Improving the Accuracy of an Affinity Prediction Method by Using Statistics on Shape Complementarity between Proteins

Tatsuya Yoshikawa,[†,‡] Koki Tsukamoto,[†] Yuichiro Hourai,[†] and Kazuhiko Fukui*[,†]

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan, and Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

Received September 1, 2008

To elucidate the partners in protein−protein interactions (PPIs), we previously proposed an affinity prediction method called *affinity evaluation and prediction* (AEP), which is based on the shape complementarity characteristics between proteins. The structures of the protein complexes obtained in our shape complementarity evaluation were selected by a newly developed clustering method called *grouping*. Our previous experiments showed that AEP gave accuracies that differed with the data composition and scale. In this study, we set a data scale ($84 \times 84 = 7056$ protein pairs) including 84 biologically relevant complexes and then designed 225 parameter sets based on four key parameters related to the grouping and the calculation of affinity scores. As a result of receiver operating characteristic analysis, we obtained 27.4% sensitivity (= recall), 91.0% specificity, 3.5% precision, 90.2% accuracy, 6.3% F-measure^{max}, and an area under the curve of 0.585. Chiefly by optimization of the grouping, AEP was able to provide prediction accuracy for a maximum F-measure that statistically distinguished 23 target complexes among 84 protein pairs. Moreover, the active sites of these complexes were successfully predicted with high accuracy (i.e., 2.37 Å in 1CGI and 2.38 Å in 1PPE) of interface RMSD. To assess the improvement in accuracy we compared the results of AEP of different data sets and of tentative methods using ZDOCK 3.0.1 or ZRANK scores.

1. INTRODUCTION

Along with nucleic acids (DNA and RNA), carbohydrates, and lipids, proteins are important materials that support vital cellular functions. They have their own specific structures and functions, depending on the types, numbers, and binding order of their constituent amino acids. It is well-known that most of the functions specific to a particular protein are related to interactions of that protein with other proteins. These protein−protein interactions (PPIs) play central roles in biological phenomena, including cellular signaling, enzyme reactions, and gene expression regulation. Previously, several reviews[1−3] have examined the characteristics of interactions in known protein complexes, and PPI maps, which can help us to discover protein partners, have been generated by experimental or computational approaches. Most of the comprehensive studies of PPI maps have been derived from genome-scale analyses using yeast two-hybrid (Y2H)[4] *in vitro* or protein-fragment complementation assay (PCA) *in vivo* in several organisms such as *Saccharomyces cerevisiae*,[5−7] *Caenorhabditis elegans*,[8] *Helicobacter pylori*,[9,10] and *Drosophila melanogaster*.[11] Moreover, in recent years, the generation of proteome-scale maps of human proteins[12] or of *Epstein−Barr virus*,[13] etc., has been reported. In computational studies, Pazos et al.[14] have constructed PPI maps of *Escherichia coli* by using their own in silico two-hybrid, and von Mering et al.[15] have compared previously reported PPIs of yeasts at the genome scale. Recently, the study of PPI has been focused on not only ascertaining their roles in the living organism but also applying this knowledge to medicinal fields such as drug design. In essence, the interactive properties associated with protein conformation need to be evaluated if we are to analyze the relationships between protein structure and function. However, the most comprehensive PPI maps based on the primary structures of proteins still do not have enough conformational information for us to thoroughly investigate PPIs. Ideally, PPI studies for drug design should utilize protein structural information. This kind of PPI study has been discussed by Smith et al.,[16] but they did not clearly show the performance of an actual system or the analysis of the results. Here, we focused on the three-dimensional (3D) structures of proteins in order to calculate PPIs via protein−protein docking. Most of the previous docking studies[3,17−22] have attempted to solve protein-docking problems; the goal of these studies has been to accurately and rapidly predict the structures of protein complexes from the 3D structures of the individual proteins by e.g. reducing the search space or roughly modeling the proteins. Exhaustive calculations using a large number of PPIs can reveal useful protein pairs and improve the accuracy of identification of unknown proteins. However, because of the computational time and cost, no such large-scale, in silico PPI predictions based on structural information have been reported.

Hence, in order to predict protein−protein affinity, we previously proposed a method of assessing the biological

* Corresponding author phone: +81 3 3599 8060; fax: +81 3 3599 8076; e-mail: k-fukui@aist.go.jp.
[†] National Institute of Advanced Industrial Science and Technology (AIST).
[‡] Osaka University.

10.1021/ci800310f CCC: $40.75
Published on Web 02/10/2009

relevance of target proteins on the basis of the shape complementarity characteristic between protein pairs. We called this method *affinity evaluation and prediction* (AEP). The protein complex structures obtained from this shape complementarity evaluation are selected by our newly developed clustering method, called *grouping*. Although the exhaustive shape complementarity search involves a large amount of computational time, we have made the following improvements to our docking simulation, which is fast and highly accurate:[23] (1) In the modeling of proteins, we have developed a new representation method to discretize target proteins into 3D grids and classify them into core, surface, and cavity area. This method, called *voxelization*, has two distinguishing features, as follows. One is that the molecular surface can be modeled as the extended solvent-accessible surface,[24] which is traced out by the probe sphere center as it rolls over the protein. We defined expanded van der Waals (vdW) radii by the probe radius according to each atomic type, where the probe radius is half the length of each atom's vdW radius. The shape complementarity is made significantly more accurate than just made voxel models by slightly reducing the thickness of the modeled molecular surface. (2) In the shape complementarity evaluation, use of the Fast Fourier Transform (FFT) algorithm enables us to considerably reduce the number of docking calculations needed. The actual calculation can be performed on massive parallel computers by using the high-performance FFT library (CONV3D),[25] which is highly optimized for specific CPU architecture, achieves high-speed performance, and is about three times faster than the widely used FFT library package FFTW.[26,27] Moreover, with the massive parallel computer IBM Blue Gene (8192 CPU cores, 2-TB memory, and 22.8 TFlops/s peak performance) at CBRC, we have successfully parallelized our accelerated docking simulation.

From the results of some previous experiments,[23,28] AEP has been found to give different accuracy depending on the protein-pair data set, the data scale ($20 \times 20 = 400$ or $54 \times 54 = 2916$ protein pairs), the grouping, and the calculation of affinity scores. In this study, we set a data scale ($84 \times 84 = 7056$ protein pairs) including 84 biologically relevant complexes, and we then designed 225 parameter sets based on four key parameters (i.e., $D_{distance}$, $D_{orientation}$, $W_{conform}$, and W_{score}) related to the grouping and the calculation of affinity scores. The purpose of this study was to improve the accuracy of AEP by parameter optimization and to interpret AEP characteristics.

2. PROTEIN−PROTEIN AFFINITY PREDICTION

In order to predict protein−protein affinity, we evaluated the statistical significance of binding likelihood based on the shape complementarity characteristics between protein pairs by grouping. The procedures used for the affinity prediction and grouping are defined below.

2.1. Affinity Prediction Procedure. AEP consists of five procedures (shown below), as outlined in Figure 1.

• Step 1 (Shape complementarity evaluation):

The shape complementarity between protein pairs is evaluated by using the 3D structural information on two proteins, called the receptor and ligand (see paragraph 3.4). The procedure starts by rotating a ligand as a probe protein. When the pitch of the rotational angle is equal to 15 degrees, the number of rotated proteins is 3600 because of rotational symmetry.[29] A target protein (receptor) and 3600 probe proteins (ligands) are discretized into the 3D grids ($l \times m \times n$) by a certain grid size (1.2 Å), where l, m, and n are natural numbers (**N**). Each grid point is assigned a structural property value based on the geometric characteristics of the protein of interest. These processes were called *voxelization and scoring* in our previous study.[23] This voxelization process is a newly developed representation method used to classify each point of protein into core, surface, and cavity areas (i.e., INNER, SURFACE, and CAVITY). After the rotated probe protein is translated with respect to the target protein, the docking algorithm calculates the product sum of the structural property values, called the *shape complementarity score* (S_{PSC}). If the grid size of a target protein is equal to ($l \times m \times n$), then ($l \times m \times n$) $\times$ 3600 shape complementarity scores can be obtained by the entire calculation for only one protein pair. In this study, we selected the top-ranked N_s ($N_s \in \mathbf{N}$, $N_s \geq 1$) scores as candidate set $X = \{x_i \mid i \in \mathbf{N}, 1 \leq i \leq N_s\}$, where the number of samplings (N_s) is set to 512; these settings yielded the maximum prediction accuracy in some preliminary experiments. Here, i is the index for each candidate.

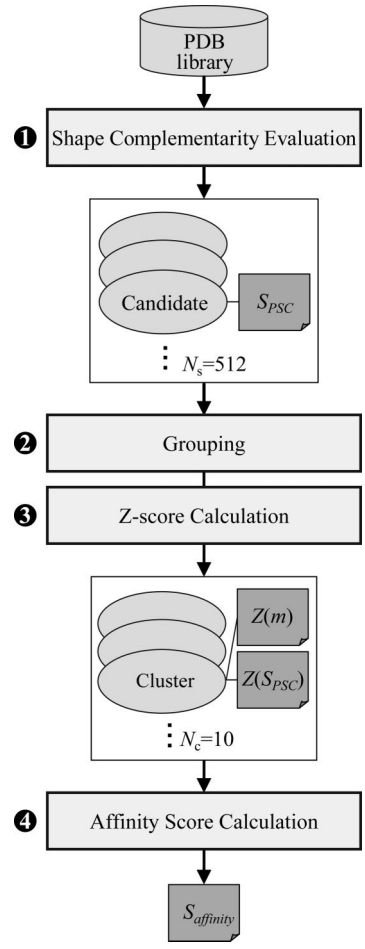

Figure 1. Flowchart for AEP, consisting four steps without an iteration procedure. After assessment of the shape complementarity score (S_{PSC}) between the 3D structures of two proteins in step 1, 512 S_{PSC} scores are classified into 10 clusters by grouping in step 2. In steps 3 and 4, an affinity score is calculated as a statistical binding likelihood based on the Z-scores of each representative's S_{PSC} score and of each cluster's density.

• Step 2 (Grouping):

The candidate set *X*obtained in step 1 is classified into clusters according to the similarity between protein structures. Here, a set of representatives y_j of each cluster c_j calculated by the grouping is defined as *Y*, where j ($j \in \mathbf{N}$, $1 \le j \le N_c$) is an index of each cluster and N_c ($N_c \in \mathbf{N}$, $N_c \ge 1$) is the number of clusters. We set the value of N_c to 10; this value had predicted accuracy very well in some preliminary experiments. The detailed steps for the grouping are shown in the following paragraph.

• Step 3 (Z-score calculation):

In this step, the statistical distribution $Z_{score}(y_j)$ and $Z_{conform}(c_j)$ are calculated, where $Z_{score}(y_j)$ is the Z-score of each cluster's representative $y_j \in Y$, of which the population is data set *X*. Then, $Z_{conform}(c_j)$ is the Z-score of the number of candidates including a representative in cluster c_j, m_j, when the data set $M = \{m_j\}$ is a population. The Z-score indicates the number of standard deviations by which an observation is above or below the mean. It enables the comparison of observations from different normal distributions.

• Step 4 (Affinity score calculation):

The measurement for affinity prediction is assessed on the basis of a statistical binding likelihood called the *affinity score* ($S_{affinity}$), which is represented as the mean value of Z_j above 3σ, defined in

$$\begin{cases} S_{affinity} = \dfrac{1}{N_{z_{Z_j>3}}} \displaystyle\sum^{N_z} Z_j \\ Z_j = W_{conform} \times Z_{conform}(c_j) + W_{score} \times Z_{score}(y_j) \end{cases} \quad (1)$$

where Z_j denotes the weighted combined Z-score of $Z_{score}(y_j)$ and $Z_{conform}(c_j)$. That is, our newly score is defined as the relative probability based on appearance frequency of candidates' structure of a ligand specific to the binding sites of a receptor. Here the weighting factors W_{score} and $W_{conform}$ for determining a balance between $Z_{score}(y_j)$ and $Z_{conform}(c_j)$ satisfy the tradeoff relationship as

$$\begin{cases} W_{conform} + W_{score} = 1 \\ 0 < W_{conform} < 1,\ 0 < W_{score} < 1 \end{cases} \quad (2)$$

For example, when the affinity score $S_{affinity}(R_1, L_1)$ of a protein pair (i.e., R_1 and L_1) is higher than that of $S_{affinity}(R_1, L_2)$, the former is evaluated as a protein pair with a significantly higher possibility of biological binding than the latter. Here, R and L are the receptor and ligand proteins, respectively, described in paragraph 3.4.

• Step 5 (Iteration of steps 1 to 4):

Steps 1 to 4 are repeated as for the required number of times. For example, if an affinity prediction is needed for 84 target and probe proteins (84 × 84), then this step is iterated 7056 times, since each target protein is checked by all 84 probe proteins.

AEP with the above flow is a kind of conversion process for calculating affinity scores (S_{PSC}) from shape complementarity scores ($S_{affinity}$). In this study, because the aim is to improve the accuracy of AEP, grouping and calculation of affinity scores are especially important in these five steps. Thus, the following paragraph gives the grouping steps in detail.

2.2. Grouping. Grouping is an original hierarchical clustering method based on a nearest-neighbor model and is used to improve the accuracy of AEP. The purpose of grouping is to classify the candidate set obtained in the shape complementarity evaluation into N_c (= 10) clusters. Grouping consists primarily of the following six steps.

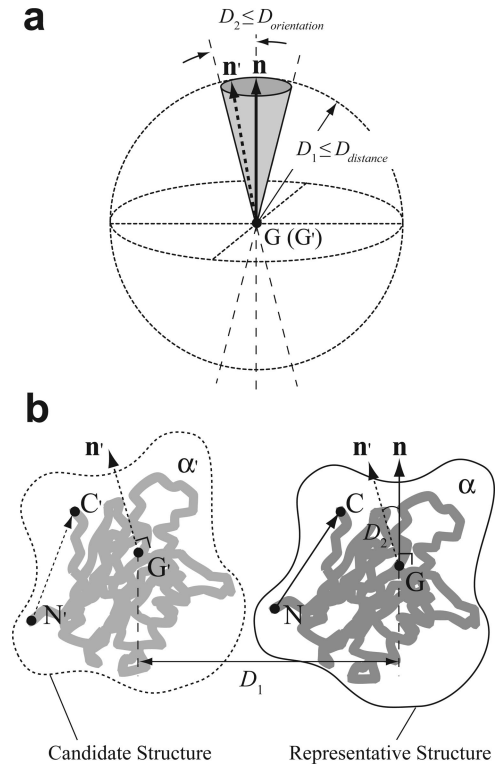

Figure 2. (a) The possible space for the normal vectors **n′** of candidates classified into a cluster. When the structural dissimilarities D_1 and D_2 are less than each threshold (i.e., $D_{distance}$ and $D_{orientation}$), respectively, the normal vector **n′** exists in the cone-shaped area of which the axis of rotation is the vector **n**. (b) The Euclidean translation distance D_1 and the rotational angle D_2 show the structural dissimilarities of a representative and some candidates in a cluster. The center of gravity of a representative is defined as point G, and the coordinates of the C_α atom in the C- and N-terminals are indicated as points C and N, respectively. Then **n** is a normal vector of a plane α including these three points (i.e., G, C, and N). Here G′, C′, N′, **n′**, and α′ are points of a candidate protein.

• Step 1 (Setup for candidate set *X*):

The procedure starts by hierarchical clustering of the shape complementarity scores by using a nearest-neighbor method, as follows. First, these N_s (= 512) scores obtained by the shape complementarity evaluation are set to the target data set $X = \{x_1,..., x_{512}\}$. For each data point, x_i is represented as $x_i = (S_i, T_i(T_x, T_y, T_z), R_i(R_\theta, R_\varphi, R_\psi))$, where S_i is the shape complementarity score (S_{PSC}) and $T_i(T_x, T_y, T_z)$ and $R_i(R_\theta, R_\varphi, R_\psi)$ are the translational distance and the rotational angle of the probe protein, respectively. All x data in *X* are sorted by descending order of S_i.

• Step 2 (Determination of representative y_j):

The datum $x_i \in X$, with a maximum of S_i, is defined as representative y_j of the j-th cluster c_j.

• Step 3 (Calculation of structural similarity between representative y_j and candidate x_i):

The distance $D_1(x_i, y_j)$ and the angle $D_2(x_i, y_j)$ between y_j and x_i ($\neq y_j$) are calculated as structural dissimilarities, *w*here x_i is not classified into any clusters. The distance D_1 is represented as a Euclidean translation distance, and the rotational angle D_2 is an angle between normal vectors **n** of x_i and **n′** of y_j, illustrated in Figure 2a. As shown in Figure

Table 1. Logical Combinations of Biological Relevance and Affinity Prediction

		biological relevance	
		true	false
affinity prediction	positive	true positive (TP)	false positive (FP)
	negative	false negative (FN)	true negative (TN)

Table 2. Evaluation Criteria for Affinity Prediction

cutoff	evaluation criteria	definition equation
[0, ∞]	AUC	$\int\{\text{sensitivity} \times (1 - \text{specificity})\}\, d(\text{cutoff})$
C_{max}[a]	sensitivity (= recall)	$n(\text{TP})/(n(\text{TP}) + n(\text{FN}))$
	specificity	$n(\text{TN})/(n(\text{FP}) + n(\text{TN}))$
	precision	$n(\text{TP})/(n(\text{TP}) + n(\text{FP}))$
	accuracy	$(n(\text{TP}) + n(\text{TN}))/(n(\text{TP}) + n(\text{FN}) + n(\text{FP}) + n(\text{TN}))$
	F-measure	$2 \times \text{precision} \times \text{recall}/(\text{precision} + \text{recall})$

[a] C_{max} is the cutoff value that maximizes the combination of sensitivity and specificity.

2b, point G denotes the center of gravity of a representative of predicted candidate (ligand) proteins, and points C and N indicate the coordinates of the C_α atom in the C- and N-terminals, respectively. Then **n** is a normal vector of plane α including these three points (i.e., G, C, and N). Here G′, C′, N′, **n**′, and α′ are the points of the candidate protein.

• Step 4 (Classification of candidate x_i into the j-th cluster c_j):

The candidate x_i is classified into the j-th cluster c_j when the structural dissimilarities D_1 and D_2 satisfy

$$\begin{cases} D_1(x_i, y_j) \le D_{distance} \\ D_2(x_i, y_j) \le D_{orientation} \end{cases} \quad (3)$$

where $D_{distance}$ and $D_{orientation}$ denote the thresholds of D_1 and D_2, respectively, as shown in

$$\begin{cases} D_{distance} > 0 \\ 0 < D_{orientation} < 180 \end{cases} \quad (4)$$

• Step 5 (Determination of the j-th cluster c_j):

The j-th cluster c_j is determined by repeating steps 3 to 4 for all the candidates $x \in X$ not classified into clusters.

• Step 6 (Determination of cluster set C):

Similarly, the procedure from steps 2 to 5 is repeated for all the clusters $c \in C$. This step ends either when N_c (= 10) clusters (c_1-c_{10}) are generated or when all the candidates x are classified into each cluster c_j.

3. EVALUATION EXPERIMENT

The purpose of the evaluation experiment was to verify the effectiveness of parameter optimization in improving the accuracy of AEP. We set the primary parts of our evaluation experiment—the affinity prediction problem, evaluation criteria, parameter sets, and bound-protein-pair data set—as follows.

3.1. Affinity Prediction Problem. The affinity prediction problem in this study was defined to statistically determine the biologically relevant protein pairs in a data set. The separation of these relevant protein pairs as positive examples from the others as negative is a binary classification problem. The scale of this problem can be represented by means of the ratio of positives in all examples. Table 1 shows the logical combinations of biological relevance and affinity prediction as a 2 × 2 contingency table. Here the biological relevance of each protein pair is based on established experimental results, and a positive or negative prediction indicates whether or not an AEP prediction result is statistically significant. The four logical combinations (i.e., TP, FN, FP, and TN) are defined in Table 1, and these numbers are represented as n(TP), n(FN), n(FP), and n(TN), respectively. The content ratio of biologically relevant pairs in a data set (i.e., the prevalence), independent of affinity prediction, is given by

$$\begin{aligned} prevalence = & \\ & (number\ of\ biologically\ relevant\ pairs)/ \\ & (number\ of\ all\ protein\ pairs) = \\ & (n(TP) + n(FN))/(n(TP) + n(FN) + \\ & n(FP) + n(TN)) \end{aligned} \quad (5)$$

3.2. Evaluation Criteria. In general, the evaluation criteria for information retrieval shown in Table 2 are used for assessing the prediction accuracy of a binary classification problem. For example, an area under the curve (AUC) can be used to measure the robustness independently of the cutoff values obtained from the receiver operating characteristics (ROC) analysis. The criterion of sensitivity is represented as the proportion of true positives in biologically relevant pairs: that is, a prediction method with high sensitivity can reveal the most biologically relevant pairs. Since the criterion of specificity is defined as the proportion of true negatives in protein pairs that are not biologically relevant, a prediction method with high specificity can efficiently distinguish protein pairs that are not biologically relevant. Then, recall (= sensitivity) for assessing exhaustive accuracy is defined as the proportion of correctly predicted protein pairs in all the biologically relevant pairs. Moreover, precision, which is the degree of reproducibility, shows the ratio of biologically relevant pairs in all the prediction results. Even if both sensitivity and specificity are quite high, the precision generally takes a low value if there is a problem with low prevalence. The criterion of "accuracy" represents the proportion of the number of correct predictions of biologically relevant or nonrelevant pairs in all the prediction results. Finally, the F-measure is defined as the harmonic mean of recall and precision, and its value becomes much higher as the values of both of these factors increase. Considering these attributes of the criteria used, we employed the F-measure to assess AEP. Since the F-measure can quantitatively gauge prediction accuracy to the prevalence of a problem, we can evaluate the recall and precision as a tradeoff in the form of only one combined value:

$$F\text{-}measure = \frac{2 \times precision \times recall}{precision + recall} \quad (6)$$

Incidentally, methodologies have not been established for PPI studies based on structural information, and results for comparison with our prediction accuracy have therefore not been published. In this study, we evaluated AEP on the basis of the rate of increase in prediction accuracy by random sampling or tentative prediction methods, using docking scores from ZDOCK 3.0.1[30,31] or scores reranked by using an optimum energy function from ZRANK.[32] When a

Table 3. Parameters and Their Values Used To Optimize Affinity Prediction

parameter	value
$D_{distance}$ (Å)	1, 5, 10, 15, 20
$D_{orientation}$ (degrees)	5, 15, 30, 45, 60
($W_{conform}$, W_{score})	(0.00, 1.00), (0.05, 0.95), (0.10, 0.90), (0.15, 0.85), (0.20, 0.80), (0.25, 0.75), (0.50, 0.50), (0.75, 0.25), (1.00, 0.00)

random sampling was applied to a prediction problem based on the data set of 7056 bound-protein pairs described in paragraph 3.4, the values of the criteria were as follows: 50% sensitivity^{random} (= recall^{random}); 50% specificity^{random}; 1.2% precision^{random} (= prevalence); 50% accuracy^{random}; 2.3% (= $2 \times 0.5 \times 0.0119/(0.5 + 0.0119)$) F-measure^{random}; and AUC^{random} 0.5. Our preliminary experiments showed that the values for the criteria using ZDOCK 3.0.1 scores were as follows: 73.8% sensitivity^{ZDOCK}(= recall^{ZDOCK}); 57.8% specificity^{ZDOCK}; 2.1% precision^{ZDOCK}(= prevalence); 58.0% accuracy^{ZDOCK}; 4.0% F-measure^{ZDOCK}; and AUC^{ZDOCK} 0.698. Moreover, the prediction accuracies of the ZRANK scores were as follows: 72.6% sensitivity$^{ZDOCK+ZRANK}$(= recall$^{ZDOCK+ZRANK}$); 67.5% specificity$^{ZDOCK+ZRANK}$; 2.6% precision$^{ZDOCK+ZRANK}$(= prevalence); 67.6% accuracy$^{ZDOCK+ZRANK}$; 5.1% F-measure$^{ZDOCK+ZRANK}$; and AUC$^{ZDOCK+ZRANK}$ 0.738. Here, in applying the ZDOCK 3.0.1 program execution option, all the default values (i.e., -N (= 2000), -S (= no at randomization), and -D (= none)) are used. Then the affinity prediction accuracies of ZDOCK 3.0.1 or ZRANK are evaluated by using the maximum value among each 2000 scores. All the previous evaluations of AEP using 20×20 or 54×54 pairs were shown in the Supporting Information.

3.3. Parameter Sets. We generated parameter sets to maximize the prediction accuracy (i.e., F-measure) of AEP. In general, although expanding the search range for optimization allows us to raise the prediction accuracy, the total number of parameter sets explodes with the increase in the number of parameters employed and in the numbers of their combinations at each setting value. Hence we have focused only on grouping and the calculation of affinity scores, which are the central procedures of AEP, as described in section 2. The four key parameters (i.e., $D_{distance}$, $D_{orientation}$, $W_{conform}$, and W_{score}) in these two procedures are set within the range of search values shown in Table 3. Here, $D_{distance}$ and $D_{orientation}$ defined in the grouping are the thresholds for assessing the structural dissimilarities between each set of candidates; they respectively represent the distance between the centers of gravity and the deflection difference of a cluster's representative structure and the candidates. Moreover, W_{score} and $W_{conform}$, defined in the calculation of affinity scores, respectively represent the weighting factors of the shape complementarity scores' distribution for each cluster's representative (Z_{score}) and the clusters' density distribution ($Z_{conform}$).

In the evaluation experiment, we designed 225 parameter sets P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) by combining the above four key parameters. Our policy of setting parameter values, as shown in Table 3, was based on the results of preliminary experiments. More specifically, in the experiment using 144 parameter sets consisting of $D_{distance} = \{5, 10, 15, 20\}$, $D_{orientation} = \{15, 30, 45, 60\}$, and ($W_{conform}$, W_{score}) = {(1, 0), (0.75, 0.25), (0.5, 0.5), (0.25, 0.75), (0, 1)}, we obtained a high level of accuracy at the end points of each search range. Hence, we reset the range chosen to the above 225 parameter sets by subdividing the interval of setting values near the sets that had shown high accuracy. Then, in order to determine the search range of $D_{distance}$, we investigated all the radii of the smallest enclosing ball of 84 ligands. Moreover, the upper limit of the search range of $D_{distance}$ was set to 20 Å, since the 25th percentile of the above 84 radii was 21.81 Å. That is, candidate (ligand) pairs in which the distance between the centers of gravity was less than each other's radii were classified into the same cluster. However, in the previous experiments using data sets of 400 or 2916 bound-protein pairs,[23,28] we set values of P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = {10, 30, 0.25, 0.75} by roughly optimization for combining two parameters of $W_{conform}$ and W_{score} with the values of $D_{distance}$ and $D_{orientation}$ respectively fixed to 10 and 30.

Table 4. 84 Target Protein Complexes

category[a] receptor/ ligand	complex[b]
antibody/antigen (A)	1AHW(0), 1BGX(1), 1BVK(2), 1DQJ(3), 1E6J(4), 1JPS(5), 1MLC(6), 1VFB(7), 1WEJ(8), 2VIS(9)
bound antibody/ antigen (AB)	1BJ1(10), 1FSK(11), 1I9R(12), 1IQD(13), 1K4C(14), 1KXQ(15), 1NCA(16), 1NSN(17), 1QFW(18), 2HMI(19), 2JEL(20), 2QFW(21)[c]
enzyme/inhibitor or substrate (E)	1ACB(22), 1AVX(23), 1AY7(24), 1BVN(25), 1CGI(26), 1D6R(27), 1DFJ(28), 1E6E(29), 1EAW(30),1EWY(31), 1EZU(32), 1F34(33), 1HIA(34), 1KKL(35), 1MAH(36), 1PPE(37), 1TMQ(38), 1UDI(39), 2MTA(40), 2PCC(41), 2SIC(42), 2SNI(43), 7CEI(44)
others (O)	1A2K(45), 1AK4(46), 1AKJ(47), 1ATN(48), 1B6C(49), 1BUH(50), 1DE4(51), 1E96(52), 1EER(53), 1F51(54), 1FAK(55), 1FC2(56), 1FQ1(57), 1FQJ(58), 1GCQ(59), 1GHQ(60), 1GP2(61), 1GRN(62), 1H1V(63), 1HE1(64), 1HE8(65), 1I2M(66), 1I4D(67), 1IB1(68), 1IBR(69), 1IJK(70), 1K5D(71), 1KAC(72), 1KLU(73), 1KTZ(74), 1KXP(75), 1M10(76), 1ML0(77), 1N2C(78), 1QA9(79), 1RLB(80), 1SBB(81), 1WQ1(82), 2BTF(83)

[a] A, AB, E, and O are abbreviations for the respective categories. [b] Each complex is assigned to a PDB ID, and its corresponding index from 0 to 83 is shown in parentheses. [c] 2QFW indicates 1QFW_HL:AB, where HL and AB are names of the receptor and ligand chains, respectively.

3.4. Bound-Protein-Pair Data Set. We made a data set consisting of 168 bound proteins from the Protein−protein Docking Benchmark 2.0 of Weng et al.[33,34] All proteins were classified into receptor (R) or ligand (L) according to Weng's definition, giving 84 of each type of protein. The total number of pairs in the data set was 7056 (receptor × ligand = 84×84), including 84 biologically relevant pairs (Table 4). These 84 complexes were classified as follows: 10 antibody−antigen (category A), 12 bound antibody−antigen (AB), 23 enzyme−inhibitor or substrate (E), and 39 other (O). Each of the complexes was assigned both a protein database identifier (PDB ID) and an index from 0 to 83. Table 4 shows the classification of the 84 PDB IDs. Whereas in the previous experiments the prevalence of biologically relevant pairs was 5.0% (= 20/400) or 1.9% (= 54/2916), the prevalence in

Table 5. Results of Evaluation for Optimizing Affinity Prediction

	evaluation criteria					
descriptive statistics	AUC	sensitivity (= recall)	Specificity	precision	accuracy	F-measure
average	0.577	0.508	0.637	0.020	0.636	0.037
standard deviation	0.028	0.219	0.227	0.006	0.222	0.010
median	0.587	0.464	0.673	0.017	0.671	0.033
maximum	0.619	0.952	0.935	0.036	0.926	0.063
P($D_{distance}$,$D_{orientation}$, $W_{conform}$,W_{score})	P(20, 5, 0.25, 0.75)	P(5, 15, 0.25, 0.75) P(1, 45, 0.50, 0.50) P(1, 30, 0.75, 0.25)	P(1, 5, 0.05, 0.95)	P(1, 5, 0.05, 0.95)	P(1, 5, 0.05, 0.95)	P(5, 15, 0.00, 1.00)
minimum	0.490	0.155	0.115	0.012	0.125	0.024
P($D_{distance}$,$D_{orientation}$, $W_{conform}$,W_{score})	P(1, 5, 1.00, 0.00)	P(1, 5, 0.75, 0.25)	P(1, 45, 0.50, 0.50)	P(1, 5, 1.00, 0.00)	P(1, 45, 0.50, 0.50)	P(1, 5, 1.00, 0.00)
kurtosis	0.256	−1.066	−0.914	−0.257	−0.911	−0.434
skewness	−1.070	0.443	−0.525	1.000	−0.525	0.918

this study was 1.2% (= 84/7056), as obtained from eq 5. Here, all the proteins were approximated by a rigid-body model for high-speed searching of some of the binding sites.

4. RESULTS AND DISCUSSION

4.1. Results of Evaluation of Affinity Prediction. To verify the accuracy improvement of AEP by parameter optimization, we performed the evaluation experiment described in the previous section. The aim of parameter optimization is to maximize the F-measure of AEP. Table 5 lists the results of the evaluation of 225 parameter sets; all the values for each evaluation result are indicated by descriptive statistics of 225 results. AEP gave a prediction accuracy value of 6.3% for the maximum F-measure (F-measure^{max}) (Table 5). This was about 2.7, 1.6, and 1.2 times better, respectively, than the values from random sampling (F-measure^{random} = 2.3%) and from tentative prediction methods using ZDOCK 3.0.1 (F-measure^{ZDOCK} = 4.0%) or ZRANK (F-measure$^{ZDOCK+ZRANK}$ = 5.1%) scores, when using an optimum parameter set of P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00). In addition, although the value of the F-measure could not be directly compared among different data scales (i.e., 20 × 20, 54 × 54, or 84 × 84 protein pairs), we found that the rate of improvement in prediction accuracy against random sampling increased significantly: that is, the 2.7 improvement rate with 84 × 84 pairs was greater than the 2.3 and 1.8 rates with 20 × 20 and 54 × 54 pairs, respectively. This indicates that the prediction accuracy of AEP was robust against fluctuations in data scales. To show a clear advantage of AEP using 84 × 84 pairs, we compared the results of evaluation by different programs and data scales, in the Supporting Information S-1. The other evaluation values for the maximum F-measure, as shown in Table 2, were AUC of 0.585, 27.4% sensitivity (= recall), 91.0% specificity, 3.5% precision, and 90.2% accuracy.

This improvement of the F-measure over that in ZDOCK 3.0.1 or ZRANK was derived from the high precision of the method. When the prediction accuracies of ZDOCK 3.0.1 were 73.8% for recall^{ZDOCK} and 2.1% for precision^{ZDOCK} (see paragraph 3.2), the numbers of true positives and false positives were n(TP)^{ZDOCK} = 62 and n(FP)^{ZDOCK} = 2941, respectively. In the case of ZRANK, these numbers were n(TP)$^{ZDOCK+ZRANK}$ = 61 and n(FP)$^{ZDOCK+ZRANK}$ = 2263. Since the prediction accuracy of AEP was n(TP) = 23 and n(FP) = 629, AEP can eliminate many false positives even if the recall measurably decreases. These results indicate that a combination of clustering of the candidates obtained from the shape complementarity evaluation and calculation of affinity scores is effective in gaining high prediction accuracy. That is, after we have classified candidates for which the distance between the centers of gravity ($D_{distance}$) is less than or equal to 5 Å and the deflection difference from each cluster's representative structure ($D_{orientation}$) is less than or equal to 15 degrees into the same cluster, we can calculate the affinity scores from the statistics on only those clusters that have representatives with shape complementarity scores over 3σ. In general, the existing probability for the scores over 3σ, according to a normal distribution with a standard deviation of σ, is 99.7%. For this reason, there is only a 0.3% probability that representatives with shape complementarity scores over 3σ will be specifically included among all the candidates. Moreover, the value of 5 Å covers about a quarter of the mean radius of candidates—that is, it is approximately equal to the size for four grids when a grid of a certain size (1.2 Å) is used in the voxelization of shape complementarity evaluation. Meanwhile, since the minimum value of the F-measure (F-measure^{min} = 2.4%) is slightly better than that from a random sampling (F-measure^{random} = 2.3%), AEP can accurately predict biologically relevant pairs independently of selection of the combination of parameters in all 225 parameter sets.

Thus although AEP gives different accuracies depending on the parameter set, in practice our results have significant differences of up to about 2.6 (= 0.063/0.024) times when 225 parameter sets are used. In the following, we closely investigated the combination of parameter sets that would yield highly accurate results. First, we divided a total of 225 parameter sets into nine blocks (B_1−B_9) so that we could analyze the changes in F-measure in each block. A block consisted of 25 parameters of which the value of $W_{conform}$ (or W_{score}) was the same (Table 6). All the blocks were assigned to indexes (1−9) in ascending order of W_{score}. By defining these nine blocks, we could investigate the optimum balance between the distribution of the shape complementarity scores of each cluster 's representative (Z_{score}) and that of the clusters' density ($Z_{conform}$) to obtain more accurate results. Figure 3 shows boxplots of the F-measures in the nine blocks. The longitudinal and quadrature axes denote the F-measure descriptive statistics (i.e., the smallest observation, lower quartile, median, upper quartile, largest observation, and outlier) and the parameter set blocks, respectively. Moreover, the dashed line parallel to the quadrature axis represents the F-measure of a random sampling (F-measure^{random} = 0.023), and the polygonal line shows fluctuations in the median F-measure in each block. Since

Table 6. Parameter Set Blocks Grouped by W_{score} (or $W_{conform}$)

parameter set block	parameter set[a] ($W_{conform}$, W_{score})
B_1	(1.00, 0.00)
B_2	(0.75, 0.25)
B_3	(0.50, 0.50)
B_4	(0.25, 0.75)
B_5	(0.20, 0.80)
B_6	(0.15, 0.85)
B_7	(0.10, 0.90)
B_8	(0.05, 0.95)
B_9	(0.00, 1.00)

[a] The 25 parameter sets in each block related to structural dissimilarities consist of a combination of $D_{orientation}$ = {1, 5, 10, 15, 20} and $D_{distance}$ = {5, 15, 30, 45, 60}.

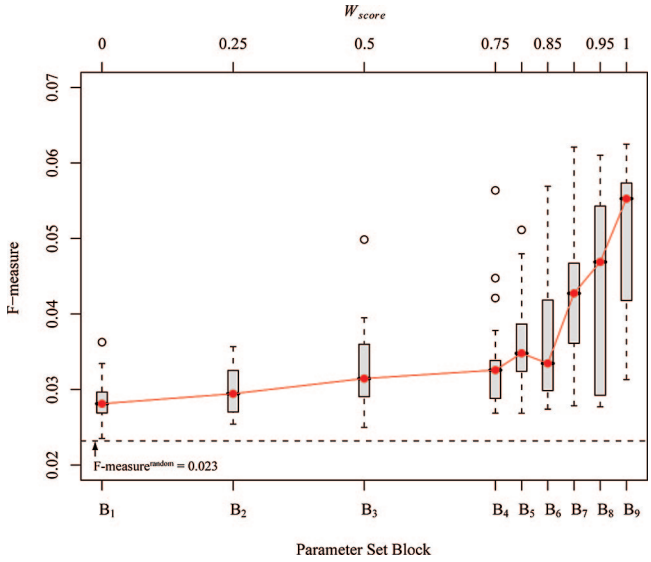

Figure 3. Boxplots of F-measures in nine blocks (B_1−B_9) consisting of 25 parameter sets of which the value of W_{score} (or $W_{conform}$) is the same. All the blocks are assigned to indexes (1−9) in ascending order of W_{score}. The longitudinal and quadrature axes denote the descriptive statistics of the F-measure (i.e., the smallest observation, lower quartile, median, upper quartile, largest observation, and outlier) and the parameter set blocks, respectively. Dashed line parallel to the quadrature axis represents the F-measure of a random sampling (F-measure^{random} = 0.023), and the polygonal line shows the fluctuations in the median F-measure in each block.

the F-measure values in each block are not normally distributed, according to the skewness of −0.9−2.1 and kurtosis of −1.6−5.8, we employed the median F-measure to assess the prediction accuracy of each block. The results (Figure 3) indicated that the median F-measure became much higher as the value of W_{score} increased. This shows that the prediction accuracy of AEP improved when W_{score} was set to a high value. In particular, the median F-measure in three blocks (B_7−B_9) was dramatically greater than those in the other blocks (B_1−B_6). In fact, the average median F-measure in B_7−B_9 was 50.3%−1.6 times that in B_1−B_6 (31.6%). That is, we found that AEP could give a high level of accuracy if we used the optimum parameters included in blocks B_7−B_9. However, in light of the definition of blocks (Table 6), selecting parameter sets in B_7−B_9 meant setting the value of W_{score} to 0.90, 0.95, or 1.00 (i.e., setting the value of $W_{conform}$ to 0.10, 0.05, or 0.00). This suggests that we obtained more accurate results because of an emphasis on the distribution of the shape complementarity scores of each

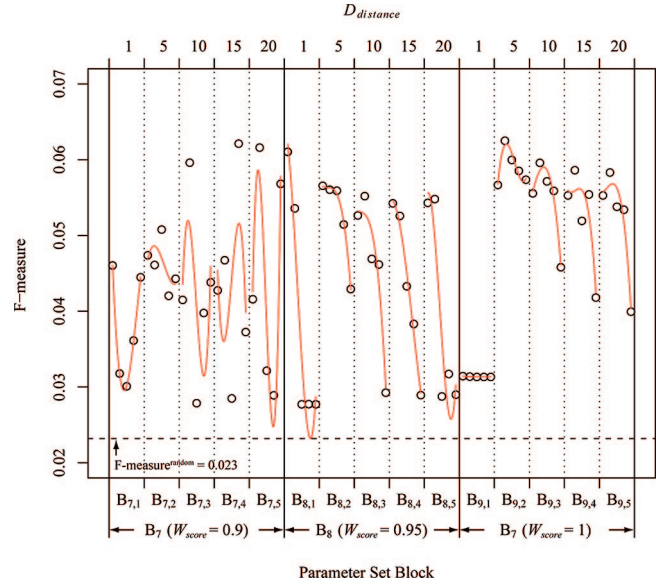

Figure 4. Prediction accuracies for F-measures in blocks B_7−B_9. The longitudinal and quadrature axes denote F-measure and 75 parameter sets included in blocks B_7−B_9, respectively. The solid and dotted lines parallel to the longitudinal axis denote the borders of blocks and of subblocks, respectively. All the blocks are assigned to indexes (1−9) in ascending order of W_{score}, and the indexes of subblocks are arranged in ascending order of $D_{distance}$. For example, the first subblock of B_8, described as $B_{8,1}$, contains five parameter sets: P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (1, 5, 0.05, 0.95), (1, 15, 0.05, 0.95), (1, 30, 0.05, 0.95), (1, 45, 0.05, 0.95), (1, 60, 0.05, 0.95). A block consists of five subblocks including parameter sets of which the value of $D_{distance}$ is the same. The curves are regression curves obtained from polynomial regression analysis for five F-measure values in each subblock.

cluster's representative (Z_{score}) rather than on that of the cluster's density ($Z_{conform}$).

We focused on interpreting the results from blocks B_7−B_9, which were all relatively accurate. Here, the combination value of $D_{distance}$ and $D_{orientation}$ was particularly analyzed to give a more accurate result. Figure 4 shows the prediction accuracy of blocks B_7−B_9; the longitudinal and quadrature axes denote the F-measure and the 75 parameter sets included in blocks B_7−B_9, respectively. The solid and dotted lines parallel to the longitudinal axis respectively denote the borders of the blocks and of the subblocks, described below. The assignment of block indexes is the same as shown in Figure 3, and the subblock index is arranged in ascending order of $D_{distance}$. For example, the first subblock of B_8, described as $B_{8,1}$, contains five parameter sets, P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (1, 5, 0.05, 0.95), (1, 15, 0.05, 0.95), (1, 30, 0.05, 0.95), (1, 45, 0.05, 0.95), and (1, 60, 0.05, 0.95). Here, the parameter sets on the quadrature axis are regularly arrayed according to the definition of blocks (Table 6). That is, one block consists of five subblocks, including parameter sets of which the value of $D_{distance}$ is the same. Each curve is a regression curve obtained from polynomial regression analysis of the five F-measure values in each subblock. High F-measure values were observed in blocks B_7−B_9, in which the value of $D_{orientation}$ was set to 15 degrees, and this trend was especially prominent in blocks B_8−B_9 (Figure 4). This observation suggests that setting the value of $D_{orientation}$ to 15 degrees produced prediction results that were highly independent of the value of $D_{distance}$. In blocks B_8−B_9, the median F-measure using parameter sets

Table 7. Median F-measures in Blocks B_8-B_9 ($D_{orientation} = 15$)

parameter set[a] ($D_{distance}$, $D_{orientation}$)	median F-measure
(1, 15)	0.042
(5, 15)	0.059
(10, 15)	0.057
(15, 15)	0.056
(20, 15)	0.057

[a] The two parameter sets related to weighting factors are ($W_{conform}$, W_{score}) = (0.05, 0.95) or (0.00, 1.00).

of $D_{orientation} = 15$ degrees was 5.6%; this was higher than the value obtained by using any other values of $D_{orientation}$. Therefore, when there is an emphasis on the value of Z_{score} rather than that of $Z_{conform}$, the grouping should be optimized so that candidates with deflection differences between each representative ($D_{orientation}$) of less than or equal to 15 degrees are classified into the same clusters.

The five F-measure values in subblock $B_{9,1}$ were very similar to each other, within the range from 31.3% to 31.4%, regardless of the value of $D_{orientation}$. This occurred because of similarities in n(TP) = 53 and n(FP) = 3235−3246 for all five subblocks. When all 7056 affinity scores for each prediction result in subblock $B_{9,1}$ were investigated, we found that more than 97.5% of scores were the same. One reason is that the value of $D_{distance}$ in the grouping was set extremely low (e.g., to 1 Å). This biased grouping generated mostly small clusters with fewer than 10 candidates; therefore, the shape complementarity score of representatives dominated over the density of each cluster. Moreover, since the 512 candidates were sorted by descending order of shape complementarity scores in the first step of grouping, when clusters with few candidates were generated the same candidates tended to be classified as representatives. The additional reason for the similarity between the values in subblock $B_{9,1}$ was that we calculated the affinity scores without the density of each cluster by using the parameter set P($W_{conform}$, W_{score}) = (0, 1).

As a further analysis, we investigated the influence of the value of $D_{distance}$ on the prediction accuracy in blocks B_8-B_9, where highly accurate results were obtained. Table 7 shows the median F-measure obtained when we used parameter sets in blocks B_8-B_9 with the value of $D_{orientation}$ set to 15 degrees. As shown in Table 7, the median F-measure using parameter sets of $D_{distance}$ (= 5 Å) was 6.0%—higher than that obtained by using any other values of $D_{distance}$ in all the 10 parameter sets of blocks B_8-B_9. Thus, using the optimum parameter sets in B_8-B_9, P($D_{distance}$, $D_{orientation}$) = (5, 15), gave the highest accuracy out of the entire 225 parameter sets. That is, after we classified candidates with distances between the centers of gravity ($D_{distance}$) less than or equal to 5 Å and deflection differences from each representative ($D_{orientation}$) of 15 degrees into the same cluster and then calculated the affinity scores with an emphasis on the distribution of the shape complementarity scores of each cluster's representative (Z_{score}) rather than of the cluster density ($Z_{conform}$), AEP provided extremely accurate predictions.

4.2. Protein−Protein Interaction Map. One of the purposes of AEP is to statistically distinguish biologically relevant pairs out of a data set with high accuracy and to thereby generate a reliable PPI map. First, we used ROC analysis to determine whether the protein pairs predicted by

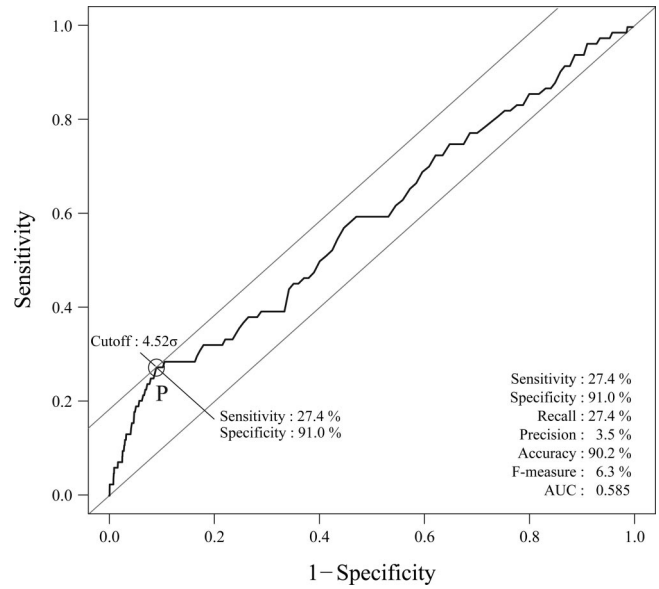

Figure 5. ROC curve of prediction results obtained by AEP when using the optimum parameter set P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00) that maximizes the F-measure. The diagonal line represents the accuracy result from a completely random sampling. Since this line divides the ROC space into areas of accurate and wrong classification, the points above and below the line are the accurate and wrong prediction results, respectively. Point P denotes the maximum accuracy of AEP at the optimum cutoff value ($C_{max} = 4.52\sigma$), and their values are shown in the bottom-right corner.

AEP were statistically significant (positive) or nonsignificant (negative). Figure 5 is an ROC curve of the prediction results obtained by using the optimum parameter set P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00) to maximize the F-measure; the diagonal line represents the accuracy of the results compared with a completely random sampling. Since this line divides the ROC space into areas of accurate or wrong classification, the points above and below the line show the accurate and wrong prediction results, respectively. Here, in order to closely plot this curve, we used the Epi package,[35] which is a program for the language and environment for statistical computing and graphics called R.[36] As shown in Figure 5, the curve lies in only the area above the diagonal line, and the value of AUC for assessing robustness is 0.585 over a random sampling ($AUC^{random} = 0.5$). That is, independently of the use of a cutoff value for affinity scores, AEP optimized to achieve the maximum F-measure provided a more accurate result than random sampling. We then evaluated the extremal accuracy of AEP. Point P in Figure 5 denotes the maximum accuracy, which occurred at 27.4% sensitivity (= recall), 91.0% specificity, 3.5% precision, 90.2% accuracy, and 6.3% $F\text{-measure}^{max}$ ($C_{max} = 4.52\sigma$). Here, the optimum cutoff value (C_{max}) is the threshold for determining positives out of the entire 7056 affinity scores. When the cutoff value is set to C_{max}, the value of $\{\text{sensitivity}^2 + (1 - \text{specificity})^2\}$ becomes the maximum.

Next, by setting the optimum cutoff value ($C_{max} = 4.52\sigma$), we generated a PPI map that showed the maximum accuracy of AEP. Figure 6 is a PPI map based on a matrix of 7056 affinity scores, obtained by using the optimum parameter set P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00). Each point is colorized: the red and purple ends of the scale bar indicate high- and low-affinity scores, respectively. The longitudinal and quadrature axe denote the indexes of the

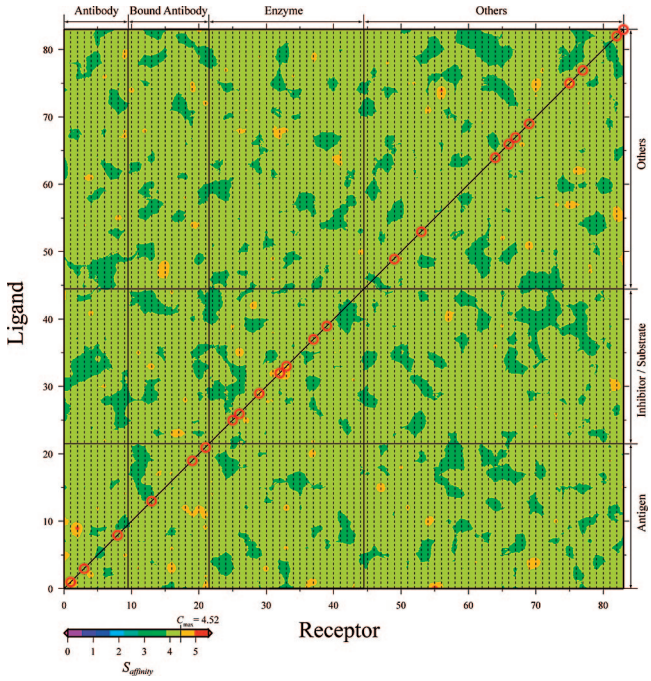

Figure 6. PPI map based on a matrix of 7056 affinity scores obtained by AEP when using the optimum parameter set P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00) that maximizes the F-measure. Each point is colorized; the red and purple ends of the scale bar indicate the high and low affinity scores, respectively. The longitudinal and quadrature axes respectively denote the indexes of the ligand (probe) and receptor (target) protein. The relationships between PDB IDs of complexes and their corresponding indexes are shown in Table 4. The straight lines parallel to either the longitudinal or the quadrature axis show the borders of each category, based on the biological functions of the complexes. The encircled points along the diagonal line indicate affinity scores above the optimum cutoff value ($C_{max} = 4.52\sigma$).

ligand (probe) and receptor (target) protein, respectively. The classification of the complexes is shown in Table 4. The straight lines parallel to the longitudinal or quadrature axis show the borders of each category, based on the biological functions of the complexes. The encircled points along the diagonal line indicate the affinity scores above the optimum cutoff value ($C_{max} = 4.52\sigma$). Since the sensitivity value was 27.4%, as mentioned above, 23 protein pairs out of 84 biologically relevant pairs were predicted as statistically significant. These complexes were classified as follows: three out of 10 antibody−antigen (category A), three out of 12 bound antibody−antigen (AB), seven out of 23 enzyme−inhibitor (E), and 10 out of 39 other (O). As the maximum difference of the true positive rate in each category was only 5.4%, we revealed that the prediction accuracy of AEP for the 84 complexes did not depend on the biological function categories of the target protein pairs. Therefore, by optimizing some of the key parameters related to the grouping and the calculation of affinity scores, we found that AEP could successfully predict 23 statistically significant pairs out of 84 complexes.

4.3. Active Sites of High-Affinity Complexes. As mentioned in section 1, in PPI studies intended for application to medicinal fields such as drug design, it is necessary to predict the active sites of complexes and to find protein pairs with high affinity in order to analyze the relationships between protein structure and function. For this reason, we attempted to verify whether the 3D structures of the protein

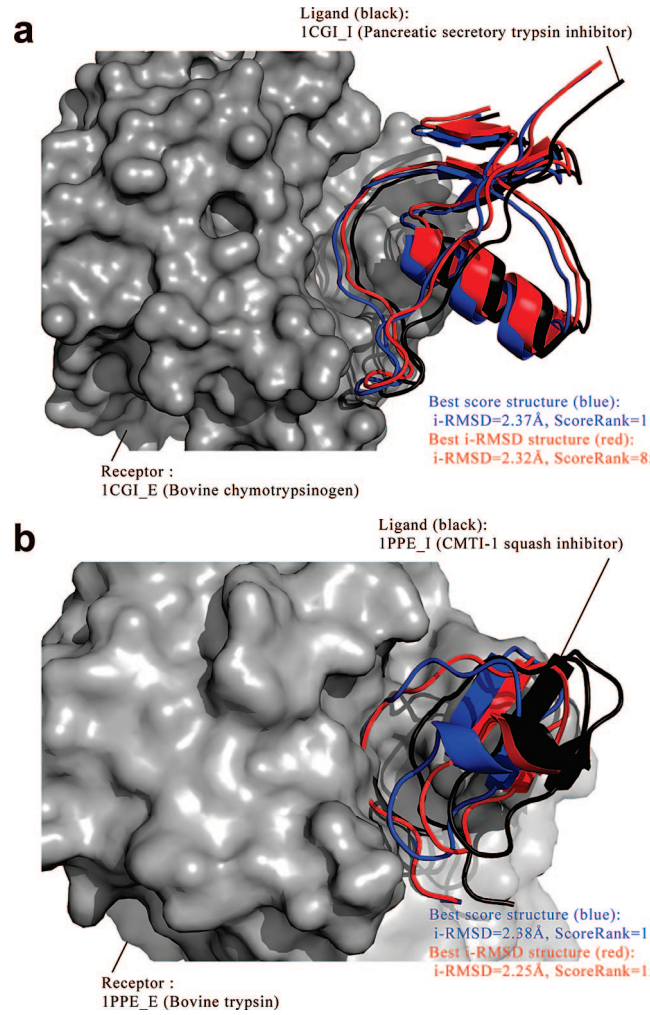

Figure 7. Active sites of 1CGI(index 26) and 1PPE(37) with high affinity (i.e., $S_{affinity}$ of 1CGI and 1PPI is 4.97σ and 4.92σ, respectively) above the optimum cutoff value ($C_{max} = 4.52\sigma$). (a) 1CGI(26) is a complex between a proenzyme of 2CGA_B (bovine chymotrypsinogen) and an inhibitor of 1HPT_ (pancreatic secretory trypsin inhibitor). (b) 1PPE(37) consists of an enzyme of 1BTP_ (bovine trypsin) and an inhibitor of 1LU0_A (CMTI-1 squash inhibitor). Both complexes are classified into category E (i.e., the receptor and ligand proteins are an enzyme and an inhibitor, respectively). The crystal structure of the receptor is illustrated by a molecular surface model, and the black, blue, and red cartoon models show the ligand of the crystal, the highest shape complementarity, and the nearest native structure, respectively.

complexes obtained by AEP were correctly predicted. We investigated two complexes out of 23 statistically significant pairs predicted by AEP. Figure 7 shows the active sites of 1CGI (index 26)[37] and 1PPE (index 37)[38] with high affinity (i.e., $S_{affinity}$ of 1CGI and 1PPI was 4.97σ and 4.92σ, respectively) above the optimum cutoff value ($C_{max} = 4.52\sigma$). 1CGI(26) is a complex between a proenzyme of 2CGA_B (bovine chymotrypsinogen)[39] and an inhibitor of 1HPT_ (pancreatic secretory trypsin inhibitor);[40] 1PPE(37) consists of an enzyme of 1BTP_ (bovine trypsin)[41] and an inhibitor of 1LU0_A (CMTI-1 squash inhibitor).[42] Both complexes are classified into category E (i.e., the receptor and ligand protein is an enzyme and an inhibitor, respectively) (Table 4). The crystal structure of the receptor is illustrated as a molecular surface model, and the black, blue, and red cartoon models show the ligand of the crystal, the highest shape complementarity, and the nearest native structure, respec-

tively. Here, we employed the interface root-mean-square deviation (i-RMSD) to assess the predicted complex structure. The i-RMSD is defined as the backbone RMSD from the reference structure of the complex for those residues making contacts across the interface within a 10-Å cutoff. In the above test case of 1CGI(26), the evaluation value of a structure with the maximum affinity score is 2.37 Å of i-RMSD (rank = 1). The best (minimum) value is 2.32 Å of i-RMSD^{min} (rank = 85), obtained from the 85th candidate structure in descending order of shape complementarity scores. In 1PPE(37), the top-ranked i-RMSD (rank = 1) is 2.38 Å, and the best i-RMSD^{min} (rank = 15) is 2.25 Å. This indicates that the top-ranked i-RMSDs are slightly worse (by 0.35 Å or 1.94 Å) than the best i-RMSD^{ZDOCK} from the ZDOCK program (1.2 Å of grid size and 6 degrees of rotational angle are used, but both the score rank and program version are not shown in the benchmark results[34]). However, since each evaluation value is better than a well-known threshold (= 2.5 Å) in protein−protein docking studies, these complexes are closely similar to the native structure. That is, the prediction is considered acceptable according to the critical assessment of predicted interactions (CAPRI) criteria.[43] These results suggest that AEP can correctly predict the active sites of protein−protein complexes as well as statistically determine biologically relevant protein pairs with high accuracy.

5. CONCLUSIONS

In this study, which was intended to improve the prediction accuracy of AEP, we investigated the effectiveness of optimizing key parameters in grouping and in the calculation of affinity scores. We set the evaluation experiment to a data scale (84 × 84 = 7056 protein pairs) that included 84 biologically relevant pairs and 225 parameter sets consisting of four key parameters ($D_{distance}$, $D_{orientation}$, $W_{conform}$, and W_{score}). As a result of parameter optimization, we obtained 27.4% sensitivity (= recall), 91.0% specificity, 3.5% precision, 90.2% accuracy, 6.3% F-measure^{max}, and AUC of 0.585 in the optimum parameter set of P($D_{distance}$, $D_{orientation}$, $W_{conform}$, W_{score}) = (5, 15, 0.00, 1.00) that maximized F-measure. The prediction accuracy was about 2.7, 1.6, and 1.2 times that of random sampling and tentative prediction methods using ZDOCK 3.0.1 or ZRANK scores, respectively. Comparison among different data scales revealed improvements in prediction accuracy rates against random sampling, i.e. 2.7 (84 × 84), 2.3 (20 × 20), and 1.8 (54 × 54). This indicates that AEP gave prediction accuracy with robustness for fluctuations in data scale. (See the Supporting Information S-1.) From a detailed analysis of the relationship between parameter sets and prediction accuracy, we found that the following two points of optimization greatly improved the prediction accuracy of AEP: (1) classifying candidates with a distance between the centers of gravity ($D_{distance}$) of less than or equal to 5 Å and a deflection difference from each representative ($D_{orientation}$) of 15 degrees into the same cluster and (2) calculating affinity scores with an emphasis on the distribution of shape complementarity scores of each cluster's representative (Z_{score}) rather than on that of cluster density ($Z_{conform}$). We successfully obtained a PPI map consisting of 7056 affinity scores from AEP, enabling us to predict 23 statistically significant pairs out of 84 complexes. Moreover, by investigating the two complexes of 1CGI(26) and 1PPE(37), consisting of enzyme (including proenzyme) and inhibitor, we found that AEP allowed us to significantly predict the active sites of protein−protein complexes and to statistically distinguish biologically relevant pairs with high accuracy.

In the future, we will improve the prediction accuracy of AEP by using unbound data sets, since the F-measure values are similar to the results from random sampling. Sacquin-Mora et al.[44] successfully predicted 7 interaction partners with unbound structures out of 10 in a small data set (i.e., 10 × 10 pairs). It can be thought that the combination of results based on our shape complementarity, weighted interaction energy of Sacquin-Mora et al., and other score function in docking simulation is more available for identification of partners with unbound structures. First, we will rerank the shape complementarity scores on the basis of an energy function such as that used in ZRANK to assess unbound protein−protein docking; we will then apply these scores to affinity prediction between unbound proteins.

ACKNOWLEDGMENT

We are sincerely grateful to Akira Nukada of the University of Tokyo for providing the optimized FFT library for IBM Blue Gene.

Supporting Information Available: Section S1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Jones, S.; Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13–20.
(2) Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177–2198.
(3) Norel, R.; Petrey, D.; Wolfson, H. J.; Nussinov, R. Examination of shape complementarity in docking of unbound proteins. *Proteins* **1999**, *36*, 307–317.
(4) Fields, S.; Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **1989**, *340*, 245–246.
(5) Ito, T.; Tashiro, K.; Muta, S.; Ozawa, R.; Chiba, T.; Nishizawa, M.; Yamamoto, K.; Kuhara, S.; Sakaki, Y. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1143–1147.
(6) Tarassov, K.; Messier, V.; Landry, C. R.; Radinovic, S.; Molina, M. M.; Shames, I.; Malitskaya, Y.; Vogel, J.; Bussey, H.; Michnick, S. W. An in vivo map of the yeast protein interactome. *Science* **2008**, *320*, 1465–1470.
(7) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M.; Johnston, M.; Fields, S.; Rothberg, J. M. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **2000**, *403*, 623–627.
(8) Walhout, A. J.; Sordella, R.; Lu, X.; Hartley, J. L.; Temple, G. F.; Brasch, M. A.; Thierry-Mieg, N.; Vidal, M. Protein interaction mapping in C. elegans using proteins involved in vulval development. *Science* **2000**, *287*, 116–122.
(9) Rain, J. C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schachter, V.; Chemama, Y.; Labigne, A.; Legrain, P. The protein-protein interaction map of Helicobacter pylori. *Nature* **2001**, *409*, 211–215.
(10) Wojcik, J.; Schachter, V. Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics* **2001**, *17*, S296−S305.
(11) Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.;

Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L., Jr.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M. A protein interaction map of Drosophila melanogaster. *Science* **2003**, *302*, 1727–1736.

(12) Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albala, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005**, *437*, 1173–1178.

(13) Calderwood, M. A.; Venkatesan, K.; Xing, L.; Chase, M. R.; Vazquez, A.; Holthaus, A. M.; Ewence, A. E.; Li, N.; Hirozane-Kishikawa, T.; Hill, D. E.; Vidal, M.; Kieff, E.; Johannsen, E. Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7606–7611.

(14) Pazos, F.; Valencia, A. In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins* **2002**, *47*, 219–227.

(15) von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Fields, S.; Bork, P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **2002**, *417*, 399–403.

(16) Smith, G. R.; Sternberg, M. J. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **2002**, *12*, 28–35.

(17) Duhovny, D.; Nussinov, R.; Wolfson, H. Efficient Unbound Docking of Rigid Molecules In *Algorithms in Bioinformatics*; Springer-Verlag: Berlin, Heidelberg, 2002; Vol. 2452, pp 185−200.

(18) Gardiner, E. J.; Willett, P.; Artymiuk, P. J. GAPDOCK: a Genetic Algorithm Approach to Protein Docking in CAPRI round 1. *Proteins* **2003**, *52*, 10–14.

(19) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281–299.

(20) Norel, R.; Sheinerman, F.; Petrey, D.; Honig, B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* **2001**, *10*, 2147–2161.

(21) Palma, P. N.; Krippahl, L.; Wampler, J. E.; Moura, J. J. BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* **2000**, *39*, 372–384.

(22) Taylor, J. S.; Burnett, R. M. DARWIN: a program for docking flexible molecules. *Proteins* **2000**, *41*, 173–191.

(23) Tsukamoto, K.; Yoshikawa, T.; Hourai, Y.; Fukui, K.; Akiyama, Y. Development of an affinity evaluation and prediction system by using the shape complementarity characteristic between proteins. *J. Bioinf. Comput. Biol* **2008**, *6*, 1133–1156.

(24) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400.

(25) Nukada, A.; Hourai, Y.; Nishida, A.; Akiyama, Y. High Performance 3D Convolution for Protein Docking on IBM Blue Gene In *Parallel and Distributed Processing and Applications*; Springer-Verlag: Berlin, Heidelberg, 2007; Vol. 4742, pp 958−969.

(26) Kiselyov, O.; Taha, W. Relating FFTW and Split-Radix In *Embedded Software and Systems*; Springer-Verlag: Berlin, Heidelberg, 2005; Vol. 3605, pp 488−493.

(27) Vuduc, R.; Demmel, J. Code Generators for Automatic Tuning of Numerical Kernels: Experiences with FFTW Position Paper. In *Semantics, Applications, and Implementation of Program Generation*; Springer-Verlag: Berlin, Heidelberg, 2000; Vol. 1924, pp 190−211.

(28) Yoshikawa, T.; Tsukamoto, K.; Hourai, Y.; Fukui, K. In Parameter tuning and evaluation of an affinity prediction using protein-protein docking, *Proceedings of the 10th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering, Sofia, Bulgaria, 2008*; World Scientific and Engineering Academy and Society (WSEAS): Stevens Point, Wisconsin, USA, 2008.

(29) Lattman, E. Optimal sampling of the rotation functio. *Acta Crystallogr., Sect. B: Struct. Sci.* **1972**, *28*, 1065–1068.

(30) Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. Integrating statistical pair potentials into protein complex prediction. *Proteins* **2007**, *69*, 511–520.

(31) Pierce, B.; Hourai, Y.; Weng, Z. ZDOCK 2.3.1 and ZDOCK 3.0.1: Using a New 3D Convolution Library to Enhance Docking Efficiency. *Bioinformatics* **2007**, . submitted for publication.

(32) Pierce, B.; Weng, Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* **2007**, *67*, 1078–1086.

(33) Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. A protein-protein docking benchmark. *Proteins* **2003**, *52*, 88–91.

(34) Mintseris, J.; Wiehe, K.; Pierce, B.; Anderson, R.; Chen, R.; Janin, J.; Weng, Z. Protein-Protein Docking Benchmark 2.0: an update. *Proteins* **2005**, *60*, 214–216.

(35) Carstensen, B.; Plummer, M.; Laara, E.; Hills, M. *Epi: A package for statistical analysis in epidemiology*; Steno Diabetes Center: Gentofte, Denmark, 2008.

(36) R Development Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2007.

(37) Hecht, H. J.; Szardenings, M.; Collins, J.; Schomburg, D. Three-dimensional structure of the complexes between bovine chymotrypsinogen A and two recombinant variants of human pancreatic secretory trypsin inhibitor (Kazal-type). *J. Mol. Biol.* **1991**, *220*, 711–722.

(38) Bode, W.; Greyling, H. J.; Huber, R.; Otlewski, J.; Wilusz, T. The refined 2.0 A X-ray crystal structure of the complex formed between bovine beta-trypsin and CMTI-I, a trypsin inhibitor from squash seeds (Cucurbita maxima). Topological similarity of the squash seed inhibitors with the carboxypeptidase A inhibitor from potatoes. *FEBS Lett.* **1989**, *242*, 285–292.

(39) Wang, D.; Bode, W.; Huber, R. Bovine chymotrypsinogen A X-ray crystal structure analysis and refinement of a new crystal form at 1.8 A resolution. *J. Mol. Biol.* **1985**, *185*, 595–624.

(40) Hecht, H. J.; Szardenings, M.; Collins, J.; Schomburg, D. Three-dimensional structure of a recombinant variant of human pancreatic secretory trypsin inhibitor (Kazal type). *J. Mol. Biol.* **1992**, *225*, 1095–1103.

(41) Odagaki, Y.; Nakai, H.; Senokuchi, K.; Kawamura, M.; Hamanaka, N.; Nakamura, M.; Tomoo, K.; Ishida, T. Unique binding of a novel synthetic inhibitor, N-[3-[4-[4-(amidinophenoxy)carbonyl]phenyl]-2-methyl-2-propenoyl]- N-allylglycine methanesulfonate, to bovine trypsin, revealed by the crystal structure of the complex. *Biochemistry (Moscow)* **1995**, *34*, 12849–12853.

(42) Taranova, N. P. Effect of whole-body x-irradiation on the composition and metabolism of rat brain lipids. *Radiobiologiia* **1975**, *15*, 821–825.

(43) Mendez, R.; Leplae, R.; Lensink, M. F.; Wodak, S. J. Assessment of CAPRI predictions in rounds 3−5 shows progress in docking procedures. *Proteins* **2005**, *60*, 150–169.

(44) Sacquin-Mora, S.; Carbone, A.; Lavery, R. Identification of protein interaction partners and protein-protein interaction sites. *J. Mol. Biol.* **2008**, *382*, 1276–1289.

CI800310F