MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins

Vardhan S.Dani[1], C.Ramakrishnan[1,3] and Raghavan Varadarajan[1,2]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012 and [2]Chemical Biology Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur P.O., Bangalore-560004, India

[3]To whom correspondence should be addressed.
E-mail: ramki@crmbu2.mbu.iisc.ernet.in

There have been several attempts to stabilize proteins through the introduction of engineered disulfide bonds. For reasons that are currently unclear, these have met with mixed success. Hence identification of locations where introduction of a disulfide cross-link will lead to protein stabilization is still a challenging task. A computational procedure, MODIP, was introduced more than a decade ago to select sites in protein structures that have the correct geometry for disulfide formation when replaced by Cys. In this study, we re-evaluated the stereochemical criteria used by MODIP for the selection and gradation of sites for modeling disulfides. We introduced steric criteria to check for energetically unfavorable non-bonded contacts with the modeled disulfide, since these can considerably offset the stabilizing effect of the cross-link. The performance of the refined procedure was checked for its ability to correctly predict naturally occurring disulfide bonds in proteins. A set of proteins in which disulfide bonds were introduced experimentally were analyzed with respect to MODIP predictions, stability and other parameters such as accessibility, residue depth, *B*-factors of the mutated sites, change in volume upon mutation and loop length enclosed by the disulfide. The analysis suggests that in addition to proper stereochemistry, stabilizing disulfides occur in regions of low depth, relatively high mobility, have a loop length greater than 25 and where the disulfide typically occupies a volume less than or equal to that of the original residues.

Keywords: disulfide bonds/protein stability/site-directed mutagenesis

Introduction

One of the main objectives of protein engineering is to produce variants with altered stability. Enzymes having greater stability at elevated temperatures or high salt concentrations are desirable for various commercial applications. Hence, rational approaches aimed at designing mutant proteins with improved stability are of great utility. Theoretical and experimental studies have revealed the factors that are important for protein stability (Matthews, 1993; Chakravarty and Varadarajan, 2000; Chakravarty *et al.*, 2002). These include hydrogen bonds, hydrophobic interactions, salt bridges and ion-pair networks, which are individually weak, but in summation can confer considerable stability to the protein. In addition, in several proteins, disulfide bridges contribute significantly to the stability of the protein. The stabilization effect is thought to be due mainly to reduction of main-chain entropy in the unfolded state. Studies on proteins lacking naturally occurring disulfides have shown the importance of these covalent bonds in folding and stability of the protein (Pace *et al.*, 1988). A single disulfide bridge can stabilize the protein by 2–5 kcal/mol (Creighton, 1988; Tidor and Karplus, 1993). Thus introduction of disulfide bridges by site-directed mutagenesis is an attractive strategy to engineer proteins with additional stability.

Several previous studies have reported experiments examining the effect of newly introduced disulfide bridges on the structure and stability of proteins. While some of these studies have resulted in striking stabilization of the mutant compared with the wild-type, many of the mutants showed moderate stabilization or even destabilization of the mutant (Wells and Powers, 1986; Mitchinson and Wells, 1989; Matsumura and Matthews, 1991; Clarke *et al.*, 1995). A good stereochemistry of the introduced disulfide is indeed essential, since atypical stereochemical features result in higher strain energy, which can offset the stabilizing effect of the disulfide cross-link (Katz and Kossiakoff, 1986). Simultaneously, loss of existing interactions after substituting with cysteine and unfavorable steric contacts can also have a destabilizing effect (Mitchinson and Wells, 1989). Hence it is important to consider these aspects while selecting novel sites for substitution with Cys residues. In addition to these factors, disulfides may also lead to stabilization of the unfolded state through favorable enthalpic interactions (Doig and Williams, 1991). However, at present, there is no way to take such effects into account satisfactorily.

With the aim of predicting sites in a protein which are geometrically suited for disulfide bridge formation, a computational procedure called MODIP (for Modeling of DIsulfide bridges in Proteins) was developed (Sowdhamini *et al.*, 1989). The evaluation of modeled sites was done through a grading system using certain conformational parameters that were found to be typical of disulfide bridges. All non-Gly residue pairs that satisfy the $C^\alpha...C^\alpha$ and $C^\beta...C^\beta$ distance criteria are selected for fixation of a hypothetical sulfur atom, using a geometric procedure (Sowdhamini *et al.*, 1989). To evaluate the modeled disulfide, a gradation scheme was used to rank the predicted disulfides based on certain stereochemical parameters (dihedral angles and S–S bond distance). Thus, a disulfide bridge modeled with dihedral angles within an ideal range of values, typical of naturally occurring disulfide bridges, was assigned grade 'A'. A site which was geometrically suitable for formation of the S–S covalent bond, but would have a somewhat distorted stereochemistry, was assigned grade 'B', and sites which were simply close enough in space potentially to allow the formation of the disulfide bond were assigned grade 'C'.

In this paper, we report the further modification of the program MODIP with some additional criteria for selection of sites and a new scheme for the evaluation of the selected sites.

The criteria employed by the previous version of MODIP were revisited using a much larger database of proteins than was initially available. In addition, the new version of MODIP checks for any kind of energetically unfavorable steric interaction in the immediate vicinity of the sulfur atoms that are introduced. Thus a modeled disulfide with good stereochemistry but severe steric hindrance will be deemed a lower priority prediction than one which is free of steric hindrance or at best has only some marginal steric contacts. The performance of MODIP was tested by its ability to predict successfully disulfides that already exist in proteins. Since this and other disulfide modeling procedures were first proposed, there have been many experimental attempts at protein stabilization through engineered disulfides. We therefore carried out a systematic comparison to find correlations between MODIP predictions and stability of mutant proteins where disulfide bridges have been experimentally introduced by site-directed mutagenesis. This analysis also points towards several additional criteria that can be applied while short-listing appropriate sites to introduce disulfide bridges.

Materials and methods

Selection of dataset for analysis of MODIP predictions

The 1997 release of the Protein Data Bank (Berman *et al.*, 2000) containing 6017 entries was chosen for the analysis. Entries of crystal structures from this database were selected to build a non-redundant dataset by using two criteria: (1) the structure should have a resolution of 2.0 Å or better and (2) the data set should not contain entries for which the sequence homology between any two polypeptide chains is more than 40%. This yielded 571 unique polypeptide chains from a total of 538 proteins.

Energy calculations

Non-bonded energies of interaction for the contacts involving the modeled sulfur atoms were calculated using an energy function, with constants as described earlier and summarized below. The potential used is of the form $Ae^{-\mu r} - Br^{-6}$, where r is the distance between the two non-bonded atoms. Parameters $(A \times 10^{-4}, B)$ are indicated in parentheses for the following atom pairs: S, N (124.2, 1038); S, O (90.5, 855.2); S, C (188.2, 1076) (Chandrasekaran and Balasubramanian, 1969). A value of 4.6 has been used for μ (Ramachandran and Sasisekharan, 1968).

Results and discussion

The 538 proteins were further classified mainly based on the presence or absence of Cys residues and disulfide bridges. Thus the final dataset of disulfide bridge containing proteins consisted of 172 entries representing 730 disulfides. A schematic diagram showing the details of selection of the dataset is shown in Figure 1. This dataset was used for further analysis of stereochemical features of disulfide bonds and MODIP predictions.

MODIP revisited

The original criteria used by the program MODIP were based on the conformational parameters of disulfide-bridged cysteine residues derived from 22 well-resolved crystal structures (Sowdhamini *et al.*, 1989; Srinivasan *et al.*, 1990). With the rapid growth in the number of crystal structures reported since then, it became relevant to ask if the parameters originally used by MODIP for modeling and evaluation of disulfide
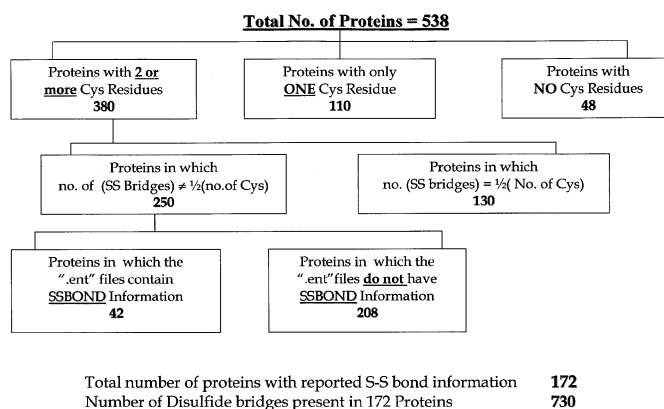

Fig. 1. Flow chart giving details regarding occurrence of cysteine residues and disulfide bridges in the current dataset.

bridges require revision on account of the increase in the data. The different parameters that were evaluated include the non-bonded distances $C_i^\alpha...C_j^\alpha$ (r^α) and $C_i^\beta...C_j^\beta$ (r^β) (where residues i and j are cysteines that are connected by an S–S bridge) and torsion angles χ^{ss}, χ_i^1, χ_j^1. Of these, r^α and r^β are used by MODIP for initial selection of residue pairs for modeling Cys residues; while the torsion angles χ^{ss}, χ_i^1, χ_j^1 are used as criteria for assigning grades to the modeled disulfides. Since the earlier criteria were postulated with the then available data on disulfide bridges, it was felt worthwhile to redo the analysis using the much larger number of examples occurring in the present data.

Parameters r^α and r^β

The distribution of these distance parameters for the two sets of data (earlier and present) is shown as histograms in Figure 2. A cursory look at the figures shows that the overall distributions and peak intervals are very similar. However, the finer structure of the distribution is somewhat altered. The program MODIP makes use of these distances as upper limit cut-offs (6.5 Å for r^α and 4.5 Å for r^β). It can be seen that these limits are not very different in the dataset of disulfide-bridged cysteines. However, in order not to miss out borderline cases, the upper limits used for r^α and r^β are slightly increased to 7.0 and 4.7 Å, respectively, in the modified version of MODIP. In addition, lower limit cut-offs of 3.8 and 3.4 Å for r^α and r^β, respectively have been introduced.

The distribution for the angular parameters χ^{ss}, χ^1 and χ^2 are shown as circular histograms in Figure 3 for both the earlier and the current data.

Parameter χ^{ss}

Figure 3a shows the distribution of this parameter in the two sets of data. The close similarity of the two histograms can easily be observed. In quantitative terms the values are more or less equally distributed around $|\chi^{ss}| = 90°$ in both cases, with a marginal increase in the percentage around $\chi^{ss} = +90°$. Thus the limits of 60 and 120° for $|\chi^{ss}|$ are still suitable for modeling purposes.

Parameters χ^1 and χ^2

Figure 3b and c show the distribution of these parameters in the two datasets. As with other amino acids, values of these χ^1 angles typically occur around ±60 and $\pm180°$ (staggered positions). Again, the histograms for the two datasets are strikingly similar, with about 65% of the examples distributed about $-60°$ for χ^1 in the two datasets, 25% around $\chi^1 = 180°$
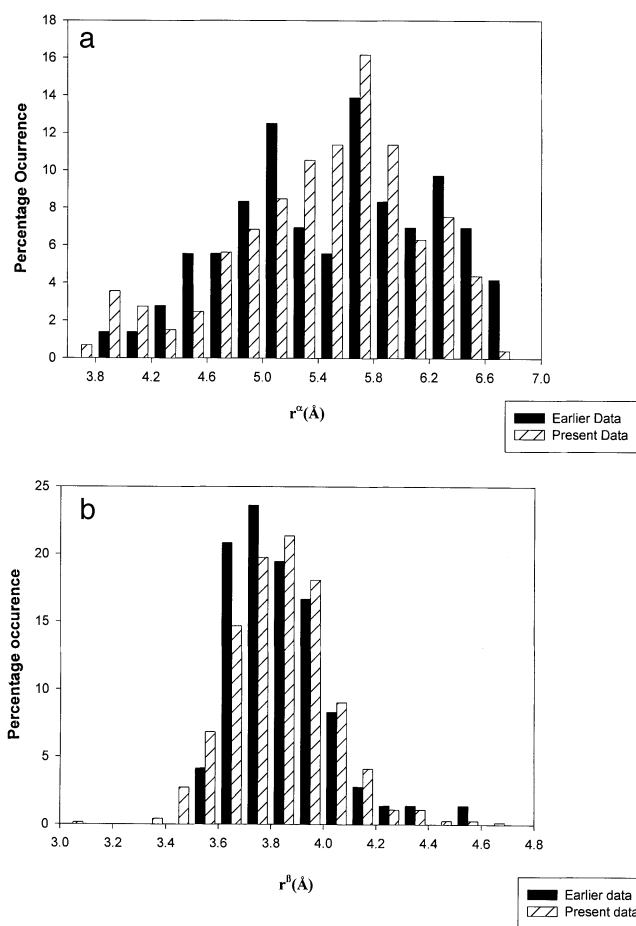
Fig. 2. Histograms showing the percentage distribution for the distance parameters: (**a**) $C_i^\alpha...C_j^\alpha$ (r^α) and (**b**) $C_i^\beta...C_j^\beta$ (r^β) for the earlier dataset (Srinivasan *et al.*, 1990) and the current dataset.

Fig. 3. Circular histograms showing the percentage distributions for the parameters: (**a**) χ^{ss} and (**b**) χ^1 (which are used in the MODIP program) and (**c**) χ^2 for the earlier (Srinivasan *et al.*, 1990) and the current datasets. The remarkable similarity between the corresponding histograms is clearly visible. The percentages for various ranges are indicated on the circumference of the histograms.

and about 12% around $\chi^1 = +60°$. In the previous version of MODIP $|\chi_i^1|$ and $|\chi_j^1|$ were restricted to values between 30 and 90° or between 150 and 180°. These limits have been retained in the present version of MODIP. Values of χ^2 were found to lie largely around -60 (~41%) and around $+60°$ (~20%) in both the present and the earlier datasets. However, these values are not used for the MODIP calculations. In summary, this analysis only goes to show that the ranges of values of different parameters preferred for the formation of disulfide bridges are very consistent between the older and the present datasets, thus justifying their use for computational modeling.

Since the procedure uses grades (A, B or C) to evaluate the stereochemistry of the modeled disulfide (Sowdhamini *et al.*, 1989), sites in proteins where existing disulfides are located should ideally be modeled with a high grade. Since the current dataset is ~300-fold larger than the earlier dataset, the program can be tested against those additional entries which are found only in the present dataset. When this is done, it is found that 100% of these naturally occurring examples are predicted by MODIP, of which an overwhelming majority (84.6%) are modeled with near ideal stereochemistry (A grade) and the modeled conformation was virtually identical with the experimentally observed conformation. In addition, we also analyzed burial of cysteine residues involved in disulfide bonds in proteins. Figure 4a shows the residue accessibilities of disulfide-bridged cysteine residues in naturally occurring disulfides. The accessibilities are typically in keeping with the general
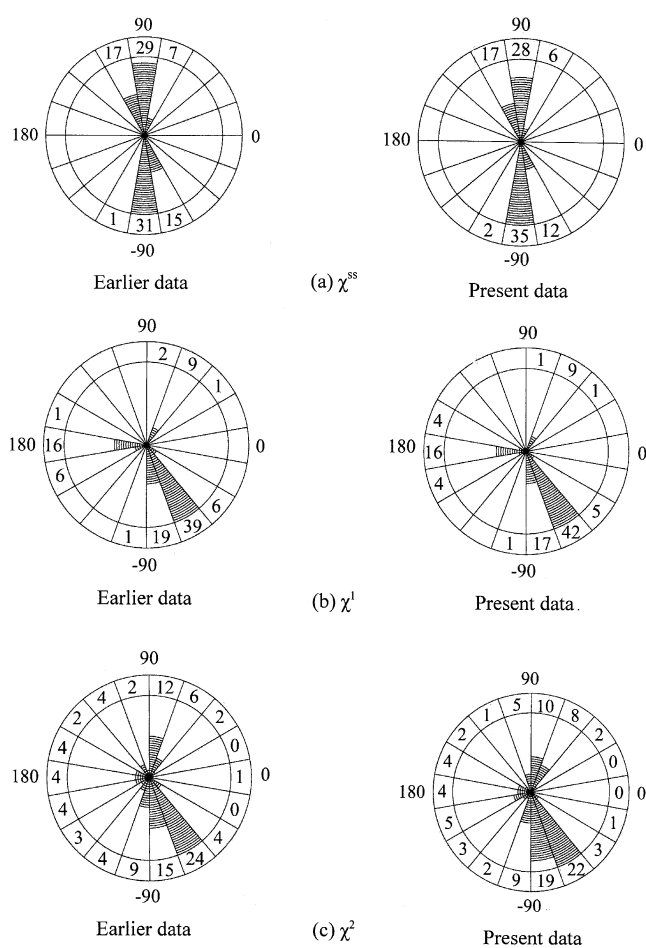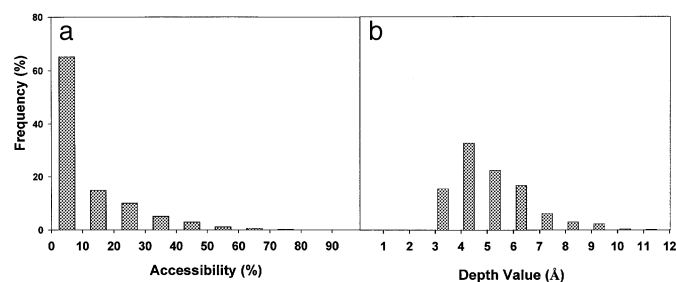

Fig. 4. Histograms showing burial of naturally occurring disulfide bridges in protein structures. (**a**) Percentage accessibility (Lee and Richards, 1971) and (**b**) residue depth (Chakravarty and Varadarajan, 1999).

notion (Thornton, 1981; Srinivasan *et al.*, 1990; Petersen *et al.*, 1999) that disulfide bonds occur at buried locations in protein structures. Recently, we have developed an alternative measure of residue burial known as residue depth. Depth is a measure of the distance of an atom or residue from the protein surface (Chakravarty and Varadarajan, 1999). It was previously shown that residue depths typically vary from about 4 to 10 Å and completely buried residues typically have depths >6 Å (Chakravarty and Varadarajan, 1999). The data in Figure 4b show that, although buried, disulfides are in fact typically located fairly close to the protein surface.

189

V.S.Dani, C.Ramakrishnan and **R.Varadarajan**

Table I. Correlation of stability data in proteins with engineered disulfides with structural features of the corresponding wild-type protein

Protein	PDB code (WT)	Position of mutation	Grade factor	Average accessibility (%)	Stability/ΔT_m/$\Delta\Delta G^{\circ a}$ (S, D, N[b]/°C/kcal/mol)	ΔVolume[c] (Å^3)	Loop length (No. of residues)	Reference
Human carbonic anhydrase II	2cba	Ala38–Ala258	5	12	D/–/–	27	220	Burton *et al.*, 2000
		Ser29–Ser197	4	0	D/–/–	17	168	
		Leu60–Ser173		32	S/–/–	52	113	
Arc repressor	1par	Asn11–Asn11′	–	34	S/40/–	–45	d	Robinson and Sauer, 2000
λ Repressor	1lmb	Tyr88(3)–Tyr88(4)	5	19	S/–/–	–186	d	Sauer *et al.*, 1986
		Tyr85(3)–Tyr85(4)	–	46	D/–/–	–186	d	
Dihydrofolate reductase	4dfr	Pro39A–Cys85A	4	6	S/–/1.8	–29	46	Villafranca *et al.*, 1987
Staphylococcal nuclease	1sno	Asp77–Asn118	4	5	D/–2/–2.3	–37	41	Hinck *et al.*, 1996
		Gly79–Asn118	–	16	S/8/1.5	17	39	
		Gln80–Lys116	–	75	S/2/1.0	–113	36	
T4 lysozyme	2lzm	Ile3–Cys97	5	13	S/4.8/–	–70	94	Matsumura and Matthews, 1991
		Thr21–Thr142	–	47	S/11/–	–34	121	
		Ile9–Leu164	–	44	S/6.4/–	–121	155	
		Ser90–Gln122	–	38	D/–0.5/–	–38	32	
		Asp127–Arg154	–	50	D/–2.4/–	–106	27	
Xylanase	1xnb	Val98–Ala152	5	12	S/–/–	–22	54	Wakarchuk *et al.*, 1994
		Ser100–Asn148	3	23	S/–/–	–14	48	
		Ser179–Ser179′	–	31	S/–/–	17	d	
Barnase	1rnb	Ala43–Ser80	2	35	S/–/2.1	22	37	Clarke *et al.*, 1995
		Thr70–Ser92	4	40	D/–/–2.9	–8	22	
		Ser85–His102	–7	18	S/–/4.3	–47	17	
Thymidylate synthase	4tms	Thr155–Glu188′	–	66	S/–/–	–55	d	Gokhale *et al.*, 1994
		Glu188–Thr155′	–	66		–55	d	
Subtilisin BPN′	2st1	Val26–Ala232	5	2	D/–/–	–21	206	Mitchinson and Wells, 1989
		Ala29–Met119	8	2	D/–/–	–49	90	
		Asp36–Pro210	10	16	D/–/–	–34	174	
		Asp41–Gly80	–	8	D/–/–	25	39	
		Val148–Asn243	4	8	D/–/–	–57	95	
		Thr22–Ser87	10	24	S/3.1/–	–8	65	Wells and Powers, 1986; Pantoliano *et al.*, 1987
		Ser24–Ser87	–	47	N/–/–	17	63	Wells and Powers, 1986
RNase 1	9rat	Ala4–Val118	5	63	S/–/2.0	–22	114	Futami *et al.*, 2000
Manganese peroxidase	1mnp	Ala48–Ala63	1	13	N/–/–	27	15	Reading and Aust, 2000
RNase H1	2rn2	Cys13–Asn44	4	28	S/11.8/2.8	–24	31	Kanaya *et al.*, 1991
Human lysozyme	1lz1	Trp64–Cys81	—	37	D/–/–	–137	17	Kanaya and Kikuchi, 1992
Aspergillus glucoamylase	3gly	Ala246–Cys320	4	0	S/4/–	4	74	Fierobe *et al.*, 1996
		Asn20–Ala27	7	42	S/1.7/–	–9	7	Li *et al.*, 1998
		Thr72–Ala471	5	52	N/0.3/–	–3	399	
Human antibody Fv	1mcp	Gln44H–Gly105L	–	22	S/–/–	–7	d	Reiter *et al.*, 1994
		Glu111H–Ile48L	–	32	S/–/–	–77	d	
Yeast iso-1-cytochrome *c*	1chh	Val20–Thr102	–	6.3	D/–/–1.6	–52	82	Betz *et al.*, 1996
Heat-labile enterotoxin	1ltt	Asn40A–Gly166A	–	2	S/6/–	17	126	van den Akker *et al.*, 1997
β-Lactoglobulin	1qg5	Cys121–Leu104	–	1	S/–/–	–70	17	Cho *et al.*, 1994
		Cys121–Ala132	–	2	S/–/–	–4	11	
Alk. protease	1alk	Gly199–Phe236	–	8	S/–/–	–50	37	Ko *et al.*, 1996
Human interferon γ	1fg9 (A and B chains)	Glu7A–Ser69A	10	24	S/15/–	–29	63	Waschutza *et al.*, 1996
		Ala17A–His111B	8	3	S/25/–	–42	d	
Barley chymotrypsin inhibitor	2ci2	Thr22–Val82	5	22	S/–/3.9	–52	60	Roesler and Rao, 2000
Troponin C N-terminal domain	4tnc	Met48–Met82	–	26	S/–/–	–126	34	Gusev *et al.*, 1991

[a]$\Delta\Delta G^{\circ}$ is the change in free energy of unfolding for the mutant relative to that for the wild-type [ΔG°(mutant) – ΔG°(wild-type)]. In some cases it is reported as $\Delta\Delta G^{H_2O}$, while in others it is calculated at the midpoint of denaturant concentration. $\Delta T_m = T_m$(mutant) – T_m(wild-type).
[b]For mutants where stability data provided are not ΔT_m/$\Delta\Delta G^{\circ}$, stability is represented by the letters S (stabilized with respect to wild-type), D (destabilized) or N (no change in stability).
[c]Calculated as 2 × (standard Voronoi volume of cystine) – (sum of standard Voronoi volumes for the pair of mutated residues). Standard Voronoi volumes of residues were obtained from the Macromolecular Geometry webpage (http://bioinfo.mbb.yale.edu/hyper/mbg/SurfaceVolumes/StdCoreVolumes.xtalatoms).
[d]The disulfide is formed between residues on different subunits (intermolecular disulfide).

Introduction of steric criteria and grade factor

The original version of the program MODIP detected geometrically suitable sites in a protein for modeling the disulfide bridge. The sites where modeled disulfides could assume a good geometry were assigned grade 'A', while other sites were assigned a 'B' or a 'C' grade; the final listing of possible sites was classified according to these grades. The predictions made by MODIP and other disulfide modeling programs (Pabo and Suchanek, 1986; Hazes and Dijkstra, 1988) have been verified experimentally with varying degrees of success (Matsumura and Matthews, 1991; Gokhale *et al.*, 1994; Clarke *et al.*, 1995; Gopal *et al.*, 1999; Robinson and Sauer, 2000). Several mutational studies have shown that, although introduction of disulfide bonds in a protein can sometimes lead to an increase in its stability, in several cases such a change leads to overall destabilization of the mutant protein with respect to the wild-type (Mitchinson and Wells, 1989; Matsumura and Matthews, 1991; Hinck *et al.*, 1996). One of the possible reasons for this could be steric hindrance of the newly introduced cysteine residues with surrounding atoms. This destabilizing effect can partially or completely offset the stabilization offered by the covalent disulfide bond. Hence, it was felt necessary to introduce a steric criterion for evaluation of the modeled disulfides. Furthermore, such an additional criterion would help in eliminating geometrically feasible but sterically unfeasible sites and this, in turn, could be expected to increase the overall stability of the protein. With this in mind, the program was upgraded to include a search for non-bonded contacts with atoms immediately surrounding the modeled sulfur atom, which may cause steric clashes with the modeled cysteine residues. All atoms which lie within a sphere of radius 4 Å, with the modeled sulfur atom as center, are picked up for energy calculations. Non-bonded energies of interaction of sulfur atoms with these atoms were evaluated.

In order to facilitate short-listing and easy selection of appropriate sites for mutation experiments, a new scheme has been introduced for evaluation of the modeled disulfide bridge. This scheme combines the former grading system with the newly introduced steric criteria. To start with, a numerical value (hereafter referred to as 'grade factor') is assigned to each stereochemical grade. Grades A, B and C are assigned grade factor values of 10, 8 and 5, respectively. A penalty is deducted from the grade factor for each predicted site that shows a positive value for non-bonded energy of interaction; for every kcal/mol (positive value) of the total non-bonded energy, a unit 'penalty' is deducted. If the positions of sulfur atoms cannot be fixed, an additional unit is deducted. The final grade factor value is thus representative of both the stereochemical and steric suitability of the modeled disulfide bridge with respect to its surrounding microenvironment. While the relative grade weightage assigned to the disulfide stereochemistry and steric short contacts is arbitrary, the final value of the grade factor indicates if there is substantial steric overlap of the modeled disulfide with its surrounding. The final output enlists the predictions in descending order of the values of the grade factor. It can be seen that if the penalty due to steric hindrance is large the final grade factor can be a negative value, even for a better stereochemical grade. For instance, if there is a predicted disulfide which can be modeled as either an 'A' or a 'B' grade, but modeling in 'A' grade alone results in severe steric hindrance, then a grade with a better net grade factor ('B', in this case) will be assigned to
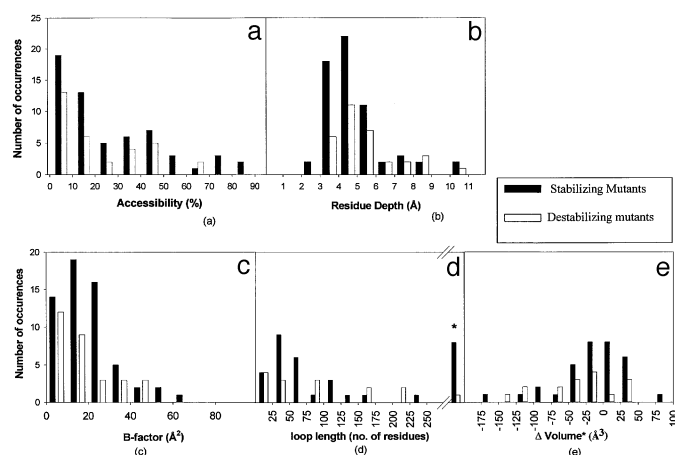

Fig. 5. Analysis of factors that result in stabilization by engineered disulfides. The number of occurrences of stabilizing and destabilizing disulfides are shown as a function of (**a**) average accessibility of the wild-type residues; (**b**) average depth of the wild-type residues; (**c**) average side chain *B*-factor of wild-type residues; (**d**) residue separation, where the asterisk indicates that the two residues are on different chains, hence an intermolecular disulfide bond is formed; (**e**) change in volume upon mutation (see footnote to Table I for definition of ΔVolume).

the predicted site. Thus a balance between stereochemical grade and steric hindrance is obtained. The output also gives the details of atoms involved in short contacts. This can be used to design additional residue substitutions to relieve these short contacts.

As a first step to check the correctness of the procedure, its ability to predict naturally existing disulfides in proteins with good grades and grade factor was tested by running the modified version of MODIP with the dataset of 172 proteins containing disulfide bonds. It is important to note that the modeling procedure does not take into account any information about the presence/absence of disulfide linkages (or their location) in a protein; disulfide-bridged cysteine residues are treated identically with any other non-Gly residue. Satisfactorily, most naturally occurring disulfide bridges were modeled with an 'A' grade (80.6%, after including the steric criteria, as against 84.5% without applying steric criteria). Thus, only a small fraction of the disulfide bridges that were previously modeled as 'A' grade are now modeled as 'B' or 'C' grades, after applying the steric criteria.

Analysis of reported disulfide mutants

In the past, several attempts have been made, with varying degrees of success, to introduce disulfide bridges in proteins with the aim of increasing stability. In a small subset of these proteins, crystal structures of the mutant proteins are also available (Pjura *et al.*, 1990; Jacobson *et al.*, 1992; Wakarchuk *et al.*, 1994; Clarke *et al.*, 1995; Velanker *et al.*, 1999). MODIP was further tested, using proteins in which stability data for disulfide engineered mutants were available; MODIP was used to analyze the crystal structure coordinates of the wild-type protein. The results of this analysis are shown in Table I and Figure 5. The criteria used by the experimentalists to select sites for mutagenesis in these proteins were not necessarily the same as those used by MODIP, although in all cases spatial proximity of the two residues to be mutated is considered.

Of the 47 mutations in 24 proteins, 30 resulted in increasing the thermodynamic stability of the protein at least by a small degree; in 14 examples the introduction of the disulfide bridges resulted in a decrease in stability and in three cases no change

191

was seen. Within this dataset, several were either not predicted at all by MODIP or were predicted only as grade C, although some of these were successful in imparting stability in the protein. For example, of the three stabilizing mutants of T4 lysozyme, only one is predicted as grade C (grade factor = 5). This suggests that, although these sites appear to be geometrically less preferred for disulfide formation in the wild-type structure, the protein structure is able to relax so that it can accommodate a disulfide bridge with good stereochemistry. The crystal structure of the disulfide mutants of T4 lysozyme supports this assertion. For the disulfides in the X-ray structures of the three stabilizing mutants, all have grade factors of 10, although two of these sites did not satisfy the MODIP criteria when the structure for the wild-type was used. The only scenario in which MODIP will not consider a site at all will be when the $C^\alpha...C^\alpha$ and $C^\beta...C^\beta$ distances of the two residues deviate considerably from standard values for disulfide bridges. Comparison of the crystal structures of the 9–164 (Pjura *et al.*, 1990) and 21–142 (Jacobson *et al.*, 1992) disulfide mutants of T4 lysozyme with the wild-type structure has revealed that backbone atoms of the region containing the respective sites have moved closer by about 2.5 Å upon mutation. A similar situation is also observed in the case of disulfide mutants of barnase (Clarke *et al.*, 1995) and *Bacillus* xylanase (Wakarchuk *et al.*, 1994). Hence even C grade disulfide can be considered as a candidate for mutagenesis, provided that the steric penalty is minimal and the sites are in relatively flexible regions of proteins.

Since different studies used different measures of stability, for the data in Figure 5 engineered disulfides are classified only as stabilizing or destabilizing. More detailed descriptions of the measured stability changes can be found in Table I. Figure 5a and b show accessibilities and residue depths at sites where disulfides have been introduced through mutation. Figure 5b indicates that stabilizing mutations are more likely to be located close to the protein surface with depths of 3–5 Å. Figure 5c shows average side-chain *B*-factors at sites selected for mutation. Stabilizing disulfides are typically in regions of medium to high mobility. The side chain *B*-factors are predominantly in the range 20–40 Å^2. Figure 5d shows the stabilization as a function of the residue separation (loop length enclosed by the disulfide, in number of residues) between the two mutated sites. As expected from theoretical considerations (Schellman, 1955; Pace *et al.*, 1988) there is little stabilization at smaller loop lengths, <25 residues. A large fraction of the stabilized mutants have loop lengths between 25 and 75 residues. Inter-subunit disulfide bridges also appear to result in stabilization in a large fraction of such examples that were studied. Figure 5e shows the correlation of stabilization with the difference in the standard residue volume before and after mutation.

Since MODIP and also other disulfide prediction programs typically yield large numbers of sites where disulfides may be introduced with acceptable stereochemistry, the problem is to predict which of these are most likely to result in protein stabilization. The data in Figure 5 and Table I suggests that amongst these predicted sites it is desirable to choose ones that are in relatively mobile regions close to the protein surface. This ensures that the protein structure can relax without incurring a significant energetic penalty. The modeled disulfides should not have significant steric overlaps with the surrounding atoms, especially with main chain atoms. Finally, the sequence separation between the mutated residues should be >25

residues and cases where the change in volume is more than +25 Å^3 or less than –75 Å^3 should be avoided. Positive volume changes typically result in steric overlap while large negative volumes are indicative of cavity formation and loss of van der Waals interactions. Both of these situations are energetically unfavorable. When selecting sites in proteins for mutagenesis, use of these additional criteria is likely to enhance the probability that the engineered disulfide will result in protein stabilization.

References

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.

Betz,S.F., Marmorino,J.L., Saunders,A.J., Doyle,D.F., Young,G.B. and Pielak,G.J. (1996) *Biochemistry*, **35**, 7422–7428.

Burton,R.E., Hunt,J.A., Fierke,C.A. and Oas,T.G. (2000) *Protein Sci.*, **9**, 776–785.

Chakravarty,S. and Varadarajan,R. (1999) *Struct. Fold. Des.*, **7**, 723–732.

Chakravarty,S. and Varadarajan,R. (2000) *FEBS Lett.*, **470**, 65–69.

Chakravarty,S., Bhinge,A. and Varadarajan,R. (2002) *J. Biol. Chem.*, **277**, 31345–31353.

Chandrasekaran,R. and Balasubramanian,R. (1969) *Biochim. Biophys. Acta*, **188**, 1–9.

Cho,Y., Gu,W., Watkins,S., Lee,S.P., Kim,T.R., Brady,J.W. and Batt,C.A. (1994) *Protein Eng.*, **7**, 263–270.

Clarke,J., Henrick,K. and Fersht,A.R. (1995) *J. Mol. Biol.*, **253**, 493–504.

Creighton,T.E. (1988) *Bioessays*, **8**, 57–63.

Doig,A. and Williams,D.H. (1991) *J. Mol. Biol.*, **217,** 389–398.

Fierobe,H.P., Stoffer,B.B., Frandsen,T.P. and Svensson,B. (1996) *Biochemistry*, **35**, 8696–8704.

Futami,J., Tada,H., Seno,M., Ishikami,S. and Yamada,H. (2000) *J. Biochem. (Tokyo)*, **128**, 245–250.

Gokhale,R.S., Agarwalla,S., Francis,V.S., Santi,D.V. and Balaram,P. (1994) *J. Mol. Biol.*, **235**, 89–94.

Gopal,B., Ray,S.S., Gokhale,R.S., Balaram,H., Murthy,M.R. and Balaram,P. (1999) *Biochemistry*, **38**, 478–486.

Gusev,N.B., Grabarek,Z. and Gergely,J. (1991) *J. Biol. Chem.*, **266**, 16622–16626.

Hazes,B. and Dijkstra,B.W. (1988) *Protein Eng.*, **2**, 119–125.

Hinck,A.P., Truckses,D.M. and Markley,J.L. (1996) *Biochemistry*, **35**, 10328–10338.

Jacobson,R.H., Matsumura,M., Faber,H.R. and Matthews,B.W. (1992) *Protein Sci.*, **1**, 46–57.

Kanaya,E. and Kikuchi,M. (1992) *J. Biol. Chem.*, **267**, 15111–15115.

Kanaya,S., Katsuda,C., Kimura,S., Nakai,T., Kitakuni,E., Nakamura,H., Katayanagi,K., Morikawa,K. and Ikehara,M. (1991) *J. Biol. Chem.*, **266**, 6038–6044.

Katz,B.A. and Kossiakoff,A. (1986) *J. Biol. Chem.*, **261**, 15480–15485.

Ko,J.H., Jang,W.H., Kim,E.K., Lee,H.B., Park,K.D., Chung,J.H. and Yoo,O.J. (1996) *Biochem. Biophys. Res. Commun.*, **221**, 631–635.

Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.

Li,Y., Coutinho,P.M. and Ford,C. (1998) *Protein Eng.*, **11**, 661–667.

Matsumura,M. and Matthews,B.W. (1991) *Methods Enzymol.*, **202**, 336–356.

Matthews,B.W. (1993) *Annu. Rev. Biochem.*, **62**, 139–160.

Mitchinson,C. and Wells,J.A. (1989) *Biochemistry*, **28**, 4807–4815.

Pabo,C.O. and Suchanek,E.G. (1986) *Biochemistry*, **25**, 5987–5991.

Pace,C.N., Grimsley,G.R., Thomson,J.A. and Barnett,B.J. (1988) *J. Biol. Chem.*, **263**, 11820–11825.

Pantoliano,M.W., Ladner,R.C., Bryan,P.N., Rollence,M.L., Wood,J.F. and Poulos,T.L. (1987) *Biochemistry*, **26**, 2077–2082.

Petersen,M.T., Jonson,P.H. and Petersen,S.B. (1999) *Protein Eng.*, **12**, 535–548.

Pjura,P.E., Matsumura,M., Wozniak,J.A. and Matthews,B.W. (1990) *Biochemistry*, **29**, 2592–2598.

Ramachandran,G.N. and Sasisekharan,V. (1968) *Adv. Protein Chem.*, **23**, 283–437.

Reading,N.S. and Aust,S.D. (2000) *Biotechnol. Prog.*, **16**, 326–333.

Reiter,Y., Brinkmann,U., Webber,K.O., Jung,S.H., Lee,B. and Pastan,I. (1994) *Protein Eng.*, **7**, 697–704.

Robinson,C.R. and Sauer,R.T. (2000) *Biochemistry*, **39**, 12494–12502.

Roesler,K.R. and Rao,A.G. (2000) *Protein Sci.*, **9**, 1642–1650.

Sauer,R.T., Hehir,K., Stearman,R.S., Weiss,M.A., Jeitler-Nilsson,A., Suchanek,E.G. and Pabo,C.O. (1986) *Biochemistry*, **25**, 5992–5998.

Schellman,J.A. (1955) *C. R. Lab. Carlsberg Ser. Chim.*, **29**, 230–259.

Sowdhamini,R., Srinivasan,N., Shoichet,B., Santi,D.V., Ramakrishnan,C. and Balaram,P. (1989) *Protein Eng.*, **3**, 95–103.

Srinivasan,N., Sowdhamini,R., Ramakrishnan,C. and Balaram,P. (1990) *Int. J. Pept. Protein Res.*, **36**, 147–155.

Thornton,J.M. (1981) *J. Mol. Biol.*, **151**, 261–287.

Tidor,B. and Karplus,M. (1993) *Proteins*, **15**, 71–79.

van den Akker,F., Feil,I.K., Roach,C., Platas,A.A., Merritt,E.A. and Hol,W.G. (1997) *Protein Sci.*, **6**, 2644–2649.

Velanker,S.S., Gokhale,R.S., Ray,S.S., Gopal,B., Parthasarathy,S., Santi,D.V., Balaram,P. and Murthy,M.R. (1999) *Protein Sci.*, **8**, 930–933.

Villafranca,J.E., Howell,E.E., Oatley,S.J., Xuong,N.H. and Kraut,J. (1987) *Biochemistry*, **26**, 2182–2189.

Wakarchuk,W.W., Sung,W.L., Campbell,R.L., Cunningham,A., Watson,D.C. and Yaguchi,M. (1994) *Protein Eng.*, **7**, 1379–1386.

Waschutza,G., Li,V., Schafer,T., Schomburg,D., Villmann,C., Zakaria,H. and Otto,B. (1996) *Protein Eng.*, **9**, 905–912.

Wells,J.A. and Powers,D.B. (1986) *J. Biol. Chem.*, **261**, 6564–6570.