Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins

Christina Redfield*

Oxford Centre for Molecular Sciences, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK

Accepted 5 March 2004
Available online 2 June 2004

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for the study of the structure, dynamics, and folding of proteins in solution. It is particularly powerful when applied to dynamic or flexible systems, such as partially folded molten globule states of proteins, which are not usually amenable to X-ray crystallography. In this article, NMR methods suitable for the detailed characterisation of molten globule states are described. The specific method used to study the molten globule is determined by the quality of the NMR spectrum obtained. Molten globules are characterised by significant levels of secondary structure. Site-specific hydrogen–deuterium exchange experiments can be used to identify residues located in regions of secondary structure in the molten globule. If spectra characterised by sharp peaks are observed for the molten globule then information about secondary structure can be obtained by analysis of $^1$H, $^{13}$C, $^{13}$C, and $^{13}$CO chemical shifts; this can be supplemented by $^{15}$N relaxation studies. For molten globules characterised by extremely broad peaks $^{15}$N-edited NMR experiments carried out in increasing concentrations of denaturants can be used to study the relative stabilities of different regions of structure. Examples of the application of these methods to the study of the low pH molten globule states of $\alpha$-lactalbumin and apomyoglobin are presented.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Molten globule; Partially folded proteins; Protein folding; Protein denaturation; Hydrogen–deuterium exchange; HSQC; $\alpha$-Lactalbumin; Apomyoglobin; Resonance assignment; Chemical shift

1. Introduction

Understanding the factors that determine how a linear polypeptide chain folds to its unique three-dimensional structure remains one of the fundamental questions in biology. In some cases, protein folding has been found to be a highly cooperative process. However, in other cases partially structured species have been observed to form early in the refolding process before the formation of the native state. These partially folded species are compact, have extensive native-like secondary structure but lack the specific tertiary side-chain packing characteristic of native structures [1,2]. For a number of proteins, similar species have been found to be stable at equilibrium under mild denaturing conditions [1,2]. The term ‘molten globule’ was first used in 1983 by Ohgushi and Wada [3] and by Ptitsyn and co-workers [4] to describe these compact partially folded species. Several studies have suggested a close resemblance between molten globule states observed at equilibrium and those formed during the early stages of refolding of several proteins, including $\alpha$-lactalbumin [5–7] and apomyoglobin [8,9]. There has been considerable interest in the detailed characterisation of the stable equilibrium molten globule states of proteins to gain insights into the transient kinetic molten globule and the possible determinants of the protein fold.

Molten globules have been found to exist for several proteins under mild denaturing conditions or upon removal of a cofactor. Acidic pH is a particularly effective method for generating molten globule states [1,2,10]. At low pH, $\alpha$-lactalbumin, equine lysozyme, RNase HI from Escherichia coli, and $\beta$-lactoglobulin all form molten globules [1]. For some proteins, including cytochrome $c$ and staphylococcal nuclease, high salt
concentrations are required at low pH to stabilise the molten globule state [1]. Removal of the heme group from myoglobin, to form apomyoglobin, leads to a native-like structure at neutral pH and a molten globule at pH 4 [1,11,12].

The lack of fixed tertiary interactions in many molten globules results in a fluctuating ensemble of structures, which inter-convert on a millisecond to microsecond time scale [13–15]. This heterogeneous character can make detailed structural studies using conventional X-ray crystallographic or NMR methods difficult. Although extreme line broadening often precludes direct NMR structural studies of these molten globules, detailed residue-specific information can be gained using indirect NMR methods. For example, hydrogen–deuterium exchange methods can be used to identify residues involved in hydrogen-bonded secondary structure in the molten globule [5,6,8,13,16–20]. 15N-edited NMR experiments carried out in increasing concentrations of denaturants can be used to characterise the relative stabilities of different regions of the molten globule structure [14,21–25]. For some molten globules, significant line narrowing can be achieved at elevated temperature. For these proteins, more conventional NMR methods can be used to assign the NMR spectrum. Analysis of 1H and 13C chemical shift data for these molten globules can provide direct information about the structure present in the molten globule [12,23,26,27]. In this article, NMR methods suitable for the detailed characterisation of molten globule states are described. These NMR methods are illustrated with examples from α-lactalbumin and apomyoglobin (Fig. 1).

2. Hydrogen–deuterium exchange methods for identifying secondary structure

Molten globule states of proteins are often characterised by a high level of secondary structure, particularly α-helices. The pH 2 molten globule of α-lactalbumin has a near-native helical content [1,2]. The molten globule of apomyoglobin at pH 4.1 has ~65% of the helical content observed for native apomyoglobin at neutral pH [8]. Hydrogen–deuterium exchange methods monitored by 2D NMR can be used to identify the specific residues involved in these regions of secondary structure.

If a protein is dissolved in D2O then backbone amide protons will exchange with deuterons. The intrinsic kinetics of the exchange process depends on pH, temperature, and the neighbouring side chains [28–30]; hydrogen–deuterium exchange is described more detail in another article in this volume [31]. If an amide is involved in a hydrogen bond in the protein then the rate of exchange may be slowed down significantly compared to the predicted intrinsic rate. The level of protection of an amide from hydrogen–deuterium exchange, the protection factor (PF), is measured as the ratio of the intrinsic exchange rate (k_{intr}) to the observed exchange rate (k_{obs}), PF = k_{intr}/k_{obs}. In native proteins, protection factors of 10^5–10^7 are not uncommon in regions of hydrogen-bonded secondary structure. In completely unfolded proteins, the observed rates of exchange are those predicted from the intrinsic rates. Molten globules contain significant levels of secondary structure and the hydrogen bonds present in them can give rise to protection from hydrogen–deuterium exchange. This was first demonstrated, in a residue-specific manner, for the low pH molten globule of guinea-pig α-lactalbumin [13].

The poor resolution and extreme line broadening often observed in the NMR spectrum of a molten globule at low temperature mean that the slowly exchanging amides located in regions of secondary structure cannot be identified directly from the NMR spectrum of the molten globule. Instead, the increased resolution and narrow lines observed in the native state spectrum of the protein are exploited. Hydrogen–deuterium exchange is allowed to proceed under solution conditions that favour the molten globule. Aliquots are taken at various times after the protein has been

Fig. 1. Schematic representation of the native structures of (A) α-lactalbumin and (B) holomyoglobin. Regions of helical structure are labelled. The diagram was generated using Molscript [56] and the X-ray coordinates for the native proteins [57,58].
dissolved in D$_2$O and exchange in the molten globule state is quenched. Suitable time points can be identified from bulk hydrogen–deuterium exchange measurements carried out using NMR or mass spectrometric methods [32]. The protein is then returned to its native state where further exchange is very slow due to high protection factors [5,6,8,13,16–20]. In the case of $\alpha$-lactalbumin, the conversion of the molten globule to the native state is achieved by raising the pH from 2 to $\sim$6 [6,13,16,17]. For apomyoglobin, the native state is reconstituted by the addition of heme coupled with an increase in pH to a value of 5.6 [8]. Under native conditions a well-resolved NMR spectrum is obtained allowing residue specific exchange rates to be measured from peak intensities measured in either 2D homonuclear (COSY/TOCSY) or heteronuclear ($^{15}$N–$^1$H HSQC) NMR experiments. The measured rates reflect the exchange protection due to hydrogen bonds present in the molten globule state of the protein. This method requires substantial amounts of protein because a separate sample is required for each time point. Molten globules are often prone to aggregation so the hydrogen–deuterium exchange experiments are generally conducted at low protein concentrations (as low as 10$\mu$M) and the samples are then concentrated for NMR analysis after the protein has been returned to native solution conditions. Because the individual samples may vary slightly in protein concentration, it is important to scale the measured peak intensities for each sample. If COSY or TOCSY spectra are used, then the intensities of non-exchangeable cross peaks from aromatic protons can be used for scaling. If $^{15}$N–$^1$H HSQC spectra are used, then a reference 1D $^1$H NMR spectrum should be collected for each sample for the purposes of scaling the peak intensities. This method is only suitable for molten globules that adopt a stable folded native structure under some solution conditions, and is only appropriate for identifying amides that are protected in both the molten globule and native states of the protein.

Amides that are protected in the low pH molten globules of bovine, guinea-pig, and human $\alpha$-lactalbumin have been identified using this approach [6,13,16,17]. For example, bovine $\alpha$-lactalbumin was dissolved at a concentration of 300$\mu$M in D$_2$O at pH 2 and aliquots taken at specific time points ranging from 30 s to 3.5 h were frozen to quench the exchange. Samples were then freeze-dried and redissolved in pH 5.5 buffer containing Ca$^{2+}$; under these conditions bovine $\alpha$-lactalbumin adopts a native structure [6]. COSY spectra were collected to identify the protected amides; examples of these spectra are shown in Fig. 2. Amides located in the native A-, B-, and C-helices of the $\alpha$-domain are protected from hydrogen–deuterium exchange in bovine $\alpha$-lactalbumin at pH 2, indicating that these helices are also formed in the molten globule state (Fig. 1). The observed protection factors (up to $\sim$100) in the bovine $\alpha$-lactalbumin molten globule are substantially lower than those observed in the native state (up to $\sim$10$^3$) [6]. It is not possible to determine if amides in the D- and C-terminal $3_{10}$ helices are formed in the molten globule state because amides in these helices are not protected in the native state of the protein [6]. Similar experiments carried out for guinea-pig $\alpha$-lactalbumin show protection in the B- and C-helices with protection factors up to $\sim$200 [13,16]. For human $\alpha$-lactalbumin, protection factors as high as $\sim$500 are observed in the A-, B-, and C-helices [17]. The higher values observed in guinea-pig and human $\alpha$-lactalbumin indicate that the helical secondary structure in the molten globules of these proteins is more stable than that in bovine $\alpha$-lactalbumin.

Similar hydrogen–deuterium exchange experiments have been used to probe the secondary structure present in the pH 4, molten globule state of apomyoglobin [8].
Exchange was allowed to proceed at pH 4.2, for 10 μM samples of apomyoglobin at 5 °C. The protein was returned to the native state by the addition of heme and an increase in pH to 5.6. Protected amides are observed in the A-, G-, and H-helices of the molten globule; amides in the B- and E-helices are not found to be protected from exchange (Fig. 1). It was not possible to determine if the C-, D-, and F-helices are formed in the molten globule state. It was not possible to determine if the C-, D-, and F-helices are formed in the molten globule because amides located in these helices are not protected from exchange in the native state of myoglobin.

3. Direct multidimensional NMR studies of molten globule states

15N- and 13C-edited 2D and 3D NMR spectroscopy is a powerful method for obtaining detailed structural information for proteins in solution. However, these methods can only be applied if the protein gives a relatively sharp peak in 15N–1H HSQC spectra. NMR resonances of residues located in compact, folded regions of molten globule states are often extremely broad due to conformational fluctuations on a millisecond to microsecond time scale [13–15]. The 15N–1H HSQC spectrum of the pH 4.1 molten globule state of apomyoglobin at 20 °C is characterised by broadened peaks [8,26]. If the temperature is increased to 50 °C, a significant sharpening of peaks is observed in the HSQC spectrum of apomyoglobin [26]. This sharpening can be attributed to modification of the complex dynamic properties of the molten globule. The observed sharpening of peaks at higher temperature allows the direct study of the pH 4.1 molten globule of apomyoglobin. These studies, by Dyson, Wright, and co-workers [8,12,26,27], have allowed a more detailed picture of the structure and dynamics of the apomyoglobin molten globule to be described than had been possible from the earlier hydrogen–deuterium exchange studies.

Spectroscopy at elevated temperature is complicated by the tendency of many molten globules to aggregate at high temperature at the concentrations required for 2D and 3D NMR studies. In the case of apomyoglobin, a relatively low protein concentration of ~300 μM was used for the NMR experiments and 10% ethanol was added to the samples for increased stability; the ethanol was not found to perturb the NMR spectrum or physical properties of the pH 4.1 molten globule [12,26,27]. A fresh sample of 13C–15N double-labelled protein was used for each 3D NMR experiment [12,27]. For some 3D experiments, which suffer from poorer sensitivity, 2H labelling at a level of ~80% was also used [27]. All NMR spectra were collected at a 1H frequency of 750 MHz to increase sensitivity.

Assignments for 1H^V, 15N, 1H^H, 13C^a, 13C^b, and 13CO resonances of the apomyoglobin molten globule were obtained using a suite of 3D triple resonance experiments [12,27]. The spectra of the apomyoglobin molten globule are similar to those of other partially folded states in that they lack significant chemical shift dispersion in the 1H^V, 1H^H, and 13C^a dimensions; assignment methods that rely on dispersion of these chemical shifts are not particularly successful. The assignment approach adopted in studies of the apomyoglobin molten globule relied on the inherent chemical shift dispersion of 15N and 13CO [33,34]. This dispersion arises from the variation of the intrinsic chemical shifts of these groups in the individual amino acids and in the sequence dependence of the random coil shifts [33,35,36]. The experiments used to assign 1H^V, 15N, 1H^H, 13C^a, 13C^b, and 13CO of apomyoglobin include 3D HNCA [37], HN(CO)CA [37], (HCA)CO(CA)NH [38], HNCO [37], HNACCB [39], and CBCA(CO)NH [40]. The HNCA experiment correlates 1H^V(i), 15N(i), 13C^a(i), and 13C^b(i−1) while HN(CO)CA correlates 1H^H(i), 15N(i), and 13C^a(i−1). The (HCA)CO(CA)NH experiment correlates 1H^V(i), 15N(i), 13C^a(i), and 13CO(i−1) while HNCO correlates 1H^V(i), 15N(i), and 13CO(i−1). The HNACCB experiment correlates 1H^V(i), 15N(i), 13C^a(i), 13C^b(i−1), 13CO(i−1) while CBCA(CO)NH correlates 1H^V(i), 15N(i), 13C^a(i−1), and 13CO(i−1). Using these pairs of experiments it was possible to identify the intra- and interresidue connections between 1H^V, 15N, and 13C^a, 13C^b, and 13CO in apomyoglobin at pH 4.1, and 50 °C [27].

A 3D 15N-edited HSQC-NOESY-HSQC [41] experiment which contains sequential 1H^V–1H^V NOE connectivities was also used to assist in the assignment process (Fig. 3). This experiment contains two indirect 15N dimensions and a 1H^V acquisition dimension. It is preferred to the more standard 3D 15N-edited NOESY-HSQC, which contains two 1H dimensions because of the better resolution obtained in the 15N dimension due to the large range of 15N random coil chemical shifts.

Further information useful in the assignment process was obtained using a pH titration from pH 2 to 4 [27]. At pH 2.3, apomyoglobin is present as the acid-unfolded state. The spectrum of this species was assigned previously using methods analogous to those described above [12,42]. Peaks in the HSQC spectrum were observed to shift in position between pH 2 and 4, indicating fast chemical exchange between the acid-unfolded and molten globule species. Thus, in some cases assignments obtained at pH 2.3 for the acid unfolded state could be transferred directly to peaks in the molten globule state at pH 4.1. In less favourable cases, the assignments at pH 2.3 were used to constrain significantly the possible assignments for peaks in the molten globule state. This approach was found to be very useful for resolving ambiguities that arise in the analysis of the 3D triple resonance data sets described above [27].
Once complete backbone and $^{13}\text{C}^\beta$ assignments were available for the apomyoglobin molten globule, these were used to obtain information about structural preferences in the molten globule state. Previous studies have indicated that there is a significant correlation between secondary chemical shifts of $^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^{13}\text{CO}$ and secondary structure in proteins [27,43–45]. In the analysis of chemical shifts, it is important to use the appropriate random coil chemical shifts in calculating the secondary chemical shifts. In the case of the $^{13}\text{CO}$, it is important to take into account the effects on the random coil chemical shifts of the preceding or following residue. It has been shown previously that $^{13}\text{C}^\alpha$ resonances are shifted downfield by 3.1 ± 1.0 ppm in $\alpha$-helical structure and are shifted upfield by 1.5 ± 1.2 ppm in $\beta$-sheets [44]. In helical structure, $^1\text{H}^\alpha$ and $^{13}\text{C}^\beta$ are observed to be shifted upfield relative to random coil values [27,43,45]. $^{13}\text{CO}$ peaks are shifted downfield in helical structure [27,45]. The secondary chemical shifts observed for apomyoglobin at pH 4.1, are plotted as a function of sequence in Fig. 3. It can be seen that large downfield shifts for $^{13}\text{C}^\alpha$ and $^{13}\text{CO}$ and large upfield shifts for $^1\text{H}^\alpha$ and $^{13}\text{C}^\beta$ are observed throughout the apomyoglobin sequence. Particularly large shifts are observed for residues located in the A-, G- and H-helical regions in native apomyoglobin; these shifts can be interpreted as indicators of helical structure in the molten globule state of apomyoglobin. This interpretation agrees with previous hydrogen–deuterium exchange studies of the apomyoglobin molten globule discussed above [8]. The observed shifts are smaller than those observed in the native state of holomyoglobin. The native state shifts can be used to calculate a fractional helicity for each residue or group of residues in the molten globule. On this basis the A-, G-, and H-helices are estimated to contain ~72, ~62, and ~53% of
their native helical content. Other regions of the sequence, the B-, C-, D-, and E-helices, are found to have smaller secondary chemical shift changes and these are indicative of smaller propensities (~25 to 40%) for helical structure at pH 4.1 [12].

Information about the presence of intramolecular hydrogen bonds in the apomyoglobin molten globule was obtained by measuring the \(^1\)H\(^N\) temperature coefficients over the temperature range of 40–50 °C. Generally the \(^1\)H\(^N\) chemical shifts of amides that are involved in intramolecular hydrogen bonds in a protein are less sensitive to temperature than amides that exchange with solvent. Unstructured peptides are characterised by temperature coefficients of ~8 ppb/K [46]. Temperature coefficients for many residues in the A-, G-, and H-helices of apomyoglobin at pH 4.1 were found to be substantially lower than the random coil values suggesting the presence of stable hydrogen bonds [27]. This is consistent with the observed \(^1\)H\(^a\), \(^13\)C\(^a\), \(^13\)C\(^b\), and \(^13\)CO chemical shift deviations for these residues discussed above and with the hydrogen–deuterium exchange data [8,27].

Additional information about the apomyoglobin molten globule was obtained using \(^15\)N relaxation measurements which are sensitive to motions of the backbone amides. \(^15\)N \(R_1\) and \(R_2\) relaxation rates and the \(^1\)H–\(^15\)N steady-state heteronuclear NOE were measured for apomyoglobin at pH 4.1, and 50°C at 750 MHz [12,27]. Because of the dynamic and fluctuating nature of at least part of the apomyoglobin structure at pH 4.1, analysis of these relaxation data using the Modelfree approach used for native globular proteins is not meaningful. Plots of \(R_1\), \(R_2\), and the \(^1\)H–\(^15\)N NOE are shown in Fig. 4. The \(^1\)H–\(^15\)N NOE is sensitive to

---

Fig. 4. Backbone \(^15\)N relaxation parameters (\(R_1\), \(R_2\), and \(^1\)H–\(^15\)N NOE) measured for apomyoglobin at pH 4.1. Black rectangles at the bottom of the top panel indicate the locations of helices in the native holomyoglobin structure [58] (reproduced from [27] with permission).
motions on the picosecond to nanosecond time scale. In the F-helix region, for example, negative NOEs characteristic of an unstructured free-flight chain are observed. By contrast, in structured regions of the molten globule the observed NOE ratio is around 0.7; this value is close to the values of 0.8–0.9 observed in the native state and indicates that backbone mobility on a fast time scale is limited for these residues [12,27]. $R_2$ is sensitive to nanosecond motions and to exchange processes on the millisecond to microsecond time scale. Higher $R_2$ rates are observed in the structured regions of the apomyoglobin molten globule indicating the dominance of slower time scale motions. These $R_2$ values may include contributions from slower exchange processes resulting from larger scale motions of the structural elements within the AGH core [27]. The $R_1$ values are rather uniform across the apomyoglobin sequence and do not provide insights into the backbone motions.

The $R_1$, $R_2$, and $^1$H–$^1$H NOE data can be analysed using the reduced spectral density mapping approach to yield $J(0)$, $J_{(0\pi)}$, and $J_{(0.87\pi)}$ [47]. However, this approach does not appear to provide additional insights into the backbone dynamics of the apomyoglobin molten globule [27]. The $J_{(0.87\pi)}$ plot provides very similar information to the $^1$H–$^15$N NOE plots and reflects the highest frequency motions. $J(0)$ is most sensitive to $R_2$ and the plots of $J(0)$ and $R_2$ versus sequence are very similar. $J_{(0\pi)}$ does not discriminate well between the faster and slower motions and, like $R_1$, is rather uniform across the sequence [27]. Thus, it is possible to obtain a significant level of information about backbone dynamics in the apomyoglobin molten globule directly from an analysis of the $R_2$ and $^1$H–$^1$H NOE data sets without the need for reduced spectral density mapping.

In native proteins, the majority of structural information is obtained from short and long-range NOEs. Sequential $^1$H–$^15$N NOEs are observed for the apomyoglobin molten globule and the intensities of these have been found to correlate with regions of helical secondary structure (Fig. 3); stronger $^3$H–$^1^$H NOEs are observed in regions where chemical shifts indicate significant population of helical $\phi - \psi$ conformational space [27]. However, no medium-range $^1$H–$^1$H NOE data are observed for apomyoglobin. A similar observation has been made for the human $\alpha$-lactalbumin molten globule at 50°C [23]. The lack of these NOEs is thought to arise from the lack of fixed tertiary interactions characteristic in the molten globule. Thus, NOE data cannot be used to characterise in any detail the tertiary structure of the globule.

Spin labelling is a promising approach for the identification of long-range contacts in unfolded and partially folded proteins. The presence of a label containing a free radical, covalently attached to a unique cysteine residue introduced by mutagenesis, will lead to a perturbation of the relaxation rates of nuclei that are located within a ~20 Å radius of the spin label [48]. This interaction can be detected by comparison of $R_1$ and $R_2$ rates measured in the presence and absence of the spin label or more simply by the comparison of peak intensities in HSQC spectra collected in the presence or absence of the spin label [48,49]. This method has been used to identify long-range interactions in the denatured state of the $\Delta$131A fragment of staphylococcal nuclease labelled with the PROXYL spin label, indicating that this species is not a random coil [48]. A similar study of the pH 2.3 acid-unfolded state of apomyoglobin labelled with the methanethiosulphonate spin label has identified the presence of transient native-like contacts [49]. This method has not yet been applied to the pH 4.1 molten globule of apomyoglobin but this approach may provide an important source of long-range structural information.

4. Probing the stability of the molten globule state using HSQC spectroscopy

In the previous section, it was shown that for apomyoglobin a significant sharpening of the HSQC spectrum is observed at 50°C and that spectra of sufficient quality are obtained to permit complete backbone $^1$H, $^13$C, and $^15$N assignment. Analysis of chemical shift and relaxation data provide detailed information about the secondary structure and backbone dynamics of the apomyoglobin molten globule. The NMR spectrum of human $\alpha$-lactalbumin is also found to sharpen significantly at elevated temperature and important information about the secondary structure of the molten globule state has been obtained from high temperature studies [23]. However, at 50°C peaks from more than 25 residues are still absent from the $^15$N–$^1$H HSQC spectrum of human $\alpha$-lactalbumin and a number of observed peaks are still significantly broadened [23]. Therefore, a full analysis of chemical shifts and dynamics, as has been carried out for apomyoglobin, has not yet been possible. Instead, a 2D NMR approach using $^15$N–$^1$H HSQC spectra collected with increasing concentrations of urea has been developed for the study of the $\alpha$-lactalbumin molten globule [14,21,22]; a similar approach has been used previously to study truncated variants of staphylococcal nuclease [24,25]. This approach has been used for the characterisation of the molten globule states of bovine and human $\alpha$-lactalbumin at pH 2 [14,21,22].

The $^15$N–$^1$H HSQC spectra of bovine and human $\alpha$-lactalbumin at pH 2, and 20°C are characterised by extreme line broadening; sharp peaks from only 2 or 3 backbone amides are observed [14,22]. With so few peaks observed in the HSQC spectra, it is not possible to apply 3D NMR methods to obtain direct structural information for the molten globules of bovine and human.
α-lactalbumin. However, the number of resolved peaks observed in the HSQC spectra of bovine and human α-lactalbumin increases as the concentration of urea is increased at 20°C (Fig. 5) [14,22]. In 2 M urea, 13 peaks are observed in the HSQC spectrum of bovine α-lactalbumin. This increases to 57 peaks in 6 M urea and to 92 peaks in 10 M urea. If the temperature is increased to 50°C then all of the 121 residues in bovine α-lactalbumin give rise to detectable peaks in the HSQC spectrum in 10 M urea. For the homologous protein, human α-lactalbumin, more strongly denaturing conditions are required. In 10 M urea at 20°C only 77 peaks are observed in contrast to 92 for bovine α-lactalbumin. When the temperature is increased to 50°C, 92 of the 122 residues give rise to observable peaks. Peaks for all 122 residues of human α-lactalbumin are only observed in 8 M guanidine HCl at 50°C [14].

The urea-titration $^{15}$N-$^1$H HSQC experiments described above can be carried out with a single sample of $^{15}$N-labelled protein [14]. The sample is made up initially at a protein concentration of ~0.5–1 mM in 95% H$_2$O/5% D$_2$O at pH 2, and an HSQC spectrum is collected. Solid urea is weighed and then added to the solution to give a final urea concentration of 1 M; for a 500 μl NMR sample, 500 μl of 1 M urea is added to a 500 μl sample of protein in 95% H$_2$O/5% D$_2$O at pH 2.

![Fig. 5. Progressive appearance of unfolded NMR resonances in $^{15}$N-$^1$H HSQC spectra of the bovine (A–D) and human (E–H) α-lactalbumin molten globule. Spectra were recorded at pH 2 in different concentrations of urea: (A,E) 2 M; (B,F) 6 M; (C,G) 10 M; all at 20°C, and (D,H) 10 M at 50°C. Peaks are labelled with their residue assignments as they become visible in the spectra. The spectra are contoured so that a cross peak defined by a single contour has 20–25% of the intensity of a cross peak corresponding to a fully unfolded residue. The peaks corresponding to Ala 22 in human α-lactalbumin and Leu 96 in bovine α-lactalbumin are folded in the $^{15}$N dimension (reproduced from [22] with permission).](image-url)
sample 30 mg of urea must be added to increase the urea concentration from 0 to 1 M. The pH of the sample is adjusted to 2 and an HSQC spectrum is collected. This procedure is repeated for each molar increment in the urea concentration. It should be noted that the volume of the NMR solution increases as urea is added. This means that larger quantities of urea must be added for each molar increment in the urea concentration and that the protein concentration of the NMR sample will decrease during the titration. If desired the protein concentration can be kept constant if two samples initially containing no urea and 10 M urea are used. Samples of intermediate urea concentration are obtained by suitable mixtures of the 0 and 10 M solutions.

When a significant number of peaks are observed in the HSQC spectrum, 15N-edited 3D methods can be applied to obtain specific assignments for these peaks. The assignment of the spectrum of α-lactalbumin at pH 2, in urea is feasible using protein labelled with only 15N because of the relatively large range of 15N chemical shifts observed in partially folded and unfolded proteins. This arises from the large spread of random coil 15N chemical shifts for the amino acids and from the large sequence-dependent effects, of up to 4.5 ppm, observed on these random coil shifts [33]. It can be seen from Fig. 5 that peaks arising from Gly, Ser, and Thr residues are separated from the peaks arising from other amino acids. Within the cluster of Thr and Ser peaks there is sufficient resolution arising from sequence-dependent shifts to give well-resolved peaks.

Strips from a 15N-edited 3D NOESY-HSQC [50,51] spectrum obtained for human α-lactalbumin at pH 2, in 5 M urea are shown in Fig. 6; intrar residue and sequential NOEs are indicated. A number of observations can be made from 3D NOE data collected at a range of urea concentrations and temperatures [14]. First, the chemical shifts of aliphatic resonances are close to values expected for unstructured peptides and are largely independent of denaturant concentrations. Second, the pattern of NOESY cross peaks at various concentrations of urea is characteristic of a highly unfolded polypeptide chain [52]. Both strong sequential 1Hα–1Hα and 1Hα–1Hβ NOEs, reflecting a random distribution between α and β conformational space, are observed for all residues irrespective of their structure in the native protein [52]. Finally, no medium range 1Hα–1Hα(i,i + 3) or (i,i + 4) NOEs, characteristic of helical structure, are observed. These observations suggest that the peaks observed in the HSQC spectra as urea is added correspond to residues in unfolded regions of the polypeptide chain.

The observation of a large number of sequential 1Hα–1Hβ NOEs is particularly useful for assignment.
because the pair of $^1H^N$ giving rise to these NOEs can be assigned unambiguously in $^{15}N$-edited NOESY spectra. If poor $^1H^N$ chemical shift dispersion is a problem then assignment can be aided using the $^1H^N$–$^1H^V$ NOEs observed in the $^{15}N$-edited HSQC-NOESY-HSQC experiment described in Section 3 for apomyoglobin [41]. Sequential assignments can be confirmed by the observation of $^1H^\alpha$–$^1H^N(i,i+1)$ and $^1H^\beta$–$^1H^N(i,i+1)$ NOEs (Fig. 6) [53]. Information about residue type can be determined from $^{15}N$-edited 3D TOCSY-HSQC spectra [50,51] and from the chemical shifts of peaks giving rise to intraresidue and sequential NOEs in the aliphatic region of the 3D NOESY-HSQC spectra.

The unfolding of a native protein to its denatured state is usually a cooperative two-state process. If this transition is followed by NMR then the HSQC peaks arising from the native state will be found to decrease in intensity and a new set of HSQC peaks corresponding to the unfolded state will be found to increase in intensity as the concentration of denaturant is increased. All peaks increase or decrease in intensity in the same way as a result of the cooperativity of the process. This behaviour is not observed for the $\alpha$-lactalbumin molten globule at pH 2. The stepwise appearance of HSQC peaks with increasing concentrations of the denaturant urea is consistent with non-cooperative unfolding. NMR resonances of residues located in compact, folded regions of the low pH molten-globule state are extremely broad due to conformational fluctuations on a millisecond to microsecond time scale. By exposing the partially folded protein to increasingly destabilizing conditions, the progressive unfolding of different structural regions can be monitored by the appearance of sharp, well-resolved peaks in the HSQC spectrum corresponding to highly dynamic, unfolded parts of the protein.

The availability of residue specific assignments permits the analysis of the pattern of unfolding to gain insights into the stability of different regions of the molten globule structure. The unfolding patterns in urea at pH 2, observed for bovine and human $\alpha$-lactalbumin are compared in Fig. 7 [14,22]. The absence of all but two or three peaks in 0 M urea indicates that both the $\alpha$- and $\beta$-domains are at least partially collapsed in the molten globules of both bovine and human $\alpha$-lactalbumin. In both proteins, resonances appearing at the lowest concentrations of denaturant corresponds to the N-terminal residues and to the $\beta$-domain of native $\alpha$-lactalbumin. Previous studies have indicated that the $\beta$-domain has a lower propensity for a native fold in the molten globule [54,55]. The residues most resistant to denaturant are clustered together in the $\alpha$-domain of the native structure of $\alpha$-lactalbumin, indicating that the molten globule has a highly stable native-like core. The

![Fig. 7. Unfolding behaviour of (A) human and (B) bovine $\alpha$-lactalbumin at pH 2. The bars indicate the urea concentration at which an HSQC peak corresponding to 20–25% of the intensity of a fully unfolded residue is first observed. The scale on the right indicates the more destabilising conditions required to unfold parts of the $\alpha$-lactalbumin molten globule, 10 M urea at 30, 40, and 50 °C and 8 M guanidine HCl at 20, 35, and 50 °C. Residues located in the $\alpha$-domain are shown in grey and those in the $\beta$-domain are shown in white. The secondary structure found in native $\alpha$-lactalbumin is summarised at the top of the Figure (reproduced from [22] with permission).](image-url)
molten globule of bovine α-lactalbumin is less stable than that of human α-lactalbumin. In 6 M urea, many residues from the α-domain of bovine α-lactalbumin are visible in the HSQC spectrum; in particular, residues from the AB loop region and from the C-helix are observed. By contrast, the HSQC spectrum of human α-lactalbumin in 6 M urea does not contain peaks from any residues in the α-domain except those from the three N-terminal residues. The stability of the α-domain core is found to differ between human and bovine α-lactalbumin. In human α-lactalbumin, residues from the C-helix are found to be less resistant to unfolding than residues from the other helices. Residues located in the A, B, D, and C-terminal 3_10 helices form a highly stable core in the absence of a folded C-helix [14]. In bovine α-lactalbumin, all four α-helices and the C-terminal 3_10 helix unfold over a relatively narrow range of urea concentrations indicating that, in contrast to human α-lactalbumin, the C-helix is a required element in the stable molten globule core [22]. These differences in behaviour between bovine and human α-lactalbumin appear to arise from amino acid substitutions found in the B- and C-helices that result in more extensive contacts between the B- and C-helix regions in bovine α-lactalbumin.

5. Concluding remarks

Molten globule states of proteins are compact, have extensive native-like secondary structure but lack the specific tertiary side-chain packing characteristic of native structures [1,2]. Several studies have suggested a close similarity between molten globule states observed at equilibrium and those formed during the early stages of refolding of several proteins, including α-lactalbumin [5–7] and apomyoglobin [8,9]. The detailed characterisation of the stable equilibrium molten globule states of proteins promises to provide valuable insights into structure and dynamics of the transient kinetic molten globule important in protein folding pathways.

NMR spectroscopy is a powerful method for the study of the structure and dynamics of proteins in solution. It is particularly powerful when applied to dynamic or flexible systems, such as molten globule states of proteins, which are not usually amenable to characterisation by X-ray crystallography. In this article, a variety of NMR methods that can be used to study molten globule states of proteins have been described and these have been illustrated using α-lactalbumin and apomyoglobin, two of the most widely studied molten globules. The specific method used to study the molten globule state of a particular protein is determined by the quality of the NMR spectrum obtained. These NMR methods have provided detailed information about the specific residues involved in secondary structure within the molten globule, about the backbone dynamics of the polypeptide chain in the molten globule state, and about the relative stabilities of different regions of structure within the molten globule. However, these NMR methods have not yet been able to identify long-range tertiary interactions that may exist in the molten globule state. NMR studies of molten globules can be challenging because many of these proteins have a tendency to aggregate under conditions suitable for 2D and 3D NMR studies. The availability of higher field spectrometers equipped with cryo-probes in the future will improve the sensitivity of NMR experiments and permit the collection of data for more dilute samples that have less tendency to aggregate. These developments may enable an even more detailed picture of the molten globule to be described in future studies.

Acknowledgments

The author thanks the many people who have been collaborators in the study of the α-lactalbumin molten globule in Oxford; these are C.M. Dobson, L. Greene, P.S. Kim, Z.-y. Peng, T. Pertinhez, C. Quezada, S. Ramboarina, B.A. Schulman, and R. Wijesinha-Bettoni.

References
