Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme

Denis Canet[1,2], Alexander M. Last[1], Paula Tito[1], Margaret Sunde[3], Andrew Spencer[4], David B. Archer[5], Christina Redfield[1], Carol V. Robinson[1,6] and Christopher M. Dobson[1,6]

Published online: 11 March 2002, DOI: 10.1038/nsb768

Hydrogen exchange experiments monitored by NMR and mass spectrometry reveal that the amyloidogenic D67H mutation in human lysozyme significantly reduces the stability of the β-domain and the adjacent C-helix in the native structure. In addition, mass spectrometric data reveal that transient unfolding of these regions occurs with a high degree of cooperativity. This behavior results in the occasional population of a partially structured intermediate in which the three α-helices that form the core of the α-domain still have native-like structure, whereas the β-domain and C-helix are simultaneously substantially unfolded. This finding suggests that the extensive intermolecular interactions that will be possible in such a species are likely to initiate the aggregation events that ultimately lead to the formation of the well-defined fibrillar structures observed in the tissues of patients carrying this mutation in the lysozyme gene.

The accumulation of amyloid fibrils in tissue is characteristic of a range of pathological conditions including type II diabetes, Alzheimer's disease and the transmissible spongiform encephalopathies[1,2]. The amyloid fibrils formed in such diseases are derived from normally soluble proteins that have undergone structural transitions to form insoluble and potentially pathogenic material[3]. A recently discovered human amyloidotic disease is associated with one of the best characterized of all proteins: lysozyme[4]. The reported cases of this non-neuropathic systemic amyloidosis result from single point mutations in the lysozyme gene, giving rise to one or the other of two variant proteins: I56T and D67H. In their native states, these variants are structurally similar to the wild type protein[5]. The D67H variant of human lysozyme is incorporated into the disease-associated amyloid fibrils in its full-length, disulfide-bonded form; this is probably also the case for the I56T variant[5]. Moreover, the wild type lysozymes from both human and hen sources have recently been shown to be converted *in vitro* to fibrils that seem to be closely similar to amyloid fibrils formed *in vivo* by the variant proteins[6–8]. The mechanism by which the soluble monomeric form of lysozyme, whose native structure includes a substantial proportion of α-helical structure, is transformed into multimeric fibrillar assemblies that are rich in β-sheet structure is not yet understood in detail. It has been proposed, however, that partially unfolded states of the protein must be populated to permit aggregation and, ultimately, fibril formation[5]; similar arguments have been advanced for other proteins, such as transthyretin[9], that are associated with amyloid diseases. To probe the nature of such species, we describe here the results of experiments designed to establish the conformational properties of one of the lysozyme variants, D67H, over a range of different solution conditions.

Partially folded states of globular proteins are usually populated only marginally under physiological conditions and, therefore, are difficult to study by the majority of biophysical techniques. Measurement of hydrogen exchange behavior, however, has been used extensively to study aspects of protein structure and dynamics in such circumstances[10]. The method represents an exquisitely sensitive approach for detecting the properties of species present at low levels relative to the fully folded states of protein molecules. This approach is effective because any protein molecule that has converted, even on a single occasion, to an unfolded or partially folded state can, under appropriate conditions, be labeled irreversibly by specific isotopes present in the solvent molecules[11,12]. Here, we use NMR in conjunction with hydrogen exchange methods at pH 5.0 and 37 °C to probe the structure of the D67H variant of lysozyme and compare the data with that of the wild type protein. We then probe the hydrogen exchange properties of wild type lysozyme and the D67H variant by mass spectrometry (MS) over a range of pH values between 6.0 and 10.0. These experiments have allowed us to measure kinetic parameters describing the equilibrium between the native state and a partly folded state of the variant protein, and to describe the properties of the latter species at the level of individual residues. The results enable us to propose a specific mechanism for a structural transition by which the soluble form of this protein begins its conversion into amyloid fibrils.

NMR reveals dynamical effects of the D67H mutation

A series of 2D NMR experiments was carried out to monitor in a site-specific manner the hydrogen exchange behavior of wild type human lysozyme and the D67H variant. The experiments were carried out in D_2O at pH 5.0 using ^{15}N-labeled proteins. This pH value was chosen because the intrinsic exchange rates of amide hydrogens are close to their minimum values. Therefore, it should allow the maximum number of amide hydrogens to be observed in a series of ^{15}N-^{1}H HSQC NMR spectra recorded

[1]Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QH, UK. [2]Present address: Oxford Glycosciences UK Ltd., The Forum, 86 Milton Park, Abingdon, OX14 RY, UK. [3]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1EW, UK. [4]Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, UK. [5]Present address: School of Life and Environmental Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK. [6]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

Correspondence should be addressed to C.V.R. *email: cvr24@cam.ac.uk* and C.M.D. *email: cmd44@cam.ac.uk*

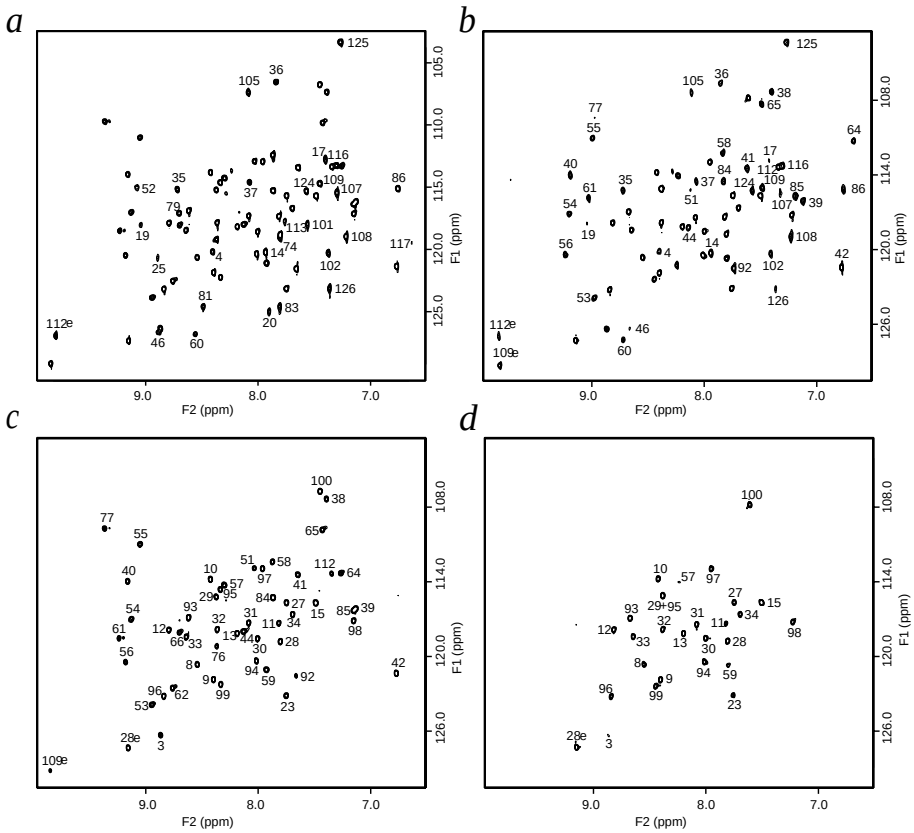

Fig. 1 ^{15}N-^{1}H HSQC NMR spectra of the wild type and D67H variant lysozymes. Spectra are recorded for the ***a***, wild type and ***b***, variant protein immediately after exposure to D_2O, and the ***c***, wild type and ***d***, variant protein after 3.5 h of exchange. Amide peaks visible immediately after exposure to D_2O are labeled in (*a*) and (*b*). Amide peaks that have not exchanged after 3.5 h are labeled in (*c*) and (*d*). The amides that have exchanged in the D67H variant but not in wild type lysozyme can be seen from the assignments to be located largely in the β-domain and the associated C-helix of the α-domain.

while hydrogen exchange is taking place. Representative spectra are shown for the two proteins immediately after exposure to D_2O and then after 3.5 h of exchange at pH 5.0 and 37 °C (Fig. 1). At the earliest acquisition time, crosspeaks from 79 and 66 amide hydrogens are visible in the HSQC spectra from the wild type protein and the D67H variant, respectively. The spectra recorded after 3.5 h illustrate qualitatively the differences in the hydrogen exchange behavior of the two proteins; the wild type protein has ~50 amide hydrogens that have not exchanged significantly after this length of time, whereas the D67H variant has only ~27.

To obtain quantitative information on the exchange rates of specific amide hydrogens in the two proteins, we have analyzed the heights of individual crosspeaks in the spectra as a function of the time of exchange. The rates resulting from this analysis were then used to obtain protection factors (Fig. 2), parameters that describe the extent to which hydrogen exchange of individual amide groups is slowed relative to that expected for the same residue in an unstructured protein[13] (see Methods). The ratios of the protection factors of the wild type protein to those of the D67H variant permit a comparison of the hydrogen exchange behavior in the two proteins. The results confirm that the effects of the D67H mutation on hydrogen exchange behavior are predominantly on amides from residues in the β-domain (defined as containing residues 41–82). For example, the exchange rates of four residues in this region increase by more than a factor of 1,000 in the D67H variant relative to the wild type protein, and four others, including residues in both main strands of the antiparallel β-sheet, increase by factors of 100 or more. In contrast, none of the amides in α-helices A, B and D show any significant difference in exchange rate between the two proteins. The exchange rates of amides in helix C are, however, significantly increased in the variant relative to the wild type protein. The exchange rate of one residue, Ala 96, is ~1,000× greater in the variant protein, and the rates of up to five amides are increased by more than a factor of 10. Helix C forms part of the α-domain but is located at the interface of the β-domain and linked to the latter by the (77-95) disulfide bond. The changes in exchange rates resulting from the D67H mutation are illustrated with respect to their position in the structure of the native protein (Fig. 3).

Unfolding of the D67H variant involves local cooperativity

To investigate the effect of the D67H mutation on the stability of the protein in greater detail, MS was used to analyze a series of samples of wild type lysozyme and the D67H variant. The labile hydrogens of both proteins were allowed to exchange with solvent in aqueous solution at 37 °C for a series of time intervals. The exchange process was investigated at four different pH values between 6.0 and 10.0, and also compared with data obtained previously at pH 5.0 (ref. 5). The structural properties of wild type lysozyme and the D67H variant under these pH conditions were found by CD spectroscopy to be similar (unpublished data). The samples were initially exposed to D_2O to replace all labile hydrogens with deuterium atoms; the subsequent exchange process involves the replacement of these deuterium atoms with hydrogen atoms from the solvent H_2O. The exchange reaction was then quenched after the desired time intervals by reducing the pH and temperature of the solution. Because experiments of this type are highly sensitive to small changes in the solution conditions, the D67H variant and the wild type protein were allowed to exchange simultaneously in the same solution[14]. By using isotopically ^{15}N-labeled wild type protein, the molecular weights of the two species are sufficiently distinct to resolve in mass spectra[15]; the data obtained for a solution of the proteins at pH 8.2 are shown as an example (Fig. 4). Samples pulse labeled for lengths of time, ranging from t = 0.4 to 45 s, are

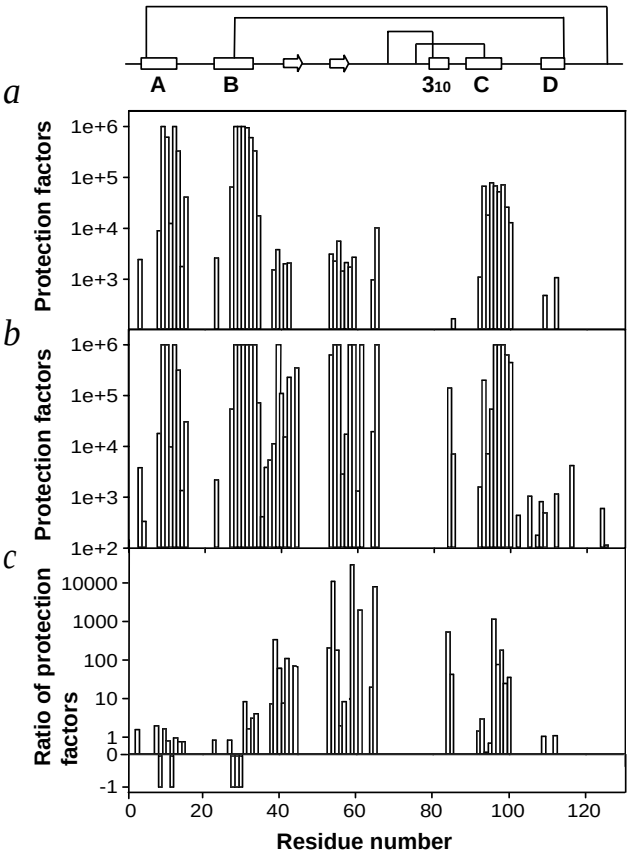

Fig. 2 Comparison of the protection factors for the D67H variant and wild type lysozyme. Protection factors are measured for ***a***, the variant and ***b***, wild type proteins. ***c***, The ratio of the protection factors of the wild type protein to those of the D67H variant. For those residues that did not show significant exchange during the time course of the experiment, we have used data recorded previously over a longer time scale for the wild type protein[31]. Amides whose exchange rates are accelerated by >10-fold relative to the wild type protein are located within residues 39–100 of the sequence.

shown along with the spectra of a sample not exposed to exchange conditions and one of a fully exchanged sample (defining the isotope labeling pattern at t = 0 and t = ∞, respectively).

The mass spectra show that the wild type protein is visible as a single peak whose mass decreases with the length of time that exchange was allowed to take place. This behavior is qualitatively similar at all the pH values studied (Fig. 5). The observation of a single peak with a gradual loss of mass over time reflects the exchange of labile hydrogens in the native wild type protein[16]. This process is usually described as hydrogen exchange under EX2 conditions, where the rate at which the unfolded structure returns to its protected state is much higher than the intrinsic exchange rate[10]. The observed exchange rate under these conditions can be interpreted using the relationship:

$$k_{obs} = k_o / k_c \times k_{ex} \quad (1)$$

where k_o is the opening and k_c the closing rate for a fluctuation permitting exchange, and k_{ex} is the intrinsic exchange rate of a labile hydrogen in the exposed form of the protein structure. To further test the validity of the EX2 description of hydrogen exchange for the wild type protein, the results of the whole series of experiments carried out during this investigation are plotted (Fig. 5) after scaling of the data sets by a factor of 10 for each unit of pH at which the exchange takes place[17]. The latter factor is required by the assumption of ideal EX2 kinetics as a result of base catalysis of the exchange reaction[13]. The close overlap of the data obtained at the different pH values indicates that this assumption is broadly consistent with the experimental measurements.

The data corresponding to the D67H variant, in contrast to those of the wild type protein, show a clear bimodal distribution of masses as the exchange process takes place (Fig. 4). Such a distribution is indicative of the limiting case of EX1 hydrogen exchange behavior[16,18], in which the closing rate of a fluctuation enabling hydrogen exchange to take place (k_c) is much lower than the intrinsic exchange rate — that is, $k_c << k_{ex}$. Hence, $k_{obs} = k_o$ (refs 16,18) and is usually observed only at high pH, where the intrinsic exchange rates are high. Under these conditions, effectively all of the labile hydrogens that are exposed will undergo exchange each time a fluctuation occurs. Such behavior is usually associated with the complete global unfolding of the protein structure. If, however, a fluctuation results in the cooperative unfolding of a region of the protein molecule, a well-resolved multimodal distribution of masses could result. In the present case, although the mass difference between the peaks is too small to correspond to complete global unfolding, the clear bimodal distribution observed at pH values between 8.2 and 10.0 indicates that a cooperative local fluctuation of a significant region of the protein takes place. At the lower pH values of 5.0 and 6.1, only a single broad peak can be observed, indicating that the intrinsic exchange rate at these lower pH values is not fast enough to compete with the closing rate, and the process ultimately obeys EX2 kinetics. The exchange data measured at the higher pH values, however, reveal a high degree of cooperativity in the fluctuations leading to exchange of the labile hydrogens in a well-defined substructure of the D67H variant.

To characterize further the specific process giving rise to this locally cooperative hydrogen exchange behavior, the areas under the two peaks in the mass spectra observed at high pH values were integrated and the fraction of the lower mass species calculated as a function of time. The time dependence of the intensity change of this peak provides a direct measure of the opening rate of the cooperative fluctuation, giving rise to exchange. The

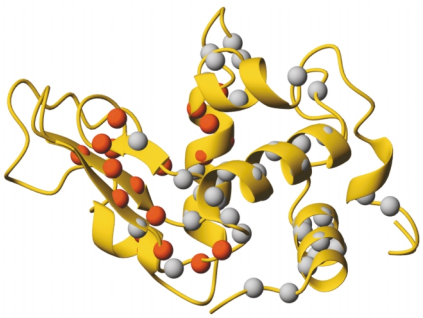

Fig. 3 Acceleration of hydrogen exchange rates in a substructure of the lysozyme variant. Backbone amides that exchanged sufficiently slowly and could be studied by NMR at pH 5.0 and 37 °C are shown in a space-filling representation. Those shown in red denote backbone amides whose exchange is accelerated by more than a factor of 15 relative to the wild type protein under these conditions. The lysozyme structure was generated from the X-ray coordinates (PDB entry 1REX) and produced using MOLMOL[35].

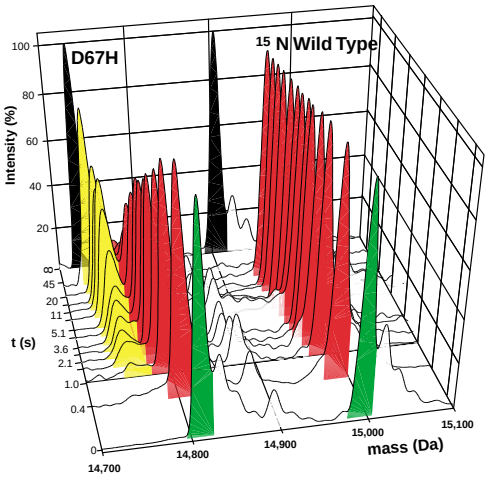

Fig. 4 Electrospray mass spectra of wild type and variant lysozymes demonstrating the unfolding dynamics. Mass spectra were recorded for an approximately equimolar mixture of ^{15}N-labeled wild type lysozyme and the D67H variant following exposure to hydrogen exchange conditions at pH 8.2 and 37 °C. The spectra represent mass transformed data from the +9 charge state. The peaks colored yellow were observed in spectra of the D67H variant but not in the wild type protein and are formed as a result of the locally cooperative unfolding event. The peaks colored red arise from the gradual loss of deuterium during the course of the exchange reaction. The peaks in the spectra of control samples recorded for time zero and time infinity are colored green and black, respectively.

kinetic data at the three pH values where bimodal distributions are evident (Fig. 5*b*) were fit to exponential functions, yielding an average time constant for the unfolding process of 5.8 ± 3.2 s. The mass differences between the two peaks were also measured as a function of time (Fig. 5*b*). As the exchange reaction progresses, the observed mass difference gradually becomes smaller. This observation reflects the existence of additional highly localized fluctuations in the structure that can give rise to hydrogen exchange. Such events may be more prevalent in the less stable D67H variant than in the wild type protein[5]. As these exchange processes reduce the total number of protected amides, the cooperative unfolding process will seem to involve a smaller number of residues as time increases. To obtain a value for the total number of hydrogens protected initially and involved in the cooperative unfolding process, the hydrogen exchange data at each pH value were extrapolated to zero time (Fig. 5*b*). The results from this procedure show that over the pH range 8.2–10.0, the average number of hydrogens involved in the cooperative fluctuation is 34.5 ± 1.7.

Local unfolding of the domain interface in the variant

To locate the regions of the structure of the D67H variant involved in the cooperative unfolding processes, a sample of the protein was exposed to hydrogen exchange conditions to generate two populations of approximately equal intensity. This sample was then subjected to peptic digestion and mass spectrometric analysis. The rationale behind this experiment is to generate a number of peptide fragments from the protein sample. From an analysis of the H/D distribution in a sufficient number of fragments, identifying regions of the protein that have undergone exchange as a result of partial unfolding should be possible[19]. Before these experiments, a reference sample of the variant protein not exposed to D_2O was subjected to peptic digestion. In the mass spectra, 19 peptides were identified that give complete sequence coverage of the protein. The assignment of these fragments was achieved by using a combination of accurate mass measurement and fragmentation strategies[20]. The results of this approach were then verified by carrying out peptic digestion of uniformly ^{15}N-labeled D67H lysozyme; the increase in mass resulting from the change in isotopic content enabled the number of nitrogen atoms in each peptide fragment to be defined[15]. The principal proteolytic fragments evident during the earliest time points of the digestion were assigned to three major regions of the protein (Fig. 6). Two of these peptide fragments are linked by native disulfide bonds that remain intact during digestion.

After identification of the peptide fragments resulting from digestion, the partially exchanged protein sample was subsequently digested and examined. From the masses that were measured, the average number of protected sites within each peptic fragment was calculated. For some peptides, two populations are clearly visible. For these peptides, the experimental isotopic distributions have been compared with simulations on the

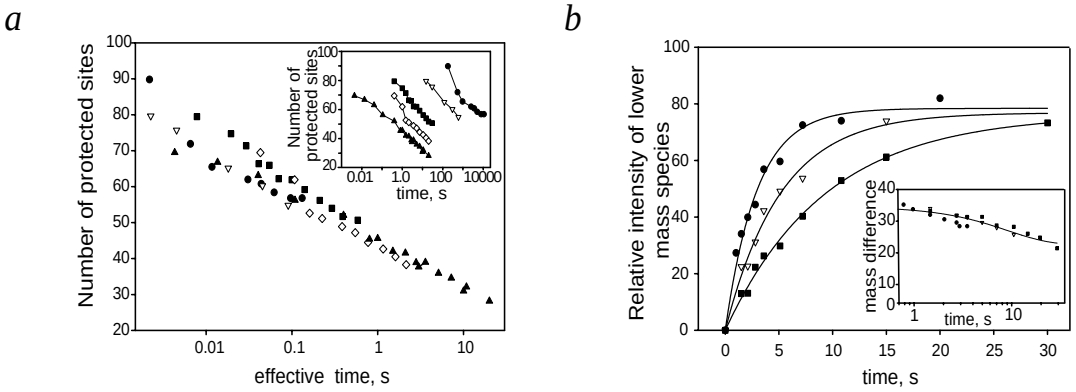

Fig. 5 Hydrogen exchange protection as a function of time. ***a***, The number of protected sites is plotted as a function of time (on a logarithmic scale) in the wild type protein at different pH values (inset) and after transposition to pH 10, assuming an EX2 mechanism (see text). pH values are 5.0 (filled circle), 6.1 (open triangle), 8.2 (filled square), 9.0 (open diamond) and 10.0 (filled triangle). ***b***, Plot of the relative intensity of the lower mass species observed in hydrogen exchange experiments involving the D67H variant (Fig. 4) as a function of time of exchange. These data give a measure of the unfolding kinetics at pH 8.2 (filled square), 9.0 (open triangle) and 10.0 (filled circle) as 9.3 s, 5.2 s and 3.0 s, respectively. Inset: A plot of the mass difference between the two species detected in the D67H variant (Fig. 4) as a function of exchange time (on a logarithmic scale). The value of this mass difference (34.5 ± 1.7), extrapolated to time = 0, represents the number of hydrogen exchange sites protected against exchange to an extent that they can be detected by MS at pH 8.2 that are involved in the cooperative unfolding event. The total number of protected hydrogens that can be detected by MS under these conditions in the native protein is 80 ± 2.

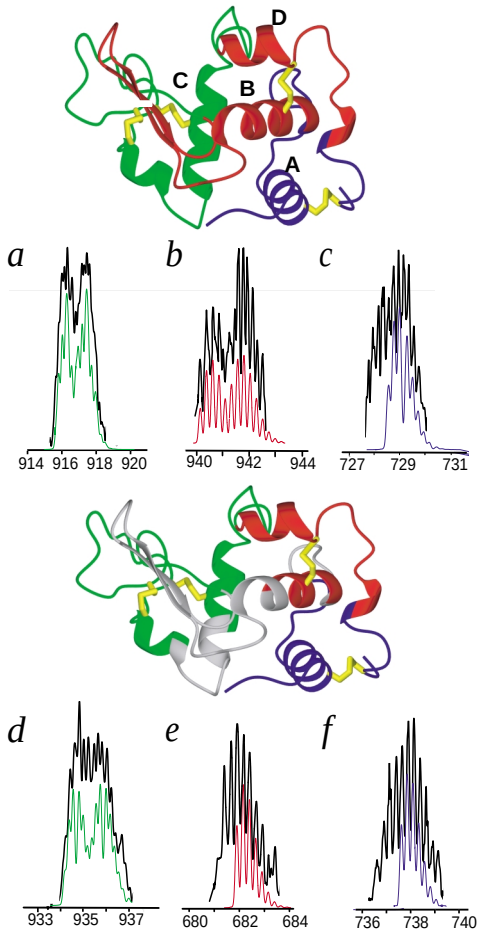

Fig. 6 Electrospray mass spectra of the partially exchanged peptide fragments of the D67H variant. The multiply-charged peptides are assigned as ***a***, (58–108)$^{+6}$; ***b***, ((26–40)–(109–123))$^{+5}$; ***c***, ((1–25)–(124–130))$^{+4}$; ***d***, ((58–84)–(93–108))$^{+5}$; ***e***, ((26–31)–(109–123))$^{+4}$; and ***f***, ((1–19)–(124–130))$^{+4}$. Before digestion, the protein was exposed to hydrogen exchange at pH 10 for a few seconds. The location of the principal peptic digestion products is shown within the native structure of lysozyme in the upper panel. Similarly, the smaller peptide fragments, derived from three regions (*d*), (*e*) and (*f*), are indicated in the lysozyme structure in the lower panel. The lysozyme structure was generated from the X-ray structure (1REX) and produced using MOLMOL[35]. The peaks in the spectra are compared with simulations of the isotope patterns color-coded for each peptide fragment. For two populations undergoing EX1 exchange, two binomial distributions of isotopes were calculated using isotope-modeling software (Masslynx, Micromass). One distribution corresponds to the natural abundance isotopes and the other to the natural abundance isotope plus the number of deuterium atoms calculated from the average mass difference between the two peaks. Addition of the two isotope distributions was then carried out using Sigmaplot v4.01 (SPSS). For EX2 exchange, only one isotope distribution was calculated, which corresponded to the isotopic abundance with the average number of deuterium atoms calculated from the m / z value of the peak. No attempt was made to model the gradual loss of hydrogen exchange protection that leads to broadening of the isotope pattern because of partial occupancy of amide sites[34]. The peptide fragments (*a*), (*b*) and (*d*) fit well to an exchange model in which the total number of protected sites undergo exchange exclusively *via* an EX1 mechanism to give two distinct hydrogen exchange populations. The peptide fragments (*c*), (*e*) and (*f*) show a single hydrogen exchange population that undergoes exchange *via* an EX2 mechanism.

basis of an EX1 exchange mechanism and the measured number of protected sites. The bimodal isotope distribution corresponding to peptide 58–108, which includes the C helix (residues 89–100) as well as residues 58–64 from the β-domain, closely matches a simulation in which 10 labile hydrogens from this region are involved in the EX1 unfolding event. However, the isotope distribution in the shorter fragment from this region (58–84)–(93–108) (peptides 58–84 and 93–108 linked by a disulfide bond) is best described by a simulation for an EX1 exchange process involving eight labile hydrogens. Thus, removal of the residues 85–92, which includes four residues from the C-helix, involves a reduction of two in the number of protected hydrogens involved in the EX1 unfolding event. Previous hydrogen exchange studies of human lysozyme have shown that only residues 85 and 92 are highly protected in this region of the sequence[21]. The present results of the MS analysis, therefore, indicate that these two residues, located in the C-helix and the loop preceding it, are involved in the cooperative unfolding transition detected by hydrogen exchange.

The peptide fragment (26–57)–(109–123), which includes the residues forming the B- and D-helices (26–34 and 109–115, respectively) of the α-domain and residues 41–57 of the β-domain, also shows a distinct bimodal distribution, corresponding to two hydrogen exchange populations involving 10 backbone amides (data not shown). The isotope distribution in an additional peptide (26–40)–(109–123), which differs by the absence of residues 41–57 in the β-domain, is also bimodal and correlates well with a simulation involving five backbone amide hydrogens. For a yet shorter peptide fragment derived from this region, (26–31)–(109–123), in which residues 32–40 are removed, the isotope distribution shows no evidence of EX1 exchange (Fig. 6). The distribution is well described by a single hydrogen exchange population having an average of four protected hydrogens that exchange *via* an EX2 mechanism. Thus, comparison of the exchange properties of a series of peptides generated from the sequence (26–57)–(109–123) indicates that the amide hydrogens undergoing EX1 exchange in this region are located within the sequence 31–57, corresponding to the four C-terminal residues of the B-helix and encompassing the antiparallel β-sheet of the β-domain. In contrast, the peptide fragments containing the sequence 26–40, which includes the residues forming helix B, are consistent with a single hydrogen exchange population with an average of eight protected amides. Overall, the MS analysis of these peptide fragments indicates that the regions involved in the cooperative unfolding process are all located within the sequence 31–104. This segment of the protein includes all the residues forming the β-domain and the C-helix of the α-domain.

Intermediate resembles species detected during folding

The hydrogen exchange experiments allow the comparison of the fluctuations in the structure of the D67H variant protein of human lysozyme with those characteristic of the wild type protein. The NMR spectra detailed above were recorded under conditions (pH 5.0 and 37 °C) where the variant protein has been found to form amyloid fibrils *in vitro* over longer periods of time[5]. Therefore, these experiments probe the species present in solution before the aggregation event. The results show that the rate of hydrogen exchange is accelerated >10-fold for 18 amide hydrogens located in the sequence between residues 39 and 100. In a second series of experiments, in which the pH of the solution was increased, the localized unfolding of a region of the variant protein was found by MS analysis to be highly cooperative. Of the 80 amide hydrogens that are substantially protected in the native structure, ~35 are located in the specific region of the protein undergoing the cooperative unfolding process. The region of the structure involved in this transition was found to

be between residues 31 and 104 by using selective proteolytic digestion procedures and then exploiting the unique power of MS to detect coherent exchange behavior[16]. This finding is consistent with the conclusion from the results of the NMR experiments; the same regions of the protein where accelerated rates of hydrogen exchange are observed at pH 5.0 are those found to be involved in the locally cooperative unfolding process. Moreover, the present results are supported by earlier observations in which an intermediate population was generated during unfolding of the variant protein and subsequently detected by CD measurements[5]. Taken together, these results indicate that a partially structured intermediate, in which the β-domain residues (41–82) and the adjacent C-helix (89–100) are substantially destabilized relative to the rest of the protein structure, is transiently populated under fibril-forming conditions.

The molecular detail we have been able to gain in this study indicates that the hydrogen bonded β-sheet structure of the native protein is substantially disrupted in the partially folded state formed by the D67H variant. The major hydrophobic core of the protein, located in the α-domain of the structure, retains its ability to protect amide hydrogens, except those in the N-terminus of the C-helix. The destabilization of the C-helix is probably associated with the existence of the disulfide bond 77-95 linking this helix to the β-domain. The unfolding process that results in formation of this transient species, therefore, involves disruption of the interface or hinge region between the two structural domains of the protein; this process is effectively the reverse of that observed to be the final step in the folding of the protein[22]. Indeed, the partially folded species characterized here, which results from this unfolding behavior, seems to be closely similar to the major intermediate that accumulates during the folding of lysozyme from its chemically denatured state[23]. For the D76H variant at pH 5.0 and 37 °C, approximately one molecule in 1,500 is in this partially folded state[5]. Even though the existence of a similar partially folded state for the wild type protein cannot be ruled out, we estimate from the hydrogen exchange data reported here that under similar conditions, fewer than one molecule in 100,000 is in an analogous state.

Origin of the amyloidogenic behavior of D67H

One of the most important observations in the present work is that the MS data show that fluctuations associated with localized unfolding of the β-domain and the adjacent C-helix are highly cooperative. In other words, although the fluctuations detected by the hydrogen exchange measurements are only transient, the regions of the structure that are affected by them are simultaneously exposed to the solvent. In the resulting partially folded species, however, the remainder of the protein is highly structured. This result provides an important clue to the possible significance of such a species in amyloidogenesis. It is clear from a range of studies that the formation of amyloid structures involves the aggregation of peptides or proteins in states that are completely or partially unfolded[9,24]. In the case of peptide fragments of proteins, such as the 40- or 42-residue Aβ peptide that is cleaved from the amyloid precursor protein, their inability to fold to a globular state in the absence of the other regions of the precursor protein results in a substantial population of species that are able to aggregate through intermolecular hydrogen bonding involving the polypeptide main chain. In the case of intact globular proteins that form amyloid structures, the most aggregation prone states are probably those that simultaneously expose significant lengths of polypeptide chain in extended structures to the solvent, enabling several interactions to occur within similar exposed regions in other molecules.

In the native state of a globular protein, the high cooperativity of the overall fold buries the polypeptide main chain within the structure and generally prevents the development of intermolecular interactions involving this region of the protein[25]. Fluctuations in the native structure that can give rise to hydrogen exchange but expose only very local regions of the polypeptide chain to solvent are unlikely to generate, even transiently, conformations of the protein able to form strong intermolecular interactions. In accord with this conclusion, conditions that readily stimulate formation of amyloid fibrils are generally those that partially denature the protein and expose a large section of the polypeptide backbone to solvent[24]. For D67H lysozyme, the amyloidogenic protein studied here, not only is the population of a partially folded intermediate much greater than that of the wild type protein under similar conditions, but the cooperativity of the local unfolding process results in most of the β-domain and the C-helix being unfolded simultaneously, albeit transiently, due to occasional fluctuations in the otherwise native structure. This combination of factors could give rise to a sufficient concentration of soluble, partially folded species capable of forming strong intermolecular interactions to initiate aggregation under milder conditions than would be needed to generate a similar concentration of completely unfolded protein molecules.

Conclusions

We believe that the common, perhaps universal, ability of polypeptides to form amyloid fibrils with similar morphological properties[25,26] indicates that the specific structure of the partially folded intermediate formed by the amyloidogenic variant is unlikely to be critical in defining the characteristic cross-β structure of the fibrillar assembly. Nevertheless, the properties of the variant are probably crucial for allowing such structures to form more readily than for the wild type protein under conditions experienced during the lifetime of an otherwise normally functional biological system[27]. The properties of the intermediate are, however, likely to be crucial in determining the regions of the polypeptide chain in the core cross-β structure and in defining the detailed manner in which the mature fibrils are assembled. This conclusion reinforces the view that the remarkable cooperativity of native protein structures is an essential evolutionary development to enable otherwise marginally stable structures to resist aggregation under conditions in which they exert their biological function[28].

Methods

Protein expression and purification. Wild type human lysozyme and the D67H mutational variant, including proteins uniformly labeled in ^{15}N, were expressed in *Aspergillus niger* and purified as described[29,30]. Guanidinium hydrochloride (GuHCl), sodium deuteroxide and deuterium chloride were obtained from Sigma. Deuterium oxide (D_2O, 99.9%) was from Fluorochem. Acetic acid-d_4 (DLM-12, 99.5%) was from Cambridge Isotope Laboratories. GuHCl was deuterated at exchangeable sites by three cycles of dilution in D_2O followed by lyophilization.

Real time 2D NMR spectroscopy of ^{15}N-labeled proteins. ^{15}N-labeled wild type human lysozyme or the D67H variant protein (5 mg) was dissolved in 575 μl of a 20 mM acetic acid solution in D_2O at pH* 5.0 (the value uncorrected for isotope effects). The exact pH values of the samples were determined at the end of each experiment and used in the calculations of protection factors. Series of gradient-enhanced ^{15}N-^1H HSQC spectra were collected at 37 °C in a

600 MHz NMR spectrometer. Each spectrum took 30 min to record, with 96 complex t_1 increments of eight transients and 1,000 data points. Sweep widths of 8,000 Hz and 1,825 Hz were used in the F2 and F1 dimensions, respectively. Data were processed with a Gaussian window function in F2 and a 90° shifted sinebell in F1. After zero filling, the digital resolution was 3.9 and 3.6 Hz pt^{-1} in F2 and F1, respectively. Hydrogen exchange rates were monitored by measuring peak heights in the HSQC spectra as a function of time. Sigmaplot v4.01 (SPSS) was used to fit exponential functions to the data. The predicted exchange rates for the same sequence in a random coil conformation[13] were divided by the measured rates to calculate protection factors. For those residues that did not show significant exchange during the time course of the experiment in both the D67H variant and the wild type protein, the rate enhancement factor is plotted as –1 (Fig. 2). For those residues that did not show significant exchange during the time course of the experiment for the wild type protein but did for the variant, we have used previous data for the wild type protein recorded at pH 5.5 and 35 °C for longer time periods[31] to calculate the rate enhancement for the variant protein. Assignments for the wild type protein were based on published ^{1}H and ^{15}N assignments[21,32]. Additional 2D COSY and NOESY spectra were recorded at 20 °C and 37 °C on samples of wild type and variant proteins prepared in D_2O or in H_2O to obtain assignments for the peaks observed in the HSQC spectra.

Mass spectrometry. The ^{14}N D67H variant and ^{15}N wild type lysozyme were deuterated at exchangeable sites by unfolding in deuterated guanidinium hydrochloride (GuDCl) followed by refolding in deuterated buffer (50 mM deuterated acetic acid, pH* 5.0). The proteins were subsequently concentrated, and the buffer was exchanged into 10 mM deuterated acetic acid, pH* 5.0, using Centriprep-10 concentrators (Amicon, Millipore). For the comparative experiments, solutions of the two proteins were mixed in equal proportions.

The pulse-labeling hydrogen exchange experiments were carried out using a Bio-Logic QFM5 mixer connected to a circulating water bath set to 37 °C. The protein solution (1 volume) was mixed with 15 volumes of 100 mM ammonia/formic acid buffer in H_2O at a given pH, and the exchange was allowed to proceed for various times between 0.4 s and 600 s. The measured pH values of the labeling pulses in the four sets of experiments were 6.1, 8.2, 9.0 and 10.0. The resulting solutions were then mixed with 7 volumes of an acetic acid solution in H_2O, whose concentration was adjusted to ~1 M to generate a pH of 3.5 for the solutions after the quench. The samples were collected and placed on ice before analysis by MS. Two control samples were prepared by mixing 1 volume of protein solution with 22 volumes of 20 mM acetic acid/ammonia buffer in H_2O at pH 5.0. One of these samples was left at room temperature for 15 min (initial protection sample) and subsequently placed on ice, whereas the other one was incubated at 80 °C for 15 min (fully exchanged sample) and subsequently placed on ice. We have also incorporated data obtained in a hydrogen exchange experiment conducted at pH 5.0 (Fig. 5)[5].

Mass spectra of pulse-labeled samples were recorded using a Micromass LC-ToF spectrometer (Micromass). Samples were introduced using a nanoflow electrospray probe with borosilicate glass capillaries[33] prepared in house. For each experiment, 2 μl of solution containing 3 μM protein was used. Mass spectra (Fig. 4) represent the +9 charge state with minimal smoothing, converted to a mass scale. Data analysis was performed using Grams 32 V5.10 (Galactic Industries). The spectra were fit with a linear baseline and a combination of Gaussian distributions to determine the relative proportions of the different species. The calculated areas for each peak from the fit were analyzed as a function of time and used to extract kinetic parameters for the exchange reactions using Sigmaplot v4.01 (SPSS). No correction was made to the number of exchangeable amides for the 6% deuterium present in the labeling pulse.

For the sequence-specific mass spectrometric analysis, spectra were recorded on a Micromass Q-ToF mass spectrometer (Micromass). The variant protein was subjected to peptic digestion at low pH, and assignment of proteolytic fragments was carried out on samples not exposed to the labeling pulse. These assignments were confirmed by carrying out similar experiments with ^{15}N-labeled protein to determine the number of nitrogen atoms in the peptide[15]. Full experimental details of the complete assignment process are described elsewhere[20]. Assigning peptides with masses up to ~600 Da to regions of the protein sequence was possible simply on the basis of mass measurement. Tandem MS approaches were used for peptides with masses between ~600 and ~3,000 Da. Peptides with masses >3,000 Da were isolated by HPLC and subjected to tryptic digestion to enable unambiguous sequence determination. Individual samples of the D67H variant protein, pulse labeled at pH 10.0 and 37 °C, were prepared manually in a similar manner as the samples pulse labeled in the quench-flow experiment. A few seconds of labeling ensured an equal proportion of the two peaks detected in the mass spectra under these conditions. Isotope modeling was carried out as described[34].

Acknowledgments

D.C. was supported by a fellowship from the European Community, C.R. is supported by a BBSRC Advanced Research Fellowship, and C.V.R. and M.S. hold Royal Society University Research Fellowships. This work is in part a contribution from the Oxford Centre for Molecular Sciences, which is supported by EPSRC, BBSRC and MRC. The research of C.M.D. is also supported in part by a programme grant from the Wellcome Trust.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 5 October, 2001; accepted 24 January, 2002.

1. Koo, E.H., Lansbury, P.T. & Kelly, J.W. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. USA* **96**, 9989–9990 (1999).
2. Perutz, M.F. Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends Biochem. Sci.* **24**, 58–63 (1999).
3. Tan, S.Y. & Pepys, M.B. Amyloidosis. *Histopathology* **25**, 403–414 (1994).
4. Pepys, M.B. *et al.* Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362**, 553–557 (1993).
5. Booth, D.R. *et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793 (1997).
6. Morozova-Roche, L.A. *et al.* Amyloid fibril formation and seeding by wild type human lysozyme and its disease-related mutational variants. *J. Struct. Biol.* **130**, 339–351 (2000).
7. Goda, S. *et al.* Amyloid protofilament formation of hen egg lysozyme in highly concentrated ethanol solution. *Protein Sci.* **9**, 369–375 (2000).
8. Krebs, M.R.H. *et al.* Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the β-domain. *J. Mol. Biol.* **300**, 541–548 (2000).
9. Kelly, J.W. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8**, 101–106 (1998).
10. Englander, S.W. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238 (2000).
11. Roder, H., Elove, G.A. & Englander, S.W. Structural characterization of folding intermediates in cytochrome *c* by H exchange labeling and proton NMR. *Nature* **335**, 700–704 (1988).
12. Raschke, T.M. & Marqusee, S. Hydrogen exchange studies of protein structure. *Curr. Opin. Biotechnol.* **9**, 80–86 (1998).
13. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Genet.* **17**, 75–86 (1993).
14. Hooke, S. Cooperative elements in protein folding monitored by electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* **117**, 7548–7549 (1995).
15. Miranker, A., Kruppa, G.H., Robinson, C.V., Aplin, R.T. & Dobson, C.M. Isotopic-labeling strategy for the assignment of protein fragments generated for mass spectrometry. *J. Am. Chem. Soc.* **118**, 7402–7403 (1996).
16. Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T. & Dobson, C.M. Detection of transient protein folding populations by mass spectrometry. *Science* **262**, 896–899 (1993).
17. Wang, L., Lane, L.C. & Smith, D.L. Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci.* **10**, 1234–1232 (2001).
18. Engen, J.R., Smithgall, W.H., Gmeiner, T.E. & Smith, D.L. Identification and localization of slow, natural, cooperative unfolding in the hematopoietic cell kinase SH3 domain by amide hydrogen exchange and mass spectrometry. *Biochemistry* **36**, 14384–14391 (1997).
19. Zhang, Z. & Smith, D.L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522–531 (1993).
20. Tito, P. & Robinson, C.V. Hydrogen exchange of proteins in partially folded states: a quadrupole time-of-flight approach. *Methods Enzymol.* **In the press** (2002).
21. Redfield, C. & Dobson, C.M. ^{1}H NMR studies of human lysozyme: spectral assignment and comparison with hen lysozyme. *Biochemistry* **29**, 7201–7214 (1990).
22. Dobson, C.M., Evans, P.A. & Radford, S.E. Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37 (1994).
23. Hooke, S.D., Radford, S.E. & Dobson, C.M. Refolding of human lysozyme — a comparison with the structurally homologous hen lysozyme. *Biochemistry* **33**, 5867–5876 (1994).
24. Dobson, C.M. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond. A/B* **356**, 133–145 (2001).
25. Chiti, F. *et al.* Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. USA* **96**, 3590–3594 (1999).
26. Guijarro, J.I.I., Sunde, M., Jones, J.A., Campbell, I.D. & Dobson, C.M. Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95**, 4224–4228. (1998).
27. Chamberlain, A. *et al.* Ultrastructural organization of amyloid fibrils by atomic force microscopy. *Biophys. J.* **79**, 3282–3293 (2000).
28. Dobson, C.M., Sali, A. & Karplus, M. Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed. Engl.* **37**, 868–893 (1998).
29. Canet, D. *et al.* Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease related variants. *Biochemistry* **38**, 6419–6427 (1999).
30. Spencer, A. *et al.* Expression, purification and characterisation of the recombinant calcium-binding equine lysozyme secreted by the filamentous fungus *Aspergillus niger*. *Protein Exp. Purif.* **16**, 171–180 (1999).
31. Woodruff, N. PhD. Thesis. Investigation of protein structure and folding by NMR spectroscopy. (University of Oxford; 1998).
32. Ohkubo, T., Taniyama, Y. & Kikuchi, N. ^{1}H and ^{15}NMR study of human lysozyme. *J. Biochem.* **110**, 1022–1029 (1991).
33. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1–8 (1996).
34. Chung, E.W. *et al.* Hydrogen exchange properties of proteins in native and denatured states monitored by mass spectrometry and NMR. *Protein Sci.* **6**, 1316–1324 (1997).
35. Koradi, R., M, B. & Wuthrich, K. MOLMOL: A program for the display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).