Characterization of the unfolded state of bovine α -lactalbumin and comparison with unfolded states of homologous proteins

JULIA WIRMER,^{1,2} HOLGER BERK,¹ RAFFAELLA UGOLINI,³ CHRISTINA REDFIELD,³
AND HARALD SCHWALBE¹

¹Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University Frankfurt, D-60439 Frankfurt, Germany

²Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

³Department of Biochemistry and Chemistry Research Laboratory, University of Oxford, Oxford OX1 3TA, United Kingdom

(RECEIVED November 11, 2005; FINAL REVISION February 21, 2006; ACCEPTED February 25, 2006)

Abstract

The unfolded states of three homologous proteins with a very similar fold have been investigated by heteronuclear NMR spectroscopy. Secondary structure propensities as derived from interpretation of chemical shifts and motional restrictions as evidenced by heteronuclear ^{15}N relaxation rates have been analyzed in the reduced unfolded states of hen lysozyme and the calcium-binding proteins bovine α -lactalbumin and human α -lactalbumin. For all three proteins, significant deviations from random-coil predictions can be identified; in addition, the unfolded states also differ from each other, despite the fact that they possess very similar structures in their native states. Deviations from random-coil motional properties are observed in the α - and the β -domain in bovine α -lactalbumin and lysozyme, while only regions within the α -domain deviate in human α -lactalbumin. The motional restrictions and residual secondary structure are determined both by the amino acid sequence of the protein and by residual long-range interactions. Even a conservative single point mutation from I to L in a highly conserved region between the two α -lactalbumins results in considerable differences in the motional properties. Given the differences in oxidative folding between hen lysozyme and α -lactalbumin, the results obtained on the unfolded states suggest that residual long-range interactions, i.e., those between the α - and the β -domain of lysozyme, may act as nucleation sites for protein folding, while this property of residual structure is replaced by the calcium-binding site between the domains in α -lactalbumin.

Keywords: NMR spectroscopy; unfolded proteins; protein folding; α -lactalbumin; lysozyme

Supplemental material: see www.proteinscience.org

A complete definition of protein folding requires the characterization of all states that a polypeptide chain can adopt as well as a description of the kinetics of protein

folding. The folded, native state and the kinetics of folding are well characterized for a large number of proteins. In contrast, the structural and dynamic properties of nonnative states such as molten globule or unfolded states of proteins are not as well understood. However, one key step toward the understanding of protein folding is the investigation of the unfolded state of proteins, which represents the starting point of protein folding. Consequently, the interest in nonnative states of proteins has increased recently, and the properties of a number of unfolded proteins have been investigated by

Reprint requests to: Harald Schwalbe, Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University Frankfurt, Marie-Curie-Str. 11, D-60439 Frankfurt, Germany; e-mail: schwalbe@nmr.uni-frankfurt.de; fax: ++49-69-29515.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051974506>.

using biophysical methods, including NMR spectroscopy. The properties of an ideal unfolded state, an ensemble of rapidly interconverting conformers, can be predicted by the so-called random coil model, which describes the protein as a polymer consisting of 20 different monomers. The structure and dynamics of the polypeptide chain can be described remarkably well based on the hypothesis that the polymer chain possesses few structural preferences except those inherent to its monomers (Fiebig et al. 1996; Smith et al. 1996a, b; Schwalbe et al. 1997).

While the overall properties of unfolded proteins are very similar to the random coil state, residual secondary and even tertiary structure are observed in nonnative states of proteins. Such residual secondary structure has been detected in a large number of proteins, including lysozyme (Schwalbe et al. 1997; Wirmer et al. 2004), apomyoglobin (Yao et al. 2001), barnase (Arcus et al. 1995), protein G (Frank et al. 1995; Sari et al. 2000), acyl-CoA-binding protein (ACBP) (Fieber et al. 2004), SH3 of drk (Blanco et al. 1998), and α -synuclein (Bussell and Eliezer 2001; Bertonecini et al. 2005). Interestingly, the observed residual secondary structure is mostly α -helical and often coincides with hydrophobic residues. A correlation between residual α -helical structure in unfolded states of some proteins, including apomyoglobin (Yao et al. 2001), and the positions of secondary structure elements in the folded state of proteins has been observed, suggesting possible nucleation sites for the initiation of protein folding. In addition, residual long-range interactions are a feature observed in unfolded proteins, including apomyoglobin (Lietzow et al. 2002), staphylococcal nuclease (Gillespie and Shortle 1997a,b), protein L (Yi et al. 2000), ACBP (Lindorff-Larsen et al. 2004), and lysozyme (Wirmer et al. 2004). While it has been shown that nonconservative single point mutations can dramatically change the ensemble of conformations present in the unfolded state and even provoke a shift from a compact ensemble to an extended state (Wirmer et al. 2004), unfolded states of closely related proteins have not yet been investigated. Such studies may allow delineation of whether the dynamics are governed by local factors encoded in the primary sequence or by more global factors. In the latter case, the question of whether the interactions in the unfolded state are driven by native or by nonnative transient interactions can be assessed.

In this report, the homologous proteins hen egg-white lysozyme (HEWL), human α -lactalbumin (HLA), and bovine α -lactalbumin (BLA) have been investigated. They have served as model systems for the study of protein folding (McKenzie and White 1991; Dobson et al. 1994; Forge et al. 1999). HEWL and both α -lactalbumins (α -LAs) share $\sim 40\%$ sequence identity and $\sim 60\%$ sequence similarity. The native structures of these proteins superimpose very well (backbone root mean square deviation [RMSD] = 1.27 Å for residues 1–116 in HEWL, corresponding to

residues 1–112 in both α -LAs), yet the proteins have entirely different functions. The Ca^{2+} -binding protein α -LA ($K_D = 2 \times 10^{-7}$ M at 37°C) is a subunit of lactose synthase (Hill and Brew 1975; Permyakov and Berliner 2000), which catalyzes the final step in lactose biosynthesis (formation of a $\beta(1 \rightarrow 4)$ glucopyranosyl linkage). Lysozyme destroys cell walls by cleavage of a $\beta(1 \rightarrow 4)$ glucopyranosyl linkage (McKenzie and White 1991) and is not a metalloprotein.

While the structure and function of HEWL, HLA, and BLA in the native state are well understood, their unfolded states are less well characterized. Differences in the oxidative folding of HEWL and the α -LAs have been reported. However, these differences are not yet completely understood (van den Berg et al. 1999; Chang and Li 2002; Chang 2004; Li and Chang 2004). One step toward the understanding of oxidative refolding is the investigation of the unfolded state of proteins in the absence of disulfide bridges (henceforth called unbranched). Here, the unbranched (disulfide reduced) states of the three proteins are compared. While unbranched HEWL (HEWL-S^{ME}) and unbranched HLA (all-Ala-HLA) have already been studied to some extent using NMR spectroscopy and heteronuclear resonance assignments are available for both (Schwalbe et al. 1997; Redfield et al. 1999; Klein-Seetharaman et al. 2002; Wirmer et al. 2004), unbranched BLA (BLA-S^{ME}) has not been studied previously. Therefore, BLA has been expressed in isotope labeled form and the ^1H , ^{13}C , ^{15}N resonances of BLA-S^{ME} have been assigned (see Supplemental Material).

Results and Discussion

Sequence homology and hydrophobicity predictions for unbranched hen lysozyme, bovine, and human α -LA

Comparison of unbranched HEWL, BLA, and HLA requires the absence of disulfide bridges in these proteins. Unbranched HEWL (HEWL-S^{ME}) and BLA (BLA-S^{ME}) were prepared by reduction and methylation of their native disulfide bridges (see Materials and Methods). Unbranched HLA was obtained by mutation of all cysteines to alanines (all-Ala-HLA) (Peng et al. 1995; Redfield et al. 1999). HEWL-S^{ME} is found to be unfolded in water (pH 2–6) and in the presence of 8 M urea at pH 2, as judged by one-dimensional ^1H NMR spectroscopy and CD measurements. By contrast, BLA-S^{ME} and all-Ala-HLA have characteristics of partially folded molten globule states in the absence of urea (Peng et al. 1995; Redfield et al. 1999); far UV CD spectra show the presence of helical secondary structure, while near UV CD spectra show no evidence of fixed tertiary interactions. In addition, the majority of resonances in NMR spectra collected in the absence of urea are broadened beyond detection due to conformational

fluctuations on a millisecond to microsecond timescale. The addition of urea to the molten globule of all-Ala-HLA results in the disruption of persistent structure and leads to a step-wise, noncooperative unfolding, which results in the appearance of sharp NMR signals (Redfield et al. 1999); BLA-S^{ME} and all-Ala-HLA are completely unfolded in 8 M urea at pH 2.

Sequence alignments of human and bovine α -LA and HEWL are shown in Figure 1. The sequence identity between HEWL and both α -LAs is $\sim 40\%$, while the sequence similarity is $\sim 60\%$. Sequence identity between HLA and BLA is 76%, while the similarity is 89%. In addition, the eight cysteine residues that form four disulfide bridges in the native state of the proteins align.

Hydrophobic regions of the sequences of BLA, HLA, and HEWL align, although there are differences if hydrophobicities are analyzed in more detail. Figure 1B shows hydrophobicity predictions using the scale of Abraham

and Leo (1987). Six regions of elevated hydrophobicity can be identified in HEWL, numbered 1–6. Eight areas of elevated hydrophobicity labeled A–H can be identified in the α -LAs. The hydrophobicity profiles of BLA and HLA are almost identical, with the exception of region B, where higher hydrophobicities are found in HLA. The patterns observed for HEWL, BLA, and HLA are very similar, although hydrophobicities in BLA and HLA are on average slightly higher than in HEWL. The most pronounced differences are found in the regions from T38 to N45 in BLA/HLA and T40 to T47 in HEWL; while increased hydrophobicity is observed in BLA and HLA (region C), rather polar residues are found in the corresponding part of the sequence in HEWL. Hydrophobicities correspond very closely for all three proteins in the region labeled 3/D from D52 to R61 in HEWL and E49 to K58 in BLA/HLA ($R^2 = 0.988$ between HLA and HEWL).

Figure 1. (A) Sequence alignment for HEWL, BLA, and HLA. Alignment of HEWL and BLA results in 111 residues overlap (5–115) and indicates 42% sequence identity (47/111) and 59% sequence similarity (65/111). Similarly, 39% sequence identity (45/115) and 59% sequence similarity (68/115) are found for the alignment of HEWL and HLA (data not shown). All residues overlap in the alignment of BLA and HLA: 76% sequence identity (93/123) and 89% sequence similarity (109/123) are found. (B) Hydrophobicity predictions using the scale of Abraham and Leo (1987). Hydrophobicities for HEWL (red circles), HLA (black circles), and BLA (green triangles). Residue numbering is according to the HEWL sequence, and HLA and BLA were aligned according to the sequence alignment. Areas of increased hydrophobicity in HEWL are indicated by numbers (1–6), while in BLA and HLA they are indicated by letters (A–H); area F is indicated not due to increased hydrophobicities but for discussion in the text. Positions of secondary structure elements in lysozyme are indicated at the top of the figure (closed circles, α -helices; open circles, 3^{10} -helices; square, β -sheets). The C-to-A mutations in HLA to obtain the unbranched protein do not change the hydrophobicity profile ($R^2 = 0.9991$).

Differences in chemical shift deviations from random coil values

Figure 2 shows a comparison of perturbations from the chemical shifts measured in small random coil peptides (Wishart et al. 1995a) for BLA-S^{ME} in 8 M urea, all-Ala-HLA in 8 M urea, and HEWL-S^{ME} in 8 M urea and in water. H^α and H^N chemical shift perturbations are shown

here, since it was previously shown for lysozyme that these perturbations are the most sensitive for monitoring deviations from the random coil (Hennig et al. 1999) and can be monitored for all three proteins investigated here. The majority of residues in the proteins fall within the range of negligible perturbations from the expected random coil chemical shift values. Significant negative deviations, which may be indicative of residual helical structure, are found in all proteins

Figure 2. Deviations of H^α (left) and H^N (right; $\Delta\delta_{H^\alpha}$ and $\Delta\delta_{H^N}$) experimental chemical shifts (δ^{ex}) from random coil chemical shifts (δ^{rc} , $\Delta\delta = \delta^{ex} - \delta^{rc}$) (Wishart et al. 1995a). (I) BLA-S^{ME} in urea. (II) All-Ala-α-LA in urea. (III) HEWL-S^{ME} in urea (reproduced from Schwalbe et al. 1997). (IV) HEWL-S^{ME} in water (reproduced from Klein-Seetharaman et al. 2002). Areas that show only small chemical shift perturbations are shaded in gray ($\Delta\delta = \pm 0.1$ ppm for H^α and $\Delta\delta = \pm 0.2$ ppm for H^N) (Wishart et al. 1992; Wishart and Sykes 1994). Residues that show significant deviations are labeled.

for H^N of G20^{BLA/HLA} (G22^{HEWL}) and for the H^N and H^α chemical shifts in the regions around W104/L105^{BLA/HLA} (W108/W111^{HEWL}) and W118/L119^{HLA} (W123^{HEWL}). Furthermore, significant deviations around W62/W63 are found in HEWL-S^{ME} both in water and in urea. These coincide with the small deviations found in BLA-S^{ME} at K62 (H^α). In contrast to the other proteins, additional chemical shift perturbations around W26 are observed in BLA-S^{ME}.

Differences of R_2 relaxation rates

The R_2 relaxation rates measured in BLA-S^{ME}, all-Ala-HLA, and HEWL-S^{ME} are shown in Figure 3. A full set of ^{15}N relaxation parameters has been published previously for HEWL-S^{ME} (Schwalbe et al. 1997). These and other data in the literature (Logan et al. 1994; Bussell and Eliezer 2001; Schwarzing et al. 2002) indicate that R_1 relaxation rates and the heteronuclear NOE vary only slightly in unfolded states of proteins, while R_2 relaxation rates show significant deviations. The deviations from the random coil relaxation rates (R_2^{rc}) differ considerably for the three proteins and can be identified by Gaussian fitting (Table S1 of Supplemental Material). Eight clusters of deviations from R_2^{rc} are found in BLA-S^{ME}, while only five clusters were identified in all-Ala-HLA. Six

clusters have been identified previously in HEWL-S^{ME} (Schwalbe et al. 1997; Wirmer et al. 2004).

Clusters identified by Gaussian fitting in BLA-S^{ME} are centered at E11 (A), W26 (B), S47 (C), W60 (D), D82 (E), L96 (F), W104 (G), and Q117 (H). The highest relaxation rates are observed in clusters B, D, F, and G ($R_2(W26) = 6.86 \text{ sec}^{-1}$, $R_2(W60) = 6.87 \text{ sec}^{-1}$, $R_2(L96) = 7.26 \text{ sec}^{-1}$, and $R_2(L105) = 6.88 \text{ sec}^{-1}$). The five clusters in all-Ala-HLA are centered around S9 (A), I27 (B), L96 (F), W104 (G), and W118 (H). The highest relaxation rates are measured in cluster B with $R_2(A28) = 7.56 \text{ sec}^{-1}$ (C28 in wild-type HLA). High R_2 rates are also found in cluster A ($R_2(S9) = 5.86 \text{ sec}^{-1}$) and in cluster G ($R_2(W104) = 6.65 \text{ sec}^{-1}$ and $R_2(L105) = 6.67 \text{ sec}^{-1}$). Six clusters were identified previously (Klein-Seetharaman et al. 2002; Wirmer et al. 2004) in HEWL-S^{ME} in water centered at L8 (1), C30 (2), S60 (3), S85 (4), W111 (5), and W123 (6). While the location of the clusters is the same in water and in 8 M urea, the clusters are generally more pronounced for HEWL-S^{ME} in water. The highest R_2 rates are found in cluster 3; however, they differ with $R_2(S60) = 7.97 \text{ sec}^{-1}$ in water and $R_2(C64) = 6.34 \text{ sec}^{-1}$ in urea. Another significant difference is found in cluster 2 in HEWL; this cluster is very pronounced in water but rather small in urea. Both α -LA and lysozyme show differences in dynamics in the absence and presence of urea. For α -LA, the difference is particularly dramatic; in the absence of urea most residues have very high relaxation rates, leading to peaks that are broadened beyond detection. In 8 M urea, broadening in α -LA is restricted to a small number of residues located in the clusters shown in Figure 3. For HEWL-S^{ME}, line broadening in the absence of urea is much less pronounced than observed for α -LA, but this broadening is observed to decrease upon the addition of urea. Thus, both α -LA and lysozyme show an increase in dynamics, as reflected by the decrease of deviations from R_2^{rc} , upon the addition of urea that reflects the disruption of temporary contacts that are more stable in water.

The clusters identified in the α -domain of HEWL-S^{ME} (1, 2, 5, and 6) are in similar regions of the sequence to clusters A, B, G, and H in BLA-S^{ME} and all-Ala-HLA. However, an additional α -domain cluster, cluster F, is present in both α -LAs. Striking differences between the three proteins are found in the β -domain: Cluster 3 in HEWL-S^{ME} has its counterpart in BLA-S^{ME} (cluster D) but not in all-Ala-HLA. Moreover, a very flat cluster around S47 (cluster C) and a cluster around residue D82 (cluster E) can be identified in BLA-S^{ME} but not in all-Ala-HLA or in HEWL-S^{ME}.

A more detailed analysis of the sequence within the corresponding clusters in the α -domain reveals that while the positions of the clusters correspond to each other, the sequence differs slightly in all cases except for cluster 5/G: In both α -LAs the sequence ¹⁰³YWLAH¹⁰⁷ is

Figure 3. ^{15}N transverse relaxation rates (R_2) as a function of residue: (I) BLA-S^{ME} in urea, (II) All-Ala-HLA in urea, (III) HEWL-S^{ME} in urea (reproduced from Schwalbe et al. 1997), and (IV) HEWL-S^{ME} in water (reproduced from Klein-Seetharaman et al. 2002). The relaxation rates expected for a random coil (R_2^{rc}) and the Gaussian fit (Supplemental Table S1) of experimentally determined relaxation rates are shown as black lines. Numbering of the deviations from R_2^{rc} corresponds to the numbering in Figure 1. Gray circles indicate the position of methylated cysteines (I, III, and IV) or cysteine to alanine mutations (II).

found corresponding to ¹⁰⁷AWVAW¹¹¹ in HEWL. Overall, mainly uncharged residues at pH 2 are found in the regions containing the clusters in the three proteins. The differences in other clusters are less obvious from the primary sequence: Cluster F in both α-LAs has no counterpart in HEWL but has very similar sequences: ⁹¹CAKKILD⁹⁷ in HLA and ⁹⁴CAKKIVSD¹⁰¹ in HEWL, while cluster 1/A is present in all three proteins but has very different amino acid primary sequences: ⁸LSQLL¹² in HLA, ⁸VFREL¹² in BLA, and ⁸LAAAM¹² in HEWL.

The unfolded state of proteins is determined by the primary sequence and long-range interactions

The results presented above, summarized in Table 1, can be used to highlight the similarities and differences between the unfolded states of the three proteins.

Backbone motional restrictions, as monitored by R₂ relaxation rates, coincide in all three proteins with hydrophobic regions of the sequence. Nonrandom secondary structural propensities, as monitored by H^N and H^α chemical shift deviations, generally occur within the clusters of hydrophobic residues, which show elevated R₂ values; the exception to this correlation is G20^{HLA/BLA}/G22^{HEWL} (see above). Interestingly, residual secondary structure is not found in all of the hydrophobic clusters identified by R₂ relaxation rates, but only in approximately half of them. It is thus not necessarily the residual secondary structure that leads to decreased flexibility in the unfolded state. Similarly, aromatic residues and specifically tryptophan residues occur within some of the hydrophobic clusters. Interestingly, while at least one aromatic residue is present in each of the major clusters in HEWL, this is not the case in the two α-LAs. Aromatic residues are absent in clusters A (all-Ala-HLA), B (all-Ala-HLA), and F (all-Ala-HLA, BLA-S^{ME}), all of which

are significant clusters. Furthermore, an aromatic residue is present at W60^{HLA}, but no cluster is observed in region D of all-Ala-HLA, as monitored by R₂ relaxation rates. Even more striking is the fact that the major difference between all-Ala-HLA and BLA-S^{ME} in region D is a conservative L^{HLA}- to-I^{BLA} replacement at residue 59.

These results clearly indicate that there are two determinants imposing motional restrictions to the unfolded polypeptide chain. These are (local) primary structure and long-range interactions. Previously, the presence of long-range interactions had already been revealed for HEWL-S^{ME} (Klein-Seetharaman et al. 2002; Wirmer et al. 2004). Nonconservative single point mutations in the unfolded state such as A9G, W62G, W62Y, W111G, and W123G disturb relaxation rates remote in sequence and thus indicate the presence of long-range interactions. Furthermore, the hydrodynamic radius (R_h) changes drastically upon mutation within the most prominent cluster. While the WT-HEWL-S^{ME} in water has a R_h of 26.9 Å, indicating a compact unfolded state, the W62G mutant has a R_h of 32.1 Å, corresponding to an extended conformation (Wirmer et al. 2004). Sequence comparison between the α-LAs and HEWL supports the hypothesis of both local and long-range structure in the unfolded state: e.g., while the sequence in cluster F is nearly identical in the three proteins (⁹¹CVKKILD⁹⁷ in BLA ⁹¹CAKKILD⁹⁷ in HLA and ⁹⁴CAKKIVSD¹⁰¹ in HEWL); the cluster is present in the α-LAs but not in HEWL-S^{ME}. Given the similarity of the sequences, it could be argued that the stabilization of the cluster in the α-LAs can be attributed to long-range interactions resulting in the deviations. However, on the other hand, subtle changes in the sequence such as ⁹⁵IL⁹⁶ in both α-LAs and ⁹⁸IV⁹⁹ in HEWL, where two β-branched amino acids are adjacent, could account for the difference. Clearly, the local primary sequence determines the absence or the presence of

Table 1. Comparison of the properties of the unfolded state of the three proteins

	Cluster 1/A				Cluster 2/B				Cluster C				Cluster 3/D			
	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar
All-Ala-HLA																W60
BLA-S ^{ME}				F9		E25 W26		W26						K62		W60
HEWL-S ^{ME}								W28						R61 W63		W62 W63
	Cluster 4/E				Cluster F				Cluster 5/G				Cluster 6/H			
	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar	Hyd	Sec	R ₂	Ar
All-Ala-HLA										Y103 W104 L105		Y103 W104		A120 E121		W118
BLA-S ^{ME}		K79								Y103 L105 A106		Y103 W104		Q117		W118
HEWL-S ^{ME}										V109 W111 R112		W108 W111		W123 R125		W123

See also Figures 1–3.
Hyd indicates increased hydrophobicity; Sec, residual secondary structure from chemical shift deviations; R₂, increased R₂ relaxation rates; and Ar, aromatic residues. Striped areas indicate moderate increases in hydrophobicity/R₂ relaxation rates, while gray areas depict significant increases in hydrophobicity/R₂ relaxation rates.

cluster D in all-Ala-HLA/BLA-S^{ME}, respectively. All clusters but cluster D correspond to each other in the two proteins; the presence of long-range interactions, therefore, can be excluded. It is thus the single amino acid replacement at residue 59, which induces cluster formation in BLA-S^{ME} and absence of any motional restriction in all-Ala-HLA.

Furthermore, the comparison reveals that the unfolded states of the three proteins differ considerably from each other and from random coil behavior, with respect to motional restrictions and nonrandom secondary structural propensities. As the native structures of the three proteins are very similar, the structural properties of the unfolded state of the proteins are determined by the primary amino acid sequence and different, presumably transient, long-range interactions and thus are independent of the native structure. One property of the primary amino acid sequence is the hydrophobicity: Areas that exhibit motional restrictions in the unfolded state often coincide with hydrophobic areas in the protein sequence and consequently also with areas with large side-chain volumes. However, this coincidence is by chance. From the data presented here, it is clear that the properties of the unfolded state are neither predetermined by the hydrophobicity (see above) or volumes of the respective amino acids in the primary sequence, as proposed earlier (Schwarzinger et al. 2002).

Rather, specific interactions between amino acids and even long-range interactions lead to deviations from the expected random coil behavior. Even the smallest differences in the sequence, such as a conservative single point mutation of an isoleucine residue to a leucine, dramatically change the motional restrictions in the unfolded state. The combination of amino acid I59W60 in BLA-S^{ME} imposes motional restrictions on the region of cluster D, whereas no cluster is present in all-Ala-HLA where the combination L59W60 is found.

Relevance of hydrophobic clusters

The unfolded states of closely related proteins may differ considerably. Such differences in reduced unfolded states could have implications for the oxidative refolding of the respective proteins. Figure 4 shows the deviations of fitted R_2 relaxation rates in the unfolded state from the expected random coil relaxation rates mapped onto the native structures of BLA, HLA, and HEWL. Clusters in BLA-S^{ME} are found throughout the protein, while clusters in all-Ala-HLA are only located in the α -domain of the protein. Similarly to BLA-S^{ME}, clusters in HEWL-S^{ME} are found in both domains of the proteins. Comparison of the regions exhibiting deviations of R_2 relaxation rates from R_2^{rc} in HEWL-S^{ME} in 8 M urea at pH 2 (Fig. 4,

Figure 4. Deviations of fitted R_2 relaxation rates from the relaxation rates expected in a random coil mapped onto the native structure: (I) BLA (Protein Data Bank [PDB] file 1HFZ; Pike et al. 1996), (II) HLA (PDB file 1A4V; Chandra et al. 1998), and (III,IV) HEWL (PDB file 193L; Vaney et al. 1996). No deviations are indicated in red, while 100% deviations are colored in blue; 100% deviation is defined independently for each of the proteins/conditions. In the case of BLA-S^{ME}, the highest deviation was scaled according to cluster G. The picture was prepared by using the program InsightII (Accelrys).

panel III; Schwalbe et al. 1997) and in water at pH 2 (Fig. 4, panel IV; Klein-Seetharaman et al. 2002; Wirmer et al. 2004) reveals that these regions overlap. Furthermore, the extent of nonrandom secondary structural propensities, as determined from chemical shift perturbations, is nearly identical (Fig. 2). Therefore, the general features of the denatured states appear similar under refolding (pH 2) and under denaturing (pH 2, 8 M urea) conditions despite the large differences in the size of the clusters. Residual structure in the unfolded state under any conditions can, therefore, be considered as the starting point of oxidative refolding under native conditions.

Analysis of the hydrophobic clusters in HEWL-S^{ME} by a combination of nonconservative single point mutations and NMR techniques revealed long-range interactions between the clusters (Klein-Seetharaman et al. 2002; Wirmer et al. 2004). These long-range interactions are stabilized by cluster 3, as well as by interactions between the clusters in the α -domain of the protein. Oxidative refolding studies of HEWL revealed that three intermediates are populated during folding in which three out of the four native disulfide bridges are formed. The 30–115 disulfide bridge is the only disulfide formed in all of the three intermediates (van den Berg et al. 1999). The interactions found within the α -domain could account for the formation of the 30–115 disulfide bridge during oxidative folding, while interactions of the cluster in the β -domain (cluster 3) with clusters in the α -domain could be important for the formation of domain–domain interactions and structuring of the β -domain. Similarly, motional restrictions and residual secondary structure in both unfolded α -LAs in regions within the α -domain could account for the formation of the 28–111 disulfide bridge during oxidative folding of these proteins (Chang and Li 2002; Li and Chang 2004).

The situation is different with respect to the regions of α -LA that are in the β -domain of the folded protein: There are no motional restrictions in the unfolded state around W60 in all-Ala-HLA, while there is cluster D in BLA-S^{ME}. As protein folding of BLA and HLA is similar (Peng and Kim 1994; Peng et al. 1995; Wu et al. 1995; Chang and Li 2002; Chang 2004) it can be assumed that cluster D in BLA-S^{ME} has no influence on the folding of BLA. Consequently, there appear to be no long-range interactions between the β - and the α -domain in the unfolded states of α -LAs. This is in strong agreement with the observations in the molten globule state of α -LAs. The molten globule of α -LA is considered to resemble the predominant folding intermediate and can be stabilized under acidic conditions. While the α -domain possesses α -helical structure with a native-like topology, the β -domain of the protein is largely unstructured (Wu et al. 1995).

The absence of interactions between the domains in the unfolded state of α -LAs is a major difference to

HEWL-S^{ME}, where cluster 3 mediates long-range interactions between the two domains. Possibly the guiding role of cluster 3 in the oxidative refolding of HEWL is replaced in α -LAs by the Ca²⁺ binding site that connects the two domains in α -LA. The importance of the Ca²⁺ binding site in oxidative folding has been demonstrated by the fact that oxidative refolding is faster and much more efficient in the presence of Ca²⁺ than in its absence (Chang and Li 2002; Chang 2004). This hypothesis is further supported by the observed sequence conservation: Cluster 3 in HEWL-S^{ME} is stabilized by aromatic interactions in all known and investigated lysozyme variants by either a YW or a WW pair (Nitta and Sugai 1989; McKenzie and White 1991). The corresponding position in α -LA contains, in contrast, an XW or XF pair where X is not an aromatic residue but instead an aliphatic residue with X = L, I, M, G, N, and D (Nitta and Sugai 1989; McKenzie and White 1991). It can be argued that the conservation of aromatic residues at position 62 in lysozymes is not due to folding properties but due to its catalytic function and its importance in substrate recognition. However, it has been shown that the bacteriolytic efficiency of lysozyme mutants containing a nonaromatic hydrophobic amino acid at position 62 is equal or even increased compared with the wild type (Maenaka et al. 1994). It is, therefore, concluded here that while the presence of a WW or a YW pair in lysozyme is necessary to promote protein folding, it is not necessary in α -LA due to the presence of the Ca²⁺-binding site. Further insight into the importance of the cluster for protein folding could arise in the future from mutational studies within the three proteins such as an I-to-L change in the two α -LAs.

A comparison of the properties of the unfolded state of BLA-S^{ME} with all-Ala-HLA and HEWL-S^{ME} shows that unfolded states of proteins differ considerably for proteins that possess very similar structures in the native state. The properties of the unfolded state are solely determined by the primary structure of the protein and long-range interactions. Even the smallest conservative mutations can considerably change the structural ensemble present in the unfolded state. A closer inspection of the properties of the three unfolded proteins in the light of oxidative protein folding data revealed that residual structure and long-range interactions in unfolded states of proteins can act as nucleation sites for protein folding. This property of residual structure can be replaced by ion-binding sites.

Materials and methods

Heterologous overexpression, purification, and refolding of BLA

BLA without any additional amino acid at the N terminus was expressed by using a construct containing an N-terminal His-tag

with a trypsin cleavage site (MGHHHHHHTEGPIGPK). The construct (in pet11a) was purchased from Entelechon, and the plasmid was cloned into *Escherichia coli* BL21(DE3)-Gold cells (Stratagene). The protein was overexpressed in inclusion bodies either in LB medium or in ^{15}N or $^{13}\text{C}/^{15}\text{N}$ labeled minimal medium. After cell lysis and washing, the inclusion bodies could be solubilized in 8 M urea in the presence of reducing agents; further purification was achieved by using a Ni-NTA column. Almost quantitative oxidative refolding of the protein was performed in three steps, starting out under moderate denaturant conditions (4 M urea) and a glutathione redox shuffling system, and ending without any denaturing and redox agent. Enzymatic cleavage using trypsin yielded BLA, which could be further purified by ion-exchange chromatography. A yield of 7 mg/L BLA was obtained. Mass spectrometry, SDS-PAGE, Edman-degradation, and NMR spectroscopy revealed the correct identity of the purified protein. It could be verified that in contrast to protein conventionally expressed in *E. coli* containing an additional methionine residue at the N terminus, the stability against denaturants of the heterologously expressed BLA is identical to that of the protein isolated from milk.

NMR samples of BLA- S^{ME}

Methylation of BLA to yield BLA- S^{ME} was performed following the procedure of Henrikson (1971): 7 mg BLA were dissolved in 8 mL denaturing buffer (6 M GdnHCl, 25% [v/v] acetonitrile [AcN], 250 mM Tris, 5 mM EDTA at pH 8.5) and the solution was purged for 1 min with N_2 . Three- and-a-half microliters (3.5 μL) β -mercaptoethanol (~ 12 -fold molar excess of cysteine residues) was added, and the sample was kept for 1 h at 50°C to reduce all disulfide bridges. The sample was cooled down to 37°C before 150 μL of 0.6 M methyl-4-nitrobenzene sulfonate (MNBS) in AcN (at least 1.5-fold excess of the total amount of $-\text{SH}$ groups in the sample) were added. After 2 h at 37°C , the sample was dialyzed against water at pH 2 (five times against 5 L, 12 h each step). The samples were subsequently lyophilized and dissolved in NMR buffer: The BLA- S^{ME} NMR samples contained 0.5 mM (^{15}N) and 1.5 mM ($^{13}\text{C}, ^{15}\text{N}$) BLA- S^{ME} in 8 M urea (pH 2), 90% H_2O , and 10% D_2O . The spectra of the protein were concentration-independent.

NMR assignment measurements

All NMR assignment measurements were acquired at 20°C on Bruker AV900 and DRX600 spectrometers equipped with TXI HCN xyz-grad probe heads, on an AV800 spectrometer equipped with a TXI HCN z-grad probe head, and on a DRX600 spectrometer equipped with a TXI HCN cryoprobe. Sequential backbone assignments were obtained by using HNCA (Grzesiek and Bax 1992), HNCACB (Wittekind and Mueller 1993), CBCACONH (Muhandiram and Kay 1994), HN(CA)CO (Clubb et al. 1992), HNCO (Grzesiek and Bax 1992), NOESY-HSQC (Marion et al. 1989a), and TOCSY-HSQC (Marion et al. 1989b) experiments. Pulsed field gradient versions of all these experiments with sensitivity enhancement for the back transfer (Schleucher et al. 1994) were used. Carrier positions were as follows: ^1H , 4.7 ppm; $^{13}\text{C}^{\text{ali}}$, 39 ppm; ^{13}CO , 172 ppm; and ^{15}N , 117.5 ppm; and the recycle delay was set to 1.5 sec. Carbon pulses were implemented as Gaussian pulse cascades (Emsley and Bodenhausen 1990); in particular, Q3 and Q5 cascades (Emsley and Bodenhausen 1992) were applied.

Data processing was done using the program Xwinnmr version 3.5 (Bruker); Linear prediction, as implemented in Xwinnmr, was used to extend the data in the ^{15}N dimension; and all data sets were zero-filled to yield final matrix sizes of $256 \times 256 \times 2048$ complex points. Data analysis was done using the program CARA (R. Keller, in prep.), downloaded from <http://www.nmr.ch>. ^1H chemical shifts were referenced directly to internal (2,2,3,3-d₄) trimethyl-3-propanoic acid, sodium salt (TMSP), while ^{13}C and ^{15}N chemical shifts were referenced indirectly calculating the true frequency by using the approach of Wishart et al. (1995b).

Chemical shift deviations

Random coil chemical shifts were extracted from NMRview by using the "Wishart peptide" scale (Wishart et al. 1995a), including sequence specific corrections (Schwarzinger et al. 2001), and used to determine deviations from random coil chemical shifts. The chemical shift deviations were calculated according to $\Delta\delta = \delta^{\text{exp}} - \delta^{\text{rc}}$, with δ^{exp} = experimental chemical shifts, δ^{rc} = random coil chemical shifts according to the method of Wishart et al. (1995a). δ^{exp} for BLA- S^{ME} were used from the assignment described here; δ^{exp} for reduced HLA (all-Ala-HLA) and δ^{exp} for HEWL- S^{ME} were taken from previous work (Schwalbe et al. 1997; Redfield et al. 1999).

R_2 relaxation rates

R_2 relaxation rates of BLA- S^{ME} and all-Ala-HLA were determined as described previously (Schwalbe et al. 1997). Relaxation rates of HEWL- S^{ME} were taken from earlier work (Schwalbe et al. 1997). All R_2 relaxation rates compared here were measured at 20°C at a proton frequency of 600 MHz. Relaxation rates were fitted using the segmental motion model as described in the literature (Schwalbe et al. 1997).

Hydrophobicity predictions

Hydrophobicity was calculated by using the protscale tool from the ExPASy (Expert Protein Analysis System) (Appel et al. 1994) molecular biology server (<http://us.expasy.org/cgi-bin/protscale.pl>) using the following sequences from the Swiss Prot database: P00709 LCA_HUMAN, P00711 LCA_BOVIN, and P00698 LYC_CHICK, each leaving out the first 19 N-terminal amino acids that belong to the precursor peptide. The Abraham and Leo (1987) scale was applied. The window size (length of the interval used for profile computation in units of residues) was seven (as this is the persistence length found in a number of unfolded proteins) (Schwalbe et al. 1997; Schwarzinger et al. 2002), and the weight at the edge of the window was set to 100%; the hydrophobicity shown is normalized from 0 to 1.

Sequence alignment

Sequence alignment was performed by using the BLAST database algorithm based on the method of Altschul et al. (1997). The same sequences were used as for the hydrophobicity predictions.

Acknowledgments

Work in the group of H.S. was supported by the state of Hesse (BMRZ) and the EU (Strategic Research Project: UPMAN). R.U. acknowledges support from an EMBO short-term fellowship.

C.R. acknowledges support from the Wellcome Trust and the EPA Cephalosporin Trust.

References

- Abraham, D.J. and Leo, A.J. 1987. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins* **2**: 130–152.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Appel, R.D., Bairoch, A., and Hochstrasser, D.F. 1994. A new generation of information retrieval tools for biologists: The example of the ExpASY WWW server. *Trends Biochem. Sci.* **19**: 258–260.
- Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M., and Fersht, A.R. 1995. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: Implications for the initiation of protein folding. *J. Mol. Biol.* **254**: 305–321.
- Bertoncini, C.W., Jung, Y.S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M., and Zweckstetter, M. 2005. Release of long-range tertiary interactions potentiates aggregation of natively unstructured α -synuclein. *Proc. Natl. Acad. Sci.* **102**: 1430–1435.
- Blanco, F.J., Serrano, L., and Forman-Kay, J.D. 1998. High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol.* **284**: 1153–1164.
- Bussell Jr., R. and Eliezer, D. 2001. Residual structure and dynamics in Parkinson's disease-associated mutants of α -synuclein. *J. Biol. Chem.* **276**: 45996–46003.
- Chandra, N., Brew, K., and Acharya, K.R. 1998. Structural evidence for the presence of a secondary calcium binding site in human α -lactalbumin. *Biochemistry* **37**: 4767–4772.
- Chang, J.Y. 2004. Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry* **43**: 4522–4529.
- Chang, J.Y. and Li, L. 2002. Pathway of oxidative folding of α -lactalbumin: A model for illustrating the diversity of disulfide folding pathways. *Biochemistry* **41**: 8405–8413.
- Clubb, R.T., Thanbal, V., and Wagner, G. 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intrareidue ^1H , ^{15}N , and $^{13}\text{C}'$ chemical shifts in ^{15}N , ^{13}C -labelled proteins. *J. Magn. Reson.* **97**: 213–217.
- Dobson, C.M., Evans, P.A., and Radford, S.E. 1994. Understanding how proteins fold: The lysozyme story so far. *Trends Biochem. Sci.* **19**: 31–37.
- Emsley, L. and Bodenhausen, G. 1990. Gaussian pulse cascades: New analytical functions for rectangular selective inversion and in-phase excitation in NMR. *Chem. Phys. Lett.* **165**: 469–476.
- . 1992. Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *J. Magn. Reson.* **97**: 135–148.
- Fieber, W., Kristjansdottir, S., and Poulsen, F.M. 2004. Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol.* **339**: 1191–1199.
- Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J., and Dobson, C.M. 1996. Towards a description of the conformations of denatured states of proteins: Comparison of a random coil model with NMR measurements. *J. Chem. Phys.* **100**: 2661–2666.
- Forge, V., Wijesinha, R.T., Balbach, J., Brew, K., Robinson, C.V., Redfield, C., and Dobson, C.M. 1999. Rapid collapse and slow structural reorganisation during the refolding of bovine α -lactalbumin. *J. Mol. Biol.* **288**: 673–688.
- Frank, M.K., Clore, G.M., and Gronenborn, A.M. 1995. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci.* **4**: 2605–2615.
- Gillespie, J.R. and Shortle, D. 1997a. Characterization of long-range structure in the denatured state of staphylococcal nuclease, I: Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**: 158–169.
- . 1997b. Characterization of long-range structure in the denatured state of staphylococcal nuclease, II: Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**: 170–184.
- Grzesiek, S. and Bax, A. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**: 432–440.
- Heinrikson, R.L. 1971. The selective S-methylation of sulfhydryl groups in proteins and peptides with methyl-p-nitrobenzenesulfonate. *J. Biol. Chem.* **246**: 4090–4096.
- Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J., and Schwalbe, H. 1999. Side-chain conformations in an unfolded protein: χ^1 distributions in denatured hen lysozyme determined by heteronuclear ^{13}C , ^{15}N NMR spectroscopy. *J. Mol. Biol.* **288**: 705–723.
- Hill, R.L. and Brew, K. 1975. Lactose synthetase. *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**: 411–490.
- Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M., and Schwalbe, H. 2002. Long-range interactions within a nonnative protein. *Science* **295**: 1719–1722.
- Li, L. and Chang, J.Y. 2004. Two-state folding of lysozyme versus multiple-state folding of α -lactalbumin illustrated by the technique of disulfide scrambling. *Protein J.* **23**: 3–10.
- Lietzow, M.A., Jamin, M., Dyson, H.J., and Wright, P.E. 2002. Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322**: 655–662.
- Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C.M., Poulsen, F.M., and Vendruscolo, M. 2004. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J. Am. Chem. Soc.* **126**: 3291–3299.
- Logan, T.M., Theriault, Y., and Fesik, S.W. 1994. Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.* **236**: 637–648.
- Maenaka, K., Kawai, G., Watanabe, K., Sunada, F., and Kumagai, I. 1994. Functional and structural role of a tryptophan generally observed in protein-carbohydrate interaction. TRP-62 of hen egg white lysozyme. *J. Biol. Chem.* **269**: 7070–7075.
- Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. 1989a. Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H - ^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1β . *Biochemistry* **28**: 6150–6156.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. 1989b. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* **111**: 1515–1517.
- McKenzie, H.A. and White Jr., F.H. 1991. Lysozyme and α -lactalbumin: Structure, function, and interrelationships. *Adv. Protein Chem.* **41**: 173–315.
- Muhandiram, D.R. and Kay, L.E. 1994. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B.* **103**: 203–216.
- Nitta, K. and Sugai, S. 1989. The evolution of lysozyme and α -lactalbumin. *Eur. J. Biochem.* **182**: 111–118.
- Peng, Z.Y. and Kim, P.S. 1994. A protein dissection study of a molten globule. *Biochemistry* **33**: 2136–2141.
- Peng, Z.Y., Wu, L.C., and Kim, P.S. 1995. Local structural preferences in the α -lactalbumin molten globule. *Biochemistry* **34**: 3248–3252.
- Permyakov, E.A. and Berliner, L.J. 2000. α -Lactalbumin: Structure and function. *FEBS Lett.* **473**: 269–274.
- Pike, A.C., Brew, K., and Acharya, K.R. 1996. Crystal structures of guinea-pig, goat and bovine α -lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure* **4**: 691–703.
- Redfield, C., Schulman, B.A., Milhollen, M.A., Kim, P.S., and Dobson, C.M. 1999. α -Lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat. Struct. Biol.* **6**: 948–952.
- Sari, N., Alexander, P., Bryan, P.N., and Orban, J. 2000. Structure and dynamics of an acid-denatured protein G mutant. *Biochemistry* **39**: 965–977.
- Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S.J., Sorensen, O.W., and Griesinger, C. 1994. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **4**: 301–306.
- Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J., and Dobson, C.M. 1997. Structural and dynamical properties of a denatured protein: Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**: 8977–8991.
- Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. 2001. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**: 2970–2978.
- Schwarzinger, S., Wright, P.E., and Dyson, H.J. 2002. Molecular hinges in protein folding: The urea-denatured state of apomyoglobin. *Biochemistry* **41**: 12681–12686.
- Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M., and Dobson, C.M. 1996a. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**: 494–506.
- Smith, L.J., Fiebig, K.M., Schwalbe, H., and Dobson, C.M. 1996b. The concept of a random coil: Residual structure in peptides and denatured proteins. *Fold. Des.* **1**: R95–R106.

- van den Berg, B., Chung, E.W., Robinson, C.V., Mateo, P.L., and Dobson, C.M. 1999. The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *EMBO J.* **18**: 4794–4803.
- Vaney, M.C., Maignan, S., Riès-Kautt, M., and Ducruix, A. 1996. High-resolution structure (1.33 angstrom) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus: Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. *Acta Crystallogr. D Biol. Crystallogr.* **52**: 505–517.
- Wirmer, J., Schlörb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T., and Schwalbe, H. 2004. Modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations. *Angew. Chem. Int. Ed. Engl.* **43**: 5780–5785.
- Wishart, D.S. and Sykes, B.D. 1994. The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4**: 171–180.
- Wishart, D.S., Sykes, B.D., and Richards, F.M. 1992. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**: 1647–1651.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995a. ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5**: 67–81.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. 1995b. ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**: 135–140.
- Wittekind, M. and Mueller, L. 1993. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α -carbon and β -carbon resonances in proteins. *J. Magn. Reson. B* **101**: 201–205.
- Wu, L.C., Peng, Z.Y., and Kim, P.S. 1995. Bipartite structure of the α -lactalbumin molten globule. *Nat. Struct. Biol.* **2**: 281–286.
- Yao, J., Chung, J., Eliezer, D., Wright, P.E., and Dyson, H.J. 2001. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40**: 3561–3571.
- Yi, Q., Scalley-Kim, M.L., Alm, E.J., and Baker, D. 2000. NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**: 1341–1351.