Three-Dimensional Solution Structure and Stability of Phage 434 Cro Protein[†,‡]

S. Padmanabhan,*[,§] M. A. Jiménez,[§] C. González,[§] J. M. Sanz,[||] G. Giménez-Gallego,[||] and M. Rico*[,§]

Instituto de Estructura de la Materia and Centro de Investigaciones Biologicas, Consejo Superior de Investigaciones Cientificas, Madrid, Spain

Received January 14, 1997; Revised Manuscript Received March 20, 1997[⊗]

ABSTRACT: ^{1}H NMR resonances of the phage 434 Cro protein were assigned using standard 2D NMR methods, and its solution structure determined using 867 distance constraints in distance geometry (DIANA) calculations ultimately refined by restrained molecular dynamics (GROMOS). In the 20 best NMR structures, the average pairwise backbone and heavy atom RMSDs are 0.63 ± 0.14 and 1.53 ± 0.15 Å, respectively, for the structurally well-defined residues 4−65. Residues 1−3 and 66−71 at the N- and C-termini are structurally disordered. The region 4−65 includes five α-helices and tight turns which define the hydrophobic core of the protein. The backbone and heavy atom RMSDs for residues 4−65 are 0.92 ± 0.12 and 1.99 ± 0.12 Å, respectively, for the NMR versus the crystal structures, but there are significant differences in the side-chain conformations and solvent accessibilities for some core residues. Analytical ultracentrifugation experiments confirm that 434 Cro is monomeric even at the high NMR concentrations. 434 Cro folding under NMR solution conditions is two-state as indicated by coincident urea denaturation curves from circular dichroism and intrinsic fluorescence measurements. They yield values for 434 Cro stability which show good correspondence to the free energy for global unfolding determined by NMR hydrogen exchange measurements for the slowest exchanging amide protons.

The bacteriophage 434 Cro and repressor are sequence-specific DNA-binding proteins which regulate the lysis−lysogeny "switch" of the phage, and have been extensively characterized genetically and biochemically (Ptashne, 1986). 434 Cro is a monomer while the repressor is a dimer in the free form, although both bind specifically to DNA as dimers (Harrison & Aggarwal, 1990). The 71-residue 434 Cro and the 69-residue monomeric N-terminal DNA-binding domain of the 434 repressor (R1−69)[1] are remarkably similar in sequence and structure, as indicated by the high-resolution crystal structures of the two proteins in the absence of DNA and in their DNA−protein co-crystal structures [see Harrison and Aggarwal (1990) and references therein]. The NMR structure of R1−69 free in solution has been determined (Neri et al., 1992a). Here we determine the NMR structure of 434 Cro free in solution and examine its stability under these solution conditions.

We are interested in the 434 Cro as a model single-domain α-helical protein for solution studies of its structure, folding, and DNA binding. The small size, the high solubility and monomeric nature in aqueous solutions, the absence of any disulfides or prosthetic groups, and the availability of a good expression system and high-resolution structural information make 434 Cro a convenient model for physical and structural studies of both folding and DNA binding. The folding and stability of 434 Cro can be compared with those of other structurally homologous, single-domain, all α-helical, monomeric proteins for which NMR and folding data are available, including R1−69, the monomeric N-terminal domain of bacteriophage λ repressor consisting of residues 6−85 (λ_{6-85}; Huang & Oas, 1995a,b, 1996), and the N-terminal domain (residues 1−76) of the bacteriophage *P22 c2* repressor [P22 c2(1−76): Sevilla et al., 1994]. It may be noted that unlike the last three mentioned 434 Cro is the whole protein and not a fragment corresponding to a specific, single domain. The essential first step is to characterize 434 Cro in solution, and this paper is focused on obtaining the ^{1}H NMR resonance assignments and the three-dimensional solution structure of 434 Cro and on comparing it with its crystal structures in the free and DNA-bound forms. We also examine the folding and the stability of 434 Cro under NMR solution conditions.

EXPERIMENTAL PROCEDURES

Expression and Purification of 434 Cro Protein. The plasmid pRW74 containing the gene for the 434 Cro protein under the control of a *tac* promoter was a generous gift from the laboratory of Professor Stephen Harrison (Harvard University). The 434 Cro purification protocol used was that of Mondragon et al. (1989a) modified as described below. 5 mL of overnight culture of *Escherichia coli* strain AB1899 *lac* (Zazo et al., 1992), containing plasmid pRW74, was innoculated into 400 mL of Terrific Broth (TB) medium (Sambrook et al., 1989) with 100 μg/mL carbenicillin at 30 °C. 434 Cro expression was induced with 0.4 mM isopropyl

[†] Supported by DGICYT Grant PB93-0189 from the Ministerio de Educación y Ciencia, Spain.

[‡] The coordinates and constraints for the 20 NMR structures have been deposited in the Brookhaven Protein Data Bank with PDB ID codes 1ZUG for coordinates and R1ZUGMR for restraints.

* Corresponding authors at Serrano 119, 28006 Madrid, Spain.

[§] Instituto de Estructura de la Materia.

[||] Centro de Investigaciones Biologicas.

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.

[1] Abbreviations: ASA, solvent-accessible surface area; CD, circular dichroism; COSY, two-dimensional (2D) correlated spectroscopy; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; λ_{6-85}, N-terminal domain of phage λ repressor (residues 6−85); NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, 2D nuclear Overhauser effect spectroscopy; R1−69, N-terminal domain of 434 repressor (residues 1−69); SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; P22 c2(1−76), N-terminal domain (residues 1−76) of phage *P22 c2* repressor; RMSD, root mean square deviation; TSP, sodium 3-trimethylsilyl(2,2,3,3′-$^2H_4$)propionate; TOCSY, 2D total correlation spectroscopy; Tris, tris(hydroxy)aminomethane; UV, ultraviolet.

S0006-2960(97)00085-8 CCC: $14.00

β-D-thiogalactoside when the optical density at 600 nm of the culture was 0.8−1.0. Cells were pelleted by centrifugation after overnight growth, resuspended in lysis buffer (200 mM KCl, 50 mM Tris-HCl, 2 mM EDTA, 5% glycerol, 1.4 mM β-mercaptoethanol, pH 7.9) at <1 mL of buffer/g of cells, and lysed in a French press. Lysed cells were diluted into three times their volume of Buffer 8 (10 mM Tris-HCl, 2 mM EDTA, 5% glycerol, 1.4 mM β-mercaptoethanol, pH 8.0) and 400 mM KCl, and centrifuged at 9000 rpm for 45 min to remove cell debris. The supernatant was decanted into three times its volume of Buffer 9 (10 mM Tris-HCl, 2 mM EDTA, 5% glycerol, 1.4 mM β-mercaptoethanol, pH 9.0) containing pre-equilibrated SP-Sephadex C-25 (Pharmacia) at 1 g of dry resin per 10 g of cell wet weight and was gently stirred for 2−3 h. 434 Cro was eluted from SP-Sephadex C-25 using from 0.1−1 M KCl gradient in Buffer 8 and concentrated with 60% ammonium sulfate. It was resuspended in minimal volume of Buffer 8 plus 100 mM KCl and purified by reverse phase HPLC on a Vydac-C_4 preparative column using a linear 0−70% acetonitrile/water/0.1% trifluoracetic acid gradient monitored at 280 nm. The major peak corresponding to 434 Cro was pooled and lyophilized. Its purity was >95% in a 15% SDS−PAGE gel with Coomassie Blue staining and in an analytical Vydac-C_4 reverse-phase column. The reverse-phase step does not alter Cro structure: its NMR spectrum in native, aqueous conditions [at 20 °C and in 100 mM KCl, 25 mM KH_2PO_4, 1 mM DTT, 0.01% NaN_3 at pH 6.0 (NMR buffer)] is identical to the NMR spectrum obtained for the protein purified by the earlier protocol in which Sephadex G-50 was used instead of reverse-phase HPLC (Mondragon et al.,-1989a). The yield of pure Cro was 5 mg/L of culture, and its concentration was determined from the absorbance at 280 nm with $\epsilon_{280} = 6970\ M^{-1}\ cm^{-1}$ (Gill & von Hippel, 1989).

Sedimentation Equilibrium. The monomeric state of 434 Cro at NMR concentrations was confirmed by sedimentation equilibrium experiments in a Beckman Optima XL-A ultracentrifuge using a Ti60 rotor and six-channel Epon charcoal centerpieces. 70 μL samples of 434 Cro in the concentration range 0.1−1.00 mM in NMR buffer were centrifuged at 40 000 rpm and at 20 °C. Radial scans were monitored at 250, 280, or 300 nm and recorded every 2 h after 10 h of centrifugation. Successive radial scans were superimposable after 18 h of centrifugation, indicating equilibrium. The apparent relative molecular mass (M_r) of 434 Cro was determined by analyzing equilibrium radial scans using the program XLAEQ (Beckman) with the partial specific volume of 434 Cro set to 0.753 mL g^{-1} as calculated from its amino acid composition (Laue et al., 1992).

NMR Spectroscopy. NMR samples were prepared by dissolving the lyophilized protein in a 9:1 H_2O:^2H_2O or in a ^2H_2O solution of NMR buffer in 5 mm diameter tubes. NMR spectra at pH 6.0/4.5 and at temperatures 10/20/30 °C were recorded on a Bruker AMX-600 NMR spectrometer with a ^1H frequency of 600.13 MHz and TSP as the chemical shift reference. 1D spectra were obtained using 32K data points, 8475 Hz spectral width (14 ppm), and were averaged over 16, 32, or 80 scans. 2D spectra with a spectral width of 8475 Hz, 1 s recycle delay, an acquisition data matrix size of 2048 × 512 (t_2 and t_1, respectively), and at least 80 scans were acquired in the phase-sensitive mode using time-proportional phase incrementation (TPPI) technique (Marion & Wüthrich, 1983). Standard phase-cycling sequences were used for COSY (Aue et al., 1976) and NOESY (Kumar et al., 1980) spectra. TOCSY spectra were obtained using the standard MLEV17 spin-lock sequence (Bax & Davis, 1985). Mixing times were 60 or 150 ms for NOESY spectra and 80, 40, or 26 ms for TOCSY spectra. Water signal suppression was achieved by presaturation or by WATERGATE (Piotto et al., 1992). Data were zero-filled, resolution-enhanced, and base line corrected using phase-shifted sine bell or square sine bell window functions (optimized for every spectrum) and processed using Bruker UXNMR software. ^1H NMR resonances were assigned using standard sequential assignment procedures (Wüthrich, 1986). Random coil αCH proton chemical shifts from Wishart et al. (1995) were used to compute the difference between the observed and random coil chemical shifts of the αCH proton for each residue—its chemical shift index or CSI (Wishart et al., 1991, 1992). $^3J_{HN\alpha}$ coupling constants of non-overlapping signals were estimated by the analysis of a TOCSY spectrum [40 ms mixing time, 4096 × 512 ($t_2 \times t_1$) data size, in 9:1 H_2O:^2H_2O NMR buffer at 20 °C] using the method of Stonehouse and Keeler (1995).

Amide Proton Exchange by NMR. To identify slow exchanging amide protons, the protein in 0.5 mL of ^1H_2O NMR buffer solution at pH 6.0/20 °C or pH 4.5/10 °C was lyophilized in an NMR tube, redissolved in 0.5 mL of ^2H_2O of equivalent pH (the pH was rechecked at the end of every exchange experiment), and consecutive 2D-TOCSY spectra with a short mixing time of 26 ms (to observe only the NH−αCH type three-bond couplings) were recorded. TOCSY peaks still present 3 and 12 h after exchange initiation were qualitatively classified as the slow and the slowest exchanging (most protected) amide protons, respectively. The exchange rates at pH 6.0 and 20 °C for the most protected amide protons were estimated using 1D NMR as follows. Exchange was initiated as described before, and immediately a continuous series of 1D spectra, each with 32K data points, 8475 Hz spectral width (14 ppm), 80 scans, and an acquisition time of 5 min, was recorded. 3 h after initiation of exchange only the slowest exchanging amide proton signals which do not overlap in the 1D spectra and which were previously identified in the TOCSY are observed. Their peak intensities were integrated after they were base line corrected, and their intensities were normalized relative to the non-exchanging 3,5-aromatic protons of the single Tyr. The integrated intensities, $I(t)$, of each of these NH signals at a given time, t, (≥3 h) after initiation of exchange, was fitted to a single exponential decay, $I(t) = I(0)\exp(-k_{ex}t)$ + base line, to determine the individual exchange rate (k_{ex}) and the intensity at time zero, $I(0)$. $I(0)$ for any NH signal in the absence of signal overlap is expected to be 1 when normalized to the two non-exchanging Tyr 3,5-aromatic protons, so only those amide protons for which the fitted value of $I(0)$ was in the range 1.03 ± 0.12 are reported. The base line term was 0 in every case except Leu54 where it is 2 because of signal overlap with two non-exchanging aromatic Trp protons. Backbone amide proton exchange in proteins is in the EX2 limit whereby $k_{ex} = (K_{op})(k_i)$ (Englander & Kallenbach, 1984). Since k_i, the intrinsic exchange rate for residue i, is known (Bai et al., 1993), K_{op}, the equilibrium constant for i going from the ensemble of states where it is protected from solvent to an ensemble where it is solvent-exposed, and $\Delta G_{op} = -RT \ln K_{op}$ can be determined.

Determination of Three-Dimensional Structure. Nonsequential and long-range NOE cross-peaks were assigned in consecutive rounds of peak assignments and distance geometry calculations as follows. Possible assignments for NOE cross-peaks were obtained using the program XEASY (Bartels et al., 1995), and their intensities were determined using integration routines in XEASY. The intensities were converted into upper distance bounds with the program CALIBA calibrated using scaling factors chosen to reproduce the correct distance limits for proton pairs separated by fixed distances, e.g., HB2−HB3 distances, αCH-methyl in Val etc. (Güntert et al., 1991). Cross-peaks which could not be integrated due to partial overlap were qualitatively classified as strong, medium, and weak and assigned distance constraints of 3.0, 4.0, and 5.0 Å, respectively. Initial structures were calculated from the NMR-derived upper bounds using the variable target function method implemented in the distance geometry program DIANA and the REDAC strategy (Güntert & Wüthrich, 1991). The best structures at the end of each cycle were used to resolve ambiguities in the cross-peak assignments with the program ASNO (Güntert et al., 1993). The program GLOMSA (Güntert et al., 1991) was used to obtain stereospecific assignments from structures obtained in the final stages of the distance geometry calculations. For protons not stereospecifically assigned, and the methyl and aromatic protons, the usual pseudo-atom corrections were applied (Wüthrich, 1986). No constraints from H bonds or from $^3J_{HN\alpha}$ values were used in the structure calculations. The best structures obtained in the final round of DIANA calculations were further refined by restrained molecular dynamics using the program GROMOS (van Gunsteren & Berendsen, 1987) using the following protocol and parameters previously used in our laboratory for structure calculations (Santoro et al., 1993; Rico et al., 1994). The five best DIANA structures were first energy minimized, and then each was subjected to a simulated annealing protocol consisting of a heating period of 5 ps, a high-temperature (1000 K) run of 20 ps from the trajectory of which one structure was extracted after every 5 ps; this was followed by a cooling procedure of 5 ps at 600 K and a final 5 ps at 300 K, averaged, and finally energy minimized. 20 final energy-minimized NMR structures were thus generated. The input constraints for the structure calculation and the atomic coordinates for the calculated structures are deposited with the Brookhaven Protein Data Bank.

The programs SYBYL version 6.1 (Tripos Inc., St. Louis, MO) and MOLMOL (Koradi et al., 1996) were used in structure analysis, molecular graphics manipulations, detection of possible H bonds, and surface area calculations (Richards, 1977). The average NMR structure was analyzed with the Shifts subroutine in MOLMOL to identify residues whose αCH could be subject to sizable ring current effects. The criteria for H-bond formation are proton−acceptor distance <2.40 Å and the donor−proton−acceptor angle <35°. Surface area calculations used the standard solvent radius of 1.4 Å. The polar and nonpolar components and the total accessible surface areas for native 434 Cro (from its NMR and crystal structures) and denatured 434 Cro (modeled as an extended structure with β-sheet dihedral angles; Livingstone et al., 1991; Myers et al., 1995) were determined using atomic radii from Richards et al. (1977) in the program VADAR, which implements the surface area calculation program ANAREA (Richmond, 1984).

CD and Tryptophan Fluorescence Experiments. CD measurements were performed in a Jasco J720 spectropolarimeter equipped with a Neslab temperature control unit and calibrated with (+)-10-camphorsulfonic acid (Chen & Yang, 1977). Fluorescence measurements were done in a Perkin-Elmer LS-50B spectrofluorometer thermostatically controlled at 293 K. The respective protein concentrations and cuvette path lengths were 12−20 μM and 1 or 2 mm for far-UV CD measurements; 100−160 μM and 1 cm for near-UV CD measurements; 5−10 μM and 1 cm for fluorescence measurements. CD spectra were collected at 20 nm/min with a bandwidth of 1 nm, 2 or 4s response time, and averaged over 6 (for far-UV) or 10 (for near-UV) scans. For intrinsic protein fluorescence measurements the samples were excited at 280 nm with an excitation slit width of 2.5 nm, and the emission spectra, which were averaged over 2 scans, were recorded using an emission slit width of 5 nm. Thermal denaturations were monitored only by far-UV CD using the ellipticity at 222 nm and a heating rate of 20 °C/h. Urea denaturations at pH 6.0 and 20 °C were monitored using the ellipticity at 222 nm (far-UV CD) or 275 nm (near-UV CD) and by intrinsic fluorescence. In the latter, the ratio of the fluorescence emission at 340 nm to 360 nm (on either side of the emission maximum for native or unfolded protein) was used since the ratio, in addition to normalizing protein concentrations, also gave improved native and unfolded base lines. Both urea- and temperature-induced folding transitions equilibrate very rapidly, as verified by samples equilibrated for a few minutes or overnight, giving identical CD or fluorescence signals. Consequently, urea denaturations were carried out by the serial addition of small aliquots of a concentrated stock solution of urea to the protein sample. A fresh stock solution of urea (ultrapure from ICN Biochemicals) in NMR buffer was prepared prior to each experiment as described by Pace et al. (1990). After each addition of urea the sample was equilibrated for about 5 min and then the CD or fluorescence signal was recorded. Each time the cuvette became full, a new sample of the protein in a urea concentration somewhat lower than the last measurement was used and the process was repeated, with the correspondence between overlapping parts of the curve being used to check for any systematic errors. Both the amount of protein required as well as errors associated with preparing individual samples in various urea concentrations are minimized by this serial addition method.

Analysis of Denaturation and Free Energy Determination. Denaturation curves were analyzed using a two-state model for the N(native)↔U(unfolded) equilibrium. The observed free energy of unfolding (ΔG_{obs}) and the unfolding equilibrium constant (K_U) are related as

$$\Delta G_{obs} = -RT \ln K_U = -RT \ln\{(y_N - y)/(y - y_U)\} \quad (1)$$

where y, y_N, and y_U are the observed, native, and unfolded CD or fluorescence signal at a given temperature (T, in degrees Kelvin) or urea concentration. Equation 1 can be expressed as

$$y = \frac{(y_{N0x} + m_{N0x}x) + (y_{U0x} + m_{U0x}x)e^{-\Delta G_{obs}/RT}}{(1 + e^{-\Delta G_{obs}/RT})} \quad (2)$$

where x corresponds to the absolute temperature in temperature denaturation and to urea concentration in urea dena-

turation. The native base line is described by the intercept at zero x, y_{N0x}, and the slope, m_{N0x}. y_{U0x} and m_{U0x} are the respective unfolded base line slope and intercept. The fraction of native protein under given solution conditions can be calculated from the observed CD or fluorescence signal (y) and the corresponding fully unfolded (y_U) and fully native (y_N) spectroscopic signals using the following expression: native fraction $= (y - y_U)/(y_N - y_U)$. The temperature dependence of ΔG_{obs} is a modified form of the Gibbs−Helmholtz equation and is described by eq 3a (Becktel & Schellman, 1987); the urea dependence of ΔG_{obs} is assumed to be linear and is given by eq 3b (Santoro & Bolen, 1988).

$$\Delta G_{obs}(T) = \Delta H_m(1 - T/T_m) - \Delta C_p[(T_m - T) + T \ln(T/T_m)] \quad (3a)$$

$$\Delta G_{obs} = \Delta G_{water} - m[\text{urea}]; \quad \Delta G_{water} = mC_m, \quad \text{for } \Delta G_{obs} = 0 \quad (3b)$$

Here T_m and C_m are the temperature and urea concentration at the midpoint of the folding transition where $\Delta G_{obs} = 0$, ΔH_m is the enthalpy at T_m, and ΔC_p is the heat capacity. m and ΔC_p are also assumed to be temperature independent. Since an accurate value for ΔC_p cannot be directly determined from only the denaturation data in this study, we used fixed values of ΔC_p in the range 0.9−1.4 kcal mol^{-1} K^{-1}. Each denaturation experiment was repeated at least twice, and the data were subjected to a six-parameter nonlinear least square fitting: four parameters for the native and unfolded base lines and m and C_m, for urea denaturation, or ΔH_m and T_m, for thermal denaturation. Curve fittings and the reported errors were obtained using Kaleidagraph (Synergy Software, PCS Inc.) or Sigmaplot (Jandel Corporation) software.

RESULTS

^{1}H NMR Assignments and Secondary Structure in 434 Cro. Sedimentation equilibrium analytical ultracentrifugation experiments of 434 Cro at concentrations between 0.1 and 1 mM and in the same solution conditions as those used in the NMR experiments yield weight-average molecular weights that are essentially identical and equal to the molecular weight calculated from the amino acid sequence (data not shown). This confirms that the protein is monomeric even at millimolar concentrations, as has been reported (Harrison & Aggarwal, 1988). Consequently, all the NMR data we report for 434 Cro are intramolecular.

434 Cro in solution conditions favoring the native folded state (20 or 30 °C and in 100 mM KCl, 25 mM KH_2PO_4, 1 mM DTT, 0.01% NaN_3 at pH 6.0; NMR buffer) shows good NMR spectral dispersion unlike several other helical proteins such as interleukin-4 (Powers et al., 1992), the P22 c2(1−76) domain (Sevilla et al., 1994), and the λ_{6-85} domain (Huang & Oas, 1995a). The spectral dispersion and general quality of the NMR data are illustrated in the NOESY spectrum of 434 Cro shown in Figure 1, and essentially complete ^{1}H NMR assignments for 434 Cro could be achieved from 2D ^{1}H NMR spectra without ^{15}N or ^{13}C isotope labeling. ^{1}H NMR resonances were assigned employing the standard strategy of identifying individual amino acid spin systems from COSY and TOCSY spectra, and sequential connectivities from NOESY spectra (Wüthrich, 1986). Continuous sequential backbone−backbone NOEs (αN, NN, and βN) facilitated assignments of most backbone proton resonances except where these were absent, such as between some residues in the stretch 42−46 and at the C-terminus (Figure 2). These few residues (Lys42, Lys66, Arg45, and the Pro α/β protons) were assigned on the basis of their unique spin systems from among the remaining unassigned resonances, from compatible NOEs in successive rounds of structure calculations, or from assignments in model peptides. For example, since structure calculations (see below) indicated that residues 66−71 are disordered, their ^{1}H NMR assignments could be inferred or confirmed from those in a synthetic water-soluble peptide fragment of 434 Cro containing these residues under identical solution conditions (Padmanabhan et al., unpublished data). Strong and distinct NOE cross-peaks between Pro44 (δ, δ') and Arg43 αCH protons and between Pro58 (δ, δ') and Asp57 αCH protons enabled the identification of the two Pro residues and showed that both are in the *trans* conformation (Wüthrich, 1986). All of the aromatic ring protons and the side-chain amide resonances of Gln and Asn were easily identified by comparing spectra obtained in 9:1 H_2O:^2H_2O and 100% ^2H_2O. The only backbone amide protons unidentified are those of Met1 and Gln2 possibly due to the pH used (=6.0). The side-chain ^{1}H NMR resonances were assigned for most residues except for some lysines. The ^{1}H NMR resonance assignments are provided in Table 1 of the Supporting Information.

The presence of five helices roughly spanning residues 4−15, 19−26, 30−38, 47−53, and 58−63 in 434 Cro can be inferred from criteria summarized in Figure 2 and include the following. (i) Strong, nearly contiguous sets of sequential $d_{NN}(i,i+1)$ NOEs, and medium range $d_{NN}(i,i+2)$, $d_{\alpha\beta}(i,i+3)$, $d_{\alpha N}(i,i+3)$, and $d_{\alpha N}(i,i+4)$ NOEs are seen in the helical regions, while few or no such NOEs are seen in the regions between helices and, in particular, in the stretch of residues 41−46 and 66−71 (Wüthrich, 1986). (ii) $^3J_{HN\alpha}$ values (estimated for 55 out of 63 possible residues excluding Met1, Gln2, two Pro, and four Gly) are <6 Hz for residues in helices and >6 Hz for interhelical N- or C-terminal residues (Wüthrich, 1986). (iii) The αCH proton chemical shifts minus the random coil value for a given residue, referred to as the chemical shift index or CSI, is negative (by more than 0.1 ppm) for residues in the helices as expected [see Wishart et al. (1991, 1992)]. CSI values were used only qualitatively because they can be altered by aromatic ring current effects: residues whose CSI may be so affected are indicated in Figure 2 and were identified by analyzing the NMR structures (see below) as described in the Experimental Procedures. (iv) Residues involved in hydrogen bond (H bond) formation such as in helices or helix termini typically exhibit slower amide proton−solvent exchange (Englander & Kallenbach, 1984). In a 2D-TOCSY spectrum at pH 6.0 and 20 °C we identified 30 NH-αCH cross-peaks 3 h after initiation of amide proton exchange, and eight of these still remained 12 h after exchange initiation. All of these corresponded to residues in helices or their C-terminii.

Solution Structure Calculations of 434 Cro. Calculations of solution structures used the procedures described in Experimental Procedures. These involved several repeated rounds of assigning non-sequential and long-range cross-peaks in NOESY spectra which were then converted with the programs XEASY and CALIBA to distance constraints

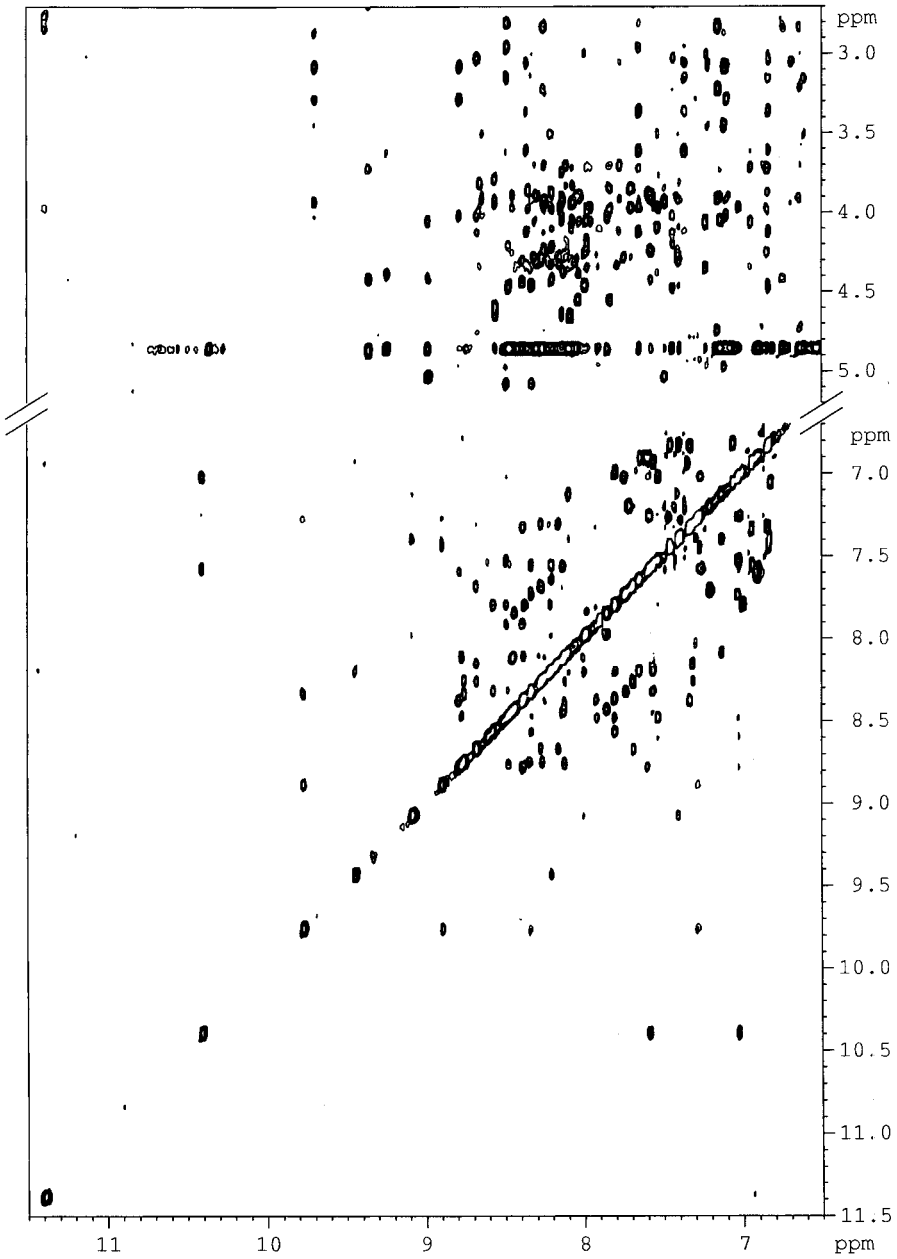

FIGURE 1: 2D ^{1}H NMR NOESY spectra (^{1}H frequency of 600.13 MHz) with mixing time 150 ms of 434 Cro (1 mM) in 100 mM KCl, 25 mM KH_2PO_4, 1 mM DTT in 9:1 H_2O:D_2O, pH 6.0, 20 °C, showing spectral dispersion in the NH−NH and αCH−NH regions. The signal with the unusually large chemical shift value of 11.38 corresponds to the ϵNH of Arg12.

for use as input in DIANA, the distance geometry program for structure calculations. Distance constraints were derived from the 150 ms mixing time NOESY spectrum cross-checked with the 60 ms NOESY spectrum for the absence of spin-diffusion contributions. The best DIANA structures were used to stereospecifically assign the methyls of three Leu residues and nine out of 30 possible $\beta\beta'$ protons (excludes all Ala, Thr, Val, Ile, Pro, and Gly residues and 14 others without distinct chemical shift assignments for the β and β' protons) with the program GLOMSA and were then subjected to another DIANA calculation; the resulting structures were refined by restrained molecular dynamics using the program GROMOS (see Experimental Procedures) to generate the final 20 energy-minimized NMR solution structures. Table 1 summarizes the numbers of the structurally relevant intraresidual, sequential, and medium- and long-range constraints and the total used for the final structures, the residual violations, and the potential energies. The large number of medium-range NOEs is consistent with the high α-helical content, while the absence of medium- or long-range constraints for the C-terminal residues 66−71 indicates that these do not have any regular structure (see also Figure 2). The average pairwise RMSD (Table 2) is high when all backbone or all heavy atoms are considered: 2.60 ± 0.96 and 3.00 ± 0.76 Å, respectively, but decrease considerably to 0.63 ± 0.14 and 1.53 ± 0.15 Å, respectively, when only residues 4−65 are considered. Thus residues 4−65 are structurally well-defined but not residues 1−3 or 66−71. Figure 3A shows the backbone conformations in the final 20 energy-minimized structures of 434 Cro superimposed for minimal RMSD over the structurally well-defined residues 4−65 with the five structurally disordered C-terminal residues 66−71 omitted for easy visualization.

Analysis of the Three-Dimensional NMR Solution Structure of 434 Cro. 434 Cro has five short helices spanning residues 4−15, 19−26, 30−38, 47−54, and 58−63 (depicted by the

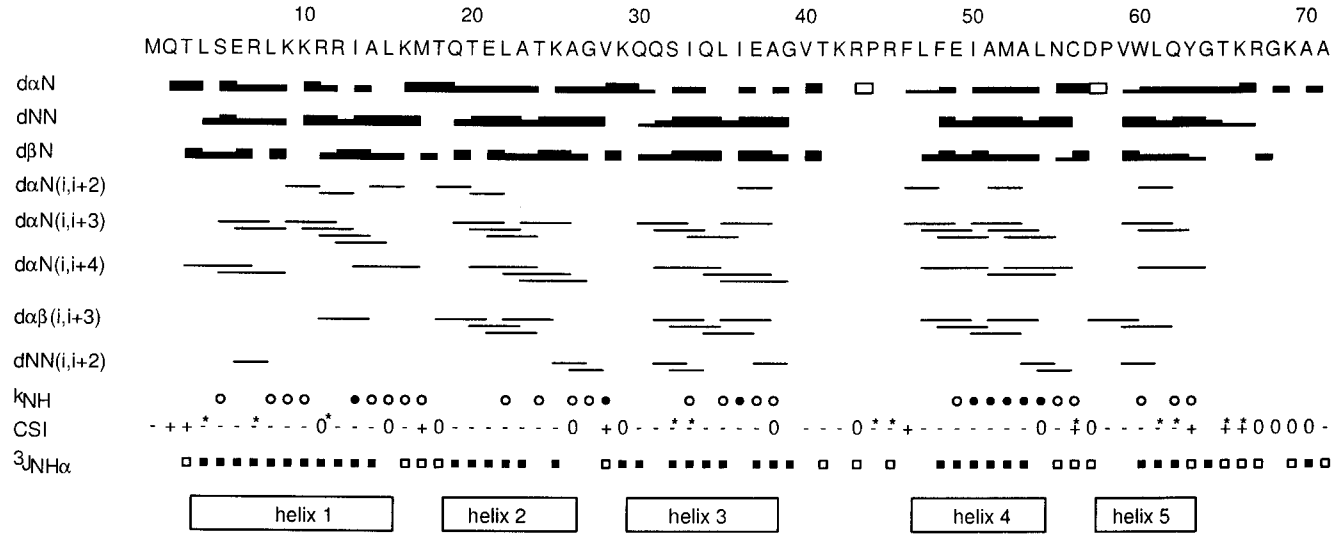

FIGURE 2: Summary of NMR data. (i) Sequential CαH−NH ($d_{\alpha N}$), NH−NH (d_{NN}), and CβH−NH ($d_{\beta N}$) NOE connectivities; medium-range CαH−NH ($d_{\alpha N}(i,i+2)$, $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$), CαH−CβH ($d_{\alpha\beta}(i,i+3)$), and NH−NH ($d_{NN}(i,i+2)$) NOE connectivities for residues i, $i+2$, $i+3$ and $i+4$. Strong, medium, and weak NOEs are indicated by thick, medium, and thin bars, respectively, and the sequential αCH NOE with the following Pro δ proton is indicated by an open bar in the row corresponding to $d_{\alpha N}$. (ii) Amide proton exchange, k_{NH}. "○" and "●" indicate residues whose amide protons were detectable in TOCSY spectra 3 and 12 h, respectively, after exchange was initiated by dissolving the lyophilized protein in ^2H_2O at pH 6.0 and 20 °C (see Experimental Procedures). (iii) The chemical shift index, CSI, is the observed minus random coil αCH chemical shift. "+" and "−", respectively, indicate positive and negative CSI of more than 0.1 ppm, "0" indicates CSI (positive and negative) ≤0.1 ppm, the blanks at positions 27, 39, and 64 are Gly whose two αCH protons have opposite CSI, and "*" indicates residues whose αCH appear to have more than ±0.1 ppm contributions from ring current effects in the average NMR structure analyzed as described in Experimental Procedures. (iv) Amide proton−α-proton vicinal spin−spin coupling constant, $^3J_{HN\alpha}$: <6 Hz (filled black squares) and ≥6 Hz (open squares).

Table 1: NMR Structure Calculations Summary

	no. of constraints			
type of constraints	total	<3.5 Å	3.5 < 4.5 Å	≥4.5 Å[a]
intraresidual ($\|i-j\|=0$)	217	80	81	56
sequential ($\|i-j\|=1$)	259	90	43	126
medium-range ($2 \le \|i-j\| \le 5$)	207	7	51	149
long-range ($\|i-j\|>5$)	184		29	155
all	867			

Residual Constraint Violations in the Final 20 Structures	
range (Å)	av no. of distance constraint violations
0.00−0.25	74.5
0.25−0.50	37.1
0.50−0.75	1.9
0.75−1.00	0.1
>1.00	0
maximum violation (Å)	0.60
av sum of violations (Å)	20.9

	av	range
total energy (kJ mol^{-1})	−4119	−4628 to −3567
Lennard-Jones energy (kJ mol^{-1})	−2663	−2766 to −2555
NOE term (kJ mol^{-1})	121	93 to 142

[a] Upper limit is 5.0 Å plus appropriate pseudo-atom corrections.

Table 2: Structure Comparisons

	RMSD (Å)			
	434 Cro[a]		434 Cro[a] vs R1−69 [f]	
residues	NMR[a]	vs X-ray[b]	vs X-ray[c]	vs NMR[d]
all residues				
backbone	2.60 ± 0.96			
heavy atoms	3.00 ± 0.76			
residues (4−65)				
backbone	0.63 ± 0.14	0.92 ± 0.12	1.30 ± 0.08	1.19 ± 0.12
heavy atoms	1.53 ± 0.15	1.99 ± 0.12		

	ΔASA (Å^2)[g]		
structure	$\Delta ASA_{nonpolar}$	ΔASA_{polar}	ΔASA_{total}
434 Cro (1−65) X-ray[b]	4110	570	4680
434 Cro (1−65) NMR[e]	3960	1040	4990
434 Cro (1−71) NMR[e]	4050	1040	5080
R1−69 (1−63) X-ray[c]	3550	280	3830
R1−69 (1−63) NMR[d]	3420	390	3810
R1−69 (1−69) NMR[d]	3430	400	3830

[a] 20 NMR structures of 434 Cro from this study. [b] 434 Cro X-ray structure (residues 1−65) from Mondragon et al. (1989a). [c] R1−69 X-ray structure (residues 1−63) from Mondragon et al. (1989b). [d] Average of the 20 NMR structures of R1−69 from Neri et al. (1992). [e] Average of the 20 NMR structures of 434 Cro. [f] For superposition of the backbone atoms only of residues 4−65 of 434 Cro and residues 2−63 of R1−69. [g] Calculated as described in Experimental Procedures.

ribbon diagram in Figure 3B). The helices assigned on the basis of backbone (ϕ,ψ) angles and hydrogen-bonding patterns are consistent with the qualitative NMR criteria discussed earlier and shown in Figure 2. Backbone (ϕ,ψ) dihedral angle distributions in the 20 final NMR structures (Figure 4) for residues in structurally well-defined helical and interhelical regions are quite narrow (small angular RMSDs; data not shown) but are especially broad (large angular RMSDs) for the N- and C-terminal residues 1−3 and 65−71 which show structural disorder. All non-glycine residues except Lys16 and Asn55 have (ϕ,ψ) in the allowed regions of the Ramachandran map (not shown). The values for Lys16 and Asn55 are not unusual since Asn, like Gly, can have (ϕ,ψ) both positive (Richardson & Richardson, 1989) as can helix-terminating residues like Lys16 which terminates helix 1 (Schellman, 1980; Nagarajaram et al., 1993, Sevilla et al., 1994). H bonds identified in the NMR structures (Table 3) on the basis of the criteria that the proton−acceptor distance is <2.40 Å and the acceptor−donor−proton angle is <35° are fully compatible with residues whose amide protons exchange slowly with solvent (Figure 2). Continuous stretches of (CO_i−NH_{i+4}) H bonds

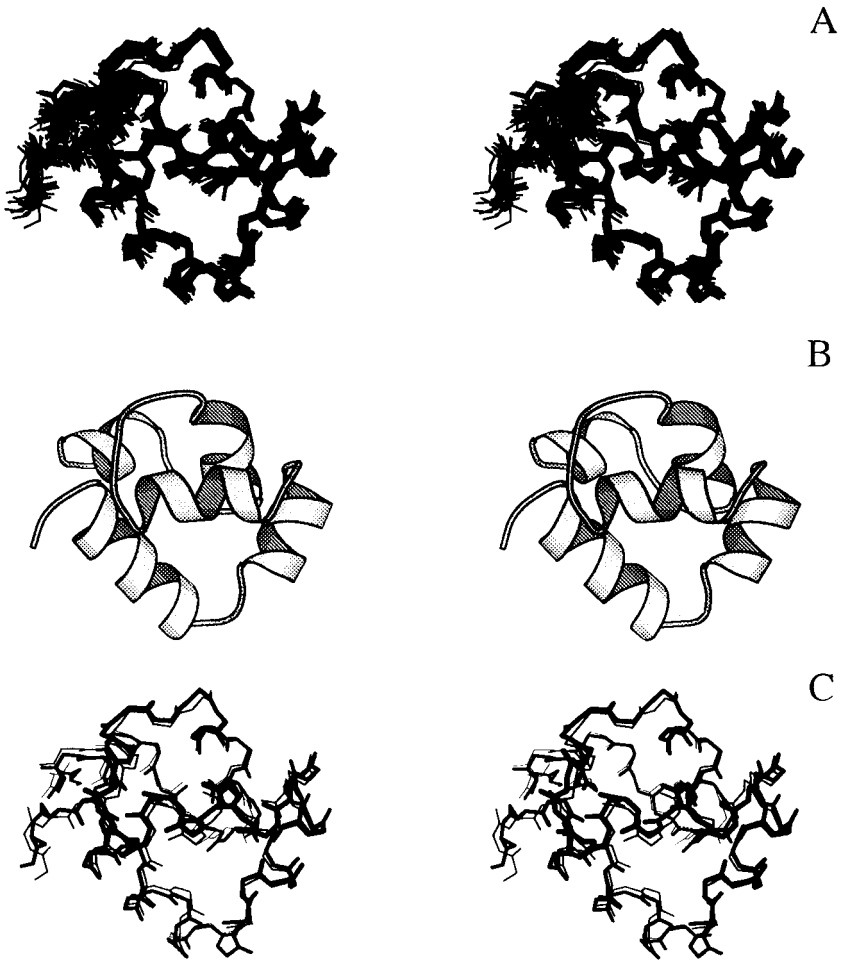

FIGURE 3: Stereoview of the polypeptide backbone of 434 Cro (residues 1−66). (A) The final 20 energy-minimized NMR structures superimposed over residues 4−65. (B) Ribbon diagram of the average of the 20 best NMR structures. (C) The average of the 20 best NMR structures (thick line) and the crystal structure (thin line) superimposed over residues 4−65.

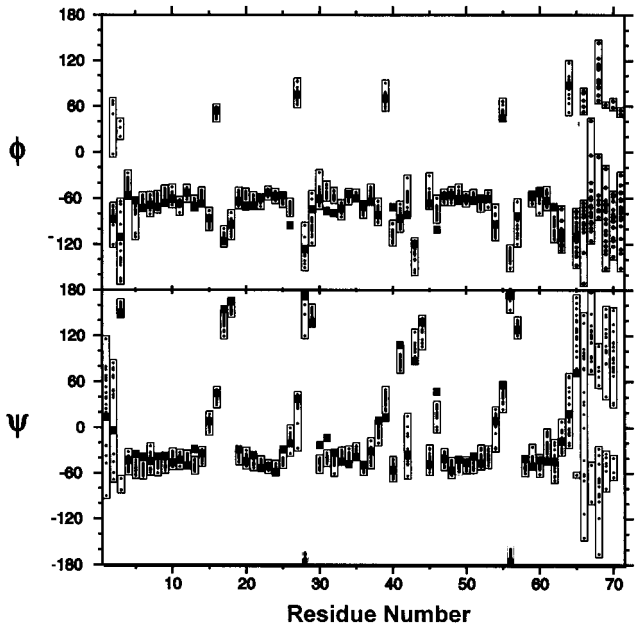

FIGURE 4: Backbone (ϕ,ψ) angles for each residue in the 20 best NMR structures and in the crystal structure. Each dot indicates the value in one NMR structure, the bar indicates the range of NMR values, and the filled square indicates the crystal structure value.

specific for α-helices as well as the (CO_{i-3}−NH_i) and (CO_{i-4}−NH_{i+1}) H bonds characteristic of a C-cap residue *i* are observed (Table 3). The N- and C-caps define helix boundaries and are the first and last residues whose Cα atoms lie on the helix cylinder axis (Richardson & Richardson, 1988) or are those which form helical H bonds but do not have the canonical helical (ϕ,ψ) angles (Presta & Rose, 1988). Helices 1, 2, and 5 have the favorable N-cap residues Thr3, Thr18, and Asp57, respectively; helices 2, 3, and 5 have the favorable C-cap residue Gly, and the Pro residue at the start of helix 5 can also be favorable (Richardson & Richardson, 1988; Presta & Rose, 1988; Doig & Baldwin, 1995). Helix 1 is the longest and the most highly charged, while helices 4 and 5 are the shortest, helix 4 also being the most hydrophobic. All the helices have several intrinsically helix-destabilizing and few helix-stabilizing Ala and Arg^+ residues (Baldwin, 1995). Figure 5 is a contact map for the 434 Cro structures. The five regions along the diagonal with the highest densities of contacts correspond to the five helices. The up-diagonal region of Figure 5 shows the interhelical contacts in the average of the 20 NMR structures of 434 Cro: there are several contacts for helix 1 with helices 3 and 5 (but not between helices 3 and 5) and also for helix 2 with helix 4. Few or no contacts occur for other sets of helices. These features are also apparent in the ribbon diagram of the 434 Cro structure shown in Figure 3B.

The side chains of 23 non-Ala/Gly/Pro residues are well-defined with low conformational flexibility, as deduced from their χ_1 angular RMSD (Figure 6A) being ≤15° and from their χ_1 angular order parameters (Figure 6B) being ≥0.95 (0 and 1 describe completely random and completely well-

Table 3: H Bonds in NMR and X-ray Structures (Residues 3−65)[a]

donor	acceptor	NMR	X-ray
Arg7 NH	Thr3 O′	18	1
Arg7 ϵNH	Leu61 O′		1
Leu8 NH	Leu4 O′	11	1
Lys9 NH	Ser5 O′	20	1
Lys10 NH	Glu6 O′	18	1
Arg11 NH	Arg7 O′	11	1
Arg12 NH	Leu8 O′	20	1
Arg11 ϵNH	Leu54 O′		1
Arg11 ηNH$_2$	Leu54 O′		1
Arg12 NH	Leu8 O′		1
Arg12 ϵNH	Glu37 ϵO		1
Arg12 ηNH$_2$	Met17 O′		1
Arg12 ηNH$_2$	Glu37 ϵO		1
Ile13 NH	Lys9 O′	14	1
Ala14 NH	Lys10 O′	20	1
Leu15 NH	Arg11 O′	20	1
Lys16 NH	Ile13 O′	20	1
Met17 NH	Arg12 O′	20	1
Gln19 ϵNH$_2$	Glu37 ϵO		1
Glu21 NH	Thr18 γO′		1
Leu22 NH	Thr18 O′	20	1
Ala23 NH	Gln19 O′	20	1
Thr24 NH	Thr20 O′	20	1
Lys25 NH	Glu21 O′	20	1
Ala26 NH	Leu22 O′	20	1
Gly27 NH	Thr24 O′	15	
Val28 NH	Ala23 O′	18	1
Gln30 ϵNH2	Thr20 γO′		1
Gln34 NH	Gln30 O′	19	
Leu35 NH	Gln31 O′	18	
Glu37 NH	Ile33 O′	20	1
Ala38 NH	Gln34 O′	20	
Gly39 NH	Ile36 O′	20	
Val40 NH	Leu35 O′	19	1
Thr41 NH	Leu35 O′	17	1
Ile50 NH	Phe46 O′	20	1
Ala51 NH	Leu47 O′	20	1
Met52 NH	Phe48 O′	20	1
Ala53 NH	Glu49 O′	20	1
Leu54 NH	Ile50 O′	20	1
Asn55 NH	Met52 O′	15	1
Cys56 NH	Ala51 O′	20	1
Trp60 NH	Asp57 δO′	20	1
Leu61 NH	Asp57 O′	19	1
Gln62 NH	Pro58 O′		1
Tyr63 NH	Val59 O′	20	1
Gly64 NH	Trp60 O′	19	1

[a] Proton−acceptor distance <2.4 Å; acceptor−donor−proton angle, <35°. H bonds are listed if present in ≥10 NMR structures (the number of NMR structures is given in column 3) or in the crystal structure (column 4).

defined χ_1, respectively; Hyberts et al., 1992). Almost all of these residues are located in regions of regular structure such as in the helices or their boundaries and include Val59, Met52, Leu8/22/35/54, all Ile and aromatic residues, Glu6/49, Arg11, Gln19/30/34, Thr24/41, and Asp57. Moreover, among these residues the fractional solvent-accessible surface areas (ASA) are <0.10 for Leu 8/22/54, Trp60, Ile33/36/50, and Gln19. The residues which have the lowest fractional ASA (<0.05) and constitute the compact hydrophobic core of 434 Cro include, in addition to Leu8/22/54, Ile33/36/50, Leu61, Ala23/51, and Pro44 (Figure 6C), and all of these except Pro44 are present in the helices. The amide protons of Leu54, Ile36, Ile50, and Ala51 are also among those most protected against solvent exchange (Figure 2, Table 4). The changes in the nonpolar, polar, and total accessible surface areas ($\Delta ASA_{nonpolar}$, ΔASA_{polar}, and ΔASA_{total}, respectively) on unfolding 434 Cro as calculated

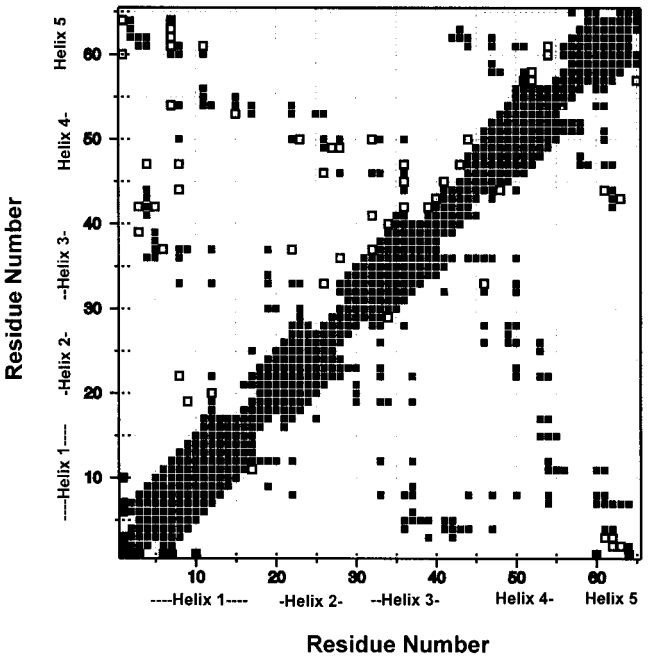

FIGURE 5: Contact plot of the interhelical interactions in 434 Cro. Contacts correspond to distances of Å. The filled squares in the up-diagonal region are contacts observed in the average of the 20 NMR structures, and in the down-diagonal region they are contacts observed in the X-ray structure. The open squares in the up-diagonal region are contacts in the X-ray structure that are not observed in the average of the 20 NMR structures, and in the down-diagonal region the open squares are contacts observed in the average of the 20 NMR structures but not in the X-ray structure. The filled squares close to or along the diagonal are intrahelical contacts or those between successive helices.

from the average of the 20 NMR structures for native 434 Cro and modeling denatured 434 Cro as an extended β-sheet (see Experimental Procedures) are listed in Table 2. These, as will be discussed later, can be related to the thermodynamic parameters describing 434 Cro folding transitions.

Amide Proton Exchange in 434 Cro. Amide proton exchange is very rapid at pH 6.0 and 20 °C, the solution conditions used for the NMR structure determination. Lower temperature (10 °C) and pH (4.5) slow down exchange in 434 Cro but induce changes in the chemical shifts which lead to considerable spectral overlap. The very rapid exchange at pH 6.0 and 20 °C could, however, be exploited to estimate the exchange rates for the most protected, slowest exchanging amide protons using 1D NMR under these solution conditions. 434 Cro has eight such very slowly exchanging amide protons at pH 6.0 and 20 °C which were identified using 2D-TOCSY as described earlier. Table 4 lists five of these eight residues whose amide proton signal intensities do not appear to have contributions from other overlapping signals in 1D spectra recorded 3 h after initiation of exchange. The observed exchange rates for these five amide protons (k_{ex}), their intrinsic exchange rates (k_i) calculated according to Bai et al. (1993), and $\Delta G_{op} = -RT \ln K_{op} = -RT \ln(k_{ex}/k_i)$ are also listed in Table 4. ΔG_{op} for the slowest exchanging amides is a measure of the free energy for global unfolding (Bai et al., 1994). The five slowest exchanging amide protons in 434 Cro listed in Table 4 have similar values for ΔG_{op} with a mean value of 3.7 ± 0.5 kcal mol^{-1}.

CD and Intrinsic Fluorescence Measurements of 434 Cro Folding and Stability. The folding and stability of 434 Cro

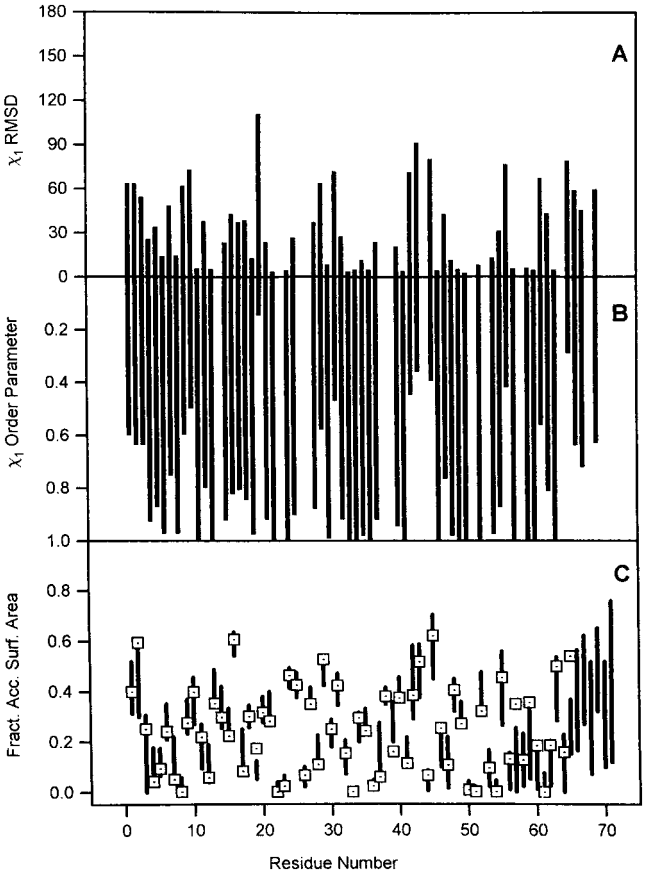

FIGURE 6: (A) χ_1 angular RMSD; (B) χ_1 angular order parameter; (C) fraction of accessible surface area for each residue in the 20 best NMR structures versus amino acid sequence. The bar indicates the range of NMR values, and the square indicates the crystal structure value in C.

Table 4: Parameters for Urea Denaturation and for the Slowest Exchanging Amide Protons of 434 Cro

	Urea Denaturation		
probe	$m^{a,b}$	C_m (M)^b	$\Delta G_{water}{}^{b,c}$
far-UV CD	1.02	3.50	3.6
near-UV CD	1.14	3.48	3.9
fluorescence	1.05	3.35	3.5
mean	1.07	3.44	3.67

	Amide Proton Exchange		
NH residue	k_i (min^{-1})	k_{ex} (min^{-1})	$\Delta G_{op} = (-RT \ln K_{op})^{b,c}$
Ile13	3.10	4.3×10^{-3}	3.8
Ile36	1.15	4.1×10^{-3}	3.3
Ile50	1.32	4.4×10^{-3}	3.3
Ala53	12.92	5.5×10^{-3}	4.5
Leu54	2.64	3.8×10^{-3}	3.8
mean			3.7

[a] In kcal mol^{-1} M^{-1}. [b] Standard errors: far-UV CD, 2% for C_m and 8% for m and ΔG_{water}; near-UV CD and fluorescence, 4% for C_m and 20% for m and ΔG_{water}; amide proton exchange, 14% for mean ΔG_{op}. [c] In kcal mol^{-1} at pH 6.0 and 20 °C in NMR buffer.

in NMR solution conditions was examined using CD and intrinsic Trp fluorescence. Protein concentrations for CD and fluorescence are lower than for NMR, but native 434 Cro is monomeric over this entire concentration range, as shown earlier in sedimentation equilibrium measurements. Native 434 Cro has the far-UV CD spectrum of a helical protein with two minima at 206 and 222 nm. The mean residue ellipticities at 222 nm, $[\Theta]_{222}$ (in deg cm^2 dmol^{-1}), are around −14 000, 0, and −6000, respectively, for native, urea-denatured (≥6 M urea) and temperature-denatured (75 °C) 434 Cro (Figure 7A). A "residual" CD signal attributed to a thermal base line effect has also been observed for several other thermally denatured proteins. Near-UV CD spectra characterize the environment and rigidity of aromatic residues. The near-UV CD spectrum of native 434 Cro has a maximum at 275 nm, whereas in ≥6 M urea no near-UV CD signals are observed (Figure 7B). The intrinsic fluorescence emission maximum on excitation at 280 nm is at 348 nm for native 434 Cro, and this shifts to 353 nm with a decrease in intensity on urea denaturation (Figure 7C).

Thermal transitions monitored by far-UV CD at protein concentrations around 20 μM are >95% reversible and equilibrate very rapidly (see Experimental Procedures). CD spectra recorded for a series of temperatures (not shown) have an isodichroic point at 205 nm, suggesting the presence of just two conformations for each residue, helix, and random chain (Holzwarth & Doty, 1965). The thermal transitions were analyzed as described in Experimental Procedures using fixed ΔC_p values between 0.9 and 1.4 kcal mol^{-1} K^{-1} [this ΔC_p range is based on the 434 Cro sequence length and the per residue ΔC_p values reported by Alexander et al. (1992) and Myers et al. (1995) for a number of proteins]. ΔC_p can also be estimated from its observed linear correlation with the change in the nonpolar and polar solvent-accessible surface areas, $\Delta ASA_{nonpolar}$ and ΔASA_{polar}, respectively, using the following expressions based on the transfer of model compounds from the liquid state to water (Spolar et al., 1992)

$$\Delta C_p = (0.32 \pm 0.04)(\Delta ASA_{nonpolar}) - (0.14 \pm 0.04)(\Delta ASA_{polar}) \quad (4a)$$

or based on the dissolution of solid model compounds (Murphy & Freire, 1992)

$$\Delta C_p = (0.45 \pm 0.02)(\Delta ASA_{nonpolar}) - (0.26 \pm 0.03)(\Delta ASA_{polar}) \quad (4b)$$

$\Delta ASA_{nonpolar}$ and ΔASA_{polar} for 434 Cro listed in Table 2 and eq 4a give $\Delta C_p \approx 1.13 \pm 0.20$ kcal mol^{-1} K^{-1} (or 1.24 ± 0.20 kcal mol^{-1} K^{-1} for the X-ray structure), while eq 4b yields $\Delta C_p \approx 1.50 \pm 0.10$ kcal mol^{-1} K^{-1} (or 1.70 ± 0.10 kcal mol^{-1} K^{-1} from X-ray data). Although eq 4b predicts a higher ΔC_p [ascribed to differences in calculating ΔASA; see Myers et al. (1995)], the range 0.9−1.4 kcal mol^{-1} K^{-1} for ΔC_p is sufficiently conservative and yields T_m around 333 K, ΔH_m = 40−60 kcal mol^{-1} (ΔH_m = 70 ± 4 kcal mol^{-1} from the slope of an approximately linear ln K_U versus $1/T$ van't Hoff plot in the narrow temperature range around T_m), and ΔG_{water} = 2.5−3.6 kcal mol^{-1} (20 °C) or 2.0−3.0 kcal mol^{-1} (37 °C). The values are tentative because ΔC_p is fixed and assumed to be temperature independent. A reliable experimental method of determining ΔC_p is by calorimetry. However, thermal transitions of 434 Cro monitored by calorimetry are irreversible, and no more than 60% of the native 222 nm far-UV CD signal is recovered on cooling the thermally denatured protein at concentrations ≥100 μM (required for calorimetry and near-UV CD). Consequently, only the thermal transitions monitored by far-UV CD (222 nm) and around 20 μM protein concentration were analyzed. ΔC_p can also be determined by a combined analysis of thermal and chemical denaturations, for example, urea denaturations at various temperatures and thermal

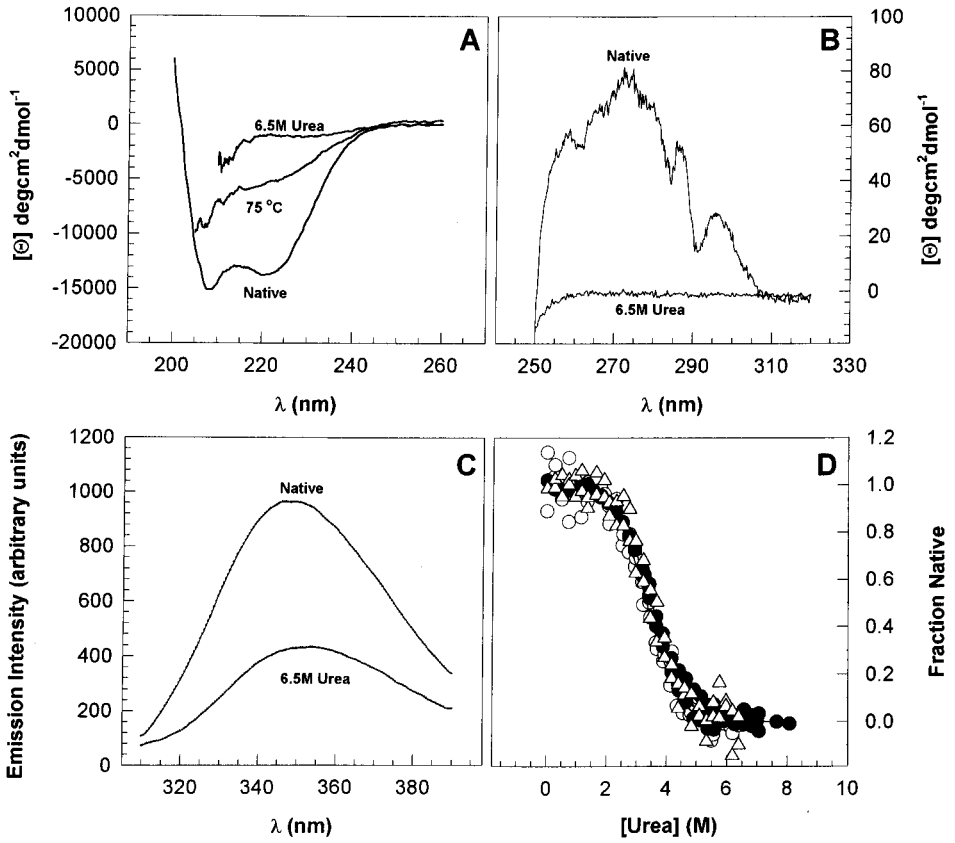

FIGURE 7: (A) Far-UV CD; (B) near-UV CD; (C) fluorescence emission spectra of native (pH 6.0, 20 °C, 0 M urea in NMR buffer), thermally denatured (pH 6.0, 75 °C, 0 M urea in NMR buffer), and urea-denatured (pH 6.0, 20 °C, 6.5 M urea in NMR buffer) 434 Cro. See text for protein concentrations. (D) Fraction of native 434 Cro versus urea concentration at pH 6.0 and 20 °C in NMR buffer monitored by ellipticity at 222 nm (●) and 275 nm (△), and by fluorescence (○).

denaturations at various urea concentrations [see Scholtz (1995) and Huang and Oas (1996)], and such experiments are currently in progress.

Urea denaturations were monitored by far-UV CD (222 nm), near-UV CD (275 nm), and intrinsic fluorescence. The urea-induced transitions equilibrate very rapidly, are reversible, and were analyzed as described in Experimental Procedures. Uncertainties in C_m (like T_m) are small but can be large for m (like for ΔH_m) and, as a consequence, for ΔG_{water} (Pace et al., 1990). This appears to be the case with the fluorescence and near-UV CD data (Table 4). Figure 7D shows that the CD and fluorescence urea denaturation data plotted as the fraction of native protein versus urea concentration are superimposable. The fitted parameters listed in Table 4 are similar within error. At 20 °C, pH 6.0, and in NMR buffer, the three spectral probes together give an average m value of 1.07 ± 0.06 kcal mol^{-1} M^{-1} and an average C_m value of 3.44 ± 0.08 M. These, by the linear extrapolation method (LEM) used in eq 3b, yield an average ΔG_{water} value of 3.7 ± 0.2 kcal mol^{-1}, which is in the range of values for ΔG_{water} at 20 °C obtained from thermal denaturations for the ΔC_p values used. Moreover, ΔG_{water} from urea denaturations is in excellent agreement with ΔG_{op} values for the slowest exchanging amide protons determined under similar solution conditions (see previous section). The average m value is close to the value 0.99 kcal mol^{-1} M^{-1} calculated by multiplying the number of helical residues in 434 Cro (=43) and the per residue m value for helix unfolding (=0.023 kcal mol^{-1} M^{-1} per residue), consistent with observations for other all-helical proteins (Scholtz et al., 1995). The m values and ΔASA also show a linear correlation (although not as good as that observed between ΔC_p and ΔASA): m (cal mol^{-1} M^{-1}) = {368 + 0.11(ΔASA$_{total}$)} (Myers et al., 1995). This expression yields estimates for m (in kcal mol^{-1} M^{-1}) of ~0.92 and 0.88 for the NMR and X-ray structures of 434 Cro, respectively, which are reasonably close to that observed experimentally.

DISCUSSION

Comparison of the NMR and X-ray Structures of 434 Cro. The backbone and heavy atom RMSDs for residues 4−65 between the X-ray structure (Mondragon et al., 1989a) and the NMR structures are, respectively, 0.92 ± 0.12 and 1.99 ± 0.12 Å (Table 2), indicating similar overall structural features (Figure 3C). However, the NMR and crystal structures do not overlap exactly and show some significant differences. The backbone (ϕ,ψ) dihedral angles (Figure 4) and H bonds (Table 3) characteristic of helices in the NMR structures parallel those in the crystal structures, but most of the side-chain H bonds in the crystal are not seen in the NMR structures. In particular, side-chain H bonds involving Arg12, Met17, Gln19, and Glu37 in the crystal, which were suggested to fix the carboxy terminus of helix 3 to the rest of the molecule (Mondragon et al., 1989a,b), were observed in none or very few of the NMR structures (Table 3). The lack of NOE constraints and the somewhat higher conformational flexibility of Arg12 and Met17 can partially explain this. It can also be the consequence of the χ_1 and fractional ASA of Gln19 and Glu37 in the crystal structure differing significantly from the NMR structures: (a) $\chi_{1,X-ray}$, $\chi_{1,NMR}$, and $\chi_{1,NMR}$ order parameters are −67°, −113 ± 13°, and 0.98, respectively, for Gln19, and −67°, −161 ± 24°, and 0.92,

respectively, for Glu37; (b) fractional ASA (X-ray) are 0.17 and 0.06 for Gln19 and Glu37, respectively, whereas the respective average fractional ASA (NMR) are 0.08 (range 0.05−0.12) and 0.22 (range 0.07−0.28) for Gln19 and Glu37. In general, the fractional ASA in the crystal for almost all other residues, especially the least solvent-accessible ones in the hydrophobic core, lie within or only slightly outside the ranges observed in the NMR structures (Figure 6C). Besides Gln19 and Glu37, other residues with well-defined side chains in the NMR structures whose χ_1 torsion angles differ from the X-ray value include Leu8 ($\chi_{1,NMR} = -169 \pm 15°$; $\chi_{1,X\text{-}ray} = 157°$), Leu22 ($\chi_{1,NMR} = 173 \pm 4°$; $\chi_{1,X\text{-}ray} = -157°$), Thr41 ($\chi_{1,NMR} = -52 \pm 4°$; $\chi_{1,X\text{-}ray} = -20°$), and, in particular, Phe46 ($\chi_{1,NMR} = 39 \pm 5°$; $\chi_{1,X\text{-}ray} = -52°$) and Leu54 ($\chi_{1,NMR} = -153 \pm 13°$; $\chi_{1,X\text{-}ray} = -66°$), the three Leu also being among the core residues with the lowest fractional ASA as discussed earlier. All the largest violations (≥1.5 Å) of the NMR distance constraints when applied to the X-ray structure involve at least one of the above residues whose side-chain torsion angles in the NMR structures differ significantly from those in the crystal (data not shown). As seen in Figure 5, the overall patterns of interhelical and other contacts are similar in the NMR and X-ray structures, but there are differences as reflected by the open squares in the up-diagonal region for X-ray structure contacts not observed in the NMR structures and vice versa in the down-diagonal region. Values for $\Delta ASA_{nonpolar}$, ΔASA_{polar}, and ΔASA_{total} obtained from the X-ray and the NMR structures for residues 1−65 (X-ray structure coordinates are available only for these structurally defined residues) listed in Table 2 indicate differences by at most 5% in $\Delta ASA_{nonpolar}$ and ΔASA_{total} values but significantly more in ΔASA_{polar} values. These, however, do not lead to appreciable differences in the predicted values for ΔC_p and *m*, the parameters describing folding transitions. The crystal structures of free 434 Cro and of each monomeric unit of the 434 Cro dimer−DNA co-crystal (Mondragon & Harrison, 1991) among themselves have global backbone and heavy atom RMSDs of 0.53 ± 0.05 and 1.48 ± 0.24 Å, respectively, for residues 4−63. DNA binding induces a few conformational changes in some of the protein residues in the DNA-binding or protein dimerization interface. However, the NMR structures of 434 Cro free in solution determined in this study show variations from the structure of each DNA-bound 434 Cro monomeric unit in the DNA−protein co-crystal which, more or less, parallel those seen with the free 434 Cro crystal structure.

Comparison of the Structures of 434 Cro and the 434 Repressor N-Terminal Domain R1−69. Residues 4−65 of the 71-residue 434 Cro protein and residues 2−63 of the 69-residue 434 repressor N-terminal fragment R1−69 show remarkable homology in their sequence and in their crystal structures with backbone RMSD < 0.8 Å (Mondragon et al., 1989b). This RMSD is comparable to that seen between the NMR structures of 434 Cro from its X-ray structure (Table 2) or between the average R1−69 NMR structure and its X-ray structure (the respective backbone and heavy atom RMSDs for residues 1−63 are 1.09 and 1.69 Å; Neri et al., 1992). We have listed in Table 2 the backbone RMSDs for residues 4−65 of our 434 Cro NMR structure versus residues 2−63 of the average R1−69 NMR structure and versus the R1−69 crystal structure for the sake of completion. The helices and other structural elements spanning the residues indicated are highly coincident in the two proteins, and they show similar patterns in the distributions of their (ϕ,ψ) angles and H bonds characteristic of helices. The C-caps Lys16 and Asn55 of helices 1 and 4 of 434 Cro whose ϕ and ψ angles are both positive have as their corresponding residues Gly in R1−69 also with both ϕ and ψ angles positive. In the two proteins, all the other helix C-caps are Gly and both have favorable but different N-cap residues for helices 1, 2, and 5. Arg12, Gln19, and Glu37 in 434 Cro are conserved in the equivalent positions in R1−69, and H bonds involving their side chains are detected in all but the 434 Cro NMR structures. Eleven of the non-Gly/Ala/Pro residues with well-defined side chains in the 434 Cro NMR structures have their equivalent residues in R1−69 well-defined in the R1−69 NMR structures, and six of these (Ile13, Gln19, Leu22, Ile33, Phe46, and Trp60 in 434 Cro) share the same identity in the two proteins. The side-chain conformation of Phe46 in the NMR structures of 434 Cro ($\chi_1 = 34 \pm 4°$) unlike in its crystal structure ($\chi_1 = -52°$) is closer to those in the R1−69 crystal ($\chi_1 = 55°$) and NMR structures ($\chi_1 = 54 \pm 7°$). Several of the hydrophobic core residues (fractional ASA ≤ 0.05) in the two proteins are in equivalent positions, and many of these (residues 22, 23, 33, 51, 54, and 61 in 434 Cro) share the same identity. The parallels in the overall packing of R1−69 and 434 Cro may also be appreciated from their ΔASA values listed in Table 2.

Folding and Stability of 434 Cro. Many small single-domain proteins lacking disulfide bonds and *cis* proline peptide bonds show two-state folding transitions, i.e., transitions involving only native and unfolded conformations without significant populations of intermediate conformations. Criteria used to ascertain two-state folding behavior include coincidence of equilibrium denaturation transitions monitored using different probes, the equivalence of the calorimetric and van't Hoff enthalpies, and the requirement that the unfolding free energy and its dependence on the denaturant estimated from equilibrium denaturation transitions is identical to that obtained from the kinetics of folding and refolding (Jackson & Fersht, 1991). For 434 Cro, equilibrium urea denaturation transitions monitored by far-UV CD, which is a probe of secondary structure, and by near-UV CD and fluorescence, which are probes of tertiary structure, yield superimposable curves (Figure 6D) and very similar estimates for the urea denaturation parameters (Table 4). The coincidence of equilibrium denaturation transitions monitored using the different spectroscopic probes demonstrates that under equilibrium NMR solution conditions 434 Cro folding is two-state.

Folding transitions of 434 Cro in the solution conditions used for the NMR experiments do not appear to be amenable to calorimetric measurements, as mentioned earlier. Measurements of the folding and unfolding kinetics of 434 Cro and their dependences on the denaturant concentration were not done, but 434 Cro most likely exhibits the very fast folding kinetics (in sub-milliseconds) observed for λ_{6-85} (which is very similar in size, structure. and sequence to 434 Cro) and possibly for the all α-helical 51-residue *lac* repressor headpiece (Huang & Oas, 1995b). The latter study led to the proposal that this rapid folding behavior may be a characteristic of all α-helical proteins. The kinetics of 434 Cro folding, whether the helical secondary structure forms simultaneously or otherwise with the tertiary structure for this one-domain, single hydrophobic core protein and

whether the individual helices are intrinsically stable, remain to be investigated. One or both of the capping residues of helices 1, 2, 3, and 5 in 434 Cro are favorable, as mentioned earlier, but all of the helices are short with several intrinsically helix-destabilizing residues and few helix-stabilizing Ala and Arg^+ residues. The intrinsic stabilities of the helices of 434 Cro and their correspondence to those in the native protein is currently being investigated in its synthetic water-soluble peptide fragments.

The two-state folding behavior under equilibrium conditions and the average m and ΔG_{water} values for 434 Cro from urea denaturation experiments parallel the behavior of the similar λ_{6-85} (Huang & Oas, 1995, 1996). 434 Cro as discussed earlier shows exceptional sequence and structural homology to R1−69, the monomeric N-terminal domain of the 434 repressor. Since the two proteins exhibit a high degree of correspondence in their structural elements and in their packing, it is highly likely that they show very similar folding behavior. NMR studies of the urea-denatured state of R1−69 showed residual structure in the segment 45−60 where the side chains of residues 54−59 appeared to form a hydrophobic cluster, and this was proposed as a possible folding nucleation site (Neri et al., 1992b). The corresponding segment in 434 Cro spans residues 47−62, and it is interesting to note that the majority of the slowest exchanging amide protons in 434 Cro (Table 4) are present within this region. Residual structure in the denatured protein if present in significant populations has implications in the thermodynamic analysis of the folding and stability of the protein. Whether any residual structure is present in urea-denatured 434 Cro and, if so, the populations, the locations in the protein, and the correspondence to that in R1−69 all remain to be examined.

The average m value for 434 Cro is on a per residue basis in the range observed for other small monomeric proteins. It is close to a value that can be predicted from studies on peptide helices as observed with other all-helical proteins consistent with the general conclusion that protein denaturation by urea is largely a consequence of the interactions between urea and peptide groups (Scholtz et al., 1995). Furthermore, as is observed for many proteins without disulfide cross-links which exhibit a simple two-state unfolding mechanism, the m value predicted from ΔASA for 434 Cro agrees reasonably well with that determined experimentally supporting the proposal that ΔASA is a main structural determinant of m values (Myers et al., 1995). The ΔG_{water} at pH 6.0 and at 20 °C (or 37 °C) is not very large for 434 Cro relative to other small proteins (Privalov & Gill, 1988) and indicates its modest stability under NMR solution conditions. λ_{6-85} has similar stability at this temperature and solution conditions at pH 8.0 but it is an isolated domain of a larger protein without the DNA-binding activity of the latter (Huang & Oas, 1996). Consequently, λ_{6-85} could lack some of the stabilizing interactions present in the whole protein. On the other hand, 434 Cro is the whole protein exhibiting the specific DNA-binding activity required for its biological regulatory function. Its modest stability under physiological conditions may therefore be among the factors modulating its *in vivo* concentration relative to that of the 434 repressor and consequently in the lysis versus lysogeny regulation of the life cycle of phage 434.

ΔG_{water} for 434 Cro agrees very well with ΔG_{op} for the most protected amide protons under the same solution conditions. Bai et al. (1994) report that $\Delta G_{op} = \Delta G_{water}$ if global unfolding controls the slowest exchanging amide protons: factors such as ^2H_2O in exchange experiments versus ^1H_2O in urea denaturation measurements, residual structure in the unfolded state which blocks amide proton exchange, and, most notably, the presence of *cis*-prolines in the native structure can make $\Delta G_{op} > \Delta G_{water}$, whereas $\Delta G_{op} < \Delta G_{water}$ if exchange is controlled by nonglobal, local unfolding processes or if residual structure in the unfolded state does not block amide proton exchange. Although the observed equivalence of ΔG_{op} and ΔG_{water} for 434 Cro may be the result of the canceling out of the different factors mentioned above, a simpler explanation is that global unfolding governs the exchange of the most solvent-protected amide protons in 434 Cro. This is consistent with what has generally been observed with other single-domain proteins which either do not have *cis*-prolines in their native structure like 434 Cro or in which the *cis*−*trans* proline isomerization effects on amide proton exchange have been accounted for. The close correspondence between ΔG_{water} from urea denaturation and ΔG_{op} for the most protected amide protons under the same solution conditions for 434 Cro also appears to validate the use of the linear extrapolation method (LEM) to determine free energies of unfolding from chemical denaturation.

In summary, the 434 Cro NMR resonance assignments and solution structure obtained in this study are the first steps for examining its structure and those of its recently reported variants (Desjarlais & Handel, 1995) in different solution conditions as well as for a structural characterization of its DNA binding in solution. Moreover, the structural similarity of 434 Cro to several other DNA-binding gene-regulatory proteins makes comparative studies of this family of helical globular proteins relevant both in understanding their folding as well as the basis for their DNA-binding specificities. We have characterized the thermodynamic stability of 434 Cro in NMR solution conditions, and further studies characterizing the folding of 434 Cro, and the structures and stabilities of its synthetic peptide fragments using NMR and other spectroscopic methods are currently in progress.

ACKNOWLEDGMENT

S.P. thanks Dr. S. Subbiah (Stanford University) for suggesting 434 Cro as a model system and for his initial help, Professor Steve Harrison (Harvard University) and, in his laboratory, Dr. David Rodgers and Marie Drottar for the gift of plasmids and purification protocols. We thank Dr. Montserrat Elías-Arnanz (CBM-Madrid), Dr. Javier Varela (CIB-Madrid), Dr. J. Santoro, Dr. M. Bruix, Eva de Alba, D. García, C. López (IEM-CSIC, Madrid), Dr. M. Gasset, and Dr. M. Menéndez (Madrid) for help and Dr. D. V. Laurents (Stanford University) for very useful discussions.

SUPPORTING INFORMATION AVAILABLE

Table listing the ^1H chemical shift assignments for 434 Cro (4 pages). Ordering information is given on any masthead page.

REFERENCES

Alexander, P., Fahnestock, S., Lee, T., Orban, J., & Bryan, P. (1992) *Biochemistry 31,* 3597−3603.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229−2246.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
Bai, Y., Milne, J. S., Mayne, L., & Englander, S. W. (1994) *Proteins: Struct., Funct., Genet. 20*, 4−14.
Baldwin, R. L. (1995) *Biophys. Chem. 55*, 127−195.
Bartels, C., Xia, T., Billeter, M., Güntert, P., & Wüthrich, K. (1995) *J. Biomol. NMR. 6*, 1−10.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859−1877.
Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10*, 1195−1207.
Desjarlais, J. R., & Handel, T. M. (1995) *Protein Sci. 4*, 2006−2018.
Doig, A. D., & Baldwin, R. L. (1995) *Protein Sci. 4*, 1325−1336.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521−655.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
Güntert, P., & Wüthrich, K. (1991) *J. Biomol. NMR 1*, 447−456.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
Güntert, P., Berndt, K. D., & Wüthrich, K. (1993) *J. Biomol. NMR 3*, 601−606.
Harrison, S. C., & Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 933−969.
Holzwarth, G., & Doty, P. (1965) *J. Am. Chem. Soc. 87*, 218 −228.
Huang, G. S., & Oas, T. G. (1995a) *Biochemistry 34*, 3884−3892.
Huang, G. S., & Oas, T. G. (1995b) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
Huang, G. S., & Oas, T. G. (1996) *Biochemistry 35*, 6173−6180.
Hyberts, S. G., Goldberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.
Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
Koradi, R., Billeter, M., & Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biophys. Biochem. Res. Commun. 95*, 1−6.
Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Sciences* (Harding, S. E., Rowe, A. J., & Horton, J. C., Eds.) pp 90−125, Royal Society of Chemistry, Cambridge, U.K.
Livingstone, J. R., Spolar, R., & Record, M. T., Jr. (1991) *Biochemistry 30*, 4237−4244.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Mondragon, A., & Harrison, S. C. (1991) *J. Mol. Biol. 219*, 321−334.
Mondragon, A., Wolberger, C., & Harrison, S. C. (1989a) *J. Mol. Biol. 205*, 179−188.
Mondragon, A., Subbiah, S., Almo, S. C., Drottar, M., & Harrison, S. C. (1989b) *J. Mol. Biol. 205*, 189−200.
Murphy, K. P., & Freire, E. (1992) *Adv. Protein Chem. 43*, 313−361.
Myers, J. K., Pace, C. N., & Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
Nagarajaram, H. A., Sowdhamini, R., Ramakrishnan, C., & Balaram, P. (1993) *FEBS Lett. 321*, 79−83.
Neri, D., Billeter, M., & Wüthrich, K. (1992a) *J. Mol. Biol. 223*, 743−767.
Neri, D., Billeter, M., Wider, G., & Wüthrich, K. (1992b) *Science 257*, 1559−1563.
Pace, C. N., Shirley, B. A., Thomson, J. A. (1990) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 311−330, IRL Press, Oxford, England.
Piotto, M., Saudek, V., & Sklenar, V. (1993) *J. Biomol. NMR 2*, 661−665.
Powers, R., Garrett, D. S., March, C. J., Frieden, E. A., Gronenborn, A. M., & Clore, G. M. (1992) *Biochemistry 31*, 4334−4346.
Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1648−1652.
Privalov, P. L., & Gill, S. J. (1989) *Adv. Protein Chem. 39*, 191−234.
Ptashne, M. (1986) *A Genetic Switch: Gene Control and Phage λ*, Cell Press and Blackwell Scientific Publications, Cambridge, MA.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648−1652.
Richardson, J. S., & Richardson, D. C. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 1−98, Plenum Press, New York.
Richmond, T. J. (1984) *J. Mol. Biol. 178*, 63−89.
Rico, M., Jiménez, M. A., González, C., De Fillippis, V., & Fontana, A. (1994) *Biochemistry 33*, 14834−14847.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd ed., Vol. 3, pp A-2, Cold Spring Harbor Laboratory Press, Plainview, NY.
Santoro, J., González, C., Bruix, M. A., Neira, J. L., Nieto, J. L., Herranz, J., & Rico, M. (1993) *J. Mol. Biol. 229*, 722−734.
Santoro, M. M., & Bolen, D. W. (1988) *Biochemistry 27*, 8063−8068.
Schellman, C. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 53−61, Elsevier, New York.
Scholtz, J. M. (1995) *Protein Sci. 4*, 35−43.
Scholtz, J. M., Barrick, D., York, E. J., Stewart, J. M., & Baldwin, R. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 185−189.
Sevilla-Sierra, P., Otting, G., & Wüthrich, K. (1994) *J. Mol. Biol. 235*, 1003−1020.
Spolar, R. S., Livingstone, J. R., Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
Stonehouse, J., & Keeler, J. (1995) *J. Magn. Reson., Ser. A 112*, 43−57.
van Gunsteren, W. F., & Berendsen, H. J. C. (1987) in *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
Wishart, D. S., Bigam, C. G., Holm, A,. Hodges, R. S., & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zazo, M., Lozano, R. M., Ortega, S., Varela, J., Díaz-Orejas, R., Ramírez, J. M., & Giménez-Gallego G. (1992) *Gene 113*, 231−238.

BI970085P