Radial Basis Function Neural Network Ensemble for Predicting Protein-Protein Interaction Sites in Heterocomplexes

Bing Wang[1,2,*], Peng Chen[3], Peizhen Wang[1], Guangxin Zhao[1] and Xiang Zhang[2]

[1]*School of Electrical & Information, Anhui University of Technology, Ma'anshan, Anhui, 243002, China;* [2]*Department of Chemistry, University of Louisville, Louisville, KY, 40202, USA;* [3]*School of Computer Engineering, Nanyang Technological University, Singapore, 639798, Singapore*

Abstract: Prediction of protein-protein interaction sites can guide the structural elucidation of protein complexes. We propose a novel method using a radial basis function neural network (RBFNN) ensemble model for the prediction of protein interaction sites in heterocomplexes. We classified protein surface residues into interaction sites or non-interaction sites based on the RBFNNs trained on different datasets, then judged a prediction to be the final output. Only information of evolutionary conservation and spatial sequence profile are used in this ensemble predictor to describe the protein sites. A non-redundant data set of heterodimers used is consisted of 69 protein chains, in which 10329 surface residues can be found. The efficiency and the effectiveness of our proposed approach can be validated by a better performance such as the accuracy of 0.689, the sensitivity of 66.6% and the specificity of 67.6%.

Keywords: Protein interaction sites, heterocomplex, radial basis function neural networks, ensemble, spatial neighboring residue, surface residue.

1. INTRODUCTION

Protein-protein interaction has become one of the most important research fields in the current molecular biology. Understanding the characteristics of interfacial sites between two interacting proteins is a necessary step to decipher the molecular recognition process and to elucidate protein function and the structure of protein complexes. The ability to predict interfacial sites and interacting protein pairs is also important in mutant design and drug design [1]. Experimental techniques, such as X-ray and NMR, detect residues in protein-protein interaction surface by determination of the structure of a protein complex. However, the number of resolved protein complexes is currently far less than the number of known protein sequences. It is necessary to develop computational approaches to identify the roles of function residues for the analyses of protein-protein interaction sites.

Some computational methods have been proposed for studying protein interaction sites by analyzing the different characteristics of amino acids present in the interfaces of two or more proteins in a protein complex [2-6]. Based on the observation that proline is the most common residue found in the flanking segments of interaction sites, Kini & Evans proposed a method to predict protein interaction sites by detecting the presence of "proline brackets" [7]. By defining and analyzing a series of residue patches on the surface of protein structures, Jones & Thornton successfully predicted the interfaces in a set of 59 structures using six parameters (solvation potential, residue interface propensity, hydropho bicity, planarity, protrusion and accessible surface area) [8, 9]. In 2004, Yan *et al.* proposed a two-stage method consisting of a support vector machine (SVM) and a Bayesian classifier to predict protein interaction sites and concluded that interface residues tend to form clusters in the primary amino acid sequence [10]. Wang *et al.* also employed machine learning methods for the prediction of protein interaction sites using features extracted from spatial sequence and evolutionary conservation scores based on a phylogenetic tree [3, 5]. Chen & Jeong proposed a random forest-based integrative model based on physicochemical properties and evolutionary conservation score, residue-based distance matrix and sequence profile [11]. Jones & Thornton found that hydrophobicity is a common characteristic of interfacial surfaces in homodimers, most of which exist in an oligomeric state [8]. Although Glaser reported that hydrophobic residues are abundant in large interfaces while polar residues are more abundant in small interacting patches [12], it still remains unknown about the residue preference for the interfaces in heterocomplexes. Therefore, how to capture the inherent properties which can differentiate interface residues and non-interface residues is the most important issue in predicting interaction sites in heterocomplexes.

In this work, a radial basis function neural network (RBFNN) ensemble-based model is proposed and realized in prediction of the protein interfacial sites. The ensemble model was built on the profiles of spatially neighboring sequences and information of evolutionary conservation. By analyzing 69 protein heterodimers, we achieved 0.689 of prediction accuracy, 66.6% of sensitivity and 67.65 of specificity. These results are in agreement with the previous findings that conserved residues in protein are most likely to be found at important functional sites on a protein [13].

*Address correspondence to this author at the School of Electrical & Information, Anhui University of Technology, Ma'anshan, Anhui, 243002, China; Tel: +86 555 2311541; E-mail: wangbing@ustc.edu

0929-8665/10 $55.00+.00

2. MATERIALS AND METHODS

2.1. Data Preparation

A dataset containing 69 protein chains was obtained from our previous work [3]. As a non-redundant dataset, there is no protein chain whose sequence identity is larger than 30% existing in the dataset. The dataset has also removed homo-complexes whose interacting surfaces are characterized by hydrophobicity, and protease-inhibitor complexes whose interfaces can be distinguished by serine and histidine active site signatures. The dataset also excludes the chains labeled as 'membrane peptides', 'small proteins' or 'coiled coils' in the SCOP classification [14].

2.2. Surface Residue and Interface Residue Definition

There are no consentaneous definitions of surface residues and interface residues in current researches [3, 10, 11, 15]. Generally, those definitions are dependent on the reduction of accessible surface area (ASA) or the distance between the target residue and the atoms in the other molecule of complex. In this paper, a residue is defined to be a surface residue if its relative ASA is at least 16% of its nominal maximum area defined by Rost and Sander [16]. ASA was computed for each residue in each protein chain using the DSSP program [17]. The coordinates of the particular chain are the only parameters used in calculation. Otherwise, ASA will be incorrect due to the other chains existing in the complex. A residue is classified to be an interface residue if the spatial distance between its alpha-carbon (CA) atom and one of CA atoms in the other chains in the complex is less than 1.2 nm. Each protein in this paper is represented by its CA trace and the other atoms are not considered. According to the above definitions, we obtained 10329 surface residues, 34.8% of which are interface residues.

2.3. Predictor Design

To build the predictors that can distinguish interaction sites from non-interaction sites in protein sequence, we extracted features based on residue spatial sequence profile and evolutionary conservation score. Sequence profile for each sequence site was obtained from the homology-derived structures of proteins (HSSP) database [18], in which each site in protein sequence was represented by a 20-dimensional vector indicating the occurrence of 20 amino acid. Evolutionary conservation score, a measurement of evolutionary rate for a given site in the sequence, was calculated using ConSurf software [12].

In our experiment, the prediction of protein-protein interaction sites was treated as a two-category classification problem. The predictor assigns each sequence site in protein chain as a target value. The target value is +1 if the target residue is classified into interface residue set, and -1 otherwise. The predictor is generated by using the RBFNN to judge whether a residue in protein sequence is located in an interface. The input vector of the network is the conservation information of a surface residue, i.e., the predictor is fed with a window of 11 residues centered by the target residue and 5 spatially neighboring residues on each side. The use of these windows means that the predictor is trained by patches that consist of 11 spatially continuous residues although the target values determined by the target residues. Therefore, the final input vector for each residue in RBFNN is a $(20+1)\times 11=231$-dimensional vector.

2.4. Constructing the RBFNN Ensemble Predictor

Radial basis function neural network (RBFNN) is an efficient approach to solve non-linear problems by using a special class of functions, named radial basis function [19-22]. Given a sufficient number of hidden units, a RBFNN is also considered as a universal approximator for any continuous functions. A standard RBFNN consists of three layers of neurons, i.e., input, hidden and output layers. The nodes within each layer are fully connected to the previous layer. The input variables are each assigned to a node in the input layer and passed directly to the hidden layer without weights. The hidden nodes contain the radial basis function, also called transfer function. Given an input pattern $\mathbf{x} = (x_1, x_2, \cdots, x_d)^T$, the output of RBFNN, $y(\mathbf{x})$, can be obtained as follows:

$$y(\mathbf{x}) = \sum_{i=1}^{c} w_i \phi_i(\mathbf{x}) \quad (1)$$

where c is the number of hidden layer's neurons, w denotes the corresponding weights and ϕ is radial basis function. Here, we consider ϕ as Gaussian function:

$$\phi_i(\mathbf{x}) = \exp(-\frac{(\mathbf{x}-\mu_i)^2}{2\sigma_i^2}) \quad (2)$$

2.5. RBFNN Ensemble

In the beginning of 1990s, Hansen and Salamon improved the generalization ability of an artificial neural network (ANN) system through ensembling artificial neural networks [23]. Later, Hansen *et al*. applied ANN ensemble for handwritten digit recognition and achieved a good results in which the accuracy is 20-25% better than that of individual ANN [24]. Other works had shown that RBFNN ensemble also improve the pattern classification performance [25, 26]. In this work, we split the original dataset randomly into 6 subsets in almost equal size due to the large dataset that consists of 10329 surface residues. We selected one of the six subsets as test dataset, and the remaining five subsets were used as training datasets. In each training dataset, we trained a RBFNN classifier, and tested the RBFNN model in the test dataset. For all protein residues, we therefore obtained five predicted results. A majority voting method was adopted to decide a prediction to be the final output, if more than half of the individual RBFNN vote the prediction. This process was repeated six times, where different data part was selected as test dataset each time. This analysis architecture is displayed in Fig. (**1**).

3. RESULTS AND DISCUSSION

3.1. Evaluation for Predictor Performance

Prediction accuracy, the ratio of the number of correctly predicted protein interaction sites to the total number of pre-

dicted interaction sites in experiment, is considered the best index for evaluating the performance of our predictors. However, only 34.8% of the surface residues are interaction sites according to their definitions in this work, which lead to a rather unbalanced distribution of positive (interaction sites) and negative (non-interaction sites) samples. Therefore, the performance of predictor is evaluated by three parameters, i.e., *specificity* (SP), *sensitivity* (SN) and *correlation coefficient* (CC) [27]. The specificity is defined as the ratio of the number of matched residues between the predicted set and the actual set over the total number of predicted residues. The sensitivity is defined as the ratio of the number of matched interaction sites over the total number of the interaction sites of the observed set.

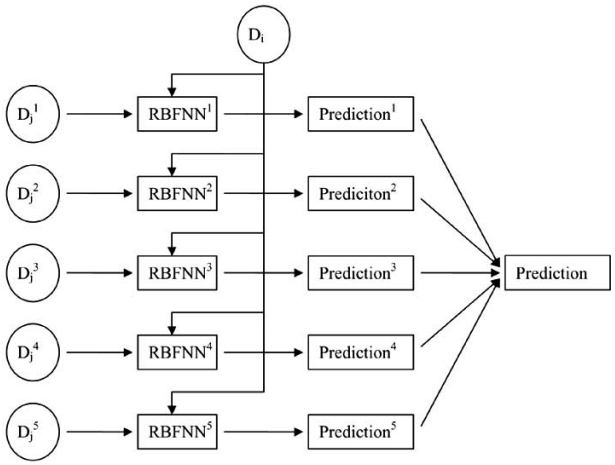

Figure 1. Workflow of RBFNN ensemble. The whole dataset D is split into six parts $\{D_1, D_2, \cdots, D_6\}$; the ensemble strategy will take one of them, i.e., D_i as test dataset, and each remaining part can be seen as a training dataset on which a RBFNN classifier is trained. Each RBFNN classifier will give one prediction result for D_i, and the final prediction is combined by each prediction of the RBFNN classifier using majority voting method.

Let TP (true positives) be the number of true positives, where residues are predicted to be interface residues that actually are interface residues, and FP (false positives) be the number of false positives, where residues are predicted to be interface residues that are in fact not interface residues. In addition, assume TN (true negative) to be the number of true negatives, and FN (false negative) to be the number of false negatives. Thus, the total number of examples N = TP + TN + FP + FN.

Then the evaluation measures can be computed as follows:

$$SN^+ = \frac{TP}{TP+FN} \quad SN^- = \frac{TN}{TN+FP}$$

$$SP^+ = \frac{TP}{TP+FP} \quad SP^- = \frac{TN}{TN+FN}$$

$$SN = (\frac{TP+FN}{N})SN^+ + (\frac{TN+FP}{N})SN^-$$

$$SP = (\frac{TP+FN}{N})SP^+ + (\frac{TN+FP}{N})SP^-$$

$$CC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}}$$

Correlation coefficient is a measure of how well the predicted class labels correlate with the actual class labels. Its range is from −1 to 1, where a correlation coefficient of 1 corresponds to perfect predictions, while a correlation coefficient of 0 corresponds to random guessing.

3.2. Classification Results of Surface Residues

For the whole dataset in this work, our predictor classified 2166 surface residues into interaction sites, of which 1486 are true positives. Therefore, the prediction accuracy is 68.6%. Based on the definitions of the evaluation measures, we achieved SP = 66.6%, SN = 67.6% and CC = 0.2458, as showed in Table **1**. The specificity for a class is the probability that a positive prediction for the class is correct, where SP^+ corresponds to the interface residues and SP^- to the non-interface residues. The sensitivity for a class is the probability of correctly predicting an example of the class. The comparatively high and balanced values of SP and SN indicate that the average ability of the predictor is independent of the class type (interface or non-interface). The values of SP and SN listed in Table **1** demonstrate that our proposed approach has a good performance of predicting interaction sites in heterocomplexes. Taking the actual class label into account, the values of SP and SN on the negative subset are higher than the ones on the positive subset. This result is related to the unequal numbers of interface residues and non-interface residues in our dataset (only 34.8% surface residues are defined as the interface residues). The correlation coefficient of 0.25 indicates that the predictor is better than a random predictor (whose correlation coefficient is 0). These results achieved here suggest that our computational methods are efficiently capable of predicting the interaction sites in protein chains.

The distribution of the number of proteins against different performance measures of RBFNN_ensemble for 69 polypeptide chains is shown in Fig. (**2**). This experiment is based on the complete dataset and majority voting. In Fig. (**2**), the horizontal axis stands for thresholds of different performance measures, including accuracy, specificity, sensitivity and correlation coefficient; the vertical axis means the number of proteins in the prediction results which satisfy different performance thresholds. Specificity indicates the probability that an actual residue is classified by a predictor into a class. All proteins have the SP value of over 20%. This means that there are at least 20% residues in each protein that can be identified correctly. For 94.2% (65 of the total 69) of the proteins, our predictor can predict at least 50% of residues into the right class. In addition, the SN values are greater than 20% for all proteins in our dataset. However, at least 50% residues were correctly classified for over 85.5% (59 of 75) of the proteins. The correlation coefficient values listed in Table **1** show how well our predictor worked. It can be seen in Fig.(**2**) that among 78.3% of the proteins the correlation coefficient values are greater than 0, which suggests

Table 1. The Comparison of the Performance Obtained by RBFNN Trained by the EM Algorithm as well as the BP Algorithm on our Dataset

	CC	SP	SN	SP^+	SP^-	SN^+	SN^-
RBFNN_EM	0.18	59.9%	58.3%	57.5%	61.8%	42.5%	74.8%
BP_ANN	0.12	54.5%	55.2%	49.2%	58.8%	44.9%	62.9%
RBFNN_ensemble	0.2458	66.6%	67.6%	57.0%	71.7%	34.8%	86.3%

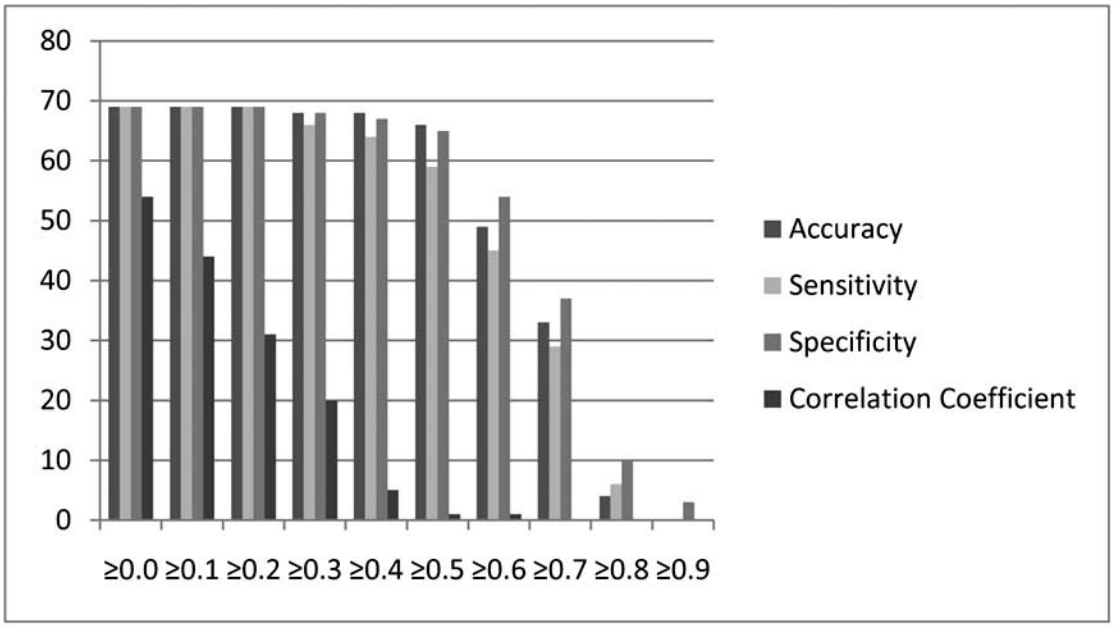

Figure 2. The statistical results of accuracy and three performance parameters.

that our predictor is indeed better than the random predictor [27]. For 95.7% of the proteins, the predicted interface residues are indeed located in the interface almost exclusively because the accuracy is above 50% [15].

3.3. Comparison to Other Methods

In previous studies, many neural networks based approaches have been adopted to infer protein interaction sites for protein sequence, such as BP_ANN, a artificial neural network model based back-propagation algorithm proposed by Farisell *et al*. [15] and RBFNN_EM, a radial basis function neural networks optimized by expectation maximization algorithm introduced by Wang *et al*. [4]. Zhou and Shan identified whether a residue is interacting residue or not using neural networks based on their structural neighbors [28]. The accuracies of these approaches are around 70% for the prediction of interactions at the residue level. The comparison of the performance between our methods and the BP_ANN approach used by Fariselli *et al.* and RBFNN_EM by Wang *et al.* is shown in Table **1**. As a result, it was found that our method exhibited a more confident effectiveness with respect to the BP_ANN and RBFNN_EM algorithm. A direct comparison of results with the experiments of Zhou is impossible due to differences in the choice of source data, and the definitions of surface residue as well as interface residue, but we note that there are two consecutive neural networks were used in the training process of Zhou's experiments, which is more complex than the networks adopted here.

Yan *et al*. trained a support vector machine to investigate this problem and achieved relatively high specificity (71%) and sensitivity (67%) [10]. In addition, Ofran and Rost attempted to identify interaction sites using neural networks based on the sequence clustering interface residues, and they reported an accuracy of 70% and 20% sensitivity [29]. These works all tackled the problem of predicting protein interaction sites based on primary sequence, and it is an important attempt in prediction of protein interface without structural information. The HSSP database was also adopted in these studies. The HSSP sequence profile is constructed by multiple sequence alignments of homologous proteins for each protein of known three-dimensional structure in the protein data bank (PDB) [30]. The information contained in the sequence profile indicated evolution conservation of the residues, which include structural conservation and functional conservation. For the study of interaction sites between the protein chains in complex, the spatial neighbor residues in surface we adopted here are more effective than the consecutive residues in primary since these residues provide with more information of functional conservation which is more responsible for forming interactions between protein partners. The similar conclusion between our method and those

similar approaches is that no information about geometric and electrostatic complementarity within interfaces is required.

3.4. Recognition of Interaction Sites

The effectiveness of our proposed method is further demonstrated by a test on a protein complex 1gla_F (PDB ID). This protein is a phosphocarrier protein, IIIGlc, which is an integral component of the bacterial phosphotransferase (PTS) system [31]. As shown in Fig. (**3**) the structure diagram was generated using RasTop (http://www.geneinfinity.org/rastop) tool. The prediction results are illustrated by sphere residues, where the green sphere denotes true positives (true interaction residues that are predicted correctly), the yellow sphere indicates the missing interface residues (true interaction residues that did not detect) and the red sphere denotes false positives. It can be found that there is only 1 observed non-interface residue that cannot be correctly predicted by our method. Ten residues that actually are interface residues had been predicted as non-interface residues; it may be related to the original dataset in which only 34.8% surface residues had been defined as interface residues. But from Fig. (**3**), it can also be found that the incorrectly predicted residues are much closed to the observed interface residues. If we magnify the threshold a little, by which we can distinguish interface from non-interface residues, the predicted incorrect residues will be taken as interface residues and thus can be predicted exactly. Among the factors that affect the experiment results, the task of how to classify a residue as a surface residue or an interface residue becomes most important. For the reliability index, we prefer to a stricter criterion in our experiment even at the cost of descent of accuracy.

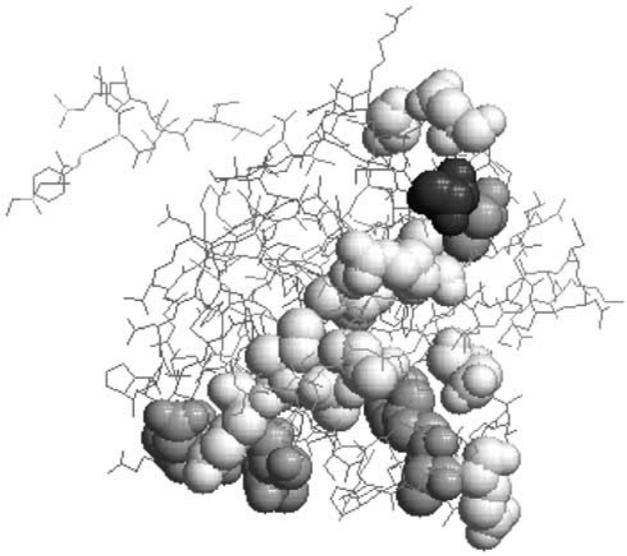

Figure 3. The visualization of prediction results in protein chain 1gla_F (PDB index). T he target protein is shown in gray wireframe, the residues related to prediction are displayed in sphere and the corresponding colors are coded as follows: white denotes interface residues identified by the predictor; grey denotes the missing interface residues in the predictor; black denotes the residues predicted as interface residues but non-interface residues actually.

CONCLUSION

In this paper, we described a novel ensemble approach of protein interaction sites predictor by using RBFNN, which uses spatial neighbor residues and their conservation HSSP profile. Generally, interfaces consist of interacting residues that belong to two different chains along with residues in their spatial vicinity. If the partner chain is removed, those interacting residues are exposed to the solvent. So we focus only on the surface residues in this study. As a database merging information from three-dimensional structures and one-dimensional sequences of proteins, HSSP database implied the residue conservation in structural context. Many of the residues on interfaces that are critical for binding together are likely to be evolutionarily conserved. This is because the pace of evolution at interfaces is slower than the rest of the protein. This slower pace of evolution at interfaces can be explained by the phenomena of co-evolution, where substitutions in one protein will result in the selection pressure for reciprocal changes in interacting partners [32]. The evolution conservation we employed in this paper provides the foundation for our designed predictor. The results obtained in this paper showed that our proposed method is a promising approach for studying protein-protein interaction, and it can facilitate those experimental investigators to validate the roles of specific residues in protein complexes.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China (No.60803107).

REFERENCES

[1] Koike, A.; Kobayashi, Y.; Takagi, T. Kinase pathway database: an integrated protein-kinase and NLP-based protein-interaction resource. *Genome Res*., **2003**, *13*(6A), 1231-43.

[2] Shi, M. G.; Huang, D. S.; Li, X. L. A protein interaction network analysis for yeast integral membrane protein. *Protein Pept. Lett*., **2008**, *15*(7), 692-9.

[3] Wang, B.; Chen, P.; Huang, D. S.; Li, J. J.; Lok, T. M.; Lyu, M. R. Predicting protein interaction sites from residue spatial sequence profile and evolution rate. *FEBS Lett*, **2006**, *580*(2), 380-4.

[4] Wang, B.; Wong, H. S.; Chen, P.; Wang, H. Q.; Huang, D. S. Predicting Protein-Protein Interaction Sites Using Radial Basis Function Neural Networks. In *International Joint Conference on Neural Networks*, **2006**; pp. 2325-2330.

[5] Wang, B.; Wong, H. S.; Huang, D. S. Inferring protein-protein interacting sites using residue conservation and evolutionary information. *Protein Pept. Lett.*, **2006**, *13*(10), 999-1005.

[6] Xia, J. F.; Han, K.; Huang, D. S. Sequence-Based Prediction of Protein-Protein Interactions by Means of Rotation Forest and Autocorrelation Descriptor. *Protein Pept. Lett*., **2010**, *17*(1), 137-145.

[7] Kini, R. M.; Evans, H. J. Prediction of potential protein-protein interaction sites from amino acid sequence. Identification of a fibrin polymerization site. *FEBS Lett*., **1996**, *385*(1-2), 81-6.

[8] Jones, S.; Thornton, J. M. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol*., **1997**, *272*(1), 133-43.

[9] Jones, S.; Thornton, J. M. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol*., **1997**, *272*(1), 121-32.

[10] Yan, C.; Dobbs, D.; Honavar, V. A two-stage classifier for identification of protein-protein interface residues. *Bioinformatics*, **2004**, *20*(Suppl 1), i371-8.

[11] Chen, X. W.; Jeong, J. C. Sequence-based prediction of protein interaction sites with an integrative method. *Bioinformatics*, **2009**, *25*(5), 585-91.

[12] Glaser, F.; Rosenberg, Y.; Kessel, A.; Pupko, T.; Ben-Tal, N. The ConSurf-HSSP database: the mapping of evolutionary conservation

among homologs onto PDB structures. *Proteins*, **2005**, *58*(3), 610-7.

[13] Livingstone, C. D.; Barton, G. J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci.*, **1993**, *9*(6), 745-56.

[14] Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **1995**, *247*(4), 536-40.

[15] Fariselli, P.; Pazos, F.; Valencia, A.; Casadio, R. Prediction of protein--protein interaction sites in heterocomplexes with neural networks. *Eur. J. Biochem.*, **2002**, *269*(5), 1356-61.

[16] Rost, B.; Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins*, **1994**, *20*(3), 216-26.

[17] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **1983**, *22*(12), 2577-637.

[18] Dodge, C.; Schneider, R.; Sander, C. The HSSP database of protein structure-sequence alignments and family profiles. *Nucleic Acids Res.*, **1998**, *26*(1), 313-5.

[19] Huang, D. S.; Du, J. X. A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. *IEEE. Trans. Neural Netw.*, **2008**, *19*(12), 2099-115.

[20] Chen, S.; Cowan, C. N.; Grant, P. M. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE. Trans. Neural Netw.*, **1991**, *2*(2), 302-9.

[21] Huang, D. S. Radial basis probabilistic neural networks: Model and application. *Int. J. Patt. Recognit.Artif. Intell.*, **1999**, *13*(7), 1083-1101.

[22] Huang, D. *Systematic Theory of Neural Networks for Pattern Recognition*. Publishing House of Electronic Industry of China: Beijing, **1996**.

[23] Hansen, L. K.; Salamon, P. Neural network ensembles. *IEEE. Trans. Pattern Anal. Machine Intell.*, **1990**, *12*(10), 993-1001.

[24] Hansen, L.; Liisberg, L.; Salamon, P. Ensemble methods for handwritten digit recognition. In *Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing*, 1992; pp. 333-42.

[25] Liu, Y.; Wang, Y.; Li, Y.; Zhang, B. F.; Wu, G. F. Earthquake prediction by RBF neural network ensemble. *Adv. Neural Networks - Isnn 2004, Pt 2*, **2004**, *3174*, 962-969.

[26] Yu, L.; Lai, K. K.; Wang, S. Multistage RBF neural network ensemble learning for exchange rates forecasting. *Neurocomputing*, **2008**, *71*(16-18), 3295-3302.

[27] Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C. A.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, **2000**, *16*(5), 412-24.

[28] Zhou, H. X.; Shan, Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins*, **2001**, *44*(3), 336-43.

[29] Ofran, Y.; Rost, B. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett.*, **2003**, *544*(1-3), 236-239.

[30] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.*, **2000**, *28*(1), 235-42.

[31] Hurley, J. H.; Faber, H. R.; Worthylake, D.; Meadow, N. D.; Roseman, S.; Pettigrew, D. W.; Remington, S. J. Structure of the regulatory complex of Escherichia coli IIIGlc with glycerol kinase. *Science*, **1993**, *259*(5095), 673-7.

[32] Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.*, **1997**, *271*(4), 511-23.

Received: April 31, 2009 **Revised: May 17, 2009** **Accepted: June 08, 2009**