Loebermann, H., Tokuoka, R., Deisenhofer, J., & Huber, R. (1984) *J. Mol. Biol. 177*, 531-556.
Long, G. L., Chandra, T., Woo, S. L. C., Davie, E. W., & Kurachi, K. (1984) *Biochemistry 23*, 4828-4837.
Lu, H., Lin, F., Chao, L., & Chao, J. (1989) *Int. J. Pept. Protein Res. 33*, 237-249.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning—A Laboratory Manual*, pp 280-281, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marshall, R. D. (1974) *Biochem. Soc. Symp. 40*, 17-26.
Messing, J. (1983) *Methods Enzymol. 101*, 20-77.
Monaci, P., Nicosia, A., & Cortese, R. (1988) *EMBO J. 7*, 2075-2087.
Murray, S., Chao, J., Lin, F., & Chao, L. (1989) *J. Cardiovasc. Pharmacol.* (in press).
Pannell, R., Johnson, D., & Travis, J. (1974) *Biochemistry 13*, 5349-5445.
Rogers, J., Kalsheker, N., Wallis, S., Speer, A., Coutelle, C. H., Woods, D., & Humphries, S. E. (1983) *Biochem. Biophys. Res. Commun. 116*, 375-382.
Sanger, F., Nicklen, S., & Coulson, A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Shai, S., Woodley-Miller, C., Chao, J., & Chao, L. (1989) *Biochemistry 28*, 5334-5343.
Shimamoto, K. Chao, J., & Margolius, H. S. (1980) *J. Clin. Endocrinol. Metab. 51*, 840-848.
Travis, J., & Salvesen, G. S. (1983) *Annu. Rev. Biochem. 52*, 655-709.
Young, R. A., & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194-1198.

High-Resolution Structure of an HIV Zinc Fingerlike Domain via a New NMR-Based Distance Geometry Approach†

Michael F. Summers,*,‡ Terri L. South,‡ Bo Kim,‡ and Dennis R. Hare*,§
Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21228, and Hare Research, Inc., 14810 216th Avenue N.E., Woodinville, Washington 98072
Received May 11, 1989; Revised Manuscript Received August 26, 1989

ABSTRACT: A new method is described for determining molecular structures from NMR data. The approach utilizes 2D NOESY back-calculations to generate simulated spectra for structures obtained from distance geometry (DG) computations. Comparison of experimental and back-calculated spectra, including analysis of cross-peak buildup and auto-peak decay with increasing mixing time, provides a quantitative measure of the consistence between the experimental data and generated structures and allows for use of tighter interproton distance constraints. For the first time, the "goodness" of the generated structures is evaluated on the basis of their consistence with the actual experimental data rather than on the basis of consistence with other generated structures. This method is applied to the structure determination of an 18-residue peptide with an amino acid sequence comprising the first zinc fingerlike domain from the *gag* protein p55 of HIV. This is the first structure determination to atomic resolution for a retroviral zinc fingerlike complex. The peptide [Zn(p55F1)] exhibits a novel folding pattern that includes type I and type II NH-S tight turns and is stabilized both by coordination of the three Cys and one His residues to zinc and by extensive internal hydrogen bonding. The backbone folding is significantly different from that of a "classical" DNA-binding zinc finger. Residues C(1)-F(2)-N(3)-C(4)-G(5)-K(6) fold in a manner virtually identical with the folding observed by X-ray crystallography for related residues in the iron domain of rubredoxin; superposition of all main-chain and Cys side-chain atoms of residues C(1)-K(6) of Zn(p55F1) onto residues C(6)-Y(11) and C(39)-V(44) of rubredoxin gives RMSDs of 0.46 and 0.35 Å, respectively. The side chains of conservatively substituted Phe and Ile residues implicated in genomic RNA recognition form a hydrophobic patch on the peptide surface.

During assembly and budding stages of the retroviral life cycle, a *gag* polyprotein identifies and forms a complex with viral RNA (Dickson et al., 1985; Bolognesi et al., 1978). Subsequent to transporting the RNA to the cell wall for budding, the *gag* polyprotein is cleaved to give, among other products, a low molecular weight nucleic acid binding protein (NABP).[1] Without exception, retroviral *gag* polyproteins and their NABP products contain either one or two conserved sequences of the type C-X_2-C-X_4-H-X_4-C (Henderson et al., 1981; Copeland et al., 1984; Karpel et al., 1987). Berg has

†This work was supported by ACS Institutional Research Grant IN-147F, by the MD Cancer Program/University of Maryland (M.F.S.), and by NIH Grants GM42561-01 (M.F.S.) and GM35620-02 (D.R.H.).
*Author to whom correspondence should be addressed. Reprint requests should be addressed to M.F.S.
‡University of Maryland Baltimore County.
§Hare Research, Inc.

[1] Abbreviations: 2D NMR, two-dimensional nuclear magnetic resonance; 2QF-COSY, double quantum filtered correlated spectroscopy; BKCALC, NOESY back-calculation program; DG, distance geometry; DSPACE, distance geometry program; HIV, human immunodeficiency virus; HMQC, heteronuclear multiple quantum coherence spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; HS, homospoil pulse; HTLV, human T-cell leukemia virus; k_{CR}, parameter used by BKCALC which governs the rate of cross relaxation; k_{ZL}, parameter used by BKCALC to control z leakage rates; MuLV, murine leukemia virus; NABP, nucleic acid binding protein; NOESY, nuclear Overhauser effect spectroscopy; RMSD, root-mean-square deviation; ROESY, rotating-frame Overhauser effect spectroscopy.

0006-2960/90/0429-329$02.50/0

suggested that these sequences may function as zinc-binding domains in a manner similar to that observed for the well-characterized zinc finger domains of transcription factor IIIA and related DNA-binding proteins (Berg, 1986) and has proposed a three-dimensional model for such a structure (Green & Berg, 1989).

Although NABPs isolated from mature virus particles lack zinc (Jentoft et al., 1988; Gorelick et al., 1988), possibly due to the presence of internal disulfides (Jentoft et al., 1989), data are accumulating that support the physiological relevance of zinc binding to the fingerlike motifs at an earlier stage of the retroviral life cycle. Site-directed mutagenesis experiments involving point mutations of the conserved Cys and His residues in the *gag* protein of murine leukemia virus (MuLV) led to the conclusion that the conserved array *within the gag polyprotein* (1) functions as a zinc-binding domain and (2) is intimately involved in retroviral RNA recognition during viral growth, possibly due to interactions between the conservatively substituted amino acids and viral RNA (Gorelick et al., 1988). Fully compatible results were obtained in a study involving point mutations of conservatively substituted aromatic and Gly residues (Meric & Goff, 1989). Given the difficulties associated with isolation of the membrane-bound *gag* polyproteins prior to budding and proteolysis, the site-directed mutagenesis experiments provide the best in vivo evidence in favor of a zinc-binding role for the fingerlike domains during the early stages (i.e., at the *gag* polyprotein level) of the retroviral life cycle.

We demonstrated previously that an 18-residue peptide having the amino acid sequence of the first fingerlike domain from the *gag* protein p55 of HIV (p55F1) binds zinc tightly and stoichiometrically, and we provided unambiguous confirmation that the Cys and His-Nϵ ligands were coordinated to Cd in ^{113}Cd-substituted p55F1 (South et al., 1989). A peptide having the sequence of the first fingerlike domain from the HTLV *gag* protein also binds zinc tightly (Blake and Summers, unpublished results), and Berg has recently obtained evidence that Zn and Co bind tightly to a retroviral fingerlike peptide having the sequence of the MuLV *gag* protein (Green & Berg, 1989). Thus, tight zinc binding appears to be a common feature of peptides containing the retroviral fingerlike sequence.

Since details regarding the structure and metal-binding properties of zinc fingerlike complexes are important for understanding molecular-level aspects of retroviral gene recognition and, additionally, provide a model that can be tested in vivo via new site-directed mutagenesis experiments, we have determined the structure of Zn(p55F1) to high atomic resolution using a novel NMR-based distance geometry (DG) approach. The amino acid sequence of p55F1 is V-K-C-F-N-C-G-K-E-G-H-I-A-R-N-C-R-A. This is the first structure determination to atomic level for a retroviral zinc fingerlike complex and serves to illustrate the significant improvement in structural resolution obtainable when 2D NOESY back-calculations are included in the refinement process.

Experimental Procedures

Computations. NMR data processing and structure calculations were carried out on VAX 8600 mainframe, VAXstation, and Personal Iris computers. NMR data were transferred via Ethernet to these computers and converted to "readable" files by an in-house program (GENET). NMR data processing and analysis, distance geometry calculations, and NOESY back-calculations were performed with FTNMR, DSPACE, BKCALC, and GNOE software packages (Hare Research, Inc.). Space-filling models were generated with the Raster3D software package (D. Bacon, CARB).

NMR Spectroscopy. NMR data were collected on a GE GN500 (500.11 MHz, ^{1}H) NMR spectrometer. Sample conditions were as follows: 15 mg of Zn(p55F1) in D_2O (99.9%, MSD) or 90% H_2O/10% D_2O solution; pH 7.0; T = 30 °C. ^{1}H and ^{13}C NMR chemical shifts were referenced to internal H_2O (4.725 ppm at 30 °C) and external 3-(trimethylsilyl)propionate (TSP), respectively. 2D NMR data were collected without sample spinning. All homonuclear 2D NMR data were processed with zero-filling to a final spectrum size of 2048 × 2048 data points. Pulse sequences and parameters used to collect and process the 2D NMR data are described below.

HOHAHA. Phase-sensitive HOHAHA spectra (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) were obtained for aqueous samples by a 1–$\bar{1}$ echo read pulse [Bax et al. (1987) and references cited therein] with read delay periods of 90 μs (τ_1) and 212 μs (τ_2) and with the carrier set to the frequency of H_2O. No water presaturation was employed. Other parameters were as follows: 2 × 350 × 2048 data matrix size (two separate sets of States–Haberkorn type data, with 350 data points in t_1 and 2048 data points in t_2); 32 scans per t_1 value; pulse delay = 1.4 s; MLEV17 mixing period = 33 ms, preceded and followed by 2.5-ms trim pulses; 7.9-kHz spin-lock field, corresponding to 32-μs 90° pulse widths; 3- and 6-Hz Gaussian filtering in the t_2 and t_1 dimensions, respectively; third-order polynomial base-line correction in the t_1 dimension prior to the t_1 Fourier transform.

ROESY. ROESY spectra (Bothner-by et al., 1984) were collected with the carrier at the H_2O (HDO) frequency using a low-rf continuous-wave spin-lock field (6.25 kHz, corresponding to 40-μs 90° pulse widths) to minimize Hartmann–Hahn artifacts. A 1–$\bar{1}$ echo read pulse was employed for H_2O solutions [Bax et al. (1987) and references cited therein]. Other parameters were as described for the HOHAHA spectra.

NOESY. For H_2O samples, severe radiation damping precluded data collection from published sequences containing the 1–$\bar{1}$ echo read pulse train (Sklenar & Bax, 1987). We found that these effects could be substantially reduced (especially for short mixing times) by modifying the phase cycling of the read pulses such that the H_2O magnetization is aligned along the $+z$ axis during the acquisition period for all scans (Blake & Summers, 1990). Optimum sensitivity was obtained by employing weak water presaturation during the recycle delay period and by incorporating a 2.5-ms homospoil pulse (followed by a 2.5-ms recovery delay period) subsequent to the beginning of the read pulse train. The homospoil pulse also served to reduce multiple quantum artifacts.

Data for quantitative interproton distance analysis were obtained from both D_2O and 90% H_2O solutions with mixing times of 0.005, 0.05, 0.1, 0.3, and 0.5 s. Read delay periods (τ_1 and τ_2) of 90 and 220 μs, respectively, were used, with a total recycle time of 2.5 s (=2.5 times the longest proton T_1 value). All NOESY data were processed with 6-Hz exponential line broadening in t_2, with 90°-shifted squared sine bell filtering in t_1, and with third-order polynomial base-line correction in the F_1 domain subsequent to the final Fourier transform. Other parameters for D_2O solutions: 2 × 512 × 2048 raw data matrix sizes; 32 scans per t_1 value. For aqueous solutions: 2 × 450 × 2048 data matrix sizes; 128 scans per t_1 value.

Phase-Sensitive 2QF-COSY. Solvent preirradiation was used to obtain 2QF-COSY data with 90% H_2O solutions. A modified version of the standard pulse sequence (Rance et al.,

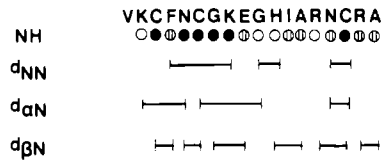


FIGURE 1: Scheme showing sequential NOESY connectivities used to assist in the assignment of the ^{1}H spectrum of Zn(p55F1). Filled, shaded, and open circles indicate low, moderate, and high sensitivity, respectively, of the amide proton signal intensity to solvent preirradiation. d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ indicate NOESY connectivities from backbone amide protons to neighboring amide protons, α-protons, and β-protons, respectively.

1983) was employed, wherein the relaxation delay period was preceded by a HS–90°(^{1}H)–HS– pulse train (HS = 10-ms homospoil pulse followed by a 20-ms delay period) to eliminate artifacts due to incomplete ^{1}H relaxation; 2 × 512 × 2048 data matrix sizes were collected and processed with 4-Hz Gaussian filtering in the t_2 dimension and 8-Hz Gaussian plus trapezoidal filtering in the t_1 dimension.

^{1}H–^{13}C HMQC. Phase-sensitive ^{1}H–^{13}C HMQC (Bax & Subramanian, 1986) spectra were obtained with broad-band ^{13}C decoupling (16 W) during the 200-ms acquisition period. Other parameters were as follows: 2 × 256 × 1024 data matrix size; 32 scans (preceded by two dummy scans) per t_1 value; pulse delay = 1.2 s; nondecoupling ^{13}C pulse widths = 33 μs; defocusing/refocusing delay periods = 3.5 ms; "Weft" delay period = 0.5 s; 6-Hz Gaussian and 90°-shifted squared sine bell filtering in the t_2 and t_1 dimensions, respectively.

NMR Signal Assignments

To facilitate the comparisons with fingerlike domains from other retroviral *gag* proteins, an internal numbering scheme is utilized here wherein the first cysteine in the zinc domain is designated as C(1). ^{1}H NMR signal assignments were made by first determining scalar connectivities within amino acid residues by COSY and HOHAHA spectroscopies and then correlating the signals of adjacent residues by 2D NOESY and ROESY NMR spectroscopies. Our approach differs only slightly from published procedures [Wuthrich (1986) and references sited therein] in that we employed ^{1}H–^{13}C HMQC spectroscopy to assist in the assignment of overlapping γ, δ, and ϵ side-chain proton signals. The complete set of COSY, HOHAHA, and ^{1}H–^{13}C HMQC spectra used to make intraresidue scalar connectivities are included in the supplementary material.

Sequential interresidue connectivities were made on the basis of observed NH(i)–NH(i+1) and/or CHα(i)–NH(i+1) Overhauser effect cross peaks. Two stretches of NH–NH connectivities were observed: F2-N3-C4-G5-K6 and N13-C14 (supplementary material). The most intense NH–NH connectivities were observed for the N3-C4, C4-G5, and N13-C14 pairs. Intense signals were also observed in this region for the side-chain amide protons of N3 and N13. Connectivities used to make sequential signal assignments are summarized in Figure 1, and ^{1}H and ^{13}C NMR chemical shift assignments are summarized in Tables I and II, respectively.

Portions of the 2D ROESY and NOESY spectra that contain amide to α and to side-chain protons are shown in Figure 2. Both spectra are shown to illustrate three points. First, since the ROESY data were obtained without water presaturation, connectivities from the amide protons that are moderately labile [e.g., A(11), N(13), and A(+2)] are observed in the ROESY spectrum but are either very weak or are completely unobserved in the NOESY spectrum. Second, because the read pulses in the ROESY spectrum (40 μs) were nearly 4 times longer than the read pulses in the NOESY spectrum (12 μs), the effective excitation profiles resulting from the 1–$\bar{1}$ echo pulses were significantly different. As a result,

Table I: ^{1}H NMR Chemical Shift Assignments for Zn(p55F1)[a]

residue	NH	CαH	CβH	others
V(−2)	*b*	3.69	2.01	γCH$_3$ 0.82, 0.88
K(−1)	8.17	4.41	1.43	γCH$_2$ 1.20
			1.50	δCH$_2$ 1.50
				ϵCH$_2$ 2.51
C(1)	8.27	4.10	1.78	
			2.79	
F(2)	8.67	4.50	3.15	H$^{2-6}$ 7.32–7.38
N(3)	9.58	4.75	2.68	amide 7.07, 8.21
			3.13	
C(4)	8.83	4.88	2.53	
			3.20	
G(5)	7.93	3.77		
		4.09		
K(6)	8.17	4.37	1.80	γCH$_2$ 1.47, 1.62
			1.89	δCH$_2$ 1.76
				ϵCH$_2$ 3.01
E(7)	8.38	4.50	1.87	γCH$_2$ 2.25, 2.28
			2.04	
G(8)	8.43	3.72		
		4.40		
H(9)	8.30	4.74	3.15	H^2 7.42
				H^4 6.87
I(10)	7.84	4.69	2.10	γCH$_2$ 1.21
				γCH$_3$ 0.87
				δCH$_3$ 0.90
A(11)	8.72	4.75	1.40	
R(12)	8.22	4.12	1.60	γCH$_2$ 1.60
			1.90	δCH$_2$ 3.15
N(13)	8.01	5.05	2.59	amide 7.00, 7.91
			3.11	
C(14)	7.65	3.77	3.02	
			3.33	
R(+1)	8.38	4.55	1.56	γCH$_2$ 1.56
			1.99	δCH$_2$ 3.19
A(+2)	8.39	4.07	1.32	

[a] Chemical shifts in ppm relative to internal H_2O (4.725 ppm). [b] Signal for terminal NH_3^+ not observed.

Table II: ^{13}C NMR Chemical Shift Assignments for Zn(p55F1)[a]

residue	Cα	Cβ	others
V(−2)	60.8	32.8	γCH$_3$ 18.9, 20.1
K(−1)	55.2	35.0	γCH$_2$ 24.8
			δCH$_2$ 29.0
			ϵCH$_2$ 41.7
C(1)	59.4	30.2	
F(2)	59.0	38.8	
N(3)	56.2	39.8	
C(4)	58.6	32.1	
G(5)	46.0		
K(6)	56.0	33.9	γCH$_2$ *b*
			δCH$_2$ 28.9
			ϵCH$_2$ 42.1
E(7)	56.0	30.9	γCH$_2$ 36.4
G(8)	45.8		
H(9)	56.0	30.1	
I(10)	59.2	40.9	γCH$_2$ 26.0
			γCH$_3$ 18.2
			δCH$_3$ 13.8
A(11)	55.1	18.7	
R(12)	58.2	30.0	γCH$_2$ *b*
			δCH$_2$ 43.2
N(13)	51.7	39.8	
C(14)	62.2	30.5	
R(+1)	55.1	30.4	γCH$_2$ *b*
			δCH$_2$ 43.1
A(+2)	54.2	20.0	

[a] Chemical shifts in ppm relative to external TMS. Sample conditions: 15 mM in Zn(p55F1); T = 25 °C; pH 7.0. [b] Unambiguous ^{13}C assignments precluded by signal overlap in the ^{1}H dimension of the ^{1}H–^{13}C HMQC spectrum.

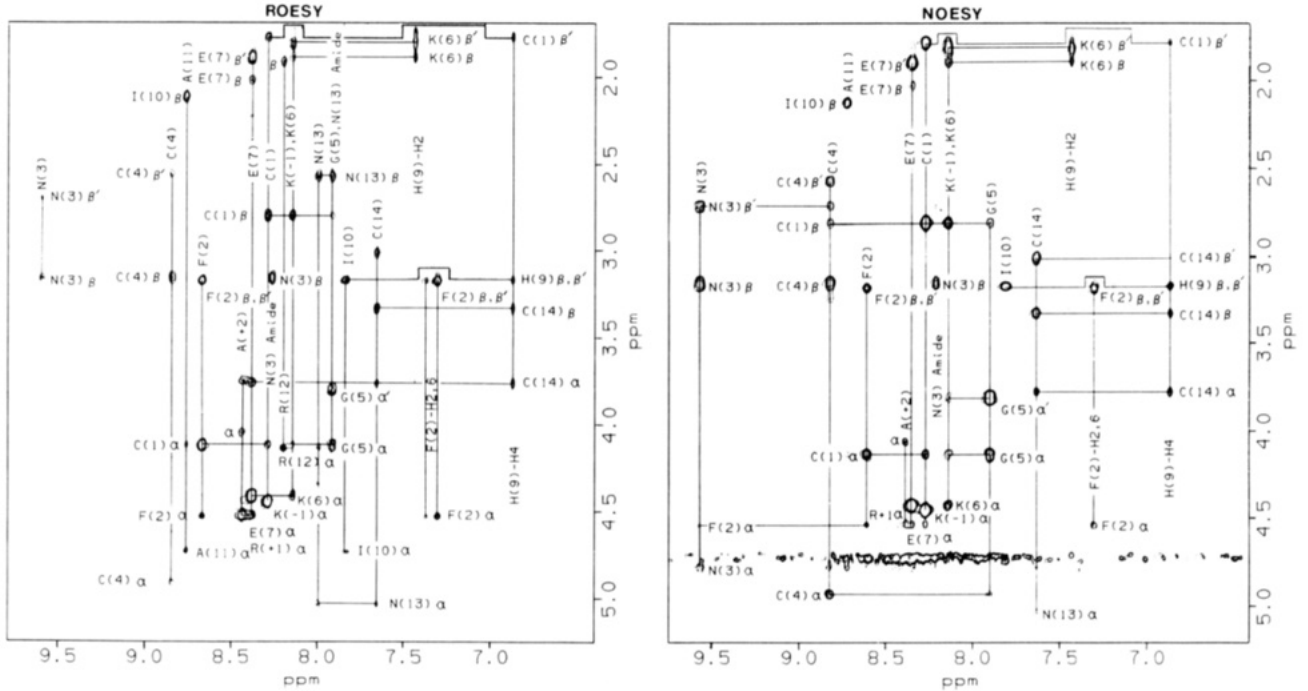
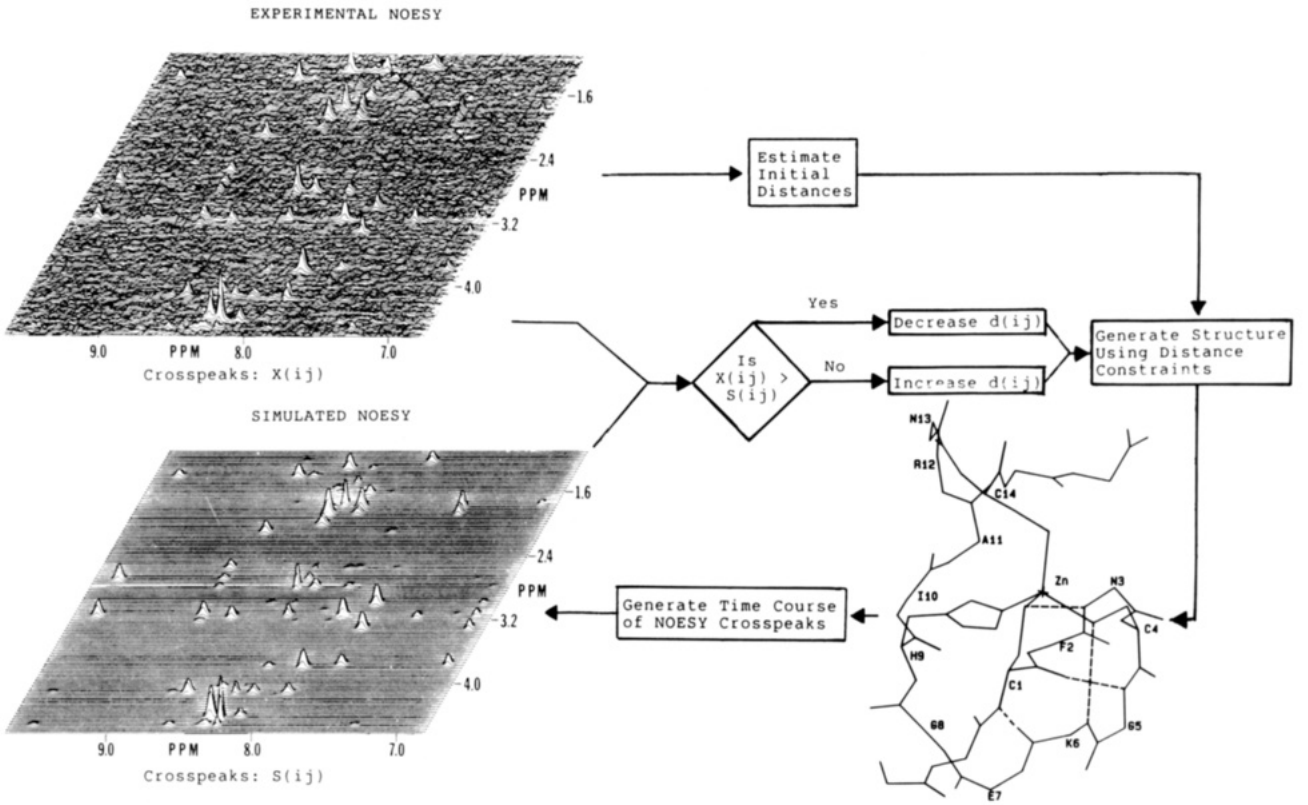


FIGURE 2: Portions of the ROESY (left, 80-ms mixing period) and NOESY (right, 300-ms mixing period) spectra obtained for Zn(p55F1) showing Overhauser effect connectivities from signals of amide and aromatic protons to α- and side-chain protons. Differences in signal intensities associated with the N(3) amide proton are due to the different excitation profiles employed (see Experimental Procedures). Comparisons of the ROESY and NOESY spectra were used to identify spin-diffusion signals in the NOESY data.

Scheme I



cross peaks in the data columns associated with the backbone amide proton of N(3) are weaker in the ROESY spectrum compared to the NOESY spectrum. Third, signals in the NOESY spectrum that result from spin-diffusion effects are readily identified by their absence (inverted phase) in the ROESY spectrum. For example, spin-diffusion peaks from the C(4) amide proton to the N(3)β' and C(1)β protons are clearly visible in the NOESY spectrum shown in Figure 2.

Structure Determination: Overview of the Approach

The general approach is shown in Scheme I. First, cross-peak intensities observed in a NOESY spectrum obtained with a short mixing time (to minimize relay effects) are

measured and used to make rough estimates of interproton distance constraints. Initially, lower constraints of 2.0 Å are used, and upper constraints of ca. 2.5, 3.5, and 4.5 Å are employed to reflect qualitatively observed strong, medium, and weak cross-peak intensities, respectively. A number of DG structures are then generated with the DSPACE distance geometry program using random initial atomic coordinates (see below), covalency constraints, and the experimental interproton distance constraints as input. To this point, our approach is similar to that utilized by others [e.g., see Wuthrich (1989) and references cited therein and Folkers et al. (1989)].

Structures generated with DSPACE are then used as input for the programs BKCALC and GNOE (see below), which together produce theoretical 2D NOESY spectra for each calculated structure. The initial calculated NOESY spectra typically contain cross peaks not observed in the experimental spectra, and minimum internuclear distance constraints are added for the appropriate proton pairs. This modified bounds file is used to generate new DG structures, and the cycle is repeated iteratively until the experimental and theoretical 2D NOESY spectra are visually identical. Further structure refinements are then made by adjusting *and tightening* the interproton distance constraints to obtain a match between the experimental and theoretical NOE buildup curves and auto peak decay curves.

After all of the initial DG structures give calculated NOESY spectra consistent with the experimental spectra, additional structures are generated from the prepared bounds file and randomly generated initial atomic coordinates. The back-calculation procedure is employed for each structure generated, and additional modifications to the experimental distance constraints are made as necessary in order that the experimental and calculated NOESY spectra for each structure match. All structures are re-refined subsequent to any modifications made to the bounds file. In this manner, final structures generated with the same boundary constraints will all produce calculated spectra that are consistent (within the experimental signal to noise) with the experimental spectra.

Distance Geometry Calculations (DSPACE). Bond lengths, bond angle lengths, and chirality dictated by the covalent primary structure, along with experimental distance boundaries determined from the NOE data, are used to create a matrix of upper and lower interproton distance bounds. Additional constraints (i.e., enforcing the triangle inequality for all triplets of atoms) are imposed. Trial distances generated by selecting random distances between the upper and lower bounds for each element are embedded in 3-space with the *metric* matrix method (Crippen, 1977; Havel et al., 1979). Initial coordinates of the resulting structure, which contain violations of the upper or lower bounds, are then refined to minimize deviation (penalty = squared sum of all bounds violations) with the entire bounds matrix. Complete refinement is achieved with two additional algorithms, one that adds a random vector of user-specified magnitude (0.2–0.3 Å) to the coordinates of each atom in order to randomize the structure followed by a simulated annealing refinement algorithm.

Simulated annealing is achieved by integration of Newton's equations of motion. Force fields for all constraints consist of simple Hook's law spring potentials. Bond distances, bond angle distances, nonbonded constraints, and experimental constraints can be weighted differently but are usually weighted equally. The algorithm initially chooses random velocity vectors for each atom and then integrates velocities into new atomic positions. Accelerations from the force field are added to the velocities as the integration proceeds. After a defined amount of motion has taken place, a new penalty list is generated. Annealing may be performed in three or four dimensions. Further refinement is generally achieved by performing conjugate gradient refinement.

Additional structures are subsequently generated by performing either random or difference embeds, which generates new atomic coordinates with interatomic distances that are either randomly chosen or chosen to be as different as possible from the distances in the current (refined) DG structure. The above refinement/simulated annealing algorithms are then repeated to minimize the constraint violations for the new structures.

2D NOESY Back-Calculations (BKCALC). Nuclear magnetic relaxation of protons in polymers of molecular mass greater than ca. 1000 daltons at fields above ca. 9 T ($\omega^2\tau_c^2 \approx 1$) is dominated by dipolar cross relaxation. For a molecule undergoing isotropic rotation, the rate at which z magnetization at spin i is transferred to spin j is given by

$$R_{ij} = \gamma^4 h^2 \tau_c / 10 r_{ij}^6$$

where τ_c is the proton–proton correlation time and r_{ij} is the interproton distance. In addition to cross relaxation, factors including chemical exchange and multiple quantum effects may also contribute to the total relaxation matrix of the spin system. In proteins, chemical exchange rates of amide protons may vary significantly, depending on their involvement in hydrogen bonding and exposure to solvent. The BKCALC algorithm employs what is basically an empirical two-parameter fitting of the spectral density, one parameter (k_{CR}) which is used to fit the cross-relaxation rate and the other (k_{ZL}) which is used to fit the "z leakage" rate. The calculation is actually slightly more complex since individual proton z leakage rate constants can be defined for exchanging protons. The z leakage rate constants quantitate loss of z magnetization during the mixing period due to all processes other than cross relaxation, including chemical exchange with H_2O.

The BKCALC algorithm calculates the time course of each proton by numerical integration of the Bloch equations. A cutoff distance is assigned, which limits the calculations to protons separated by 5–6 Å. For each pair of protons within the cutoff distance, NOE time course calculations are performed that include the contributions of all other protons that are within the cutoff distance from either proton in the pair. z magnetization is initially set to $-1/2$ for one proton of the pair, and to zero for all other protons (equals the average state of the spins at the beginning of the NOESY mixing period). The exchange of z magnetization between spins in the set is then calculated iteratively as a function of mixing time, on the basis of the interproton distances and the differences in z magnetization between each proton pair. A small time increment (0.001 s) is used to achieve good integration, and the iterative nature of the calculations accurately accounts for spin diffusion. This method generates results similar to diagonalization of the rate matrix (Redfield, 1957) but is computationally faster and allows inclusion of terms that account for nonuniform behavior (i.e., different z leakage rates).

In practice, a unique z leakage rate constant is used initially, which is determined by measuring the decay of total spectral magnetization with increasing mixing time. Using this generic z leakage rate and an initial DG structure generated with rough distance constraints, an estimate of the cross-relaxation rate constant is made by iterative fitting of auto-peak and cross-peak volumes as a function of mixing time for proton pairs with known internuclear separations (e.g., geminal protons). Individual z leakage rate constants are then assigned for each proton and are independently adjusted to give the best

fitting of the experimental auto-peak decay and cross-peak buildup curves.

The output file generated by BKCALC contains a list of normalized cross- and auto-peak intensities for the selected mixing time values. Plots of these data provide the theoretical buildup curves used for comparison with the experimental curves. Alternatively, this file may be used as input to the GNOE (Generate a NOESY spectrum) program, which generates data points for a complete two-dimensional spectrum to compare directly with the experimental spectrum. With this approach, peak volumes may be measured for the experimental and calculated spectra with identical procedures.

Stereospecific Assignments. Initially, the upfield and downfield signals of geminal protons are arbitrarily assigned to the *pro-S* and *pro-R* protons, respectively, and an option in DSPACE is employed which allows the stereospecific assignments for geminal protons to "float" in order to achieve the lowest penalty values. After low-penalty DG structures have been generated that exhibit simulated NOESY buildup curves in agreement with the experimental curves, the stereospecific assignments are established by sequentially turning off the floating option and enforcing correct geminal chiralities. Two conditions may result. If the initial stereospecific assignments were correct, then turning off the floating assignment option has no effect on the penalty or back-calculated spectra. If DSPACE had reversed the assignments in order to achieve the minimum penalty, then turning off the floating option results in a significant increase in the penalty. In these instances, the proton assignments are reversed in the BKCALC assignment file and in the interproton distance constraints file. It is important to note that, by this procedure, stereospecific assignments are obtained as a byproduct of the structure determination and are not used a priori to establish the structure (Weber et al., 1988).

Determination of the Zn(p55F1) Structure

Treatment of Overhauser Effect Data. Although ROESY spectroscopy is potentially very attractive for obtaining quantitative Overhauser effect information, Hartmann–Hahn effects may significantly affect cross-peak intensities for protons that are strongly coupled to each other and have similar chemical shifts (Bax et al., 1986). In addition, the nonlinear spin-lock field strengths must also be taken into account when used in quantitative analyses (Bax, 1988). For these reasons, NOESY spectra were used here for quantitative distance calculations. Since it was not possible to identify all spin-diffusion peaks on the basis of the shape of their buildup curves due to poor signal to noise obtained at the lower mixing times, spin-diffusion signals were identified by ROESY spectra.

An essential prerequisite for quantitative analysis of 2D NOESY spectra is the generation of flat spectral base lines. Flat base lines in the F_2 domain were obtained by adjusting the delay period between the final read pulse and the acquisition period, resulting in spectra that require no first-order phase corrections. However, phasing in the F_1 domain required the application of first-order phase corrections (up to 35°), which resulted in base-line curvature. The F_1 curvature was largely (but not completely) eliminated with a third-order polynomial base-line correction algorithm applied in the F_1 domain.

Volumes of all resolved and partially overlapping cross and auto peaks were measured for the ten 2D NOESY spectra obtained in D_2O and 90% H_2O solutions. Volumes for well-resolved diagonal peaks in the 5-ms (mixing time) NOESY spectra were used to normalize the volume data such that auto-peak volumes of the nonexchangeable protons obtained from the 5-ms NOESY spectra had values of 1.0. Normalized auto- and cross-peak volumes were then plotted as a function of increasing mixing time. Examples of the normalized volumes associated with the C(4) residue are shown in Figure 3. The complete set of cross- and auto-peak volume curves is given in the supplementary material.

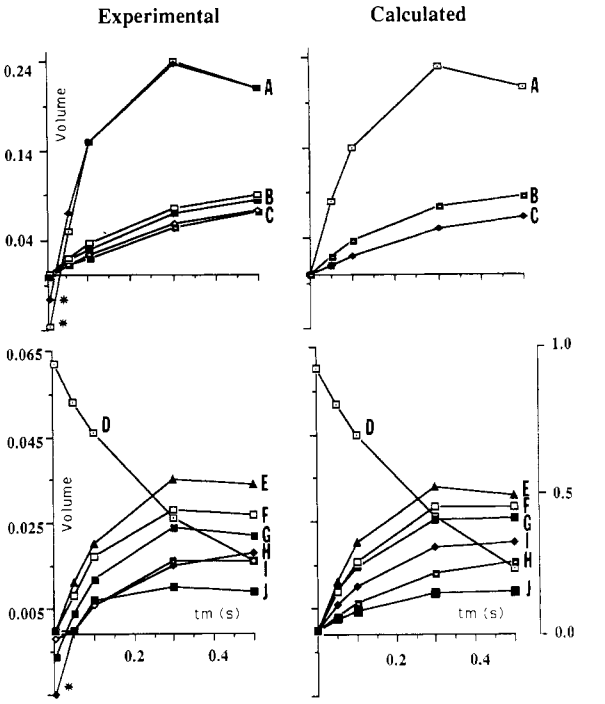


FIGURE 3: Experimental and calculated NOESY buildup curves for proton signals associated with Cys(4): (A) HβR(4):HβS(4); (B) Hα(4):HβS(4); (C) Hα(4):HβR(4); (D) HN(4) auto-peak decay; (E) HN(4):HN(5); (F) HN(4):HβR(3); (G) HN(4):HN(3); (H) HN(4):Hα(4); (I) HN(4):HβR(4); (J) HN(4):HβR(1). Data influenced by multiple quantum artifacts are indicated with an asterisk (*). The vertical scale on the right is for curve D.

Generation of Initial Structures. As described above, an initial bounds file was prepared from a qualitative assessment of the 100-ms NOESY spectrum. Zn–S and Zn–N bond distances of 2.3 and 2.1 Å, respectively, were included as constraints, but no bond angle constraints involving zinc were employed. Approximately one out of five trial structures refined to low-penalty values (less than 0.5 Å^2); the remaining structures exhibited penalties between 2 and 18 Å^2. The high-penalty structures could be further refined to low-penalty structures by first employing extensive annealing or by reversing the handedness of the folding followed by additional annealing. Initial low-penalty structures generally exhibited nontetrahedral (in some cases, square-planar) zinc coordination geometries.

2D NOESY Back-Calculations. An initial estimate of the generic z leakage rate (0.8 s^{-1}) was determined by fully integrating the 2D NOESY spectra obtained in D_2O solutions and measuring the rate of loss of total spectral magnetization as a function of the increasing mixing time. Back-calculations were then performed repeatedly for a representative low-penalty DG structure with a single z leakage rate constant (0.8 s^{-1}) and estimates of k_{CR}. A final value of k_{CR} (−60 s^{-1}) afforded accurate simulation of the geminal and Asn side-chain amide proton buildup curves for spectra obtained in both aqueous and D_2O solutions (see Figure 3 and the supplementary material; note, the value for k_{CR} may be scaled differently in other versions of BKCALC). For D_2O data, N–H and other exchangeable protons are not included in the

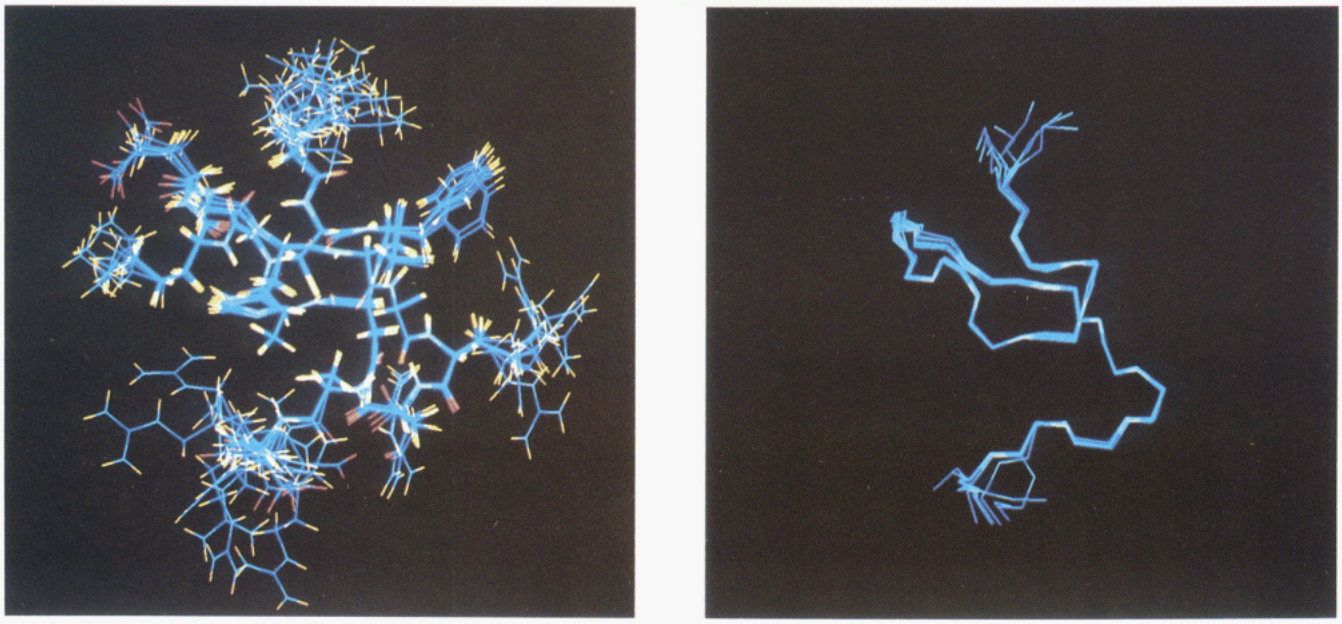


FIGURE 4: Superposition of DG structures 1–8: (left) all atoms; (right) backbone atoms only.

back-calculations. Final values of k_{ZL} (supplementary material) were obtained by adjusting values for individual protons to get the best fits for the auto-peak decay and cross-peak buildup curves (Figure 3 and supplementary material).

Qualitative Fitting of the 100-ms NOESY Spectrum. Each back-calculated 100-ms NOESY spectrum was compared with the experimental 100-ms NOESY spectrum. For comparisons with data obtained with $1\text{-}\bar{1}$ echo read pulses, data rows in the calculated NOESY spectra were scaled with the relationship

$$f(v) = \sin \omega\tau_1(1 - \cos \omega\tau_2)/2$$

($\tau_1 = 114$ μs, $\tau_2 = 220$ μs) in order to simulate the excitation profile of the $1\text{-}\bar{1}$ echo read pulse (allowing for 90- and 220-μs read delays and 12-μs pulse widths). Back-calculated spectra obtained for high-penalty (>2 Å) structures differed significantly in appearance from the experimental spectra, and these structures were subjected to additional annealing/refinement to reduce the penalty as described above. For the low-penalty structures, initial back-calculated spectra contained a number of cross peaks that were not observed in the experimental spectra, and lower distance constraints (which allowed for a minimal interproton distance of 4.0 Å) were added to the experimental constraints file for appropriate proton pairs. The modified experimental constraints file was then used with DSPACE to generate a new set of structures, and the back-calculated spectrum for each new structure was again compared with the experimental spectrum. This cycle was repeated iteratively until the experimental and calculated 100-ms NOESY spectra for all DG structures exhibited similar cross-peak patterns.

Structure Refinement—Quantitative Fitting of the NOESY Data. Once similar looking back-calculated and experimental NOESY spectra were obtained, experimental and theoretical buildup curves were compared. Compensations for differences were made by making appropriate adjustments to the interproton distance constraints. Additional annealing and refinement was performed to minimize the bounds violations, and the buildup curves for the refined structures were compared with the experimental buildup curves. This procedure was repeated iteratively until the experimental and theoretical curves were visually similar. Since the NOE buildup is very sensitive to interproton distance, especially for short (2–3 Å) distances, tight distance constraints as low as ±0.05 Å were required to obtain the best fits.

Two types of artifacts occasionally obtained in the NOESY data had to be accounted for when the buildup curves were being fit. First, slight deviations of the base line from true zero afforded non-zero volumes for data obtained with the 5-ms mixing period. All of the 2D NOESY data were processed identically, and the regions of the spectra that contained these deviations were observed consistently in the 5–500-ms NOESY spectra. Minor corrections for these artifacts were made by shifting the experimental curves such that the NOE intensities at 5-ms mixing times were zero. Note, however, that all of the buildup curves shown in Figure 3 (and the supplementary material) are presented as *raw, noncorrected data* to illustrate the fact that the base-line offsets, where observed, were very small. An example of this type of artifact is illustrated by curve G in Figure 3, where the HN(4):HN(3) buildup curve is negatively shifted due to negative base-line distortions in this region of the NOESY spectra.

A second type of artifact results from incomplete suppression of zero quantum signals at mixing times of 100 ms or less. In most cases, small zero quantum signals observed in the 5-ms NOESY spectra gave peak volumes close to zero due to the antiphase nature of the multiplets. Note that, as previously reported (Macura et al., 1982), these artifacts decay rapidly with increasing NOESY mixing time and, except for geminal Gly protons, were completely undetected in our spectra at mixing times of 100 ms or greater. However, a few multiple quantum signals gave rise to significant non-zero cross-peak volumes at the lower mixing times as evidenced by the negative volumes (5 ms) observed for the HBR(4)–HBS(4) and HN(4)–HA(4) cross peaks shown in Figure 3 (curves A and H, respectively). These artifacts were readily identified and distinguished from base-line distortion artifacts by their antiphase peak patterns and, of course, were only observed for scalar-coupled protons. Best fits were made for these curves by ignoring the 5- and 50-ms data points and fitting only the 100-, 300-, and 500-ms points. The geminal Gly protons exhibited multiple quantum artifacts in the 100-ms NOESY data as well, and fits for these curves were made by using only the 300- and 500-ms data.

Generation of Final DG Structures 1–8. The difference-embed method, followed by simulated annealing and refining,

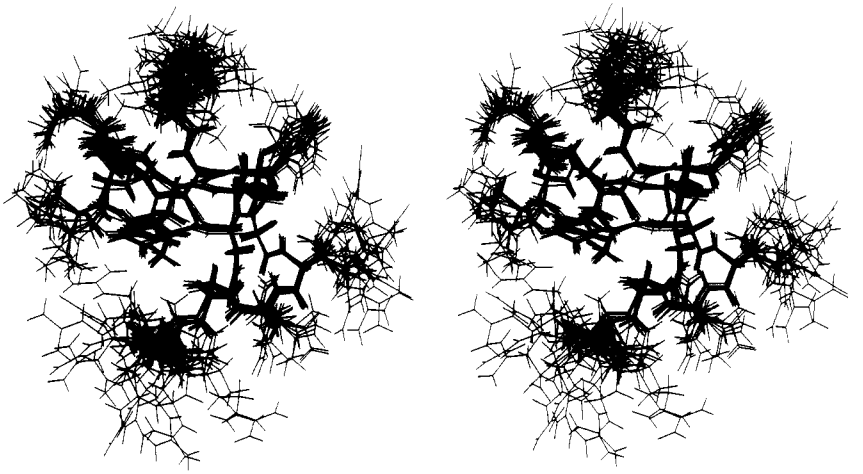


FIGURE 5: Superposition of DG structures 1–16 showing the high similarity between structures generated with (DG 1–8) and without (DG 9–16) zinc covalency constraints.

was used to generate seven additional low-penalty DG structures. Some of the new structures gave BKCALC spectra that contained cross peaks absent in the experimental spectra. Appropriate lower distance constraints were added, and all of the DG structures were subjected to further refinement using the modified constraints file. All of the final DG structures (1–8) were generated by using the final interproton distance constraints listed in Table III. *No hydrogen bonding or ligand–Zn–ligand bond angle constraints were utilized at any point in the refinement process.* Each of the final structures refined to a total penalty of less than 0.21 Å^2, by use of 226 interproton distance constraints with bounds for the 72 observable distances typically in the range of ±0.05 Å (Table III and supplementary material). The largest single bounds violation was 0.078 Å^2. Anomalous data were obtained only for the C-terminal Ala, which gave intraresidue NOE buildup curves that were all very strong and could not be fit with any combination of z leakage or cross-relaxation rate constants. In this case, the NOESY data reflect rapid exchange between multiple internal conformations for the C-terminal Ala. It is important to note that BKCALC *fails* to accurately simulate experimental NOESY data obtained for systems in rapid conformational equilibria, regardless of the choices of z leakage and cross-relaxation rate parameters, and this provides evidence for multiple local conformations.

Except for signals associated with the C-terminal Ala, *back-calculated NOESY data were consistent with the experimental NOESY data for all of the final DG structures.* In fact, the buildup of spin-diffusion peaks also matched, even though constraints to try to fit these signals were not included in the experimental constraints file. In both the experimental and back-calculated data, the buildup of spin-diffusion peaks was about half as fast as the buildup of direct NOE peaks for NOEs involving geminal protons. Calculated NOE buildup profiles associated with Cys(4) of DG structure 1 are shown in Figure 3. The complete set of back-calculated NOE data for DG structure 1 is given in the supplementary material. Except for the C-terminal Ala, all residues gave self-consistent inter- and intraresidue NOE buildup profiles, and no inconsistent NOEs were observed between three or more protons that would be indicative of dynamic exchange between multiple structures.

Generation of DG Structures without Zinc Constraints. Eight additional structures (DG 9–16) were generated without the zinc constraints in order to assess the contribution of these constraints in determining the peptide folding. The experimental interproton distance constraints used to generate DG 9–16 were identical with the constraints used to generate DG 1–8. DG structures 9–16 refined to total penalties of 0.23 Å^2 or less (largest penalty <0.09 Å^2), and all gave back-calculated NOESY data that agreed with the experimental data.

RESULTS

Degree of Refinement. For DG structures 1–8, pairwise RMS deviations for backbone atoms of residues C(1)–C(14), for all backbone atoms, and for all atoms are in the ranges 0.07–0.26, 0.20–0.83, and 1.07–1.90 Å, respectively (supplementary material). Superpositioned DG structures 1–8 are shown in Figure 4, illustrating the high degree of refinement obtained for the C(1)–C(14) backbone atoms of Zn(p55F1) and the significantly reduced (or lacking) refinement for the C- and N-terminal backbone atoms and for certain side-chain atoms. Poorer refinement resulted from insufficient (or lacking) NOE constraints due to either proton chemical exchange (amides) or severe signal overlap (protons on longer side chains).

DG structures 9–16 were superpositioned onto DG structures 1–8, with pairwise RMSDs for main-chain atoms of residues C(1)–C(14) in the range 0.20–0.39 Å (Figure 5 and supplementary material). The folding exhibited by DG 1–8 and DG 9–16 was virtually identical, the only visible differences being a slight twist of the His aromatic ring and a minor change in the relative positions of the type II tight turn (described below). In light of the extremely small differences between these structures, it is clear that the *global folding and high degree of refinement obtained for DG structures 1–8 is the result of the NOE-derived distance constraints utilized and is not due to imposed covalency constraints to zinc.*

Description of the Structure. Residues C(1)–K(6) fold in a manner nearly identical with the folding observed via X-ray crystallography for related residues in the iron-binding domain of rubredoxin (Adman et al., 1975; Watenpaugh et al., 1980). Superposition of all backbone and Cys side-chain atoms of residues C(1)–K(6) of Zn(p55F1) (DG structure 1) onto the respective atoms of residues C(6)–Y(11) and residues C(39)–V(44) of rubredoxin gives RMSDs of 0.46 and 0.35 Å, respectively. The N(3) and C(4) backbone amide protons of Zn(p55F1) are oriented in a manner consistent with hydrogen bonding to the C(1) sulfur in what has been described as a type I NH–S tight turn, Figure 6 (Adman et al., 1975). The N–H^E side-chain amide proton of N(3) appears to be hydrogen

Table III: Interproton Distance Constraints Used To Generate DG Structures[a]

α–α
ha[3] ha[12] 2.75 2.85
ha[2] ha[9] 2.75 3.15
ha[1] har[10] 4.0 #

Amide–Amide
hn[3] hn[6] 4.0 #
hn[3] hn[8] 4.0 #
hn[4] hn[5] 2.95 3.10
hn[4] hn[13] 3.8 #
hn[5] hn[6] 2.40 2.45
hn[5] hn[7] 3.5 #
hn[6] hn[7] 2.40 2.45
hn[6] hn[8] 4.0 #
hn[7] hn[8] 2.8 2.9
hn[8] hn[9] 3.5 #
hn[9] hn[10] 2.5 #
hn[15] hn[16] 2.1 2.2

Amide–α
hn[3] ha[2] 2.13 2.17
hn[3] ha[3] 2.73 2.78
hn[4] ha[3] 2.00 2.05
hn[4] ha[4] 2.7 2.85
hn[4] ha[9] 4.0 #
hn[5] ha[3] 4.0 #
hn[5] ha[4] 3.2 3.52
hn[5] ha[5] 2.8 #
hn[6] ha[6] 2.75 2.90
hn[7] ha[6] 2.8 #
hn[7] har[7] 2.26 2.31
hn[7] has[7] 2.9 #
hn[8] ha[5] 4.0 #
hn[8] ha[8] 2.75 2.85
hn[8] har[7] 3.0 #
hn[8] has[7] 2.7 3.0
hn[9] ha[8] 2.00 2.05
hn[9] ha[9] 2.5 2.8
hn[12] ha[15] 4.0 #
hn[13] ha[3] 3.0 3.8
hn[16] ha[15] 2.8 3.1
hn[16] ha[16] 2.6 2.8

Amide–β
hn[3] hbr[3] 2.42 2.47
hn[3] hbs[3] 2.44 2.49
hn[4] hbr[3] 3.8 #
hn[4] hbs[3] 3.8 #
hn[4] hbr[4] 2.03 2.07
hn[5] hbr[3] 4.0 #
hn[5] hbs[3] 4.0 #
hn[5] hbr[5] 2.42 2.52
hn[5] hbs[5] 2.42 2.52
hn[5] hbs[15] 4.0 #
hn[5] hbr[6] 4.0 #
hn[6] hbr[3] 2.9 3.0
hn[6] hbs[3] 3.8 #
hn[6] hbr[5] 2.5 2.6
hn[6] hbs[5] 3.5 #
hn[6] hbr[6] 2.65 2.70
hn[6] hbs[6] 2.8 #
hn[7] hbr[3] 2.8 2.9
hn[7] hbs[3] 3.8 #

Amide–β
hn[7] hbs[8] 4.0 #
hn[7] hbs[16] 4.0 #
hn[7] hbr[16] 4.0 #
hn[8] hbr[3] 2.60 2.70
hn[8] hbr[6] 4.0 #
hn[8] hbr[8] 2.33 2.38
hn[8] hbs[8] 2.6 #
hn[9] hbr[8] 4.0 #
hn[9] hbs[9] 2.1 2.2
hn[9] hbr[9] 3.0 #
hn[9] hbr[14] 4.0 #
hn[12] hbr[11] 2.45 2.65
hn[12] hb[12] 2.8 #
hn[12] hbr[15] 4.0 #
hn[12] hbs[15] 4.0 #
hn[13] hb[12] 2.15 2.25
hn[13] hbs[6] 3.8 #
hn[13] hbr[6] 3.8 #
hn[13] hbs[15] 3.8 #
hn[13] hbs[16] 3.8 #
hn[15] hbs[16] 3.2 #
hn[16] hbs[5] 4.0 #
hn[16] hbr[11] 4.0 #
hn[16] hbs[16] 2.35 2.45
hn[16] hbr[16] 2.35 2.45
hn[18] hbr[16] 3.8 #
hn[18] hbr[17] 3.8 #

Amide–γ
hn[3] hgr[2] 4.0 #
hn[8] hgr[8] 2.75 3.00
hn[9] hgr[8] 4.0 #
hn[12] hglr[12] 2.15 2.2
hn[12] mg2[12] 3.8 #
hn[12] mdl[12] 3.8 #
hn[13] hgr[2] 4.0 #
hn[13] hglr[12] 3.8 #
hn[13] hgls[12] 3.8 #
hn[13] mg2[12] 3.8 #
hn[13] mdl[12] 3.8 #
hn[17] hgr[2] 4.0 #
hn[18] hbr[17] 4.0 #
hn[18] hbs[17] 4.0 #
hn[18] hgr[17] 4.0 #
hn[18] hgs[17] 4.0 #
hn[9] hgr[9] 3.8 #
hn[9] hgs[9] 3.8 #

α–Others
ha[1] hb[1] 2.35 2.55
ha[2] hbr[2] 2.23 2.6
ha[2] hbs[2] 2.23 2.6
ha[3] hbr[3] 2.7 #
ha[3] hbs[3] 2.30 2.35
ha[5] hbr[5] 2.8 #
ha[5] hbs[5] 2.35 2.40
ha[5] ha[6] 4.0 #
ha[6] hbs[6] 2.30 2.40
ha[6] hbr[6] 2.35 2.50
has[7] hbr[6] 4.0 #
ha[9] hbr[9] 2.2 2.6
ha[9] hbs[9] 2.2 2.6
har[10] hbr[2] 4.0 #

α–Others
har[10] hbs[2] 4.0 #
har[10] hgr[2] 4.0 #
har[10] hgs[2] 4.0 #
har[10] hdr[2] 4.0 #
har[10] hds[2] 4.0 #
har[10] her[2] 4.0 #
har[10] hes[2] 4.0 #
ha[11] hbr[6] 4.0 #
ha[12] hb[12] 2.40 2.48
ha[13] hb[12] 3.8 #
ha[13] hbr[16] 3.0 #
ha[13] hbs[16] 2.40 2.45
ha[14] hdl[4] 4.0 #
ha[14] hbs[5] 3.8 #
ha[14] hd2e[5] 4.0 #
ha[15] hb[12] 4.0 #
ha[15] hbs[16] 3.8 #
ha[15] mg2[12] 4.0 #
ha[15] hbr[15] 2.3 2.8
ha[15] hbs[15] 2.8 #
ha[16] hb[12] 3.8 #
ha[16] hbr[11] 4.0 #
ha[16] hbs[11] 4.0 #
ha[16] hbs[16] 2.35 2.40
ha[16] hbr[16] 2.8 #

Histidine Aromatics
hd2[11] has[7] 4.0 #
hd2[11] ha[9] 4.0 #
hd2[11] ha[15] 3.8 #
hd2[11] ha[16] 2.70 2.75
hd2[11] hbs[3] 3.2 3.3
hd2[11] hbs[6] 4.0 #
hd2[11] hbr[11] 2.65 2.80
hd2[11] hn[12] 3.9 #
hd2[11] hbs[15] 4.0 #
hd2[11] hbs[16] 2.70 2.75
hd2[11] hbr[16] 3.8 #
hd2[11] hglr[12] 3.8 #
hel[11] ha[5] 4.0 #
hel[11] har[7] 4.0 #
hel[11] has[7] 4.0 #
hel[11] ha[8] 4.0 #
hel[11] hbs[5] 4.0 #
hel[11] hbr[8] 3.1 3.6
hel[11] hbs[8] 2.7 2.9
hel[11] hn[8] 4.0 #
hel[11] hn[9] 4.5 #
hel[11] hbr[14] 4.0 #
hel[11] hgr[8] 4.0 #
hel[11] hgs[8] 4.0 #
hel[11] har[10] 4.0 #
hel[11] has[10] 4.0 #

Methyls
mb[13] hn[4] 3.30 3.90
mb[13] hn[5] 3.2 3.4
mb[13] hd2e[5] 3.8 #
mb[13] hbs[5] 3.55 3.60
mb[13] hbr[5] 4.0 #
mb[13] hn[6] 4.0 #
mb[13] hn[7] 4.0 #
mb[13] hn[13] 2.58 2.62

Methyls
mb[13] ha[14] 3.55 3.70
mb[13] hn[15] 4.0 #
mb[13] hn[16] 3.8 #
mb[13] hbr[4] 3.1 3.3
mb[13] hdl[4] 4.6 #
mb[18] hbr[6] 3.5 3.7
mb[18] hbs[6] 4.0 #
mb[18] hd2z[5] 4.2 #
mb[18] hd2e[5] 4.2 #
mgr[1] hbr[2] 3.8 #
mgr[1] hbs[2] 3.8 #
mgs[1] hbr[2] 3.8 #
mgs[1] hbs[2] 3.8 #
mgr[1] hgr[2] 3.8 #
mgr[1] hgs[2] 3.8 #
mgs[1] hgr[2] 3.8 #
mgs[1] hgs[2] 3.8 #
mgr[1] har[10] 4.0 #
mgr[1] has[10] 4.0 #
mgs[1] har[10] 4.0 #
mgs[1] has[10] 4.0 #

β–β
hb[12] hbs[5] 4.0 #
hb[12] hbr[5] 3.8 #
hbr[3] hbr[6] 4.0 #
hbr[3] hbs[6] 4.0 #
hbs[3] hbr[6] 4.0 #
hbs[3] hbs[6] 4.0 #
hbr[6] hbr[8] 4.0 #
hbr[6] hbs[8] 4.0 #
hbs[16] hbs[15] 4.0 #
hbr[16] hbs[15] 4.0 #
hbs[15] hgr[14] 4.0 #
hbs[15] hgs[14] 4.0 #
hbs[15] hdr[14] 4.0 #
hbs[15] hds[14] 4.0 #
hbr[15] hgr[17] 4.0 #
hbr[15] hgs[17] 4.0 #
hbr[15] hdr[17] 4.0 #
hbr[15] hds[17] 4.0 #
hd2e[5] hbr[5] 2.28 2.33
hd2e[5] hbs[5] 3.0 #
hd2e[5] hn[6] 4.0 #
hd2z[5] hbr[5] 3.0 #
hd2z[5] hbs[5] 3.0 #

80-ms ROESY Spectrum[b]
hn[11] hbs[3] 2.0 3.0
hn[11] ha[11] 2.8 #
hn[15] ha[15] 2.6 #
hn[15] ha[14] 2.8 #
hn[15] hbs[14] 2.5 3.5
hn[15] hbr[14] 2.6 #
hn[15] hbr[15] 2.8 #
hn[15] hbs[15] 2.0 2.8
hn[15] hgls[12] 3.0 4.0
hn[10] ha[9] 2.0 2.2
hn[10] har[10] 2.6 #
hn[10] has[10] 2.6 #
hn[10] hn[11] 2.4 3.0

[a] In this table only, residues (in brackets) are numbered 1–18, beginning with the N-terminal Val. Numbers for pairs of protons represent lower and upper (# = infinite) distance constraints. Atom definitions: h = single proton; m = methyl; a, b, c, d = α, β, γ, and δ, respectively; r = *pro-R* and s = *pro-S* stereochemistry. [b] Estimates based on the 80-ms ROESY spectrum.

bonded to the C(14) sulfur, consistent with the low chemical exchange rate observed for this proton.

The C(4) α-carbon serves as a corner between the type I NH–S tight turn and a short, orthogonally directed type II NH–S tight turn, with hydrogen bonding between the amide of K(6) and the C(4) sulfur (Figure 6). The carbonyl oxygen of K(6) is pointing in a direction consistent with hydrogen bonding to the amide hydrogen of C(1), and the amide hydrogen of G(5) exhibits hydrogen bonding to the C(1) carbonyl oxygen. With this folding, the β-protons of K(6) are in the vicinity of the H(9) imidazole H^2 proton, and the K(−1) α-proton is in the vicinity of the E(7) α-proton. The folding implicitly directs the hydrophobic phenyl group of F(2) away from the body of the peptide and into the solvent.

The G(8) α-carbon provides a corner for a 3_{10} turn, with the *pro-R* α-H in close proximity to the E(7) carbonyl oxygen.

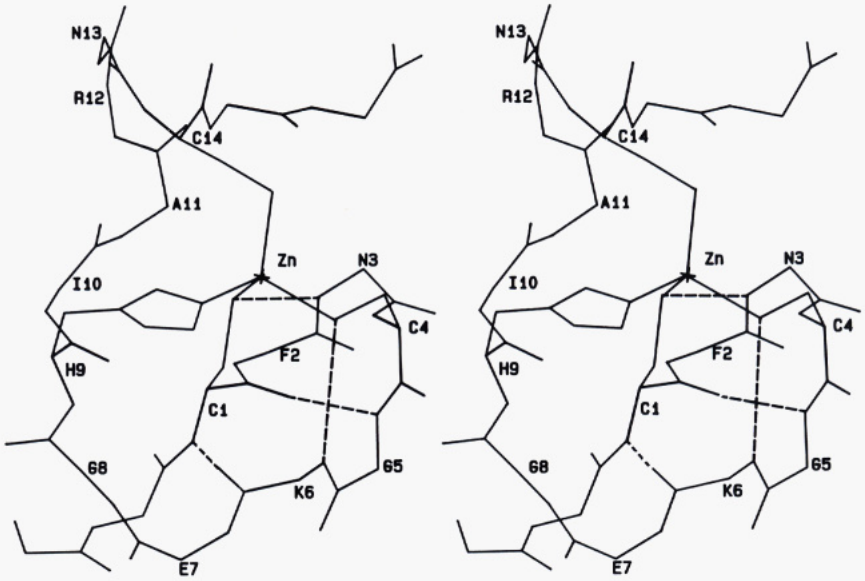


FIGURE 6: Stereo diagram of DG structure 1 showing all backbone atoms and the side-chain atoms (except protons) for residues C(1), C(4), H(9), and C(14). Dashed lines indicate hydrogen bonds.

Table IV: Ligand–Zn–Ligand Bond Angles (deg) in Zn(p55F1) DG Structures

structure	S1–Zn–S4	S1–Zn–S14	S4–Zn–S14	N–Zn–S1	N–Zn–S4	N–Zn–S14
1	106	111	114	97	115	112
2	105	112	116	96	113	112
3	106	114	117	96	113	109
4	106	109	114	97	117	112
5	106	106	113	97	117	116
6	106	107	114	96	115	117
7	106	110	115	96	116	112
8	106	110	112	97	117	112

The backbone atoms including the carbonyl of G(8) through the amide of I(10) form a β-like stretch. None of the backbone atoms within this stretch appear to be involved in hydrogen bonding. The imidazole Nϵ nitrogen of H(9) is coordinated to Zn, and the hydrophobic side chain of I(10) is directed away from the body of the peptide and into the solvent. As shown in Figure 7, the methyl groups of I(10) (treated as spheres) are relatively close to the hydrophobic side chain of F(2).

The A(11) α-carbon serves as a corner that leads to a type II tight turn. This tight turn is comprised of residues A(11)–C(14) and is stabilized by apparent hydrogen bonding between the A(11) carbonyl oxygen and the C(14) amide proton and by coordination of the C(14) sulfur to Zn. The methyl group of A(11) is located directly above the F(2) and N(3) backbone atoms, and the pseudoplane made by the backbone atoms of the type II tight turn is nearly perpendicular to the pseudoplane made by the backbone atoms of the type I NH–S tight turn (see Figure 7). With this folding, the positively charged side chain of R(12) and the amide of N(13) are positioned above the hydrophobic side chains of residues F(2) and I(10) (Figure 7).

The backbone amide protons of V(−2), K(−1), and R(+1) were not observed in the 2D NOESY spectra due to rapid exchange; as a result, the conformation of the N- and C-terminal residues of Zn(p55F1) were determined with less accuracy as indicated by the backbone deviation data (see above). Although not implicitly invoked in the calculations, all of the final structures exhibit tetrahedral or near-tetrahedral Zn coordination geometries, with ligand(1)–Zn–ligand(2) bond angles in the range 96° to 117° (Table IV).

Residues C(1)–K(6) fold in a manner predicted recently for a retroviral zinc finger (Green & Berg, 1989). However, the

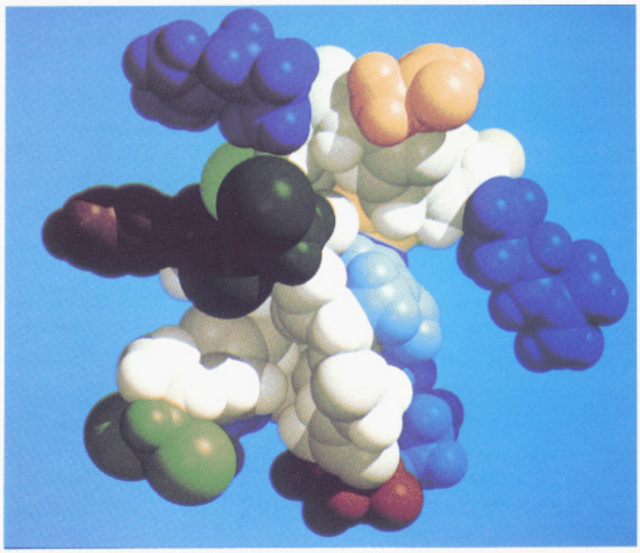


FIGURE 7: Space-filling representation of Zn(p55F1) DG structure 1 showing the spatial relationships between the side chains of the conservatively substituted amino acid residues that have been implicated in genomic RNA recognition (methyl groups displayed as spheres): I(10), dark green, center; F(2), brown, center left; A(11), light green, upper left; R(12), dark blue, upper left; N(13), orange, top. The Shapely model coloring scheme was employed (backbone atoms are white; side chains for basic, acidic, and hydrophobic residues are blue, red, and green, respectively.

folding observed here for residues E(7)–C(14) is significantly different from that in the model, where the His was predicted to bind Zn via the Nδ rather than the Nϵ nitrogen. We cannot rule out the possibility that Nδ binding could occur in retroviral

zinc fingers with amino acid sequences different from that of Zn(p55F1).

Discussion

Limitations and New Directions of the Approach. Two significant advances in the area of NMR-based structure determination are provided by this new approach. By obtaining back-calculated NOESY spectra for the DG structures, semiquantitative evaluation of the degree that the structures reflect the actual experimental data is possible. In addition, this approach allows the use of tighter distance constraints, thereby narrowing the "range" (and increasing the resolution) of structures generated with distance geometry. In the future, it should be possible to compare more quantitatively the experimental and calculated NOESY data, which will increase the requirements for obtaining NOESY data that lack zero-quantum artifacts and F_1 base-line roll. In this regard, new methods for reducing F_1 base-line roll at both the data collection (Davis, 1989) and processing (Marion & Bax, 1989) stages have been described.

In addition, pairwise RMSD values alone are not sufficient for evaluating the quality of the structure refinement. Clearly, tighter inter- and intraresidue distance constraints can significantly reduce conformational variations in the DG structures. In the future, error analyses that account for (1) the signal to noise in the experimental NOESY spectra and (2) the estimated errors associated with the $1/r^6$ relationship should provide a better indication of the degree of determination (resolution) of the final structure.

Biological Implications of the Zn(p55F1) Structure. Two of the most striking structural features exhibited by Zn(p55F1) are (1) the presence of at least seven internal hydrogen bonds within the 14-residue zinc-binding domain and (2) the similarity of the first six residues to related residues in the iron domain of rubredoxin. The extensive hydrogen bonding is consistent with the high thermal stability observed for Zn(p55F1) (South et al., 1989). Interestingly, Zn(p55F1) and rubredoxin contain a Gly residue immediately following the C-X_2-C sequences. This Gly (position 5 in *gag* proteins) is conservatively substituted in retroviral *gag* proteins. The presence of a sterically nondemanding Gly at this position appears to stabilize type II NH–S folding (Adman et al., 1975). Steric requirements associated with the 3_{10} turn at position 8 (Figure 7) may also explain the conservative substitution of Gly at this position in retroviral *gag* proteins. *Thus, the structural constraints associated with the zinc-binding motif provide a rationale for the conservatively substituted Gly residues at positions 5 and 8*, and this adds further evidence for a biologically relevant zinc fingerlike motif.

The backbone folding exhibited by Zn(p55F1) implicitly positions the side chains of the conservatively substituted F(2), I(10), R(12), and N(13) residues in the same general spatial location. All of these side chains are directed away from the body of the peptide and into the solvent. As a consequence, the aromatic ring of F(2) and the methyl groups of I(10) form a hydrophobic surface patch (see Figure 7). Adjacent to the hydrophobic patch is the positively charged, basic side chain of R(12) and the amide side chain of N(13). The orientation of these conservatively substituted side chains and preliminary nucleic acid binding experiments (South et al., 1990) are consistent with the proposal that these residues participate in gene recognition by binding to specific RNA structures (Gorelick et al., 1988; Meric & Goff, 1989).

The structure of Zn(p55F1) differs considerably from structures predicted (Berg, 1988) and found (Parraga et al., 1988; Lee et al., 1989) for the classical DNA-binding zinc finger motif. The short amino acid sequence between C(4) and H(9) in retroviral fingerlike domains precludes the formation of a nucleic acid binding finger structure. Instead, it appears that the functionally important nucleic acid binding regions of the retroviral motif are the four-residue tight turns, which may be best described as "knuckles" rather than fingers.

Acknowledgments

Technical assistance from Dr. O. Herzberg (CARB, Gaithersburg, MD), J. Suess (UMBC), and Dr. H. R. Summers and helpful discussions with Drs. R. L. Karpel (UMBC) and L. E. Henderson (Frederick Cancer Research Facility) are gratefully acknowledged. We thank Dr. J. Berg (Johns Hopkins) for providing a preprint of his paper, which stimulated the comparison of our structure with the rubredoxin structure.

Supplementary Material Available

Four tables containing the DSPACE experimental constraints (Table SI), BKCALC assignment file (Table SII), and pairwise RMS deviation data (Tables SIII and SIV) and 35 figures presenting a complete set of experimental and calculated (DG structure 1) NOESY cross-peak buildup (Figures S1–S23) and auto-peak decay (Figures S24–S31) curves and HOHAHA (Figure S32), 2QF-COSY (Figure S33), ^1H–^{13}C HMQC (Figure S34), and ROESY spectra (Figure S35) (42 pages). Ordering information is given on any current masthead page. Coordinates for DG structures 1–16 have been submitted to the Brookhaven Protein Data Bank.

Registry No. p55F1, 123963-60-0.

References

Adman, E., Watenpaugh, K. D., & Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4854.
Bax, A. (1988) *J. Magn. Reson. 77*, 134.
Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565.
Bax, A., Sklenar, V., & Summers, M. F. (1986) *J. Magn. Reson. 70*, 327.
Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511.
Berg, J. M. (1986) *Science 232*, 485.
Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 99.
Blake, P., & Summers, M. F. (1990) *J. Magn. Reson.* (in press).
Bolognesi, D. P., Montelaro, R. C., Frank, H., & Schafer, W. (1978) *Science 199*, 183.
Bothner-by, A., Stephens, R. L., Lee, J. T., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521.
Copeland, T. D., Morgan, M. A., & Oroszlan, S. (1984) *Virology 133*, 137.
Crippen, G. M. (1977) *J. Comput. Phys. 24*, 96.
Davis, D. G. (1989) *J. Magn. Reson. 81*, 603.
Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820.
Diakun, G. P., Fairall, L., & Klug, A. (1986) *Nature 324*, 698.
Dickson, C., Eisenman, R., Fan, J., Hunter, E., & Teich, N. (1985) in *RNA Tumor Viruses*, 2nd ed. (Weiss, R., Teich, N., Varmus, H., & Coffin, J., Eds.) Vol. 1 and 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dott, J., Kohler, S., & Gronenborn, A. M. (1989) *Biochemistry 28*, 2601.
Gorelick, R. J., Henderson, L. E., Hanser, J. P., & Rein, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8420.
Green, L. M., & Berg, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4047.

Havel, T. F., Crippen, G. M., & Kuntz, I. D. (1979) *Biopolymers 18*, 73.
Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. W., & Oroszlan, S. (1981) *J. Biol. Chem. 256*, 8400.
Jentoft, J. E., Smith, L. M., Fu, X., Johnson, M., & Leis, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7094.
Jentoft, J. E., Smith, L. M., & Secnik, J. (1989) *Biophys. J. 55*, 584a.
Karpel, R. L., Henderson, L. E., & Oroszlan, S. (1987) *J. Biol. Chem. 262*, 4961.
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 565.
Macura, S., Ernst, R. R., & Wuthrich, K. (1982) *J. Magn. Reson. 47*, 351.
Marion, D., & Bax, A. (1989) *J. Magn. Reson.* (in press).
Meric, C., & Goff, S. P. (1989) *J. Virol. 63*, 1558.
Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., & Klevit, R. E. (1988) *Science 241*, 1489.
Rance, M., Sorensen, O. W., Bodenhausen, B., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479.
Redfield, A. G. (1957) *IBM J. Res. Dev. 1*, 19.
Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469.
South, T. L., Kim, B., & Summers, M. F. (1989) *J. Am. Chem. Soc. 111*, 395.
South, T. L., Kim, B., Hare, D. R., & Summers, M. F. (1990) *Biochem. Pharm.* (in press).
Watenpaugh, K. D., Sieker, L. C., & Jensen, L. H. (1980) *J. Mol. Biol. 138*, 615.
Weber, P. L., Morrison, R., & Hare, D. (1988) *J. Mol. Biol. 204*, 483.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wuthrich, K. (1989) *Science 243*, 45.

Electrical Potential of Transfer RNAs: Codon–Anticodon Recognition†

Kim A. Sharp,*,‡ Barry Honig,‡ and Stephen C. Harvey§

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, New York 10032, and Department of Biochemistry, University of Alabama at Birmingham, Birmingham, Alabama 35294

Received July 19, 1989; Revised Manuscript Received August 31, 1989

ABSTRACT: Calculations of the electrostatic potentials were made around yeast elongator phenylalanine, aspartate tRNAs, and yeast initiator methionine tRNA in aqueous solution at physiological ionic strength. The calculations were carried out with a finite difference algorithm for solving the nonlinear Poisson–Boltzmann equation that incorporates the screening effects of the electrolyte, the exclusion of ions by the molecule, the molecular shape, and the different polarizabilities of the solvent and the tRNA. The initiator tRNA is surrounded by uniformly spaced contours of negative potential. The elongator tRNAs are also surrounded by a similar contour pattern except in the anticodon region where there is a pronounced "hole" in the potential surface. This hole is caused by an invagination of the potential contours, which also results in an increase in the local field strength. The effect of this hole is that the anticodon region in the elongator tRNAs is the least negative, or conversely the most positive, region of the molecule. This hole, which is not found when simple Coulombic potentials are used, is due both to the structure of the elongator tRNA anticodon loops and to the different polarizabilities of the solvent and tRNA. The existence of the potential hole in elongator tRNAs may account in part for their ability to associate with other negatively charged macromolecules, in particular mRNA. Moreover, it suggests that the anticodon loop of elongator tRNAs is the energetically most favorable point of approach of mRNA to tRNA.

Electrostatic interactions play a central role in interactions between biological molecules (Harvey, 1989; Rogers, 1986; Honig et al., 1986). In traditional modeling studies on macromolecules, electrostatic interactions are often described by Coulomb's law (McCammon & Harvey, 1987). However, this approch neglects three important electrostatic factors. First, the polarizability of water is much greater than that of macromolecules, the dielectric constant of the former being about 80 and that of the latter being around 2–4 (Honig et al., 1986; Gilson & Honig, 1986; Nakamura et al., 1988). The effect of solvent screening and presence of a dielectric discontinuity results in the electrostatic forces being dependent not only on the distribution of charge but also on the shape of the macromolecular surface (Warwicker & Watson, 1982; Zauhar & Morgan, 1985; Gilson et al., 1985; Klapper et al., 1986). Second, when the ionic strength of the solvent is nonzero, the electrostatic screening arising from the mobile ions cannot be calculated from Coulomb's law, since the ionic positions are not fixed and can only be described statistically. Third, the ions in the solvent are excluded from the interior of the molecule.

Several recent investigations based on efficient finite difference and finite element numerical solutions to the linear Poisson or Poisson–Boltzmann equation provided more rigorous calculations of electrostatic effects in protein systems (Warwicker & Watson, 1982; Zauhar & Morgan, 1985; Gilson et al., 1985; Klapper et al., 1986). This approach has been used with considerable success in the study of protein structure and enzyme activity (Klapper et al., 1986; Rogers & Sternberg, 1984; Rogers et al., 1985; Warwicker et al., 1985; Gilson &

† This research was supported by grants from the NIH (GM-30518 to B.H.), the NSF (DMB-8706551 to S.C.H.), and the ONR (N00014-86-K0483 to B.H.).
* To whom correspondence should be addressed.
‡ Columbia University.
§ University of Alabama at Birmingham.

0006-2960/90/0429-340$02.50/0