Peptide Models of Protein Folding Initiation Sites. 1. Secondary Structure Formation by Peptides Corresponding to the G- and H-Helices of Myoglobin†

Jonathan P. Waltho,‡ Victoria A. Feher, Gene Merutka, H. Jane Dyson,* and Peter E. Wright*

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received February 8, 1993; Revised Manuscript Received April 2, 1993

ABSTRACT: Myoglobin has been extensively studied as a model system for protein folding *in vitro*. As part of an ongoing study of myoglobin folding, we have synthesized a series of peptide fragments corresponding to portions of the sequence of the sperm whale protein. The conformational preferences of these peptides have been investigated by circular dichroism and nuclear magnetic resonance spectroscopy in aqueous solution. In this paper we describe the folding propensities of two peptides (Mb-G and Mb-H), corresponding to the G- and H-helix segments of the myoglobin sequence. The Mb-G peptide shows evidence of a very small population of helical conformations in aqueous solution, both by CD and NMR. By contrast, the monomeric Mb-H peptide is found by CD to adopt a significant population (ca. 30%) of ordered helix and by NMR to populate helical conformations in rapid dynamic equilibrium with unfolded states. The Mb-H peptide undergoes a well-characterized, concentration-dependent monomer–tetramer equilibrium. At peptide concentrations greater than 1 mM there is an increase in the population of helix, to approximately 85% according to the CD spectrum, through self-association to form a tetramer. Both medium-range NOE connectivities and a CD spectrum characteristic of ordered helix are observed at low peptide concentrations, establishing that helical conformations are present in the monomeric state of Mb-H. The relative helicity at various sites throughout the Mb-H peptide has been estimated using a novel method for assessing the distribution of helical populations based on the relative magnitudes of medium-range $d_{\alpha\beta}(i,i+3)$ NOE connectivities. The population of ordered helix is seen to be highest in the center of the peptide sequence; the ends of the peptide show evidence of pronounced fraying.

INTRODUCTION

The mechanism of folding of proteins into their native conformations remains one of the central unsolved problems of molecular biology. As a result of intensive research efforts in many laboratories, a generalized pathway by which folding occurs is now emerging (for reviews see Creighton, 1988; Wright et al., 1988; Baldwin, 1989; Montelione & Scheraga, 1989; Kuwajima, 1989; Kim & Baldwin, 1990; Ptitsyn, 1991; Jaenicke, 1991). The folding of a polypeptide chain in vitro into a native, biologically active conformation apparently involves a self-assembly process (Anfinsen, 1973) and frequently occurs in a short period of time, of the order of seconds or minutes (Creighton, 1985; Kim & Baldwin, 1982). This implies that the folding of a protein proceeds not through a random search of all configurations but along a defined pathway with structured intermediates (Levinthal, 1968; Wetlaufer, 1973). Therefore the detection and structural characterization of folding intermediates are of fundamental importance to our understanding of the mechanism of protein folding. These transient structures, however, are not easy to characterize due to the highly cooperative process of protein folding. To circumvent these difficulties, proteins have been examined under partially denaturing conditions, such as high temperature, low pH, or weak denaturants. Under such conditions, several proteins were found to exist as compact "molten-globule states" (Kuwajima, 1977; Ohgushi & Wada, 1983), in which there is secondary structure, but ordered tertiary interactions are not present (for recent reviews, see Christensen & Pain, 1991; Dobson, 1991). Refolding experiments using hydrogen-exchange pulse-labeling coupled with NMR[1] spectroscopy have been useful in delineating the nature of folding pathways and in directly showing the existence of folding intermediates (Udgaonkar & Baldwin, 1988; Roder et al., 1988; Matouschek et al., 1990; Bycroft et al., 1990; Radford et al., 1992; Lu & Dahlquist, 1992). These studies establish the formation of secondary structure within a few milliseconds of initiation of protein folding. The role of secondary structural elements formed even earlier in the folding process in directing the course of folding and in stabilization of folding intermediates remains to be elucidated. Such structures are formed too rapidly in the intact protein to be studied by presently available methods. We have therefore turned to peptide fragments of proteins as models to investigate the earliest events of the protein folding process (Wright et al., 1988).

Secondary structures, especially helices and reverse turns, have been observed in short linear peptide fragments in aqueous solution [reviewed in Dyson and Wright (1991)], lending considerable support to the notion that secondary structural elements may play an important role in the initiation of protein folding (Kim & Baldwin, 1982). The secondary structural

† Supported by Grant GM38794 (P.E.W. and H.J.D.) and postdoctoral award GM14526 (G.M.) from the National Institutes of Health.

‡ Present address: Department of Molecular Biology and Biotechnology, University of Sheffield, P.O. Box 594, Sheffield, S10 2UH U.K.

* To whom correspondence should be addressed.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; Mb, myoglobin; Mb-G, Mb-H, synthetic peptides corresponding to the G- and H-helixes of myoglobin respectively; Mb-F1, peptide corresponding to the C-terminal fragment of myoglobin obtained by cyanogen bromide cleavage; NOE, nuclear Overhauser effect; HPLC, high-performance liquid chromatography; 2QF COSY, double quantum filtered phase-sensitive two-dimensional correlated spectroscopy; 2Q, two-dimensional double-quantum spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; TOCSY, two-dimensional total correlated spectroscopy; $d_{\alpha N}(i,j)$, $d_{NN}(i,j)$, etc., intramolecular distance between the protons $C^{\alpha}H$ and NH, NH, and NH, etc., on residues i and j; $^3J_{HN\alpha}$, NH–$C^{\alpha}H$ coupling constant; ppm, parts per million; TSS, (trimethylsilyl)propanesulfonic acid; TFE, 2,2,2-trifluoroethanol.

0006-2960/93/0432-6337$04.00/0

elements found in peptides, studied under conditions where they are devoid of tertiary interactions, are in rapid dynamic equilibrium with unfolded states (Wright et al., 1988; Dyson & Wright, 1991). Successful attempts have been made to stabilize secondary structural elements in peptide fragments (Oas & Kim, 1988; Ghaderi & Choi, 1990; Satterthwait et al., 1990). We present in this and the following two papers the results of an extensive study of peptide fragments corresponding to a region of the myoglobin molecule that has been implicated in the early folding events by a number of studies.

The mechanism of refolding of myoglobin has been the subject of extensive investigation. A number of the original studies on the conformations of peptide fragments of proteins were carried out using fragments of myoglobin (Epand & Scheraga; 1968; Hermans & Puett, 1971). Myoglobin contains a well-defined helical hairpin towards the carboxy terminus, between helices G (residues 100–118) and H (residues 124–150). Theoretical studies of the myoglobin sequence prompted the proposal (Matheson & Scheraga, 1978) that this region should constitute a folding initiation site. The hairpin structure formed in the intact protein between the G- and H-helices was therefore proposed as an intermediate in the folding of myoglobin (Gerritsen et al., 1985). An intermediate state of apomyoglobin formed at acid pH has been shown to form a partly folded structure containing the A-, G-, and H-helices (Hughson et al., 1990; 1991). It has recently been demonstrated that this same intermediate is formed in less than 5 ms on the kinetic folding pathway of apomyoglobin (P. A. Jennings and P. E. Wright, submitted for publication).

Since the G–H helical hairpin appears to be present in an early myoglobin folding intermediate, the earliest folding initiation events may involve one or more of the possible secondary structural elements encoded in the sequence. Initial folding events do not necessarily give rise to structures which are retained in the final folded form of the protein; examples exist of structures found in peptides which are not seen in the intact protein (Dyson et al., 1985). However, in the main, it does appear that the protein primary sequence codes for secondary structure found in the native folded protein; this was recently demonstrated using series of peptide fragments from two different protein folds: the peptides derived from the β-sheet protein plastocyanin showed conformational preferences predominantly for extended (β) structures in solution (Dyson et al., 1992b), while those derived from the four-helix bundle protein myohemerythrin showed considerable propensity for helical or nascent helical structures (Dyson et al., 1992a). To identify transient structures that might be involved in the earliest initiation events in the folding of apomyoglobin, we have studied the conformational preferences of peptides corresponding to two helical segments, the G- and H-helices (this paper) and of the intervening sequence, termed the G–H turn (Shin et al., 1993a). A number of earlier studies on peptides derived from myoglobin concentrated on the C-terminal cyanogen bromide fragment (residues 132–153, termed Mb-F1), which contains the three extra nonhelical residues at the C-terminus of the myoglobin molecule. Optical methods failed to detect helical structure in this peptide (Epand & Scheraga, 1968; Hermans & Puett, 1981), but small conformational preferences for helical structure were found by NMR (Waltho et al., 1989, 1990). The present studies demonstrate that a peptide corresponding to the entire H-helix contains a considerably greater population of helical structure in solution than Mb-F1, as a consequence of the removal of nonhelical residues at the C-terminus and the addition of residues at the N-terminus that form part of the H-helix in the native protein.

Secondary structure in potential folding initiation sites not only should be formed in a threshold population in short peptide fragments (which model the unfolded protein in the absence of tertiary stabilizing interactions) but also should be retained in longer fragments of the protein involved in the next stage of folding, if not in the final folded state. To model this second stage of folding, which we envision as a process of coalescence and stabilization of secondary structure elements, without, as yet, significant specific tertiary interactions, we have constructed a series of larger peptides, carefully designed to be monomeric in solution. The behavior of these peptides is described in the final paper of this series (Shin et al., 1993b).

MATERIALS AND METHODS

Sample Preparation. Two peptides were synthesized, one corresponding to residues 124–150 of sperm whale myoglobin, termed Mb-H, and one corresponding to residues 101–118, termed Mb-G. The sequences are

Mb-G: ^{101}IKLEFISQAIIHVLHSR^{118}

Mb-H: ^{124}GADAQGAMNKALELFRKDIAAKYKELG^{150}

Both peptides were blocked at the N-terminus with an acetyl group and at the C-terminus with a primary amide group. The residue numbering of sperm whale myoglobin is retained throughout. The peptides were synthesized on an Applied Biosystems Model 430A synthesizer using Boc chemistry and purified using a Vydac C_{18} semipreparative reverse-phase column on a Hitachi HPLC system, with a solvent gradient between water and acetonitrile, both solvents containing 0.1% v/v trifluoroacetic acid. Samples used for NMR studies were >95% pure by analytical reverse-phase HPLC.

All of the residues are identical to those found in the sequence of sperm whale myoglobin, with the exception of the glutamic acid residue at position 109. This residue was inadvertently substituted with glutamine due to reference to an incorrect sequence. However, subsequent studies using peptides which contain the correct residue at position 109 (Shin et al., 1993a) reveal no difference in behavior from that of the Mb-G peptide described here.

NMR Spectroscopy. NMR samples were prepared in the concentration range 10 μM to 10 mM in 95% ^1H_2O/5% ^2H_2O or 99.98% ^2H_2O and the pH was adjusted to 4.0. Most experiments were performed on 0.6–1.0 mM samples for Mb-H peptide and 0.4–0.8 mM samples for Mb-G at 278 K. These concentrations were chosen such that the peptides were monomeric. Chemical shifts were referenced to an internal standard of dioxane at 3.75 ppm. Probe temperature was calibrated using methanol by the method of VanGeet (1969).

NMR spectra were recorded on a Bruker AM500 spectrometer equipped with digital phase shifting hardware. Two-dimensional spectra were acquired in the phase-sensitive mode with quadrature detection in both dimensions. Standard pulse sequences and phase cycling were employed for all experiments, and solvent suppression was achieved through preirradiation of the water resonance for 1.5 s. Double quantum filtered scalar correlated spectroscopy (2QF COSY) experiments (Rance et al., 1983) were recorded typically with a t_{1max} of 100 ms and a t_{2max} of 400 ms, with 64 scans per t_1 increment. Double quantum (2Q) experiments (Braunschweiler et al., 1983; Rance & Wright, 1986) were recorded without

quadrature detection in the $t1$ dimension with a t_{1max} of 50 ms and a t_{2max} of 400 ms and 128 scans per t_1 increment. Total correlated spectroscopy (TOCSY) experiments (Braunschweiler & Ernst, 1983; Rance, 1987) were recorded with WALTZ-16 spin-locking of longitudinal magnetization for periods of 50–100 ms. Two-dimensional nuclear Overhauser effect spectra (NOESY) (Bodenhausen et al., 1984) were recorded with mixing times in the range 100–400 ms. Solvent saturation was applied during the mixing time. 2Q, NOESY and TOCSY experiments were acquired with Hahn-echo observation (Rance & Byrd, 1983) with a t_{1max} of 50–80 ms and t_{2max} of 328 ms and with 64–256 scans per t_1 increment.

Data were processed using the program FTNMR (Hare Research). Lorentzian–Gaussian or shifted sine-bell apodization and zero-filling were applied prior to Fourier transformation, and subsequent baseline corrections were applied in one or both dimensions. Temperature coefficients for Mb-H were calculated from the amide proton chemical shifts obtained from a series of 2QF COSY spectra at temperatures 278, 288, and 298 K. The NH and C$^{\alpha}$H resonances of Gly 124 and Gly 129 are coincident at 278 K, the temperature at which the resonances were sequentially assigned. The NH resonances are resolved at 288 and 298 K, and the temperature coefficients obtained from the two sets of data were assigned as shown. However, no NOESY data were obtained at the higher temperatures, so there is ambiguity between the temperature coefficient values for the two residues.

CD Spectroscopy. CD spectra were recorded on an AVIV 61DS spectropolarimeter calibrated with a standard solution of 10-camphorsulfonic acid (Johnson, 1985). Samples were prepared in the concentration range 10 μM to 6.5 mM in quartz cells with path lengths in the range 0.01–10 mm. Peptide concentrations were determined by quantitative amino acid analyses. The mean residue ellipticity, in deg cm^2 dmol^{-1}, was calculated from the relationship $[\theta] = \theta \times 1/(10lcN)$, where θ is the observed ellipticity, l is the pathlength in cm, c is the molar concentration of peptide, and N is the number of amino acids in the sequence. Spectra obtained for the concentration-dependence studies were an average of three scans recorded with a 0.50-nm bandwidth, a 0.25-nm step size, and a 2.0-s time constant. Data were smoothed with a third-order polynomial function. All other data were obtained using a 1.5-nm bandwidth, a 0.5-nm step size, and a 4.0-s time constant.

Calculation of Thermodynamic Parameters. The thermodynamic parameters ΔH and ΔS for the unfolding of the H-helix peptide with temperature were calculated by a least-squares fit to eq 1 to the ellipticity measured at 222 nm as a function of temperature for 14 temperatures between 275 and 333 K.

We assume that the transition is a simple two-state conversion F $\rightleftharpoons$ U, where F represents a folded (helical) state and U an unfolded state, and that ΔH and ΔS are temperature independent. The fraction of folded form χ_F at any temperature can be described in terms of the experimental θ_{222}:

$$\chi_F = \frac{\theta_{expt} - \theta_U}{\theta_F - \theta_U}$$

where θ_F and θ_U are the ellipticities of the folded and unfolded forms respectively and θ_{expt} is the experimentally observed ellipticity. Since χ_F is also related to the equilibrium constant K for the F $\rightleftharpoons$ U transition:

$$\chi_F = 1/(K+1) \qquad \Delta G = -RT \ln K$$

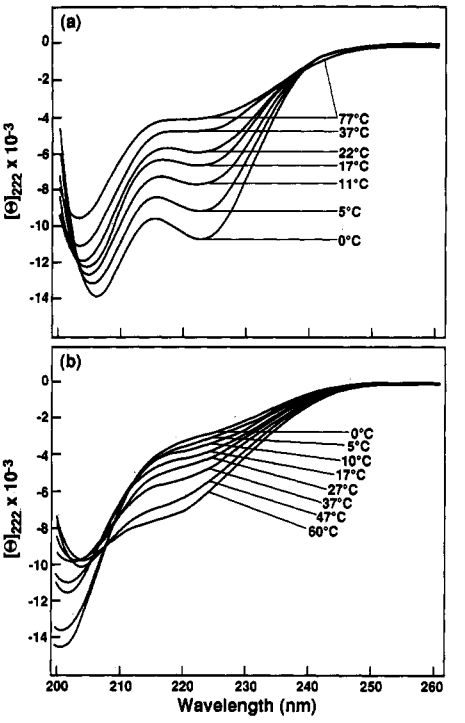

FIGURE 1: Dependence of the ultraviolet CD spectrum on temperature. (a) Mb-H, 14.5 μM in 1 mM sodium phosphate buffer, pH 5.08, in a 10-mm cell. (b) Mb-G, 16.9 μM in 1 mM sodium phosphate buffer, pH 4.92, in a 10-mm cell.

we can relate ΔH and ΔS to the ellipticity using

$$\theta_{expt} = \theta_U + \frac{\theta_F - \theta_U}{1 + \exp(\Delta S/R - \Delta H/RT)} \qquad (1)$$

RESULTS AND DISCUSSION

CD of the Monomeric Species. Both Mb-H and Mb-G exhibit concentration-dependent CD spectra at peptide concentrations above about 1 mM. The CD spectrum is independent of concentration for Mb-H over the concentration range 100 μM to 1 mM and for Mb-G over the concentration range 14–500 μM. For low concentrations of Mb-H, the ellipticity at 222 nm [Θ_{222}] is −11 000 deg cm^2 dmol^{-1} at 0 °C, which suggests a population average of helical conformations of approximately 30%, using [Θ_{222}] of 0 deg cm^2 dmol^{-1} and −36 000 deg cm^2 dmol^{-1} to represent 0% and 100% helix respectively (Chen et al., 1974). At a Mb-H concentration of 6.5 mM, the highest concentration used in the CD experiments, [Θ_{222}] increases to −29 000 deg cm^2 dmol^{-1}. This behavior is associated with tetramer formation and is treated in detail in a later section.

The CD spectrum of Mb-H is temperature dependent; the ellipticity at 222 nm becomes less negative with increasing temperature (Figure 1a), consistent with temperature-induced unfolding of helical conformations. This behavior is typical of small peptides that partially populate helical conformations and is similar, for example, to that of the ribonuclease C-peptide (Shoemaker et al., 1987a). The data fit well to a simple two-state model of unfolding with ΔH and ΔS being temperature independent (ΔH = 51 ± 9 kJ mol.$^{-1}$, ΔS = 190 ± 30 J K^{-1} mol^{-1}).

The CD spectrum of Mb-G at low concentration and low temperature shows little evidence for helix formation ([Θ_{222}] = −3000 deg cm^2 dmol^{-1} at 278 K, corresponding to less than 10% helix). As the temperature is increased, the CD spectrum

Table I: Resonance Assignments and $^3J_{HN\alpha}$ Coupling Constants for Mb-G (0.8 mM, pH 4.0, 278 K) and Mb-H (1.0 mM, pH 4.0, 278 K)

	chemical shift (ppm)						
	NH	C$^\alpha$H	C$^\beta$H	C$^\gamma$H	C$^\delta$H	other	$^3J_{HN\alpha}$(Hz)[a]
				(a) Mb-G			
I^{101}	8.29	3.99	1.74	1.17, 1.48	0.85	0.77(C$^\gamma$H$_3$)	6.3
K^{102}	8.46	4.28	1.69	1.28, 1.37	1.64	2.94 (C$^\epsilon$H$_2$) 7.62 (N$^\zeta$H$_3$)	7.1
Y^{103}	8.29	4.52	2.94, 3.03		7.13	6.81 (C$^\epsilon$H$_2$)	7.1
L^{104}	8.20	4.24	1.62, 1.50	1.52	0.86, 0.91		7.3
E^{105}	8.27	4.18	1.92, 1.93	2.18, 2.28			6.6
F^{106}	8.34	4.57	3.05, 3.10		7.22	7.34, 7.31 (C$^\epsilon$H$_2$,C$^\zeta$H)	6.9
I^{107}	8.13	4.07	1.79	1.16, 1.44	0.85	0.85 (C$^\gamma$H$_3$)	7.5
S^{108}	8.38	4.34	3.85, 3.91				6.7
Q^{109}	8.51	4.27	1.97, 2.13	2.39	6.98, 7.67 (N$^\delta$H)		7.0
A^{110}	8.30	4.24	1.35				6.2
I^{111}	8.17	4.04	1.81	1.16, 1.44	0.85	0.74 (C$^\gamma$H$_3$)	7.7
I^{112}	8.26	4.06	1.78	1.16, 1.44	0.82	0.83 (C$^\gamma$H$_3$)	7.7
H^{113}	8.73	4.73	3.13, 3.20		8.65	7.30 (C$^\epsilon$H)	
V^{114}	8.38	4.01	1.99	0.85, 0.92			7.4
L^{115}	8.55	4.30	1.47, 1.62	1.54	0.82, 0.90		7.1
H^{116}	8.72	4.71	3.19, 3.26		8.65	7.34 (C$^\epsilon$H)	
S^{117}	8.52	4.42	3.83, 3.89				6.4
R^{118}	8.67	4.31	1.80, 1.92	1.66, 1.71	3.22	7.27 (N$^\epsilon$H)	6.9
				C-terminal NH$_2$: 6.50, 6.95			
				(b) Mb-H			
G^{124}	8.46	3.94					
A^{125}	8.57	4.29	1.40				5.3
D^{126}	8.57	4.63	2.80, 2.74				6.8
A^{127}	8.29	4.27	1.44				5.5
Q^{128}	8.48	4.25	2.16, 2.07	2.41			6.0
G^{129}	8.45	3.95					
A^{130}	8.26	4.25	1.46				5.1
M^{131}	8.40	4.38	2.11	2.64, 2.56		2.10 (C$^\epsilon$H$_3$)	6.0
N^{132}	8.46	4.62	2.87				5.9
K^{133}	8.37	4.17	1.89	1.45, 1.55	1.70	2.99 (C$^\epsilon$H$_2$)	6.1
A^{134}	8.18	4.23	1.49				5.8
L^{135}	8.13	4.31	1.77, 1.70	1.70	0.98, 0.90		6.2
E^{136}	8.17	4.18	2.09	2.43, 2.34			6.3
L^{137}	8.07	4.19	1.68	1.66	0.93, 0.87		6.8
F^{138}	8.16	4.51	3.24, 3.16		7.27	7.30, 7.30 (C$^\epsilon$H$_2$, C$^\zeta$H)	6.7
R^{139}	8.27	4.11	1.87, 1.81	1.65	3.22		6.2
K^{140}	8.29	4.18	1.89	1.48, 1.50	1.70	2.99 (C$^\epsilon$H$_2$)	6.7
D^{141}	8.34	4.58	2.76				7.0
I^{142}	8.05	3.99	1.87	1.38, 1.13	0.78	0.89 (C$^\gamma$H$_3$)	7.4
A^{143}	8.20	4.24	1.45				6.2
A^{144}	8.12	4.22	1.43				6.5
K^{145}	8.11	4.16	1.75	1.28, 1.37	1.63	2.94 (C$^\epsilon$H$_2$)	6.1
Y^{146}	8.10	4.51	3.07, 2.99		7.14	6.82 (C$^\epsilon$H$_2$)	6.4
K^{147}	8.09	4.10	1.83				6.5
E^{148}	8.33	4.25	2.11, 2.02	2.42			6.6
L^{149}	8.43	4.33	1.73, 1.63	1.69	0.94, 0.89		6.4
G^{150}	8.48	3.87, 3.91					

[a] Average of two values, one calculated from peak-to-peak separation in the multiplet structure of COSY cross peaks, and the other using the method of Kim & Prestegard (1989).

changes, apparently with formation of β-structure or polyproline II-type helix (Woody, 1985) (Figure 1b). These changes are consistent with those observed at higher concentrations in NMR experiments and are probably due to association processes which are discussed more fully in a later section.

NMR Spectra: Proton Resonance Assignments. Complete assignment of proton resonances for the monomeric forms of the peptides was achieved using the sequence-specific resonance assignment procedure (Billeter et al., 1982), modified as suggested by Chazin and Wright (1987). There is little chemical shift dispersion in the C$^\alpha$H region of the NMR spectra of these peptides. Thus the problems encountered in making the resonance assignments are comparable to those experienced with small helical proteins. Overlap in the C$^\alpha$H–C$^\beta$H region of spectra of both peptides permitted the observation of only a limited number of resolved cross peaks. However, complete assignments were made by utilizing the favorable dispersion of the amide proton resonances in TOCSY spectra in 90% ^1H_2O/10% ^2H_2O. In most cases, the $T_{1\rho}$ relaxation times were sufficiently long to permit the observation of four-bond coupling from the C$^\beta$H protons into the aromatic rings of tyrosine, phenylalanine, and histidine in TOCSY spectra with long spin-locking periods (100 ms). Such connectivities have previously been observed in double-quantum spectra of proteins (Dalvit et al., 1987). Direct and remote cross peaks in 2Q spectra of the peptides in ^2H_2O solution were used to identify degeneracy in geminal proton pairs and to distinguish vicinal versus geminal ambiguities within the cross peaks observed in TOCSY spectra. Spin systems were linked in a sequence-specific manner using $d_{\alpha N}(i,i+1)$, $d_{NN}(i,i+1)$, and $d_{\beta N}(i,i+1)$ connectivities in NOESY spectra. NOEs were also used to determine connections between primary amide groups and their adjacent methylene groups in asparagine and glutamine residues and to confirm connections between C$^\beta$H and aromatic rings. The assignments for Mb-G and Mb-H are summarized in Table I, and the amide proton chemical shifts are plotted for each residue of Mb-H in Figure 2a.

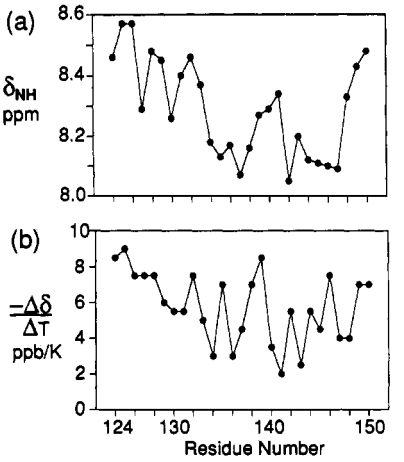

FIGURE 2: (a) Chemical shifts of amide protons of Mb-H, plotted against residue number. (b) Temperature coefficients of backbone amide protons of Mb-H, calculated from least-squares fits to the linear temperature dependence of the amide proton chemical shift, plotted against residue number.

Coupling Constants and Temperature Coefficients. Measurements of $^3J_{HN\alpha}$ were made using the method of Kim and Prestegard (1989) and from peak-to-peak distances in resolution-enhanced 2QF COSY spectra, zero-filled to give high digital resolution in ω_2. The results obtained by both methods were closely comparable. The $^3J_{HN\alpha}$ values obtained for Mb-G and Mb-H are shown in Table I. The $^3J_{HN\alpha}$ values typically observed in helices within proteins mostly fall in the range 4–5 Hz, whilst those in extended conformations fall in the range 8–10 Hz (Wüthrich, 1986). For Mb-H, coupling constants are mostly in a range intermediate between that observed for random coil peptides (7–8 Hz) and that for helix (average value 6.4 Hz; standard deviation 0.45 Hz). The average $^3J_{HN\alpha}$ for Mb-G is higher (7.0 Hz; standard deviation 0.45 Hz) than for Mb-H, consistent with the lower CD estimate of helix for Mb-G. Many of the $^3J_{HN\alpha}$ coupling constants in Mb-G fall within the accepted range for random coil. It is noteworthy that $^3J_{HN\alpha}$ values for alanine residues are consistently lower, and those for isoleucine consistently higher, than for neighboring residues, as has been previously noted (Wüthrich, 1986).

Temperature coefficients ($\Delta\delta/\Delta T$) of amide proton resonances have been shown in a number of cases to correlate with the population of hydrogen bonds between amide protons and nonaqueous acceptors (Rose et al., 1985), and have been found to correlate with the populations of folded forms of peptides (Dyson et al., 1988a). Because of the unusual behavior of Mb-G as the temperature is raised, no temperature coefficients were obtained for this peptide. Temperature coefficients were determined for Mb-H using the NH-C$^\alpha$H cross peaks of 2QF COSY spectra at low concentration at 278, 288, and 298 K and are shown in Figure 2b. Values significantly lower than the ~8 ppb/K found in random-coil peptides are seen for residues in the center of the peptide, in the region which NOE intensities indicate to be the most helical (see later), with a periodic variation to higher values. Similar variation in temperature coefficients has been observed in other helical peptides in the monomeric state (Zhou et al., 1992; Dyson et al., 1992a). In addition to solvent protection effects due to hydrogen bonding in helical conformers, it is likely that local sequence variations also affect the temperature coefficient, for example, the side chain of a Glu residue can hydrogen-bond to its amide proton (Ebina & Wüthrich, 1984) and a cluster of hydrophobic side chains can cause quite extreme solvent protection of amide protons (Dyson et al., 1992a,b).

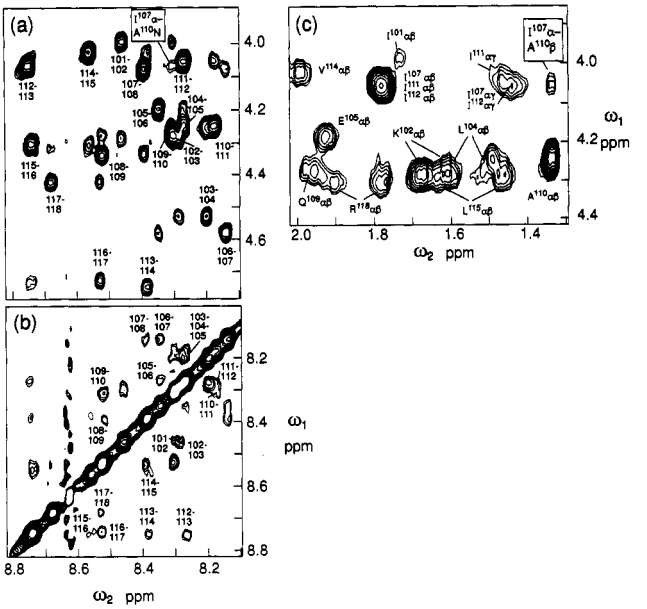

FIGURE 3: (a) NH–C$^\alpha$H region; (b) NH–NH region, of a 500-MHz Hahn-echo NOESY spectrum of Mb-G (0.78 mM) in 90% ^1H_2O/10% ^2H_2O, at pH 4.0 at 278 K. The mixing time τ was 400 ms. Sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOE connectivities are indicated, and peaks of weaker intensity represent intraresidue $d_{\alpha N}(i,i)$ NOEs. (c) C$^\alpha$H–C$^\beta$H region of a 500-MHz Hahn-echo NOESY spectrum (τ = 380 ms) of Mb-G (0.8 mM) in ^2H_2O at pH 4.0 and 278 K. The $d_{\alpha\beta}(i,i+3)$ NOE connectivity between Ile^{107} and Ala^{110} is indicated by a box.

NOE Evidence for Helix in Mb-H and Mb-G. The interpretation of NOE connectivities to provide insights into the conformational preferences of small linear peptides, which rapidly fluctuate over an ensemble of conformations in solution, has been discussed in detail elsewhere (Wright et al., 1988; Dyson & Wright, 1991). Of particular relevance to the present work, the observation of a series of sequential $d_{NN}(i,i+1)$ NOE connectivities plus medium-range $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$ and $d_{\alpha\beta}(i,i+3)$ NOEs (Wüthrich et al., 1984) is diagnostic of helical structures in the conformational ensemble. At the ends of the helix, where fraying might be expected, some $d_{\alpha N}(i,i+2)$ NOEs may also occur, indicating population of 3_{10} helix (Wüthrich et al., 1984), "nascent helix" (Dyson et al., 1988b) or turn conformations.

Regions of the NOESY spectrum of monomeric Mb-H containing the sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ and intraresidue $d_{\alpha N}(i,i)$ NOE cross peaks are shown for Mb-G and Mb-H in Figures 3 and 4 and a schematic representation of the resolvable NOEs for Mb-G and Mb-H, including the medium-range NOEs observed, is given in Figure 5. For Mb-H, the NMR evidence is strong for the presence of a substantial population of ordered helical forms: an extensive set of $d_{NN}(i,i+1)$ and $d_{\alpha\beta}(i,i+3)$ NOE connectivities is observed throughout the peptide. Resonance overlap largely precludes observation of $d_{\alpha N}(i,i+2)$ and $d_{\alpha N}(i,i+3)$ NOEs for Mb-H. For Mb-G, the evidence is less strong for ordered helical conformations: the only medium-range NOE connectivities observed are $d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+3)$ NOEs between Ile^{107} and Ala^{110}.

Self-Association of Mb-G and Mb-H. Since our primary interest is in events that precede extensive tertiary interactions in the folding of proteins, it is imperative to ascertain whether the peptides are monomeric under the conditions of study. To investigate this, one-dimensional NMR spectra and CD spectra were recorded for Mb-H and Mb-G for peptide concentrations ranging from approximately 10 μM to over 10 mM. For Mb-

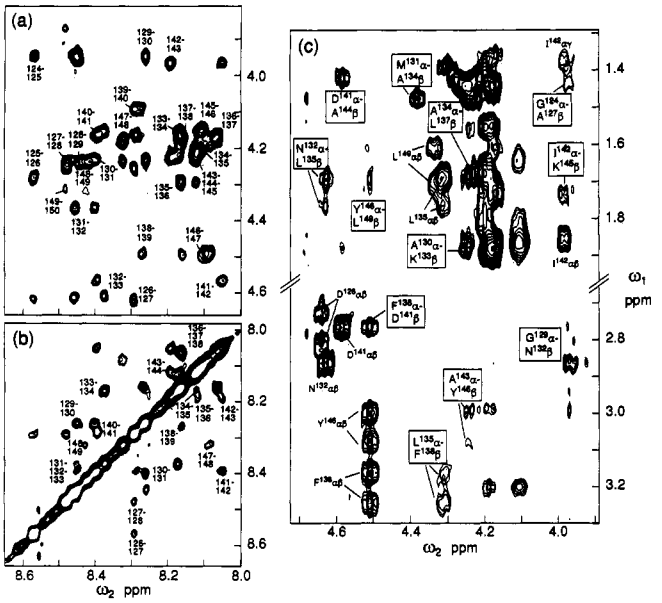

FIGURE 4: (a) NH–C$^{\alpha}$H region; (b) NH–NH region of a 500-MHz Hahn-echo NOESY spectrum (τ = 300 ms) of Mb-H (1.0 mM) in 90% ^1H_2O/10% ^2H_2O at pH 4.0 and 278 K. Sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOE connectivities are indicated, and peaks of weaker intensity represent intraresidue $d_{\alpha N}(i,i)$ NOEs. (c) C$^{\alpha}$H–C$^{\beta}$H region of a 500-MHz Hahn-echo NOESY spectrum (τ = 400 ms) of Mb-H (0.8 mM) in ^2H_2O at pH 4.0 and 278 K. $d_{\alpha\beta}(i,i+3)$ NOE connectivities are boxed.

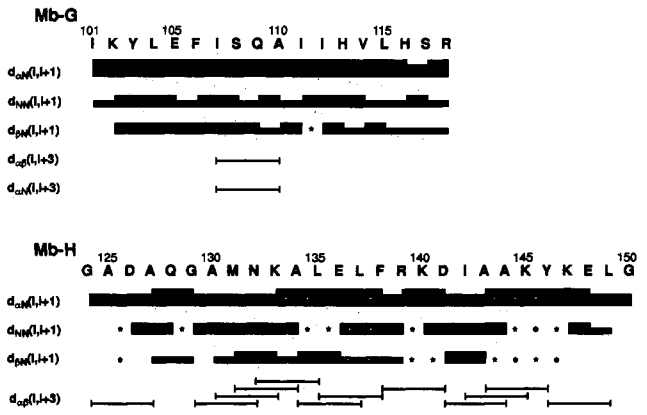

FIGURE 5: Schematic diagram showing observed NOE connectivities between the backbone amide, C$^{\alpha}$H and C$^{\beta}$H protons for Mb-G and Mb-H. NOE connectivities that are unobservable as a result of resonance overlap are indicated by black asterisks. White asterisks indicate connectivities for which strong but ambiguous cross peaks are observed.

H, the line width of the proton resonances, the chemical shift of certain resolved resonances, and [Θ_{222}] varied significantly with concentration, as shown in Figure 6. Clearly, Mb-H undergoes a self-association process in which populations of associated states become significant at concentrations greater than 1 mM.

The extent of association of the peptide can be determined by fitting an equilibrium curve to the experimental data plotted as a function of peptide concentration. Of the three parameters shown in Figure 6, the NMR chemical shift measurement can be made with the most accuracy and precision, since the C$^{\delta}H_3$ resonance of Ile^{142} is shifted ~0.3 ppm upfield upon increasing the concentration of Mb-H from 0.2 to 10 mM and is resolved throughout this concentration range. An indication of the stoichiometry of the monomer–oligomer equilibrium was obtained by fitting the data to equilibrium functions for different monomer–oligomer equilibria. For each calculation,

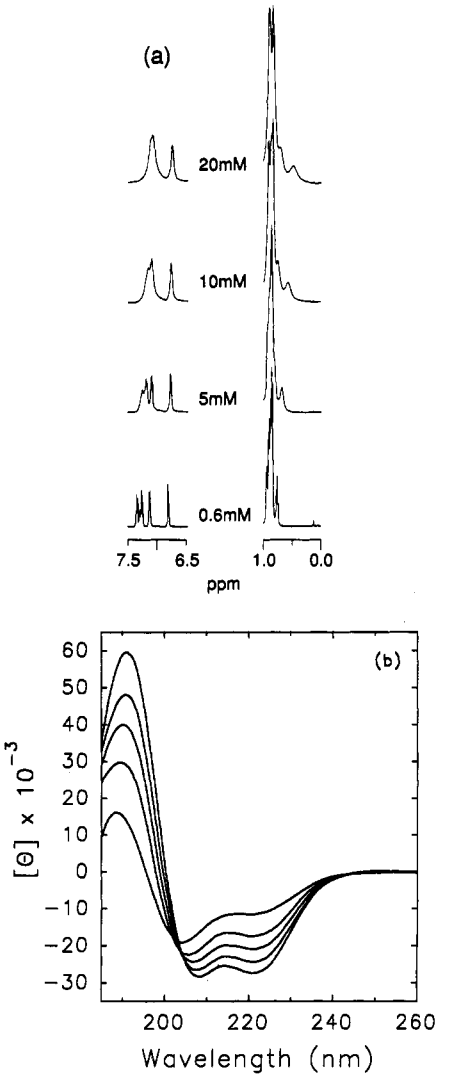

FIGURE 6: (a) Concentration dependence of the NMR spectrum of Mb-H in ^2H_2O at pH 4.0 and 278 K. The regions shown include the C$^{\delta}H_3$ resonance of Ile^{142} near 0.5 ppm and the C$^{\epsilon}H_2$ resonance of Tyr^{146} at ~6.8 ppm. (b) Concentration dependence of the CD spectrum of Mb-H at pH 4.0 and 278 K. Concentrations were (in descending order at 222 nm) 0.025, 2.26, 3.25, 4.64, and 6.5 mM. Cell path length was varied from 0.01 to 1.0 mm to compensate for the concentration changes.

the values of three parameters, δ_M, δ_N, and K_N, the chemical shifts of the monomer and of the oligomer and the equilibrium constant for the monomer–N-mer equilibrium respectively, were allowed to float. The chemical shift data are plotted in Figure 7a, together with nonlinear least-squares best-fit curves for monomer–dimer, –trimer, –tetramer, –pentamer, and –hexamer equilibria (inset). The data at high concentration are not sufficiently well-determined to allow quantitative distinction to be made between the various curves, but a visual inspection of the curvature of the concentration dependence in the vicinity of 1 mM indicates that the best fit is to a monomer–tetramer equilibrium, with a corresponding association constant of 4.4×10^6 M^{-3}, and limiting values of the chemical shift of 0.783 ppm (monomer) and 0.329 ppm (tetramer). Attempts were made to measure the molecular weight by gel filtration, but these, while clearly showing association, could not give a reliable measure of the stoichiometry, due to dissociation of the oligomer under the chromatography conditions.

The upfield shift of the I^{142} C$^{\delta}H_3$ methyl group in the monomer relative to random coil shifts probably reflects the

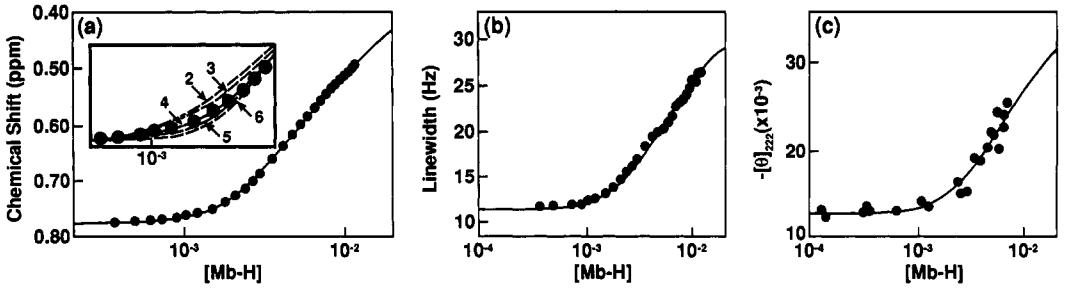

FIGURE 7: Concentration dependence of spectroscopic parameters for Mb-H. (a) Chemical shift of the $C^{\delta}H_3$ resonance of Ile^{142}. (b) Line width of the $C^{\epsilon}H_2$ resonance of Tyr^{146}, including the ~8 Hz coupling between the $C^{\delta}H$ and $C^{\epsilon}H$ protons. (c) $[\Theta_{222}]$ at 0 °C. Solid lines are least-squares fits assuming a monomer–tetramer equilibrium. The inset to (a) shows a closer view of the region near 1 mM, with least-squares fits assuming monomer–dimer, monomer–trimer, monomer–pentamer, and monomer–hexamer equilibria shown as dotted lines.

proximity of the aromatic side chains of Phe^{138} and Tyr^{146} in the helical structure; these interactions are stabilized and strengthened in the tetramer. The resonance of the I^{142} $C^{\delta}H_3$ in native myoglobin also experiences an upfield ring-current shift, to –0.10 ppm (Dalvit & Wright, 1987), the major contribution to which comes from Phe^{138}.

Below 1 mM, the line widths of the NMR resonances are characteristic of a monomeric species in this molecular weight range at 5 °C (Waltho et al., 1989); the line widths increase monotonically with concentration above 1 mM. This increase follows the same concentration dependence as the variation in chemical shift and can be fitted to equilibrium curves in the same way. The overall width of the $C^{\epsilon}H$ resonance of Tyr^{146} (including contributions from coupling) is plotted as a function of concentration in Figure 7b. A nonlinear least-squares fit to the monomer–tetramer equilibrium gives an association constant of 4.7×10^6 M^{-3}, in excellent agreement with that obtained by analysis of the chemical shift data. The calculated line width of this resonance is 2.8 Hz for the monomer, and 26 Hz for the tetramer, an indication of the sensitivity of line width to the changes in correlation time caused by association in these systems. Correlation times estimated from NOE buildup curves for the fixed distance $C^{\delta}H$–$C^{\epsilon}H$ protons of Y^{146} (data not shown) are consistent with the measured linewidths, assuming that the tetramer consists of a compact array similar to a four-helix bundle. Since the line width was not found to be temperature dependent, we infer that contributions from intermediate exchange-broadening phenomena within the associated state can be neglected.

The molar ellipticity at 222 nm ($[\Theta_{222}]$) shows similar behavior as a function of concentration (Figure 7c). The association constant calculated for the monomer–tetramer equilibrium from the CD data is 3.9×10^6 M^{-3} with a limiting value of $[\Theta_{222}]$ for the tetramer of –46 000 deg cm^2 $dmol^{-1}$. However, the limiting $[\Theta_{222}]$ may not be very accurate because of experimental uncertainty in the molar ellipticities at high concentrations.

The behavior of Mb-G is quite unlike that of Mb-H. Initial NMR studies at a concentration of 3 mM showed an unusual temperature dependence. With increase of temperature from 278 to 318 K, the line widths of all resonances increase, and the integrated intensities decrease by a factor of 5. These changes are reversed on cooling. At the higher temperatures the viscosity of the solution is increased substantially, consistent with the peptide forming a gel phase, which also appears to occur upon addition of sodium chloride (~200 mM). The changes in the CD spectrum with increasing temperature (Figure 1b) are consistent with formation of β-like secondary structure, but this could not be confirmed by NMR due to the apparent aggregation. At 278 K no line width, chemical shift, or intensity changes are observed over the concentration range 0.4–0.8 mM, establishing conditions under which CD and NMR studies can be made without interference from self-association processes.

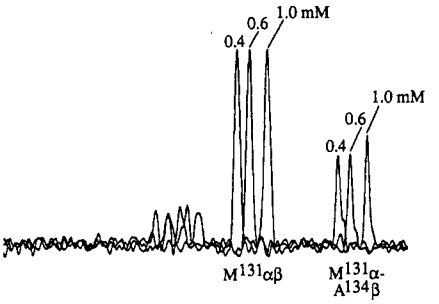

FIGURE 8: Cross sections (columns) through the $C^{\alpha}H$ resonance of Met^{131} from a series of NOESY spectra (τ = 400 ms) in ^2H_2O at pH 4.0 and 278 K, at peptide concentrations of 0.4 mM, 0.6 mM, and 1.0 mM. The cross sections are normalized according to the height of the $d_{\alpha\beta}(i,i)$ resonance and are offset for clarity.

Oligomeric Forms Do Not Contribute to the Helical NOEs of Monomeric Mb-H. The intensity of the nuclear Overhauser effect is dependent upon the correlation time of the molecule. In practice this means that NOEs for larger molecules build up more rapidly. The presence of even very small amounts of fully helical oligomeric forms in rapid equilibrium with monomer could contribute to the observed NOEs. Thus it is important to establish that the NOEs observed at low concentrations of Mb-H faithfully reflect the conformational preferences of the monomer and do not contain contributions from transferred NOEs through exchange with oligomeric species.

A series of NOESY spectra were acquired at concentrations of 6.0, 1.0, 0.6, and 0.4 mM, in order to assess the contribution of the oligomer population to the NOESY spectrum observed for monomeric Mb-H. Cross sections comparing the $d_{\alpha\beta}(i,i+3)$ and intraresidue $d_{\alpha\beta}(i,i)$ intensities for M^{131} are shown in Figure 8. The difference in the chemical shifts of the $C^{\beta}H$ resonances of M^{131} and A^{134} between the concentrations 0.4, 0.6, and 1.0 mM shown in Figure 7 is negligible, as it is for the I^{142} $C^{\delta}H_3$ (Figure 7a). A small (~10%) increase in intensity is observed for the $d_{\alpha\beta}(i,i+3)$ NOE compared to the $d_{\alpha\beta}(i,i)$ NOE for the sample at 1.0 mM (Figure 8). We ascribe this to the appearance of a small proportion of oligomeric forms in the conformational ensemble, even at concentrations as low as 1.0 mM: as a result of the differences in correlation time between the monomeric and associated states, small contributions to NOE intensity from associated states are observable even at concentrations where chemical shift, line width, or $[\Theta_{222}]$ indicate that the peptide is almost completely monomeric (Figure 7). It is highly significant that these

experiments unequivocally confirm the presence of $d_{\alpha\beta}(i,i+3)$ NOEs in the monomeric state of Mb-H. At a concentration of 0.4–0.6 mM, the NOEs observed faithfully represent the conformational ensemble of the monomeric Mb-H peptide and cannot be ascribed either to transferred NOEs or to intermolecular NOEs arising from the presence of oligomeric forms. Such a rigorous proof that helix exists in a monomeric peptide has not been previously reported using NMR methods.

NOE Ratios as a Measure of Local Helix Population in Mb-H. For short linear peptides, which fluctuate over an ensemble of conformations in aqueous solution, the sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs by themselves provide no information on secondary structure (Wright et al., 1988; Dyson & Wright, 1991). However, the relative intensities of these sequential NOEs do provide a measure of the relative population of backbone dihedral angles in the α- and β-regions of (ϕ,ψ) space (Dyson & Wright, 1991). In the β-region (extended-chain), the $d_{\alpha N}(i,i+1)$ distance is very short (ca. 2.2 Å in an ideal β-sheet) whereas the $d_{NN}(i,i+1)$ distance is long (ca 4.3 Å) (Wüthrich et al., 1984). The opposite is true in the α_R region of (ϕ,ψ) space, where the $d_{\alpha N}(i,i+1)$ distance is ca. 3.4 Å and the $d_{NN}(i,i+1)$ distance is ca. 2.3 Å. The magnitude of the NOE is proportional to the inverse sixth power of the internuclear distance, and the observed magnitude of a given NOE in a conformationally averaged system is a weighted average according to the populations of the various conformers. In practical terms, the differences in the $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ distances between conformers with backbone dihedral angles in the α and β regions of (ϕ,ψ) space mean that the relative intensities of the sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs for any pair of residues will vary according to the populations of α and β backbone conformations at this position in the sequence.

Figures 3–5 show that both $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ sequential NOEs are observed throughout both Mb-G and Mb-H, in all positions for which the appropriate pairs of resonances can be resolved. The ratio R of the intensity of $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ NOEs at each position

$$R = [\mathbf{I}(NH_i,NH_{i+1})]/[\mathbf{I}(C^{\alpha}H_i,NH_{i+1})]$$

where **I** represents the NOE intensity, can be calculated using peak volume integrals or peak heights in cross sections of the NOESY spectrum. Examples for each peptide are shown in Figure 1 of the supplementary material (see paragraph at end of paper). The values of R calculated for Mb-H using volume integrals from a NOESY spectrum are large, above 1.0 in parts of the peptide, indicative of a substantial population of conformations in the α region of (ϕ,ψ) space. However, the buildup rate of $d_{\alpha N}(i,i+1)$ NOEs in monomeric Mb-H is very rapid, indicating that the peptide also samples conformations of the extended-chain (β) type in the conformational ensemble. The values of the ratio R calculated for Mb-G are substantially lower than for Mb-H, about 0.3 on average. The Mb-G peptide also samples backbone conformations in both the α and β regions of (ϕ,ψ) space, but the population of α conformations is lower than for Mb-H, consistent with the lower proportion of helix present in Mb-G indicated by the CD spectrum and by the relative absence of medium-range NOEs (Figure 5).

Attempts have previously been made to quantitate actual populations of helix from NOE data solely on the basis of the relative intensities of the sequential interresidue NOEs (Bradley et al., 1990). However, such estimates are likely to be inaccurate since the $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ NOEs reflect the population of both folded and unfolded conformers; the latter may include states that have backbone dihedral

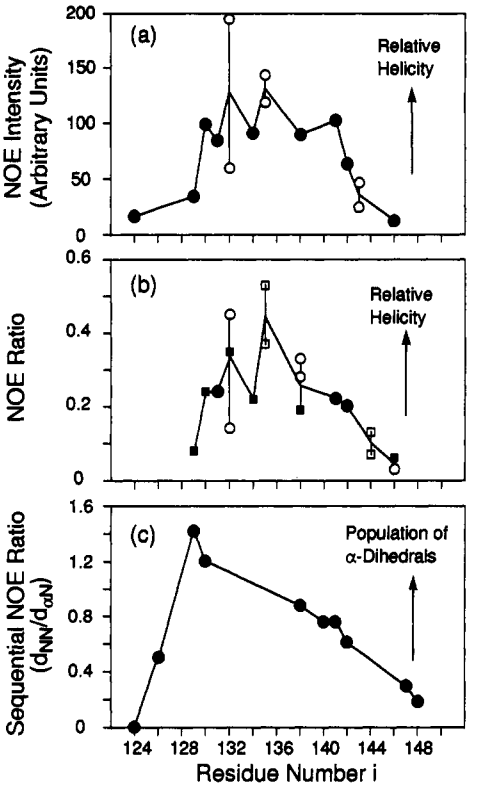

FIGURE 9: (a) Intensity of the $d_{\alpha\beta}(i,i+3)$ NOE connectivity (volume integrals in arbitrary units obtained from the NOESY spectrum shown in Figure 4c), plotted against number of residue i for Mb-H. The volumes are scaled according to the number of equivalent protons contributing to the NOE. Filled circles represent NOEs observed between $C^{\alpha}H$ and a single $C^{\beta}H$ resonance (scaled appropriately where two or more magnetically equivalent $C^{\beta}H$ protons are involved); empty circles indicate NOEs to residues with inequivalent $C^{\beta}H_2$ protons giving rise to two resolved resonances (the solid line is drawn to the average NOE intensity for the two protons.) (b) Ratio of the intensity of the scaled $d_{\alpha\beta}(i,i+3)$ NOE to the intensity of the $d_{\alpha\beta}(i,i)$ NOE (circles) or the $d_{\alpha\beta}(i+3,i+3)$ NOE (squares). The filled and empty symbols have the same meaning as in part (a). In cases of inequivalent $C^{\beta}H_2$ protons at the $i+3$ position, the NOE ratios were calculated only between $\alpha\beta(i,i+3)$ and $\alpha\beta(i+3,i+3)$ cross peaks corresponding to the same $C^{\beta}H$ frequencies. (c) Ratio of the intensities of the sequential $d_{NN}(i,i+1)/d_{\alpha N}(i,i+1)$ NOEs, calculated from volume integrals obtained from the NOESY spectrum in Figure 4a,b, plotted against number of residue i.

angles in the α region of (ϕ,ψ) space. By this simple method, nascent helical and even turn conformations would contribute to the estimated fraction of helix, since they give rise to sizeable $d_{NN}(i,i+1)$ NOE connectivities. The behavior of Mb-H illustrates this point, as we show in the following analysis.

The relative intensities of the $d_{\alpha\beta}(i,i+3)$ NOEs can provide insight into the relative helicity at each residue in the peptide. The intensities of the $d_{\alpha\beta}(i,i+3)$ NOEs, scaled for the number of protons involved, are plotted in Figure 9a. It is difficult to quantitate the local helix population on the basis of these NOEs alone, since the $d_{\alpha\beta}(i,i+3)$ distance is rather variable (2.5–4.4 Å) in regular helices in protein X-ray structures (Wüthrich et al., 1984), where side chain χ_1 angles are often fixed by packing in the tertiary structure. However, in the case of an isolated peptide helix, conformational transitions between different rotamers are to be expected (and are indeed observed in solvated molecular dynamics simulations; D. A. Case, personal communication) such that the average $d_{\alpha\beta}(i,i+3)$ distance is closer to the lower limit of the range observed in proteins. Although precise quantitation of the population of helix is not possible from these medium-range NOEs alone, they do provide a reliable qualitative guide to

the relative helicity throughout the peptide. To correct for possible variations in effective correlation time, especially at the ends of the peptide, the ratios of the intensities of the $d_{\alpha\beta}(i,i+3)$ and intraresidue $d_{\alpha\beta}(i,i)$ or $d_{\alpha\beta}(i+3,i+3)$ NOEs can be used (Figure 9b). In the case of the Mb-H peptide, both the direct $d_{\alpha\beta}(i,i+3)$ NOE intensities and the NOE ratios provide the same qualitative view of the local helical propensities.

Population of Helix in Mb-G. The only unambiguous $d_{\alpha\beta}(i,i+3)$ NOE in the spectrum of monomeric Mb-G is that between I^{107} and A^{110} (Figure 5). The ratio $I[d_{\alpha\beta}(107,110)]/I[d_{\alpha\beta}(110,110)]$ for this NOE has a value of 0.14, consistent with the lower $d_{NN}(i,i+1)/d_{\alpha N}(i,i+1)$ ratios observed for Mb-G, and the general agreement between all of the spectroscopic methods that Mb-G contains considerably less helix than Mb-H. In addition, the detection of only a single $d_{\alpha\beta}(i,i+3)$ NOE for Mb-G, compared to the substantial number seen for Mb-H is almost certainly the result of the lower population of helical forms in Mb-G.

Conformational Equilibrium of Mb-H. It is notable that both the absolute magnitude of the $d_{\alpha\beta}(i,i+3)$ NOEs (scaled for the number of protons) and the ratios of $d_{\alpha\beta}(i,i+3)$ to the intraresidue NOEs vary considerably for different regions of Mb-H, being largest in the center of the peptide and smallest at the ends (Figure 9). We interpret these data in terms of a model in which the peptide rapidly interconverts between an ensemble of conformations that include both helical and unfolded states at each amino acid. The population of helical conformations at any given residue is directly reflected in the magnitude of the $d_{\alpha\beta}(i,i+3)$ NOE. The highest population of helix in the Mb-H peptide is seen from Figure 9a and b to be between residues 130–145, with decreasing populations, i.e., fraying, toward the N and C termini of the peptide. The central, most helical region of the peptide is also that in which the amide proton temperature coefficients are lowest (Figure 2b), suggesting that backbone hydrogen bonding plays a significant role in the stabilization of the helical structure. In addition, the amide proton resonances of most residues in the region 134–147 are shifted significantly upfield relative to those of the N- and C-terminal residues which have chemical shifts close to random coil values (Figure 2a). The periodic variation in NH chemical shifts reported for amphipathic helices in proteins (Kuntz et al., 1986) or in peptides in TFE (Zhou et al., 1992; Jimenez et al., 1992) are not observed for Mb-H in aqueous solution. The mean population of helix between residues 130–145 appears to be of the order of 40%, since the average $^3J_{HN\alpha}$ for these residues is 6.3 ± 0.5 Hz, which is not quite midway between the coupling constants expected for helix (~4 Hz; Wüthrich, 1986) and for a random coil peptide with backbone dihedral angles in the β-region of conformational space (~8 Hz).[2] This is in good agreement with the CD estimate of overall mean helicity (30%) for Mb-H, since the NMR data suggest that 16 out of 27 residues are of the order of 40% helical, i.e., approximately 25% helix overall.

The amide temperature coefficients and the chemical shift data for Mb-H (Figure 2) suggest that states involving intramolecular hydrogen bonding of backbone amides begin to be populated at around residue 134. This is the position to be expected for the first $(i,i+4)$ hydrogen bond if residue 130 is the first to participate in ordered helix, as indicated by the medium-range NOE data shown in Figures 5 and 9. A more extensive discussion and a comparison of the amide proton chemical shifts in different peptides containing part or all of the sequence of the myoglobin H-helix is presented in the third paper in this series (Shin et al., 1993b).

A picture similar in general but significantly different in detail is obtained by considering the sequence dependence of the $d_{NN}(i,i+1)/d_{\alpha N}(i,i+1)$ ratio, shown in Figure 9c, which provides a measure of the relative population of backbone dihedral angles at each residue in the α- and β-regions of conformational space. In the C-terminal region of the peptide, the ratios involving the medium-range NOEs and the sequential NOEs are consistent, indicating both a high population of α-dihedrals and helix. In the N-terminal region of the peptide (residues 124–129), the observable $d_{\alpha\beta}(i,i+3)$ NOEs and NOE ratios are small but the $d_{NN}(i,i+1)/d_{\alpha N}(i,i+1)$ ratio is apparently large. The $^3J_{HN\alpha}$ coupling constants for Mb-H are consistently low in the center of the peptide, and these low values extend as far as residue 125, next to the N terminus (Table I). These data suggest the presence of a significant population of unfolded states with backbone dihedral angles in the α region of conformational space with only a low population of helix. The amide proton temperature coefficients in the N-terminal region are uniformly large (average 7.7 ± 1.0 ppb/K for residues 124–129; Figure 2b), suggesting that they are solvent-exposed and do not participate in hydrogen-bonded secondary structure.

The N-terminal portion of Mb-H thus appears to consist largely of unfolded conformations in aqueous solution which are unusual in that they populate α dihedrals. No $d_{\alpha N}(i,i+2)$ NOE connectivities suggestive of nascent helical conformations are observed in this region. The behavior in the presence of TFE of Mb-H and a 25-residue peptide (Mb-GH25) that includes the N-terminal portion of the H-helix sequence is consistent with these results (Shin et al., 1993a,b).

Residues 124–130 of sperm whale myoglobin include two glycines and an aspartate, which would be expected to be helix-breaking under the conditions of the experiments. In addition, there are three alanines and a glutamine, helix-promoting residues, interspersed with the helix-breakers. This mixture of residue types may explain the properties of this region of the peptide: the alanines promote a high population of backbone conformations in the α region of conformational space, but the helix-breaking residues prevent formation of stable, ordered helix. This is confirmed by the behavior of a peptide with Gly^{129} substituted by Glu, where helicity increases from 30% to 50% (Hughson et al., 1991). A completely helical conformation is stabilized in this region in the native sequence of the folded protein and in associated states of the H-helix peptide. We therefore conclude that the local amino acid sequence reduces the propensity for helix formation in the N-terminal portion of Mb-H in the isolated peptide, although tertiary interactions in the native protein or in associated states can stabilize helical conformations for residues 124–130.

Other Factors That Affect Helicity in the H-Helix Sequence. A shorter fragment of the H-helix of sperm whale myoglobin (Mb-F1; residues 132–153) was previously investigated for evidence of helix formation (Waltho et al., 1989). Medium-range NOE connectivities clearly demonstrate the presence of helical conformations in Mb-F1, although the CD spectrum shows that the helix is considerably less populated in Mb-F1 than in Mb-H. We have shown in the present work that the most helical region of Mb-H includes residues 130–145; the Mb-F1 sequence consists of residues 132–153. Destabilization of the helix in Mb-F1 is probably due primarily

[2] The mean $^3J_{HN\alpha}$ coupling constant is 7.9 ± 0.6 for residues that apparently populate predominantly the β-region of (ϕ,ψ) space (i.e., no $d_{NN}(i,i+1)$ NOEs are observed) in a series of unstructured peptides derived from plastocyanin (Dyson et al., 1992b).

to the truncation of the region of maximum helicity of Mb-H (130–145) by elimination of residues 130 and 131. In addition, the N-terminus is blocked by acetylation in Mb-H, while the termini are unblocked in Mb-F1. Blocking of the N and C termini in peptides has the effect of promoting helix by removing unfavorable interactions of the charged termini with the helix dipole (Shoemaker et al., 1987a). This factor exerts a measurable effect on the conformational ensemble of Mb-F1: studies of terminally blocked Mb-F1 peptides show greater helix formation both by NMR and CD (Waltho et al., 1990; J. P. Waltho, V. A. Feher, and P. E. Wright, unpublished observations).

Time Scales for Initiation of Protein Folding. For many proteins refolding from a denatured state occurs very rapidly, within seconds. Exceptions to this occur when folding involves *cis–trans* isomerism of X-Pro bonds (Brandts et al., 1975) or the rearrangement of disulfides (Creighton, 1988). Initiation of protein folding through formation of secondary structure (for example, formation of β-turns or helices) is likely to occur on the submicrosecond time scale: the stabilization of secondary structure through tertiary interactions commences within 10 ms or less, as shown by NMR pulse labeling experiments (Roder et al., 1988; Udgaonkar & Baldwin, 1988, 1990; Matouschek et al., 1990; Bycroft et al., 1990; Radford et al., 1992; Lu & Dahlquist, 1992).

The monomeric Mb-H peptide spontaneously forms highly populated helical structures under conditions that favor protein folding; from this we infer that the H-helix of intact myoglobin will spontaneously populate helical conformations irrespective of the state of folding of the remainder of the polypeptide chain. Thus, using the terminology of Baldwin and co-workers (Shoemaker et al., 1987b), the H-helix may be considered to be an autonomous folding unit, capable of folding independently of tertiary interactions with other parts of the polypeptide chain.

The NMR experiments place a lower limit on the rate of interconversion of folded and unfolded forms of the Mb-H peptide: since only one set of resonances is observed, exchange must be fast on the NMR time scale, i.e., faster than 10^4 s^{-1} based on the upfield shift of the Ile^{142} $C^{\delta}H_3$ resonance. Indeed, rates measured for helix-coil transitions of homopolypeptides are of the order of 10^6–10^7 s^{-1} (Lumry et al., 1964; Cummings & Eyring, 1975), giving a probable upper limit on the rate of folding of sequences such as the H-helix. It thus seems highly likely that an equilibrium population of helical conformations will be formed in the H-helix region during folding of apomyoglobin, long before the appearance, in less than 5 ms (P. A. Jennings and P. E. Wright, submitted for publication), of stabilized secondary structure in the A-, G-, and H-helices in a compact folding intermediate.

CONCLUSIONS

The data obtained from both CD and NMR spectroscopy show that the peptide Mb-H in a monomeric state in aqueous solution populates ordered helical conformations, in rapid equilibrium with unfolded states, throughout the entire peptide. From the relative intensities of the medium-range NOE connectivities, the helicity is seen to be greatest in the central part of the peptide. An important advantage of NMR is that information is obtained on the precise location of secondary structure within the sequence. In addition, the NOE is very sensitive to even quite small populations of folded structures with characteristic short interproton distances, even in the presence of conformational averaging (Wright et al., 1988; Dyson & Wright, 1991).

In this paper we have used synthetic peptides derived from the sequences of the G- and H-helices of sperm whale myoglobin to investigate possible initiating events in the folding pathway of myoglobin. A rigorous study of the conformational preferences of each peptide has revealed that Mb-H populates helical states in water solution in the monomeric state and that it undergoes a concentration-dependent association to a tetramer with a helix content approaching 100%. The Mb-H peptide thus spontaneously adopts a conformation in aqueous solution that, apart from end fraying and some sequence-specific effects at the N terminus, is close to that observed for the same sequence in the folded protein. On the other hand, Mb-G shows little propensity for helix formation in aqueous solution, although an ordered helix can be readily induced by the addition of TFE (Shin et al., 1993a,b). Helix-coil transitions of Mb-H occur on a submillisecond or submicrosecond timescale, consistent with a role in the early initiation events of protein folding.

ACKNOWLEDGMENT

We thank Afshin Karimi for assistance with CD spectra and Linda Tennant for expert technical assistance.

SUPPLEMENTARY MATERIAL AVAILABLE

Cross sections of NOESY spectra of Mb-G and Mb-H (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Anfinsen, C. B. (1973) *Science 181*, 223–230.
Baldwin, R. L. (1989) *Trends Biochem. Sci. 14*, 291–294.
Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.
Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.
Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A., & Kuntz, I. D. (1990) *J. Mol. Biol. 215*, 607–622.
Brandts, J. F., Halvorson, H. R., & Brennan, M. (1975) *Biochemistry 14*, 4953–4963.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.
Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature (London) 346*, 488–490.
Chazin, W. J., & Wright, P. E. (1987) *Biopolymers 26*, 973–977.
Chen, Y.-H., Yang, J.-T., & Chau, K. H. (1974) *Biochemistry 13*, 3350–3359.
Christensen, H., & Pain, R. H. (1991) *Eur. Biophys. J. 19*, 221–229.
Creighton, T. E. (1985) *J. Phys. Chem. 89*, 2452–2459.
Creighton, T. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5082–5086.
Cummings, A. L., & Eyring, E. M. (1975) *Biopolymers 14*, 2107–2114.
Dalvit, C., & Wright, P. E. (1987) *J. Mol. Biol. 194*, 313–327.
Dalvit, C., Wright, P. E., & Rance, M. (1987) *J. Magn. Reson. 71*, 539–543.
Dobson, C. M. (1991) *Curr. Opin. Struct. Biol. 1*, 22–27.
Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519–538.
Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., & Lerner, R. A. (1985) *Nature (London) 318*, 480–483.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988a) *J. Mol. Biol. 201*, 161–200.
Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., & Lerner, R. A. (1988b) *J. Mol. Biol. 201*, 201–217.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992a) *J. Mol. Biol. 226*, 795–817.
Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992b) *J. Mol. Biol. 226*, 819–835.
Ebina, S., & Wuthrich, K. (1984) *J. Mol. Biol. 179*, 283–288.
Epand, R. M., & Scheraga, H. A. (1968) *Biochemistry 7*, 2864–2872.
Gerritsen, M., Chou, K.-C., Nemethy, G., & Scheraga, H. A. (1985) *Biopolymers 24*, 1271–1291.
Ghaderi, M. R., & Choi, C. (1990) *J. Am. Chem. Soc. 112*, 1630–1632.
Hermans, J., & Puett, D. (1971) *Biopolymers 10*, 895–914.
Hughson, F. M., Barrick, D., & Baldwin, R. L. (1991) *Biochemistry 30*, 4113–4118.
Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.
Jaenicke, R. (1991) *Biochemistry 30*, 3147–3161.
Jimenez, M. A., Blanco, F. J., Rico, M., Santoro, J., Herranz, J., & Nieto, J. L. (1992) *Eur. J. Biochem. 207*, 39–49.
Johnson, W. C., Jr. (1985) *Methods Biochem. Anal. 31*, 61–163.
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.
Kim, Y., & Prestegard, J. H. (1989) *J. Magn. Reson. 84*, 9–13.
Kuwajima, K. (1977) *J. Mol. Biol. 114*, 241–258.
Kuwajima, K. (1989) *Proteins 6*, 87–103.
Levinthal, C. (1968) *J. Chem. Phys.-Chem. Biol. 65*, 44–45.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749–4756.
Lumry, R., Legare, R., & Miller, W. G. (1964) *Biopolymers 2*, 489–500.
Matheson, R. R., & Scheraga, H. A. (1978) *Macromolecules 11*, 819–829.
Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., & Fersht, A. R. (1990) *Nature (London) 346*, 440–445.
Montelione, G. T., & Scheraga, H. A. (1989) *Acc. Chem. Res. 22*, 70–76.
Oas, T. G., & Kim, P. S. (1988) *Nature (London) 336*, 42–48.
Ohgushi, M., & Wada, A. (1983) *FEBS Lett. 164*, 21–24.
Osapay, K., & Case, D. A. (1991) *J. Am. Chem. Soc. 113*, 9436–9444.
Ptitsyn, O. B. (1991) *FEBS Lett. 285*, 176–181.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature (London) 358*, 302–307.
Rance, M., & Byrd, R. A. (1983) *J. Magn. Reson. 52*, 221–240.
Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.
Rance, M. (1987) *J. Magn. Reson. 74*, 557–564.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophy. Res. Commun. 117*, 479–485.
Roder, J., Elöve, G. A., & Englander, S. W. (1988) *Nature (London) 335*, 700–704.
Rose, G. D., Gierasch, L. M., & Smith, J. A. (1985) *Adv. Protein Chem. 37*, 1–106.
Satterthwait, A. C., Chiang, L. C., Arrhenius, T., Cabezas, E., Zavala, F., Dyson, H. J., Wright, P. E., & Lerner, R. A. (1990) *Bull. W. H. O. 68*, 17–25.
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., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987a) *Nature (London) 326*, 563–567.
Shoemaker, K. R., Fairman, R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987b) *Cold Spring Harbor Symp. L11*, 391–398.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature (London) 335*, 694–699.
Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197–8201.
VanGeet, A. L. (1969) *Anal. Chem. 40*, 2227–2229.
Waltho, J. P., Feher, V. A., & Wright, P. E. (1990) In *Current Research in Protein Chemistry* (Villafranca, J. J., Ed.) pp 283–293, Academic Press, San Diego.
Waltho, J. P., Feher, V. A., Lerner, R. A., & Wright, P. E. (1989) *FEBS Lett. 250*, 400–404.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.
Zhou, N. E., Zhu, B.-Y., Sykes, B. D., & Hodges, R. S. (1992) *J. Am. Chem. Soc. 114*, 4320–4326.