J|A|C|S
ARTICLES
Published on Web 02/20/2004

Determination of an Ensemble of Structures Representing the Denatured State of the Bovine Acyl-Coenzyme A Binding Protein

Kresten Lindorff-Larsen,[†] Sigridur Kristjansdottir,[‡] Kaare Teilum,[‡] Wolfgang Fieber,[‡] Christopher M. Dobson,[†] Flemming M. Poulsen,[‡] and Michele Vendruscolo*[,†]

Contribution from the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom, and Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received October 27, 2003; E-mail: mv245@cam.ac.uk

Abstract: The denatured state of a protein contains important information about the determinants of the folding process. By combining site-directed spin-labeling NMR experiments and restrained computer simulations, we have determined ensembles of conformations that represent the denatured state of the bovine acyl-coenzyme A binding protein (ACBP) at three different concentrations of guanidine hydrochloride. As the experimentally determined distance information corresponds to weighted averages over a broad ensemble of structures, we applied the experimental restraints to a system of noninteracting replicas of the protein by using a Monte Carlo sampling scheme. This procedure permits us to sample ensembles of conformations that are compatible with the experimental data and thus to obtain information regarding the distribution of structures in the denatured state. Our results show that the denatured state of ACBP is highly heterogeneous. The high sensitivity of the computational method that we present, however, enabled us to identify long-range interactions between two regions, located near the N- and C-termini, that include both native and non-native elements. The preferential formation of these contacts suggests that the sequence-dependent patterns of helical propensity and hydrophobicity are important determinants of the structure in the denatured state of ACBP.

Introduction

Providing a detailed characterization of structural features in non-native states of proteins is becoming increasingly important in molecular biology. A large number of proteins are believed to contain long "natively unfolded" regions,[1] that is, sequences that do not by themselves adopt a single, well-defined conformation under physiological conditions.[2,3] Residual structure in the denatured state may also affect both protein stability[4,5] and folding behavior.[6,7] It has also been suggested that partially unfolded conformations populated at equilibrium may be the precursors for formation of amyloid fibrils and toxic prefibrillar protein aggregates.[8,9]

Small-angle X-ray scattering studies indicate that individual structures of denatured proteins vary from being rather compact to highly expanded.[10] Additional detailed structural information about denatured states can be obtained by nuclear magnetic resonance (NMR) spectroscopy.[11−13] NMR experiments are able to provide data such as NOEs, chemical shifts, and scalar coupling constants that are commonly employed in structure determination of native states, but which also have been applied to denatured states.[14−19] More recently, paramagnetic relaxation enhancement (PRE) from site-directed spin-labeling[20−22] and residual dipolar couplings[23] have also proved to be useful in

[†] University of Cambridge.
[‡] University of Copenhagen.

(1) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inf.* **2000**, *11*, 161−171.
(2) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739−756.
(3) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54−60.
(4) Shortle, D. *FASEB J.* **1996**, *10*, 27−34.
(5) Robic, S.; Guzman-Casado, M.; Sanchez-Ruiz, J. M.; Marqusee, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11345−11349.
(6) van Gunsteren, W. F.; Bürgi, R.; Peter, C.; Daura, X. *Angew. Chem., Int. Ed.* **2001**, *40*, 351−355.
(7) Daggett, V.; Fersht, A. R. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497−502.
(8) Chiti, F.; Taddei, N.; Bucciantini, M.; White, P.; Ramponi, G.; Dobson, C. M. *EMBO J.* **2000**, *19*, 1441−1449.
(9) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507−511.
(10) Millet, I. S.; Doniach, S.; Plaxco, K. W. *Adv. Protein Chem.* **2002**, *62*, 241−262.
(11) Shortle, D. R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 24−30.
(12) Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M. *Fold. Des.* **1996**, *1*, R95−R106.
(13) Barbar, E. *Biopolymers* **1999**, *51*, 191−207.
(14) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661−2666.
(15) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977−8991.
(16) Zhang, O.; Kay, L. E.; Shortle, D.; Forman-Kay, J. D. *J. Mol. Biol.* **1997**, *272*, 9−20.
(17) Blanco, F. J.; Serrano, L.; Forman-Kay, J. D. *J. Mol. Biol.* **1998**, *284*, 1153−1164.
(18) Mok, Y. K.; Kay, C. M.; Kay, L. E.; Forman-Kay, J. D. *J. Mol. Biol.* **1999**, *289*, 619−638.
(19) Kortemme, T.; Kelly, M. J. S.; Kay, L. E.; Forman-Kay, J.; Serrano, L. *J. Mol. Biol.* **2000**, *297*, 1217−1229.
(20) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158−169.
(21) Teilum, K.; Kragelund, B. B.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *324*, 349−357.
(22) Lietzow, M. A.; Jamin, M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 655−662.
(23) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489.

obtaining information about denatured states. Dynamic properties can be extracted from heteronuclear spin relaxation experiments[24,25] and may provide further information about structural features in the denatured state.

A major difficulty in the structural interpretation of NMR studies of denatured states is the fact that the experimental data are averages over a broad ensemble of structures. For example, in NOE and PRE experiments, the measured distances are weighted averages that are biased toward close contacts, because of the r^{-6} dependence of dipole−dipole interactions. The use of such average distances as restraints in structure calculations of denatured states is problematic, and it is critical to take this averaging specifically into account. Therefore, standard structure determination procedures where experimentally determined restraints are interpreted in terms of a single molecular structure cannot be applied to denatured states.[26,27]

Several computational approaches have been developed to solve the less dramatic but still important averaging problem involved in the determination of native structures. One of these methods[27] has recently been extended to structure determination of the denatured state ensemble. The central idea in these methods is that experimental data must be compared to averages calculated over an ensemble of structures. These computational techniques can be divided into three categories. The first and most straightforward approach is to pregenerate multiple conformations and subsequently assign weights to each conformation so that the ensemble average matches the experimentally determined results. In the case of native states, these initial structures can be determined using either all or a subset of the available restraints.[28−30] For a denatured state ensemble, large numbers of random conformations have been used to form the pregenerated ensemble.[27] This approach requires that the ensemble of initial conformations includes the structures in the specific state under investigation as no new conformations are added during the assignment of the relative weights.

The second method, referred to as molecular dynamics with time-averaged restraints (MD-TAR),[31−36] is based on the equivalence between time and ensemble averages expected in sufficiently long simulations. In MD-TAR, time averages are calculated over conformations generated during a molecular dynamics (MD) trajectory, by introducing a memory function weighted exponentially in the molecular mechanics force-field.

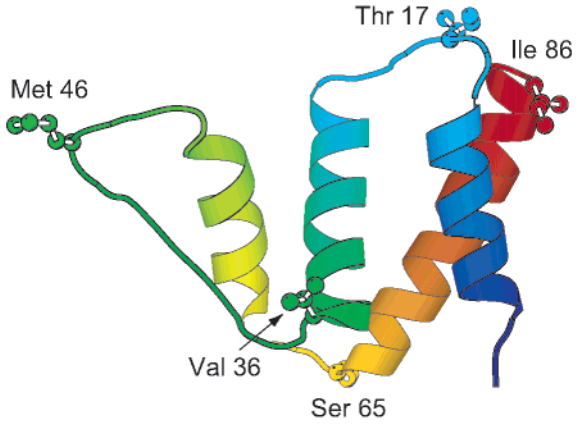

Figure 1. NMR structure of the native state of ACBP. The protein is colored from its N (blue) to C (red) terminus. The four helices are spanned by the following residues: A_1, 3−15; A_2, 21−36; A_3, 52−62; and A_4, 65−84. The side chains of five residues that were individually mutated to cysteine are shown and labeled. The figure was prepared using Bobscript.[70]

The method requires that the time span over which averages are taken is sufficient to sample the conformational space available.

In the third method, known as "ensemble averaged refinement" in NMR structure determination[37−41] or "multiconformer refinement" in X-ray structure determination,[36,42,43] several noninteracting copies of the molecule are simulated in parallel, and at each step of the simulation the experimental restraints are compared to ensemble averages calculated across the multiple copies. As the simulation proceeds, each copy may diverge and explore different regions of conformational space as long as the ensemble average does not violate the restraints. So far, this procedure has mainly been applied to the refinement of native state structures.

In this work, we describe an approach, the Monte Carlo replica sampling (MCRES) method, that considerably extends the scope of the multiconformer technique and allows for efficient sampling of very heterogeneous ensembles of structures. We have applied the method to experimental PRE data from the denatured state of the bovine acyl-coenzyme A binding protein (ACBP). ACBP is an 86 residue four-helix bundle protein whose folding process has been studied extensively.[44−46] The native state structure of ACBP is shown in Figure 1. The denatured state of ACBP is known to contain a broad range of protein conformations,[21,47] and it is therefore essential to take

(24) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Durchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722.
(25) Choy, W. Y.; Shortle, D.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 1748−1758.
(26) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170−184.
(27) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011−1032.
(28) Kim, Y.; Prestegard, J. H. *Biochemistry* **1989**, *28*, 8792−8797.
(29) Ulyanov, N. B.; Schmitz, U.; Kumar, A.; James, T. L. *Biophys. J.* **1995**, *68*, 13−24.
(30) Lukin, J. A.; Kontaxis, G.; Simplaceanu, V.; Yuan, Y.; Bax, A.; Ho, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 517−520.
(31) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(32) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(33) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. Time averaged distance restraints in NMR based structural refinement. In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*; Hoch, J. C., Ed.; Plenum Press: New York, 1991.
(34) Gros, P.; van Gunsteren, W. F.; Hol, W. G. J. *Science* **1990**, *249*, 1149−1152.
(35) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479.
(36) Burling, F. T.; Brünger, A. T. *Isr. J. Chem.* **1994**, *34*, 165−175.
(37) Scheek, R. M.; Torda, A. E.; Kemmink, J.; van Gunsteren, W. F. Structure determination by NMR: The modeling of NMR parameters as ensemble averages. In *Computational aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*; Hoch, J. C., Ed.; Plenum Press: New York, 1991.
(38) Bonvin, A. M. J. J.; Rullmann, J. A. C.; Lamerichs, R. M. J. N.; Boelens, R.; Kaptein, R. *Proteins* **1993**, *15*, 385−400.
(39) Kemmink, J.; van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993**, *230*, 312−322.
(40) Bonvin, A. M. J. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.
(41) Görler, A.; Ulyanov, N. B.; James, T. L. *J. Biomol. NMR* **2000**, *16*, 147−164.
(42) Kuriyan, J.; Osapay, K.; Burley, S. K.; Brunger, A. T.; Hendrickson, W. A.; Karplus, M. *Proteins* **1991**, *10*, 340−58.
(43) Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brünger, A. T. *Science* **1996**, *271*, 72−77.
(44) Kragelund, B. B.; Osmark, P.; Neergaard, T. B.; Schiødt, J.; Kristiansen, K.; Knudsen, J.; Poulsen, F. M. *Nat. Struct. Biol.* **1999**, *6*, 594−601.
(45) Teilum, K.; Kragelund, B. B.; Knudsen, J.; Poulsen, F. M. *J. Mol. Biol.* **2000**, *301*, 1307−1314.
(46) Teilum, K.; Maki, K.; Kragelund, B. B.; Poulsen, F. M.; Roder, H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9807−9812.
(47) Thomsen, J. K.; Kragelund, B. B.; Teilum, K.; Knudsen, J.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *318*, 805−814.

ensemble averaging into account in structural interpretation of the PRE data. The results in this study show that the MCRES method permits us to sample the denatured state efficiently and to obtain structural information that could not otherwise be extracted from the experimental data.

Methods

Experimental Methods. Spin-labeling experiments were carried out as described previously.[21] In short, five site-directed cysteine insertion mutants (T17C, V36C, M46C, S65C, and I86C) of ACBP were expressed, purified, and labeled individually with the nitroxide spin-label MTSL. $^{15}N-^{1}H$ HSQC experiments were recorded on a Varian NMR spectrometer operating at a proton frequency of 750 MHz. Spectra were recorded with the nitroxide in both its oxidized (paramagnetic) and its reduced (diamagnetic) state; the latter was obtained by reduction with ascorbate.[21] The denatured state was stabilized by addition of either 1.6, 1.9, or 3.0 M guanidine hydrochloride (GuHCl). NMR peak intensities were converted into distance restraints as described previously.[21,26] Distance restraint values of 280, 303, and 269 ($|i - j| > 1$) were obtained from the data at 1.6, 1.9, and 3.0 M GuHCl, respectively (see Supporting Information).

^{1}H PG-SLED (pulse gradient stimulated echo longitudinal encode-decode) NMR experiments[48,49] were performed on wild-type ACBP to determine the hydrodynamic radius (R_h) of the denatured state. Dioxane was used as an internal standard.[48,49]

Computational Methods. Monte Carlo simulations were carried out using a C_α model of the protein chain.[50,51] The C_α pseudoatoms are separated by a fixed bond-length of 3.8 Å from their neighbors and have a hard-sphere diameter of 5 Å. Apart from these two types of interactions and the distance restraints described below, our model does not contain any additional interaction potential. In MCRES, an ensemble of N_{rep} noninteracting replicas of the molecule are simulated in parallel. Each step in the Monte Carlo simulation consists of a crank-shaft move[50,51] of a randomly selected residue in each of the N_{rep} members of the ensemble. For each pair of residues, we calculate the ensemble averaged distances as the average over the N_{rep} replicas according to $d_{ij}^{calc} = (N_{rep}^{-1} \sum_{k=1}^{N_{rep}} r_{ij,k}^{-6})^{-1/6}$. The calculated distances are compared to the PRE derived experimental restraints using a penalty function of the form: $E = \sum_{i,j}(d_{ij}^{exp} - d_{ij}^{calc})^2$. In the calculation of E, distances were allowed to vary within ±5 Å in accord with previous estimates of the uncertainty of PRE derived distance information.[26,52] In the case of distance restraints ≥20 Å, the lower bound was set to 20 Å, and no upper bound was enforced. Each Monte Carlo step is evaluated according to the Metropolis criterion[53] using a pseudo temperature of 10^{-3}. At this temperature, the average pseudo-energy, E, obtained is about 10^{-3} with fluctuations also of the order of 10^{-3}. The small value of the energy and of its fluctuations ensures that the restraints are satisfied during the simulations. The number of Monte Carlo steps varied between different simulations, depending on which observables were recorded. Convergence of simulations was determined by monitoring several sets of molecular properties including the radius of gyration (R_g) and individual pairwise distances. After we established that 20 replicas were sufficient to sample the denatured state (see Results), long simulations were run for 2×10^7 Monte carlo steps, the first half of which was discarded to ensure equilibration. Structures were saved every 10^4 Monte Carlo steps so that the final ensembles consisted of 20 000 uncorrelated structures (1000 per replica).

In the calculation of contact probabilities, we define residues as being in contact if their pairwise distance is below 8.5 Å.[50,51] All-atom models were generated from the C_α structures using MAXSPROUT[54] with default settings. These structures were energy minimized for 200 steps using CHARMM[55] with the EEF1 potential.[56] The global features of the structures did not change as a result of this energy minimization (not shown).

A 2 ns native state CHARMM/EEF1 MD all-atom simulation was carried out at 300 K. The EEF1 interaction energy between each pair of residues was calculated from this simulation as described previously.[57] Four 10 ns high temperature (400, 500, 600, and 700 K) MD simulations were also performed using CHARMM/EEF1, and each was initiated from a single highly expanded structure. A total of 103 all-atom structures from the 700 K simulation were selected to sample a range of values of R_g. The R_h value of each of these 103 structures was estimated using HYDROPRO[58] with default settings and automatic extrapolation to zero bead-size. R_g was calculated from the C_α coordinates to allow for a comparison to the values calculated from the C_α model.

Results

For most proteins, the denatured state is populated at a very low level under near-physiological conditions. It is therefore usually studied by destabilizing the native state either by addition of denaturant,[15,19,24,59] by heat,[60] by mutation,[61] or by removal of binding partners.[62] Evidence for nonrandom residual structure in the GuHCl unfolded state of ACBP was recently found from site-directed spin-labeling measurements.[21] In this work, we extended these experiments to additional GuHCl concentrations, enabling us to explore the extent to which the structural features of the denatured state are affected by the presence of GuHCl. The experiments include the site-specific attachment of a nitroxide spin-label to one of five genetically engineered cysteine residues. The five residues (see Figure 1) are distributed along the sequence of ACBP and were chosen to avoid drastic perturbations in the chemical nature of the amino acid. The *m*-values for the spin-labeled mutants,[21] which are correlated to the degree of exposure of surface area during unfolding,[63] are approximately the same as that of wild-type ACBP (14.7 ± 0.1 kJ mol^{-1} M^{-1}), with the *m*-value of M46C (12.6 ± 0.4 kJ mol^{-1} M^{-1}) being the most dissimilar. The overall similarity of the *m*-values suggests that the denatured states of the mutational variants are not significantly different from that of the wild-type protein.

As the native and denatured states are in slow exchange on the NMR time-scale, we can study the denatured state even under conditions where the native state is significantly populated. This is illustrated in Figure 2 in which we show $^{15}N-^{1}H$ HSQC NMR spectra of the diamagnetic form of the MTSL-labeled T17C mutant of ACBP in 3.0 M GuHCl and 1.6 M GuHCl. In the spectrum recorded in 3.0 M GuHCl, we observe

(48) Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 199−203.
(49) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424−16431.
(50) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641−645.
(51) Davis, R.; Dobson, C. M.; Vendruscolo, M. *J. Chem. Phys.* **2002**, *117*, 9510−9517.
(52) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355−5365.
(53) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(54) Holm, L.; Sander, C. *J. Mol. Biol.* **1991**, *218*, 183−194.
(55) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(56) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.
(57) Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. *J. Mol. Biol.* **2002**, *324*, 151−163.
(58) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719−730.
(59) Freund, S. M.; Wong, K. B.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10600−10603.
(60) Arcus, V. L.; Vuilleumier, S.; Freund, S. M.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 305−321.
(61) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M.; Alonso, D. O.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863−867.
(62) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1997**, *36*, 3959−3970.
(63) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138−2148.

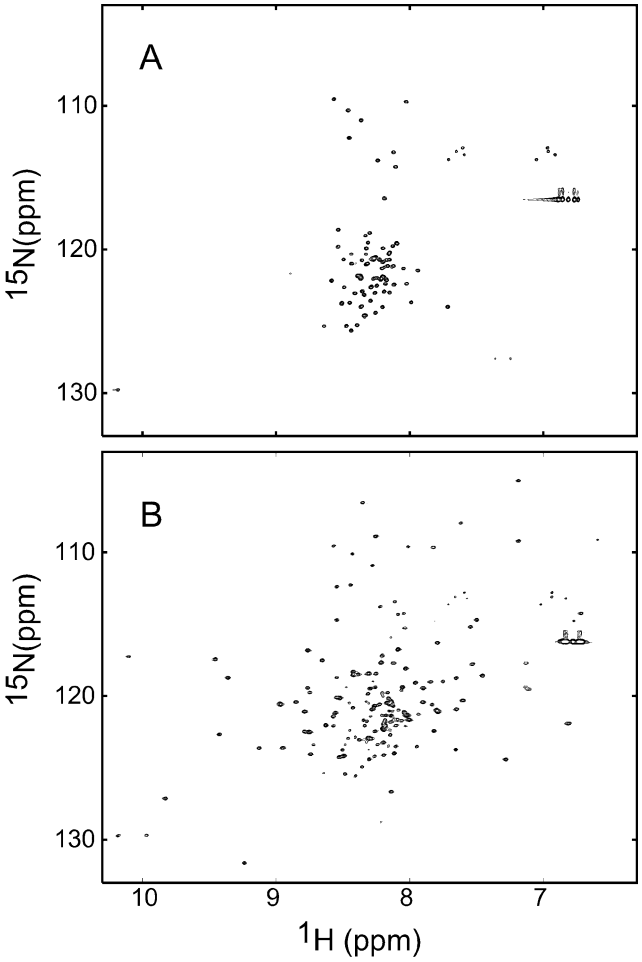

Figure 2. HSQC spectra of the diamagnetic form of the spin-labeled T17C mutant of ACBP recorded in (A) 3.0 M GuHCl and (B) 1.6 M GuHCl. In the spectrum recorded in 3.0 M GuHCl, we observe one set of peaks corresponding to the denatured state of ACBP. In the spectrum recorded at 1.6 M GuHCl, the native state is populated to approximately 60% and gives rise to an extra set of peaks in addition to those from the denatured state.

one set of peaks corresponding to the denatured state of ACBP. In the spectrum recorded in 1.6 M GuHCl, we observe an additional set of peaks corresponding to the native state of ACBP. The peaks corresponding to the denatured state in the spectrum recorded at 1.6 M GuHCl spectrum could be assigned and analyzed separately based on the assignments of the denatured state[21] and of the native state in the absence of denaturant.

The interaction between the paramagnetic nitroxide radical and nearby protons causes broadening of their NMR signals by an increase in transverse relaxation rate. In the simplest case, this effect is dependent on the electron−proton distance as r^{-6}. By measuring peak intensities in two ^{15}N−^{1}H HSQC NMR spectra, with the nitroxide in both its paramagnetic (oxidized) and its diamagnetic (reduced) form, it is therefore possible to obtain distance information between the spin-label and the various amide protons in the protein.[20,21,52,64,65] The intensity ratio I_{ox}/I_{red} is a convenient indicator of this interaction. A ratio of unity indicates no relaxation enhancement and thus an average distance between electron and proton larger than ∼20 Å.[21,52] A ratio of zero indicates complete elimination of the amide proton signal and thus a very strong electron−proton interaction.

We performed PRE experiments on ACBP in 1.6, 1.9, and 3.0 M GuHCl. The resulting I_{ox}/I_{red} intensity ratios are shown in Figure 3. The population of the denatured state under these three denaturant concentrations can be estimated from GuHCl induced unfolding curves for the five mutants, analyzed under the assumptions of a two-state equilibrium.[21] In 1.6 M GuHCl, the denatured state is populated between 40% (T17C) and 90% (S65C), and in 1.9 M GuHCl, it is between 80% (T17C) and 99% (S65C). In 3.0 M GuHCl, the denatured state is >99.9% populated for all mutants. We have also performed experiments under iso-stability conditions[21](S.K. and F. M.P., unpublished) in which the denatured state was equally populated (at 45%, 65%, 95%, and 100%) for all five variants; only the results obtained at a fixed denaturant concentration are, however, discussed here.

The I_{ox}/I_{red} ratios were converted into distance restraints as described previously.[21,26,52,64,65] With these data as restraints, we performed MCRES calculations of the denatured state of ACBP using a C_α model of the protein chain. This model is computationally fast and permits us to sample the denatured state efficiently. Although the model lacks a detailed molecular force-field, the use of experimental restraints allows us to sample conformations that are experimentally relevant and to examine the topological features of the denatured state.

We first examined whether there are conformations that, either individually or as ensembles, satisfy the experimental distance restraints. In the Monte Carlo sampling, we found that even in simulations with $N_{rep} = 1$ it is possible to find structures that satisfy all of the available restraints within error bounds. However, as in the simulations with $N_{rep} = 1$, we require that all restraints are satisfied in a single structure; such simulations can only sample a small fraction of the denatured state ensemble. By increasing N_{rep}, it becomes possible to sample additional conformations that, when distances are averaged, are also fully compatible with the experimentally observed data. This is a result of the r^{-6} averaging of the distances across the replicas such that short distances in only a few replicas permit longer distances in the remainder. An important parameter in MCRES is therefore N_{rep}, the number of replicas used in the simulations. We examined values of N_{rep} ranging from 1 to 50 and predicted that the overall calculated structural ensembles will expand as a function of the number of replicas.

In Figure 4A, we show the distribution of R_g values obtained with values of N_{rep} in the range from 1 to 50. In the one-replica simulations, only compact conformations are sampled, whereas, as more replicas are included, structures with larger R_g are observed. To examine whether the expanded structural ensembles calculated with multiple replicas are consistent with independent data, we determined R_h experimentally using NMR diffusion experiments.[48,49] Such experiments report on the translational diffusion coefficient which is directly related to R_h.[49,66] In 3.0 M GuHCl, we found $R_h = 25.1 \pm 0.3$ Å by measuring the translational diffusion coefficient of denatured ACBP relative to that of dioxane. This value is slightly lower

(64) Gaponenko, V.; Howarth, J. W.; Columbus, L.; Gasmi-Seabrook, G.; Yuan, J.; Hubbell, W. L.; Rosevear, P. R. *Proteins* **2000**, *9*, 302−309.
(65) Donaldson, L. W.; Skrynnikov, N. R.; Choy, W.-Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 9843−9847.
(66) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Volume Part II: Techniques for the Study of Biological Structure and Function*; W. H. Freeman & Co.: New York, 1980.

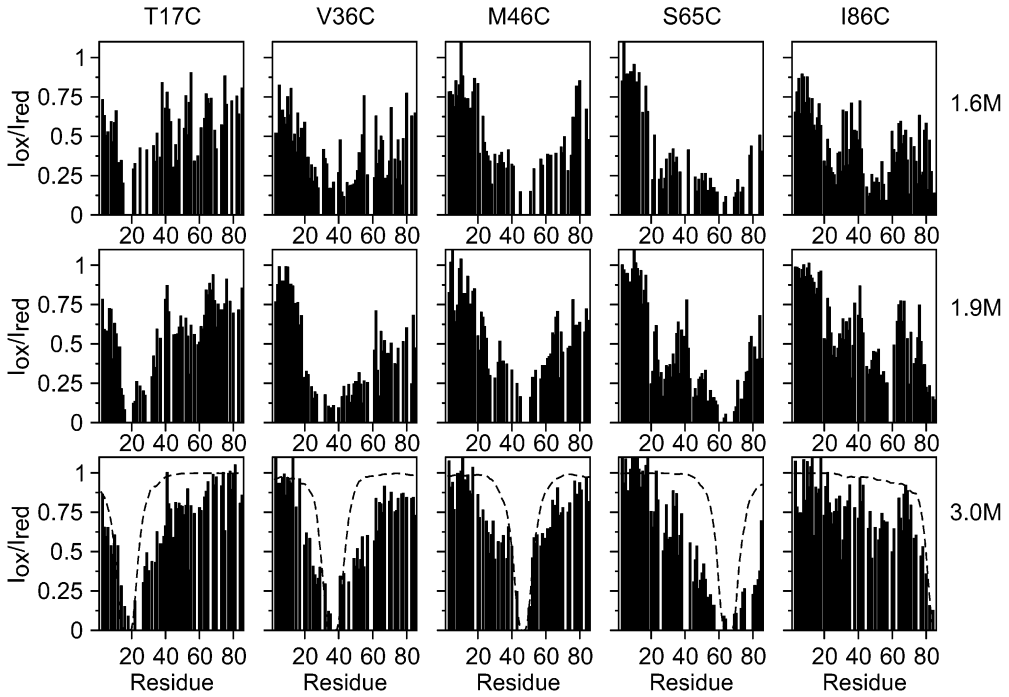

Figure 3. Site-directed spin-labeling reveals residual structure in the denatured state. Each plot shows peak intensity ratios between HSQC NMR spectra with the nitroxide in its paramagnetic and in its diamagnetic state. The five plots correspond to different sites for attachment of the spin-label. The experiments were performed at three different GuHCl concentrations: top, 1.6 M; middle, 1.9 M; and bottom, 3.0 M. The dashed lines in the 3.0 M plots show the expected ratios in a random coil model.[21]

than the 28 Å predicted from the length dependence of R_h for a series of denatured proteins and peptides.[49] For comparison, the R_h of the native state was determined to be 17.6 ± 0.2 Å.

To convert the data from the structural ensembles into a form which can be directly compared to the diffusion experiments, we use a method previously described for the SH3 domain from drkN.[67] First, we generated an ensemble of 103 all-atom conformations of ACBP using high-temperature MD simulations. For each of these 103 structures, we calculated the R_g of the C_α atoms and estimated R_h using the program HYDRO-PRO.[58] The results from this procedure suggest a phenomenological relationship between the NMR diffusion data and R_g values. This relationship is shown in Figure 4B in which each point corresponds to one of the 103 ACBP conformations obtained from the MD simulations. The results indicate that, for ACBP, R_h^{-1} and R_g^{-1} are approximately linearly related. Thus, we estimate the hydrodynamic radius of the calculated ensembles as $R_h^{calc} = \langle 0.0212 + 0.430 R_g^{-1} \rangle^{-1}$, where R_g and R_h are both measured in angstroms. A numerically similar expression was derived for the SH3 domain,[67] and we obtained essentially identical results ($R_h^{calc} = \langle 0.0212 + 0.428 R_g^{-1} \rangle^{-1}$) using TRADES[68] sampling of ACBP conformations. In Figure 4C, we plot our calculated estimates of R_h as a function of N_{rep}. It is evident that the average molecular size in the calculated denatured state ensembles increases as a function of N_{rep}. The molecular size, however, converges to a value between 25 and 26 Å at all three denaturant concentrations. These predictions are in close agreement with the value of $R_h = 25.1 \pm 0.3$ Å determined experimentally in 3.0 M GuHCl, thus confirming that the multiple replica sampling approach is able to sample the expanded conformations relevant to the denatured state. We carried out additional diffusion experiments at several different denaturant concentrations and observed a slight expansion of the denatured state of about 0.44 ± 0.06 Å/M when the GuHCl concentration was increased. A similar trend was observed in the calculated values of R_h (Figure 4C), but the small differences observed are on the same order of magnitude as the experimental error.

The multiple replica sampling method can also be tested through a cross-validation procedure using the experimental data.[36,40,43] For each of the datasets at 1.6, 1.9, and 3.0 M, we performed cross-validation calculations with 1−50 replicas. For each denaturant concentration, we divided the distance restraints into five randomly selected subsets. We then carried out a series of simulations, in each case using only four of the five subsets and leaving out the last for cross validation. We then calculated an R-score defined as $R_{free} = \sum(d_{ij}^{calc} - d_{ij}^{exp})^2 / \sum d_{ij}^{exp}$, where "free" indicates that the summation includes only the distances that were not used as restraints. In Figure 5, we show the average of the five R-scores from each subset as a function of the number of replicas. The results show that the prediction of the experimental distance information in the test set improves significantly when multiple replicas are included in the simulations. The largest improvement occurs when going from one to two replicas, but further addition of replicas up to about 20 causes an additional decrease in the R-score at all three denaturant concentrations. The R-score then remains approximately constant at higher values of N_{rep}.

To study the denatured state in more detail, we performed long simulations (10^7 Monte Carlo steps for each replica after an initial relaxation phase of the same length) with $N_{rep} = 20$ using each of the datasets at the three denaturant concentrations.

(67) Choy, W. Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101−112.
(68) Feldman, H. J.; Hogue, C. W. *Proteins* **2000**, *39*, 112−131.

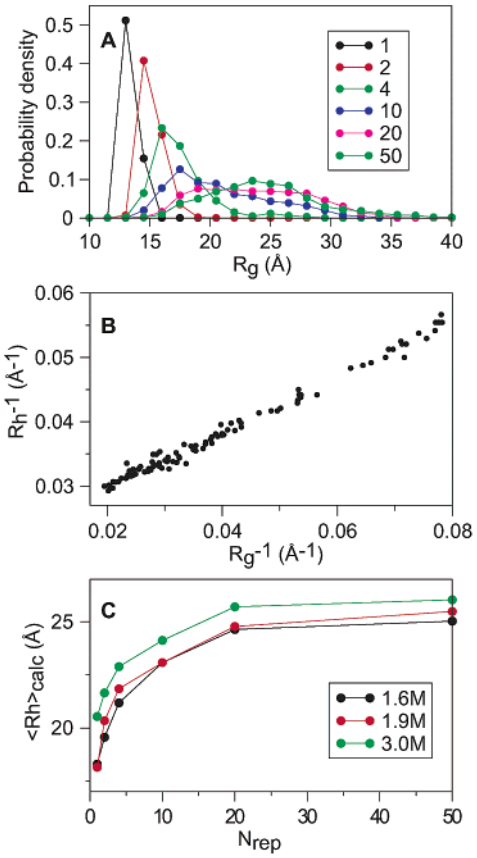

Figure 4. Comparison of calculated and experimental estimates of the average molecular dimensions of denatured ACBP. (A) Distribution of R_g in simulations using 1.6 M GuHCl data and with 1−50 replicas. (B) An empirical linear relationship between R_g^{-1} and R_h^{-1}. Each point on the plot corresponds to a different all-atom ACBP conformation for which we calculated R_g from the position of the C_α-atoms and estimated R_h using HYDROPRO.[58] C: R_h^{calc} as a function of N_{rep} at three denaturant concentrations.

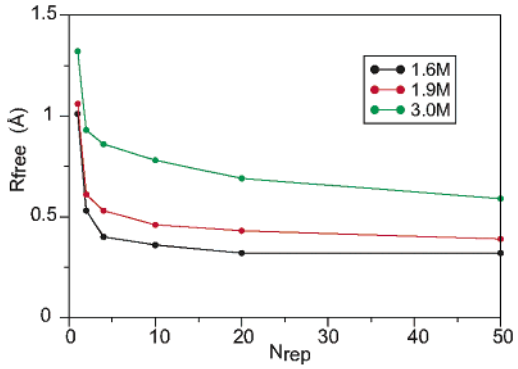

Figure 5. Cross-validation of the experimentally derived distance restraints. R_{free} was calculated as described in the text. Use of multiple replicas up to about 20 significantly improves the prediction of experimentally determined distances used in the cross-validation.

We chose $N_{rep} = 20$ as a compromise between having sufficient replicas to sample expanded conformations (see above) and otherwise minimizing the computational requirements. In Figure 6, we show a set of structures selected from the 1.6 M denatured state ensemble to represent the wide range of compactness observed. To highlight the considerable volume taken up by side-chain and backbone atoms, we have, for this figure, reconstructed all-atom models of the denatured state of ACBP

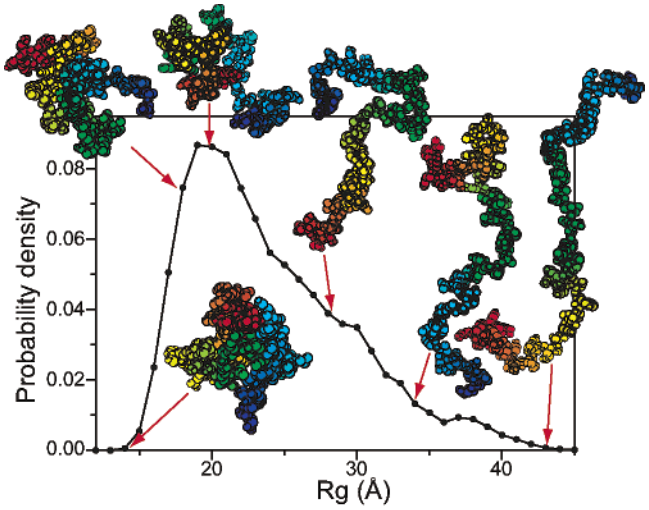

Figure 6. Examples of structures from the denatured state of ACBP in 1.6 M GuHCl. Structures are colored from their N (blue) to C (red) terminus. The radius of gyration of each conformation is indicated by arrows on the R_g histogram obtained from 20 000 conformations from the 1.6 M GuHCl denatured state ensemble. The relative ruggedness of this curve at high values of R_g is presumably due to incomplete sampling of these very rare and highly expanded conformations. We estimate the error in the probability density at $R_g > 35$ Å to be ∼10−20%. The structures were prepared using Bobscript.[70]

from the C_α atom coordinates (see Methods). The figure clearly shows that the ensemble contains a very broad range of conformations and that many different interactions are formed in the denatured state.

Detailed examination of the structures in the calculated denatured state ensembles shows that there are only weak preferences for forming particular interactions. Therefore, we determined the residual structure in the denatured state. The residual structure is defined by the structural features that are more dominant than in a specified reference state representing a fully random polypeptide chain. To determine the latter, we performed a long simulation (6×10^8 Monte Carlo steps) of the C_α chain without distance restraints but with retention of the excluded volume associated with the hard-sphere radii of the pseudoatoms. Using this state as a reference, we calculated the relative contact probabilities of each pair of residues. In Figure 7B−D, we show the result of these calculations as maps whose entries are $-\ln(p_{ij}/p_{ij}^{ref})$, where p_{ij} is the contact probability from the replica simulations and p_{ij}^{ref} is from the reference state simulation. To minimize artifacts arising from the fact that the five spin-label sites have many more distance restraints than the other residues, we averaged the map locally over a window of ±3 residues. Additionally, in Figure 7A, we show an "energy map"[57] representing the native state ensemble of ACBP. In this map, the entries are the EEF1[56] interaction energies between each pair of residues averaged over the native state ensemble.

Using the unbiased simulation as a reference in this way, we can detect a significant degree of long-range residual structure in the denatured state of ACBP at each concentration of denaturant. In particular, residues in the regions spanned approximately by 15−35 and 60−86 show in each case a nonrandom tendency to interact; these two regions correspond approximately to helices A_2 and A_4 in the native state structure. A more detailed comparison with the "energy map" shows that

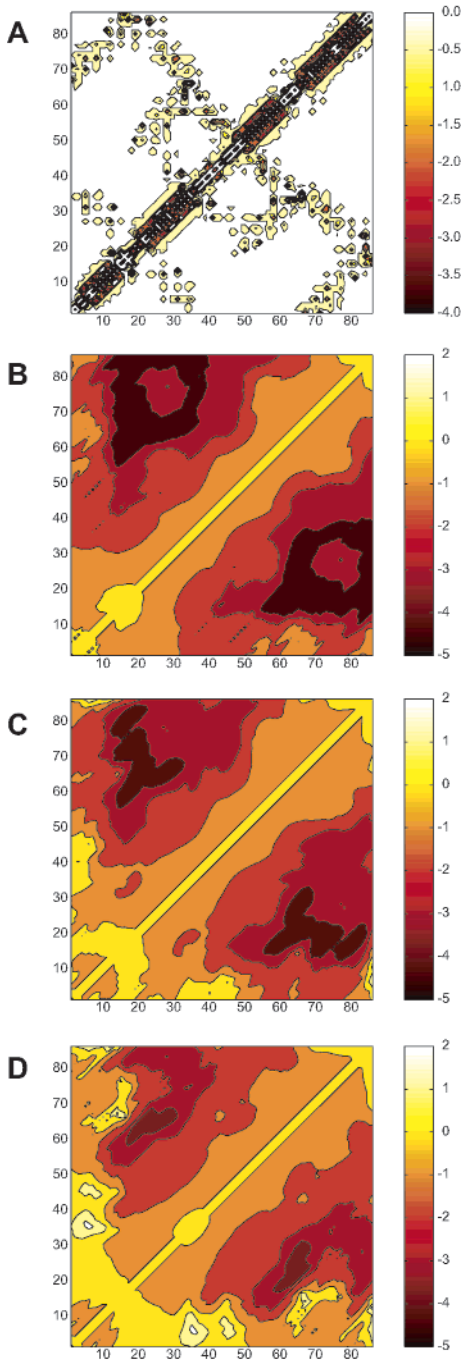

Figure 7. Residual structure in the denatured state contains both native and non-native elements. In (A), we show an "energy map"[57] of the native state of ACBP. Entries are the calculated pairwise energies between residues, averaged across the native state ensemble. The colorbar to the right indicates the energy scale (in kcal/mol) with more negative energies representing stronger contacts. Residual structures in the denatured state at 1.6, 1.9, and 3.0 M GuHCl are shown in (B), (C), and (D), respectively. The entries in these plots are $-\ln(p_{ij}/p_{ij}^{\mathrm{ref}})$. The colorbars to the right indicate the scale. Negative entries indicate contacts that are more persistent in the ACBP denatured state than in a random coil model.

the residual structure involves both native and non-native interactions. Finally, we observe that the extent of residual structure in the denatured state increases as the denaturant concentration is decreased, with observable differences even between 1.6 and 1.9 M GuHCl and even larger changes in 3.0 M GuHCl. The most persistent interactions appear to be those involving residues 20−30 and 60−65, as these interactions are formed at all three denaturant concentrations with approximately equal probability.

MD simulations have previously been used in combination with NMR experiments to obtain structural information on denatured states of proteins.[69] MD simulations at the experimentally relevant temperature, and with an atomistic description of solvent and denaturant, are extremely demanding of computational time; such simulations are therefore usually carried out at high temperatures to increase the rate at which different conformations are sampled. We performed a series of MD simulations at 400, 500, 600, and 700 K to sample unfolded conformations. Analysis of each of the ensembles showed that more expanded conformations are sampled at the higher temperatures. By comparison with the experimentally determined R_h values, we chose the MD simulation at 600 K ($R_h^{\mathrm{calc}} = 25.4$ Å, calculated from R_g as described above) for further examination. We calculated the r^{-6} averaged distances from the 600 K MD simulation and compared the results to each set of PRE derived distance restraints by calculation of *R*-scores. We find the best match to the experimental dataset measured in 1.6 M GuHCl, for which we obtain $R = 0.63$ Å. For this dataset, we find 22 distance violations (taking error bounds into account); only seven of these are, however, larger than 1 Å, and none is larger than 5 Å. Inspection of these violations reveals that, although the MD sampling underestimates the interaction between residues 20−30 and the C-terminal region, the distributions of pairwise distances from the MD simulation in general agree with those obtained from MCRES. Together, these results show that high-temperature MD simulations, validated by comparison with the experimental R_h value, are complementary to the MCRES analysis and reinforce the conclusions drawn from the latter about the overall distribution of structures in the denatured state of ACBP.

Discussion

We have used experimental PRE data as restraints in a Monte Carlo sampling procedure (MCRES) to determine ensembles of structures representing the denatured state of ACBP at three different GuHCl concentrations. The approach takes into account the fact that the experimental results provide interresidue distances averaged over a large ensemble of heterogeneous conformers. By simulating noninteracting replicas in parallel, we have shown that MCRES allows for efficient sampling of such broad ensembles of structures.

The experimental PRE data described here are averages over a large number (2×10^{17}) of molecules. As ensembles of this size cannot be simulated, it is important to demonstrate that averages taken over a much smaller number of molecules are sufficient to capture the broad character of the denatured state ensemble. We have used two complementary methods to evaluate the effect of the number (N_{rep}) of replicas on the properties of the denatured state of ACBP obtained from the calculations. First, we compared the molecular dimensions, measured experimentally by the hydrodynamic radius R_h, to the predictions from PRE restrained simulations and found a very good agreement when 20 replicas or more are used. Second,

(69) Wong, K. B.; Clarke, J.; Bond, C. J.; Neira, J. L.; Freund, S. M.; Fersht, A. R.; Daggett, V. *J. Mol. Biol.* **2000**, *296*, 1257−1282.
(70) Esnouf, R. M. *J. Mol. Graphics Modell.* **1997**, *15*, 132−134.

full cross-validation of the experimental restraints shows that multiple-replica sampling of the denatured state significantly increases the quality of the structural ensembles. Taken together, these results show that the ensembles that we have determined provide a self-consistent representation of the structural features in the denatured state of ACBP. We note that our result that 20 replicas are needed in the case of ACBP may not generalize to other systems. We expect that the optimal value of N_{rep} can be affected by several factors including the broadness of the ensemble, the number of restraints, and the tightness with which these restraints are employed. However, as the denatured state of ACBP constitutes a very broad ensemble of structures, we expect that 20 replicas will be sufficient in most cases. Further, the methods that we present to determine an optimal value of N_{rep} are general and should be applied when possible.

The calculations reveal a very broad ensemble of structures whose compactness values range from that of the native state to almost fully extended conformations (Figure 6). At the same time, the MCRES method is extremely sensitive and can be used to determine even small preferences for specific types of nonrandom structure. When the structural ensembles are compared to a random-coil model with excluded-volume, we find clear evidence for long-range, residual structure in ACBP. In particular, we observe interactions between the two regions consisting of residues 15−35 and 60−86. Interestingly, this finding is consistent with a recent observation that the mutation Ile27Ala causes significant changes in residual dipolar couplings in the C-terminal region of ACBP (W.F. and F.M.P., submitted).

The hydrophobicity profile and predicted propensities for formation of helices and turns in ACBP have previously been described[21] and provide an explanation for the existence of the long-range residual structure that we observe. The segment 20−30 is the most hydrophobic region of the ACBP molecule. Furthermore, there is a high helical propensity and intermediate hydrophobicity in the C-terminal region corresponding to the amphipathic helix A_4 in the native state (residues 65−84). Between these two regions is a stretch of amino acids that display a high propensity for formation of turns. We therefore suggest that the long-range residual structure is caused by interactions between the region of high hydrophobicity (20−30) and high helical propensity (60−80). The high propensity for turn formation in the intervening sequence may in addition disrupt persistent contacts involving these residues. Finally, the N-terminal residues display both low helical propensity and low hydrophobicity, presumably causing these regions of the sequence to have little residual structure.

By comparing the ensembles obtained at three different denaturant concentrations, we observe changes in the degree of residual structure. This result emphasizes an important issue in studies of the denatured state. On one hand, one must stabilize the denatured state relative to the native state to populate it sufficiently for its characterization, but, on the other hand, one will thereby perturb its structure. This observation applies not only to stabilization by denaturant but also by heat, pH, or mutation. One advantage of using a denaturant is that its concentration can be varied continuously. Further, the effects of denaturants are expected to be global and not through one or a few specific amino acid residues. From our calculations, we observe long-range interactions in ACBP at all three denaturant concentrations, although these are less pronounced in 3.0 M GuHCl. These results suggest strongly that these interactions are present, and probably even more persistent, under native conditions. In particular, interactions between residues in the vicinity of 20−30 and 60−65 are approximately equally prevalent at all three GuHCl concentrations. Furthermore, recent experimental results on the acid denatured state of ACBP have revealed that this state is structurally quite similar to the denaturant-induced unfolded state (S.K., W.F., and F.M.P., unpublished data). This finding suggests that the structural features that we observe are not specific to the denaturant-stabilized denatured state.

The propensity to form persistent structure in the denatured state of a protein may affect the events leading to the formation of the native state. The appearance of structure during the folding of ACBP has been examined using a range of biophysical experiments. The early stages of ACBP folding have been studied using quenched-flow hydrogen-exchange[45] and rapid-mixing techniques.[46] The results of these experiments show that, after dilution of denaturant, ACBP rapidly forms a partially collapsed kinetic intermediate characterized by the partial protection of amide hydrogens in helix A_4 against solvent exchange. The subsequent formation of the rate-limiting transition state ensemble has been investigated by the protein engineering method.[44] The results of this analysis show that residues from helices A_1 and A_4 interact in the transition state and form additional hydrophobic contacts with residues in helix A_2.

Taken together, these experiments suggest that the folding of ACBP involves the initial formation of helical structure in A_4 and the subsequent formation of long-range interactions between helices A_1, A_2, and A_4 in the transition state. The presence of a long-range residual structure in the denatured state shows that A_2 and A_4 have intrinsic propensity to interact and stabilize each other. The existence of this structure prior to the initiation of the folding reaction may partially compensate for the entropic cost of bringing the N- and C-terminal residues together in the transition state. Finally, the observation that the structural features of the denatured state are more stable at lower concentrations of denaturant suggests that the collapsed intermediate state formed early in ACBP folding may correspond to the dominant conformations of the equilibrium denatured state under native conditions.

Conclusions

Advances in experimental techniques have made it possible to obtain detailed information about biologically relevant nonnative conformations of proteins. Such states include natively unfolded states, molten globules, and other partially unfolded conformations that may act as precursors for toxic protein aggregation. The increased interest in such states has resulted in a series of detailed biophysical characterizations. However, due to the heterogeneous nature of these states, such information has been difficult to convert into structural data. In this work, we have shown that it is possible to use experimental NMR data, here from paramagnetic relaxation enhancement, as restraints in procedures that can result in the determination of the ensemble of conformations corresponding to the denatured state, by simulating in parallel a relatively small number of molecular models. We anticipate that the extension of this approach to other types of NMR data, such as NOEs, chemical

shifts, and residual dipolar couplings, will permit the characterization of the structures of several non-native states of proteins including the proteins that appear to contain intrinsically disordered regions even under native conditions.

Acknowledgment. K.L.-L. is supported by the Danish Research Agency. W.F. is the recipient of an Erwin-Schrödinger Fellowship (J2229-B07) from the Austrian Science Foundation (FWF). The research of C.M.D. is supported in part by a Program Grant from the Wellcome Trust. F.M.P. is supported by the John and Birthe Meyer Foundation. M.V. is a Royal Society University Research Fellow. Research in Cambridge is supported in part by a Program Grant of the Leverhulme Trust.

Supporting Information Available: Distance restraints obtained from PRE experiments in 1.6, 1.9, and 3.0 M GuHCl (TXT). This material is available free of charge via the Internet at http://pubs.acs.org.

JA039250G