Inferring Protein-Protein Interactions (PPIs) Based on Computational Methods

Shuichi Hirose

Nagase & Co. Ltd. Research & Development Center, 2-2-3 Murotani,
Nishi-ku, Kobe, Hyogo,
Computational Biology Research Center, Advanced Industrial Science and
Technology, 2-4-7 Aomi, Koto-ku, Tokyo,
Japan

1. Introduction

Proteins are involved in many essential cellular processes, such as metabolism and signalling. They function by interacting with other molecules within the cell. Thus, protein interaction is one of the important keys to understand protein functions. As a consequence of the development of high-throughput experimental methods for detecting protein interactions, large volumes of data are now available. Although the data are valuable, there are limitations to their application. Therefore, computational methods are helpful tools for predicting protein interactions. With the increase in genome sequence data, the importance of computational methods in this field is growing more and more.

Another important factor to understand protein function is flexibility, because a protein molecule is not a rigid body. Flexible regions are often necessary for proteins to perform their functions, e.g. by enabling their flexible conformations to interact with other molecules and proteins. Therefore, it is important to understand the relationship between protein flexibility and protein interactions. In accordance with the increasing numbers of available protein structures, several databases that deal with protein flexibility have been built. Computational methods for analyzing protein motion are also being developed, for applications to PPI (protein-protein interaction) data.

The aim of this chapter is to provide a review on PPI prediction by computational techniques. In the first half of this chapter, the concepts and applications of several methods for inferring PPIs are introduced. They use genomic information based on evolutionary events. In the second half, the databases and prediction methods that deal with protein flexibility are introduced, and the possibility of inferring PPIs from protein flexibility will be discussed.

2. Computational methods to infer PPIs

The prediction of PPIs can be regarded as a binary classification problem, whereby the aim is to identify pairs of proteins as either interacting or non-interacting. PPIs can be divided into three types (Brown *et al.*, 2010). The first is direct protein interactions, which involve

direct physical contacts between proteins. The second is indirect functional association. In this case, the interacting protein pair does not have a direct physical contact, but it indirectly interacts, such as in the formation of a complex. The third is a member of biological pathway. In this case, the protein pairs do not form a complex, but their interactions occur in a logical order (for instance, in a signalling pathway).

Proteins exert their biological functions by participating in a PPI network. Protein interactions, as well as their biochemical functions, work as a type of selective pressure during evolution. Therefore, they influence the genome structure associated with the protein interactions. Conversely, analyzing the changes in the patterns on the genome makes it possible to infer PPIs. Several evolutionary events function as factors that impact the evolution of PPIs, including horizontal gene transfer, operon structure, co-evolution, co-expression, and lineage-specific gene loss. Several methods that infer PPIs using these various types of evolutionary information have been developed (Valencia and Pazos, 2002; Skrabanek *et al.*, 2008). In this section, the principles and applications of PPI prediction methods are introduced. The different PPI prediction methods (Shoemaker and Panchenko, 2007) are listed in Table 1. First, the genomic inference methods (Rosetta stone, conservation of gene neighborhood, and phylogenetic profile), which predict functional association using genomic context, are presented. Next, the methods (mirror tree, *in silico* two-hybrid system) based on co-evolution are introduced, which are applicable to domain interactions as well as protein interactions. Finally, the sequence signature and machine learning based-methods are presented.

Method	Interaction Type	Interaction
Rosetta stone	Indirect functional association	Protein
Conservation of gene neighborhood	Indirect functional association	Protein
Phylogenetic profiles	Indirect functional association	Protein/domain
Mirror tree	Indirect functional association	Protein/domain
In silico two-hybrid system	Direct physical interaction	Protein/domain
Sequence signature	Direct physical interaction	Protein/domain
Supervised classification	Direct physical interaction	Protein/domain

Table 1. Summary of PPI prediction methods.

The second column represents the interaction type predicted by the method. The third column shows whether the method is designed to predict protein or domain interaction.
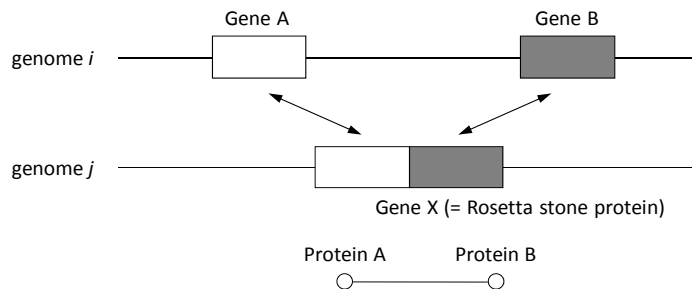
2.1 Rosetta stone

The Rosetta stone approach infers PPIs by comparing different genomes. It is often observed that two proteins that interact in genome *i* have a homologous protein that is fused into one protein in genome *j* (Fig. 1). These two gene products are functionally related in many cases (Enright *et al.*, 1999). The fused protein is called a 'Rosetta stone protein', since it serves as a key for unlocking the functional relationship between two genes that are encoded independently in the genome. The Rosetta stone approach estimates functionally related protein pairs based on such a concept. The benefit of this approach is applicable to all

genomes, including those of Eukaryote. Not surprisingly, the inference of a protein interaction is restricted to the case where the gene fusion can be detected.

Hence, the approach searches for the proteins that are conserved between different organisms. The following two points must be considered, in order to obtain higher prediction accuracy. First, the proteins that interact with many other proteins, such as the HRG domain, and the CBS domain, which binds to DNA, should be removed. Second, the analysis is focused on the case where the pairs of genes that are fused together are orthologous. As an extension of the Rosetta stone approach, it can predict a functionally related gene cluster by combining several results. Four proteins, A, B, C, and D, are considered to be functionally related if the Rosetta stone proteins of the A-B, B-C, and C-D pairs are found.

Applying the Rosetta stone approach to many genomes revealed 6,809 potentially interacting pairs in *Escherichia coli* and 45,502 pairs in yeast (Marcotte *et al*., 1999). The two proteins in each pair have significant sequence similarity to a single fused protein in another genome. Some proteins interact with several other proteins, and these connections apparently represent functional interactions, such as complexes or pathways.

Gene X is the Rosetta stone protein, indicating that protein A and protein B are functionally related.

Fig. 1. The concept of the Rosetta stone approach.

2.2 Conservation of gene neighborhood

The genome comparison among Bacteria or Archaea indicated that the gene order and the operon structure are not conserved on the genome. This is because they have changed with evolutionary events, such as recombination, gene disruption, gene formation, and horizontal gene transfer. These phenomena suggest that the gene order is basically not subjected to selective pressure. However, the gene order or gene clusters on a genome are conserved if the gene products physically interact with each other, such as by complex formation, or if the proteins are transcribed as a single unit (Dandekar *et al*., 1998). Briefly, it is often observed that the genes encoding proteins that either form a complex via physical interaction or work together in the same pathway are encoded in the same operon in different genomes. Thus, the gene order is conserved among different genomes, although the operon structure is fundamentally unstable during evolution (Fig. 2(A)). The conservation of the gene neighborhood approach infers proteins that are involved in the same biological process, using genome information. Many of the functionally related genes predicted by this approach encode proteins that either interact with each other directly, participate in the formation of the same complex, or work in the same metabolic pathway.

The conserved clusters of genes in an operon are detected by various concepts, such as Run or BBH (bidirectional best hit) (Overbeek *et al.*, 1999) (Fig. 2(B)). A set of genes is called a "run" if they all occur on the same strand and the gaps between adjacent genes are 300 bases or less. Any pair of genes occurring within a single run is called "close". If gene Xi in genome i is closest to Xj in genome j and Xj is closest to Xi, then Xi and Xj are called BBH. Genes (Xi, Yi) from genome i and genes (Xj, Yj) from genome j form a PCBBH (pair of close bidirectional best hit) if two pairs of BBHs are considered. The conservation of gene neighborhood approach uses such virtual operons and orthologs to infer PPIs. That is, two orthologous groups are considered to have a connection if they co-occur in the same potential operon two or more times. The advantage of this approach is that the conservation of gene order or gene co-occurrence in the Run is stricter than the Rosetta stone and phylogenetic profile approaches, and it can cover a wider range of genes. However, the application of this approach is limited to Bacteria or Archaea that have operon structures.

Snel *et al.* reported 3,033 orthologous groups with 8,178 pairwise significant associations, by comparing 38 genomes (Snel *et al.*, 2002). Among them, 88% of the 516 small, disjointed clusters, containing 2.7 orthologous groups on average, have a more homogeneous functional composition, in terms of the COG functional category. They are regarded as functional modules.
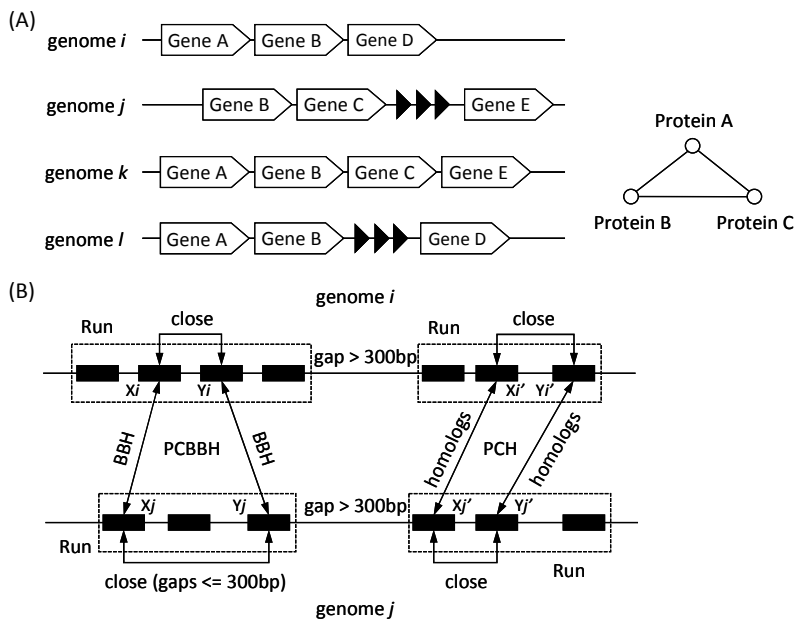
Different boxes signify different genes. The triangles represent genes that lack a conserved gene order. Protein A, protein B, and protein C, which line up in the same order among different organisms, are considered to be functionally associated.
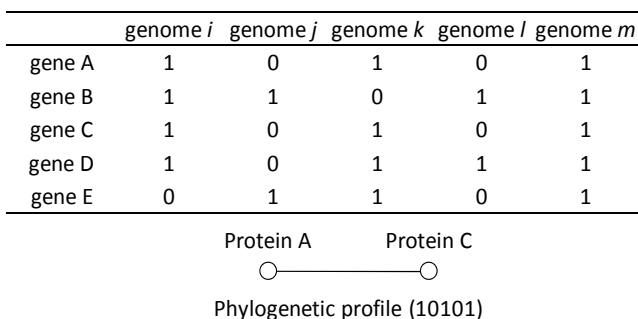
Fig. 2. Illustration of (A) the concept of the conservation of gene neighborhood approach and (B) the definitions of BBH, PCBBH, and PCH (pairs of close homologs).

2.3 Phylogenetic profile

This approach is based on a concept derived from the lineage-specific gene loss. The genes encoding proteins that interact with each other co-occur in different genomes. If one gene is absent in a genome, and then the other gene that interacts with it also is lost. On the basis of this hypothesis, the phylogenetic profile approach infers PPIs from genome comparisons. The phylogenetic profile approach is based on the co-occurrence of gene pairs, while the conservation of gene neighborhood approach is based on the gene order or co-occurrence of genes. The advantage is that it is applicable to Eukaryote, since it is not necessary to consider operon structure. In addition, this approach is different from the prediction method based on the operon, in that the rate of predicted genes that belong to the same biological process is higher. The approach has two disadvantages. The first point is that the analysis targets are limited to the organisms with completely sequenced genomes, because whether a certain gene is actually encoded in the genome must be known. The second is that this approach is not applicable for the proteins encoded in all organisms that are analysis subjects.

The functional relationship between two genes is detected by comparing their phylogenetic profiles (Fig. 3) (Pellegrini *et al.*, 1999). A phylogenetic profile is constructed for each protein, as a vector of N elements, where N is the number of genomes. Each position of the profile represents whether the protein that is homologous to the target protein is absent (signified by 0) or present (1) in each genome. Consequently, the phylogenetic distribution is shown by a long binary number along with each genome. A functionally related protein pair is detected by searching for the same phylogenetic distribution patterns. This method is applicable to domains as well as proteins (Pagel *et al.*, 2004).

Pellegrini *et al.* applied a phylogenetic profile approach to the *Escherichia coli* genome and 16 other fully sequenced genomes, in order to predict the functions of uncharacterized proteins. When the function of a protein is assumed to be the same as that of its neighbors in the phylogenetic-profile space, 18% of the neighbor keywords overlapped the known keywords of the query protein. This indicates that the phylogenetic profile approach has the ability to assign functions to uncharacterized proteins.

	genome *i*	genome *j*	genome *k*	genome *l*	genome *m*
gene A	1	0	1	0	1
gene B	1	1	0	1	1
gene C	1	0	1	0	1
gene D	1	0	1	1	1
gene E	0	1	1	0	1

Protein A   Protein C
○————————————○
Phylogenetic profile (10101)

Protein A and protein C are considered to interact with each other, since they have the same profile (10101).

Fig. 3. An example of the phylogenetic profile approach.

2.4 Mirror tree

Pairs of physically contacting proteins co-evolve, such as insulin and its receptors (Fryxell, 1996). Co-evolution refers to the phenomenon in which the evolution found in one protein has a considerable effect on the evolution of its partner protein, in order to maintain the protein interaction. Therefore, the amino acid substitutions are expected to occur at the same time in the interacting proteins. As a result, the two phylogenetic trees drawn for the interacting proteins show a greater degree of similarity than those drawn for proteins without interactions (Goh *et al.*, 2000). The mirror tree approach infers two protein/domain interaction pairs, using the similarity between the phylogenetic trees as an indicator. The advantage of this approach is that it can be applied to an organism whose genome has not been completely sequenced. Conversely, the approach is not applicable to a gene that shows a species-specific loss. In addition, the applications of this approach are limited to the cases where high-quality and complete multiple sequence alignments, including sequences from the common organisms, can be obtained.

The similarity between two proteins/domains can be quantified as follows (Fig. 4) (Pazos and Valencia, 2001). First, for two proteins or domains, the multiple sequence alignments are built using orthologous proteins that are collected from N organisms. Next, the distant matrices are constructed from the genetic distances among all sequences, based on the multiple sequence alignment. The correlation coefficient between the two distance matrices is calculated. The value can be considered as an indicator that shows the intensity of co-evolution. Hence, if the value is close to one, it is judged that the two phylogenetic trees , and the two proteins are considered to interact. The mirror tree approach does not depend on the method used to construct the phylogenetic tree, since it does not compare them directly.
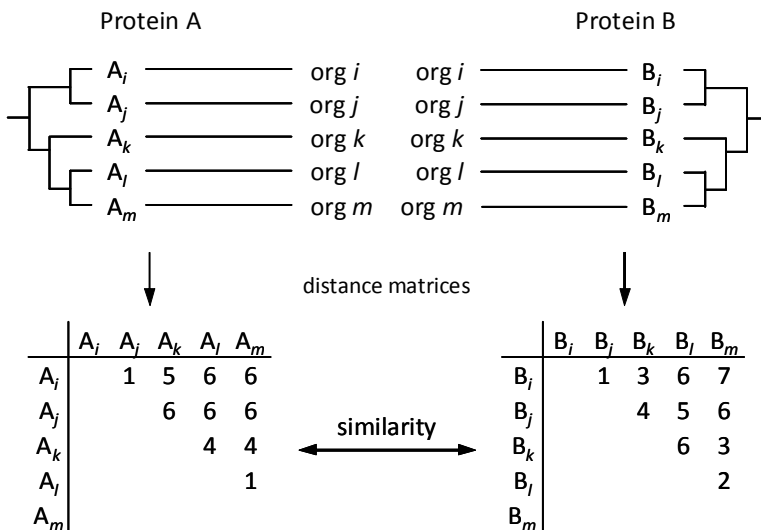
The trees have the same number of leaves and the same organisms in the leaves.

Fig. 4. The flow of the mirror tree approach.

The mirror tree method was applied to six protein families of ligand-receptor pairs, to predict the interaction partners (Goh and Cohen, 2002). Consequently, 79% of all known binding partners on average were detected. In addition, potentially new binding partner in the syntaxin/Unc-18 protein and TGF-β/TGF-β receptor families were found among previously characterized proteins.

2.5 *In silico* two-hybrid system

The *in silico* two-hybrid system infers physical contact sites by computing the correlation coefficient of amino acid variation between two sites, using the multiple sequence alignments for protein pairs (Göbel *et al.*, 1994). That is, in the residue pairs that are in physical contact or related functionally, the amino acids tend to change at the same time. This type of correlated mutation is called co-variation. The similarity of the variation patterns is thought to be related to compensatory mutation. The *in cilico* two-hybrid system infers PPIs by expanding this concept. This system can detect an interaction accompanied by physical contact, and estimates the protein binding sites as well as interacting protein pairs. Meanwhile, the main limitation of this system is the requirement of high quality alignments that include a wide range of common organisms encoding the two proteins, in the same manner as the mirror tree approach.

The *in silico* two-hybrid system quantifies the degree of co-variation between pairs of residues (Fig. 5) (Pazos and Valencia, 2002). First, a multiple sequence alignments are built
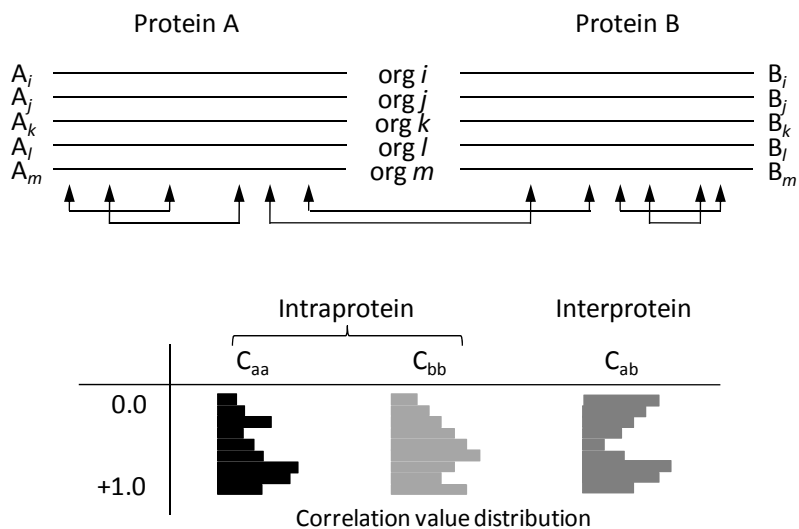
On the top, the alignments are built for two different proteins (protein A and protein B), including the corresponding sequences from different organisms (org $i, j, k...$). On the bottom, the distributions of the correlated coefficients for the pair of residues internal to the two proteins (C_{aa} and C_{bb}) and for the pair of residues from each of the two proteins (C_{ab}) are represented.

Fig. 5. A schematic representation of the *in silico* two-hybrid system.

using orthologs derived from the common organism for two proteins. Next, the correlation coefficients between all combinations of sites in a protein are computed, and the frequency distribution of the values that are computed between two sites is investigated. Similarly, the correlation coefficients between all combinations of two sites from different proteins are calculated, and the frequency distribution is computed. Finally, the interaction index score is computed by using the three frequency distributions of correlation coefficients. If the value is close to one, then the two proteins are considered to interact.

Pazos *et al.* applied this system to four test sets: 1) 14 two-domain proteins with a tight intradomain interaction, from the PDB, 2) 53 proteins including 31 known interactions, 3) 195 pairs with 15 possible interactions, derived from 749 predicted interactions, and 4) 321 pairs, 17 of which are known to interact, from the SPIN database. As a result, it discriminated between true and false interactions in a significant number of cases.

2.6 Sequence signature

The sequence signature approach, which predicts interacting proteins based on domain information, has developed separately from the methods using genome comparison or protein sequence analysis. This approach utilizes sequence and/or structure motifs in order to discriminate interacting proteins. In this approach, the characteristic pairs of sequence signatures are prepared from a database including experimentally determined interacting proteins, where one protein contains one sequence-signature and its interacting partner contains the other sequence-signature. The pairs that occur with high frequency are termed "correlated sequence-signatures", and they can be used for the prediction of putative interacting partners. The prediction result provides the pairs of protein/domain groups that include the correlated sequence-signature, while the other methods described above predict one-on-one protein pairs. Combining this approach with other techniques can yield higher performance.

In this approach, the sequence-signature of the signature combinations must be constructed to identify the correlated sequence-signatures (Fig. 6) (Sprinzak and Margalit, 2001). First, the experimentally determined interacting protein pairs are collected. Then, the sequence-signatures, defined by a motif database such as InterPro, are identified for each sequence. Each entry (*a*,*b*) in the table shows the number of protein pairs, composed of one protein containing signature *a* and its partner containing signature *b*. Next, the occurrence frequencies of the sequence-signature are converted into the log-odds. The sequence-signature with a positive log-odds value is considered to be observed more frequently in the interactive pairs. Therefore, they are regarding as having a correlated sequence-signature. Finally, this approach searches for the protein or domain pairs that contain the correlated sequence-signature.

An example of applying the Myb domain and the Bromodomain that are correlated sequence-signature to the yeast *S. cerevisiae* is shown (Sprinzak and Margalit, 2001). There are 19 and 10 protein sequences containing the Myb domain and the Bromodomain, respectively. Therefore, in this case, 190 protein interaction pairs are predicted, out of which five interactions were already known.
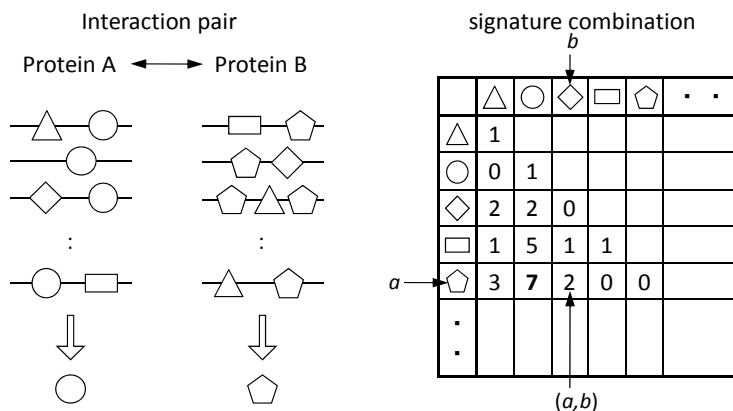
In the left panel, each row contains the sequences of the pair of protein A and protein B. Each sequence has a sequence-signature, illustrated by shapes. In the right panel, a contingency table of the signature combination is described, where each entry (*a*,*b*) in the table shows the number of protein pairs. For example, the sequence-signature pair represented by a square and a pentagon appears in two pairs of interacting proteins. The most abundant pair of sequence-signatures is indicated by bold type.

Fig. 6. A scheme for detecting correlated sequence-signatures in interacting proteins.

2.7 Supervised classification

The PPI prediction can be defined as a binary classification problem. Therefore, a statistical model or machine learning method can be applied to the problem of determining whether a pair of proteins is interacting or non-interacting. The K-Nearest Neighbor (KNN) (Qi *et al*., 2006), Naïve Bayesian (NB) (Jansen *et al*., 2003; Lu *et al*., 2005), support vector machines (SVM) (Lo *et al*., 2005), Artificial Neural Networks (ANN) (Ma *et al*., 2007), and Random Forest (RF); (Chen and Liu, 2005; Qi *et al*., 2005) methods were previously applied to this problem. The advantage of these methods is to use data that integrated different datasets. Datasets that do not directly measure PPI, such as sequence and structure information, can be used to infer PPIs. Conversely, the weak point is that the predictive performance varies widely, depending on the quality of the dataset and the selection of statistical methods.

In a statistical model, protein pairs are expressed by N dimensional vectors, where N is the number of features. For example, gene co-expression, GO biological process similarity, MIPS functional similarity, and essentiality are used as features in Jansen's work (Jansen *et al*., 2003). In addition, sequence information, such as homology and domain data, is used. Two points must be considered when the prediction model is built. The first is that it is necessary to pay attention to the quality of the experimental data used for training and evaluating statistical model, since the performance of the prediction model strongly depends on them. A high-throughput experimental method, such as Yeast Two-Hybrid (Y2H), Mass Spectrometry and Tandem Affinity Purification (MS TAP), and gene co-expression, can detect proteomic-wide PPIs, yielding vast amounts of protein interaction data within the cell. However, these data are often noisy, incomplete, and low-reproducible, since they contain contradictory values. The second is that the selection of an appropriate classification technique is an important task.

The statistical model was developed to infer PPIs in the human and yeast genomes (Lee *et al.*, 2004; Rhodes *et al.*, 2005). Qi *et al.* applied six different classifiers (RF, KNN, NB, Decision Tree, Logistic Regression, and SVM) to predict PPIs, and among them, the RF classifier exhibited the highest performance (Qi *et al.*, 2006). In addition, gene expression is the most important feature for prediction.

3. Computational methods to infer protein flexibility

A protein molecule is not a rigid body. The scale of protein motions is very broad: motions range from local fluctuations, such as those seen in loop regions, to global ones involving changes in the relative positions of rigid domains. Protein motion is often necessary for proteins to perform their specific biological functions. For example, a protein possesses certain conformations in order to interact with its partner protein in many cases. Therefore, structural flexibility is an important feature to consider for understanding protein functions.

Experimental methods that analyze protein dynamics have been developed. Nuclear magnetic resonance (NMR) is a powerful experimental technique (Williams, 1989). NOEs and relaxation experiments provide information related to picosecond-microsecond-scale motions of the backbone atoms (Chill *et al.*, 2004; Gitti *et al.*, 2005). Also, model-free analysis enables quantitative determination of the fluctuations and slow conformational changes of the backbone amide vectors (Lipari and Szabo, 1982a; Lipari and Szabo, 1982b). Although NMR provides a detailed view of protein dynamics, it is time-consuming and suffers from size limitations.

In contrast, computational methods are useful to calculate the dynamics of proteins for which structures are available. They are divided into two types of method. One method compares the structures of a protein crystallized under different conditions or different conformers obtained by NMR. The structural differences indicate flexible regions (Shatsky *et al.*, 2002; Ye and Godzik, 2004). Another computational method is to simulate protein dynamics by methods such as Normal Model Analysis (NMA) and Molecular Dynamics (MD). With the increasing number of available protein structures and the development of high-performance computers, databases that treat protein dynamics have been developed (Table 2). Some databases are introduced below.

ProMode

ProMode is a database including NMA results from analyses performed with a full-atom model for many proteins. It displays realistic three-dimensional motions at an atomic level, using a free plug-in, Charm. In addition, the dynamic domains and their mutual screw motions defined from NMA results are displayed.

MolMovDB

The database of macromolecular movements (MolMovDB) is a collection of quantitative data for flexibility and a number of graphical representations. The motions are generated from alignments of pairs of structures from the Protein Data Bank (PDB). The motions are divided into various classes (e.g. 'hinged domain' or 'allosteric'), according to the type of conformational change.

DynDom database

DynDom, a domain motion analysis program, analyzes the conformational change in terms of dynamic domains, interdomain screw axes, and interdomain bending regions, by comparing two structures when at least two X-ray conformers are available. The DynDom database displays details on the conformational changes obtained from the DynDom analysis results.

iGNM

The database contains visual and quantitative information on the collective modes predicted by the Gaussian Network Model (GNM) for the structure in the PDB. The output includes the equilibrium fluctuations of residues and comparisons with X-ray crystallographic B-factors, the sizes of residue motions in different collective modes, the cross-correlations between the residue fluctuations or domain motions, and other useful information.

Database name	HTTP address	Description	Reference
ProMode	http://cube.socs.waseda.ac.jp/pages/jsp/index.jsp	Large-scale collection of animations of the normal mode vibrating proteins with the full-atom models.	Wako *et al.*, 2004
MolMovDB	http://www.molmovdb.org/	Visualization and classification of molecular motions according to their size and mechanism.	Echols *et al.*, 2003
DynDom database	http://fizz.cmp.uea.ac.uk/dyndom/	Collection of domains, hinge axes and hinge bending residues in proteins.	Lee *et al.*, 2003
*i*GNM	http://ignm.ccbb.pitt.edu/	Static and animated images for describing the conformational mobility of proteins by computing the GNM dynamics.	Yang *et al.*, 2005

Table 2. List of databases that deal with conformational changes.

4. PPI prediction from protein flexibility

A Structural flexibility is an important characteristic of protein that is frequently related to their functions, as reviewed in section 3. Flexible regions are often necessary for proteins to bind a ligand or another protein. When we focus on the motion of a protein backbone segment, the movement can be classified conceptually into two forms: internal motion and external motion (Nishikawa and Go, 1987). An internal motion is the deformation of the segment itself, while an external motion involves only rotational and translational motions, as a rigid body. The segment fluctuates as a rigid body by changes in the dihedral angles of the flanking residues. For this reason, internal and external motions are considered to be fundamentally different.

This section introduces a means for the calculation of internal and external motions in a protein, by the constriction of statistical models, called "FlexRetriver", and its application to PPI data (Hirose *et al.*, 2010).

4.1 Development of a method for predicting internal and external motions

This subsection introduces the RF-based method for predicting the internal and external motions defined by the NMA from the sequence information.

4.1.1 Calculation of internal and external motions

Using FEDER/2 (Wako *et al.*, 2004), the NMA was performed for the energy-minimized conformation, with the PDB data as the starting conformation. In the NMA, the mean-square displacement of atom a, $<D_a^2>$, in the thermal fluctuations is given as the sum of contributions from individual modes

$$< D_a^2 >= \sum_{k=1}^{N} D_{ak}^2 ,$$

where D_{ak} is the displacement vector of atom a in the k-th normal mode, and N is the number of dihedral angles used as independent variables, i.e., the number of normal modes.

In this study, two conformations for a nine-residue segment in each normal mode are considered. The displacement vector of atom a by this purely translational and rotational motion is designated as D_{ak}^e, and the residual one is designated as D_{ak}^i. Then, D_{ak} is decomposed as

$$D_{ak} = D_{ak}^e + D_{ak}^i .$$

The superscripts e and i respectively stand for external and internal. The mean square deviation of atom a is given as

$$< D_a^2 >= \sum_k \left| D_{ak}^e \right|^2 + \sum_k \left| D_{ak}^i \right|^2 + \sum_k 2 D_{ak}^e \cdot D_{ak}^i$$

$$= < \left| D_a^e \right|^2 > + < \left| D_a^i \right|^2 > + 2 < \left| D_a^e \cdot D_a^i \right| > .$$

The third term on the right-hand side of this equation is usually much smaller than the first two terms. Therefore, the mean-square deviation of atom a is decomposed approximately into external (first term) and internal (second term) ones. In this case, we are interested in the main-chain fluctuation; for simplicity, only the C_α atom in this decomposition is considered. This means that we selected data for the C_α atoms from the results obtained using NMA with a full-atom model.

4.1.2 Dataset

The dataset was created by selecting protein chains from ProMode, as follows. Proteins with a root mean square deviation (RMSD) of more than 2Å between the energy-minimized structure and the PDB structure were excluded. Protein chains with redundant SCOP IDs were excluded, multi-domain proteins defined by SCOP were then removed. Next, some proteins were discarded so that the maximum pairwise sequence identity was limited to 25%. The resulting dataset comprised 481 chains (87,236 residues).

We calculated the internal and external motions using NMA with a full-atom model for all proteins in the dataset. Raw NMA values were normalized to correct for the variability among proteins in the dataset.

4.1.3 Structure-specific protein mobility propensity

The protein mobility propensities of amino acids are associated with their secondary structures and accessible surface areas (ASAs). The protein mobility propensity was divided into three types of protein mobility: the high and low groups comprised amino acids with normalized NMA scores higher than 1 and lower than -1, respectively, while the normal group comprised amino acids with normalized NMA scores between 1 and -1. The structure-specific protein mobility propensity ($SpecProg(n,s,g)$) was calculated as

$$SpecProp(n,s,g) = log_2\ freq(n,s,g)/freq(n,s)\ ,$$

where $freq(n,s,g)$ and $freq(n,s)$ respectively represent the relative frequencies of amino acid n in the g protein mobility group of the s state dataset and in the s dataset. The s state indicates a secondary structure or ASA.

The results of the structure-specific protein mobility propensity are shown in Fig. 7. For most amino acids, the protein mobility propensity pattern (in the high, normal, and low groups in the same type of secondary structure) depended on the type of secondary structure (Fig. 7(A)). For example, for both motions, the high mobility propensity of proline (Pro) was low in alpha helices and beta sheets, but high in other structures. This might be because Pro is a secondary structure breaker and its amide nitrogen cannot form a hydrogen bond. On the other hand, the low mobility propensity of hydrophobic amino acids tended to be high in
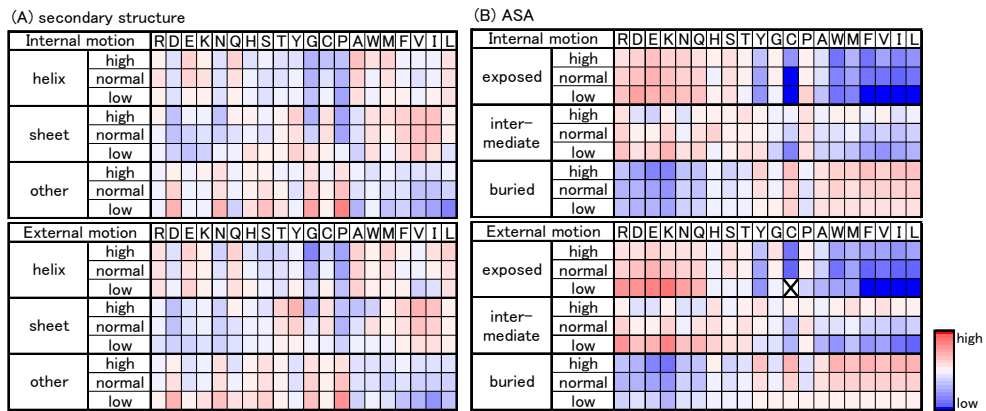
The upper and lower tables represent the results of internal and external motions, respectively. The terms high, normal, and low stand for the protein mobility of the high, normal and low states, respectively. The protein mobility propensity is colored with a gradient from negative (blue) to positive (red). The other in secondary structure is a region without helix and sheet. The cross mark signifies that no data exist.

Fig. 7. The protein mobility propensity associated with (A) secondary structures and (B) ASA.

alpha helices and beta sheets, but low in other structures. The distribution of high mobility propensity in other structures is similar to that of the propensity in hinge regions (Flores *et al.*, 2007). Similarly, the protein mobility propensity pattern changes, depending on the ASA (Fig. 7(B)). The high mobility propensity became higher with increasing ASA, as seen for hydrophilic amino acids. In contrast, high mobility propensity of hydrophobic amino acids was lower with increasing ASA. The external motion might be more strongly influenced by the ASA, as compared to the internal motion. Altogether, these results strongly suggest that the secondary structure and the ASA influence the degrees of the internal and external motions.

4.1.4 Construction of a prediction method

A method for predicting internal and external motions was built by applying RF, which is a type of supervised classification algorithm. The sequence in the sliding window (with sizes of 11 and 17 residues for internal and external motions, respectively) was encoded by using paired amino acid information, corresponding to the variable. The variables were obtained by adding two features, which are derived from the amino acid pairs of the central amino acid with the other amino acids in the window. In total, 18 features were defined, and they were divided into four groups, designated as physicochemical, mobility, secondary structure (predicted by psipred (Jones, 1999) or PHD (Rost, 1996)), and ASA (predicted by sable (Adamczak *et al.*, 2005) or RVPnet (Ahmad *et al.*, 2003)). The profile-based predictors (psipred and sable) have higher prediction accuracy than the amino-acid propensity-based predictors (PHD and RVPnet). The value of a feature of an amino acid was set to one if the amino acid satisfied a feature's definition, and to zero otherwise.

The RF algorithm was used to build a prediction model for classifying amino acids into the three classes: flexible, intermediate, and rigid. Three RF prediction models were trained for the three categories of window location: the center of a secondary structure (CS), the remote area from a secondary structure (RS), and the periphery of a secondary structure (PS). The RF prediction models classified the windows into the three classes, and their prediction results were attributed to the central residue in the window. The results of the classification obtained from the RF were then converted into a score.

4.1.5 Prediction performance

The prediction results were assessed on a residue basis, by which the predicted score in the sequence was compared to the normalized NMA score. The prediction performance was evaluated by using three criteria: the mean absolute error (MAE), correlated coefficient (CC), and Receiver Operating Characteristic (ROC) curves. The MAE was defined as the absolute difference between two values. The MAE value approaches 0 as the prediction improves. The CC was also computed between two values. The CC ranges from -1 to 1, and a large, positive value represents a better prediction. The ROC curve was obtained by plotting the false positive rate against the true positive rate. A larger area under the ROC curve (AUC) indicates a more robust algorithm.

The prediction performance of FlexRetriever was compared with those of three published methods (PROFbval (Schlessinger *et al.*, 2006), POODLE-S (Shimizu *et al.*, 2007), FlexPred

(Kuznetsov and McDuffie, 2008)), and the naïve model. The naïve model is based on the simple idea that protein motion tends to be large in a coil or loop region and small in a secondary structure. The FlexRetriver, which implemented psipred and sable, yielded the lowest MAE and the highest CC among all prediction methods for both motions (Table 3). However, it is noteworthy that the distribution of CC varied widely. Additionally, in AUC, FlexRetriever exhibited the best performance among all methods.

(A) Internal motion			
Method	MAE	CC	AUC
FlexRetriever (PHD & RVPnet)	0.621	0.482	0.765
FlexRetriever (psipred & sable)	<u>0.605</u>	<u>0.525</u>	<u>0.786</u>
Naïve model (PHD)	0.988	0.248	0.653
Naïve model (psipred)	0.952	0.293	0.672
PROFbval	0.743	0.367	0.693
POODLE-S	-	-	0.730
FlexPred	-	-	0.741
(B) External motion			
Method	MAE	CC	AUC
FlexRetriever (PHD & RVPnet)	0.571	0.541	0.777
FlexRetriever (psipred & sable)	<u>0.542</u>	<u>0.597</u>	<u>0.806</u>
Naïve model (PHD)	0.970	0.262	0.661
Naïve model (psipred)	0.929	0.320	0.681
PROFbval	0.608	0.547	0.784
POODLE-S	-	-	0.783
FlexPred	-	-	0.777

The CC and MAE were estimated by performing a five-fold cross validation test. The highest scores in each criterion are underlined. PHD and psipred in parentheses signify the secondary structure predictor. Similarly, RVPnet and sable represent the ASA predictor. "-": scores could not be calculated.

Table 3. Comparison of prediction performance.

4.2 Applying FlexRetriever to PPI data

In this study, we utilized the set of 20 proteins that undergo large conformational changes upon association (> 2Å C_α RMSD) created by Dobbins *et al.*, with which they demonstrated the relationship between normal mode fluctuations and conformational change (Dobbins *et al.*, 2008). They regarded protein motions as being associated with their functions, because they are observed along with the PPI. We compared the internal motion with the observed conformational change region, because it was defined as the deformation of a segment itself. To begin with, we present three typical results, in which the observed conformational change regions are located in a binding site, a hinge region, and other regions. We will then discuss the overall results.

i.   Ecotin

Ecotin, a homodimeric protein, is an inhibitor of a group of homologous serine proteases, such as trypsin, chymotrypsin, and elastase. One dimeric inhibitor binds to a protease molecule. From a comparison of two structures determined with different crystalline environments, an inherent flexible loop was identified in the binding site with trypsin. It was necessary for its inhibitory function (Shin *et al.*, 1996). FlexRetriever predicted high internal motion for the corresponding loop (Fig. 8(A)).

ii.   Fab fragment

The fragment antigen binding (Fab fragment) region is the site where an antibody binds to antigens. It is a heterodimer of the heavy and light chains in each of the two composed domains. The hinge region between the two domains changed its conformation when Fab bound to hemagglutinin derived from a flu virus (Fleury *et al.*, 1998). FlexRetriever predicted high internal motion at the hinge region in each chain (Fig. 8(B)).

iii. Erythropoietin

Erythropoietin (EPO) is a hormone produced primarily in the kidneys. It has a four-helical bundle topology with two long loops, and binds to the extracellular domain of the EPO receptor. The CD loop, which is located in a region remote from the binding site, changed its conformation (Cheetham *et al.*, 1998). FlexRetriever predicted high internal motion for the corresponding loop (Fig. 8(C)).
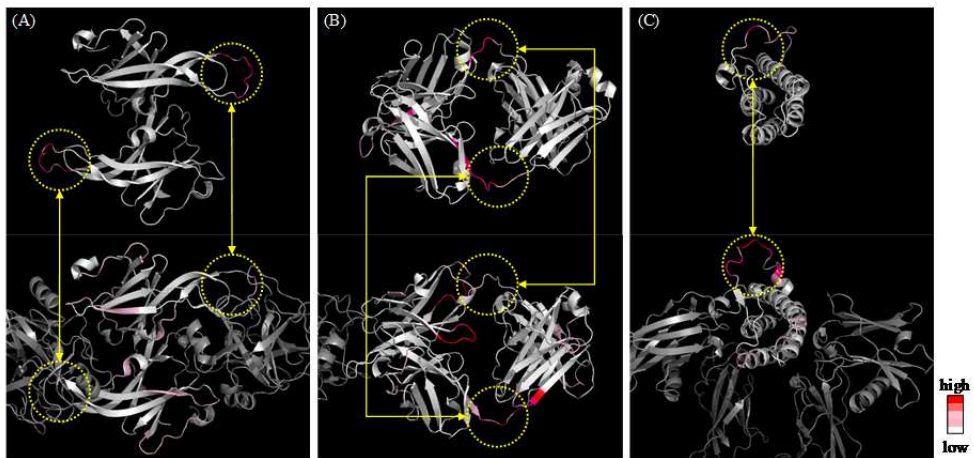
The observed degrees of conformational change and the predicted scores for internal motion are mapped, respectively, with a gradient from zero (white) to a high score (dark red) onto their structures in the upper and lower sections. The regions enclosed with a yellow dotted line are the regions with observed conformational changes. The free-state and complex-state structures are displayed, respectively, in the upper and lower sections.

Fig. 8. Example of the relationship between the predicted internal motions and the observed conformational changes of (A) ectin, (B) Fab fragment, and (C) erythropoietin.
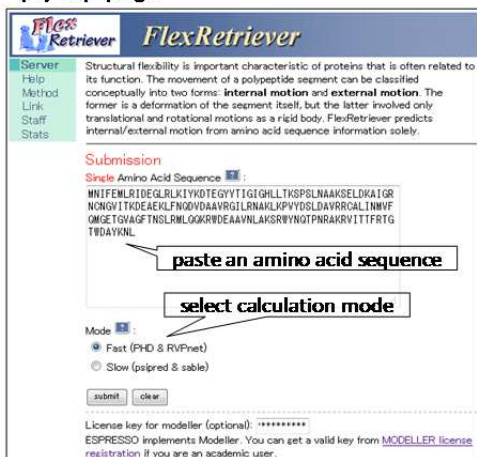
Overall results

When FlexRetriever was applied to a set of 20 proteins, three or more consecutive residues with high internal scores were regarded as candidates for the regions undergoing conformational changes. From a comparison between the observed conformational change regions with the predicted high internal motion regions, at least one overlap was found in 85% of the proteins studied. If the analysis object was limited to the 16 proteins that interact with only one partner, then the overlap was observed in 15 proteins (94% of the proteins studied). These observations suggest that FlexRetriever is a sensitive method for the detection of protein motions related to PPIs, including binding sites.

4.3 Web server

The presented method is implemented in the FlexRetriever server, which has been designed with a user-friendly interface to provide easily interpretable prediction results.

The server accepts the submission of a single amino acid sequence with less than 1,000 amino acids in the FASTA format (Fig. 9(A)). The user is asked to choose a calculation mode.
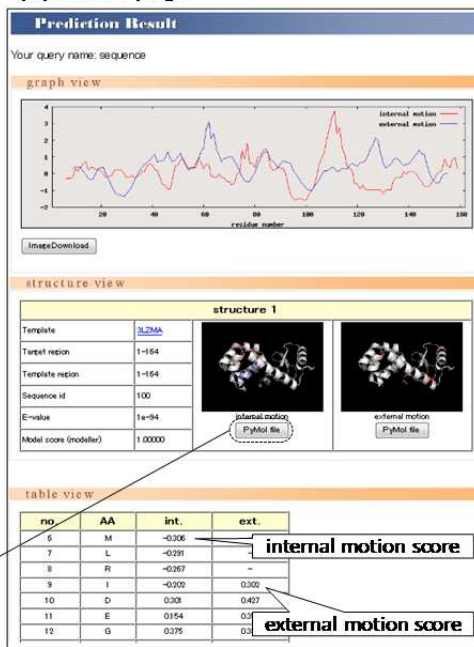
On the result page, a graph is displayed on the top, and two structures on which the scores of the internal and external motions are mapped are shown in the middle. They can be downloaded as a PyMol file. The table with the raw scores is displayed below the structures.

Fig. 9. Images of FlexRetriever's (A) top page and (B) result page.

The calculation time of the fast mode, which uses PHD for secondary structure prediction and RVPnet for ASA prediction, is shorter than that of the slow mode, but its performance is poorer.

The results page is divided into three sections (Fig. 9(B)). The first section (graph view) provides the graph which contains the prediction results of both motions. The second section (structure view) presents the degrees of internal and external motions on the three-dimensional structure. The third section (table view) lists the amino acids and the raw scores of their internal and external motions.

FlexRetriever is available at http://mbs.cbrc.jp/FlexRetriever and is free.

5. Conclusion

This chapter provides an overview of the computational methods to infer pairs of interacting proteins and to study the relevance of protein flexibility. Genomic information and experimental data are now readily available, and thus computational methods will become more important tools in the field of analyzing or inferring PPIs. In addition, as a novel attempt to predict PPIs, we have presented an efficient algorithm for predicting flexible regions in proteins, and shown its application to PPIs. The tool is expected to be useful for inferring motions associated with PPIs.

6. References

Adamczak, R.; Porollo, A. & Meller, J. (2005) Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins*, Vol.59, No.3, (May 15), pp. 467-475, ISSN 0887-3585

Ahmad, S.; Gromiha, M.M & Sarai, A. (2003) Real value prediction of solvent accessibility from amino acid sequence. *Proteins*, Vol.50, No.4, (Mar 1), pp. 629-635, ISSN 0887-3585

Brown, F.; Zheng, H.; Wang, H.Y. & Azuaje, F. (2010) From experimental approaches to computational techniques: A review on the prediction of protein-protein interaction. *Advances in Artificial Intelligence*, Vol.2010, ISSN 16877470

Chill, J.H.; Quadt, S.R. & Anglister, J. (2004) NMR backbone dynamics of the human type I interferon binding subunit, a representative cytokine receptor. *Biochemistry*, Vol.43, No.31, (Aug 10), pp. 10127-10137, ISSN 0006-2960

Cheetham, J.C.; Smith, D.M.; Aoki, K.H.; Stevenson, J.L.; Hoeffel, T.J.; Syed, R.S.; Egrid, J. & Harvey, T.S. (1998) NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat Struct Biol*, Vol.5, No.10, (Oct 1998), pp. 861-868, ISSN 1072-8368

Chen, X.W. & Liu, M. (2005) Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics*, Vol.21, No.24, (Dec 15), pp. 4394-4400, ISSN 0973-2063

Dandekar, T.; Snel, B.; Huynen, M. & Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*, Vol.23, No.9, (Sep 1998), pp. 324-328, ISSN 0968-0004

Dobbins, S.E.; Lesk, V.I. & Sternberg, M.J. (2008) Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc Natl Acad Sci U S A*, Vol.105, No.30, (July 29), pp. 10390-10395, ISSN 0027-8424

Echols, N.; Milburn, D. & Gerstein, M. (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res*, Vol.31, No.1, (Jan 1), pp. 478-482, ISSN 0305-1048

Enright, A.J.; Iliopoulos, I.; Kyrpides N.C. & Ouzounis, C.A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature*, Vol.402, No.6757, (Nov 4), pp. 86-90, ISSN 0028-0836

Fleury, D.; Wharton, S.A.; Skehel, J.J.; Knossow, M. & Bizebard, T. (1998) Antigen distortion allows influenza virus to escape neutralization. *Nat Struct Biol*, Vol.5, No.2, (Feb 1998), pp. 119-123, ISSN 1072-8368

Flores, S.C.; Lu, L.J.; Yang, J.; Carriero, N. & Gerstein, M.B. (2007) Hinge Atlas: relating protein sequence to site of structural flexibility. *BMC Bioinformatics*, Vol.8, No.167, (May 22), ISSN 1471-2105

Fryxell, K.J. (1996) The coevolution of gene family trees. *Trends Genet*, Vol.12, No.9, (Sep 1996), pp. 364-369, ISSN 0168-9525

Gitti, R.K.; Wright, N.T.; Margolis, J.W.; Varney, K.M.; Weber, D.J. & Margolis, F.L. (2005) Backbone dynamics of the olfactory marker protein as studied by 15N NMR relaxation measurements. *Biochemistry*, Vol.44, No.28, (Jul 19), pp. 9673-9679, ISSN 0006-2960

Göbel, U.; Sander, C.; Schneider, R. & Valencis, A. (1994) Correlated mutations and residue contacts in proteins. *Proteins*, Vol.18, No.4, (Apr 1994), pp. 309-317, ISSN 0887-3585

Goh, C.S.; Bogan, A.A.; Joachimiak, M.; Walther, D. & Cohen, F.E. (2000) Co-evolution of proteins with their interaction partners. *J Mol Biol*, Vol.299, No.2, (Jun 2), pp. 283-293, ISSN 0022-2836

Goh, C.S. & Cohen, F.E. (2002) Co-evolutionary analysis reveals insights into protein-protein interactions. *J Mol Biol*, Vol.324, No.1, (Nov 15), pp. 177-192, ISSN 0022-2836

Hirose, S.; Yokota, K.; Kuroda, Y.; Wako, H.; Endo, S.; Kanai, S. & Noguchi, T. (2010) Prediction of protein motions from amino acid sequence and its application to protein-protein interaction. *BMC Struct Biol*, Vol.10, No.20, (Jul 13), ISSN 1472-6807

Jansen, R.; Yu, H.; Greenbaum, D.; Kluger, Y.; Krogan, N.J.; Chung, S.; Emili, A.; Snyder, M.; Greenblatt, J.F. & Gerstein, M. (2003) A Baysian networks approach for predicting protein-protein interactions from genomic data. *Science*, Vol. 302, No.5644, (Oct 17), pp. 449-453, ISSN 0036-8075

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, Vol.292, No.2, (Sep 17), pp. 195-202, ISSN 0022-2836

Kuznetsov, I.B. & McDuffie M. (2008) FlexPred: a web-server for predicting residue positions involved in conformational switches in proteins. *Bioinformation*, Vol.3, No.3, (Nov 5), pp. 134-136, ISSN 0973-2063

Lee, I.; Date, S.V.; Adai, A.T. & Marcotte E.M. (2004) A probabilistic functional network of yeast genes. *Science*, Vol.306, No.5701, (Nov 26), pp. 1555-1558, ISSN 0036-8075

Lee, R.A.; Razaz, M. & Hayward, S. (2003) The DynDom database of protein domain motions. *Bioinformatics*, Vol.19, No.10, (Jul 1), pp.1290-1291, ISSN 1460-2059

Lipari, G. & Szabo, A. (1982a) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*, Vol.104, No.17, (August 1982), pp. 4546-4559, ISSN 0002-7863

Lipari, G. & Szabo, A. (1982b) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc*, Vol.104, No.17, (August 1982), pp. 4559-4570, ISSN 0002-7863

Lo, S.L.; Cai, C.Z.; Chen, Y.Z. & Chung, M.C. (2005) Effect of training datasets on support vector machine prediction of protein-protein interactions. *Proteomics*, Vol.5, No.4, (Mar 2005), pp. 876-884, ISSN 1615-9861

Lu, L.J.; Xia, Y.; Paccanaro, A.; Yu, H. & Gerstein, H. (2005) Assessing the limits of genomic data integration for predicting protein networks. *Genome Res*, Vol.15, No.7, (Jul 2005), pp. 945-953, ISSN 1088-9051

Ma, Z.; Zhou, C.; Lu, L.; Lu, L.; Ma, Y.; Sun, P. & Cui, Y. Predicting protein-protein interactions based on BP neural network, *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops*, pp. 3-7, ISBN 978-1-4244-1604-2, Fremont, CA, Nov 2-4, 2007

Marcotte, E.M.; Pellegrini, M.; Ng, H.L.; Rice D.W.; Yeates, T.O. & Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequence. *Science*, Vol.285, No.5428, (Jul 30), pp. 751-753, ISSN 0036-8075

Nishikawa, T. & Go, N. (1987) Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins*, Vol.2, No.4, (1987), pp. 308-329, ISSN 0887-3585

Overbeek, R.; Fonstein, M.; D'Souza, M.; Pusch, G.D. & Maltsev, N. (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A*, Vol.96, No.6, (Mar 16), pp. 2896-2901, ISSN 0027-8424

Pagel, P.; Wong, P. & Frishman, D. (2004) A domain interaction map based on phylogenetic profiling. *J Mol Biol*, Vol.344, No.5, (Dec 10), pp. 1331-1346, ISSN 0022-2836

Pazos, F. & Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Engineering*, Vol.14, No.9, (Sep 14), pp. 609-614, ISSN 0269-2139

Pazos, F. & Valencia, A. (2002) *In silico* two-hybrid system for the selection of physically interacting protein pairs. *Proteins*, Vol.47, No.2, (May 1), pp. 219-227, ISSN 0887-3585

Pellegrini, M.; Marcotte, E.M.; Thompson, M.J.; Eisenberg, D. & Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*, Vol.96, No.8, (Apr 13), pp. 4285-4288, ISSN 0027-8424

Rost, B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol*, Vol.266, (1996), pp. 525-539, ISSN 0076-6879

Qi, Y.; Klein-Seetharaman, J & Bar-Joseph, Z. (2005) Random forest similarity for protein-protein interaction prediction from multiple sources. *Pac Symp Biocomput*, pp. 531-542, ISSN 1793-5091

Qi, Y.; Bar-Joseph, Z. & Klein-Seetharaman, J. (2006) Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins*, Vol.63, No.3, (May 15), pp. 490-500, ISSN 0887-3585

Rhodes, D.R.; Tomlins, S.A.; Varambally, S.; Mahavisno, V.; Barrette, T.; Kalyana-Sundaram, S.; Ghosh, D.; Pandey, A. & Chinnalyan, A.M. (2005) Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol*, Vol.23, No.8, (Aug 2005), pp. 951-959, ISSN 1087-0156

Schlessinger, A.; Yachday, G. & Rost, B. (2006) PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics*, Vol.22, No.7, (Apr 1), pp. 891-893, ISSN 1460-2059

Shatsky, M.; Nussinov, R. & Wolfson, H.J. (2002) Flexible protein alignment and hinge detection. *Proteins*, Vol.48, No.2, (Aug 1), pp. 242-256, ISSN 0887-3585

Shimizu, K.; Hirose, S. & Noguchi, T. (2007) POODLE-S: web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix. *Bioinformatics*, Vol.23, No.17, (Sep 1), pp. 2337-2338, ISSN 1460-2059

Shin, D.H.; Song, H.K.; Seong, I.S.; Lee, C.S.; Chung C.H & Suh, S.W. (1996) Crystal structure analyses of uncomplexed ecotin in two crystal forms: implications for its function and stability. *Protein Sci*, Vol.5, No.11, (Nov 1996), pp. 2236-2247, ISSN 0961-8368

Shoemaker, B.A. & Panchenko, A.R. (2007) Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol*, Vol.3, No.4, (Apr 27), pp. e43, ISSN 1553-734X

Skrabanek, L.; Saini, H.K.; Bader, G.D. & Enright, A.J. (2008) Computational prediction of protein-protein interactions. *Mol Biotechnol*, Vol.38, No.1, (Jan 2008), pp. 1-17, ISSN 1073-6085

Snel, B.; Bork, P. & Huynen, M.A. (2002) The identification of functional modules from the genomic association of genes. *Proc Natl Acad Sci U S A*, Vol.99, No.9, (Apr 30), pp. 5890-5895, ISSN 0027-8424

Sprinzak, E. & Margalit, H. (2001) Correlated sequence-signatures as markers of protein-protein interaction. *J Mol Biol*, Vol.311, No.4, (Aug 24), pp.681-692, ISSN 0022-2836

Valencia, A. & Pozos, F. (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol*, Vol.12, No.3, (Jun 2002), pp. 368-373, ISSN 0959-440X

Wako, H.; Kato, M. & Endo, S. (2004) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, Vol.20, No.13, (Sep 1), pp. 2035-2043, ISSN 1460-2059

Williams, R.J. (1989) NMR studies of mobility within protein structure. *Eur J Biochem*, Vol.183, No.3, (Aug 15), pp. 479-497, ISSN 0014-2956

Yang, L.W.; Liu, X.; Jursa, C.J.; Holliman, M.; Rader, A.J.; Karimi, H.A. & Bahar, I. (2005) *i*GMN: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics*, Vol.21, No.13, (Jul 1), pp. 2978-2987, ISSN 1460-2059

Ye, Y. & Godzik, A. (2004) Database searching by flexible protein structure alignment. *Protein Sci*, Vol.13, No.7, (Jul 2004), pp. 1841-1850, ISSN 0961-8368
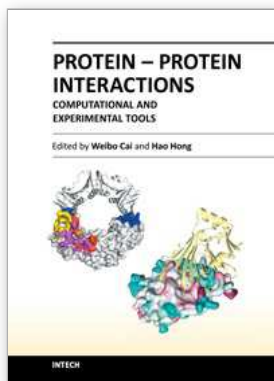
Protein-Protein Interactions - Computational and Experimental Tools

Edited by Dr. Weibo Cai

Proteins are indispensable players in virtually all biological events. The functions of proteins are coordinated through intricate regulatory networks of transient protein-protein interactions (PPIs). To predict and/or study PPIs, a wide variety of techniques have been developed over the last several decades. Many in vitro and in vivo assays have been implemented to explore the mechanism of these ubiquitous interactions. However, despite significant advances in these experimental approaches, many limitations exist such as false-positives/false-negatives, difficulty in obtaining crystal structures of proteins, challenges in the detection of transient PPI, among others. To overcome these limitations, many computational approaches have been developed which are becoming increasingly widely used to facilitate the investigation of PPIs. This book has gathered an ensemble of experts in the field, in 22 chapters, which have been broadly categorized into Computational Approaches, Experimental Approaches, and Others.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH

open science | open minds