Protein Science (1992), *1*, 91-106. Cambridge University Press. Printed in the USA.

Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance

MARK R. HURLE,[1,3] CHARLES D. EADS,[1,4] DAVID A. PEARLMAN,[1,5] GEORGE L. SEIBEL,[1,6] JOHN THOMASON,[1] PHYLLIS ANNE KOSEN,[1] PETER KOLLMAN,[1] STEPHEN ANDERSON,[2,7] AND IRWIN D. KUNTZ[1]

[1] Department of Pharmaceutical Chemistry, The University of California at San Francisco, San Francisco, California 94143
[2] Department of Cardiovascular Research, Genentech, South San Francisco, California 94080

(RECEIVED June 27, 1991; ACCEPTED September 6, 1991)

Abstract

Structural perturbations due to a series of mutations at the 30-51 disulfide bond of bovine pancreatic trypsin inhibitor have been explored using NMR. The mutants replaced cysteines at positions 30 and 51 by alanine at position 51 and alanine, threonine, or valine at position 30. Chemical shift changes occur in residues proximate to the site of mutation. NOE assignments were made using an automated procedure, NASIGN, which used information from the wild-type crystal structure. Intensity information was utilized by a distance geometry algorithm, VEMBED, to generate a series of structures for each protein. Statistical analyses of these structures indicated larger averaged structural perturbations than would be expected from crystallographic and other information. Constrained molecular dynamics refinement using AMBER at 900 K was useful in eliminating structural movements that were not a necessary consequence of the NMR data. In most cases, statistically significant movements are shown to be those greater than approximately 1 Å. Such movements do not appear to occur between wild type and A30A51, a result confirmed by crystallography (Eigenbrot, C., Randal, M., & Kossiakoff, A.A., 1990, *Protein Eng. 3*, 591-598). Structural alterations in the T30A51 or V30A51 mutant proteins near the limits of detection occur in the β-loop (residues 25-28) or C-terminal α-helix, respectively.

Keywords: bovine pancreatic trypsin inhibitor proteins; molecular dynamics; mutants; NMR

Nuclear magnetic resonance (NMR) spectroscopy is an increasingly important tool for evaluating the structures of small proteins in solution (Wüthrich, 1986; Kaptein et al., 1988). It has become possible to determine de novo protein structures, largely because of methodologies for assigning the chemical shifts of virtually all protons in a small protein (Billeter et al., 1982; Wüthrich et al., 1982). Several spectral parameters reflect structural features. These include the nuclear Overhauser effect (NOE), chemical shifts, and coupling constants. The geometric consequences of these data cannot be realized directly. Instead, some form of model building is required. In this paper, we use distance geometry (DG) (Kuntz et al., 1989) and molecular dynamics (MD) (Scheek et al., 1989) as two approaches toward the determination of three-dimensional structures.

We apply the methods to explore the structural consequences of a series of amino acid substitutions in the well-studied protein, bovine pancreatic trypsin inhibitor (BPTI). BPTI is an ideal molecule for this study (Kassell, 1970). It is small (58 residues), stable ($T_m > 100$ °C), soluble, and is therefore well suited for NMR studies. The chemical shifts of nearly all protons have been assigned by Wüthrich's laboratory (Wagner et al., 1987a,b), and it is one of the first protein molecules to have its structure determined by NMR methods. There are several

Reprint requests to: Irwin D. Kuntz, Department of Pharmaceutical Chemistry, The University of California at San Francisco, San Francisco, California 94143.

[3] Present address: Department of Macromolecular Science, SmithKline Beecham, King of Prussia, Pennsylvania 19406.

[4] Present address: Procter and Gamble, Miami Valley Labs, P.O. Box 398707, Cincinnati, Ohio 45239-8707.

[5] Present address: Vertex Corporation, 40 Allston St., Cambridge, Massachusetts 02139.

[6] Present address: Department of Physical and Structural Chemistry, SmithKline Beecham, King of Prussia, Pennsylvania 19406.

[7] Present address: Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854.

high resolution crystal structures available (Wlodawer et al., 1987). The disulfide formation and rearrangement steps along the folding pathway have been determined (Creighton, 1974, 1977; Creighton & Goldenberg, 1984).

Recent work in our laboratory and others has focused on determining the effects of amino acid substitutions at several sites, including the cysteine residues involved in disulfide bonds, on the folding pathway (Berman-Marks et al., 1987; Goldenberg, 1988) and stability (Hurle et al., 1990) of the protein. Of special interest among these mutants is a set involving cysteine residues 30 and 51, which form a buried disulfide bridge in the native protein structure (Fig. 1; Kinemage 1). The antiparallel β-sheet and the C-terminal α-helix are joined by this linkage. A significant amount of work has been carried out to study the folding of these mutants from reduced (Berman-Marks et al., 1987; P.A. Kosen, unpubl.) and oxidized (Hurle et al., 1990) unfolded forms. Both kinetics and stability are altered by the mutations. Oas and Kim (1988) have synthesized fragments of BPTI that form a stable, native-like structure upon formation of the 30–51 disulfide bond. Thus, the structural consequences of mutations at these positions are of great interest.

We report here studies of BPTI derivatives in which the cysteine residue at position 51 is replaced by alanine, and the cysteine residue at position 30 is replaced with alanine, valine, or threonine (A30A51, V30A51, and T30A51, respectively). Because the disulfide is buried, substitutions at these positions introduce changes in the packing of the hydrophobic core. In a study parallel to the one reported here, Kossiakoff and coworkers (Eigenbrot et al., 1990) have carried out a high resolution X-ray structure study of the A30A51 mutant. Their analysis indicates that the structural changes are quite small for this mutant. X-ray analyses of the V30A51 and T30A51 molecules have not been completed because of differing crystallization properties, suggesting that larger structural changes exist in these molecules. NMR is not dependent on the crystallization properties of the molecule and can provide information not presently available from crystallography.

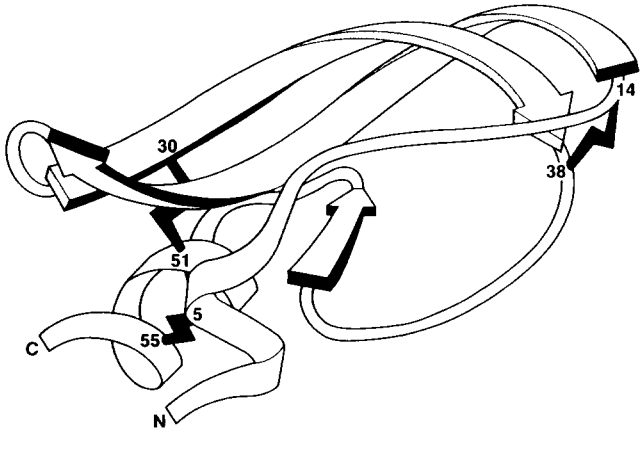

Fig. 1. Backbone structure BPTI, modified from Richardson (1981).

Results

Sequential assignment

Assignment of proton chemical shifts was based on published values for the wild-type protein (Wagner et al., 1987a,b). For each mutant, peak coordinates in the fingerprint region of the double-quantum filtered correlated spectroscopy (2QF-COSY) spectrum were compared to positions of the wild-type fingerprint peaks. By analogy to the wild type, the majority of fingerprint peaks could thereby be assigned. In a few cases the NOE spectroscopy (NOESY) spectrum had to be consulted to identify characteristic patterns of sequential connectivity to resolve ambiguity. The chemical shifts of protons for each side chain were identified using homonuclear Hartman–Hahn (HOHAHA) spectra, which give chemical shifts for most protons in each spin system, and with 2QF-COSY spectra. Again, peak identification relied heavily on published chemical shift values and was verified where possible using COSY data in the usual way. Wild-type chemical shift values determined in our laboratory and mutant chemical shift differences are given in Tables 1 and 2 (supplementary information).

Figure 2 shows the difference in the chemical shift for amide and C^{α} protons between the wild type and the three mutants. In the C^{α} plots, it is clear that a large number of atoms are not affected by the amino acid substitution, whereas the remainder show an unambiguous chemical shift perturbation. A similar interpretation of the amide proton chemical shifts holds, except there is greater variability in the chemical shifts in all regions of the molecule. This is due not only to structural differences between the molecules, but also to the greater sensitivity of amide proton chemical shifts to experimental conditions.

A molecular graphics representation of the crystal structure of the molecule in which residues with large chemical shift differences are highlighted (see Kinemage 2) indicates that the measurable chemical shift perturbations occur only in the region surrounding the amino acid substitutions. Thus, the chemical shift perturbations resulting from mutations are a sensitive and rapid method for evaluating the spatial location of structural perturbations caused by mutations if a structure is known. Major chemical shift markers for BPTI tertiary structure, including the high field resonances of Pro 9, Gly 37, Asn 44, and the α-carbon proton of residue 51, are unaffected by these mutations, suggesting that the overall tertiary structures are identical. However, a structural interpretation of the chemical shift changes has not been attempted because theories for chemical shifts are currently insufficiently developed.

Table 1. *Summary of sequence-specific proton resonance assignments for wild-type BPTI (pH 4.6, 36 °C)*

Residue	NH	α	β	β′	Other assignments
Arg 1		4.37	1.89	1.80	1.48[a] (γ) 1.35[a] (γ′) 3.08[a] (δ) 2.88[a] (δ′) 7.05[a] (ε)
Pro 2		4.33	2.03	0.92	1.87 (γ) 1.59 (γ′) 3.73 (δ) 3.59 (δ′)
Asp 3	8.66	4.25	2.77		
Phe 4	7.81	4.60	3.35		7.00 (δ) 7.37 (ε) 7.32 (ζ)
Cys 5	7.45	4.35	2.86	2.74	
Leu 6	7.58	4.51	1.84		1.71 (γ) 0.96 (δ) 0.86 (δ′)
Glu 7	7.51	4.59	2.26	2.17	2.62 (γ) 2.57 (γ′)
Pro 8		4.64	2.44	1.84	2.11 (γ) 3.98 (δ) 3.70 (δ′)
Pro 9		3.72	0.22	0.09	1.27 (γ) 0.15 (γ′) 3.34 (δ) 2.91 (δ′)
Tyr 10	7.80	4.95	2.96		7.33 (δ) 7.09 (ε)
Thr 11	8.95	4.53	4.05		1.38 (γ)
Gly 12	7.16	3.90 3.26			
Pro 13		4.56	2.17	2.11	1.98 (γ) 3.63 (δ) 3.59 (δ′)
Cys 14	8.69[b]	4.56	3.47	2.79	
Lys 15	7.98	4.41	2.09	1.58	1.27[b] (γ)
Ala 16	8.22	4.29	1.18		
Arg 17	8.18	4.32	1.60		1.48[b] (γ) 1.31 (γ′) 3.13 (δ) 7.18[a] (ε)
Ile 18	8.11	4.20	1.87		1.37 (γCH_2) 0.97 ($\gamma' CH_2$) 0.97 (γCH_3) 0.69 (δ)
Ile 19	8.68	4.31	1.95		1.47 (γCH_2) 1.40 ($\gamma' CH_2$) 0.72 (γCH_3) 0.68 (δ)
Arg 20	8.39	4.70	1.81	0.81	1.81[b] (γ) 1.32 (γ′) 3.48 (δ) 3.03 (δ′) 7.44 (ε)
Tyr 21	9.19	5.69	2.70		6.71 (δ) 6.79 (ε)
Phe 22	9.78	5.27	2.91	2.82	7.16 (δ) 7.24 (ε) 7.32 (ζ)
Tyr 23	10.55	4.29	3.47	2.73	7.18 (δ) 6.33 (ε)
Asn 24	7.77	4.61	2.86	2.17	7.90 (δ) 7.09 (δ′)
Ala 25	8.77	3.76	1.56		
Lys 26	7.91	4.07	1.89		1.54[b] (γ) 1.46 (γ′) 1.73[a] (δ) 3.05 (ε) 7.53[a] (ζ)
Ala 27	6.82	4.29	1.18		
Gly 28	8.13	3.92 3.61[b]			
Leu 29	6.82	4.74	1.73	1.43	1.43 (γ) 0.85 (δ) 0.76 (δ′)
Cys 30	8.39	5.61	3.67	2.67	
Gln 31	8.77	4.83	2.15	1.73	2.23 (γ) 1.90 (γ)
Thr 32	8.04	5.29	4.04		0.59 (γ)
Phe 33	9.36	4.86	3.09	2.96	7.09 (δ) 7.12 (ε) 7.00 (ζ)
Val 34	8.36	3.92	1.95		0.81 (γ) 0.71 (γ′)
Tyr 35	9.39	4.89	2.66	2.51	7.78 (δ) 6.69 (δ′) 6.87 (ε) 6.78 (ε′)
Gly 36	8.60	4.31 3.25			
Gly 37	4.34	4.23 2.91			
Cys 38	7.76	4.95	3.96	3.03	
Arg 39	9.07	3.94	2.27		1.59 (γ) 3.28 (δ) 7.29 (ε)
Ala 40	7.38	4.08	1.20		
Lys 41	8.32	4.45	2.26	1.65	
Arg 42	8.32	3.67	1.03	0.37	1.49 (γ) 1.23 (γ′) 2.84 (δ) 2.69 (δ′) 7.11 (ε)
Asn 43	7.20	5.06	3.34	3.26	7.97 (δ) 7.78 (δ′)
Asn 44	6.77	4.90	2.78	2.51	
Phe 45	9.94	5.13	3.41	2.79	7.40 (δ) 7.87 (ε) 7.64 (ζ)
Lys 46	9.90	4.38	2.10	1.99	1.62 (γ) 1.48 (γ′)
Ser 47	7.47	4.54	4.12	3.87	
Ala 48	8.14	3.15	1.04		
Glu 49	8.55	3.88	2.02	1.86	2.36 (γ) 2.22 (γ′)
Asp 50	7.84	4.29	2.88	2.73	
Cys 51	7.00	1.70	3.18	2.88[b]	
Met 52	8.57	4.16	2.06	1.97	2.68 (γ) 2.17 (ε)
Arg 53	8.29	3.98	1.93	1.87	1.73 (γ) 1.62 (γ′) 3.21 (δ) 7.26 (ε)
Thr 54	7.40	4.08	4.00		1.61 (γ)
Cys 55	8.24	4.62[b]	2.25	2.01	
Gly 56	7.95	3.85			
Gly 57	8.18	3.97 3.81			
Ala 58	7.93	4.02	1.32		

(continued)

Table 1. *Continued*

[a] An assignment not reported previously by Wagner et al. (1987b) for the NMR spectrum of BPTI at pH 4.6 and 36 °C.

[b] The assigned chemical shift differs by ±0.05 ppm or more compared to that reported by Wagner et al. (1987b).

Assignment of NOE cross-peaks

The cross-peaks in the NOESY spectrum were assigned using an automated procedure, NASIGN, based on knowledge of the crystal structure of the wild-type molecule. The procedure involves the following steps. The peak coordinates and intensities are collected into a file. For each peak in the file, the program compiles a list of all pairs of protons that could have given rise to the peak based on the chemical shifts with a user-specified tolerance. For each of the candidate pairs of protons, the program then consults the crystal structure and calculates the interproton distance. If the interproton distance is less than a user-specified cutoff, then that peak is flagged in the output. Finally, those peaks that can be ascribed to only one pair of protons within the above criteria are written to the file of uniquely assigned peaks used in the structural analysis.

The chemical shift tolerance and distance cutoff were varied systematically to find the combination leading to the maximum number of uniquely assigned NOESY peaks. These values are 0.015 ppm and 4.75 Å, respectively. For our acquisition parameters, 0.015 ppm is close to the digital resolution, and 4.75 Å is close to the upper limit of distances measurable by NMR. Thus these parameters are reasonable. These parameters allowed us to assign uniquely 711, 870, 634, and 888 cross-peaks for the wild type, A30A51, V30A51, and T30A51 proteins, respectively. For the A30A51 mutant, a NOESY data set was acquired in D_2O as well, leading to the assignment of several peaks that were obscured by solvent saturation and t_1 noise in the H_2O spectra.

This analysis is based on the assumption that the solution structures of the wild-type and mutant molecules are

Table 2. *Differences in proton resonance assignments for the 30–51 mutants*[a]

Residue	NH	α	β	β′	Other assignments
Arg 1					0.06t (γ′)
					0.05a (δ)
					0.15a, 0.06v (δ′)
					0.14a, 0.11t (ε)
Asp 3	0.05a				
	0.08t				
Cys 5		0.06t			
	0.07v	0.07v			
Leu 6	0.06t				
	0.05v	−0.08t,v			
Glu 7	0.06a				
Pro 9		−0.05a,t	0.09a		0.06a, 0.05v (γ′) −0.05a (δ)
			−0.08t	−0.07t	−0.24a, −0.13t (δ′)
			−0.12v	−0.13v	
Tyr 10	−0.05v,t				0.15a (ε)
Pro 13		0.05t			
Arg 17					−0.15t, −0.16v (γ) −0.05t (γ′)
Ile 19			−0.05v		−0.05t, −0.07v (γCH$_3$)
Arg 20				−0.05a,v	−0.07a (δ) 0.05a (δ′)
		−0.05t		−0.06t	0.06a (ε)
		−0.08v	−0.05v		
Tyr 21	−0.08a	−0.05a	(2.61a)[b]	(2.48a)[b]	
	−0.16t	−0.17t	−0.5t		0.13t, 0.14t (δ)
	−0.28v	−0.15v	(2.68v)[b]	(2.64v)[b]	
Phe 22	0.10a	0.08a,t	(2.86a)[b]		
	−0.08t				
	−0.06v	0.07v	0.07v		−0.07v (δ)
Tyr 23		−0.16a	−0.10a	−0.32a	−0.30a,t −0.35v (δ)
	−0.23t		−0.14t	−0.33t	0.08a, 0.11t,v (ε)
	−0.35v		−0.09v	−0.36v	
Asn 24	0.07a	0.05a		−0.05a	
	0.25t	0.06t	0.05t		
	0.30v		0.05v		0.08v (δ)
Ala 25	0.19a	0.10a			
	0.07t	0.15t			
		0.14v	0.06v		
Lys 26	0.05a				
	0.12t,v				
Ala 27					
Gly 28	0.06a				
Leu 29	−0.07a		−0.07a	−0.08a	
		0.06t	−0.06t	−0.08t	
	−0.06v	−0.07v	−0.20v	−0.18v	−0.07v (γ)
Ala 30[c]	−0.29a[c]	−0.52a	(1.30a)[b]		
Thr 30[c]	−0.23t[c]	−0.77t	(4.00t)[b]		(1.29t)[b] (γ)
Val 30[c]		−1.14v	(2.02v)[b]		(1.10v, 0.97v)[b]
Gln 31	−0.13a		(1.60a)[b]		0.1a, 0.12t, 0.08v (γ)
	−0.37t		0.16t	−0.12t	0.06a (γ′)
	−0.28v	0.07v	0.06v	−0.11v	(6.93t, 7.48t)[b] (ε,ε′)
					(6.95v, 7.42v)[b] (ε,ε′)
Thr 32	0.05a	−0.10a			
	−0.07t	0.17t			
	−0.10v	0.27v			0.11v (γ)
Phe 33	−0.08a				0.08a (ζ)
			−0.05t		
	−0.06v		−0.11v		−0.07v (δ)
Val 34	0.06a				
Tyr 35		0.05a			0.09a (ε), −0.07a (ε′)
Gly 36	0.05a				
Arg 39	0.05t				
Ala 40	0.05t				

(*continued*)

Table 2. *Continued*

Residue	NH	α	β	β′	Other assignments
Lys 41			−0.06t		(1.32a,v 1.31t)[b] (γ)
					(1.47v,t)[b] (γ′)
					(1.95t, 1.98v)[b] (δ)
					(3.04t, 3.03v)[b] (ε)
Arg 42	0.06a		−0.06t		0.05a (ε)
Asn 43				−0.05a	−0.05, −0.20a (δ,δ′)
				−0.09t	0.17t (δ)
				−0.06v	0.20a (δ)
Asn 44	−0.05t				(7.81t,v)[b] (δ)
					(3.39t, 3.38v)[b] (δ′)
Phe 45	0.15a		0.06a	0.05a	0.07a, −0.07a (δ,ζ)
			0.07t	0.07t	0.09t, −0.05t (δ,ζ)
			0.06v	0.08v	0.11v (δ)
Lys 46		0.06a,v,t			(1.80t,v)[b] (δ)
					(3.06t)[b] (ε)
Ser 47		−0.18a	−0.08a	−0.05a	
	0.06t	−0.13t	−0.05t		
	0.07v	−0.11v	−0.05v		
Ala 48	−0.26a				
	−0.19t	0.23t			
	−0.10v	0.27v			
Glu 49	−0.05a	−0.06a			
	−0.09t				
	−0.05v				
Asp 50				0.05a	
	−0.10t			0.08t	
	−0.12v			0.09v	
Ala 51[c]		(1.83a)[b]	(1.03a)[b]		
		(1.84t)[b]	(1.07t)[b]		
		(1.87v)[b]	(1.07v)[b]		
Met 52	−0.48a				(2.72a, 2.66t, 2.56v)[b] (γ)
	−0.48t	−0.19t			(2.60a, 2.62t, 2.53v)[b] (γ′)
	−0.42v	−0.24v			−0.13t (ε)
					−0.07v (ε)
Arg 53	0.16a				
	0.05t				
	0.07v				
Thr 54		−0.06a			
	0.07t				
	0.12v				
Cys 55	−0.41a	−0.65a	−0.16a	−0.09a	
	−0.31t	0.06t,v	(2.15t)[b]		
	−0.36v		(2.24v)[b]		
Gly 56	−0.12a	−0.07a			
	−0.15t	(3.75t, 3.79t)[b]			
	−0.12v	(3.70v, 3.86v)[b]			

[a] The mutant associated with the reported change is indicated by the letter following, a, t, or v for A30A51, T30A51, and V30A51, respectively. Only changes larger than 0.05 ppm are shown.
[b] Values in parentheses are actual chemical shifts, not differences.
[c] Residue has changed from wild type.

similar to the crystal structure of the wild type. There are several reasons to believe this assumption is valid. The chemical shifts of all molecules are quite similar, precluding the possibility of a large conformational change. The mutants all bind to chymotrypsin without being proteolyzed (a chymotrypsin affinity column is part of the purification procedure), indicating structural integrity. Approximately 75% of all observed NOESY peaks could be assigned using the protocol, which reinforces the validity of the assumption. This approach does not preclude the detection of structural differences among the mutants and wild-type solution structures for two reasons. First, the NOESY peak intensities were interpreted in terms of distances independently of the crystal struc-

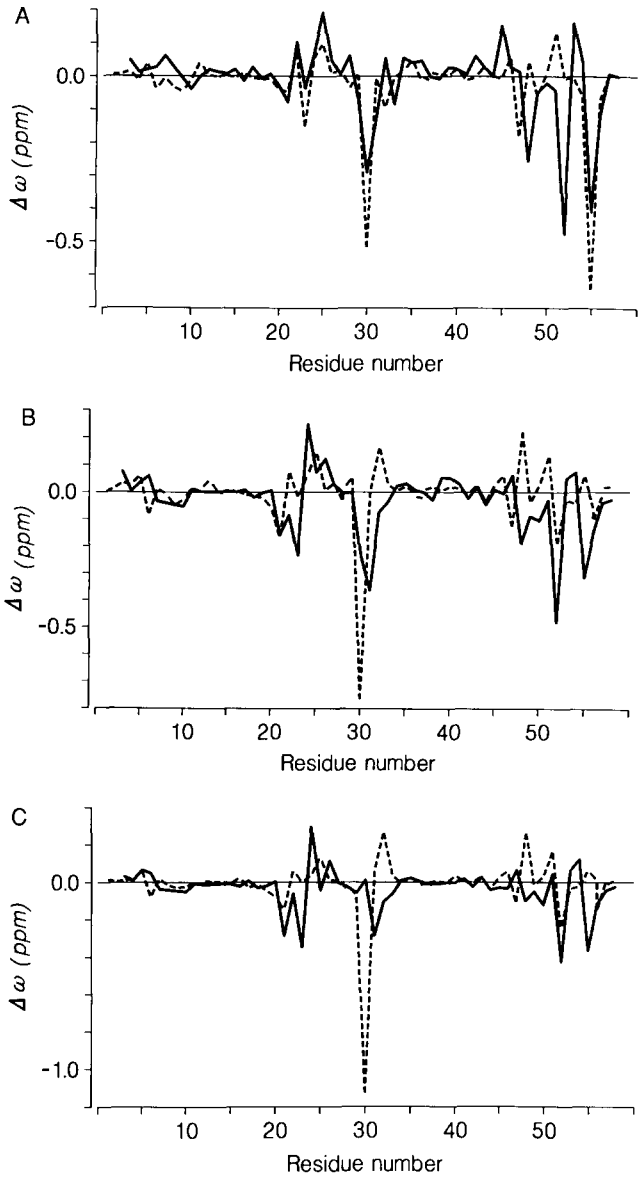

Fig. 2. Difference in chemical shifts between wild-type BPTI and BPTI mutants (**A**) A30A51, (**B**) T30A51, and (**C**) V30A51 for amide (—) and C^{α} (---) protons. Chemical shifts were acquired at 36 °C, pH 4.6, against a 3-(trimethylsilyl)tetradeutero sodium propionate (TSP) reference and are reported as parts per million.

ture. Second, a second iteration of the use of NASIGN using a longer cutoff distance (6 Å) and excluding previously assigned peaks was included. Approximately 25 and 50 additional peaks were thereby assigned in the V30A51 and T30A51 data sets, respectively, virtually all in the vicinity of the mutation.

Assignment of distance constraints

The changes in the NOE peak assignments and their respective intensities observed within the group of four proteins presumably reflect structural, dynamic, and electronic differences between the proteins. In order to put these differences on a structural basis, the conversion of NOE peak intensity to distance is an important issue. The more limiting we make these distance constraints, the smaller will be the conformational space that will describe the structure. We need to make sure that our constraints are physically reasonable, in light of the factors affecting NOE peak intensities.

Even for a small protein, many NOESY peaks overlap each other at 500 MHz. For the four proteins studied here, approximately 40% of the total number of assigned peaks are overlapping with other peaks. Because the maximum observable distance of an NOE peak intensity is approximately 4.5 Å, we have set the maximum distance constraint of all overlapping peaks at this distance.

The remaining isolated peaks, which form the bulk of the assignments, are not easily converted to respective distances. Figure 3 shows there is a significant spread in the relationship between the sixth root of the intensity of a wild-type isolated NOE peak and its distance in the wild-type crystal structure, due to measurement error, spin diffusion, or dynamic effects. The spread for peaks of identical intensity is 1.5–2 Å.

Because the mutations are located at one end of the long axis of the molecule, we expect that most NOE intensities should not change. The structure of most of the protein is not affected as reported by chemical shifts (Kinemage 2), thus the correlation time should remain constant. Figure 4 shows the distribution of the standard deviation of the percentage change in the sixth root of intensity, i.e., change in distance, for the isolated NOE peaks (~200) assigned in at least three of the four proteins. It can be seen that 90% of the intensities are perturbed by a factor that corresponds to a distance change of less than 10%. This suggests a measurement error of somewhat less than that amount. Note that 10% of 4.5 Å (maximum observable distance) is 0.45 Å. Given the spread of Figure 3, an estimate of 0.5 Å for error in con-

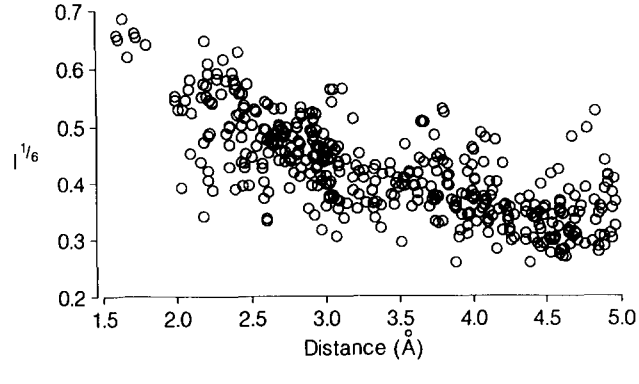

Fig. 3. Sixth root of observed intensity of isolated NOE peaks determined for wild-type BPTI plotted against the distance between the assigned protons in the wild-type crystal structure (4PTI). Experimental conditions were identical to those of Figure 2.

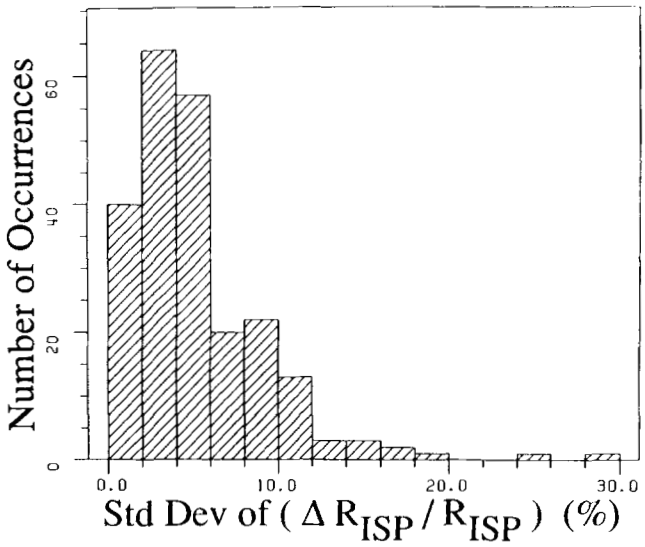

Fig. 4. Distribution of standard deviations of percentile changes in the sixth root intensities of NOE peaks in the set of four proteins. Intensities were scaled before determination of standard deviation of sixth root. Scaling was performed by summing the intensities of all assigned isolated NOE peaks in common in all four proteins (wild-type, A30A51, T30A51, and V30A51) and then multiplying the intensities of each protein set by the appropriate scaling factor. The set of intensities reported in the figure includes all assigned isolated peaks that occur in a minimum of three out of the four proteins.

version to distance constraints seems reasonable as noted below.

Distance geometry

For each of the data sets, a set of at least 10 structures was generated using DG as described above. In approximately $\frac{1}{3}$ of the structures, the handedness of the structure was incorrect. These structures had larger final error functions and were discarded. The remaining structures, for all molecules, had an organization of secondary structural elements in line with that of the published crystal structures. This result was expected based on the success of using the crystal structure to aid in assigning NOESY cross-peaks. The organization of secondary structural elements is similar for all of the DG structures of the wild-type and three mutant proteins. Table 3 shows that the backbone (C, N, C^{α}) atoms of the DG structures have a root-mean square (RMS) spread among themselves (1.5 Å) comparable to the RMS distance of each DG structure to the wild-type crystal structure (1.7 Å), with the slight exception of the V30A51 DG structures. A more detailed analysis is required to determine whether these RMS differences represent real structural differences among the mutant and wild-type structures (see below).

Although it has been the practice to generate a collection of structures to characterize the uncertainties of an NMR data set, an alternative approach is to calculate an average structure. During the calculation of the average coordinate for each of the atoms, one can also calculate some measure of the spread in coordinates for each atom. Thus, by reporting the average coordinates and the spread, one can characterize a collection of structures with a single structure.

One can also use the average structure to compare the sets of DG structures from the four proteins. Table 3 demonstrates that the average DG structure is located near the center of the envelope of DG structures for a particular protein, as the average RMS of the backbone of each DG structure to the average DG structure (1 Å) is approximately 60% of the total spread for each protein. However, Table 4 shows that the average DG structures for the four proteins have higher RMS deviations to each other (1.5 Å), suggesting that structural differences between these proteins are greater than the spread of each of the structures.

Table 3. *Distance geometry*

Genotype	NOESY distances	Number of structures[a]	Avg. RMS[b] to wild type X-ray[c]	RMS[b] spread[d]	Avg. RMS[b] to avg. DG structure[e]	DG error[f] of each structure[g]	DG error[f] of DG-minimum structure
Wild type	711	9	2.03	1.49	0.99	167	29
A30A51	870	10	1.58	1.58	1.05	122	6
T30A51	888	13	1.50	1.30	0.89	85	11
V30A51	634	15	2.11	1.59	1.08	131	24

[a] With correct handedness of helical segments.
[b] Root-mean square distance (Å) between corresponding N, C, and C^{α} atoms.
[c] 4PTI (reference).
[d] Average RMS distance (Å) for each structure–structure pair.
[e] Determined as defined in the text.
[f] Defined as sum of squares of violations, where violations are defined as $(D_{ij}^2 - D_{\text{bound}}^2)$ where D_{bound} are the NMR and holonomic distance constraints.
[g] Average error in $Å^4$ for all DG structures.

Table 4. *Root-mean square values of backbone mutant structures*[a]

Genotype	Wild type	A30A51	T30A51	V30A51
Avg. DG structures				
Wild type	–	1.50	1.49	2.33
A30A51	–	–	1.17	1.49
T30A51	–	–	–	1.83
V30A51	–	–	–	–
Avg. 298 K MD structures[b]				
Wild type	–	1.15	1.31	1.54
A30A51	–	–	1.04	1.21
T30A51	–	–	–	1.13
V30A51	–	–	–	–
Avg. 900 K MD structures[b]				
Wild type	–	0.87	0.96	1.57
A30A51	–	–	0.99	1.62
T30A51	–	–	–	1.68
V30A51	–	–	–	–

[a] Root-mean square distance (Å) between corresponding N, C, and C^{α} atoms.
[b] Averaged over the 16–25-ps period of the simulation.

Certain properties of the average structure will be energetically unreasonable. All bond lengths and bond angles and some interatomic distances are likely to average to unacceptable values. A solution to this problem is to resubmit the averaged structure to minimization of the DG error function for the original constraints using the DG optimization routine, followed by reminimization by AMBER.

For our data, the resulting DG-minimized average structure met the distance constraints better than any of the contributing structures, having exceptionally low error values (Table 1, supplementary information). This structure was utilized in subsequent refinement with molecular dynamics.

Molecular dynamics

The constraint potential

In order to determine whether the differences in the average-DG structures of the four proteins are required by the NMR constraints, we explored whether these differences were stable during MD simulations. The program used was a modified version of the AMBER MD module, which, in addition to the usual potential energy function, includes a potential term that takes into account the NOE constraints.

The constraint potential is a flat-bottomed well of energy zero with parabolic sides. The lower boundary was arbitrarily set at 1.8 Å in the T30A51 and V30A51 data sets, which is below the sum of the van der Waals radii. The upper boundary is the same distance as was used for the distance geometry calculations. Outside of the flat region, the potential increases as $E = k * (r - r_B)^2$, where r is the distance between the protons, r_B is the nearest boundary distance, and k is the force constant. For the proteins in which large conformational changes were expected (T30A51 and V30A51), the potential was modified at large distances. At distances greater than 1.5 Å above the upper boundary the potential increases linearly with a slope equal to that at $r_B + 1.5$ Å. This is so that large distance violations may be corrected without excessive distortion of nearby structure.

The relative importance of terms in the potential energy force field depends on temperature. At low temperature, relatively weak interactions such as the van der Waals attraction play an important role. At higher temperatures, the kinetic energy overwhelms these interactions. The behavior then becomes dominated by strong interactions such as covalent bonds and the applied NMR constraints. Thus, we expect that high temperature structures will reflect mainly covalent structures constrained by the NMR data. Comparison of the differences in the mutant proteins using this technique with those determined by DG will determine which changes are a necessary consequence of the data, and whether there are truly sampling differences between the techniques.

Description of typical MD simulation

The results of a typical MD simulation at 298 K on T30A51, the mutant with the most constraints, will be described in order to illustrate the concept of an average MD structure and its variability. All of the simulations described in this paper were 25 ps in length, and structures were recorded every 1 ps. Unless otherwise indicated, all simulations start with the single DG-minimized structure, which has subsequently been energy minimized using the AMBER module.

The dependence of the constraint violation energy on time is shown in Figure 5. During the first picosecond, the value of the constraint force constant is increased geometrically from 0.01 to the final value of 15 kcal mol^{-1} $Å^{-2}$. This results in an increase in the constraint energy during the first picosecond. The energy then decreases to its final value within 5 ps. The backbone RMS between the MD structure and the wild-type structure remains reasonably constant (1.5 Å) throughout the simulation (data not shown). However, closer inspection shows that there is indeed a structural change during the simulation.

Figure 6 shows the RMS value between each picosecond structure and the previous one during the length of the simulation. One observes that, at early times, the RMS between each structure is large, approximately 1.5 Å. However, as the simulation proceeds, the structures become more and more similar, until the RMS between structures approaches 0.5 Å over the range 16–25 ps. The structures appear to be moving into a region of confor-

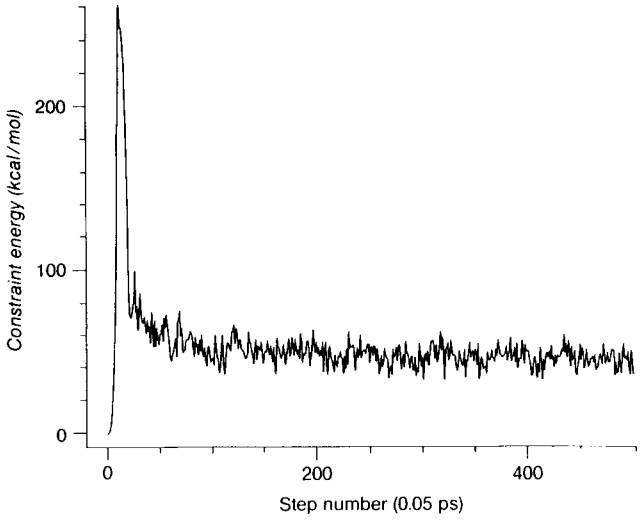

Fig. 5. Dependence of constraint violation energy on simulation time for an MD simulation with constraints, performed at 298 K, for T30A51.

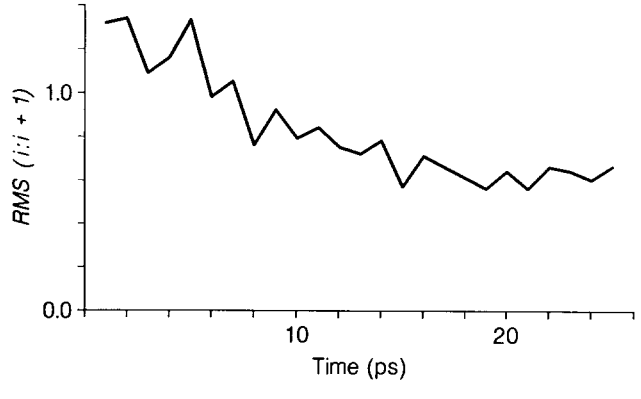

Fig. 6. Backbone RMS change for each picosecond step structure generated during the 25-ps MD simulation of T30A51 in Figure 5.

mational space that is displaced from the X-ray or DG structures, yet is rather restricted. Based on the energies of these structures, they appear to satisfy the NMR constraints well (Table 5). Thus, averaging structures that are located in this region of conformational space, and determining their variability, should be a good representation of the structures sampled by the MD simulation. Further, we can evaluate the structures generated by the different mutants in an identical manner to our evaluation of the DG structures. Using the above methodology, average structures were computed for all four proteins at 298 K and 900 K. Behavior of the constraint energies and RMS vs. time were very similar for the four mutants. Table 4 shows that as the temperature of the simulation is raised, the average structures of the wild-type, A30A51, and T30A51 molecules become more similar. The

Table 5. *Constraint energies of structures*

Minimized structure[a]	Energy[b] of constraint set T30A51	Energy[b] of constraint set V30A51	Potential energy[c]
X-ray	925	467	−677
T30A51 DG-minimum	183	–	−345
V30A51 DG-minimum	–	538	−274
T30A51 298 K MD	30	–	−648
V30A51 298 K MD	–	22	−726
T30A51 900 K MD	47	–	−415
V30A51 900 K MD	–	22	−452

[a] For MD structures, evaluation of energies was performed as follows. RMS calculations were performed between each of the single structures and the average structure. The single structure with the lowest RMS value was subsequently minimized with MIN-NOE from AMBER.
[b] kcal/mol, defined as $\Sigma\ 15*(D_{ij} - D_{\text{bound}})^2$, when $D_{ij} > D_{\text{bound}}$.
[c] kcal/mol, from AMBER.

V30A51 structure still appears to differ significantly from the other three structures.

Discussion

It has been repeatedly shown that NMR measurements, coupled with any of several reconstruction programs, yield reasonably accurate (1–2 Å error) three-dimensional structures for small proteins (Kuntz et al., 1989). The typical residual uncertainties are 5–10 times as large as those derived from crystallography, thus it is worthwhile considering the major sources of these uncertainties.

First, the data themselves are limited in number compared to a typical diffraction experiment, and they are also biased strongly toward short distances. Second, the conversion of NOE intensities to distance constraints requires assumptions about the nature of internal motion and the adequacy of the model used to characterize spin-spin interaction (Havel et al., 1979). Third, the reconstruction methods each contribute to the uncertainty. Distance geometry does not provide structures of low potential energy and has specific sampling problems (e.g., Havel, 1990). Molecular dynamics uses approximate potentials with a poor compensation for solvent and electrostatic interactions. Finally, all the sampling approaches to date are based on incorrect statistical mechanics, as the experimentally derived constraints are applied to single structures rather than to an ensemble of structures (Altman & Jardetzky, 1989).

Given this wealth of sources of both random and systematic error, we have searched for measures of the reliability of any given set of structures produced by DG and/or MD. Mutant proteins are useful in this regard, as their structures should be very similar, especially in regions far removed from the mutation; however, the NOE data sets can be quite different due to the different over-

laps caused by small perturbations in chemical shift. Further, protons whose chemical shifts are unchanged give direct evidence for areas of the protein whose structure is unaffected by the mutation; these regions act as checks for the examination of the reliability of the reconstruction method.

Analysis of DG structures

The RMS differences between the four average DG structures (Table 4) appear to be larger than the variation of each set (Table 3), suggesting that there are structural differences between the mutants. The concept of an average position for each atom, along with its variability, allows us to pinpoint the locations of these changes. The variability, CL, is defined as the radius of a sphere in which there is a 95% probability that the true value is within that sphere, determined from the standard deviation in position and assuming a spherical distribution of atomic position in each of the structures. This method of defining variability is used because our question is whether any apparent structural perturbation is greater than that explainable by experimental uncertainty. Note that this variability is a measure of the preciseness of the location of the average position as opposed to describing a fluctuation in the structures, where the fluctuation describes the volume of space, V, with radius R, occupied by all of the structures (R = average distance between each structure and the average position). For comparisons of the DG and MD structures determined in this study, though, these numbers are very similar. For a spread of 10 structures, $CL \sim 0.7R$.

One simple method of evaluating difference between structures is to measure the difference in position, $D_{A\text{-}A'}$, between corresponding atoms A and A′ in two RMS-aligned structures. In order to evaluate the significance, S, of each difference, it must be compared with the errors in the location of the average positions, $CL_{A\text{-}A'} = (CL_A^2 + CL_{A'}^2)^{1/2}$. Each change in position can be evaluated as to its significance, that is, whether the distance change is greater than the error, by using the t-test. Because we are using the 95% confidence limit for the variability, the t term is accounted for in the calculation. Thus, movements that are 95% significant are ones where $S = D_{A\text{-}A'}/CL_{A\text{-}A'} > 1$.

Figure 7A shows the positional difference for the three mutant DG structures vs. the wild-type DG structure, averaged over the backbone residues C, C^α, and N. One can observe that many differences are greater than 1 Å; furthermore, Figure 7B shows that most of these changes are significant. This method of analysis does not inform us as to the direction of movement, however.

Another method of comparing structures that also gives directional information is the difference distance matrix. This involves computing all the backbone atom-atom distances in protein A, doing the same for the protein B, and determining the differences in the distances in proteins A and B, and then averaging over the entire residue-residue pair to determine a distance change ΔD_{AB}. The experimental error for this measurement, CL_{AB}, is approximately the square root of the sums of the squares of the four variabilities (two residues in proteins A and B). Significant movements are again changes where the ratio is greater than one.

Figure 8 illustrates the distance changes between the wild-type DG structure and the A30A51, T30A51, and V30A51 DG structures in the upper triangles of A, B, and C, respectively. Changes that are significant are indicated in the corresponding lower triangles. One can make several observations from Figures 7 and 8: (1) The

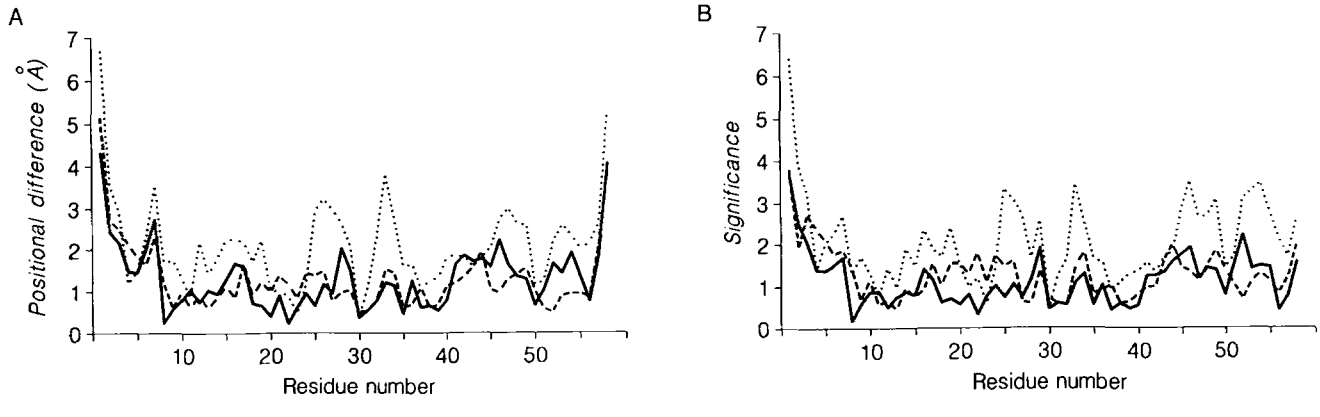

Fig. 7. **A**: Positional difference in Å, averaged over the N, C, and C^α atoms of each residue, between corresponding residues in the average DG structures of wild-type BPTI and the mutants A30A51 (——), T30A51 (– – –), and V30A51 (· · ·). **B**: Significance value, defined in the text as positional difference divided by the error, of positional changes shown in A. Values are determined for each pair of corresponding atoms and then averaged over the residue.

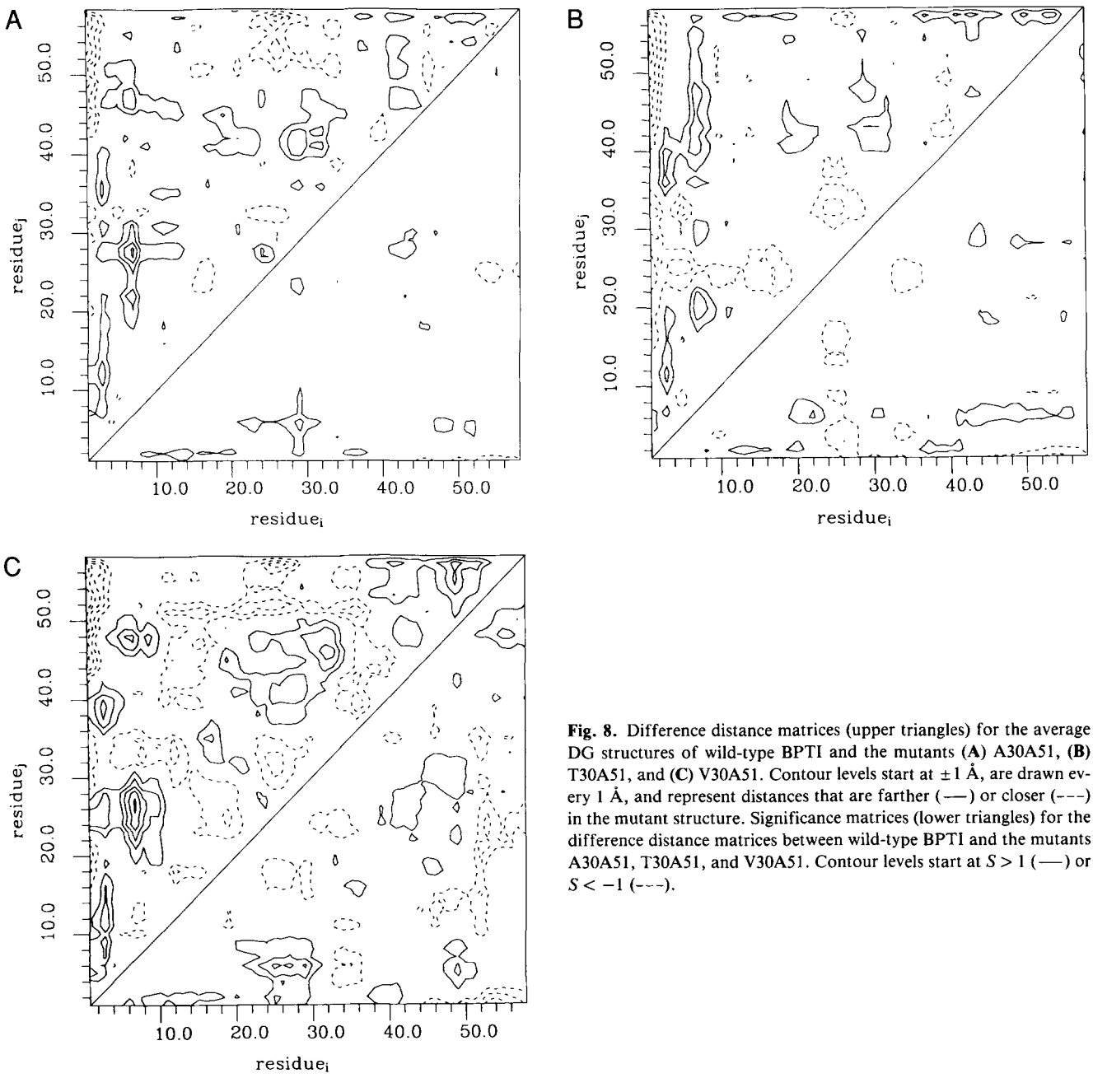

Fig. 8. Difference distance matrices (upper triangles) for the average DG structures of wild-type BPTI and the mutants (**A**) A30A51, (**B**) T30A51, and (**C**) V30A51. Contour levels start at ±1 Å, are drawn every 1 Å, and represent distances that are farther (—) or closer (---) in the mutant structure. Significance matrices (lower triangles) for the difference distance matrices between wild-type BPTI and the mutants A30A51, T30A51, and V30A51. Contour levels start at $S > 1$ (—) or $S < -1$ (---).

differences in the DG structures are quite large in some areas of the structure, larger than would normally be associated with mutational effects (Matthews, 1987), even in regions not including the termini. (2) These differences are not exclusive to the site of mutation, contrary to expectations from chemical shift differences (Fig. 2). (3) The differences from wild type progress in the order A30A51, T30A51, V30A51. (4) By statistical tests, most of these changes are significant.

It seems that real changes of the magnitudes exhibited by the A30A51 and wild-type DG structures are highly unlikely, based on the fact that the X-ray analysis of the A30A51 and wild-type BPTI molecules show that their structures have an all-atom deviation of 0.26 Å (Eigenbrot et al., 1990). The even larger changes in the T30A51 and V30A51 DG structures may also be untenable. The most likely possibility is that these changes are due to insufficient sampling of three-dimensional conformational space by the limited number of structures generated by the distance geometry algorithm, and thus the average position from the DG structure does not represent the average position of all accessible conformational space. Thus, we need to examine structures from other methods of analysis to determine if the structural perturbations

seen in the DG structures are a necessary consequence of the NMR data.

Analysis of MD structures

Searching conformational space

The initial question is whether the MD simulation does a good job of searching the conformational space accessible with the constraints. The variability of the atoms does increase with temperature for the MD simulations. The average CL_A for the T30A51 900 K MD structures (0.72) is larger than that of the DG structures (0.60) or the 298 K MD structures (0.40). In addition, after high temperature refinement the centers of these variabilities, the average positions, become more similar (Table 4). Because there are no chemical shift changes due to mutation in most of the protein (Fig. 2), we expect little structural change in these regions; thus, the MD result is more physically realistic. The constraint energy violations of the 900 K MD, 298 K MD, and DG structures are similar (Table 5), so all methods are searching conformational space accessible to the constraints. Thus, it seems a reasonable conclusion that MD simulation at 900 K samples the accessible conformational space better than DG (Scheek et al., 1989).

We also know that the 900 K simulation with constraints is reflective of primarily the NMR data, as a simulation without constraints on the wild-type protein results in a disordered structure with an RMS to the X-ray structure of >10 Å (data not shown). Further, the structures resulting from 900 K simulations starting from the wild-type X-ray structure are virtually identical to those starting from the separate DG structures (data not shown), suggesting that although the MD simulation may be pathway-dependent, the average structures from those pathways are not. This demonstrates that most of the differences in the DG structures are not a necessary consequence of the NMR data.

Structural consequences of mutation

From the variability of the backbone atoms during MD simulation at 900 K, one can state that given typical data sets of 10–15 NOEs per residue and no coupling constant information, the atomic positions can be defined to within 1 Å for the backbone atoms. Thus, the structural movements that occur upon mutation must be greater than 1 Å in order to be detected reliably. Further, all reconstruction methods should show that these movements are significant. Thus, Figures 9–11 show positional difference plots (A) and significance plots (B) for the three mutants A30A51, T30A51, and V30A51, respectively, vs. wild type for each reconstruction method. We examine these figures for consistent differences.

The A30A51 mutation appears to have very few effects on the structure greater than 1 Å, as determined by NMR (Fig. 9A). Further, the differences that do occur do not appear to be significant as most S values are <1 (Fig. 9B). This conclusion is supported by the crystallography study performed by Eigenbrot et al. (1990), in which they determined that the backbone atoms of the wild-type and A30A51 mutant were virtually identical, with an RMS of less than 0.2 Å.

Crystallography studies of the T30A51 mutant have been hampered by the inability of the protein to crystallize in useful orientations (C. Eigenbrot et al., pers. comm.). Because the amino acid mutation at position 30 is buried, this suggests that the effect of the mutations is propagated to the protein surface, thus affecting the crystallization properties. The NMR structure of T30A51 appears to bear this out. Figure 10A shows that the positions of the N-terminal α-helix (residues 3–7) and the β-

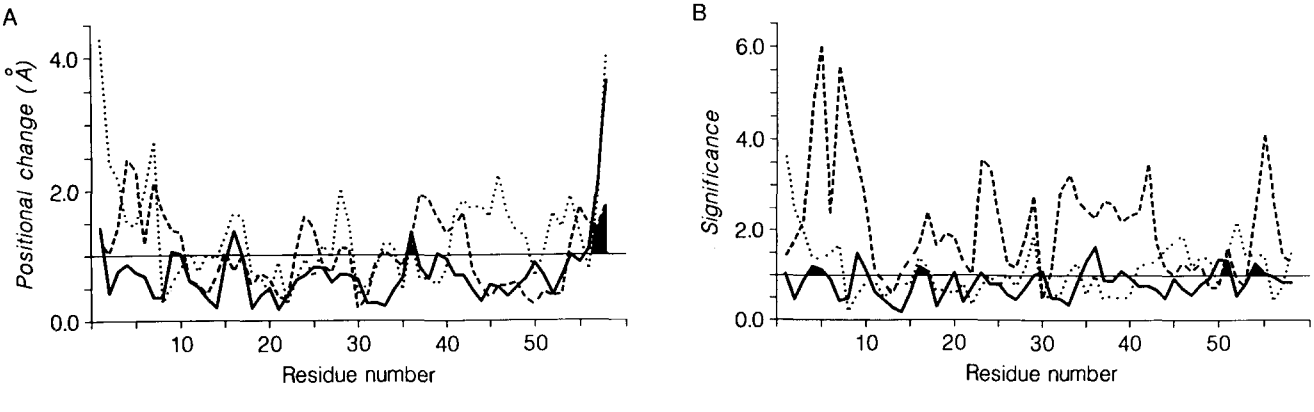

Fig. 9. **A**: Positional difference in A, averaged over the N, C, and C^{α} atoms of each residue, between corresponding residues in the average structures of A30A51 and wild-type BPTI using DG (· · ·), 298 K MD simulation (– – –), and 900 K MD simulation (—). **B**: Significance value, defined in the text as positional difference divided by the error, of positional changes shown in A. Values are determined for each pair of corresponding atoms and then averaged over the residue. Shaded regions represent regions where all changes are greater than (A) 1 Å or (B) $S > 1$.

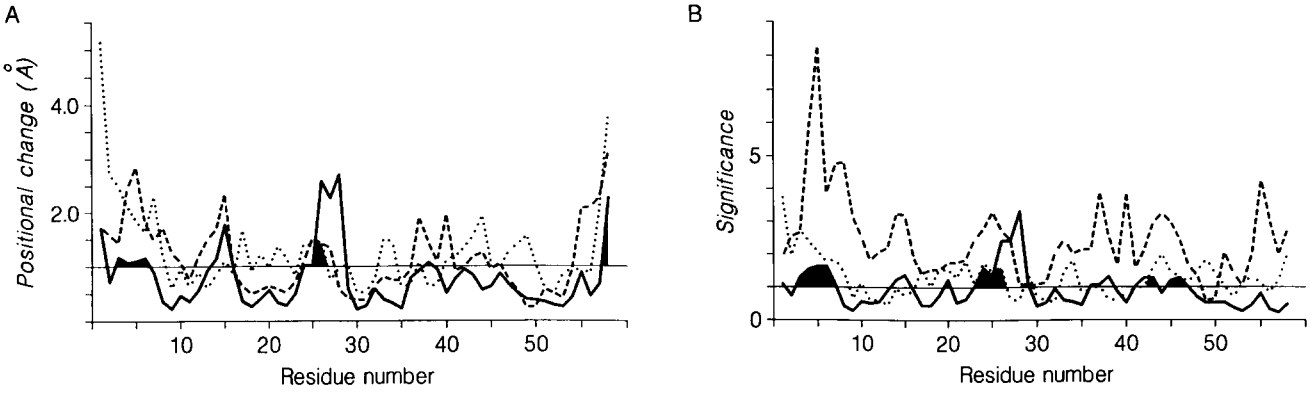

Fig. 10. **A**: Positional difference in A, averaged over the N, C, and C^{α} atoms of each residue, between corresponding residues in the average structures of T30A51 and wild-type BPTI using DG (· · ·), 298 K MD simulation (– – –), and 900 K MD simulation (——). **B**: Significance value, defined in the text as positional difference divided by the error, of positional changes shown in A. Values are determined for each pair of corresponding atoms and then averaged over the residue. Shaded regions represent regions where all changes are greater than (A) 1 Å or (B) $S > 1$.

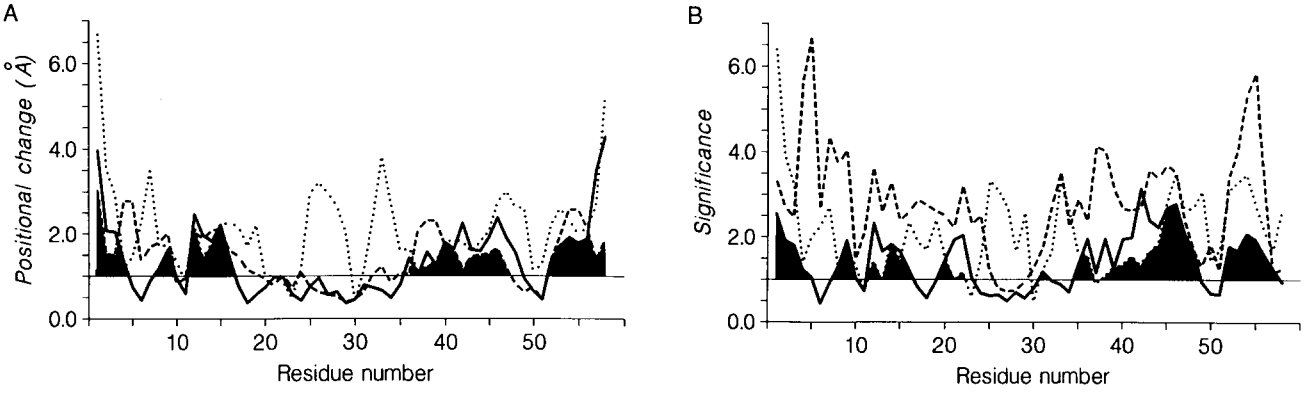

Fig. 11. **A**: Positional difference in A, averaged over the N, C, and C^{α} atoms of each residue, between corresponding residues in the average structures of V30A51 and wild-type BPTI using DG (· · ·), 298 K MD simulation (– – –), and 900 K MD simulation (——). **B**: Significance value, defined in the text as positional difference divided by the error, of positional changes shown in A. Values are determined for each pair of corresponding atoms and then averaged over the residue. Shaded regions represent regions where all changes are greater than (A) 1 Å or (B) $S > 1$.

loop (25–28) between the two strands appear to be affected upon mutation. Both of these movements are also significant in all of the structures (Fig. 10B) suggesting that this change is not an artifact of the method of conformational sampling. Difference distance matrix plots demonstrate that these regions move closer in the mutant (data not shown). Further, this loop is in a crystal contact region of the wild-type molecule and thus may be responsible for the differing crystallization properties.

The response of the protein to the V30A51 mutation appears to be somewhat more complex (Fig. 11A,B). Unfortunately, interpretation of any of the contour maps appears to be made difficult by the effect of 25-ps simulations on the end-to-end distances in proteins with fewer constraints; thus, comparison of V30A51 and wild-type is not straightforward. A region of the molecule that appears to have large consistent differences between V30A51 and wild type in the DG and MD simulations is the C-terminal α-helix. The N-terminal portion of the helix (45–47) appears to move away from the β-sheet, while the C-terminal portion (52–53) appears to move toward this sheet (difference distance matrix not shown), suggesting that the helix undergoes a slight rotation centered near the site of mutation. At present, this interpretation is tentative at best.

Conclusions

For an NOE data set typically available to the NMR spectroscopist, we have attempted to determine the extent to which the intensity information can be related to structural perturbations. We have examined several routes of

analysis. By using real data from mutant proteins for which we have structural information from other sources, we had reasonable expectations of structural similarity. This allowed us to evaluate the results from the various methods. We have determined that molecular dynamics refinement at 900 K with constraint information derived from NOE intensities is a necessary complement to a DG algorithm.

At present, the ability to use typical data sets determined by NMR to distinguish the structural effects of mutation appears to be limited to changes greater than 1 Å. We have shown that changes in the A30A51 structure are less than this limit, a result consistent with that from crystallography (Eigenbrot et al., 1990). Changes in the T30A51 and V30A51 appear to occur upon mutation, but are near this detection limit.

Further reduction in this uncertainty can be expected with the use of relaxation matrix methods to establish tighter bounds on intensities (Borgias & James, 1989), coupling constant information to extract further stereospecific assignments (Wagner et al., 1987b), and higher field instrumentation to limit the number of overlapping peaks and increase the number of detectable peaks.

Materials and methods

Materials

The BPTI double mutants Cys 30 → Ala/Cys 51 → Ala (A30A51), Cys 30 → Val/Cys 51 → Ala (V30A51), and Cys 30 → Thr/Cys 51 → Ala (T30A51) were expressed and purified in a manner identical to that described earlier (Hurle et al., 1990). Wild-type BPTI (Trasylol®) was obtained from Mobay Chemical Corporation.

Spectroscopy

All two-dimensional NMR spectra were acquired at 36 °C, pH 4.6, at approximately 8 mM protein concentration, unbuffered. Spectra in D_2O were acquired under identical conditions with a pH meter reading of 4.6, unadjusted for the deuterium isotope effect. Spectra were acquired on a General Electric GN-500 spectrometer. Phase sensitive mode with quadrature detection in both dimensions was used. Nuclear Overhauser effect spectra were acquired with the procedure of States et al. (1982), with a mixing time of 160 ms in all cases. Double-quantum filtered COSY (Piantini et al., 1982; Rance et al., 1983; Shaka and Freeman, 1983) were acquired with time-proportional phase incrementation of the first pulse (Redfield and Kunz, 1975; Marion and Wüthrich, 1983). HOHAHA spectra with MLEV-17 (Bax and Davis, 1985) were acquired after modification of the spectrometer by inclusion of a 6-W amplifier as a transmitter. A mixing time near 70 ms was used in all cases. The large H_2O resonance was eliminated with decoupler saturation during the relaxation delay. The spectral width was 7,042 Hz for all spectra. Double-quantum filtered COSY spectra were acquired with 1,024 t_1 increments. Other spectra were acquired with 512 t_1 increments, and all spectra were zero filled to at least 1,024 × 1,024 during processing. A variety of apodization functions were used. Data processing was carried out with programs initially developed at Dr. R. Kaptein's laboratory at the University of Groningen, Groningen, The Netherlands, and then adapted for our environment by Drs. R.M. Scheek, S. Manogaran, and M. Day in our laboratory. Baseline correction was carried out in both dimensions for all spectra.

Computational methods and assignments

Chemical shift assignments

The NMR data used for these studies were easily acquired. Determination of chemical shifts for protons in these molecules was straightforward, because in most cases shifts were close to the published values for the wild type. This eliminated the need to use the time-consuming sequential assignment procedure. Given prior knowledge of the crystal structure, the process of assigning NOESY cross-peaks to acquire distance information was rapid using an automated procedure, NASIGN (see the Results section). The amount of data acquired for each mutant (~14 cross-peaks/residue) is typical of some of the more highly constrained NMR structures published to date. These assignments could be obtained within a few minutes of computer time. Thus, we base our analysis on these easily measured parameters: chemical shift perturbations and NOE intensities.

Distance constraints and distance geometry

The NOESY cross-peaks were integrated by presenting a contour plot of the two-dimensional spectrum on a graphics terminal. Cursor-oriented commands were used to draw a box around each peak recognized by eye. A computer program integrated the points within the box to determine the intensity and calculated the first moment to determine the chemical shift coordinates. Peaks suffering from overlap sufficient to impair the integration were treated as weak and therefore were given the longest upper bound (4.5 Å). Approximately 40% of all peaks were of this nature.

Isolated NOESY peaks were classified into three groups according to intensity. Strong, medium, and weak peaks correspond to distance upper bounds of 2.5, 3.5, and 4.5 Å, respectively. The criteria used for the classification was based on the isolated spin pair approximation as follows. The intensities of cross-peaks between geminal protons of known distance were collected, and an average conversion factor relating the distance to the inverse sixth root of the intensity was determined. Intensities corresponding to the upper-bound distance classifications were then calculated, and the NOESY peaks were assigned accordingly, with a 0.5-Å leeway added to

account for measurement uncertainty, conformational averaging, and spin diffusion effects. (This leeway is justified in the Results section.)

For unresolvable protons, the upper bound was increased by 1.8 Å for protons bound to the same carbon or by 4.2 Å for symmetry-related protons across an aromatic ring, corresponding to the interproton distance. For such peaks, the relaxed constraint was applied to all of the protons involved. For resolvable protons bound to the same carbon, the protons were arbitrarily associated with the corresponding peaks. The chirality of such carbon atoms was not constrained during DG calculations. Constraints corresponding to hydrogen bonding were also applied when elements of regular secondary structure could be identified based on patterns of connectivity in NOESY spectra. The bounds for such bonds are $r \leq 3.3$ Å for the N-O distance and $2.0 \leq r \leq 2.3$ Å for the H-O distance. A computer program, PREDG, was developed to handle the extensive bookkeeping associated with the intensity-to-distance constraint conversion. Distance geometry calculations were carried out using the VEMBED program (Kuntz et al., 1989), a vectorized version of EMBED (Havel et al., 1983), and run at the San Diego Supercomputer Center. The normal procedures of bound smoothing, structure embedding, and optimization were employed (Crippen, 1981; Crippen and Havel, 1988; Kuntz et al., 1989). On the average, eight structures were obtained in 34 min of Cray XMP-48 time.

Molecular dynamics

Molecular dynamics was carried out using modifications of version 3.0 of the AMBER modeling package (AMBER [UCSF] 3.0, 1986). MIN-NOE and MD-NOE refinements were performed with AMBER-NMR (1989). The NMR distance bounds were identical to those used in the DG calculations. The constraints were implemented as a flat potential within the range of allowed distances. Outside this range, the constraint violation was calculated as a Hook's law parabolic potential, with the option of a linear extension at specified energy for upper-bound violations. Simulations were carried out on Sun SparcStation 1 computers and in all cases required approximately 2 h/ps of simulation. The calculations used the SHAKE algorithm, a distance-dependent dielectric of $4r_{ij}$, a 1-fs step size, and a 15-kcal $Å^{-2}$ Hook's law force constant. For simulations of mutants that were not known to have structures similar to wild type (V30A51 and T30A51), the constraint potential was geometrically ramped from 0.01 to the final value during the first picosecond of the simulation, so that the structure could more easily adapt to the new constraint potential.

Acknowledgments

The authors thank Mike Brochier for performing the fermentations, and the San Diego Supercomputer Center for computing resources for the DG algorithm.

This work was supported by Genentech, Inc., NSF grant DMB8606901, and NIH grant GM19267 (to I.D.K.), NIH GM29072 (to P.K.), NIH grant RR01695 (to I.D.K.), DARPA grant N00014-86-K-0757 (to Robert Langridge), NIH Postdoctoral Fellowship GM13119-01 (to M.R.H.), and a grant from Biotechnology Resources Education Program (Molecular Biology Institute, U.C.L.A.).

References

Altman, R.B. & Jardetzky, O. (1989). Heuristic refinement method for determination of solution structure of proteins from nuclear magnetic resonance data. In *Methods in Enzymology* (Oppenheimer, N.J. & James, T.L., Eds.), pp. 218-245. Academic Press, San Diego.

AMBER 3.0 (1986). Written by Singh, U.C., Weiner, P.K., Caldwell, J.W., & Kollman, P.A. Department of Pharmaceutical Chemistry, University of California, San Francisco.

Bax, A. & Davis, D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson. 65*, 355-360.

Berman-Marks, C., Naderi, H., Kosen, P.A., Kuntz, I.D., & Anderson, S. (1987). Refolding of bovine pancreatic trypsin inhibitor mutants lacking cysteines 30 and 51. In *Protein Structure, Folding and Design 2* (Oxender, D.L., Ed.), pp. 335-340. Alan R. Liss Inc., New York.

Billeter, M., Braun, W., & Wüthrich, K. (1982). Sequential resonance assignments in protein proton nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances in single crystal protein conformations. *J. Mol. Biol. 155*, 321-346.

Borgias, B.A. & James, T.L. (1989). Two-dimensional nuclear Overhauser effect: Complete relaxation matrix analysis. In *Methods in Enzymology* (Oppenheimer, N.J. & James, T.L., Eds.), pp. 169-183. Academic Press, San Diego.

Creighton, T.E. (1974). Intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J. Mol. Biol. 87*, 579-602.

Creighton, T.E. (1977). Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J. Mol. Biol. 113*, 275-294.

Creighton, T.E. & Goldenberg, D.P. (1984). Kinetic role of a metastable native-like two-disulfide species in the folding transition of bovine pancreatic trypsin inhibitor. *J. Mol. Biol. 179*, 497-526.

Crippen, G.M. (1981). *Distance Geometry and Conformational Calculations*. Wiley, Chichester, England.

Crippen, G.M. & Havel, T.F. (1988). *Distance Geometry and Molecular Conformation*. Research Studies Press Ltd., Tauton, England, and Wiley, Chichester, England.

Eigenbrot, C., Randal, M., & Kossiakoff, A.A. (1990). Structural effects induced by removal of a disulfide bridge; the X-ray structure of the C30A/C51A mutant of BPTI at 1.6 Ångstrom. *Protein Eng. 3*, 591-598.

Goldenberg, D.P. (1988). Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. *Biochemistry 27*, 2481-2489.

Havel, T.F. (1990). The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains. *Biopolymers 29*, 1565-1585.

Havel, T.F., Crippen, G.M., & Kuntz, I.D. (1979). Effects of distance constraints on macromolecular conformation. II. Simulation of experimental results and theoretical predictions. *Biopolymers 18*, 73-81.

Havel, T.F., Kuntz, I.D., & Crippen, G.M. (1983). The theory and practice of distance geometry. *Bull. Math. Biol. 45*, 665-720.

Hurle, M.R., Marks, C.B., Kosen, P.A., Anderson, S., & Kuntz, I.D. (1990). Denaturant-dependent folding of bovine pancreatic trypsin inhibitor mutants with two intact disulfide bonds. *Biochemistry 29*, 4410-4419.

Kaptein, R., Boelens, R., Scheek, R.M., & Van Gunsteren, W.F. (1988). Protein structures from NMR. *Biochemistry 27*, 5389-5395.

Kassell, B. (1970). Naturally occurring inhibitors of proteolytic enzymes. *Methods Enzymol. 11*, 839-852.

Kuntz, I.D., Thomason, J.F., & Oshiro, C.M. (1989). Distance geometry. In *Methods in Enzymology* (Oppenheimer, N.J. & James, T.L., Eds.), pp. 159-203. Academic Press, San Diego.

Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of proton-proton spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun. 113*, 967-974.

Matthews, B.W. (1987). Genetic and structural analysis of the protein stability problem. *Biochemistry 26*, 6885-6888.

Oas, T.G. & Kim, P.S. (1988). A peptide model of a protein folding intermediate. *Nature 336*, 42-48.

Piantini, O.W., Sorenson, O.W., & Ernst, R.R. (1982). Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc. 104*, 6800-6801.

Rance, M., Sorenson, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., & Wüthrich, K. (1983). Improved spectral resolution in COSY proton NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun. 117*, 479-485.

Redfield, A.G. & Kunz, S.D. (1975). Quadrature Fourier NMR detection. *J. Magn. Reson. 19*, 250-254.

Richardson, J. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem. 34*, 167-339.

Scheek, R.M., van Gunsteren, W.F., & Kaptein, R. (1989). Molecular dynamics simulation techniques for determination of molecular structures from NMR data. In *Methods in Enzymology* (Oppenheimer, N.J. & James, T.L., Eds.), pp. 204-217. Academic Press, San Diego.

Shaka, A.J. & Freeman, R. (1983). Simplification of NMR spectra by filtration through multiple quantum coherence. *J. Magn. Reson. 51*, 169-173.

States, D.J., Haberkorn, R.A., & Ruben, D.J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson. 48*, 286-292.

Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N., & Wüthrich, K. (1987a). Protein structures in solution by NMR and distance geometry: The polypeptide fold of BPTI using DISGEO and DISMAN. *J. Mol. Biol. 196*, 611-639.

Wagner, G., Bruhwiler, D., & Wüthrich, K. (1987b). Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin inhibitor by heteronuclear two-dimensional nuclear magnetic resonance. *J. Mol. Biol. 196*, 227-231.

Wlodawer, A., Nachman, J., Gilliland, G.L., Gallagher, W., & Woodward, C. (1987). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol. 198*, 469-480.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. Wiley, New York.

Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982). Sequential resonance assignments as a basis for determination of spatial protein structure by high resolution proton nuclear magnetic resonance. *J. Mol. Biol. 155*, 311-319.