J|A|C|S
ARTICLES
Published on Web 12/03/2003

On the Extended β-Conformation Propensity of Polypeptides at High Temperature

Wei Yuan Yang,[§] Edgar Larios,[‡] and Martin Gruebele*[,†,‡,§]

Contribution from the Departments of Chemistry and Physics, and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Illinois 61801

Received May 7, 2003; E-mail: gruebele@scs.uiuc.edu

Abstract: At room temperature, natural polypeptides exposed to high concentrations of a strong denaturant nearly attain the circular dichroism spectra characteristic of random coils. As temperature is increased, the spectra begin to show the signature of a substantial fraction of extended chain, the structure common in β-sheets. This structural propensity at high temperature is not altered by concentration changes over a greater than 1000-fold range, so it is not caused by aggregation. Four proteins with different folds and varying amounts of α-helical and β-sheet secondary structure, in the presence or absence of denaturant, all were subject to extended chain formation upon heating. This effect arises naturally from the steric constraints associated with polypeptides and is probably counteracted, not enhanced, by hydrophobic interactions. Molecular dynamics simulations in the 298−1000 K range reveal an attractive potential of mean force in the extended chain region of the Ramachandran diagram, which broadens as the temperature is raised. We also demonstrate a direct correlation between extended structure content and the rate of aggregation kinetics. Thus pre-existing extended structure could funnel proteins into aggregates.

Introduction

Since the initial proposal by Pauling and Corey,[1] the extended structure that makes up the bulk of β-sheets has turned up in many proteins. Yet the materials richest in β-sheets are protein aggregates such as the amyloid fibers associated with many protein misfolding diseases.[2,3] β-sheet fibers can be made from proteins in many fold classes.[4] Heat-induced aggregation frequently produces structures similar to natural amyloids and has led to the development of many convenient model systems for misfolding and aggregation.[4−6]

Heat unfolding of proteins is easily rationalized because increased entropy of the protein backbone and side chains should be favored at high temperature. More perplexing is the observation that heat unfolding often is accompanied by aggregation. Different proteins have varying tendencies to aggregate, but an increased aggregation probability at high temperature is a generic property of proteins. This problem has limited the use of temperature tuning for folding studies of many proteins and requires experiments to be operated at low concentrations to avoid aggregation problems.

There are several reasons why heat aggregation may seem counterintuitive. Protein association greatly reduces the entropy compared to free chains in solution. This reduction in entropy disfavors aggregation as the temperature is increased. The dominant interaction that induces aggregation cannot be the same as the main driving force that folds proteins (hydrophobicity), or it would prevent thermal unfolding in the first place. Finally, if the underlying principles in folding and aggregation are the same, why is there such a large difference in inherent β-sheet propensity for aggregates compared to folded proteins? Proteins rich in α-helical structure are far more common than analogous aggregates.

In this paper, we examine the denaturation of four proteins with very different folds and secondary structure content: the lambda repressor N-terminal fragment λ_{6-85},[7] an all-α protein with 80 residues; N-terminal domain of U1A,[8] an $\alpha+\beta$ sandwich with 102 residues; ubiquitin (76 residues), containing mainly antiparallel β-sheets with some helices; and phosphoglycerate kinase, with 415 residues in two $\beta-\alpha-\beta$ domains with mainly parallel β-sheets. Circular dichroism spectra reveal a strong propensity for formation of an extended chain structure, typical of β-sheets, as the temperature is increased. This propensity is independent of concentration, as shown by tethering together two monomeric proteins to increase the bimolecular collision rate. By applying singular value decomposition to circular dichroism spectra, we show that the same trend persists in the absence of denaturant. Two different incubation experiments show that an extended structure can be a precursor for the formation of aggregates and that the amount of extended structure is directly correlated with the aggregation rate. Potentials of mean force from full-atom molecular dynamics

[†] Department of Chemistry.
[‡] Department of Physics.
[§] Center for Biophysics and Computational Biology.

(1) Pauling, L.; Corey, R. B. *PNAS* **1951**, *37*, 729−740.
(2) Prusiner, S. B. *PNAS* **1998**, *95*, 13363−13383.
(3) Tan, S. Y.; Pepys, M. B. *Histopathology* **1994**, *25*, 403−414.
(4) Fandrich, M.; Fletcher, M. A.; Dobson, C. M. *Nature* **2001**, *410*, 165−166.
(5) Dong, A.; Randolph, T. W.; Carpenter, J. F. *J. Biol. Chem.* **2000**, *275*, 27689−27693.
(6) Gursky, O.; Aleshkov, S. *Biochim. Biophys. Acta* **2000**, *1476*, 93−102.
(7) Huang, G. S.; Oas, T. G. *Biochemistry* **1995**, *34*, 3884−3892.
(8) Lu, J.; Hall, K. B. *Biophys. Chem.* **1997**, *64*, 111−119.

10.1021/ja0360081 CCC: $25.00

simulations of λ_{6-85} reveal a similar propensity for a local extended structure in the Ramachandran diagram.[9] Thus, the propensity for forming the extended structure common in β-sheets exists a priori, not as a consequence of aggregation. Heat denaturation funnels proteins toward β-sheet aggregates.

Experimental and Computational Methods

Experimental Procedures. The wild-type genes used in this study were obtained from Terrence Oas (λ_{6-85}), Anne Baranger (U1A F56W), Tracy Handel (ubiquitin), and Maria Mas (yeast PGK). Genes for lambda repressor, U1A, and PGK were cloned into PET-15b or PET-28b vectors for His-tag purification. The linked U1A dimer gene was constructed by PCR duplicating the gene, which was inserted into the vector of the single U1A. Single tryptophan fluorescent mutants of λ_{6-85} (several; see Figure 4B), ubiquitin (F45W/I61A), and PGK (W308F) were prepared using the Stratagene Quickchange site-directed mutagenesis kit. U1A and λ_{6-85} were expressed in Rosetta (DE3) pLysS cells (Novagen) and purified through an Ni-NTA column. Lambda repressor was further purified through a Sephacryl S-200 HR column. Purification of yeast phosphoglycerate kinase and ubiquitin was done following the exact protocols described elsewhere.[10,11] Unlike the single proteins harvested from the cell lysate supernatant, linked U1A was found to exist predominantly in inclusion bodies and was therefore purified under denaturing conditions (6 M GuHCl).

Circular dichroism measurements as a function of wavelength and temperature were taken on a JASCO J-715 spectrometer with a standard 1 cm path length capped cuvette. Additional measurements with a 100 μm path length were made by using demountable cells from Starna; these short path length cells allowed us to access wavelengths as low as 205 nm in the presence of several molar denaturant (guanidinium hydrochloride). Fourier transform infrared spectra were taken in the amide I′ band (1600−1700 cm^{-1}) as a function of temperature and concentration. For IR spectroscopy, proteins were lyophilized and redissolved in D_2O. A 75 μm Mylar spacer sandwiched between two CaF_2 windows set the sample path length.

Aggregation kinetics of the lambda repressor fragment were measured by incubating protein at 80 °C in 1 M GuHCl or 0 M GuHCl for comparison. Multiple identical samples were prepared in 1.5 mL Eppendorf tubes and incubated simultaneously (without any stirring or disturbances to the protein solution at any time). At different times after incubation started, samples were taken out and subjected to 15 min of centrifugation at 6000 × *g*. The supernatant after the centrifugation was cooled to 20 °C, and the CD spectrum at 222 nm was recorded. The loss of CD signal at 222 nm caused by precipitation was taken to represent the aggregation kinetics. To account for any salt effects, 1 M CsCl was used instead of 1 M GuHCl and did not cause the trend observed for GuHCl.

All buffers contained 50 mM sodium or potassium phosphate. Protein concentrations differ widely among the various measurements and are listed in the figures or captions illustrating individual experiments.

Molecular Dynamics. Starting coordinates for molecular dynamics (MD) simulations were taken from the native structure of the λ-repressor N-terminal domain (PDB data bank entry 1LMB). MD was carried out using NAMD2,[12] based on the force field charm22.[13] VMD[14] was used for visualization purposes as well as quantitative analysis of the simulation. The feature *SOLVATE* of VMD was used to solvate the protein in a 60 × 60 × 60 $Å^3$ box filled with TIP3 water molecules. Deleting TIP3 molecules within 2.4 Å of the protein and minimizing the energy for 8000 steps resulted in a system containing 20 024 atoms. Periodic boundary conditions were used for all simulations. Unless indicated, all trajectories were carried out in the NVT ensemble.

An unfolded structure for λ_{6-85} was obtained by subjecting the molecule to high temperature. Starting from the folded structure, the system was continuously simulated at 1000 K for 2.21 ns, at 1300 K for 0.72 ns, and finally at 2000 K for 0.65 ns. The final structure of this trajectory was used as the unfolded molecule. The radius of gyration was 27 Å (native value: 11.8 Å), and all the long helices were completely disrupted.

The local extended structure was characterized using Ramachandran plots of the (ψ,ϕ) distributions, obtained by molecular dynamics at three different temperatures. The extended β-region was taken from (ψ = 70° to 180°, φ = −180° to −70°), and the α-region, from (ψ = −90° to −10° φ = −150° to −70°), to provide a conservative estimate of any transitions from the α region to the extended region. Starting out from the unfolded structure obtained at high temperature, the system was simulated at 450 K for 4.4 ns, at 700 K for 2.4 ns, and at 1000 K for 2.5 ns. To test if proper equilibration of the (ψ,ϕ) distribution was accomplished at 450 K, we computed an additional 4.7 ns trajectory at 450 K, starting with a more compact initial structure ($R_g \approx 15$ Å; obtained by propagating the $R_g \approx 27$ Å unfolded state for an additional 14 ns at lower (298 K) temperature). Finally, to construct the potential of mean force at room temperature, we simulate native λ_{6-85} for 1 ns at 298 K. Coordinates of the system were saved every 0.5 ps for further analysis. Potentials of mean force were calculated from the summed population density for all residues ρ by using the formula PMF = $-kT$ ln(ρ).

Results

Unfolded U1A Increasingly Adopts Monomeric Extended Forms at High Temperature. Previous studies of U1A revealed that this protein aggregates at neutral pH when heat denatured,[8] leading to the formation of gel-like precipitates. We find that, at high concentrations of GuHCl (4.5−5.7 M, 50 mM phosphate buffer at pH 7), U1A does not aggregate and exhibits a CD spectrum typically considered as "random coil" at low temperature (1 °C). But as the temperature is raised, the unfolded state shifts more and more toward a β-sheetlike CD signal.[15] Figure 1 shows the circular dichroism spectrum of folded monomeric U1A for comparison with the temperature dependence of the circular dichroism spectrum in 4.5 and 5.7 M GuHCl solutions. The spectra are concentration independent in the 5 μM to 47 μM protein concentration range. In 4.5 M GuHCl at 1 °C, the samples have very small CD values in the 220−240 nm range, characteristic of irregular backbone conformations which have lost the delocalized electric and magnetic dipoles associated with ordered structures. The extended structure increasingly appears upon raising the temperature. This is particularly evident in the 80 °C short-path length spectrum extending to 205 nm. Because GuHCl disrupts side chain interactions, the β-sheetlike CD signature we see most likely results from local extended β-conformation.

High-Temperature U1A Extended Structure Does Not Result from Aggregation. Although high concentrations of GuHCl usually prevent protein aggregation, we checked whether extended β-conformation in the U1A unfolded ensemble at high

(9) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(10) Osvath, S.; Gruebele, M. *Biophys. J.* **2003**, *85*, 1215−1222.
(11) Lazar, G. A.; Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1997**, *6*, 1167−1178.
(12) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.
(13) MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(14) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(15) Greenfield, N.; Fasman, G. D. *Biochemistry* **1969**, *8*, 4108−4116.

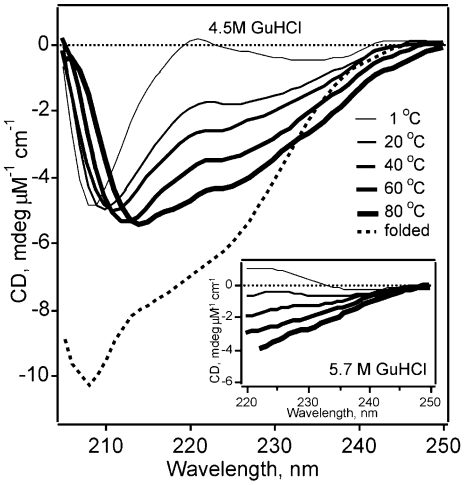

Figure 1. U1A CD spectrum in 4.5 M GuHCl between 1 and 80 °C in comparison to the folded protein. Measurements were done with 47 μM protein in a 100 μm path length cuvette. Inset: same protein in 5.7 M GuHCl solution. Measurements were made with 7.7 μM protein in a 1 cm path length cuvette. All CD data were smoothed over 3 nm intervals. The spectrum contains increasing β-like extended structure at higher temperature.

temperatures is an intrinsic property of the single protein or is induced by interaction with other U1A molecules in the solution. To confirm that the extended structure at high temperature is not caused by aggregation, U1A was subjected to two additional experiments besides the simple concentration dependence reported above:

The temperature scans of the circular dichroism spectrum were repeated at pH 2 (20 mM HCl in 6 M GuHCl buffer). It has been shown that low pH prevents U1A aggregation,[8] and thus properties observed under these kinds of conditions will be intrinsic to the monomeric protein and free of intermolecular effects. We find that the spectrum and increasing extended β-structure trend monitored at CD_{222nm} are exactly the same as those in the pH 7 solvent condition (Figure 2).

We also linked two U1A molecules together with glycine linkers. A 10-residue linker increases the bimolecular collision rate of U1A by confining protein pairs and reducing their translational entropy. The linker has the advantage that only binary collisions are increased in a controlled manner because the overall concentration of constructs can be kept small. An effective binary concentration > 30 mM is obtained (using $2/(6 \times 10^{23})/(4\pi(\mathbf{r}_{linker} + \mathbf{r}_{U1A})^3/3$, with $\mathbf{r}_{linker} = 1.8$ nm and $\mathbf{r}_{U1A} = 1$ nm). If the extended structure observed is caused by interacting U1As, the huge increase in effective concentration (compared to <8 μM protein used in single U1A measurements) should greatly enhance the amount of extended structure signal seen. We found that the absolute signal and trend at CD_{222nm} are identical to that of a single U1A at twice the concentration (Figure 2). This further confirms that the propensity of exhibiting extended β-structures at high temperatures is intrinsic to the single protein.

High-Temperature Extended Structure Induces Aggregation at High Concentration. In addition to the tethering experiments, ordinary concentration dependences can also show that a protein forms an extended structure at low concentrations where aggregation is not significant. Moreover, a concentration scan can reveal how the extended β structure funnels the protein toward aggregation at high temperature. Infrared spectroscopy

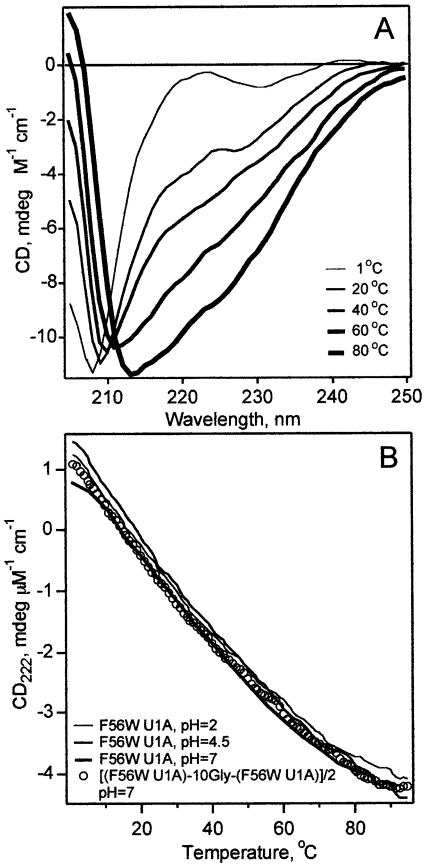

Figure 2. (A) Linked U1A ([F56W U1A]-10Gly-[F56W U1A]) CD spectrum between 1 and 80 °C. Measurements were done with 24 μM protein in a 100 μm path length cuvette in 6M GuHCl buffer as indicated by **r**. (B) Comparison of the trend of increase in β-structure monitored at CD_{222nm} for U1A in different pH buffers. Measurements with single U1A were made with 5.8 μM protein in a 1 cm path length cuvette. For the linked U1A, 2.9 μM protein was used instead. Temperature scans were smoothed over 3 °C intervals. No significant difference is observed, even though linked U1A molecules suffer a binary collision frequency comparable to >30 mM single protein solution.

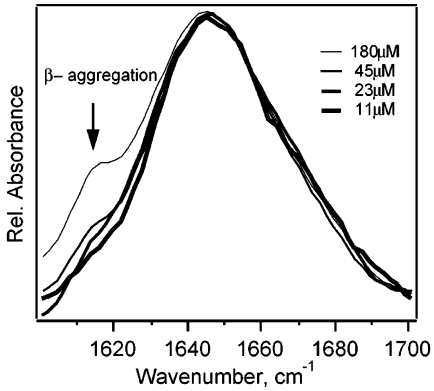

Figure 3. FTIR spectrum for the lambda repressor variant A37G A49G at four different concentrations after 30 min of incubation at 60 °C. The aggregate peak is not strongly evident at concentrations below 180 μM. The spectra were smoothed over a 6 cm^{-1} interval.

is the best tool to distinguish α-helix, extended structures that are hydrogen bonded to the solvent and aggregates.

Figure 3 shows the amide I′ region of A37G A49G λ_{6-85}. At 30 °C, the spectrum has the characteristic helical peak at 1650 cm^{-1}. As the temperature is raised in low concentration solutions (11 μM), the broad peak shifts toward 1645 cm^{-1}, characteristic of the unfolded extended structure with increased hydrogen

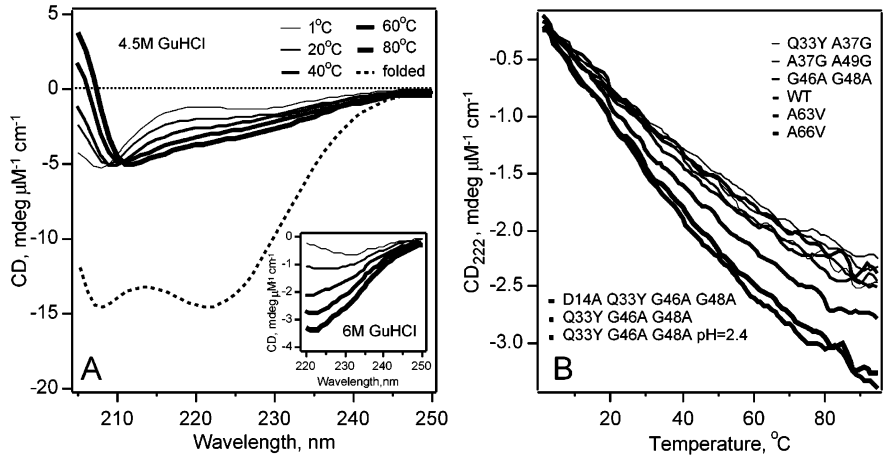

Figure 4. (A) Lambda repressor with mutation Q33Y, G46A, and G48A CD spectrum in 4.5 M GuHCl between 1 and 80 °C in comparison to the folded protein. Measurements were done with 115 μM protein in a 100 μm path length cuvette. Inset: same protein in 6 M GuHCl solution. Measurements were made with 4.9 μM protein in a 1 cm path length cuvette. (B) Comparison of the trend of increase in extended structure monitored at CD_{222nm} for different lambda repressor mutants using 4.9 μM protein in 1 cm path length cuvettes. Traces and legends are sorted from thinnest to thickest in descending order.

bonding to the solvent. When the concentration is raised to 180 μM, a new peak appears at 1615 cm^{-1}, characteristic of β-sheet aggregates.[16] This peak further grows when the sample is incubated at 80 °C for 30 min.

The extended β-structure is thus already thermodynamically favored in the monomeric protein at high temperature and precedes the formation of aggregates, which in addition require high concentration.

The Effect is General: λ_{6-85}, Ubiquitin, PGK, and Many Other Proteins Show the Same Propensity. To provide some evidence that the formation of an extended structure at high temperature is a general phenomenon, we conducted the same tests on three additional proteins. In particular, the N-terminal λ-repressor fragment provides a critical confirmation of the U1A studies because it forms no β-sheets in its native state.

λ_{6-85} in 4.5 to 6 M GuHCl shows the same type of behavior as that of U1A (Figure 4). Although its CD signal is not as close to the typical "random-coil spectrum" as that of U1A at 1 °C, the same shift from a random coil to an extended structure spectrum is observed as a function of temperature. We observed the same trend, monitored at CD_{222nm} (Figure 4), for all the mutants of λ_{6-85} that we tested. We also investigated yeast phosphoglycerate kinase and ubiquitin, whose folds differ from one another and which are in a different fold class from λ_{6-85}.[17] Both show the same shift in CD spectra toward β-like extended conformations and exhibit the characteristic of increasing negative CD_{222nm} at higher temperatures (Figure 5). This shows that this property is robust to mutations and occurs for several different folds. The rate of increase is different for different mutants, and we note that the two variants whose extended structure content increases most rapidly with temperature (a Q33Y G46A G48A triple mutant and a quadruple mutant with an additional D14A mutation) are also the most aggregation prone,[18] correlating increase in extended β-structure content with protein aggregation.

Besides the four proteins we studied here, the literature contains accounts of numerous other proteins whose negative

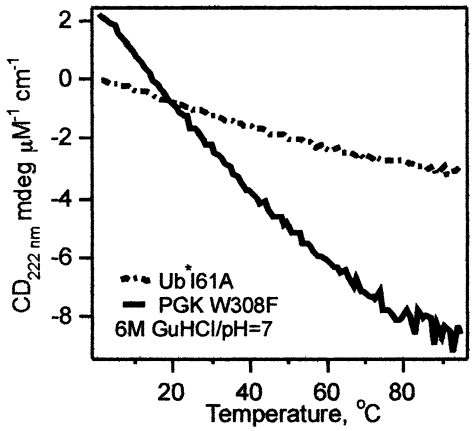

Figure 5. PGK W308F and Ubiquitin F45W/I61A in 6 M GuHCl. As temperature is increased, both show a steady increase in the extended β-structure content (negative CD at 222 nm), similar to that seen for λ-repressor and U1A.

CD_{222nm} in denaturants (both urea and GuHCl) increases in magnitude at high temperatures. The examples include proteins such as LacI DBD,[19] Prothymosin α,[20] Endo-β-1,3-glucanase,[21] Bovine brain A1, Ribonuclease A, apomyoglobin, cytochrome C, hen egg-white lysozyme, staphylococcal nuclease, and collagen,[22] as well as short peptides, for example Glu-Lys-Lys-Glu-Gln-Ala[22] and homopolypetides.[23] The prototypical amyloid peptide, $A\beta_{1-40}$ peptide, which is largely unfolded at low temperature, also forms an extended structure at increased temperature.[6] Thus, many different classes of peptides and folded protein form an extended structure at higher temperature.

The Propensity Is Not Induced by Denaturants. Under conditions favoring the native state, the trend toward extended structure with increasing temperature is usually masked by the presence of the much sharper sigmoidal unfolding transition. Nevertheless, the formation of extended β-structures still occurs and can be revealed by analyzing the CD spectrum as a function

(16) Meersman, F.; Smeller, L.; Heremans, K. *Biophys. J.* **2002**, *82*, 2635−2644.
(17) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536−540.
(18) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193−197.
(19) Felitsky, D. J.; Record, M. T. J. *Biochemistry* **2003**, *42*, 2202−2217.
(20) Gast, K.; Zirwer, D.; Damaschun, G. *Eur. Biophys. J.* **2003**, *31*, 586−594.
(21) Chiaraluce, R.; Van Der Oost, J.; Lebbink, J. H.; Kaper, T.; Consalvi, V. *Biochemistry* **2002**, *41*, 14624−14632.
(22) Privalov, P. L.; I., T. E.; Venyaminov, S. Y.; Griko, Y. V.; Makhatadze, G. I.; Khechinashvili, N. N. *J. Mol. Biol.* **1989**, *205*, 737−750.
(23) Tiffany, M. L.; Krimm, S. *Biopolymers* **1973**, *12*, 575−587.

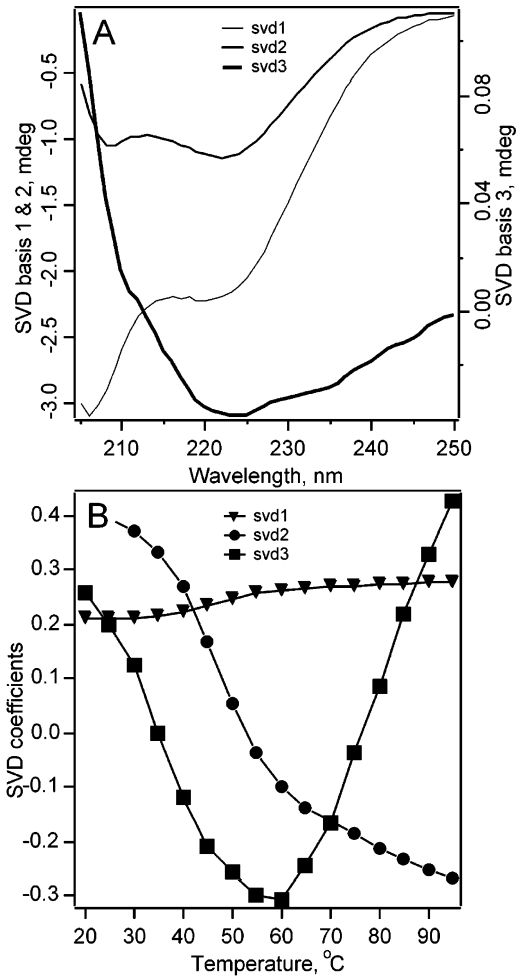

Figure 6. Singular value decomposition (SVD) analysis of the lambda repressor CD spectrum (variant A37G A49G, 1.8 μM) as a function of temperature. (A) The three largest [basis function]*[singular value] from the SVD analysis. All three basis were shifted to have $CD_{250nm} = 0$ mdeg. The third basis function reveals the growth of extended structure. (B) SVD coefficients corresponding to the three basis functions as a function of temperature. The extended structure component (3) increases rapidly at higher temperatures past the unfolding transition midpoint.

of temperature with singular value decomposition.[24] Singular value decomposition extracts the linearly independent components of the spectrum that contribute to the spectral changes as a function of temperature. The first and second components track the usual two-state unfolding transition. The third component shows a steady increase of the extended β-structure signal past the midpoint of the unfolding transition. Because the lambda repressor fragment does not tend to aggregate at high temperatures even in the absence of denaturant, we used it for this particular measurement in Figure 6.

The calculated increase of the extended β-like structure in 0 M GuHCl is about a third of the change observed when protein is situated in denaturant. For the lambda repressor under native solvent conditions, the formation of extended structure is attenuated by other interactions not present in high denaturants, most likely the residual helix-forming propensity in the unfolded state, and side chain hydrophobic interactions. Nevertheless, extended β-structure formation is still a major factor as the temperature of the unfolded ensemble is increased, comparable in magnitude to the case where denaturant is present.

(24) Henry, E. R.; Hofrichter, J. *Methods Enzymol.* **1992**, *210*, 129−192.

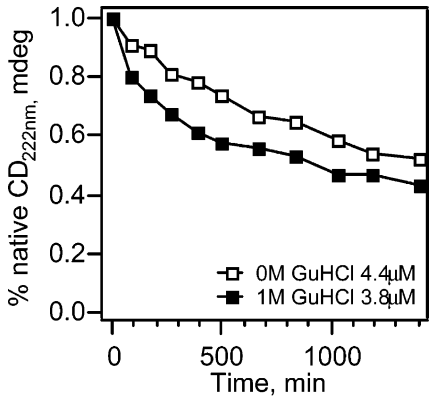

Figure 7. Aggregation kinetics of λ-repressor (Q33Y G46A G48A) at 80 °C in 0 M and 1 M GuHCl. The protein is completely unfolded in both situations (i.e., in the baseline of the sigmoidal temperature unfolding titration curve). The rate of aggregation in 1 M GuHCl turns out to be faster than that in 0 M GuHCl, when there is less extended structure present.

Increasing extended β-content without denaturant has also been observed by Shi et al. in a nonfolding 7-alanine peptide by NMR spectroscopy.[25] It has also been seen in acid-induced partially denatured α-synuclein,[26] the Sup35p prion determining region,[27] and nonfolding short peptides[28] using circular dichroism. All these observations further confirm that GuHCl is not required to induce the extended β-structure propensity at high temperatures.

λ_{6-85} Aggregates Faster When There Is More Extended Structure. We propose one last experiment to observe directly the speed-up of aggregation upon formation of a more extended structure (Figure 7). This is not easily carried out because perturbing the protein ensemble requires tuning interactions (such as hydrophobic effect) and invariably complicates the analysis. However, we were able to locate a solution condition that demonstrates the correlation between the aggregation rate and the amount of extended structure. Usually, addition of guanidinium to proteins already unfolded at low temperature tends to reduce aggregation. However, because guanidinium enhances formation of extended states at high temperature, as found above, it can speed up the rate of aggregation. λ_{6-85} at 80 °C, whether situated in 0 M GuHCl or in 1 M GuHCl, is highly unfolded (K_{eq} for unfolding exceeds 20 in either case, based on thermodynamic fits of the titration curves). Guanidinium decreases hydrophobic and electrostatic interactions and is generally used to reduce aggregation. However, we see a speed-up of the aggregation kinetics in 1 M GuHCl, when compared to that in 0 M GuHCl. As shown above, guanidinium favors an extended structure in the unfolded ensemble, and this effect dominates here.

MD Simulation of λ_{8-85} Confirms the Formation of a More Extended Structure as Temperature Increases. We looked at the protein with the highest helix-forming propensity to determine whether the increase in extended structure with temperature also occurs in molecular dynamics simulations. Although global parameters such as the contact order or the radius of gyration R_g cannot be expected to equilibrate in a few nanoseconds of simulation even at high temperatures, we are

(25) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *PNAS* **2002**, *99*, 9190−9195.
(26) Uversky, V. N.; Li, J.; Fink, A. L. *J. Biol. Chem.* **2001**, *276*.
(27) Scheibel, T.; Lindquist, S. L. *Nat. Struct. Biol.* **2001**, *8*, 958−962.
(28) Tiffany, M. L.; Krimm, S. *Biopolymers* **1972**, *11*, 2309−2316.

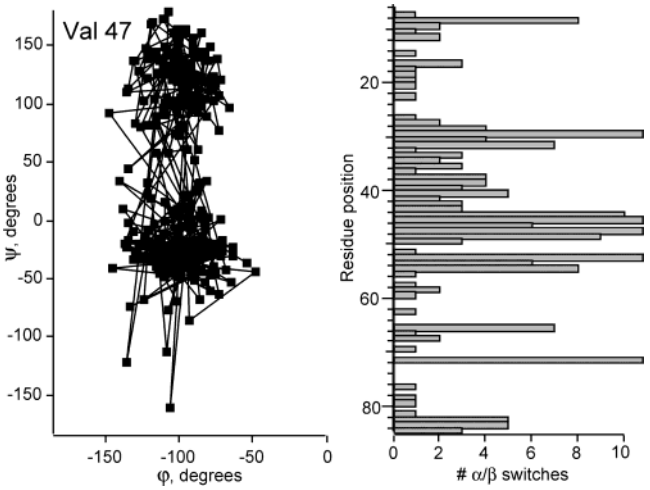

Figure 8. (A) Residue Val 47 of λ_{6-85} illustrates that local backbone structure is fully sampled in the last nanosecond of the first 450 K trajectory. (B) The number of α region to β region (on the Ramachandran plot) transitions each residue executes during the last nanosecond of the trajectory.

interested here in the local backbone conformation. The two 450 K simulations, starting from a more extended structure ($R_g \approx 27$ Å) and a more compact structure ($R_g \approx 15$ Å), nearly converge in R_g after 5 ns (to 18 Å and 20 Å, respectively) and, more importantly, yield nearly indistinguishable distributions of (ψ,φ) angles when the final 1 ns is plotted in a Ramachandran plot. Moreover, most residues on both trajectories switch back and forth between the α, β, and other regions several times during the last nanosecond of simulation, indicating that (ψ,φ) are fully sampled on an individual residue basis (Figure 8). Only a few residues have a sufficiently strong helical propensity to remain in the helical region of the Ramachandran plot throughout the simulation.

Figure 9 shows the potentials of mean force as a function of Ramachandran angles. Trajectories at 298 K (folded), 450 K, 700 K, and 1000 K (not shown, but continuing the same trend of decreasing α/β well-depth ratio) were sampled. The PMFs were obtained from the last 1 ns of each trajectory. The well in the β-region of the Ramachandran plot deepens smoothly with temperature compared to the α-region peak and broadens at the same time. This smooth behavior is one more sign that local backbone structure is sufficiently sampled by the trajectories. To test further for artifacts that might be caused by insufficient sampling, we also divided each trajectory into three and constructed the (ψ,φ) distribution for each division. The distributions obtained at a given temperature were very similar to each other, and differed from those at other temperatures as shown in Figure 9.

Thus the molecular dynamics simulations, within the limitations imposed by classical trajectories and empirical force fields not optimized for all temperatures and solvent densities, show the increasing number of samples with a broadening area of the extended β-sheet region as the temperature is raised. Comparison of the potentials of mean force at 450, 700, and 1000 K shows that population density parity between the α-helical and the β-structure regions increases. This raises the total probability of residues located in β-structure regions. The temperature dependence of the unfolded protein between 450 and 1000 K is not to be confused with the protein unfolding process of λ_{6-85}, which also boosts an increase in β-content through the loss of native α-helical structures (see table of contents figure).

Discussion

Upon examining the secondary structure types in proteins, one finds the β-strand (the extended form of the backbone) a good scaffold for self-association. On the extended β-strand backbone, the C═O and N─H moieties are unprotected and have the potential for intermolecular H-bonding on both sides of the strand, unlike in α-helices where they are intramolecularly H-bonded to their third neighbor. This self-perpetuation in two dimensions rather than just one is a possible reason for the observation that protein aggregates are more often found locked into β-type structures.

We thus have a chicken-and-egg problem: does aggregation impose an extended structure upon the peptide chain, or does the chain have an intrinsic propensity for forming extended β-like structures, which naturally funnels proteins toward β-sheet aggregates as protein concentration is increased? At least for heat aggregation, which serves as an important model for the formation of β-like fibers, the problem is resolved by the sequence of experiments described in the Results section. Monomeric proteins, such as U1A, λ-repressor, ubiquitin, and PGK, investigated here, form extended structures before aggregating as the temperature is raised. Numerous other proteins we cited from the literature also show an increase in the

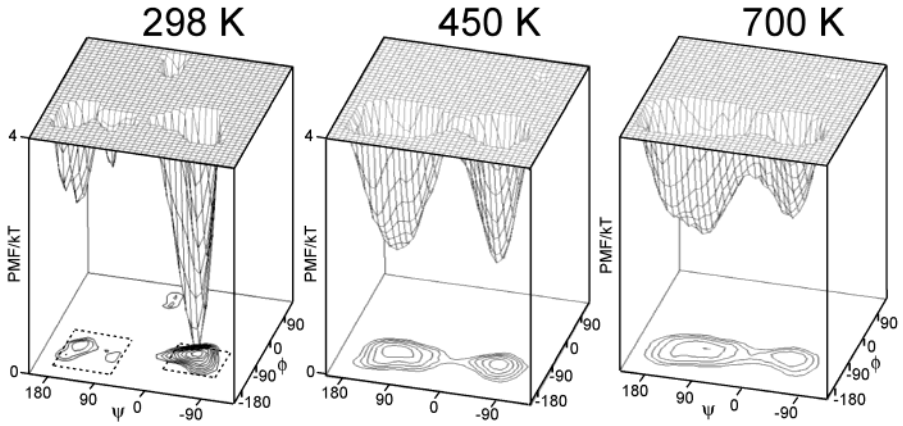

Figure 9. Potential of mean force at three temperatures from 298 K (folded) to 700 K (highly unfolded) for lambda repressor fragment. The peak ratio in the extended (β-structure) and α-helical regions grows steadily as the temperature is increased. The distribution also broadens out from the peaks centroids as temperature is increased. The dashed lines in the 298 K plot indicate the definitions for the α and β regions we used.

extended β-structure CD signature upon heating.[6,19−23] The presence of extended structure subsequently leads to a higher probability of forming random intermolecular H-bonds. This increases the chance of a protein aggregating when the concentration is raised. Indeed, we see from our aggregation kinetics measurements that a more extended structure within the unfolded ensemble speeds up aggregation.

At high temperatures, one would expect conformational entropy (ultimately limited by steric interactions) to control the amount of extended population. At low temperatures, enthalpic effects of intraprotein contacts become increasingly important, and the picture could become more complicated than outlined above. We consider both entropic and enthalpic effects and the resulting temperature dependence of the extended structure propensity in more detail.

The intrinsic sterics of backbone-side chain interactions favor extended structures. It is often assumed that protein structure becomes increasingly like a random coil as temperature is raised. However, large side chains in particular are subject to steric clashes, heavily confining the possible conformations a protein can take on. This makes the behavior of proteins deviate from a Gaussian random chain, leading to overexpanded conformations.[29] In proteins, β-strands are the most extended form and place the side chains in a staggered arrangement that avoids steric clashes with the backbone toward the N terminus. Thus steric constraint is at least as dominant as other types of interactions in guiding local structure propensity.

Ramachadran's plot of the (ψ,ϕ) backbone dihedral angles of an alanine dipeptide demonstrated that only two main secondary structure classes, the α-helix and β-sheet, are observed.[9] Among these two, β-structures occupy the larger fraction of the allowed configuration (Ramachandran) space. In addition, one must not neglect the fact that a two-dimensional (ψ,ϕ) plot for a given amino acid neglects degrees of freedom possessed by the side chains, for example, the residue's side chain torsion angles. The number of conformational microstates that an amino acid side chain can take on in a Ramachandran plot will contribute an effective entropy $S_{\mathrm{eff}}(\psi,\phi)$, increasing the probability in regions of the plot where S_{eff} is larger. The β-region, which allows side chains to move most freely, has the largest S_{eff}. Both the larger range of backbone angles and the increased mobility of side chains contribute to the larger phase space volume and lower PMF of the β-region (Figure 9).

With the major allowed phase space allocated to β-structures in proteins, β-structure propensity at high temperatures is naturally high. This is indeed what we found in our MD simulations as well as in our experiments (where proteins were seen to adopt extended β-structure CD signals). One interesting issue that can be addressed by the MD simulations is what $S_{\mathrm{eff}}(\psi,\phi)$ looks like throughout the Ramachandran space. If S_{eff} were the same in the β and α regions, a protein initially in the α well (Figure 9 top left) should have equally deep PMFs in the α and β regions as $T \rightarrow \infty$. Going from 298 to 450 to 700 K to 1000 K, we find that the β PMF well becomes much deeper than the α well. The peak ratios (or population densities) continue to increase even from 700 to 1000 K, confirming that the side chains are allowed to adopt many more configurations in the extended form.

Simulations by Street et al. are in agreement with our high temperature picture in Figure 9 and from the experiments. They explicitly demonstrated that when any amino acid is placed in an Ala-X-Ala context, 33−48% of the sterically allowed conformations are β-structures[30] (excluding the special residues glycine and proline). In addition, Pappu et al. found that sterics between residue i and ($i + 3$, $i + 4$, $i + 5$, $i + 6$) in longer poly-alanine peptides provide strong additional constraints to the dihedral angle peptide bonds can take on, further reducing the allowed phase space near the α-helix region (and the coil region nearby) while leaving the β region unaffected.[31] If these calculations had been done on amino acids with bulkier side chains, the bias away from the α-helical region would have been even stronger. With this in mind, extended β-structure easily accounts for more than half of the total allowed conformations.

At room temperature, unfolded U1A, λ-repressor, ubiquitin, and PGK monitored by CD do not sample as much extended β-structure as expected from its share of the fraction of the sterically allowed conformations. The same holds true for many other proteins when unfolded at room temperature.[32] This implies that the enthalpy $H_{\mathrm{eff}}(\psi,\phi)$ for β-structures in a Ramachandran plot has to be higher than that of α-helical and other secondary structures. The staggered side chains that increased S_{eff} are again most likely responsible: β-structures extend side chains away from other residues closer to the N-terminus, thereby decreasing the chance for inter-residue interactions (van der Waals, hydrogen bonding, or mediated by water molecules). The smaller probability of random side chain contacts in the unfolded state makes these structures relatively higher in enthalpy. Together with the fact that α-helices form local side chain contacts very well, we can write down $H_{\mathrm{eff}}(\beta) > H_{\mathrm{eff}}(\alpha)$ for the unfolded structural substates.

The shift toward extended structures at higher temperatures seen in our experiment could also be explained by a temperature dependence of H_{eff} (e.g., the hydrophobic effect decreases at low temperatures in water, resulting in cold denaturation).[33] We believe that such contributions, if any, are smaller than the entropic effects for several reasons. Proteins do unfold at high temperatures, so hydrophobicity is much less of a factor compared to backbone entropic effects; a switchover in the ordering of $H_{\mathrm{eff}}(\beta)$ and $H_{\mathrm{eff}}(\alpha)$ would be required, which is physically hard to rationalize. The increase of extended structure in 6 M GuHCl as the temperature is raised is actually larger than that in 0 M GuHCl, indicating that the attractive interactions are acting against the increase of extended structure at all temperatures in accordance with our proposal that $H_{\mathrm{eff}}(\beta) > H_{\mathrm{eff}}(\alpha)$. This again rules out hydrophobicity as the main underlying interaction.

Conclusions

From our observation of intrinsic extended β-structure propensity in different monomeric proteins, both in the presence and absence of denaturants over a wide range of concentrations, two important conclusions can be drawn that may be of interest in other areas.

(29) Damaschun, G.; Damaschun, H.; Gast, K.; Zirwer, D. *Biochemistry (Moscow)* **1998**, *63*, 259−275.
(30) Street, A. G.; Mayo, S. L. *PNAS* **1999**, *96*, 9074−9076.
(31) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *PNAS* **2000**, *97*, 12565−12570.
(32) Eliezer, D.; Chung, J.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2000**, *39*, 2894−2901.
(33) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521−533.

The first, already alluded to several times, is that steric constraints in the states of highest conformational entropy funnel monomeric proteins toward extended β-structures prior to aggregation. The effect is most pronounced in the presence of denaturant, but an effect of similar magnitude (30%) is preserved even in the absence of denaturants. At lower temperatures, the effect is partly masked by more favorable alternative side chain interactions during folding but persists even there. The effect is not mainly due to hydrophobic interactions and occurs in many structural classes of proteins. It is a generic property of proteins, not limited to those classically associated with fiber or aggregate formation or those with extensive β-sheets in the native state.

Second, the unfolded state contains a temperature-dependent structure. This could be one reason for the commonly observed non-Arrhenius type folding kinetics in proteins.[18,34−36] How important this factor is compared to structural changes in the transition state will have to be examined more closely. Certainly any movement of the unfolded state needs to be considered when using ϕ-value analysis (free energy derivatives) to study transition state structures.[37,38] It has been pointed out that ϕ_T analysis (probing transition state motion via temperature changes) provides only a *relative* transition state coordinate,[39] but it is often conveniently assumed in any kind of ϕ-value analysis that the position of the folded and of the unfolded state along the folding reaction coordinate is fixed.[40] It would be fruitful to characterize more carefully the changes in the unfolded state upon different perturbations, if an absolute interpretation of ϕ-value analysis results is desired.

Acknowledgment. This research was supported by a grant from the National Science Foundation, MCB-0316925. The computational resources were supported by an NIH Center grant to Klaus Schulten (P41 RR05969-05), and we thank Klaus Schulten for making time on a Linux cluster available for the MD calculations. E.L. was supported by an IBM Predoctoral Fellowship, and M.G. thanks the University of Illinois for an Alumni Scholarship.

JA0360081

(34) Scalley, M. L.; Baker, D. *PNAS* **1997**, *94*, 10636−10640.
(35) Schindler, T.; Schmid, F. X. *Biochemistry* **1996**, *35*, 16833−16842.
(36) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2000**, *298*, 283−292.
(37) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(38) Sánchez, I. E.; Kiefhaber, T. *Biophys. Chem.* **2003**, *100*, 397−407.
(39) Ervin, J.; Gruebele, M. *J. Biol. Phys.* **2002**, *28*, 115−128.
(40) Matouschek, A.; Fersht, A. R. *PNAS* **1993**, *90*, 7814−7818.