Predicting the Oxidation State of Cysteines by Multiple Sequence Alignment

Article *in* Bioinformatics · April 2000

DOI: 10.1093/bioinformatics/16.3.251 · Source: PubMed

2 authors, including:

István Simon
Hungarian Academy of Sciences
17 PUBLICATIONS **1,387** CITATIONS

SEE PROFILE

BIOINFORMATICS

Vol. 16 no. 3 2000
Pages 251–256

Predicting the oxidation state of cysteines by multiple sequence alignment

András Fiser [1,2,*] and István Simon [1]

[1] Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest H-1518, PO Box 7, Hungary

Received on April 21, 1999; revised on September 27, 1999; accepted on October 4, 1999

Abstract

Motivation: *Protein sequences found in databanks usually do not report post translational covalent modifications such as the oxidation state of cystein (Cys) residues. Accurate prediction of whether a functionally or structurally important Cys occurs in the oxidized or thiol form would be helpful for molecular biology experiments and structure prediction.*

Results: *A new method is presented for predicting the oxidation state of Cys residues based on multiple sequence alignments and on the observation that Cys tends to occur in the same oxidation state within the same protein. The prediction of the redox state of Cys performs above 82%. The oxidation state of Cys correlates with the cellular location of the given protein within the cell, but the correlation is not perfect (up to 70%). We also perform a statistical analysis of the different redox states of Cys found in secondary structures and buried positions, and of the secondary structures linked by disulfide bonds. The results suggest that the natural borderline lies between the different oxidation states of Cys rather than between the half cystines and cysteins.*

Availability: *A web server implementing the prediction method is available at http://guitar.rockefeller.edu/~andras/cyspred.html*

Contact: *fisera@rockefeller.edu*

Introduction

The oxidation state of cysteine plays an important role in protein structure and function. In its thiol form, cysteine is the most reactive amino acid under physiological conditions, and is often used for adding fluorescent groups and spin labels (Creighton, 1993). In the oxidized forms, cysteines form disulfide bonds, which are the primary covalent cross links found in proteins and which stabilize the native conformation of a protein. Thus accurate predictions of the oxidation state of cysteines would have numerous applications, for example in engineering when stabilizing cystines or reactive thiol groups (Clarke and Fersht, 1993; Zhou *et al.*, 1993; Eder and Wilmanns, 1992; Matsumura and Matthews, 1991), in locating key reactive thiol groups in enzymatic reactions (Ritov, 1995) or in determining topologies to aid three-dimensional structure predictions (Simon *et al.*, 1991).

Due to the efficiency of DNA sequencing techniques and the worldwide genome projects, the sequence databases are approximately 100-fold larger than the three-dimensional ones and the gap is growing rapidly. Statistics on recently sequenced organisms yield 22–56% of putative reading frames with unknown function (Cedano *et al.*, 1997). Unfortunately, the sequence databases do not report post-translational covalent modifications such as the oxidation of cysteines forming disulfide bridges. If the protein is physically available disulfide bridges can be revealed by time-consuming experimental methods (Kremser and Rasched, 1994; Xue *et al.*, 1994; Morris and Pucci, 1985; Tannhauser *et al.*, 1984; Hartley *et al.*, 1965).

There are two published methods predicting disulfide bond forming Cys residues (Muskal *et al.*, 1990; Fiser *et al.*, 1992). Both methods used sequence information alone hidden in the specific sequence environment of cysteines and half cystines. One method employed a neural network (NN) to recognize disulfide-bonded Cys while the other method performed a statistical analysis of the amino-acid frequencies in the sequence environment of Cys. The NN prediction, tested on a smaller independent dataset, achieved 81% accuracy, whereas the statistical method (which used a four-fold bigger dataset) performed at 71% prediction accuracy as tested by the jack-knife procedure.

In this paper we assess the differences between the statistical frequencies of oxidized and reduced cysteines on the surface and in the interior of proteins, as well as the frequencies of different oxidation states of Cys in secondary structural elements. We also analyze the types of secondary structural elements linked by a disulfide

*To whom correspondence should be addressed at The Rockefeller University.
[2] Laboratories of Molecular Biophysics, The Rockefeller University, Box 270, 1230 York Avenue, New York, NY 10021-6399, USA.

bridge, and the correlation between the cellular location of a protein and the oxidation state of its cysteines.

A method is introduced to predict the covalent state of Cys from multiple sequence alignments. The results demonstrate that the analysis of multiple sequence alignments is an efficient tool to distinguish oxidized cysteines from those with reactive sulfhydryl groups.

Materials and methods

A database of 81 protein alignments was used in this analysis. The protein structures were selected from the Protein Data Bank (PDB) (Abola *et al.*, 1987) by a two-step procedure. First, the sequences of all PDB proteins longer than 50 residues and having a crystallographic resolution better than 2.5 Å were compared by calculating the correlation coefficient of dipeptide frequencies. A set of 101 proteins remained after requiring that any pair should have a dipeptide frequency correlation smaller than 0.4. In the second step, every pair of proteins in the filtered set was compared by a rigorous sequence comparison method (Smith and Waterman, 1981; Barton, 1993) followed by cluster analysis, which yielded the final 81 proteins. The four-letter PDB codes and chain identifiers are as follows: 155C, 1acx, 1alc, 1bbpA, 1cc5, 1eca, 1fkf, 1fnf, 1fnr, 1gcr, 1gp1A, 1hdsB, 1hip, 1hoe, 1lrd4, 1paz, 1pcy, 1phh, 1prcC, 1rbp, 1rhd, 1rnh, 1sn3, 1tgs, 1tpkA, 1wsyB, 256bA, 2alp, 2azaA, 2cab, 2cd4, 2cdv, 2cpp, 2fxb, 2gn5, 2lh7, 2liv, 2ltnA, 2or1L, 2pabA, 2rnt, 2rspA, 2secI, 2snlE, 2sns, 2sodB, 2ssi, 2stv, 2ts1, 2utgA, 3adk, 3b5c, 3cla, 3fxc, 3gapB, 3lzm, 3sgbI, 451C, 4bp2, 4fd1, 4fxn, 4hhbA, 4pep, 4pfk, 4ptp, 4tnc, 5cts, 5cytR, 5ebx, 5rubA, 5rxn, 6ldh, 6tmnE, 7pti, 8adh, 8atcB, 8catA, 8dfr, 9pap, 9rsaA and 9wgaA. Each of these sequences was compared with the PIR database (George *et al.*, 1986) by the program SCANPS (http://www2.ebi.ac.uk/scanps/). Sequences that gave a probability lower than 10^{-6} were used to produce the multiple sequence alignments by the method of Barton and Sternberg (1987). The number of sequences in the alignment varied between 3 and 499 with a median of 28.

Accessible surface areas were calculated by the program DSSP (Kabsch and Sander, 1983) and converted to relative accessibilities by dividing by the accessibility of the residue in a Gly-X-Gly tripeptide (Rose *et al.*, 1985). Two relative accessibility classes were considered: buried ($A \leq 0.25$) and exposed ($A > 0.25$).

Conservation scores based upon the physico-chemical properties of the amino acids were calculated for each position in each alignment according to Livingstone and Barton (1993). Such conservation scores range from 0 to 10 and count the number of the properties shared at a position, where the properties are: Hydrophobic, Positive, Negative, Polar, Charged, Small, Tiny, Aliphatic, Aromatic, Proline and their negation (e.g. *not* Hydrophobic). For each position in each protein this score was then divided by the average conservation of the protein to give a relative conservation score C_r. We refer to a position as 'conserved' if $C_r > 1$, i.e. the conservation of the given position is higher than the average conservation of the sequence.

Results

Conservation, occurrence and distribution of Cys residues

The average relative conservation score (C_r) and average relative accessibility are shown for the 20 residues (Figure 1). Cys is by far the most conserved residue both on the surface and in the interior, which reflects its crucial role in structure stabilization and biochemical functions. Figure 2 shows the distribution and occurrence of residues: Cys is the second rarest residue on the surface (after Trp), presumably because it has the most reactive side chain (Creighton, 1993). The distribution of half cystines and cysteines between the surface and the interior is almost identical (84% and 80% are buried, respectively). If we group liganded cysteines and half cystines as a new category of 'bonded' cysteines, this result is slightly modified: 79% and 80% of the bonded Cys and 'free' (i.e. non-bonded) cysteines are buried, respectively. The difference in the relative conservation score between the bonded and free forms of Cys is significant, the standard deviations are about the difference, bonded: 1.56 ($\delta = 0.53$) and free: 1.13 ($\delta = 0.44$). Of the protein set, 51% contained only free cysteines, 27% only half cystines, 5% both forms of cysteine, and 15% neither form. By chance the number of half cystines (plus liganded ones) and free cysteines in the set was exactly equal: 148 half-cystines (plus nine liganded cysteines) and 157 free cysteines.

Exceptions: Proteins containing both cysteines and half Cystines

In the few proteins which have both cysteine and cystine residue, the cysteines are usually bonded, e.g. 1CC5, (155C) cytochromes, where two Cys are bonded to the heme group, while the other two occur in disulfide bonds. Other examples include: 2AZA, the electron transport protein, azurin, where the 'free' Cys is in the active site, forming a ligand to Cu together with two His and a Met, and 9PAP, the sulfhydryl proteinase papain, where Cys can be found in the active site. The only protein in our set which contains both the bonded and free forms of Cys is 2SOD, which is an oxidoreductase (Cu, Zn superoxide dismutase) (Bordo *et al.*, 1994). In this molecule there are three Cys, two of which form a disulfide bond (Cys144–Cys55), while the third (Cys6) is free. The calculated relative accessible surface area for this latter Cys is 0.000.

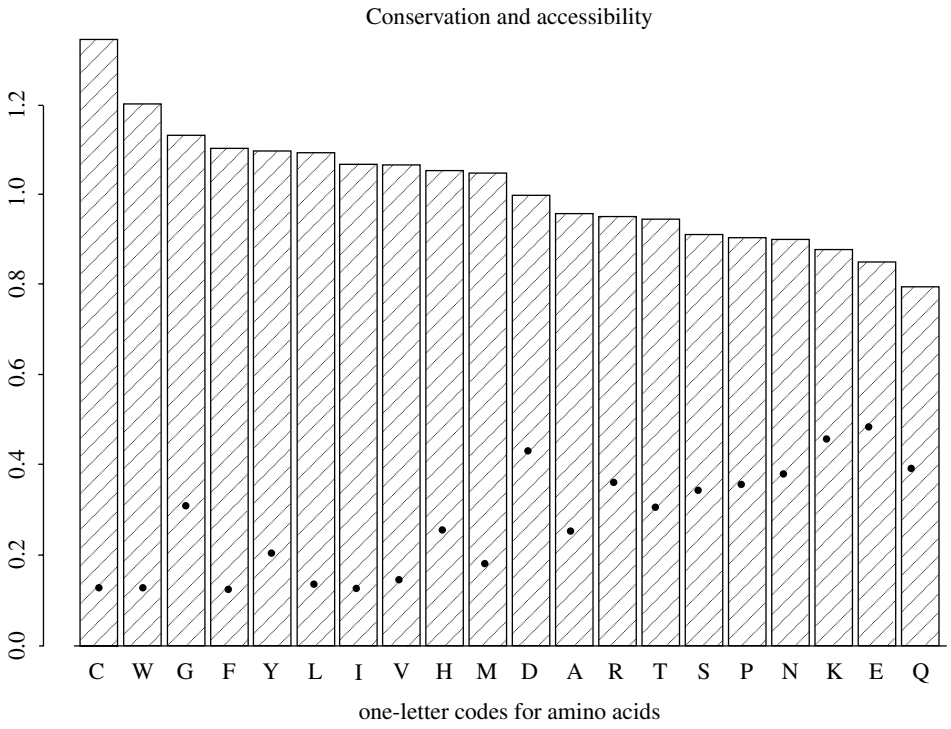

Fig. 1. Average relative conservation and accessibility of the 20 amino acids (see Materials and methods). The shaded bars represent the average relative conservation score and the black dots correspond to the average relative accessibilities of each of the 20 residues. The amino acids are indicated underneath by their one-letter code.

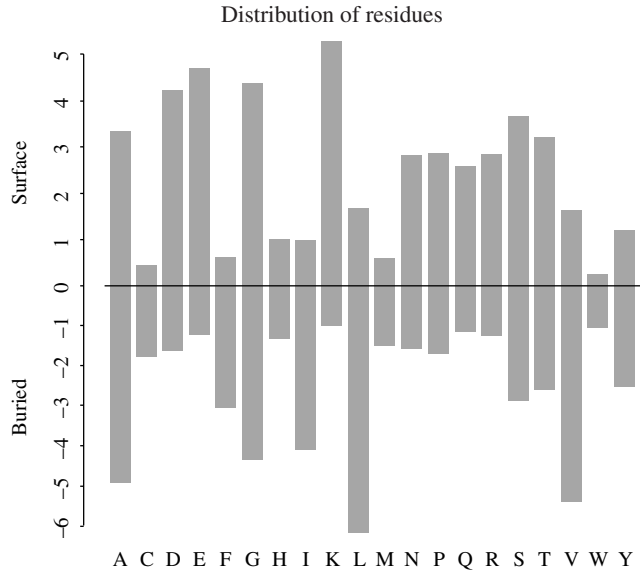

Fig. 2. Distribution of the 20 amino acids between exposed and buried positions in the analyzed 81 protein structures. The length of the bars correspond to the percentage of the residues in the overall composition. The amino acid one-letter codes are underneath.

Thus, the statistical correlations appear more pronounced if cysteines are grouped as bonded versus free rather than as cysteine versus cystine.

We checked this hypothesis on a larger, less strictly selected data set (including lower resolution x-ray structures with a crystallographic R factor less than 25%). This larger set contained 233 proteins: 161 (69.1%) had only free cysteines, 24 (10.3%) had only half cystines, 33 (14.2%) had neither and 15 (6.4%) contained both forms of Cys. Investigating this latter group, we found that at least half of the proteins containing both cysteines and half-cystines occur in the same oxidation state, as in the cytochromes, sulfhydryl proteinases, electron transport proteins cited above. This is true also for protein complexes (e.g. 1ATN), endodeoxyribonuclease complex with actin. The actin forms a 1 : 1 complex with DNase I, where the actin has four sulfhydryl groups and the DNase has two disulfide bridges, but that situation corresponds to two separately folded molecules.

In the two analyzed datasets only a few percent (2–4%) of the proteins contain both redox types of cysteines in the same molecule, but a specific role is often suspected for the free cysteines such as interdomain links, heavy atom binding sites, active site, etc. The number of Cys in these proteins is not even.

Table 1. The distribution of different oxidation states of cysteines among different cellular location (extracellular, intracellular and periplasmic)

Cys state	Protein location		
	Intracellular% (46)	Extracellular% (30)	Periplasmic% (5)
free	78 (36)	10 (3)	40 (2)
SS	0 (0)	70 (21)	20 (1)
Bonded	4 (2)	0 (0)	0 (0)
−	13 (6)	17 (5)	20 (1)
Free + SS	2 (1)	0 (0)	0 (0)
Bonded + SS	2 (1)	3 (1)	20 (1)
Free + bonded	0 (0)	0 (0)	0 (0)

Free: Cysteines, free thiols; SS: Cystines, i.e. Cys in disulfide bridges; Bonded: Cys liganded to a prosthetic group or in an active center; −: proteins without any Cys; and combination of the above subgroups (free + SS; bonded + SS; free + bonded)

Thornton (1981) mentions that the free thiols are unstable outside the cell, i.e. Cys predominantly occur in disulfide bridges (see also Fahey *et al.* (1977)). In the intracellular environment the thiols are kept reduced by glutathione, but once outside, they are very reactive and may even cause polymerizations. We grouped the proteins in our survey according to their cellular locations into three subgroups (intracellular, extracellular and periplasmic), and checked the occurrence of different types of Cys in the subgroups (Table 1). The cellular location shows a high correlation with the oxidation state of the cysteines, but this correlation is not exclusive. In our set among the intra- and extracellular proteins only one contains both type of Cys (Table 1) while 10% of extracellular proteins contain free Cys. Inside the cell, none of the proteins contains disulfide bridges according to our data set. Only a few periplasmic proteins are present in our data set but they already show the most variation in the redox state of Cys.

Prediction methods

In the light of the conservation analyses of multiple sequence alignment results we can set up a simple and efficient prediction approach. As discussed, the covalent state of Cys is determined almost exclusively by the location of the proteins: in our representative set there is hardly a single protein where oxidized and reduced cysteine occur together, except when a cysteine is covalently bonded to heteroatoms, prosthetic groups or other amino acids in active sites, etc. so that the cysteine is also oxidized. Therefore, we use the criterion that if a larger fraction of the predicted Cys belongs to one oxidation state (with high conservation score to the group of oxidized, and lower conservation score to the group of reduced Cys) then the other Cys in the same molecule can be assumed to be in the same oxidation state.

If the predicted number n of reduced and oxidized cysteines is equal in a protein, one must take the relative conservation score into account. To compare them, one must take the average of the relative conservation scores for the predicted bonded cystines and predicted free cysteines, then take the logarithm of these averages and compare their absolute value. Mathematically, if

$$\ln\left[\mathrm{abs}\left(\frac{\sum_{i=1}^{n} C^i_{r/\mathrm{oxidized}/}}{n} - \mathrm{Mean}\right)\right] - \ln\left[\mathrm{abs}\left(\frac{\sum_{i=1}^{n} C^i_{r/\mathrm{reduced}/}}{n} - \mathrm{Mean}\right)\right] > 0$$

then the cysteines are predicted as oxidized, otherwise as reduced. The overall mean for relative conservation score is 1.27. This normalized score is not sensitive for the number of sequences or for their similarities in the alignment.

The efficiency of the prediction was tested by the jack-knife procedure. One alignment was removed in each step and the averages of the relative conservation of half cystines and cysteines were calculated from the remaining ones. The average of the two gave the threshold; if the larger fraction of the Cys in the tested protein fell below the threshold, every Cys was predicted to be in the reduced state, otherwise every Cys was considered as oxidized. In case of an equal number of predicted oxidized and reduced Cys, we considered the absolute average deviation from the actually determined threshold, which was determined from the remaining part of the set, as described above. In this case it is possible to define the covalent state 75.8% of cysteines (119 good versus 38 bad prediction) and 89.8% of half Cystines (141 good versus 16 bad prediction), in a jack-knife test (overall average 82.8%). If we used a constant threshold (1.27) throughout the test the overall efficiency rose above 84%.

The amino acid composition of proteins with oxidized and reduced cysteines shows clear differences. In the case of oxidized cysteines, the occurrence of serine, threonine and glutamine is favoured while in case of reduced cysteines the occurrence of glutamate, histidine, lysine and arginine are higher. The former group includes H-bond forming residues, while positively charged residues are prominent in the latter group. These observations agree with our earlier observations about the specific sequence environment of Cys (Fiser *et al.*, 1992).

Three types of mispredictions can occur. First, some free cysteines are strongly conserved for functional reason as in the case of bilin binding protein (1BBPA), plastocyanin (1PCY) and Ferredoxin (3FXC). Second, both forms of Cys occur in the same protein, e.g. the oxidoreductase (2SOD). Third, misprediction can occur when the normalized conservation of the cysteines is near

Table 2. Conservation and frequency of bonded and non-bonded cysteines in secondary structural elements. The first column refers to overall database statistics considering all type of residues

Secondary structures	All residues frequency %	Non-bonded Cys		Bonded Cys	
		Frequency %	Conservation	Frequency %	Conservation
Helix	32	31	1.137	25.0	1.770
Sheet	23	26	1.033	28.7	1.444
Coil	45	42	1.350	47.0	1.587

the prediction threshold, as in cytochrome C (1CC5), serine protease (2ALP) and azurin (2AZAA).

If the C_r values of cysteines and cystines are very close to each other (near the mean) this usually results from an uninformative alignment that has either too few sequences or the sequences are very similar and C_r does not vary enough along the sequence.

Occurrence in secondary structures, linked structures

The frequencies of oxidized and reduced cysteines in secondary structure elements are very similar. Cys occur most often in coil structures, and these Cys are also the most conserved ones, but conserved bonded cysteines also occur in helixes (Table 2). The high frequency of bonded half cystines in coil structures is not connected with a distortion of the regular structure by disulfide links: such a situation occurs only twice among our examples where the preceding and following residues occur in the same secondary structural elements, while the central Cys is in coil structure according to the DSSP definition (Kabsch and Sander, 1983).

Among the secondary structural elements which are linked by disulfide bonds, the most predominant is the coil–coil linkage followed, in order, by the 'regular structure'–coil ones (sheet–coil, helix–coil), 'regular structure'–'different regular structure' (e.g. sheet–helix) while the sheet–sheet or helix–helix linkage (i.e. the 'regular structure'–'similar regular structure') is quite rare (6.6%) (Table 3). This latter one occurs only five times, and three of them in the same molecule: bovine pancreatic prophospholypase (4BP2). It seems that the high frequency of disulfide bridges between linked coils or a coil and a regular structure is more general and important for the protein three dimensional structures. This result agrees with the experimental observation of Matsumura and Matthews (1989, 1991) who introduced non-native disulfide bonds into T4 lysosyme: if the disulfide bond was located in a regular (α-helical) structural region, which was the most rigid part of the protein, there was no observable increase in stability, while in the case of flexible regions, the engineered disulfide bond increased the thermostability of the molecule. The disulfide links in these structural elements appear to be as important as conserved salt bridges between Asp residues which occur mainly in coil structures (Fiser *et al.*, 1996).

Table 3. The frequency of secondary structural elements linked by disulfide bonds

Types of linked secondary structures	Frequency	Percentage
Helix – Helix	5	6.6
Helix – Sheet	15	19.7
Helix – Coil	13	17.1
Sheet – Sheet	5	6.6
Sheet – Coil	19	25.0
Coil – Coil	19	25.0
Total	76	100

Conclusion

Cysteine is by far the most conserved residue, both on the surface and in the interior of proteins. We showed that the relative conservation scores of the oxidized and reduced forms of Cys differ significantly, which permits the oxidation state of Cys to be predicted from multiple sequence alignments. This prediction method performs above 82% accuracy in a jack-knife test of 81 proteins. The results suggest that the different level of conservation of cysteines in different covalent states is highly informative and this is independent from the average conservation of the whole protein. Functionally important Cys residues that are covalently bonded to prosthetic groups or active-site residues appear to correlate better with disulfide-bonded cysteines rather than with free cysteines. These results suggest that the natural borderline lies between the different oxidation states of Cys rather than between the half cystines and cysteines. There is a high (but not perfect) correlation between the oxidation state of Cys and the location of the protein within the cell; intra-cellular proteins all have reduced cysteines, whereas one-tenth of extracellular proteins also have reduced cysteines. The role of disulfide bonds in structure stabilization is supported by the observation that

they most often connect two coil regions or a coil with a regular secondary structure.

Acknowledgements

We thank Dr Geoff Barton for allowing part of this work to be done in his laboratory, at the Laboratory of Molecular Biophysics, University of Oxford where A.F. has been awarded with 9 month Soros fellowship. We also thank William J. Wedemeyer, Dr Peter Tompa and Gábor E. Tusnády for their critical comments on the manuscript. The authors thank the British–Hungarian project, OTKA F017645, T012890 and T017652. A.F. is a Burroughs Wellcome Fellow.

References

Abola,E.E., Bernstein,F.C., Bryant,S.H., Koetzle,T. and Weng,J. (1987) Protein data bank. In Allen,F.H., Bergerhoff,G. and Sievers,R. (eds), *Crystallographic Databases — Information, Content, Software Systems, Scientific Applications.* Data Commission of the International Union of Crystallography, Bonn, pp. 107–132.

Barton,G.J. (1993) Alscript: A tool to format multiple sequence alignments. *Prot. Eng.*, **6**, 37–40.

Barton,G.J. and Sternberg,M.J.E. (1987) A strategy for the rapid multiple alignment of protein sequences: Confidence levels from tertiary structure comparisons. *J. Mol. Biol.*, **198**, 327–337.

Bordo,D., Dijnovic,K. and Bolognesi,M. (1994) Conserved patterns in the cu,zn superoxide dismutase family. *J. Mol. Biol.*, **238**, 366–386.

Cedano,J., Aloy,P., Pérez-Pons,J. and Querol,E. (1997) Relation between amino acid composition and cellular location of proteins. *J. Mol. Biol.*, **266**, 594–600.

Clarke,J. and Fersht,A. (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry*, **32**, 4322–4329.

Creighton,T. (1993) *Proteins: Structures and Molecular Properties*. 2nd edn, Freeman,W.H. (ed.), New York.

Eder,J. and Wilmanns,M. (1992) Protein engineering of a disulfide bond in a beta/alpha-barrel protein. *Biochemistry*, **31**, 4437–4444.

Fahey,R., Hunt,J. and Windham,G. (1977) On the cysteine and cystine content of proteins. Differences between intracellular and extracellular proteins. *J. Mol. Evol.*, **10**, 155–160.

Fiser,A., Cserző,M., Tüdős,E. and Simon,I. (1992) Different sequence environments of cysteines and half cystines in proteins: Application to predict disulfide forming residues. *FEBS Letters*, **302**, 117–120.

Fiser,A., Simon,I. and Barton,G. (1996) Conservation of amino acids in multiplealignments: Asp shows unsusal conservation. *FEBS Letters*, **397**, 225–229.

George,D.G., Barker,W.C. and Hunt,L.T. (1986) The protein identification resource. *Nucl. Acids Res.*, **14**, 11–15.

Hartley,B., Brown,J., Kauffman,D. and Smilie,L. (1965) Evolutionary similarities between pancreatic proteolytic enzymes. *Nature*, **1**, 1157–1159.

Kabsch,W. and Sander,C. (1983) A dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kremser,A. and Rasched,I. (1994) The adsorption protein of filamentous phage fd: Assignment of its disulphide bridges and identification of the domain incorporated in the coat. *Biochemistry*, **33**, 13954–13958.

Livingstone,C. and Barton,G. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comp. Appl. Biosci.*, **9**, 745–756.

Matsumura,M. and Matthews,B. (1989) Control of enzyme activity by an engineered disulfide bond. *Science*, **243**, 792–794.

Matsumura,M. and Matthews,B. (1991) Stabilization of functional proteins by introduction of multiple disulfide bonds. *Meth. Enz.*, **202**, 336–355.

Morris,H. and Pucci,P. (1985) A new method for rapid assignment of s-s bridges in proteins. *Biochem. Biophys. Res. Commun.*, **126**, 1122–1128.

Muskal,S., Holbrook,S. and Kim,S. (1990) Prediction of the disulfide-bonding state of cysteine in proteins. *Prot. Eng.*, **3**, 667–672.

Ritov (1995) Antioxidant paradoxes of phenolic compounds: peroxyl radical scavenger and lipid antioxidant, etoposide (vp-16), inhibits sarcoplasmic reticulum ca(2+)-atpase via thiol oxidation by its phenoxyl radical. *Arch. Biochem. Biophys.*, **321**, 140–152.

Rose,G., Geselowitz,A., Lesser,G., Lee,R. and Zehfus,M. (1985) Hydrophobicity of amino acids in globular proteins. *Science*, **229**, 834–838.

Simon,I., Glasser,L. and Scheraga,H. (1991) Calculation of protein conformation as an assembly of stable overlapping segments: application to bovine pancreatic trypsin inhibitor. *Proc. Natl Acad. Sci.*, **88**, 3661–3665.

Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.

Tannhauser,T., Konishi,Y. and Scheraga,H. (1984) Sensitive quantitative analysis of disulfide bonds in polypeptides and proteins. *Anal. Biochem.*, **138**, 181–188.

Thornton,J. (1981) Disulphide bridges in globular proteins. *J. Mol. Biol.*, **151**, 261–287.

Xue,J., Kalafatis,M., Silveira,J., Kung,C. and M.,K.G. (1994) Determination of the disulfide bridges in factor va heavy chain. *Biochemistry*, **33**, 13109–13116.

Zhou,N., Kay,C. and Hodges,R. (1993) Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-stranded helical coiled-coil. *Biochemistry*, **32**, 3178–3187.