Peptide Models of Protein Folding Initiation Sites. 3. The G–H Helical Hairpin of Myoglobin†

Hang-Cheol Shin,‡ Gene Merutka, Jonathan P. Waltho,§ Linda L. Tennant, 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: As part of an extensive dissection of the folding pathway of myoglobin, a series of peptides corresponding to fragments of sperm whale myoglobin have been synthesized, and their conformational preferences investigated using circular dichroism and nuclear magnetic resonance spectroscopy in aqueous solution and in solvent mixtures containing water and trifluoroethanol. The behavior of short fragments corresponding to the sequences of the G- and H-helices of myoglobin and to the turn region between these helices has been described in accompanying papers. At the next level of complexity, peptide model compounds have been synthesized to explore the longer-range interactions which may take place in protein folding after initial secondary structure formation has occurred. A series of disulfide-bridged dimeric peptides containing the complete sequences of the G- and H-helices of myoglobin were synthesized and their conformational preferences examined. CD spectra indicate that disulfide-bridged peptides consisting of two H-helix sequences (Mb-HssH) and of one G- and one H-helix (Mb-GssH) are highly helical in water solution, as a result of intermolecular association. A 51-residue peptide, Mb-GH51, encompassing the entire G–H helical hairpin of myoglobin, including the turn sequence between the two helices, has been successfully synthesized by standard methods. This peptide was designed to be monomeric in aqueous solution. Mb-GH51 does not appear from CD spectra to contain any additional helix in water solution above what would be expected from an equimolar mixture of the G- and H-helix peptides. NMR spectra indicate that the turn conformation observed in shorter peptide fragments is retained in Mb-GH51 in high population. The addition of TFE results in the formation of some helix, though not as much as might be expected even from a simple combination of the ellipticities of the component helical peptides in TFE. Current experimental and theoretical studies of myoglobin folding implicate the G–H helical hairpin in the earliest stages of folding: the present results imply that other parts of the polypeptide chain may be participating in these early events to a greater extent than heretofore imagined.

INTRODUCTION

The early events in the folding of proteins must involve considerable restriction in the conformational space available to the polypeptide chain. The exact nature of this conformational restriction has been a matter of some controversy. Kim and Baldwin (1982) suggested that secondary structure formation preceded collapse of the polypeptide chain into a compact structure, while Chan and Dill (1990) argued on the basis of the thermodynamics of the hydrophobic effect for an initial collapse of the polypeptide followed by organization of secondary structure within the hydrophobic core of the molecule. A crucial observation which has lent weight to the hypothesis of early formation of secondary structure is the observation of secondary structure in peptide fragments of proteins in the absence of the stabilizing tertiary interactions found in the folded protein structure (Wright et al., 1988; Montelione & Scheraga, 1989; Kim & Baldwin, 1990). Early attempts to identify secondary structure formation within fragments of proteins were generally unsuccessful [see, for example, Epand and Scheraga (1968)]. This led to the conclusion that long-range interactions occurring within proteins were essential for the formation of secondary structure and that short linear peptides were largely unstructured in aqueous solution. A major reason for the dearth of observations of secondary structure in short peptides in aqueous solution was the absence of a sensitive probe for small populations of structured conformers within a conformational ensemble. However, one exception did exist—the C-peptide of ribonuclease A (residues 1–13) was found to populate helical conformations to approximately 30% (as assayed using circular dichroism) in aqueous solution at 1.7 °C (Brown & Klee, 1971). Since that time, many peptide fragments of proteins have been found to populate a variety of folded conformations in aqueous solution, in dynamic equilibrium with fully unfolded forms [reviewed in Dyson and Wright (1991)].

Both experimental and theoretical evidence implicates the G–H helical hairpin region of myoglobin as part of an early folding intermediate (Ptitsyn & Rashin, 1975; Cohen et al., 1980; Gerritsen et al., 1985; Bashford et al., 1988; Hughson et al., 1990, 1991; Chelvanayagam et al., 1992). Recent quenched flow pulse labeling experiments show that a compact intermediate containing the A-, G-, and H-helices is formed on the folding pathway of apomyoglobin within the dead time of the flow-quench apparatus (P. A. Jennings and P. E. Wright, submitted for publication). Hydrogen exchange pulse labeling is the most rapid technique presently available for structural characterization of protein folding intermediates; the formation of the A–G–H intermediate occurs more rapidly than can be quantitated by this method. We therefore turn to peptide model systems to elucidate the factors that contribute to its formation and stabilization.

A great deal of information is available from the previous studies (Waltho et al., 1993; Shin et al., 1993) on the pro-

† 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: Institute of Biological Sciences, Hanhyo Institutes of Technology, CPO Box 1751, Seoul, South Korea.

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

0006-2960/93/0432-6356$04.00/0

Table I: Peptides Synthesized[a]

	100 105 110 115 120 125 130 135 140 145 150
Mb (SW)	...P I K Y L E F I S E A I I H V L H S R H P G D F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G...
Mb-G19	Ac-P I K Y L E F I S E A I I H V L H S R-NH_2
Mb-H27	Ac-G A D A Q G A M N K A L E L F R K D I A A K Y K E L G-NH_2
Mb-G21	Ac-P I K Y L E A K S E A I I H V L H S G G C-NH_2
Mb-H30	Ac-C G F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G-NH_2
Mb-HssH	Ac-C G F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G-NH_2
	Ac-C G F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G-NH_2
Mb-GssH	Ac-P I K Y L E A K S E A I I H V L H S G G C-NH_2
	Ac-C G F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G-NH_2
Mb-GH51	Ac-P I K Y L E A K S E A I I H V L H S R H P G D F G A D A Q G A K N K A L E L Y R K D I A A K Y K E L G-NH_2

[a] Underlined letters show those residues which were altered from the sperm whale myoglobin sequence in order to maximize the water-solubility of these peptides.

pensities of local sequences contained within the 51-residue G–H hairpin segment for spontaneous secondary structure formation in aqueous solution. The isolated H-helix peptide is ~30% helical in aqueous solution and residues His^{119}–Asp^{122} form a well-populated turn. In addition, the G-helix segment of the Mb-GH25 peptide,[1] which is unfolded in aqueous solution, forms ordered helix in TFE, which may in this case model the effects of tertiary contacts in stabilizing helix. The G–H hairpin region of myoglobin thus contains a number of potential initiation sites for folding. Our next objective was to design a peptide model of the G–H helical hairpin to investigate interactions between elements of secondary structure that might lead to stabilization of a folding intermediate.

The complete G–H helical hairpin has been modeled in two ways, in the absence and in the presence of the G–H turn sequence. Covalent coupling of two synthetic peptides, corresponding to the G- and H-helices of myoglobin, has been achieved using appropriately placed disulfide linkages; this construct does not contain the G–H turn sequence. The complete 51-residue peptide encompassing the G-helix, the G–H turn, and the H-helix sequences, slightly modified to increase peptide solubility and to ensure monomeric behavior, has been successfully synthesized and characterized. These studies constitute part of an extensive investigation of myoglobin folding which aims to dissect its folding mechanism using synthetic peptide models to elucidate partially structured states that are likely to be formed during the early stages of folding, together with NMR hydrogen-exchange pulse labeling experiments on the intact protein for structural characterization of intermediates formed during the later stages of the folding process.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; Mb, myoglobin; Mb-G19, Mb-G21, Mb-GH5, Mb-GH25, Mb-GH51, Mb-H10, Mb-H27, Mb-H30, synthetic peptides of varying lengths corresponding to fragments of the G- and H-helices and the G–H hairpin region of myoglobin; Mb-GssG, Mb-HssH, Mb-GssH, disulfide-bridged homo- and heterodimers, respectively, of the G- and H-helix peptides of myoglobin; NOE, nuclear Overhauser effect; HPLC, high-performance liquid chromatography; 2QF COSY, double quantum-filtered two-dimensional correlated spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; ROESY, rotating-frame NOESY spectroscopy; 2Q, double quantum spectroscopy; R-COSY, relayed COSY 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.

MATERIALS AND METHODS

Peptide Synthesis. The peptides shown in Table I were synthesized using *t*-BOC chemistry as previously described (Shin et al., 1993). The composition of all peptides was checked by amino acid analysis. The synthesis and purification of a peptide of the size of Mb-GH51 was not a trivial task. Each amino acid in the sequence was added by double coupling of the activated amino acid derivative followed by capping of unreacted peptide with acetic anhydride. This procedure served to minimize formation of impurities that differ from the desired material by a single deletion and which are therefore difficult to separate in later purification steps. The peptide was removed from the resin with low/high HF cleavage and was purified on a Hitachi HPLC system with a Vydac C18 semipreparative column, using a linear water/acetonitrile gradient containing 0.05% trifluoroacetic acid (TFA). The final product consisted of a tailing peak on analytical HPLC, but this material gave by mass spectrometry (FAB-MS) species corresponding to the predicted molecular weight of Mb-GH51 (5704 M+H). No significant species corresponding to sequence deletions were observed. In addition the proton NMR spectrum suggested a high purity (>95%).

Specific Disulfide Formation. For the peptides Mb-GssG, Mb-GssH, and Mb-HssH, a protocol was developed to ensure that the specific cross-link required was made in high yield, excluding the formation of other species. This was especially important for the formation of the heterodimer Mb-GssH. The reagent aldrithiol-2 (2,2′-dithiodipyridine; Aldrich Chemical Co.) was employed to derivatize the cysteine thiol of Mb-G21, which was then coupled in a rapid reaction with the thiol form of Mb-H30. All solutions were saturated with argon gas, and the reactions were carried out under argon to prevent air oxidation of the thiols. Due to solubility problems, Mb-G21 thiol was dissolved first in a small quantity of concentrated acetic acid, then a solution of 0.1 M Tris-acetate containing 4 M urea at pH 7 was added. The urea was necessary to ensure that the peptide was an unfolded monomer. A saturated solution of aldrithiol-2 in the same buffer (0.1 ml) was then added; the reaction was complete in 5 min at room temperature. The pH was lowered to 2.0 with glacial acetic acid and the reaction mixture passed through a column of Sephadex G15 equilibrated in 0.2 M acetic acid to remove unreacted aldrithiol reagent. The peptide band was lyophilized, then redissolved in acetic acid followed by 0.1 M Tris, 4 M urea, pH 7.0 as before. The Mb-G21-aldrithiol adduct was mixed with Mb-H30-thiol in the ratio 1:1. Reaction takes place immediately.

The reaction was stopped by lowering the pH to 3.0 with concentrated acetic acid. The cross-linked Mb-GssH was purified using reversed-phase HPLC. Removal of unreacted aldrithiol reagent is crucial for good yields of the heterodimer, because it mediates thiol exchange in the Mb-G21-Mb-H30 reaction mixture, allowing formation of homodimers which are preferentially formed under the reaction conditions.

Circular Dichroism Spectroscopy. Circular dichroism spectroscopy was performed as described previously (Waltho et al., 1993).

Proton NMR Spectroscopy. Samples of Mb-GH51 were prepared for NMR spectroscopy in 90% ^1H_2O/10% ^2H_2O, and the pH was adjusted to a value between 4 and 5 with small additions of concentrated HCl or NaOH. Values reported are meter readings uncorrected for the deuterium isotope effect.

NMR experiments were performed on a Bruker AM500 spectrometer and processed as described previously (Shin et al., 1993). Double-quantum filtered COSY (2QF COSY; Rance et al., 1983) and phase-sensitive NOESY (Jeener et al., 1979) experiments were used to obtain proton resonance assignments for Mb-GH51. Hahn-echo pulses (Rance & Byrd, 1983) were used in the NOESY spectra for better baselines and solvent suppression. Extensive resonance overlap meant that complete assignments of all side-chain resonances could not be made, but backbone resonance assignments are complete. The solvent resonance was decoupled by irradiation during the relaxation delay and during the mixing time of NOESY experiments. For all experiments, the transmitter offset was placed in the center of the spectrum, coinciding with the water resonance. Spectral widths were 7000 Hz for 4K real points, or 12 500 Hz for 8K real points in ω_2.

RESULTS AND DISCUSSION

Design of a Peptide Model of the Myoglobin Folding Intermediate. The sequences of all peptides studied are shown in Table I, together with the portion of the sequence of sperm whale myoglobin from which they are derived. Except for a small number of amino acid sequence changes detailed below, the sequences are identical to those of sperm whale myoglobin. The 51-residue sequence in sperm whale myoglobin contains hydrophobic residues which interact with the A-, E-, and F-helices in the folded protein. Retaining these residues in the Mb-GH51 peptide could potentially reduce solubility or lead to oligomerization, as observed for Mb-H at high concentration (Waltho et al., 1993) and for the disulfide-cross-linked peptides (see later section). To ensure that the peptide was both soluble and monomeric, several hydrophobic residues were substituted with reference both to comparative sequence data for myoglobins (Dayhoff et al., 1972), in which such substitutions appear in myoglobins from other species, and to the X-ray crystal structure of sperm whale myoglobin (Kuriyan et al., 1986). Four residues were replaced: Phe^{106}/Ala, Ile^{107}/Lys, Met^{131}/Lys, and Phe^{138}/Tyr. The former two substitutions were also made in Mb-G21 and Mb-GssH peptides to improve solubility; since the hydrophobic side chains of residues 106 and 107 form part of the packing surface for the heme group in intact myoglobin, they are unlikely to be required for interhelical G–H contacts. Met^{131} and Phe^{138} form part of the hydrophobic surface of the G–H hairpin which packs onto the rest of the molecule and were replaced to prevent dimerization. None of these substitutions would affect the G–H interhelical contacts observed in the myoglobin crystal structure. As a secondary consideration, substitutions were made with residues that favor helix formation.

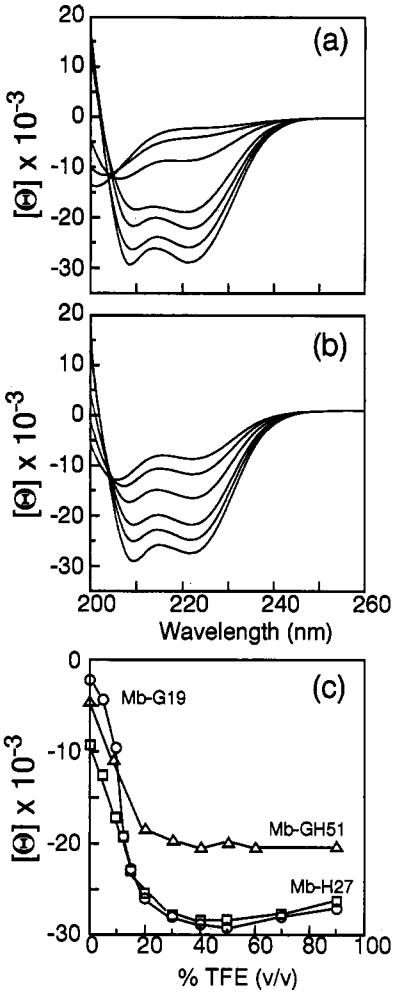

FIGURE 1: Effect of TFE on the CD spectrum of (a) Mb-G19 (17 μM), [TFE] = 0%, 5%, 10%, 12.5%, 15%, 20%, 50% (v/v) and (b) Mb-H27 (14.5 μM), [TFE] = 0%, 5%, 10%, 15%, 20%, 50% (v/v). (c) Ellipticity at 222 nm plotted as a function of TFE concentration for Mb-G19 (17 μM), Mb-H27 (15 μM), and Mb-GH51 (58 μM). All of the CD spectra were acquired at pH 5.0 ± 0.1, 278 K.

Conformational Preferences of Mb-G19 and Mb-H27 in TFE. As part of the design process, we first examined the propensity for helix formation in the peptides corresponding to the isolated G- and H-helical segments. The effect of TFE on the conformational preferences of Mb-G19 and Mb-H27 was measured by CD spectroscopy. Both Mb-G19 and Mb-H27 undergo concentration-dependent oligomerization at concentrations above 0.5–1 mM (Waltho et al., 1993) but are essentially monomeric at lower concentrations. The two peptides have very different helical propensities in aqueous solution: Mb-G19 adopts only a low population of helical conformations (<5%) at μM concentrations, while monomeric Mb-H27 shows a substantial amount (~30%) of helical structure (Waltho et al., 1993). By contrast, considerable stabilization of ordered helix was observed in the C-terminal end of the G-helix sequence but not in the N-terminal end of the H-helix sequence upon addition of TFE to peptide Mb-GH25 (Shin et al., 1993).

Titrations of Mb-G19 and Mb-H27 with TFE were monitored by CD and the results are shown in Figure 1. Mb-G19 shows a dramatic increase in helical content between 0% and 20% (v/v) TFE, with little change in ellipticity at concentrations above 20% (v/v). For Mb-H27, the increase in helix content with TFE is less steep, since it already contains ~30% helix by CD at 0% TFE. Probably coincidentally, the ellipticity ([Θ_{222}]) of Mb-H27 at TFE concentrations above

15% closely parallels that of Mb-G19, and the final $[\theta_{222}]$ obtained in the plateau region above 30% TFE is identical for the two peptides (−28 000 deg cm^2 dmol^{-1}). This is illustrated in Figure 1c. This value indicates that both peptides are approximately 80% helical at maximum in TFE (assuming 0 and −36 000 deg cm^2 dmol^{-1} for 0 and 100% helix respectively (Chen et al., 1974)). There is no further increase in helicity on increasing the TFE concentration to 90%; in fact, there is a slight decrease, which might even indicate some destabilization of helix at high concentrations of TFE.

Mixing of Mb-G19 and Mb-H27 in Water and TFE Solutions. Mixing experiments were performed to see whether the isolated G- and H-helix peptides can interact to stabilize additional secondary structure. Each peptide solution was placed in a separate compartment of a tandem cell and the CD spectra were measured before and after mixing. No change in helix content was observed on mixing, either in water or in solutions containing various concentrations of TFE. While no change in $[\theta_{222}]$ might in principle be expected at high TFE concentrations, assuming that the helix content of the individual helices is maximal, an augmentation of the helix content in water or at intermediate TFE concentrations might be expected if the two helical peptides were to associate to form nativelike structure in solution. The results indicate that such interactions do not occur under these conditions; similar observations have been made by Hughson et al. (1991).

Disulfide-Cross-Linked Peptides. The mixing experiments show clearly that the isolated Mb-G19 and Mb-H27 peptides do not interact in solution under the conditions of the CD experiments to form adducts with increased helical content. Therefore, new peptides were synthesized to evaluate the role of helix orientation and to increase the effective concentration of the G- and H-helices. This was achieved in two ways, one by synthesizing modified Mb-G and Mb-H peptides containing cysteine residues which could be site-specifically cross-linked, and the other by synthesizing a 51-residue peptide spanning the entire G–H hairpin sequence of sperm whale myoglobin. An advantage of using both approaches is that the effects of the turn sequence between the helical segments, well-populated in the shorter peptides Mb-GH5 and Mb-GH25, can, in principle, be evaluated.

Two cysteine-containing peptides, Mb-G21 and Mb-H30, were designed so that when cross-linked, the spacing between the two helical sequences would be similar to that in the native sequence. Mb-G21 contains two glycine residues, one replacing the C-terminal Arg of Mb-G19, followed by a C-terminal cysteine residue, giving a total of two extra residues at the C-terminus. Mb-H30 contains three extra residues at the N-terminus: an N-terminal cysteine, followed by glycine and the phenylalanine residue which takes up this position in the native protein. This residue was included, at possible risk to the solubility of the peptide, because it is apparently involved in extensive hydrophobic interactions with the G-helix in the intact protein (Kuriyan et al., 1986; Takano, 1977a,b; Hanson & Schoenborn, 1981).

Specific peptide cross-linking was achieved using 2,2′-dithiodipyridine (aldrithiol-2), to maximize formation of the desired dimer. Formation of Mb-HssH and Mb-GssH was facile, but problems with solubility of Mb-G21 prevented the formation of the analogous cross-linked peptide Mb-GssG in any reasonable yield.

The CD spectra of Mb-HssH and Mb-GssH show evidence of extensive helix formation in aqueous solution (Figure 2). The line widths of resolved ^{1}H resonances are considerably greater than would be expected for peptides of this size (Figure 3) and are comparable to those of tetrameric Mb-H27 (Waltho et al., 1993). In addition, the NMR spectrum of Mb-HssH closely resembles that of Mb-H27 at high concentration: chemical shifts, including that of an upfield-shifted methyl group (shown in Figure 3) identified as Ile^{142} in tetrameric Mb-H (Waltho et al., 1993), are very similar, indicating that the environment of the protons is similar in the two systems. We conclude from these results that the disulfide-bridged peptides Mb-HssH and Mb-GssH spontaneously associate in solution, probably forming dimers (on the basis of the similar NMR line widths observed for tetrameric Mb-H27). Since the peptides are extensively helical by CD, we conclude that the most likely structure in solution is probably some form of four-helix bundle. The behavior of these Mb peptides is similar to that observed for melittin, which associates to form a helical tetramer (Brown et al., 1980; Terwilliger & Eisenberg, 1982a,b).

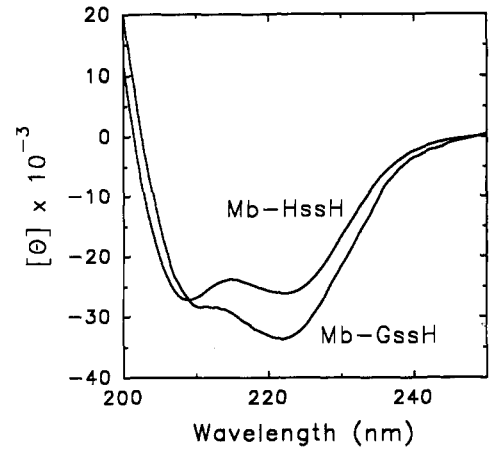

FIGURE 2: CD spectra of Mb-GssH (9 μM) and Mb-HssH (22 μM), at pH 5.0, 289 K.

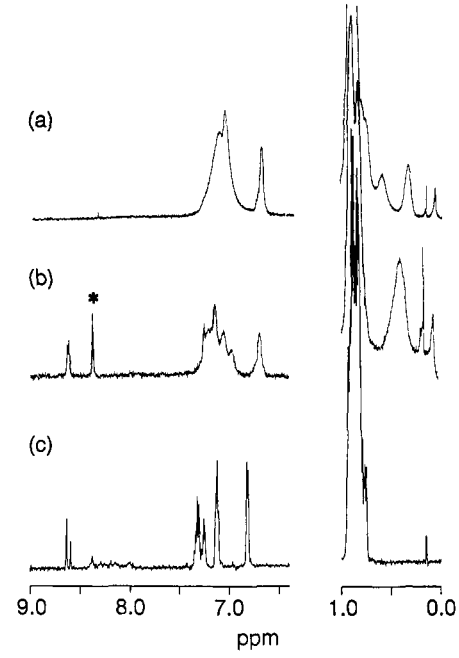

FIGURE 3: Portions of 500-MHz one-dimensional NMR spectra at 298 K, pH 4.0 in 99% ^2H_2O of (a) Mb-HssH (0.5 mM), (b) Mb-GssH (0.01 mM), and (c) Mb-GH51 (1.0 mM) showing differences in line width for aromatic proton resonances. Note that Mb-HssH contains six aromatic residues (4 Phe, 2 Tyr), Mb-GssH contains 7 (3 Phe, 2 Tyr, 2 His) and Mb-GH51 contains six (1 Phe, 3 Tyr, 2 His). The asterisk indicates the resonance of a small-molecular weight impurity.

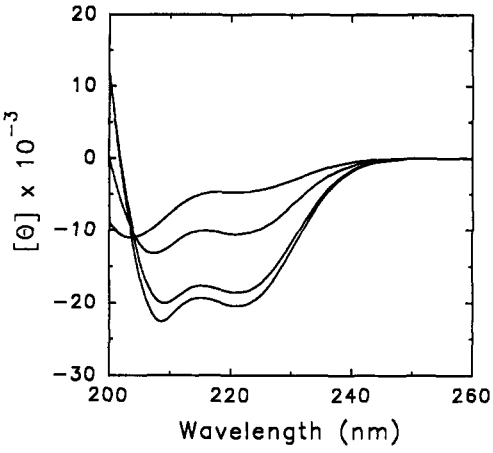

FIGURE 4: CD spectrum of Mb-GH51 (5.8 μM) as a function of TFE concentration (0%, 10%, 20%, 50% v/v) at 278K in 1 mM sodium phosphate, pH 5.0 ± 0.1.

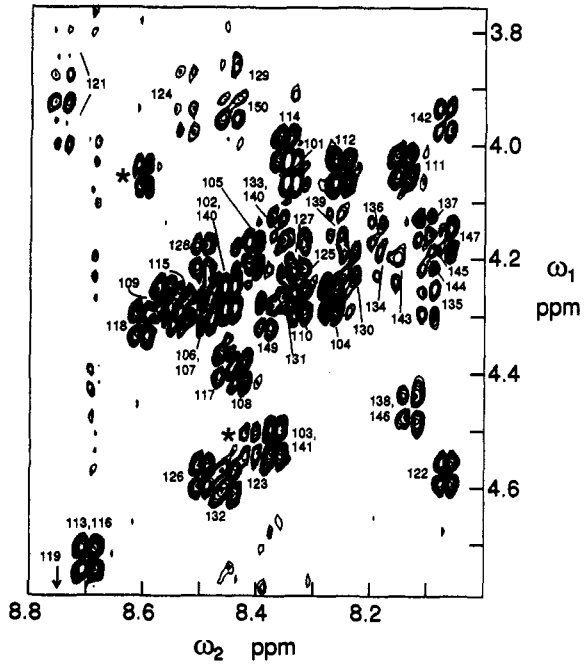

FIGURE 5: Portion of a 500-MHz 2QF COSY spectrum of Mb-GH51 in 90% ^1H_2O/10% ^2H_2O at pH 4.6 and 278 K showing cross peaks between NH and C$^\alpha$H. Positive and negative contour levels have been plotted without discrimination. Asterisks refer to peaks that arise from a small amount of impurity of lower molecular weight.

Solution Conformation of Mb-GH51. Mb-GH51 encompasses the entire G–H helical hairpin of myoglobin, residues 100–150. Its sequence is homologous to that of the cross-linked peptide Mb-GssH, but it includes the turn sequence of Mb-GH5 instead of the disulfide-containing linkage of Mb-GssH. This peptide was synthesized as an isolated helical hairpin in water solution, to explore the role of the central reverse turn in helix formation, stabilization, and termination.

The CD spectrum of Mb-GH51 is independent of peptide concentration from 10 μM to 2 mM, and a series of 1D NMR spectra at concentrations between 100 μM and 4 mM showed no difference in chemical shifts or line widths. We therefore conclude that Mb-GH51 behaves as a monomer under these conditions.

The CD spectrum of Mb-GH51 as a function of TFE concentration is shown in Figure 4. In water solution, the [Θ_{222}] value for Mb-GH51 is approximately midway between that of Mb-G21 and Mb-H27: there is no evidence for stabilization of additional helix over that present in the isolated

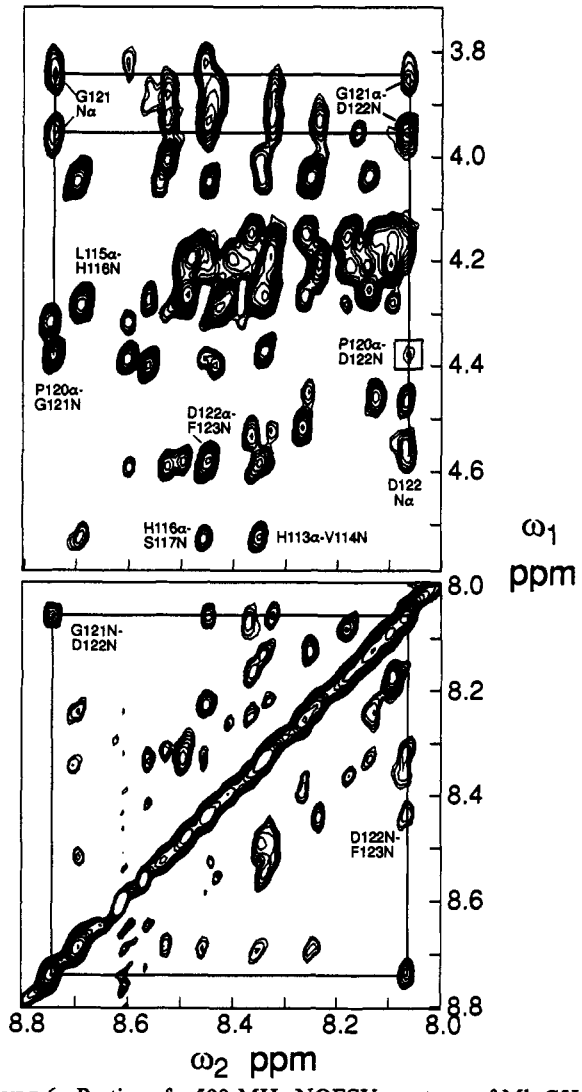

FIGURE 6: Portion of a 500-MHz NOESY spectrum of Mb-GH51 in 90% ^1H_2O/10% ^2H_2O at pH 4.6 and 278 K, mixing time 300 ms. The $d_{\alpha N}(i,i+2)$ NOE connectivity between Pro^{120} and Asp^{122} is shown in the box, plotted the same contour level as the rest of the figure.

G- and H-helix peptides, a result reminiscent of that obtained for the shorter Mb-GH25 peptide (Shin et al., 1993). The addition of TFE results in the stabilization of helix in Mb-GH51 (Figure 4), but the ellipticity at 222 nm is smaller than expected on the basis of the CD data for the two component peptides, Mb-G19 and Mb-H27. This most likely reflects a contribution of positive ellipticity at 222 nm from a central reverse turn, which serves to terminate the G-helix even in the presence of TFE (Shin et al., 1993), as well as low helicity in the N-terminal part of the H-helix as seen for both the H-helix peptide (Mb-H27) and Mb-GH25 (Waltho et al., 1993; Shin et al., 1993).

The ^1H NMR spectrum of Mb-GH51 in aqueous solution is typical of a peptide without a dominant folded conformation. Chemical shifts mostly fall within the ranges expected for random coil peptides, and coupling constants are averaged. Resonance overlap is a problem for such a long peptide. Using a combination of COSY and NOESY spectroscopy, the resonances of Mb-GH51 have been assigned, at least as far as the C$^\beta$H protons, and in some cases further out along the side chain. The fingerprint region of the 2QF COSY spectrum in 90% ^1H_2O/10% ^2H_2O is shown in Figure 5 and part of the NOESY spectrum in Figure 6. A Pro^{120}–Asp^{122} $d_{\alpha N}(i,i+2)$ NOE is observed (Figure 6), showing that the turn conformation involving residues 119–122 is highly populated. In

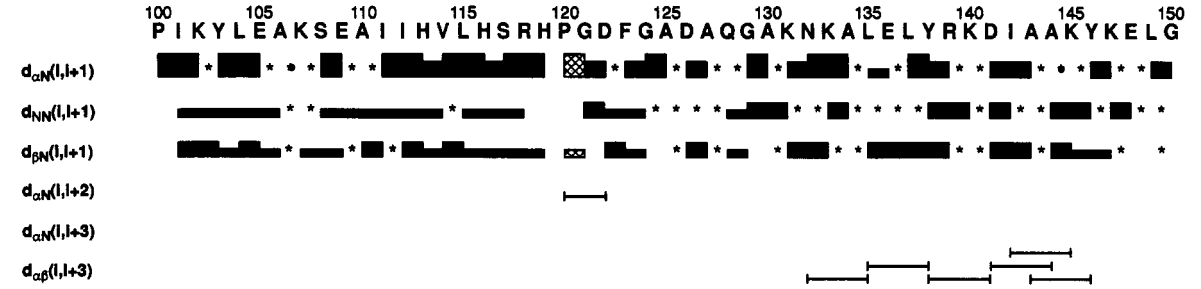

FIGURE 7: Summary of unambiguous short- and medium-range NOEs observed for Mb-GH51 in 90% ^1H_2O/10% ^2H_2O at pH 4.6, 278 K. The thickness of the bars corresponds to the intensity of the NOE; hatched bars represent connectivities to the $C^\delta H$ of proline, as a substitute for the NH. Asterisks denote connectivities obscured by resonance overlap. Possible medium-range NOEs that would be obscured by overlap with strong sequential or intraresidue cross peaks are omitted.

addition, the resonances of the turn residues have the same chemical shifts in Mb-GH5, Mb-GH25 and Mb-GH51 in water solution (Shin et al., 1993), a further indication that the turn conformation is present in the larger peptide.

As summarized in Figure 7, a number of unambiguous medium-range NOE connectivities are observed in the NOESY spectrum of Mb-GH51 in water solution. These are found largely in the middle of the H-helix portion of the peptide, indicating a significant population of helical conformations; medium-range NOEs are absent from the G-helix portion and the segment immediately following the central turn sequence, suggesting that an ordered helix is not formed. This behavior is consistent with that of the isolated G- and H-helix peptides under the same conditions (Waltho et al., 1993). Observation of unambiguous long-range NOE connectivities for Mb-GH51, for example, between the ends of the peptide if a helical hairpin structure were present, is made difficult by the extensive overlap of the proton resonances, but since all of the evidence points to a rather low population of ordered helix for a large portion of the peptide in water solution, such connectivities are probably not to be expected. No NMR measurements of Mb-GH51 in TFE solutions were attempted; serious difficulties with overlap and broad resonance lines preclude assignment under these conditions for a peptide of the length of Mb-GH51.

Does Mb-GH51 Form a Helical Hairpin Structure? NMR provides convincing evidence that local amino acid sequences in the 51-residue G–H sequence of myoglobin adopt very similar conformations in peptides of widely different length (this work; Waltho et al., 1993; Shin et al., 1993). Not only are the patterns of NOEs, particularly the medium-range connectivities, very similar for a given sequence in the different peptides, but the chemical shifts of the proton resonances are virtually identical. This is illustrated in Figure 8 for the amide proton chemical shifts in aqueous solutions of Mb-G, Mb-H (Waltho et al., 1993), Mb-GH5, Mb-GH25 (Shin et al., 1993), and Mb-GH51 (this paper). The only significant differences between peptides occur either at the termini or where amino acid substitutions have been made (asterisks in Figure 8). In particular, the amide proton chemical shifts for the turn residues Gly^{121} and Asp^{122} are almost invariant. Since these resonances are highly sensitive to turn population (Dyson et al., 1988; H. J. Dyson, L. Bolinger, and P. E. Wright, unpublished), it is clear that the turn is retained in all of the peptides. The excellent correspondence between NH chemical shifts in Mb-H and Mb-GH51 is strong evidence that a similar propensity for helix exists in the two peptides and that no further stabilization of secondary structure has occurred in Mb-GH51 due to the formation of an intramolecular helical hairpin structure. The chemical shift of the $C^\delta H_3$ of Ile^{142} is also a very sensitive measure of helix formation in Mb-H, which becomes >80% helical at high peptide concentrations

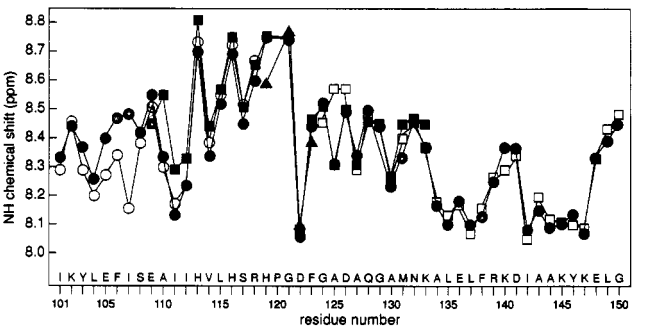

FIGURE 8: Chemical shifts of amide protons from peptides Mb-G (Waltho et al., 1993) (open circle), Mb-H (Waltho et al., 1993) (open square), Mb-GH5 (Shin et al., 1993) (filled triangle), Mb-GH25 (Shin et al., 1993) (filled square), and Mb-GH51 (Table II) (filled circle), plotted as a function of residue number. The sperm whale myoglobin sequence is shown at the bottom of the figure. Asterisks indicate residues that were replaced in various peptides.

(Waltho et al., 1993); this resonance is shifted upfield upon helix formation, which brings the side chains of Ile^{142} and Phe^{138} into close proximity (Waltho et al., 1993). The chemical shift observed for Ile^{142} $C^\delta H_3$ in Mb-GH51 (0.77 ppm) is essentially the same as that for monomeric Mb-H (0.78 ppm; Waltho et al., 1993), providing further evidence that additional helix stabilization through intramolecular hairpin formation does not occur in Mb-GH51 in aqueous solution. We therefore conclude that Mb-GH51 does not form a helical hairpin structure as a monomer in aqueous solution. This observation has profound implications for our understanding of the mechanism of folding of apomyoglobin.

Folding Studies on Apomyoglobin. Perhaps for historical reasons, because it was the first protein whose structure was solved by X-ray crystallography, myoglobin has been the focus of a number of studies exploring the relationship between the primary structure and the three-dimensional structure of the folded protein. Fragments of myoglobin were among the first peptide fragments of proteins to be investigated for conformation in solution (Epand & Scheraga, 1968; Hermans & Puett, 1971). The results obtained in these studies generally showed little secondary structure formation; more recent studies of the same peptides gave identical results by CD, but NMR revealed a significant population of local helical structures (Waltho et al., 1989, 1990). Clearly the site specificity and sensitivity of the NMR experiment are an advantage for peptide conformational studies, although the NMR data must be interpreted with caution because of conformational averaging (Dyson & Wright, 1991).

Helix packing has been explored in detail for myoglobin (Cohen et al., 1980; Richmond & Richards, 1978; Gerritsen et al., 1985) and possible folding pathways have been predicted based on packing considerations (Cohen et al., 1980; Ptitsyn & Rashin, 1975; Bashford et al., 1988; Chelvanayagam et al.,

Table II: Proton Resonance Assignments for Mb-GH51 (278 K, pH 4.6, 90% ^1H_2O/10% ^2H_2O)

	chemical shift (ppm)					
	NH	$C^\alpha H$	$C^\beta H$	$C^\gamma H$	$C^\delta H$	other
P^{100}		4.37	2.28, 1.87			
I^{101}	8.33	4.04	1.77	1.18, 1.49	0.85	0.75 ($C^\gamma H_3$)
K^{102}	8.44	4.27	1.69	1.43		
Y^{103}	8.37	4.52	2.94, 3.01		7.11	6.79 ($C^\epsilon H_2$)
L^{104}	8.26	4.27	1.53, 1.59	1.53	0.91, 0.84	
E^{105}	8.40	4.20	1.96, 2.04	2.33		
A^{106}	8.47	4.27	1.40			
K^{107}	8.48	4.29	1.84, 1.78	1.44		
S^{108}	8.42	4.40	3.85, 3.91			
E^{109}	8.55	4.27	2.08, 1.95			
A^{110}	8.33	4.28	1.37			
I^{111}	8.13	4.03	1.80	1.49, 1.18	0.85	0.78 ($C^\gamma H_3$)
I^{112}	8.24	4.04	1.77	1.43, 1.15	0.82	0.82 ($C^\gamma H_3$)
H^{113}	8.70	4.72	3.14, 3.21			
V^{114}	8.34	4.00	2.01	0.91, 0.85		
L^{115}	8.52	4.29	1.54, 1.46	1.61	0.86, 0.92	
H^{116}	8.69	4.71	3.16, 3.25			
S^{117}	8.45	4.39	3.81			
R^{118}	8.60	4.31	1.80, 1.72	1.57, 1.63		
H^{119}	8.75	5.00	3.24, 3.10			
P^{120}		4.38	2.32, 1.95			
G^{121}	8.74	3.84, 3.94				
D^{122}	8.06	4.58	2.62			
F^{123}	8.44	4.59	3.01, 3.20			
G^{124}	8.52	3.88, 3.92				
A^{125}	8.31	4.29	1.39			
D^{126}	8.49	4.58	2.69, 2.74			
A^{127}	8.34	4.19	1.42			
Q^{128}	8.50	4.20	2.06, 2.14	2.39	7.02, 7.65 ($N^\delta H$)	
G^{129}	8.44	3.92				
A^{130}	8.23	4.20	1.43			
K^{131}	8.33	4.21	1.84			
N^{132}	8.46	4.60	2.85	7.09, 7.79 ($N^\gamma H$)		
K^{133}	8.37	4.14	1.84, 1.88	1.46, 1.52		
A^{134}	8.17	4.21	1.46			
L^{135}	8.10	4.28	1.69, 1.52	1.68	0.96, 0.91	
E^{136}	8.18	4.15	2.08	2.29, 2.36		
L^{137}	8.10	4.14	1.62, 1.68			
Y^{138}	8.13	4.47	3.15, 3.30		7.13	6.80 ($C^\epsilon H_2$)
R^{139}	8.25	4.15	1.86	1.69, 1.63		
K^{140}	8.37	4.14	1.84			
D^{141}	8.37	4.54	2.73			
I^{142}	8.08	3.96	1.85	1.13, 1.34	0.77	0.88 ($C^\gamma H_3$)
A^{143}	8.15	4.22	1.44			
A^{144}	8.09	4.20	1.44			
K^{145}	8.10	4.16	1.73			
Y^{146}	8.13	4.48	2.98, 3.06		7.12	6.80 ($C^\epsilon H_2$)
K^{147}	8.07	4.17	1.81, 1.77	1.34, 1.38		
E^{148}	8.33	4.20	2.09, 2.00	2.35		
L^{149}	8.39	4.31	1.70	1.69	0.93, 0.88	
G^{150}	8.45	3.93				

N-terminal CH_3 2.10; C-terminal NH_2 7.18, 7.48

1992). Conformational energy calculations (Gerritsen et al., 1985) show that minimum energy arrangements of the helices A, G, and H of myoglobin invariably give the same nativelike structure. This suggests that a single packing arrangement is favored by these helices and that the G-H helical hairpin, with the A-helix packed against it, is a likely intermediate in the folding of myoglobin.

Such a structure has indeed been inferred from hydrogen exchange studies of the intermediate state of apomyoglobin observed at pH ~4 (Hughson et al., 1990). Protons from the A, G, and H helices were found to be protected in the low-pH form, while those from the B and E helices exchanged freely. The factors stabilizing the A-G-H folded region were thought to include nativelike hydrophobic interactions, but later mutagenesis studies (Hughson et al., 1991) suggest that the hydrophobic interactions are nonspecific and that this partly folded form of apomyoglobin may properly be termed a "molten globule intermediate".

The folding pathway of apomyoglobin has recently been studied experimentally using quenched flow hydrogen exchange pulse labeling methods (P. A. Jennings and P. E. Wright, submitted for publication). Most of the slowly-exchanging amide protons in the A-, G-, and H-helices are protected from exchange by formation of a compact folding intermediate within the 6-ms dead time of the quenched-flow apparatus. The formation of helical structure in helix-forming peptides has been found to be a prerequisite for slowed amide exchange (Rohl et al., 1992). Thus, protection of amide protons early in the folding pathway of myoglobin is an indication that secondary structure formation, in the form of helices, is rapid, a conclusion borne out by the observations described in the present three papers. Work with homopolypeptides suggests that helix formation can occur on a submicrosecond time scale (Lumry et al., 1964; Cummings & Eyring, 1975).

Stabilization of Helices in Folded Proteins and Peptides. Considerable research has been performed to elucidate the intrinsic factors that stabilize helix in monomeric peptides in aqueous solution, in an effort to understand, at least in part, helix formation in proteins. This work has been reviewed recently (Scholtz & Baldwin, 1992). For the myoglobin peptides considered here, it is likely that a number of stabilizing factors are operative, including specific hydrophobic interactions, specific electrostatic interactions, interaction of charges with the helix dipole, etc. However, for myoglobin, helix stabilization also appears to be intimately linked to the actual process of folding, i.e., to factors external to the local peptide sequence.

Such external factors have been noted previously for helical peptides. For example, sequences corresponding to helices in myohemerythrin were found to form nascent helix in water solution (Dyson et al., 1988, 1992a): the nascent helix structure was stabilized to form extended, ordered helix in the presence of TFE, by both CD and NMR criteria. By contrast, no helix is formed even in TFE in peptides which exhibit no tendency to populate backbone conformations in the α region of (ϕ,ψ) space (Dyson et al., 1992b; Zhou et al., 1992). Interestingly, the C-helix peptide of myohemerythrin, which contains two histidine residues near the N-terminus that ligate the diiron prosthetic group in the native protein, forms nascent helix in water solution and ordered helix in TFE solution only in the C-terminal half of the peptide, yet the N-terminal region is present as regular helix in the intact protein. In this case we infer that the N-terminal sequence has low probability of helix formation, as reflected in the failure to adopt helical conformations in TFE (Dyson et al., 1988). The helix is apparently stabilized in the intact protein by external factors, in this case presumably by coordination of the histidine side chains to the binuclear iron center in the folded holoprotein.

Additional interactions from distant parts of the polypeptide chain in native myoglobin must also play an important role in stabilizing the G- and H-helices during folding. Neither NMR nor CD are able to detect additional helix stabilization or packing interactions between the G- and H-helix portions of Mb-GH51 and Mb-GH25 in water solution. Similar negative results were obtained when the individual peptides were mixed. On the other hand, considerable stabilization of helix is observed for the disulfide-linked peptide Mb-GssH which from NMR experiments appears to dimerize, presumably to form a four-helix bundle. Similar stabilization of

helix occurs upon dimerization of Mb-HssH or upon formation of a tetramer of the Mb-H peptide (Waltho et al., 1993). We thus conclude that interactions between the G-helix and the H-helix segments of the polypeptide chain alone are insufficient to stabilize helical secondary structure beyond that observed in isolated peptides; additional interactions with other polypeptide chains in the Mb-GssH and Mb-HssH dimers or in the Mb-H tetramer are required for extensive helix stabilization. Since the G- and H-helices of apomyoglobin are folded and stabilized in a compact intermediate formed on the folding pathway of apomyoglobin (P. A. Jennings and P. E. Wright, submitted for publication), we infer that long-range interactions involving distant parts of the myoglobin polypeptide chain within a compact state are required to stabilize the G-H helical hairpin structure during protein folding.

The present work establishes clearly that the G-H helical hairpin is not stable in isolation. A number of calculations of solvent accessible surface area for myoglobin helices showed that the largest buried area on formation of helix pairs was for the helix G-helix H interaction (Richmond & Richards, 1978; Bashford et al., 1988; Gerritsen et al., 1985). These results have led several workers to propose that association of the G- and H-helices is an important early step on the myoglobin folding pathway (Lim & Efimov, 1977; Cohen et al., 1980; Bashford et al., 1988; Chelvanayagam et al., 1992). The free energy decrease upon association of helix G and helix H estimated from solvent accessible surface area calculations is large (~17 kcal/mol) (Bashford et al., 1988). It is therefore somewhat surprising that the helices do not associate readily in peptides such as Mb-GH51. This apparent discrepancy may be a reflection of the fact that these simple calculations overestimate the net stabilization free energy associated with burial of hydrophobic surface. Electrostatic calculations indicate that the free energy cost of burying polar groups such as backbone peptide bonds can effectively compensate for the stabilization due to burial of hydrophobic groups upon folding of a protein (Yang et al., 1992), consistent with the small values of the overall free energy of stabilization observed for most proteins. In addition, interactions between two isolated elements of secondary structure may involve loose and nonspecific hydrophobic contacts such as are believed to occur in molten globules, rather than the specific close packing of side chains characteristic of the tertiary structure of a folded protein. Given that the free energy of stabilization of the molten globule state of apomyoglobin is only ~2 kcal/mol (P. A. Jennings & P. E. Wright, submitted for publication), it is perhaps not surprising, in retrospect, that an isolated G-H helical hairpin is unstable. The low helix-forming propensity of the residues at the N-terminus of the H-helix sequence may also play a role in destabilizing the helical hairpin state. Experiments with peptide models are in progress to assess the effects of stabilizing the isolated G- and H-helices and the effect of association with the A-helix on stability of hairpin structures.

CONCLUSIONS

The overall conclusion from the work described in the present series of papers is that secondary structural elements can be formed at a number of sites within the 50-residue sequence spanning the G- and H-helices of myoglobin. These structures persist in peptide fragments of varying lengths, an indication of their potential as folding initiation sites in the intact protein sequence. However, we consistently find that long-range tertiary contacts appear to be necessary for the stabilization of a helical hairpin structure in the G-H sequence. The isolated H-helix peptide forms a tetramer at high concentrations, with concomitant formation of additional helix, and the disulfide-bridged peptides are highly helical in water solution, apparently due to stabilization by formation of aggregates, probably dimers. The G-helix also requires external interactions for stabilization; Mb-G19 does not form ordered α helix in aqueous solution, but the helix is strongly stabilized by TFE, as well as in aggregates such as Mb-GssH and in the folded protein. The peptide Mb-GH51, however, was designed to be strictly monomeric, and no intermolecular interactions were observed in water solutions. There is also no evidence of intramolecular interaction between the two helical halves of the peptide, perhaps due to the helix-destabilizing effects of the two glycine residues at the N-terminus of the H-helix. The chemical shifts are mostly consistent with those of peptides which populate predominantly unfolded conformations, and there is no NOE or CD evidence for stabilization of additional helix resulting from packing of the two helical segments. Given the helix stabilization observed for the H-helix at high concentration, caused by tetramerization, and that of Mb-GssH, caused by dimerization, it seems likely that the formation of a folded helical hairpin requires additional interactions from polypeptide sequences external to the isolated G-H sequence. These results imply that the G-H helical hairpin is not formed as a stable, autonomous folding unit during refolding of apomyoglobin, but is stabilized at an early stage ($<\sim$1 ms) by interactions with other regions of the polypeptide chain.

ACKNOWLEDGMENT

We thank Drs. Don Bashford, Barry Honig, and Patricia Jennings for insightful discussions.

REFERENCES

Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D., & Weaver, D. L. (1988) *Proteins 4*, 211–227.
Brown, J. E., & Klee, W. A. (1971) *Biochemistry 10*, 470–476.
Brown, L. R., Lauterwein, J., & Wüthrich, K. (1980) *Biochim. Biophys. Acta 622*, 231–244.
Chan, H. S., & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6388–6392.
Chelvanayagam, G., Reich, Z., Bringas, R., & Argos, P. (1992) *J. Mol. Biol. 227*, 901–916.
Chen, Y.-H., Yang, J.-T., & Chau, K. H. (1974) *Biochemistry 13*, 3350–3359.
Cohen, F. E., Sternberg, M. J. E., Phillips, D. C., Kuntz, I. D., & Kollman, P. A. (1980) *Nature 286*, 632–634.
Cummings, A. L., & Eyring, E. M. (1975) *Biopolymers 14*, 2107–2114.
Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5, National Biochemical Research Foundation, Washington, DC.
Dyson, H. J., & Wright, P. E. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 519–538.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988) *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.
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.
Hanson, J. C., & Schoenborn, B. P. (1981) *J. Mol. Biol. 153*, 117–146.
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.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
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.
Kuriyan, J., Wilz, S., Karplus, M., & Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133–154.
Lim, V. I., & Efimov, A. V. (1977) *FEBS Lett. 78*, 279–283.
Lumry, R., Legare, R., & Miller, W. G. (1964) *Biopolymers 2*, 489–500.
Montelione, G. T., & Scheraga, H. A. (1989) *Acc. Chem. Res. 22*, 70–76.
Ptitsyn, O. B., & Rashin, A. A. (1975) *Biophys. Chem. 3*, 1–20.
Rance, M., & Byrd, R. A. (1983) *J. Magn. Reson. 52*, 221–240.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophy. Res. Commun. 117*, 479–485.
Richmond, T., & Richards, F. M. (1978) *J. Mol. Biol. 119*, 537–
Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M., & Baldwin, R. L. (1992) *Biochemistry 31*, 1263–1269.
Scholtz, J. M., & Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 95–118.
Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., & Dyson, H. J. (1993) *Biochemistry 32*, 6348–6355.
Takano, T. (1977a) *J. Mol. Biol. 110*, 537–568.
Takano, T. (1977b) *J. Mol. Biol. 110*, 569–584.
Terwilliger, T. C., & Eisenberg, D. (1982) *J. Biol. Chem. 257*, 6010–6015.
Terwilliger, T. C., & Eisenberg, D. (1982) *J. Biol. Chem. 257*, 6016–6022.
Waltho, J. P., Feher, V. A., & Wright, P. E. (1990) In *Current Research in Protein Chemistry*, (Villafranca, J. J., Ed.) Academic Press, San Diego, pp 283–293.
Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 6337–6347.
Waltho, J. P., Feher, V. A., Lerner, R. A., & Wright, P. E. (1989) *FEBS Lett. 250*, 400–404.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167–7175.
Yang, A.-S., Sharp, K. A., & Honig, B. (1992) *J. Mol. Biol. 227*, 889–900.
Zhou, N. E., Zhu, B.-Y., Sykes, B. D., & Hodges, R. S. (1992) *J. Am. Chem. Soc. 114*, 4320–4326.