Folding propensities of synthetic peptide fragments covering the entire sequence of phage 434 Cro protein

S. PADMANABHAN, M. ANGELES JIMÉNEZ, AND MANUEL RICO

Instituto de Estructura de la Materia, Consejo Superior de Investigaciones Científicas, Serrano, 119, 28006-Madrid, Spain

(RECEIVED December 30, 1998; ACCEPTED May 10, 1999)

Abstract

The phage 434 Cro protein, the N-terminal domain of its repressor (R1–69) and that of phage λ (λ_{6-85}) constitute a group of small, monomeric, single-domain folding units consisting of five helices with striking structural similarity. The intrinsic helix stabilities in λ_{6-85} have been correlated to its rapid folding behavior, and a residual hydrophobic cluster found in R1–69 in 7 M urea has been proposed as a folding initiation site. To understand the early events in the folding of 434 Cro, and for comparison with R1–69 and λ_{6-85} , we examined the conformational behavior of five peptides covering the entire 434 Cro sequence in water, 40% (by volume) TFE/water, and 7 M urea solutions using CD and NMR. Each peptide corresponds to a helix and adjacent residues as identified in the native 434 Cro NMR and crystal structures. All are soluble and monomeric in the solution conditions examined except for the peptide corresponding to the 434 Cro helix 4, which has low water solubility. Helix formation is observed for the 434 Cro helix 1 and helix 2 peptides in water, for all the peptides in 40% TFE and for none in 7 M urea. NMR data indicate that the helix limits in the peptides are similar to those in the native protein helices. The number of side-chain NOEs in water and TFE correlates with the helix content, and essentially none are observed in 7 M urea for any peptide, except that for helix 5, where a hydrophobic cluster may be present. The low intrinsic folding propensities of the five helices could account for the observed stability and folding behavior of 434 Cro and is, at least qualitatively, in accord with the results of the recently described diffusion-collision model incorporating intrinsic helix propensities.

Keywords: 434 Cro; helices; peptide conformation; peptide NMR; protein folding

The folding of a protein to its native, biologically active conformation remains an open question. One view holds that folding proceeds through defined pathways with discrete intermediates (Baldwin, 1989; Kim & Baldwin, 1990), while a more recent view is that many unstructured conformations coalesce or funnel into smaller populations of different partly folded conformations including secondary structures such as α -helices and compact structures resulting from hydrophobic interactions (Bryngelson et al., 1995; Wolynes et al., 1995; Dill & Chan, 1997). Thus, from either

viewpoint, intrinsically stable helices, which can form in tens of nanoseconds (Williams et al., 1996; Thompson et al., 1997), can populate conformational ensembles not only of the transition state but those preceding it, as has been proposed in the microsecond timescale folding of λ_{6-85} , the small 80-residue α -helical N-terminal domain of the λ repressor, and some of its variants (Burton et al., 1997, 1998). Moreover, local interactions, such as those involved in the formation of secondary structural elements like α -helices, are thermodynamically important in protein stability (Muñoz & Serrano, 1996), and possibly so in increasing the speed of folding (Viguera et al., 1996; Burton et al., 1998; Plaxco et al., 1998). It is therefore important to know the intrinsic stabilities of secondary structures and the local interactions affecting their formation. These are most clearly examined under conditions where long-range tertiary interactions are absent as would be the case in peptide fragments. Peptide fragments of proteins and synthetic designed peptides have proved very successful in understanding helix formation and stability (see, for example, Baldwin, 1995), and helical propensities of peptide fragments of several proteins have been examined to understand the earliest events in the folding of these proteins. Such studies include, among others, families of peptides that together constitute the entire sequence of the all helical proteins myohemerythrin (Dyson et al., 1992), myoglobin

Reprint requests to: S. Padmanabhan, Instituto de Estructura de la Materia-CSIC, Serrano 119, 28006 Madrid, Spain; e-mail: empadhu@rocky2.roca.csic.es.

Abbreviations: 1D, one-dimensional; 2D, two-dimensional; CD, circular dichroism; COSY, scalar coupled correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; Fmoc, 9-fluorenyl methoxycarbonyl; HPLC, high-performance liquid chromatography; $^3J_{\text{HN}\alpha}$, amide proton- α -proton vicinal spin-spin coupling constant; λ_{6-85} , the N-terminal domain of the λ repressor excluding the first five residues; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement and exchange spectroscopy; OPfp, pentafluorophenyl; R1–69, the N-terminal DNA-binding domain of the 434 repressor; ROESY, rotating frame exchange spectroscopy; TFA, 2,2,2-trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; TFE- d_3 , deuterated trifluoroethanol; TOCSY, total correlation spectroscopy; TSP, 3-trimethyl-silyl(2,2,3,3'- $^2\text{H}_4$)-propionate.

(Reymond et al., 1997), and the C-terminal, helical domain of thermolysin (Jiménez et al., 1993; Rico et al., 1994). The present study examines synthetic peptide fragments that together cover the entire sequence length of another helical protein—the bacteriophage 434 Cro.

Cro and repressor proteins regulate the lytic-lysogenic switch in the life cycle of many phages by sequence-specific DNA-binding (Ptashne, 1986). The 71-residue Cro and the 69-residue DNA-binding N-terminal domain of the repressor (R1–69) of phage 434, as well as λ_{6-85} mentioned earlier, are small, highly water-soluble monomers with remarkably homologous structures consisting of five helices (Mondragon et al., 1989a; Padmanabhan et al., 1997). They are devoid of disulfides or prosthetic groups and make excellent model systems for individual and comparative physical and structural studies of both folding and DNA-binding. The homology in the sequence and in the secondary structure among the three proteins is illustrated in Figure 1. A two-state folding behavior without the accumulation of folding intermediates has been demonstrated for 434 Cro (Padmanabhan et al., 1997) and for λ_{6-85} (Huang & Oas, 1995a) and is probably the case for R1–69. λ_{6-85} folds in the submillisecond timescale, and its helical nature and the intrinsic stabilities of its helices have been proposed to play important roles in determining its rapid folding and its transition state conformations (Huang & Oas, 1995b; Burton et al., 1997, 1998). In R1–69, a hydrophobic cluster that persists in 7 M urea has been suggested as a folding nucleation site (Neri et al., 1992). However, to our knowledge, the conformations of only two peptide fragments for this group of proteins have been previously examined: a peptide corresponding to helix 1 of λ_{6-85} in aqueous solutions (Marqusee & Sauer, 1994), and a peptide corresponding to residues 44–63 of R1–69 in 7 M urea solutions (Neri et al., 1992). Consequently, we chose to examine the conformations of five synthetic peptides covering the entire sequence of 434 Cro using

CD and NMR to understand the early folding of this protein and to compare it with R1–69 and λ_{6-85} . The peptides are examined in aqueous solutions at pH and salt concentrations used in the NMR structure and stability measurements of the whole native protein. They are studied in 7 M urea where the urea denaturation of the native protein is complete (Padmanabhan et al., 1997) but where residual structure has been reported for the homologous R1–69 as mentioned earlier (Neri et al., 1992). The helix contents of most peptide fragments of proteins increase in solutions containing trifluoroethanol (TFE), a solvent considered to mimic the hydrophobic, solvent-excluded environment of the collapsed state where tertiary interactions are absent and which may form early in the protein folding pathway (Nelson & Kallenbach, 1986, 1989; Lehman et al., 1990; Segawa et al., 1991; Dyson et al., 1992; Sonnichsen et al., 1992; Jiménez et al., 1993; Hamada et al., 1995; Kemmink & Creighton, 1995a; Muñoz et al., 1995b; Yang et al., 1995; Reymond et al., 1997; Hirota et al., 1998; Myers et al., 1998). The peptide helix contents typically reach a maximum by about 40% (by volume) TFE/water, a condition where the intrinsic amino acid helix propensities were recently determined (Rohl et al., 1996). Consequently, we also examined the 434 Cro peptide fragments in 40% TFE.

Results

Peptide design, solubility, and monomeric nature

The peptides 1Cro, 2Cro, 3Cro, 4Cro, and 5Cro (Table 1) correspond to residues 1–16, 16–29, 29–43, 42–57, and 51–70, respectively, in 434 Cro and so include residues constituting helices 1, 2, 3, 4, and 5 in the native protein as seen in the X-ray (Mondragon et al., 1989b) and NMR structures (Padmanabhan et al., 1997) (Table 1; Fig. 1). Residues flanking the termini of each native protein helix were also included in the corresponding peptide to minimize helix end-fraying effects and to maximize the number of charged or polar groups for enhanced water-solubility of the peptide. The N-terminus is unblocked in all peptides. The C-terminal carboxyl in 1Cro, 2Cro, 3Cro, or 5Cro is amidated but free in 4Cro, since the stretch of hydrophobic residues in the native 4Cro sequence could cause poor water solubility. Furthermore, 4Cro contains several of the N-terminal polar and charged residues flanking helix 4 in the native protein, and the Ser replaces the native Cys to avoid complications arising from oxidation of Cys–SH group. The nonnative Tyr included at the N- or the C-terminus of peptides 1Cro, 2Cro, and 3Cro, the Tyr in 4Cro, which replaces the native Phe in 4Cro, and the native Trp and Tyr in 5Cro serve in the accurate determination of peptide concentration by UV absorbance (see Materials and methods). 5Cro contains a large number of the residues flanking helix 5 in the native sequence (including some present in the C-terminal region of 4Cro) to mirror the structurally homologous region in R1–69 (see Fig. 1), which is reported to have residual structure in 7 M urea (Neri et al., 1992). The C-terminus of 5Cro contains all but the final Ala of the C-terminus of native 434 Cro where they are structurally disordered (Mondragon et al., 1989b; Padmanabhan et al., 1997). To avoid oxidation of the native Cys retained in 5Cro, it was always examined in the presence of 1 mM DTT.

1Cro, 2Cro, 3Cro, and 5Cro are water soluble with no visible aggregation even at NMR concentrations (1–2 mM) and solution conditions used in this study. Their monomeric nature in aqueous solutions was confirmed by: (1) the thermally reversible,

Fig. 1. Sequences of 434 Cro, R1–69, and λ_{6-85} showing the correspondence between their helical regions (shown by boxes with thick lines with helix number indicated at the bottom). Residue numbers for each protein stretch are shown on the right and the vertical bars indicate matching residues in two or in all three proteins.

Table 1. Peptide sequences and CD data

Peptide name	Sequence ^a	[Θ] ₂₂₂ (deg cm ² dmol ⁻¹) ^b		
		H ₂ O ^c	40% TFE ^d	7 M urea ^e
1Cro	MQTL SERLKKRR IALKY*-amide	-6,800 ± 400	-16,300 ± 300	1,200 ± 200
2Cro	Y*KMT QTELATKA GVK-amide	-3,100 ± 300	-14,000 ± 800	-75 ± 50
3Cro	Y*K QQS IQ LIEA GVTKR-amide	-1,600 ± 300	-16,200 ± 1,000	-45 ± 50
4Cro ^f	TKRPRFLY* ET AMALNS*D	—	-9,200 ± 200	670 ± 200
5Cro	AMALNCDPVWLQY GTKRGKA-amide	-1,800 ± 500	-10,600 ± 300	175 ± 50

^aResidues in boldface and underlined are helical in the native 434 Cro protein. Tyr followed by an asterisk in peptides 1Cro, 2Cro, and 3Cro is absent in the native sequence and replaces the native Phe in peptide 4Cro. Ser followed by an asterisk in peptide 4Cro replaces the native Cys. No peptide is blocked at the N-terminus.

^bAll solutions contain 100 mM KCl, 25 mM KH₂PO₄ (pH 6.0), 1 mM DTT, and CD measurements are at 2 °C.

^c100% water.

^d40% TFE:water (volume:volume).

^e7 M urea.

^fPoor water solubility.

concentration-independent (10–300 μ M peptide) mean molar residue ellipticity values at 222 nm, [Θ]₂₂₂ (in deg cm² dmol⁻¹), in far-UV CD measurements (see below); (2) the agreement between the apparent molecular mass (M_r) determined in analytical ultracentrifugation measurements at NMR peptide concentrations ($\approx$ 1 mM) and the molecular mass (M) estimated from the sequence or by mass spectrometry: $M_r/M = 1.003, 0.969, 1.032$, and 0.903 for 1Cro, 2Cro, 3Cro, and 5Cro, respectively (data not shown). 4Cro has poor water solubility despite the sequence modifications discussed above and reliable data for this peptide could not be obtained in water solutions. It is, however, soluble and monomeric in 40% TFE as well as in 7 M urea, as suggested by its concentration-independent far-UV CD signal and narrow NMR signals.

Far UV-CD: Data and analysis

Figure 2A shows the CD spectra of the peptides in NMR buffer (pH 6.0, 100 mM KCl, 25 mM KH₂PO₄, 1 mM DTT, 0.01% NaN₃, as used in native 434 Cro solution structure and stability studies; Padmanabhan et al., 1997) and at 2 °C, the lower temperature being used to stabilize helix formation. Two minima, one around 206 nm and the other at 222 nm characteristic of helices (Holzwarth & Doty, 1965), are observed for 1Cro. The 222 nm minimum is considerably smaller for 2Cro and not discernible for 3Cro and 5Cro in water solutions. The strong 222 nm minimum and the appearance of the maximum at 192 nm characteristic of helices indicate significant helix formation in all the peptides in 40% TFE (Fig. 2B). Table 1 lists [Θ]₂₂₂ for each peptide and solution condition, and Table 2 the corresponding helix contents estimated from [Θ]₂₂₂ (see Materials and methods). Aromatic residues (Phe, Tyr, Trp) can contribute to the measured [Θ]₂₂₂ (Chakrabartty et al., 1993), but quantitative relationships between this contribution and helix content and length, sequence position, and orientation of the aromatic group are not known. An approximate correction that assumes a linear relation between the aromatic contribution and helix content (as in Chakrabartty et al., 1994) yields the helix contents listed parenthetically in Table 2. The helix contents predicted for the peptides studied here using algorithms based on the modified Lifson-Roig (mL-R) (Rohl et al., 1996), and AGADIR (Muñoz & Serrano, 1997) models for helix-coil transi-

tions are also listed in Table 2. Both algorithms were used for predictions in water solutions, but only the mL-R algorithm for those in 40% TFE since the relevant parameters are available only for this model (Rohl et al., 1996).

Fig. 2. CD spectra of peptides 1Cro, 2Cro, 3Cro, 4Cro, and 5Cro at 2 °C, pH 6.0 in NMR buffer. **A:** Water. **B:** 40% TFE:water (by volume).

Table 2. Peptide helix contents from CD and NMR and those predicted using AGADIR and modified Lifson-Roig algorithms^a

Peptide	H ₂ O ^b				40% TFE ^b			7 M Urea ^b	
	CD	NMR	mL-R	AGADIR	CD	NMR	mL-R ^c	CD	NMR
1Cro	21 (26)	29	8	22	48 (59)	52	33	0	4
2Cro	11 (13)	14	<1	3	42 (51)	31	13	2	1
3Cro	6 (7)	5	<1	1	49 (58)	45	16	2	3
4Cro ^d	—	—	10	5	28 (38)	30	29	0	3
5Cro	7 (8)	17	<1	<1	31 (36)	21	1	0	7

^a% Helix at 2 °C estimated as described in Materials and methods; in parentheses are the CD values obtained after correction for aromatic contributions (see text).

^bColumn headings H₂O, 40% TFE, and 7 M urea are, respectively, solutions made up in 100% H₂O (CD) or 9:1 H₂O: D₂O (NMR), 40% (by volume) TFE:water (CD) or 40% TFE-d₃:water (by volume) (NMR), and 7 M urea in 100% H₂O (CD) or 9:1 H₂O: D₂O (NMR), with all solutions containing 100 mM KCl, 25 mM KH₂PO₄ (pH 6.0), 1 mM DTT.

^cItalicized to indicate that side-chain interactions were not considered in these calculations.

^dPoor water solubility.

Helix formation in water, 40% TFE and 7 M urea solutions from NMR data

2D-¹H NMR COSY, DQF-COSY, TOCSY, NOESY, and ROESY spectra were recorded at 2 °C and pH 6.0 in NMR buffer for all peptides in water (except for the poorly water soluble 4Cro), in 40% TFE and in 7 M urea. The general quality of the NMR data

in the three solution conditions is illustrated in representative 2D-NOESY spectra obtained for 1Cro (Fig. 3). The spectral dispersion is reasonably good and, in general, improves with increasing helix content. Thus, a complete assignment of the ¹H-NMR resonances could be achieved for each peptide and condition studied using the standard strategy (Wüthrich, 1986). (¹H NMR assignments are tabulated in Supplementary material in Electronic Appendix.)

Fig. 3. 2D-NOESY spectra (mixing time 150 ms) of peptide 1Cro at 2 °C, pH 6.0 in NMR buffer containing 10% by volume ²H₂O and TSP. **A:** Water. **B:** 40% TFE-d₃. **C:** 7 M urea.

NOESY and ROESY cross-peak patterns are identical for the mixing times used confirming the absence of spin-diffusion contributions. The number of NOEs increases with helix content, being maximum in 40% TFE and least in helix-destabilizing 7 M urea (Fig. 4). Differences in solvent, peptide concentrations, and conformational averaging exclude any quantitative comparisons of NOE intensities, qualitatively represented as strong, medium, or weak in Figure 4. The pattern of intensities found for the sequential NOEs together with those corresponding to medium-range ($i, i + 3$) and ($i, i + 4$) NOEs indicate significant populations of α -helical conformations for 1Cro in water solutions (Fig. 4), and the few $\alpha N(i, i + 2)$ NOEs observed for residues at or close to the peptide N- and C-termini probably reflect nascent or partial 3_{10} -helical conformation in these regions (Dyson & Wright, 1991). Few of the medium-range NOEs are observed in water solutions of 2Cro and virtually none for 3Cro, while for 5Cro the observed $\alpha N(i, i + 2)$ and $\alpha N(i, i + 3)$ NOEs of the Pro8 α -proton and the $\alpha N(i, i + 3)$ NOE of the Val9 α -proton may suggest a nascent helical character. We observed a large number of medium-range NOEs for all the peptides in 40% TFE (including all of those identified in water) consistent with significantly enhanced helix formation in 40% TFE. Essentially no medium-range NOEs were observed in 7 M urea indicating absence of backbone conformations in the α region of the (ϕ, ψ) space. Pro4 in 4Cro and Pro8 in 5Cro are both predominantly *trans* in all the conditions examined, as shown by the NOEs between the Pro δ -protons and the α -proton of the preceding residue—Arg3 in 4Cro and Asp7 in 5Cro.

$^3J_{HN\alpha}$ (the amide proton- α -proton vicinal spin-spin coupling constant) is related to the backbone dihedral angle ϕ and a stretch of at least three successive residues with $^3J_{HN\alpha} < 6.0$ Hz indicates helical conformations (Wüthrich, 1986). ($^3J_{HN\alpha} \approx 4.8, 5.6$, and 8.5 Hz, respectively, for α -helical, 3_{10} -helical, and extended β -sheet residues in a recent report; Smith et al., 1996). $^3J_{HN\alpha}$ were estimated from 1D spectra with high digital resolution if possible or from 2D spectra as described in Materials and methods. The values reflect an ensemble average and their accuracy is limited by the digital resolution (1–1.5 Hz) in the 2D-NMR experiments and so were used only qualitatively (summarized in Fig. 4 for each residue and peptide except Gly, Pro, and those resonances that overlap in the 2D spectra or have poor low signal-to-noise). $^3J_{HN\alpha}$ is ≥ 6 Hz for most residues in 7 M urea, and for those close to the N- or C-termini in water or 40% TFE indicating that these residues are not in a helical conformation. In water solutions, helical structure in 1Cro and 2Cro is suggested by stretches of residues with $^3J_{HN\alpha} \leq 5$ Hz, but not in 3Cro where $^3J_{HN\alpha} \geq 6$ Hz for all residues. In 5Cro, $^3J_{HN\alpha}$ could not be estimated for the central stretch of residues in any of the solution conditions used, but is generally ≥ 6 Hz for some other residues. Stretches of residues with $^3J_{HN\alpha} \leq 5$ Hz are observed in 40% TFE for 1Cro, 2Cro, 3Cro, and 4Cro consistent with helical structure.

The presence of continuous stretches of residues with α -CH resonances shifted upfield relative to random coil values provide additional evidence for helical conformation (Wishart et al., 1991, 1992). Figure 5 shows these upfield shifts (CSI values) for each peptide and solution condition, and the corresponding ones in the native protein (from NMR assignments in Padmanabhan et al., 1997) were estimated using C $^{\alpha}$ H random coil chemical shift values determined in GG-X-GG peptides at 4 °C and pH 5.0 (Merutka et al., 1995) conditions that most closely approximate those used in this study (2 °C, pH 6.0). (Other model peptides and/or solution conditions, e.g., various TFE concentrations (Merutka et al., 1995);

25 °C, pH 5.0, 1 M urea and at least 50 mM salt (Wishart et al., 1995); 30 °C and pH 7.0 (Bundi & Wüthrich, 1979), report similar random coil chemical shifts for C $^{\alpha}$ H and side chain, but not amide protons, indicating that these are negligibly affected by temperature, urea, or TFE.) Random coil C $^{\alpha}$ H values of 4.65 ppm (not 4.34 ppm) for the Arg preceding Pro in 4Cro, and 4.90 ppm (not 4.63 ppm) for Asp preceding Pro in 5Cro were used, since Pro (relative to Gly or Ala) causes downfield shifts in the random coil C $^{\alpha}$ H values of its preceding residue by, on average, 0.29 ppm (Wishart et al., 1995). The N-terminal residue in all the peptides and solution conditions and the C-terminal residue in 4Cro (the only one with a free C-terminal carboxyl) have large negative CSI possibly due to charge effects. For all peptides in 7 M urea, CSI is within ± 0.05 ppm for almost every residue and is probably non-zero because of sequence rather than structure-based effects. 4Cro and 5Cro in 7 M urea have CSI $> +0.1$ ppm for some residues close to the aromatic residues and so susceptible to ring current effects. Relative to those in 7 M urea, CSI in water solutions are very similar in 3Cro and 5Cro, slightly more negative in 2Cro and significantly more so in 1Cro indicating that 1Cro, and to a smaller extent, 2Cro are helical in water solutions but not 3Cro or 5Cro. In all peptides in 40% TFE, large negative CSI are observed consistent with the high helix contents. Helix contents estimated using the upfield C $^{\alpha}$ H shifts (see Materials and methods) and from CD are in overall agreement (Table 2): the largest discrepancies occur for 5Cro or in 40% TFE possibly due to aromatic side-chain effects on both the C $^{\alpha}$ H and the CD signals, the uncertainties associated with these effects increasing with helix content and the number of aromatic groups.

Helical residues for the peptides studied here lie within the following residue limits (depicted at the top of each panel in Fig. 5): Gln2–Tyr17 in 1Cro, Thr4–Gly13 in 2Cro, Lys2–Gly12 in 3Cro, Pro4–Asn15 in 4Cro, and Pro8–Gly14 in 5Cro. This assignment was based on the following combined NMR criteria: (1) residues involved in a continuous stretch of $\alpha N(i, i + 3)$, $\alpha N(i, i + 4)$, $\alpha \beta(i, i + 3)$, or $NN(i, i + 2)$ NOEs are helical with the first and last residues with these NOE connectivities defining the helix limits; (2) a continuous stretch of residues with $^3J_{HN\alpha} \leq 5$ Hz are helical, the first and last residues of the stretch constituting the helix limits; (3) a continuous stretch of residues with negative CSI form a helix whose boundary residues have CSI close to zero or positive. The assignment was largely based on NMR data obtained in 40% TFE solutions where helix contents are greatest, and because some of the peptides show little or no helix formation in water solutions. The nature of the helices in the two solution conditions is similar since all of the medium-range NOEs characteristic of helices observed in water solutions are also seen in 40% TFE, in accord with the conclusion that TFE enhances helix contents without altering helix limits (Nelson & Kallenbach, 1989).

Side-chain interactions from NMR data

The presence of NOEs between two side chains indicates proximity between them in a significant fraction of the molecules and for a significant fraction of the time. This hints at possible interactions between the two side chains and the consequent existence of non-random structure. The observed nonsequential main-chain–side-chain or side-chain–side-chain ^1H – ^1H NOEs are shown in Figure 4 (a table in Supplementary material in Electronic Appendix also lists side-chain–side-chain NOEs observed in the different condi-

A

B

C

Fig. 4. Figure continues on facing page.

tions and the corresponding NOE if identified in the native protein NMR structure). Side-chain NOEs are maximum for all the peptides in 40% TFE where the helix contents are maximum. In water solutions, there are no side-chain NOEs for 3Cro, a few for the slightly helical 2Cro, and several more for the more helical 1Cro. Most side-chain NOEs observed in water are also seen in 40% TFE except for 5Cro. Some, though not all, of the medium-range side-chain–side-chain NOEs in 40% TFE correspond to those in the native protein. Nonsequential side-chain NOEs were essentially absent in 7 M urea consistent with overall random structure for all the peptides except 5Cro. The latter shows a significant number of side-chain NOEs in water and 7 M urea solutions but absent in 40% TFE or in the native protein. This suggests that 5Cro adopts similar nonrandom and (from NMR and CD data) nonhelical struc-

tures in water or in 7 M urea different from those in 40% TFE or in the native protein.

Discussion

Structural propensities of 434 Cro peptide fragments

The α -helix is the major secondary structural element in 434 Cro. The intrinsic stability of each of its five native helices was evaluated by using CD and NMR to examine the structural propensities in five synthetic peptides—1Cro, 2Cro, 3Cro, 4Cro, and 5Cro corresponding to the native protein helices 1, 2, 3, 4, and 5, respectively, and together covering the entire protein sequence. The peptides are all monomeric in the solution conditions employed so

Fig. 4 (continued from facing page). Summary of NMR data in water, 40% TFE- d_3 , and 7 M urea solutions at 2 °C, pH 6.0 in NMR buffer. Sequential C α H–NH ($d_{\alpha N}$), NH–NH (d_{NN}), and C β H–NH ($d_{\beta N}$) NOEs (mixing time, 150 ms) shown as thick, medium, and thin bars indicate strong, medium, and weak NOE connectivities. Medium-range C α H–NH ($d_{\alpha N}(i, i + 2)$, $d_{\alpha N}(i, i + 3)$, $d_{\alpha N}(i, i + 4)$), C α H–C β H ($d_{\alpha\beta}(i, i + 3)$), and NH–NH ($d_{NN}(i, i + 2)$), and side-chain–side-chain ($dsch(i, i + 3)$, $dsch(i, i + 4)$) NOE connectivities for residues $i, i + 2, i + 3$, and $i + 4$ are indicated. $^3J_{HN\alpha}$ (closed circles ≤ 5 Hz, open circles ≥ 6 Hz) is the amide–C α H proton–proton vicinal spin–spin coupling constant. **A:** 1Cro. **B:** 2Cro. **C:** 3Cro. **D:** 4Cro. **E:** 5 Cro.

Fig. 5. The difference between the observed (δ_{obs}) and random coil (δ_{rc}) $C^\alpha H$ chemical shift (in ppm) for each residue in the peptides in water (open circles, "H₂O"), 40% TFE-d₃/water (open squares, "TFE"), or 7 M urea in water (crosses, "Urea"), and for the native protein at 20°C (filled circles, "Native"). Other conditions are the same as in Figures 3 and 4. Abscissa show peptide sequence with nonnative residues underlined and followed by an asterisk. Lines connecting the symbols are to aid in viewing, the dotted line representing zero shifts. Helix limits based on NMR data (see text) are indicated at the top of each panel by the dashed line for the peptide and by the line with arrowheads for the native protein. **A:** 1Cro. **B:** 2Cro. **C:** 3Cro. **D:** 4Cro. **E:** 5 Cro.

that any structure formation in them is intrinsic without involving long-range tertiary interactions. The peptides containing 434 Cro helices 1 and 2 but not 3 and 5 show helix formation in water solutions (the helix 4 peptide is not water soluble). Isolated peptide helices in water solutions are typically exposed to the solvent water molecules and so to hydrogen bonding interactions, which compete with those involved in helix formation. On the other hand, individual helices in proteins could be intrinsically stable in a solvent-excluded environment such as in the hydrophobically collapsed state, which, according to some models, forms early in the folding pathway and where specific tertiary interactions are absent. The cosolvent TFE has frequently been used to mimic such a solvent-excluded environment free of specific tertiary interactions, and all the five 434 Cro peptides show significant helix formation

in 40% TFE. By contrast, no helix formation is observed for any of these peptides in helix-destabilizing 7 M urea.

Algorithms based on models for helix-coil transitions, which incorporate various types of amino acid side-chain interactions known to affect helices (Rohl et al., 1996; Muñoz & Serrano, 1997), predict helix contents that agree with the relative order of helix forming tendencies observed experimentally. Thus, both models predict the 434 Cro helix 1 to have the highest intrinsic stability in water solutions as is observed. The predicted intrinsic stability of helix 4 in water solutions, though not verifiable experimentally because the corresponding peptide is not water soluble, is consistent with the observation that several residues in this helix in the native protein have amides with the slowest exchange with solvent (Padmanabhan et al., 1997). The experimental and predicted helix

contents agree less at a quantitative level, the discrepancy being greater with the mL-R algorithm than with AGADIR (Table 2), especially for those in 40% TFE. The quantitative differences between the model predictions and from experiment are probably the consequence of one or all of the following. First, both algorithms have large uncertainties associated with predicting helix contents that are very low or are very high (Rohl et al., 1996; Muñoz & Serrano, 1997). Second, the majority of input parameters for the mL-R algorithm is determined at 1 M salt and at 0°C different from those used in this study (0.1 M, 2°C), and while temperature and salt effects appear to be taken into account in AGADIR, they are not, as yet, accounted for in the mL-R algorithm. The most significant cause for the discrepancies is, probably, the way in which side-chain interaction energies are accounted for: the mL-R algorithm includes experimentally determined and quantified side-chain interactions but clearly not those that remain to be identified or quantified; AGADIR, on the other hand, includes a larger number of both experimentally determined as well as assumed values for several side-chain interactions. Either algorithm would underpredict the helix content if a helix-stabilizing side-chain interaction is not included, while an overprediction of helix content results when a helix-destabilizing interaction is not considered. The lower predicted helix contents could thus largely be the consequence of not including or assigning lower values for some side-chain interactions. This is the case in applying the mL-R algorithm in 40% TFE where side-chain interactions (although they persist in TFE (Albert & Hamilton, 1995; Muñoz et al., 1995b; Myers et al., 1998; Padmanabhan et al., 1998)) were not considered because these, to our knowledge, have yet to be quantitated.

NMR data suggest that the helical regions in the five peptides extend over residues 2–17 in 1Cro, 4–13 in 2Cro, 2–12 in 3Cro, 4–15 in 4Cro, and 8–14 in 5Cro, and that the peptide helices are limited to the same regions in both water and in 40% TFE solutions. The 434 Cro peptide helices are thus short (helix 1 with about 12 residues or three helical turns is the longest, and helices 4 and 5 with less than two helical turns are the shortest). They would therefore be expected to have low intrinsic stabilities in water solutions with helix 1 being the most stable, and this is what is experimentally observed. The helix boundaries conform to the residue occurrences, interactions, or both reported to be favorable at helix termini (Presta & Rose, 1988; Richardson & Richardson, 1988; Lyu et al., 1992, 1993; Doig & Baldwin, 1995; Muñoz et al., 1995a): Pro as in 4Cro and 5Cro can be a favorable helix start site (Richardson & Richardson, 1988); the large number of backbone–side-chain and side-chain–side-chain NOEs between Met3 and Leu8 in 2Cro is consistent with the presence of the “hydrophobic staple” motif, an N-terminal ($i, i + 5$) helix-stabilizing nonpolar interaction (Muñoz et al., 1995a); Thr3 in 1Cro, Thr4 in 2Cro, and Asp7 in 5Cro are favorable helix N-cap residues, while Gly13 in 2Cro, Gly12 in 3Cro, Asn15 in 4Cro, and Gly14 in 5Cro are favorable C-caps (the N- and C-caps are the first and last residues in a helix whose C α atoms lie on the axis of the helical cylinder or, alternately, those that form helical H-bonds without having the canonical helical (ϕ, ψ) angles). The $\alpha N(i, i + 2)$ NOEs of the helix termini residues are consistent with the conformational behavior observed for the end residues in protein helices (Némethy et al., 1967; Baker & Hubbard, 1984; Barlow & Thornton, 1988). Moreover, the capping residues show NOE cross peaks (Fig. 4) involving their backbone amide and side-chain protons typical of the capping box motif in 40% TFE and in water solutions (when helical): Thr3–Glu6 in 1Cro, Thr4–Glu7 in 2Cro, Gln3–Ile6 in

3Cro, Pro4–Leu7 in 4Cro, and Pro8–Leu11 and Val9–Gln12 in 5Cro at the N-caps (Lyu et al., 1992, 1993; Muñoz et al., 1995a); and Ala14–Tyr17 in 1Cro, Lys11–Val14 in 2Cro, Glu10–Val13 in 3Cro, and Gln12–Thr15 in 5Cro at the C-caps (Prieto & Serrano, 1997).

Peptide helices are known to be stabilized by various types of interactions between pairs of nonpolar, polar, or charged side-chains spaced ($i, i + 3$) and ($i, i + 4$) (see, for example, Baldwin, 1995) whose presence can be deduced from the occurrence of NOEs between their side chains. Side-chain NOEs between Glu6 and Lys9 or Lys10 in 1Cro and several between various nonpolar pairs in 1Cro, 3Cro, 4Cro, and 5Cro (Fig. 4) are consistent with the presence of possible ($i, i + 3$) and ($i, i + 4$) interactions between oppositely charged side chains (Scholtz et al., 1993, and references therein) or nonpolar pairs (Padmanabhan et al., 1998, and references therein). A cluster of NOEs involving nonpolar and aromatic residues lying between Pro8 and Thr15 in 5Cro (helix 5 in native 434 Cro) seen in water and in 7 M urea but not in 40% TFE or in the native protein suggests that these hydrophobic residues cluster to form a residual nonhelical, structure in 7 M urea and water solutions. Some of the residues in this segment of 5Cro also show large deviations from random coil chemical shifts in 7 M urea and water solutions consistent with a nonrandom structure (Fig. 5), but almost all of these are close to a Trp or Tyr (5Cro has one of each), both of which can significantly affect the chemical shifts of residues adjacent to them even in very short random coil peptides (Merutka et al., 1995; Wishart et al., 1995). The intrinsic fluorescence and wavelength for the emission maximum of Trp in the putative hydrophobic cluster in 5Cro in water and in 7 M urea are the same as for solvent-exposed Trp in model unstructured peptides and in the protein in 7 M urea (data not shown). Urea is believed to denature largely by favorable interactions with peptide and aromatic groups (Nandi & Robinson, 1984; Scholtz et al., 1995) so that formation of a residual structure in 7 M urea (as well as in water) involving both the aromatic residues in 5Cro would imply that the side-chain interactions involved are strong. Interactions between hydrophobic residues, in particular those involving aromatic groups, have been reported for peptide fragments of other proteins (Dyson et al., 1992; Neri et al., 1992; Jiménez et al., 1993; Lumb & Kim, 1994; Itzhaki et al., 1995; Kemmink & Creighton, 1995b; Muñoz et al., 1995b) as well as in model α -helical peptides (Padmanabhan & Baldwin, 1994a, 1994b; Muñoz et al., 1995a; Stapely et al., 1995; Viguera & Serrano, 1995; Padmanabhan et al., 1998) in water and/or in 7 M urea solutions. These interactions and the structures they stabilize have been frequently proposed as initiation or nucleation sites in the protein folding pathway, but their exact roles in the acquisition of native structure remain to be demonstrated.

Comparison with helices in R1–69 and λ_{6-85}

The three-dimensional structure of the 71-residue 434 Cro is, as pointed out earlier, strikingly similar to those of the 69-residue R1–69 and the 80-residue λ_{6-85} . The sequence alignment of the three proteins in Figure 1 illustrates the extent of similarity in their helices and other elements of secondary structure, as well as in their sequences. Seven residues are conserved in all three proteins with six of these being in the region about helices 2 and 3, the helix-turn-helix, sequence-specific DNA-binding motif of these proteins. 434 Cro and λ_{6-85} (two in helix 1, one in helix 5) share

in common an additional three residues as do R1-69 and λ_{6-85} (two in helix 3, one in helix 4). Just under half of all residues in 434 Cro and R1-69 are conserved, almost all in the helical regions and in the loop between helices 3 and 4 (Fig. 1). The helix termini in R1-69 and λ_{6-85} , like in 434 Cro, are demarcated by favorable residues and interactions such as those mentioned earlier. These sequence similarities therefore suggest that the folding behavior of R1-69 and λ_{6-85} would likely parallel that of 434 Cro. A comparison at the level of intrinsic helix stabilities in 434 Cro, R1-69, and λ_{6-85} would, therefore, be relevant and informative.

Experimental data on intrinsic secondary structure propensities in R1-69 and λ_{6-85} are, to our knowledge, available only for two peptides. The first is a fragment of R1-69 extending over residues 44 to 63 and spanning its helices 4 and 5, which was examined only in 7 M urea because it was insoluble in water (Neri et al., 1992). Based on structures calculated using NOEs from ^{15}N and ^{13}C labeled samples of this fragment and the identification of residues with large chemical shift deviations (>0.1 ppm) from random coil values, it was concluded that the hydrophobic side chains of residues 54-59 in R1-69 cluster to form a residual structure in 7 M urea, which is preserved in the native protein and so is a possible folding initiation site (Neri et al., 1992). Of these residues in R1-69 only Trp and its succeeding Leu residue are conserved in the equivalent region in 434 Cro (Fig. 1). As discussed earlier, this region in peptide 5Cro also appears to have a residual structure in 7 M urea but it may not be native like because not all the observed medium-range NOEs are seen in the native protein or in 40% TFE. The importance of this residual structure in the folding of 434 Cro or of R1-69 remains to be resolved. The second peptide for which experimental data are available is an N-acetylated peptide corresponding to the helix 1 residues 9-23 of λ_{6-85} , which has significant helix content (41%) in aqueous solutions at 1 °C, pH 7.0, and 0.1 M KCl (Marqusee & Sauer, 1994). Helix 1 in λ_{6-85} thus appears to be even more stable than the 434 Cro helix 1. In the absence of experimental data, only predictions of helix contents using the mL-R and AGADIR algorithms are

possible for the other helices in R1-69 and λ_{6-85} . These predictions, though not always in quantitative agreement with experimental data (as seen here for the 434 Cro peptides), do provide a relative order of helix forming tendencies that agree qualitatively with experiment. The R1-69 and λ_{6-85} sequences equivalent to the 434 Cro peptides studied here (based on the alignment in Fig. 1) are predicted to have the helix contents listed in Table 3: in R1-69 only helix 2 is predicted to have some intrinsic stability; in λ_{6-85} helices 1 and 4, like in 434 Cro, would be the most stable. The possible implications of the different helix propensities in three proteins will be discussed next.

Peptide structural propensities and the structure, folding, and stability of 434 Cro

The helical regions in the isolated peptide fragments of 434 Cro determined in this study show excellent correspondence with those in the native protein (Mondragon et al., 1989b; Padmanabhan et al., 1997; also see Fig. 1). This parallels the results obtained in numerous other studies of peptide fragments of various proteins (Nelson & Kallenbach, 1989; Lehrman et al., 1990; Segawa et al., 1991; Sancho et al., 1992; Jiménez et al., 1993; Muñoz et al., 1995b; Reymond et al., 1997), although some exceptions to such a correspondence have been reported (Hamada et al., 1995; Hirota et al., 1998). Most of the NMR data characteristic of helices as well as NOE cross peaks characteristic of helix termini observed in the peptides were also observed in the native protein (Padmanabhan et al., 1997). Some of the side-chain-side-chain NOEs seen in the peptides are also seen in the native protein, but not several others. The strong packing and tertiary interactions present in the protein but absent in peptides would very likely alter the average disposition, and consequently, the corresponding side-chain-side-chain NOEs especially for those buried in the interior of the protein. These packing and tertiary interactions are also the predominant determinants of the stability of the native structure of a protein. Consequently, correlating the stability of isolated secondary struc-

Table 3. Helix contents predicted for peptides corresponding to each helix in R(1-69) and λ_{6-85} ^a

Protein	Helix	Sequence ^a	mL-R		AGADIR
			H ₂ O	TFE ^b	H ₂ O
R1-69	1	SISSRVKSKRIQLG-amide	<1	8	1
	2	GLN Q AEL AQ KVGGT-amide	7	33	9
	3	T Q SIEQL EN GKTKR-amide	<1	9	3
	4	TKRPRFL P ELASALGV	1	4	<1
	5	AS ALGVSV D WL LN GTSDSNVR-amide	<1	2	<1
λ_{6-85}	1	SLT Q E Q LEDAR RL KAIYEK-amide	38	59	58
	2	LS Q ESVAD K MGMG-amide	<1	4	4
	3	G Q SGVGA L FN G INAL-amide	<1	2	<1
	4	NALN A YNA A LLAKILK	62	78	20
	5	AKILKSP S IAREIR-amide	<1	2	<1

^aThese sequences are designed to correspond to the 434 Cro peptide sequences (Table 1) and the sequence alignment shown in Figure 1. In the sequences above for helix 2 and helix 5 of λ_{6-85} residues, 25-30 and 71-76 were excluded. An ionic strength of 0.125 M, pH 6.0, and 275 K was used where required in the calculations.

^bItalicized to indicate that side-chain interactions were not considered in these calculations for the expected helix contents.

tures with that of the whole protein is not straightforward. Some compact helical proteins are stable even as their isolated helices are intrinsically unstable (Reymond et al., 1997; Spector et al., 1998). Yet, in some other proteins, enhancing helix stability by introducing stabilizing mutations in solvent-exposed regions has been reported to augment the protein's stability (Burton et al., 1996; Villegas et al., 1996). In particular, the thermodynamic stability of λ_{6-85} was enhanced from around 3.2 kcal mol⁻¹ at 37 °C by almost 1.8 kcal mol⁻¹ (Huang & Oas, 1995a; Burton et al., 1996) when the two highly helix-destabilizing Gly residues in helix3 of the wild-type protein were replaced by Ala, the most helix-stabilizing residue (Rohl et al., 1996). The maximum stability of the structurally similar 434 Cro is over 1 kcal mol⁻¹ less than wild-type λ_{6-85} (Padmanabhan et al., 1997). Helix 3 in 434 Cro does not contain the two Gly present in wild-type λ_{6-85} , but is of the same length, and as discussed previously, the intrinsic helix stabilities of helix 3 and of other helices in λ_{6-85} are similar or greater than those in 434 Cro (Table 3). Thus, it is not unreasonable to speculate that the lower intrinsic helix propensities in 434 Cro contribute to its modest stability. The stability of R1-69 and its intrinsic helix propensities (as yet unreported to our knowledge) and examination of peptide fragments and mutants of 434 Cro with specific changes in the wild-type sequence to alter helix stabilities without affecting packing or overall tertiary structure could throw more light on the correlation between intrinsic helix propensities and stability of the protein.

The native protein helices in 434 Cro can be induced in its isolated peptides by TFE suggesting (as elsewhere; Reymond et al., 1997) that these helices may be stabilized as the protein folds. The intrinsic helix propensities of 434 Cro may thus be important in its folding pathway. Various experimental studies have emphasized the correlation between the folding rates of a protein with strong local interactions such as in the formation of helices (Viguera et al., 1996; Burton et al., 1996, 1997, 1998). Very recently, Plaxco et al. (1998) attempted to theoretically correlate the structure of a protein to its folding rate using the concept of contact order—a measure of the number of local interactions (such as intrahelical interactions) relative to nonlocal, long-range interactions. They reported that proteins with low contact order—highly helical proteins like 434 Cro, λ_{6-85} , and R1-69—would have fast folding rates, although it has since been argued that contact order may be just one of many factors influencing the rate of folding (Jackson, 1998). Wild-type λ_{6-85} has been shown experimentally to fold in the submillisecond time range ($k_f = 4,900 \pm 600$ s⁻¹ at 37 °C), and the folding is enhanced by almost 20-fold ($k_f = 88,000 \pm 11,000$ s⁻¹ at 37 °C) in its mutant with the two Gly → Ala replacements described above, and the enhanced folding rate entirely accounts for the increased stability of the mutant (Burton et al., 1996). Furthermore, Burton et al. (1997, 1998) demonstrated a correlation between folding rates and various helix-destabilizing Ala to Gly substitutions in λ_{6-85} , a correlation that they could successfully describe using a diffusion-collision model that quantitatively incorporated intrinsic helix propensities. λ_{6-85} and 434 Cro being structurally similar have nearly the same contact order, but recent kinetic studies indicate that the folding of 434 Cro is more than two orders of magnitude slower than λ_{6-85} ($k_f \approx 33$ s⁻¹; D.V. Laurents, pers. comm.). Given the relationship between intrinsic helix stabilities and folding rates discussed above, the slower observed folding of 434 Cro relative to λ_{6-85} could be a reflection, at least in part, of the low intrinsic helix stabilities in 434 Cro. A detailed examination of the kinetics of folding of 434

Cro is currently underway and will be reported elsewhere. Moreover, a knowledge of the interplay between the helices, as well as structure-stability data on the unfolded and any partially folded states may also shed additional light on the role of the stability of secondary structures in the folding pathway. We are currently examining these questions for 434 Cro.

Materials and methods

Peptide synthesis

Each peptide was synthesized in a separate fine mesh polypropylene bag (Propyltex fabric 74 mm; Tetko Inc., Chicago, Illinois) employing the multiple solid-phase synthesis procedure (Houghten, 1985) on a 0.03–0.05 mmol scale with 0.1–0.15 g of a 0.49 mEq/g polystyrene Fmoc resin (Novabiochem, Läufelfingen, Switzerland) and the active ester coupling procedure using the OPfp esters of the Fmoc amino acids (Bachem, Bubendorf, Switzerland). Peptides were cleaved from the resin with 95:5 TFA:anisole mixture for 2–4 h to yield crude peptides amidated at the C-terminus, which were purified on a preparative Vydac C18 reverse-phase column using water-acetonitrile (0.1% TFA) gradients and lyophilized. Peptide 4Cro (Fig. 1 shows the peptide sequences and names) with the C-terminal free carboxyl was synthesized by Neosystem Laboratoire (Strasbourg, France). Peptides were >95% pure as judged by analytical Vydac C18 reverse-phase HPLC and their identities were confirmed by NMR and laser desorption mass spectrometry (calculated and observed molecular weights in Daltons are, respectively, 2,134.59 and 2,136.15 for 1Cro, 1,668.97 and 1,669.32 for 2Cro, 1,862.15 and 1,863.01 for 3Cro, 2,025.3 and 2,025.4 for 4Cro, and 2,223 and 2,225.03 for 5Cro). Peptide concentrations were determined from absorbance measurements in 6 M guanidium hydrochloride: at 275 nm for 1Cro, 2Cro, 3Cro, and 4Cro, which contain one Tyr ($\epsilon_{275} = 1,450$ M⁻¹ cm⁻¹; Brandts & Kaplan, 1973), and at 280 nm for 5Cro with one Trp and one Tyr ($\epsilon_{280} = 6,970$ M⁻¹ cm⁻¹, Edelhoch, 1967; Gill & von Hippel, 1989).

CD and fluorescence measurements

CD measurements were made in a Jasco-720 spectropolarimeter equipped with a Neslab control unit and calibrated with (+)-10-camphor-sulfonic acid (Chen & Yang, 1977) in 1 mm pathlength cuvettes, 0.2 nm step size, 4 s response time, and averaged over six scans. Mean molar residue ellipticities at 222 nm ($[\theta]_{222}$ in deg cm² dmol⁻¹) were measured using peptide concentrations between 50–300 μ M at 2 °C and pH 6.0 in NMR buffer: 100 mM KCl, 25 mM KH₂PO₄, 1 mM DTT, 0.01% NaN₃, or in NMR buffer in 40% (by volume) TFE in water. Helix contents were estimated using $[\theta]_{222}$ values of +640 deg cm² dmol⁻¹ for 0% helix and -42,500(1-3/N) for a peptide with N residues for 100% helix (Rohl et al., 1996). The most recent versions of the programs and parameters implementing the modified Lifson-Roig, mL-R, helix-coil theory (Rohl et al., 1996) and AGADIR (Muñoz & Serrano, 1997) were used to predict helix contents for comparison with the measured values. AGADIR was accessed from the web (<http://www.embl-heidelberg.de/ExternalInfo/serrano/>) and the mL-R algorithm was from the laboratory of Prof. R.L. Baldwin (Stanford University), and the programs run using described procedures. Side-chain interaction energies in aqueous solutions whose values have been experimentally determined are included in the

corresponding mL-R algorithm, but not for that in 40% TFE where values for side-chain interaction energies are not available. Intrinsic Trp fluorescence measurements of peptide 5Cro at concentrations between 5–10 μM were done in a Perkin-Elmer LS-50B spectrofluorometer thermostatically controlled at 293 K using a 1 cm pathlength cuvette, an excitation wavelength of 280 nm and a slit width of 2.5 nm and recording the emission spectra using a slit width of 5 nm and two scans.

Sedimentation equilibrium

Sedimentation equilibrium experiments in a Beckman Optima XL-A ultracentrifuge using a Ti60 rotor and two or six channel Epon charcoal centerpieces with adaptors where necessary. 30–70 μL peptide samples in the NMR concentration range (0.6–1.5 mM) and in NMR buffer were centrifuged at 40,000 rpm and at 4 °C to equilibrium (>15 h). This is indicated by superimposable radial scans (monitored at 250, 280, or 300 nm) recorded every 2 h after 10 h of centrifugation. The apparent relative molecular masses (M_r) were determined by analyzing equilibrium radial scans using the program XLAEQ or the program EQASSOC for a single molecular species. The following partial specific volumes for the peptides were used and were calculated from the amino acid compositions as described by Laue et al. (1992): 0.757 mL g⁻¹ for 1Cro; 0.748 mL g⁻¹ for 2Cro; 0.744 mL g⁻¹ for 3Cro; set to 0.730 mL g⁻¹ for 4Cro; and set to 0.730 mL g⁻¹ for 5Cro.

NMR spectroscopy

All NMR experiments used peptide samples at pH 6.0 in 5 mm diameter NMR tubes and were done at 2 °C in a Bruker AMX-600 NMR spectrometer with a ¹H frequency of 600.13 MHz. Samples were prepared by dissolving lyophilized peptide in 9:1 H₂O:²H₂O or 40% deuterated TFE (TFE-d₃) (volume: volume) or 7 M urea in NMR buffer with sodium TSP as the internal chemical shift reference. One-dimensional spectra were obtained using 16 K data points, 6,000 Hz spectral width and averaged over 16 scans. The same spectral width and a 1 s recycle delay were used to acquire 2D spectra in the phase-sensitive mode using time-proportional phase incrementation (TPPI) technique (Marion & Wüthrich, 1983) and presaturation of the water signal. Standard phase-cycling sequences and data acquisition matrix size of 2,048 × 512 (t_2 and t_1 , respectively) were used for COSY (Aue et al., 1976), NOESY (Kumar et al., 1980), and ROESY (Bothner-By et al., 1984) spectra, with NOESY mixing times of 200 and 150 ms in aqueous and 40% TFE/water solutions, respectively. A data acquisition size of 4,096 × 512 ($t_2 \times t_1$) was used for DQF-COSY (Rance et al., 1983), and for TOCSY spectra (mixing time 80 ms) the latter recorded using the standard MLEV17 spinlock sequence (Bax & Davis, 1985). The number of scans were 8 or 16 for the COSY and TOCSY experiments, and 48–80 scans for the NOESY and ROESY experiments. Data were zero-filled, multiplied using a phase-shifted square-sine bell window function optimized for every spectrum, and processed using the Bruker UXNMR software. ¹H resonances were assigned using standard assignment procedures (Wüthrich, 1986). ³J_{H_Nα} coupling constants for nonoverlapping signals were determined by analyzing TOCSY spectra using the method of Stonehouse and Keeler (1995) or from the separations between antiphase absorptive and dispersive peak extrema in DQF-COSY analyzed as described by Kim and Prestegard (1989). These were independently verified where possible in 1D spectra with

high S/N and digital resolution. The difference between the observed C^αH proton chemical shift for each residue and its random coil values (from Merutka et al., 1995) in water and TFE solutions is its chemical shift index or CSI (Wishart et al., 1991, 1992). Helix populations can be estimated from C^αH shifts (Jiménez et al., 1993; Rizo et al., 1993), and this was done using the expression, $1/N[\sum\{(\delta_{obs} - \delta_{rc})/(\delta_{hel} - \delta_{rc})\}]$. Here, δ_{obs} , δ_{rc} , and δ_{hel} are the observed, random coil and 100% helical values for the C^αH chemical shifts with $(\delta_{hel} - \delta_{rc}) = -0.39$ ppm (Wishart et al., 1991), and N is the number of residues in the peptide excluding the N- and the C-terminal residues whose chemical shifts were not considered because they are susceptible to end effects.

Supplementary material in Electronic Appendix

Tables listing the ¹H NMR chemical shifts (relative to TSP) of peptides 1Cro, 2Cro, 3Cro, and 5Cro in aqueous solutions (Tables 1A–1D), and of all the peptides in 40% TFE and 7 M urea solutions (Tables 2A–2E and 3A–3E, respectively), and a table (Table 4) listing side-chain–side-chain NOEs in 434 Cro peptides and the corresponding NOE if identified in the native protein NMR structure.

Acknowledgments

This work was supported by grant PB93-0189 from the Dirección General de Política Científica del Ministerio de Educación (Spain) and from project B104-CT97-2086 from the European Community. We thank Dr. D.V. Laurents for valuable discussions, Dr. Jorge Santoro for the programs for calculating coupling constants, Dr. Guillermo Giménez-Gallego and the Protein Chemistry Division in CIB-Madrid for access to HPLC, laser-desorption mass spectrometer, the CD spectropolarimeter and fluorimeter, Dr. German Rivas for the analytical ultracentrifugation and Apolo Gómez and Cristina López for technical assistance.

References

- Albert JS, Hamilton AD. 1995. Stabilization of helical domains in short peptides using hydrophobic interactions. *Biochemistry* 34:984–990.
- Aue WP, Bartholdi E, Ernst RR. 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J Chem Phys* 64:2229–2246.
- Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44:97–179.
- Baldwin RL. 1989. How does protein folding get started? *Trends in Biochem Sci* 14:291–294.
- Baldwin RL. 1995. α -helix formation by peptides of defined sequence. *Biophys Chem* 55:127–135.
- Barlow DJ, Thornton JM. 1988. Helix geometry in proteins. *J Mol Biol* 201:601–619.
- Bax A, Davis DG. 1985. MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson* 65:355–360.
- Bothner-By AA, Stephens RL, Lee JM, Warren CD, Jeanloz RW. 1984. Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame. *J Am Chem Soc* 106:811–813.
- Brandts JF, Kaplan KJ. 1973. Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitors, insulin, and pancreatic trypsin inhibitor. *Biochemistry* 12:2011–2024.
- Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet* 21:167–195.
- Bundi A, Wüthrich K. 1979. ¹H NMR parameters of the common amino acid residues measured in aqueous solution of the linear peptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* 18:285–297.
- Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG. 1997. The energy landscape of a fast-folding protein mapped by Ala-Gly substitutions. *Nature Struct Biol* 4:305–310.
- Burton RE, Huang GS, Daugherty MA, Fulbright PW, Oas TG. 1996. Microsecond protein folding through a compact transition state. *J Mol Biol* 263:311–322.

- Burton RE, Myers JK, Oas TG. 1998. Protein folding dynamics: Quantitative comparison between theory and experiment. *Biochemistry* 37:5337–5343.
- Chakrabarty A, Kortemme T, Baldwin RL. 1994. Helix propensities of amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci* 3:843–852.
- Chakrabarty A, Kortemme T, Padmanabhan S, Baldwin RL. 1993. Aromatic side-chain contribution to the far-ultraviolet circular dichroism of helical peptides and its effects on the measurements of helix propensities. *Biochemistry* 32:5560–5565.
- Chen GC, Yang JT. 1977. Two point calibration of circular dichrometer with d-10-camphorsulfonic acid. *Anal Lett* 10:1195–1207.
- Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nature Struct Biol* 4:10–19.
- Doig AD, Baldwin RL. 1995. N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci* 4:1325–1336.
- Dyson HJ, Merutka G, Waltho JP, Lerner RA, Wright PE. 1992. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding I. Myohemerythrin. *J Mol Biol* 226:795–817.
- Dyson HJ, Wright PE. 1991. Defining solution conformations of small linear peptides. *Annu Rev Biophys Chem* 20:519–538.
- Edelhoch H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* 6:1948–1954.
- Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182:319–326.
- Hamada D, Kuroda Y, Toshiaki T, Goto Y. 1995. High helical propensity of peptide fragments derived from β -globulin, a predominantly β -sheet protein. *J Mol Biol* 254:737–746.
- Hirota N, Mizuno K, Goto Y. 1998. Group additive contributions to the alcohol-induced α -helix formation of melittin: Implication for the mechanism of the alcohol effects on proteins. *J Mol Biol* 275:365–378.
- Holzwarth G, Doty P. 1965. The ultraviolet circular dichroism of polypeptides. *J Am Chem Soc* 87:218–228.
- Houghten RA. 1985. General method for the rapid solid-phase synthesis of large numbers of peptides: Specificity of antigen-antibody interaction at the level of individual amino acids. *Proc Natl Acad Sci USA* 82:5131–5135.
- Huang GS, Oas TG. 1995a. Structure and stability of monomeric λ -repressor. NMR evidence for two-state folding. *Biochemistry* 34:3884–3892.
- Huang GS, Oas TG. 1995b. Sub-millisecond folding of monomeric λ -repressor. *Proc Natl Acad Sci USA* 92:6878–6882.
- Itzhaki L, Neira JL, Ruiz-Sanz J, de Prat Gay G, Fersht A. 1995. Search for nucleation fragments in smaller fragments of chymotrypsin inhibitor 2. *J Mol Biol* 254:289–304.
- Jackson SE. 1998. How do small single-domain proteins fold? *Folding Design* 3:R81–R91.
- Jiménez MA, Bruix M, González C, Blanco FJ, Nieto JL, Herranz J, Rico M. 1993. CD and ^1H -NMR studies on the conformational properties of peptide fragments from the C-terminal domain of thermolysin. *Eur J Biochem* 207:39–49.
- Kemmink J, Creighton T. 1995a. Effects of trifluoroethanol on the conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor. *Biochemistry* 34:12630–12635.
- Kemmink J, Creighton T. 1995b. The physical properties of local interactions of tyrosine residues in peptides and unfolded proteins. *J Mol Biol* 245:251–260.
- Kim PS, Baldwin RL. 1990. Intermediates in the folding reaction of small proteins. *Annu Rev Biochem* 59:631–660.
- Kim Y, Prestegard JH. 1989. Measurement of vicinal couplings from cross peaks in COSY spectra. *J Mag Reson* 84:9–13.
- Kumar A, Ernst RR, Wüthrich K. 1980. A two-dimensional nuclear Overhauser enhancement (2D NOE) for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biophys Biochem Res Commun* 95:1–6.
- Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE, Rowe AJ, Horton JC, eds. *Analytical ultracentrifugation in biochemistry and polymer sciences*. Cambridge, UK: Royal Society of Chemistry. pp 90–125.
- Lehrman SR, Tuls JL, Lund M. 1990. Peptide α -helicity in aqueous TFE: Correlation with predicted α -helicity and the secondary structure of the corresponding regions of bovine growth hormone. *Biochemistry* 29:5590–5596.
- Lumb KJ, Kim PS. 1994. Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor. *J Mol Biol* 236:412–420.
- Lyu PC, Wemmer DE, Zhou HX, Pinker RJ, Kallenbach NR. 1993. Capping interactions in isolated α -helices: Position-dependent substitution effects and structure of a serine-capped peptide helix. *Biochemistry* 32:421–425.
- Lyu PC, Zhou HX, Jelveh N, Wemmer DE, Kallenbach NR. 1992. Position-dependent stabilizing effects in α -helices: N-terminal capping in synthetic model peptides. *J Am Chem Soc* 114:6560–6562.
- Marion D, Wüthrich K. 1983. Application of phase-sensitive two-dimensional correlated spectroscopy (COSY) for measurement of proton–proton spin–spin coupling constants. *Biochem Biophys Res Commun* 113:967–974.
- Marqusee S, Sauer RT. 1994. Contributions of a hydrogen bond/salt-bridge network to the stability of secondary and tertiary structure in λ repressor. *Protein Sci* 3:2217–2225.
- Merutka G, Morikis D, Brüschweiler R, Wright PE. 1995. Random coil ^1H NMR chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR* 5:14–24.
- Mondragon A, Subbiah S, Almo SC, Drottman M, Harrison SC. 1989a. Structure of the amino-terminal domain of the phage 434 repressor at 2.00 Å resolution. *J Mol Biol* 205:189–200.
- Mondragon A, Wolberger C, Harrison SC. 1989b. Structure of phage 434 Cro protein at 2.35 Å resolution. *J Mol Biol* 205:179–188.
- Muñoz V, Blanco FJ, Serrano L. 1995a. The hydrophobic-staple motif and a role for loop-residues in α -helix stability and folding. *Nature Struct Biol* 2:380–385.
- Muñoz V, Serrano L. 1996. Local and nonlocal interactions in protein folding and stability—An experimentalist's point of view. *Folding Design* 1:R71–R77.
- Muñoz V, Serrano L. 1997. Development of the multiple sequence approximation within the AGADIR model of α -helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41:495–509.
- Muñoz V, Serrano L, Jimenez MA, Rico M. 1995b. Structural analysis of helices encompassing all α -helices of three α/β parallel proteins: Che-Y, flavodoxin, p21-ras: Implications for α -helix stability and the folding of α/β parallel proteins. *J Mol Biol* 247:648–669.
- Myers JK, Pace CN, Scholtz JM. 1998. Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci* 7:383–388.
- Nandi PK, Robinson DR. 1984. Effects of urea and guanidine hydrochloride on peptide and nonpolar groups. *Biochemistry* 23:6661–6668.
- Nelson JW, Kallenbach N. 1986. Stabilization of the ribonuclease S-peptide in trifluoroethanol solutions. *Proteins Struct Funct Genet* 1:211–217.
- Nelson JW, Kallenbach N. 1989. Persistence of the alpha-helix stop signal in S-peptide in TFE solutions. *Biochemistry* 28:5256–5261.
- Némethy G, Phillips DC, Leach SJ, Scheraga HA. 1967. A second right-handed helical structure with the parameters of the Pauling-Corey α -helix. *Nature* 214:363–365.
- Neri D, Billeter M, Wider G, Wüthrich K. 1992. NMR determination of residual structure in a urea-denatured protein. *Science* 257:1559–1563.
- Padmanabhan S, Baldwin RL. 1994a. Helix-stabilizing interactions between tyrosine and leucine or valine when the spacing is ($i, i + 4$). *J Mol Biol* 241:706–713.
- Padmanabhan S, Baldwin RL. 1994b. Tests for helix-stabilizing interactions between various nonpolar amino in alanine-based peptides. *Protein Sci* 3:1992–1997.
- Padmanabhan S, Jiménez MA, González C, Sanz JM, Giménez-Gallego G, Rico M. 1997. Three-dimensional solution structure and stability of phage 434 Cro protein. *Biochemistry* 36:6424–6436.
- Padmanabhan S, Jiménez MA, Laurens DV, Rico M. 1998. Helix-stabilizing nonpolar interactions between tyrosine and leucine in aqueous solutions: 2D- ^1H NMR and CD studies in alanine-lysine peptides. *Biochemistry* 37:17318–17330.
- Plaxco KW, Simons KT, Baker D. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277:985–994.
- Presta LG, Rose GD. 1988. Helix signals in proteins. *Science* 240:1648–1652.
- Prieto J, Serrano L. 1997. C-capping and helix stability: The Pro C-capping motif. *J Mol Biol* 274:276–288.
- Ptashne M. 1986. *A genetic switch: Gene control and phage λ* . Cambridge, MA: Cell Press and Blackwell Scientific Publications.
- Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. 1983. Improved spectral resolution in COSY ^1H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun* 117:479–485.
- Reymond MT, Merutka G, Dyson HJ, Wright PE. 1997. Folding propensities of peptide fragments of myoglobin. *Protein Sci* 6:706–716.
- Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of helices. *Science* 240:1648–1652.
- Rico M, Jiménez MA, González C, De Fillippis V, Fontana A. 1994. NMR solution structure of the C-terminal fragment 255–316 of thermolysin: A dimer formed by subunits having the native structure. *Biochemistry* 33:14834–14847.
- Rizo J, Blanco FJ, Kobe B, Bruch MD, Gierasch LM. 1993. Conformational behavior of *Escherichia coli* OmpA signal peptides in membrane mimetic environments. *Biochemistry* 32:4881–4894.

- Rohl CA, Chakraborty A, Baldwin RL. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci* 5:2623–2637.
- Sancho J, Neira JL, Fersht A. 1992. An N-terminal fragment of barnase has residual structure similar to that in a refolding intermediate. *J Mol Biol* 224:749–758.
- Scholtz JM, Barrick D, York EJ, Stewart JM, Baldwin RL. 1995. Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc Natl Acad Sci USA* 92:185–189.
- Scholtz JM, Qian H, Robbins VH, Baldwin RL. 1993. The energetics of ion-pair and hydrogen bonding interactions in a helical peptide. *Biochemistry* 32:9668–9676.
- Segawa SI, Fukuno T, Fujiwara K, Noda Y. 1991. Local structures in unfolded lysozyme and correlation with secondary structures in the native conformation: Helix forming or breaking propensity of peptide segments. *Biopolymers* 31:457–509.
- Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. 1996. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol* 255:494–506.
- Sonnichsen FD, Van Eyk JE, Hodges RS, Sykes BD. 1992. Effect of trifluoroethanol on protein secondary structure: An NMR and CD study using a synthetic actin peptide. *Biochemistry* 31:8790–8798.
- Spector S, Kuhlman B, Fairman R, Wong E, Boice JA, Raleigh DR. 1998. Cooperative folding of a protein mini domain: The peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex. *J Mol Biol* 276:479–489.
- Stapely BJ, Rohl CA, Doig AD. 1995. Addition of side chain interactions to the modified Lifson-Roig helix-coil transition theory: Application to energetics of Phe-Met interactions. *Protein Sci* 4:2383–2391.
- Stonehouse J, Keeler J. 1995. A convenient and accurate method for the measurements of the values of spin-spin coupling constants. *J Magn Reson* 112 (Series A):43–57.
- Thompson PA, Eaton WA, Hofrichter J. 1997. Laser temperature jump study of the helix coil kinetics of an alanine peptide interpreted with a “kinetic zipper” model. *Biochemistry* 36:9200–9210.
- Viguera AR, Serrano L. 1995. Side-chain interactions between sulfur-containing amino acids and Phe in α -helices. *Biochemistry* 34:8771–8779.
- Viguera AR, Villegas V, Avilés FX, Serrano L. 1996. Favorable native-like local interactions can accelerate protein folding. *Folding Design* 2:23–33.
- Villegas V, Viguera AR, Avilés FX, Serrano L. 1996. Stabilization of proteins by rational design of α -helix stability using helix/coil transition theory. *Folding Design* 1:29–34.
- Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB. 1996. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry* 35:691–697.
- Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigation of nearest-neighbor effects. *J Biomol NMR* 5:67–81.
- Wishart DS, Sykes BD, Richards FM. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311–333.
- Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index and protein secondary structure. *Biochemistry* 31:1647–1651.
- Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science* 267:1619–1620.
- Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: J. Wiley & Sons, Inc.
- Yang JJ, Buck M, Pitkeathly M, Kotik M, Haynie DT, Dobson CM, Radford SE. 1995. Conformational properties of four peptides spanning the sequence of hen lysozyme. *J Mol Biol* 252:483–491.