Multiple native-like conformations trapped via self-association-induced hydrophobic collapse of the 33-residue β -sheet domain from platelet factor 4

Elena ILYINA and Kevin H. MAYO*

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

Native platelet factor 4 (PF4) (70 residues) has a hydrophobic three-stranded anti-parallel β -sheet domain on to which is folded an amphipathic C-terminal α -helix and an aperiodic N-terminal domain. The 33-amino acid β -sheet domain from PF4 (residues 23–55) has been synthesized and studied by c.d. and n.m.r. At 10 °C and low concentration, peptide 23–55 appears to exist in aqueous solution in a random-coil distribution of highly flexible conformational states. Some preferred conformation, however, is observed, particularly within a relatively stable chain reversal from Leu-45 to Arg-49. As the peptide concentration and/or temperature is increased, a new conformational state(s) appears and intensifies as slowly exchanging (600 MHz ^1H -n.m.r.

chemical-shift time scale) random-coil resonances disappear. Hill plots of the concentration-dependence indicated mostly tetramer formation as found in native PF4. Although apparent resonance linewidths in aggregate state(s) are of the order of 100 Hz, sequence-specific assignments for most resonances could be made. N.m.r./nuclear Overhauser effect structural analysis indicates the formation of multiple native-like anti-parallel β -sheet conformations, kinetically trapped via subunit-association-induced hydrophobic collapse and stabilized by low-dielectric electrostatic interactions among/between Gly-28 and Lys-50 in opposing subunits. Results are discussed in terms of protein folding.

INTRODUCTION

Platelet factor (PF4) is a 70-residue protein [1] with a three-stranded anti-parallel β -sheet domain on to which is folded a C-terminal amphipathic α -helix and an aperiodic N-terminal region [2]. Figure 1 gives the general backbone folding scheme of the anti-parallel β -sheet domain. The N-terminal region is covalently constrained to the β -sheet domain by the presence of two cystine disulphide bridges [3]. Depending on solution conditions, PF4 can form homodimers and tetramers which are in slow exchange on the 600 MHz ^1H -n.m.r. chemical-shift time scale [4]. Reduction of the disulphides shifts the folding equilibrium into a monomeric molten-globule state which is more stable at lower pH [5]. Raising the pH induces a native-like folding process that is thermodynamically linked to dimerization [5]. On the basis of recent evidence that the protein-folding core correlates well with the presence of the five to ten longest-lived backbone NHs [6], it appears that the anti-parallel β -sheet domain with six to eight longest-lived NHs forms the PF4 folding core [7].

It is generally accepted that polypeptides fold into a native conformation via a self-assembly process [8] which occurs along some defined pathway(s) with structured intermediates [9]. Study of these highly transient intermediates normally requires the use of indirect methods such as n.m.r. proton/deuteron trapping [10–12], kinetic refolding of protein-engineered mutants [13] or microcalorimetric thermodynamic dissection [14]. In addition, peptide fragment conformational analysis has been useful for elucidating potential folding-initiation sequences, for example, in myoglobin [15–17] and cytochrome *c* [18]. If a particular amino acid sequence forms some stable conformation, it probably folds in a similar fashion in larger peptide fragments (or proteins) that

include that sequence and therefore may define a folding-initiation site. Secondary structures, especially helices and reverse turns, are considered to play a significant role as folding-initiation sites [19].

In PF4, the turn conformation about residues Leu-45–Arg-49 has been recognized as one such folding-initiation site [7]. Moreover, this relatively well-defined turn conformation is apparently stabilized by hydrophobic interactions from residues flanking both N- and C-terminal sides of the turn [7]. Although not dismissing the proposal that secondary-structure formation precedes (and perhaps directs) the collapse of the polypeptide chain into a compact structure [20], the hydrophobically stabilized PF4 folding-initiation site supports the Chan and Dill [21] hypothesis concerning initial collapse of a hydrophobic core and the ‘hydrophobic zipper’ model of protein folding [22]. As short linear peptides derived from structural units in native PF4 show specific conformational constraints, the purpose of the present study is to investigate folding potential in the entire PF4 β -sheet domain.

MATERIALS AND METHODS

Peptide synthesis

A 33-residue peptide comprised of amino acids 23–55 from human PF4 was synthesized on a Milligen Bioscience 9600 automated peptide synthesizer. The procedures used were based on Merrifield solid-phase synthesis utilizing Fmoc-BOP chemistry [23]. After the sequence has been obtained, the peptide support and side-chain protection groups were cleaved with acid (trifluoroacetic acid and scavenger mixture). Crude peptides were

Abbreviations used: PF4, platelet factor 4; 2D-n.m.r., two-dimensional n.m.r. spectroscopy; DQF-COSY, double-quantum-filtered shift-correlated 2D-n.m.r. spectroscopy; HOHAHA, homonuclear Hartman–Hahn 2D-n.m.r. spectroscopy; n.O.e., nuclear Overhauser effect; NOESY, nuclear Overhauser effect 2D-n.m.r. spectroscopy; TPPI, Time Proportional Phase Incremented.

* To whom correspondence should be addressed.

Figure 1 Schematic representation of PF4 backbone folding

PF4 peptide 23–55 in an anti-parallel β -sheet fold found for native PF4 is shown. The one-letter code for amino acids is used. Arrows indicate probable backbone amide hydrogen bonds. The C-terminus has been synthesized in the amide form as shown.

analysed for purity on Hewlett–Packard 1090M analytical h.p.l.c. apparatus using a reversed-phase C18 VyDac column. Peptides were generally about 90% pure. Further purification was achieved on a preparative reversed-phase h.p.l.c. C18 column using an elution gradient of 0–60% acetonitrile with 0.1% trifluoroacetic acid in water. Peptide 23–55 was then analysed for amino acid composition on a Beckman 6300 amino acid analyser by total hydrolysis of samples using 6M HCl at 110 °C for 18–20 h. N-terminal sequencing confirmed the purity of the peptide.

As peptide 23–55 was excised from the central portion of native PF4, the Leu55 C-terminal carboxylate was synthesized in

the amide form. To avoid potential problems with cysteine oxidation, Cys-36 and Cys-52 (Figure 1) were replaced by serines.

N.m.r. measurements

For n.m.r. measurements, freeze-dried peptide was dissolved in either $^2\text{H}_2\text{O}$ or $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ (9:1, v/v). Protein concentration was normally in the range 1–5 mM. pH was adjusted by adding microlitre quantities of NaOH or ^2HCl to the peptide sample. All n.m.r. spectra were acquired on a Bruker AMX-600 n.m.r. spectrometer.

For sequence-specific resonance assignments, double-quantum-filtered shift-correlated n.m.r. spectroscopy (DQF-COSY) [24,25] and two-dimensional (2D) homonuclear magnetization-transfer (HOHAHA) spectra, obtained by spin-locking with an MLEV-17 sequence [26] with a mixing time of 60 ms, were used to identify most spin systems completely. Nuclear Overhauser effect n.m.r. spectroscopy (NOESY) experiments [27,28] were performed to connect spin systems sequentially and later to identify nuclear Overhauser-effect (n.O.e) connectivities. All 2D-n.m.r. spectra were acquired in the Time Proportional Phase Incremented (TPPI) [29] or States-TPPI [30,31] phase-sensitive mode. The water resonance was suppressed by direct irradiation (0.8 s) at the water frequency during the relaxation delay between scans as well as during the mixing time in NOESY experiments.

The majority of the 2D-n.m.r. spectra were collected as 256 or 512 t_1 experiments, each with 1 k or 2 k 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 off-line 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 1 k in the t_1 dimension before 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 [32].

The fraction of peptide in the aggregated state relative to that in the monomeric state was estimated by using the ‘cut and weigh’ method. Initially, an n.m.r. spectral trace of peptide 23–55 mostly in the aggregated state (more than 95%) was used to subtract out the fractional population in the monomeric state. This could be done because the NH resonance region from 8.7 p.p.m. to 9.6 p.p.m. contains only aggregate-state resonances, and the remainder of the downfield aggregate-state resonance intensity which overlapped with monomer-state resonances could be scaled fairly accurately to that intensity. The scaled aggregate-state intensity was then traced on to a given spectrum and then cut out and weighed. The error in these estimates is about $\pm 10\%$ and somewhat better when the fraction of aggregate is between 0.2 and 0.8.

C.d. spectroscopy

C.d. spectra were recorded 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 baseline subtraction. Spectra were recorded from 5 to 65 °C in the presence of 10 mM phosphate buffer, pH 5.3, over a 180–250 nm range using a 0.5 mm 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 eight times, and an equally signal-averaged solvent baseline was subtracted. C.d. spectra were analysed by the method of Sreerama and Woody [33].

RESULTS

C.d. spectra

As a function of temperature from 5 to 65 °C, peptide 23–55 (low concentration, 14 μ M) gives c.d. spectra (Figure 2) with negative ellipticity bands at 199 and 222 nm. Irrespective of temperature, each c.d. trace appears to give typical random-structure distributions as predicted for mostly unstructured peptides [34–36]. In 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. Increasing the temperature tends to make the 199 nm band more positive and the 222 nm band more negative. The fractional change is greater at lower

temperature as shown in the top inset in Figure 2. By 45–50 °C, fractional changes are relatively small and appear to be levelling out. This observation may indicate a role for the hydrophobic effect [37] in stabilizing various conformational populations. Peptide 23–55 is at least 50 % hydrophobic. On the other hand, such temperature effects in c.d. spectra of short linear peptides may be due merely to the physical properties of the system rather than to conformational stabilization, for even though the ratio $[\Theta_{199}]/[\Theta_{222}]$ changes with temperature, the ratios $[\Theta_{199}]^{5^\circ\text{C}}/[\Theta_{199}]^{65^\circ\text{C}}$ and $[\Theta_{222}]^{5^\circ\text{C}}/[\Theta_{222}]^{65^\circ\text{C}}$ are almost constant.

C.d. spectra analysed by the method of Sreerama and Woody [33] provide fractional conformational populations shown in Figure 2 (bottom inset). Since this approach is based on a structural database for folded proteins, caution in interpreting these distributions should be exercised. Fractional distributions show that α -helicity increases fairly linearly from 5° to 65 °C at the expense of other conformational populations. Interestingly, unordered structure tends to decrease and not increase as expected if any unfolding were occurring. As will become evident below, aggregation at this peptide concentration can account for no more than about 0.1 % of the total peptide population. Therefore any effects of temperature are within the apparent random-coil distribution.

N.m.r. spectroscopy

Self-association of peptide 23–55

The bottom trace of Figure 3(a) shows an n.m.r. spectrum for peptide 23–55 at 10 °C and at a concentration of 2.3 mM, about 150 times higher than that used for c.d. measurements. Resonance linewidths are relatively narrow as expected for a peptide of this size. The n.m.r. spectrum, moreover, is characteristic of a random-coil distribution of conformational states as indicated by the above c.d. data. Increasing the temperature (Figure 3a) induces new NH and α H resonances to appear about 0.5–1.0 p.p.m. downfield of random-coil ones. The new NH and α H resonance envelopes are considerably broader and increase in intensity as random-coil resonances decrease. These observations are consistent with the presence of a slow exchange (600 MHz ^1H -n.m.r. chemical-shift time scale) between conformational states.

As increasing the peptide concentration (Figure 3b) enhances the population of these broader resonances at the expense of random-coil resonances, intermolecular associations must be occurring. Plotting the natural logarithm of n.m.r.-derived fractional populations (concentrations) in one state versus those in the other state gives a linear curve with a slope of about 4 for both 30 and 40 °C (see inset to Figure 3b). This indicates that the new conformational state is primarily tetrameric under these conditions, although some dimer formation or higher-order aggregates cannot be excluded. Native PF4 is known to exist in solution in a distribution of slowly exchanging (^1H -n.m.r. chemical-shift time scale) monomer–dimer–tetramer states, with the tetramer state also predominating [4]. In fact, the pH-dependence of peptide 23–55 association (Figure 3c) follows the same trend as observed for native PF4 [4], supporting the idea that peptide 23–55 preferentially forms tetramers. For the remainder of the paper, these slowly exchanging conformational states will be referred to as the monomeric state for non-associated random-coil-like populations and the tetrameric state for the subunit-associated state.

Increasing the pH stabilizes the tetrameric state (see Figure 3c insert) as it does in native PF4 [4]. For native PF4, it was proposed that the ionization state of Glu-28 modulates subunit association, although other carboxylates could also have been

Figure 2 C.d. spectra of peptide 23–55

Far-u.v. c.d. spectra for PF4 peptide 23–55 are shown as mean residue ellipticity versus wavelength (nm). Peptide concentration was 14 μ M in 10 mM potassium phosphate, pH 5.3. The temperature was varied from 5 to 65 °C. Other experimental variables are given in the Materials and methods section. The bottom inset shows the fractional conformational populations derived by using the method of Sreerama and Woody [33]; ∇ , unordered; Δ , turn; $\circ$, β -strand; and $\square$, α -helix. The top inset gives the fractional change in ellipticity for the 199 (■) and 222 (●) nm bands.

Figure 3 N.m.r. spectra for peptide 23-55

N.m.r. spectra of peptide 23-55 as a function of temperature (a), peptide concentration (40 °C) (b) and pH (40 °C) (c) are presented. For (a) and (c) the peptide concentration was 10 mg/ml. The inset in (b) shows the natural logarithm of the aggregate-state concentration versus the monomer concentration for two temperatures, 30 °C and 40 °C. The inset in (c) gives the intensity ratio of tetramer to monomer versus pH. Aggregate- and monomer-state fractions were estimated as described in the Materials and methods section.

responsible for this observation. In peptide 23-55, however, only Glu-28 and Asp-54 carboxylates are present and the C-terminal Leu-55 residue has been synthesized in the amide form. It is unlikely that the terminal Asp-54 ionization state would greatly influence subunit association in this way, given the structure of

native tetrameric human PF4 [38] where Glu-28 and Lys-50 from four subunits are engaged in electrostatic interactions within the low-dielectric protein core [2]. Results with peptide 23-55 therefore support the idea that Glu-28 has a significant role in the native PF4 subunit-association process.

Table 1 Resonance assignments for PF4 monomeric peptide 23–55

Conditions were 10 mM NaCl, pH 5.3, and 283 K.

Residue	Chemical shift (p.p.m.)					
	NH	C α H	C β H	C γ H	C δ H	Other
His-23	—	4.85	3.14, 3.06	—	—	8.20(2H) 7.11(4H)
Ile-24	—	4.16	—	—	—	—
Thr-25	8.47	4.26	4.08	1.11	—	—
Ser-26	8.39	4.34	3.75	—	—	—
Leu-27	8.35	4.15	1.50	1.50	0.76	—
Glu-28	8.27	4.14	1.89	2.12	—	—
Val-29	8.22	3.94	1.90	0.79	—	—
Ile-30	8.39	4.00	1.71	1.31, 1.06	0.76	—
Lys-31	8.44	4.18	1.66	1.35	1.56	2.87(C α H $_2$) 7.49(N ϵ H $_3$)
Ala-32	8.36	4.14	1.30	—	—	—
Gly-33	8.27	4.05, 3.91	—	—	—	—
Pro-34	—	4.29	2.13, 1.71	1.86	3.50	—
His-35	8.64	4.64	3.15, 3.06	—	—	8.43(2H) 7.17(4H)
Ser-36	8.32	4.65	3.80	—	—	—
Pro-37	—	4.42	2.24, 1.78	1.93	3.75, 3.65	—
Thr-38	8.13	4.15	—	1.11	—	—
Ala-39	8.27	4.15	1.28	—	—	—
Gln-40	8.32	4.15	1.88	2.26	6.85, 7.52 (N δ H)	—
Leu-41	8.21	4.24	1.53	1.53	0.79	—
Ile-42	8.14	3.99	1.75	1.39, 1.08	0.76	—
Ala-43	8.32	4.23	1.29	—	—	—
Thr-44	8.08	4.18	4.11	1.11	—	—
Leu-45	8.19	4.26	1.58	1.42	0.80	—
Lys-46	8.40	4.18	1.69	1.29	1.56	2.87(C α H $_2$) 7.49(N ϵ H $_3$)
Asn-47	8.42	4.55	2.75	7.59, 6.87 (N γ H)	—	—
Gly-48	8.31	3.81	—	—	—	—
Arg-49	7.99	4.22	1.69	1.49	3.09	7.15(N ϵ H)
Lys-50	8.41	4.26	1.66	1.32	1.58	2.87(C α H $_2$) 7.49(N ϵ H $_3$)
Ile-51	8.28	4.09	1.75	1.38, 1.07	0.78	—
Ser-52	8.37	4.38	3.72	—	—	—
Leu-53	8.37	4.25	1.49	1.49	0.76	—
Asp-54	8.30	4.48	2.64, 2.46	—	—	—
Leu-55	8.26	4.16	1.55	1.55	0.79	—

The well-known ‘hydrophobic’ effect [37], moreover, must contribute significantly to formation of tetrameric peptide 23–55. As the temperature is increased, tetramer populations increase dramatically as a result of hydrophobic stabilization, which shifts the normally fast-folding/unfolding equilibrium to the slow ^1H -n.m.r. exchange regime, making the lifetime of the tetrameric state longer than about 10 ms. PF4 peptide 23–55 contains 48 % non-polar residues (excluding glycine). If the partially hydrophobic arginine and lysine residues are counted, hydrophobic character is increased further.

N.m.r. conformational analysis of monomeric peptide 23–55

At 10 °C and relatively low concentration, peptide 23–55 exists mostly (more than 95 %) in the monomeric state. Sequence-specific resonance assignments derived for this state by standard 2D-n.m.r. techniques [32] are given in Table 1. Analysis of HOHAHA and DQF-COSY spectra allowed grouping of most spin systems, which were connected in a sequential fashion by using NOESY spectra. Some ambiguities were clarified by varying temperature and pH. This could be carried out, however, only to a limited extent because of the self-association discussed above. His-23 αH and βH_2 resonances could be found only in upfield regions of 2D spectra. His-23 and His-35-C2/C-4 proton resonances were initially identified from analysis of HOHAHA

spectra in which these long-range scalar connectivities showed up. The αN fingerprint region from a HOHAHA spectrum and the NH–NH region from a typical NOESY spectrum of monomeric peptide 23–55 are shown in Figures 4(a) and 4(b). The HOHAHA αN region is presented to show the relatively narrow resonances usually found for a peptide of this size. Although monomeric peptide 23–55 has mostly random-coil c.d. and n.m.r. spectral characteristics, some conformationally meaningful n.O.e.s can be identified (see Figure 4). In particular, several key n.O.e.s within the Leu-45–Arg-49 sequences support the formation of a chain reversal, as was found for the shorter PF4 peptide 38–57 [7]. In particular, Gly-48/Arg-49 and Arg-49/Lys-50 d_{NN} , Arg-49/Asn-47 $d_{\text{N}\beta}$ and Arg-49/Leu-45 $d_{\text{N}\delta}$ n.O.e.s are observed. However, some n.O.e.s within the Leu-45–Arg-49 sequence which were observed in peptide 38–57 are not observed in this longer fragment. The same is true for the Leu-41–Thr-44 nascent helix conformation previously reported in peptide 38–57. In peptide 23–55, only the Ile-42/Ala-43 d_{NN} and Asn-47/Ala43 $d_{\text{N}\beta}$ n.O.e.s were observed. Apparently, conformational populations have shifted somewhat relative to peptide 38–57 [7]. This indicates that using a longer peptide does not guarantee the observation of larger preferred conformational populations.

Chemical-shift differences, $\Delta\delta$, from random-coil chemical shifts [32] are given in Table 2. $\Delta\delta$ NH mean values are mostly positive, i.e. downfield shifted, with the greatest changes observed

Figure 4 HOHAHA and NOESY spectra of monomeric peptide 23–55

The α H–NH and NH–NH resonance regions from HOHAHA (a) and NOESY (b) contour plots respectively are shown for peptide 23–55 recorded at pH 5.3 and 10 °C. Data were collected in $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ (9:1, v/v). Protein concentration was 5 mg/ml. Some 512 hypercomplex free induction decays containing 1 k words each were collected and processed as discussed in the Materials and methods section. The NOESY mixing time was 0.5 s, and the HOHAHA spin-lock time was 60 ms. 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 before Fourier transformation. Labelling of resonances (one-letter code) is as discussed in the text.

Table 2 Chemical-shift changes and NH temperature factors for PF4 monomeric peptide 23–55

NH temperature factors were calculated as the slope of δ_{NH} versus temperature in Kelvin. Average $\Delta\delta$ was calculated as a simple average using all the values in a given column. Numbers in parentheses are the β -sheet/ α -helix chemical shift index as given by Wishart et al. [39]; a series of +1 indicates β -sheet tendency, a series of –1 indicates α -helix tendency and 0 indicates other.

Residue	Chemical-shift difference (p.p.m.)				Temperature factor (p.p.b./K)
	NH	α H	β H ₂	γ H ₂	
His-23	–	0.47* (+1)	0.08*	–	–
Ile-24	–	0.18 (+1)	–	–	–
Thr-25	0.48*	0.16 (0)	0.11*	0.13	3.0
Ser-26	0.26	0.09 (–1)	0.12	–	2.5
Leu-27	0.18	0.02 (0)	0.10	0.11	2.9
Glu-28	0.15	0.10 (–1)	0.05	0.06	1.2
Val-29	0.03	0.01 (0)	0.02	0.07	–
Ile-30	0.45	0.02 (0)	0.06	0.08	0.5
Lys-31	0.28	0.07 (–1)	0.06	0.15	3.6
Ala-32	0.36	0.04 (–1)	0.16	–	2.1
Gly-33	0.13	0.19 (0)	–	–	3.0
Pro-34	–	0.10 (–1)	0.1	0.08	–
His-35	0.48	0.26 (0)	0.14	–	3.3
Ser-36	0.19	0.40 (+1)	0.17	–	2.8
Pro-37	–	0.23 (0)	0.21	0.15	–
Thr-38	0.14	0.05 (–1)	–	0.13	3.4
Ala-39	0.27	0.05 (–1)	0.14	–	3.2
Gln-40	0.16	0.03 (–1)	0	0.13	3.1
Leu-41	0.04	0.11 (0)	0.13	0.13	3.7
Ile-42	0.20	0.01 (0)	0.10	0.16	3.0
Ala-43	0.32	0.13 (–1)	0.15	–	2.9
Thr-44	0.09	0.08 (–1)	0.14	0.13	2.6
Leu-45	0.02	0.13 (0)	0.18	0.19	3.5
Lys-46	0.24	0.07 (–1)	0.09	0.06	2.6
Asn-47	–0.08	0.05 (–1)	0.17	–	3.2
Gly-48	0.17	0.09 (–1)	–	–	5.8
Arg-49	–0.03	0.09 (–1)	0.05	0.04	7.0
Lys-50	0.25	0.15 (0)	0.06	0.12	2.6
Ile-51	0.34	0.11 (+1)	0.10	0.15	4.1
Ser-52	0.24	0.13 (–1)	0.09	–	1.9
Leu-53	0.20	0.12 (0)	0.09	0.10	3.1
Asp-54	0.14	–0.03 (–1)	0.05	–	6.4
Leu-55	0.09	0.03 (0)	0.15	0.16	2.2
Average	0.19	0.10	0.10	0.11	–

* $\Delta\delta$ calculated as PF4 peptide chemical shift minus random-coil chemical shift given by Wüthrich [32].

for Thr-25, Val-29, Ile-30, Ala-32, His-35, Leu-41, Ala-43, Leu-45, Asn-47, Arg-49, Lys-50, Ile-51 and Leu-55. $\Delta\delta$ α H mean values are also mostly positive, and ten of these are greater than 0.2 ppm. The structural nature of these chemical-shift changes is unknown. In native PF4, anti-parallel β -sheet α H and NH chemical shifts are shifted considerably further downfield than these. However, some transient native-like conformations are probably represented in this dynamic-equilibrium conformational ensemble. If one can take the relative chemical-shift change from native chemical shifts as a measure of fractional native-like conformation, then, for example, α H resonance-shift differences suggest a contribution of about 10 % from native-like PF4 conformation. Viewing α H chemical-shift changes in terms of the Wishart et al. [39] chemical-shift index for α -helix/ β -sheet tendency (Table 2) allows little to be said with regard to regions of secondary-structure preference. In this scale, a value of +1 or –1 over several sequential residues indicates formation

Figure 5 HOHAHA spectra of tetrameric peptide 23–55

The NH- α H/upfield fingerprint (a) and α H-upfield (b) resonance regions from a HOHAHA contour plot are shown for peptide 23–55 (30 mg/ml, pH 5.3, 50 °C and 10 mM NaCl). Data were collected and processed as described in the legend to Figure 4 and in the Materials and methods section. The spin-lock time was 50 ms. Sequence-specific resonance assignments (one-letter code) are indicated as discussed in the text.

of β -sheet or α -helix respectively. The only series of -1 values occurs within the sequence Lys-46–Arg-49. This sequence does form a helix-like chain reversal in native PF4 and in PF4 peptide 38–57 [7].

Temperature factors (Table 2) are unusually low for several NH resonances. In a 33-residue peptide, this may be the result of either NH hydrogen-bonding [32] or clustering of hydrophobic side chains [40]. To explain these rather low NH temperature factors by hydrogen-bonding, one would expect considerable structure stability, which is not observed in monomeric peptide 23–55 n.m.r./n.o.e. parameters. The more probable explanation is hydrophobic side-chain clustering, as concluded from results on PF4 peptide 38–57 [7].

N.m.r. conformational analysis of peptide 23–55 tetrameric state

The NH/ α H chemical-shift dispersion noted above for the peptide 23–55 tetrameric state appears remarkably similar to that observed for native PF4 where residues 23–55 fold in an anti-parallel β -sheet conformation [2,38]. The similarity in chemical shifts suggests that, even at this level of analysis, tetrameric peptide 23–55 forms native-like folds. Moreover, if one assumes that the newly folded state represents only one

tetramer species, monomer–tetramer equilibrium constants calculated from fractional populations (Figure 3) and the total peptide 23–55 concentration are in the same range as those found for native PF4 [4].

In order to identify structures in this apparently folded peptide 23–55 tetramer, HOHAHA and NOESY spectra were accumulated at high peptide concentration (30 mg/ml) and at 50 °C, conditions under which the tetrameric state is favoured (at least 90%) over the monomeric state. HOHAHA contour plots are shown in Figures 5(a) (downfield region) and 5(b) (upfield region). At this peptide concentration, even the presence of only 10% monomer makes monomer-state cross-peaks readily apparent because of their narrower linewidths. As monomer populations are so much lower than tetramer populations, increased monomer resonance broadening over that observed at 10 °C (Figure 4) is due to chemical exchange. For the sake of analysis and discussion, monomeric peptide 23–55 cross-peak positions have been shaded-in in Figure 5(a). Remaining cross-peaks arise from tetrameric peptide 23–55. Assigned tetramer resonances have been labelled as discussed below. These cross-peaks are considerably broader and more highly dispersed than monomer resonances. In fact, as mentioned above, the chemical-shift dispersion is similar to that observed for native PF4. For

Figure 6 HOHAHA spectrum of native PF4

The NH- α H/upfield resonance region from a HOHAHA contour plot is shown for native PF4. Data were collected and processed as described in the legend to Figure 4 and in the Materials and methods section. This PF4 sample is a chimaeric mutant containing the first eight interleukin 8 residues. There is no difference with regard to native PF4 folding and function. The spin-lock time was 40 ms. Sequence-specific resonance assignments (one-letter code) are indicated. Sample conditions were pH 5.9, 50 °C, 10 mM NaCl. Only cross-peaks for residues 23–55 have been labelled.

comparison, a HOHAHA spectrum of native PF4 is given in Figure 6. Only cross-peaks originating from residues 23–55 are labelled.

Although the eventual goal is to analyse the conformation of folded peptide 23–55 in the tetrameric state, sequence-specific assignments must be made first. In some cases, resonance assignments were relatively straightforward and could be approached in the standard way [32]. In other cases, analysis was complicated by resonance broadening which for a number of cross-peaks appears to be asymmetric. We shall return to this point later. Wherever possible, spin systems were initially grouped or partially grouped by HOHAHA spectral analysis (Figure 5). Combination of a larger population, shorter n.O.e. mixing times and larger rotational correlation times for peptide 23–55 tetramer had the advantageous effect of attenuating monomer cross-peak intensities in NOESY spectra (Figures 7a and 7b), which simplified analysis. The standard assignment approach worked well for residues Lys-46, Asn-47, Gly-48, Arg-49 and Lys-50 as traced out in Figure 7(a). Asn-47, Gly-48 and Arg-49 cross-peaks are most well resolved and least broadened (see Figure 5a as well). Two possibly reasons for this are greater internal mobility and

more clearly defined conformations. Cross-peak overlap and multiplicity made other assignments feasible by comparison with cross-peak shifts found in native PF4 (see Figure 6). For example, Ser-26, Lys-50, Cys-36 and Cys-52 (in peptide 23–55, they are Ser-36 and Ser-52) are well resolved in native PF4. HOHAHA spectra of tetrameric peptide 23–55 (Figure 5) show cross-peaks in the same region. Serine spin systems in particular can be identified from their α H- β H₂ AMX fragments. The Lys-50 α H chemical shift can be identified in Figure 5(b) by following that spin system from its ϵ H₂ resonance back to its α H resonance. Assignments for Glu-28, His-35 and Asp-54 followed from a similar analysis. Having made the Ser-52 and Asp-54 assignments, α N ($i, i+1$) NOESY cross-peaks were identified running from Ser-52 to Leu-53 to Asp-54 (see Figure 7a).

As it was observed that tetramer peptide 23–55 side-chain resonance chemical shifts generally vary little from their native PF4 counterparts, initial assignments for Thr-25, Val-29, Ile-30, Ile-42, Ala-43, Thr-44 and Ile-51 were made by following side-chain chemical shifts to respective α N cross-peaks. α N ($i, i+1$) n.O.e.s consistent with these assignments could then be identified from among several partially overlapping cross-peaks. Two additional assignments, i.e. Leu-41 and Leu-45, followed as traced out in Figure 7(a). Several $d_{\beta\alpha}$ ($i, i+1$) n.O.e.s supported these assignments. As no other upfield resonances overlapped with those of Pro-34 and Pro-37, proline resonance assignments were made by comparison with those in native PF4 and later confirmed by identifying δ H- α H ($i, i-1$) n.O.e.s from Gly-33 and Ser-36 respectively (Figure 7b). It appears therefore that both prolines exist predominantly in the trans-peptide state as both proline δ H- α H ($i, i-1$) n.O.e.s from Gly-33 and Ser-36 are relatively strong (see Figure 7b). Assignments for Ile-24, Lys-31, Ala-32, Thr-38, Ala-39 and Gln-40 are either ambiguous or could not be made. Their cross-peaks are either too broad, resonate under H₂O/²H₂O or overlap with monomer resonances. For Ala-32 this was surprising, because in native PF4, its α N cross-peak is relatively well resolved and shifted upfield from the H₂O/²H₂O resonance (see Figure 6).

Histidine C-2 and C-4 resonances were grouped in HOHAHA spectra and by analysis of the pH-dependence of their C-2 resonances (Figure 8). Under these solution conditions, the tetrameric state is predominant through most of the titration, although some monomer-state histidine C-2 populations could be resolved at lower pH values. Histidine C-2 titration in the tetrameric state was measured on this sample. For the monomeric state, the titration curve was recorded at 10 °C and at a lower concentration at which tetramer populations are minimal. By using the Henderson–Hasselbalch equation and assuming an overall chemical-shift change of 1 p.p.m. between fully protonated and fully unprotonated states [41], the two histidines in the tetrameric state gave different pK_a values of 5.2 and 6.3. The one with the lower pK_a also broadened greatly around its titration midpoint (see HOHAHA spectrum in Figure 8). This C-2/C-4 resonance broadening near pK_a values is usually the result of exchange between acidic and basic forms of histidine [42]. In peptide 23–55, apparent broadening may also be the result of multiple conformations (see below). In native tetramer PF4, similar pK_a and resonance broadening behaviour were noted for His-23 and His-35 [4]; pK_a values were 4.2 and 5.8 respectively. His-23 and His-35 C-2 and C-4 assignments for tetrameric peptide 23–55 were made by analogy with those in native PF4 (as labelled in Figure 8). Similarly reduced His-23 pK_a values suggest similar conformational environments because in random-coil monomeric peptide 23–55, both His-23 and His-35 pK_a values are essentially the same, i.e. 6.5 (Figure 8). In native PF4, Arg-22, Lys-46, Arg-49, Lys-50 and the four C-terminal lysines are

Figure 7 NOESY spectrum of tetrameric peptide 23–55

The NH- α H/upfield (a), and α H-upfield (b) resonance regions from a NOESY contour plot are shown for peptide 23–55 (30 mg/ml, pH 5.3, 50 °C and 10 mM NaCl). Data were collected in $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ (9:1, v/v). Some 512 hypercomplex free induction decays containing 1 k words each were collected and processed as discussed in the Materials and methods section. The mixing time was 0.2 s. The data were zero-filled to 1024 in t_1 . Raw data were multiplied by a 30° -shifted sine-squared function in t_1 and t_2 before Fourier transformation. Labelling of resonances (one-letter code) is as discussed in the text.

proximal to His-23, and in the tetrameric state these residues form a ring of positive charge which could explain the lower pK_a of His-23.

Table 3 shows NH and α H chemical-shift differences between random coil [32] and peptide 23–55 tetrameric chemical shifts. Only unambiguously assigned resonances are shown, and, when multiple α N cross-peaks are observed, the major species is used. Clearly, those residues that directly participate in the folded structure as β -sheet strands are downfield-shifted by about 0.5–1.0 p.p.m. These include Ser-26–Ile-30, Leu-41–Leu-45 and Lys-50–Asp-54. In general, observed NH and/or α H chemical shifts are closer to those found in native PF4 (see Figure 6) than in monomeric peptide 23–55. The formation of β -sheet structure is also correlated with α H chemical-shift changes based on the Wishart et al. [39] chemical-shift conformational index (see Table 3). A series of +1 values, which indicates β -sheet formation [39], is observed within the three sequence stretches given above.

As mentioned earlier, residues in tetrameric peptide 23–55

show either greatly broadened or multiple cross-peaks arising mostly from the NH shift dimension. Notice the Ser-52, Thr-44, His-35 and Glu-28 multiple cross-peaks. This is most clearly exemplified in Figure 5 with Ser-52 α N and β N cross-peaks. Although quite broadened, one major and several minor cross-peaks are apparent. These almost certainly indicate multiplicity rather than simply being broadened, as spin systems can be followed through to their respective βH_2 side-chain resonances (see Figure 5). This multiplicity must be due to the presence of multiple folded conformations.

At this level of analysis, conformational multiplicity, resonance broadening and cross-peak overlap make a complete structural derivation impossible. Several structurally informative n.O.e.s, however, are identified in Figure 7. In particular, three α H- α H n.O.e.s are evident between Ile-42–Ser-52, Thr-44–Lys-50 and Ser-26–Ile-30 (Figure 7b). In combination with the observation of greatly downfield-shifted NH and α H resonances (Table 3), their appearance, as in native PF4, indicates the formation of

Figure 8 His-23 and His-35 assignments and pK_a values

His-23 and His-35 C-2 chemical shifts are plotted against pH. Data are given for both monomeric (10 °C, 3 mg/ml) and tetrameric (50 °C, 10 mg/ml) peptide 23–55. For monomer-state chemical shifts recorded at 10 °C, the ordinate axis was shifted downfield by 0.4 p.p.m. In this way, His-23 and His-35 C-2 chemical shifts overlapped at low pH values. Solid lines connect data points as visual aids. A H_2O HOHAHA spectrum of tetrameric peptide 23–55 (50 °C, 30 mg/ml, pH 5.3, 10 mM NaCl) showing the histidine C-2/C-4 resonance region is given in the inset. His-23 and His-35 cross-peaks have been labelled (one-letter code) as discussed in the text. M, Monomer; T, tetramer.

anti-parallel β -sheet conformation. The Ile-42–Ser-52 and Thr-44–Lys-50 αH – αH proximities can be seen in the native backbone-folding scheme shown in Figure 1, while the Ser-26–Ile-30 αH – αH n.O.e. indicates AB dimer formation as observed in bovine PF4 [2] and human interleukin 8 [43] and PF4 (K. H. Mayo, S. Barker, E. Ilyina, T. J. Daly, J. K. Barry and G. J. La Rosa, unpublished work). n.O.e.s from Gly-48 αH to Thr-25 γH_2 and from His-35 C-4 to Arg-49 βH_2 and γH_2 are also observed in native PF4 between A(B) and D(C) subunits (K. H. Mayo, S. Barker, E. Ilyina, T. J. Daly, J. K. Barry and G. J. La Rosa, unpublished work), further supporting native-like peptide 23–55 tetramer formation. With the presence of multiple conformations, it is impossible to say whether observed n.O.e.s arise from one or more states. As it appears that all folded-state resonances arise from native-like structures, it is reasonable to assume that observed n.O.e.s represent some weighted average over all possible conformations. n.O.e. build-up curves for tetramer-state n.O.e.s show maximal negative n.O.e.s at about 150 ms, followed by diminished n.O.e. intensities most probably caused by spin diffusion at longer mixing times. This would be expected for a compact protein of molecular mass 15 kDa, i.e. tetrameric peptide 23–55.

DISCUSSION

In aqueous solution, sufficient information is present in the 33-

residue peptide to form native-like PF4 β -sheet folds and quaternary structural associations. As monomeric peptide 23–55 exhibits the c.d. and n.m.r. spectral characteristics of a random-coil conformational distribution, the folding process is thermodynamically linked to subunit association which, in any event, is an integral part of protein folding interactions between domains which differ from interactions between subunits of a multimeric protein primarily in the covalent continuity of the peptide backbone [44]. Increasing the temperature stabilizes the predominant tetrameric state, indicating that the subunit-association enthalpy term must be positive. Assuming that tetramer is the only aggregate state formed and disregarding the presence of multiple conformational states, the overall association enthalpy can be estimated from van't Hoff plots of $\ln K_{eq}$ (taken from monomer and tetramer fractional populations) versus $1/T$ to be +38 kJ/mol of subunit. With an association free energy of about –17 kJ/mol of subunit (at 30 °C, pH 5.3), the association entropy is about +130 kJ·mol⁻¹·K⁻¹/subunit. Although one might argue with the magnitudes of these thermodynamic parameters, the sign of their values is irrefutable. According to Klotz [45] and Ross and Subramanian [46], hydrophobic interactions and ionic charge neutralization both give rise to positive enthalpic and entropic contributions to association free energies. Therefore peptide 23–55 subunit-association-induced folding results primary hydrophobic collapse which has been suggested to mediate compact intermediate tertiary-structure formation

Table 3 Chemical-shift changes for PF4 tetrameric peptide 23–55

Values in parentheses are the β -sheet/ α -helix chemical shift index as given by Wishart et al. [39]; a series of +1 indicates β -sheet tendency, a series of -1 indicates α -helix tendency and 0 indicates others.

Residue	Chemical-shift difference (p.p.m.)		
	NH	α H	
His-23	—	—	—
Ile-24	—	—	—
Thr-25	1.16*	—	—
Ser-26	0.10	0.75*	(+1)
Leu-27	0.96	1.17	(+1)
Glu-28	0.54	1.04	(+1)
Val-29	0.85	0.98	(+1)
Ile-30	—	—	—
Lys-31	—	—	—
Ala-32	—	—	—
Gly-33	—	—	—
Pro-34	—	—	—
His-35	0.26	0.35	(+1)
Ser-36	-0.67	0.62	(+1)
Pro-37	—	—	—
Thr-38	—	—	—
Ala-39	—	—	—
Gln-40	—	—	—
Leu-41	0.83	1.08	(+1)
Ile-42	1.39	1.10	(+1)
Ala-43	1.22	1.03	(+1)
Thr-44	0.55	0.75	(+1)
Leu-45	0.64	1.05	(+1)
Lys-46	0.15	-0.10	(0)
Asn-47	-0.8	0.01	(+1)
Gly-48	-0.03	-0.22	(0)
Arg-49	-0.48	0.07	(+1)
Lys-50	-0.05	1.09	(+1)
Ile-51	1.46	0.64	(+1)
Ser-52	0.02	1.19	(+1)
Leu-53	1.08	0.59	(+1)
Asp-54	0.32	0.42	(+1)
Leu-55	—	—	—

* $\Delta\delta$ calculated as PF4 peptide chemical shift minus random-coil chemical shift given by Wüthrich [32].

[21]. Although the phrase 'hydrophobic collapse' implies that only hydrophobic residues are involved, Glu-28 and Lys-50 clearly play a crucial role in the peptide 23–55-folding process. As these electrostatic interactions occur in a low-dielectric medium, i.e. within the folded protein, it is worth emphasizing that hydrophobic collapse should be used as a general term for all low dielectric-type interactions.

In cystine-reduced PF4, which contains the entire PF4 amino acid sequence without the conformationally constraining disulphide bonds, a molten-globule monomeric state collapses into a native-like folded state also via subunit association [5]. There, however, only the dimer state could be detected. Although the tetrameric state is predominant for PF4 peptide 23–55, the presence of some peptide 23–55 dimers cannot be excluded. Differences in preferred aggregation states between PF4 peptide 23–55 and reduced PF4 (residues 1–70) must lie within the N- and C-terminal sequences. Interestingly, the concentration range within which aggregate states are observed is similar in both cases. Moreover, their pH-dependences follow the same general trends. Much the same can be said for native PF4 [4]. These observations suggest that the overall forces that hold peptide

Figure 9 Structural model for tetrameric peptide 23–55

A structural model for tetrameric peptide 23–55 has been generated by taking the backbone/side-chain atom co-ordinates (see ref. [38]) for native PF4 and deleting residues 1–22 and 56–70 as discussed in the text. Glu-28 and Lys-50 are shaded in to highlight their positions within the folded tetrameric peptide core. Some surface-exposed residues have also been labelled (one-letter code).

23–55, reduced PF4 and native PF4 subunits together are in some ways similar. The temperature-dependence of aggregate-state populations, however, varies considerably from PF4 species to PF4 species. For peptide 23–55 tetrameric state, fractional populations increase dramatically with increasing temperature. For reduced PF4, dimer-state fractional populations increase only slightly between 10 and 35 °C before decreasing rapidly over the temperature range 40–60 °C [5]. For native PF4, temperature effects run in opposite directions for dimer and tetramer formation [47], with dimer populations increasing and tetramer populations decreasing with increasing temperature. Of course, such trends relate to the sum total of forces that hold the respective subunits together. Therefore, not surprisingly, differences do exist within intersubunit domains in each case. For example, in native PF4, the N- and C-terminal carboxylates ionically interact between subunits and additional hydrogen bonds are formed. In peptide 23–55, these groups are absent.

Figure 9 shows a model for one possible peptide 23–55 tetrameric state. As peptide 23–55, like native PF4, predominantly forms tetramers, this Figure was generated from the backbone co-ordinates of residues 23 to 55 in native tetrameric human PF4 [38]. N- and C-terminal segments were merely deleted. Glu-28 and Lys-50 (highlighted) from the four subunits interact electrostatically within the tetramer core. Notice that in this conformation, other charged residues, i.e. His-23, His-35, Lys-46, Arg-49 and Asp-54, are on the surface of the tetramer which also contains many hydrophobic groups, some of which are screened by these charged residues. As already discussed, chemical-shift differences, intra- and inter-subunit n.O.e.s, pH-titration and thermodynamic data are consistent with this general conformational model. Specifically, NH and α H chemical shifts

in the peptide 23–55 tetrameric state indicate native-like β -sheet folding which is supported by the n.O.e. data. pH-titration data on peptide 23–55 indicate the importance of electrostatic interactions in the subunit-association process, whereas thermodynamic data indicate a significant role for the hydrophobic effect. Moreover, specific native-like subunit associations are evident from AB-type dimer α H– α H n.O.e.s between Ser-26 and Ile-30 and from AC-type dimer γ H₃– α H n.O.e.s between Thr-25 and Gly-48.

Relative to native PF4, conformational differences do exist. Cross-peaks are generally broader and more asymmetric, especially in the NH chemical-shift dimension. Within a folded protein, NH shifts are particularly sensitive to conformational changes and hydrogen-bonding. In some cases, the observation of multiple cross-peaks suggests that hydrophobic collapse of peptide 23–55 kinetically traps multiple folds via the subunit-association process. The Ser-52 cross-peak distribution (Figure 5), for example, suggests the presence of at least five such trapped conformers. One of these is apparently in the majority (about 50%) and others are about 5–20% populated. Each of these is in slow exchange (¹H-n.m.r. time scale) with the monomeric state and with each other. It is also probable that numerous shorter-lived conformational substrates exist within each of these multiple potentials. Gly-28, His-35 and Thr-44 provide similar examples. On the other hand, not all cross-peaks that are reasonably resolved show much multiplicity. For Asn-47, Gly-48 and Arg-49, broadening is least apparent, consistent with this sequence forming either a relatively well-defined chain reversal conformation or demonstrating increased mobility. The presence of multiple near-equiengetic conformational states implies that, although there is sufficient information within this truncated PF4 peptide sequence to form native-like β -sheet folds and quaternary structure, there is insufficient information for the peptide to decide on one 'unique' conformation. In this respect, N- and/or C-terminal sequences are apparently necessary to provide guidance on the PF4 protein-folding pathway.

Hydrophobic collapse of monomeric peptide 23–55, moreover, suggests that even within the distribution of rapidly interconverting monomeric conformers, some native-like tertiary folds must be present. Conformational preferences are noted in monomeric peptide 23–55, particularly within the Leu-41–Lys-50 sequence where a relatively stable chain reversal (Lys-45–Arg-49) and nascent helix (Leu-41–Thr-44) exist. The Leu-45–Arg-49 chain reversal which is found in native PF4 has been recognized as a folding-initiation site for PF4 [7]. Reverse turns and chain reversals, usually thought of as likely sites for initiation of protein folding, bring more distant parts of the chain into closer proximity, thereby directing subsequent folding events [9]. In this respect, formation of the native-like Leu-45–Arg-49 chain reversal in monomeric peptide 23–55 before hydrophobic collapse supports the Kim and Baldwin [20] 'framework' folding model. On the other hand, this does not exclude some initial hydrophobic side-chain stabilization allowing for organization of the relatively well-formed Leu-45–Arg-49 chain reversal [21] as proposed for PF4 peptide 38–57 [7]. In fact, folding of the PF4 peptide 23–55 β -sheet domain may be explained by the hydrophobic zipper hypothesis [22]. In this model, 'hydrophobic contacts act as constraints that bring other contacts into spatial proximity, which then further constrain and zip up the next contacts, etc.' [22].

Subunit-associated peptide 23–55 may be called a molten globule [48–51] in the sense that it has no fixed tertiary structure, but does show multiple conformational states which have considerable secondary structure and are relatively compact. The lifetimes of the collapsed state of peptide 23–55, being of the order

of 10 ms or more, however, are longer than those observed for most currently identified tertiary-structural molten-globule states which, because of intermediate exchange with random-coil conformers, give only somewhat broadened n.m.r. resonances. Therefore differences may be a matter of molten-globule-state lifetimes which for peptide 23–55 are increased because of subunit association. Alternatively or in addition, a hierarchy of molten-globule states may exist which are related to backbone/side-chain rearrangements [52,53] or reshuffling [54]. Such rearrangements have been proposed to occur on a much slower time scale (normally after molten-globule formation), leading to tighter packing inside the protein core [55,56]. In cystine-reduced PF4, the molten-globule monomeric state was observed to collapse into a native-like folded state also via subunit association [5]. Although that monomeric state fits the criteria for a molten globule, its dimeric state may also be an ill-defined native-like compact molten globule. Both might be considered quaternary-structural molten-globule-folding intermediates about which little is known.

CONCLUSIONS

Three main conclusions may be drawn about PF4 folding: (1) in the absence of more than half the amino acid sequence, native-like β -sheet folds can form; (2) folding is thermodynamically linked to subunit-association-induced hydrophobic collapse; (3) the hydrophobic collapse traps multiple conformational states. This folding process, moreover, implies that the random-coil conformational distribution present before folding must have native-like conformational populations within its dynamic-equilibrium ensemble.

N.m.r. experiments were performed at the University of Minnesota high-field n.m.r. 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 and to Cindy Fields for N-terminal peptide sequencing.

REFERENCES

- Deuel, T. F., Keim, P. S., Farmer, M. and Heinrikson, R. L. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2256–2258
- St. Charles, R., Walz, D. A. and Edwards, B. F. P. (1989) *J. Biol. Chem.* **264**, 2092–2099
- Holt, J. C., Harris, M. E., Holt, A., Lange, E., Henschen, A. and Niewiarowski, S. (1986) *Biochemistry* **25**, 1988–1996
- Mayo, K. H. and Chen, M. J. (1989) *Biochemistry* **28**, 9469–9478
- Mayo, K. H., Barker, S., Kuranda, M. J., Hunt, A. J., Myers, J. A. and Malone, T. E. (1992) *Biochemistry* **31**, 12255–12265
- Kim, K.-S., Fuchs, J. A. and Woodward, C. K. (1993) *Biochemistry* **32**, 9600–9608
- Ilyina, E., Milius, R. and Mayo, K. H. (1994) *Biochemistry* **33**, 13436–13444
- Anfinsen, C. B. (1973) *Science* **181**, 223–227
- Jaenicke, R. (1991) *Biochemistry* **30**, 3147–3161
- Udgaonkar, J. B. and Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699
- Roder, H., Elöve, G. A. and Englander, S. W. (1988) *Nature (London)* **335**, 700–704
- Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L. and Fersht, A. R. (1990) *Nature (London)* **346**, 488–490
- Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. and Fersht, A. R. (1990) *Nature (London)* **346**, 440–445
- Freire, E., Murphy, K. P., Sanchez-Ruiz, J. M., Galisteo, M. L. and Privalov, P. L. (1992) *Biochemistry* **31**, 250–256
- Walther, J. P., Feher, V. A., Merutka, G., Dyson, H. J. and Wright, P. E. (1993) *Biochemistry* **32**, 6337–6347
- Shin, H. C., Merutka, G., Walther, J. P., Wright, P. E. and Dyson, H. J. (1993) *Biochemistry* **32**, 6348–6355
- Shin, H. C., Merutka, G., Walther, J. P., Tennant, L. L., Dyson, H. J. and Wright, P. E. (1993) *Biochemistry* **32**, 6356–6364
- Wu, L. C., Laub, P. B., Eloee, G. A., Carey, J. and Roder, H. (1993) *Biochemistry* **32**, 10271–10276
- Dyson, H. J. and Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 519–538

- 20 Kim, P. S. and Baldwin, R. L. (1982) *Annu. Rev. Biochem.* **51**, 459–489
- 21 Chan, H. S. and Dill, K. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6388–6392
- 22 Dill, K. A., Fiebig, K. M. and Chan, H. S. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1942–1946
- 23 Stewart, J. M. and Young, J. D. (1984) *Solid Phase Peptide Synthesis* (Merrifield, L., ed.), 2nd edn., p. 135, Pierce Chemical Co., Rockford, IL
- 24 Piantini, U., Sørensen, O. W. and Ernst, R. R. (1982) *J. Am. Chem. Soc.* **104**, 6800–6805
- 25 Shaka, A. J. and Freeman, R. (1983) *J. Magn. Reson.* **51**, 161–169
- 26 Bax, A. and Davis, D. G. (1985) *J. Magn. Reson.* **65**, 355–360
- 27 Jeener, J., Meier, B., Backman, P. and Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546–4550
- 28 Wider, G., Macura, S., Anil-Kumar, Ernst, R. R. and Wüthrich, K. (1984) *J. Magn. Reson.* **56**, 207–234
- 29 Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113**, 967–975
- 30 States, D. J., Haberkorn, R. A. and Ruben, D. J. (1982) *J. Magn. Reson.* **48**, 286–293
- 31 Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.* **85**, 393–399
- 32 Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley–Interscience, New York
- 33 Sreerama, N. and Woody, R. W. (1993) *Anal. Biochem.* **209**, 32–44
- 34 Bierzyński, A., Kim, P. S. and Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2470–2474
- 35 Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M. and Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2349–2353
- 36 Waterhous, D. V. and Johnson, W. C. (1994) *Biochemistry* **33**, 2121–2128
- 37 Tanford, C. (1961) *Physical Chemistry of Macromolecules*, pp. 286–296, Wiley, New York
- 38 Zhang, X., Chen, L. and Bancroft, D. P. (1994) *Biochemistry* **33**, 8361–8366
- 39 Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) *Biochemistry* **31**, 1647–1651
- 40 Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A. and Wright, P. E. (1992) *J. Mol. Biol.* **226**, 819–835
- 41 Jardetsky, O. and Roberts, G. C. K. (1981) *NMR in Molecular Biology*, Academic Press, New York
- 42 Patel, D. J., Woodward, C. K. and Bovey, F. A. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 599–602
- 43 Clore, G. M., Appella, E., Yamada, M., Matsushima, K. and Grønenborn, A. M. (1989) *Biochemistry* **29**, 1689–1696
- 44 Wettlaufer, D. B. (1981) *Adv. Protein Chem.* **34**, 61–92
- 45 Klotz, I. M. (1973) *Ann. N. Y. Acad. Sci.* **226**, 18–24
- 46 Ross, P. D. and Subramanian, S. (1981) *Biochemistry* **20**, 3096–3102
- 47 Chen, M.-J. and Mayo, K. H. (1991) *Biochemistry* **30**, 6402–6411
- 48 Ptitsyn, O. B. (1987) *J. Protein Chem.* **6**, 273–293
- 49 Ptitsyn, O. B. (1991) *FEBS Lett.* **285**, 176–181
- 50 Chaffotte, A., Giullou, Y., Dellepierre, M., Hinz, H. J. and Goldberg, M. E. (1991) *Biochemistry* **30**, 8067–8074
- 51 Chaffotte, A. F., Cadieux, C., Giullou, Y. and Goldberg, M. E. (1992) *Biochemistry* **31**, 4303–4308
- 52 Segawa, S.-I. and Sugihara, M. (1984) *Biopolymers* **23**, 2473–2488
- 53 Goldenberg, D. P. and Creighton, T. E. (1985) *Biopolymers* **24**, 167–182
- 54 Vaucheret, H., Signon, L., Le Bras, G. and Garel, J. R. (1987) *Biochemistry* **26**, 2785–2790
- 55 Ptitsyn, O. B. and Semisotnov, G. V. (1991) in *Conformations and Forces in Protein Folding* (Nail, B. T. and Dill, K. A., eds.), pp. 155–168, American Association for the Advancement of Science, Washington, D.C.
- 56 Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 84–91