Synthetic Peptides Probe Folding Initiation Sites in Platelet Factor-4: Stable Chain Reversal Found within the Hydrophobic Sequence LIATLKNGRKISL†

Elena Ilyina, Robert Milius, and Kevin H. Mayo*

Department of Biochemistry, Biomedical Engineering Center, University of Minnesota, 420 Delaware Street, S.E., Minneapolis, Minnesota 55455

Received May 25, 1994; Revised Manuscript Received July 26, 1994⊗

ABSTRACT: Platelet factor-4 (PF4) is a 70-residue protein which contains a 3-stranded antiparallel β-sheet domain on to which is folded a C-terminal α-helix and an aperiodic N-terminal region. In this study, three peptides derived from the β-sheet (residues 24−46 and 38−57) and helix (residues 57−70) domains have been synthesized and studied in aqueous solution by CD and NMR. While peptides 24−46 and 56−70 demonstrate some weak conformational preferences, peptide 38−57 maintains a relatively well-defined, NOE-rich chain reversal sequence, L45-K46-N47-G48-R49-K50, which apparently is stabilized by hydrophobic side-chain interactions from the flanking sequences L41−L45 and I51−L53. Some helix-like conformational populations are noted in the native PF4 I42-A43-T44-L45 β-strand segment. NOE-based distance geometry calculations yield native-like conformations within the L45−K50 sequence. Among 40 structures, backbone RMS deviations range from 0.5 to 1.2 Å, and compared to the same sequence in native PF4, the average RMS deviation is 1.1 Å. These results suggest that β-sheet/turn residues L41−L53 present a folding initiation site on the PF4 folding pathway.

In aqueous solution, proteins fold on a time scale that runs from microseconds to milliseconds to seconds and sometimes longer [see review by Jaenicke (1991)]. At the shortest times, "early folding intermediates" contain considerable secondary structure (Kuwajima, 1992) with an expanded nonglobular or collapsed unfolded form (Baldwin, 1991). These highly transient intermediates appear to precede (Ptitsyn, 1991; Chaffotte et al., 1991, 1992) formation of a more compact molten globule state (Ptitsyn, 1987). Hydrophobic collapse has been suggested to mediate compact intermediate formation followed by local rearrangements (Segawa & Sugihara, 1984; Goldenberg & Creighton, 1985) or reshuffling (Vaucheret et al., 1987) on a slower time scale that leads to tighter packing inside the protein core (Ptitsyn & Semisotnov, 1991; Baldwin, 1993).

Intermediate structures that may form on the folding pathway can be either stabilized or destabilized by additional interresidue interactions as the folding process proceeds. By investigating short linear peptides derived from specific structural units in the native protein, i.e., helix, turn, or β-sheet/turn regions, conformational preferences in early folding intermediates may be identified under equilibrium conditions. When solvent conditions are the same as native folding conditions, it may be concluded that relatively "stable" conformations probably do exist early on in the protein folding conformational ensemble. That being the case, it is possible to identify structures which either form first or at least are stable intermediates. These intermediates may or may not remain as part of the native protein structure. Recently, this approach has been exemplified in the α-helix protein myoglobin (Waltho et al., 1993; Shin et al., 1993a,b). Although numerous studies have been done to understand α-helix protein folding [see review by Jaenicke (1991)], much less is known about how β-sheet proteins fold (Buechner et al., 1991; Ropson et al., 1992; Varley et al., 1993). For interleukin-1β, Varley et al. (1993) have indicated that folding of this β-sheet protein follows a sequential mechanism with kinetically identifiable intermediates. Since a β-sheet represents more nonlocal interactions, it is likely that unfolded proteins have a higher tendency to form more local interactions like turns or helices which fold faster (Gruenewald et al., 1979) than β-sheets (Finkelstein, 1991).

Platelet factor-4 (PF4)[1] is a 70-residue (Deuel et al., 1977), mostly β-sheet protein. In the native state, PF4 has a three-stranded antiparallel β-sheet domain onto which is folded a C-terminal α-helix and an aperiodic N-terminal region (St. Charles et al., 1989). Figure 1 gives the general backbone folding scheme for PF4. The N-terminal region (not shown in Figure 1) is covalently constrained to the β-sheet domain by two cystine disulfide bridges (Holt et al., 1986). On the basis of recent evidence that the protein folding core is correlated with the presence of the 5−10 longest-lived backbone NHs (Kim et al., 1993), it appears that the antiparallel β-sheet domain which contains the six to eight longest-lived NHs in PF4 (Mayo et al., unpublished results), as well as in sequentially and conformationally homologous IL-8 (Clore et al., 1990), NAP-2 (Mayo et al., 1994), and Gro-α (Fairbrother et al., 1993), forms the protein folding core. Depending on solution conditions, PF4 can form homodimers and tetramers which are in slow exchange

† This work was supported by a grant from the National Heart Lung, and Blood Institute.

* Address correspondence to this author.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1994.

[1] Abbreviations: PF4, platelet factor-4; NMR, nuclear magnetic resonance; 2D-NMR, two-dimensional NMR spectroscopy; DQF-COSY, 2D-NMR double quantum-filtered correlated spectroscopy; HOHAHA, 2D-NMR homonuclear Hartman−Hahn spectroscopy; NOE, nuclear Overhauser effect; NOESY, 2D-NMR nuclear Overhauser effect spectroscopy; rf, radio frequency; FID, free induction decay; CD, circular dichroism; DG, distance geometry; RMSD, root mean square deviation.

0006-2960/94/0433-13436$04.50/0

on the chemical shift time scale (Mayo & Chen, 1989). Reduction of the disulfides shifts the folding equilibrium into a monomer molten globule state which is more stable at lower pH values (Mayo et al., 1992). Raising the pH induces a native-like folding process that is thermodynamically linked to dimerization (Mayo et al., 1992).

The purpose of the present study is to investigate the propensity of short linear peptides derived from specific native PF4 structural units to form stable conformation(s) in aqueous solution. The structural units are identified in Figure 1 as two segments of antiparallel β-sheet (residues 24−46 and 38−57) and the C-terminal α-helix domain (residues 56−70). The aperiodic N-terminal domain has not been considered. Peptides defining these domains were synthesized and investigated by CD and NMR.

MATERIALS AND METHODS

Peptide Synthesis. Peptides representing amino acid residues 24−46, 38−57, and 56−70 from human platelet factor 4 (PF4) were synthesized on a Milligen Biosearch 9600 automated peptide synthesizer. The procedures used were based on Merrifield solid phase synthesis utilizing Fmoc-BOP chemistry (Stewart & Young, 1984). For the internal sequences 24−46 and 38−57, the C-terminus carboxylate group has been synthesized in the amide form. To avoid complications due to cysteine oxidation, C36 and C52 (shown in Figure 1) have been replaced by serines in these synthetic peptides. After the sequence had been obtained, the peptide support and side chain protection groups were acid (trifluoroacetic acid and scavenger mixture) cleaved. Crude peptides were analyzed for purity on a Hewlett-Packard 1090M analytical HPLC using a reverse phase C18 VyDac column. Peptides generally were about 90% pure. Further purification was done on a preparative reverse-phase HPLC C-18 column using an elution gradient of 0−60% acetonitrile with 0.1% trifluoroacetic acid in water. Peptides then were analyzed for amino acid composition on a Beckman 6300 amino acid analyzer by total hydrolysis (6 N HCl at 110 °C for 18−20 h) and by mass spectrometry. Final peptide purity was greater than 95%.

NMR Measurements. For NMR measurements, freeze-dried peptides were dissolved either in D_2O in H_2O/D_2O (9:1). Protein concentration was in the range of 0.1 − 5 mM. The pH was adjusted to pH 5.3 by adding microliter quantities of NaOD or DCl to the protein sample. For most experiments, the temperature was controlled at 10 °C. All NMR spectra were acquired on a Bruker AMX-600 NMR spectrometer.

For sequence-specific resonance assignments, 2D-homonuclear magnetization transfer (HOHAHA) spectra, obtained by spin-locking with a MLEV-17 sequence (Bax & Davis, 1985) and a mixing time of 60 ms, were used to identify spin systems. NOESY experiments (Jeener et al., 1979; Wider et al., 1984) were performed to sequentially connect spin systems and to identify distance constraints. Double quantum-filtered COSY data (Piantini et al., 1982; Shaka & Freeman, 1983) were acquired to resolve $^3J_{\alpha N}$ coupling constants. All 2D-NMR spectra were acquired in the TPPI (Marion & Wuthrich, 1983) or States-TPPI (States et al., 1982; Marion et al., 1989) phase-sensitive mode. The water resonance was suppressed by direct irradiation (0.8 s) at the water frequency during the relaxation delay betweeen scans as well as during the mixing time in NOESY experiments.

The majority of the 2D-NMR spectra were collected as 256 or 512 t_1 experiments, each with 1K or 2K complex data points over a spectral width of 6 kHz in both dimensions with the carrier placed on the water resonance. For HOHAHA (COSY) and NOESY spectra, normally 16 and 64 scans, respectively, were time-averaged per t_1 experiment. The data were processed directly on the Bruker AMX-600 X-32 or offline on a Bruker Aspect-1 workstation with the Bruker UXNMR program. Data sets were multiplied in both dimensions by a 30−60° shifted sine-bell function and zero-filled to 1K in the t_1 dimension prior to Fourier transformation.

The temperature dependence of backbone NH chemical shifts was followed from 10 to 50 °C at pH 5.3 using HOHAHA spectra where cross-peaks were assigned or checked at various temperatures. NH chemical shifts were plotted versus temperature, and slopes were derived from least-squares fits of these data. For a short, linear peptide, a shallower slope generally indicates a relatively more hydrogen-bonded or solvent-protected NH proton (Wüthrich, 1986).

Circular Dichroism. Circular dichroic (CD) spectra were measured on a JASCO JA-710 automatic recording spectropolarimeter coupled with a data processor. Curves were recorded digitally and fed through the data processor for signal-averaging and base-line subtraction. Spectra were recorded from 5 to 65 °C in 10 mM phosphate buffer, pH 5.3, over a 180−250 nm range using a 0.01 cm or 0.1 cm path-length, thermally-jacketed quartz cuvette. Temperature was controlled by using a NesLab water bath. Peptide concentration was varied from 0.014 to 0.14 mM. The scan speed was 100 nm/min. Spectra were signal-averaged 8 times, and an equally signal-averaged solvent base line was subtracted. CD spectra were analyzed in terms of fractional conformational populations by the methods of Sreerama and Woody (1993) and of Chen et al. (1974).

Distance Geometry Calculations. NOE-derived interproton distance constraints were used in distance geometry structure calculations with the BIOSYM suite of programs. Starting from the linear amino acid sequence, 40 DG structures were calculated. No NOE distance violations were noted. Resulting structures were compared to the T44−K50 folded sequence from the native PF4 structure. RMSD values are reported relative to that structure and to the average DG structure of T44−K50.

RESULTS

Three peptides derived from PF4 have been synthesized: residues 24−46, 38−57, and 56−70. Each peptide represents one structural unit in native PF4 (see Figure 1). C36 and C52 have been replaced by serines. The following sections give CD and NMR results for each of these peptides.

CD Spectra. CD spectra for all three peptides are given as a function of temperature in Figure 2A,B,C. The experimental error in any of these CD traces, i.e., standard deviation from three separate runs, is no more than ±1.0 in molar ellipticity. For peptides 22−46 and 38−57, standard deviations are closer to 0.5−0.7. Irrespective of temperature, each of these peptides gives similar CD traces which indicate the presence of considerable random structure distributions as predicted for short linear peptides (Bierzynski et al., 1982; Shoemaker et al., 1985; Waterhous & Johnson, 1994). In

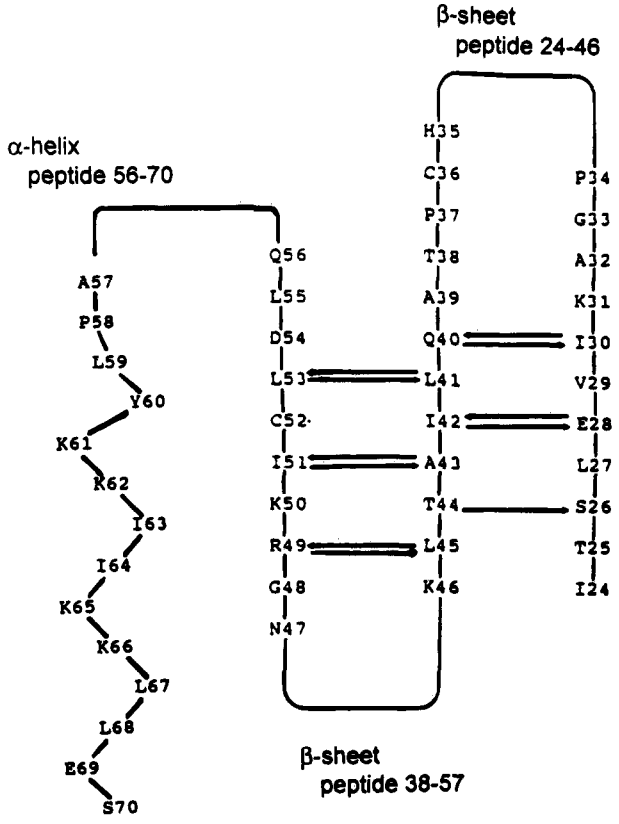

FIGURE 1: Schematic of PF4 backbone folding. A schematic representation showing backbone folds in natve PF4 is given. Arrows indicate long-lived backbone amide hydrogen bonds. L41, A43, L45, I51, and L53 are the longest-lived NHs, and presumably define the folding core in PF4. Synthetic peptides under investigation in this study are indicated as described in the text.

aqueous solution, such peptides are usually a heterogeneous population of low-energy conformers whose ϕ,ψ angles, representing helix, β-strand, turn, and unordered structure, are part of the dynamic equilibrium ensemble. For peptides 24−46 and 38−57, it appears that this ensemble average contains some non-random-coil conformational populations. Relative to peptide 56−70, the 198 and 220 nm bands are more negative, and the 185 nm band is more positive. This suggests increased α-helix and/or β-strand ϕ,ψ angular populations in these two peptides.

CD spectra analyzed by the method of Sreerama and Woody (1993) provide conformational fractional populations shown in the Figure 2 inserts. Since this analysis is done on short linear peptides and the Sreerama and Woody approach is based on a structural database for folded proteins, caution in interpreting these distributions should be exercised. Standard deviations from three individual runs are about ±0.03 in fractional populations. For peptides 24−46 and 38−57, the α-helicity increases from 5 to 25 °C at the expense of other conformational populations. Fractional distributions remain mostly unchanged above 25 °C. Increasing the temperature appears to stabilize conformational populations. The precise reason(s) for this is (are) unknown. Both peptides contain about 50% hydrophobic residues, suggesting that the hydrophobic effect may account for some increased stability. Since at a given temperature no significant CD spectral changes are noted on varying the peptide concentration between 0.14 and 14 μM, aggregation plays no apparent role in this process.

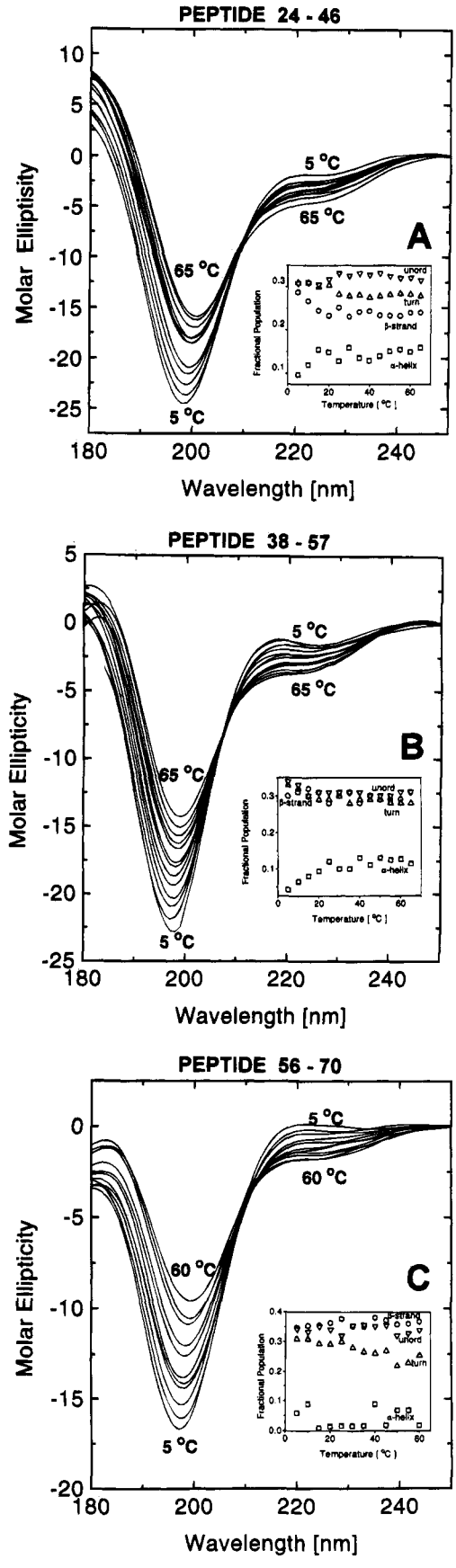

FIGURE 2: CD spectra for PF4 peptides. Far-ultraviolet circular dichroic spectra for PF4 peptides 24−46, 38−57, and 56−70 are shown as mean residue ellipticity versus wavelength (nm). Peptide concentrations were 1.4−14 μM in 10 mM potassium phosphate, pH 5.3. The temperature was varied from 278 to 338 K. Other experimental variables are given under Materials and Methods.

Table 1: Resonance Assignments for PF4 24−46 (10 mM NaCl, pH 5.3, 283 K)

	chemical shift (ppm)					
	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	other
I24		3.86				
T25	8.60	4.33	4.09	1.11		
S26	8.57	4.33	3.76			
L27	8.34	4.22	1.5	1.5	0.78	
E28	8.22	4.14	1.86	2.18		
V29	8.17	3.91	1.92	0.8		
I30	8.26	3.97	1.7	1.31, 1.06		0.78($C^{\gamma}H_3$)
K31	8.42	4.18	1.66, 1.57	1.32	1.56	2.85($C^{\epsilon}H_2$),
A32	8.3	4.24	1.29			7.51($N^{\zeta}H_3$)
G33	8.22	4.04, 3.91				
P34		4.29	2.13, 1.71	1.86	3.5	
H35	8.67	4.64	3.17, 3.08			
S36	8.36	4.64	3.76			
P37		4.42	2.23, 1.88	1.94	3.75, 3.65	
T38	8.15	4.15		1.11		
A39	8.27	4.15	1.28			
Q40	8.33	4.14	1.88	2.24	6.85, 7.51 ($N^{\delta}H$)	
L41	8.23	4.215	1.5	1.5	0.78	
I42	8.14	3.99	1.74			0.79($C^{\gamma}H_3$)
A43	8.34	4.23	1.29			
T44	8.09	4.14	4.09	1.10		
L45	8.21	4.24	1.53	1.53	0.78	
K46	8.25	3.98	1.68, 1.62	1.33	1.56	2.85($C^{\epsilon}H_2$)

Table 2: Resonance Assignments for PF4 38−57 (10 mM NaCl, pH 5.3, 283 K)

	chemical shift (ppm)					
	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	other
T38		4.03	3.74	1.09		
A39	8.80	4.26	1.28			
Q40	8.50	4.17	1.87	2.25	6.84, 7.55 ($N^{\delta}H$)	
L41	8.39	4.25	1.50	1.50	0.79	
I42	8.21	4.05	1.74	1.39, 1.08	0.89	0.79($C^{\gamma}H_3$)
A43	8.43	4.26	1.28			
T44	8.12	4.20	4.07	1.09		
L45	8.26	4.30	1.57, 1.40	1.40	0.79	
K46	8.43	4.17	1.69	1.28	1.57	2.87($C^{\epsilon}H_2$),
N47	8.46	4.55	2.74	7.59, 7.87		7.47($N^{\zeta}H_3$)
G48	8.33	3.82		($N^{\gamma}H$)		
R49	8.00	4.22	1.69	1.49	3.09	7.14 ($N^{\epsilon}H$)
K50	8.41	4.24	1.67	1.29		2.87($C^{\epsilon}H_2$),
						7.47($N^{\zeta}H_3$)
I51	8.28	4.09	1.71	1.38, 1.06	0.9	0.76($C^{\gamma}H_3$)
S52	8.35	4.37	3.75			
L53	8.37	4.22	1.50	1.50	0.79	
D54	8.28	4.47	2.63, 2.46			
L55	8.18	4.18	1.53	1.53	0.79	
Q56	8.29	4.17	2.03, 1.91	2.26	6.85, 7.52 ($N^{\delta}H$)	
A57	8.10	4.14	1.28			

Table 3: Resonance Assignments for PF4 56−70 (10 mM NaCl, pH 5.3, 283 K)

	chemical shift (ppm)					
	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	other
Q56		3.91	2.01	2.33	6.85, 7.61	
A57	8.77	4.53	1.28		($N^{\delta}H$)	
P58		4.26	2.12, 1.61	1.91	3.73	
L59	8.26	4.13	1.43, 1.34		0.77	
Y60	8.11	4.44	2.88, 2.84	6.99	6.6 ($C^{\epsilon}H_2$)	
K61	8.04	4.11	1.56	1.21		2.87($C^{\epsilon}H_2$), 7.55($N^{\zeta}H_3$)
K62	8.19	4.10	1.62	1.29		2.87($C^{\epsilon}H_2$), 7.55($N^{\zeta}H_3$)
I63	8.28	4.01	1.70	1.39	1.07	0.72($C^{\gamma}H_3$)
I64	8.36	4.05	1.72	1.36	1.07	0.75($C^{\gamma}H_3$)
K65	8.41	4.18	1.63	1.32		2.87($C^{\epsilon}H_2$), 7.55($N^{\zeta}H_3$)
K66	8.38	4.16	1.63	1.31		2.87($C^{\epsilon}H_2$), 7.55($N^{\zeta}H_3$)
L67	8.32	4.22	1.50	1.50	0.81	
L68	8.27	4.24	1.50	1.50	0.81	
E69	8.42	4.21	1.97, 1.85	2.18		
S70	8.26	4.27	3.77			

For peptide 56−70, CD molar ellipticities are more attenuated and yield noiser fractional distributions. In this respect, it seems reasonable to take a simple average, for example, 3% helix. Since this peptide forms α-helix in native PF4, the Chen et al. method (1974) also was employed to derive the fraction of helicity by using the mean residue ellipticity at 220 nm, $[\theta_{220}]$, and the equation:

$$\theta = (f_H - ik/N)\theta_H \quad (1)$$

where θ is the observed mean residue ellipticity at 220 nm wavelength, θ_H is the maximum mean residue ellipticity of a helix of infinite length, f_H is the fraction of helix in the molecule (12/15), i is the number of helical segments (1), N is the total number of residues (15), and k is a wavelength-dependent constant (2.6 at 220 nm). The expected value of the mean residue ellipticity for 100% helicity of peptides of chain length 12 residues was determined to be −14 542 deg·cm²/dmol. Using this method, the calculated helicity as a function of temperature ranges from 0% to 4%. The average value compares favorably with that derived from the Sreerama and Woody analysis.

^{1}H-NMR Assignments. For each peptide, sequence-specific ^{1}H-NMR assignments were made by using standard 2D-NMR methods (Wüthrich, 1986). Spin systems were initially identified in HOHAHA spectra. Unique residues were specifically assigned at this stage and used as starting points for sequential assignments via analysis of NOESY spectra. Chemical shift values for these three peptides are listed in Tables 1−3.

NMR Conformational Analyses. (A) Peptide 24−46. In native PF4, the amino acid sequence running for S26 to L45 folds in a antiparallel β-sheet structure with a chain reversal between residues A32 and T38 (see Figure 1).

In general, peptide 24−46 NOESY spectra show few conformationally meaningful NOEs. d_{NN} NOES are noted, however, from I42 to T44 and between H35 and S36 (Figure 3), whereas N-terminal segment d_{NN} NOEs are not apparent up to a mixing time of 500 ms. In native PF4, a medium-sized H35-S36 d_{NN} NOE is observed, while no d_{NN} NOEs are found within the I42−T44 sequence. CD data do suggest 10−15% (about 3 or 4 out of 23 residues) α-helicity in peptide 24−46. The I42−T44 segment could be involved in nascent helix formation (Dyson et al., 1988).

αH and NH chemical shift differences, $\Delta\delta$, indicate significant deviations from random coil strtucture (Figure 4). T25, A32, G33, H35, S36, P37, A43, and L45 αH resonances are shifted greater than 0.1−0.2 ppm down field relative to those expected for a random coil conformation (Wüthrich, 1986). The T25 shift difference could be due to an N-terminal effect, while other αH shift differences suggest preferred conformational populations within the A32−P37 sequence. The presence of two prolines, P34 and P37, and the H35−S36 d_{NN} NOE support the idea of a chain reversal within this sequence. V29 and I30 NH temperature factors (Figure 4) are abnormally small, indicating NH solvent protection. This could be the result of intramolecular hydrogen bond formation (Wüthrich, 1986) or hydrophobic

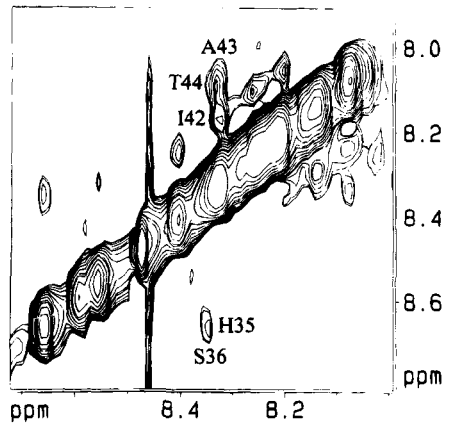

FIGURE 3: NOESY of peptide 24−46. The NH−NH resonance regions from a NOESY contour plot are shown for peptide 24−46, pH 5.3, 10 °C. Data were collected in 90% ^1H_2O/10% ^2H_2O. Protein concentration was 5 mg/mL. 512 hypercomplex FIDS containing 1K words each were collected and processed as discussed under Materials and Methods. The mixing time was 0.5 s. The data were zero-filled to 1024 in t_1. The raw data were multiplied by a 30° shifted sine-squared function in t_1 and t_2 prior to Fourier transformation. Labeling of resonances is as discussed in the text.

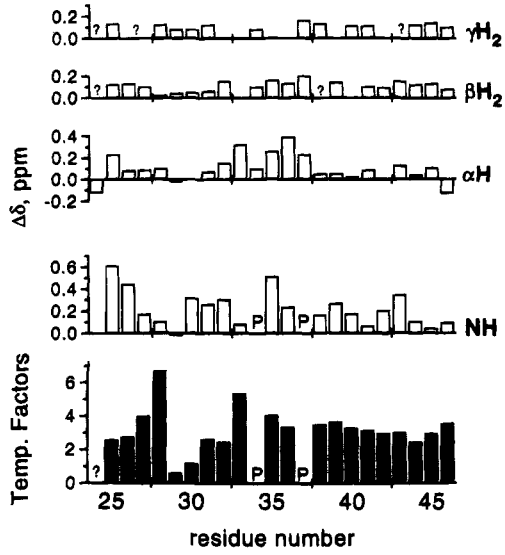

FIGURE 4: Chemical shift differences and temperature factors for 24−46. NH and αH chemical shift differences were determined by subtracting a particular peptide 24−46 proton resonance from its respective random coil chemical shift (Wuthrich, 1986). In this respect, a positive difference refers to a downfield chemical shifted resonance in the peptide. Temperature factors were calculated from the slope of NH chemical shifts versus temperature in kelvin. Routinely, NH chemical shifts for eight temperature points (every 5 or 10 K from 278 to 323 K) were measured. Lesat-squares fit residuals were generally greater than 0.98. Resulting temperature factors are plotted in this figure as parts per billion (ppb) per kelvin (K), i.e., ppb K^{-1}, versus the peptide residue number.

side-chain clustering (Dyson et al., 1992). Hydrogen bond formation seems a less likely explanation given the absence of folded structure.

(B) Peptide 38−57. In native PF4, this sequence forms a two-stranded antiparallel β-sheet structure, i.e., T38−L45 and R49−L53, with a chain reversal from L45 to R49. A second chain reversal is found from D54 to A57 which then leads into the C-terminal helix.

Figures 5 and 6 give NOESY contour plots for the αH/upfield-NH and NH−NH regions, respectively, for peptide 38−57. Some sequential αN NOEs are traced out and

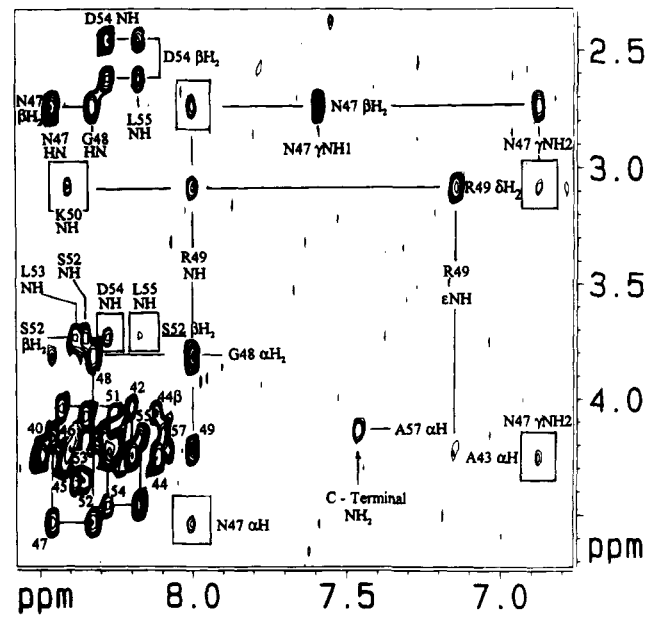

FIGURE 5: NOESY of peptide 38−57. The αCH−NH fingerprint resonance region from a NOESY contour plot is shown for peptide 38−57, pH 5.3, 10 °C. Data were collected and processed as described in the legend to Figure 3 and under Materials and Methods. The mixing time was 0.5 s. Sequential resonance assignments are traced out, and some longer range NOEs are indicated. Labeling of resonances is as discussed in the text.

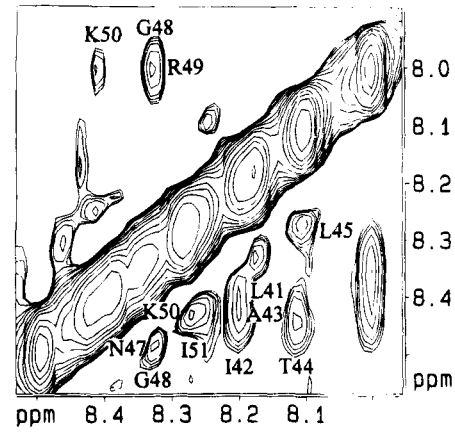

FIGURE 6: NOESY of peptide 38−57. The NH−NH resonance region from a NOESY contour plot is shown for peptide 38−57, pH 5.3, 10 °C. Data were collected and processed as described in the legend to Figure 5 and under Materials and Methods.

labeled (Figure 5). Several conformationally meaningful interresidue NOEs are labeled and/or boxed-in. Most of these originate from within the L45−R49 fragment. In particular, a series of d_{NN} NOEs are observed from N47 to I51, with the strongest one being between G48 and R49. In addition, $d_{\alpha N}(i,i+2)$ and $d_{\beta N}(i,i+2)$ NOEs between N47 and R49, and a $d_{\gamma N}$ NOE between L45 and R49, strongly suggest the presence of a preferred chain reversal within the R49−L45 sequence. These and other NOEs are listed in Table IV. Given the fact that this is a short linear 20-residue peptide, considerable conformational flexibility is expected, and multiple conformations could be represented in observed NOEs as the peptide samples a wide range of conformational space. Within the L45−K50 sequence, however, NOEs are consistent with a discrete family of local folds. This leads one to conclude that this sequence maintains a relatively well-defined chain reversal.

Table 4: NOEs for Peptide 38−57[a]

from	to	size	from	to	size
T38 NH			G48		
A39 NH	T38 αH	vw	NH	K46 βH_2	w
	T38 βH	s		K46 αH	w
Q40 NH	A39 αH	s		N47 βH_2	m
	A39 βH_3	m		L45 βH_2	vw
L41 NH	Q40 βH_2	m		R49 NH	m
	Q40 γH_2	w	αH	R49 βH_2	vw
	I42 NH	w	αH	L45 δH_3	vw
I42 NH	I42 βH	s	R49		
	L41 βH_2	m/w	NH	N47 αH	w
	I42 γH_2	m	NH	R49 GH_2	w
	I42 γH_3	s	NH	N47 βH_2	w
	I42 δH_3	s	NH	R49 βH_2	s
	A43 NH	w	NH	R49 γH_2	s
A43 NH	I42 γH_2	w	NH	L45 δH_2	w
	I42 δH_2	w	NH	K50 NH	w
	142 δH	w	side NH	R49 βH_2	m
	I42 γH_3	w	side NH	R49 γH_2	vw
	T44 NH	w	side NH	I51 $\gamma H_3/\delta H_3$	vw
T44 NH	A43 βH_3	m	K50 NH	R49 δH_2	w
	T44 γH_3	s		K50 βH_2	m
	T44 βH	s		I51 NH	w
	I42 δH_3	m	I51		
	L45 NH	w	NH	R49 γH_2	w
L45 NH	T44 γH_3	m		K50 βH_2	w
	L45 βH_2	m		I51 βH	m
	L45 γH_2	m		I51 $\gamma H_3/\delta H_3$	m
	L45 δH_3	m	βH	R49 δH_2	vw
	R49 βH_2	vw	S52		
K46 NH	L45 βH_2	vw	NH	I51 βH	s
	L45 γH_2	vw		I51 γH_2	m
N47				O51 $\gamma H_3/\delta H_3$	m
NH	K46 βH_2	w		S52 βH_3	s
NH	G48 αH_2	w	βH_3	L55 δH_3	w
NH	G48 NH	w	L53 NH	S52 βH_3	s
side NH1	N47 NH	m/w		L53 $\beta H_2/\gamma H$	s
side NH1	N47 βH_2	s		L53 δH_3	m
side NH1	K46 γH_2	vw	D54 NH	S52 βH_3	m
side NH1	L45 γH	vw		D54 βH_2	s
side NH1	R49 γH_2	vw		L53 βH_2	s
side NH2	R49 δH_2	vw	L55 NH	D54 βH_2	m
side NH2	L45 γH	w		S52 βH_3	vw
side NH2	R49 γH_2	w		L55 βH_2	s
side NH2	R49 αH	w	Q56 NH	L55 βH_2	vw
side NH2	N47 βH_2	m		L55 δH_3	vw
side NH2	N47 NH	w		Q56 βH1	s
side NH2	R49 NH	w		Q56 βH2	m/w
side NH2	A43 αH	vw		Q56 γH_2	m/w
side NH2	A43 βH_3	vw	A57 NH	Q56 βH1, βH2	vw
				Q56 NH	vw
				A57 βH_3	s
			O=CNH$_2$	A57 αH	m

[a] Distances used are s, 2.2−2.8 Å; m 2.2−3.5 Å; w, 2.2−4.2 Å; and vw, 2.2−5 Å. For each side chain involved in an NOE, an additional 0.5 Å was added to compensate for motional uncertainty.

Two other local conformational populations are noteworthy. The segment S52−L55 shows some chain reversal character, with $d_{\beta N}(i,i+2)$ and $(i,i+3)$ NOEs being observed between S52−D54 and S52−L55, respectively. A series of d_{NN} NOEs from L41 to L45 and N47 γNH_2/A43 αH, βH_3 NOEs suggest the presence of a small population of nascent helix conformation from L41 to N47. Moreover, the number of NOEs observed from N47 γNH_2 suggests reduced N47 side-chain mobility and the presence of a preferred side-chain conformation. Asparagine has been shown to form the C-cap position in helices (Zhou et al., 1994), presumably by hydrogen bond formation between the side-chain γNH_2 protons and backbone carbonyls. This apparent nascent

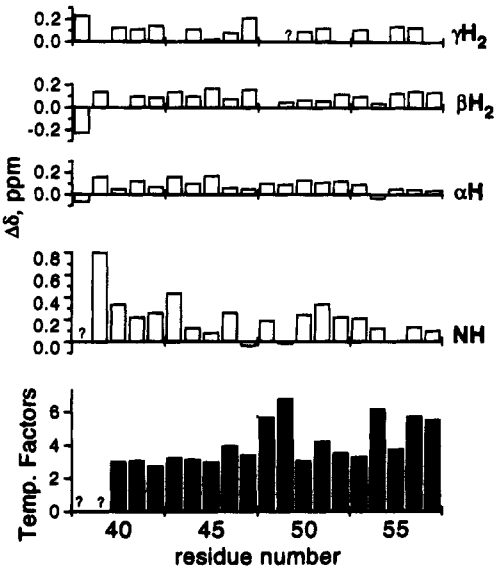

FIGURE 7: Chemical shift differences and temperature factors for 38−57. NH and αH chemical shift differences were determined by subtracting a particular peptide 38−57 proton resonance from its respective random coil chemical shift (Wuthrich, 1986). In this respect, a positive difference refers to a downfield chemical shifted resonance in the peptide. Temperature factors were calculated as described in the legend to Figure 4 and under Materials and Methods. Resulting temperature factors are plotted as parts per billion (ppb) per kelvin (K), i.e., ppb K^{-1}, versus the peptide residue number.

"helix" conformation is found within the same sequence noted above in peptide 24−46.

A plot of the NH temperature factors for peptide 38−57 (Figure 7) shows that most lie between 3 and 4 ppb/K. This is relatively small and comparable for what one might expect of a hydrogen-bonded group in a folded protein (Wüthrich, 1986) or from clustering of hydrophobic side chains which also can cause quite extreme solvent protection of amide protons (Dyson et al., 1992). At the very least, these results suggest solvent protection of these NHs. G48, R49, D54, Q56, and A57 are most exposed to solvent.

Figure 7 also gives αH and NH chemical shift differences, Δδ, between peptide 38−57 resonances and random coil resonances (Wüthrich, 1986). Many values are signficantly larger than the mean value, supporting the presence of non-random-coil structure. The largest deviations generally correlate with those residues which give the lowest temperature factors.

Distance geometry calculations were performed in order to assess various possible conformational families and compare resulting structures with those known to form in native PF4. NOEs listed in Table IV were used in a standard distance geometry protocol (Biosym suite of programs). A total of 40 structures were calculated twice, both times starting from extended chain conformations. Resulting structures were screened for large deviations in normally allowed Ramachandran phase space. As expected, a chain reversal is found from L45 to R49; 20 selected structures are overlayed with respect to the T44−K50 sequence (Figure 8). The average RMSD value from the native conformation for this sequence is 1.58 Å. For the L45−R49 sequence, the RMSD value drops to 1.1 Å. Such small deviations are surprising coming from a short linear peptide and indicate considerable conformational preference within this region. Moreover, no distance violations from input NOE constraints

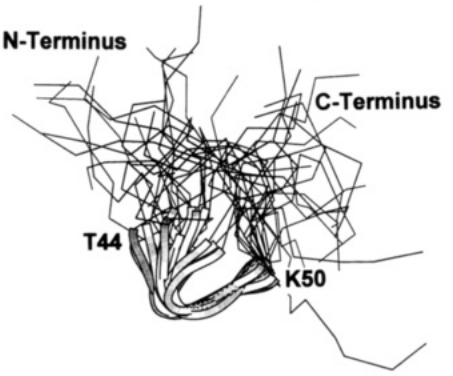

FIGURE 8: Distance geometry structures for peptide 38−57. Using NOE-derived distance constraints for peptide 38−57, distance geometry calculations yielded structures overlayed in this figure. At left is shown the superposition of 20 such generated structures overlayed by RMS minimization of the structural coordinates for backbone atoms of residues T44−K50. The program MOLSCRIPT (Kraulis, 1991) was used to produce this figure.

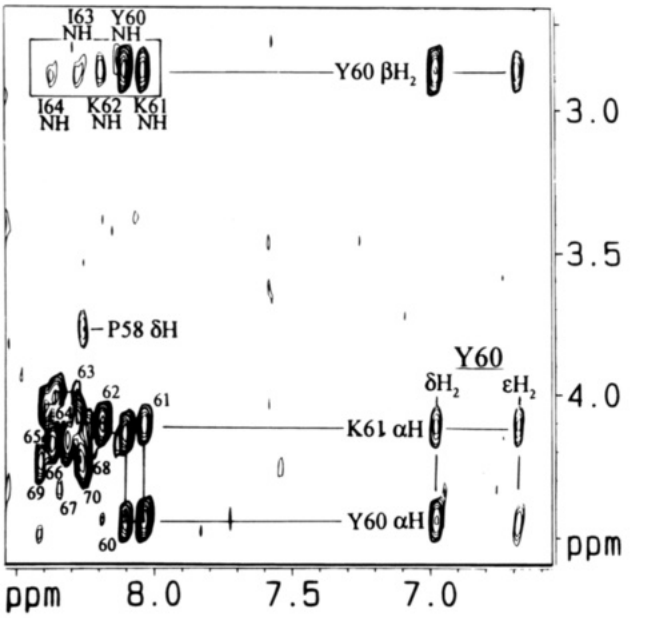

FIGURE 9: NOESY spectra for peptide 56−70. The αCH−NH fingerprint (top panel) and NH−NH (bottom panel) resonance regions from a NOESY contour plot are shown for peptide 56−70, pH 5.3, 10 °C. Data were collected and processed as described in the legend to Figure 3 and under Materials and Methods. The mixing time was 0.5 s. Sequential resonance assignments are traced out, and some longer range NOEs are indicated. Labeling of resonances is as discussed in the text.

were found, supporting the idea that observed NOEs are consistent with one conformational family.

It is important to note that NOE data and DG calculations also show a chain reversal at L53−L55 and a helix-like conformation from I42 to L45. A chain reversal from D54 to A57/P58 does exist in native PF4. Formation of the I42−L45 helix-like segment allows for additional hydrophobic interactions among I42, A43, L45, I51, and L53.

(*C*) *Peptide 56−70.* In native PF4, this C-terminal peptide forms an amphipathic α-helix which folds over onto a hydrophobic surface of the three-stranded β-sheet structure (see Figure 1).

In aqueous solution, several conformationally constraining NOEs are consistent with weak helix formation. NOESY data in Figure 9 show Y60/K62 $d_{\alpha N(i,i+2)}$, $d_{\beta N(i,i+2)}$, Y60/I63 $d_{\beta N(i,i+3)}$, and Y60/I64 $d_{\beta N(i,i+4)}$ NOEs suggesting helix-like

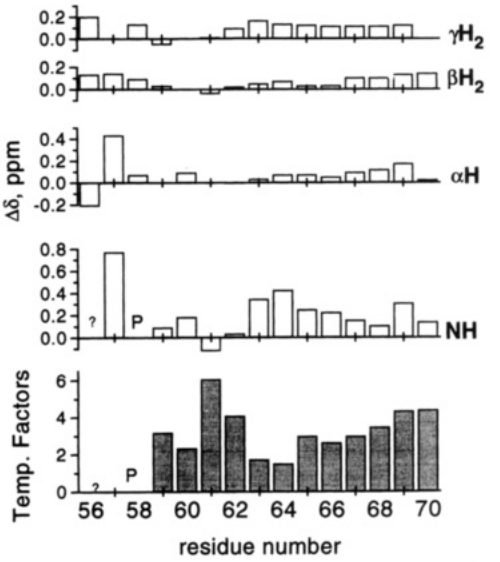

FIGURE 10: Chemical shift differences and temperature factors for 56−70. NH and αH chemical shift differences were determined by subtracting a particular peptide 56−70 proton resonance from its respective random coil chemical shift (Wuthrich, 1986). A positive difference refers to a downfield chemical shifted resonance in the peptide. Temperature factors were calculated as described in the legend to Figure 4 and under Materials and Methods. Resulting temperature factors are plotted as parts per billion (ppb) per kelvin (K), i.e., ppb K^{-1}, versus the peptide residue number.

structure from Y60 to I64. Three very weak d_{NN} NOEs (L59/Y60, K61/K62, and L68/E69) (data not shown) support the presence of nascent helix formation. Due to resonance overlap, other potentially observable d_{NN} NOEs could not be identified.

Chemical shift differences, Δδ, from random coil positions (Figure 10) indicate little change in αH shifts, but more substantial shifts in NH resonances. In fact, the largest NH chemical shift differences are found for A57, I63, I64, and E69. For A57, it is unclear whether N-terminal effects contribute to observed Δδ values, but for I63, I64, and E69, Δδ NH values suggest conformational contributions. Moreover, assuming helix structure, the NHs of I63 and I64 would be positioned to hydrogen bond to carbonyls of Y60 and K61. In further support of this, temperature factors (Figure 10) for I63 and I64 are reduced relative to other NH groups.

DISCUSSION

Each of the three PF4 peptides studied gave some conformationally meaningful NMR/CD parameters indicating the presence of populations of preferred conformation within a highly dynamic equilibrium ensemble. Peptide 24−46, which forms one antiparallel β-sheet structural unit in native PF4, shows NMR parameters characteristic of structural preference, possibly a native-like chain reversal, within the $A^{32}GPHSP^{37}$ sequence. Reduced temperature factors suggest hydrophobic side-chain clustering, while narrow NMR line widths indicate rapid exchange (on the 600 MHz chemical shift time scale) among low-energy conformers as normally observed in short linear peptides. Helix-like NOEs withn the $L^{41}IATLK^{46}$ sequence indicate the presence of non-native structural populations. The same sequence in peptide 38−57 shows similar nascent helix-like NOE patterns. The significance of this will be discussed later. NMR/CD data on peptide 56−70 suggest very weak, nascent helix formation within the native PF4 helix domain, residues L59−S70.

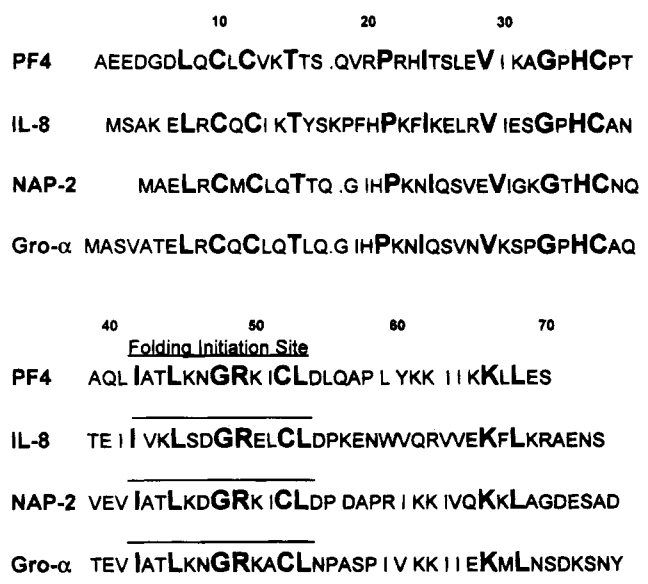

FIGURE 11: PF4 family sequence comparisons. Amino acid sequences of PF4, interleukin-8 (IL-8), neutrophil activating peptide-2 (NAP-2), and growth-related protein-α (Gro-α) are shown. The four conserved cystines are aligned for all sequences. Identical residues are enlarged in boldface letters.

Although these results allow one to suggest that these weak conformational populations probably do form relatively early on the PF4 folding pathway, neither of these two peptides (24−46 and 56−70) aparently possesses sufficient sequence to allow for significant chain collapse and conformational stabilization. On the other hand, peptide 38−57, which forms the second major antiparallel β-sheet structural unit in native PF4, does form relatively stable conformational populations in aqueous solution. The sequence $L^{45}KNGRK^{50}$ provides a significant number of local NOEs. Distance geometry calculations based on these NOE constraints yield a family of turn/chain reversal structures that gives average backbone RMS deviations from the native conformation of less than 1.1−1.58 Å for the sequence T44−K50. Reverse turns or chain reversals are usually thought of as likely sites for initiation of protein folding since they are determined by short-range interactions, limit the conformational space available to the polypeptide chain, and bring more distant parts of the chain together to direct subsequent folding events (Jaenicke, 1991). Since these turn/chain reversal conformational populations in peptide 38−57 are relatively stable in aqueous solution, it is likely that they will persist in longer fragments and, therefore, in the native sequence during initial folding steps.

Comparison of homologous sequences from PF4, IL-8, NAP-2, and Gro-α (Figure 11) shows that 18 residues are identical. Of these, 4 are cystines which form the 2 disulfide bridges apparently late in the folding process (Mayo et al., 1992), and 10 are hydrophobic. Of these 18 residues, 6 are found within the L41−L53 sequence. If one were to include conserved residue types, L41−L53 would be the most highly conserved sequence in this protein family. Second, backbone amide H/D exchange results for PF4 (Mayo et al., unpublished results), IL-8 (Clore et al., 1990), NAP-2 (Mayo et al., 1994), and Gro-α (Fairbrother et al., 1993) indicate that this part of the antiparallel β-sheet segment contains the most slowly exchanging NH protons. Kim et al. (1993) recently have proposed that the most slowly exchanging amides define the protein folding core. Therefore, it is logical to propose that the L41−L53 domain with K45−R49 chain reversal is a (if not the) major folding initiation site in PF4.

In peptide 38−57, hydrophobic side-chain clustering is suggested by reduced NH temperature factors and the hydrophobic nature of this sequence. Moreover, some peptide sequences from other proteins that we have worked with are less hydrophobic and yet are relatively insoluble in water alone. The observation supports the idea that hydrophobic side-chain solvent screening is, in this case, conformationally dependent. Out of its 20 residues, peptide 38−57 has 4 Leu, 3 Ala, and 2 Ile. In the native structure, these hydrophobic residues pair in an antiparallel β-sheet fold, i.e., L41 with L53, A43, with I51, and L45 with R49. It may be proposed that since the L45−R49 chain reversal is relatively stable, hydrophobic effects from "sticky ends" contribute to this stability as promoted in the "hydrophobic zipper" model of protein folding (Chan & Dill, 1990; Dill et al., 1993).

Of further note is the appearance of a helix-like population from L41 to N47. This conformation would tend to continue the chain reversal from N47 into the native state β-strand residues T38−L45. One question which arises is whether or not this non-native conformation is involved in protein folding (Darby et al., 1992; Weissman & Kim, 1991; Matthews, 1993). Chan and Dill (1990) have proposed that protein folding may be driven by "hydrophobic collapse." Formation of this helix-like conformation in peptide 38−57 would promote increased hydrophobic interactions among I42, A43, L45, I51, and L53 without apparent disruption of the L45−R49 chain reversal. Moreover, some nascent helix-like conformation also has been observed within the same sequence in peptide 24−46. We propose, therefore, that at least in the case of PF4, such a non-native helix-like conformation may be relevant to the folding pathway.

The following PF4 folding sequence may be proposed. The segment L41−L53, with a distinct chain reversal from L45 to R49, folds first. This key native-like conformation is apparently stabilized by hydrophobic interactions from flanking N- and C-terminal sequences. The relatively stable L45−R49 chain reversal sets up antiparallel β-sheet folding from T38 to L53. The P34−P37 sequence contributes to formation of a second chain reversal. The remainder of the β-sheet (see Figure 1) is then formed since the longest-lived NHs are found within the three-stranded β-sheet domain and not only from L41 to L53. The C-terminal helix and N-terminal domains then fold onto and are stablized by the three-stranded, antiparallel β-sheet hydrophobic surface. The two cystine disulfide bonds form last in order to structurally stabilize the native PF4 conformation (Mayo et al., 1992). In addition, it appears that the final stages of PF4 folding are thermodynamically linked to subunit association (Mayo et al., 1992).

ACKNOWLEDGMENT

NMR experiments were performed at the University of Minnesota High Field-NMR Laboratory. Peptides were synthesized at the Microchemical Facility, Institute of Human Genetics, University of Minnesota. We are very grateful to Dinesha S. Walek for her expertise in peptide synthesis.

REFERENCES

Baldwin, R. L. (1991) *Chemtracts: Biochem. Mol. Biol. 2*, 379–389.

Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84–91.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bierzynski, A., Kim, P. S., & Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2470–2474.

Buechner, J., Renner, M., Lilie, H., Hinz, H. J., Jaenicke, R., Kiefhabel, T., & Rudolph, R. (1991) *Biochemistry 30*, 6922–6929.

Chaffotte, A., Giullou, Y., Dellepierre, M., Hinz, H. J., & Goldberg, M. E. (1991) *Biochemistry 30*, 8067–8074.

Chaffotte, A. F., Cadieux, C., Giullou, Y., & Goldberg, M. E. (1992) *Biochemistry 31*, 4303–4308.

Chan, H. S., & Dill, K. A. (1990) *Proc. Natl. Acad. U.S.A. 87*, 6388–6392.

Chen, Y. H., Yang, T. Y., & Chau, K. M. (1974) *Biochemistry 13*, 3350–3359.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 48*, 45–148.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 1689–1696.

Darby, N. J., van Mierlo, C. P. M., Scott, G. H., Neuhaus, D., & Creighton, T. E. (1992) *J. Mol. Biol. 224*, 905–911.

Deuel, T. F., Keim, P. S., Farmer, M., & Heinrikson, R. L. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2256–2258.

Dill, K. A., Fiebig, K. M., & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1942–1946.

Dyson, H. J., Rance, M., Houghton, R. A., Wright, P. E., & Lerner, R. A. (1988) *J. Mol. Biol. 201*, 201–217.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992) *J. Mol. Biol. 226*, 819–835.

Fairbrother, W. J., Reilly, D., Colby, T., & Horuk, R. (1993) *FEBS Lett. 330*, 302–306.

Finkelstein, A. V. (1991) *Proteins: Struct., Funct., Genet. 9*, 23–27.

Goldenberg, D. P., & Creighton, T. E. (1985) *Biopolymers 24*, 167–182.

Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., & Klump, H. (1979) *Biophys. Chem. 9*, 137–147.

Holt, J. C., Harris, M. E., Holt, A., Lange, E., Henschen, A., & Niewiarowski, S. (1986) *Biochemistry 25*, 1988–1996.

Jaenicke, R. (1991) *Biochemistry 30*, 3147–3161.

Jeener, J., Meier, B., Backman, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4550.

Kim, K.-S., Fuchs, J. A., & Woodward, C. K. (1993) *Biochemistry 32*, 9600–9608.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631–660.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.

Kuwajima, K. (1992) *Curr. Opin. Biotechnol. 3*, 462–467.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–975.

Marion, D., Ikura, M., Tschudin, R., & Bax, A (1989) *J. Magn. Reson. 85*, 393–399.

Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653–683.

Mayo, K. H., & Chen, M. J. (1989) *Biochemistry 28*, 9469–9478.

Mayo, K. H., Barker, S., Kuranda, M. J., Hunt, A. J., Myers, J. A., & Maione, T. E. (1992) *Biochemistry 31*, 12255–12265.

Mayo, K. H., Yang, Y., Daly, T. J., Barry, J. K., & La Rosa, G. J. (1994) *Biochem. J.* (in press).

Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6805.

Ptitsyn, O. B. (1987) *J. Protein Chem. 6*, 273–293.

Ptitsyn, O. B. (1991) *FEBS Lett. 285*, 176–181.

Ptitsyn, O. B., & Semisotnov, G. V. (1991) in *Conformations and Forces in Protein Folding* (Nall, B. T., & Dill, K. A., Eds.) pp 155–168, American Association for the Advancement of Science, Washington, D.C.

Robson, I. J., & Frieden, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7222–7226.

Segawa, S.-I., & Sugihara, M. (1984) *Biopolymers 23*, 2473–2488.

Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 161–169.

Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., & Dyson, H. J. (1993a) *Biochemistry 32*, 6348–6355.

Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J., & Wright, P. E. (1993b) *Biochemistry 32*, 6356–6364.

Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.

Sreerama, N., & Woody, R. W. (1993) *Anal. Biochem. 209*, 32–44.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–293.

St. Charles, R., Walz, D. A., & Edwards, B. F. P. (1989) *J. Biol. Chem. 264*, 2092–2099.

Stewart, J. M., & Young, J. D. (1984) *Solid Phase Peptide Synthesis*, 135, 2nd ed., Pierce Chemical Co., Rockford, IL.

Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., & Clore, G. M. (1993) *Science 260*, 1110–1113.

Vaucheret, H., Signon, L., Le Bras, G., & Garel, J. R. (1987) *Biochemistry 26*, 2785–2790.

Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 6337–6347.

Waterhous, D. V., & Johnson, W. C. (1994) *Biochemistry 33*, 2121–2128.

Weissman, J. S., & Kim, P. S. (1991) *Science 253*, 1386–1393.

Wider, G., Macura, S., Anil-Kumar, Ernst, R. R., & Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207–234.

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

Zhou, H. X., Lyu, P. C., Wemmer, D. E., & Kallenbach, N. R. (1994) *J. Am. Chem. Soc. 116*, 1139–1140.