Available online at www.sciencedirect.com

SCIENCE DIRECT®

Biochemical and Biophysical Research Communications 318 (2004) 142–147

BBRC

www.elsevier.com/locate/ybbrc

Prediction of the disulfide-bonding state of cysteines in proteins based on dipeptide composition

Jiang-Ning Song,[a,b,*] Ming-Lei Wang,[a,b] Wei-Jiang Li,[a,b] and Wen-Bo Xu[c]

[a] The Key Laboratory of Industrial Biotechnology, Ministry of Education, Southern Yangtze University, Wuxi 214036, China
[b] School of Biotechnology, Southern Yangtze University, Wuxi 214036, China
[c] School of Information Technology, Southern Yangtze University, Wuxi 214036, China

Received 24 March 2004

Abstract

In this paper, a novel approach has been introduced to predict the disulfide-bonding state of cysteines in proteins by means of a linear discriminator based on their dipeptide composition. The prediction is performed with a newly enlarged dataset with 8114 cysteine-containing segments extracted from 1856 non-homologous proteins of well-resolved three-dimensional structures. The oxidation of cysteines exhibits obvious cooperativity: almost all cysteines in disulfide-bond-containing proteins are in the oxidized form. This cooperativity can be well described by protein's dipeptide composition, based on which the prediction accuracy of the oxidation form of cysteines scores as high as 89.1% and 85.2%, when measured on cysteine and protein basis using the rigorous jackknife procedure, respectively. The result demonstrates the applicability of this new relatively simple method and provides superior prediction performance compared with existing methods for the prediction of the oxidation states of cysteines in proteins.

Keywords: Cysteine; Disulfide-bonding state; Protein folding; Dipeptide composition; Bioinformatics; Cooperativity

Disulfide bonds are primary covalent crosslinks between cysteine side chains that play very important roles in the native structures of globular proteins. Such bonds can stabilize protein spatial conformation and ensure that protein will perform its biochemical function [1]. The correct formation of disulfide bonds is the crucial step in the folding pathway [2,3]. Many theoretical and experimental studies indicated that disulfide bridges can increase the conformational stability of proteins mainly by reducing the conformational entropy of the unfolded state and constraining the unfolded conformation [4–9]. Several analyses of the characteristics of disulfide bonds and detailed conformational analysis of cysteines, as well as amino acid neighbors in proteins have been performed [10,11]. But information of such important bonds cannot be derived directly from amino acid sequences. Previously numerous researches on disulfide bridges were reported, most of them were mainly time-consuming experimental works [12–18].

Disulfide-bonding pattern information can help understand structural properties of proteins and identify which family a protein belongs to, giving important insights into its biological functions. More recently, Chuang et al. [19] found that there exists a very close relationship between the disulfide-bonding patterns and protein structures, based on which it is feasible to discriminate structure similarities and identify protein homologs. van Vlijmen et al. [20] constructed a comprehensive database of disulfide-bonding patterns and developed search method to find related protein homologs with similar disulfide patterns. In protein folding prediction, the localization of disulfide bridges can strongly reduce the search in the conformational space [21,22]. Thus, the accurate predictions of disulfide connectivity in proteins would have potentially important applications, both in introducing engineered disulfide bonds to increase the conformational stability of proteins and helping locate disulfide bridges to aid three-dimensional structure predictions.

Methodologies related to the prediction of disulfide bridges can be decomposed into two steps. First, the

[*] Corresponding author. Fax: +86-510-580-6493.
E-mail address: sjnbeckham@yahoo.com.cn (J.-N. Song).

0006-291X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2004.03.189

disulfide-bonding state of each cysteine is predicted from protein amino acid sequence, a typical binary classification problem. Subsequently, the second step is to locate the actual disulfide connectivity from candidate oxidized cysteines, which has received relatively scant attention in the published literature. Fariselli and Casadio [22] presented a method based on the weighted graph representation of disulfide bridges and achieved 17 times accuracy higher than that of a random predictor in the case of proteins with 4 disulfide bonds. Then a neural network based approach was adopted to solve the pairing problem and received satisfactory results for the simplest cases (2 or 3 disulfide bonds in one protein) [23]. More recently, Vullo and Frasconi proposed a novel machine learning method based on extended recursive neural networks (RNN) to predict the disulfide connectivity patterns in cysteine-rich proteins [24]. They further improved the prediction performance by incorporating evolutionary information in the form of multiple alignment profiles.

This paper focused on the first task of the prediction of the disulfide-bonding state of cysteines in proteins, i.e., to predict which cysteines in protein sequence are oxidized. Concerning this topic, theoretical investigations emerged in recent years. Muskal et al. [25] predicted the disulfide-bonding states of cysteines by means of neural networks. They used local sequences, i.e., the flanking amino acid sequences of cysteines as input and an overall accuracy of 80% was achieved. By using additional evolutionary information, higher success ratio can be obtained [26]. Fiser et al. [27] also used local sequence information but they employed statistical method. Their method performed at 71% prediction accuracy. Since disulfide bridges are crucial to maintain proper structures of proteins, oxidized cysteines that take part in disulfide bonds should be more conserved than free cysteines. Based on this idea, multiple sequence alignment was used to predict the oxidation state of cysteines, the success rate of which is about 80% [28,29]. If a constant threshold was used, the overall prediction accuracy could rise above 84% [29]. Mucchielli-Giorgi et al. [30] used logistic functions learned with subsets of proteins with similar amino acid compositions to predict the disulfide-bonding state and reached success rates close to 84%.

Support vector machine based predictor that operated at two stages (a multi-class classifier at the protein level and a binary classifier at cysteine level) was suggested by Ceroni et al. [31]. They achieved 85% accuracy measured by fivefold cross-validation. Martelli et al. implemented a hybrid system (hidden neural network) that combines a hidden Markov model (HMM) and neural networks (NN). After 20-fold cross-validation procedure, the predictor accuracy scores as high as 88% and 84%, measured on cysteine and protein basis, respectively [32,33].

In this paper, we presented a novel approach by means of a linear discriminator based on their dipeptide composition. Predictions based only on single amino acid composition may lose some sequence-order information, but incorporating this information may improve prediction performance. Dipeptide composition can be considered as another simple representative form of protein's incorporating neighborhood information. For this case, comparatively higher prediction accuracy using this approach could be improved to 89.1% and 85.2% in a jack-knife test, when measured on cysteine and protein basis, respectively.

Materials and methods

The protein dataset. 8114 cysteine-containing protein chain structures were used in this work, which were taken from the PISCES Culled PDB (available at http://www.fccc.edu/research/labs/dunbrack/pisces/) [34], a protein sequence culling server, which is a representative dataset of accurately resolved non-homologous Protein Data Bank (PDB) [35] structures. The PISCES Culled PDB list used in this paper was generated on 27 April, 2003. Information about disulfide bonds was extracted directly from the SSBOND records of the PDB entries.

All structures in this list had resolution better than 2.5 Å, determined only by X-ray diffraction with a crystallographic *R* factor less than 25%. Sequence identity between each pair of the sequences was less than 25%. Chains with sequence length shorter than 50 amino acids and without any cysteines were excluded. After this filtering procedure, the total number of protein chains in the final dataset is 1856 (protein chains with cysteines that are interchain disulfide bonded are included as "free" cysteines). Less than 15% of PDB sequences in this dataset have more than five disulfide bridges (see Fig. 1).

According to whether containing intra-chain disulfide bonds, the 8114 cysteine-containing protein chains were divided into two classes, which are called OXICYS and REDCYS for convenience. Proteins in REDCYS have no intra-chain disulfide bridges, all cysteines are in reduced form. Every protein in OXICYS has at least one disulfide bridge. Among the total 1856 protein chains, there are 459 chains belonging to OXICYS and 1397 chains belonging to REDCYS, with totally 2354 cysteine-containing segments in the disulfide-bonded state

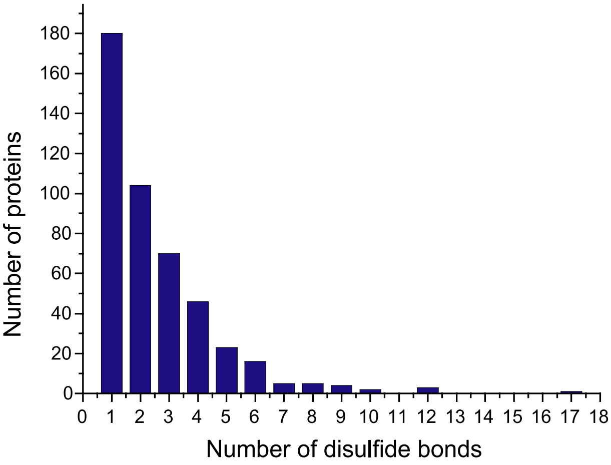

Fig. 1. Distribution of disulfide bridges per protein sequences in the dataset. Chains are grouped according to the number of disulfide bonds.

forming 1177 disulfide bonds and 5760 in the non-disulfide-bonded state.

Prediction methods. The 459 OXICYS proteins have 2719 cysteines, of which 2354 take part in intra-chain disulfide bonds. That is to say, almost all (87%) cysteines in OXICYS proteins are oxidized. While 5395 cysteines in 1397 REDCYS proteins are all in free form. This key fact that cysteines (REDCYS) and half cysteines (OXICYS) rarely co-occur was also noticed by other researchers [30,31]. This is an obvious cooperation phenomenon that cannot be elucidated by only local sequences near cysteines. The cooperativity is a global characteristic that reflects properties concerning protein structure, and there must be some global sequence information to account for it. In the present paper, we proposed a new two-class predictor for predicting the oxidation state of cysteines in proteins by means of a linear discriminator, which explores the dipeptide composition of protein sequence.

Predicting the disulfide-bonding state of cysteines based on 400 dipeptide composition. For a protein k in the dataset, we define a characteristic index Q_k,

$$Q_k = \begin{cases} +1, & \text{if protein } k \text{ belongs to OXICYS class,} \\ -1, & \text{if protein } k \text{ belongs to REDCYS class.} \end{cases} \tag{1}$$

We try to predict the characteristic index Q_k of protein k by means of its 400 dipeptide composition $p_{ab}^{(k)}$. We use the linear function of $p_{ab}^{(k)}$ to approximate Q_k, namely,

$$Q_k = \sum_{ab} v_{ab} p_{ab}^{(k)}, \tag{2}$$

where ab stands for one dipeptide (a and b could be the same or different amino acid), and the summation runs over all the 400 types of dipeptides. The parameters v_{ab} are constants for all proteins. In order to choose the parameters v_{ab} that best fit the dataset, we minimize

$$Z = \sum_k \left(Q_k - \sum_{ab} v_{ab} p_{ab}^{(k)} \right)^2, \tag{3}$$

by letting $\partial Z / \partial v_{cd} = 0$ for all dipeptides cd, which leads to

$$\sum_{ab} \left(\sum_k p_{ab}^{(k)} p_{cd}^{(k)} \right) v_{ab} = \sum_k Q_k p_{cd}^{(k)}, \tag{4}$$

where the summations on k run over all protein sequences in the dataset. By solving Eq. (4), the fitted parameters v_{ab} can be obtained.

With these parameters one can calculate the quantity Q for a given protein with dipeptide content p_{ab} as follows:

$$Q = \sum_{ab} v_{ab} p_{ab}, \tag{5}$$

which is designed to approach the characteristic index of the protein (+1 for OXICYS and −1 for REDCYS). To test the fitness, we computed the following cumulative distributions:

$$F_{\text{OXICYS}}(Q_0) = \frac{\text{The number of OXICYS proteins with } Q \geqslant Q_0}{\text{The number of all OXICYS proteins}}, \tag{6}$$

and

$$F_{\text{REDCYS}}(Q_0) = \frac{\text{The number of REDCYS proteins with } Q < Q_0}{\text{The number of all REDCYS proteins}}. \tag{7}$$

Measurement accuracy. The prediction quality was examined using the jack-knife test (leave-one-out procedure), an objective and rigorous testing procedure. In comparison with subsampling test or independent dataset test, the jack-knife test is thought to be more rigorous and reliable [36]. During the process of jack-knife test, each protein was singled out in turn as a test protein with the remaining proteins used as training set to calculate the test sample's v_{ab} parameters and predict the class (OXICYS class or REDCYS class). The prediction quality was evaluated by the overall prediction accuracy and prediction accuracy for each cysteine and each protein chain.

Denote n_{xy} the number of proteins that are predicted as x class and in fact they belong to y class, where x, y = o (OXICYS), or r (REDCYS). Therefore, the overall prediction accuracy is

$$Q2 = P/N = \frac{n_{\text{oo}} + n_{\text{rr}}}{n_{\text{oo}} + n_{\text{rr}} + n_{\text{oo}} + n_{\text{rr}}}, \tag{8}$$

where P is the total number of correctly predicted cysteines and N is the total number of cysteines.

The other measure of prediction accuracy is Matthew's correlation coefficient (MCC) [37] between the observed and predicted cysteines, based on the cysteine basis or between the observed and predicted proteins, based on the protein basis, as given by

$$\text{MCC}(s) = \frac{p(s)n(s) - u(s)o(s)}{\sqrt{(p(s) + u(s))(p(s) + o(s))(n(s) + u(s))(n(s) + o(s))}}. \tag{9}$$

Here, for each class s (OXICYS class or REDCYS class), $p(s)$, and $n(s)$ are the total number of correct predictions and correctly rejected assignments, respectively, and $u(s)$ and $o(s)$ are the number of under- and over-predictions. The more MCC is, usually the higher the prediction reliability is.

The accuracy for each discriminated class s is evaluated as

$$Q(s) = \frac{p(s)}{p(s) + u(s)}. \tag{10}$$

For sake of further explanation, when s refers to the OXICYS class and REDCYS, respectively, Eq. (10) equivalent to the following Eq. (11):

$$Q_{\text{oxi}} = \frac{n_{\text{oo}}}{n_{\text{oo}} + n_{\text{or}}}, \quad Q_{\text{red}} = \frac{n_{rr}}{n_{\text{rr}} + n_{\text{ro}}}, \tag{11}$$

where Q_{oxi} and Q_{red} are the success rates for OXICYS and REDCYS class, respectively. $p(s)$ and $u(s)$ are the same as in Eq. (9).

Also, the probability of correct predictions $P(s)$ is calculated as

$$P(s) = \frac{p(s)}{p(s) + o(s)}, \tag{12}$$

where $n(s)$ and $o(s)$ are the same as in Eq. (9).

Finally, the prediction accuracy per protein is

$$Q2_{\text{prot}} = \frac{P_{\text{p}}}{N_{\text{p}}}, \tag{13}$$

where P_{p} is the number of proteins whose cysteines are all correctly predicted and N_{p} is the total number of proteins.

Results and discussion

Cumulative distribution of Q values

The results are depicted in Fig. 2.

Fig. 2 shows clearly that the Q value is a good index to distinguish the two classes of proteins. Therefore, the classification of a protein can be predicted based on its Q value: If $Q > Q_{\text{c}}$ then the protein is predicted as OXICYS, otherwise REDCYS, where Q_{c} is a critical value. From Fig. 2, we could also observe that $Q_0 = 0$ is usually not the best-fitted critical value, i.e., in most cases 0 and Q_{c} do not match each other. The highest prediction accuracy may be achieved at the value of Q_0 less than zero.

Cysteine conservation and sequence environment conducive to disulfide bond formation

Cysteines tend to be more conserved in proteins when they pair to form disulfide bridges, which may reflect

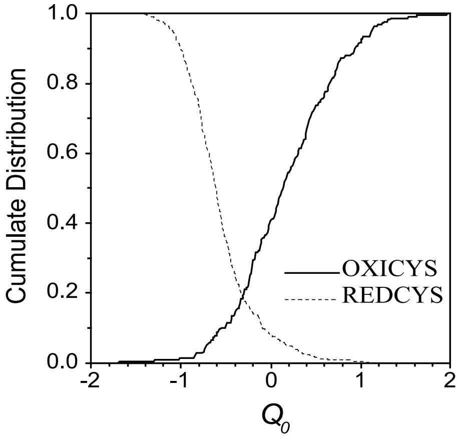

Fig. 2. Cumulative distributions of the Q values for OXICYS and REDCYS proteins.

Table 1
Distinguished dipeptide patterns that favor disulfide bond formation and non-disulfide bond formation

Distinguished dipeptide patterns	
Disulfide bond formation favoring	
AX	X = any of 20 amino acids, except A, Q, W, and Y
NX	X = any of 20 amino acids, except A, H, L, and Y
QX	X = any of 20 amino acids
YX	X = one of 20 amino acids, except Q and Y
Non-disulfide bond formation favoring	
XC	X = any of 20 amino acids, except G, I, K, and M
XG	X = any of 20 amino acids, except M
XI	X = any of 20 amino acids, except G, I, M, and T
XM	X = any of 20 amino acids, except M
XT	X = any of 20 amino acids, except C, G, and M

their crucial and essential role in maintaining protein structure stability and biological functions. As reported previously, the amino acid composition of oxidized and reduced cysteines shows clear differences. In the case of oxidized cysteines, serine (S), threonine (T), and glutamine (Q) are residues highly conducive to disulfide bond formation while histidine (H), glutamate (E), lysine (K), and arginine (R) are more frequently found in case of reduced cysteines [28].

In the present paper instead of using amino acid composition, we use protein's dipeptide composition to reveal the hidden information conducive to disulfide bond formation. By solving Eq. (4), the fitted parameters v_{ab} can be obtained. The larger the v_{ab} value of one corresponding dipeptide, the stronger the relevance of that dipeptide to be distinguished mark for the disulfide bond (non-disulfide bond) formation. Four hundred dipeptide contribution weights to disulfide formation are shown in Fig. 3. Darker bars highlight the relative dipeptides more conducive to disulfide-bonding state, while lighter bars indicate the corresponding dipeptide more conducive to non-disulfide bond formation.

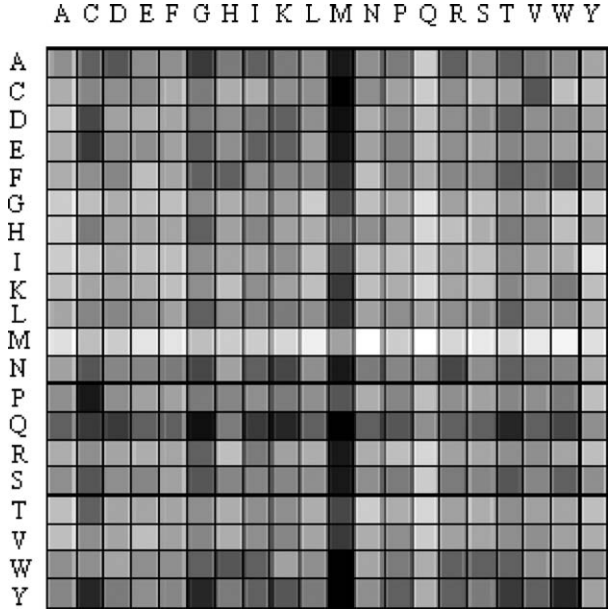

Fig. 3. Four hundred dipeptide contribution weights to disulfide bond formation. Bar's dark shades indicate high propensity to form disulfide bonds of the corresponding dipeptide while light shades indicate low propensity for disulfide bond formation.

Noticeably and interestingly, there exist some distinguished dipeptide patterns that strongly favor the disulfide bond formation and non-disulfide bond formation of cysteines, respectively, as displayed in Table 1.

It can be seen from Table 1 that alanine (A), asparagine (N), glutamine (Q), and tyrosine (Y) are four favorable residues conducive toward disulfide bond formation when located in the first position in dipeptide pairs, while cysteine (C), glycine (G), isoleucine (I), methionine (M), and threonine (T) are five residues more frequently involved in the environment of non-disulfide bond when in the second position in dipeptide pairs.

Prediction performance based on 400 dipeptide composition

As is shown in Table 2, a remarkable improvement has been achieved in prediction quality. Predictions based only on single amino acid composition may lose some sequence-order information hidden in the protein sequence, but incorporating this information may improve prediction performance. On the other hand, dipeptide composition can be considered as another representative form of proteins incorporating sequence neighborhood information [39]. Dipeptide composition representation can be regarded as a sort of 2-gram method. This method extracts and computes the occurrences of two consecutive residues from a sequence string in a sliding window fashion. Therefore, the counts of all 2-gram dipeptide patterns are 400 parameters

Table 2
Prediction accuracy (%) of the predictor based on the 400 dipeptide composition by the jack-knife test

Method	Prediction accuracy (%)					
	Q_c	MCC	Q_{oxi}	Q_{red}	$Q2$	$Q2_{prot}$
400 dipeptide composition	−0.2	67.6	93.2	71.9	87.2	83.1
	−0.1	70.7	92.2	79.3	89.1	85.2
	0	68.8	90.7	81.5	88.7	84.9
	0.1	68.6	89.2	86.8	88.8	85.2
	0.2	65.1	87.4	88.4	87.6	84.1

(dimension vectors), which can be used to represent the protein sequence. Using dipeptide composition (2-gram method) for protein sequence encoding, we can incorporate some important sequence-order information, while the dimension of the feature vectors is still not very high [39], compared to the total 8114 cysteines.

As was previously pointed out before [30–33], a predictor that exploits both local sequence context and global protein features would be relevant for improving the prediction performance. Concerning this topic, to some extent, dipeptide encoding can be considered as a representation combining local with global information in protein sequences.

As seen from Table 2, when the "natural" value 0, of Q_c was simply chosen, we obtained $Q2 = 88.7\%$, $Q_{oxi} = 90.7\%$, $Q_{red} = 81.5\%$, MCC = 68.8%, and when scored on a protein basis (only those protein chains are accepted for which the predictions of all the disulfide-bonding or non-disulfide-bonding states of the cysteines in the protein sequence are correctly predicted), the accuracy rate $Q2_{prot} = 84.9\%$. Moreover, after fine-tuning of Q_c, the prediction score can be slightly improved. Remarkably, if $Q_c = -0.1$ is chosen, then $Q2 = 89.1\%$, $Q2_{prot} = 85.2\%$, $Q_{oxi} = 92.2\%$, $Q_{red} = 79.3\%$, and MCC = 70.7%.

This prediction score can be further improved slightly by avoiding using those dipeptides with the absolute v_{ab} value less than 15 (data not shown). In the case of those dipeptides, to our method and dataset, 15 is the most appropriate v_{ab} value when better accuracy was achieved. The reason that there exist some lower absolute v_{ab} values may possibly owe to the computationally statistical fluctuation.

Even though it is difficult to compare all the existing methods tested on the different databases, it could be claimed that the accuracy achieved by this method tested on a most recently enlarged dataset outperforms or is at least comparable to those previously developed and described methods to predict the disulfide-bonding state of cysteines. The higher prediction scores (89.1% and 85.2%, on the cysteine and protein basis, respectively) obtained by our approach may be possibly due to the interior quality in our method for incorporating both the global information and the local context in protein sequences. The results can demonstrate the applicability and efficiency of this relative simple method.

Combining with other methods and incorporating other features

Several ways may further improve the prediction performance. On the one hand, it should be pointed out that the linear combination of amino acid or dipeptide contents may not be the best function for the purpose of prediction. Although the above results have demonstrated the capability of the simple linear discriminator to effectively discriminate the two cysteine classes (OXICYS and REDCYS), use of more complex functions (for example, the non-linear polynomial functions) can possibly lead to better prediction results than the linear discriminator based classifiers. This aspect is worthy of a deeper investigation.

On the other hand, single prediction methods do have limitations. A possible alternative strategy is to combine other complementary methods, such as neural networks [26,28], combinational logistic functions [30], support vector machines [31,38], and Hidden Markov Models [32,33]. Integration of other different methods incorporating more sequence-order information and evolutionary information together with global features and local sequence context may be likely to further improve prediction performance. Considering the conservation of disulfide bonds and the cysteines in proteins, it is anticipated to combine several methods to use protein primary sequence and three-dimensional structure information and construct the multi-strategy approach to perfect the task of disulfide-bonding state of cysteines.

Conclusion

In this paper, a novel and efficient binary classifier based on the dipeptide composition is presented to reveal the hidden information of the disulfide-bonding states of cysteines. This novel approach provides superior prediction performance compared with the existing methods. Our studies support the phenomena that the oxidation of cysteines exhibits obvious cooperativity and demonstrate that dipeptide contents carry much information about disulfide-bonding. This cooperativity can be well described by protein's dipeptide composition, based on which the prediction accuracy of the oxidation form of cysteines scores as high as 89.1% and 85.2%, when measured on cysteine and protein basis using the rigorous jack-knife procedure, respectively. It indicates that whether cysteines should form disulfide bonds depends not only on the global but also on the local structural features of proteins. This finding indicates that global structural feature of proteins, as well as the local sequence environment of cysteines is important determinant of whether the cysteines should form disulfide bridges. The present study demonstrates the applicability of this new relatively simple method and provides superior prediction performance compared with existing methods for the prediction of the oxidation states of cysteines in proteins. The higher prediction scores obtained by our approach may be possibly due to the interior quality in our method for incorporating both the global information and the local sequence-order information in protein sequences.

Acknowledgments

The authors thank Dr. Gexin Zhang and Yanrui Ding for helpful discussions and valuable comments on the manuscript. This work was supported by the Open Project Program of the Key Laboratory of Industrial Biotechnology, Ministry of Education, China (No. 200102).

References

[1] K.D. Wittrup, Curr. Opin. Biotechnol. 6 (1995) 203–208.
[2] T. Creighton, in: Proteins: Structures and Molecular Properties, W.H. Freeman (Eds.), second ed., New York, 1993.
[3] T. Creighton, Philos. Trans. R. Soc. Lond. B. 348 (1995) 5–10.
[4] S.F. Betz, Protein Sci. 2 (1993) 1551–1558.
[5] J. Skolnick, A. Kolinski, A.R. Ortiz, J. Mol. Biol. 265 (1997) 217–241.
[6] V.I. Abkevich, E.I. Shakhnovich, J. Mol. Biol. 300 (2000) 975–985.
[7] J. Clarke, A.M. Hounslow, C.J. Bond, A.R. Fersht, V. Daggett, Protein Sci. 9 (2000) 2394–2404.
[8] W.J. Wedemeyer, E. Welkler, M. Narayan, H.A. Scheraga, Biochemistry 39 (2000) 4207–4216.
[9] E. Welker, M. Narayan, W.J. Wedemeyer, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 98 (2001) 2312–2316.
[10] P.M. Harrison, M.J.E. Sternberg, J. Mol. Biol. 244 (1994) 448–463.
[11] M.T. Petersen, P.H. Jonson, S.B. Petersen, Protein Eng. 12 (1999) 535–548.
[12] H. Morris, P. Pucci, Biochem. Biophys. Res. Commun. 126 (1985) 1122–1128.
[13] M. Matsumura, B. Matthews, Science 243 (1989) 792–794.
[14] M. Matsumura, B. Matthews, Methods Enzymol. 202 (1991) 336–355.
[15] J. Eder, M. Wilmanns, Biochemistry 31 (1992) 4437–4444.
[16] N.E. Zhou, C.M. Kay, R.S. Hodges, Biochemistry 32 (1993) 178–187.
[17] A. Kremser, I. Rasched, Biochemistry 33 (1994) 13954–13958.
[18] J. Xue, M. Kalafatis, J. Silveira, C. Kung, K.G. Mann, Biochemistry 33 (1994) 13019–13116.
[19] C.C. Chuang, C.Y. Chen, J.M. Yang, P.C. Lyu, J.K. Hwang, Proteins Struct. Funct. Genet. 53 (2004) 1–5.
[20] H.W. van Vlijmen, A. Gupta, L.S. Narasimhan, J. Singh, J. Mol. Biol. 335 (2004) 1083–1092.
[21] E.S. Huang, R. Samudrala, J.W. Ponder, J. Mol. Biol. 290 (1999) 267–281.
[22] P. Fariselli, R. Casadio, Bioinformatics 17 (2001) 957–964.
[23] P. Fariselli, P.L. Martelli, R. Casadio, A neural network-based method for predicting the disulfide connectivity in proteins, in: Knowledge Based Intelligent Information Engineering Systems and Allied Technologies (KES 2002), vol .1, IOS Press, Amsterdam, 2002, pp. 464–468.
[24] A. Vullo, A. Passerini, Bioinformatics 20 (2004) 653–659.
[25] S.M. Muskal, S.R. Holbrook, S.H. Kim, Protein Eng. 3 (1990) 667–672.
[26] P. Fariselli, P. Riccobelli, R. Casadio, Proteins Struct. Funct. Genet. 36 (1999) 340–346.
[27] A. Fiser, M. Cserzo, E. Tudos, I. Simon, FEBS Lett. 302 (1992) 117–120.
[28] A. Fiser, I. Simon, Bioinformatics 16 (2000) 251–256.
[29] A. Fiser, I. Simon, Methods Enzymol. 353 (2002) 10–21.
[30] M.H. Mucchielli-Giorgi, S. Hazout, P. Tuffery, Proteins Struct. Funct. Genet. 46 (2002) 243–249.
[31] A. Ceroni, P. Frasconi, A. Passerini, A. Vullo, J. VLSI Signal Process. 35 (2003) 287–295.
[32] P.L. Martelli, P. Fariselli, L. Malaguti, R. Casadio, Protein Eng. 15 (2002) 951–953.
[33] P.L. Martelli, P. Fariselli, L. Malaguti, R. Casadio, Protein Sci. 11 (2002) 2735–2739.
[34] G. Wang, R.L. Dunbrack Jr., Bioinformatics 19 (2003) 1589–1591.
[35] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[36] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979, p. 322, 381.
[37] B.W. Matthews, Biophys. Acta 405 (1975) 442–451.
[38] P. Frasconi, A. Passerini, A. Vullo, A two stage SVM architecture for predicting the disulfide bonding state of cysteines, in: Proceedings of IEEE Neural Network for signal processing conference, IEEE Press, New York, 2002.
[39] Y. Huang, Y. Li, Bioinformatics 20 (2004) 21–28.