Molecular Hinges in Protein Folding: the Urea-Denatured State of Apomyoglobin†

Stephan Schwarzinger,‡ Peter E. Wright,* and H. Jane Dyson*

Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

Received May 24, 2002; Revised Manuscript Received August 8, 2002

ABSTRACT: Unfolded apomyoglobin in 8 M urea at pH 2.3 displays distinct regions with different backbone mobility, as monitored by NMR relaxation. These variations in backbone mobility can be correlated with intrinsic properties of the amino acids in the sequence. Clusters of small amino acids such as glycine and alanine show increased backbone mobility compared to the average. In contrast, local hydrophobic interactions that persist in urea denaturant cause some restriction of backbone motions on a picosecond to nanosecond time scale. The model derived from the behavior of apoMb in urea depends only on the most fundamental properties of the local amino acid sequence, and thus provides a feasible paradigm for the initiation of folding.

The earliest events in protein folding are thought to play a key role in directing the folding process, but are experimentally inaccessible because they occur so rapidly and involve large conformational ensembles with widely variable structures. NMR[1] quench flow methods can be used to obtain residue-specific information on processes that occur later in the folding process, but are limited by the 6−10 ms dead time of the apparatus. Faster methods, such as pressure jump, temperature jump, or continuous flow methods, employing probes such as fluorescence, circular dichroism, or infrared spectroscopy, do not generally provide information at amino acid resolution. One possible means to obtain information on the earliest events in folding is to use the NMR spectra of equilibrium models of unfolded and partly folded states to determine conformational preferences on a per-residue basis.

In the case of apomyoglobin (apoMb), a number of equilibrium models are available, some of which appear to correspond closely to intermediates in the kinetic folding pathway. For example, the equilibrium molten globule formed at pH 4.1 is found to be very similar to the kinetic molten globule intermediate observed in quench-flow experiments (*1*). NMR characterization at equilibrium of the molten globule (*2*) and acid-denatured (*3*) states of apoMb, together with a comparison of the behavior of the folded apoprotein and the heme-containing holoprotein (*4*), has provided a model for the folding pathway of myoglobin that is consistent in all respects with the known kinetic folding process. Even the acid denatured state, however, exhibits residual helix-like structure and local hydrophobic cluster formation (*3*). We therefore undertook the NMR characterization of apoMb denatured in urea at pH 2.3 to study the behavior of the polypeptide under fully denaturing conditions, which constitute the starting state for most kinetic folding experiments.

MATERIALS AND METHODS

$^{13}C/^{15}N$ doubly labeled sperm whale apomyoglobin for assignment purposes and ^{15}N labeled protein for relaxation studies were expressed and purified as described previously (*5*). NMR samples were prepared from HPLC-purified and lyophilized protein and typically contained 300 μM apoMb in a solution of 8 M urea in 90% ^1H_2O/10% ^2H_2O. The pH was adjusted to 2.3 ± 0.05 using dilute hydrochloric acid. The final chloride concentration was approximately 200 mM. The probe temperature for all NMR experiments was calibrated to 20 °C using neat methanol (*6*).

NMR experiments were carried out on a Bruker Avance DMX 750 MHz spectrometer with a triple resonance gradient probe head. Backbone assignment was carried out using multidimensional heteronuclear NMR methods as previously described (*3*, *4*). Because of the extremely poor dispersion of the C^α and C^β resonances, the dispersion of the ^{15}N and ^{13}CO resonances was employed to make backbone assignments (*7*), utilizing pulsed field gradient HNCA, HNCO (*8*), (HCA)CO(CA)NH (*9*), HNCACB (*10*), and CBCA(CO)NH (*11*) spectra. ^{15}N-HSQC-TOCSY and ^{15}N-HSQC-NOESY (*12*) were used in some cases to verify assignments. Some ambiguous assignments were resolved using samples selectively labeled at Ala, Arg, Gly, His, Lys, Met, Phe, Ser, and Tyr. Spectra were processed using NMRPipe (*13*) and analyzed with NMRView (*14*). Backbone ^{15}N relaxation measurements were carried out as described by Farrow et al. (*15*). Duplicate points were used to obtain an estimate of the error. T_1 relaxation times were measured with delays of 10 (2×), 50, 100, 200 (2×), 400, 600, 800, 1100, 1500 (2x), 2000, 3000, and 5000 ms, and T_2 experiments were carried

† This work was supported by Grant GM57374 from the National Institutes of Health and by an Erwin Schrödinger Fellowship (FWF J1736-CHE) to S.S.

* Corresponding authors. Address: Department of Molecular Biology, The Scripps Research Institute, MB-2, 10550 North Torrey Pines Road, La Jolla, CA 92037. Tel: (858) 784-9721. Fax: (858) 784-9822.

‡ Present address: Lehrstuhl Biopolymere, Universität Bayreuth, Germany.

[1] Abbreviations: NMR, nuclear nagnetic resonance; apoMb, apomyoglobin (without heme); HSQC, heteronuclear single quantum coherence; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; CD, circular dichroism; AABUF, average area buried upon folding.

10.1021/bi020381o CCC: $22.00
Published on Web 09/25/2002

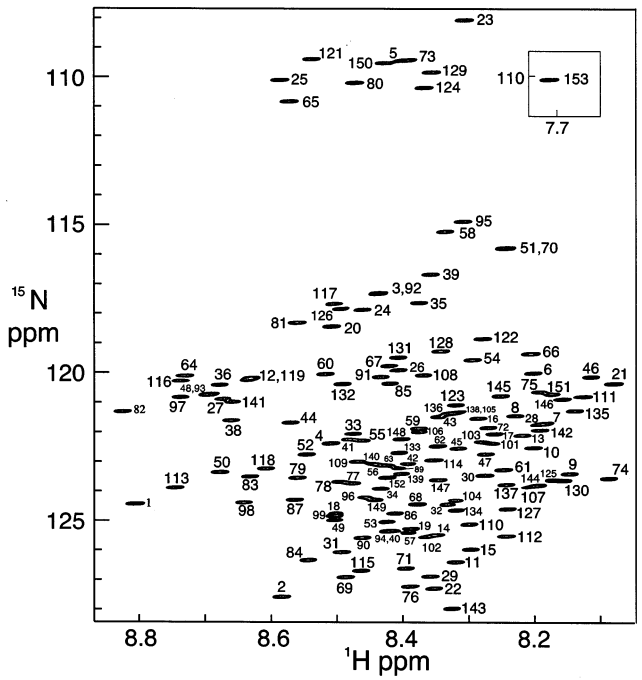

FIGURE 1: 750 MHz ^{1}H-^{15}N HSQC spectrum of apomyoglobin in 8 M urea, pH 2.3, at 20 °C. Assignments are shown. Inset: cross-peak for G153, shifted from its normal position on the far right of the spectrum.

out with delays of 6 (2×), 102, 202, 302 (2×), 350, 402, 450, 502 (2×), 550, 602, 750, and 902 ms. The rates were fit with the program Curvefit (*16*) to a two-parameter single-exponential decay. Heteronuclear NOE experiments were carried out using a 3.6 s saturation period and an additional 3 s recycle delay. The relaxation data were analyzed using the reduced spectral density mapping approach (*17*).

Analysis of the primary structure was carried out employing the parameters of Rose et al. (*18*) for the average area buried upon folding and the parameters of Kyte and Doolittle (*19*) for hydrophobicity. Hydrophobic cluster analysis was performed using HCA_Draw (*20*). Sequence alignments were done using CLUSTAL W (*21*) with sets of myoglobin sequences deposited in the PIR database. Glycine and alanine residues were analyzed for conservation and type of substitution, respectively, using web-based tools (*22*).

RESULTS

Urea Denaturation of apoMb at pH 2.3. The circular dichroism (CD) spectrum of apomyoglobin at pH 2.3 in the absence of urea shows a small amount of residual helical structure, which is found by NMR to be present in parts of the sequence corresponding to the A, D/E, and H helices of the folded protein (*3*). This residual helical structure is lost as urea is added. The ellipticity at 222 nm changes from -3500 deg cm^2 dmol^{-1} in the absence of urea, to -1500 deg cm^2 dmol^{-1} in 8 M urea. The ^{1}H-^{15}N HSQC spectrum (Figure 1) shows slightly less dispersion in the ^{1}H dimension than that of acid-unfolded apomyoglobin (*3*). However, sufficient resolution was present in the ^{15}N and ^{13}CO dimensions of spectra of the doubly labeled protein that complete backbone resonance assignments could be made. This was aided by the narrow line widths characteristic of unfolded proteins and by the spectral resolution afforded by the 750 MHz spectrometer.

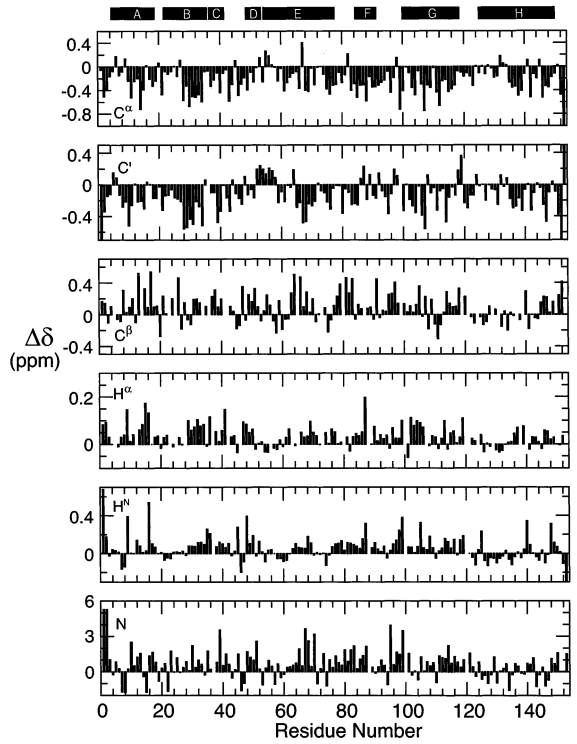

FIGURE 2: Secondary chemical shifts for apomyoglobin unfolded in 8 M urea, pH 2.3, 20 °C. Black bars at the top of the figure represent the positions of the helices A−H in the fully folded holoprotein.

Conformational Preferences from Chemical Shifts. With resonance assignments in hand, it is possible to determine a number of NMR parameters that give insights into the nature of the conformations present in the ensemble of the urea-denatured state. The chemical shifts themselves can be used to determine local backbone structural propensities, and the flexibility of the backbone can be probed by measuring ^{15}N-heteronuclear nuclear Overhauser effects (NOEs) and ^{15}N T_1 and T_2 relaxation experiments.

Deviations of the backbone chemical shifts from random-coil values measured for GGXGG peptides in 8 M urea at pH 2.3 (*23*) corrected for local sequence effects (*24*) are very small (Figure 2), confirming that the protein is fully denatured and showing that the residual helical structure found at pH 2.3 is no longer present. The secondary chemical shifts of the ^{13}C$^{\alpha}$, ^{1}H$^{\alpha}$, and ^{13}CO nuclei in some regions of the protein are consistent with a slightly increased population of backbone dihedral angles in the broad β region of ϕ,ψ space in the urea-denatured state, relative to the random-coil reference peptides. The tendency toward extended backbone conformations is most marked in regions rich in β-branched or bulky hydrophobic side chains, e.g., in the B helix region (I^{28}LIRLFK^{34}). Most other denatured proteins that have been studied (*25*−*28*) also show a slight preference for backbone dihedral angles in the β-region of (ϕ,ψ) space, but these proteins invariably have a high content of β-structure in their native states.

Small but consistent differences from the average behavior are found in the secondary chemical shifts of H$^{\alpha}$, C$^{\alpha}$, and CO in the D-helix region (residues 52−58). Although the secondary shifts are very close to the random coil values, they appear to show a very slight preference for backbone dihedral angles in the α region of (ϕ,ψ) space. It is important

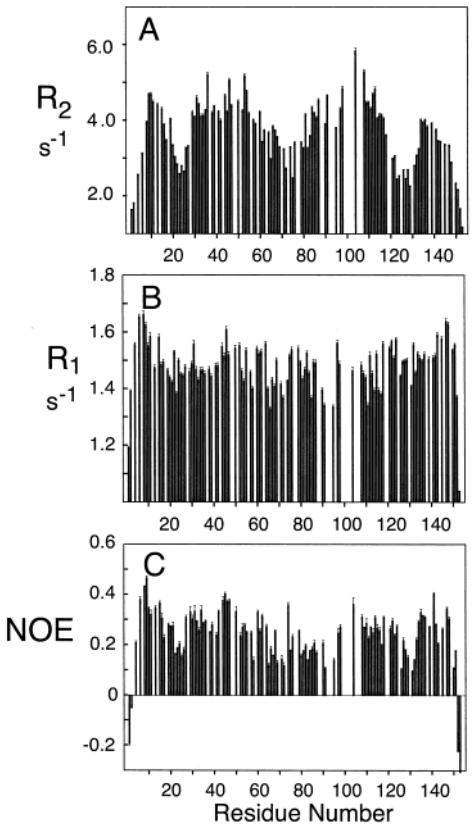

FIGURE 3: Relaxation data for apomyoglobin unfolded in 8 M urea, pH 2.3, 20 °C. (A) Relaxation rate R_2 (=$1/T_2$). (B) Relaxation rate R_1 (=$1/T_1$). (C) {^{1}H}-^{15}N heteronuclear NOE.

to emphasize that such small, yet clearly present, consecutive stretches of conformational preference can only be detected if effects on the chemical shift arising from the local amino acid sequence are considered (*24*). Chemical shifts are the only means to detect such small populations, which cannot be detected by methods based on NOEs or coupling constants.

Backbone Dynamics of Urea-Denatured Apomyoglobin. {^{1}H}-^{15}N heteronuclear NOE data and ^{15}N T_1 and ^{15}N T_2 relaxation data were recorded at a single magnetic field, 750 MHz (Figure 3). The high field is necessary to obtain sufficient resolution for the measurements. All of the data indicate flexibility throughout the polypeptide, as expected for a denatured protein. Some deviations from the average are present, including an increase in flexibility at the chain termini. Small values of R_2 and negative {^{1}H}-^{15}N NOEs for residues 2−4 and 150−153 indicate large amplitude fluctuations of the polypeptide backbone in these regions on a subnanosecond time scale. R_2 shows distinct minima around residues 21−25 and 121−129 and a broad minimum between residues 60 and 80. R_1 is relatively featureless, with a uniform average value of 1.48 s^{-1}. The heteronuclear NOE shows slightly reduced values in some regions, indicating increased flexibility on a nanosecond time scale.

DISCUSSION

Conformational Preferences in Urea-Denatured Apomyoglobin. The addition of urea denatures the residual helical structure that was found in apomyoglobin at pH 2.3 (*3*). In addition, there is no sign of the residual long-range interactions that were indicated by anomalously high R_2 values in the absence of urea (*3*). An 8 M urea solution at pH 2.3 presumably causes denaturation of a protein by two different mechanisms. First, the low pH causes an increase in the charge state, leading to broken salt-bridges and electrostatic repulsion. For apoMb, the charge at pH 2.3 is approximately +36, with 7 Asp and 14 Glu uncharged but with 19 Lys, 4 Arg, and 12 His, together with the N-terminus, all positively charged. Among the positively charged residues, the histidines are of particular importance for the folding and unfolding processes, since they titrate in the pH region of the transition between the pH-induced molten globule state at pH ∼4 and the folded apoprotein at pH ∼6. In fact, the interaction H24−H119 seems to be important for formation of the molten globule state (*29*).

The presence of 8M urea in the solution achieves the additional denaturation that distinguishes this state from that present at pH 2.3 in water (*3*). Although the mechanism of denaturation by urea has been under investigation for decades, there is still no commonly accepted model. Recent investigations suggest that the denaturation process occurs as a consequence of altered solvent properties rather than by specific binding of urea to the solute (see ref *30* for a review). Interaction of urea (or guanidine) with the protein serves to preserve the solubility of the denatured protein, whereas with other denaturants such as NaSCN unfolding and precipitation of the protein often occurs (*30*). Urea interacts with proteins in two ways, by hydrogen bonding to the protein backbone and by preferred solvation of hydrophobic groups. The backbone interaction is thought to be caused by one urea molecule replacing—thus freeing and increasing the entropy of—two water molecules in the immediate solvation sphere around the solute while still providing hydrogen bonds to both solvent and solute. This is in agreement with molecular mechanics simulations where elevated urea concentration is found in the inner solvation sphere (*31*). It is also in agreement with NMR-binding studies, which reveal only short contact times for urea on the protein surface, indicating only weak, and probably nonspecific interactions (*32*, *33*). Preferential solvation of hydrophobic residues may be mediated by differences in the energy required for cavity formation in urea and water solutions. Additionally, urea changes macroscopic properties of the solution such as viscosity (doubles compared to water), dielectric constant (increases compared to water thus electrostatic interactions have a shorter range), and surface tension (increases compared to water), which all have an effect on the balance of forces causing proteins to fold and unfold (*34*).

Correlation of Backbone Motions with Intrinsic Molecular Properties. Long-range interactions play a vital role in the stabilization of natively folded proteins but are largely absent in unfolded states under denaturing conditions. Local interactions and the intrinsic properties of the local amino acid sequence are likely to be the major determinants of deviations from random-coil behavior in unfolded proteins. The plot of R_2 versus residue number (Figure 3A) deviates significantly from the flattened bell-shaped profile expected for a random-coil polypeptide (*25*, *35*). The three distinct minima (residues 21−25, 60−80, and 121−129) all correspond to regions that are rich in Gly and Ala residues (Figure 4B).

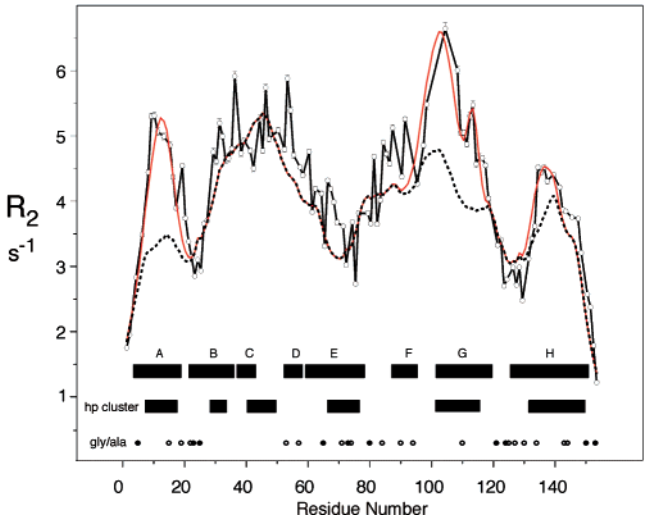

FIGURE 4: Relaxation rate R_2 for apomyoglobin unfolded in 8 M urea, pH 2.3, 20 °C (open circles joined by thin lines), plotted as a function of residue number. The dashed line shows the R_2 values calculated according to the simple model incorporating only side chain radius of gyration and with persistence length $\lambda_j = 7$ except for Gly and Ala ($\lambda_j = 1$). Note that use of $\lambda_j = 7$ for all residues has only a slight effect on the depth of the minima. The red line shows the result of the calculation incorporating both the radius of gyration and four clusters centered at residues 12 ($w_c = 5$; $R_{cluster} = 1.87$), 103 ($w_c = 7$; $R_{cluster} = 1.85$), 113 ($w_c = 2.5$; $R_{cluster} = 1.3$), and 135 ($w_c = 4$; $R_{cluster} = 0.8$). Calculated values have been scaled for comparison with the experimental data. Black bars represent the positions of the helices A−H in the fully folded holoprotein. Lower panels: hydrophobic clusters found by HCA (*20*) (black bars); positions of glycine residues (filled circles) and alanine residues (open circles) in the sequence of sperm whale apomyoglobin.

Glycine residues are unique among amino acids because of the missing side chain, resulting in a much higher degree of conformational freedom than any other residue. A correlation has been noted previously between the occurrence of glycine-rich regions and increased backbone flexibility in unfolded and denatured proteins (*36*−*38*). Alanine, which also causes increased backbone flexibility, although to a lesser extent than glycine, is also small compared to other amino acids.

Motional Model for Urea-Denatured ApoMb. The observation of significant decreases in R_2 in regions rich in Gly and Ala residues suggests that side chain size plays a significant role in determining the backbone dynamics of urea unfolded apomyoglobin. The R_2 relaxation data were analyzed using a simple model, in which sequence-dependent variations in backbone motions were simulated using the radius of gyration R_g of the different side chains (*39*) to describe the intrinsic flexibility of each residue. The intrinsic correlation time τ for each amino acid was assumed to be proportional to R_g^3, based on Stokes Law. The decreased conformational freedom associated with proline was modeled by increasing its effective side chain R_g to 2.0 Å (from 1.25 Å, *39*). The effect of neighboring amino acids on the motions of each residue was assumed to decay exponentially (*25*). The sequence dependence of the R_2 relaxation rate was then calculated from the expression:

$$R_{2i} \propto \sum_{j=1}^{N} \tau j \cdot e^{(-|i-j|)/(\lambda_j)} \ldots \quad (1)$$

where R_{2i} is the relaxation rate at each residue, N is the total number of residues, τ_j is the intrinsic correlation time of residue j, and λ_j is the persistence length for segmental motion of the polypeptide chain. The persistence length was taken to be 7 amino acids for all residues (*25*), except for glycine and alanine for which the value of λ_j, was set to 2.

Figure 4 shows the R_2 relaxation rates simulated using this simple model to describe sequence-dependent backbone motions (hatched line). While the profile of the simulated R_2 values reproduces many of the general features of the experimental data, particularly the minima between residues 21 and 25 and residues 121 and 129 and the broad minimum from residues 60−80, there are several regions where this simple model cannot explain the experimental R_2 values. In particular, the maxima between residues 4 and 20, residues 97 and 118, and residues 133 and 142 cannot be fit by a simple model based only on side chain size. This suggests that additional interactions within local clusters of residues influence the motion of the polypeptide backbone in these regions. The contributions of these local interactions to R_2 were modeled as a Gaussian distribution (*40*), and the sequence dependence of R_2 was simulated using the following expression:

$$R_{2i} = k \cdot \sum_{j=1}^{N} \tau j \cdot e^{(-|i-j|)/(\lambda_j)} + \sum_{cluster} R_{cluster} \cdot e^{(-(i-n_c)^2)/((w_c)^2)} \ldots \quad (2)$$

where K is an empirical scaling constant, $R_{cluster}$ is the intrinsic R_2 relaxation rate for residues in each cluster, centered at residue n_c, and w_c is the half width of the cluster. Inclusion of four clusters, centered at residues 12, 103, 113, and 135, in the model resulted in remarkably good agreement between the experimental and simulated R_2 profiles (Figure 4, thick red line). Interestingly, each of these clusters corresponds to regions with large surface area buried upon folding (see following section). There are two additional maxima in the buried surface area plot (see Figure 5), centered on residues 30 and 46. However, inclusion of clusters at these positions in the model calculation does not appear to be necessary to fit the experimental R_2 data. As noted for the acid-denatured state in the absence of urea (*3*), there is no correlation of the backbone motions with the hydrophobicity of the side chain (Figure 4B).

Analysis of Relaxation Data: Correlation with Buried Surface Area. Relaxation data for folded proteins are commonly analyzed by the "model-free" approach (*41*, *42*). However, this approach is inapplicable for partially folded and unfolded systems because of its underlying assumption of a single correlation time for all of the molecules in the ensemble. We have therefore used reduced spectral density function mapping (*17*) to analyze the relaxation data. This analysis yields spectral densities at three different frequencies, $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ (Figure 5A−C). Variations in $J(0)$ with apomyoglobin sequence (Figure 5A) reflect variations in nanosecond time scale backbone motions; in particular, there are three well-defined minima in the $J(0)$ profile, between residues 22 and 25, residues 72 and 76, and residues 123 and 129, that correspond to the Gly/Ala-rich regions. In the case of urea unfolded apomyoglobin, there is no evidence for microsecond−millisecond time scale motions that affect $J(0)$. The plot of $J(\omega_N)$ versus residue

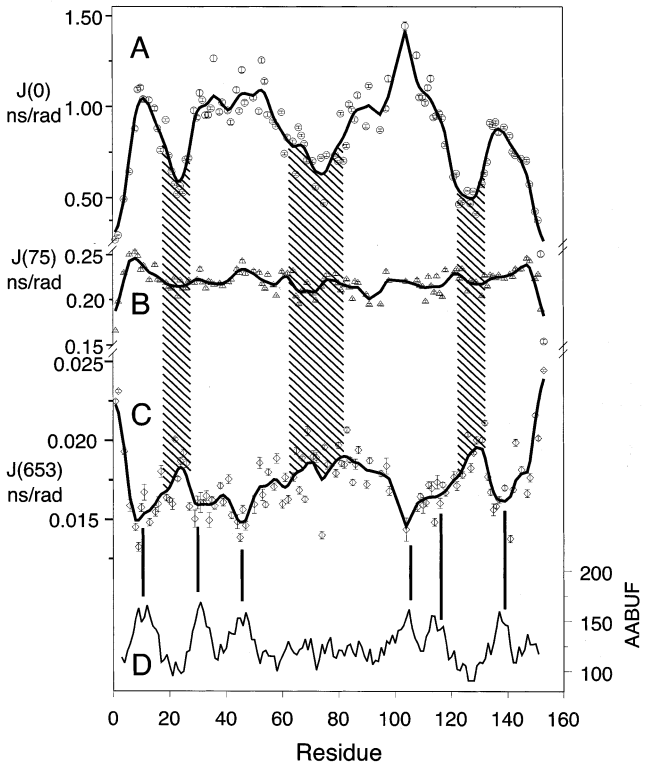

FIGURE 5: Calculated spectral densities for apomyoglobin in 8 M urea at pH 2.3. (A) $J(0)$, (B) $J(\omega_N)$ [$J(75)$], and (C) $J(\omega_H)$ [$J(653)$]. The data were smoothed with a kernel function with a bandwidth of 5, using the program Axum (bold lines). (D) average area buried upon folding (*18*). Vertical hatched boxes show the correlation between minima of $J(0)$ and maxima of $J(\omega_H)$. Black vertical lines show the correlation between minima of $J(\omega_H)$ and maxima of AABUF.

number (Figure 5B) is relatively featureless, except at the termini, and does not correlate in any obvious way with intrinsic properties of the amino acid sequence. Significant variations are observed in $J(0.87\omega_H)$, however (Figure 5C). Minima in the smoothed plot of $J(0.87\omega_H)$ versus residue number correlate with maxima in a plot (Figure 5D) of the average area buried upon folding (AABUF) (*18*). This suggests the persistence of local hydrophobic interactions, even in urea denaturant, that restrict backbone motions occurring on a subnanosecond time scale. The largest values of in $J(0.87\omega_H)$ are observed at the chain termini and in the regions that are rich in glycine and alanine.

Effect of Urea on the Denatured State of ApoMb in Acid. For the acid denatured state of apoMb in the absence of urea, a correlation was observed between $J(\omega_N)$ and AABUF (*3*). This correlation suggests that local interactions between hydrophobic side chains, for which AABUF is a measure, are responsible for mobility restrictions on a ps-ns time-scale under these conditions. Because denaturation at pH 2.3 is chiefly ascribed to the loss of attractive and the increase of repulsive electrostatic interactions, the contributions of local hydrophobic interactions to the formation of residual structure and to the restriction of backbone dynamics can explain the correlation of AABUF with residual secondary shift and relaxation data in the pH 2.3 state (*3*). In addition, it appears from $J(0)$ that exchange processes involving helices A and G of the acid-unfolded state are present (*3*), and recent experiments using spin labels suggest that the A and G helices, although positioned at opposite ends of the polypeptide chain, may be in transient contact at pH 2.3 (*43*). Again, these contacts are most likely a result of hydrophobic interactions.

In 8 M urea, on the other hand, hydrophobic interactions are weakened. This does not mean that no hydrophobic interactions are possible in urea. On the contrary, hydrophobic clusters stable in highly concentrated urea have been reported (*27*), inferred from the presence of microsecond−millisecond exchange contributions in the backbone dynamics, and a structure has been calculated on the basis of NOE data for such a cluster in concentrated urea solutions (*44*). For urea-denatured apoMb at pH 2.3, persistent local hydrophobic interactions leading to backbone motional restriction are evident in the $J(0)$, R_2, and $J(0.87\omega_H)$ values. As for the acid-unfolded state, these effects are localized to regions of the sequence with large values of AABUF. However, in the urea-denatured state, the microsecond−millisecond motions are absent, and the $J(0)$ values for the A and G helix regions are comparable to those of the rest of the protein. A similar loss of microsecond−millisecond motions upon addition of urea was observed by Fong et al. (*45*), who found that slow motions in a hydrophobic cluster of Ig 18′ at medium urea concentrations vanished upon increase of urea concentration.

Gly and Ala Molecular Hinges. The correlations shown in Figures 4 and 5 indicate that the most important influence on the motion of the apoMb backbone in 8M urea, pH 2.3, is the size and shape of individual amino acid side chains. The excellent agreement between experimentally observed values and theoretical parameters describing configuration, size, and physicochemical properties of side chains, confirms that, in an unfolded or denatured protein, the primary sequence is the primary determinant for the backbone mobility, as well as for the fundamental local interactions detected by the CSI analysis.

A high proportion of Gly and Ala residues is present in the most flexible regions of apoMb in 8M urea, pH 2.3 (Figure 4). Since the majority of the glycine residues (except G23 and G65) are exposed to solvent in the fully folded myoglobin, a rather large variability of these residues between species might be expected. However, a multiple sequence alignment reveals that glycine residues are very well conserved in myoglobins (though not in hemoglobins or leghemoglobin). Much of this conservation can be attributed to a requirement for Gly in densely packed areas, or for a residue that is capable of forming positive ϕ angles. However, no structural or steric requirements are apparent for A15, A19, A22, G23, G121, G124, and G129; yet these residues with small side chains remain identical or are substituted conservatively with small side chains in more than 70% if myoglobin sequences. This raises the possibility that they play a role in the folding process.

The function of glycine/alanine rich segments during folding could be to act as a flexible "molecular hinge" in the polypeptide chain, allowing other parts of the protein, perhaps containing nascent structure, to reposition relative to each other more efficiently. Such a highly mobile region allows sampling of backbone conformational space to a greater extent than in regions of the sequence with a concentration of bulky side chains and could facilitate formation early in the folding process of local hydrophobic clusters between the regions of the polypeptide that ultimately

form the A and B helices or the G and H helices.

Implications for Protein Folding. Protein folding is generally believed to occur by a progressive collapse of the polypeptide chain, with a concomitant increase in topological order, and formation and stabilization of secondary and tertiary structure. Here we demonstrate that, even in 8 M urea, local hydrophobic interactions persist that result in transient hydrophobic cluster formation and restriction of the polypeptide backbone motions. The observation that hydrophobic interactions persist in local regions of the polypeptide in denaturant supports a model for protein folding in which the initial collapse following removal of the denaturant occurs through transient interactions between hydrophobic clusters that are proximal in the sequence. Direct evidence for interactions, under partially denaturing conditions, between the clusters identified in the present dynamics measurements has recently been obtained by paramagnetic spin labeling experiments (*43*). Our data also show that in the absence of long- and even medium-range interactions, the backbone dynamics of a denatured protein can be explained by intrinsic properties of the individual amino acids, including the effective radius of gyration of the amino acid side chains and the average area buried upon folding. Thus, potential folding initiation sites, which direct the earliest stages of compaction driven by local hydrophobic collapse, can be identified from simple properties of the amino acid sequence.

ACKNOWLEDGMENT

We thank Drs. Mike Lietzow, Chiaki Nishimura, Ronaldo Mohana Borges, David Eliezer, and Silvia Cavagnero for stimulating discussions and Linda Tennant for expert technical help. S.S. thanks Philipp Neudecker for help with computer programming and Professor Paul Rösch for his support.

REFERENCES

1. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
2. Eliezer, D., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 2894−2901.
3. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) *Biochemistry 40*, 3561−3571.
4. Eliezer, D., Yao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
5. Jennings, P. A., Stone, M. J., and Wright, P. E. (1995) *J. Biomol. NMR 6*, 271−276.
6. Van Geet, A. L. (1970) *Anal. Chem. 42*, 679−680.
7. Yao, J., Dyson, H. J., and Wright, P. E. (1997) *FEBS Lett. 419*, 285−289.
8. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., B 103*, 203-216.
9. Löhr, F., and Rüterjans, H. (1995) *J. Biomol. NMR 6*, 189−197.
10. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201−205.
11. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
12. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
13. Delaglio, F., Grzesiek, S., Vuister, G. W., Guang, Z., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
14. Johnson, B. A., and Blevins, R. A. (1994) *J. Chem. Phys. 29*, 1012−1014.
15. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
16. Viles, J. H., Duggan, B. M., Zaborowski, E., Schwarzinger, S., Huntley, J. J., Kroon, G. J., Dyson, H. J., and Wright, P. E. (2001) *J. Biomol. NMR 21*, 1−9.
17. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) *J. Biomol. NMR 6*, 153−162.
18. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) *Science 229*, 834−838.
19. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105−132.
20. Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. (1987) *FEBS Lett. 224*, 149−155.
21. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
22. http://www.bork.EMBL-Heidelberg.DE/Alignment/
23. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) *J. Biomol. NMR 18*, 43−48.
24. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) *J. Am. Chem. Soc. 123*, 2970−2978.
25. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36*, 8977−8991.
26. Zhang, O., and Forman-Kay, J. D. (1995) *Biochemistry 34*, 6784−6794.
27. Meekhof, A. E., and Freund, S. M. V. (1999) *J. Mol. Biol. 286*, 579−592.
28. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) *Protein Sci. 4*, 2605−2615.
29. Geierstanger, B., Jamin, M., Volkman, B. F., and Baldwin, R. L. (1998) *Biochemistry 37*, 4254−4265.
30. Schiffer, C. A., and Dotsch, V. (1996) *Curr. Opin. Biotechnol. 7*, 428−432.
31. Jorgensen, W. L., Duffy, E. M., and Tirado-Rives, J. (1993) *Philos. Trans. R. Soc. London, Ser. A 345*, 87−96.
32. Dötsch, V., Wider, G., Siegal, G., and Wüthrich, K. (1995) *FEBS Lett. 366*, 6−10.
33. Liepinsh, E., and Otting, G. (1994) *J. Am. Chem. Soc. 116*, 9670−9674.
34. Graziano, G. (2001) *J. Phys. Chem. B 105*, 2632−2637.
35. Hu, Y., MacInnis, J. M., Cherayil, B. J., Fleming, G. R., Freed, K. F., and Perico, A. (1990) *J. Chem. Phys. 93*, 822−836.
36. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) *Biochemistry 36*, 2390−2402.
37. Buck, M., Schwalbe, H., and Dobson, C. M. (1996) *J. Mol. Biol. 257*, 669−683.
38. Penkett, C. J., Redfield, C., Jones, J. A., Dodd, I., Hubbard, J., Smith, R. A. G., Smith, L. J., and Dobson, C. M. (1998) *Biochemistry 37*, 17054−17067.
39. Levitt, M. (1976) *J. Mol. Biol. 104*, 59−107.
40. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) *Science 295*, 1719−1722.
41. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
42. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
43. Lietzow, M. A., Jamin, M., Dyson, H. J., and Wright, P. E. (2002) *J. Mol. Biol. 322*, 655−662.
44. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) *Science 257*, 1559−1563.
45. Fong, S., Bycroft, M., Clarke, J., and Freund, S. M. (1998) *J. Mol. Biol. 278*, 417−429.

BI020381O