Protein Engineering vol.3 no.2 pp.95–103, 1989

Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis

R.Sowdhamini[1], N.Srinivasan[1], Brian Shoichet[2], Daniel V.Santi[2,3], C.Ramakrishnan[1] and P.Balaram[1,4]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, and Departments of [2]Pharmaceutical Chemistry and [3]Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94143, USA

[4]To whom correspondence should be addressed

A computer modeling procedure for assessing the stereochemical suitability of pairs of residues in proteins as potential sites for introduction of cystine disulfide crosslinks has been developed. Residue pairs with $C^{\alpha}-C^{\alpha}$ distances of $\leq$6.5 Å and $C^{\beta}-C^{\beta}$ distances of $\leq$4.5 Å are chosen for geometrical fixation of S atoms using the program MODIP. The stereochemistry of the modeled disulfides is evaluated using limits for the structural parameters of the various torsion angles and S–S bond length in the disulfide bridge. The ability of the procedure to correctly model disulfides has been checked with examples of cystine peptides of known crystal structures and 103 disulfide bridges from 25 available protein crystal structures determined at $\leq$2 Å resolution. An analysis of results on three proteins with engineered disulfides, T4 lysozyme, dihydrofolate reductase and subtilisin, is presented. Two positions for the introduction of 'stereochemically optimal' disulfides are identified in subtilisin.

Key words: disulfide bonds/disulfide stereochemistry/computer modeling of disulfides/site-directed mutagenesis

Introduction

The introduction of disulfide bonds into proteins by site-directed mutagenesis affords a means of stabilizing native, folded conformations (Villafranca *et al.*, 1983; Perry and Wetzel, 1984; Wetzel, 1987). Covalent disulfide cross-links are a convenient means of limiting mobility of specific segments of a polypeptide chain (Anfinsen and Scheraga, 1975; Creighton, 1988). Recent reports of incorporation of new disulfide bonds into proteins include the examples of subtilisin (Wells and Powers, 1986; Pantoliano *et al.*, 1987; Mitchinson and Wells, 1989), T4 lysozyme (Matsumara and Matthews, 1989; Wetzel *et al.*, 1988), dihydrofolate reductase (Villafranca *et al.*, 1983, 1987) and λ-repressor (Sauer *et al.*, 1986). The choice of the pair of residues for replacement by Cys is of primary importance in determining the structural characteristics of the mutant protein. In general, for purposes of stabilization, the disulfide link must be introduced at positions which are stereochemically optimal for formation of an unstrained cross-link. In this report we outline a procedure which conveniently permits an evaluation of the various possibilities for incorporation of disulfide bridges into proteins of known crystal structure. This approach differs from that reported by Pabo and Suchanek (1986) but is related to that described by Hazes and Dijkstra (1988), which appeared subsequent to the completion of this study.

Materials and methods

The definition of the various dihedral angles in a cystine disulfide bridge is indicated in Figure 1(a). The standard parameters used in all calculations, obtained by the averaging of known cystine peptide structures, are $C^{\beta}-S$ ($r_{\beta S}$) = 1.87 Å, S_i-S_j (r_{SS}) = 2.04 Å, $C^{\alpha}C^{\beta}S$ = 114° and $C^{\beta}S_iS_j$ = 104°.

Stereochemical fixing of S atoms

The necessary geometrical relationships are indicated in Figure 1(b). The coordinates of C_i^{α}, C_i^{β}, C_j^{α} and C_j^{β} were obtained from published peptide crystal structures or from the Protein Data Bank (Bernstein *et al.*, 1977), using only structures refined to a resolution of $\leq$2 Å.

The distances $r_{\alpha_i S_i}$ and $r_{\beta_j S_i}$ can be computed using the known values for $r_{\alpha\beta}$, $r_{\beta S}$ and r_{SS}. The atom S_i will lie on the surface of a sphere with C_i^{β} as the center and $r_{\beta S}$ (known) as the radius. In addition S_i must lie on the surface of a sphere with C_i^{α} as the center and radius $r_{\alpha_i S_i}$. The intersection surface for these two spheres is a circle. Thirdly S_i must also lie on the surface of a sphere with C_j^{β} as center and radius $r_{\beta_j S_i}$. This sphere and the earlier circle can intersect, permitting fixing of atom S_i in space. Three situations are possible: (i) no intersection (S_i fixing not possible); (ii) intersection at one point only (rare); and (iii) intersection at two points (most common). Similarly, there will generally be two positions for S_j, leading to four possible S–S geometries.

The parameters r_{SS}, χ_{SS} and $\chi^1_{i,j}$ are computed for all four

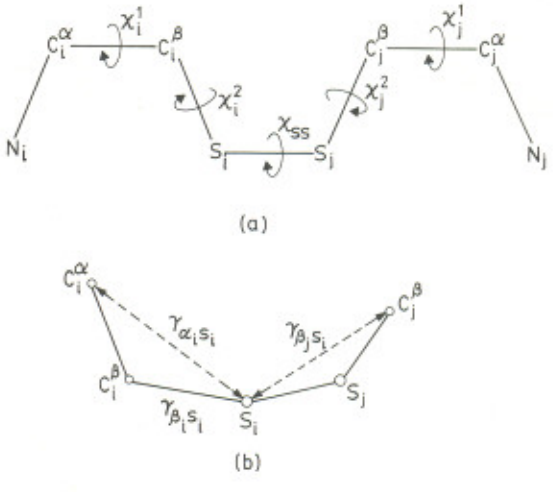


Fig. 1. (**a**) Definition of the various dihedral angles in a cystine disulfide bridge. (**b**) Definition of interatomic distances used in geometrical fixing of S atoms.

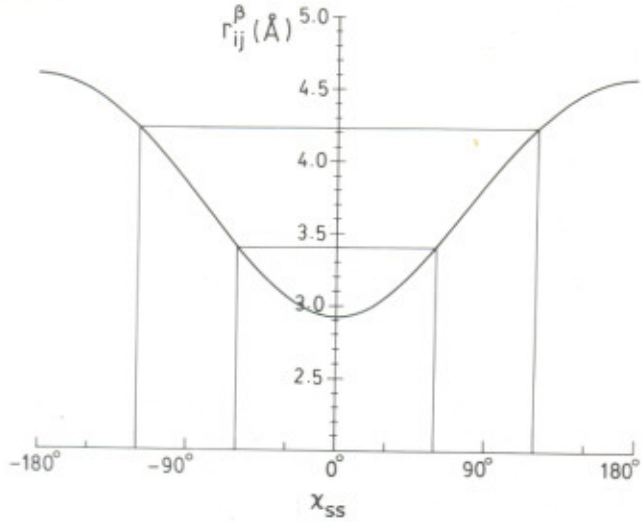


Fig. 2. Variation of the C_i^β–C_j^β distance (r_{ij}^β) as a function of the dihedral angle C_i^β–S_i–S_j–C_j^β (χ_{SS}). Vertical lines mark the range of acceptable χ_{SS} values, $\pm 90°$ ($\pm 30°$), while horizontal lines indicate range of acceptable r_{ij}^β values.

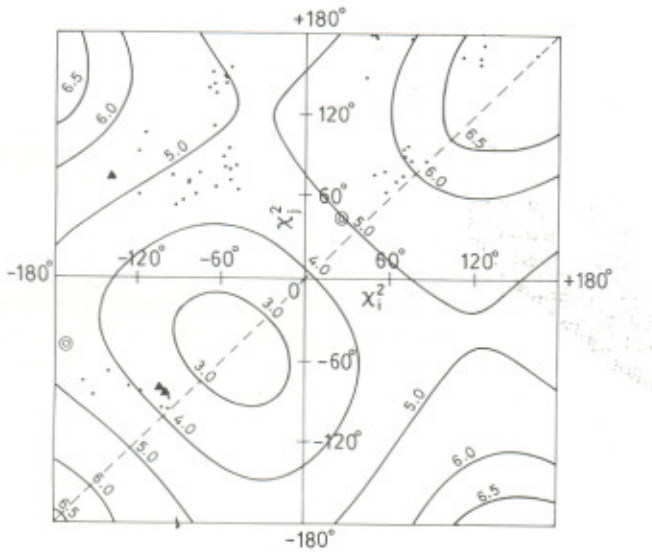


Fig. 3. Variation of the C_i^α–C_j^α distance (r_{ij}^α) as a function of the dihedral angles C_i^α–C_i^β–S_i–S_j (χ_i^2) and C_j^α–C_j^β–S_j–S_i (χ_j^2) for a fixed χ_{SS} value of 90°. Contours of equal r_{ij}^α are drawn at 1 Å intervals. Experimentally observed χ^2 values are represented on the plot. ● Protein structures, ▲ peptide structures. ⓒ Values in engineered protein disulfides subtilisin 24–87 and T4 lysozyme 3–97 (see text and Table VI). The plot is symmetric since χ_i^2 and χ_j^2 are indistinguishable in the calculation. Experimental points are therefore represented only on one side of the diagonal.

cases. Stereochemical limitations imposed on the modeling procedure are considered in the subsequent section.

Results and discussion

Figure 2 shows a plot of the distance C_i^β–C_j^β (r_{ij}^β) as a function of the disulfide dihedral angle χ_{SS}. For peptide and protein disulfides, crystallographically observed χ_{SS} values lie largely around $\pm 90°$ ($\pm 30°$). This limits the range of acceptable r_{ij}^β

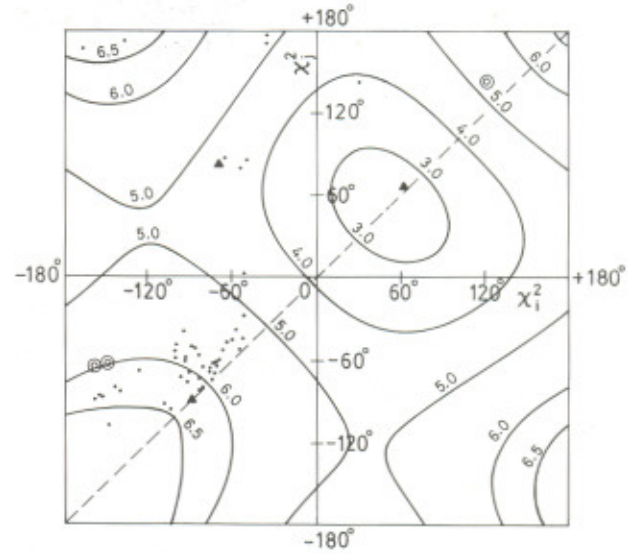


Fig. 4. Same as in Figure 3, for a fixed χ_{SS} value of $-90°$. ⓒ indicates engineered disulfides in dihydrofolate reductase 39–85 (two independent molecules) and subtilisin 22–87 (see text and Table VI).

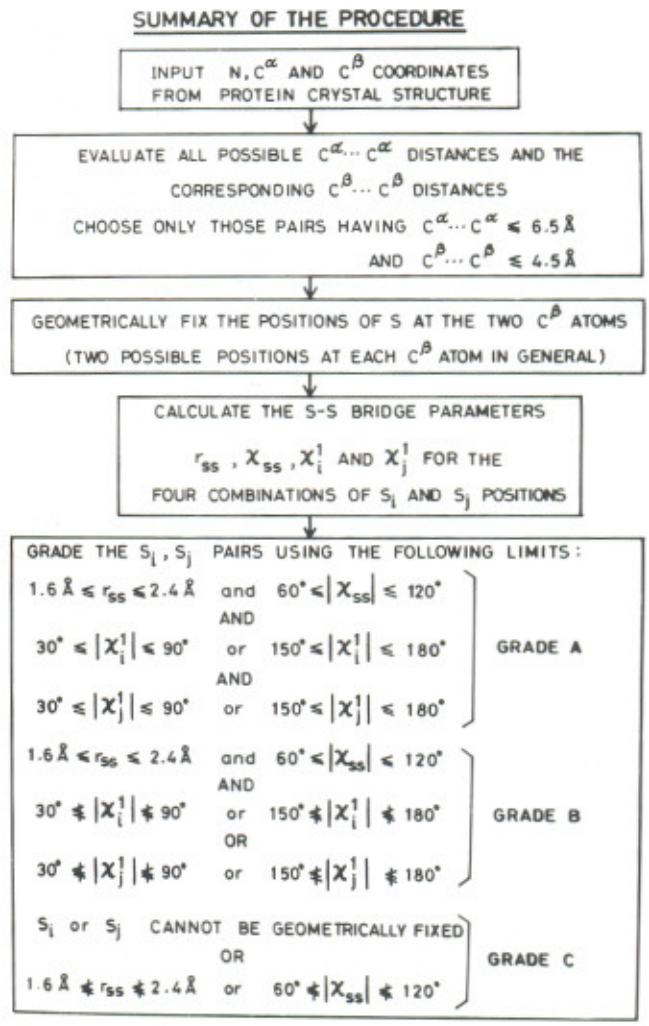


Fig. 5. Summary of modeling procedure.

Table I. Comparison of observed and modeled disulfides[a]: protein trypsin inhibitor (5PTI)

Cys5	Cys55						
	r_{SS}(Å)	χ_{SS}(°)	χ_i^1(°)	χ_j^1(°)	χ_i^2(°)	χ_j^2(°)	
Observed	2.0430	−82.816	−62.223	−64.480	−75.496	−65.953	
Combination1	2.1226	−75.778	−56.922	−63.941	−79.629	−68.294	**
Shift of S atoms (Å)	SI = 0.1728	SJ = 0.1414					
Combination 2	3.5516	160.848	−56.922	167.317	−124.837	93.618	−*
Shift of S atoms (Å)	SI = 0.1728	SJ = 2.9708					
Combination 3	3.5014	151.701	169.664	−63.941	100.260	−118.196	−*
Shift of S atoms (Å)	SI = 3.0073	SJ = 0.1414					
Combination 4	1.660	.758	169.664	167.317	114.504	103.168	−*
Shift of S atoms (Å)	SI = 3.0073	SJ = 2.9708					

[a]Symbols in the last column indicate quality of modeled disulfide. ** corresponds to a grade A case, *− to a grade B case and −* and −− to a grade C case. See text and Figure 5 for further discussion of grading scheme.

to 3.4−4.25 Å. The distance r_{ij}^{α} (C_i^{α}−C_j^{α}) depends on the dihedral angles χ_i^2, χ_j^2 and χ_{SS}. For fixed χ_{SS} values r_{ij}^{α} can be represented in the form of a two-dimensional χ_i^2, χ_j^2 contour map. Such maps have been computed for fixed values of χ_{SS} of ±70, ±90 and ±110°. The results for $\chi_{SS} = +90°$ and $\chi_{SS} = -90°$ are illustrated in Figures 3 and 4 respectively. The plots obtained using χ_{SS} values of ±70 and ±110° are qualitatively similar. The experimentally determined χ_i^2, χ_j^2 values for representative cystine bridges in peptides and proteins are also indicated in Figures 3 and 4. A clustering of points in the region $\chi_{i(j)}^2 \simeq -80°$ and $\chi_{j(i)}^2 \sim -70°$ is observed. It is seen that an upper limit of ≤6.5 Å may be safely placed on the acceptable r_{ij}^{α} distance. Thus, the elementary condition for considering residue positions in proteins, as potential sites for cysteine introduction, to generate *unstrained disulfides* is $r_{ij}^{\alpha} \leq 6.5$ Å; $r_{ij}^{\beta} \leq 4.5$ Å. Note that the upper limit is slightly more than that obtained from Figure 2. Minor bond angle distortions can then be accommodated. The mean bond angles obtained from 11 amino acid and peptide structures are C^{α} C^{β} S 114.1 ± 2.7° (range 106.8−119.1°) and C^{β} S S 104.2 ± 1.9° (range 100.0−107.9°). The corresponding values from a data set of 72 disulfides in non-homologous proteins are 114.1 ± 3.1° (range 106.1−122.7°) and 104.0 ± 3.1° (range 92.6−113.3°).

Figure 5 summarizes the procedure used to model disulfide bridges which has been coded into the computer program MODIP (Modelling Of Disulfides In Proteins), which runs on an IBM PC-XT or compatible system. A listing of all pairs of residues in a protein of known crystal structure which satisfy the above distance criteria may be obtained as the first step in this program. The feasibility of inserting an S−S bridge of satisfactory stereochemistry between the chosen C^{β} positions is then examined. Theoretically, four possible S−S links may be obtained, corresponding to the two distinct positions of each S atom. Computation of the cystine bridge stereochemical parameters (r_{SS}, χ_{SS}, χ_i^1, χ_j^1, χ_i^2 and χ_j^2) then permits an evaluation of the geometry of the modeled S−S bridge. The limiting ranges for the parameters used in determining the acceptability of a particular S−S bridge are as follows:

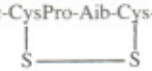
$$r_{SS} = 1.6-2.4 \text{ Å}, \quad |\chi_{SS}| = 60-120°$$

$$|\chi_i^1| = 30-90° \text{ or } 150-180°$$

Table II. Comparison of the best modeled (M) disulfides with observed (O) structures for cystine peptides.

Peptide[a]		r_{SS}(Å)	χ_{SS}(°)	χ_i^1(°)	χ_j^1(°)	χ_i^2(°)	χ_j^2(°)
1	M	2.00	96.4	−47.6	−62.9	−105.2	−80.5
	O	2.03	99.0	−49.0	−61.0	−104.0	−81.0
2	M	1.98	83.6	75.9	47.2	66.0	−146.0
	M	1.87	75.0	−169.5	−70.0	−135.6	76.4
	O	2.04	81.5	−169.1	−71.3	−138.7	73.3
3	M	2.20	107.8	−52.0	−59.9	−106.0	−89.9
	O	2.03	101.0	−55.0	−55.0	−101.0	−84.0
4	M	2.30	−100.0	42.1	37.7	67.9	73.5
	O	2.02	−93.5	47.5	44.1	61.6	66.2
5	M	1.87	−64.1	−60.3	−60.3	155.6	155.6
	M	2.08	−78.8	70.0	70.0	−89.3	−89.3
	O	2.04	−82.6	68.7	68.7	−88.4	−88.4
6	M	2.25	−79.3	−50.5	−82.0	−83.5	−69.0
	O	2.04	−79.7	−54.8	−85.3	−82.3	−70.0
7[b]	M	2.40	102.60	163.25	178.92	−119.00	−171.77
	M	2.10	−89.33	58.22	178.92	119.17	−98.00
	O	2.04	−89.68	59.23	177.41	117.16	−97.89

[a]1, (Boc-Cys-Ala-Cys-NHMe)$_2$ (Karle *et al.*, 1989);
2, Boc-CysPro-Aib-Cys-NHMe (Ravi *et al.*, 1983);



3, Boc-Cys-Val-Aib-Ala-Leu-Cys-NHMe (Karle *et al.*, 1988);

4, cyclo (Cys-Cys) (Varughese *et al.*, 1981); 5, cystine.2HCl (Jones *et al.*, 1974); 6, cystine.2HBr dihydrate (Rosenfield and Parthasarathy, 1975); 7, cyclo (Cys-Gly-Pro-Phe)$_2$ (Kopple *et al.*, 1988).

[b]Coordinates for the observed structure were calculated from the published internal parameters.

and

$$|\chi_j^1| = 30-90° \text{ or } 150-180°$$

No constraints were applied on $\chi_{i(j)}^2$ in view of the large spread of observed χ^2 values in native disulfides.

In order to test that this procedure successfully modeled S−S bridges, cystine peptides of known crystal structure and native disulfide bonds in proteins, for which high resolution crystal structures and atomic coordinates are available, were considered.

Table III. Comparison of the best modeled (M) disulfides with observed (O) structures for an illustrative set of protein disulfides

Protein	Disulfide bonded residues		r_{SS}(Å)	χ_{SS}(°)	χ_i^1(°)	χ_j^1(°)	χ_i^2(°)	χ_j^2(°)
Trypsin inhibitor	5–55	(M)	2.12	−75.8	−56.9	−63.9	−79.6	−68.3
		(O)	2.04	−82.8	−62.2	−64.5	−75.5	−66.0
	14–38	(M)	2.29	99.4	−76.1	66.3	102.3	−120.1
		(O)	2.03	95.1	−71.9	61.0	105.9	−114.3
	30–51	(M)	2.34	−98.4	−61.3	179.8	−98.0	−93.8
		(O)	2.02	−89.9	−72.3	178.1	−102.8	−95.9
Human lysozyme	6–128[a]	(M)	2.33	81.0	170.3	−175.0	44.1	38.8
		(M)	2.14	−72.0	−70.9	−60.2	−46.2	−40.9
		(O)	2.07	−65.3	−68.9	−60.0	−51.8	−44.3
	30–116[a]	(M)	2.15	−101.0	−78.6	−166.4	166.2	146.3
		(M)	2.19	−102.5	−172.0	−57.4	−91.7	−71.8
		(O)	2.06	−95.7	−175.4	−58.1	−95.7	−72.3
	65–81[b]	(M)	sulfur cannot be fixed					
		(O)	2.08	95.0	62.6	−71.0	81.2	−58.1
	77–95	(M)	1.91	75.2	−64.8	−177.6	−176.9	50.7
		(O)	2.04	83.0	−71.0	−176.3	−178.0	48.3

[a]In these cases two acceptable disulfide bridges were obtained by modeling. One of them closely corresponds to the experimentally determined stereochemistry.
[b]This disulfide can be satisfactorily modeled upon relaxation of the angle $C^{\beta}SS$. See Table VI and text.

Using only $N_{i,j}$, $C^{\alpha}_{i,j}$ and $C^{\beta}_{i,j}$ coordinates of non-glycyl residues as inputs to the MODIP program, the modeled S−S positions were compared with experimentally determined values. Table I shows a typical partial listing, obtained for the protein bovine pancreatic trypsin inhibitor. In general, two of the four possibilities were clearly incompatible with the range of acceptable values of parameters for cystine bridges, indicated above. In an overwhelming majority of cases only one possibility compares very closely to the observed values as determined by the shift of the modeled atoms S_i and S_j from the observed positions and comparison of stereochemical parameters. Tables II and III provide a summary of typical results obtained for peptides and proteins respectively. Figure 6 illustrates the comparison of observed and modeled S−S geometries for representative disulfides in peptides and proteins.

Satisfactory modeling of S−S bridges was possible in the cases of all the cystine peptides examined. In only one case of a peptide in Table II (peptide 2) was the same chirality obtained for two stereochemically acceptable disulfide positions. However, an examination of non-bonded contacts revealed that there are unacceptably short contacts between S atoms and the carbonyl oxygens of the same Cys residue for the χ_{SS} value of 83.6°, thus rationalizing the observed S−S geometry. Two satisfactory positions were also observed for the hydrochloride salt of the amino acid cystine.

An encouraging feature of the modeling of disulfide in peptides is that perfectly acceptable positioning of the S−S bond was possible in the cases of peptides 1 [$(Boc\text{-}Cys\text{-}Ala\text{-}Cys\text{-}NHMe)_2$], 3 (Boc-Cys-Val-Aib-Ala-Leu-Cys-NHMe) and 4 [cyclo(cystine)].

In these three cases unusually short $C^{\alpha}-C^{\alpha}$ distances are observed. In peptides 1 and 3 disulfide bridging is accomplished across the nearest neighbours on an antiparallel β-sheet structure, and the observed distances are: 1, $r_{ij}^{\alpha} = 3.88$ Å and $r_{ij}^{\beta} = 4.03$ Å; 3, $r_{ij}^{\alpha} = 4.04$ Å and $r_{ij}^{\beta} = 4.05$ Å. In peptide 4 disulfide bridging is achieved across the 1 and 4 positions of a dioxopiperazine ring with the amide bonds constrained to a *cis*

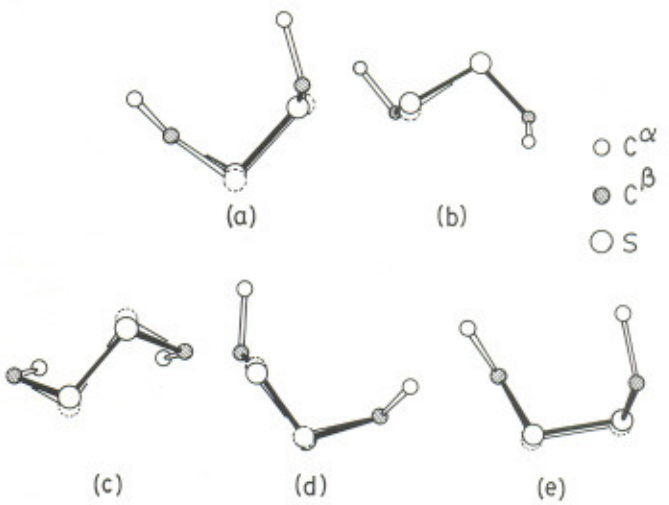
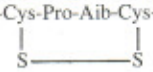


Fig. 6. Comparison of observed and modeled disulfides. C^{α} and C^{β} positions are identical in both cases. The darkened bonds for the $C^{\beta}-S-S-C^{\beta}$ segment correspond to the experimentally observed structure, while the open bonds represent the modeled segment. The modeled S atoms are shown as broken circles, while the observed S positions are full circles. (a) Crambin (4–32); (b) human lysozyme (30–116); (c) 4, cyclo (cystine) (d) 2, Boc-Cys-Pro-Aib-Cys-NHMe, (e) 1, $[Boc\text{-}Cys\text{-}Ala\text{-}Cys\text{-}NHMe]_2$.



geometry. In this case the relevant distances are $r_{ij}^{\alpha} = 2.80$ Å and $r_{ij}^{\beta} = 3.87$ Å.

In proteins a total correspondence of observed and modeled structures has not been obtained. A summary of the results is presented in Table IV. Modeled disulfides which satisfied all stereochemical criteria outlined earlier (Figure 5) were characterized as grade A. If both r_{SS} and χ_{SS} values fell within acceptable limits and χ^1 values outside the normal limits, the disulfides were classified as grade B. Cases where either r_{SS} or χ_{SS} fell outside the set limits or where the modeling procedure failed to geometrically fix the S atoms were categorized as grade C. Three cases (carboxypeptidase 138–161, wheatgerm

Table IV. Summary of the modeling results as applied to disulfide bonds in proteins

Protein[a,b]	PDB code	Resolution (Å)	No. of disulfide bonds	Quality of modeled disulfides		
				Grade A	Grade B	Grade C
Actinidin	2ACT	1.7	3	3	0	0
Alpha-lytic protease	2ALP	1.7	3	2	1	0
Azurin[c]	2AZA	1.8	2	1	0	1
Bence–Jones protein (lambda, variable domain)	2RHE	1.6	1	1	0	0
Bence–Jones protein (REI, variable domain)	1REI	2.0	1	1	0	0
Crambin	1CRN	1.5	3	3	0	0
Carboxypeptidase A	5CPA	1.54	1	0	0	1
Erabutoxin	2EBX	1.4	4	4	0	0
Glutathione reductase	3GRS	1.54	1	0	0	1
Immunoglobulin FAB	1FB4	1.9	6	5	0	1
Insulin[c]	1INS	1.5	6	6	0	0
Lysozyme (HEW)	6LYZ	2.0	4	3	0	1
Lysozyme (Human)	1LZ1	1.5	4	3	0	1
Ovomucoid, third domain	2OVO	1.5	3	2	0	1
Papain	9PAP	1.65	3	3	0	0
Pencillopepsin	2APP	1.8	1	0	1	0
Phospholipase A-2	1BP2	1.7	7	5	1	1
Proteinase A	2SGA	1.5	2	0	0	2
Rat mast cell protease[c]	3RP2	1.9	6	3	1	2
Ribonuclease A	5RSA	2.0	4	3	1	0
Scorpion neurotoxin	1SN3	1.8	4	4	0	0
Trypsin inhibitor	5PTI	1.0	3	3	0	0
Trypsin complex with *p*-amidinophenyl pyruvate	1TPP	1.4	6	4	1	1
Trypsin complex with pancreatic trypsin inhibitor	2PTC	1.9	9	8	0	1
Wheatgerm agglutinin	3WGA	1.8	16	14	1	1

[a]Literature citations for the various structures are: 2ACT (Baker and Dodson, 1980); 2ALP (Fujinaga *et al.*, 1985); 2AZA (Norris *et al.*, 1983); 2RHE (Furey *et al.*, 1983); 1REI (Epp *et al.*, 1975); 1CRN (Teeter, 1984); 5CPA (Rees *et al.*, 1983); 2EBX (Bourne *et al.*, 1985); 3GRS (Karplus and Schulz, 1987); 1FB4 (Marquart *et al.*, 1980); 1INS (Dodson *et al.*, 1979); 6LYZ (Diamond, 1974); 1LZ1 (Artymiuk and Blake, 1981); 2OVO (Bode *et al.*, 1985); 9PAP (Kamphuis *et al.*, 1984); 2APP (James and Sielecki, 1983); 1BP2 (Dijkstra *et al.*, 1981); 2SGA (James *et al.*, 1980); 3RP2 (Reynolds *et al.*, 1985); 5RSA (Wlodawer *et al.*, 1986); 1SN3 (Almassy *et al.*, 1983); 5PTI (Wlodawer *et al.*, 1984); 1TPP (Marquart *et al.*, 1983); 2PTC (Huber *et al.*, 1974); 3WGA (Wright, 1987).

[b]In this study homologous proteins (*cf.* actinidin and papain), independent determinations (*cf.* trypsin complexes) and homologous subunits (*cf.* wheatgerm agglutinin) have been considered separately, since precise details of disulfide stereochemistry vary and these serve as a check on the modeling procedure.

[c]There are two molecules in the asymmetric unit and atom numbering is done with chain identifiers A and B. Both have been considered for the present study.

agglutinin A17–A31 and proteinase A 42–58) where the modeled S_i, S_j positions varied appreciably from the observed positions were also listed under grade C. It may be noted that while the classification scheme outlined in Figure 5 is based purely on stereochemical parameters, in the comparisons with native disulfides agreement with observed S_i, S_j positions has also been considered. Of the 103 S–S bridges from 25 independent protein structures examined (Table IV), 81 were in grade A, seven in grade B and 15 in grade C. All the grade A and B models compared very well with those observed in the crystal structures, with only small shifts of the atomic coordinates for S_i and S_j (0.4 Å). This suggests that the simple procedure described here successfully models most of the observed S–S bridges in protein crystal structures. The 15 grade C cases were further re-examined (Table V).

The A168–A182 disulfide in rat mast cell protease has a very highly distorted χ_{SS} value of 158.9° in the crystal structure. This is indeed modeled with a highly unfavorable χ_{SS} value of 170°, the calculated S_i, S_j positions showing only small shifts of 0.15 or 0.61 Å from the experimental structure. Interestingly, no such distortion is noted for the same disulfide bridge in the second molecule (B) present in the crystallographic asymmetric unit which is also modeled well by the present procedure. Similarly, the χ_{SS} value of the 58–63 disulfide in glutathione reductase is −133.4°, far from the ideal value of −90° for a left-handed disulfide. The modeled disulfide compares well with the observed, as seen by small S_i, S_j shifts of 0.53 and 0.22 Å respectively. The modeled disulfide has a χ_{SS} value of −134.2° and thus falls into grade C. Four cases, azurin B3–B26, ovomucoid third domain 24–56, phospholipase A-2 61–91 and the immunoglobulin FAB fragment 22–96, fall into a borderline category, with modeled r_{SS} value falling just outside (<0.1 Å) the upper limit of 2.4 Å. Three cases, carboxypeptidase A 138–161, wheatgerm agglutinin A17–A31 and proteinase A (*Streptomyces griseus*) 42–58, are modeled poorly, with appreciable shifts of the S_i, S_j atoms (0.9–1.7 Å) compared to the crystal structure. These disulfides show significant distortions of the bond angles $C^{\alpha}C^{\beta}S$ and $C^{\beta}SS$ in the observed structures. In all three cases, variation of the bond angles by ±4° resulted in good agreement between modeled and observed disulfides. In six cases the sulfur atoms could not be stereochemically fixed using the standard bond lengths and bond angles. However,

Table V. Listing of grade C disulfides in proteins obtained by modeling procedure

Protein	Disulfide bonded residues	Reason for grade C
Azurin (2AZA)	B3–B26	r_{SS} = 2.42 Å
Carboxypeptidase A (5CPA)	138–161	large shifts of S_i and S_j (>0.4 Å) in the modeled disulfide[a]
Lysozyme (HEW) (6LYZ)	76–94	fixing of sulfur was not possible geometrically[b]
Lysozyme (human) (1LZ1)	65–81	fixing of sulfur was not possible geometrically[c]
Ovomucoid, third domain (2OVO)	24–56	r_{SS} = 2.41 Å
Rat mast cell protease (3RP2)	A42–A58	fixing of sulfur was not possible geometrically[c]
	A168–A182	χ_{SS} = 170°[d]
Glutathione reductase (3GRS)	58–63	χ_{SS} = −134.2°
Phospholipase A-2 (1BP2)	61–91	r_{SS} = 2.49 Å
β-Trypsin (1TPP)	42–58	fixing of sulfur was not possible geometrically[c]
β-Trypsin complex with inhibitor (2PTC)	E42–E58	fixing of sulfur was not possible geometrically[c]
Immunoglobulin FAB (1FB4)	22–96	χ_{SS} = 59.7°
Proteinase A (2SGA)	42–58	large shifts of S_i and S_j (>0.4 Å) in the modeled disulfide[a]
	191–220	sulfur could not be fixed geometrically[c]
Wheatgerm agglutinin (3WGA)	A17–A31	large shifts of S_i and S_j (>0.4 Å) in the modeled disulfide[a]

[a]In these cases, variation of the bond angles $C^\alpha C^\beta S$ and $C^\beta SS$ by ±4° resulted in a good fit of the modeled and the observed disulfides.
[b]The observed value of $C^\alpha C^\beta S$ of residue 76 was 14° more than the ideal value of 114°.
[c]A slight increase in the value of $C^\beta SS$ leads to satisfactory modeling of S–S bridge.
[d]The value of χ_{SS} observed in the protein itself is 158°.
[e]Increases in the values of $C^\alpha C^\beta S$ and $C^\beta SS$ leads to satisfactory modeling of S–S bridge.

Table VI. Comparison of parameters in reported (R) and modeled (M) structures for four engineered disulfides

Protein[a]	Residues		Parameters[b]						
			χ_{SS}(°)	χ_i^1(°)	χ_j^1(°)	χ_i^2(°)	χ_j^2(°)	r_{ij}^α(Å)	r_{ij}^β(Å)
Dihydrofolate	39–85	(R) Mol.I	−86.0	−147.0	−78.0	−149.0	−63.0		
Reductase		(R) Mol.II	−76.0	−159.0	−81.0	−158.0	−65.0		
(4DFR)		(M)	sulfur could not be fixed geometrically[c]					6.10	4.36
T4 lysozyme	3–97	(R)	134.0	117.0	−162.0	25.0	43.0		
(2LZM)		(M)[d]	162.0	−84.9	−153.5	−9.7	50.1	5.69	4.63
Subtilisin BPN′	22–87	(R)	−98.0	53.0	−49.0	121.0	143.0		
(1SBT)		(M)	−66.9	31.8	−81.9	136.8	152.7	4.88	3.09
	24–87	(R)	96.0	−65.0	−157.0	−50.0	−171.0		
		(M)	sulfur could not be fixed geometrically[c]					4.65	4.46

[a]Literature citations for the structure of native proteins are: 4DFR (Bolin *et al.*, 1982); 2LZM (Weaver and Matthews, 1987) and 1SBT (Alden *et al.*, 1971).
[b]Reported values of DHFR and subtilisin mutants taken from the crystallographic analysis of the engineered mutants: DHFR (Villafranca *et al.*, 1987) and subtilisin (Katz and Kossiakoff, 1986). Reported values for the T4 lysozyme mutant were obtained from a computer graphics modeling study done by Perry and Wetzel (1984).
[c]Fixing of sulfur atoms was not possible, even after relaxation was applied.
[d]Since the r_{ij}^β observed in the crystal structure is 4.63 Å, the r_{ij}^β criterion was relaxed for this case alone.

altering the bond angles at S and C^β up to ±4° from the standard values ($C^\beta SS$ = 104°, $C^\alpha C^\beta S$ = 114°), in increments of 0.5°, permitted satisfactory modeling in four cases, with grade A models being obtained. In the case of rat mast cell protease, the A42–A58 disulfide could be fixed with poor stereochemical parameters. The only case where this did not yield proper geometrical fixing is the 76–94 disulfide in HEW lysozyme. In this case, the observed $C^\alpha C^\beta S$ angles are highly distorted (116 and 128°).

Analysis of reported examples of engineered disulfides

The three cases for which mutant disulfides have been made and crystallographic coordinates for the native protein are available in the Protein Data Bank, namely T4 lysozyme (Perry and Wetzel, 1984; Wetzel *et al.*, 1988), dihydrofolate reductase, DHFR (Villafranca *et al.*, 1983, 1987) and subtilisin BPN′ (Wells and Powers, 1986; Pantoliano *et al.*, 1987), were examined. In all cases, the chosen positions for Cys introduction were identified on the basis of the r_{ij}^α, r_{ij}^β criteria, with the exception of the 3–97 disulfide in T4 lysozyme, where r_{ij}^β = 4.63 Å. Table VI summarizes the results of attempts of model disulfides across these positions, using the available C^α, C^β coordinates. For subtilisin BPN′, the 22–87 disulfide bond had poor r_{SS} and χ_{SS} values and the chosen positions would be classified as grade C. A satisfactory disulfide bond could not be fixed between residues 24 and 87. A crystallographic study of these two disulfides has indeed revealed that they are strained, with χ_i^2, χ_j^2 values being appreciably different from those observed in native protein disulfides (Katz and Kossiakoff, 1986; Figures 3 and 4). Although there is some controversy in the literature on the effect of the

Fig. 7. Alpha carbon tracing of subtilisin using the coordinates of Alden *et al.* (1971). The modeled 35−92 and 22−121 disulfide bonds are indicated. The active site residues Asp32, His64 and Ser221 are marked and side-chain atoms are shown. A stereoview is shown at the bottom. This figure was drawn using CAPLOT, a program developed for the HP 1000 system at Bangalore.

engineered disulfide bonds on the stabilization of subtilisin to autolysis and thermal inactivation (Wells and Powers, 1986; Pantoliano *et al.*, 1987), it appears safe to conclude that the two S−S bonds do not confer any appreciable degree of extra stability to the mutant protein. Indeed, the present modeling studies, together with the earlier crystallographic report (Katz and Kossiakoff, 1986), suggest that the chosen positions for Cys introduction lead to 'stereochemically non-ideal' disulfides. The inherent strain then results in unusual dihedral angles at the cystine bridge. In principle, the stabilization resulting from the lowered entropy of the unfolded form could be offset by an unfavorable enthalpic contribution in the folded form (Matthews, 1987; Wetzel, 1987), as a consequence of non-optimal disulfide introduction. In the present analysis of the subtilisin BPN′ structure using a 2.5 Å coordinate set (Alden *et al.*, 1971), two stereochemically optimal positions for disulfide introduction were identified. These are between Val28 and Val121 (modeled parameters are: $r_{ij}^{\alpha} = 4.54$ Å, $r_{ij}^{\beta} = 4.23$ Å, $r_{SS} = 1.98$ Å, $\chi_i^1 = -79.9°$ $\chi_i^2 = +6.4°$ $\chi_{SS} = -109.9°$, $\chi_j^2 = 139.5°$ $\chi_j^1 = 35.0°$) and between Ile35 and Ala92 ($r_{ij}^{\alpha} = 6.26$ Å, $r_{ij}^{\beta} = 4.33$ Å, $\chi_i^1 = 81.8°$, $\chi_i^2 = -158.3°$, $\chi_{SS} = 105.7°$, $\chi_j^2 = -133.5°$ and $\chi_j^1 = -69.7°$). These positions also satisfy other desirable conditions, i.e. the chosen residues are neutral, uninvolved in functional aspects and fairly far apart in the primary sequence resulting in a relatively large disulfide loop. Figure 7 shows an alpha carbon tracing of the subtilisin structure, indicating the modeled disulfide positions. We are unaware of attempts to engineer these disulfides. Subsequent to the submission of this manuscript, engineered disulfides have been reported between residues 26−232, 29−119, 36−210, 41−80 and 148−243 in subtilisin BPN′. None of the disulfide mutants was substantially more stable than the wild type enzyme (Mitchinson and Wells, 1987).

In the case of DHFR a stereochemically optimal disulfide bond could not be modeled across positions 39 and 85. Interestingly, this engineered disulfide did not enhance the stability of DHFR to thermal denaturation and was not readily formed from the dithiol form of the enzyme. Furthermore, the geometry of the disulfide determined by crystallographic studies (Villafranca *et al.*, 1987) is significantly different from those obtained for native protein disulfides (see Table VII). From Figure 4 it is seen that the χ_i^2, χ_j^2 values are appreciably different from protein observations.

In T4 lysozyme a disulfide has been introduced between residues 3 and 97 (Perry and Wetzel, 1984). Using the available coordinate set (resolution: 1.7 Å; Weaver and Matthews, 1987), this pair is not identified by our procedure since the values of r_{ij}^{β} is 4.63 Å (slightly greater than our limiting distance of 4.5 Å). Relaxation of this limit permitted modeling of the S−S bridge. However, none of the four alternatives obtained satisfied the stereochemical criteria for a strain-free position. Indeed, a computer graphics simulation of the 3−97 disulfide bridge by Perry and Wetzel (1984) yielded stereochemical parameters (Table VII) which are quite different from those generally observed in proteins. It is seen from Figure 3 that these reported χ_i^2, χ_j^2 values fall in a unique, unpopulated region of $\chi_i^2-\chi_j^2$ space. Interestingly, the 3−97 disulfide has been shown to enhance the thermal stability of the mutant protein as compared to the wild-type (Perry and Wetzel, 1984). However, a recent report suggests that stabilization is not due to 'control over the thermodynamics of the reversible unfolding equilibrium' (Wetzel *et al.*, 1988); rather, the covalent cross-link appears to act as a constraint on the thermally unfolded state, limiting other pathways of loss of activity. Thermal inactivation of T4 lysozyme appears to involve cysteine oxidation (Cys54 and Cys97), resulting in disulfide linked oligomeric forms (Perry and Wetzel, 1984). The enhanced thermal stability of the 3−97 disulfide mutant may then be due to the relative inaccessibility of Cys54 in the thermally unfolded state of the mutant protein compared to the wild type. An interesting feature of the analysis of T4 lysozyme is that no 'stereochemically optimal' positions for S−S introduction were obtained. This observation compares well with a statement in the original study that 'in no case was a simulated disulfide identical in structure to any protein disulfides characterized by X-ray diffraction' (Perry and Wetzel, 1984).

In all three examples discussed above, disulfide introduction has not resulted in a dramatic increase in stability to thermal unfolding. Indeed, all the mutant disulfides considered appear to be 'stereochemically non-optimal'. The strain accompanying S−S bond formation is then distributed in the stereochemical parameters describing the cystine bridge or in local distortions

of the protein structure. In the two cases where crystal structures of engineered proteins are available (subtilisin and DHFR), the disulfide bridges appear to be inherently strained (Katz and Kossaikoff, 1986; Villafranca *et al.*, 1987).

The method outlined in this report for modeling disulfide bonds into proteins of known three-dimensional structure adopts a strategy different from that described by Pabo and Suchanek (1986) in that positions suitable for disulfide bridging are identified without recourse to a comparison with known conformations in proteins. Indeed, the approach has been tested by modeling >100 accurate (≤2 Å) disulfide structures available in the Protein Data Bank, using only C^{α} and C^{β} coordinates. The present approach is more closely related to the recent report of Hazes and Dijkstra (1988), which presented an explicit analysis of fewer examples. However, energy minimization of an initially modeled disulfide is not adopted in our approach, since this procedure is fraught with uncertainty when only a few local interactions are considered. For example, the 168–182 disulfide in molecule B of rat mast cell protease and the 56–95 disulfide in papain are predicted as high energy disulfides by Hazes and Dijkstra (1988). Both bonds are comfortably modeled by the present procedure. Furthermore, the strained S–S bond (A168–A182) in rat mast cell protease is correctly modeled by the present approach, whereas energy minimization leads to a geometry distinctly different from the observed structure (Hazes and Dijkstra, 1988). Despite this reservation, energy minimization procedures using an appreciable number of atoms in the vicinity of the modeled disulfide can be a powerful tool, but this would entail extensive computational resources. The present procedure on the other hand is programmed for a personal computer and can be quickly used to provide a set of positions for disulfide bond introduction. Our approach has been applied to the enzyme thymidylate synthase from *Lactobacillus casei*, using a 2.8 Å coordinate set (Hardy *et al.*, 1987), in order to determine a site for substitution that would permit S–S bond formation to Cys244, thereby leaving only a single free sulfhydryl group in the molecule, i.e. the active site residue Cys198. Ile171 satisifies the modeling criteria, which provide only a rough guideline, in view of the relatively low resolution structure presently available. Work in progress is aimed at characterizing the mutant protein with the 171–244 disulfide bridge (H.Balaram, D.V.Santi, S.Agarwalla and P.Balaram, unpublished data). The present study has been specifically directed at identifying 'stereochemically optimal' positions for disulfide bridging, which may be important for protein stabilization by covalent cross-linking. This approach is particularly attractive where high resolution crystal structures are available for the protein of interest. For a typical protein of ~100 residues in length, only ~2–3 'ideal' (grade A) positions are obtained. It should, however, be emphasized that disulfide bonds can in fact be introduced at 'non-ideal' positions, but in such cases stereochemical distortion of the disulfide bridge itself or the protein backbone in the vicinity of the cross-link may energetically offset any stabilization gained by reducing the entropy of the unfolded protein structure. The modeling procedure has also served to fix disulfide bridges in small cystine peptides, obtaining excellent agreement with known crystal structures. This may prove valuable in the theoretical conformational analysis of a growing number of highly active, disulfide-bridged analogs of biologically active peptides (Hruby *et al.*, 1984). The increasing availability of NMR-determined three-dimensional structures of proteins and peptides, of admittedly low resolution (Wüthrich, 1986; Kaptein *et al.*, 1988), will undoubtedly stimulate further exercises in limiting conformational mobility by engineering disulfides. Extensions of this procedure to geometrically localizing stereochemically acceptable disulfide bridges using only C^{α} coordinates are presently under investigation.

Acknowledgements

This research was supported by a grant from the Department of Science and Technology, Government of India and by NIH grant CA14394. RS is the recipient of a fellowship from the Council of Scientific and Industrial Research, India. The use of facilities in the Computer Graphics Laboratory, University of California at San Francisco, in the initial stages of the project is gratefully acknowledged.

References

Alden,R.A., Birktoft,J.J., Kraut,J., Robertus,J.D. and Wright,C.S. (1971) *Biochem. Biophys. Res. Commun.*, **45**, 337–344.
Almassy,R.J., Fontecilla-Camps,J.C., Suddath,F.L. and Bugg,C.E. (1983) *J. Mol. Biol.*, **170**, 497–527.
Anfinsen,C.B. and Scheraga,H.A. (1975) *Adv. Protein Chem.*, **29**, 205–299.
Artymiuk,P.J. and Blake,C.C.F. (1981) *J. Mol. Biol.*, **152**, 737–762.
Baker,E.N. and Dodson,E.J. (1980) *Acta Crystallogr.*, **A36**, 559–572.
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.
Bode,W., Epp,O., Huber,R., Laskowski,M.,Jr and Ardelt,W. (1985) *Eur. J. Biochem.*, **147**, 387–395.
Bolin,J.T., Filman,J., Matthews,D.A., Hamlin,R.C. and Kraut,J. (1982) *J. Biol. Chem.*, **257**, 13650–13662.
Bourne,P.E., Sato,A., Corfield,P.W.R., Rosen,L.S., Birken,S. and Low,B.W. (1985) *Eur. J. Biochem.*, **153**, 521–527.
Creighton,T.E. (1988) *BioEssays*, **8**, 57–63.
Diamond,R. (1974) *J. Mol. Biol.*, **82**, 371–391.
Dijkstra,B.W., Kalk,K.H., Hol,W.G.J. and Drenth,J. (1981) *J. Mol. Biol.*, **147**, 97–123.
Dodson,E.J., Dodson,G.G., Hodgkin,D.C. and Reynolds,C.D. (1979) *Can. J. Biochem.*, **57**, 469–479.
Epp,O., Lattman,E.E., Schiffer,M., Huber,R. and Palm,W. (1975) *Biochemistry*, **14**, 4943–4952.
Fujinaga,M., Delbaere,L.T.J., Brayer,G.D. and James,M.N.G. (1985) *J. Mol. Biol.*, **184**, 479–502.
Furey,W.,Jr, Wang,B.C., Yoo,C.S. and Sax,M. (1983) *J. Mol. Biol.*, **167**, 661–692.
Hardy,L.W., Finer-Moore,J.S., Montfort,W.R., Jones,M.O., Santi,D.V. and Stroud,R.M. (1987) *Science*, **235**, 448–455.
Hazes,B. and Dijkstra,B.W. (1988) *Prot. Engng*, **2**, 119–125.
Hruby,V.J., Krstenasky,J.L. and Cody,W.L. (1984) *Annu. Rep. Med. Chem.*, **19**, 303–312.
Huber,R., Kukla,D., Bode,W., Schwager,P., Bartels,K., Deisenhofer,J. and Steigemann,W. (1974) *J. Mol. Biol.*, **89**, 73–101.
James,M.N.G. and Sielecki,A.R. (1983) *J. Mol. Biol.*, **163**, 299–361.
James,M.N.G. Sielecki,A.R., Brayer,G.D., Delbaere,L.T.J. and Bauer,C.-A. (1980) *J. Mol. Biol.*, **144**, 43–88.
Jones,D.D., Bernal,I., Frey,M.N. and Koetzle,T.F. (1974) *Acta Crystallogr.*, **B30**, 1220–1227.
Kamphuis,I.G., Kalk,K.H., Swarte,M.B.A. and Drenth,J. (1984) *J. Mol. Biol.*, **179**, 233–256.
Kaptein,R., Boclens,R., Sceek,R.M. and Van Gunsteren,W.F. (1988) *Biochemistry*, **27**, 5389–5395.
Karle,I.L., Kishore,R., Raghothama,S. and Balaram,P. (1988) *J. Am. Chem. Soc.*, **110**, 1958–1963.
Karle,I.L., Flippen-Anderson,J.L., Kishore,R. and Balaram,P. (1989) *Int. J. Peptide Protein Res.*, **34**, 37–41.
Karplus,P.A. and Schulz,G.E. (1987) *J. Mol. Biol.*, **195**, 701–729.
Katz,B.A. and Kossiakoff,A. (1986) *J. Biol. Chem.*, **261**, 15480–15485.
Kopple,K.D., Wang, Yu-S., Cheng,A.G. and Bhandary,K.K (1988) *J. Am Chem. Soc.*, **110**, 4168–4176.
Marquart,M. and Deisenhofer,J., Huber,R. and Palm,W. (1980) *J. Mol. Biol.*, **141**, 369–391.
Marquart,M., Walter,J., Deisenhofer,J., Bode,W. and Huber,R. (1983) *Acta Crystallogr.*, **B39**, 480–490.
Matsumara,M. and Matthews,B.W. (1989) *Science*, **243**, 792–794.
Matthews,B.W. (1987) *Biochemistry*, **26**, 6885–6888.
Mitchinson,C. and Wells,J.S. (1989) *Biochemistry*, **28**, 4807–4815.
Norris,G.E., Anderson,B.F. and Baker,E.N. (1983) *J. Mol. Biol.*, **165**, 501–521.
Pabo,C.O. and Suchanek,E.G. (1986) *Biochemistry*, **25**, 5987–5991.

Pantoliano,M.W., Ladner,R.C., Bryan,P.N., Rollence,M.L., Wood,J.F. and Poulos,T.L. (1987) *Biochemistry*, **26**, 2077–2082.
Perry,L.J. and Wetzel,R. (1984) *Science*, **226**, 555–557.
Ravi,A., Prasad,B.V.V. and Balaram,P. (1983) *J. Am. Chem. Soc.*, **105**, 105–109.
Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) *J. Mol. Biol.*, **168**, 367–387.
Reynolds,R.A., Remington,S.J., Weaver,L.H., Fischer,R.G., Anderson,W.F., Ammon,H.L. and Matthews,B.W. (1985) *Acta Crystallogr.*, **B41**, 139–147.
Rosenfield,R.E.,Jr and Parthasarathy,R. (1975) *Acta Crytallogr.*, **B31**, 816–819.
Sauer,R.T., Hehir,K., Stearman,R.S., Weiss,M.A., Teitler-Nilsson,A., Suchanek,E.G. and Pabo,C.O. (1986) *Biochemistry*, **25**, 5992–5998.
Teeter,M.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6014–6018.
Varughese,K.I., Lu,C.T. and Kartha,G. (1981) *Int. J. Peptide Protein. Res.*, **18**, 88–102.
Villafranca,J.E., Howell,E.E., Voet,D.H., Strobel,M.S., Ogden,R.C., Abelson,J.N. and Kraut,J. (1983) *Science*, **222**, 782–788.
Villafrance,J.E., Howell,E.E., Oatley,S.J., Xuong,N.-h. and Kraut,J. (1987) *Biochemistry*, **26**, 2182–2189.
Weaver,L.H. and Matthews,B.W. (1987) *J. Mol. Biol.*, **193**, 189–199.
Wells,J.A. and Powers,D.B. (1986) *J. Biol. Chem.*, **261**, 6564–6570.
Wetzel,R. (1987) *Trends Biochem. Sci.*, **12**, 478–482.
Wetzel,R., Perry,L.J., Baase,W.A. and Becktel,W.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 401–405.
Wlodawer,A., Walter,J., Huber,R. and Sjolin,L. (1984) *J. Mol. Biol.*, **180**, 307–329.
Wlodawer,A., Borkakoti,N., Moss,D.S. and Howlin,B. (1986) *Acta Crystallogr.*, **B42**, 379–387.
Wright,C.S. (1987) *J. Mol. Biol.*, **197**, 501–529.
Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.

Received on May 2, 1989; accepted on August 31, 1989