Chain Length Dependence of Apomyoglobin Folding: Structural Evolution from Misfolded Sheets to Native Helices[†]

Clement C. Chow,[§] Charles Chow,[§] Vinodhkumar Raghunathan,[‡] Theodore J. Huppert,[§] Erin B. Kimball,[§] and Silvia Cavagnero*[,§]

Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, Wisconsin 53706, and Department of Chemistry, University of Washington, Seattle, Washington 98195

Received December 5, 2002; Revised Manuscript Received April 12, 2003

ABSTRACT: Very little is known about how protein structure evolves during the polypeptide chain elongation that accompanies cotranslational protein folding. This in vitro model study is aimed at probing how conformational space evolves for purified N-terminal polypeptides of increasing length. These peptides are derived from the sequence of an all-α-helical single domain protein, Sperm whale apomyoglobin (apoMb). Even at short chain lengths, ordered structure is found. The nature of this structure is strongly chain length dependent. At relatively short lengths, a predominantly non-native *β*-sheet conformation is present, and self-associated amyloid-like species are generated. As chain length increases, α-helix progressively takes over, and it replaces the *β*-strand. The observed trends correlate with the specific fraction of solvent-accessible nonpolar surface area present at different chain lengths. The C-terminal portion of the chain plays an important role by promoting a large and cooperative overall increase in helical content and by consolidating the monomeric association state of the full-length protein. Thus, a native-like energy landscape develops late during apoMb chain elongation. This effect may provide an important driving force for chain expulsion from the ribosome and promote nearly-posttranslational folding of single domain proteins in the cell. Nature has been able to overcome the above intrinsic misfolding trends by modulating the composition of the intracellular environment. An imbalance or improper functioning by the above modulating factors during translation may play a role in misfolding-driven intracellular disorders.

Significant progress has been made in recent years in understanding the physical principles and the mechanisms governing the folding of single domain proteins (*1*−*7*). Since Anfinsen's key experiments (*8*), most of the work in this area has focused on the mechanisms by which sequence encodes structure. Folding is usually envisaged as the convergence of an ensemble of disordered conformations (i.e., the unfolded state) toward a lower energy spatially ordered compact structure. The initial conditions of these experiments involve a chemically or photochemically generated denatured state. The process is then followed as a function of time, once the solution has rapidly been switched to conditions thermodynamically favoring the folded conformation. The physical forces acting on the polypeptide chain on its way to the native conformation are then probed, and experimental information is gained on the energy landscapes of the folded protein. These in vitro experiments assume that the pertinent species to be studied is the full-length polypeptide.

A poorly explored aspect in protein folding is how precisely structure develops as a function of chain length, starting from the N-terminus and proceeding toward the C-terminus. This is a biologically relevant question since, within the ribosomal machinery of the cell, proteins are vectorially synthesized from the N-terminus toward the C-terminus. Most importantly, translation rates in both prokaryotic and eukaryotic systems are far slower than the typical folding rates of single domain proteins at physiologically relevant temperatures (Table 1). The above kinetic argument suggests that there is ample chance for conformational equilibration to take place cotranslationally, before synthesis of the full length chain has been completed.

While it is clear that other complicating factors (e.g., presence of chaperones, molecular crowding) may play a role during cotranslational folding/misfolding events in the intracellular environment, this work takes a model system approach and investigates how polypeptide conformation and energy landscapes evolve with chain elongation under native-like conditions (i.e., in the absence of any denaturing agents). It is important to emphasize upfront that this strategy is not aimed at representing the complex conditions found in the cell's natural milieu (e.g., high viscosity and molecular crowding, ribosome, dynamic equilibrium with cotranslationally active chaperones), but rather, at establishing physical principles and expected trends of the polymer chain encoding a protein sequence as it gets elongated. These

[†] This work was supported by the American Heart Association (Grant 0160492Z), the National Science Foundation (Grant 0215368), by the Research Corporation (Research Innovation Award), and by the Milwaukee Foundation (Shaw Scientist Award).

* To whom correspondence should be addressed. Phone: (608) 262-5430. Fax: (608) 262-5430. E-mail: cavagnero@chem.wisc.edu.

[‡] University of Washington.

[§] University of Wisconsin.

10.1021/bi0273056 CCC: $25.00 © 2003 American Chemical Society
Published on Web 05/20/2003

Table 1: Kinetic Parameters for the Observed Translation Rates of Prokaryotic and Eukaryotic Species Both in Vivo and in Cell-Free Systems

Translation Rates

system	translation rate (amino acids/s)	translation time for a 200 residue protein (s)	*T* (°C)	ref
prokaryotic, in vivo	15−20	10−13	37	*67, 68*
eukaryotic, in vivo	3−4	50−65	37	*67*
prokaryotic, cell-free system	2.5−3.5	55−80	37	*69*
eukaryotic, cell-free system (rabbit reticulocyte)	3−5	40−65	37	*69*
eukaryotic, cell-free system (wheat germ)	1−1.5	130−200	22−27	*69*

In Vitro Protein Folding Rates

protein type	folding rate (s^{-1})	half-life (s)	*T* (°C)	ref
single domain, α helix	$1-10^6$	$7 \times 10^{-7}-1$	37−42	*1*
single domain, *β* sheet	0.1−150	$5 \times 10^{-3}-7$	20−25	*5*
multidomain (tail-spike protein, luciferase)	$3-10 \times 10^{-5}$	$1-3 \times 10^4$	10	*70, 71*

principles can then be exploited to devise tests that more rigorously take the cell's environment into account.

The chain length dependence of homopolymer and block-copolymer model polypeptides has been studied before. Scheraga (*9*, *10*), Blout (*11*), and Loh (*12*) investigated the secondary structure evolution of α-helical, polyproline II helical, and *β*-sheet polypeptides. Hodges and Litowski (*13*), and Toniolo et al. (*14*) studied the chain length dependence of coiled coil and 3−10 helix formation, respectively. A different approach has been adopted by Wright and Dyson (*15*−*18*), and Carey and Tasayco (*19*−*21*), who have analyzed the folding of protein segments corresponding to individual secondary structure modules of globular proteins with the goal of learning about locally versus globally driven secondary structure formation in folding. Fersht and co-workers examined the behavior of polypeptide fragments of increasing length derived from the chymotrypsin inhibitor II (CI-2) (*22*, *23*) and barnase (*24*) sequences.

The target protein of this work is apomyoglobin (apoMb),[1] a well-characterized all α-helical single domain protein (Figure 1a) whose full length chain in vitro folding pathways have been well-studied (*25*−*29*). The A, G, and H helices form first as part of a molten globule intermediate (Figure 2a). The additional secondary and tertiary structure modules are subsequently formed within a cooperative step characterized by single exponential kinetics (*25*). We have examined 36, 77, and 119 amino acid long N-terminal fragments. Lengths were designed to match ends of individual helices in the native state (Figure 1b). All species were compared to the 153 amino acid full length protein. Figure 2 (panels b and c) illustrates two possible limiting models by which a polypeptide may fold as its chain elongates from the N- to the C-terminus. The first model (panel b) postulates that the elongating polypeptide chain is largely dynamic and structurally disordered until most of the polypeptide chain has been synthesized. According to the alternative limiting model (panel c), both secondary and tertiary interactions develop in concert and are extremely close to the corresponding structures found in the full length protein. The present study aims at testing the above models and the existence of any alternative folding/misfolding motifs.

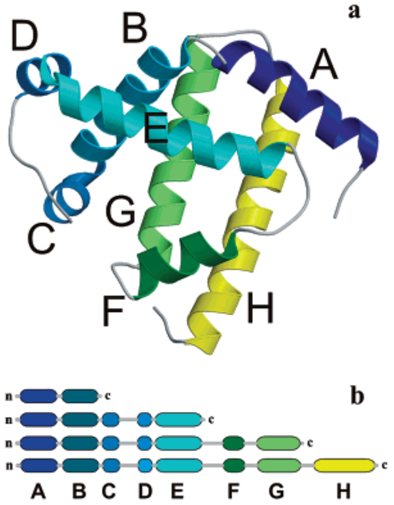

FIGURE 1: (a) Three-dimensional structure of Sperm whale myoglobin (apoMb). The letters denote corresponding helices found in the native structure. Coordinates were derived from the X-ray structure of carbonmonoxymyoglobin (*72*). The image was created using the MOLSCRIPT (*73*) and RASTER3D (*74*) softwares. (b) Schematic representation of the N-terminal apoMb fragments studied in this work. The length and color coding of the different segments matches that of the helices found in myoglobin's native state (see panel a).

MATERIALS AND METHODS

N-Terminal Fragment Preparation. The wild type apoMb gene was obtained from Steve Sligar (University of Illinois at Urbana Champaign) and subsequently subcloned into a pET blue-1 vector (Novagen, Madison, WI). Briefly, the apoMb gene insert was obtained by PCR, blunted by T4 DNA polymerase, and finally subcloned into the pETblue-1 vector at the EcoR V restriction site. The genes for the 1−77 and 1−119 apoMb N-terminal fragments were produced by engineering stop codon point mutations into appropriate regions of the parent plasmid (QuickChange Site-Directed Mutagenesis Kit; Stratagene, La Jolla, CA). The gene for the 1−36 fragment was generated by introducing a Met point mutation in the parent gene at a position corresponding to amino acid 36. After expression and reverse phase HPLC

[1] Abbreviations: apoMb, apomyoglobin; GnHCl, guanidine hydrochloride; CD, circular dichroism; TFA, trifluoroacetic acid; HFIP, hexafluoro2-propanol; far-UV CD, circular dichroism in the far ultraviolet region; FT-IR, Fourier transform infrared.

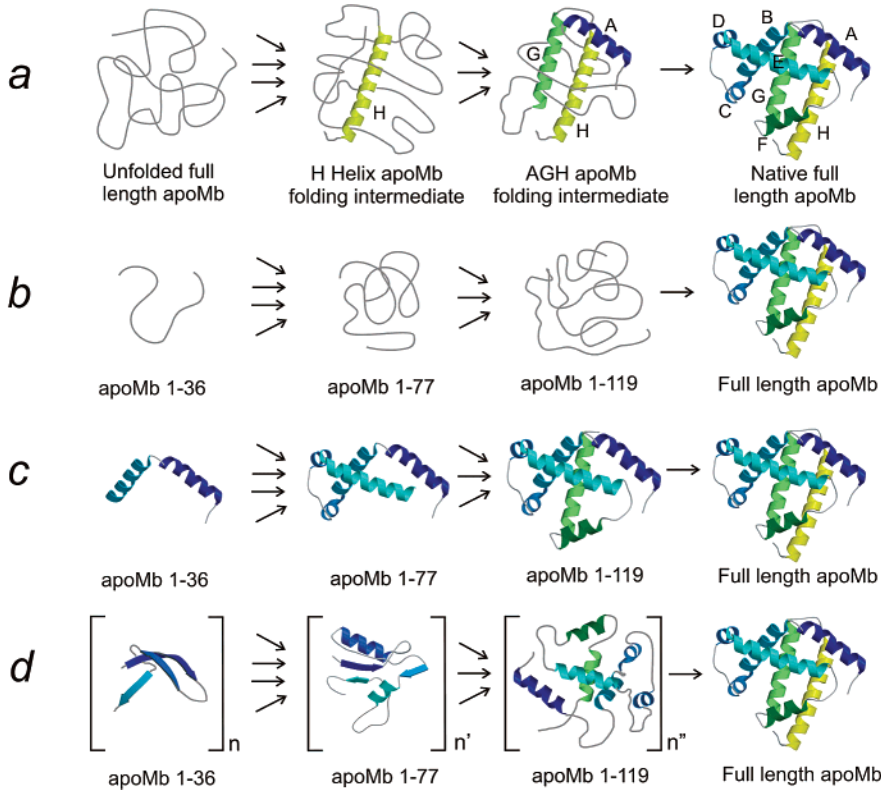

FIGURE 2: Schematic illustration of apoMb folding models based either on experimental evidence or on hypothetical limiting mechanisms. (a) Experimentally determined folding pathways for full-length apomyoglobin starting from a urea unfolded state (*17*, *25*). (b) A hypothetical limiting mechanism supporting no defined secondary and tertiary structures formation until nearly the entire chain has been synthesized. (c) An additional limiting folding mechanism illustrating that structure may evolve upon increasing chain length just as in the full-length protein. (d) Proposed chain length dependence of apoMb folding/misfolding based on the experimental evidence collected in this work.

purification, cyanogen bromide cleavage was performed. Expression was done in LB medium at 37 °C. Protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG, 1 mM, at OD_{600} 0.5). Growth was then continued for at least 5 hrs until saturation. Plasmids were transformed in *Escherichia coli* Tuner (DE3) pLac1 competent cells (Novagen). Cell lysis and purification were performed according to known protocols for wild-type full-length protein (*27*, *30*). Identity of the desired fragments was confirmed by high-resolution time-of-flight electrospray mass spectrometry. All masses were within ±1 atomic mass unit from the theoretical value.

Cyanogen Bromide Cleavage. Each milligram of protein was treated with 0.1 mL of 0.1 M HCl and 2.5 mg of cyanogen bromide. The mixture was stirred at room temperature for about 14 h and then lyophilized. The dried reaction product was dissolved in 0.5 mL of hexafluoro-2-propanol (HFIP) and then brought to 40% v/v HFIP/0.1% trifluoroacetic acid (TFA) water. Reverse phase HPLC purification was then performed. Pure fractions were characterized by electrospray mass spectrometry. Four different species were generated. Briefly, the C-terminal homoserine resulting from the cleavage at the Met site was found in both the lactone and the unesterified open form. Each of these species was found to either contain or lack the N-terminal Met. Only the pure N-terminal Met-containing lactone form was used.

Sample Preparation for Spectroscopic Measurements. Lyophilized powders containing pure peptides/proteins were dissolved in 10 mM sodium acetate buffer (pH 6.0), and the solution pH was then readjusted. Samples were centrifuged for 30 s on a tabletop centrifuge to remove traces of dust and any other solid particles. Supernatants were then equilibrated overnight at 4 °C and used for all the spectroscopic measurements below.

UV-Vis Absorption Measurements and Determination of Peptide Concentrations. Extinction coefficients in 5 M GnHCl at 280 nm were calculated for all species from amino acid sequence (*31*). Known amounts of these samples were then diluted in excess refolding buffer, and absorbances were measured. These values were then utilized to determine extinction coefficients in 10 mM sodium acetate at pH 6.0. Light scattering components that fit a λ^{-4} analytical expression were subtracted out, and the residual absorbances (at 280 nm) were utilized to determine sample concentration. Measurements were done on an HP 8452A diode array spectrophotometer.

Size Exclusion Chromatography. Size exclusion chromatography was performed at 25 °C on a LCC-500 FPLC system from Pharmacia (Piscataway, NJ) equipped with a LKB Uvicord SII detector, operating at 226 nm. A Superdex 75 HR 10/30 column was used. The column was calibrated with the low molecular weight calibration kit from Pharmacia. In addition, Aprotinin (6.5 kDa; USB, OH) was also used as a standard. Samples were dissolved in 10 mM sodium acetate buffer (pH 6.0) and eluted in 10 mM sodium acetate, 100 mM NaCl (pH 6.0). Concentrations of injected samples were 34 μM (1-36 apoMb), 42 μM (1-77 apoMb), 123 μM (1-119 apoMb), and 59 μM (full length apoMb).

Circular Dichroism. CD spectra were recorded at 25.0 °C on an Aviv spectropolarimeter (Model 202SF) in 10 mM sodium acetate buffer (pH 6.0). Either 0.1 or 0.2 cm path length cuvettes were used. Data were collected as single scans with a 1 nm bandwidth, 1 nm step size, and a 100 ms time constant. Data averaging time was 20 s (60 s was used for $<15\ \mu$M samples). Sample concentrations ranged from 9 to 12 μM.

Fluorescence. Trp fluorescence emission data were acquired on a Hitachi F-4500 spectrofluorometer. Samples (5 μM) were prepared in 10 mM sodium acetate buffer at pH 6.0. Emission spectra were recorded between 295 and 450 nm (λ_{ex}: 280 nm). Slit widths for both excitation and emission were 5 nm. Data were recorded at 25 °C. Fluorescence-detected GnHCl titrations were performed on solutions containing different denaturant concentrations. Accurate GnHCl concentrations were determined by refractometry (*32*). Fluorescence emission titration data were reported as total peak area between 295 and 450 mn. Curve fitting of the fluorescence-detected urea titration data was done according to an equation for an equilibrium three-state unfolding model. A complete derivation of this equation is provided as Supporting Information.

Infrared Spectroscopy. Fourier transform infrared (FT-IR) spectra were recorded on a Nicolet 740 FT-IR spectrophotometer equipped with a TGS detector and dry air purging. 256 scans were acquired for each sample with 2 cm^{-1} resolution. A CaF_2 cell with a 50 μm Teflon spacer was used. The presence of traces of trifluoroacetic acid is known to severely interfere with FT-IR data collection in the amide I region (*33*, *34*). Therefore, the samples were dissolved in 10 mM HCl and lyophilized to remove traces of residual TFA from HPLC purification, following published protocols (*35*, *36*). The powders were then dissolved in D_2O and incubated overnight at 4 °C followed by lyophilization. Samples were then dissolved in D_2O again, and pH* (uncorrected electrode reading) was then adjusted to 5.6 prior to data collection. Data were recorded at 25 °C.

Thioflavin T Assay. Samples were prepared by mixing 1.92 mL of peptide solution at a 10 μM concentration with 80 μL of 100 μM Thioflavin T in 10 mM sodium acetate buffer adjusted at the appropriate pH (*37*). Emission spectra were recorded at room temperature on the spectrofluorometer by Molecular Kinetics (equipped with MOS-250 optics and data acquisition interface) over the 450−520 nm wavelength range. Excitation wavelength was set to 450 nm. Excitation and emission slit widths were set to 5 and 20 nm, respectively.

Hydropathy and Nonpolar Surface Area Calculations. Cumulative hydropathies were calculated according to Kyte and Doolittle (*38*). The SURFACE RACER program (*39*) was used for the non polar surface area calculations of Figure 10b. In all cases, each chain length was generated in the extended chain form by the Biopolymer module of Insight II 2000 (Accelrys).

RESULTS

Far-UV Circular Dichroism Spectroscopy. Far-UV circular dichroism (CD) analysis of the N-terminal fragments (Figure 3) reveals that ordered structure is present in all species and random coil population is very low at all chain lengths even for the shortest fragments. The full length protein contains predominantly helical structure. However, as chain length decreases the doublet typical of α-helix is progressively replaced by a singlet centered at 217 nm, indicating a complete change in secondary structure content. The CD data deconvolution program CONTINLL (*40*, *41*) has been used to gain insights into the evolution of secondary structure as a function of chain length (Table 2). This program cannot be relied upon for quantitative data analysis because of the presence of putative β-strand and(or) turn, which are poorly represented in the experimental data sets used by the algorithm. Nonetheless, CONTINLL is a very valuable tool for qualitative data assessment. The program output shows that the fraction of random coil disordered conformation is low at all chain lengths, and as polypeptide length increases, β-strand (or turn) content decreases to be progressively replaced by helical conformation. It is known that CD singlets centered at around 217 nm may be due to either β-turn or β-strand.

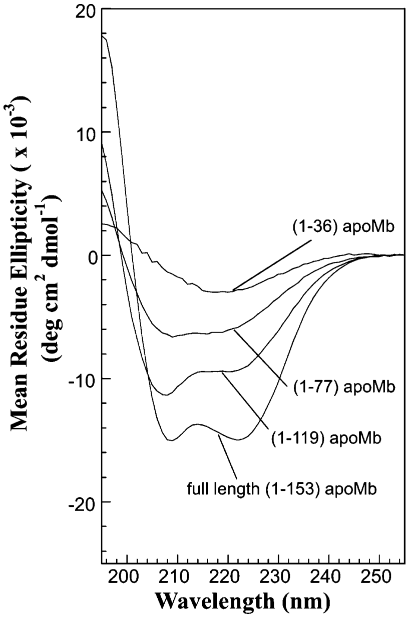

FIGURE 3: Far-UV circular dichroism (CD) spectra of N-terminal apoMb fragments. The full length protein displays a doublet characteristic of α-helices (minima at 208 and 222 nm, respectively). The doublet progressively turns into a singlet centered at 217 nm as chain length decreases.

Table 2: Deconvolution of Far-UV CD Data by the program CONTINLL[a]

	Deconvolution of Far-UV CD Data			
apoMb N-terminal fragment	α	β	β + t	rc
(1−36) apoMb	0.04	0.40	0.62	0.34
(1−77) apoMb	0.16	0.29	0.51	0.33
(1−119) apoMb	0.30	0.18	0.40	0.30
full-length (1−153) apoMb	0.50	0.04	0.28	0.21

[a] The symbols α, β, β + t, and rc denote α-helix, β-strand, β-strand + turn, and random coil, respectively.

As seen in Figure 3 and Table 2, a significant fraction of the helicity found in full-length apomyoglobin is acquired upon transition from the 1−119 fragment to the full length protein. This suggests that the last few amino acids play an important role in consolidating native-like structure and in reshaping the polypeptide energy landscape so that it assumes

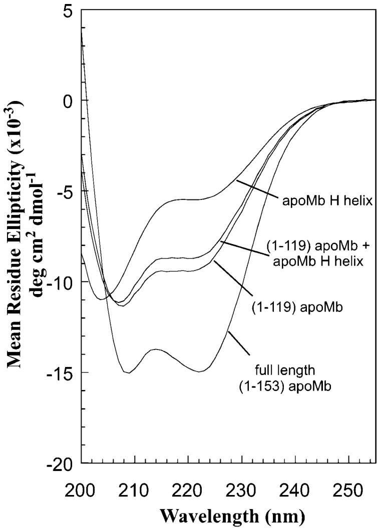

FIGURE 4: Superposition of far-UV CD spectra of full length apoMb (1−153 apoMb), its N-terminal fragment lacking amino acids corresponding to the last C-terminal helix (1−119 apoMb), and the peptide corresponding to the apoMb C-terminal H helix (*17*). The figure also reports the spectrum of an imaginary polypeptide resulting from the summation of the spectra of the C-terminal helix peptide and (1−119) apoMb. The mean residue ellipticities for this hypothetical species have been recalculated to take the increased chain length into account.

native-like features. To further investigate this evidence, we have replotted the far-UV CD data of the 1−119 fragment and full-length apoMb in Figure 4. This figure contains a comparison between the CD spectra of three species: full-length apoMb, 1−119 apoMb, and the calculated spectrum resulting from addition of the far-UV CD data for 1−119 apoMb and H helix peptide (*17*). The H helix peptide is about 25% intrinsically helical, and it comprises only the residues belonging to the apoMb H helix. Chain lengths have been readjusted to properly calculate molar ellipticity values on a per residue basis. The [1−119apoMb-H helix] CD spectrum is representative of an imaginary polypeptide containing most of the apoMb amino acids but whose nature is such that the H helix is not allowed to interact with the remaining C-terminal portion of the protein. Comparison of this spectrum and that of the full length protein shows that nearly half of the overall helicity of full length apoMb is gained as a result of the interaction between the last portion of the chain and the preexisting fragment. This indicates that the C-terminal portion of the amino acid chain plays an important role in promoting cooperativity in secondary structure formation.

Fourier Transform Infrared Spectroscopy. To discriminate whether the nonfull length polypeptides contain either turn or sheet-like secondary structure (see CD data above), we have performed additional studies by Fourier transform infrared (FT-IR) spectroscopy in the amide I region. The data are presented in Figure 5. As chain length decreases, FT-IR reveals, in full agreement with the CD data, that the helical conformation of the full length species gets converted into a different kind of secondary structure. The predominantly β-strand nature of the secondary structure found at short chain lengths is established by wavelength maxima of the observed spectral features, particularly the peaks centered at about 1621 and 1683 cm^{-1}. The second derivative data provided as Supporting Information confirm these results.

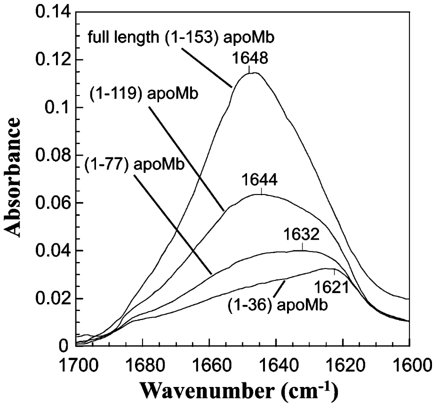

FIGURE 5: Fourier transform infrared (FTIR) absorbance spectra of N-terminal apoMb fragments. The data show a progressive maximum intensity red shift (from 1648 to 1621 cm^{-1}) as chain length decreases. Sample concentrations range from 0.8 to 7.4 mg mL^{-1}. Data have been concentration-corrected to reflect individual absorption intensities at an identical molar concentration. Solution buffer is 10 mM sodium acetate at pH 6.0.

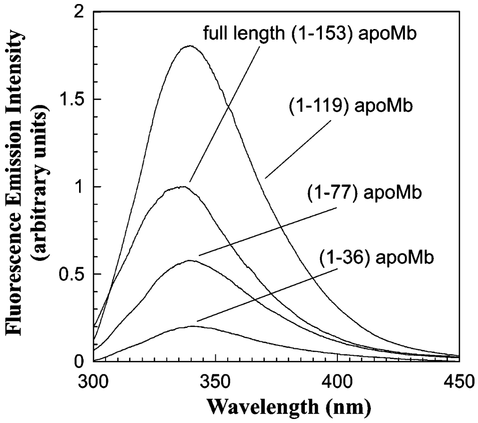

FIGURE 6: Trp fluorescence emission spectra of N-terminal apoMb fragments. (λ_{ex} = 280 nm).

The peak centered at 1621 cm^{-1} is also strongly suggestive of extended β-sheet content (*42*). In summary, the data show that secondary structure progressively evolves from β-sheet to α-helix as chain length increases. The β-sheet secondary structure acquired by the polypeptide at shorter chain lengths is to be regarded as non-native and misfolded since no β-strand is present in the native full length protein. As further discussed in the next section, these results are in interesting contrast with both the full length apoMb folding pathways and the chain length-dependent secondary structure acquired by barnase and chymotrypsin inhibitor-2 (CI2).

Fluorescence Emission Spectroscopy. The Tryptophan fluorescence emission spectra of the N-terminal fragments and full length apoMb are shown in Figure 6. All samples contain the same number of fluorophores (i.e., Trp 7 and Trp 14). These are the only two Tryptophans present in wild-type apoMb. There are large variations in emission intensity for the different species, with a notable hyperfluorescence exhibited by the 1−119 fragment. In all cases, experiments have been performed in duplicate, and concentrations have been carefully measured. The apoMb fluorescence is known to result from the delicate balance between the degree of solvent exposure of both fluorophores and the proximity of Trp 7 to Lys 79. This positively charged amino acid exerts some intramolecular quenching. The fragments and full

Table 3: Tryptophan Fluorescence Wavelength Emission Maxima

apoMb N-terminal fragment	Trp fluorescence, maximum emission wavelength (λ_{ex} = 280 nm) (nm)
unfolded full-length apoMb (in 3 M GnHCl)	349.8
(1−36) apoMb	340.2
(1−77) apoMb	339.0
(1−119) apoMb	339.6
full length (1−153) apoMb	337.4

length apoMb are quite different structurally. It is therefore expected that both the above effects will be present to an unpredictable degree and contribute to modulate the overall fluorescence intensities. Therefore, it is not straightforward to derive specific structural information from the fluorescence intensity differences of the various species. Relative wavelength shifts are much more informative in this case since fluorescence emission wavelengths are known to be dominated by the dielectric constant of the solvent (assuming constant refractive index and nearly constant ground and excited state dipole moments) and the radius of the fluorophore cavity. The Lippert equation quantitatively relates these parameters (*43*). As a result of the above, fluorescence emission wavelength maxima are frequently used to estimate the polarity of the environment surrounding the fluorophore. Therefore, these values are the best spectroscopic probes we currently have in hand to estimate the overall degree of compaction at different chain lengths. Table 3 illustrates the emission wavelengths corresponding to the maximum intensity for the different species as a function of residue number. One initial thing to notice is the relatively large wavelength differences between native and GnHCl-unfolded full length apoMb. The observed red shift by the unfolded state correlates with a solvent-exposed relatively unstructured conformational ensemble. The native state is blue-shifted because of its compact conformation characterized by a smaller radius of gyration relative to the unfolded state. The fluorescence emission wavelength data of Table 3 indicate that the N-terminal fragments do not behave as extended chains and have a degree of compaction similar to that of the full-length protein even at short chain lengths.

Size Exclusion Chromatography and CD Concentration Dependence. A crucial test pertaining to the physical characterization of the fragments is the measurement of their association state in solution. We have done gel filtration analysis of the different species (Figure 7). The 1−119 fragment is mostly present as a dimer/trimer. The 1−77 fragment, on the other hand, appears as a higher order aggregate, and the 1−36 species shows up as weakly self-associated. The peak for the 1−36 species overlaps with residual peaks because of slight buffer/solvent differences between running buffer and injected sample. More insights are provided by additional gel filtration experiments where the samples have been analyzed both in the presence and in the absence of the column and its associated precolumn filter (1 μM pore size). The peak areas have then been measured and compared to test for sample losses within the column setup. The shortest 1−36 and 1−77 fragments display very significant losses (at least 50-fold reduction of overall signal in the presence of the column), while the 1−119 and full length apoMb display negligible losses. Sample losses because of large supramolecular sizes, and consequent inability to run through gel filtration columns, is a well-known feature of some heavily aggregated species in the amyloid field (*44*). The concentration dependence of the far-UV CD signal of the fragments is provided as Supporting Information. The flat profile of the plots for the 1−77 (up to about 50 μM) and 1−119 (up to 100 μM) species is consistent with no variations in the association state over that concentration range. The 1−36 fragment is heavily associated even at low concentrations, and it does not maintain a constant association state over the 1−50 μM concentration range. Figure 8 displays equilibrium GnHCl unfolding titrations aimed at testing the stability and the degree of cooperativity in the unfolding transitions of the fragments. It is clear that most of the fragments display negligible cooperativity upon unfolding. Additionally, they do not appear to be significantly more stable than the full length protein. The 1−36 fragment forms solid particulates over time, especially at pH 6 and higher.

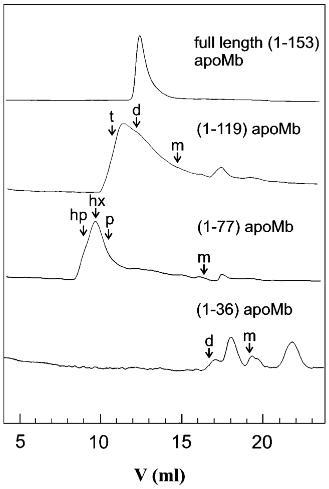

FIGURE 7: Size exclusion chromatography analysis of apoMb N-terminal fragments. The symbols m, d, t, p, hx, and hp denote predicted elution volumes for monomer, dimer, trimer, pentamer, hexamer, and heptamer, respectively, based on column calibration. Samples were dissolved in 10 mM sodium acetate buffer (pH 6.0). The pH was then readjusted to 6.0 (except for the 1−36 sample, which was readjusted to pH 5.8 to avoid the rapid formation of insoluble aggregates that takes place at pH 6.0) followed by injection onto an FPLC Superdex 75 column.

Thioflavin T Fluorescence Emission Assay. The fluorescence emission by thioflavin T has been measured in the presence of the different fragments as a function of pH. The results, shown in Figure 9, are consistent with a remarkable enhancement of the fluorescence signal as pH increases. Even at pH 7, the observed enhancement is at least 30-fold relative to the control sample, indicating the presence of amyloid-like species even under very mild conditions. The 1−36 fragment has the highest fluorescence at pH 6, consistent with its higher tendency to form macroscopic aggregates under these conditions, relative to the other species. At higher pH, the effect is enhanced, possibly because of reduced electrostatic repulsion as the pH gets closer to the isoelectric point for each individual fragment. We have performed electron microscopy experiments (negative staining, data not shown), which indicate that the 1−36 and 1−77 fragments have protofibrillar morphology. Further studies by static and dynamic light scattering are currently underway. These data

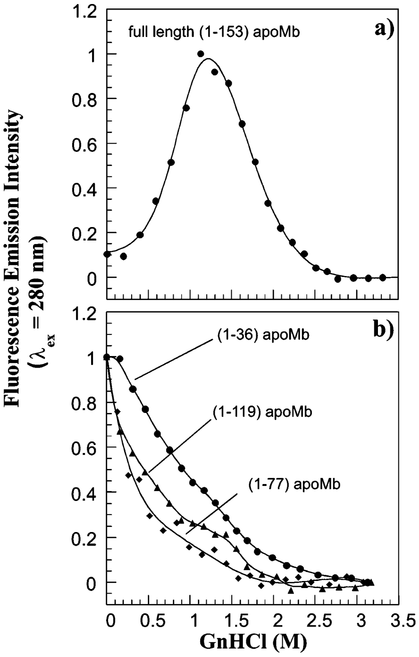

FIGURE 8: (a) Guanidinium hydrochloride (GnHCl) titration of full-length apoMb monitored by Trp fluorescence (λ_{ex} = 280 nm). The emission maximum intensity has been normalized to a value of 1. Curve fitting has been performed according to an equation derived for a three-state equilibrium unfolding model derived by an analytical treatment similar to that used for two-state unfolding (*75*). (b) GnHCl titration of apomyoglobin N-terminal fragments monitored as in panel a. The fluorescence emission intensity of the fragments at 0 M GnHCl has been normalized to 1 for all species. Sample concentrations are as in Figure 6.

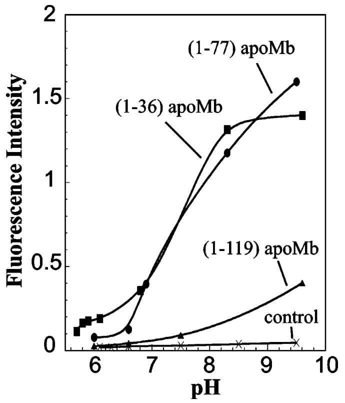

FIGURE 9: Thioflavin T assay of N-terminal fragments as a function of pH. The control sample consists of a solution equivalent to the sample solution except that no polypeptide was added.

show that amyloid-like formations are generated from nonfull length polypeptides. As the polypeptide energy landscapes becomes more native-like concomitantly with chain elongation, these misfolded and potentially dangerous amyloid-like species are no longer populated. As further discussed later, this highlights the fact that, even at room temperature and neutral pH, potentially amyloidogenic species belonging to the amino acid sequence of a nonpathogenic protein can be generated.

Solvent-Accessible Nonpolar Surface Area Calculations. To refine our data interpretation, we have performed hydropathy and nonpolar surface area calculations (Figure 10). Panel a shows the cumulative hydropathy scores according to Kyte and Doolittle. The peptide lengths chosen for this work coincide with the local maxima on the plot. These maxima reflect the fact that in some regions of the chain hydrophobic stretches are immediately followed by hydrophilic segments. The B, E, and G helices of full length apoMb are clearly quite hydrophobic. Panel b represents calculated ratios of nonpolar solvent-accessible surface areas to total solvent exposed surface areas at each polypeptide chain length. Fully extended chains have been assumed. This plot is therefore most useful to evaluate trends to shield solvent-exposed hydrophobic surface. The apoMb chain initially starts as very hydrophobic, progressively become more hydrophilic until about half of its full length, and then finally levels off to an approximately constant average hydropathy. Identical trends are also seen for average chain hydropathy (data available as Supporting Information). There is a correlation between cumulative chain hydropathy, fraction of solvent-exposed surface area, and tendency to misfold. As seen above, 1−36 apoMb heavily self-associates and misfolds, 1−77 apoMb also misfolds, and the 1−119 fragment is about 50% dimer and 50% higher order aggregate. The general observed trend can be summarized by stating that apoMb polypeptides self-associate and give rise to misfolded structured conformations proportionately with the value of their cumulative hydropathies. All species with 65% or more solvent accessible nonpolar surface area whose cumulative hydropathies are less negative than about −35 tend to self-associate at around room temperature.

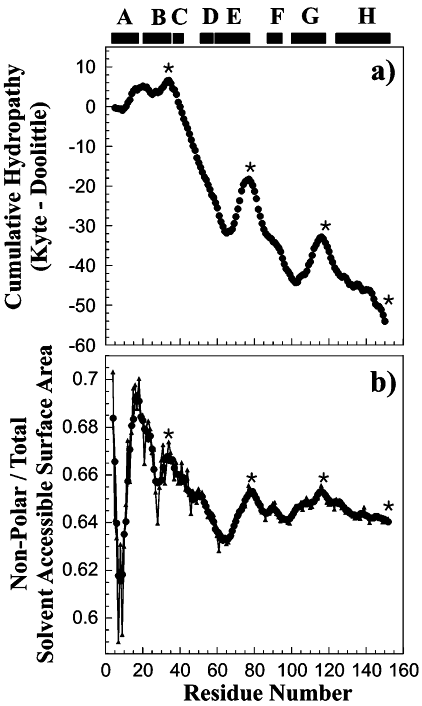

FIGURE 10: (a) Cumulative hydropathy scores for apoMb as a function of chain length according to Kyte and Doolittle (*38*). Data represent average scores over a five residue sliding window. (b) Ratio of nonpolar solvent accessible surface area to total solvent accessible surface area as a function of chain length. The (●) data points result from averaging over five contiguous amino acids. The (▲) data points derive from the output for each single residue. Asterisks denote the chain length of the N-terminal fragments examined in this work.

Similar plots for barnase and CI2, two extremely hydrophilic proteins (data not shown), are consistent with the fact that nonfull length CI2 and barnase remain unstructured and do not give rise to self-association (*23*, *45*).

DISCUSSION

This study reports on the structural trends experienced by purified N-terminal apoMb polypeptides of increasing length in buffered solution (pH 6.0) at room temperature. Motivation for the work was provided by the desire to address fundamental conformational issues relevant to the folding of cotranslationally elongating polypeptides (derived from the apoMb sequence) in the intracellular environment. As an initial step in this direction, we have employed an in vitro model system, which by design, is devoid of cell-related complicating factors (e.g., interaction with chaperones or the ribosome, molecular crowding, geometrical constraints). While this model system bears little similarity to the actual translational machinery, it allows gaining insights about the fundamental conformational trends that an elongating polypeptide chain experiences in the absence of the supporting cellular machinery. It is a first step toward understanding which issues may be important (or not important) in addressing cotranslational protein folding/misfolding. This work also attempts to provide a more biologically relevant parallel to the in vitro experimental folding studies on full-length proteins initiated by Anfinsen (*8*) and later flourished into a rich and mature research field (*1*, *4*, *46*). The kinetic issues that justify a study under equilibrium conditions are illustrated in Table 1 and have already been discussed in the introductory paragraphs.

Our data show that the N-terminal apoMb peptides are rich in β-sheet at short chain lengths, and they become progressively enriched in α-helical content as chain elongates. These folding/misfolding motifs are significantly different from those observed upon full length apoMb refolding from a urea unfolded state (Figure 2a), where only α-helical conformation is observed at all stages of folding (*25*, *26*). A working model illustrating some key findings on the chain length dependence of apoMb folding is shown in Figure 2d. The proposed mechanism substantially departs from the two limiting chain length dependent folding schemes of Figure 2b,c. Non-native extended β-strands are present at shorter chain lengths. This secondary structure phases out as chain elongates while, at the same time, a corresponding increase in α-helical secondary structure takes place.

The observed trend toward misfolded β-sheet formation is proportional to the fraction of overall nonpolar character by the different polypeptide chains (Figure 10). Quite interestingly, N-terminal shorter portions of apoMb have a considerably higher fraction of nonpolar content than longer portions of the chain comprising more residues closer to the C-terminus. It has been known for a long time that, in general, β-sheets are somewhat more effective than α-helices at burying nonpolar surface in aqueous environments (*47*). It is also well-known that the recruiting of nonpolar amino acids away from interaction with solvent is a dominant force in protein folding and stability (*48*, *49*). We argue here that the same is also likely to be true for misfolded structures. Therefore, the predominance of a non-native β-sheet at shorter N-terminal chain lengths may result from the higher efficiency in the burial of nonpolar surface by the β-sheet secondary structure. As the polypeptide gets longer and the overall fraction of nonpolar surface decreases, the peptide chain is able to sample an increased number of degrees of freedom. Under these conditions, the ability to form a monomeric nonpolar core able to stabilize intrinsic local helix propensities (*15*−*18*) takes over.

Figures 3 and 5 and Table 2 show that the progression toward appearance of α-helical structure is gradual, and it is concurrent with the disappearance of self-associated β-sheet structure. This result is in interesting contrast a related work by Fersht and Neira on barnase (*45*) and Chymotrypsin Inhibitor 2 (*23*), two mixed α/β proteins. These proteins do not misfold at any chain length and fold only when the very last few residues are added. Therefore, the behavior of apomyoglobin is substantially different. We also detect that a large fraction (c.a. 50%) of folded helical conformation is cooperatively gained upon addition of the amino acids corresponding to the very last C-terminal apoMb helix (Figure 4). These final C-terminal amino acids are particularly important for native structure formation. In addition to apoMb, barnase, and CI2, the significance of the last few C-terminal amino acids in promoting native-like fold has also been observed for ribonuclease A (*50*). In this case, correct folding enables formation of native disulfide bridges. Therefore, achieving a nearly native-like landscape late upon chain elongation may be a general concept common to several single domain proteins. This suggests that cotranslational protein folding leading to nativelike structure may only significantly take place toward the end of translation, for small single domain water soluble proteins.

Self-association clearly plays an important role, particularly at shorter chain lengths (Figure 7, and Figure 2 of Supporting Information). Since extended β-sheets are most effective at burying nonpolar surface (*47*), the presence of aggregation does not come as a surprise per se, and it is plausible to postulate that aggregation may result from the inability to effectively pack the protein's hydrophobic core at the intramolecular level. Our data support the hypothesis and explicitly show that the need to bury hydrophobic groups by self-association is less stringent at relatively long chain lengths. The 1−119 fragment is 50% populated as a dimer at equilibrium. At this chain length, the combination of intramolecular partial folding and dimerization is probably sufficient to bury the nonpolar area made accessible by the lack of the amino acids corresponding to the C-terminal H helix. The data we present in the manuscript are consistent with the above hypothesis. The implications of the above for cotranslational protein folding are significant. The presence of a tendency to self-associate at shorter chain lengths highlights the likely need by nature to invoke the action of cotranslationally active chaperones during the early stages of translation. Further studies in the presence of chaperones (*51*) will be able to shed more light on this issue.

The highly nonpolar short apoMb fragments give rise to amyloid-like species, as seen by a largely positive thioflavin T assay (Figure 9). These amyloid-like aggregates form from N-terminal fragments of apoMb, an inherently nonpathogenic protein, under extremely mild conditions (i.e., at room temperature and physiologically relevant pH). This provides a valuable complement and contrast to the study by Dobson and co-workers (*52*), which shows that full length myoglobin forms amyloid fibrils under rather nonphysiological conditions (pH 9, borate buffer, 65 °C). In our case, only the short N-terminal fragments give rise to an extended β-sheet. On the other hand, once chain length elongates, native α-helix progressively takes over, under mild solution conditions. One common take-home lesson from our and Dobson's work is that apomyoglobin, and possibly other proteins, form amyloid-like aggregates when conditions are such that their energy landscape is not supporting a monomeric folded state (*52*−*54*). Dobson's work highlights the fact that these aberrant landscapes are generated under particular pH/temperature combinations. In our case, we show that this is the case for apoMb even under very mild conditions. An essential requirement for this to happen is that chain elongation (from N- to C-terminus) has not been completed yet.

This study does not take the presence of molecular crowding and other complex cell environment conditions into account. It is therefore premature to conclude that any apoMb cotranslational self-association may take place in vivo. On the other hand, it is interesting to note that excluded volume theory and experimental studies suggest that molecular crowding enhances self-association (*55*−*58*). Additionally, we have performed an approximate calculation about the expected spatial proximity of nascent polypeptide chains within the polysome (assuming ribosomes to be in close contact). By taking ribosome size/geometry (*59*, *60*) and proximity of polypeptide chains within polysomes into account, we estimate that aggregation should not be possible until at least 48 amino acids protrude out of the ribosomal channel. As suggested by published electron micrographs mapping translating ribosomes (*61*, *62*), intermolecular chain contacts are also possible among nascent chains belonging to distant regions of the same mRNA or to different mRNAs. Evidence of cotranslational self-association has been reported in the literature for two different proteins (*63*, *64*). Additionally, apoMb is made as a heme-containing holoprotein in eukaryotic cells. Spirin reported that nascent chains of β-globin, a closely related protein, are able to cotranslationally incorporate heme when the nascent chain is 80 amino acid or longer (*65*). Finally, cotranslationally active chaperones are known to play an important role for some proteins during translation in vivo by binding to nascent chains and preventing hydrophobic sites from becoming solvent-exposed (*66*). All of the above factors may play a significant role in the intracellular context.

Perhaps the most important finding of this work is to highlight the in vitro chain length dependent progressive conformational trends for apoMb. The results are strongly suggestive of what may happen to nascent polypeptide chains in the absence of cotranslationally active chaperones and the rest of the translation-supporting cell machinery. Therefore, this study underscores the importance of this machinery in healthy cells and the potentially damaging effects that may arise during translation of proteins with nonpolar N-terminal portions, in case of pathogenic physiological imbalances. Additional work is necessary to probe the generality of these findings and compare them to the secondary and tertiary structure elements sampled cotranslationally by nascent polypeptides in the cell.

ACKNOWLEDGMENT

We thank Lisa Jungbauer for technical assistance, Rebecca Turcotte for help in preparing Table 1, and Andreas Matouscheck, Sam Gellman, Ruth Saecker, and Cavagnero group members for helpful discussions. We are also grateful to Oleg Tsodikov for assistance in using the Surface Racer program.

SUPPORTING INFORMATION AVAILABLE

Derivation of curve fitting equations, CD concentration dependence data, FT-IR second derivative spectra, and mean residue hydropathy data. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Myers, J. K., and Oas, T. G. (2002) *Annu. Rev. Biochem. 71*, 783−815.
2. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 327−359.
3. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem. 37*, 868−893.
4. Brockwell, D. J., Smith, D. A., and Radford, S. E. (2000) *Curr. Opin. Struct. Biol. 10*, 16−25.
5. Capaldi, A. P., and Radford, S. E. (1998) *Curr. Opin. Struct. Biol. 8*, 86−92.
6. Socci, N. D., Onuchic, J. N., and Wolynes, P. G. (1998) *Proteins 32*, 136−158.
7. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) *J. Mol. Biol. 277*, 985−994.
8. Anfinsen, C. B. (1973) *Science 181*, 223−227.
9. Ingwall, R. T., Scheraga, H. A., Lotan, N., Berger, A., and Katchalski, E. (1968) *Biopolymers 6*, 331−368.
10. Poland, D., and Scheraga, H. A. (1970) p 30, Academic Press, New York.
11. Doyle, B. B., Traub, W., Lorenzi, G. P., and Blout, E. R. (1971) *Biochemistry 10*, 3052−3060.
12. Giesler, M., Thorgerson, M., Masterson, L., and Loh, A. P. (2001) *Biophys. J. 80*, 1678 (Pt. 2).
13. Litowski, J. R., and Hodges, R. S. (2001) *J. Pept. Res. 58*, 477−492.
14. Yasui, S. C., Keiderling, T. A., Formaggio, F., Bonora, G. M., and Toniolo, C. (1986) *J. Am. Chem. Soc. 108*, 4988−4993.
15. Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J., and Wright, P. E. (1993) *Biochemistry 32*, 6356−6364.
16. Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., and Dyson, H. J. (1993) *Biochemistry 32*, 6348−6355.
17. Waltho, J. P., Feher, V. A., Merutka, G., and Dyson, H. J. (1993) *Biochemistry 32*, 6337−6347.
18. Reymond, M. T., Merutka, G., Dyson, H. J., and Wright, P. E. (1997) *Protein Sci. 6*, 706−716.
19. Kang, X. S., and Carey, J. (1999) *J. Mol. Biol. 285*, 463−468.
20. Tasayco, M. L., and Carey, J. (1992) *Science 255*, 594−597.
21. Wu, L. C., Grandori, L., and Carey, J. (1994) *Protein Sci. 3*, 369−371.
22. Neira, J. L., and Fersht, A. R. (1996) *Fold. Des. 1*, 231−241.
23. De Prat Gay, G., Ruiz-Sanz, J., Neira, J. L., Corrales, F. J., Otzen, D. E., Ladurner, A. G., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 968−979.
24. Neira, J. L., and Fersht, A. R. (1999) *J. Mol. Biol. 287*, 421−432.
25. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
26. Jennings, P. A., Dyson, H. J., and Wright, P. E. (1994) in *Protein Structure and Protein Substrate Interactions* (Doniach, S., Ed.) pp 7−18, Plenum Press, New York.
27. Cavagnero, S., Dyson, H. J., and Wright, P. E. (1999) *J. Mol. Biol. 285*, 269−282.
28. Cavagnero, S., Schwartzinger, S., Dyson, H. J., and Wright, P. E. (2001) *Biochemistry 40*, 14459−14467.
29. Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry 39*, 11227−11237.
30. Eliezer, D., and Wright, P. E. (1996) *J. Mol. Biol. 263*, 531−538.
31. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.

32. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1990) in *Protein Structure—A Practical Approach* (Creighton, T. E., Ed.) pp 311–330, IRL Press, Oxford.
33. Surewicz, W. K., Mantsch, H. H., and Chapman, D. (1993) *Biochemistry 32*, 389–394.
34. Surewicz, W. K., Mantsch, H. H., and Chapman, D. (1989) *J. Mol. Struct. 214*, 143–147.
35. Dieudonne, D., Mendelsohn, R., Farid, R. S., and Flach, C. R. (2001) *Biochim. Biophys. Acta 1511*, 99–112.
36. Simonetti, M., and Bello, C. D. (2001) *Biopolymers 62*, 95–108.
37. Naiki, H., Higuchi, K., and Hosokawa, M. (1989) *Anal. Biochem. 177*, 244–249.
38. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
39. Tsodikov, O. V., Record, M. T., and Sergeev, Y. V. (2002) *J. Comput. Chem. 23*, 600–609.
40. Sreerama, N., and Woody, R. W. (1993) *Anal. Biochem. 209*, 32–44.
41. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem. 282*, 252–260.
42. Singh, B. R. (2000) *Infrared Analysis of Peptides and Proteins*, Oxford University Press, New York.
43. Lakowicz, J. R. (1999) pp 186–189, Kluwer Academic/Plenum Publishers, New York.
44. Pallitto, M., and Murphy, R. M. (2001) *Biophys. J. 81*, 1805–1822.
45. Neira, J. L., and Fersht, A. R. (1999) *J. Mol. Biol. 285*, 1309–1333.
46. Baldwin, R. L. (1999) *Nat. Struct. Biol. 6*, 814–817.
47. Chothia, C. (1976) *J. Mol. Biol. 105*, 1–14.
48. Kauzmann, W. (1959) *Adv. Prot. Chem. 14*, 1–63.
49. Dill, K. A. (1985) *Biochemistry 24*, 1501–1509.
50. Taniuchi, H. (1970) *J. Biol. Chem. 245*, 5459–5468.
51. Frydman, J., and Hartl, U. F. (1996) *Science 272*, 1497–1502.
52. Fandrich, M., Fletcher, M. A., and Dobson, C. M. (2001) *Nature 410*, 165–166.
53. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. M., Dobson, C. M., Radford, S. E., Blake, C. C. F., and Pepys, M. B. (1997) *Nature 385*, 787–793.
54. Dobson, C. M. (2002) *Nature 418*, 729–730.
55. Zimmerman, S. B., and Minton, A. P. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 27–75.
56. Minton, A. P. (1981) *Biophys. J. 78*, 101–109.
57. Minton, A. P. (2000) *Curr. Opin. Struct. Biol. 10*, 34–39.
58. Van Den Berg, B., Wain, R., Dobson, C. M., and Ellis, R. J. (2000) *EMBO J. 19*, 3870–3875.
59. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905–920.
60. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) *Science 285*, 2095–2104.
61. Martin, K. A., and Miller, O. L. (1983) *Dev. Biol. 98*, 338–348.
62. Yoshida, T., Wakiyama, M., Yazaki, K., and Miura, K. (1997) *J. Electron Microsc. 46*, 503–506.
63. Nicholls, C. D., McLure, K. G., Shields, M. A., and Lee, P. W. K. (2002) *J. Biol. Chem. 277*, 12937–12945.
64. Gilmore, R., Coffey, M. C., Leone, G., McLure, K., and Lee, P. W. (1996) *EMBO J. 15*, 2651–2658.
65. Komar, A. A., Kommer, A., Krasheninnikov, I. A., and Spirin, A. S. (1997) *J. Biol. Chem. 272*, 10646–10651.
66. Frydman, J. (2001) *Annu. Rev. Biochem. 70*, 603–647.
67. Netzer, W. J., and Hartl, F. U. (1997) *Nature 388*, 343–349.
68. Netzer, W. J., and Hartl, F. U. (1998) *Trends Biochem. Sci. 23*, 68–73.
69. Fedorov, A. N., and Baldwin, T. O. (1998) *Methods Enzymol. 290*, 1–17.
70. Clark, P. L., and King, J. (2001) *J. Biol. Chem. 276*, 25411–25420.
71. Herbst, R., Schafer, U., and Seckler, R. (1997) *J. Biol. Chem. 272*, 7099–7105.
72. Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133–154.
73. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
74. Merrit, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505–524.
75. Pace, C. N. (1990) *TibTech 8*, 93–98.

BI0273056