Strong hydrophobic nature of cysteine residues in proteins

Nozomi Nagano[a,b], Motonori Ota[a], Ken Nishikawa[a,*]

[a] National Institute of Genetics, Yata, Mishima, Shizuoka 411-8540, Japan
[b] Biochemistry and Molecular Biology Department, University College London, Gower Street, London WC1E 6BT, UK

Received 7 July 1999

Abstract The differences between disulfide-bonding cystine (Cys_SS) and free cysteine (Cys_SH) residues were examined by analyzing the statistical distribution of both types of residue in proteins of known structure. Surprisingly, Cys_SH residues display stronger hydrophobicity than Cys_SS residues. A detailed survey of atoms which come into contact with the sulfhydryl group (sulfur atom) of Cys_SH revealed those atoms are essentially the same in number and variety as those of the methyl group of isoleucine, but are quite different to those of the hydroxyl group of serine. Moreover, the relationships among amino acids were also determined using the 3D-profile table of known protein structures. Cys_SH was located in the hydrophobic cluster, along with residues such as Met, Trp and Tyr, and was clearly separated from Ser and Thr in the polar cluster. These results imply that free cysteines behave as strongly hydrophobic, and not hydrophilic, residues in proteins.

Key words: Amino acid index; Disulfide bonding; Three-dimensional profile; Protein structure

1. Introduction

Rapid accumulation of protein 3D structural data provides an opportunity to statistically analyze the behavior of amino acid residues to higher accuracy. In particular it may be interesting to focus on the types of amino acids that account for a small fraction of the overall protein composition, for which accurate analysis has only recently become feasible. The Cys residue is one such amino acid. The fraction of Cys is small in proteins, especially when the Cys composition is subdivided into two subgroups: free cysteine (denoted here as Cys_SH), and disulfide-bonding half-cystine (Cys_SS). Because these two forms of Cys seem quite different in their physico-chemical properties, statistical analysis should be conducted for each type.

A disulfide bond (-SS-) is covalently formed between two half-cystine residues by an oxidation reaction. However, as the intracellular milieu is maintained in the reduced state, disulfide bond formation is usually prohibited in intracellular proteins [1], and disulfide bonds are almost exclusively found in extracellular proteins [2,3]. In contrast to the inert nature of disulfide bonds, the sulfhydryl groups (-SH) of free cysteine residues are polarized and sometimes participate in metal binding such as in iron-sulfur clusters. Because of the apparent chemical resemblance of the sulfhydryl (-SH) and hydroxyl (-OH) groups, it is commonly perceived that free cysteine (Cys_SH) residues behave in a similar manner to Ser residues in globular proteins [4]. However, in this present study we show that this is not the case and Cys_SH behaves much like the hydrophobic residue Met, and the aromatic residues Trp and Tyr.

2. Materials and methods

A data set of 378 non-redundant protein chains, having less than 30% sequence identity, was prepared from release 83 of the Brookhaven Protein Data Bank (PDB) [5]. A total of 1471 Cys residues, collected in this data set, were further divided into several forms (Table 1). Disulfide-bonding residue pairs were identified according to the DSSP algorithm [6] and these were divided into 'intra-chain' and 'inter-chain' disulfide bonds. Of these, only 610 'intra-chain' disulfide-bonding half-cystines (denoted here as Cys_SS) were used in the following analysis. The remaining cysteines consisted of 742 free cysteines (denoted here as Cys_SH) and 73 metal-binding cysteines, which were thus classified if a metal atom was located within 3 Å of the sulfur atom of the cysteine residue. Metal binding cysteines were excluded from further analysis.

The hydrophobicity of Cys_SH and Cys_SS, or their partitioning nature in globular proteins, was expressed in terms of the number of neighboring heavy (non-H) atoms within a sphere [7], and compared with other types of amino acids. More precise distributions of heavy atoms around the sulfur atom of Cys_SH were also analyzed and compared with corresponding distributions found around the oxygen atom of Ser side chains. Various physico-chemical indices, including the properties of amino acids to form secondary structures, P_α, P_β and P_c [8], and the accessible surface area defined by Rose et al. [9], were estimated for all amino acid types including Cys_SH and Cys_SS.

3. Results and discussion

3.1. Hydrophobicity of disulfide-bonding cystine and free cysteine residues

There have been various types of amino acid indices derived to date [10,11]. According to the original definition by Rose et al. [9], we have re-estimated the hydrophobicity index in terms of the solvent-accessible surface area for 20 amino acids. The results shown in Table 2 are almost identical to those originally obtained by Rose et al., if two types of Cys residues (Cys_SS and Cys_SH) are not distinguished. The mean solvent-accessible surface area, denoted by $\langle A \rangle$, was nearly the same for both Cys_SS and Cys_SH, being 15 Å^2. When the mean fractional area loss [9] was calculated, the definition of which is given in Table 2, Cys_SH showed the largest index (0.89) among all residues, even greater than the aliphatic residues Ile (0.88) and Leu (0.87) (Table 2). The mean fractional area loss for Cys_SS (0.83) is, on the other hand, similar to that for Met (0.84) and the aromatic residues Tyr (0.81) and Trp (0.85).

Table 2 shows another hydrophobic scale defined in terms of the number of surrounding heavy atoms within a sphere (see Section 2). In the original definition of Ota and Nishikawa [7], the number of surrounding atoms (Sn) is sub-divided

*Corresponding author. Fax: (81) (559) 81-6889.
E-mail: knishika@genes.nig.ac.jp

PII: S0014-5793(99)01122-9

Table 1
Subdivision of Cys in the data set

Cys type	Number of residues	Composition (%) among all residues[a]
All cysteine	1471	1.44
Non-disulfide-bonding cysteine	847	0.83
Free cysteine (Cys_SH)	742	0.73
Metal-binding cysteine	73	0.07
No side chain information	32	–
Disulfide-bonding cystine	624	0.61
Intra-chain disulfide-bonding cystine (Cys_SS)	610	0.60
Inter-chain disulfide-bonding cystine	14	-

[a]The total number of residues of all proteins in the data set is 101 905.

into nine classes, and the more residues of a particular amino acid type that are buried within a protein molecule, the larger their Sn value. We therefore expect smaller values for hydrophilic residues and larger values for hydrophobic residues. The relative order of amino acid hydrophobicities seems very similar between the two scales. In particular, the Sn value of Cys_SH (6.82) is again larger than that of Cys_SS (6.12), and is comparable with those of Ile (7.03), Leu (6.91) and Phe (6.79). The correlation coefficient between these two scales is as high as 0.983.

The two kinds of hydrophobicity indices suggest that Cys_SH behaves as a strongly hydrophobic residue in a globular protein, even stronger than Cys_SS. It should also be noted that the hydrophobicity indices of Cys_SH never resemble Ser or Thr and that Cys_SH also does not share the propensity of Ser and Thr to form α-helices (see Table 2). The helical propensity of Cys_SH (1.05) is similar to that of Ile (1.05), Trp (1.01) and Phe (0.99). On the other hand, Cys_SS has a small helical propensity value (0.49) comparable to that of the helix breaker Gly (0.46). The reason for this may be that the rigidity of α-helices prevents the formation of disulfide bonds.

Table 2
Amino acid indices re-estimated from a data set of 378 non-redundant proteins

	Hydrophobicity by		Secondary structure propensity by Chou-Fasman		
	Rose et al. [a]	Ota-Nishikawa	P_α	P_β	P_c
Cys_SS	0.83	6.12	0.49	1.33	1.19
Cys_SH	0.89	6.82	1.05	1.24	0.86
Ala	0.78	5.62	1.47	0.75	0.80
Arg	0.66	4.71	1.22	0.93	0.88
Asn	0.63	4.22	0.72	0.62	1.35
Asp	0.58	3.89	0.85	0.52	1.30
Gln	0.61	4.34	1.31	0.74	0.91
Glu	0.55	3.80	1.39	0.74	0.86
Gly	0.71	4.61	0.46	0.65	1.51
His	0.73	5.45	0.87	1.03	1.07
Ile	0.88	7.03	1.05	1.82	0.61
Leu	0.87	6.91	1.33	1.17	0.71
Lys	0.56	3.71	1.14	0.80	1.00
Met	0.84	6.60	1.32	0.99	0.79
Phe	0.87	6.79	0.99	1.42	0.82
Pro	0.66	4.42	0.45	0.42	1.61
Ser	0.68	4.50	0.78	0.82	1.22
Thr	0.71	4.81	0.77	1.19	1.07
Trp	0.85	6.47	1.01	1.24	0.89
Tyr	0.81	6.34	0.95	1.43	0.85
Val	0.87	6.66	0.94	1.93	0.63

[a]$(A^0-\langle A\rangle)/A^0$, the mean fractional area loss defined by Rose et al. [9], where $\langle A\rangle$ is the mean solvent accessible surface and A^0 is the stochastic standard state accessibility. Values of A^0 for Cys_SS and Cys_SH were obtained from Thornton [2], while the remaining values are those of Rose et al. [9].

3.2. Detailed distribution analysis

To focus on this matter more precisely, we have analyzed in detail the distributions of heavy atoms around a given central atom, such as the sulfur atom of Cys_SH. In this analysis, heavy atoms were distinguished as carbon, oxygen, nitrogen and sulfur, and the number of each type of atom was counted within a sphere of radius 6 Å of the central atom. We have selected three atoms as central atoms, oxygen (Oγ) of Ser, carbon (Cδ) of Ile, sulfur (Sγ) of Cys_SH, and the average number of surrounding heavy atoms for each of these central atoms is shown in Fig. 1. Although no large differences are seen in the number of surrounding O, N and S atoms for each of the central atoms, a clear difference is observed in the number of surrounding carbons. The sulfur atom of Cys_SH has almost the same distribution of surrounding carbons as does the Cδ atom of Ile, but the Oγ of Ser has a lesser number of carbons. Moreover, the similarities in number of surrounding atoms between Cys_SH and Ile did not change even when the surrounding carbons were subdivided into aliphatic, aromatic and carbonyl carbons (data not shown). These results, consistent with the previous data in Table 2, demonstrate that

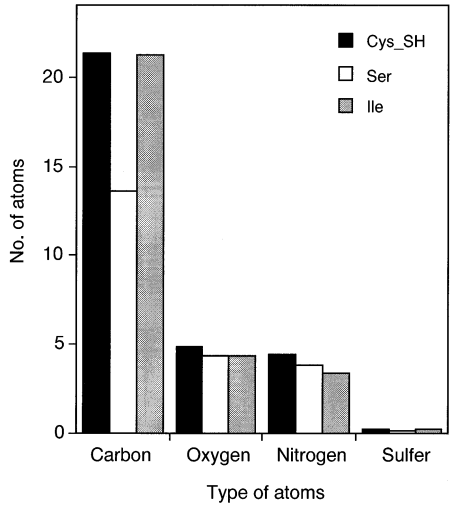

Fig. 1. The number of atoms observed in the vicinity of the side chain of either free cysteine (Cys_SH), Ser, or Ile residues. The average number of surrounding C, O, N and S atoms located within a distance of 6 Å from the Sγ atom of Cys_SH (black bar), the Oγ atom of Ser (white bar) and the Cδ atom of Ile (gray bar) are indicated, respectively. Those atoms belonging to the same central residue or its nearest neighbors were excluded from the counting.

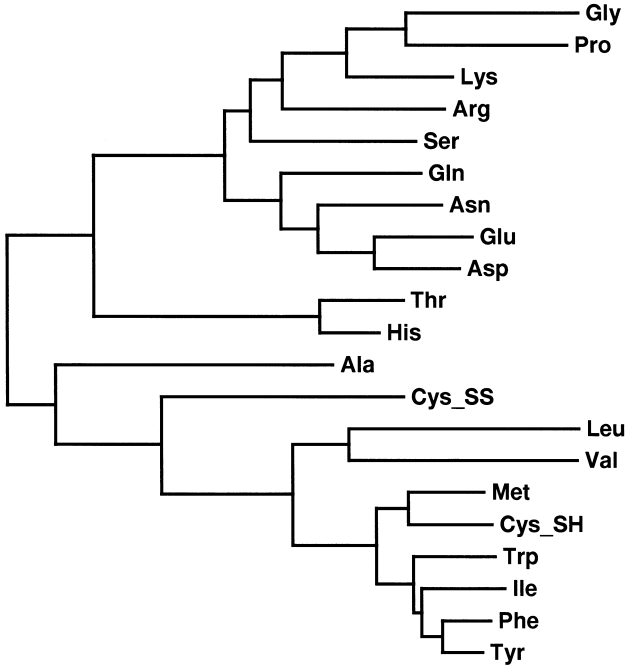

Fig. 2. Dendrogram showing the amino acid relationships. This dendrogram was computed for all proteins in the data set using 3D profiles [7] as follows. First, substitution scores S(x, y) were calculated from the 3D profile table, $p(i, x)$, as defined for amino acid type x at residue site i. Supposing x is the amino acid type of the native sequence at position i, substitution of x by y at position i is expressed as $p(i,y)-p(i,x)$. Subsequently, substitution scores, independent of site, were calculated by $S(x, y)=\langle p(i,y)-p(i,x)\rangle$, where $\langle\ \rangle$ implies that the average is taken over all sites in which the native amino acid type is x. Considering the 21 native amino acid types (i.e. distinguishing Cys_SS from Cys_SH), the form of $S(x, y)$ becomes 21×20 matrix. Next, the correlation coefficient between any two amino acids, x and z, was to be calculated using cor$[x, z]=\mathrm{cor}(S(x, y), S(z, y))$, where y varies among the 20 types of amino acid. However, considering $S(x, x)=S(z, z)=0$, the sum was taken for 19 pairs, i.e. 18 normal pairs having common y plus an additional pair of $S(x, z)$ and $S(z, x)$. The correlation coefficient so obtained was converted to a measure of distance between two amino acids (x and z) using dist$[x, z]=\arccos[\mathrm{cor}[x, z]]$. These distances were used to construct a dendrogram using the neighbor joining method [13].

the environment of the sulfhydryl group of Cys_SH is as hydrophobic as that of a methyl group of Ile.

3.3. Overall properties of amino acids empirically deduced from structural data

The present analyses have hitherto been conducted with a focus on the hydrophobicity of amino acids. This raises the question of whether free cysteine (Cys_SH) is more like aliphatic/aromatic amino acids, and less like serine (Ser), in a more general sense. To investigate this matter, we have employed an analysis similar to that of Taylor [4], where relationships (similarity and dissimilarity) among the amino acids, deduced from the Dayhoff mutation matrix [12], are presented as a Venn diagram. As the Dayhoff mutation matrix reflects various physico-chemical aspects of amino acids, the empirical relationship deduced from it must be more generally descriptive of the nature of the particular amino acids than just a hydrophobicity index. The Venn diagram obtained in the study by Taylor showed the relationship among 21 amino acid types, distinguishing Cys residues as being either Cys_SS or Cys_SH. Disulfide-bonding Cys_SS was categorized as (hydrophobic, small), while Cys_SH was categorized as (polar, hydrophobic, tiny). The relative position of Cys_SH was found to be close to that of Ser and Thr, leading to the traditional view that "the reduced form (Cys_SH) has similar properties to serine, while the oxidized form (Cys_SS) may be more equivalent to Val" [4].

It would be interesting to know if the same kind of amino acid relationship is still observed from the large amount of structural data now available. We have repeated the study of Taylor, just replacing the Dayhoff mutation matrix with the corresponding matrix obtained from the 3D profile table for proteins of known structure (see the legend of Fig. 2). The relationships among the amino acids obtained thus are represented as a dendrogram in Fig. 2. It is noted that the amino acids have clustered into two groups of polar and hydrophobic nature, corresponding respectively to the upper and lower halves of Fig. 2. Most importantly, Cys_SH is located in the hydrophobic cluster, amongst Met, Trp and Tyr, and is clearly separated from Ser and Thr which are in the polar cluster. Except for the location of Cys_SH, the overall relationships of amino acids in Fig. 2 seem consistent with those of Taylor's Venn diagram.

In conclusion, the present study has shown that free cysteine (Cys_SH) residues, as well as disulfide-bonding cystines (Cys_SS), behave like strongly hydrophobic residues in proteins. This apparent hydrophobic nature of Cys_SH seems to conflict with the polarized nature of the sulfhydryl group. However, we need to consider the fact that the sulfhydryl group is inactive toward water molecules. More specifically, unlike the hydroxyl group (-OH), the sulfhydryl group (-SH) has essentially no ability to form a hydrogen bond with water [4]. This may explain why the -SH group is both active in metal binding and disulfide bond formation, and is hydrophobic in water.

Acknowledgements: We are grateful to Thomas D. Andrews for the critical reading of the manuscript. This work was supported by a grant-in-aid from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Ziegler, D.M. and Poulsen, L.L. (1977) Trends Biochem. Sci. 2, 79–81.
[2] Thornton, J.M. (1981) J. Mol. Biol. 151, 261–287.
[3] Nishikawa, K., Kubota, Y. and Ooi, T. (1983) J. Biochem. (Tokyo) 94, 997–1007.
[4] Taylor, W.R. (1986) J. Theor. Biol. 119, 205–218.
[5] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[6] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[7] Ota, M. and Nishikawa, K. (1997) Protein Eng. 10, 339–351.
[8] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 211–222.
[9] Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. and Zehfus, M.H. (1985) Science 229, 834–838.
[10] Nakai, K., Kidera, A. and Kanehisa, M. (1988) Protein Eng. 2, 93–100.
[11] Tomii, K. and Kanehisa, M. (1996) Protein Eng. 9, 27–36.
[12] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O., Ed.), Vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Washington, DC.
[13] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.