JOURNAL OF
PEPTIDE SCIENCE

Limited tendency of α-helical residues to form disulfide bridges: a structural explanation

ALFONSO DE SIMONE,[a] RITA BERISIO,[b,c] ADRIANA ZAGARI,[a,b] and LUIGI VITAGLIANO[b,c]*

[a] Dipartimento delle Scienze Biologiche, Sezione Biostrutture, and CNISM, Università degli Studi di Napoli ``Federico II'', I-80134 Napoli, Italy
[b] Istituto di Biostrutture e Bioimmagini, CNR, I-80134 Napoli, Italy
[c] Centro Interuniversitario di Ricerca sui Peptidi Bioattivi (CIRPEB), I-80134 Napoli, Italy

Abstract: Disulfide bridges have an enormous impact on the structure of a large number of proteins and polypeptides. Understanding the structural basis that regulates their formation may be important for the design of novel peptide-based molecules with a specific fold and stability. Here we report a statistical analysis of the relationships between secondary structure and disulfide bond formation, carried out using a large database of protein structures. Our analyses confirm the observation sporadically reported in previous investigations that cysteine residues located in α-helices display a limited tendency to form disulfide bridges. The very low occurrence of the disulfide bond in all α-chains compared to all β-chains indicates that this property is also evident when proteins with different topologies are investigated. Taking advantage of the large database that endorsed the analysis on relatively rare motifs, we demonstrate that cysteine residues embedded in 3_{10} helices present a good tendency to form disulfide bonds. This result is somewhat surprising since 3_{10} helices are commonly assimilated into α-helices. A plausible structural explanation for the observed data has been derived combining analyses of disulfide bond sequence separation and of the length of the different secondary structure elements. Copyright © 2006 European Peptide Society and John Wiley & Sons, Ltd.

INTRODUCTION

Disulfide bridges play a fundamental role in the folding and the stabilization of a large number of proteins and polypeptides [1]. Although disulfide bonds are traditionally considered as either structural or catalytic, 'allosteric' disulfide bridges capable of controlling protein function by triggering conformational changes upon their formation/breakage have recently been characterized [2]. Disulfide bond formation occurs in the periplasmic space in bacterial cells and on the endoplasmic reticulum in eukaryotes; in both cases it is catalyzed/regulated by a system of enzymes [3,4]. Disulfide bonds are also involved in the formation of certain structural motifs, such as the cystine knot in a number of proteins [5] and the CxxC motif in redox-active proteins [6].

Given the importance of disulfides in protein folding, stability and function, a deep knowledge of the physico-chemical principles that regulate their formation is of primary importance in structural biology. General characteristics of disulfide bonds have been examined in numerous studies [1,7–14]. One of the first broad

analyses of protein disulfide bridges [7] examined the distribution of disulfides in different topologies, cystine conformations and structural folds. From this study, it was noted that cysteine residues are remarkably conserved in homologous proteins and that the loss of a disulfide was usually associated with mutations of both cysteine residues [7]. Early statistical analyses of protein databases also provided preliminary indications that disulfide bridges are relatively rare among residues located in α-helices [15]. With the increase of the structural content of the Protein Data Bank (PDB) [16], almost every aspect related to disulfide bond formation has been analyzed. In particular, disulfide bond formation has been used to engineer proteins with additional stability via site-directed mutagenesis [17,18]. Furthermore, the strong conservation of disulfide patterns has found applications in the recognition of evolutionary relationships among distantly related protein homologs [19,20]. More generally, it has been shown that proteins with similar disulfide patterns have similar structure and function even in the cases of poor sequence similarity [21].

Recent investigations have further dealt with the relationship between the oxidation state of cysteines and their secondary structure [12,22]. In this respect, an analysis of the dependence of disulfide conformational properties on the local secondary structure has been reported [12]. However, the existence of a bias in the

* Correspondence to: L. Vitagliano, Istituto di Biostrutture e Bioimmagini, Consiglio Nazionale delle Ricerche, Via Mezzocannone 16. I-80134 Napoli, Italy; e-mail: luigi.vitagliano@unina.it

WILEY
InterScience®
DISCOVER SOMETHING GREAT

tendency of cysteine residues embedded in different secondary structural elements to form disulfides was not discussed. An independent investigation [22] confirmed the original observation of Thornton and Gardner [15] that cysteines in α-helix have a low tendency to form disulfide bonds. Although these authors have not provided a structural explanation for this observation, they fruitfully exploited these findings in the development of strategies for predictions of disulfide bridge formation [22]. Taking advantage of the explosive growth of protein structure records in the PDB [16], we extended the statistical analysis of disulfide occurrence depending on the secondary structure by including rare motifs such as the 3_{10} helix. We also provide a structural explanation for the observed data.

MATERIALS AND METHODS

Notation

The term *half-cystine* designates Cys residues participating in the formation of disulfide bridges. Cys residues not involved in disulfides are denoted as *reduced cysteine*. Finally, the term *cysteine* is used if the oxidation state of the residue is not relevant to the subject discussed.

Taking into account the secondary structure location of the half-cystine involved, disulfide bonds are classified as HH (α-helix–α-helix), (HE) (α-helix–β-sheets), HG (α-helix–3_{10} helix), HC (α-helix–coil), and so on.

The Database

The May 2006 release of the PDB [16] was used for the statistical surveys. A set of nonredundant protein chains with sequence identity lower than 25% was extracted from the PDB using the program PISCES [23]. From this ensemble, a dataset, hereafter denoted as DATA25, was generated by considering only crystallographic structures refined at a resolution higher than 2.5 Å and to a crystallographic R-factor value lower than 0.22. The application of these selection criteria led to the sorting of 3121 protein chains from 2948 PDB entries.

Identification of Secondary Structure Elements and Disulfide Bridges

The identification of the secondary structure elements was performed using the DSSP algorithm [24] implemented in the program PROCHECK [25]. Residues belonging to α-helices, β-sheets and 3_{10} helices were identified from the labels H (or h), E (or e) and G (or g), respectively. All other residues were assigned to coil structures. These data were also used for the evaluation of secondary structure element lengths.

Two independent approaches were used to detect intra- and interchain disulfide bridges. The former were selected from the single chain database DATA25 using the program PROCHECK. The interchain bridges were preliminarily identified using the information contained in the header section of the PDB file (record SSBOND) and then analyzed individually.

All α- and all β-protein chains were identified on the basis of the secondary structure assignment made by PROCHECK [25]. In particular, the presence of α-helical residues (H) and the simultaneous absence of β-sheet residues (E) were checked for the recognition of all α-chains. These criteria were reversed in the identification of all β-chains.

In the evaluation of the occurrence of half-cystines depending on their location in long α-helices (longer than 12 residues), helices were divided into an N-terminal region (first four residues), a C-terminal region (last four residues) and a central region (remaining residues).

RESULTS AND DISCUSSION

Disulfide Bridge Occurrence and Secondary Structure Assignment

The statistical analyses on secondary structure elements and disulfide bridges here reported were conducted using the structural information contained in the May 2006 release of the PDB. Using a database of 3121 nonredundant protein chains (DATA25), selected with the criteria described in the 'Materials and Methods' section, 1195 intrachain disulfide bonds were identified. The percentage of the protein entries containing at least a disulfide bridge in the PDB release here used is 26.7%, a value that is in line with previous statistical surveys. It is worth mentioning, however, that this percentage of protein chains rises up to 41.1% in the nonredundant database DATA25. As expected, the number of interchain disulfide bonds identified in the DATA25 chains is much smaller (as small as 76).

As reported in Table 1, the distribution of secondary structure elements in the DATA25 follows well-known trends. Indeed, the percentage of residues located in α-helices, β-sheets and 3_{10} helices is 39.3, 26.0 and 5.1%, respectively. The analysis of the distribution of reduced cysteines among secondary structure elements indicates that this residue does not present any particular preference for a specific motif. Indeed, the percentages of reduced cysteine residues in α-helices, β-sheets and 3_{10} helices are 37.9, 33.7 and 2.8%.

The assignment of the secondary structure for the 2390 half-cystines involved in intrachain disulfide bridges is reported in Tables 1 and 2. From Table 2 it is evident that disulfide bond formation is compatible with half-cystine location in every combination of secondary structure elements. As an obvious consequence of the somewhat rare occurrence of 3_{10} helices, we found only two bridges formed by half-cystines both embedded in 3_{10} helices (GG bridges, see notation in 'Materials and Methods').

In order to unveil the existence of any bias of the half-cystine distribution among different secondary structure elements, we compared the observed values with those expected on the basis of the abundance of reduced cysteine residues in each secondary structure element and assuming their random association (Table 2). This comparison clearly shows that, in line

Table 1 Total secondary structure content and cysteine distribution in DATA25

	Number of residues	Number of reduced cysteines	Number of half-cystines involved in interchain disulfide bonds
α-Helix (H + h)[a]	281 784	2622 (540)[b]	50
β-Sheet (E + e)[a]	186 713	2333 (795)[b]	28
3_{10} Helix (G + g)[a]	36 419	296 (148)[b]	1
Coil regions	212 406	1659 (907)[b]	73
Total	717 322	6910 (2390)[b]	152

[a] The notation is taken from the program PROCHECK [25].
[b] The number of half-cystines involved in intrachain disulfide bonds is given in parenthesis.

Table 2 Percentage of disulfide bridges between half-cystine residues embedded in various secondary structure elements. The number of the detected bridges is given in parenthesis

	H		E		G		C	
Detected[a]								
H	9.5	(114)	8.2	(98)	3.4	(41)	14.5	(173)
E	8.2	(98)	15.2	(185)	3.8	(46)	23.5	(281)
G	3.4	(41)	3.8	(46)	0.2	(2)	4.8	(57)
C	14.5	(173)	23.5	(281)	4.8	(57)	16.6	(198)
Expected[b]								
H	21.3		19.8		2.5		13.3	
E	19.8		18.4		2.3		12.3	
G	2.5		2.3		0.3		1.6	
C	13.3		12.3		1.6		8.3	

[a] The analysis has been limited to the intrachain disulfide bridges, which represent the vast majority of disulfide bridges.
[b] Expected frequencies have been computed on the observed reduced cysteine frequencies in secondary structure elements (Table 1). In particular, all possible associations of the reduced cysteine residues were considered. Percentages were computed from the ratio between the possible bonds for a given disulfide type and the total number of possible bridges.

with previous reports [15,22], the actual occurrence of disulfide bonds involving α-helical half-cystine residues is significantly lower than expected. For example, the percentage of HH bridges is less than half of that predicted (9.5% *vs* 21.3%), whereas only a slight decrease (15.5% *vs* 18.4%) is observed in the number of EE bonds. On the other hand, CC bridges are highly over-represented among the observed disulfide bridges (16.6% *vs* 8.3%). Since both predicted and observed GG bridges are very uncommon, any consideration on these bonds is precluded. The examination of the mixed

bonds formed by residues belonging to different structure elements indicates that they follow the same trend identified from the analysis of HH, EE and CC bonds. Along this line, we observe a reduced occurrence of HE bonds, in line with the reduced frequencies of HH and EE. Mixed bonds involving coil residues generally display higher frequencies than expected. This is likely due to the high percentage of coil half-cystine residues involved in disulfide bridge formation (Table 1). In contrast to GG, the frequency of mixed bonds involving residues located in 3_{10} helices, although limited, is significant. Interestingly, all these mixed bonds, HG, EG and GC, display higher observed frequencies than expected. This finding suggests that, unlike those in α-helix, cysteine residues located in 3_{10} helices have a good tendency to form disulfide bridges. This result is corroborated by the observation that the proportion between the total number of half-cystines and reduced cysteines is very different in 3_{10} helices (148 *vs* 296) and α-helices (540 *vs* 2622) (Table 1).

In order to check whether a reminiscence of these findings exists at the protein topology level, we analyzed the occurrence of disulfide bridges in all α- and all β-proteins. Following the procedure illustrated in the 'Materials and Methods' section, we identified 483 (total number of residues 56 812) and 192 (total number of residues 28 071) all α- and all β-chains, respectively. As a consequence of the difference in the amount of protein chains in the two classes, the number of reduced cysteines is much larger (519 *vs* 197) in the all α-dataset. On the other hand, the number of half-cystines in all β-chains is almost double that found in all α-ones (448 *versus* 280). These data indicate that the negative bias exhibited by cysteine residues embedded in α-helices toward the formation of disulfide bridges is conserved in structures with different topologies.

We extended these analyses with the evaluation of average hydrophobicities of secondary structure elements containing cysteine residues in different oxidation states. Following the procedure previously adopted for estimating the average hydrophobicities of secondary structural elements [26], we preliminarily identified α-helices and β-strands of DATA25 which contained either reduced cysteines or half-cystines. For each fragment, an average hydrophobicity was calculated using the amino acid hydrophobicity scale reported by Eisenberg and McLachlan [27]. Histograms showing the distribution of average hydrophobicities are reported in Figure 1. Although the histograms computed for the various classes of fragments are largely overlapping, some subtle albeit significant differences may be recognized. Indeed, β-strands embedding reduced cysteine residues present slightly higher average hydrophobicities compared to those of β-strands containing half-cystine residues (Figure 1(b) and (c)). A similar trend is displayed by α-helices embedding either reduced or oxidized cysteines (Figure 1(e) and
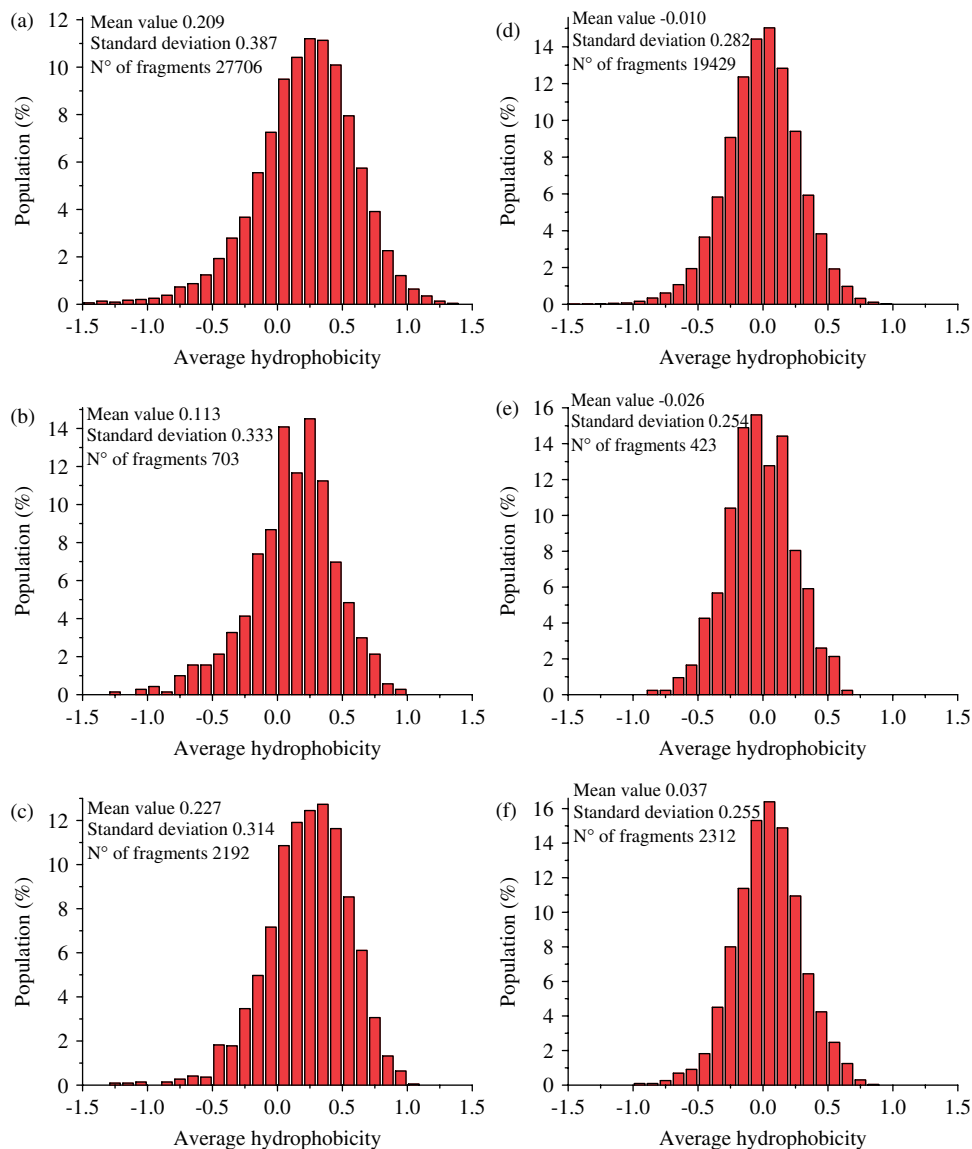
Figure 1 Distributions of average hydrophobicities of β-strands (a), (b) and (c) and α-helices (d), (e) and (f). The global distribution of average hydrophobicities of the β-strands and α-helices selected from DATA25 is shown in panels (a) and (d), respectively. Hydrophobicities of β-strands containing half-cystines and reduced cysteines is reported in panel (b) and (c), respectively. Hydrophobicities of α-helices containing half-cystines and reduced cysteines is reported in panel (e) and (f), respectively. The same hydrophobicity value has been assigned to both reduced cysteines and half-cystines.

(f)), although very small differences are observed in this case. The data here reported confirm and extend the observation that reduced cysteines are slightly more buried than half-cystines [22]. These findings indicate that cysteine burial may occasionally be a way to protect this residue from unwanted oxidation.

The Conformational Preferences of Half-cystine Residues: Analysis of Side Chain Rotamers

In order to obtain insights into the structural determinants of half-cystine conformational preferences, we carried out an extensive analysis of their side chain conformation as a function of the secondary structure adopted. For comparative purposes, the investigation

was extended to rotamers of reduced cysteines. As shown in Figure 2, the analysis of the χ_1 dihedral angle of reduced cysteines demonstrates that in all cases a prevalence of the g^- conformer ($\chi_1 \approx -60°$) is observed. However, the relative abundance of the g^- conformers is higher in both α- and 3_{10} helices when compared to coil and β-sheet cysteines. Notably, the analysis of χ_1 angles in half-cystine residues indicates a high preference for the g^- conformer even for coil and β-sheet half-cystines (Figure 3). Similar values of χ_2 and χ_3 dihedral angles are also observed for half-cystine residues located in different secondary structure elements (Figure 3). Altogether, these data clearly indicate that the observed low tendency of α-helical cysteine residues to participate
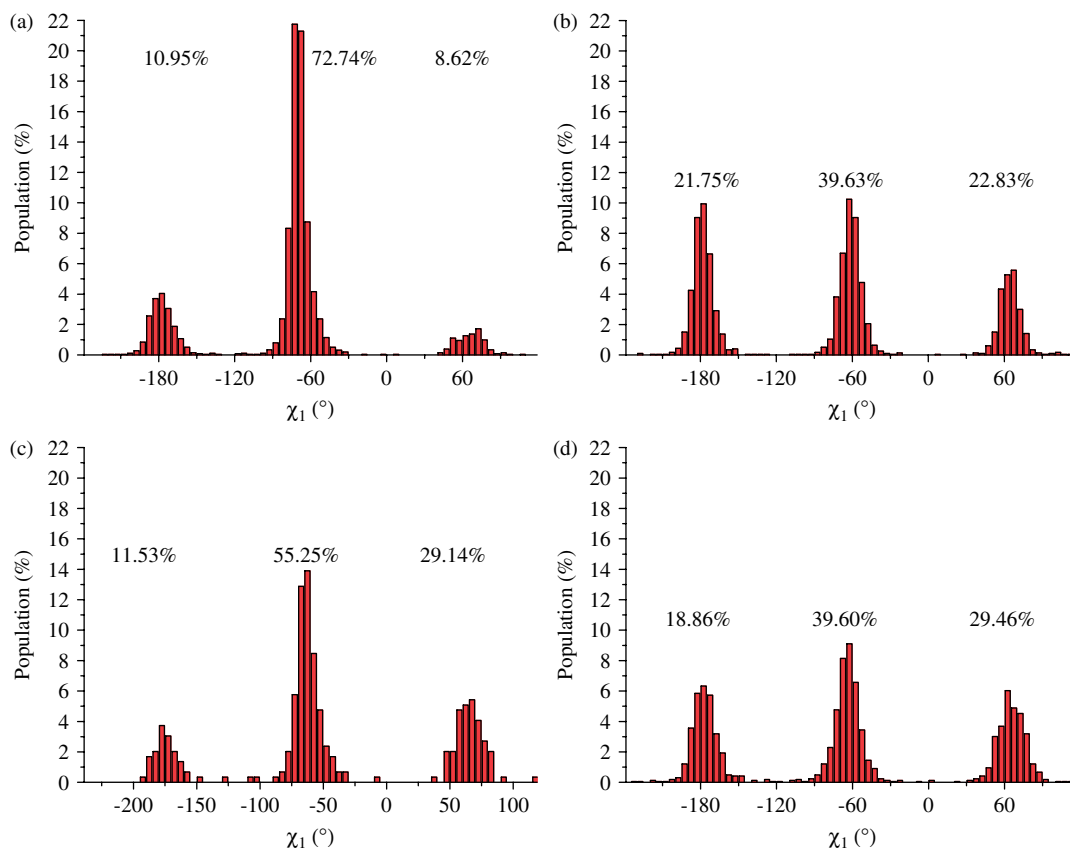
Figure 2   Side chain conformations of reduced cysteine residues embedded in α-helices (a), β-sheets (b), 3_{10} helices (c) and coil regions (d).

to disulfide bridge formation cannot be explained by invoking conformational restrictions imposed by the helical structure. This consideration is corroborated by the observation that α-helical cysteines may participate to the formation of complex and rare disulfide bridge networks, such as those formed by couples of consecutive half-cystines. Indeed, although only four cases of consecutive bridges were identified in the current release of the PDB (entry codes 1WCT, 1BSR, 2F5Y, 1M1J), one of these (1BSR) is formed by four half-cystines located in α-helices [28].

Structural Basis of the Reduced Tendency of α-Helical Cysteines to Form Disulfide Bridges

An alternative approach to look for the structural determinants of the observed tendencies of cysteine residues located in different secondary structure elements to form disulfide bonds was based on the analysis of the sequence separation of the half-cystines. Distributions of sequence distances observed for the bridges HH, EE and CC are reported in Figure 4. A comparison of the three diagrams clearly indicates that short-range HH disulfide bridges are disfavored. Indeed, with the exception of the Cys-X-X-Cys motif frequently found in the disulfide bond (Dsb) isomerases, HH bridges with sequence separation shorter than 20

are under-represented when compared to EE and CC bonds (Figure 4). In this context, it is worth noting that one of the half-cystines of the Cys-X-X-Cys motif found in the Dsb is labeled as h by PROCHECK [25], thus indicating that this residue does not display the typical dihedral angles of a regular α-helix. Indeed, when the analysis is restricted to the sole half-cystines labeled as H, the peak at sequence separation equal to two disappears (data not shown).

The number of HH and EE disulfide bonds with a sequence separation of 25–40 residues is similar (46 *versus* 48). On the other hand, EE disulfide bonds formed by Cys separated by 5–15 residues largely outnumber the corresponding HH bonds (124 *vs* 58). These data clearly demonstrate that the *missing* HH bonds are those involving half-cysteines with a limited sequence separation. The rod-like shape of the α-helix and the geometrical requirements of the disulfide bond prevent the formation of short-range (intrahelix) disulfide bridges for residues located in the central portion of this secondary structure elements. Although similar trends are expected for the other rigid and regular motifs, such as β-sheets and 3_{10} helices, their shorter lengths (Figure 5) reduce the impact of these effects. Indeed, the number of residues located in the central region of the secondary structure elements,
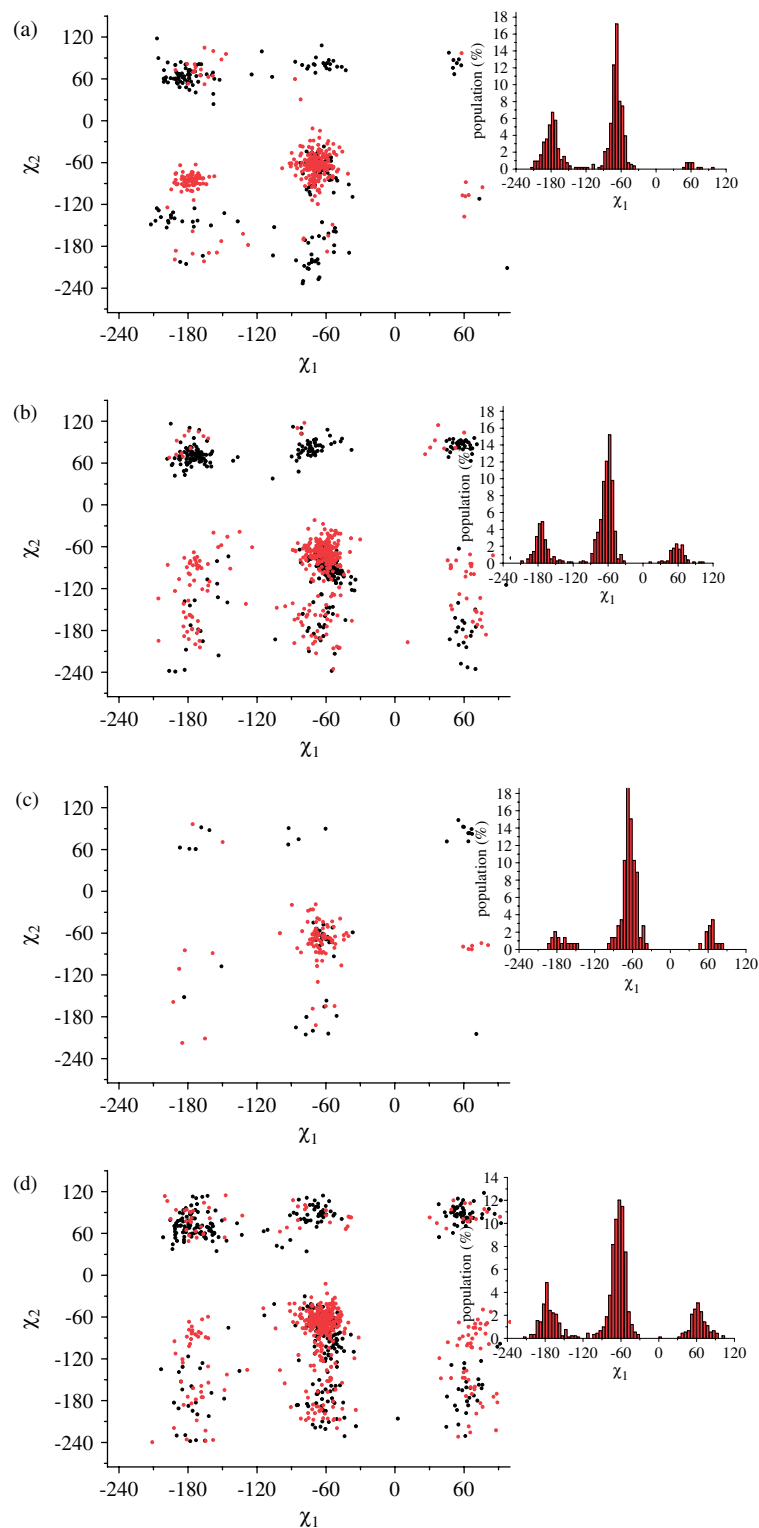
Figure 3 Side chain conformations of half-cystine residues embedded in α-helices (a), β-sheets (b), 3_{10} helices (c) and coil regions (d). Bridges with positive and negative χ_3 are represented in black and red, respectively. Projections of the distributions onto the χ_1-axis are shown in the insets.

whose participation in short-range bridges is precluded, is higher for α-helical structures. This suggestion is confirmed by the analysis of the occurrence of half-cystines *versus* reduced cysteines in the central

and terminal regions of α-helices. A survey carried out on 11 952 α-helices of DATA25 longer than 12 residues shows that the percentage of half-cystines over the total number of cysteines are 26.5, 13.7 and 20.5%
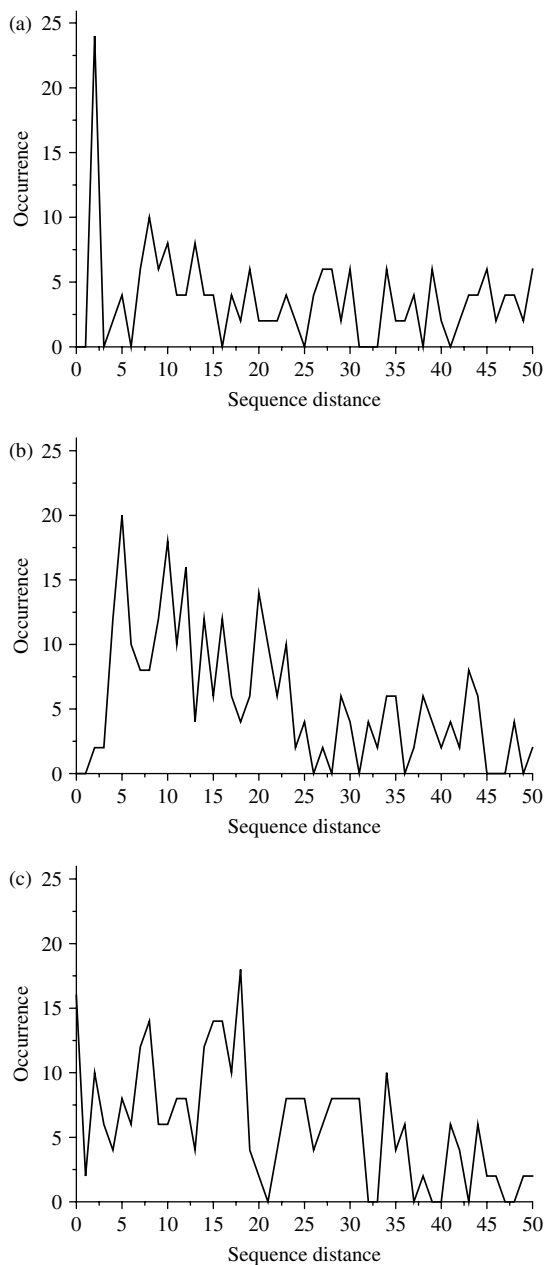
Figure 4 Sequence separation of disulfide bridges of the type HH (a), EE (b) and (CC) (c). Sequence separation is defined as the number n of residues in the Cys-$(X)_n$-Cys motif.

Figure 5 Distributions of lengths of α-helices (a), β-strands (b) and 3_{10} helices (c) in DATA25.

for N-terminal, central and C-terminal regions of α-helices, respectively (see 'Materials and Methods' for definitions).

This hypothesis is indirectly supported by the observation that there is no bias against the presence of α-helical half-cystines in interchain disulfide bonds (Table 1). Although the number of observed interchain bridges is rather limited, the ratio of half-cystines over reduced cysteines is higher for α-helices when compared to β-sheets. Evidently, the limited ability of α-helical cysteines to form bridges with a short sequence separation influences interchain disulfide bond formation.
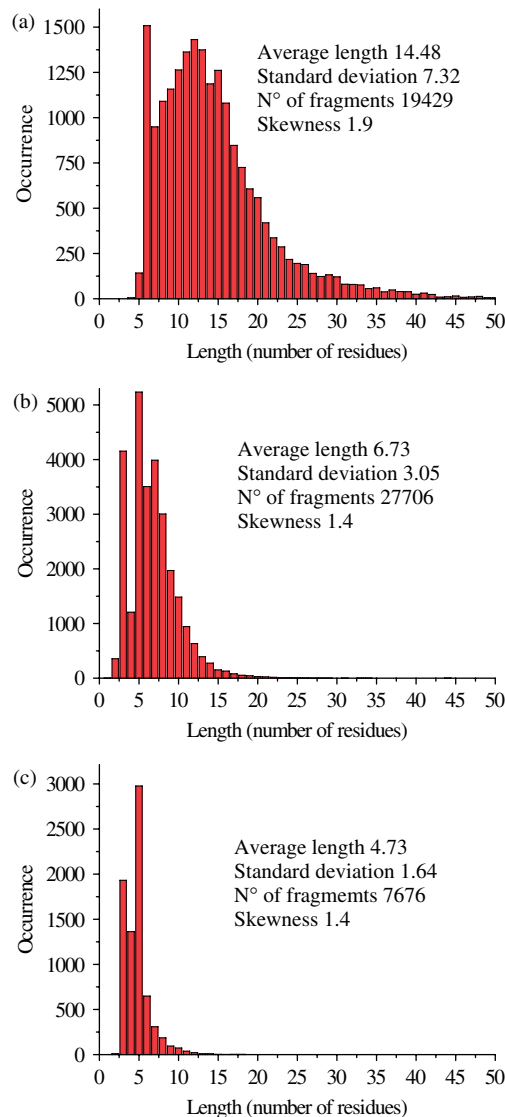
In conclusion, our analyses confirm that α-helical cysteines have a low tendency to form disulfide bridges even when compared to cysteines located in similar motifs such as 3_{10} helices. A plausible explanation for the observed trends, based on simple considerations on the shape and on the length of secondary structure elements, is also provided.

Acknowledgements

REFERENCES

1. Bulaj G. Formation of disulfide bonds in proteins and peptides. *Biotechnol Adv.* 2005; **23**: 87–92.

2. Schmidt B, Ho L, Hogg PJ. Allosteric disulfide bonds. *Biochemistry* 2006; **45**: 7429–7433.

3. Skorko-Glonekv J, Sobiecka A. Periplasmatic disulfide oxidoreductases from bacterium escherichia coli–their structure and function. *Postepy Biochem.* 2005; **51**: 459–467.

4. Tu BP, Weissman JS. Oxidative protein folding in eukaryotes: mechanisms and consequences. *J. Cell Biol.* 2004; **164**: 341–346.

5. Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. *Toxicon* 2001; **39**: 43–60.

6. Iqbalsyah TM, Moutevelis E, Warwicker J, Errington N, Doig AJ. The cxxc motif at the n terminus of an α-helical peptide. *Protein Sci.* 2006; **15**: 1945–1950.

7. Thornton J. Disulphide bridges in globular proteins. *J. Mol. Biol.* 1981; **151**: 261–287.

8. Srinivasan N, Sowdhamini R, Ramakrishnan C, Balaram P. Conformations of disulfide bridges in proteins. *Int. J. Pept. Protein Res.* 1990; **36**: 147–155.

9. Petersen MT, Jonson PH, Petersen SB. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* 1999; **12**: 535–548.

10. Carugo O, Cemazar M, Zahariev S, Hudaky I, Gaspari Z, Perczel A, Pongor S. Vicinal disulfide turns. *Protein Eng.* 2003; **16**: 637–639.

11. Esposito L, De Simone A, Zagari A, Vitagliano L. Correlation between omega and psi dihedral angles in protein structures. *J. Mol. Biol.* 2005; **347**: 483–487.

12. Bhattacharyya R, Pal D, Chakrabarti P. Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng. Des. Sel.* 2004; **17**: 795–808.

13. Benham CJ, Jafri MS. Disulfide bonding patterns and protein topologies. *Protein Sci.* 1993; **2**: 41–54.

14. Fariselli P, Casadio R. Prediction of disulfide connectivity in proteins. *Bioinformatics* 2001; **17**: 957–964.

15. Thornton JM, Gardner SP. Protein motifs and data-base searching. *Trends Biochem. Sci.* 1989; **14**: 300–304.

16. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The protein data bank. *Acta Crystallogr.* 2002; **D58**: 899–907.

17. Thangudu RR, Vinayagam A, Pugalenthi G, Manonmani A, Offmann B, Sowdhamini R. Native and modeled disulfide bonds in proteins: knowledge-based approaches toward structure prediction of disulfide-rich polypeptides. *Proteins* 2005; **58**: 866–879.

18. Vinayagam A, Pugalenthi G, Rajesh R, Sowdhamini R. Dsdbase: a consortium of native and modelled disulphide bonds in proteins. *Nucleic Acids Res.* 2004; **32**: D200–D202.

19. 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.

20. van Vlijmen HW, 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.

21. Gupta A, Van Vlijmen HW, Singh J. A classification of disulfide patterns and its relationship to protein structure and function. *Protein Sci.* 2004; **13**: 2045–2058.

22. Cheng J, Saigo H, Baldi P. Large-scale prediction of disulfide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. *Proteins* 2006; **62**: 617–629.

23. Wang G, Dunbrack RL Jr. Pisces: a protein sequence culling server. *Bioinformatics* 2003; **19**: 1589–1591.

24. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; **22**: 2577–2637.

25. Laskowski RA, MacArthur MW, Moss MD, Thornton JM. Procheck: a program to check the stereochemical quality of protein structure. *J. Appl. Crystallogr.* 1993; **26**: 283–291.

26. Berisio R, Loguercio S, De Simone A, Zagari A, Vitagliano L. Polyproline helices in protein structures: a statistical survey. *Prot. Pept. Lett.* 2006; **13**: 847–854.

27. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. *Nature* 1986; **319**: 199–203.

28. Mazzarella L, Capasso S, Demasi D, Di Lorenzo G, Mattia CA, Zagari A. Bovine seminal ribonuclease: structure at 1.9 Å resolution. *Acta Crystallogr.* 1993; **D49**: 389–402.