Vol. 160, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1989 Pages 109-114

Modification of protein stability by introduction of disulfide bridges and prolines : Geometric criteria for mutation sites

V.N.Balaji §,*, Azita Mobasser §, and Shashidhar N.Rao $,[1]

§ Discovery Research, Allergan Pharmaceuticals Inc., 2525 Dupont Drive, Irvine CA 92715

$ Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco CA 94143

Received February 1, 1989

We define geometrical parameters to characterize disulfide bridges using x-ray crystal structure data on small molecules and use them to suggest replacements of amino acids by cysteines in order to introduce disulfide bridges to increase thermal stability in proteins. We also define geometric parameters to identify target amino acids for replacements by prolines in order to conserve desired structural attributes in the vicinity of disulfide mutations leading to further structural and thermal stability of proteins. The geometric criteria are applied to the serine protease, subtilisin, to model stereochemically favorable disulfide mutants without altering the active site geometry, implying conservation of native biological activity. © 1989 Academic Press, Inc.

The three dimensional structures of proteins are stabilized by non-covalent and covalent interactions [1]. Disulfide bridges between cysteine residues are the most dominant of the covalent interactions known, and are a subject of a number of studies aimed at improving thermal stability of proteins [2-6]. Molecular modeling strategies for incorporating disulfide bridges in proteins have been enunciated recently [7].

We present **a preliminary report** on a method to identify the most probable cysteine mutation sites to promote energetically favored disulfide bridges, by a combination of crystal structure data analysis of small molecules containing disulfides, molecular modeling and energy minimization [8]. The criteria presented require that the conformations of the active site of the

*To whom reprint requests should be addressed.

[1]Present address: G.D. Searle and Co., 4901 Searle Parkway, Skokie, IL 60077.

0006-291X/89 $1.50
Copyright © 1989 by Academic Press, Inc.
All rights of reproduction in any form reserved.

protein change very little upon incorporation of disulfides. We also outline a method to identify amino acid residues to be replaced by prolines, since the latter are known to stabilize the secondary structures such as turns and loops by virtue of the rigid nature of their pyrrolidine ring [9]. We apply our method to the cysteine lacking serine protease subtilisin, to identify amino acid residues which could be replaced by cysteines (to form disulfides) and prolines to render further structural stability.

METHODS

The conformational parameters used to define a disulfide moiety between residues i and j (Fig. 1) were evaluated in 36 disulfide containing compounds from the Crystal Structure Data Base (CSDB) [10], to determine the optimal geometric conditions for disulfide formation. In addition, average bond length C^{β}-S and bond angle C^{β}-S-S were also evaluated from the x-ray structures of the compounds. Based on these evaluations, it was considered feasible to introduce cysteines at residues i and j of a protein to introduce a disulfide between them if : (a) $|i-j| > 10$, (b) τ_1-τ_2 less than or equal to 20^{o}, (c) θ within $\pm$ 50^{o} of 150^{o} or -150^{o} and (d) dij $\leq$ 5 Å. The first of these conditions was chosen to avoid local rigidity in the conformations of the secondary structures due to the formation of the nearest neighbor disulfides and to avoid disruption of the folding profile of the polypeptide. The criteria for the replacement of amino acid residues by prolines are : (a) $-80^{o} < \phi < -60^{o}$ and (b) the residue should lie on the surface and not in the interior core of the protein (Fig. 1b). The residue should be at least 10 Å away from the active site and at least four residues away along the polypeptide backbone from any disulfide introduced through the above criteria.

The crystal structure of the native protein (Katz and Kossiakoff, personal communications) and the disulfide mutants were energy refined using the program AMBER-UCSF [11] employing the force field parameters presented by Weiner *et. al.* [12], in order to rid the model of crystal related distortions in internal parameters. The modeling of disulfides and the computer graphics analysis of the refined structures were carried out using MOGLI [13] and MacroModel [14]. In addition, Balasubramanian and distance plots [15-16] were also used to analyse the secondary structure profiles of the models.

RESULTS AND DISCUSSIONS

The crystal structure analysis for the parameters listed in Fig. 1 finds that the minimum and maximum values of d_{ij} are 3.45 and 4.40 Å, respectively. The pseudo-bond angles τ_1 and τ_2 range from 70^{o} to 150^{o} and have distribution maxima around 75 and 125. Fig. 2 shows an approximate linear correlation between τ_1 and τ_2 ($\tau_1 \approx \tau_2$). χ^3 is restricted to around $\pm 85^{o}$, consistent with *ab initio* calculations [17] which indicate isoenergy minima around $\pm 90^{o}$. Fig. 3 shows that χ^3 is inversely correlated to θ. For example, corresponding to χ^3 values of around -85^{o} and $+85^{o}$, θ varies predominantly around 150^{o} and -150^{o}, respectively. In view of these correlations, we consider replacement of amino acid by cysteines only when τ_1 and τ_2 are nearly equal and θ is around 150^{o} or -150^{o}.

Application of the above described geometric and residue criteria to identify possible proline and cysteine replacements in energy refined native subtilisin led to 11 unstrained disulfides and 10 sites of proline mutation (Table 1). Using the data in Table 1, we have modeled a number of disulfide and proline mutants of subtilisin. In this paper, we present only the mutant with

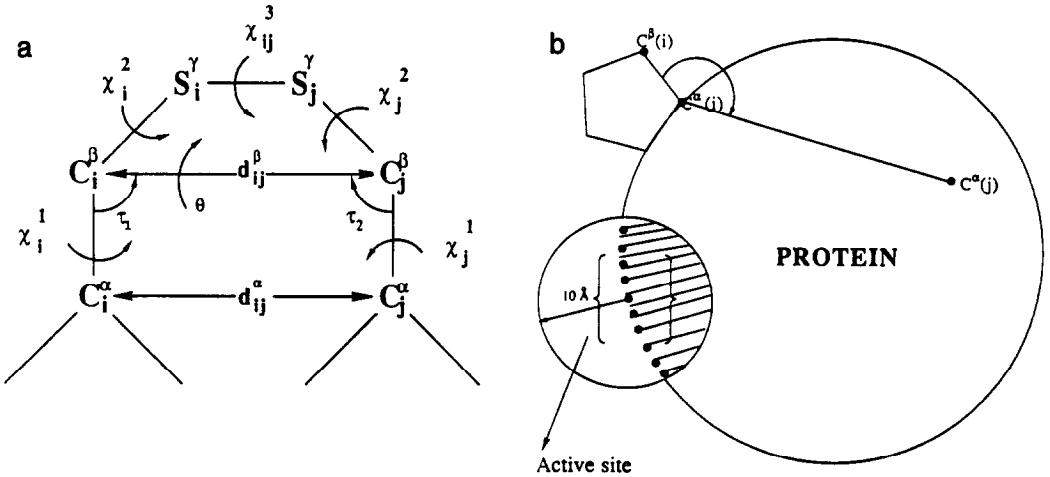

Figure 1 : Schematic representation of (a) the disulfide moiety with the symbols for the conformational parameters defining it : $d_{ij} = C^{\beta}(i)...C^{\beta}(j)$; psuedo bond angles $\tau_1 = C^{\alpha}(i)\text{-}C^{\beta}(i)\text{-}C^{\beta}(j)$ and $\tau_2 = C^{\beta}(i)\text{-}C^{\beta}(j)\text{-}C^{\alpha}(j)$; $\theta = C^{\alpha}(i)\text{-}C^{\beta}(i)\text{-}C^{\beta}(j)\text{-}C^{\alpha}(j)$ and the proper dihedral angles in accordance with the IUPAC-IUB [19] nomenclature **and** (b) the geometric criteria for identifying amino acid residues which could be replaced by proline. Any residue within 10 Å of the active site (hatched region) can not be a candidate for proline replacement. The target residue should lie on the surface : that is, firstly the $C^{\alpha}\text{-}C^{\beta}$ vector in the target residue should point away from the center of the protein and secondly the pseudo bond angle $C^{\beta}(T)\text{-}C^{\alpha}(T)\text{-}C^{\alpha}(i)$, where T is the target residue and i goes from 1 to the number of residues, except for T, should be greater than 60^{o} (only if $C^{\alpha}(i)\text{-}C^{\alpha}(T)$ distance is less than 10 Å). This lower limit for the pseudo bond angle was arrived at by an extensive analysis of the crystal stuctures in the protein data bank (details to be published elsewhere).

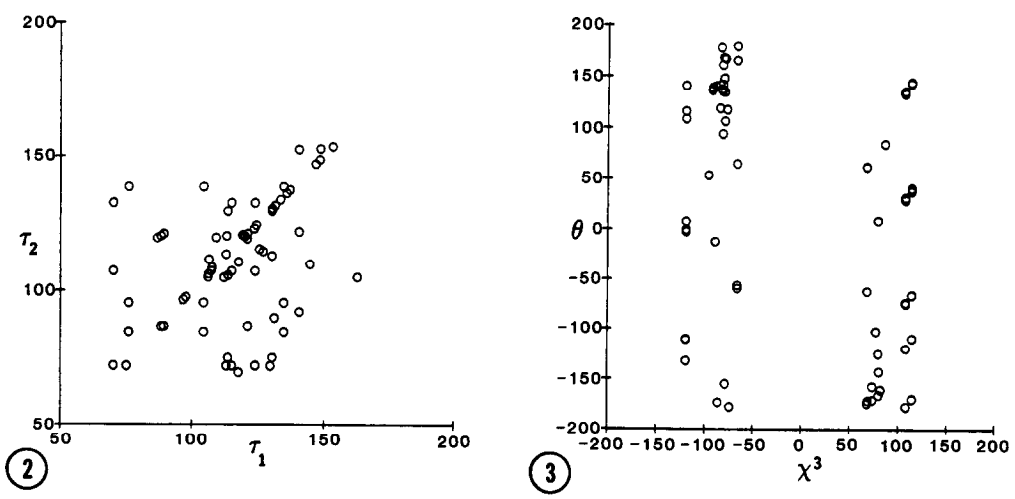

Figure 2 : Plot of τ_1 vs. τ_2 where each square represents a structure in the crystal structure data base (CSDB). τ_1 and τ_2 are approximately linearly correlated.

Figure 3 : Plot of θ vs. χ^3 where each square represents a structure in the CSDB.

TABLE I

Amino acid residues of subtilisin which could be replaced by cysteines to form disulfides (a) and proline to increase the structural and thermal stability conferred by disulfide incorporation (b)

a

Number	Residues i	Name	j	Name	(i-j)	Distance $C^{\beta}..C^{\beta}$	Angles τ_1	τ_2	$(\tau_1-\tau_2)$	θ_2
1	26	VAL	235	LEU	209	4.59	133.3	126.3	7.0	108.3
2	27	LYS	120	ASP	93	4.64	77.7	90.6	12.9	129.3
3	47	GLY	91	TYR	44	4.82	77.2	67.3	9.9	149.4
4	50	MET	109	ASN	59	4.55	130.7	120 7	10.0	148.7
5	73	ALA	88	ALA	15	4.03	127.5	108.3	19.2	-108.4
6	85	ALA	233	LEU	148	4.78	101.0	82.2	18.7	-135.8
7	121	VAL	231	ALA	110	4.82	134.4	130.8	3.6	125.6
8	164	THR	191	SER	27	4.14	93.6	84.3	9.3	-145.1
9	196	LEU	264	GLY	68	4.23	85.5	94.0	8.5	-127.2
10	234	ILE	274	ALA	40	4.81	66.9	64.6	2.3	-146.3
11	254	THR	266	GLY	12	3.99	112.1	92.2	19.8	139.1

b

RESIDUE	SER9	SER105	ASP140	SER162	GLN185	THR242	ARG247	GLN251	LYS256
ϕ (in ^{o})	-67	-77	-65	-76	-76	-70	-76	-70	-73

In (a), the residue pairs correspond to the disulfides criteria $d_{ij} \leq 5$ Å, $\tau_1-\tau_2 \leq 20^{o}$ and θ within $\pm 50^{o}$ of $+ 150^{o}$ or -150^{o}.

disulfide across residues 73 and 88, as an example. Details of the investigations on other mutants will be published elsewhere.

Computer graphics examination of the models of energy refined native and 73-88 mutant subtilisin shows no gross structural changes in the active site region, upon disulfide incorporation. This is not surprising, since these residues are at least 15 A away from the catalytic triad (SER221, ASP32 and HIS64). The root mean square deviations for the polypeptide backbone atoms in the energy refined native and mutant structures are around 0.2 Å. The parameters θ and χ^3 have values of respectively -110^{o} and -115^{o} in the refined mutant structure and these values are well within the ranges observed from x-ray data.

Our approach differs from that in [7] as follows : (a) In the earlier approach, the disulfide geometric parameters (bond lengths, bond angles and torsions) were taken only from protein crystal structures, while we have investigated a large number of small molecular crystal structures with better resolved x-ray data (lower R factor) than proteins. (b) The protein disulfides include some highly strained cases, formed under the bias of tertiary folding pattern of

the proteins, whereas the disulfides in our study contain far fewer strained cases. This is significant as it is the aim of our study to identify only those mutations which do not disrupt the active site conformation and structure and hence the biological activity of the protein. (c) The computational aspect of our approach is much simpler than in the approach proposed earlier [7]. In the latter, disulfides are identified by comparison of spatial relationships of residues with those in a disulfide data base from protein data bank. We, on the other hand, avoid the definition of local coordinate systems and need to evaluate only a few simple "bond length and angle" parameters.

Recent crystal structure studies on the 24-87 and 22-87 disulfide mutants of subtilisin [18] show considerable strain in the polypeptide backbone. In fact, they observe two short contacts between the backbone nitrogen and the cysteine sulfur in the 22-87 mutant. These structures correspond to conformationally strained values of C^{β}-S torsion. In contrast, our model built and energy refined 73-88 disulfide mutant has values of $\chi 1_{73}$ (-40°), $\chi 2_{73}$ (-75°), $\chi 3$ (-115°), $\chi 2_{88}$ (93°), $\chi 1_{88}$ (30°) which lie in the energetically favorable ranges (unpublished data).

CONCLUSIONS

We have outlined model building strategies to introduce disulfides and prolines in proteins by defining geometric parameters in a model disulfide moiety and identifying conditions for proline substitution. The defined parameters have been measured from small molecule x-ray data. Our approach enables disulfide and proline incorporation with minimal distortion in the polypeptide backbone and retention of the active site and hence the biological activity of the protein. Our approach will be useful for genetically engineering mutant proteins which are thermally stable and have desired biological activity and hope that this study will be an inspiration along such lines.

ACKNOWLEDGMENTS

We thank Professor Peter Kollman, Drs. M.E.Wolff, M. Garst and S.Profeta, Jr., for many discussions and suggestions, Drs A.Kossiakoff and B.Katz for subtilisin coordinates. SNR thanks the National Cancer Institute (grant CA-25644 to Peter Kollman, School of Pharmacy, University of California, San Francisco) for research support. Some of the initial energy minimization studies were carried out on FPS264 at the University of California, San Francisco, which was purchased through grants from the NSF (DMB-84-13762) and NIH (RR-02441). This support is very much appreciated.

REFERENCES

1) Fersht, A.R. (1985) Enzyme Structure and Mechanism, Second Edition, W.H.Freeman and Company, New York.

2) Sporn, M.B., Roberts, A.B., Wakefield, L.M. and Assoian, R.K. (1986) Science, **233**, 532-534.

3) Estell, D.A., Graycar, T.P., Miller, J.V., Powers, D.B., Burnier, J.P., Ng, P.G. and Wells, J.A. (1986) Science **233**, 659-633.

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

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

6) Wetzel, R. (1987) Tre. In Biol. Sci. **12**, 478-482.

7) Pabo, C.O. and Suchanek, E.G. (1986) Biochemistry **25**, 5992-5998.

8) Balaji, V.N., Rao, S.N., Dietrich, S.W. and Kollman P.A. Poster presentation based on the early stages of this work was made at the UCLA symposium on "Peptide and Protein Structures" held at Steamboat Springs, Colorado from April 4 to April 10, 1987.

9) Venkatachalam, C.M. (1968) Biopolymers **6**, 1425-1436.

10) Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R. and Watson, D.G. (1979) Acta Crystallogr. **B35**, 2331-2339.

11) Singh, U.C., Weiner, P.K., Caldwell, J.W. and Kollman, P.A. (1985) AMBER(UCSF) version 3.0 (developed for FPS) Department of Pharmaceutical Chemistry, University of California, San Francisco CA 94143.

12) Weiner, S.J., Kollman, P.A., Case, D., Singh, U.C., Ghio, C., Alagona, J., Profeta, Jr., S. and Weiner, P.K, (1984) J. Am. Chem. Soc. **106**, 765-785.

13) MOGLI version 2.0, Evans and Sutherland Computer Corporation, Salt Lake City, Utah 84108.

14) MacroModel v1.1. The Columbia University Molecular Modeling System, Columbia University, New York, (1986).

15) Balasubramanian, R. (1977) Nature **266**, 856-857.

16) Liljas, A. and Rossman, M.G. (1974) Annu. Rev. Biochem. **43**, 475-507.

17) Ha, T.K. (1985) J. Mol. Str. **122**, 225-234.

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

19) IUPAC-IUB Joint Commission on Biochemical Nomenclature (1983) Eur. J. Biochem., **131**, 9-15.