Detection of non-native hydrophobic interactions in the denatured state of lysozyme by molecular dynamics simulations

2005 J. Phys.: Condens. Matter 17 S1617

(http://iopscience.iop.org/0953-8984/17/18/017)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) S1617–S1626 doi:10.1088/0953-8984/17/18/017

Detection of non-native hydrophobic interactions in the denatured state of lysozyme by molecular dynamics simulations

Emanuele Paci[1] **and Michele Vendruscolo**[2]

[1] Institute of Molecular Biophysics and School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK
[2] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Received 18 November 2004, in final form 21 January 2005
Published 22 April 2005
Online at stacks.iop.org/JPhysCM/17/S1617

Abstract

The presence of native and non-native hydrophobic clusters of amino acids has been detected experimentally for the protein hen egg white lysozyme even under strongly denaturing conditions. We characterize the structure of these hydrophobic clusters by two series of molecular dynamics simulations. The first series shows that in the wild type protein a non-native cluster formed by four tryptophan residues (W62, W63, W108 and W111) is formed with significant probability under denaturing conditions. In contrast, as observed experimentally and indicated by the second series of simulations presented here, the same cluster is formed much more rarely in the W62G mutant, which reduces the hydrophobicity of the polypeptide chain in the interface region between the two structural domains in the native state of the protein. These results support the observation that non-native interactions may play an important role in the folding process of complex proteins by stabilizing intermediate states in which hydrophobic groups are sequestered from the solvent and thus prevented from initiating aggregation.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The characterization of the residual structure in non-native states of proteins is a problem that has recently been much studied, as the residual structure present in these states may affect the stability [1, 2], as well as the folding [3–6] and misfolding [7–9] behaviour. Partially folded states of proteins have also been considered because of their association with cellular processes, such as translocation across membranes, signal transduction and protein synthesis [10, 11].

It is well established that globular proteins have a certain propensity to adopt expanded conformations and be highly dynamic even under physiological conditions [12, 13]. For

proteins that are normally found in a folded state such non-native conformations are, however, difficult to study experimentally since they are not significantly populated [14, 15]. Thus, experimental investigations are often carried out using denaturants [16–19], reduction of disulfide bridges [17], pH changes [16, 20], sequence truncations [21] or point mutations [22] to destabilize the native structure relative to the denatured state. In a recent study, Klein-Seetharaman *et al* [23] used a combination of NMR spectroscopy and site-directed mutagenesis to observe the presence of an extended network of hydrophobic interactions in urea-denatured states of the protein hen egg white lysozyme. Six tryptophan residues (W28, W62, W63, W108, W111 and W123) were found to be important in stabilizing this network, in particular W62, which forms both native and non-native as well as short-range and long-range interactions. It was also shown that non-native interactions made by this residue slow down the folding process. This result was demonstrated by the observation that the W62Y mutant exhibits a 13-fold increase in the folding rate with respect to the wild type when the disulfide bonds are intact [24]. Similarly, the W108Y mutant increases the folding rate seven-fold [24]. Moreover, chemical modifications of W62 increase the propensity of the protein to misfold [7]. Klein-Seetharaman *et al* [23] detected residual structure in the highly denatured state of lysozyme by measuring the chemical shifts of the H^{N} and the H_α protons and the ^{15}N transverse relaxation rates, R_2. Changes of chemical shifts upon denaturation are indicative of local structural rearrangements, including those involving secondary elements [23, 25]. In addition, changes in the relaxation rates report on the effects of mutations on the dynamics of the protein [23].

In the native structure, W62 is exposed to the solvent and is highly dynamic [26]. In contrast, this residue, together with the other tryptophan residues, is buried in the denatured state [27] and during the early stage of folding [28]. A very important result recently reported by Schwalbe and collaborators [23] is that in the denatured state of lysozyme residues W62 and W63, which are in the β-domain, are in proximity of the native-like hydrophobic cluster in the α-domain involving residues W108 and W111 (see figure 1). In the same study, it was also shown that in the mutant W62G this extended core is destabilized.

In this work we use molecular dynamics simulations to provide detailed structural models for the conformations populated by hen egg white (HEW) lysozyme and by its W62G mutant under denaturing conditions, including the non-native hydrophobic clusters detected experimentally. In agreement with the findings of Klein-Seetharaman *et al* we found that the non-native cluster formed by residues W62, W63, W108 and W111 is formed with significant probability. This cluster is instead observed much more rarely in our simulations of the W62G mutant.

2. Methods

The crystal structure of HEW lysozyme at 1.33 Å resolution (PDB entry 193L [29]) was used as the starting point of the simulations (figure 1). We performed two sets of simulations, the first for the WT and the second for the W62G mutant. In both cases the eight cysteine residues were methylated to reproduce the experimental set up of Klein-Seetharaman *et al* [23]. The exploration of the non-native region of the conformational space was carried out by a biased molecular dynamics scheme similar to that used by Paci *et al* [30]. All simulations were performed at room temperature in implicit solvent (EEF1) [31]. The energy was initially minimized in order to remove steric clashes, in particular those associated to the methylation of the cysteine residues. The temperature of the system was then raised to 300 K in 0.6 ns and a further equilibration at this temperature was carried out for 4 ns. The average RMSD from the native structure taken over the last 1.0 ns of simulation was of 2.7 (3.1) Å for C_α (all atoms), indicating a moderate instability of the native state when the disulfide bonds are

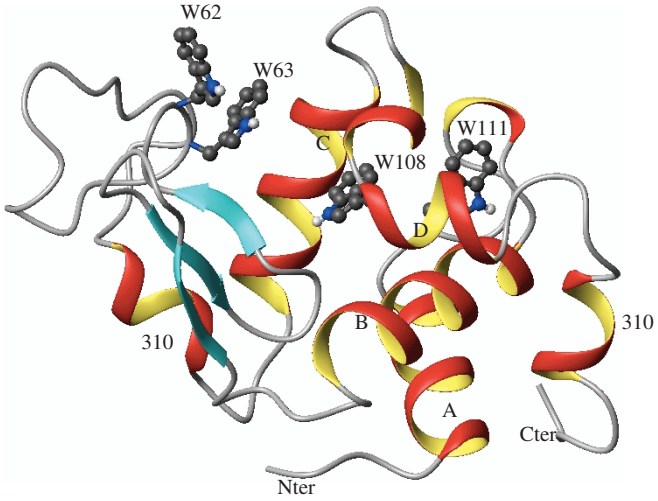

Figure 1. X-ray structure (193L.PDB) of HEW lysozyme [29]. The four tryptophan residues (W62, W63, W108 and W111) that form a non-native cluster in the denatured state [23] are indicated, as well as the secondary structure.

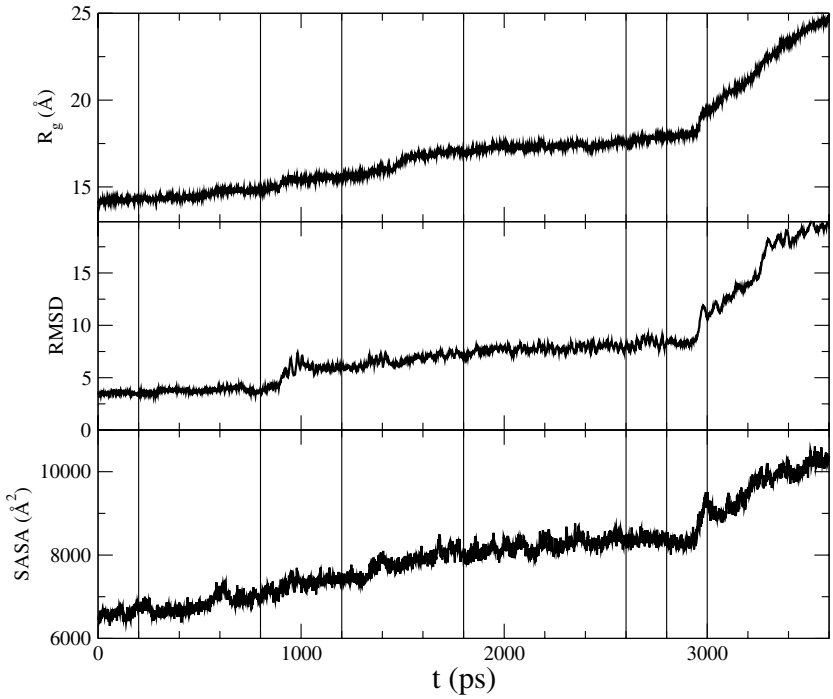

Figure 2. Time series for a 3.6 ns simulation of the W62G mutant with a bias that forces R_g to increase. The conformations that were selected from this trajectory are indicated by vertical lines and they are characterized by an RMSD between 3.5 and 11 Å and an R_g value between 14.3 and 19.5 Å.

reduced. We did not, however, try to characterize this effect on the native state in more detail here as we were interested in sampling the unfolded state. A further simulation of 3.6 ns

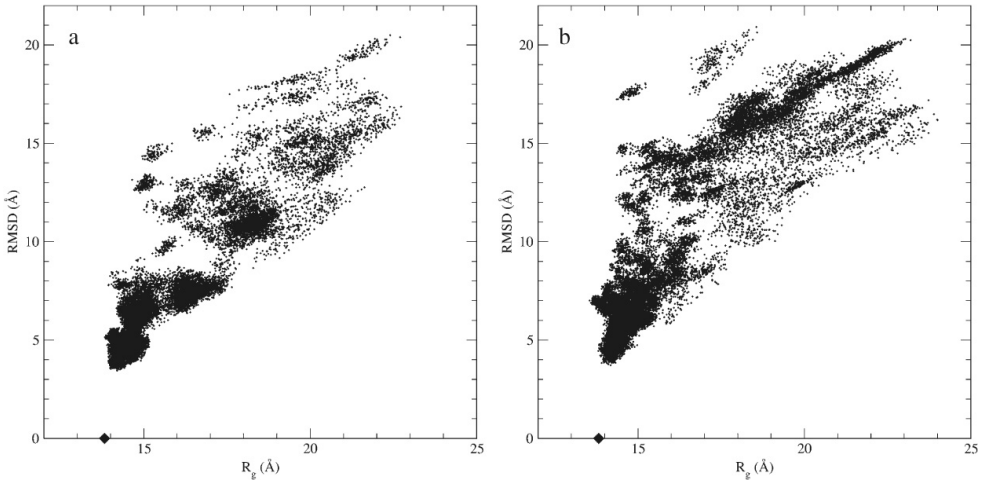

Figure 3. Scatter plots of the RMSD from the native structure and the R_g for 22 400 conformations, which illustrate the broadness of the conformational space explored by the simulations presented in this work. The circle corresponds to the native state. (a) WT; (b) W62G.

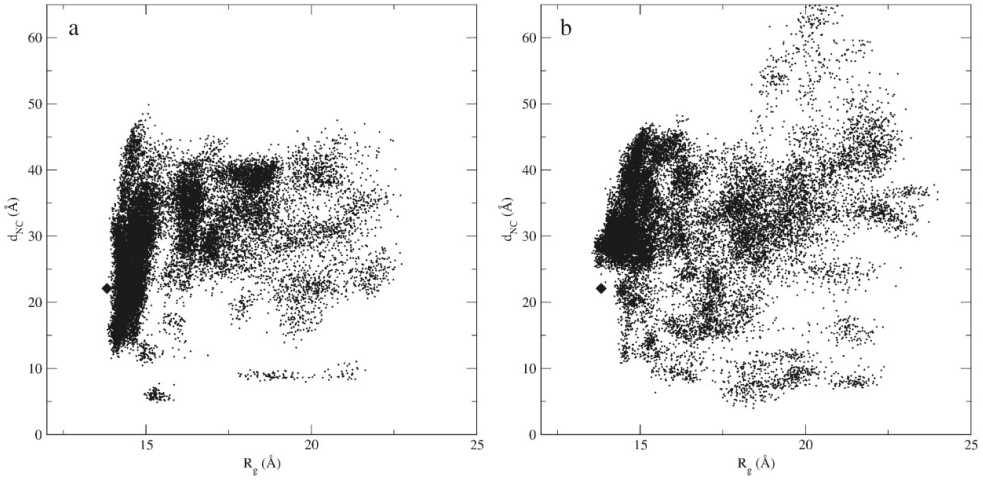

Figure 4. Scatter plots of the end-to-end and the R_g for 22 400 conformations. The circle indicates the native state. (a) WT; (b) W62G.

was then performed using a biased molecular dynamics scheme that included a perturbation increasing the RMSD [30]. The radius of gyration (R_g), the solvent accessible surface and the RMSD from the native conformation are shown as a function of time in figure 2 in the case of the W62G mutant; similar results were obtained for the WT (data not shown). Seven conformations were selected along the trajectory, and shown by the vertical lines in figure 2. We used these conformations as starting points for a series of 1.6 ns simulations with a perturbation that forced the number of contacts between heavy atoms of the side-chains to decrease. Further, a configuration every 0.4 ns was used as the starting point to perform eight sequential simulations of 0.4 ns with a perturbation increasing the RMSD from the initial structure. This procedure enabled us to generate, in 104.4 ns, 32×7 conformations

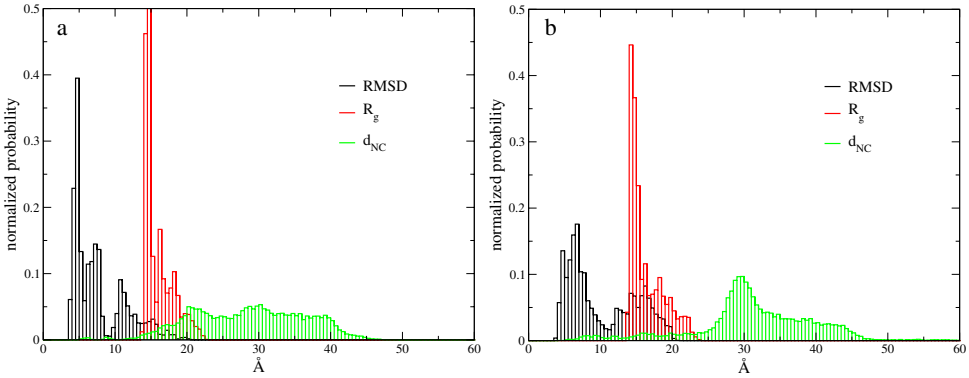

Figure 5. Probability distribution of the RMSD from the native structure, the R_g and the end-to-end distance. (a) WT; (b) W62G.

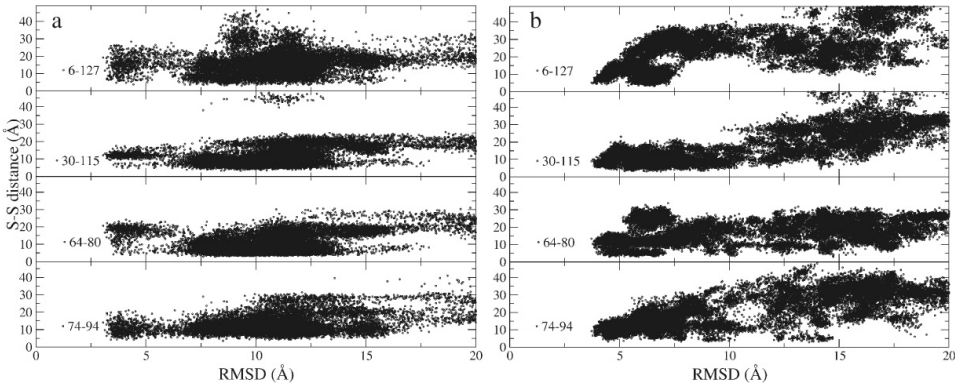

Figure 6. S–S distances as a function of the RMSD from the native structure. (a) WT; (b) W62G.

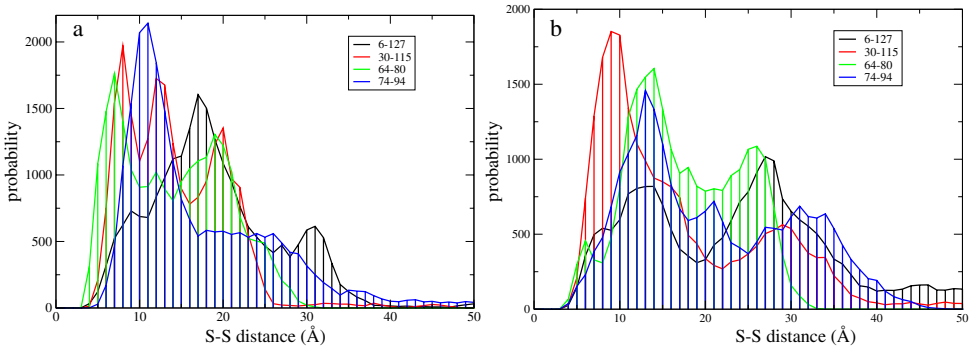

Figure 7. Probability distribution of the S–S distances. (a) WT; (b) W62G.

(coordinates and velocities) that represent a variety of possible conformations of lysozyme. These conformations ranged between 4.6 and 22.3 Å in RMSD from the native structure and between 14.3 and 23.9 in R_g. From each of these 224 conformations a 1.2 ns unbiased simulation was started and the results (after disregarding the first 0.2 ns) were used for the analysis. We thus performed 224 ns of unbiased sampling, spanning the conformational space

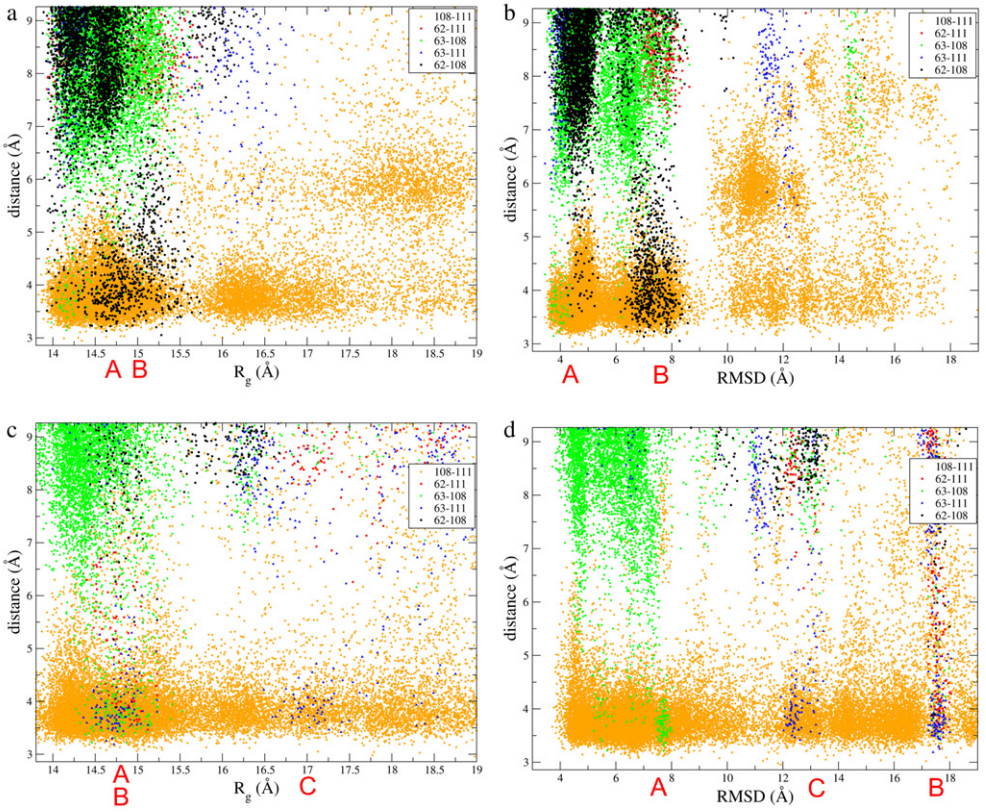

Figure 8. Distance between pairs of residues for 22 400 conformations plotted as a function of R_g ((a) and (c)) and of the RMSD ((b) and (d)) from the native structure, respectively. (a) and (b): WT. We observe the collapse of residues W62 and W63 with residue W108 and the formation of a hydrophobic non-native core; the corresponding structures are compact, with an R_g value ranging between 14.5 and 15.5 and characterized by an RMSD from the native structure between 6 and 8 Å. We suggest that these structures might correspond to the state characterized by Schwalbe and co-workers [23]. (c) and (d): W62G. During the simulations we observed three different types of conformations: (A) in some compact structures, residue W63 and W108 are in contact; (B) in a compact but non-native-like structure both residues G62 and W63 make a close contact with residues W108 and W111; (C) in a less compact structure (R_g = 17.4 Å), residues W63 and W111 are found at a distance between 4.4 and 7.9 Å. Residues W108 and W111 are found to be almost always in contact in our simulations, even in very non-native structures. Structures shown in figures 9 and 10 correspond to regions indicated with letters A and B.

starting from a relative heterogeneous set of 224 initial configurations. The total length of the simulations performed in the present study was of 373.2 ns.

3. Results

3.1. Sampling of conformational space

The sampling procedure that we used was designed to generate a set of highly heterogeneous conformations, representative of the state of lysozyme under denaturing conditions. Our approach was based on a scheme proposed by Paci *et al* [30]. In the present implementation, three different perturbations were applied in turn. The first perturbation forces an increase in

the radius of gyration, the second induces a progressive breaking of native contacts and the third increases the distance from the initial conformation (see section 2). We do not expect, however, the procedure that we described to provide a sampling of the conformational space with the correct statistical weights. Methods to perform accurate calculations of free energy landscapes with detailed all-atom models are computationally considerably demanding and have been applied so far only to proteins smaller than lysozyme [32, 33]. More recently an alternative methodology was proposed based on the use of NMR experimental data as restraints in computer simulations and applied to the determination of the free energy landscape of α-lactalbumin, a protein structurally similar to lysozyme [13, 34]. Our aim in this study is to use the biased molecular dynamics scheme described above to explore the conformational space of lysozyme in order to provide structural models to complement the experimental observations made by Schwalbe and co-workers [23].

The ensemble of conformations that we generated was heterogeneous and not correlated to the initial conformations. The RMSD as a function of R_g is shown in figure 3. The end-to-end distance varies broadly for any given value of R_g, suggesting that conformations with the same R_g value can be topologically very different (figure 4). The distributions of the RMSD from the native structure, R_g and end-to-end distance are shown in figure 5. For both the WT and the W62G mutant the distributions (for the RMSD in particular) are bimodal, suggesting a sharp transition, possibly related to the specific unfolding mechanism of lysozyme, which is associated with the relative movements of the α and β domains. In the WT, disulfide-bond (S–S) distances are well preserved even for very expanded states (figure 6). In the W62G mutant, only the cysteine pairs 64–80 and 74–94 are found often at a distance compatible with the presence of a disulfide bond. The probability distribution of S–S distances is shown in figure 7. In both the WT and the W62G mutant, the S–S distance 30–115 is short in the majority of the conformations, despite the large distance of the two residues in the sequence.

3.2. Characterization of the non-native hydrophobic clusters

In the simulations of the WT we observed that residues W62 and W63 have the tendency to form a non-native hydrophobic cluster with residues W108 and W111. In contrast, in the W62G mutant, the cluster formed by residues G62, W63, W108 and W111 is observed only very rarely. Interestingly, however, in this case such a cluster can be formed also in very compact non-native structures.

In order to illustrate these results, we analysed the distance between pairs of residues for 22 400 conformations (i.e. one every 10 ps along the unbiased simulations) plotted as a function of R_g and of the RMSD from the native structure (figure 8). In the WT (figure 8, top) we observe a persistent cluster involving residues W62, W63 and W108 at R_g between 14.5 and 15.5 Å and RMSD from the native state between 6 and 9 Å. This cluster, which is formed in very non-native but compact structures, might correspond to that observed by Schwalbe and co-workers [23]. A representative structure for this cluster is shown in figure 9(B). An incomplete formation of this cluster occurs for less denatured states with an RMSD from the native state of less than 5 Å, and it is shown in figure 9(A). In the W62G mutant (figure 8, bottom), we observed the formation of a hydrophobic cluster that involves only residues W63 and W108 and is similar to case (B) for the WT; this cluster is marked (A) in figure 8 and is represented by the structure (A) in figure 10. For the W62G mutant we also observe the formation of a hydrophobic cluster involving residues W63, W108 and W111 in compact non-native structures (with an RMSD of about 18 Å from the native state). This case is indicated with (B) in figure 8 and is represented by the structure (B) in figure 10.

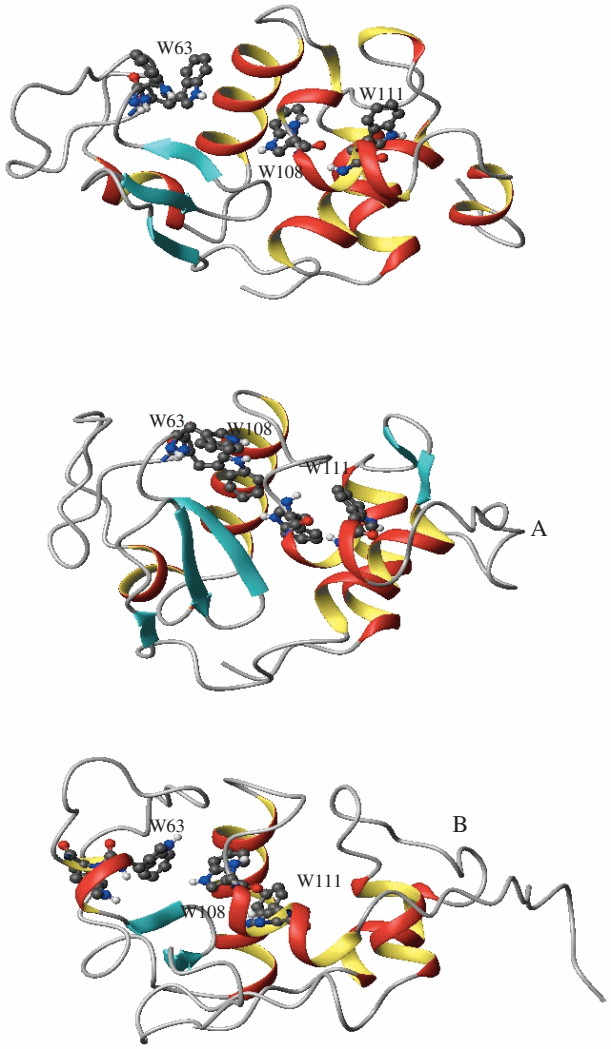

Figure 9. WT: (top) Native state of HEW lysozyme. (A) The structure found in the simulations presented in this work for which W63 is closest to W108 and W111, and (B) the structure for which W62 is closest to W108 and W111.

4. Conclusions

We used molecular dynamics simulations to characterize the non-native region of the conformational space of hen egg white lysozyme. We found a certain tendency for the formation of the non-native hydrophobic cluster formed by four tryptophan residues (W62, W63, W108 and W111) that was previously detected experimentally [23]. The probability of formation of this cluster decreases in a mutant (W62G) that reduces the hydrophobicity of one of the residues at the interface between the α and the β domain.

These results support the suggestion that non-native hydrophobic interactions may play an important role in the folding process of complex proteins by sequestering temporarily hydrophobic groups in specific intermediate states [23]. If left exposed to the solvent for too long during the folding process, such residues could potentially trigger aggregation events.

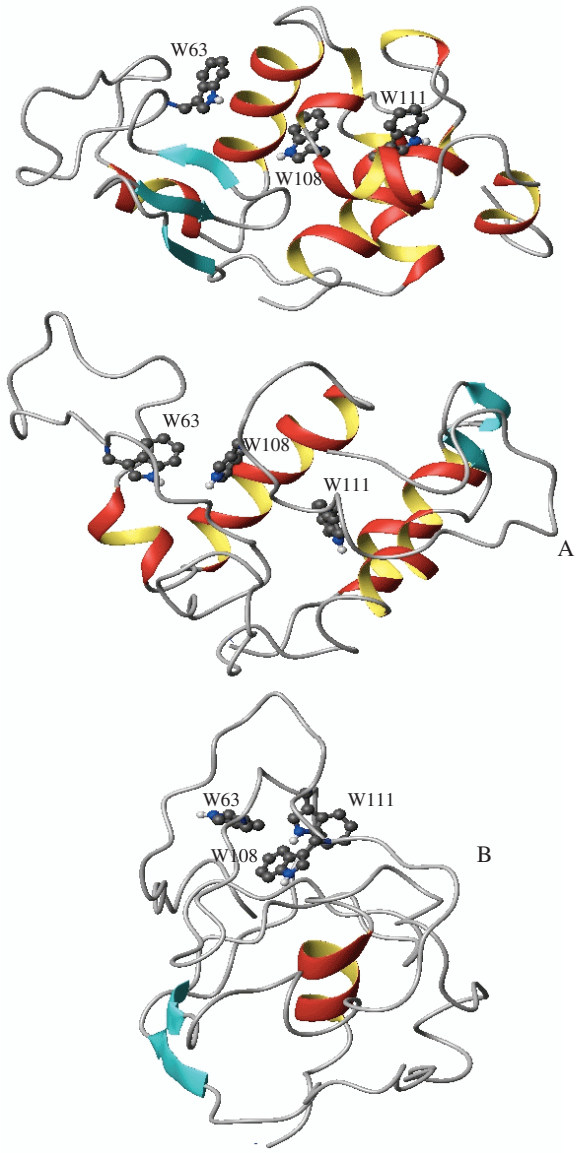

Figure 10. W62G: (top) Native structure of HEW lysozyme. (A) A structure found in the present simulations for which residues W63 and W108 are close, and (B) a structure for which residues W63 and W111 are close.

These results raise the intriguing possibility that sequences for which transient aggregation is promoted in folding intermediates have been selected through evolution in order to minimize the risk of permanent aggregation.

Acknowledgments

We thank Harald Schwalbe and Julia Wirmer for very useful discussions. All the simulations were performed while EP was at the Department of Biochemistry of the University of Zurich. The research of MV is supported by the Royal Society and by the Leverhulme Trust.

References

[1] Shortle D 1996 *FASEB J.* **10** 27–34
[2] Ding K, Louis J M and Gronenborn A M 2004 *J. Mol. Biol.* **335** 1299–307
[3] Daggett V and Fersht A R 2003 *Trends Biochem. Sci.* **28** 18–25
[4] Crowhurst K A and Forman-Kay J D 2003 *Biochemistry* **42** 8687–95
[5] Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson C M, Poulsen F M and Vendruscolo M 2004 *J. Am. Chem. Soc.* **126** 3291–9
[6] Capaldi A P, Kleanthous C and Radford S E 2002 *Nat. Struct. Biol.* **9** 209–16
[7] Ueda T, Yamada H, Aoki H and Imoto T 1990 *J. Biotechnol.* **108** 886–92
[8] McParland V J, Kalverda A P, Homans S W and Radford S E 2002 *Nat. Struct. Biol.* **9** 326–31
[9] Dedmon M M, Lindorff-Larsen K, Christodoulou J, Vendruscolo M and Dobson C M 2005 *J. Am. Chem. Soc.* **433** 476–7
[10] Uversky V N and Fink A L 2004 *Biochim. Biophys. Acta* **1698** 131–53
[11] Dyson H J and Wright P E 2002 *Curr. Opin. Struct. Biol.* **12** 54–60
[12] Clarke J and Itzhaki L S 1998 *Curr. Opin. Struct. Biol.* **8** 112–8
[13] Vendruscolo M, Paci E, Karplus M and Dobson C M 2003 *Proc. Natl Acad. Sci. USA* **100** 14817–21
[14] Choy W Y, Mulder F A, Crowhurst K A, Muhandiram D R, Millett I S, Doniach S, Forman-Kay J D and Kay L E 2002 *J. Mol. Biol.* **316** 101–12
[15] Korzhnev D M, Salvatella X, Vendruscolo M, Nardo A A D, Davidson A R, Dobson C M and Kay L E 2004 *Nature* **430** 586–90
[16] Arcus V L, Vuilleumier S, Freund S M, Bycroft M and Fersht A R 1995 *J. Mol. Biol.* **254** 305–21
[17] Schwalbe H, Fiebig K M, Buck M, Jones J A, Grimshaw S B, Spencer A, Glaser S J, Smith L J and Dobson C M 1997 *Biochemistry* **36** 8977–91
[18] Teilum K, Maki K, Kragelund B B, Poulsen F M and Roder H 2002 *Proc. Natl Acad. Sci. USA* **99** 9807–12
[19] Lindorff-Larsen K, Vendruscolo M, Paci E and Dobson C M 2004 *Nat. Struct. Mol. Biol.* **11** 443–9
[20] Yao J, Chung J, Eliezer D, Wright P E and Dyson H J 2001 *Biochemistry* **40** 3561–71
[21] Alexandrescu A T, Abeygunawardana C and Shortle D 1994 *Biochemistry* **33** 1063–72
[22] Mayor U, Guydosh N R, Johnson C M, Grossmann J G, Sato S, Jas G S, Freund S M, Alonso D O, Daggett V and Fersht A R 2003 *Nature* **421** 863–7
[23] Klein-Seetharaman J, Oikawa M, Grimshaw S B, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith L J, Dobson C M and Schwalbe H 2002 *Science* **295** 1719–22
[24] Rothwarf D M and Scheraga H A 1996 *Biochemistry* **35** 13797–807
[25] Cornilescu G, Delaglio F and Bax A 1999 *J. Biomol. NMR* **13** 289–302
[26] Buck M, Boyd J, Redfield C, MacKenzie D A, Jeenes D J, Archer D B and Dobson C M 1995 *Biochemistry* **34** 4041–55
[27] Lyon C E, Jones J A, Redfield C, Dobson C M and Hore P J 1999 *J. Am. Chem. Soc.* **121** 6505–6
[28] Hore P J, Winder S L, Roberts C H and Dobson C M 1997 *J. Am. Chem. Soc.* **119** 5049–50
[29] Young A C, Tilton R F and Dewan J C 1994 *J. Mol. Biol.* **235** 302–17
[30] Paci E, Smith L J, Dobson C M and Karplus M 2001 *J. Mol. Biol.* **306** 329–47
[31] Lazaridis T and Karplus M 1999 *Proteins* **35** 133–52
[32] Boczko E M and Brooks C L III 1995 *Science* **269** 393–6
[33] Guo Z, Brooks C L III and Boczko E M 1997 *Proc. Natl Acad. Sci. USA* **94** 10161–6
[34] Vendruscolo M and Paci E 2003 *Curr. Opin. Struct. Biol.* **13** 82–7