improved by all the three rescoring methods to roughly similar extents. Although Exp3 gives only slightly smaller average RMSD than MaxN and MaxP, it is promising that the number of improved cases for Exp3 is noticeably larger than MaxN and MaxP, 99 compared to 69 and 61.

Application to 11 scoring functions

Success rates of the different rescoring functions (colony energy with Exp3, MaxN, and MaxP) when applied to the 11 scoring functions with the ensemble of conformations of Wang *et al.*[43] are compared in Table IV. The success rate is measured as the percentage of cases in which the predicted conformations are within 2 Å of the native conformations. The results of the MaxN and MaxP methods are taken from Ref. 37. The maximum and the range of success rates are shown for MaxN and MaxP, respectively, when the RMSD criterion for clustering is varied. The two parameters, the characteristic length of the probability function and the temperature parameter can be varied for Exp3. As presented in Figures 5 and 6, the results are not very sensitive to these parameters, and the parameters were not optimized for the test set of 163 complexes. The parameters were optimized here for the best overall performance over the 11 scoring functions and the set of 100 complexes, resulting in the characteristic length = 2 Å and the temperature = 2400 K. This set of parameters also performs slightly better than the results presented in the previous subsections for rescoring of AutoDock with 163 complexes.

It is noted that there are improvements when the new rescoring method is combined with 9 out of the 11 scoring functions, although the conformation ensembles used here are not tuned for each of the rescoring function.

Table III

Comparison of the Current Rescoring Method (Used With the Exp3 Probability Function) With Other Statistical Methods, MaxN and MaxP

	Exp3	MaxN	MaxP
Average RMSD (standard deviation) (Å)	1.63 (1.51)	1.69 (1.36)	1.73 (1.37)
No. cases of decreased RMSD	99	69	61
No. cases of increased RMSD	25	29	29
No. cases of equal RMSD	39	65	73

Table IV
Success Rate When Applied to 11 Scoring Functions

Type of the scoring function	Success rate of the original scoring function (%)	Exp3 (%)[a]	MaxN (%)[b]	MaxP (%)[c]
LigScore	74	73	74	74–82
PLP	76	72	79	77–82
PMF	52	58	72	54–71
LUDI	67	78	77	67–79
F-Score	74	81	77	76–82
G-Score	42	71	70	44–67
D-Score	26	69	69	33–67
ChemScore	35	68	67	41–74
AutoDock	62	78	75	65–75
DrugScore	72	83	75	73–79
X-Score	66	78	74	66–75

[a]With characteristic length = 2.0 Å, T = 2400 K.
[b]With cluster-RMSD optimized for each scoring function.
[c]With cluster-RMSD varied in 0.5–4 Å.

The success rates of Exp3 with a fixed set of parameters are higher in 7 out of 11 scoring functions than those of MaxN for which the cluster-RMSD parameter is individually optimized for each scoring function. The success rates of Exp3 of the same parameter set are higher in 9 and 5 scoring functions than the minimum and maximum success rates of MaxP, respectively. The degree of improvement depends on the type of the scoring function, and the overall performance for a single fixed set of parameters seems to be comparable to the optimized results of MaxN or MaxP method. The advantage of the colony energy is that the rescoring method does not require explicit clustering, and is much less sensitive to parameters than other clustering methods.

CONCLUSIONS

A rescoring scheme for protein–ligand binding mode prediction was described, and demonstrated to be successful in improving prediction accuracy. The rescoring algorithm may be applied when multiple conformations are sampled with a preexisting docking program. The rescoring function is equivalent to the colony energy[39] which naturally takes entropic effect into account. In this article, the colony energy expression was derived from the configuration integral by introducing a probability function. Additional cost for the rescoring is negligible, compared to the time spent for conformational sampling. The prediction accuracy for 99 out of 163 complexes was improved by 1.18 Å on average, that for 39 complexes remained the same, and that for 25 complexes became worse only by 0.41 Å when AutoDock is used to sample conformations. When a set of conformations generated with AutoDock by Wang *et al.*[43] were rescored with 11 different scoring functions, the degree of improvement drops, but consistent improvement in accuracy is still observed. When compared with other rescoring methods that involve explicit clustering, the new rescoring method performs better for the conformational ensembles of 163 complexes generated by us, and comparable for the conformational ensembles of 100 complexes generated by Wang *et al.*[43] and applied to 11 different scoring functions. It is notable that the conformations for the 163 complexes were generated with rescoring in mind, and those for the 100 complexes were generated by Wang *et al.*[43] for a different purpose.

Entropic effect has been effectively considered in many algorithms for protein structure prediction and other applications,[46,47] where sampled conformations are clustered, and a representative structure in the largest cluster is selected to be the best one. The rescoring scheme introduced here may be useful for such applications as protein structure prediction or prediction of protein–protein or protein–nucleic acids interactions.

ACKNOWLEDGMENTS

The authors thank Dong-Seon Lee for help with ligand preparation and Woojin Lee for advice with AutoDock runs.

REFERENCES

1. Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. J Med Chem 2002;45:1412–1419.
2. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
3. Swanson JMJ, Henchman RH, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 2004;86:67–74.
4. Åqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Protein Eng 1994;7:385–391.
5. Hansson T, Åqvist J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. Protein Eng 1995;8:1137–1144.
6. Åqvist J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. J Comput Chem 1996;17:1587–1597.
7. Smith RH Jr, Jorgensen WL, Tirado Rives J, Lamb ML, Janssen PA, Michejda CJ, Smith MBK. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. J Med Chem 1998;41:5272–5286.
8. Wang W, Wang J, Kollman PA. What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? Proteins 1999;34:395–402.
9. Miyamoto S, Kollman PA. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. Proteins 1993;16:226–245.
10. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein–ligand recognition. Acc Chem Res 2002;35:430–437.

DOI 10.1002/prot

11. Kong X, Brooks CL III. λ-Dynamics: a new approach to free energy calculations. J Chem Phys 1996;105:2414–2423.
12. Guo Z, Brooks CL, III. J Rapid screening of binding affinities: application of the λ-dynamics method to a trypsin-inhibitor system. Am Chem Soc 1998;120:1920–1921.
13. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
14. Shaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. Proteins 1998;33:295–310.
15. Claußen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
16. Kua J, Zhang Y, McCammon JA. Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. J Am Chem Soc 2002;124:8260–8262.
17. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161–1182.
18. Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 1999;42:5100–5109.
19. Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. I. Evaluation of different docking/scoring combinations. J Med Chem 2000;43:4759–4767.
20. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002;20:281–295.
21. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–1662.
22. Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput-Aided Mol Des 2001;15:411–428.
23. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
24. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
25. Abagyan RA, Totrov MM, Kuznetsov DA. ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J Comput Chem 1994;15:488–506.
26. Accelrys, Inc. Cerius2, version 4.11. Available at http://www.accelrys.com/.
27. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2:317–324.
28. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions. I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput-Aided Mol Des 1997;11:425–445.
29. Muegge I, Martin YC. A general and fast scoring function for protein–ligand interactions: a simplified potential approach. J Med Chem 1999;42:791–804.
30. Zhang C, Liu S, Zhu Q, Zhou Y. A knowledge-based energy function for protein–ligand, protein–protein, and protein-DNA complexes. J Med Chem 2005;48:2325–2335.
31. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model. I. Method. J Comput Chem 2000;21:132–146.
32. Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins 1999;34:4–16.
33. Verkhivker GM, Bouzida D, Gehlaar DK, Rejto PA, Schafffer L, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Hierarchy of simulation models in predicting structure and energetics of the Src SH2 domain binding to tyrosyl phosphopeptides. J Med Chem 2002;45:72–89.
34. Källiblad P, Mancera RL, Todorov NP. Assessment of multiple binding modes in ligand–protein docking. J Med Chem 2004;47:3334–3337.
35. Ruvinsky AM, Kozintsev AV. New and fast statistical-thermodynamic method for computation of protein–ligand binding entropy substantially improves docking accuracy. J Comput Chem 2005;26:1089–1095.
36. Ruvinsky AM, Kozintsev AV. Novel statistical-thermodynamic methods to predict protein–ligand binding positions using probability distribution functions. Proteins 2006;62:202–208.
37. Ruvinsky AM. Role of binding entropy in the refinement of protein–ligand docking predictions: analysis based on the use of 11 scoring functions. J Comput Chem 2007;28:1364–1372.
38. Salaniwal S, Manas ES, Alvarez JC, Unwalla RJ. Critical evaluation of methods to incorporate entropy loss upon binding in high-throughput docking. Proteins 2007;66:422–435.
39. Xiang Z, Soto CS, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 2002;99:7432–7437.
40. Block P, Sotriffer CA, Dramburg I, Klebe G. AffinDB: a freely accessible database of affinities for protein–ligand complexes from the PDB. Nucleic Acids Res 2006;34:D522–D526.
41. RCSB PDB database. Available at http://www.rcsb.org/pdb/.
42. Tripos, Inc. SYBYL version 7.2. Available at http://www.tripos.com/.
43. Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 2003;46:2287–2303.
44. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
45. Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38:305–320.
46. CASP6 proceedings. Proteins 2005;61:1–236.
47. CAPRI articles. Proteins 2005;60:1–323.