Special Issue
"Protein-Ligand Interactions"

Vol. 13, pp. 105–115 (2016)
doi: 10.2142/biophysico.13.0_105

Biophysics and Physicobiology
https://www.jstage.jst.go.jp/browse/biophysico/

Regular Article

Specificity of broad protein interaction surfaces for proteins with multiple binding partners

Nobuyuki Uchikoga[1], Yuri Matsuzaki[2], Masahito Ohue[3] and Yutaka Akiyama[2,3]

[1] *Department of Physics, Faculty of Science and Engineering, Chuo University, Bunkyo-ku, Tokyo 112-8551, Japan*
[2] *Education Academy of Computational Life Sciences, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, Japan*
[3] *Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, Japan*

Received December 30, 2015; accepted April 29, 2016

Analysis of protein-protein interaction networks has revealed the presence of proteins with multiple interaction ligand proteins, such as hub proteins. For such proteins, multiple ligands would be predicted as interacting partners when predicting all-to-all protein-protein interactions (PPIs). In this work, to obtain a better understanding of PPI mechanisms, we focused on protein interaction surfaces, which differ between protein pairs. We then performed rigid-body docking to obtain information of interfaces of a set of decoy structures, which include many possible interaction surfaces between a certain protein pair. Then, we investigated the specificity of sets of decoy interactions between true binding partners in each case of alpha-chymotrypsin, actin, and cyclin-dependent kinase 2 as test proteins having multiple true binding partners. To observe differences in interaction surfaces of docking decoys, we introduced broad interaction profiles (BIPs), generated by assembling interaction profiles of decoys for each protein pair. After cluster analysis, the specificity of BIPs of true binding partners was observed for each receptor. We used two types of BIPs: those involved in amino acid sequences (BIP-seqs) and those involved in the compositions of interacting amino acid residue pairs (BIP-AAs). The specificity of a BIP was defined as the number of group members including all true binding partners. We found that BIP-AA cases were more specific than BIP-seq cases. These results indicated that the composition of interacting amino acid residue pairs was sufficient for determining the properties of protein interaction surfaces.

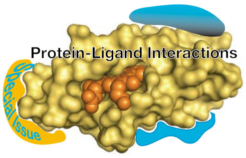
Key words: protein-protein interaction, interaction profile, rigid-body docking process, postdocking process, hub protein

Protein-protein interaction (PPI) networks provide key information for understanding living cells. In yeast, PPI networks have been classified as free-scale networks [1,2]. This type of network has proteins with multiple partners, called hub proteins, including calmodulin, p53, and kalirin, among others [3].

Large-scale PPI prediction provides information regarding the presence or absence of protein interactions. However, when all-to-all PPI predictions are performed, some false-positive pairs can be found [4–6] owing to our lack

Corresponding author: Nobuyuki Uchikoga, Department of Physics, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.
e-mail: uchikoga@phys.chuo-u.ac.jp

◀ ***Significance*** ▶



Three proteins interacting with multiple ligand proteins (true partners) were analyzed to determine the specificity of interaction surfaces among true partners. Then, to observe interaction surfaces, broad interaction profiles (BIPs) were defined as the sum of interaction profiles of decoys, generated by the rigid-body docking process. Two types of BIPs were introduced; one was involved in amino acid sequences (BIP-seq), and the other was composed of interacting amino acid pairs (BIP-AAs). Notably, the specificity of BIP-AAs for true ligand proteins was higher than that of BIP-seqs.

of knowledge regarding PPI pairs. A typical benchmark dataset, e.g., protein-protein docking benchmark dataset ver. 5.0 [7], includes some proteins that exhibit interactions with multiple partners, such as alpha-chymotrypsin, cyclin-dependent kinase 2 (CDK2), and actin. The accuracy of PPI prediction thus depends on whether we have knowledge of interacting protein pairs. When we predict two proteins as binders and we do not have information on the specific binding of the pair, the prediction will be evaluated as false-positive; this problem is directly associated with our understanding of PPI mechanisms.

The protein surface is one of the most important factors in protein sciences for understanding PPI mechanisms. We can determine the properties of a protein surface from the tertiary structure of a protein. Databases of protein surface geometries and electrostatic potentials, such as the eF-site [8], are important for predicting the functions of proteins or determining the details of interaction mechanisms. When investigating PPI mechanisms, more information is obtained from analysis of the surface of each protein and the structures of protein complexes, which are associated with their amino acid sequences and physicochemical properties. Additionally, computational simulations provide information for understanding PPI mechanisms. Molecular dynamic simulations provide information regarding numerous states of protein interactions using *ab initio* calculations [9,10]. Docking simulations can also be used to predict protein complex structures by generating many candidate complex structures (decoys) that exhibit various interaction states. In general, for PPI predictions, near-native structures are evaluated from decoy sets, and most decoys are ignored as false-positives. For example, in drug design, near-native structures are refined to generate high-resolution predicted docking structures. After this process, we can observe the interaction mechanisms in detail with the protein surface area obtained from the small number of near-native structures. In contrast, based on a comprehensive view of protein interaction surfaces, a decoy set, generated by the docking process, includes information for many possible interaction surfaces of protein pairs. Thus, we can obtain information regarding broad protein interaction surfaces or information of interaction surfaces. In this work, we focused not on each protein interaction surface as 'local' protein interaction surface but on the sum of protein interaction surfaces derived from decoys as 'broad' protein interaction surfaces. A set of interaction surfaces of a protein pair, 'local' protein interaction, may differ from sets of pairs with different docking partners because of differences in the shapes of protein surfaces and physicochemical properties of exposed amino acid residues. However, it is unclear how different 'broad' protein interaction surfaces are among multiple partners or between partners and nonbinders.

Therefore, we attempted to obtain an indicator of broad protein interaction surfaces for discriminating true partners of a receptor protein from other nonbinders using docking decoy sets. Then, we introduced profiles of broad protein interaction surfaces. Profile methods are applied to protein-small molecules or protein-protein interactions in postdocking analysis, including cluster analysis [11–13]. Profiles can easily be compared, and other properties can be added, e.g., flags of donors, acceptors, cations, anions, and aromatic residues, for protein-small molecule interactions [11,12]. To investigate PPIs, it is favorable to set interacting amino acid residue pairs for elements of an interaction profile. This type of profile results in better classification of decoys in cluster analysis compared with cases measured by root mean square deviation (RMSD) [13].

Therefore, in this work, we examined three proteins, i.e., alpha-chymotrypsin (PDB-chainID: 1ACB-E), CDK2 (1BUH-A), and actin (1ATN-A), which had multiple binding partners and were deposited in protein-protein docking benchmark dataset ver. 5.0 [7]. These proteins, used as docking receptors, interacted with multiple true partners, as described in Table 1. To generate decoy sets, rigid-body docking was performed for each docking receptor using MEGADOCK ver. 4.0 [14], and proteins were docked with 44 different ligand proteins, including their true partners. We then introduced the concept of broad interaction profiles

Table 1 List of proteins used in docking processes. Parentheses indicate PDBIDs of unbound states

Docking receptors		True partner proteins		Receptors in crystal structures[f]
Names	PDB IDs	Names	PDB IDs	
Alpha-chymotrypsin	1ACB (2CGA)	eglin C PSTI[a] BPTI[b]	1ACB (1EGL) 1CGI (1HPT) 1EAW (9PTI)	Alpha-chymotrypsin Alpha-chymotrypsinogen Matriptase
Actin	1ATN (1IJJ)	DNase I Gelsolin DBP[c]	1ATN (3DNI) 1H1V (1D0N) 1KXP (1KW2)	Actin
CDK2[d]	1BUH (1HCL)	CKSHS1 CDKN3[e]	1BUH (1DKS) 1FQ1 (1FPZ)	CDK2

a: pancreatic secretory trypsin inhibitor, b: pancreatic trypsin inhibitor, c: human vitamin-D binding protein, d: cyclin-dependent kinase 2, e: cyclin-dependent kinase inhibitor 3, f: names of receptors in X-ray crystal structure data with each true partner

(BIPs), which were made by assembling interaction profiles of decoys and provided information of broad protein interaction surfaces. These profiles of protein pairs were compared in cluster analysis, allowing us to observe differences in their profiles.

Materials and Methods

Docking process

For obtaining decoy sets, we performed docking processes using MEGADOCK ver. 4.0, an FFT-grid-based exhaustive rigid-body docking tool with multiparallel calculations [14,15]. In this work, docking processes were performed on an Intel Xeon E7-4870 CPU (2.4 GHz, 10 cores) at the National Institute of Genetics. A total of 2000 docking decoys were used for analysis.

Tertiary structures of protein pairs

We examined proteins interacting with multiple binding proteins and used the tertiary structures of proteins for docking processes. We defined binding partners within the protein docking benchmark by searching "low-throughput-like" direct interactions derived from public databases (Dr. Vachiranee Limph, personal communication, data retrieved on August 2011). Protein pairs were screened in multiple PPI databases (BIND [16], BioGRID [17], DIP [18], HPRD [19], IntAct [20], MINT [21], MPact [22], MPPI [23]), and high-throughput-like data found only in references that reported 50 or more interactions were eliminated. Then, three proteins, i.e., alpha-chymotrypsin, actin, and CDK2, referred to as the receptor proteins, were chosen from protein-protein docking benchmark dataset ver. 5.0 [7]. Table 1 shows the interactions of each receptor protein with multiple interacting proteins. Alpha-chymotrypsin interacts with Eglin C, pancreatic secretory trypsin inhibitor (PSTI), and pancreatic trypsin inhibitor (BPTI), which form various quaternary structures. For example, for BPTI, x-ray crystal structures are available showing a complex with matriptase (PDBID: 1EAW). We used the tertiary structure of alpha-chymotrypsin (1ACB) as the docking receptor. For actin and CDK2, 1ATN and 1BUH were used as the docking receptors. We also used unbound states of tertiary structures for the docking processes, corresponding with ligands of the bound state [7]; PDBIDs for these are shown in Table 1. In order to observe differences in protein interaction surfaces between true ligand partners and other nonbinders, the following 44 ligands were also docked to generate decoy sets (chain IDs are shown in parentheses): 1ACB(I), 1AK4(D), 1ATN(D), 1AVX(B), 1AY7(B), 1B6C(B), 1BUH(B), 1BVN(T), 1CGI(I), 1D6R(I), 1DFJ(I), 1E6E(B), 1E96(B), 1EAW(B), 1EWY(C), 1F34(B), 1FC2(D), 1FQ1(B), 1FQJ(B), 1GCQ(C), 1GHQ(B), 1GRN(B), 1H1V(G), 1HE1(A), 1HE8(A), 1I2M(B), 1IBR(B), 1KAC(B), 1KTZ(B), 1KXP(D), 1KXQ(A), 1M10(B), 1MAH(F), 1PPE(I), 1QA9(B), 1SBB(B), 1TMQ(B), 1UDI(I), 1WQ1(G), 2BTF(P), 2PCC(B), 2SIC(I), 2SNI(I), and 7CEI(B). For unbound states, PDBID and chain IDs are as follows: 1EGL, 1E6J(P), 3DNI, 1BA7(B), 1A19(B), 1IAS(A), 1DKS(A), 1HOE, 1HPT, 1K9B(A), 2BNH, 1CJE(D), 1HH8(A), 9PTI, 1CZP(A), 1F32(A), 1FC1(A), 1FPZ(E), 1FQI(A), 1GCP(B), 1LY2(A), 1RGP, 1D0N, 1HE9(A), 1E8Z(A), 1A23(A), 1F59(A), 1F5W(B), 1M9Z(A), 1KW2(B), 1PPI, 1M0Z(B), 1FSC, 1LU0(A), 1CCZ(A), 1SE4, 1B1U(A), 2UGI(B), 1WER, 1PNE, 1YCC, 3SSI, 2CI2(I), and 1M08(B) (same order as the bound state list). To calculate conformation changes between bound and unbound states, we used the TMscore program [24].

BIPs

First, we generated each profile from a decoy generated by the rigid-body docking process (Fig. 1). A profile was composed of elements when an interaction residue pair was detected, and the value of the corresponding element was set to 1. Otherwise, the value was set to 0. To determine interaction residue pairs, we used the dimplot command of the LIGPLOT program, whose default cut-off distance between nonhydrogen atoms is 3.9 Å [25], which is longer than the distance between hydrogen and acceptor atoms (2.5 Å) [26].

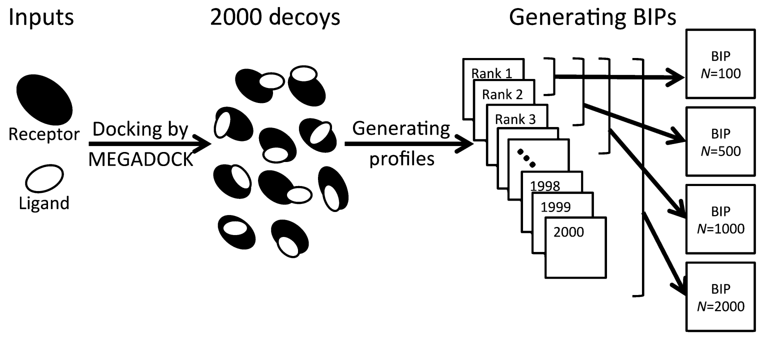


Figure 1 Flowchart for generating BIPs. After generating 2000 decoys by the rigid-body docking process with MEGADOCK, their tertiary structures were converted to interaction profiles with a reranking process. Broad interaction profiles (BIPs) were generated by assembling certain numbers (*N*-values) of profiles.

This type of profile was introduced for comparisons between decoys, as described by Uchikoga and Hirokawa [13]. After generating profiles of decoys, to investigate broad protein interacting surfaces, we introduced the profiles involved in all interaction surfaces included in a decoy set, named BIPs. Each corresponding element of the interaction profile of decoys was added to generate BIPs by assembling decoys after reranking with ZRANK [27].

Decoys with low interaction energy scores were regarded as high-ranking decoys. Then, the number of assembled decoys could influence the results of cluster analysis because a set of high-ranking decoys was expected to generate more specific profiles involved in true partners than those generated by low-ranked decoys. In this work, the *N*-value was defined as the number of top ranked decoys for generating BIPs. Therefore, the BIP depended on a pair of docking proteins and the number of assembled decoys, i.e., the *N*-value.

We used two types of BIPs. The first type of BIP included each element corresponding to the interacting amino acid residue pairs along with amino acid sequences. If two docking protein pairs had lengths of L_a and L_b amino acid sequences, the number of elements in this type of profile was $L_a \times L_b$. We called this type BIP-seq. For example, when elements of a profile of a decoy are $P_1(i, j)$, $P_2(i, j)$, ..., $P_N(i, j)$ with *i*-th and *j*-th amino acids in receptor and ligand proteins, respectively, and N is the number of decoys, the element of a BIP with an *N*-value was $BIP(N; i, j) = \sum_{k=1}^{N} P_k(i, j)$. Elements of profiles of a decoy, $P_k(i, j)$, were 0 or 1. Therefore, elements of BIPs were not only 0 or 1, but could also be integers. The other type of BIP was composed of amino acid residue types, including 400 (20×20) elements. This type was called BIP-AA. Elements of BIP-AA with an *N*-value could be described as $BIP_{AA}(N; m, n) = \sum_{k=1}^{N} \sum_{(i,j)\in S} P_k(i, j)$, where m and n are types of amino acids in receptor and ligand proteins, respectively. The condition of the second summation is $S = \{(i, j) \mid aa(i) = m \ \& \ aa(j) = n\}$, where $aa(i)$ is a type of amino acid located at the *i*-th amino acid residue in a sequence. Then, $BIP_{AA}(m, n) \neq BIP_{AA}(n, m)$, indicating that, for example, the element of interaction between ALA belonging to the receptor protein and GLY of the other protein was different from that between GLY of the receptor and ALA of the other protein.

Calculation of distance between BIPs and cluster analysis

For cluster analysis, it was necessary to compare all protein pairs between BIPs. We use the Tanimoto coefficient (Tc) to measure the similarities between profiles, which were converted to distance by calculating $1 - Tc$. Because the elements of BIPs could be integers, the following equation was used to calculate the Tanimoto coefficient between BIP_a and BIP_b:

$$Tc = \frac{\sum_{k=1}^{L} BIP_a(N; k) BIP_b(N; k)}{\sum_{k=1}^{L} BIP_a^2(N; k) + \sum_{k=1}^{L} BIP_b^2(N; k) - \sum_{k=1}^{L} BIP_a(N; k) BIP_b(N; k)}$$

where k is the element index of BIPs, and L is the number of elements (Table 2 in [28]). This equation indicates that each BIP was used as a vector. For example, when $BIP(N; 0, 0)$ was converted to $BIP(N; 1)$, and $BIP(N; 0, 1)$ was converted to $BIP(N; 2)$, etc., we obtained the L-dimension BIP vector. Then, the numerator and the third term in the denominator corresponded to the inner product between BIP_a and BIP_b.

Table 2 List of the number of ligands in a classified group including all true partners (*M*-values)

		Tanimoto Distance							
Receptor protein		Bound state				Unbound state			
		N-value				*N*-value			
		100	500	1,000	2,000	100	500	1,000	2,000
Alpha-chymotrypsin	BIP-seq	5	20	17	43	41	42	43	34
	BIP-AA	20	8	12	11	22	26	9	7
Actin	BIP-seq	41	40	39	28	44	23	16	16
	BIP-AA	41	21	14	12	25	27	21	13
CDK2	BIP-seq	41	16	32	43	44	44	44	43
	BIP-AA	43	25	43	39	15	43	43	43
		Euclidean Distance							
Alpha-chymotrypsin	BIP-seq	6	16	35	14	44	12	6	19
	BIP-AA	22	24	20	10	25	15	5	8
Actin	BIP-seq	44	44	42	41	25	12	13	15
	BIP-AA	41	22	23	12	17	43	10	13
CDK2	BIP-seq	41	14	15	42	43	43	43	43
	BIP-AA	40	43	39	39	40	33	40	38

The other terms corresponded to the squares of the length of vectors BIP_a and BIP_b, respectively. This indicator of Tanimoto distance was used to classify interaction profiles for effective exploration of docking spaces [29]. However, in the case of integer vectors of BIPs, this indicator cannot satisfy triangle inequality. Therefore, we only discussed the specificity of true partners (*M*-values) without detailing the topologies of tree diagrams. Additionally, we also calculated distances between BIPs as Euclidean distances: $Ed=\sqrt{\sum_{k=1}^{L}(BIP_a(N;\ k)-BIP_b(N;\ k))^2}$ in order to observe *M*-values.

When comparing between BIP-AAs, L was 400 because of the 20 types of amino acid residues in each docked protein. From this equation, because different pairs had different numbers of elements of BIP-seqs, we calculated *Tc* between BIPs using common receptor information. Therefore, when comparing BIP-seqs, if docked proteins had lengths of amino acid sequences of $L=L_a$, which is the length of common receptor amino acid sequences, the elements of the BIP vector were $BIP(N;\ i)=\sum_{j=1}^{L_b}BIP(N;\ i, j)$, where L_b is the length of an amino acid sequence of a ligand protein. On the other hand, in the case of BIP-AAs, although all profiles included the same elements constructed in a 20×20 shape, we used information for receptors in BIP-AAs. Then, distance matrices were generated for docking results of pairs, which included each receptor and all 44 ligands.

After distance calculations for all combinations of BIPs, we performed hierarchical clustering using software for statistical analysis (R ver. 3.2.1.) with the function 'hclust' with the group average method option.

Specificity of BIPs (*M*-value) with the T_{min}-value

To investigate differences between BIPs of true partners and other proteins, we examined the specificity of BIPs. After cluster analysis, a group including all true partners was evaluated. For this process, a tree diagram was divided using various thresholds from 0.01 to 0.5 with 0.01 step in R ver. 3.2.1 software. Next, we obtained the minimum threshold, T_{min}-value by iterative searching from the smallest *T*-value (0.01) until all target ligands were classified into the same group. At the same time, we counted the number of members in the group, i.e., the *M*-value. If the *M*-value was small, BIPs of true partners were more similar than BIPs of other docking ligands, indicating BIPs with higher specificity. When the *M*-value was 44, corresponding to the nonspecific BIP, the T_{min}-value was indicated as *N/A* (Fig. 2).

Results and Discussion

A decoy set generated by the rigid-body docking process includes information for protein surfaces. Decoys are generated using the docking score involved in shape complementarity, electrostatic parameters, and desolvation-like parameters in MEGADOCK ver. 4.0 [14]. Each decoy provides information regarding the quaternary structure. Thus, a set of decoys can be used for the postdocking analysis, e.g., the reranking process, and can provide information regarding broad interaction surfaces in a pair of proteins. In order to observe information regarding broad protein surfaces, we introduced the concept of BIPs, which were generated by assembling high-ranking decoys after reranking.

BIP depended on protein pairs and the number of assembled decoys (i.e., *N*-values). We used the BIP of a receptor molecule to calculate their similarities. Even if an identical receptor protein was used for the docking process, the BIP pattern depended on various ligand proteins, indicating that they were not the same.

Cluster analysis and specificity of BIPs including true partners

To investigate differences in the interacting surfaces with various ligands, we performed cluster analysis using BIPs. Figure 2A shows tree diagrams with the BIP of alpha-chymotrypsin as a bound state receptor. The number of assembled decoys for generating a BIP (*N*-values) was changed from 100 to 2000, and we examined ligand proteins known as the true partners of a certain receptor. We then counted the number of ligands classified into a group, including all true partner ligands interacting with the receptor (i.e., the *M*-value). The specificity of BIPs including true partners corresponded to the *M*-values. When $N=100$, the distance was greater than that when $N\neq100$ for all receptor cases (Fig. 2). These data indicated that BIPs for which *N* was 100, involving a small number of high-ranking decoys, were more different in cases with multiple ligands than in cases with larger *N*-values.

Figure 2 also shows tree diagrams generated using the two types of BIPs and various *N*-values for all receptor proteins, alpha-chymotrypsin, actin, and CDK2. For example, in the case of bound alpha-chymotrypsin (PDBID-chainID: 1ACB-E), which is known to interact with Eglin C (1ACB-I), PSTI (1CGI-I), and BPTI (1EAW-B), we obtained a group composed of five ligands ($M=5$), including all three true partners of alpha-chymotrypsin in the case of BIP-seq cases in which $N=100$. When the *N*-value was 2000, most ligand cases were classified into a single group, including all true partners ($M=43$), indicating that this situation (BIP-seq) did not allow for the distinction of true partners from other ligand cases. Additionally, the case of $N=100$ for BIP-seq was consistent with the results of reranking decoys because high-ranking decoys generated specific protein interaction surfaces for true partners. Notably, cluster analysis using BIP-AAs showed fewer group members (*M*-values) for all *N*-values than for BIP-seq cases, except in the case of $N=100$. For cases in which $N=500$, when the *M*-value was low ($M=8$), the specificity of BIP-AAs required more low-ranked decoys compared with the BIP-seq case of $N=100$. Moreover, every T_{min}-value in BIP-AA cases was smaller than that in BIP-seq cases, indicating that BIP-AAs had smaller distances between each other than BIP-seq cases, e.g., alpha-chymotrypsin in

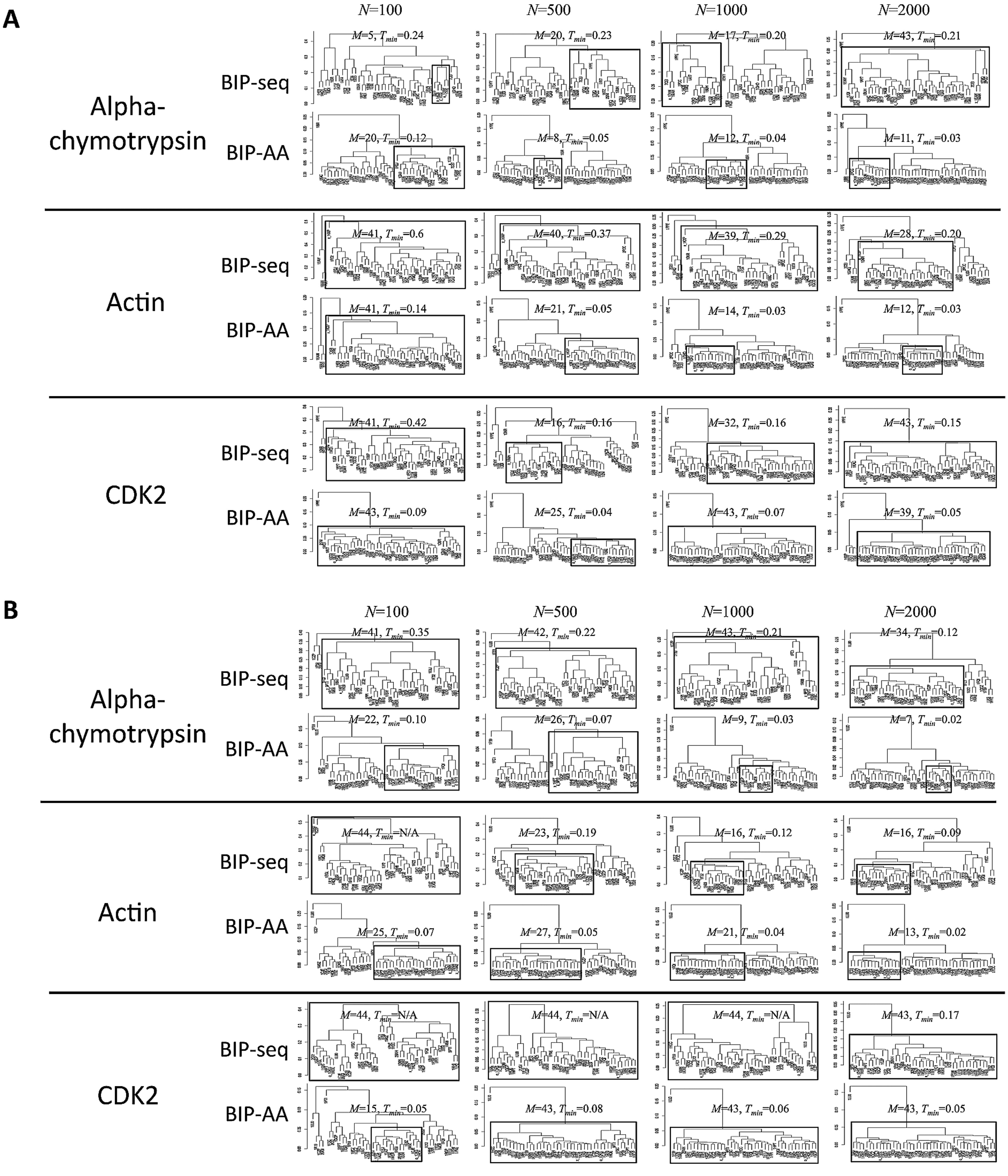


Figure 2 Tree diagrams from cluster analysis by BIP similarities as Tanimoto distances. Each square includes all true partners. *M*-values are the numbers of ligands included in each square with minimum threshold T_{min}-values. The true partners with the quaternary structure of the bound state and the corresponding unbound state protein of PDBIDs are marked as "o_". The other true partners are marked as "x_".

Figure 2 (other cases of T_{min}-values are not shown). *M*-values for all receptor protein cases are shown in Table 2. In the bound state for BIP-seq cases, we found the lowest *M*-values at *N*=2000 for actin, indicating that more decoys were need for highly specific BIP-seq cases compared with those for other receptor cases. For CDK2, the lowest *M*-values were found at *N*=500, which was comparable to the *N*-value of alpha-chymotrypsin.

In the unbound state (Fig. 2B and Table 2), the lowest *M*-values for BIP-seq cases were greater than those for bound cases. The lowest *M*-values for both types of BIPs were found at *N*=2000 for all receptor cases, except CDK2 in the case of BIP-AA. However, the *M*-value decreased as the *N*-values increased. Therefore, it was possible to obtain lower *M*-values when the *N*-values were higher. Moreover, the lowest *M*-value for CDK2 in the case of BIP-AA in the unbound state was found at *N*=100, indicating that this number of decoys was sufficient to obtain highly specific BIP-AA information. In BIP-AA cases, we also found that the lowest *M*-values in the unbound state were lower than those in the bound state, except for actin. However, analysis of actin indicated comparable results between *M*=12 in the bound state and *M*=13 in the unbound state. Comparisons between BIP-seq and BIP-AA cases in the unbound state revealed that the lowest *M*-values of BIP-AAs were lower than those of BIP-seq cases for each receptor case. We also obtained *M*-values by calculating Euclidean distances of BIPs (Table 2). In this case, the lowest *M*-values in each case were found to be comparable to those of Tanimoto distance cases. However, for alpha-chymotrypsin, smaller *M*-values were found in the unbound state, particularly for BIP-seq. For actin, in the bound state of BIP-seq, larger *M*-values were observed than those in Tanimoto distance cases, indicating that the specificity of interaction surfaces of true partners was reduced. For CDK2, we found specificity (low *M*-values) for BIP-seq in the bound state, although nonspecificity was found for BIP-AA in the bound state.

In the case of the BIP-AA of actin, *M*-values in the unbound state were reduced compared with those in the bound state owing to differences between the tertiary structures of receptor proteins. The RMSDs ranged from 13.27 Å to 19.21 Å in the bound state for all combinations of the three receptors in the tertiary complex structures, i.e., 1ATN-A, 1H1V-A, and 1KXP-A. On the other hand, receptors in the unbound state were similar because of the use of identical structures (1IIJ-B). Thus, structural deviations of receptors influenced their protein surfaces. Additionally, low-specificity BIPs were found in the bound state, indicating that these cases exhibited larger *M*-values than those in the unbound state. However, for BIP-AAs, the lowest *M*-values were comparable between the bound (*M*=12) and unbound states (*M*=13), indicating that interacting protein surfaces were more specific than those in cases of other ligands. Notably, the results for CDK2 were different. RMSDs were calculated using two receptors in tertiary complex structures, yielding RMSDs of 12.46 Å in the bound state and 8.49 Å in the unbound state. Thus, these data provided information regarding the number of high-ranking decoys suitable for achieving the most specific BIP-AA in the unbound state.

Physicochemical properties of protein-interacting interfaces

Because BIPs included information for interacting amino acid pairs, we could evaluate the physicochemical properties of BIPs. Tables 3 and 4 show the net charge and hydrophobic parameters of the three receptors and their true partners. The hydrophobic properties of ligands decreased to negative values as *N*-values increased because increasing the number of interacting residue pairs was associated with the number of assembled decoys. In receptors in the bound state (Table 3), negative hydrophobic properties were also found in ligand cases, indicating that these ligands had hydrophilic properties, with the exception of 1ACB (alpha-chymotrypsin) when *N*=100. The condition of *N*=100 in 1ACB yielded more hydrophobic receptors. We also observed interaction residues using x-ray crystal complex structure data. Figure 3A shows interaction sites and aligned amino acid sequences for alpha-chymotrypsin (1ACB), alpha-chymotrypsinogen (1CGI), and matriptase (1EAW). Interestingly, the amino acid sequences and interaction sites were similar, as were the physicochemical properties of these interacting residues, as determined using hydropathy parameters assigned by Kyte and Doolittle [30] and electrostatic charges (Arg, Lys, and His: +1; Asp and Glu: –1; and other amino acids: 0) [31]. The net charges and the hydrophobic parameters were defined as the sum of parameters assigned to their residues composing the interaction site. The net charges (*E* values) of the interaction sites were 10.0, 10.0, and 11.0, and the hydrophobic parameters (*H* values) were 25.7, 50.7, and 61.30 for the receptors 1ACB, 1CGI, and 1EAW, respectively. For the ligand sides, the following *E* and *H* values were obtained for 1ACB, 1CGI, and 1EAW, respectively: –2.0 and 16.40, –12.0 and –145.0, and 66.0 and –242.60. In the unbound state of alpha-chymotrypsin (1ACB), BIPs showed hydrophilic properties (Table 4). These results were associated with the lowest *M*-value at *N*=2000 for the unbound case (Table 2), implying that more decoys were necessary to obtain BIPs specific to the true partners. In the other cases, actin (1ATN) had interaction sites with physicochemical properties (*E* and *H* values) for 1ATN, 1H1V, and 1KXP as follows: 18.0 and –35.90, –7.0 and –22.90, and 4.0 and –167.60, respectively. In this case, the ligands had *E* and *H* values of 6.0 and –83.80, 0.0 and –43.50, and 7.0 and –41.30 for 1ATN, 1H1V, and 1KXP, respectively, in the absence of interaction sites. The interaction sites of receptors were similar, except for that of 1ATN (Fig. 3B). For actin in the bound state, the ligand of 1ATN was consistent at all *N*-values, indicating the positive net charge of 1ATN receptor and ligand of BIPs at all *N*-values (Table 3). Moreover, for the 1ATN pair, assembled high-ranking decoys had native-like interaction properties, and *N*-values were sufficient for obtaining specific BIPs.

The BIPs of CDK2 (1BUH) exhibited hydrophilic and positive net charges in receptor proteins in both the bound and unbound states (Tables 3 and 4). In the unbound state, the ligand sides of BIPs exhibited positive net charges for both ligands (Table 4). In contrast, negative net charges for the 1BUH ligand and positive net charges for the 1FQ1 ligand were found in the bound state (Table 4). For the receptors, the interaction sites exhibited *E* and *H* values of –18.0 and –45.40 for 1BUH and –18.0 and –110.90 for

Table 3 Physicochemical properties of the bound state

PDBID			Net charge		Hydrophobicity		Number of interacting residue pairs
Receptor	Ligand	*N*-value	Receptor	Ligand	Receptor	Ligand	
1ACB	1ACB	100	82	320	392	–1428	3111
	1CGI		250	–23	365.4	–3708.9	2935
	1EAW		105	794	137.7	–2664.8	2765
	1ACB	500	478	1566	–139.2	–9223.6	14692
	1CGI		1012	4	–1556.8	–17940.2	13539
	1EAW		420	3786	–2240.4	–14204.09	13323
	1ACB	1000	1062	2970	–5023.4	–18962.71	28487
	1CGI		1803	0	–7359	–37655.49	26563
	1EAW		886	7247	–8434.5	–29224.91	26576
	1ACB	2000	2392	5668	–21928.2	–41395.51	55915
	1CGI		3022	195	–24285.7	–78172.12	52376
	1EAW		2136	14199	–26546.41	–60223.99	53624
1ATN	1ATN	100	219	139	–4507.6	–3973.2	3743
	1H1V		109	121	–4634.1	–4588	3864
	1KXP		3	–1	–4614.3	–2721.1	4255
	1ATN	500	387	894	–25735.1	–21367.5	18486
	1H1V		202	798	–23722.7	–21893.59	18350
	1KXP		–33	–4	–23312.5	–22638.71	19491
	1ATN	1000	582	2058	–52365.68	–45610.88	36613
	1H1V		71	1451	–51153.1	–44999.39	36763
	1KXP		–159	22	–50730.7	–49853.99	39312
	1ATN	2000	230	4587	–113680.8	–97211.3	75000
	1H1V		–709	2824	–108758.3	–96134.17	74603
	1KXP		–1028	–130	–112267.11	–111733.74	80426
1BUH	1BUH	100	377	–221	–3076.9	–4293	2843
	1FQ1		398	325	–4381.1	–3722.9	3375
	1BUH	500	1924	–701	–16989.3	–19486.2	13554
	1FQ1		2306	1489	–23405.9	–19446.6	16929
	1BUH	1000	3778	–838	–34932	–37541.2	26969
	1FQ1		4434	3003	–47284.79	–39759.89	34296
	1BUH	2000	7430	–703	–74073.6	–76021.42	54585
	1FQ1		9194	6874	–100196.3	–83359.69	70848

1FQ1, respectively (Fig. 3C). Interaction sites on ligands had *E* and *H* values of –15.0 and –106.60 for 1BUH and 1.0 and –99.30 for 1FQ1, respectively, in the absence of interaction sites. Although the positive net charges in the ligand of 1FQ1 were close to zero, these parameters appeared to be consistent with the physicochemical properties of BIPs in the bound state (Table 3), in contrast to the results for the unbound state shown in Table 4, indicating that large *M*-values were found as nonspecific BIPs in the unbound state (Table 2). However, we found that *M*=15 for BIP-AA in the unbound state, suggesting that high-ranking decoys generated specific BIP-AAs for true partners.

We observed the physicochemical properties of BIPs under different conditions. Particularly in the bound state, consistent properties were found with actual interaction sites in some cases. In contrast, more decoys appeared to be necessary to obtain specific BIPs in the unbound state.

Conclusion

In this work, we examined proteins that could bind with other multiple proteins, e.g., hub proteins. Generally, hub proteins have disordered regions, interacting with various proteins through large conformation changes [2]. In this work, we examined globular-type proteins selected from a docking benchmark dataset. These proteins can be categorized into three types in terms of conformational changes between the bound and unbound states, i.e., difficult, moderately difficult, and rigid-body [7]. Alpha-chymotrypsin and actin are classified into the difficult type, with RMSDs

Table 4 Physicochemical properties of the unbound state

PDBID			Net charge		Hydrophobicity		Number of interacting residue pairs
Receptor	Ligand	*N*-value	Receptor	Ligand	Receptor	Ligand	
1ACB	1ACB	100	58	–5	–204.5	–2073.2	2052
	1CGI		81	273	–162.3	–4592.4	2126
	1EAW		61	674	–366.9	–3512.1	2087
	1ACB	500	321	–120	–2411.8	–10528.2	9534
	1CGI		387	977	–617.5	–19018.4	9720
	1EAW		241	2916	–2067.9	–14417.5	9673
	1ACB	1000	676	–289	–5191.5	–19787.2	18498
	1CGI		700	1669	–2326.6	–35923.51	18394
	1EAW		511	5334	–4467.3	–26688.4	18413
	1ACB	2000	1300	–598	–10210	–36639.5	35574
	1CGI		1427	2502	–6335	–66801.51	34904
	1EAW		1101	9754	–9848.8	–49854.51	35037
1ATN	1ATN	100	–120	333	–3963	–4298.8	3291
	1H1V		–164	53	–5074.8	–5474.2	3613
	1KXP		–199	–42	–4733.2	–5231.8	3579
	1ATN	500	–566	1437	–18381.8	–19863.8	14534
	1H1V		–711	–173	–20845.99	–25009.81	16389
	1KXP		–665	–496	–20928.3	–26045.5	15840
	1ATN	1000	–1173	2545	–36763.21	–39000.9	27991
	1H1V		–1481	–142	–40160.8	–47217.21	31144
	1KXP		–1220	–1088	–38860.49	–50446.61	30127
	1ATN	2000	–2468	4780	–70724.91	–74852.81	53554
	1H1V		–3070	–194	–77328.7	–89695.49	59364
	1KXP		–2573	–2230	–75127.91	–97115.8	57889
1BUH	1BUH	100	479	161	–3238.3	–3294	2361
	1FQ1		308	365	–3496.7	–3946	2782
	1BUH	500	1795	722	–13457.1	–17625.7	10622
	1FQ1		1485	1516	–14542.7	–18653.81	12177
	1BUH	1000	3128	1254	–24678.2	–34111.51	19918
	1FQ1		2885	2905	–28007.2	–36008.7	23171
	1BUH	2000	5802	2160	–46804.9	–65434.8	37755
	1FQ1		5221	5405	–52693.5	–66785.11	43747

between interfaces of the bound and unbound states of 2.26 and 3.28, respectively. CDK2 is a rigid-body type, with an RMSD of 0.75. Profile methods may be useful for studies of proteins with large conformation changes. For example, interaction fingerprints have been applied to docking problems using calmodulin with large conformation changes, indicating that profile methods are useful for analysis of protein interaction surfaces [13].

In summary, we observed the specificities of BIPs for true partners and the physicochemical properties of BIPs; these properties were not directly related to the PPI predictions. However, the information provided by analysis of broad possible protein interfaces can yield clues for understanding PPI mechanisms.

Acknowledgement

This work was supported by MEXT KAKENHI (grant number 15K00407). Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics and TSUBAME 2.5 at Global Science and Information Center, Tokyo Institute of Technology through an HPCI system research project (project ID: hp140173).

Conflicts of Interest

N. U., Y. M, M. O., and Y. A. declare that they have no conflicts of interest.

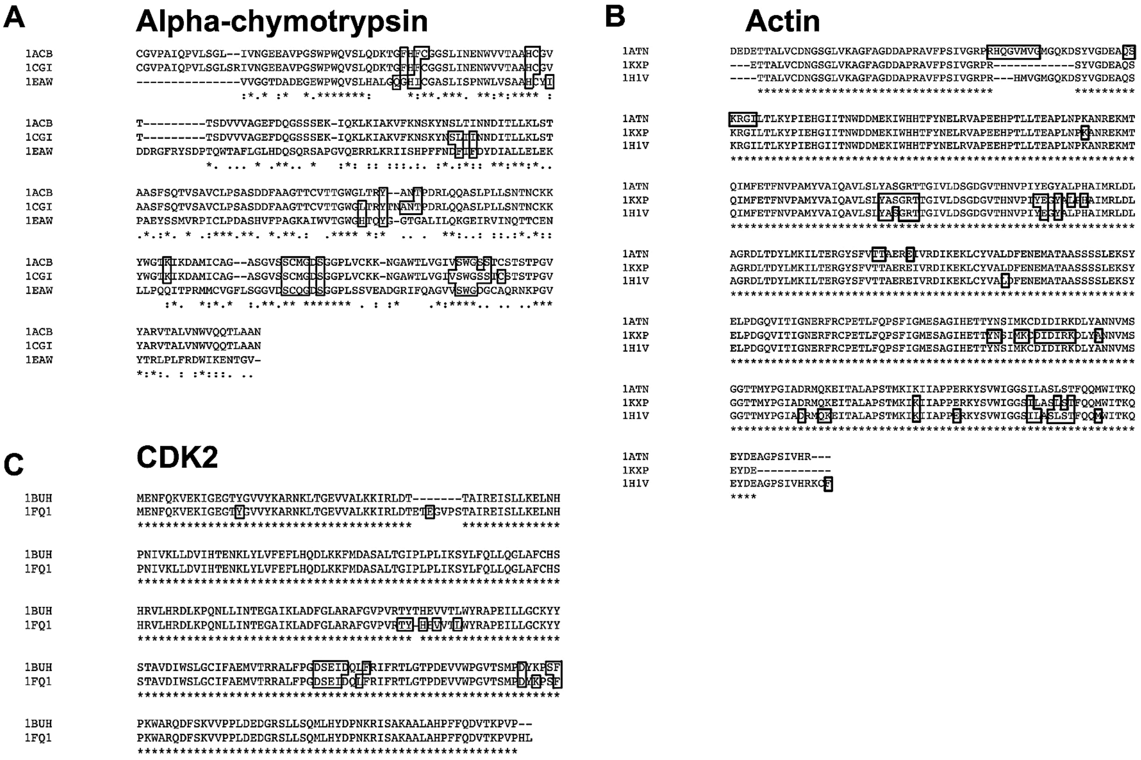


Figure 3 Alignment of receptors and interaction sites with x-ray crystal complex structures. Each receptor indicates PDBIDs corresponding to receptors in Table 1. Squares are interaction sites extracted from complex structure data by LIGPLOT [25].

Author Contributions

N. U. and Y. A. directed the entire project. N. U., Y. M., and M. O. wrote the manuscript. N. U. performed profile analysis. Y. M and M. O. prepared the data sets. All authors read and approved the final manuscript.

References

[1] Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).

[2] Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148 (2005).

[3] Patil, A., Kinoshita, K. & Nakamura, H. Hub promiscuity in protein-protein interaction networks. *Int. J. Mol. Sci.* **11**, 1930–1943 (2010).

[4] Matsuzaki, Y., Ohue, M., Uchikoga, N. & Akiyama, Y. Protein-protein interaction network prediction by using rigid-body docking tools: application to bacterial chemotaxis. *Protein Pept. Lett.* **21**, 790–798 (2014).

[5] Ohue, M., Matsuzaki, Y., Uchikoga, N., Ishida, T. & Akiyama, Y. MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data. *Protein Pept. Lett.* **21**, 766–778 (2014).

[6] Acuner-Ozbabacan, S. E., Keskin, O., Nussinov, R. & Gursoy, A. Enriching the human apoptosis pathway by predicting the structures of protein-protein complexes. *J. Struct. Biol.* **179**, 338–346 (2012).

[7] Vreven, T., Moal, I. H., Vangone, A., Pierce, B. G., Kastritis, P. L., Torchala, M. *et al.* Updates to the integrated protein-protein interaction benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *J. Mol. Biol.* **427**, 3031–3041 (2015).

[8] Kinoshita, K. & Nakamura, H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.* **12**, 1589–1595 (2003).

[9] Frembgen-Kesner, T. & Elcock, A. H. Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys. J.* **99**, L75–L77 (2010).

[10] Visscher, K. M., Kastritis, P. L. & Bonvin, A. M. J. J. Non-interacting surface solvation and dynamics in protein-protein interactions. *Proteins* **83**, 445–458 (2015).

[11] Marcou, G. & Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **47**, 195–207 (2007).

[12] Deng, Z., Chuaqui, C. & Singh, J. Structural Interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **47**, 337–344 (2004).

[13] Uchikoga, N. & Hirokawa, T. Analysis of protein-protein docking decoys using interaction fingerprints: application to the reconstruction of CaM-ligand complexes. *BMC Bioinformatics* **11**, 236 (2010).

[14] Ohue, M., Shimoda, T., Suzuki, S., Matsuzaki, Y., Ishida, T. & Akiyama, Y. MEGADOCK 4.0: an ultra-high-performance

protein-protein docking software for heterogeneous supercomputers. *Bioinformatics* **30**, 3281–3283 (2014).

[15] Matsuzaki, Y., Uchikoga, N., Ohue, M., Shimoda T, Sato, T., Ishida, T. *et al.* MEGADOCK 3.0: a high-performance protein-protein interaction prediction software using hybrid parallel computing for petascale supercomputing environments. *Source Code Biol. Med.* **8**, 18 (2013).

[16] Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K. *et al.* The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.* **33**, D418–D424 (2005).

[17] Chair-Aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43**, D470–D478 (2015).

[18] Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., Eisenberg, D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32**, D449–D451 (2004).

[19] Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S. *et al.* Human Protein Reference Database-2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2009).

[20] Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F. *et al.* The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–363 (2014).

[21] Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40**, D857–D861 (2012).

[22] Guldener, U., Munsterkotter, M., Oesterheld, M., Pagel, P., Ruepp, A. & Mewes, HW. MPact: The MIPS protein interaction resource on yeast. *Nucleic Acids Res.* **34**, D436–D441 (2006).

[23] Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, D., Frishman, G. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834 (2005).

[24] Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).

[25] Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).

[26] McDonald, I. & Thornton, J. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).

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

[28] Willet, P., Barnard, J. M. & Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **38**, 983–996 (1998).

[29] Uchikoga, N., Matsuzaki, Y., Ohue, M., Hirokawa, T. & Akiyama, Y. Re-docking scheme for generating near-native protein complexes by assembling residue interaction fingerprints. *PLoS ONE* **8**, e69365 (2013).

[30] Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

[31] Fauchere, J. L., Charton, M., Kier, L. B., Verloop, A. & Pliska, V. T. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.* **32**, 269–278 (1988).