

Simulations of Human Lysozyme: Probing the Conformations Triggering Amyloidosis

George Moraitakis and Julia M. Goodfellow

School of Crystallography, Birkbeck College, University of London, London WC1E 7HX, United Kingdom

ABSTRACT A natural mutant of human lysozyme, D67H, causes hereditary systemic nonneuropathic amyloidosis, which can be fatal. In this disease, insoluble β -stranded fibrils (amyloids) are found in tissues stemming from the aggregation of partially folded intermediates of the mutant. In this study, we specifically compare the conformation and properties of the structures adopted from the induced unfolding, at elevated temperature, using molecular dynamics. To increase the sampling of the unfolding conformational landscape, three 5 ns trajectories are performed for each of the wild-type and mutant D67H proteins resulting in a total of 30 ns simulation. Our results show that the mutant unfolds slightly faster than the wild-type with both wild-type and mutant proteins losing most of their native secondary structure within the first 2 ns. They both develop random transient β -strands across the whole polypeptide chain. Clustering analysis of all the conformations shows that a high population of the mutant protein conformations have a distorted β -domain. This is consistent with experimental results suggesting that this region is pivotal in the formation of conformations prone to act as “seeds” for amyloid fiber formation.

INTRODUCTION

Human lysozyme is a 130-residue protein found in secretions (e.g., saliva, sweat, and mucus) and more generally in leukocytes and kidneys. It is an enzyme that hydrolyzes preferentially the β -1,4 glucosidic linkages between *N*-acetylmuramic acid and *N*-acetylglucosamine that occur in the mucopolysaccharide cell wall structure of certain microorganisms (Chipman and Sharon, 1969). The wild-type human lysozyme has been crystallized and its structure elucidated by Artymiuk and Blake (1981) at 1.5 Å resolution (Fig. 1, *left*). Its native structure consists of two domains: an α -domain that has four α -helices (A–D) and one 3_{10} helix, and a β -domain, which consists mainly of an antiparallel β -sheet and a long loop. The active site is located in the cleft that is formed between these two domains. The protein contains four disulphide bonds of which two are located in the α -domain, one in the long loop of the β -domain, and one that connects the two domains.

There are two known natural mutations of the human lysozyme: D67H and I56T (Pepys et al., 1993). They both cause autosomal dominant hereditary nonneuropathic systemic amyloidosis. This is a condition whereby there is tissue deposition in viscera and other body cavities of normally soluble autologous proteins as insoluble fibrils called amyloids. The core of these fibrils structure consists of β -sheet with the strands perpendicular to the long axis of the fiber (Pepys, 1996). Amyloids can be formed from proteins of diverse sequence, fold, and function and are known to lead to serious medical conditions such as Alzheimer's disease

and spongiform encephalopathies (Kelly, 1998). Although the mechanism of fibrillogenesis is not clear, it has been suggested that it is related to changes in stability and tendency to aggregate due to mutations. More specifically, Booth et al. (1997) proposed that a partially folded transient population of the amyloidogenic proteins, which lacks global cooperativity, undergoes structural transformation (a helix-to-sheet transition) and creates the first template, the “seed”, for further protein deposition and fibril formation. In a recent paper, Morozova-Roche et al. (2000) show that the presence of these “seeds” for both wild-type and the two natural mutants of human lysozyme facilitate the formation of fibrils.

We focus on the behavior of the D67H variant. The crystallized structure of this mutant (Booth et al., 1997) at 1.75 Å resolution is quite similar to that of the wild-type (Fig. 1, *right*). However, because of the mutation, the network of hydrogen bonds that stabilizes the β -domain is destroyed. This results in a large concerted movement of the β -sheet and the long loop within the β -domain. This distortion of the β -sheet propagates partially downstream to Ile⁵⁶, the residue at the interface of the two domains, for which increased B-factors have been found.

To understand the behavior of the partially folded structures that may trigger amyloid formation, we explore the conformations adopted during the induced unfolding of the wild-type and mutants at high temperatures (500 K) using molecular dynamics (MD) simulations. The temperature denaturation of proteins using MD is considered as one of the most straightforward computational experiments (Brooks, 1998). The great advantage of unfolding simulations is that they allow the investigation of the conformational properties at every point along the unfolding pathway (Li and Daggett, 1994). There are many examples of their use (Daggett, 2000) and these have led to detailed insight of experimental results of protein unfolding (Li and Daggett, 1998; Alonso and Daggett, 2000) and the study of

Submitted March 18, 2002, and accepted for publication September 18, 2002.

Address reprint requests to Julia M. Goodfellow, School of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK. Tel.: 44-01793-413208; E-mail: julia.goodfellow@bbsrc.ac.uk.

© 2003 by the Biophysical Society

0006-3495/03/04/2149/10 \$2.00

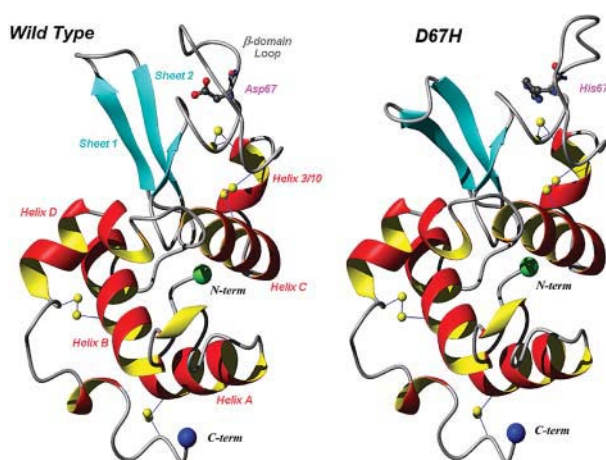


FIGURE 1 Crystal structures of the wild-type (left) and the D67H (right) human lysozyme and the mutation site (residue 67) and Ile⁵⁶, located in the interface of the α - and β -domains, are shown as ball and sticks. The sulfur atoms of the disulphide bonds as well as the N- and C-termini atoms are shown as spheres.

the folding pathway (Karplus and Sali, 1995). Lysozyme from hen egg white has been extensively investigated by means of unfolding simulations (Mark and van Gunsteren, 1992; Hünenberger et al., 1995; Williams et al., 1997; Kazmirski and Daggett, 1998; Gilquin et al., 2000) with the aim to investigate several issues such as the stability and folding of the protein. The human and hen forms of lysozyme have 60% sequence similarity but very similar 3D structure.

In our study, we attempt to examine and compare the structure and properties of the partially folded intermediates obtained from the induced unfolding of the wild-type and the D67H form.

MODEL AND SIMULATION DETAILS

The crystal structures of the wild-type (1REX) and the mutant (1LYY) lysozyme are the starting point of the simulations. Hydrogen atoms are added and the final system includes solvent molecules that are represented by the SPC216 model (Berendsen et al., 1981). All atoms are explicitly represented. The solvated model is contained in a rectangular box ($70 \times 70 \times 70$ Å) using periodic boundaries conditions (Allen and Tildesley, 1987). The building of the protein models and all their simulations are carried out with the GROMACS suite of programs (Berendsen et al., 1995). The GROMOS 96 force field is used to describe the atomic interactions (van Gunsteren et al., 1996). For the correct treatment of long-range electrostatics, we make use of the particle mesh Ewald summation algorithm (Darden et al., 1993). The high frequency degrees of freedom from the covalent bonds of hydrogen atoms are constrained using the LINCS algorithm (Hess et al., 1997) and thus the time step is increased from 1 to 2 fs. The system is coupled to an external temperature bath with a separate bath for the solvent and the solute (Berendsen et al., 1984). The regulation of the pressure is achieved by means of a pressure bath. Data on the trajectories are saved every 0.2 ps. For these simulations, we use an in-house multi-processor Origin 2000 with four CPUs. The total central processing unit time for all simulations was ~ 48 days.

To minimize the initial system, we use a combination of the conjugate gradients and steepest descent methods in which after every 50 steps of

conjugate gradients, one step of steepest descent is performed. The minimization is terminated when the overall force of the system is 100 N or after 5000 steps. This protocol is first used to minimize hydrogen atom and water molecules positions and then extended to the whole system. To further optimize the arrangement of the solvent around the protein and alleviate high energy regions (hot spots), the water molecules only are assigned initial velocities from a Gaussian distribution generated from a random seed and then warmed up from 50 K to 300 K over 10 ps of MD. This is followed by equilibration of the water molecules at 300 K for 30 ps. Then the whole system (including the previously constrained protein) is assigned initial velocities from a Gaussian distribution generated by a random seed for 50 K, warmed up to 300 K for 10 ps, and finally equilibrated at this temperature for 1000 ps. The 1000 ps trajectory at 300 K serves as a control simulation. The last conformation of the system obtained from this trajectory is used as the starting structure for the unfolding simulation in which the system is warmed up to 500 K for 10 ps and is subsequently maintained at this temperature. To enhance the sampling of the unfolding pathway, long multiple trajectories are performed by assigning velocities generated from different seed numbers. Three 5000 ps high temperature (500 K) and three 1080 ps control (300 K) simulations are performed for each lysozyme form.

MD analysis

In the analysis, the root mean-square deviation (RMSD), solvent-accessible surface area and secondary structure content plots are calculated using the analysis programs provided by GROMACS. The secondary structure analysis is carried out using the Kabsch and Sander algorithm (1983) incorporated in their DSSP program. The interresidue distances and the clustering analysis is performed with software written in-house. The percentages of secondary structure content per region across all trajectories are derived as follows: First, the percentage in each of the four secondary structure conformation (helix, strand, turn, coil) contained in the specified regions is calculated for each structure sampled from the three unfolding trajectories. Then, four sums of percentages across all trajectories are obtained and each one is divided by the total number of the 500 K snapshots to derive the percentages shown in Table 1.

The number of pairwise distances of all residues within 8 Å (closest atoms distance) is calculated for each conformation in the trajectory and for the crystal structure. For the latter, all these residue contacts are referred to as native contacts as opposed to the nonnative contacts of pairs of residues that are situated within the 8 Å cutoff but are not present in the native structure (crystal structure).

The distance matrix used in the clustering is constructed from the column vectors of a properties' matrix. Here the properties' matrix of the conformations observed during the trajectories contains four rows corresponding to four properties: 1), the number of native residue contacts, 2), the number of

TABLE 1 Secondary structure content in lysozyme during the unfolding simulations

Regions (residues)	Helix (%) [*]		Strand (%) [*]		Turn (%) [*]		Coil (%) [*]	
	WT	D67H	WT	D67H	WT	D67H	WT	D67H
Helix A (5–14)	55	50	1	2	26	35	18	13
Helix B (25–36)	23	32	4	5	36	36	37	27
Strand 1 (43–46)	1	0	20	17	21	21	58	62
Strand 2 (51–54)	0	5	35	20	16	35	49	40
Loop (60–78)	4	5	11	8	50	56	35	31
Helix 3/10 (81–85)	23	12	2	3	44	47	31	38
Helix C (90–100)	40	34	2	3	35	35	24	28
Helix D (110–115)	30	24	6	5	43	35	21	36

^{*}The percentage is calculated from accumulative sums of secondary structure within given regions across all conformations sampled during the trajectories.

nonnative residue contacts, 3), the number of residues in secondary structure elements (α -helix, β -sheet, β -turn), and 4), the number of residues in random coils. The reason that those properties have been chosen is because it has been observed from our plots that they change with time during the unfolding unlike other properties such as solvent-accessible surface area (SASA), R_g , and number of H-bonds that oscillate around their initial value. Thus, the latter cannot discriminate easily between conformations obtained in different times during the unfolding).

The Mahalanobis distance (Mahalanobis, 1930) is used for the construction of the distance matrix. The elements of this matrix are calculated by the following formula:

$$D_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{C}^{-1} (\mathbf{x}_i - \mathbf{x}_j).$$

Here, \mathbf{x} is a column vector of the properties' matrix and \mathbf{C} is the covariance matrix (each vector has four elements so a 4×4 covariance matrix is generated). The use of Mahalanobis distance removes several of the limitations of Euclidian distances used by others for classifications of MD simulations (Kazmirski et al., 1999) as it automatically accounts for the scaling of the coordinate axes and also it corrects for correlation between the different features (e.g., residue contacts and secondary structure content). It is also valid regardless if the variables are normally distributed or not (Otto, 1998). In the distance matrix, all the conformations sampled (i.e., from all trajectories) are considered.

We use hierarchical clustering to group the conformations by transforming a set of data points with a given measurement for dissimilarity (the distance matrix) into a sequence of nested partitions (a dendrogram). We use agglomerative methods to build the dendrogram. These involve starting with each data point as a single cluster and subsequently merge these together. In particular we make use of the group average distance hierarchical algorithm: At each step, we seek the shortest distance of a pair of clusters in the distance matrix and merge them together. A new distance matrix is created in which the new distances of all the clusters from the newly created matrix are based on the average distance of their members (Everitt, 1993; Späth, 1980). This process is continued until no distance in the distance matrix occurs below a set cutoff. We have selected this cutoff to be the point in the process at which a cluster formed contains more than 20% of the conformations. This is a practical but somewhat arbitrary choice to limit the number of clusters formed and thus simplify the analysis. Our results using this clustering algorithm have been verified against other statistical packages such as S-PLUS (Venables and Ripley, 1999).

RESULTS AND DISCUSSION

Three 5-ns trajectories at high temperature are performed for the wild-type (WT 1, WT 2, WT 3) and three for the mutant (D67H 1, D67H 2, D67H 3) to increase the sampling of the unfolding conformational space (Kazmirski et al., 1999). Clustering techniques are used to probe for the most populated average conformations. Analysis of these conformations is carried out to ascertain the structural features of any seeds that have the potential to lead to fiber formation.

Analysis of the individual trajectories

At 300 K we observe that the RMSD based on the $C\alpha$ atoms plateaus around 0.2 nm from the crystal structure for all trajectories (Fig. 2) and thus agrees with previous results on the equilibration of a native protein structure at this temperature by Kazmirski et al. (1999). At 500 K, we find that

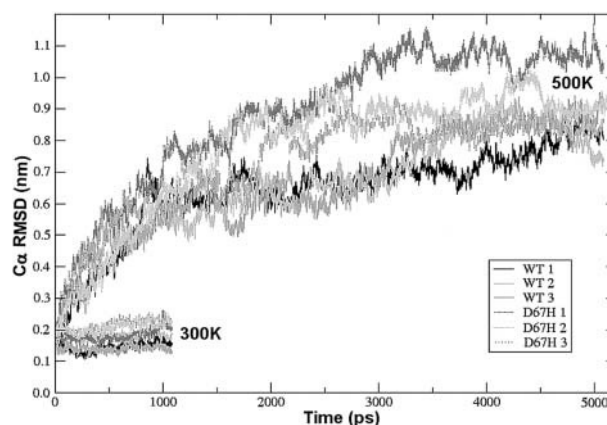


FIGURE 2 Plot of root mean-square deviation (in nm) for $C\alpha$ atoms as a function of time (ps) with respect to the crystal conformation. Results are shown for both unfolding (500 K) and control trajectories (300 K).

all simulations have similar rates of increase in RMSD for the first 1000 ps. However, after this point, the D67H mutant trajectories exhibit a larger increase in RMSD compared to that of the wild-type trajectories. The values of the RMSD near 5000 ps seem to converge for all trajectories.

The root mean square fluctuations of the atoms across all the simulations of the mutant and the wild-type form have been calculated for the $C\alpha$ atoms (Fig. 3). We observe that during unfolding, D67H tends to have higher fluctuations in most of the regions. This difference is most notable in the β -domain and, in particular, in a region just after the mutation site (residue 67) that belongs to the loop in this domain. This region of the loop demonstrates also a sharp difference in fluctuation even at 300 K. Such behavior is expected as the hydrogen bonding network connecting this loop region with

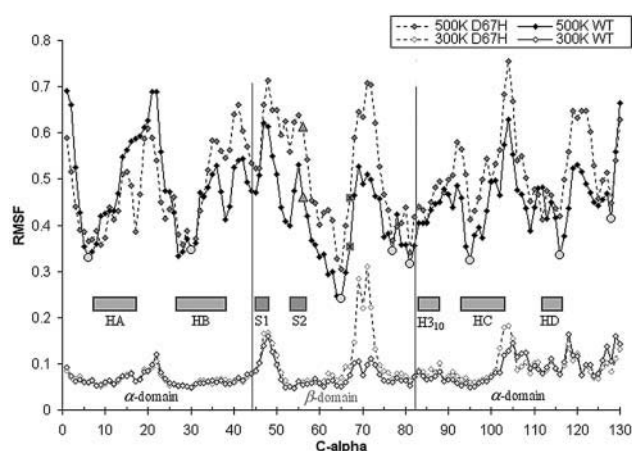


FIGURE 3 Root mean-square fluctuations of the $C\alpha$ atoms during the unfolding and 300 K simulations. The light gray triangles represent the $C\alpha$ atom at Ile⁵⁶ whereas the squares with the cross represent the mutation site $C\alpha$ atoms. The atoms of the residues involved in disulphide bonds are also shown (yellow circles). Regions of β -strand and α -helix of the protein are shown.

the S1 and S2 regions in the wild-type is disrupted due to the mutation, thus leading to a more unrestrained motion of the loop even at normal temperature. The mutation site is more flexible in the mutant during the simulations than in the wild-type. The higher flexibility of the β -domain in D67H implies that it is less stable than that of the wild-type, which is in agreement with the work of Booth et al. (1997) where the calculated B-factors for this domain are increased.

Interresidue distances change dramatically during the unfolding simulations, with residues originally distant in the native structure now being found in close proximity and vice versa. Residue contacts can be classified as native contacts (those found in the native structure) and nonnative contacts, which are formed during the simulation but not found in the native structure. For the unfolding simulations at 500 K, we have observed that the trajectories have a similar rate of loss of native contacts and formation of nonnative contacts for the first 1000 ps. Subsequently, the mutant simulations show a slight increase in loss of native contacts and also an increased formation of nonnative, with all trajectories converging by 5000 ps. For the control simulations at 300K, we observe that the number of interresidue contacts remains fairly constant. Thus, results from these contact plots are consistent with the analysis of RMSD.

The total and hydrophobic residues SASA has been calculated for both lysozyme domains and for the whole protein. Plots for all atoms and for backbone amides and carbonyls have been generated. In all cases SASA does not change greatly with time in any trajectory at 500 K, for the wild-type nor for the mutant. During the unfolding of a protein, it is expected that more residues become accessible to solvent. Although a slight increase is observed for all the trajectories between 0 and 1000 ps in the SASA of hydrophobic residues, most of the time it is around the average value of 40 nm². The total SASA increases slightly in the first 500 ps and then acquires a steady value. We noted also a small jump near 4500 ps that is common for all the trajectories. These results for SASA are consistent with the analysis of the radius of gyration against time, which similarly shows no major change with time during the simulations.

The reason for the lack of overall change of the radius of gyration and SASA may be the presence of four disulphide bonds. In the classical force field, which is used in MD simulations, the disulphide bonds are kept intact irrespective of the increased temperature conditions. In a separate set of simulations (data not shown) in which the disulphide bonds of the two forms are reduced, it is observed that at high temperature both SASA and the radius of gyration increase after 1.5 ns. At the same time, the percentage of secondary structure decreases faster than in the normal simulations. One effect of the disulphide bonding is that the residues involved in this are less flexible, as illustrated from our root mean-square fluctuation plots (Fig. 3), where the C α of these residues have the lowest fluctuations. This is consistent with the crystallographically derived B-factors for the structure.

We have also calculated the average SASA per residue across all conformations sampled to examine the pattern of solvent exposure in the different regions of the protein. We observe that most of the mutant β -domain residues have increased SASA compared to the wild-type ones (Fig. 4). Particularly the mutated residue, His⁶⁷, shows ~ 2.6 times larger SASA than Asp⁶⁷. The second half of helix C and the loop that follows up to helix D have also increased SASA in the mutant. On the contrary, the N-terminal part of the α -domain and particularly the second half of helix B have smaller solvent exposure compared to the wild-type.

For all the trajectories, we find that the secondary structure profile changes significantly during the simulation. Initially, most residues are in a specific secondary structural conformation, but the number of such residues drops gradually and reaches a plateau after ~ 4 ns. At the same time, random coil conformations become gradually predominant. Analysis of the time course of secondary structure shows that a), the original β -strands are destroyed before the helices; b), the helical elements are eventually replaced by coils, bends, and turns; and then c), transient β -strands and bridges appear in many areas across the polypeptide chain.

The changes of the secondary structure content during the unfolding in separate regions of the lysozyme are depicted in Table 1. As the regions selected are entirely composed of one of the four conformations in the native structure, we can have a rough estimation of a), how much of the original conformation is retained during the unfolding, and b), what other conformations are acquired during the simulation.

Our first observation is that the original secondary conformation is converted mostly to coils and turns during the simulations. Helix A, in particular, has $\sim 50\%$ of the time helical components for both the wild-type and mutant whereas helix 3₁₀ has the least retention of its original conformation. In very few cases, there is partial conversion from helix to strand and even less from strand to helix. We can

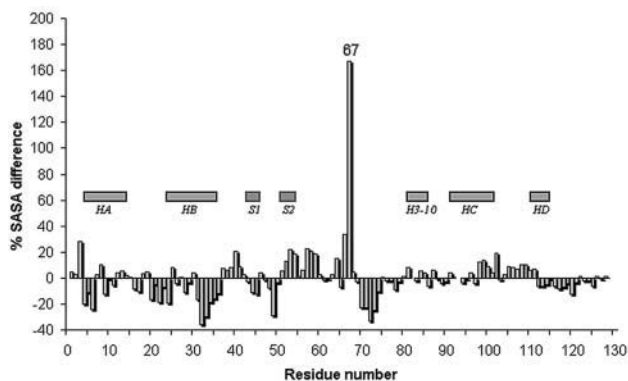


FIGURE 4 Percentage of difference ($\%D_{SASA}$) of D67H average SASA per residue from WT across all conformations sampled. The percentage is calculated from $\%D_{SASA} = ((SASA_{Mut}/SASA_{WT}) - 1) \times 100$, unless the SASA of the mutant is smaller than the WT, in which case we invert the ratio in the equation and make the percentage negative. All atoms are taken into account for the calculations of SASA. The residue 67 bar is shown.

also observe that the WT structures tend to have higher percentages of the original secondary structure conformation than the mutant structure with the exception of helix B. Conversely, the turn content is higher for the mutant in all the regions.

The hydrogen bonding across the protein, during unfolding, has also been examined, although in part it is related to the interresidue contacts analysis described previously. The total number of H-bonds is initially ~ 100 and then, after 500 ps, it drops to around an average of 75 (within the range of 60–90) for the rest of the trajectory. This applies to all of the trajectories at 500 K. The explanation for this observation may be related to the changes in the secondary structure; the native structure is rich in helices and strands, both of which involve many hydrogen bonds. During the unfolding, we have observed that these elements are converted to coils (no hydrogen bonds) and turns (contain smaller hydrogen bonding network).

Clustering of the trajectories' conformations and analysis

More informative analysis of multiple trajectories can be obtained by accumulating all snapshots from each trajectory and clustering into those with similar properties. This is particularly useful since we cannot discriminate which trajectory is more “significant” than the others (Worth et al., 1998). All the conformations generated from the three trajectories of each lysozyme form are clustered (using a hierarchical clustering procedure) to generate average structures ranked by the population of the cluster they represent. This clustering is based on residue contacts and secondary structure content (as described in methods). The resulting conformations are shown in Fig. 5.

Analysis of the wild-type trajectories leads to 21 clusters (Fig. 6 A). The most populated, WC1, contains 20.4% of the total population, followed by two clusters with the order of 14% and two more of 10%. Five more clusters ranging between 3 and 7% are found and the rest are below 2%. For the D67H mutant, where 21 clusters are assigned using the same cutoff criteria as the WT (Fig. 6 B), the distribution is slightly different. The largest group contains 20.6% and followed by three clusters, each of $\sim 16\%$ of the total population. One cluster of $\sim 12\%$ and one of 7% exist, and three more groups ranging from 2 to 4% are assigned. The rest have populations ranging between 0.46 and 0.07%. A small proportion of the conformations of both lysozyme forms are grouped in clusters (named WC8 and MC8) that both contain elements from the initial stages of the unfolding trajectory.

The average structural features of these clusters are given in Table 2. The most populated conformations for both the WT and the mutant have lost most of their secondary structure content and the original native contacts. It appears though, that the WT clusters are usually populated by conformations retaining more native contacts (i.e., a higher fraction

of the initial conformation) than its mutant counterpart. Additionally, the populated mutant clusters contain conformations usually with greater nonnative contacts. In many of the clusters there are conformations that contain a high proportion of residues in random coils. However, the WT clusters appear to contain a higher percentage of secondary structure elements (helices and sheets) than the mutant.

The major impact of unfolding can be seen in plots of interresidue contact distances. These contact maps for the average structures of the two most populated clusters and the crystal structures are plotted in a 2D color map of all-against-all residue distances (Fig. 7). In the most populated cluster of the mutant (MC1), the β -domain (especially in the region of the β -strands) exhibits a very different set of contacts compared with those in crystal structure and with those in other clusters from the mutant trajectories.

Specifically, in the MC1 cluster, the region around Ile⁵⁶, located in the interface between the α - and β -domain, loses contact with helix B and helix 3₁₀. Those two contacts are present in the crystal structures and 300 K simulations of both lysozyme forms and they exist as well in the average structure of the most populated clusters of the wild-type. The distribution of these distances in the conformational space sampled with the simulations is shown in Fig. 8. We observe that at 500 K, for the Ile⁵⁶-helix B contact, the mutant exhibits a characteristic two-peak distribution. One of the peaks is near the value of the crystal structure and the WT distance distribution. The second peak, however, lies at a greater distance apart. In a similar fashion but with less intensity, we observe a twin-peak distribution of the Ile⁵⁶-helix 3₁₀ distance for the MC1. A comparison of the means and standard deviations of the plots are shown in Table 3. The mean and the standard deviations of the distributions of the unfolding simulations have been compared using standard statistical methods (*t*-test and F-test) and found that their difference is significant (both tests fall within the 95% confidence interval).

Comparison with experimental data and other simulations

Comparisons with experimental data can be made by consideration of the RMSD and residue contacts against time. Thus, we observe that at 500 K, the RMSD of the mutant increases at a higher rate than the RMSD of the wild-type after 1000 ps, an indication that the mutant is more susceptible to the artificial high temperature applied. This implies that the mutant's conformation may be more sensitive to conditions that affect structural integrity (e.g., temperature, pH) and thus it is potentially easier to unfold and in this sense may be less stable than the wild-type. Using this logic, these results are in agreement with data from circular dichroism (Booth et al., 1997) and stopped-flow fluorescence experiments (Canet et al., 1999), which show that both natural mutant forms of lysozyme are less thermostable than the wild-type.

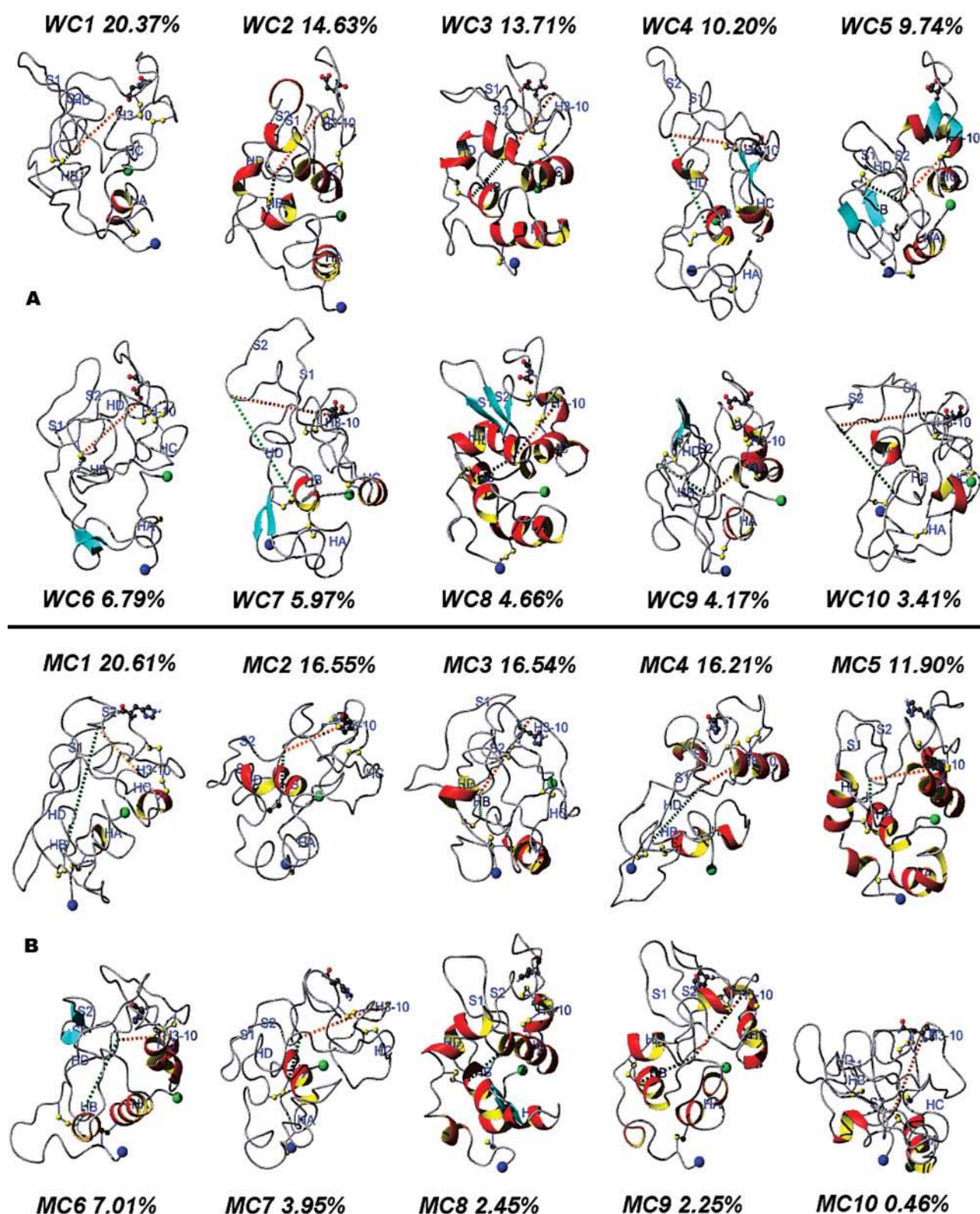


FIGURE 5 Average structures of the 10 most populated clusters. The mutation site (residue 67) is shown as ball and sticks. The N- and C-terminal are shown as green and blue spheres, respectively. The disulphide links are also shown as spheres. The clusters for the WT (A) are denoted as WC1, WC2, etc., whereas the ones from the mutant (B) are denoted as MC1, MC2, etc. The models have been obtained after least-square fitting each of these conformations to the respective initial models (crystal structure + hydrogens). Thus, the orientation and shape of the structures are relative to the ones shown in Fig. 1. The dotted lines show the distance between Ile⁵⁶ and helix 3₁₀ and the distance between Ile⁵⁶ and helix B.

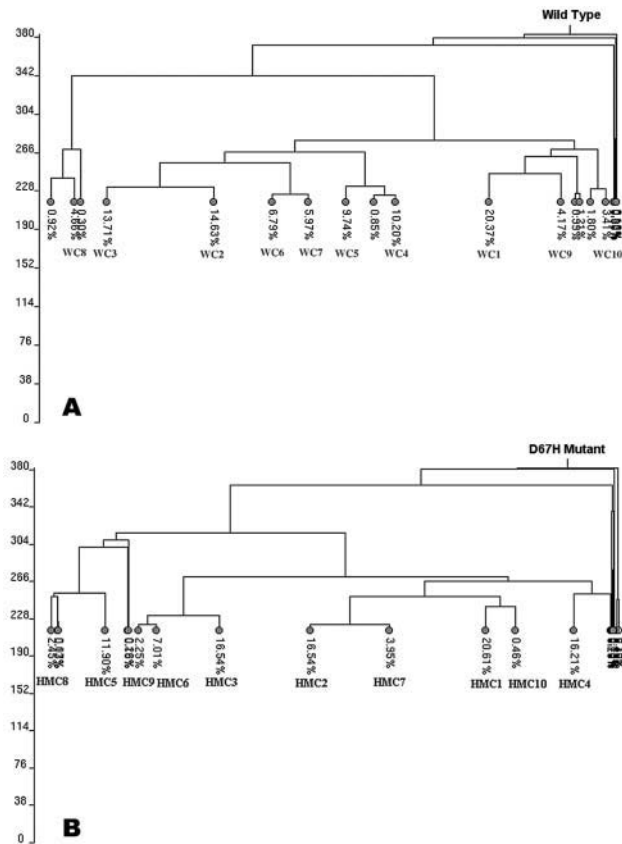


FIGURE 6 Dendrograms of (A) the wild-type and (B) the mutant, produced using hierarchical clustering. For clarity, only the branches up to the cutoff point are shown. The green circles represent the clusters derived. The 10 biggest clusters are labeled (e.g., WC1).

Canet et al. (2002) have very recently published their results from 2D NMR and mass spectroscopy in conjunction with hydrogen exchange methods on the structure of the wild-type and D67H mutant. Comparison of the protection factors per residue show that the β -domain and part of the helix C appear to be less protected from H exchange in the mutant, which is consistent with our findings that these two regions have higher SASA than the wild-type. Our simulations also show that there is a larger population of intermediates than the wild-type in the D67H mutant that have simultaneously the β -domain and helix C unfolded.

In high temperature (498 K) unfolding MD simulations carried out in hen egg white lysozyme (Kazmirski and Daggett, 1998), the partially folded intermediates identified show increased nonnative structure. This is consistent with our findings for the most populated clusters of conformations sampled during the high temperature simulations whereby nonnative contacts are increased across the polypeptide chain. Another common feature is that the radius of gyration of the intermediates increases slightly ($\sim 10\%$) in comparison to the 300 K simulations. These intermediate states show a slight change in nonpolar SASA in comparison to the crystal structure, and again, this is consistent with our findings.

TABLE 2 Average structural properties of the (A) wild-type and (B) D67H mutant clusters given as the numbers of residues in a particular secondary structure conformation

Cluster	Population (%)	Native contacts (% retained)	Nonnative contacts	No. residues in helix or sheet or turn*	No. residues in coil*
A. Wild-type					
WC1	20.37	140 (38)	240	31	52
WC2	14.63	182 (49)	175	55	42
WC3	13.71	191 (52)	173	51	48
WC4	10.20	117 (32)	267	39	56
WC5	9.74	164 (44)	237	50	45
WC6	6.79	116 (31)	238	32	61
WC7	5.97	107 (29)	247	33	54
WC8	4.66	270 (73)	94	78	28
WC9	4.17	148 (40)	265	34	54
WC10	3.41	108 (19)	283	30	49
Crystal [†]	—	370 (100)	0	89	19
B. D67H mutant					
MC1	20.61	109 (31)	297	38	53
MC2	16.54	132 (37)	269	33	50
MC3	16.54	126 (35)	250	33	57
MC4	16.21	108 (30)	256	38	49
MC5	11.90	221 (62)	142	69	33
MC6	7.01	159 (45)	183	48	50
MC7	3.95	133 (37)	256	29	49
MC8	2.45	273 (76)	94	71	31
MC9	2.25	208 (58)	132	52	43
MC10	0.46	121 (34)	326	34	56
Crystal [†]	—	357 (100)	0	93	18

*The rest of the 130 residues are in β -bends and β -bridges.
[†]For comparison, the structural features of the respective crystal structures are shown.

During induced unfolding at 500 K, both wild-type and mutant lysozyme demonstrate increasing loss of their original secondary structure. The results from the analysis of secondary structure content against time show that most of the original secondary structure elements are lost within the first 2.5 ns. An interesting feature is that the β -domain is destabilized first and then the α -domain follows. This is in agreement with previous unfolding simulations of hen egg white lysozyme at 500 K (Kazmirski and Daggett, 1998). Also data from refolding experiments (using stopped-flow amide hydrogen exchange and mass spectroscopy by Miranker et al., 1991, 1993) show that $\sim 80\%$ of the refolding molecules have their amide hydrogen atoms in the α -domain protected before those in the β -domain. Pepys et al. (1993) also showed that the β -domain is more unstable than the α -domain.

The destabilized secondary structure in both wild-type and mutant lysozyme forms gives rise to random transient β -turns across the whole polypeptide chain. β -turns do not generally constitute stable structural elements (Tobias et al., 1990), which explains their transient nature in our simulations. However, based on both experimental evidence (Zimmerman and Scheraga; 1977, Dyson et al., 1992), and

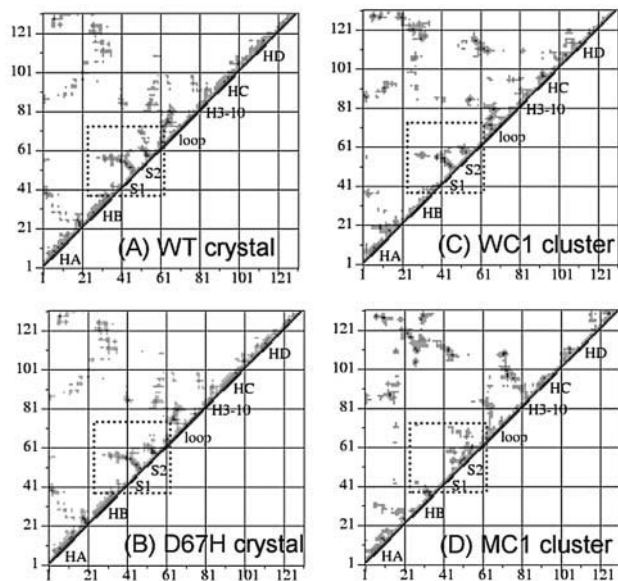


FIGURE 7 Contact maps of (A) the WT crystal structure, (B) the mutant crystal structure, (C) the WC1 cluster, and (D) the MC1 cluster. Both axes represent residue number. The labels on the diagonal correspond to regions of the molecule. The cutoff for a pair of residues to be in contact is 8 Å between their Cα atoms. The rectangular area shows the contacts inside part of the β-domain region. The shade key is shown (distances in Å).

also lattice simulations (Skolnick and Kolinski, 1991), it is argued that β-turns may play a role in the initial stages of the formation of helices and sheets during folding and thus they appear to direct folding pathways while tending to adopt conformations that minimize the “local” conformational free energy of the residues in the turn (Yang et al., 1996). In our results (Table 1) we observe that the sampled conformations of the mutant contain slightly higher percentage of turns than the WT in most of the regions. Using Fourier transform infrared spectroscopy data, Booth et al. (1997) propose that the partly folded forms of the lysozyme variants associate through the unstable β-domain to form the initial seed for the generation of amyloids. From Table 1, we observe that for this β-domain (strand 1, strand 2, and loop), the partially unfolded intermediates of the mutant have higher β-turn content, which may be an indication of the relative instability in this region in the mutant lysozyme compared to the wild-type.

Booth and collaborators (1997) suggest that the motion in the β-domain due to the mutation may not be the direct cause for amyloidogenicity. Rather, changes in the interface region of the two domains, transmitted from the disturbed β-sheet, may be the actual cause. In particular Ile⁵⁶, found to have increased B-factors for both amyloidogenic mutants of human lysozyme, could play an important role in the mutant’s amyloidogenic properties. In our unfolding simulations, the fluctuation for this residue is higher in the variant and thus in agreement with these experimental results.

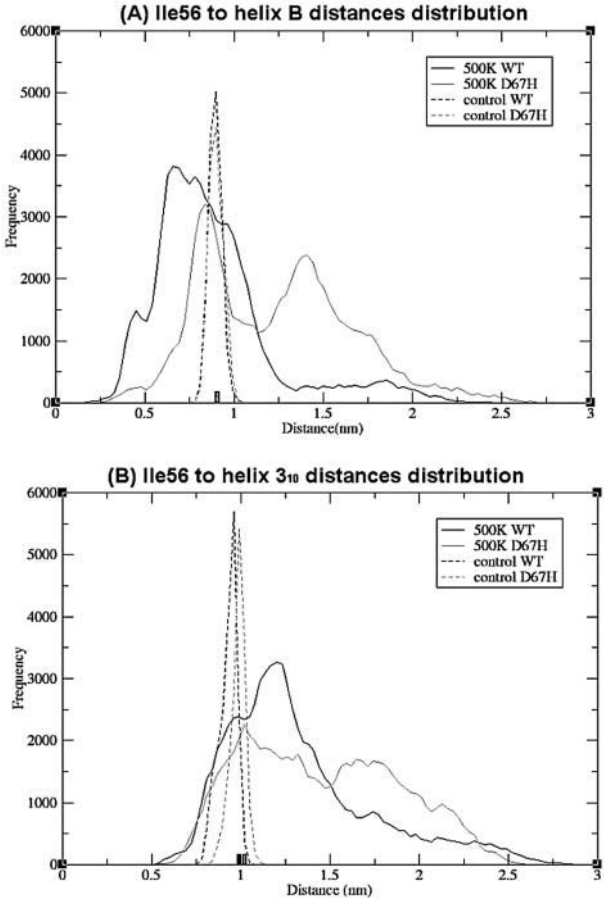


FIGURE 8 Distribution of the distances between (A) residue 56 (Ile) to helix B and (B) between residue 56 (Ile) and helix 3₁₀ in the sampled conformations of the unfolding and control trajectories. The distances at the native structure are shown as black (wild-type) and red (mutant) rectangles just above the x axis.

CONCLUSIONS

Our RMSD and residue contacts results indicate also that the unfolding of the mutant appears to occur faster than the wild-type, and we observe that it takes longer for the wild-type to lose its native structure under our artificial higher temperature conditions. In addition to that, in the conformational space sampled, it seems that the wild-type retains higher content of its original secondary structure relative to the

TABLE 3 Mean and standard deviation of the distance distribution (in nm) of Ile ⁵⁶ -helix B and Ile ⁵⁶ -helix 3 ₁₀ contacts				
Simulation	Ile ⁵⁶ -helix B (nm)		Ile ⁵⁶ -helix 3 ₁₀ (nm)	
	WT	D67H	WT	D67H
Crystal	0.90	0.91	0.99	1.01
Control	0.91 ± 0.04	0.91 ± 0.04	0.94 ± 0.05	1.00 ± 0.04
Unfolding	0.89 ± 0.35	1.26 ± 0.43	1.34 ± 0.42	1.47 ± 0.44

mutant. Both wild-type and mutant simulations include conversions of the original secondary structure conformations to coils and β -turns. These β -turns, however, seem to be more predominant in the β -domain of the mutant, a region that has been suggested to be involved in fibrillogenesis (Booth et al., 1997; Pepys et al., 1993). Turns are thought to occur in abundance before formation of helices and sheets during folding (Yang et al., 1996), and thus the overall higher content in turns in the sampled unfolding conformational area of the mutant may be another indication that it unfolds faster than the wild-type, based on our simulation results.

Also in the β -domain we observe higher atomic fluctuations in D67H than the wild-type, implying that this region is more flexible (linked to stability), and in that sense is in agreement with the work of Booth et al. (1997) linking the reduced stability of this domain with fibrillogenesis. This point is also enforced by the finding that the residues of the D67H β -domain experience higher solvent exposure during our simulations. In other amyloidogenic proteins, (e.g., transthyretin), the formation of amyloids has been correlated to increased residue accessibility to the solvent (Schormann et al., 1998).

The clustering technique has led to some interesting results. The most populated clusters of the wild-type retain near half of their native contacts compared with one-third for D67H mutant. The lack of retention of native contacts is a useful measure of the degree of unfolding. So, it appears that most of the conformations of the mutant are farther away from their native structure than those of the wild-type, and thus, the mutant protein is more susceptible to changes under unfolding conditions.

In our unfolding simulations, we attempt to identify features that distinguish the partially unfolded conformations of the WT and D67H human lysozyme. Although there is a vast diversity in the conformational features sampled for both the forms, the alternative distances of the Ile⁵⁶ and two regions of the α -domain seems to differentiate a number of conformations of the mutant from the wild-type. Although the significance of this finding cannot directly be linked at this time to amyloidogenesis, it does demonstrate clearly that one of the effects of the mutation at residue 67 is the resulting distortion of the important region at the interface of the two domains.

We thank the Biotechnology and Biological Sciences Research Council for computer hardware and the Engineering and Physical Sciences Research Council for a studentship to G.M.

This work has been carried out within the Biotechnology and Biological Sciences Research Council-sponsored Bloomsbury Centre for Structural Biology.

REFERENCES

- Allen, M. P., and D. J. Tildesley. 1987. *Computer Simulations of Liquids*. Clarendon, Oxford.
- Alonso, D. O. V., and V. Daggett. 2000. Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains. *Biophysics*. 97:133–138.
- Artymiuk, P. J., and C. C. F. Blake. 1981. Refinement of human lysozyme at 1.5 Å resolution. Analysis of non-bonded and hydrogen-bond interactions. *J. Mol. Biol.* 152:737–762.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for waters in relation to protein hydration. In *Intermolecular Forces*. B. Pullman, editor. D. Reidel, Dordrecht, The Netherlands. 331–342.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
- Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 91:43–56.
- Booth, D. R., M. Sund, V. Bellotti, C. V. Robinson, W. L. Hutchinson, P. E. Fraser, P. N. Hawkins, C. M. Dobson, S. E. Radford, C. C. F. Blake, and M. B. Pepys. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*. 385:787–793.
- Brooks, C. L., 3rd. 1998. Simulations of protein folding and unfolding. *Curr. Op. Struct. Biol.* 8:222–226.
- Canet, D., A. M. Last, P. Tito, M. Sunde, A. Spencer, D. B. Archer, C. Redfield, C. V. Robinson, and C. M. Dobson. 2002. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Biol.* 9:308–315.
- Canet, D., M. Sunde, A. M. Last, A. Miranker, A. Spencer, C. V. Robinson, and C. M. Dobson. 1999. Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. *Biochemistry*. 38:6419–6427.
- Chipman, D. M., and N. Sharon. 1969. Mechanism of lysozyme action. *Science*. 165:454–465.
- Daggett, V. 2000. Long timescale simulations. *Curr. Op. Struct. Biol.* 10: 160–164.
- Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: A N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
- Dyson, H. J., J. R. Sayre, G. Merutka, H. C. Shin, R. A. Lerner, and P. E. Wright. 1992. Folding of peptide fragments comprising the complete sequence of proteins models for initiation of protein folding. II. Plas-tocyanin. *J. Mol. Biol.* 226:819–835.
- Everitt, B. S. 1993. *Cluster Analysis*. John Wiley & Sons, New York.
- Gilquin, B., C. Guilbert, and D. Perahia. 2000. Unfolding of hen egg lysozyme by molecular dynamics simulations at 300K: insight into the role of the interdomain interface. *Proteins*. 41:58–74.
- Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.* 18:1463–1472.
- Hünenberger, P. H., A. E. Mark, and W. F. van Gunsteren. 1995. Computational approaches to study protein unfolding: hen egg white lysozyme as a case study. *Nature: Struct. Funct. Genet.* 21:196–213.
- Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.
- Karplus, M., and A. Sali. 1995. Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* 5:58–73.
- Kazmirski, S. L., and V. Daggett. 1998. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. *J. Mol. Biol.* 284:793–806.
- Kazmirski, S. L., A. Li, and V. Daggett. 1999. Analysis methods for comparison of multiple molecular dynamics trajectories: applications to protein unfolding pathways and denatured ensembles. *J. Mol. Biol.* 290: 283–304.

- Kelly, J. W. 1998. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* 8:101–106.
- Li, A., and V. Daggett. 1994. Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA.* 91:10430–10434.
- Li, A., and V. Daggett. 1998. Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol.* 275:677–694.
- Mahalanobis, P. C. 1930. On tests and measures of groups divergence I. *Journal of the Asiatic Society of Bengal.* 26:541.
- Mark, A. E., and W. F. van Gunsteren. 1992. Simulation of the thermal denaturation of HEW-lysozyme: trapping the molten globule state. *Biochemistry.* 31:7745–7748.
- Miranker, A., S. E. Radford, M. Karplus, and C. M. Dobson. 1991. Demonstration by NMR of folding domains in lysozyme. *Nature.* 349:633–636.
- Miranker, A., C. V. Robinson, S. E. Radford, R. T. Alpin, and C. M. Dobson. 1993. Detection of transient protein folding populations by mass spectroscopy. *Science.* 262:896–900.
- Morozova-Roche, L. A., J. Zurdo, A. Spencer, W. Noppe, V. Receveur, D. B. Archer, M. Joniau, and C. M. Dobson. 2000. Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants. *J. Struct. Biol.* 130:339–351.
- Otto, M. 1998. *Chemometrics: Statistics and Computer Application in Analytical Chemistry.* Weinheim, Cambridge.
- Pepys, M. B., P. N. Hawkins, D. R. Booth, D. M. Viguishin, G. A. Tennent, A. K. Soutar, N. Totty, O. Nguyen, C. C. F. Blake, C. J. Terry, T. G. Feest, A. M. Zalin, and J. J. Hsuan. 1993. Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature.* 206:553–557.
- Pepys, M. B. 1996. Amyloidosis. In *The Oxford Textbook of Medicine*, Vol. 2. D. J. Weatherall, J. G. G. Ledingham, and D. A. Warrell, editors. Oxford University Press, Oxford.
- Schormann, N., J. R. Murrell, and M. D. Benson. 1998. Tertiary structures of amyloidogenic and non-amyloidogenic transthyretin variants: new model for amyloid fibril formation. *Amyloid.* 5:175–187.
- Skolnick, J., and A. Kolinski. 1991. Dynamic Monte Carlo simulations of a new lattice model of globular protein folding, structure and dynamics. *J. Mol. Biol.* 221:499–531.
- Späth, H. 1980. *Cluster Analysis Algorithms for Data Reduction and Classification of Objects.* Ellis Horwood, Chichester, UK.
- Tobias, D. J., S. F. Sneddon, and C. L. Brooks. 1990. Reverse turns in blocked dipeptides are intrinsically unstable in water. *J. Mol. Biol.* 216:783–796.
- van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. O. Scott, and I. G. Tironi. 1996. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Hochschulverlag AG an der ETH, Zurich, Switzerland.
- Venables, W. N., and B. D. Ripley. 1999. *Modern Applied Statistics with S-PLUS*, 3rd ed. Springer, New York.
- Williams, M. A., J. M. Thornton, and J. M. Goodfellow. 1997. Modelling protein unfolding: hen egg-white lysozyme. *Protein Eng.* 10:895–903.
- Worth, G. A., F. Nardi, and R. C. Wade. 1998. Use of multiple molecular dynamics trajectories to study biomolecules in solution: the YTGP peptide. *J. Phys. Chem. B.* 102:6260–6272.
- Yang, A., B. Hitz, and B. Honig. 1996. Free energy determinants of secondary structure formation: III. β -turns and their role in protein folding. *J. Mol. Biol.* 259:873–882.
- Zimmerman, S. S., and H. A. Scheraga. 1977. Local interactions in bends of proteins. *Proc. Natl. Acad. Sci. USA.* 74:4126–4129.