Vol.2011-MPS-86 No.13
Vol.2011-BIO-27 No.13
2011/12/1

Prediction of protein residue contacts using discriminative random field

Mayumi Kamada,[†1,†2] Morihiro Hayashida,[†1]
Jiangning Song[†3,†4] and Tatsuya Akutsu[†1]

Understanding interaction between proteins provides a clue to the mechanisms of protein function. Protein residues at interacting sites have co-evolved with those at the corresponding residues in the partner protein to keep their interactions. Therefore, mutual information between residues calculated from multiple sequence alignments of homologous proteins is considered to be useful for identifying contact residues in interacting proteins. The discriminative random field (DRF) is a special type of conditional random fields and can recognize some specific characteristic regions in an image. Since the matrix consisted of correlation between residues can be regarded as an image, we propose a prediction method for protein residue contacts using DRF models with correlation scores between residues based on mutual information. In this work, we perform computational experiments for several interactions between Pfam domains and discuss the results.

1. Introduction

Protein-protein interactions is a crucial clue to understanding the biological systems and molecular networks, several investigations that have been conducted[1)–3)]. Proteins interact with other molecules at specific sites, to understand their interaction, knowing interacting protein residues is one of important steps. In evolutionary process of organisms, coevolution has been conceived as occurring in important sites such as between interacting proteins[4)], that is, it can be considered that protein residues at important sites for interactions have been simultaneously mutated to keep their interactions. In fact, it was confirmed from comparison of putatively orthologous proteins between *S. cerevisiae* and *C. elegans* that interacting proteins evolve at similar evolutionary rates[5)]. It means that interacting residues have been mutated at the same time. Therefore relationship of mutual dependence between coevolving residues can be used as a good clue for predicting protein residue contacts. Mutual information (MI) between residues, which is calculated from the distribution of amino acids in multiple sequence alignment (MSA) for homologous proteins, can represent a quantity of dependence relationship between two residues. Several prediction methods have been conducted using MI between residues. Weigt et al. proposed Direct Information (DI) that is an improvement of MI, and estimated direct residue contacts between sensor kinase and response regulator proteins from the DI calculated by using message passing[6)]. Burger and van Nimwegen developed a prediction method based on a Bayesian network method by constructing a dependence tree where a node corresponds to a position of protein sequence alignments[2)]. However, comparative studies have shown that predicting protein residue contacts is one of challenging tasks.

In the field of image analysis, Markov random fields (MRFs) have been well studied, for instance, for texture segmentation, a deformable contour model, called EigenSnake, and matching to multiple overlapping objects[7)–9)]. Also in the field of bioinformatics, MRFs have been used for protein function prediction from protein-protein interaction networks[10),11)]. In our previous work, we modeled protein-protein interactions based on domain-domain interactions using conditional random fields (CRFs), and developed prediction methods, which outperformed existing methods based on probabilistic models with domains[?)]. Kumar and Hebert proposed discriminative random fields (DRFs) to model spatial interactions in images based on CRFs[12)].

The matrix that consists of all MI between two positions in multiple sequence alignments can be considered as an image. Therefore, in this work, we make use of information about coevolving residues and propose a DRF-based method for predicting residue-residue interactions. Many algorithms have been proposed for the measures of coevolving residues, we use not only the original MI but also improved MI, called RCW-MI and ZNMI, for our method. Furthermore, we

†1 Bioinformatics Center, Institute for Chemical Research, Kyoto University
†2 Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
†3 Department of Biochemistry and Molecular Biology, Monash University
†4 Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences

Vol.2011-MPS-86 No.13
Vol.2011-BIO-27 No.13
2011/12/1

perform computational experiments, and the results suggest that the DRF-based method is useful compared with that using the corresponding MRF model.

2. Methods

In this section, we propose a discriminative random field (DRF)-based method for predicting contact residues. The input data are two amino acid sequences. Then, homologous sequences are collected for each sequence, correlation between two residues based on mutual information is calculated, and the probability that two residues interact with each other is estimated according to our proposed DRF model. For training parameters of the DRF model, several pairs of protein sequences and interacting residues are given.

2.1 Measures of coevolving residue

In our proposed method, information on correlated mutation between protein residues is one of important inputs. Mutual information (MI) for distributions of amino acids at two positions of protein sequence alignments is widely used for analysis of correlated mutation, and is calculated using only individual and joint frequencies of amino acids between columns. Figure 1 (a) shows the calculation of MI. There are two sequences A and B, and multiple alignments are calculated for each of sequences in some adequate way. Let $p_i(a), p_{ij}(a,b)$ be the observed frequency of amino acid $a \in \mathcal{A}$ at position i and that of amino acids $a, b \in \mathcal{A}$ at positions i and j, respectively, where $\mathcal{A}$ be the alphabet set indicates 20 amino acids and 1 character that represents undetermined amino acids. Then, mutual information $m_{i,j}$ between two positions i and j is calculated as follows.

$$m_{i,j} = H_i + H_j - H_{i,j}, \tag{1}$$

where H_i and H_j denote the marginal entropies at positions i and j, respectively, that is, $H_i = -\sum_{a\in\mathcal{A}} p_i(a) \log p_i(a)$, and $H_{i,j}$ denotes the joint entropy $H_{i,j} = -\sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}} p_{ij}(a,b) \log p_{ij}(a,b)$.

However it is well known that phylogenetic and stochastic noise generally occurs among aligned positions of MSA because of common ancestry and random drift[13),14)]. Thus, several approaches to improve MI for avoiding those noises have been developed. In this work, we use two types of improved MI, called RCW-MI and ZNMI.

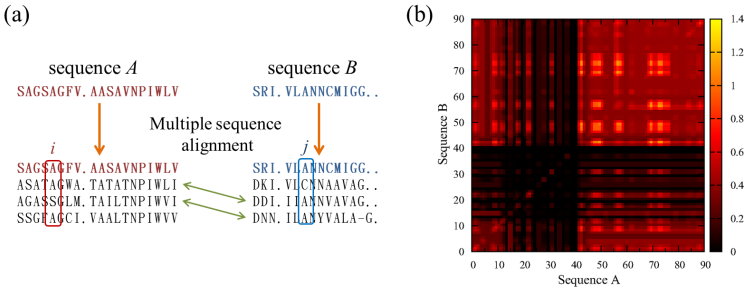

Fig. 1 (a)Illustration on calculation of mutual information between two positions in multiple alignments for sequences A and B. Sequences belonging to the same organism are connected.(b)Example of matrix of mutual information between two residues and concept of this method. For the left figure, the brighter the color of (i, j) is, the higher the value of mutual information is.

- **RCW MI**
  With increasing the probability of two sites sharing the conserved pattern, the probability of non-coevolving sites having high scores will increase. That is, the sites having a common pattern have much higher possibility of causing false positive pairs. As the method for avoiding this effect, RCW-MI (Row and Column Weighed Mutual Information) was proposed by Gouveia-Oliveira and Pedersen[15)]. RCW-MI is a weighting of MI matrix. The weighting can be performed excluding the top hits of every row/column to accommodate for more than two-way coevolution.
  $$RCW_{ij} = \frac{m_{i,j}}{\frac{m_{.,j}+m_{i,.}+m_{j,.}+m_{.,i}-2m_{i,j}}{2n-2}}. \tag{2}$$
  where, $m_{.,j}$ denotes the sum of values of j-th column of MI matrix, and n is the length of sequence.
- **ZNMI**
  To reduce the correlation between MI and the product of the variances of the column MI, Brown and Brown proposed ZNMI[16)] based on NMI, which is normalized MI by joint entropy. They make assumption that the column NMI distribution can be approximated by Gaussian distribution, $N(\mu, \sigma^2)$, parameterized by the column NMI mean and variance. Suppose that the NMI distribution of i-th column can be given as $N(\mu_i, \sigma_i^2)$ and the NMI distribution of j-th column can be given as $N(\mu_j, \sigma_j^2)$, it is straightforward

to show that,

$$N(\mu_i, \sigma_i^2) \times N(\mu_j, \sigma_j^2) = N\left(\frac{\mu_i\sigma_j^2 + \mu_j\sigma_i^2}{\sigma_i^2 + \sigma_j^2}, \frac{\sigma_i^2\sigma_j^2}{\sigma_i^2 + \sigma_j^2}\right). \quad (3)$$

Then, ZNMI are obtained by calculating of z-score for the product $\mathrm{NMI}_{i,j}$ (the right side of Eqn. 3).

2.2 Discriminative Random Field Models for Residue Contacts

Figure. 1(b) shows an example of the matrix of the original MI between two sequences, where the matrix can be considered as an image. Therefore, we make use of an image processing technique, discriminative random field (DRF) proposed by Kumar and Hebert[12], for prediction of interacting residues.

DRF is based on conditional random fields (CRFs)[17]. Let $G(V, E)$ be a graph with a set of vertices V and a set of edges E, where each vertex $s \in V$ is related with a random variable x_s, and observation y_s. Then, $(\boldsymbol{x}, \boldsymbol{y})$ is a conditional random field if the random variables x_s follow the Markov property under the conditions y_s according to the graph G, that is, $P(x_s|\boldsymbol{x}_{\{t\in V|t\neq s\}}, \boldsymbol{y}) = P(x_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y})$, where $\mathcal{N}_s$ denotes the set of vertices adjacent to the vertex s in the graph G. As well as CRFs, DRFs require $P(\boldsymbol{x}|\boldsymbol{y}) > 0$ for all $\boldsymbol{x}$, and are represented by the following formula

$$P(x_s|\boldsymbol{x}_{\mathcal{N}_s}, \boldsymbol{y}) = \frac{1}{Z_s}\exp\left\{-U_s(\boldsymbol{x}, \boldsymbol{y})\right\}, \quad (4)$$

where $U_s(\boldsymbol{x}, \boldsymbol{y})$ is a potential function concerning the vertex s, and Z_s is the normalization constant defined by $\sum_{x_s}\exp\left\{-U_s(\boldsymbol{x}, \boldsymbol{y})\right\}$. In the framework of DRFs, it is assumed that only up to pairwise clique potentials are nonzero, and the potential function is defined as follows.

$$U_s(\boldsymbol{x}, \boldsymbol{y}) = \alpha A(x_s, \boldsymbol{y}) + \beta\sum_{t\in\mathcal{N}_s} I(x_s, x_t, \boldsymbol{y}), \quad (5)$$

where $A(x_s, \boldsymbol{y})$ and $I(x_s, x_t, \boldsymbol{y})$ are the unary and binary potential functions, and called association potential and interaction potential, respectively, each random variable x_s takes 1 or -1, $\alpha \in \{0, 1\}$, and β is a variable. Let $\boldsymbol{w}$ and $\boldsymbol{v}$ be parameter vectors, and $\boldsymbol{f}_s$ and $\boldsymbol{g}_{st}$ be vector-valued functions that map observations $\boldsymbol{y}$ to feature vectors with the same size as parameter vectors. Then, the association potential $A(x_s, \boldsymbol{y})$ can be considered as a gain obtained only from the vertex s and the observations $\boldsymbol{y}$, and is defined as

$$A(x_s, \boldsymbol{y}) = -\log\left(\sigma\left(x_s\boldsymbol{w}^T\boldsymbol{f}_s(\boldsymbol{y})\right)\right), \quad (6)$$

where $\sigma(x)$ is the logistic function defined by $\frac{1}{1+e^{-x}}$, and $\boldsymbol{w}^T$ denotes the transpose of $\boldsymbol{w}$. It means that the DRF model includes generalized linear models (GLM), where other functions such as the probit function can be used as the link function of the DRF. On the other hand, the interaction potential $I(x_s, x_t, \boldsymbol{y})$ can be considered as a gain obtained from the relationship between vertices s and t, and is defined as

$$I_1(x_s, x_t, \boldsymbol{y}) = Kx_sx_t + (1-K)\left(2\sigma\left(x_sx_t\boldsymbol{v}^T\boldsymbol{g}_{st}(\boldsymbol{y})\right) - 1\right), \quad (7)$$

where $0 \leq K \leq 1$, or simply defined as

$$I_2(x_s, x_t, \boldsymbol{y}) = x_sx_t\boldsymbol{v}^T\boldsymbol{g}_{st}(\boldsymbol{y}). \quad (8)$$

Note that the set of parameters θ in DRF models consists of $\boldsymbol{w}, \boldsymbol{v}, \beta$, and K.

In order to determine a DRF model, we must design vector-valued functions $\boldsymbol{f}_s$ and $\boldsymbol{g}_{st}$. Kumar and Hebert used histograms of luminance values $(\boldsymbol{y})$ in neighbor pixels at some scales for recognition of man-made structures in an image[12]. For our purpose, we use random variables $r_{ij}(\in \{1, -1\})$ that represent residue contacts instead of x_s, where $r_{ij} = 1$ means residues between position i and j interact with each other, otherwise $r_{ij} = -1$. Here, the set of vertices V consists of pairs of positions (i, j), and we use $\mathcal{N}_{ij} = \{(i-1, j), (i, j-1), (i, j+1), (i+1, j)\}$ as adjacent vertices to (i, j) (see Fig. 2). Furthermore, we use mutual information m_{ij} between positions i and j as observations $\boldsymbol{y}$. Then, we define vector-valued functions $\boldsymbol{f}_{ij}$ and $\boldsymbol{g}_{ij,kl}$ that map $\boldsymbol{m}$ to feature vectors as follows.

$$\boldsymbol{f}_{ij}(\boldsymbol{m}) = \left(1, m_{i,j}, \frac{1}{2}(m_{i,j-1} + m_{i,j+1}), \frac{1}{2}(m_{i-1,j} + m_{i+1,j})\right)^T, \quad (9)$$

$$\boldsymbol{g}^{(h)}_{ij,kl}(\boldsymbol{m}) = \begin{cases} 1 & (h = 1) \\ |\boldsymbol{f}^{(h)}_{ij} - \boldsymbol{f}^{(h)}_{kl}| & (h = 2, 3, 4) \end{cases}, \quad (10)$$

where $\boldsymbol{g}^{(h)}$ denotes the h-th element of vector $\boldsymbol{g}$, and $|x|$ denotes the absolute value of x. r_{ij} is related with multiple observations m_{ij}, the relationship between mutual information m_{ij} and random variable r_{ij} is represented in the DRF framework as Fig. 2.

On the other hand, in the MRF framework, r_{ij} is related with only an observation m_{ij}. We define the following feature vector for comparison of random fields.

$$\boldsymbol{f}^0_{ij}(\boldsymbol{m}) = \left(1, m_{i,j}\right)^T \quad (11)$$

Vol.2011-MPS-86 No.13
Vol.2011-BIO-27 No.13
2011/12/1

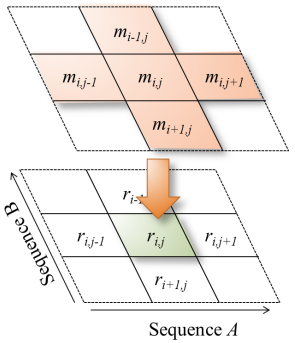

Fig. 2 Adjacent residue pairs for (i, j) and Relationship between mutual information m_{ij} and random variable r_{ij} in the DRF framework.

2.3 Parameter Estimation

We estimate parameters $\theta = \{\boldsymbol{w}, \boldsymbol{v}, \beta, K\}$ by maximizing pseudo-likelihood function as in[12]. Suppose that N pairs of multiple alignments for protein sequences and interacting residues $\boldsymbol{r}^{(n)}(n = 1, \ldots, N)$ for each pair of proteins are given. We calculate mutual information $\boldsymbol{m}^{(n)}$ for each pair. Then, the logarithm of pseudo-likelihood function is given as

$$L(\theta) = \log \prod_{n=1}^{N} \prod_{i} \prod_{j} P(r_{ij}^{(n)} | \boldsymbol{r}_{\mathcal{N}_{ij}}^{(n)}, \boldsymbol{m}^{(n)}, \theta) \tag{12}$$

$$= \sum_{n=1}^{N} \sum_{i} \sum_{j} \Big\{ -U_{ij}(\boldsymbol{r}^{(n)}) - \log \sum_{r_{ij}^{(n)} \in \{1,-1\}} \exp\Big\{ -U_{ij}(\boldsymbol{r}^{(n)}) \Big\} \Big\}. \tag{13}$$

In order to maximize $L(\theta)$, we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[18] method, which is one of quasi-Newton methods that uses partial differentials and approximates the Hessian matrix by some efficient method. For that purpose, partially differentiating $L(\theta)$ with respect to each parameter is required.

2.4 Contact Decision

After estimating parameters, for new pairs of residues, we decide whether or not residues in each pair interact with each other. For that purpose, we use Iterated Conditional Modes (ICM)[19], which iteratively updates random variables $r_{ij} \in \{1, -1\}$ until each variable cannot be changed using the following.

$$r_{ij}^{(t+1)} = \mathrm{argmax}_{r_{ij} \in \{1,-1\}} P(r_{ij} | \boldsymbol{r}_{\mathcal{N}_{ij}}^{(t)}, \boldsymbol{m}, \theta), \tag{14}$$

where $r_{ij}^{(t)}$ denotes the value of random variable r_{ij} at step t.

Table 1 Details of interacting domain pairs in each clan group .

Clan23		Clan58		Clan79	
PDBcode	#	PDBcode	#	PDBcode	#
1G29	188 × 188 (106)	1JZ4	295 × 295 (72)	1HRP	105 × 96 (119)
1KSF	195 × 162 (4)	1UZ1	443 × 443 (8)	1FL7	105 × 105 (18)
1IQP	159 × 90 (63)	1UR4	364 × 364 (4)	1MKK	79 × 79 (23)
1IQP	90 × 90 (201)	1AQ0	306 × 306 (70)	1M4U	226 × 105 (19)
1XXH	191 × 309 (33)	1PX8	482 × 482 (88)	1ES7	104 × 104 (60)
1OJL	222 × 222 (80)	1XSI	430 × 430 (13)	1FL7	96 × 96 (29)
1XXH	309 × 309 (46)	1OGS	494 × 494 (32)	1B98	125 × 125 (83)
1X6V	159 × 159 (51)	1VRX	319 × 319 (31)	1AOC	173 × 173 (78)
1HQC	197 × 197 (36)	1UKP	423 × 423 (56)		
1FL9	128 × 128 (16)	1ODZ	300 × 300 (26)		
1KO4	156 × 156 (37)	1UR8	404 × 404 (29)		
1NLY	298 × 298 (140)	1SMA	359 × 359 (6)		
		1W2V	346 × 346 (44)		
		1O7A	318 × 318 (33)		

* # columns show the lengths of each sequence of interaction pair, and the number of contact residues as indicated by ().

3. Computational Experiments

3.1 Data and Implementation

To get protein residue interaction data, we used the files, 'int_pfamA.txt' and 'interaction.txt', from Pfam database (version 21.0)[20]. The former includes 6,079 interacting domain pairs, and the latter includes information of interacting residue pairs between domains. In this work, we used three datasets belong to different superfamilies, which are registered as AAA (CL0023), Glyco_hydro_tim (CL0058) and Cytine-knot (CL0079) in Pfam, and each group includes 12, 14, 8 interaction domain pairs, respectively. "AAA" is P-loop containing nucleoside triphosphate hydrolase superfamily, "Glyco_hydro_tim" is Tim barrel glycosyl hydrolase superfamily, and "Cytine-knot" is Cystine-knot cytokine superfamily. Where we excluded pairs that contain less than 2 interacting residues and contain

Table 2 Results on average sensitivities for test datasets using three types of inputs for MRF model with feature vector $\boldsymbol{f}_{ij}^0$, DRF model with $\boldsymbol{f}_{ij}$.

	MI		RCW-MI		ZNMI	
	$\mathcal{A}$	$\mathcal{H}$	$\mathcal{A}$	$\mathcal{H}$	$\mathcal{A}$	$\mathcal{H}$
dataset: AAA						
MRF model with $\boldsymbol{f}_{ij}^0$	55.94%	53.84%	47.95%	46.25%	29.7%	32.54%
DRF model with $\boldsymbol{f}_{ij}$	57.63%	52.01%	52.47%	49.69%	30.98%	35.51%
dataset: Glyco_hydro_tim						
MRF model with $\boldsymbol{f}_{ij}^0$	51.48%	56.04%	51.85%	49.18%	34.36%	35.26%
DRF model with $\boldsymbol{f}_{ij}$	52.44%	57.08%	54.47%	49.30%	36.43%	34.26%
dataset: Cytine-knot						
MRF model with $\boldsymbol{f}_{ij}^0$	37.46%	44.58%	38.92%	32.53%	41.76%	42.07%
DRF model with $\boldsymbol{f}_{ij}$	40.98%	47.60%	39.33%	34.73%	49.53%	50.97%

$\mathcal{A}$ is the alphabets set representing 20 amino acids and $\mathcal{H}$ is the set of alphabets indicate hydrophobic or hydrophilic of amino acids.

less than 5 sequences for multiple alignments. Table 1 shows the details of the datasets. Since each sequence included from 79 to 494 residues and the number of residue pairs was more than 79×79=6,241, it is considered to be enough for estimating parameters. However, the number of interacting residues (positive examples) is too few in a pair of domains compared with that of non-interacting residues (negative examples). Therefore, we selected uniformly at random the same number of negative examples as that of positive examples. For the calculation of mutual information between residues, we used multiple alignment data provided in the file 'Pfam-A.full' in Pfam database.

The calculation of the original MI is strongly affected by the alphabet chosen to represent the protein sequence[21]. To investigate the effect, we used two types of alphabet sets representing amino acids. One is not classified, that is, each alphabet indicates a distinct amino acid, and this set is denoted by $\mathcal{A}$. Another is hydropathy-based classification. It classifies 20 amino acids into 2 groups, hydrophobic (G, A, P, V, L, I, M, W and E) and hydrophilic amino acids (R, N, D, E, Q, H, K, S, T, C and Y), it is denoted by $\mathcal{H}$.

We used libLBFGS (version 1.10) with default parameters to estimate the parameters θ, which is a C implementation of the limited memory BFGS method[22], and is available on the web page, `http://www.chokkan.org/software/liblbfgs/`.

3.2 Results

In order to evaluate the proposed DRF-based method, we performed computational experiments using two types of vector-valued functions $\boldsymbol{f}_{ij}^0$ and $\boldsymbol{f}_{ij}$, and two types of classification of amino acids, 20 amino acids and hydropathy-based classification. We performed leave-one-out cross validation, where one dataset was used for test and the remaining datasets were for training. This process was repeated, and the numbers of repeated times were the number of datasets, that is, 12, 14 and 8, respectively. We calculated the conditional probabilities $P(r_{ij} = 1|\boldsymbol{r}_{\mathcal{N}_{ij}}, \boldsymbol{m}, \theta)$ and sensitivity scores, which are measured as TP/(TP+FN), and then took the average.

First, we set $\alpha = 1$ and $\beta = 0$. It means that DRF models contained only the association potential $A(r_{ij}, \boldsymbol{m})$. Table 2 shows the results on the average sensitivities for test datasets using the original MI, RCW-MI and ZNMI for the MRF model with feature vector $\boldsymbol{f}_{ij}^0$ and the DRF model with $\boldsymbol{f}_{ij}$, respectively. For AAA and Glyco_hydro_tim datasets, the result by DRF model with MI was better than those by the other correlation indexes. On the other hand, for Cytine-knot dataset, the result by DRF model with ZNMI was better than those of the others. It seemed that MI was more useful to our prediction model than RCW-MI and ZNMI. It may suggest that the algorithm for calculating score used as observations is dependent on probabilistic model used to prediction. and it has to be chosen with the consideration for the interaction type, homodimer or heterodimer, and the number of amino acids in each sequence. Moreover, the parameters of our model should be estimated for homodimers and heterodimers independently. It is hard to say which classification of amino acids is the best in these experiments.

Next, we set $\alpha = 0$ and $\beta = 1$, which means DRF models with only the interaction potential $I(r_{ij}, r_{kl}, \boldsymbol{m})$. However, the BFGS method for parameter estimation did not converge for the potential I. It can be considered because the interaction potentials in DRFs were originally developed for smoothing images, the neighbor pixels often have similar color to each other. However, pairs of neighbor residues are not always similar, that is, even if residues at positions (i, j) interact, it might be difficult to determine whether or not residues at $(k, l) \in \mathcal{N}_{ij}$ interact. On the other hand, it is considered from the results that the association potential in DRF is useful for predicting interacting residues, information between neighbor residues is useful.

4. Conclusion

We proposed models for predicting protein residue contacts using the discriminative random field, which is a special type of conditional random fields. In order to make use of DRFs, the correlation scores between residues based on mutual information were given as observations in the potential of DRFs, where mutual information was calculated from multiple sequence alignments of homologous proteins. To validate the proposed method, we performed computational experiments using leave-one-out cross validation and calculated the average sensitivities. The results suggest that our proposed DRF-

based method is useful for prediction of protein residue contacts compared with that based on the corresponding Markov random field model. It means that correlation index between neighbor residues is useful for the contact prediction. Additionally, the original MI was more useful than other two scores, RCW-MI and ZNMI, for our proposed model. On the other hand, interaction potentials were not useful because DRFs have been originally developed for image analyses. The problem of predicting residue contacts is one of difficult problems, and it cannot be said that the prediction accuracy by our method was good. However, our method leaves much to be improved in points of the modification of observations and potential function. The selection of which correlation scores we use has strongly depended on datasets, we will introduce new score for coevolving residues. In addition, we can introduce some parameters representing properties for each amino acid in the potential function. Because the results imply that the number of parameters was not sufficient for explaining protein residue contacts.

References

1) White, R.A., Szurmant, H., Hoch, J.A. and Hwa, T.: Features of protein-protein interactions in two-component signaling deduced from genomic libraries, *Methods Enzymol*, Vol.422, pp.75–101 (2007).
2) Burger, L. and van Nimwegen, E.: Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method, *Molecular Systems Biology*, Vol.4, p.165 (2008).
3) Halabi, N., Rivoire, O., Leibler, S. and Ranganathan, R.: Protein sectors: Evolutionary units of three-dimensional structure, *Cell*, Vol.138, pp.774–786 (2009).
4) F.Pazos, M. Helmer-Citterich, G.A. and Valencia, A.: Correlated mutations contain information about protrein-protein interaction, *J Mol Bio*, Vol.271, pp.511–523 (1997).
5) Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. and Feldman, M. W.: Evolutionary rate in the protein interaction network, *Science*, Vol.296, pp.750–752 (2002).
6) Weigt, M., White, R. A., Szurmant, H., Hoch, J. A. and Hwa, T.: Identification of direct residue contacts in protein-protein interaction by message passing, *Proc. Natl. Acad. Sci. USA*, Vol.106, pp.67–72 (2009).
7) Li, S.Z.: *Markov random field modeling in image analysis*, Springer-Verlag London, 3 edition (2009).
8) Derin, H. and Elliott, H.: Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.9, No.1, pp.39–55 (1987).
9) Li, S.Z. and Lu, J.: Modeling Bayesian estimation for deformable contours, *Proc. Seventh IEEE International Conference on Computer Vision*, pp.991–996 (1999).
10) Deng, M., Zhang, K., Mehta, S., Chen, T. and Sun, F.: Prediction of protein function using protein-protein interaction data, *Journal of Computational Biology*, Vol.10, No.6, pp.947–960 (2003).
11) Deng, M., Chen, T. and Sun, F.: An integrated probabilistic model for functional prediction of proteins, *Journal of Computational Biology*, Vol.11, pp.463–475 (2004).
12) Kumar, S. and Hebert, M.: Discriminative random fields, *International Journal of Computer Vision*, Vol.68, No.2, pp.179–201 (2006).
13) Codoner, F.M. and Fares, M.A.: Improved residue contact prediction using support vector machines and a large feature set, *Evol. Bioinform*, Vol.4, pp.29–38 (2008).
14) Wollenberg, K.R. and Atchley, W.R.: Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap, *PNAS*, Vol. 97, pp.3288–3291 (2000).
15) Gouveia-Oliveira, R. and Pedersen, A.: Finding coevolving amino acid residues using row and column weighting of mutual information and multi-dimensional amino acid representation, *PLoS ONE*, Vol.5, p.e10779 (2010).
16) Brown, C.A. and Brown, K.S.: Validation of Coevolving Residue Algorithms via Pipeline Sensitivity Analysis: ELSC and OMES and ZNMI, Oh My!, *PLoS ONE*, Vol.5, p.e10779 (2010).
17) Lafferty, J., McCallum, A. and Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proc. Int. Conf. on Machine Learning* (2001).
18) Bertsekas, D.P.: *Nonlinear Programming*, Athena Scientific (1999).
19) Besag, J.: On the statistical analysis of dirty pictures, *Journal of Royal Statistical Soc.*, Vol.B-48, pp.259–302 (1986).
20) Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L.L., Eddy, S.R. and Bateman, A.: The Pfam protein families database, *Nucleic Acids Research*, Vol.38, pp.D211–D222 (2010).
21) Hemmerich, C. and Kim, S.: A study of residue correlation within protein sequences and its application to sequence classification, *EURASIP Journal on Bioinformatics and Systems Biology*, p.9 (2007).
22) Nocedal, J.: Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, Vol.35, No.151, pp.773–782 (1980).