Protein Similarity Comparison Based On Linear Neural Network And Multiple Parameters

Mimi Yin [1], Pingping Zhang [1], Hua Liu [2],Zhuye Gao[3],Jianhua Zhang [1, 2*]

[1] *School of Electrical Engineering*, Zhengzhou University, No. 100 Science Avenue, Zhengzhou, Henan,450001, China
[2]Biomedical Engineering Technology and Data Mining Research Institution, Zhengzhou University, No. 100 Science Avenue, Zhengzhou, Henan,450001, China
[3] *Education Department*, Xiyuan Hospital of China Academy of Chinese Medical Sciences,No.1 Xiyuan Playground, Hai Dian District,Beijing,100091,China

Abstract —The aim of the present study was to develop a new algorithm of protein structure similarity in which the similarity of many aspects of a protein were considered and the object of calculation is protein pdb data file from NMR (Nuclear Magnetic Resonance) not the sequence. Nine parameters (S1~S9) were selected for predicting protein similarity and they were similarities of spatial structure (density), atoms number, amino acids number, amino acid type, proportion of C element, proportion of N element, proportion of O element, spatial position of P atom and spatial position of S atom in the protein respectively. Assume that the relationship of the similarity (S) and S1~S9 was linear, then a linear neural network was used to optimize coefficients in this linear model. More than 500 pairs proteins data which collected from RCSB PDB were used to train this model. The performance of this model was evaluated and compared with BLAST in the end. The coefficients of each variables were obtained by a neural network and the formula of compute the overall similarity of two proteins was: S=0.3198S1+0.0343S2+0.0279S3+0.0618S4+ 0.0653S5+0.1062S6+0.1032S7+0.1477S8+0.1480S9-0.0142, where S1 - S9 are similarities of the spatial structure (density), atom number, amino acid number, amino acid type, the proportion of C element, the proportion of N element, the proportion of O element, spatial position of P atom and spatial position of S atom in the protein respectively. The study presented a new algorithm of protein structure similarity based on multiple parameters and linear neural network and it can be used to compute the structure similarity of any two proteins under conditions which are not applicable to BLASTp.

Keywords - Protein; Similarity; Multiple parameters; Neural network; Mathematical Model

I. INTRODUCTION

It is well known that the structure of a protein determines its function and many studies on the analysis of protein similarity have been reported. The main reason for determining the similarity of proteins is that structural similarity can be used to deduce functional similarity as well as predict the functions of some unknown proteins. Of course, most scholars obtain protein similarity by basic local alignment search tool (BLAST), but this method is always not feasible to new-found proteins, for there is no adequate homology information about them. To this end, it is indispensable to design an algorithm to compare structural similarity without relying on sequence homology, for the protein structures can be gained through the nuclear magnetic resonance (NMR) or X-ray diffraction experiment. However, there are still many other important aspects of protein similarity apart from structure that can be investigated, including the size of the protein in space, the distance between atoms, molecular corners, and the establishment of sphere model. All of these aspects can be used to calculate the similarity of proteins. Zou [1] divided the space occupied by a protein into spherical coordinates after transforming the Euclidean coordinates and counted the number of C, N and O atoms in each part of the space for estimation of the protein similarity. Harrison [2] divided the protein space evenly and recorded the statistical information of each part, changing the 3D structure diagram into a histogram. Cui [3] compared the 3D structure of proteins on the basis of their fractal characteristics. Xu and Dong [4] used the chaos game representation (CGR) technique for comparison of the protein similarity. Kotlovyi [5] had made a comparison of the parameters of the skeleton carbon atom curve in the two proteins, like curvature, torsion, translation variant et al., and similarity of the two segments of protein chains was decided within the scope of comparison. Xu[6] proposed that comparisons of protein similarity should be carried out within a multiple features framework. In other words, the similarity of proteins should be analyzed based on multiple aspects. This paper introduced an algorithm which compares the similarity of two proteins based on their structure files--pdb files. Nine aspects of a pair of proteins were considered, and each aspect of the similarity of proteins based on a BP neural network algorithm was weighted, thereby generating a comprehensive assessment of protein similarity by using weighted summation.

II. METHODS

A. Analysis

Nine parameters were selected for comparison. They were similarities of spatial structure (density), atoms number, amino acids number, amino acid type,proportion of C atom, proportion of N atom, proportion of O atom, spatial position of P atom and spatial position of S atom in the protein respectively [7-9]. Protein similarity comparison is to seek their functional similarity. So the selection of these nine parameters all affects the play of protein function to some extent. For example, proteins with different densities and sizes have different activities; proteins with different types of amino acids or different C, N and O components are bound to own different functions. In addition, P or S atom number in each protein is generally less, but their positions always have an important impact on certain function of the protein. All parameters are determined by all authors and several biology professors' discussion. A flowchart of the algorithm is given in Figure 1.

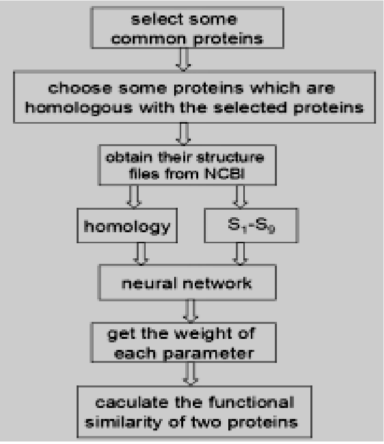

select some common proteins
choose some proteins which are homologous with the selected proteins
obtain their structure files from NCBI
homology
S_1-S_9
neural network
get the weight of each parameter
caculate the functional similarity of two proteins

Fig.1 The Flowchart of the Algorithm

B. Similarity of the proteins

1) Spatial density similarity (S_1)

To assess the density similarity of proteins, first, the geometrical center of each protein was determined so that the Euclidean coordinates obtained from the PDB structural file could be transformed into ordinary coordinates. Then, the distance from each of the atoms to the geometrical center was measured and the protein was divided into many layers in accordance with the distance. The number of atoms in each layer was counted and the number of atoms in the corresponding layer in other protein was also counted. The similarity of each layer in the two proteins was computed. This method can be described that the geometrical centers of two proteins are made to coincide so that the densities in the different regions can be compared. Assuming that the proteins were divided into n layers as follows, clearly, more layers will be more exact in division of the regions, in favor of a more exact estimation of the density similarity but increasing the computation amount. So the value of n should be decided according to the actual situation. Assuming that the number of atoms in the ith layer in one protein is n_1 and that of the second protein is n_2, then the similarity was calculated as following formula,

$$\mathrm{sim}_{\mathrm{i}} = 1 - \frac{|n_1 - n_2|}{n_1 + n_2} \tag{1}$$

In this way, the density similarity of each layer in the two proteins can be calculated. Weights were then added to the similarity of each layer and the overall density similarity was calculated by the weighted summation method. It is reasonable to suppose that the layers that contain the most atoms will be more likely to determine properties of the protein. Based on this assumption, the more atoms the layer owns, the higher the weight of this layer is, so the proportion of the atoms number in each layer determined the weight of this layer. Of course, it is maybe different in each layer for two proteins, so the average would be taken. Hence, each layer was weighted as formula 2 and the S_1 was computed as formula 3.

$$w_i = \left(\frac{l_{1i}}{n_1} + \frac{l_{2i}}{n_2}\right)/2, \quad \mathrm{i} = 1{,}2{,}\ldots\ldots \mathrm{n} \tag{2}$$

$$\mathrm{S}_1 = \lim_{\mathrm{n}\to\infty} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{w}_{\mathrm{i}}\mathrm{sim}_{\mathrm{i}} \tag{3}$$

where n_1 is the total number of atoms of the first protein,n_2 is that of the second protein and l_i is the number of atoms in the i th layer.

2) Similarity of atoms number, amino acids number and type (S2-S4)

The total number of atoms in a molecule determines its size and weight and the number and type of amino acids determine its function. Therefore, similarity of amino acids type is an important parameter that was weighted with a high value. The similarity of the three parameters, atoms number as well as amino acids number and type, was computed using equation (1) above.

3) Similarity of the C, N and O atoms (S5-S7)

The C, N and O atoms make up close to 90% of the atoms in a protein. Therefore, the similarity in the proportion of these atoms in a protein was also used to assess protein similarity. Similarly, the similarity of proportions of these atoms was calculated as,

$$\mathrm{sim} = 1 - \frac{|\mathrm{p}_1 - \mathrm{p}_2|}{\mathrm{p}_1 + p_2} \tag{4}$$

where p_1 is the proportion of the atom in one protein and p_2 is that of the corresponding atom in the other protein.

4) *Similarity of the P and S atoms (S8-S9)*

Although the P and S atoms make up a very small proportion of the atoms in proteins, they are extremely important. On average, there are only several P and S atoms in a single protein and their functions in a protein depend on their positions in the spatial structure of the protein. Therefore, the following method was developed to compare the similarity of the P and S atoms in a pair of proteins: If no P/S atoms were present in the two proteins, the similarity parameter was set as 1.0, and if there were P/S atoms in one protein but not in the other, the similarity parameter was set as 0. When there were P/S atoms in the two proteins, the layer in which they were found was determined. If the P/S atoms in the two proteins were in the same layer, the similarity was set as 1.0, and if the P/S atoms in the two proteins were in adjacent layers, the similarity was set as 0.8. Under all other conditions, the similarity was set as 0.

C. *Determination of the Weights of Each Parameter Similarity*

1) *Training sample data*

From the RCSB-PDB (http://www.rcsb.org/pdb/home/home.do), Some common proteins such as S100 protein family, ACYP1 [PDB:2K7K], TNF-18 [PDB:2R32] (denoted as “sample” proteins) were selected to put into the BLASTP searches respectively and the information of their homology was obtained then, including the numbers of the homology and the names of proteins (denoted as “comparison” proteins) which were homologous with the “sample” proteins. Next, the tertiary structure files (format of “pdb”) of every “sample” or “comparison” proteins were obtained from the RCSB-PDB according to their names. From the pdb files, only the information about the types of amino acids and atoms in each of the proteins as well as the Euclidean coordinates of the proteins were extracted. Then, the similarity of each parameter between each “comparison” protein and its corresponding “sample” protein could be calculated. Finally, a more than 600 data set which prepared to train the neural network model were obtained and a fraction of data were shown in Table I to display the data form, where S_1-S_9 is the similarity of nine parameters of two proteins; H is the homology obtained by BLAST. In neural network, S_1-S_9 can be seen as input vectors and S can be seen as output vector.A fraction of data were shown in Table I to display the data form.

TABLE I. PART OF SAMPLE DATA USED FOR TRAINING THE NEURAL NETWORK

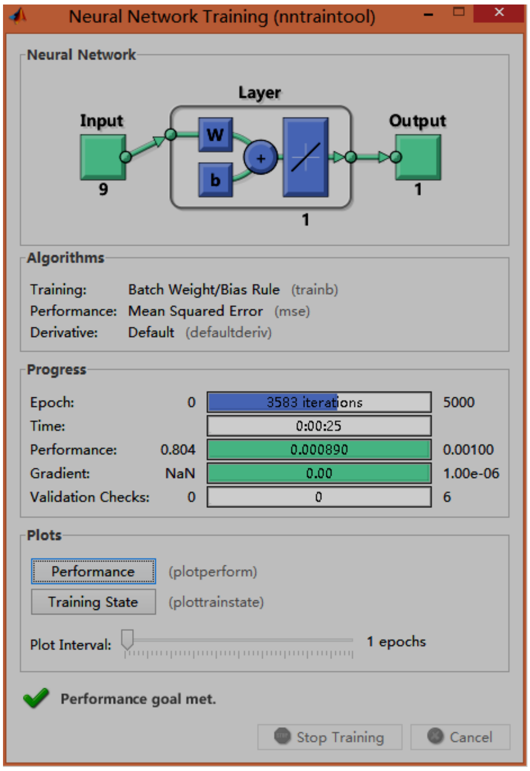
Protein1	Protein1	H	S1	S2	S3	S4	S5	S6	S7	S8	S9
2M3W	2LLU	1.0000	0.5142	0.5010	1.0000	0.9744	0.9987	0.9969	0.9987	1.0000	0.0000
2M3W	2LHL	0.9900	0.9681	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8000	0.8000
2M3W	2LLS	0.9900	0.9743	0.9986	1.0000	0.9730	0.9992	0.9943	0.9981	1.0000	1.0000
2M3W	2JPT	0.9800	0.9459	0.9955	1.0000	1.0000	0.9976	0.9958	0.9889	0.0000	0.8000
2M3W	2LLT	0.9900	0.9552	0.9997	1.0000	1.0000	0.9997	0.9997	0.9931	1.0000	1.0000
2M3W	1K2H	0.9400	0.9823	0.9990	1.0000	1.0000	1.0000	0.9946	0.9978	0.8000	0.0000
4DUQ	2RGI	0.9600	0.9846	0.9926	0.9780	0.8148	0.9907	0.9883	0.9860	1.0000	1.0000
3M0W	3C1V	1.0000	0.9453	0.9464	0.9838	0.7692	0.9959	0.9968	0.9773	1.0000	0.0000
3M0W	3ZWH	0.9600	0.9527	0.9691	0.9838	0.9756	0.9996	0.9741	0.9899	0.0000	1.0000
3M0W	4ETO	1.0000	0.9826	0.9840	0.9945	0.9091	0.9961	0.9760	0.9945	1.0000	0.8000
3M0W	4CFQ	0.9700	0.6875	0.6854	0.9775	0.8163	0.6661	0.6854	0.6660	0.8000	1.0000
2LUC	1NSH	0.8900	0.9589	0.9823	0.9806	1.0000	0.9980	0.9823	0.9856	1.0000	1.0000
1K9P	1K8U	0.9800	0.9937	0.9979	1.0000	0.9362	0.9988	0.9979	0.9908	0.8000	1.0000
4AQJ	1PSR	1.0000	0.9627	0.9626	0.9796	0.8889	0.9955	0.9932	0.9928	1.0000	1.0000
4AQJ	2WND	0.9700	0.9870	0.9948	1.0000	0.9091	0.9990	0.9948	0.9985	1.0000	0.8000
4AQJ	4AQI	0.9300	0.9896	0.9961	1.0000	0.9565	0.9971	1.0000	0.9904	0.8000	1.0000
2LE9	2H2K	1.0000	0.3983	0.3783	0.9630	0.9756	0.6594	0.6702	0.6534	0.0000	1.0000
2LE9	1YUT	1.0000	0.6694	0.6709	1.0000	0.9756	0.9998	0.9989	0.9984	1.0000	1.0000

2LE9	2KI4	0.9900	0.9512	0.9953	1.0000	0.9767	0.9998	0.9989	0.9969	1.0000	0.8000
2LE9	1YUR	1.0000	0.6837	0.6709	1.0000	0.9756	0.9998	0.9989	0.9984	1.0000	1.0000
4AQI	4AQJ	0.9400	0.9896	0.9961	1.0000	0.9565	0.9971	1.0000	0.9904	0.0000	1.0000

2) Neural network training

Currently, the most commonly used neural network models are feed forward network model, back propagation network model, back-propagation algorithm (BP) model, support vector machine network model, feedback network model and self-organizing network model [10,11]. Perceptron is one of simple feed forward neural network and can be classified into two kinds-multilayer perceptron and single layer perceptron,they are all Binary linear classifier.Single layer perceptron is the most simple neural network,It contains input layer and output layer and the two layers are connected.Assumed that the the model of S and S1~S9 is linear,so the single layer perceptron was used to modeling a equation.The training process of is a weight adjustment process and the weight of each parameter could be obtained.In this study, the number of inputs is nine and the number of output is 1, so a 1*9 weight matrix was generated finally using the nine parameters described above.

The process of the training is shown in Figure 2. In Figure 2, the 'Neural Network' section presented the number of nodes of input layer and output layer. The 'Algorithms' section displayed the training function and error function (mean squared error), etc. The 'Progress' section revealed the training times, time, error and error gradient, etc. The 'Plots' section was a graphical representation of the parameters. From figure2, the process was run more than 3000 times to obtain a mean value, so that robust results were obtained. The final weights of the parameters used in this study are shown in Table II. To test the reliability of the network, the result of the weights and some test data were input to it and conducted a simulation. The result of the simulation is shown in Figure 3. As can be seen, the actual outputs were very close to the expected outputs, suggesting that the result of the training was good and the network was reliable.



Algorithms

Training: Batch Weight/Bias Rule (trainb)
Performance: Mean Squared Error (mse)
Derivative: Default (defaultderiv)

Progress

Epoch:	0	3583 iterations	5000
Time:		0:00:25	
Performance:	0.804	0.000890	0.00100
Gradient:	NaN	0.00	1.00e-06
Validation Checks:	0	0	6

Plots

Performance (plotperform)
Training State (plottrainstate)
Plot Interval: 1 epochs

✓ Performance goal met.

Fig.2 Training Process

TABLE II. THE COEFFICIENTS OF PARAMETERS AFTER TRAINING

Parameter	Coefficients
Protein structural similarity（spatial density distribution）	0.3183
Total atom number similarity	0.0343
Amino acid number similarity	0.0204
Amino acid type similarity	0.0603
Ratio of nitrogen similarity	0.0653
Ratio of carbon similarity	0.1062
Ratio of oxygen similarity	0.1002
Spatial position of phosphorus atom similarity	0.1477
Spatial position of sulfur atom similarity	0.1480
B(constant)	-0.0142

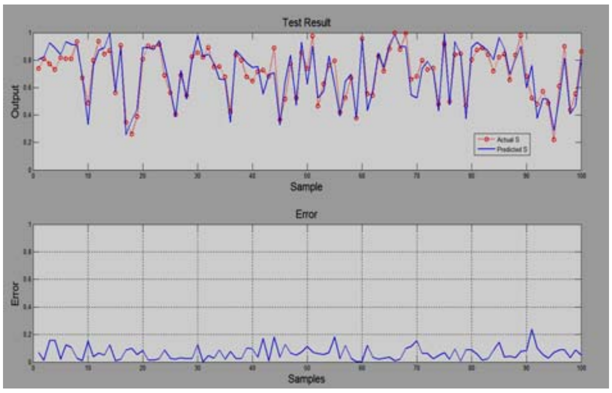


Fig.3 Result of the Test

3) *Overall Similarity of Two Proteins*

From the above, the similarity of the nine characteristics of a protein was calculated using MATLAB and the weighted parameters were calculated using a BP neural network. So the overall similarity of two proteins was obtained after weight summation according to the following formula (5).

$$S=\sum_{i=1}^{k} w_i s_i + B, \quad i=1,2,\ldots.9. \tag{5}$$

4) *Test the Practicability*

To verify the practicability of the algorithm, several pairs of protein that share different similarity were chosen as the test pairs.There are some literature reported about their similarity[12,13]. The similarity between each pair of proteins was calculated respectively by the algorithm in this paper. Moreover, BLAST was used to obtain the sequential homology of each pair of proteins. The results are shown in Table III, where S is the overall similarity calculated with the weighted sum. S_1-S_9 are the similarities of each parameter in this order: number of atoms; number of amino acids; amino acid type; proportion of the C, N and O atoms; position of P and S atoms; and spatial density. H from BLAST means the homology obtained from BLAST tool.

II. TABLE.III THE RESULT OF SIMILARITY COMPUTED BY THE ALGORITHM AND BLAST

Protein1	Protein2	S1——S9	S	H from BLAST
109L	110L	0.9890, 1.0000, 1.0000, 0.9889, 0.9932, 0.9723, 1.0000, 1.0000, 0.9861	0.9775	0.99
101M	102M	0.9930, 0.9992, 0.8800, 0.9880, 0.9792, 0.9619, 0.8000, 1.0000, 0.9835	0.9374	0.99
1AAX	101M	0.5223, 0.4628, 0.9630, 0.9907, 0.9787, 0.9802, 0.8000, 0, 0.6780	0.6720	0.32
1AAX	102M	0.0990, -0.1623, 0.9200, 0.9974, 0.9995, 0.9813, 1.0000, 0, 0.6846	0.6766	0.32
3B94	2R30	0.8147,0.9829,1.0000,0.9977,0.9864,0.9900,1.0000,1.0000 ,0.8534	0.9305	0.98
3B94	2R32	0.5521,0.9943,1.0000,0.9889,0.9988,0.9679,0 ,1.0000,0.7032	0.7248	0.70
3B94	4DB5	0.6615,0.7143,0.1500,0.9974,0.9770,0.9793,1.0000 ,1.0000,0.7713	0.8303	0.72

3WD5	2TNF	0.0309,1.0000,0.8421,0.9873,0.9383,0.9946,1.0000, 1.0000, 0.4101	0.7151	0.79
3WD5	3IT8	0.9975,1.0000,1.0000,0.9972,0.9879,0.9797,1.0000,1.0000,0.9858	0.9784	0.99

III. RESULTS

The study presented an algorithm of protein similarity comparison based on multiple parameters. 9 parameters which are related to the similarity of protein were considered and they are similarities of spatial structure (density), atoms number, amino acids number, amino acids type, proportion of C atoms, proportion of N atoms, proportion of O atoms, spatial position of P atoms and spatial position of S atoms in the protein, respectively. The weights of each parameter were obtained by a BP neural network. The overall similarity of two proteins was calculated according to the following formula:$S= 0.3198S_1+0.0343S_2+0.0279S_3+0.0618S_4+0.0653S_5+ 0.1062S_6+0.1032S_7+0.1477S_8+0.1480S_9-0.0142$, where S_1--S_9 were the 9 parameters in order above. After application in a few proteins with different similarity, it was confirmed that the model can be used for protein similarity calculation.

IV. DISCUSSION

An algorithm for assessing protein similarity based on the comparison of multiple parameters was introduced in this paper. The similarity of nine characteristics of proteins was considered, thus overcoming some of the previous shortcomings of methods that considered only protein structure. Xu[6] also used a method that could compare protein similarity in multiple frameworks by establishing a fuzzy matrix with only four parameters (number of skeleton C_α atoms, number of mutant atoms, number of hydrophilic particles and number of helices). From Table II, it was concluded that the density similarity of the proteins (regarded as atomic density similarity of protein) played a vital role in the overall similarity because it accounted for about one-third of the weight. This is a reasonable finding because the protein pairs were selected based on their high sequence identity, which would have influenced their density or structure. In addition, the weight of the amino acid kinds similarity was found to be higher than either the weight of the amino acid number similarity or the weight of total number of atoms similarity (Table II), which could be explained as the kinds of amino acids in a protein are more important than their numbers, for the more types are, the more complicated the protein functions are. Furthermore, the weights of the P/S similarity were higher than that of the C, N and O atoms similarity, which indicated that the P/S atoms are probably more important to the function of proteins than the C, N and O atoms.

The calculation results of the algorithm are basically close to but not exactly same with BLAST. The comparison objects of BLAST are amino acid sequences of two proteins. However, the calculation object in this paper of the algorithm is protein pdb file from RCSB PDB. The pdb files in this database are obtained from NMR and X-ray diffraction and reflect the real spatial structure of the protein. A kind of protein owns just one amino acid sequence, but when in a different environment or combined with different ligands, they will have different structures. In other words, even if the BLAST results show the homology of two kinds of amino acid sequences is very high, the structures of two proteins may be quite different from each other under different conditions. That is why this article started from protein spatial structure to calculate the similarity of protein.

In the previous work,authors applied the algorithm to calculated the similarity of a uncommon gastric cancer protein p42.3.For p42.3 was a new-found protein,BLAST can't obtained its helpful homology information,authors applied the presented algorithm to calculated an sorted the similarity of p42.3 and the proteins which with similar structural domain with it.From the structure similarity information,several proteins (such as S100A11)were found that with similar function with p42.3 and the function regulation path was predicted.The following molecular biology experiments proved this path,which indicated the algorithm presented in this paper can be applied in functional analysis.Details see the literature[14],[15] please.

V. CONCLUSIONS

The study presents a relatively scientific and comprehensive algorithm of protein similarity based on multiple parameters. Nine aspects of proteins are considered and each of their weight is obtained after a series of calculations, which addresses some of the limitations of other methods that take a single aspect into consideration when compared protein similarity. The algorithm can be used to compute the similarity of any two proteins and has proved a relatively accurate result.

As can be seen from Table III, most of the similarities calculated by the algorithm were close to BLAST homology, and several of them had some difference. The reason may be that the algorithm is mainly used to calculate the similarity between the two kinds of protein structure, while BLAST just takes the compatibility of two amino acid sequences into account. Some literatures have already reported

There there are high similarity between 109L and 110L as well as 101M and 102M. While in terms of protein 1AAX and 101M or 102M, similarity between them is relatively low [12, 13]. Their homology obtained from BLAST also proves this point. The results in Table III are basically in coincident with this fact, and it further verifies that algorithm in this paper is feasible.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGMENT

This work is supported by the Youth Found of Natural Science Foundation of China (Grant No.: 813D3150) and Special Scientific Research of Chinese Medicine Industry, China (No.201007001). In addition,thanks should be given to Youyong Lv from the Molecular Biology Laboratory,Peking University for his generous help in discussing this research and for the valuable suggestions for this algorithm.

REFERENCES

[1] Zou BJ,Zhang Q,Liang LM."Similarity compare of protein structure divided based on spheric polar coordinates". *Journal of Computer Aided Design and Computer Graphics*,vol.21,pp. 606-611,2005.

[2] Harrison A, Pearl F, Sillitoe I,et al. "Recognizing the fold of a protein structure",*Bioinformatics* ; vol.19,pp.1748-1759,2003.

[3] Chenyang C,Donghui W,Xin Y, "3D protein structures similarity matching based on fractal features". *In Photonics Asia 2004: Information Processing and Data Storage 11*, 2004.

[4] Xu Z, Dong HW. "Similarity compare of protein based on CGR". *Computer Engineering*,vol.36,pp. 233-234,2010.

[5] Kotlovyi V, et al. "Protein structural alignment for detection of maximally conserved regions". *Biophysical Chemistry*,vol. 105,pp. 595- 608,2003.

[6] Xu Z, Dong HW. "Similarity comparison and classification of protein under the multiple feature framework". *Engineering Graph Journal* vol.1,pp. 185-190,2010.

[7] Wang X, et al. "Statistical analysis of protein spatial structure". *Journal of Shanxi Datong University* ,vol.24pp.3-8,2010.

[8] Zhang SS, et al. "Statistical analysis and application of the protein structure". *Journal of Wuhan Inst Technology* ,vol.32.pp. 45-48,2010.

[9] Shi OY, et al. "Disulfide bond prediction integrated protein secondary structure information".*Application Research Computer*,vol. 28,pp.2049-2077,2011.

[10] University, Chengdu,2007.

[11] Wu CY. "The research and application on neural network". Northeast Agricultural University, Yangling,2009.

[12] Zhang JH, Chen ZT. "Analyzing influence on the conformation of single-chain antibody with the differential length of linkers". *African Journal of Microbiology Research*, vol.5,pp.5737-5744, 2011.

[13] Gao HL. "The Analysis of Protein Spatial Structure Similarity Comparison". Dalian Jiaotong University,Dalian, 2012.

[14] Jianhua Z, Chunlei L, Zhigang S,Rui X, Li S,et al."p42.3 gene expression in gastric cancer cell and its protein regulatory network analysis",*Theoretical Biology and Medical Modelling*, vol.9(1), pp.53, 2012.

[15] Zhang JH, Ma W, Shang ZG, Xing R,Shi L,"Analysis of the p42.3 protein structure and regulatory network",*Chinese Science Bulletin*, vol.58(8), pp.869-872, 2013.