PROTEINS: Structure, Function, and Bioinformatics 67:255–261 (2007)

SHORT COMMUNICATION

Analycys: A Database for Conservation and Conformation of Disulphide Bonds in Homologous Protein Domains

Ratna R. Thangudu,[1] Priyanka Sharma,[1] N. Srinivasan,[2] and Bernard Offmann[1]*
[1]*Laboratoire de Biochimie et Génétique Moléculaire, Université de La Réunion, BP 7151, 15 Avenue René Cassin, 97715 Saint Denis Messag Cedex 09, La Réunion, France*
[2]*Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka 560 012, India*

ABSTRACT **Disulphide bonds in proteins are known to play diverse roles ranging from folding to structure to function. Thorough knowledge of the conservation status and structural state of the disulphide bonds will help in understanding of the differences in homologous proteins. Here we present a database for the analysis of conservation and conformation of disulphide bonds in SCOP structural families. This database has a wide range of applications including mapping of disulphide bond mutation patterns, identification of disulphide bonds important for folding and stabilization, modeling of protein tertiary structures and in protein engineering. The database can be accessed at: http://bioinformatics.univ-reunion.fr/analycys/. Proteins 2007;67:255–261.** © 2007 Wiley-Liss, Inc.

Key words: disulphide bonds; conservation; mutation; homologous proteins

INTRODUCTION

Protein sequencing is complete only when the information about post-translational modifications is fully annotated. Cysteines in proteins assume a very important role in structure and function because of their free thiol group and ability to covalently bond with another cysteine to form disulphide bonds. Among others, disulphide bond is a prominent post-translational covalent modification. Often cysteine is found in the active site of proteins.[1] In addition, the covalent state of Cys provides information about the possible cellular location of proteins. Many small proteins simply cannot form stable native structures without disulphide bonds. Information on the covalent status of cysteine residues can also greatly enhance the performance of comparative and ab initio modeling studies.[2] Because of the importance attached, cysteines with free thiol group participating in catalytic or regulatory activities are conserved among homologous proteins. In addition, disulphide bonds are believed to be strongly conserved in homologous proteins.

Several methods have been developed to identify the oxidation state of the cysteines in protein sequences[3–16] and also to identify their connectivity patterns[17–25] in multiple disulphide bonds containing proteins. Tools are available to model disulphide bonds into proteins by estimating the local stereochemical compatibility to accommodate a disulphide bond.[26,27] Often such a stereochemical analysis of protein structure helps in highlighting the spatially close free thiol groups that are forced to form disulphide bonds upon oxidation *in vitro*. DSDBASE is one such database which lists all the native and modeled disulphide bonds in all the known protein structures.[28] Many of the existing methods[29–31] are strongly dependent on the conservation of disulphide bonds in order to find similarities between distant homologues based on the connectivity patterns. However disulphide bond mutations (usually associated with loss of both the cysteines) are not uncommon in homologous proteins. Hence it is necessary to understand the conservation of disulphide bonds in protein structural families. In addition not all the disulphide bonds in a protein may be equally important and they could have different roles. Presence of the disulphide bond in all the members of the family (with varying degrees of sequence identity and source organisms) might indicate their importance in folding. A disulphide bond of poor conservation might have a protein specific function such as imparting additional stability or a functional role. Classifying disulphide bonds based on their conservation will help in identifying the most important among them and thus in protein engineering and folding studies.

Several folding studies with reduction or mutation of disulphide bonds clearly indicated that these reducible disulphides does not radically change the structure or

*Correspondence to: Bernard Offmann, Laboratoire de Biochimie et Génétique Moléculaire, Université de La Réunion, BP 7151, 15 Avenue René Cassin, 97715 Saint Denis, Messag Cedex 09, La Réunion, France. E-mail: bernard.offmann@univ-reunion.fr

Received 6 March 2006; Revised 12 September 2006; Accepted 1 November 2006

Published online 6 February 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21318

function but do contribute to the stability of the protein.[29–31] Absence of a cross link in a member of a family does not always change the basic structure of that protein. Additional disulphide bonds (Cys22-Cys157, Cys128-Cys232) found in trypsin can be fitted easily into elastase and α-chymotrypsin models with almost no alteration of their structures.[32] The non-conservation of disulphide bonds thus appears to contradict the conventional wisdom that these cross-links are important to the protein folding process.[33]

Non-conservation of disulphide bond in homologous proteins thus raise several questions concerning the complementary stabilizing interactions, alteration in the local structure, and their global implications like function and stability. The understanding of evolutionary dynamics of disulphide bonds will bring important insights into the structural divergence of homologues. The present study thus aimed at providing a convenient platform to address how the mutations of disulphide bonds affect the structure of homologous proteins which is not being addressed by other existing platforms like DSDBase[28] that only focus on listing all the stereochemically compatible residue pairs for the introduction of disulphide bond in a given protein. Analycys database, on the other hand, is a rigorous tool to analyze the stereochemistry, topological equivalency, solvent accessibility, and residue depth of both the disulphide bonds and their substitutions in case of mutation, in related proteins. Our database is a rich source of information that helps in addressing these questions and serves as a handy tool for protein engineering and folding studies and also for structure explorers.

MATERIAL AND METHODS

Dataset

Release 1.67 of SCOP[34] containing 2630 families with over 65,122 domains is used for our analysis. Disulphide bond information of the domains is extracted from the SSBOND record of their original PDB files. Only intra-domain disulphide bonds are considered in the current study.

Of these, 580 families have at least one member with disulphide bond. Ignoring multidomain and membrane classes resulted in 557 families with at least one member with a disulphide bond. Filtering criteria such as 95% sequence identity cut off and a resolution better than 2.5 Å resulted in 300 families with at least two members per family. For NMR-derived protein structures without any crystallographically determined structures, only the first model in the ensemble of their NMR structure are considered for superimposition.

Structural Superimposition

Structure alignments have been performed using STAMP (v. 4.2) package of programs.[35] STAMP aligns the set of homologous protein structures and generates a structure based multiple sequence alignment among others. Topologically equivalent regions correspond to regions of conserved secondary structure within a family of structures compared. Genuine topological equivalencies from multiple structural alignments of homologous proteins are derived by a method of Rossman and Argos[36] implemented internally in the structural alignment program by calculating the probability of residue structural equivalencies.

Distinct Disulphide Bonds in a Family

For every given family, the positions of disulphide bonds in homologous members are marked on the block of multiple structure-based sequence alignment. The distinct positions of these disulphide bonds in the multiple alignment block are counted as distinct disulphide bonds in the family [Fig. 1(a)]. When there is no disulphide bond found in a member at a position, a relaxation has been applied by allowing an arbitrary shift of 4 residues on either side of the cysteine positions for counting the distinct disulphide bonds. This way all the distinct disulphide bonds in a given family are identified and their conservation is estimated for each member and also for the whole family.

Grading of Conservation

To understand the evolution of disulphide bonds in homologous proteins, their conservation is quantified and graded. A highly conserved disulphide bond is present in more than 70% of the family members while medium and low conserved disulphide bonds are present in 30–70% and $<30\%$ of the members of the family respectively.

Backbone Conformation, Solvent Accessibility, and Residue Depth

The backbone conformation and secondary structure assignment of half-cysteine are derived using the program, DSSP.[37] However, a three state assignment is used instead of the default seven states. All the helix types (H, G, I), strand types (E, B) and turns (S, T) are grouped in to H, E, and C respectively.

Solvent accessible surface area is calculated using the method of Lee and Richards[38] as implemented in the program NACCESS.[39] The relative accessibility of the half-cysteine is its percentage solvent accessibility compared to an Ala-Cys-Ala tripeptide in extended conformation.

Residue depth is calculated by using a program DPX.[40] This program gives the depth of each atom in protein, defined as distance (in Å) from the closest solvent accessible atom and calculates the mean residue depth. These depth values are critically dependent on local parameters like protein size, shape and structure. However it is important to understand whether there exists a correlation among solvent accessibility, conservation and depth of half-cysteines. Hence the half-cysteine in each protein are classified as core, intermediate or surface based on the residue depth values of residues in the protein.

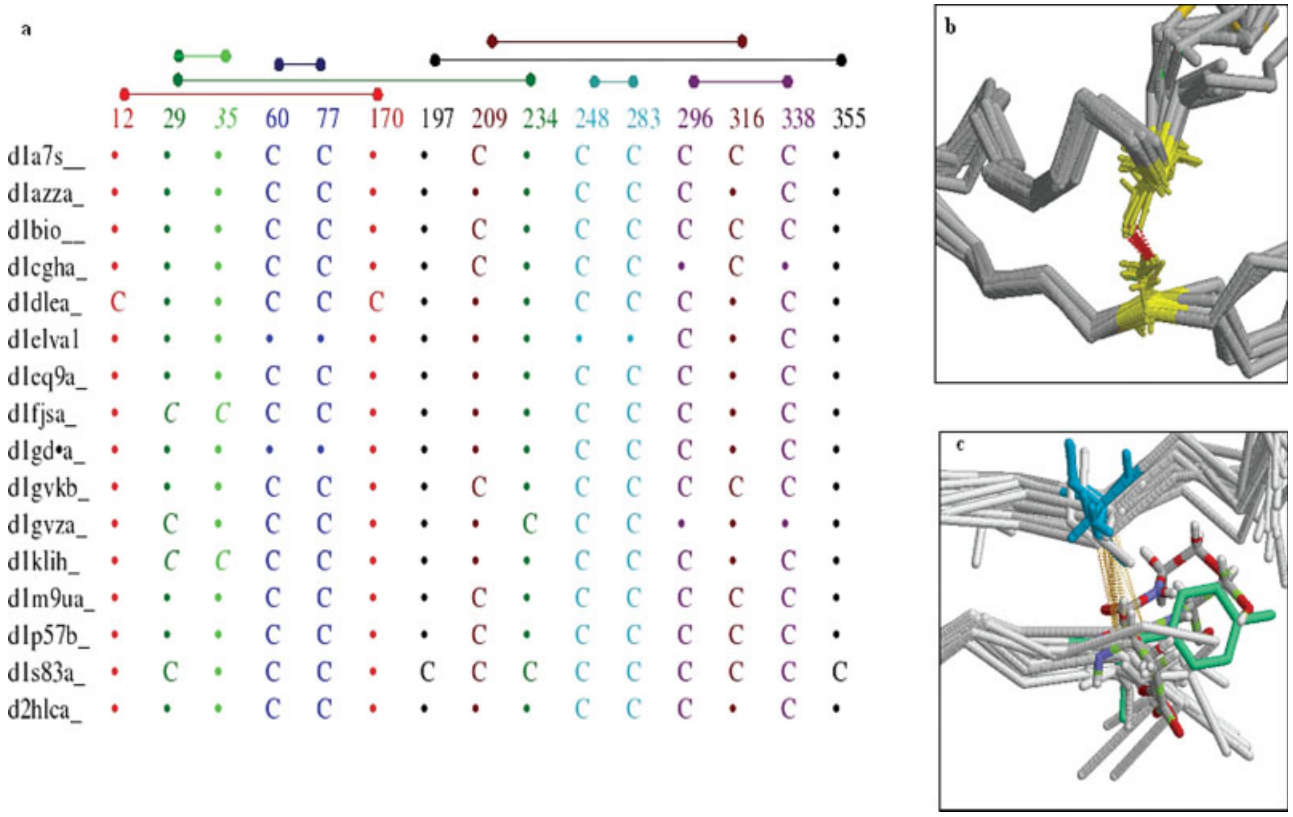

Fig. 1. Illustration of distinct (conserved and non-conserved) disulphide bonds in trypsin-like eukaryotic serine proteases (SCOP code: b.47.1.2). (**a**) Numbers indicate the approximate positions of cysteines in the structure-based sequence alignment. When a cysteine is not conserved, it is represented as a dot. Disulphide bonded cysteines are appropriately connected with their counterparts. Each connection is a distinct disulphide bond in the family irrespective of its level of conservation. (**b**) A highly conserved disulphide bond (60-77 in the alignment), which is present in all the members. Disulphide bond is represented as red dotted lines and cysteines are in yellow. (**c**) A medium conserved disulphide bond (209-316 in the alignment). The absence of disulphide bonds in other members does not bring about any drastic conformational changes locally. See the sidechains of the equivalent residues for cysteines in the members without this disulphide bond.

RESULTS

Currently, a total of 300 families are present in the Analycys database with at least two members per family and at least one of them annotated with a disulphide bond. The whole analysis is carried out on protein domains, the structural, functional and evolutionary units of proteins. Inter-chain and inter-domain disulphide bonds are not included in our database, since these are often problematic in large-scale analysis of proteins. Additionally, inter-domain and inter-subunit disulphide bonds, although important, are rare and are unlikely to change the course of our current analysis. The distribution of disulphide bonds in different fold classes is depicted in Table I. In about 180 families, there is not enough structural information to analyze the conformation and conservation of the disulphide bonds. Disulphide bonds are common in all fold classes although in varying degrees. However only 22% of the families in SCOP have disulphide bonds. A closer look at these families might give clues about their preference. Small proteins with fewer than 70 residues are usually rich in disulphide bonds. Disulphide bonds in such proteins perhaps improve their stability,[41] since the hydrophobic forces alone are not sufficient.[42] Anecdotal evidence suggests that they are present in most small proteins.[43,44] A vast majority (67%) of the small proteins in SCOP 1.67 have disulphide bonds (Table I).

TABLE I. Fold Distribution of Disulphide Bond Containing Families Under Study in SCOP

	Number of disulphide bond containing families		Number of members per family in ANALYCYS		
SCOP Class[a]	SCOP	ANALYCYS	2	3–10	>10
All α (550)	65	38	12	17	9
All β (529)	167	89	19	46	24
α/β (593)	111	58	8	30	20
α + β (650)	106	59	16	30	13
Small proteins (162)	108	56	9	35	12
Total (2630)	580	300	64	158	78

[a]Total number of families present in the respective SCOP class in parentheses.

On the contrary all α proteins showed least preference for these cross-links, probably because of the steric hindrances. The repertoire of cysteine conformations suggests

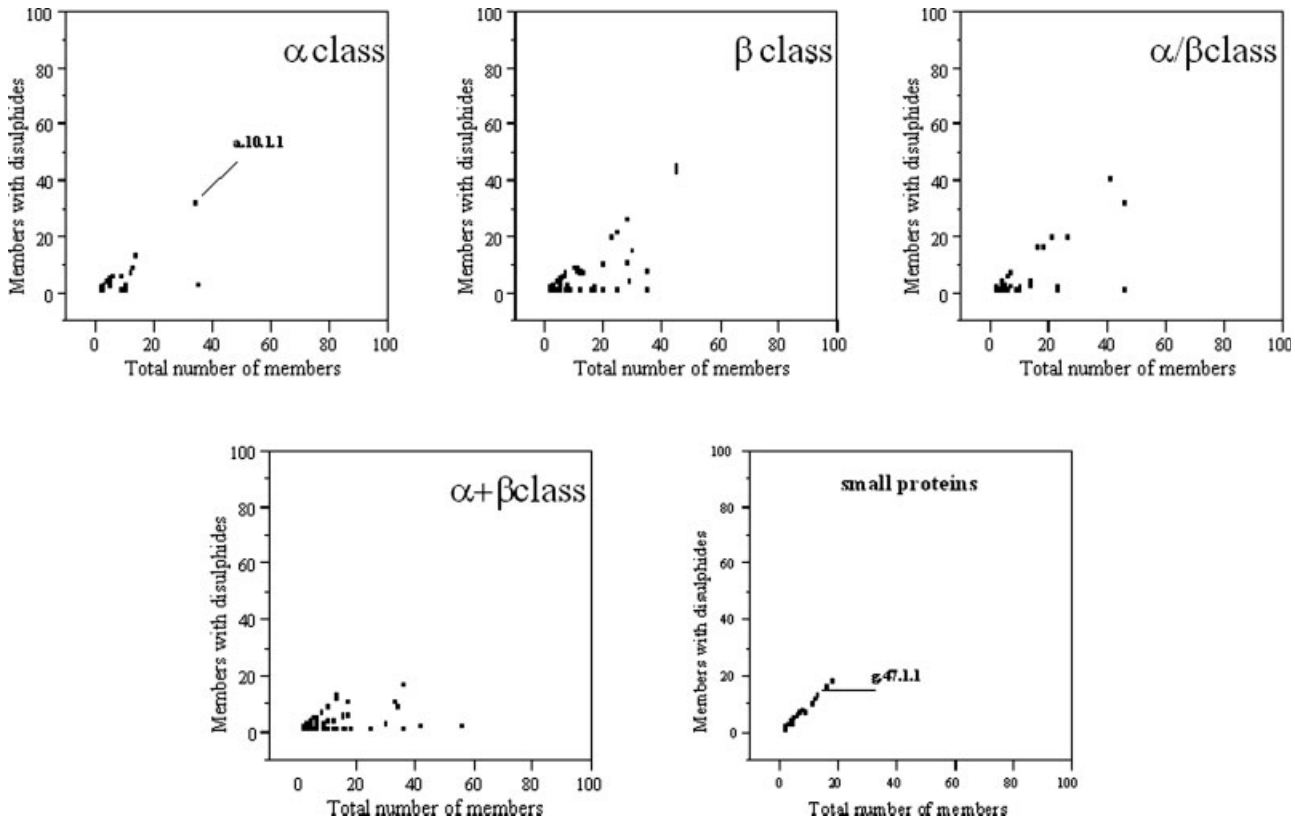

Fig. 2. Distribution of total number of members in each family against total number of members with disulphides. Each dot represents a family and some are labeled for the purpose of illustration.

preferences in disulphide distribution between/within beta sheets.[45] The secondary structural distribution of disulphide bridges highlighted their preference for stabilizing strands and loops (Thangudu et al., submitted). The evolutionary plasticity of loop regions is greater than that of the protein core[46] and this explains the greater cross link frequency in loop regions of homologous proteins.

Mutations of Disulphide Bonds During Evolution are Widespread and are Independent of Fold Class

A simple estimate of proteins with and without disulphide bonds in related proteins (members of same family) which share common fold and function appears to contradict the conventional wisdom that disulphide bridges are well conserved during evolution. Except for small proteins where they play a major role in their fold stabilization, modest conservation of disulphide bonds appears to be common feature in all the other fold classes (Fig. 2).

Relaxation Brought Better Conservation

A disulphide bond is considered strictly conserved in a multiple alignment if both the cysteines involved are placed at identical alignment positions in all the sequences in the alignment. However alignment of structures with the goal of creating an optimal multiple alignment might result in alignments between any given pair of structures with slightly displaced cystine residues.[47] This is due to either inherent structural differences or high sequence divergence in the homologous family members.[47] However such differences do not always change the function. Structural differences in homologous proteins are usually concentrated in loop regions and most often disulphide bonds, when present stabilize such regions. In order to account for such structural differences, a relaxation has been applied for calculating the conservation. A total of 34,752 pair-wise comparisons are possible within the members of 300 families. The number of conserved disulphide bonds in these homologous pairs rose from 19,500 to 26065 when a relaxation is applied. This increase suggests the structural variation in homologous proteins.

In all, only 54% of all the disulphide bonds compared between the homologous pairs are conserved. Hence conservation of disulphide bonds in homologous proteins is not a rule and non-conservation is more common than expected. This observation bolsters the notion that all disulphide bonds are not equally important and a non-conserved disulphide bond might have functions other than folding and stabilization.

Distribution of High, Medium, and Low Conserved Disulphide Bonds in SCOP Classes

The distribution of high, medium and low conserved disulphide bonds, in different fold classes, clearly estab-

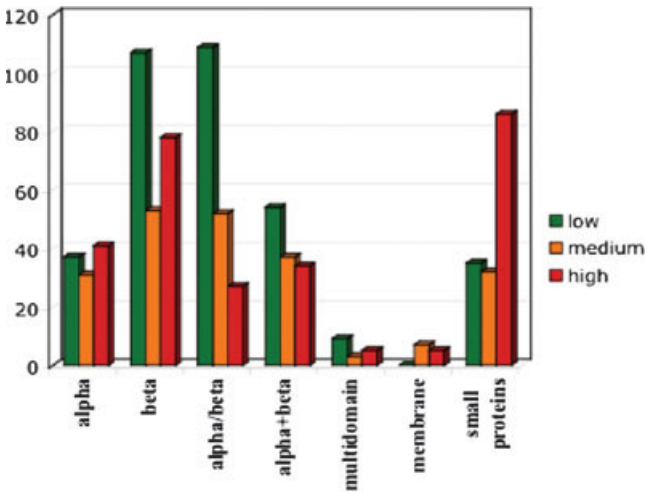

Fig. 3. Distribution of high, medium, and low conserved disulphide bonds in SCOP classes. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

lishes known but interesting facts. Figure 3 clearly depicts the state of the disulphide bond conservation in homologous proteins. Majority of the disulphide bonds in small proteins are highly conserved re-establishing the fact that disulphide bonds play a strong role in their structural stabilization. Surprisingly in other fold classes, almost 50% of the disulphide bonds are poorly conserved. Mutations of disulphide bond are widespread across all the fold classes. The current study thus helps in quickly identifying the disulphide bonds that might play a crucial role in protein folding.

DISCUSSION

We note that non-conservation of disulphide bonds is not uncommon in homologous proteins. The variable conservation of these cross-links suggests not all of them are equally important in a family of proteins. Besides playing a prominent role in folding and stabilization of the fold, disulphide bonds have protein specific roles. Identification of conserved and non-conserved disulphide bonds greatly helps in understanding the mechanism of folding, protein engineering and stabilization studies, etc.

Analycys is a database for analyzing conservation and conformation of disulphide bonds in homologous proteins. The database can be browsed or searched and organized as SCOP structural families. A table displaying the status (presence or absence) of distinct disulphide bonds in each member of the family forms the major feature of the database and serves as a disulphide mutation map of the family. Disulphide bond mutation mapping of homologous proteins help in refining distant homology prediction methods based on disulphide bond connectivity patterns by providing alternate patterns.

Knowing that members of a family usually share a well-conserved fold, it would be interesting to see the state of the distinct disulphide bonds in terms of topological equivalency and the solvent accessibility. Ideally one would expect a highly conserved disulphide bonds share a highly conserved environment [Fig. 1(b)]. However often disulphide bonds occur in flexible regions like loops. For a better understanding of disulphide bond environment in homologous members, these details are highlighted in easily readable 2D format including case and color coding of the cystines.

When a disulphide is poorly conserved in homologous proteins, the interactions among the substituted residues evoke a certain interest [Fig. 1(c)]. Stability imparted by such a disulphide bond is compounded by other local structural adjustments associated with sequence differences (Thangudu et al., submitted). The structural variation in the immediate vicinity of the mutated disulphide bond is not obvious and does not reflect on the overall structure of the protein.[32,33] Here we provide, in the database, the topological status and solvent accessibility of the substituted residues.

The database also lists the bridge stereochemical parameters (conveniently graded according to MODIP criteria[27]) for all the disulphide bonds and allows the user for rigorous assessment of the local environment.

Main Features of the Database

- Search and browse, according to SCOP classification
- Disulphide bond conservation status for a family
- Structural state of the disulphide bonds and the substituted residues (when not conserved) in terms of solvent accessibility and topological state
- A thorough analysis of bridge stereochemistry
- For single member families with not enough structural information, the structural domains are linked to their respective PFAM[48] domains, which might help potential users to evaluate sequence conservation of disulphide bonds.

Additional Features

- A structure based sequence alignment file is available for download in text format or can be viewed online as JOY[49] annotated.
- Taxonomic position of the family members is also provided to understand the role of evolution on the conservation of disulphide bonds.
- A sequence similarity matrix (generated through ClustalW[50]) for every family is provided to correlate sequence and structure conservation and their effect on disulphide bonds.

CONCLUSIONS

In this paper, we describe Analycys, a database for conservation and conformation of disulphide bonds in homologous proteins and highlight its usefulness. The database is a rich source of physicochemical information about three-dimensional environment of disulphide bonds. Such a database comes as a handy tool for protein folding, protein engineering, and modeling studies and should be of

value to biochemists and biologists. The database is located at http://bioinformatics.univ-reunion.fr/analycys/.

ACKNOWLEDGMENTS

T.R.R. is supported by a PhD grant from the Conseil Régional de La Réunion. N.S. is an International Senior Fellow of the Wellcome Trust, UK. He also thanks the authorities of Université de La Réunion for the visiting professorship. We thank Drs. Alexandre G. de Brevern and R. Sowdhamini for useful comments.

REFERENCES

1. Fiser A, Simon I, Barton GJ. Conservation of amino acids in multiple alignments: aspartic acid has unexpected conservation. FEBS Lett 1996;397:225–229.
2. Simon I, Glasser L,Scheraga HA. Calculation of protein conformation as an assembly of stable overlapping segments: application to bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 1991;88:3661–3665.
3. Muskal SM, Holbrook SR, Kim SH. Prediction of the disulfide-bonding state of cysteine in proteins. Protein Eng 1990;3:667–672.
4. Fiser A, Cserzo M, Tudos E, Simon I. Different sequence environments of cysteines and half cystines in proteins. Application to predict disulfide forming residues. FEBS Lett 1992;302:117–120.
5. Fariselli P, Riccobelli P, Casadio R. Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. Proteins 1999;36:340–346.
6. Fiser A, Simon I. Predicting redox state of cysteines in proteins. Methods Enzymol 2002;353:10–21.
7. Frasconi P, Passerini A, Vullo A. A two-stage SVM architecture for predicting the disulfide bonding state of cysteines. Proceedings of the IEEE NNSP International Workshop: special session on signal processing and neural networks for bioinformatics. Martigny, Switzerland, Sep 4–6, 2002.
8. Martelli PL, Fariselli P, Malaguti L, Casadio R. Prediction of the disulfide bonding state of cysteines in proteins with hidden neural networks. Protein Eng 2002;15:951–953.
9. Martelli PL, Fariselli P, Malaguti L, Casadio R. Prediction of the disulfide-bonding state of cysteines in proteins at 88% accuracy. Protein Sci 2002;11:2735–2739.
10. Mucchielli-Giorgi MH, Hazout S, Tuffery P. Predicting the disulfide bonding state of cysteines using protein descriptors. Proteins 2002;46:243–249.
11. Ceroni A, Frasconi P, Passerini A, Vullo A. Predicting the disulfide bonding state of cysteines with combinations of kernel machines. J VLSI Signal Process Syst Signal Image Video Technol 2003;35:287–295.
12. Dosztanyi Z, Magyar C, Tusnady GE, Cserzo M, Fiser A, Simon I. Servers for sequence-structure relationship analysis and prediction. Nucleic Acids Res 2003;31:3359–3363.
13. Chen YC, Lin YS, Lin CJ, Hwang JK. Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. Proteins 2004;55:1036–1042.
14. Martelli PL, Fariselli P, Casadio R. Prediction of disulfide-bonded cysteines in proteomes with a hidden neural network. Proteomics 2004;4:1665–1671.
15. Passerini A, Frasconi P. Learning to discriminate between ligand-bound and disulfide-bound cysteines. Protein Eng Des Sel 2004;17:367–373.
16. Song JN, Wang ML, Li WJ, Xu WB. Prediction of the disulfide-bonding state of cysteines in proteins based on dipeptide composition. Biochem Biophys Res Commun 2004;318:142–147.
17. Fariselli P, Casadio R. Prediction of disulfide connectivity in proteins. Bioinformatics 2001;17:957–964.
18. Vullo A, Frasconi P. A recursive connectionist approach for predicting disulfide connectivity in proteins. Proc. 18th ACM Symposium on Applied Computing (SAC 2003), PP. 66–71, 2003.
19. Baldi P, Cheng J, Vullo A. Large scale prediction of disulphide bond connectivity. Adv Neural Inf Process Syst 2004;17:97–104.
20. Vullo A, Frasconi P. Disulfide connectivity prediction using recursive neural networks and evolutionary information. Bioinformatics 2004;20:653–659.
21. Chen YC, Hwang JK. Prediction of disulfide connectivity from protein sequences. Proteins 2005;61:507–512.
22. Ferre F, Clote P. Disulfide connectivity prediction using secondary structure information and diresidue frequencies. Bioinformatics 2005;21:2336–2346.
23. Ferre F, Clote P. DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 2005;33:W230–W232. Web server issue.
24. Tsai CH, Chen BJ, Chan CH, Liu HL, Kao CY. Improving disulfide connectivity prediction with sequential distance between oxidized cysteines. Bioinformatics 2005;21:4416–4419.
25. Zhao E, Liu HL, Tsai CH, Tsai HK, Chan CH, Kao CY. Cysteine separations profiles on protein sequences infer disulfide connectivity. Bioinformatics 2005;21:1415–1420.
26. Hazes B, Dijkstra BW. Model building of disulfide bonds in proteins with known three-dimensional structure. Protein Eng 1988;2:119–125.
27. Sowdhamini R, Srinivasan N, Shoichet B, Santi D, Ramakrishnan C, Balaram P. Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. Protein Eng 1989;3:95–103.
28. Vinayagam A, Pugalenthi G, Rajesh R, Sowdhamini R. DSDBASE: a consortium of native and modelled disulphide bonds in proteins. Nucl Acids Res 2004;32(90001):D200–D202.
29. Chuang CC, Chen CY, Yang JM, Lyu PC, Hwang JK. Relationship between protein structures and disulfide-bonding patterns. Proteins 2003;53:1–5.
30. Mas JM, Aloy P, Marti-Renom MA, Oliva B, Blanco-Aparicio C, Molina MA, de Llorens R, Querol E, Aviles FX. Protein similarities beyond disulphide bridge topology. J Mol Biol 1998;284:541–548.
31. van Vlijmen HWT, Gupta A, Narasimhan LS, Singh J. A novel database of disulfide patterns and its application to the discovery of distantly related homologs. J Mol Biol 2004;335:1083–1092.
32. Thornton JM. Disulphide bridges in globular proteins. J Mol Biol 1981;151:261–287.
33. Kreisberg R, Buchner V, Arad D. Paired natural cysteine mutation mapping: aid to constraining models of protein tertiary structure. Protein Sci 1995;4:2405–2410.
34. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
35. Russell RB, Barton GJ. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. Proteins 1992;14:309–323.
36. Rossmann MG, Argos P. Exploring structural homology of proteins. J Mol Biol 1976;105:75–95.
37. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
38. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
39. Hubbard SJ, Thornton JM. Naccess V2.1.1, atomic solvent accessible area calculations. Department of Biochemistry and Molecular Biology, University College, London; 1993.
40. Pintar A, Carugo O, Pongor S. DPX: for the analysis of the protein core. Bioinformatics 2003;19:313–314.
41. Creighton TE. Disulphide bonds and protein stability. Bioessays 1988;8:57–63.
42. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry 1985;24:1501–1509.
43. Wetlaufer DB. Folding of protein fragments. Adv Protein Chem 1981;34:61–92.
44. Richardson JS, Richardson DC. The de novo design of protein structures. Trends Biochem Sci 1989;14:304–309.

45. Harrison PM, Sternberg MJE. The disulphide [β]-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. J Mol Biol 1996;264:603–623.
46. Panchenko AR, Wolf YI, Panchenko LA, Madej T. Evolutionary plasticity of protein families: coupling between sequence and structure variation. Proteins 2005;61:535–544.
47. Scheeff ED, Bourne PE. Structural evolution of the protein kinase-like superfamily. PLoS Comput Biol 2005;1:e49.
48. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32:D138–D141. Database issue.
49. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP. JOY: protein sequence-structure representation and analysis. Bioinformatics 1998;14:617–623.
50. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.