Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 1993

Impact Factor: 2.63 · DOI: 10.1002/prot.340160404 · Source: PubMed

10 AUTHORS, INCLUDING:

Nobuhiro Go
Japan Atomic Energy Agency
238 PUBLICATIONS **9,743** CITATIONS
SEE PROFILE

Fuyuhiko Inagaki
Hokkaido University
326 PUBLICATIONS **8,671** CITATIONS
SEE PROFILE

Mitiko Go
Cornell University
110 PUBLICATIONS **2,785** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Genetics 16:341–356 (1993)

Secondary Structural Features of Modules M2 and M3 of Barnase in Solution by NMR Experiment and Distance Geometry Calculation

Teikichi Ikura,[1] Nobuhiro Gō,[1] Daisuke Kohda,[2] Fuyuhiko Inagaki,[2] Hiroshi Yanagawa,[3] Masuyo Kawabata,[4] Shun-ichiro Kawabata,[4] Sadaaki Iwanaga,[4] Tosiyuki Noguti,[5] and Mitiko Gō[5]

[1]Department of Chemistry, Faculty of Science, Kyoto University, Kyoto 606, Japan; [2]Tokyo Metropolitan Institute of Medical Science, Tokyo 113, Japan; [3]Mitubishi Kasei Institute of Life Sciences, Machida, Tokyo 194, Japan; [4]Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan; and [5]Department of Biology, Faculty of Science, Nagoya University Nagoya 464-01, Japan

ABSTRACT **Proteins consist of structural units such as globular domains, secondary structures, and modules. Modules were originally defined by partitioning a globular domain into compact regions, each of which is a contiguous polypeptide segment having a compact conformation. Since modules show close correlations with the intron positions of genes, they are regarded as primordial polypeptide pieces encoded by exons and shuffled, leading to yield new combination of them in early biological evolution. Do modules maintain their native conformations in solution when they are excised at their boundaries? In order to find answers to this question, we have synthesized modules of barnase, one of the bacterial RNases, and studied the solution structures of modules M2 (amino acid residues 24–52) and M3 (52–73) by 2D NMR studies. Some local secondary structures, α-helix, and β-turns in M2 and β-turns in M3, were observed in the modules at the similar positions to those in the intact barnase but the overall state seems to be in a mixture of random and native conformations. The present result shows that the excised modules have propensity to form similar secondary structures to those of the intact barnase.**

Key words: ribonuclease, synthesized peptide, protein folding, protein conformation, molecular evolution, exon shuffling

INTRODUCTION

It is well known that eukaryotic genes are generally split by introns.[1,2] In cases of genes encoding proteins it is interesting to see where in three-dimensional structures of proteins the intron positions are located. Some are located in segments connecting globular domains,[3,4] but more frequently within globular domains.[5,6] During efforts to find structural characteristics of segments within a globular domain encoded by exons, a concept of modules was introduced from a purely structural point of view.[7] Modules were defined as compact structures which contiguous segments take within globular domains or proteins.[8] Sizes of modules thus defined have been found to be distributed in the range from about 10 to 40 residues.[7] Junctions of such modules have been found to have good correlations with positions of introns in hemoglobin,[8] lysozyme,[9,10] triose phosphate isomerase,[7,11] and ovomucoid third domain.[12] These correlations between exons and modules yielded a view that all modules were encoded by their corresponding exons in the ancestral genes for ancestral proteins, and new proteins may have been produced during the process of the biological evolution by assortment of various modules through the mechanism of exon shuffling.[7,13] Even though introns have been found in bacterial genes encoding tRNA,[14,15] prokaryotic genes are usually not split by introns. This fact is interpreted as suggesting that most introns have been lost in the contemporary prokaryotic genes and those still remaining were originated even before the emergence of eukaryotes.[16]

In the view prompted by the correlations between exons and modules, introns act as mediators of exon shuffling through crossing over on themselves[13] and exon shuffling in the gene corresponds to module

Abbreviations used: TLC, thin layer chromatography; HPLC, high performance liquid chromatography; TFE, trifluoroethanol; NMR, nuclear magnetic resonance; DQF-COSY, double-quantum-filtered correlated spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; RMSD, root mean square displacement; CD, circular dichroism; BPTI, bovine pancreatic trypsin inhibitor; RNase, ribonuclease.

Received December 18, 1992; revision accepted March 11, 1993.

Address reprint requests to Dr. Mitiko Gō, Department of Biology, Faculty of Science, Nagoya University Nagoya 464-01, Japan.

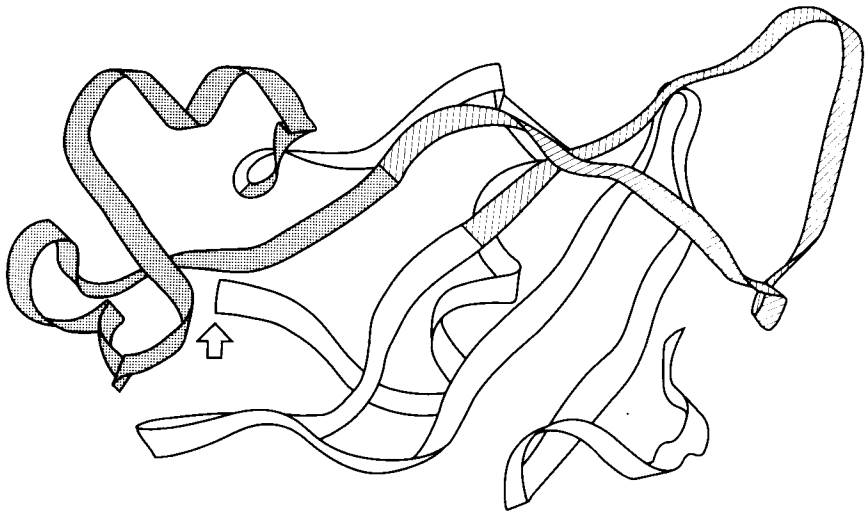


Fig. 1. Tertiary structure of barnase. Trace of main chain is represented by a ribbon model. The N-terminus is indicated by an arrow. Modules M2 and M3 are stippled and hatched, respectively. The presentation is based on the atomic coordinates determined by the X-ray crystallography.[21]

shuffling in proteins.[7] For modules to have the role as parts assembled into various proteins, compactness of module three-dimensional structures and their relative stabilities are expected to be essential factors. Therefore, we should ask whether or not a module identified in an existing globular protein retains its native three-dimensional structure in the protein, even when excised from the protein and isolated by itself.[17] For this purpose we prepared the second and third modules of barnase, a bacterial RNase, by a synthetic method, and carried out 2D NMR measurements of each module in solution. Since introns have not been found in protein-encoding prokaryotic genes, except phage thymidylate synthase gene,[18] we have identified the module boundaries from the three-dimensional structure of this protein.

MATERIALS AND METHODS

Barnase is an extracellular endoribonuclease with weak specificity produced by *Bacillus amyloliquefaciens*.[19] It is a monomer consisting of 110 amino acid residues[20] and is one of the smallest globular proteins containing no disulfide bonds. Its tertiary structure has been determined by X-ray crystallography.[21,22] Barnase has been decomposed into six modules, M1 to M6, by using the centripetal profile. Module boundaries have been identified as 24, 52, 73, 88, and 98.[23] The sizes of modules of barnase vary from 10 to 28 amino acid residues. Conformation of barnase and that of the portion corresponding to the second and third modules, M2 and M3, are shown in Figure 1. Polypeptide fragments corresponding to the two modules were synthesized on an Applied Biosystems (Foster City, CA) 430 A peptide synthesizer using *t*-butoxycarbonyl chemistry with chemicals and the program supplied by the manufacturer. One residue overlapped at each boundary of modules; i.e., M2:24–52 and M3:52–73 were synthesized. After completion of the synthesis, the resulting peptides were cleaved by trifluoromethansulfonic acid as described in a manual supplied by Applied Biosystems. The crude peptides were purified by reversed phase HPLC using a Cosmosil $_5C_{18}$(10 × 250 mm, Nacalai tesque, Kyoto). The sequence and purity of the resulting polypeptide were confirmed by TLC, analytical HPLC, amino acid analysis, and their sequences by a gas-phase protein sequencer (Applied Biosystems model 470 A).

Modules M2 and M3 were dissolved at 9 and 10 mM concentrations in 50% d_3-TFE and 50% H_2O, respectively. The pHs were adjusted both to an uncorrected glass electrode reading of pH* = 3.0 using a Radiometer PHM86 and Ingold thin-glass electrode. All the NMR spectra were measured on a JEOL JNM GX500 ^1H 500-MHz NMR spectrometer at 28°C. 2,2-Dimethyl-2-silapentane-5-sulfonate was used as an internal standard of chemical shifts. DQF-COSY,[24] HOHAHA,[25] and NOESY[26,27] spectra were recorded in the phase-sensitive mode.[28] The water resonance was suppressed by selective irradiation during the relaxation delay. For the measurements of NOESY spectra, the water resonance was also suppressed by means of the DANTE pulse.[29] All the two-dimensional spectra were recorded with 512 × 2048 data points and with a spectral width of 6,000 Hz. HOHAHA spectra were recorded with mixing time of 35 msec. NOESY spectra were recorded with mixing times of 150 and 300 msec. Typically, 64 scans were accumulated for each t_1 value with relaxation delay of 1.0 sec. The digital

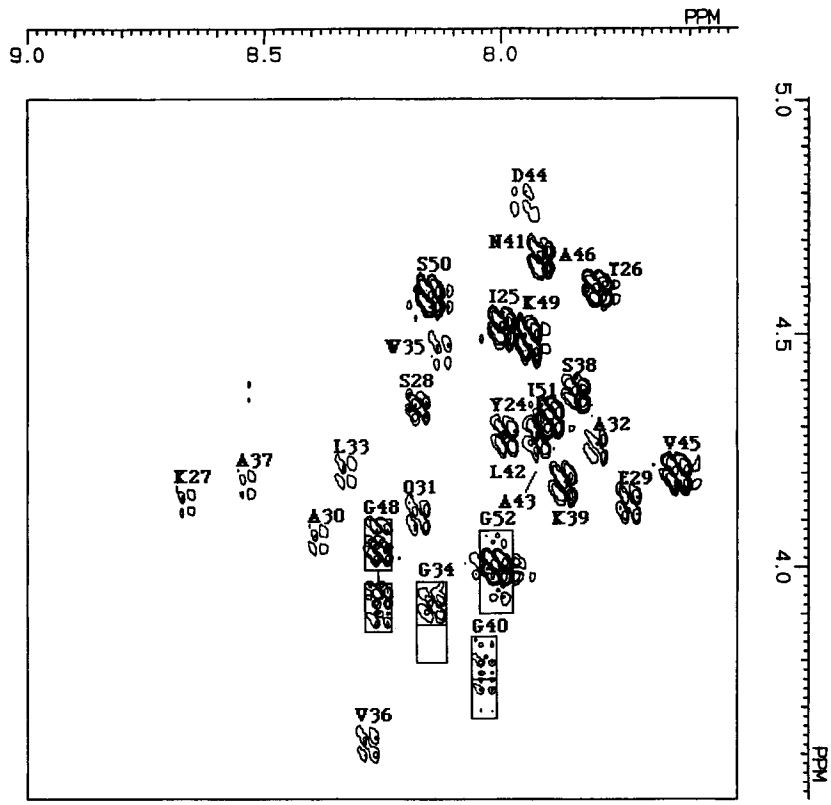


Fig. 2. Fingerprint region of DQF-COSY spectrum of module M2 with assignments of the cross-peaks. For a Gly residue, two cross-peaks C^{α} H, NH and C^{α} 'H, NH are enclosed by boxes and connected by a vertical line.

resolution was usually 5–6 Hz/point in both dimensions by zero-filling in the t_1 dimension. A phase-shifted sine bell function for DQF-COSY and HOHAHA or a Lorentzian-Gaussian function for NOESY was applied for both t_1 and t_2 dimensions. After Fourier transformations, t_1 ridges were reduced by the subtraction method.[30] Distance geometry calculations were made by using the computer program DADAS[31] with a TITAN computer at Kyoto University and an Apollo DN10000 computer at the Tokyo Metropolitan Institute of Medical Science.

RESULTS

NMR Experiment

Sequence-specific resonance assignments of M2 and M3 were made according to the method established by Wüthrich and his co-workers.[32] First, the NMR resonances were assigned to the spin systems of specific amino acid types using DQF-COSY and HOHAHA. Second, the spin systems were aligned along the primary structure based on the distance information from the NOESY spectrum.

Barnase M2

Figure 2 shows the fingerprint region of a DQF-COSY spectrum of M2, where correlations of the chemical shifts of all the amide and α protons were observed. The cross-peaks between an α proton and side chain protons were developed as shown in Figure 3, where the spin system of each amino acid type is connected by solid lines. Each amino acid residue gives a characteristic pattern which in turn gives an assignment of an amino acid type. Thus, all observable protons were assigned to characteristic spin systems and to specific types of amino acid residues.

Next, the spin systems were aligned according to the amino acid sequence of M2 using distance connectivity by NOESY. Figure 4 shows the NOESY spectrum of the amide proton region of M2, where many NOE connectivities between amide protons of adjacent residues were observed. However, connectivities were interrupted at several residues. These interruptions were connected by using $d_{\alpha N}$ and/or $d_{\beta N}$ connectivities. The connectivities thus obtained were also confirmed on the basis of the spin system identifications of the side chains already established through the analyses of the HOHAHA and DQF-COSY spectra. Thus, combined with all the 2D NMR data, the complete assignments of M2 in solution were accomplished, and are shown in Table I. Figure 5 summarizes the distance connectivities for M2.

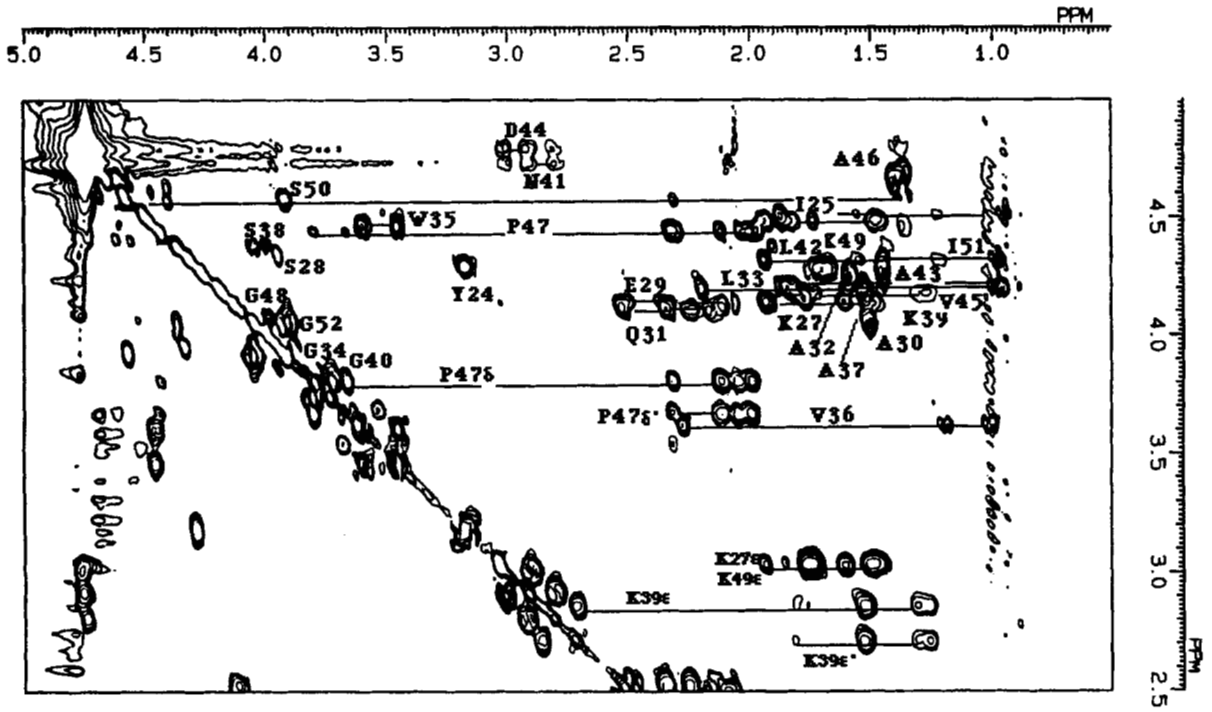


Fig. 3. HOHAHA spectrum of module M2 in the cross-peak region of the α proton and aliphatic proton resonances with a mixing time of 35 msec. The spin system of each amino acid residue is connected by a solid line at the chemical shifts of the α proton resonance. Spin systems developed from $C^{\delta}H_2$ of Pro47 and $C^{\epsilon}H_2$ of Lys27, 39, 49, are also shown as solid lines. There are several cross-peaks from impurities.

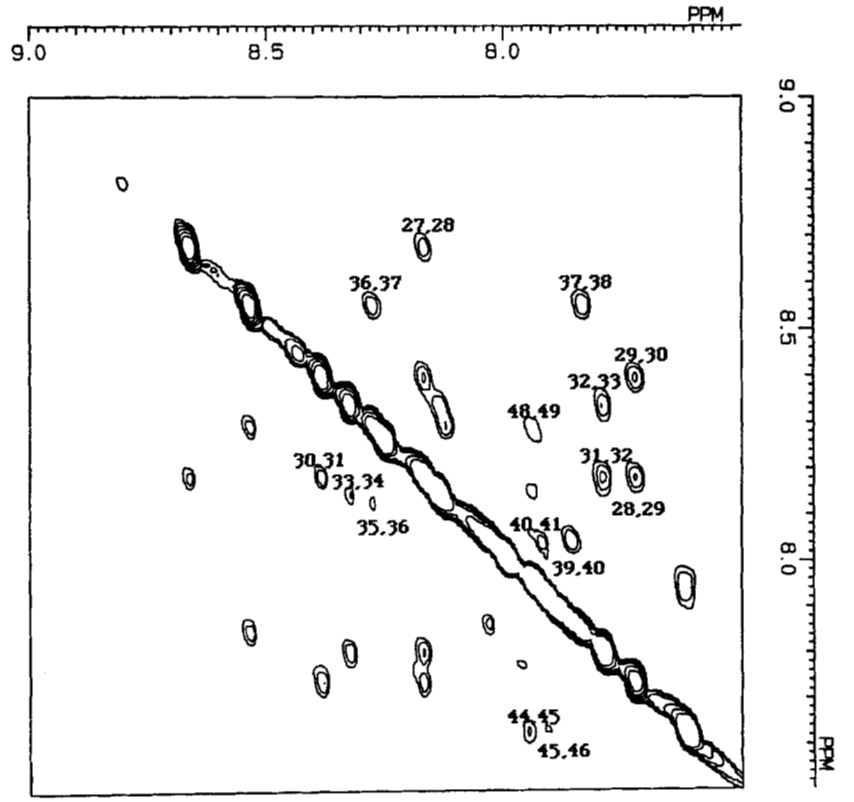


Fig. 4. Amide proton region of the NOESY spectrum of module M2 with a mixing time of 150 msec. NOESY cross-peaks due to the d_{NN} connectivities are labeled with the residue numbers of corresponding amino acids.

TABLE I. Chemical Shifts of Proton Resonances of Barnase M2 at 28°C and pH 3.0*

Residue	Chemical shift (ppm)			
	NH	C^{α} H	C^{β} H	Others
Tyr24	7.99	4.28	3.17,3.17	C2,6H 7.16 C3,5H 6.88
Ile25	7.99	4.51	1.86	C^{γ} H 1.56,1.22 $C^{\gamma}{}_{M}$H (0.94) C^{δ} H (0.94)
Thr26	7.80	4.59	4.49	C^{γ} H 1.35
Lys27	8.66	4.12	1.92,1.84	C^{γ} H 1.59,1.49 C^{δ} H 1.74,1.74 C^{ϵ} H 3.04,3.04 N^{ζ} H 7.59
Ser28	8.17	4.34	3.94,3.94	
Glu29	7.73	4.12	2.35,2.08	C^{γ} H 2.50,2.50
Ala30	8.38	4.04	1.50	
Gln31	8.17	4.11	2.23,2.14	C^{γ} H 2.52,2.34 N^{ϵ} H 7.13,6.63
Ala32	7.79	4.25	1.57	
Leu33	8.33	4.19	1.84	C^{γ} H 1.57 C^{δ} H 0.90,0.90
Gly34	8.14	3.94,3.86		
Trp35	8.12	4.46	3.60,3.46	N1H 9.80 C2H (7.19) C4H 7.61 C5H 7.10 C6H (7.19) C7H 7.45
Val36	8.27	3.60	2.27	C^{γ} H 1.18,0.99
Ala37	8.53	4.16	1.52	
Ser38	7.83	4.36	4.05,4.00	
Lys39	7.86	4.16	1.81,1.75	C^{γ} H 1.30,1.25 C^{δ} H 1.52,1.52 C^{ϵ} H 2.86,2.71 N^{ζ} H 7.34
Gly40	8.03	3.80,3.73		
Asn41	7.93	4.72	2.91,2.81	N^{δ} H 7.38,6.64
Leu42	7.99	4.27	1.74,1.68	C^{γ} H 1.66 C^{δ} H 0.97,0.91
Ala43	7.91	4.26	1.44	
Asp44	7.95	4.78	2.99,2.91	
Val45	7.62	4.18	2.19	C^{γ} H 0.97,0.97
Ala46	7.91	4.64	1.39	
Pro47		4.44	2.31,1.99	C^{γ} H 2.11,2.04 C^{δ} H 3.80,3.66
Gly48	8.25	4.05,3.91		
Lys49	7.93	4.48	1.92,1.84	C^{γ} H 1.59,1.49 C^{δ} H 1.74,1.74 C^{ϵ} H 3.04,3.04 N^{ζ} H 7.59
Ser50	8.14	4.57	3.91,3.91	
Ile51	7.89	4.31	1.94	C^{γ} H 1.55,1.22 $C^{\gamma}{}_{M}$H (0.97) C^{δ} H (0.97)
Gly52	8.00	4.04,3.96		

*Chemical shifts are measured to ± 0.01 ppm. The chemical shift values in parentheses are approximate because of spectral overlaps.

Some of the NOESY peaks are overlapping and could not be identified as between unique pair of protons. Therefore, the absence of connectivities in this figure means either real absence of NOE or inability of identifying NOE because of an overlap. In a region from Ser28 to Val36, most of NH_i-NH_{i+1} NOEs are intense and several α_i-NH_{i+3} and α_i-β_{i+3} NOEs are observed, while α_i-NH_{i+1} NOEs are weak. In a region from Ala46 to Lys49, an NH_{48}-NH_{49} NOE and an α_{47}-NH_{49} NOE are observed, and α_{47}-NH_{48} NOE is stronger than the other α_i-NH_{i+1} NOEs. These facts indicate that M2 may have a helical conformation and a turn of type II in the regions from Ser28 to Val36 and from Ala46 to Lys49, respectively.

Barnase M3

The fingerprint region of the DQF-COSY spectrum of M3 in solution is shown in Figure 6, where correlations of the chemical shifts of most of the amide and α protons were detected. However, amide and α protons of a glycine, Gly52 or Gly53, were not detected, presumably due to fast exchange with the solvent. In addition to these cross-peaks, cross-peaks of C^{δ} H_2, of arginine residues and C^{ϵ} H_2, N^{ζ} H of lysine residues, respectively, were also observed due to the acidic pH condition of the solution, where exchange rates of N^{ϵ} H and N^{ζ} H protons with solvent are slow. Module M3 contains three arginine and two lysine residues. Although it is generally diffi-

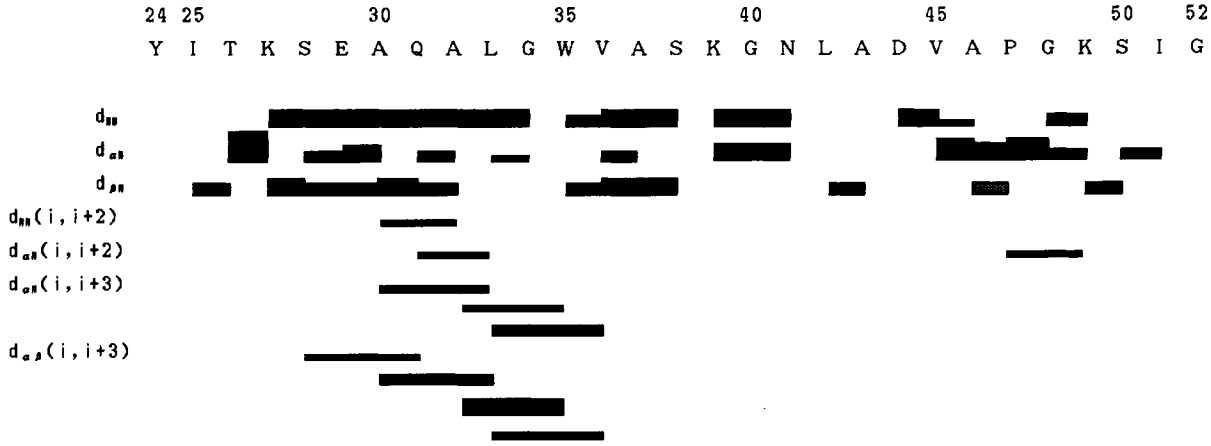


Fig. 5. Distance connectivities observed for module M2. Instead of $d_{\alpha N}$ and $d_{\beta N}$, $d_{\alpha\delta}$ and $d_{\beta\delta}$ were used for the connectivity of Ala46-Pro47. The height of the bars indicates the approximate magnitude of the NOESY cross-peaks.

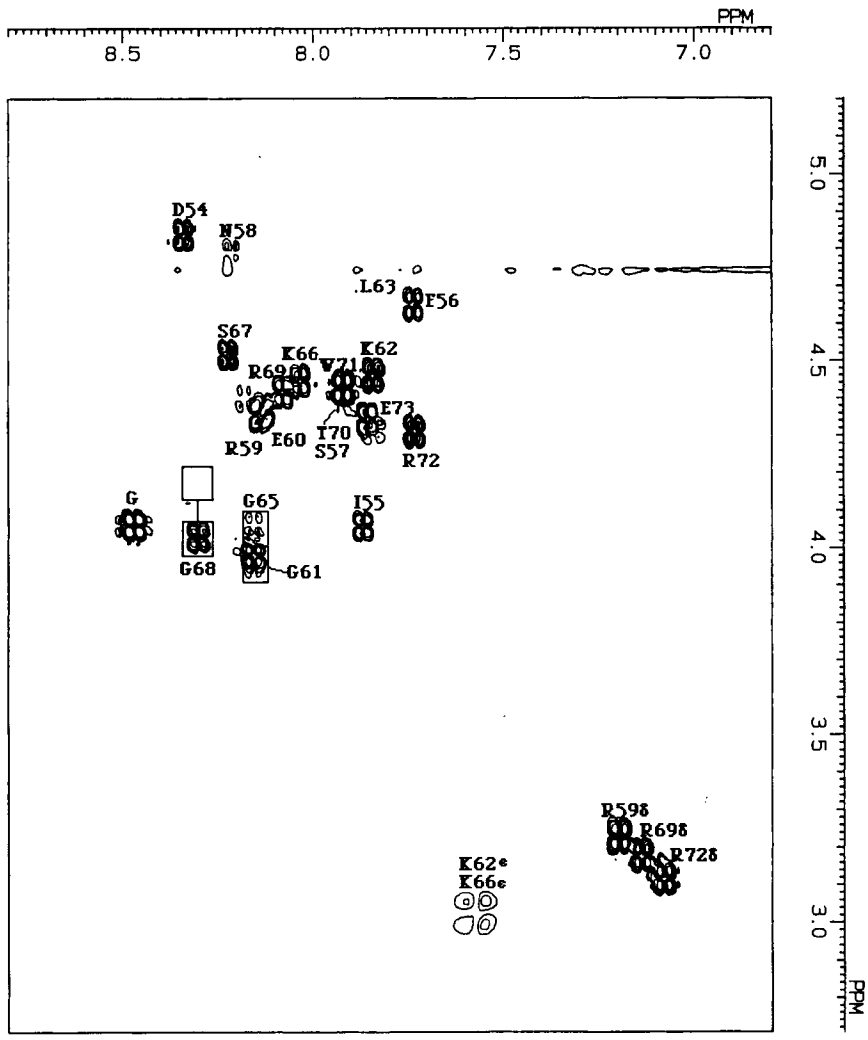


Fig. 6. Fingerprint region of DQF-COSY spectrum of Barnase M3 with assignments of the cross-peaks. For a Gly residue, two cross-peaks C^{α} H, NH and C^{α} 'H, NH are enclosed by boxes and connected by a vertical line. The cross-peaks C^{δ} H_2, N^{ε} H of Arg residues and C^{ε} H_2, N^{ζ} H of Lys residues are also shown.

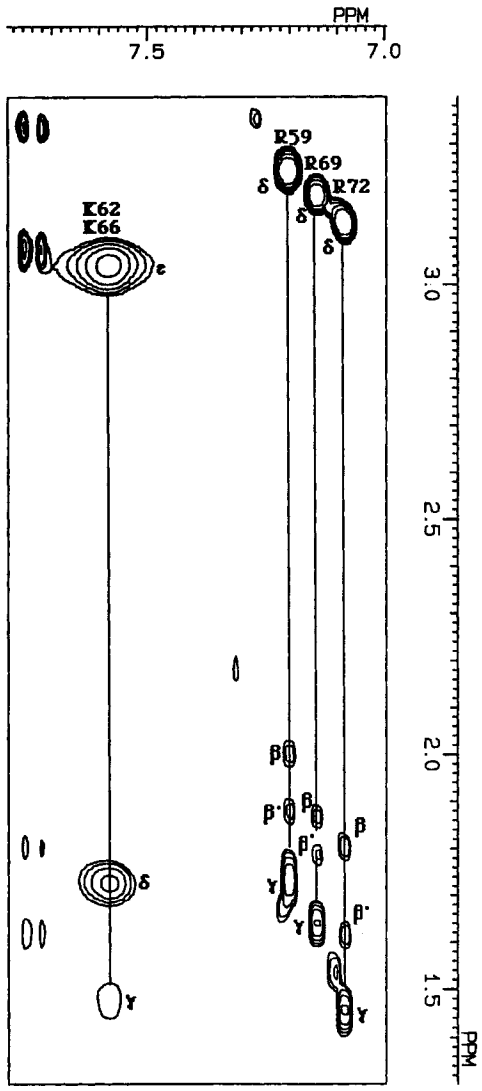


Fig. 7. HOHAHA spectrum of module M3 in the cross-peak region of aliphatic proton, N^{ϵ} H of Arg residues and aliphatic proton, N^{ζ} H of Lys residues with a mixing time of 35 msec. The spin system of each of the Arg and Lys residues is connected by a solid line with the assignment of each proton resonance.

cult to assign the side-chain proton resonances of these residues, magnetization transfer starting from N^{ϵ} H/N^{ζ} H to other side-chain protons of arginine/lysine residues was observed in the HOHAHA spectrum of M3 in solution. Thus, the spin systems of each arginine and lysine residue were assigned as shown in Figure 7. Figure 8 shows the cross-peaks between an α proton and side-chain protons in the HOHAHA spectrum, by which the chemical shift correlations were further extended to side-chain protons. It is to be noted that magnetization transfers from N^{ϵ} H of arginine to C^{δ} H_2, C^{β} H_2, and C^{γ} H_2, and from N^{ζ} H of lysine to C^{ϵ} H_2, C^{δ} H_2, and C^{γ} H_2 were also observed in HOHAHA (Fig. 7). However, overlapped cross-peaks at about ω_1 = 4.40 ppm made the assignment of chemical shifts of several side-chain protons difficult. Nevertheless, all the observable protons were assigned to characteristic spin systems and to specific types of amino acid residues.

Next, the spin systems were aligned according to the amino acid sequence of M3 using distance connectivity by NOESY. Figure 9 shows the NOESY spectrum of the amide proton region of M3 in solution. A few cross-peaks due to d_{NN} connectivities were sequentially connected. The other connections were done by using $d_{\alpha N}$ and/or $d_{\beta N}$ connectivities. The connectivities thus obtained were also confirmed on the basis of the spin system identifications of the side chains already established through the analyses of the HOHAHA and DQF-COSY spectra. Thus, combined with all the 2D NMR data, the se-

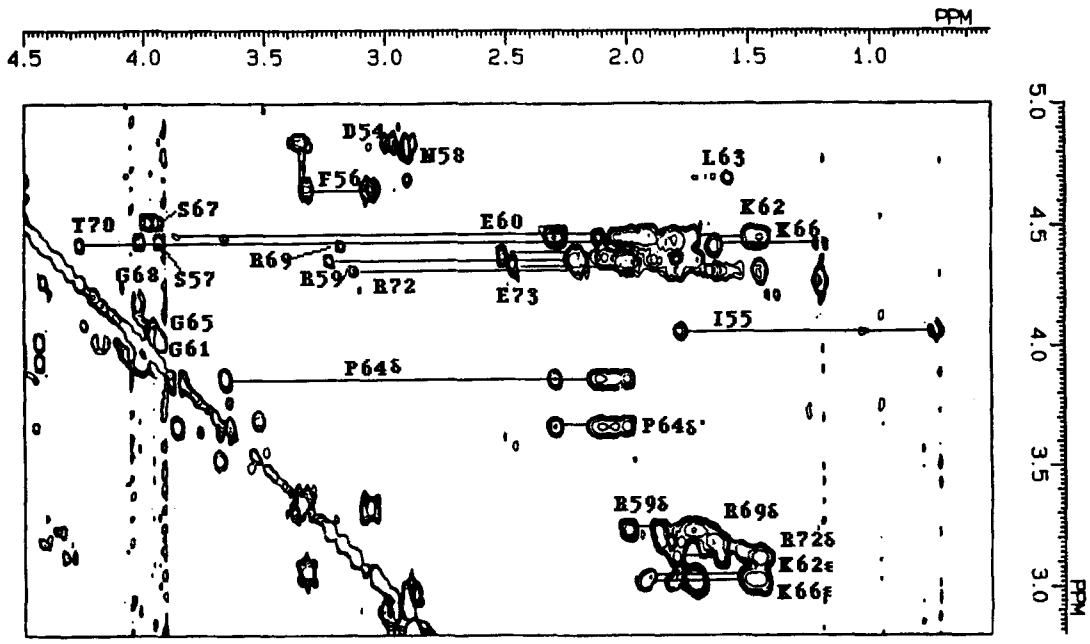


Fig. 8. HOHAHA spectrum of module M3 in the cross-peak region of the α proton and aliphatic proton resonance with a mixing time of 35 msec. The spin system of each amino acid residue is connected by a solid line at the chemical shifts of the α proton resonance. Spin systems developed from C^{δ} H_2 of Pro64, C^{δ} H_2 of Arg59, 69, 72, and C^{ϵ} H_2 of Lys62, 66, are also shown as solid lines. There are several cross-peaks from impurities.

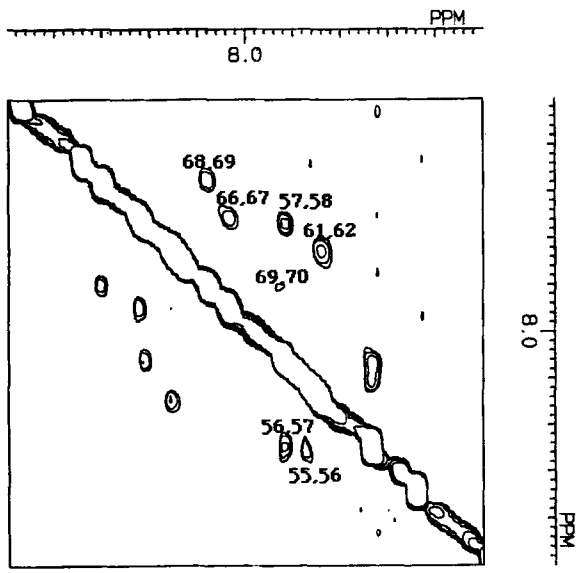


Fig. 9. Amide proton region of the NOESY spectrum of module M3 with a mixing time of 150 msec. NOESY cross-peaks due to the d_{NN} connectivities are labeled with the residue numbers of corresponding amino acids.

quential assignments of M3 in solution were accomplished, except for Gly52 and Gly53. Though a set of chemical shifts for glycine was observed at NH = 8.47 ppm and C^{α} Hs = 4.07, 4.07, there were two glycine residues unassigned. Therefore, it was impossible to accomplish the assignment of them. Table II shows a list of chemical shifts for M3 in solution. Figure 10 summarizes the distance connectivities for M3. For the same reason as in Figure 5, absence of connectivities may be due to inability of identifying NOE because of an overlap. Only one α_i-NH_{i+2} NOE is observed between Ser67 and Arg69. No NH_i-NH_{i+2}, α_i-NH_{i+3}, and α_i-β_{i+3} NOEs are observed.

Chemical shifts

The chemical shifts in peptides are sensitive to local electronic structure, protonation equilibria, hydrogen bonding, and shielding effects from the conformation. The proximity of positive and negative charges, aromatic rings, and groups with anisotropic magnetic susceptibility are remarkable among such effects. In these cases, the chemical shifts deviate from the random-coil values.[33] We examined the deviations of the chemical shifts presented in Tables I and II from random-coil values. Several chemical shifts clearly deviated. For example, most of α- and β-protons between Tyr24 and Gly40 in M2 have distinct values. On the other hand, there were some protons which had the value nearly equal to the corresponding random-coil chemical shifts. We enumerate the following protons as examples; α-protons of Asp44, Val45, Pro47; side-chain protons of Lys27, Leu42, Val45, Pro47, Lys49 in M2; α-protons of Phe56, Gly61, Pro64, Ser67; and

TABLE II. Chemical Shifts of Proton Resonances of Barnase M3 at 28°C and pH 3.0*

Residue	Chemical shift (ppm) NH	C^{α} H	C^{β} H	Others
Gly52	—	—		
Gly53	—	—		
Asp54	8.34	4.83	2.97,2.92	
Ile55	7.87	4.07	1.79	C^{γ} H 1.10,1.04 $C^{\gamma}{}_{M}$H (0.72) C^{δ} H 0.78
Phe56	7.74	4.66	3.34,3.07	C2,6H 7.28 C3,5H (7.34) C4H (7.34)
Ser57	7.91	4.43	4.02,3.93	
Asn58	8.22	4.78	2.91,2.91	N^{δ} H 7.42,6.77
Arg59	8.14	4.35	2.01,1.88	C^{γ} H 1.73,1.73 C^{δ} H 3.24,3.24 N^{ϵ} H 7.19
Glu60	8.13	4.36	2.21,2.10	C^{γ} H 2.52,2.52
Gly61	8.15	3.99,3.99		
Lys62	7.84	4.47	1.91,1.79	C^{γ} H 1.48,1.48 C^{δ} H 1.73,1.73 C^{ϵ} H 3.04,3.04 N^{ζ} H 7.58
Leu63	7.90	4.69	1.68,1.68	C^{γ} H 1.59 C^{δ} H 0.97,0.97
Pro64		4.45	2.30,2.00	C^{γ} H 2.12,2.05 C^{δ} H 3.87,3.66
Gly65	8.15	4.04,3.98		
Lys66	8.03	4.45	1.93,1.82	C^{γ} H 1.48,1.48 C^{δ} H 1.73,1.73 C^{ϵ} H 3.02,3.02 N^{ζ} H 7.58
Ser67	8.22	4.52	3.96,3.96	
Gly68	8.30	4.17,4.03		
Arg69	8.08	4.42	2.00,1.87	C^{γ} H 1.64,1.64 C^{δ} H 3.18,3.18 N^{ϵ} H 7.14
Thr70	7.92	4.42	4.27	C^{γ} H 1.20
Trp71	7.89	4.44	3.34,3.34	N1H 9.87 C2H (7.23) C4H 7.64 C5H 7.18 C6H (7.23) C7H 7.47
Arg72	7.73	4.31	1.79,1.61	C^{γ} H 1.45,1.45 C^{δ} H 3.10,3.10 N^{ϵ} H 7.08
Glu73	7.85	4.35	2.20,2.02	C^{γ} H 2.46,2.46

*Chemical shifts are measured to ± 0.01 ppm. The chemical shift values in parentheses are approximate because of spectral overlaps. Chemical shifts for Gly52 and Gly53 are not assigned.

Fig. 10. Distance connectivities observed for module M3. Instead of $d_{\alpha N}$ and $d_{\beta N}$, $d_{\alpha\delta}$ and $d_{\beta\delta}$ were used for the connectivity of Leu63-Pro64. The height of the bars indicates the approximate magnitude of the NOESY cross-peaks.

side-chain protons of Lys62, Lys66, Trp71 in M3. Figure 11 shows the deviations of chemical shifts of α- and β-protons, which were obtained by subtraction of the corresponding random-coil chemical shifts from the observed shifts.

Distance Geometry Calculation

The distance geometry calculations were made on the basis of the distance constraints obtained from the analysis of the NOESY spectra. The constraints

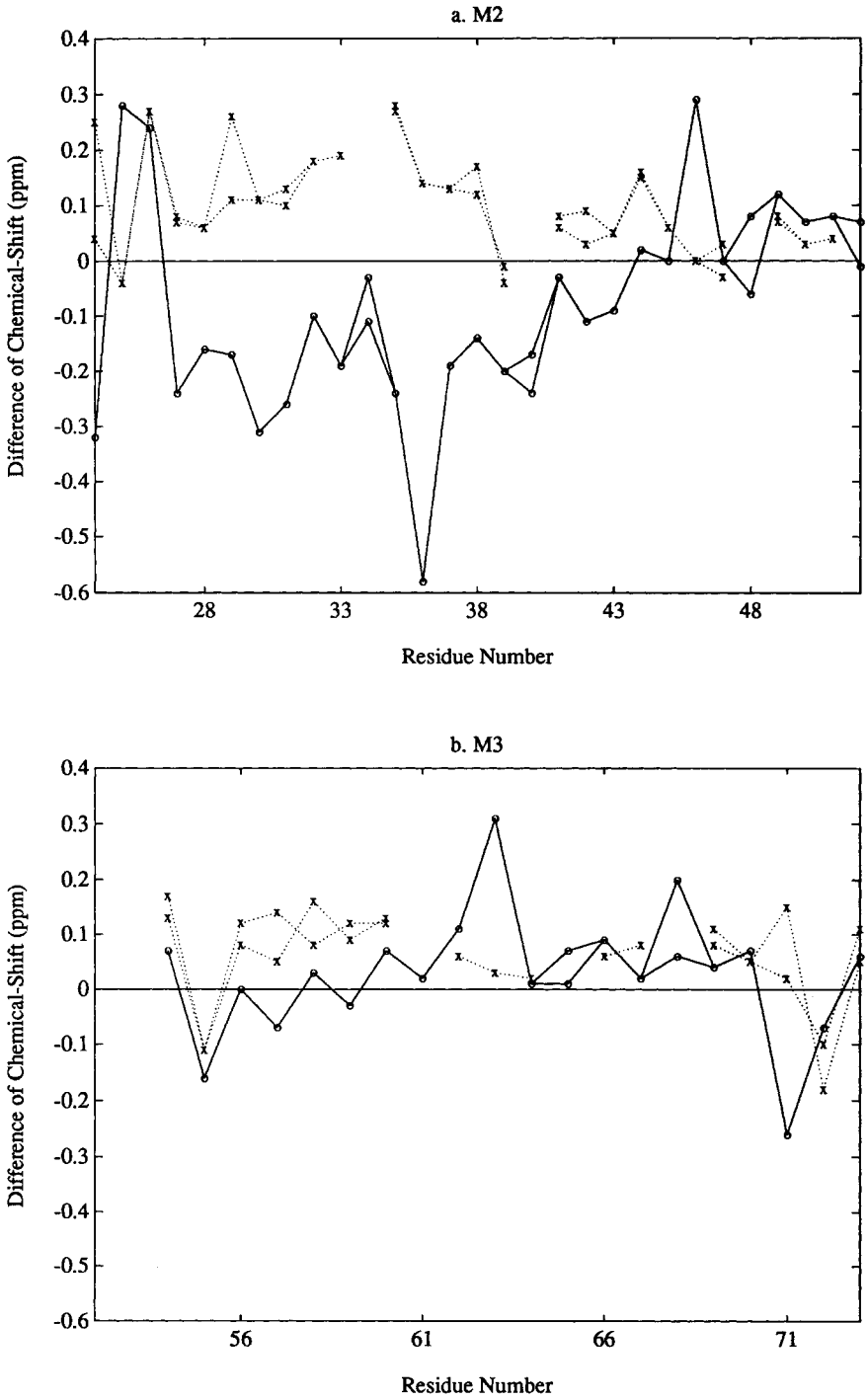


Fig. 11. Conformation-dependent chemical shifts of α-protons (marked as o) and β-protons (marked as x) presented as functions of residue numbers for (a) module M2 and (b) module M3. These conformation-dependent chemical shifts were obtained by subtraction of the corresponding random-coil chemical shifts from the measured chemical shifts. The chemical shift values of the α-protons of Gly52 and Gly53 for M3 could not be determined because of overlap problems.

were obtained by classifying NOESY cross-peak intensities into several categories based on counting the contour levels in the NOESY spectrum. In the case of M2, NOESY cross-peaks were divided into three categories for intraresidue NOEs, four for sequential NOEs, and one for long-range NOEs. In the case of M3, NOESY cross-peaks were divided into three categories for intraresidue and sequential NOEs, and one for long-range NOEs. Upper bound constraints for each case were summarized in Table

TABLE III. Upper Bound Constraints for Distance Geometry Calculations for M2 and M3(Å)*

	M2				M3		
	Strong	Medium strong	Medium	Weak	Strong	Medium	Weak
Intraresidue	2.5		3.0	4.0	2.5	3.0	4.0
Sequential	2.5	3.0	3.5(3.0)	4.0	2.5	3.0	4.0
Long-range				4.0			4.0

*The value in the parentheses is applied to a distance constraint between a pair of α and amide protons.

TABLE IV. Statistics of Violations of the Residual Distance Constraints in the Conformations of M2 Computed From the NOE Data Using DADAS*

Conformation (residual value of target function, Å^2)	Numbers of violations (Å)			
	In distance constraints			In steric constraints
	(> 0.1 Å	> 0.3 Å	> 0.5 Å)	(> 0.1 Å)
1 (0.331)	1	1	0 (0.49)	1 (0.56)
2 (0.332)	1	1	0 (0.49)	1 (0.14)
3 (0.352)	1	1	0 (0.50)	0
4 (0.377)	0	0	1 (0.53)	0
5 (0.379)	2	0	1 (0.52)	1 (0.15)
6 (0.381)	1	1	0 (0.46)	0
7 (0.384)	3	1	0 (0.49)	0
8 (0.385)	3	1	0 (0.49)	1 (0.12)
9 (0.392)	2	1	0 (0.49)	0

*The largest violations are given in parentheses in the columns for violations.

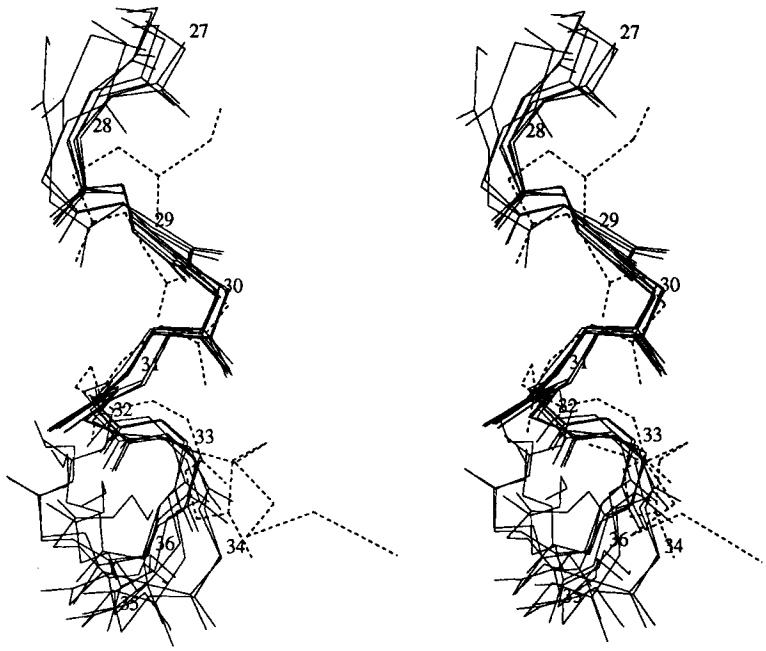
III. Since stereo-specific resonance assignments could not be done for the individual protons in methylene groups or methyl groups, pseudo-atom representations were used where appropriate. A total of 92 intraresidue, 40 sequential, and 21 long-range distance constraints were used in the distance geometry calculation for M2. For M3, 75, 27, and 7 distance constraints were used in the same way.

Two hundred initial conformations and 100 initial conformations were generated from randomly selected dihedral angles for M2 and M3, respectively. After a number of distance geometry calculations, the nine conformations for each peptide which best satisfied the experimental distance constraints were selected for further study and comparison.

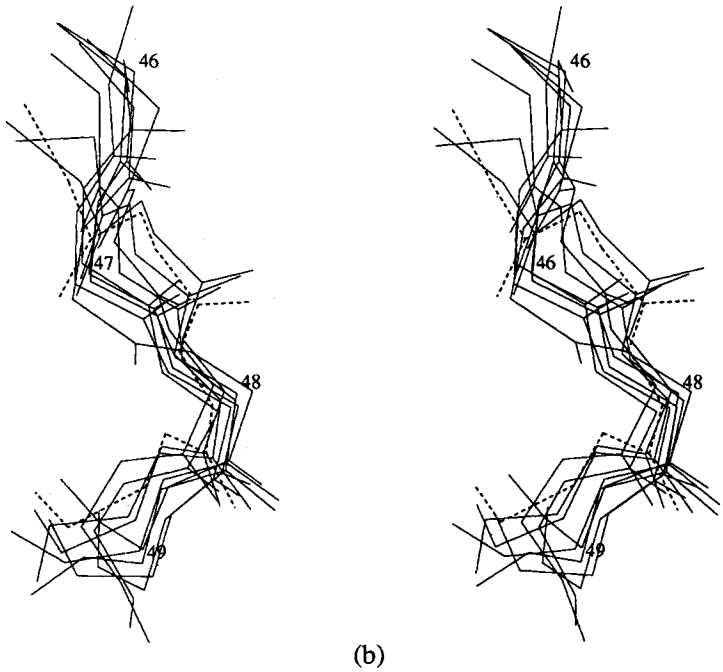
Table IV shows the residual violations of the distance constraints for M2. The largest violations of the distance constraints for NOEs in the nine conformations were less than 0.53Å. The steric violations larger than 0.15Å were not observed except in one conformation (conformation 1). The residual values of the target function were 0.331 ~ 0.392Å^2. Figure 12 shows the superposition of the nine backbone conformations for two segments of (a) Lys27-Val36 and (b) Ala46-Lys49 of M2, which are drawn as solid lines. The RMSDs for the backbone atoms for each pair of the nine conformations were calculated. The average RMSDs for two segments of Lys27-Val36 and Ala46-Lys49 of M2 were 1.2 and 1.1Å, respectively. In these conformations, segment Lys27-Val36 takes a stretched α-helical conformation, and segment Ala46-Lys49 takes a turn-like conformation. In the conformation of the intact barnase revealed by the X-ray crystallographic analysis,[21] the segment of Lys27-Ala32 is α-helical, and the segment of Ala46-Lys49 is a β-turn of type II. This indicates that the conformations of these segments in the intact barnase are well conserved in the conformations of the segments in M2. The RMSDs calculated between the average conformation of M2 and X-ray conformation of the intact barnase are 1.0Å for the segment of Ser28-Leu33 and 1.4Å for Ala46-Lys49 (Fig. 12).

According to Szilágyi and Jardetzky, the α-protons of aliphatic amino acids in α-helical conformations shift upfield by 0.2–0.6 ppm on average.[34] In the present study, the deviations of chemical shifts of α-protons of Lys27-Val36 are between −0.58 and −0.10 ppm except one α-proton of Gly34 (Fig. 11). This result supports that the segment takes a helical conformation.

The other segments, Tyr24-Thr27, Ala37-Val45, and Ser50-Gly52, took no well-defined conformations by distance geometry calculations. The fact indicates that these segments take random-coils, or that they take several well-defined conformations and their conformational changes occurred rapidly. Judging from the chemical shifts, we consider the C-terminal segment of Ser50-Gly52 to be random-coil, since all of the chemical shifts are nearly equal



(a)



(b)

Fig. 12. Stereo view of superposition of the two segments (**a**) from Lys27 to Val36 and (**b**) from Ala46 to Lys49 in the ten conformations of module M2, which consist of nine conformations of M2 by NMR and distance geometry calculation (solid lines) and the corresponding segment in the conformation of the intact barnase revealed by X-ray crystallographic analysis (dashed line). The average backbone RMSDs in the nine NMR conformations are (a) 1.2Å for a segment of Lys27-Val36 and (b) 1.1Å for a segment of Ala46-Lys49. The backbone RMSDs between the average NMR conformation and the X-ray conformation are (a) 1.0Å for a segment of Ser28-Leu33 and 1.4Å for a segment of Ala46-Leu49. The N, C^{α}, C′, and O atoms are drawn. The C^{α} atoms are labeled with the residue numbers of corresponding amino acids in the intact barnase.

TABLE V. Statistics of Violations of the Residual Distance Constraints in the Conformations of Barnase M3 Computed From the NOE Data Using DADAS*

Conformation (residual value of target function, $\times 10^{-1}$ Å^2)	Numbers of violations (Å)	
	In distance constraints (> 0.0 Å)	In steric constraints (> 0.1 Å)
1 (0.013)	5 (0.03)	1 (0.16)
2 (0.034)	7 (0.06)	0
3 (0.096)	6 (0.06)	0
4 (0.134)	1 (0.00)	1 (0.20)
5 (0.156)	3 (0.02)	0
6 (0.346)	1 (0.00)	2 (0.40)
7 (0.875)	3 (0.00)	1 (0.11)
8 (0.876)	4 (0.06)	3 (0.23)
9 (0.935)	3 (0.03)	4 (0.19)

*The largest violations are given in parentheses in the columns for violations.

to the random-coil values (Fig. 11). However, we cannot judge the conformations of the other segments.

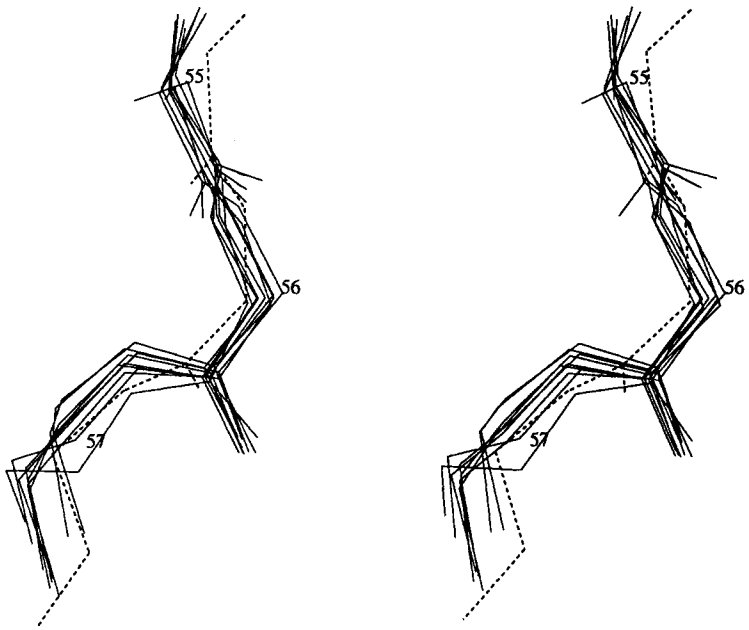
Table V shows the residual violations of the distance constraints for M3. The largest violations of the distance constraints for NOEs in the nine conformations were less than 0.06Å. The steric violations larger than 0.40Å were not observed. The residual values of the target function were less than 0.100Å^2. Figure 13 shows the superposition of the nine backbone conformations for the two segments of (a) Ile55-Ser57 and (b) Lys66-Arg69 of M3, which are drawn as solid lines. The average RMSDs for the two segments of Ile55-Ser57 and Lys66-Arg69 of M3 were 0.6 and 1.2Å, respectively. Compared with the X-ray conformation of intact barnase, the RMSDs are 1.0Å for Ile55-Ser57 and 1.8Å for Lys66-Arg69 (Fig. 13). This indicates that the conformations of these segments in the intact barnase are well conserved in the conformations of the segments of M3.

The other segments, Gly52-Asp54, Asn58-Gly65, and Thr70-Glu73, took no well-defined conformations by distance geometry calculations. Moreover we cannot find the distinct difference between well-defined segments and undetermined segments in point of the chemical shifts. Therefore we have no judgment about the conformations of these segments.

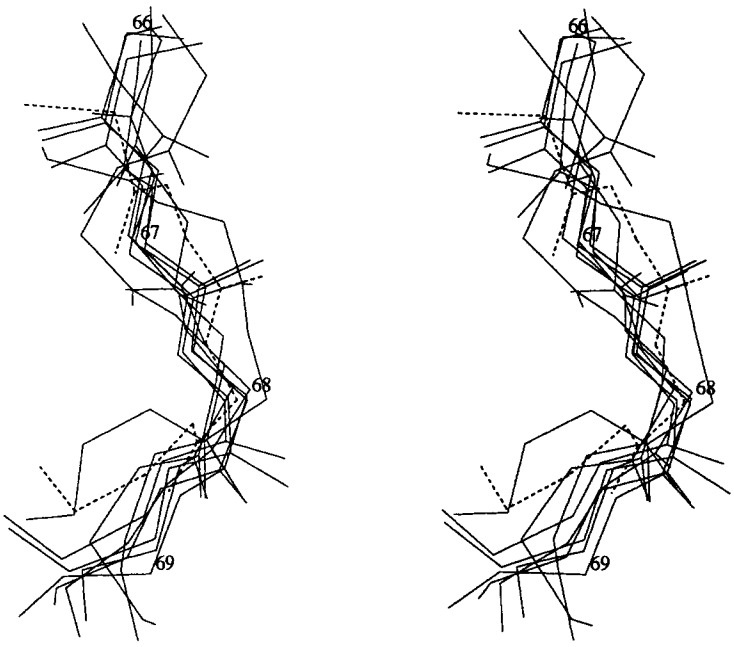
DISCUSSION

In the crystal structure of intact barnase, M2 is made up from a short stretch of a β-strand at Tyr24-Ile25, an α-helix at Lys27-Ala32, a turn-like structure at Ala37-Gly40, an α-helix at Leu42-Val45, a β-turn at Ala46-Lys49, and another short stretch of a β-strand at Ser50-Ile51. The two short stretches of the β-strands form a parallel β-sheet bridging over N- and C-termini of the module. Structures similar to the first α-helix and the β-turn were found in the corresponding locations in the isolated M2. Module M3 in the crystal structure of intact barnase has two β-turns at Asn58-Gly61 and in Lys66-Arg69, and an anti-parallel β-sheet consisting of segments, Ile55-Phe56 and Trp71-Glu73. The isolated M3 were found to form a β-turn in the segment Lys66-Arg69 and an extended structure in Ile55-Ser57. In summary, about half of the secondary structures in the intact protein were found to be formed in both of the isolated modules, M2 and M3. It is possible that the remaining secondary structures are formed in the solution condition studied, but in too small a population to be detected by NMR. The helical region in the isolated M2 is Lys27-Val36, whereas the corresponding regions in the X-ray crystal structure and in the NMR solution structure[35] of the intact barnase are Lys27-Ala32 and Thr26-Gly34, respectively. Except for this slight discrepancy of boundaries of the helical region, observed secondary structures were in the locations corresponding to those in the intact barnase. The results of this study indicate that the amino acid sequences of the two modules, M2 and M3, contain structural information that specifies at least about half of the secondary structures in each module. Boundaries of the secondary structures appear to be specified up to a few residues. We used water/TFE mixtures for the solvent. This solvent is known to have an effect of stabilizing α-helices or, in other words, increasing a population of α-helices. But it does not have an effect of changing the locations of α-helices. For example, Sancho et al. studied an N-terminal fragment of barnase consisting of residues 1 to 36 both in water and in 35% TFE by CD and 2D-NMR.[36] In water two helices in the regions corresponding to those in the intact barnase were observed by NMR, but not by CD. In 35% TFE, population of molecules with α-helices increased, but α-helices were found only in the same regions. The same effect of TFE was observed in studies of peptide fragments of lysozyme,[37] myohemerythrin,[38] and plastocyanin.[39]

Recent 2D-NMR studies revealed that peptide fragments of proteins in water or water/TFE mixtures form secondary structures corresponding to those in the intact proteins.[36,38–40] Those structures are in rapid equilibrium with unfolded states and marginally stable to be detected. The conformation of the region of Lys27-Ala32 of M2 is an α-helix in the crystal structure of the intact barnase. In this region α_i-NH_{i+1}NOEs were observed, as well as NH_i-NH_{i+1}NOEs that are characteristic for an α-helical region. The obtained α-helical structure indicates that NOE connectivity should exist between residues 34 and 35, although the connectivity could not be identified due to an overlap. The α_i-NH_{i+1} distance varies from 3.5Å of α-helix to 2.2Å of β-strand, i.e., the shorter the α_i-NH_{i+1} distance, the more stretched the peptide backbone. The observed



(a)



(b)

Fig. 13. Stereo view of superposition of two segments (**a**) from Ile55 to Ser57 and (**b**) from Lys66 to Arg69 in the ten conformations of module M3, which consist of nine conformations of M3 by NMR and distance geometry calculation (solid lines) and the corresponding segment in the conformation of the intact barnase revealed by X-ray crystallographic analysis (dashed line). The average backbone RMSDs in the nine NMR conformations are (a) 0.6Å for a segment of Ile55-Ser57 and (b) 1.2Å for a segment of Lys66-Arg69. The backbone RMSDs between the average NMR conformation and the X-ray conformation are (a) 1.0Å for a segment of Ile55-Ser57 and (b) 1.8Å for a segment of Leu66-Arg69. The N, C^{α}, C′, and O atoms are drawn. The C^{α} atoms are labeled with the residue numbers of corresponding amino acids in the intact barnase.

intensities of α_i-NH_{i+1} NOEs were calibrated to use the upper bound constraint of 3.0Å on the α_i-NH_{i+1} distances in the distance geometry calculations. Though the distance of 3.0Å is small for the α_i-NH_{i+1} of α-helix, we used the unmodified value of 3.0Å because there is no a priori reason to assume an α-helix before the distance geometry calculation. Because of this constraint the stretched α-helical conformation was derived from the distance geometry calculations. Since α_i-NH_{i+1} distance is shorter in other conformations than in the α-helix, the observation of α_i-NH_{i+1} NOEs is possibly a result of rapid equilibrium between the α-helix and random coils.

We studied the hydrogen-bond network and hydrophobic interactions within and between modules in barnase by analyzing its crystal structure. Hydrogen bonds are formed mainly within modules.[23] This finding shows that each module has the propensity within itself for taking its native conformation. The native-like secondary structures observed in the isolated modules is consistent with this propensity. Tertiary structures of the two isolated modules were, however, not uniquely determined by the distance geometry calculation, showing that the isolated modules do not have the definite tertiary structures stable enough to be detected. The isolated modules exist probably in a state where their secondary structures are in a dynamic equilibrium with unfolded states. Matouschek et al. studied the folding pathway of barnase by protein engineering and NMR.[41] They found that secondary structures were formed more rapidly than the overall folding. The study by Sancho et al. of the N-terminal fragment of barnase consisting of residues 1 to 36, which contains module M1 and an N-terminal half of the module M2, showed that the fragment had residual helical structures similar to that in a refolding intermediate.[36] The equilibrium state of isolated modules may thus correspond to an intermediate state of the folding of the intact barnase. Oas and Kim studied the solution structure of two fragments of BPTI connected by a disulfide bond.[17] They found the complex of the fragments formed a tertiary structure similar to the corresponding part of the native structure of the intact BPTI. At the interface of the two fragments exists a hydrophobic core. Their experimental results demonstrated that the interactions between the fragments were essential for the stability of the tertiary structure of the fragments. Hydrophobic cores exist between modules in barnase.[23] Interactions between modules should contribute to increase the stability of the tertiary structure of each module.

Present study shows that the excised and isolated modules M2 and M3 of barnase have a propensity to take a locally similar conformation in solution with that of intact barnase. This propensity should be an important feature of modules advantageous as parts recruited into a globular protein through exon shuffling in early evolution. The propensity was probably enhanced by interactions with other modules when several modules are combined into a globular protein. In prebiological evolutionary period, the selection pressure worked possibly on stability of independent modules encoded by mini genes. After an assembly of modules into a globular domain or protein, the selection pressure might be switched on the interaction between modules, particularly on hydrophobic interactions which contribute to make the assembly into a stable globular conformation.

Amino acid replacements which occurred after involvement of modules into an assembly probably changed the conformation of the module slightly and made the connected modules stable. Particularly the amino acid replacements should happen at the joints and interface of the modules. Thus, some of the contemporary modules are unstable in solution when they are disconnected from each other. Compact structure similar to the original modules, however, still seems maintained as a remnant in a globular protein.

ACKNOWLEDGMENTS

We thank Drs. G. Dodson and C. Hill for providing the X-ray atomic coordinates of barnase, Dr. S. Ohno for his comments on the purification of synthesized modules, Mr. S. Ono, Dr. F. Tokunaga, Ms. S. Kajiyama, and Ms. Y. Nishina for their assistance in purifying the modules and amino acid analysis. Part of this work was supported by Grants-in-Aid from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

1. Blake, C.C.F. Exons and the evolution of proteins. Int. Rev. Cytol. 93:149–185, 1985.
2. Gilbert, W. Genes-in-pieces revisited. Science 228:823–824, 1985.
3. Patthy, L. Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell 41:657–663, 1985.
4. Südhof, T.C., Russel, D.W., Goldstein, J.L., Brown, M.S., Sanchez-Pescador, R., Bell, G.I. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. Science 228:893–895, 1985.
5. Hunkapiller, T., Hood, L. Diversity of the immunoglobulin gene superfamily. Adv. immunol. 44:1–63, 1989.
6. Brändén, C-I., Eklund, H., Cambillau, C., Pryor, A.J. Correlation of exons with structural domains in alcohol dehydrogenase. EMBO J. 3:1307–1310, 1984.
7. Gō, M., Nosaka, M. Protein architecture and the origin of introns. Cold Spring Harbor Symp. Quant. Biol. 52:915–924, 1987.
8. Gō, M. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90–92, 1981.
9. Gō, M. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. U.S.A. 80:1964–1968, 1983.
10. Isaacs, N., Machin, K.J., Masakuni, M. The three-dimensional structure of the goose type lysozyme from the egg-white of the black swan, *Cygnus atratus*. Aust. J. Biol. 38:13–22, 1985.
11. Gilbert, W., Marchionni, M., McKnight, G. On the antiquity of introns. Cell 46:151–154, 1986.

12. Gō, M. Protein structures and split genes. Adv. Biophys. 19:91–131, 1985.
13. Gilbert, W. Why genes in pieces? Nature 271:501, 1978.
14. Xu, M.-Q., Kathe, S.D., Goodrich-Blair, H., Nierzwicki-Bauer, S.A., Shub, D.A. Bacterial origin of a chloroplast intron: Conserved self-splicing group I introns in cyanobacteria. Science 250:1566–1570, 1990.
15. Kuhsel, M.G., Strickland, R., Palmer, J.D. An ancient group I intron shared by eubacteria and chloroplasts. Science 250:1570–1573, 1990.
16. Darnell, J.E., Doolittle, W.F. Speculations on the early course of evolution. Proc. Natl. Acad. Sci. U.S.A. 83:1271–1275, 1986.
17. Oas, T.G., Kim, P.S. A peptide model of a protein folding intermediate. Nature 336:42–48, 1988.
18. Chu, F.K., Maley, G.F., Maley, F., Belfort, M. Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. U.S.A. 81:3049–3053, 1984.
19. Nishimura, S., Nomura, M. Ribonuclease of *Bacillus subtilis*. J. Biochem. 46:161–167, 1959.
20. Hartley, R.W., Barker, E.A. Amino-acid sequence of extracellular ribonuclease (Barnase) of *Bacillus amyloliquefaciens*. Nature New Biol. 235:15–16, 1972.
21. Mauguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C., Jack, A. Molecular structure of a new family of ribonucleases. Nature 297:162–164, 1982.
22. Baudet, S., Janin, J. Crystal structure of a barnase-d (GpC) complex at 1.9Å resolution. J. Mol. Biol. 219:123–132, 1991.
23. Noguti, T., Sakakibara, H., Gō, M. Localization of hydrogen-bonds within modules in barnase. Proteins 16:357–363, 1993.
24. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., Wüthrich, K. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. 117:479–485, 1983.
25. Bax, A., Davis, D.G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. 65:355–360, 1985.
26. Jeener, J., Meier, B.H., Bachmann, P., Ernst, R.R. Investigation of exchanged processes by two-dimensional NMR spectroscopy. J. Chem. Phys. 71:4546–4553, 1979.
27. Macura, S., Huang, Y., Suter, D., Ernst, R.R. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. J. Magn. Reson. 43:259–281, 1981.
28. States, D.J., Haberkorn, R.A., Ruben, D.J. A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. J. Magn. Reson. 48:286–292, 1982.
29. Zuiderweg, E.R.P., Hallenga, K., Olejniczak, E.T. Improvement of 2D NOE spectra of biomacromolecules in H_2O solution by coherent suppression of the solvent resonance. J. Mol. Magn. Reson. 70:336–343, 1986.
30. Klevit, R.E. Improving two-dimensional NMR spectra by t_1 ridge subtraction. J. Magn. Reson. 62:551–555, 1985.
31. Braun, W., Gō, N. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. J. Mol. Biol. 186:611–626, 1985.
32. Wüthrich, K., Wider, G., Wagner, G., Braun, W. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J. Mol. Biol. 155:311–319, 1982.
33. Bundi, A., Wüthrich, K. ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers 18:285–297, 1979.
34. Szilágyi, L., Jardetzky, O. α-proton chemical shifts and secondary structure in proteins. J. Magn. Reson. 83:441–449, 1989.
35. Bycroft, M., Ludvigsen, S., Fersht, A.R., Poulsen, F.M. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. Biochemistry 30:8697–8701, 1991.
36. Sancho, J., Neira, J.L., Fersht, A.R. An N-terminal fragment of barnase has residual helical structure similar to that in a refolding intermediate. J. Mol. Biol. 224:749–758, 1992.
37. Segawa, S., Fukuno, T., Fujiwara, K., Noda, Y. Local structures in unfolded lysozyme and correlation with secondary structures in the native conformation: Helix-forming or -breaking propensity of peptide segments. Biopolymers 31:497–509, 1991.
38. Dyson, H.J., Merutka, G., Waltho, J.P., Lerner, R.A., Wright, P.E. Folding of peptide fragments comprising the complete sequence of proteins: Models for initiation of protein folding I. Myohemerythrin. J. Mol. Biol. 226:795–817, 1992.
39. Dyson, H.J., Sayre, J.R., Merutka, G., Shin, H-C., Lerner, R.A., Wright, P.E. Folding of peptide fragments comprising the complete sequence of proteins: Models for initiation of protein folding II. Plastocyanin. J. Mol. Biol. 226:819–835, 1992.
40. Wright, P.E., Dyson, H.J., Lerner, R.A. Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. Biochemistry 27: 7167–7175, 1988.
41. Matouschek, A., Kellis, J.T., Jr., Serrano, L., Bycroft, M., Fersht, A.R. Transient folding intermediates characterized by protein engineering. Nature 346:440–445, 1990.