Biophysical Chemistry 40 (1991) 293–301
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01577

Mechanism of protein folding

III. Disulfide bonding

Kazunori Watanabe *, Atsuhiko Nakamura, Yoshinori Fukuda and Nobuhiko Saitô
Department of Applied Physics, Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169, Japan

Received 29 November 1990
Revised manuscript received 1 February 1991
Accepted 5 February 1991

Lysozyme; Phospholipase; Erabutoxin; Disulfide bond; Lampshade (locus of H atom of S–H in cysteine residue)

It was shown in lysozyme and phospholipase, and generally in proteins with disulfide bonds, that after the formation of secondary structures the hydrophobic interactions between the key pairs responsible for folding tertiary structures bring several cysteine residues close together. Among the possible combination of cysteine residues some definite pairs are realized in the tertiary structure. In the Appendix to this paper an algebraic relation is given which must be satisfied for two cysteine residues to make a disulfide bond. This relation is too strict to be applied to real problems, where the two cysteines come close together, but the distance is still too great to yield a disulfide bond. In this case the two residues can attract each other by disulfide formation potential. A geometrical graphic representation is given for the locus of the H atom of the SH group in the cysteine residue. This looks like a lampshade and provides us with a guide to select the correct choice among cysteine pairs. This method is applied to lysozyme and phospholipase to supplement the discussion of the preceding paper (T. Yoshimura, H. Noguchi, T. Inoue and N. Saitô, Biophys. Chem. 40 (1991) 277).

1. Introduction

In the preceding papers [1,2] (referred to as I, and II, hereafter), especially in II, we presented a method of refolding lysozyme and phospholipase, where the disulfide bonding played a role to fasten the polypeptide chain. Assuming the formation of α-helices and β-sheets, we considered the hydrophobic interactions which are responsible for the construction of short-distance local structures. Then some of the pairs of cysteine residues came close together to easily form disulfide bonds. As mentioned in II, the force acting between two cysteines for the formation of a disulfide bond is rather long range, but this is not sufficient for the disulfide bonding, because it is a chemical bond, which requires certain geometrical relations to be satisfied (see Fig. 1). In fact as shown in II, there are several cysteines which lie within the range of interaction for possible disulfide bonding, but among them some selected pairs only are realized in the protein. For example in lysozyme at the 2300th step cysteines Cys 76, Cys 80 and Cys 94 should be considered because they are close together (Fig. 2). They were shown in Fig. 5 in II. The native disulfide bonds are S_1 (Cys 76–Cys 94) and S_2 (Cys 64–Cys 80), but the other possible pairs S_1', S_1'', S_2', and S_2'' are not bonded, although these four pairs are close respectively. In II, we introduced the potential (1) (see below) for disulfide bonding successively between the cysteine pairs S_1 and S_2, which are found in the native structure, anticipating the results discussed here. In the present paper we want to show how to

---

* To whom correspondence should be addressed.

0301-4622/91/$03.50

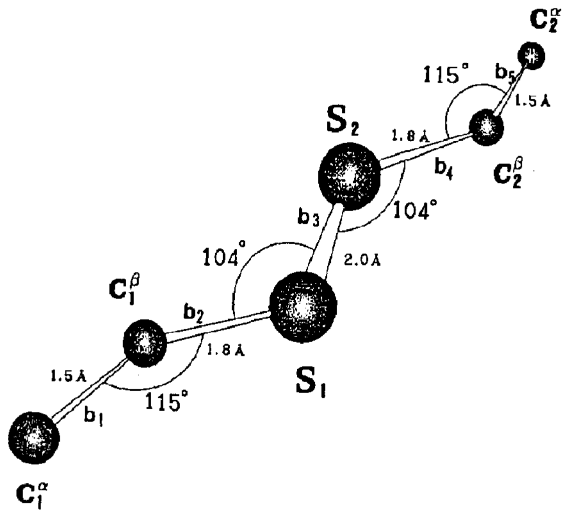


Fig. 1. Values of parameters for disulfide bond.

select the correct pairs for disulfide bonding among many possible candidates. A graphical representation of the geometrical condition is discussed in Section 2. Its applications to real proteins are given in Section 3 for lysozyme and in Section 4 for phospholipase. Section 5 is devoted to discussion. An algebraic relation is derived in the Appendix for a disulfide bond.

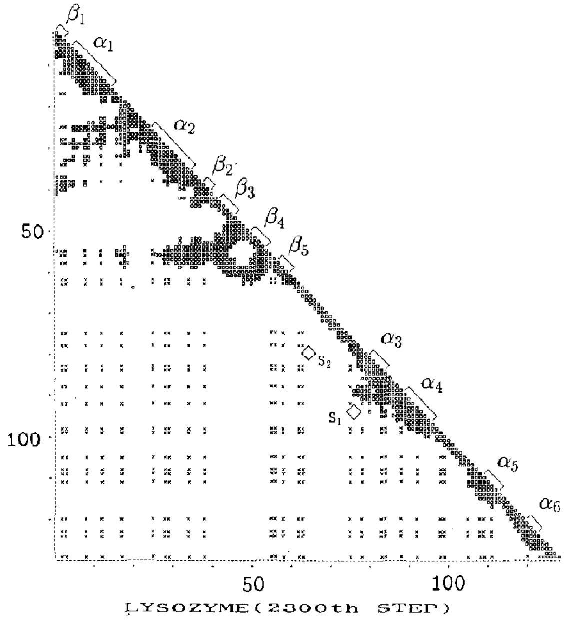


Fig. 2. A refold structure of lysozyme (2300th step without S–S).

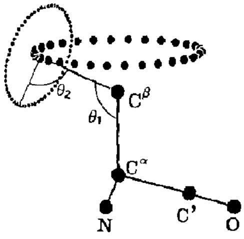


Fig. 3. Locus of H atom of SH in a cysteine (lampshade).

2. Graphical representation for disulfide bonding

While folding proceeds, two cysteine residues can be bonded only when the geometrical relation between the two is satisfied besides the distance of the two. In order to get an idea for the geometrical relation, let us look at Fig. 3 for the cysteine residue. The S atom lies on a cone around the C^α–C^β bond with bond angle θ_1 and H atoms lie on a cone around C^β–S with bond angle θ_2. Thus the combined locus of the H atom is a surface similar to a lampshade, a zone on a sphere of radius of C^β–H. In this figure and in the figures

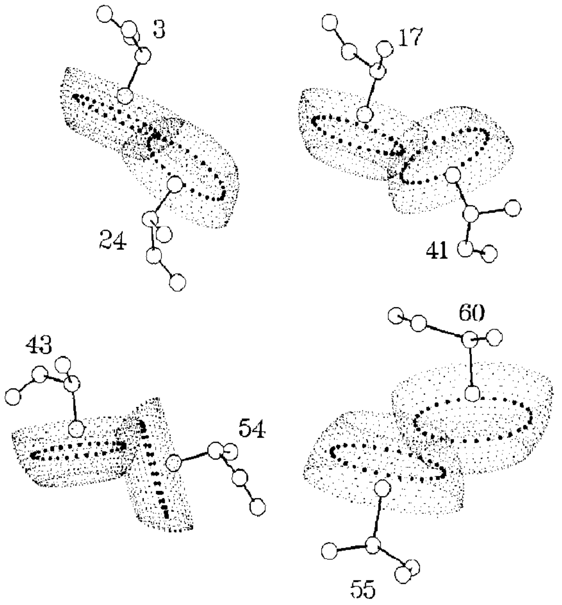


Fig. 4. Lampshades for bonded cysteines in erabutoxin.

to follow for lampshades the bond length SH is taken as 1.0Å, one half of S–S bond.

Figure 4 illustrates the loci of H atoms for bonded cysteines in native erabutoxin, as examples. Erabutoxin is composed of 62 amino acid residues and has eight cysteines and four disulfide bonds (Cys 3–Cys 24, Cys 17–Cys 41, Cys 43–Cys 54, Cys 55–Cys 60). Figure 5 shows the loci of the H atoms or the lampshades for nonbonded cysteine pairs which are located rather near (C^{α}–C^{α} < 10 Å). One can easily see that in Fig. 4 the lampshades of bonded cysteines intersect face to face to form disulfide bonds, while in Fig. 5, the lampshades of two cysteines not only are apart more or less for making a contact, but most of them are not face to face. One may notice, however, that the cysteine pairs (24–60) and (17–24) are face-to-face though not bonded. In the case of (24–60) the disulfide bond (60–55) is formed in the early stage of folding by the proximate rela-

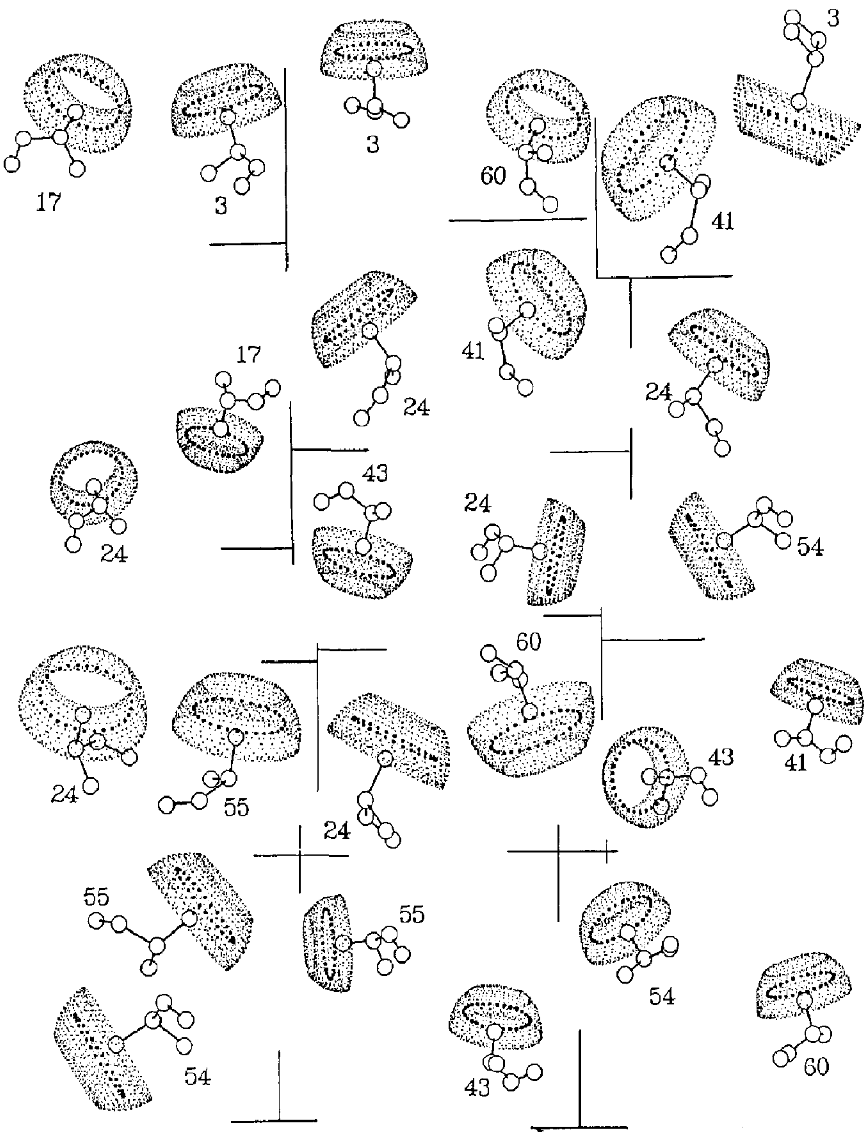


Fig. 5. Lampshades for nonbonded cysteine pairs in erabutoxin.

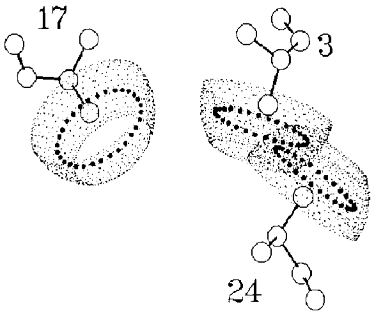


Fig. 6. Lampshades of Cys 3, 17 and 24 in erabutoxin.

tion of Cys 60 and Cys 55, and consequently the pair (24–60) cannot be bonded, while in the case of (17–24) Cys 3 is getting in between Cys 17 and Cys 24, thus preventing the disulfide bonding between 17 and 24, and allowing the 3–24 disulfide bond (Fig. 6). Other examples are given in Fig. 7 for ribonuclease A. The mutual relation of two lampshades is varied by the small changes of dihedral angles of the residues in between two cysteines. This variation is usually approximated by a small translation in certain direction. The intersection of the lampshades realized in this way is necessary, but furthermore face-to-face orientation of the lampshades (even when they are separated) is also necessary, because disulfide bonding is then possible without steric hindrance by parts of the main chain carrying the cysteines. When two lampshades are interacting, we have to study the geometrical conditions (as discussed in the Appendix) and steric hindrance of the main chain atoms. But in real protein, the face-to-face orientation of lampshades, at the intermediate folding state, seems to play a useful guide for the selection of cysteine pairs for disulfide bonding, as shown in Figs. 4, 5 and 6. This implies that the pairs of cysteines shown in Fig. 5 cannot bond, provided we take into account the situations in (24–60) and (17–24) mentioned above. These studies indicate that the correct disulfide pairs can be identified by drawing the lampshades of the cysteines at the folding stage where they come close but did not form a disulfide bond, while packing the secondary structures of a polypeptide chain by the method of introducing appropriate hydrophobic interactions as already explained in the previous paper I and II. We can introduce the potential, E, between the relevant cysteines as function of their distance, r (in Å), as

$$E = \begin{cases} 20 & r < 3.2 \\ -10\left(1-(r-4.2)^2\right) & 3.2 \le r < 5.0 \\ -3.6\left(1-(r-5.0)^2/25\right) & 5.0 \le r < 10.0 \\ 0 & 10.0 \le r \end{cases} \tag{1}$$

(see also in II) provided that the lampshades are face-to-face and that it is easy to make an intersection, but otherwise we do not have to consider the energy (1), because they cannot form a disulfide bond. This program will be performed on lysozyme and on phospholipase in Sections 3 and 4.

3. Disulfide bonding in lysozyme

Lysozyme has four disulfide bonds S_1, S_2, S_3 and S_4 as mentioned in Section 1. An intermediate structure at the 2300th step is shown in Fig. 2 (One step means a process of searching for the state of minimum energy by the method of Bremermann described in I by changing a set of

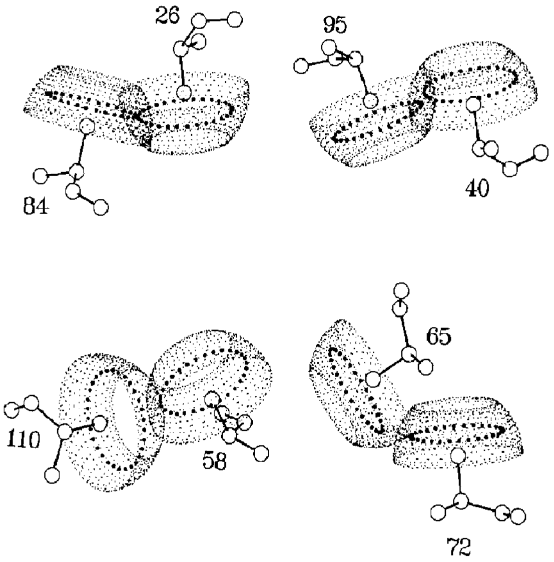
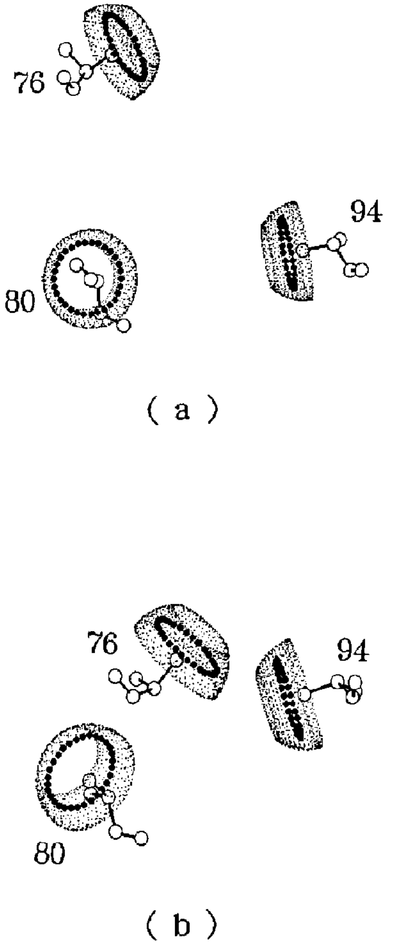


Fig. 7. Lampshades of bonded cysteine pairs in ribonuclease A.



(a)

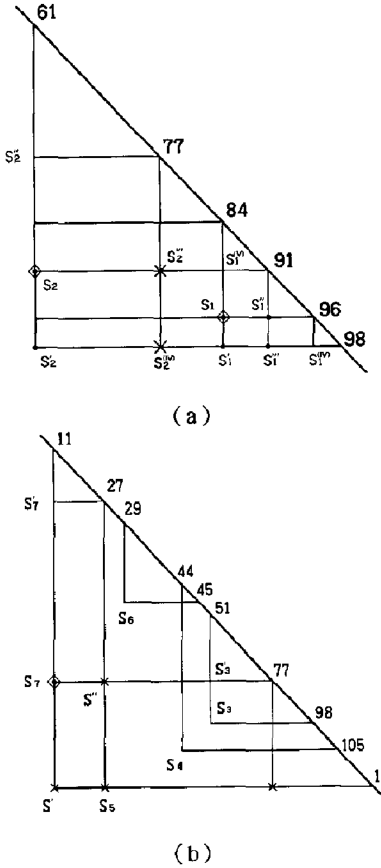
(b)

Fig. 8. (a) and (b). Lampshades for Cys 76, 80 and 94 (2300th step) from two different directions. It can easily be seen that Cys 76 and Cys 94 are face-to-face positioned.

dihedral angles of selected amino acid residues. Usually it takes 3×10^{-3} s/residue for computation by HITAC, Computer Center, University of Tokyo) One sees that Cys 76, 80 and 94 are rather close to each other. Figure 5 in II showed the six possible pairs of these cysteines. The native ones are S_1, S_2. On looking at Fig. 2, the distances between α-carbons for S_1, S_1' and S_1'' are about 15Å, and we have to select first which one is the native among these three pairs. Then we draw the lampshades for these three pairs. They are shown in Fig. 8. One can clearly select S_1 whose shades are face to face to form a disulfide bond, while the shades of the other S' and S'' turn to other directions making it difficult to form disulfide bonds. Then we introduce the disulfide bonding energy (eq. 1) between the cysteines of S_1 pairs. These bonds can fasten the chain, and shorten the distance of S_2 pairs, so one may examine the lampshades of S_2, S_2' and S_2''. The remaining pair, however, clearly is S_2 and we therefore introduce the disulfide bonding energy for the S_2 pair only. The final result is given in Fig. 6 of II (6900th step).

4. Phospholipase

Phospholipase has seven disulfide bonds. As described in II, we first fold the parts between β_1, β_2 and α_5. The possible cysteine pairs are shown in Figs. 9 (a) and (b). At the 400th step, (see also Fig. 10 in II) the mutual relations of Cys 84, 91,



(a)

(b)

Fig. 9. (a) and (b). Cysteine pairs of disulfide bonds in phospholipase.

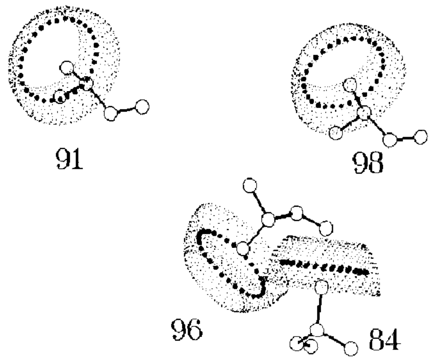


Fig. 10. Lampshades of Cys 84, 91, 96 and 98 (400th step).

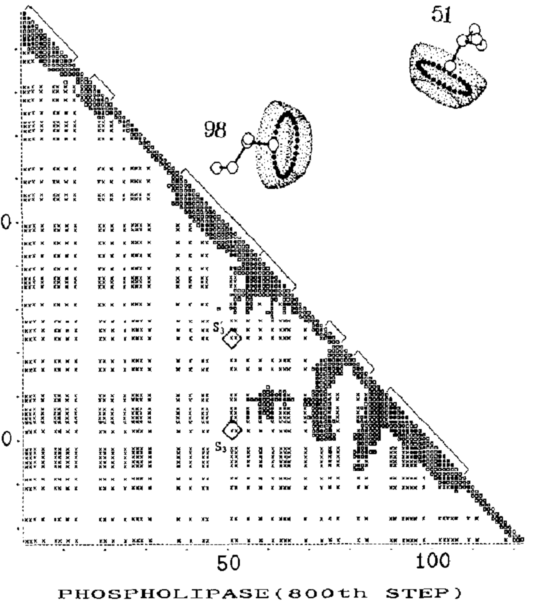


Fig. 12. Distance map and lampshades of Cys 51 and 98 for S_3 (800th step).

96 and 98 for S_1, S_1', S_1''', and $S_1^{(IV)}$ are illustrated in Fig. 10. The S_1 pair (84–96) only can form a disulfide bond. Then we introduce the energy (eq. 1) between the S_1 pair cysteines. At the 600th step, the cysteine pairs S_2''(61–77), S_2'''(77–91) and $S_2^{(IV)}$(77–98) see (Fig. 9a) are too separated to form disulfide bonds, the distances between the centers of cysteines being 27.6Å, 22.5Å, and 19.9Å, respectively. Since S_1(84–96) is bonded, we have only to consider S_2, S_2' and S_1''' pairs. They are illustrated in Fig. 11. The S_2 pair is found easy to be bonded and thus the energy equation (eq. 1) can be applied to Cys 61 and Cys 91. At the 800th step, (see Fig. 12) Cys 51 is separated more than 18Å from Cys 77, 84 and 96, and only S_3(51–98) pair is still close enough together (15.3Å) and has

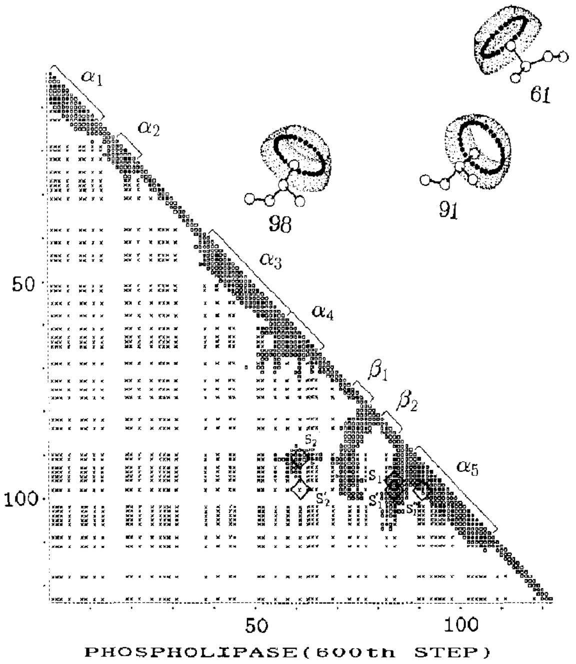


Fig. 11. Distance map and lampshades of Cys 61, 91 and 98 (600th step).

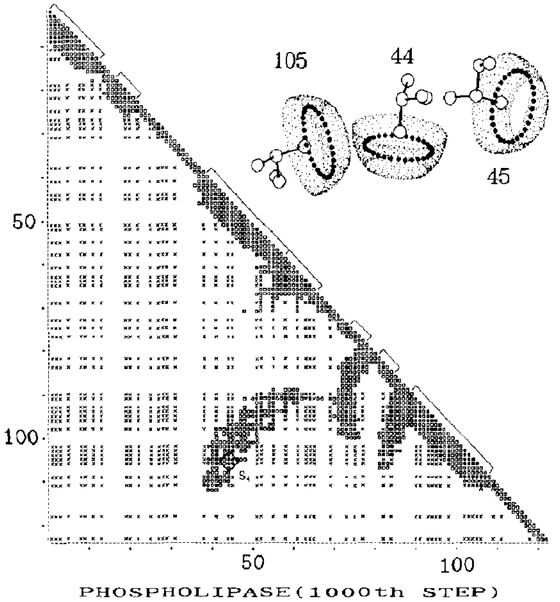


Fig. 13. Distance map and lampshades of Cys 44, 45 and 105 for S_4 (1000th step).

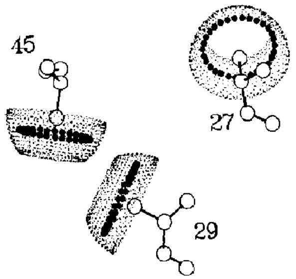


Fig. 14. Lampshades of Cys 27, 29 and 45 (1600th step).

a good mutual relation for disulfide bonding (Fig. 12). Again the energy equation (eq. 1) can be applied. At the 1000th step, the S_4(44–105) pair only is close enough and has a good mutual relation as shown in Fig. 13, and the disulfide bond S_4 is thus completed. At the 1600th step, the S_5(27–123) pair is too separated (53Å), and the lampshades for Cys 27, 29 and 45 are shown in Fig. 14. We see that S_6(29–45) pair can be bonded. At the 1800th step, the only remaining free cysteines and possible candidates are 11, 27, 77 and 123, of which only the pairs S_7(11–77) and S_7'(11–27), as shown in Fig. 15, are the real candidates, since S'(11–123), S''(27–77) and S_5(27–123) are too separated. The lampshades for S_7 and S_7' are shown in Fig. 15. One safely select S_7 pair. The S_7 pair can be bonded at the 2000th step, as was shown in Fig. 12 of II, but the S_5 pair still remains unbonded.

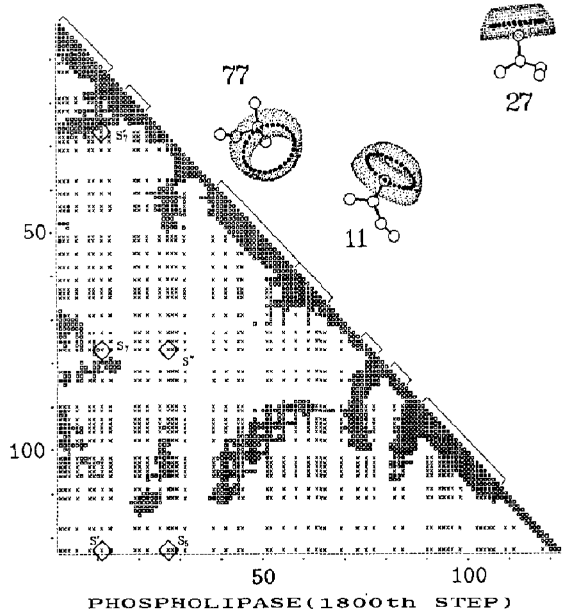


Fig. 15. Distance map and lampshades of phospholipase (1800th step).

5. Discussion

The disulfide formation is possible only when the appropriate conditions for distance and the orientational relation are satisfied, as mentioned in the Introduction. This consideration is necessary when we discuss the process of protein folding as discussed above. In addition to this, an artificial insertion of disulfide bonds into proteins of known three-dimensional structure is supposed to enhance the stability of proteins. Thus a method of selection of sites where a disulfide bond could be engineered for protein modification has attracted much attention. For example, Pabo et al. [3] and Pantoliano et al. [4] employed the crystallographic data of proteins to accomodate a disulfide bond into a protein of known three-dimensional structure. This method requires a sufficiently large size of protein data. On the other hand Hazes and Dijkstra [5] noticed that when a disulfide bond formation is possible, the positions of S atoms should be determined by the coordinates of N, α-carbon, and carbonyl carbon of two cysteines, and they have developed a computer program for this purpose. These methods, however, are not applicable for selecting proper cysteine pairs during the process of folding where the two cysteines come close together by the formation of the local structure through hydrophobic interactions as described in I and II, but are still too far apart for chemical bonding. We have presented a useful guide on how to select the correct pairs of cysteines to form a disulfide bond, by drawing lampshades around the cysteine residues. In the present studies we assumed that the cysteine residue is rigid sphere situated in the direction of C^{β}–C^{α} bond with a radius of 2.36 Å (see Table A.I. in I). Therefore no consideration was made for the determination of the positions of S atoms. To do this the computer programs developed by several authors cited above [3–5] will be useful, or the algebraic relation given in the Appendix can be employed to search for the plausible values of the rotational angles around C^α, C^β and S.

This work was supported by a Special Coordination Fund from the Science and Technology Agency of Japan and by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

Appendix

Algebraic relation for disulfide bonding

When we fold the polypeptide chain in accordance with the general principle of protein architecture by considering hydrophobic interaction mentioned in the previous papers I [1] and II [2], the coordinates of all the atoms in the main chain together with the β-carbons can be determined. In this situation a certain pair of cysteine residues happen to come close to each other with their α- and β- carbons at the specified positions. Then we pose a problem to provide a criterion whether or not the two cysteine residues really make a disulfide bond, satisfying geometrical restrictions of the chemical bonds (Fig. 1), when the coordinates of C^α and C^β of both cysteines are known. The five chemical bonds connecting two α-carbons C_1^α and C_2^α) are denoted by vectors $\boldsymbol{b}_i$ $(i = 1, 2, \ldots, 5)$ and let every chemical bond (say i) be attached by its own coordinate system (i) with z_i-axis in the direction of the ith bond, x_i-axis in the plane determined by the ith bond and the preceding $i-1$th bond (For $i=1$, N–C_1^α bond is taken as the preceding bond). The origin is taken at the connecting atom of $i-1$th and ith bonds. The azimuthal angle ϕ_i of z_{i+1}-axis with respect to the coordinate system (i) is measured from the trans position around z_i-axis shown in Fig. A.1. Then the vector $\boldsymbol{l}_{i+1}$ $(l_{i+1}, m_{i+1}, n_{i+1})$ in $i+1$-system is transformed into a vector $\boldsymbol{l}(l_i, m_i, n_i)$ in i-system by the relation. (see Fig. A.1)

$$\boldsymbol{l}_i = \mathbf{T}_i' \boldsymbol{l}_{i+1}$$

$$\mathbf{T}_i' = \begin{pmatrix} \cos\theta_i \cos\phi_i & \sin\phi_i & \sin\theta_i \cos\phi_i \\ \cos\theta_i \sin\phi_i & -\cos\theta_i & \sin\theta_i \sin\phi_i \\ \sin\theta_i & 0 & -\cos\theta_i \end{pmatrix} \tag{A.1}$$

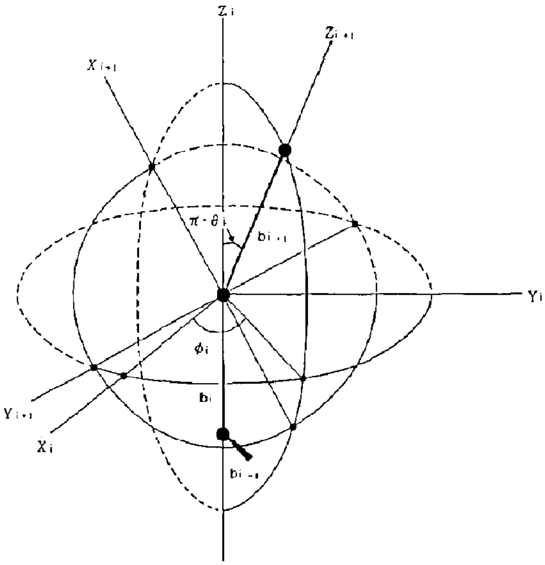


Fig. A.1. Coordinate systems.

where θ_i is the bond angle between ith and $i+1$th bonds. However, one notice that in this representation the orientation of the bonds are correctly described, but the shift of the origin is not. To give the coordinates in i-system of the point represented by the vector $\boldsymbol{l}_{i+1}$ of $i+1$-system, we augment the matrix $\mathbf{T}'$ by one row and one column

$$\mathbf{T}_i = \begin{pmatrix} & & & \alpha \\ & T_i' & & \beta \\ & & & \gamma \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.2}$$

where α, β, and γ are the components of the shift of the origin. Then the vector $(l_{i+1}, m_{i+1}, n_{i+1}, 1)^t$ in the $i+1$-system is transformed to

$$\begin{pmatrix} l_i+\alpha \\ m_i+\beta \\ n_i+\gamma \\ 1 \end{pmatrix} = \mathbf{T}_i \begin{pmatrix} l_{i+1} \\ m_{i+1} \\ n_{i+1} \\ 1 \end{pmatrix} \tag{A.3}$$

where $(l_i+\alpha,\ m_i+\beta,\ n_i+\gamma)$ are the coordinates in i-system of the point $(l_{i+1}, m_{i+1}, n_{i+1})$ in the $i+1$-system. This representation is useful for the present purpose. Let the transformation matrix from 1-sytem to (xyz)-system be

$$\mathbf{T}_0 = \begin{pmatrix} & & & x_0 \\ & \mathbf{T}_0' & & y_0 \\ & & & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.4}$$

where $\mathbf{T}_0'$ is an orthogonal transformation, and (x_0, y_0, z_0) are the coordinates of the origin (C_1^α) of 1-system. By virtue of these transformation, the vector $(0, 0, \alpha)$ in $i = 5$-system can be described in (x, y, z)-system as

$$\begin{pmatrix} x_5 \\ y_5 \\ z_5 \\ 1 \end{pmatrix} = \mathbf{T}_0\mathbf{T}_1\mathbf{T}_2\mathbf{T}_3\mathbf{T}_4 \begin{pmatrix} 0 \\ 0 \\ a \\ 1 \end{pmatrix} \tag{A.5}$$

where (α, β, γ) in these matrices are

$$\begin{array}{ll} (0, 0, b) & \text{for } T_4 \text{ and } T_2 \\ (0, 0, c) & \text{for } T_3 \\ (0, 0, a) & \text{for } T_1 \\ a = 1.5, & b = 1.8,\ c = 2.0 \end{array} \tag{A.6}$$

respectively (see Fig. 1), and x_5, y_5, z_5 are the coordinates of C_2^α in the (x, y, z)-system. The relation contains four independent variables ϕ_1, ϕ_2, ϕ_3 and ϕ_4. If there exists no solution for real ϕ's, then disulfide bonding between these two cysteines is impossible. We do not enter into this representation any more, because this strict condition is not practical in real cases where the relative positions of two cysteines are easily changed. A more practical method is presented in Section 2. However, eq. (A.5) can be used to check the final result obtained by the procedure described in Section 2 to see whether or not the selected pairs can form a disulfide bond, when neccessary.

References

1 N. Saitô, T. Shigaki and M. Yamamoto, Proteins: Struct. Funct. Genet. 3 (1988) 199.

2 T. Yoshimura, H. Noguchi and T. Inoue and N. Saitô, Biophys. Chem. 40 (1991) 277.

3 C.O. Pabo and E.G. Suchanek, Biochemistry 25 (1986) 5987.

4 M.W. Pantoliano, R.C. Ladner, P.N. Bryan, M.L. Rollence, J.F. Wood, and T.L. Poulos, Biochemistry 26 (1987) 2077.

5 B. Hazes and B.W. Dijkstra, Protein Eng. 2 (1988) 119.