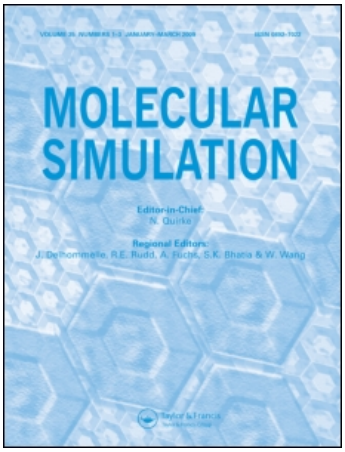

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Comparison of Generalised Born/Surface Area with Periodic Boundary Simulations to Study Protein Unfolding

A. G. Purkiss[a]; J. T. Macdonald[a]; J. M. Goodfellow[a]; C. Slingsby[a]
[a] School of Crystallography, Birkbeck College, London, UK

To cite this Article Purkiss, A. G. , Macdonald, J. T. , Goodfellow, J. M. and Slingsby, C.(2004) 'Comparison of Generalised Born/Surface Area with Periodic Boundary Simulations to Study Protein Unfolding', Molecular Simulation, 30: 5, 333 — 340
To link to this Article: DOI: 10.1080/08927020410001667566
URL: http://dx.doi.org/10.1080/08927020410001667566

Molecular Simulation, Vol. 30 (5), 30 April 2004, pp. 333–340

Taylor & Francis
Taylor & Francis Group

Comparison of Generalised Born/Surface Area with Periodic Boundary Simulations to Study Protein Unfolding

A.G. PURKISS*, J.T. MACDONALD, J.M. GOODFELLOW and C. SLINGSBY

School of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

(Received January 2004; In final form January 2004)

Simulations comparing the rapid unfolding behaviour of the model protein barnase under explicit and implicit solvent systems have been undertaken in order to validate a faster implicit solvent method for studying proteins which are kinetically stable *in silico*. A comparison is made between all-atom explicitly solvated simulations of barnase undertaken using Particle Mesh Ewald electrostatic interactions with all-atom implicit solvent simulations undertaken using the generalised born/surface area (GBSA) method with a long non-bonded cut-off. The two explicit solvent unfolding trajectories appear to explore slightly different pathways showing the importance of having statistically valid ensembles which are not accessible from a single trajectory. The 500 K GBSA trajectory is unsuitable for exploring intermediate structures on the unfolding pathway of barnase, as the protein almost immediately jumps to a predominately random coil conformation. However, dropping the temperature to 400 K gives rise to trajectories where the protein is unable to climb out of the energy well containing the first intermediate state, in a reasonable timescale. A similar pattern to the explicit solvent unfolding trajectories is seen in 450 K GBSA runs, with the intermediate states differing between trajectories. The development of computer simulation methods suitable for application to more kinetically stable proteins will offer insight into the atomic detail of the conformational changes associated with protein unfolding diseases.

Keywords: Barnase; Generalised born/surface area; Protein molecular dynamics simulation; Protein unfolding

INTRODUCTION

Methods are required for revealing the molecular details that underlie the protein unfolding diseases. A reduction in stability increases the chance that a protein will adopt a partially unfolded structure that may lead to loss of function, aggregation and harmful depositions [1,2]. Strategies are emerging for preventing these diseases by considering the energy landscape during protein unfolding and/or aggregation [3–5]. Experimental molecular details of the processes involved in protein unfolding are, however, only available for a few fast folding model systems such as chymotrypsin inhibitor 2 (CI2) [6,7]. This body of experimental data has been used to benchmark [8] and improve [9] theoretical molecular simulations of protein unfolding. Simulation protocols are needed that expand the range of target proteins to include those that are either kinetically or thermodynamically more stable than the model systems as well as those that are known to play a role in disease.

The stability of a protein has two components: thermodynamic (as measured by the free energy change (ΔG) from the native folded conformation to a denatured, random coil state) and kinetic (as measured by k_u, the rate at which the native protein changes to the unfolded state). For a simple two state pathway, the unfolding rate is governed by the difference in energy between the folded state to the transition state maximum on the free energy landscape. Proteins having high thermodynamic, but low kinetic stability will rapidly interconvert between folded and unfolded forms at equilibrium, whereas for proteins with a high kinetic stability interconversion will be slow. Folding kinetics are influenced by protein size and by the complexity of the fold topology [10,11]. Proteins that are kinetically stable are difficult to study, both *in vitro* and *in silico*, due to the presence of kinetic traps on the unfolding pathway or off-pathway intermediates [12].

*Corresponding author. Tel.: +44-20-7631-6869. Fax: +44-20-7631-6803. E-mail: a.purkiss@mail.cryst.bbk.ac.uk

ISSN 0892-7022 print/ISSN 1029-0435 online © 2004 Taylor & Francis Ltd
DOI: 10.1080/08927020410001667566

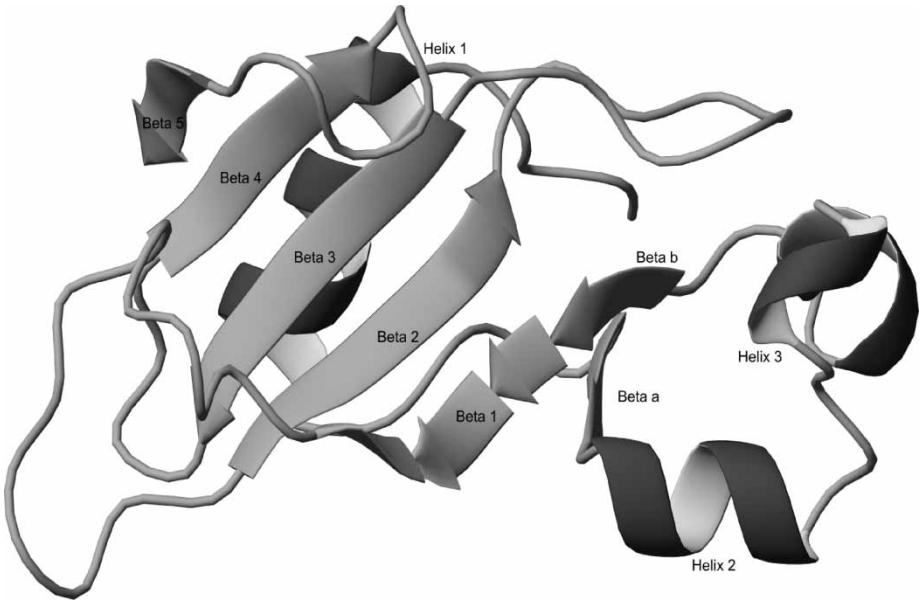

FIGURE 1 Secondary structure of barnase from the X-ray crystal coordinates.

The influence of a protein's folding and stability upon disease causation can be due to a range of physical effects [9]. Although it has been shown that point mutations of peptides and proteins can alter their conformation and/or aggregation properties [13,14], in diseases such as age-related cataract, the rate at which the protein unfolds will play a major role [15]. It is possible that stabilised intermediates on unfolding pathways can seed the formation of insoluble, light-scattering aggregates [16], as has been seen with other systems [17].

To circumvent the slow unfolding of kinetically stable proteins *in silico*, we have undertaken some simulations comparing the behaviour of the model protein barnase under both implicit and explicit solvent systems. Implicit solvation allows the use of large non-bonded cut-offs, avoiding artefacts arising from use of short cut-offs or of the EWALD summation. Implicit solvent methods also allow multiple trajectories to be generated more quickly, allowing a greater area of phase space to be explored and so giving better statistical validity [18]. The different energy landscapes explored in folding a small peptide have been investigated with both implicit and explicit solvent methodologies [19]. However, in order to undertake unfolding simulations on stable globular proteins, verification of the use of implicit solvent conditions is required as many of the changes in structure take place beneath the surface, unlike in the peptide simulations. Some work on the folding and unfolding of globular proteins using implicit solvation has been undertaken by, for example, Ferrara *et al.* [20] and Paci *et al.* [21], using biasing terms to force the protein to change structure.

Barnase is a fast (un)folding protein that has been widely studied in the field of protein stability, both *in vitro* and *in silico* [8,22–25]. The protein is a 110 residue ribonuclease from *Bacillus amyloliquefaciens*, with a theoretical pI of 8.88, a k_u of $1.2 \times 10^5\,s^{-1}$ (by extrapolation to zero denaturant from GdmHCl denaturation), and a $\Delta G_{D\text{-}N}$ of 8.2 kcal/mol at 33°C at pH 6.3 [26]. Structurally, it is a single domain mixed α/β-protein with three α-helices at the N-terminus flanking a central β-sheet (Fig. 1). There is a large hydrophobic core between α-helix 1 and the β-sheet, with other smaller hydrophobic cores also present.

Barnase has been shown by NMR, stopped-flow kinetics and ϕ-value analysis [25,27–29] to have at least one intermediate on the (un)folding pathway, with a rate determining step between the intermediate and native states:

$$D \leftrightarrows I_2 \leftrightarrows I_1 \leftrightarrows (t.s.) \leftrightarrows N$$

although whether the folding process is two state or three state is controversial [23,24]. NMR spectroscopy of barnase unfolded using pH, high temperature and urea [25] suggests that there is residual native structure present in the regions of α-helix 1, α-helix 2 and β-strands 3–4, in agreement with high temperature (498 K) simulations of the protein [22,30] by Li and Daggett [22] showing the presence of a second intermediate. A complex of barnase and barstar has been investigated using implicit solvent methods [31], but this study described the protein-protein interactions and not unfolding.

Here we compare all-atom solvated simulations of barnase undertaken at 500 K using Particle Mesh Ewald (PME) for long range electrostatics with all-atom implicit solvent simulations undertaken at

400, 450 and 500 K using the generalised born/surface area (GBSA) method. The similarities and differences between the methods and the previously published simulations on barnase are discussed.

METHODS

The X-ray crystal structure coordinates of barnase (pdb-code: pdb1rnb) were used in preference to the NMR coordinates previously studied [22] using computer simulation methods. The crystal structure has an additional two stranded β-sheet formed from sections of the N-terminal part of the protein, one β-strand coming between helix 1 and helix 2 (labelled a) and the other twisted in relation to the start of the first β-strand of the NMR structure and labelled b (Fig. 1). Missing side chain atoms and hydrogens were added using the TLEAP program; the missing N-terminal alanine was not modelled. The default protonation state for the side chains was used. For the solvated simulations, a box of TIP3P waters was built with a minimum distance of 9.0 Å from protein to the box edge, two chloride ions were added to neutralise the system.

All simulations were undertaken using the AMBER 7 [32,33] version of SANDER. The simulations were run on a Beowulf cluster using either 4 or 8 Athlon processors running at 1.3 GHz in dual-processor nodes. For the GBSA simulations [34,35], energy minimization of the protonated structure was undertaken in two stages: the positions of the hydrogens were first minimised to an rmsd of 0.01 Å, then all atoms were minimized for 5000 steps. The 'standard' AMBER 7 version of the GBSA methodology was used, with the following parameters for both energy minimization and molecular dynamics: an 18 Å non-bonded cut-off (short test trajectories showing this to be the best compromise between the computational time taken and the accuracy of the electrostatic energy calculated), an internal dielectric of 1.0 and an external dielectric of 78.5. After minimization, the protein was gradually warmed from 50 K (velocities assigned from a Maxwellian distribution) to 300 K over 20 ps, with a 1fs time step. The 300 K control trajectory continued from this warmed structure. Random structures during the 300 K trajectory were warmed to 400 or 450 K over 50 ps and used to study the unfolding pathway of the protein over a time scale of 10 ns. A trajectory at 500 K was also undertaken.

For the explicit solvent simulations, the energy minimization was undertaken as for the GBSA runs. To allow the water molecules to adjust to the protein, a four-stage start up protocol was undertaken, first warming the solvent molecules (with the protein coordinates fixed) from 50 to 300 K over 5 ps and then keeping the solvent at 300 K for 5 ps. The whole system was then returned to 50 K, new velocities assigned and warmed to 300 K over 10 ps, followed by 20 ps at 300 K at the start of the control trajectory. The explicitly solvated simulations used a cut-off of 8 Å with PME used for modelling long-range electrostatics. The warming stages were undertaken under constant volume conditions, switching to constant pressure (of 1 atm) when the temperature was no longer being raised. Structures at random points during the 300 K-control trajectory were warmed to 500 K over 25 ps, to generate the unfolding trajectories.

Analysis of the resultant coordinates was undertaken using cluster analysis [36]. Four types of data were derived from structures taken at 10 ps intervals: the radius of gyration (from the AMBER 7 program CARNAL [33]); the percentage of residues with defined secondary structure (α-helix, 3/10-helix, β-strand or β-bridge) calculated with STRIDE [37]; the total solvent accessible surface area (from NACCESS [38]); the number of retained native contacts (calculated with the program CONTAX [39], a native contact being defined as a C_α to C_α distance <8.0 Å between residues not adjacent in sequence). These data were clustered using Mahalanobis distances and hierarchical group averaging. Graphs were produced using XMGRACE, molecular pictures were produced using MOLMOL [40] and rendered using POVRAY version 3.5.

RESULTS

Explicit Solvent Simulations

For the explicit solvent system, two 500 K trajectories of over 5 ns were generated, taking about 5.3 days per ns on 8 processors. At the start of the trajectory, the first 500 K dataset shows a rapid increase in backbone root mean square deviation (rmsd) before settling down to a value of about 5 Å for 2 ns; this is followed by another jump to about 7 Å which is maintained for another nanosecond before a final increase to over 10 Å where the simulation stays until the end of the trajectory (Fig. 2a, blue datasets). The structures at 5 and 7 Å rmsd appear to be the first and second intermediates on the unfolding pathway, seen previously [22], with the structures at 10 Å rmsd being representative of the final unfolded random coil ensemble. (The term random coil is used here to describe a fluctuating collection of unfolded molecules, with little residual secondary structure and where all residues are exposed to solvent at some point). The second 500 K trajectory shows a slower increase to around 5 Å followed by a more gradual increase to around 7 Å before the sudden jump at the end of the dataset. Both simulations provide good evidence for the first intermediate with an rmsd of 5 Å, while the second intermediate is less well defined.

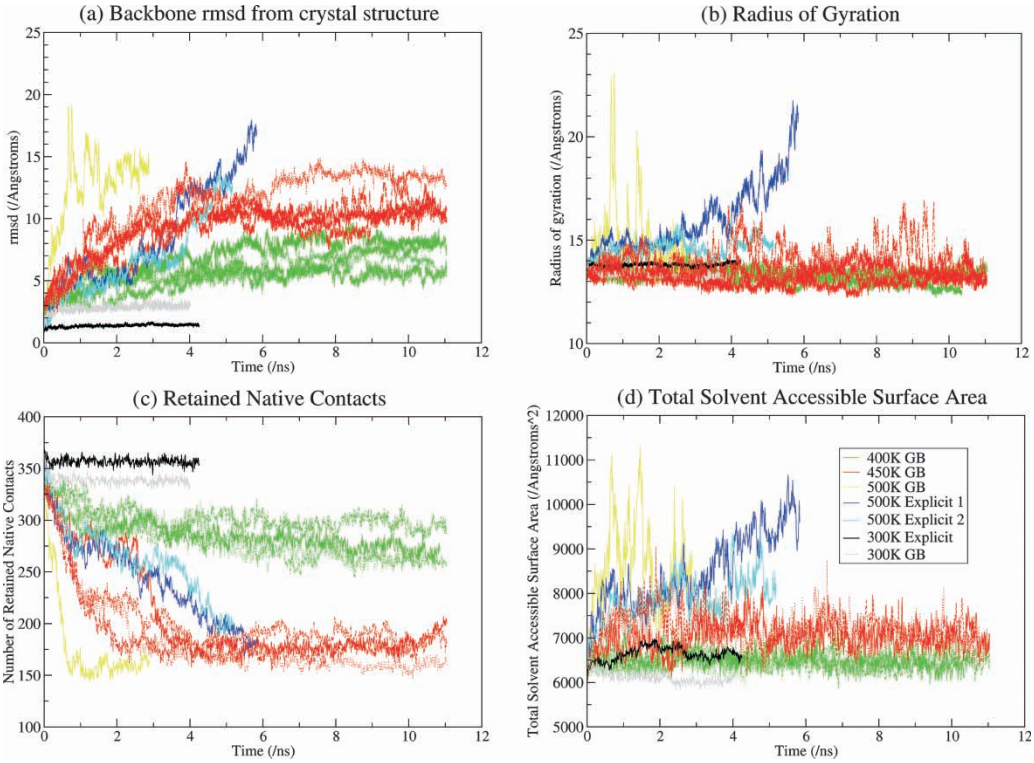

FIGURE 2 The four different assessments of structural change that were used to analyse the unfolding and control trajectories from data taken at 10 ps intervals are indicated in (a)–(d).

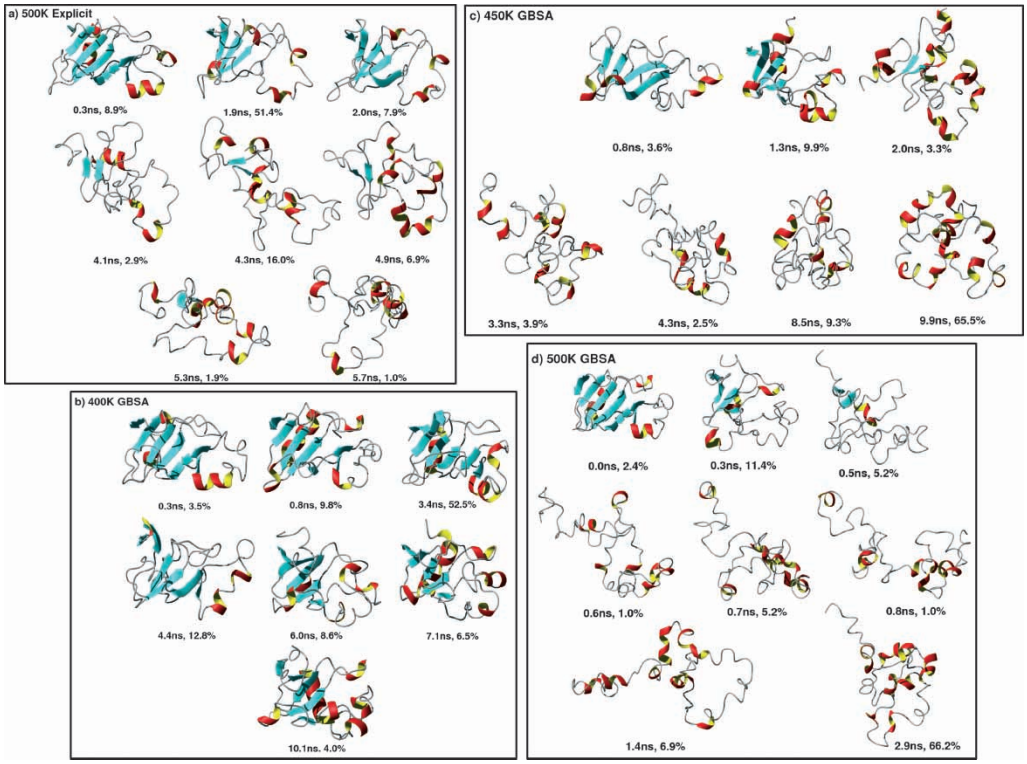

FIGURE 3 Average structures from cluster analysis to show unfolding pathways under different conditions: (a) 500 K explicit solvent simulation. (b) 400 K GBSA simulation. (c) 450 K GBSA simulation. (d) 500 K GBSA simulation. The time from the beginning of the simulation and the size of the relevant cluster is indicated beneath each structure.

Analysis of the native contact data for these 500 K explicit solvent trajectories shows that the first intermediate structure has lost approximately 20% of contacts seen in the crystal structure and that another 20% are lost during the transition to the final random coil ensemble. Of those contacts remaining at the end of each trajectory, most are "local" contacts within helical conformations. In contrast, the explicit solvent control simulation at 300 K shows little change in native contacts, as would be expected.

Cluster analysis was undertaken using the four types of data described in the methodology section, with a distance cut-off of 225. The clusters representative of the early and middle part of the unfolding pathways have structures present from both trajectories, but the clusters representative of the later parts of the unfolding pathway (i.e. the part of the ensemble in random coil conformations) only contain structures from a single trajectory. These clusters highlight the differences between the two trajectories showing substantial changes early in the course of the unfolding. The second trajectory structures are closest to the average structures shown in Fig. 3a, with the first trajectory providing the most distant data points.

The near native structures at the start of both trajectories are similar (Fig. 3a, cluster 2). However, the first intermediate structures have subtle differences, with trajectory 1 showing the N-terminal α-helices 2 and 3 opening out from the surface of the β-sheet (seen in Fig. 3a) with the two short β-strands, a and b, being lost. Trajectory 2 shows a smaller movement of the helices and retention of the short β-strands for longer. These structural differences are visible in the radius of gyration data (Fig. 2b), with the first trajectory continuing to expand after the first intermediate is reached, whilst in the case of the second trajectory, the radius stabilises after the first intermediate is reached. In both trajectories, the changes in the helical sub-domain of the protein are followed in time by gradual loss of the β-sheet (seen in Fig. 3a), before the protein reaches a disordered state. This final change from residual β-strand structure to random coil is not well defined by cluster analysis, as the trajectories have started to diverge. In summary, there appear to be subtle differences in the "first intermediate" structures (assigned by comparison to the intermediate seen by Li and Daggett [22]) between the two trajectories and the second intermediates are not well defined.

Generalised Born Simulations

For the 400 K GBSA simulations, the four different trajectories all show an immediate jump in rmsd to around 3 Å followed by a gradual increase over about 5 ns to an rmsd of 5–8 Å (Fig. 2a, green datasets). At this point, the rmsd of the trajectories reaches a plateau, with a variation in amplitude of up to 4 Å. The four different 450 K GBSA trajectories show a rapid jump in rmsd to 5–7 Å at the start of the simulation. This value is maintained for 1–3 ns before gradually rising again to over 10 Å (Fig. 2a, red datasets); one trajectory can be seen to rise to 13 Å. The timing and magnitude of this final change is different between the trajectories.

The structures of the snapshots in the plateau regions (rmsd of about 6 Å) of both 400 and 450 K GBSA simulations (Fig. 3b,c) are similar to those seen in the first intermediate structure from the explicit solvent unfolding trajectories, but with greater loss of β-strand structure.

The single 500 K GBSA trajectory shows an extremely rapid rise in rmsd to about 15 Å after 1 ns (Fig. 2a, yellow dataset) with the protein rapidly reaching a random coil structure with few retained native contacts (Fig. 2c) and only a fleeting sign of an intermediate structure with a small plateau in rmsd (at about 7 Å) after approximately 0.3 ns.

The 300 K GBSA control trajectory shows an rmsd of approximately 2.5 Å, greater than the explicit solvent control trajectory, but stable throughout the trajectory (Fig. 2a). The native contact data from the 300 K GBSA control trajectory show the loss of about 5% of native contacts when compared to the 300 K explicit solvent trajectory (Fig. 2c). These differences do not appear to affect the secondary and tertiary structure of the protein, the losses being contacts on the surface of the protein.

The four 400 K GBSA trajectories show a gradual loss of native contacts with the final number being similar to the first intermediate structure seen in the 500 K explicit solvent simulations (Fig. 2c). The four 450 K GBSA trajectories follow slightly different routes, one trajectory following the pattern of the 500 K explicit solvent datasets, whilst the other trajectories show a much greater initial change, with differing losses of native contacts in the early parts of the different trajectories (Fig. 2c). Examination of the structures from the different trajectories shows this is due to loss of the two-stranded β-sheet in the N-terminal end of the protein early in the most rapidly changing trajectory. The 500 K GBSA run has evidence of a short lived intermediate state, with the loss of 20% of native contacts, which rapidly loses another 40% of the native interactions and stays in this state until the end point of the run (Fig. 2c).

Cluster analysis of the four 400 K trajectories (Fig. 3b) gave seven major clusters (containing at least 1% of the data, using a distance cut-off of 250). The first cluster at 0.3 ns (3.5% of the data) has a near native secondary structure; the cluster at 0.8 ns (9.8%) shows disruption of the N-terminal helical sub-domain, with alterations in the β-sheet interaction between the sub-domains. The cluster at 3.4 ns (52.5%) has the majority of the data and shows

further changes in the helical sub-domain. The cluster at 4.4 ns (12.8%) shows further changes in the helical sub-domain and the first disruption of the β-sheet sub-domain. The last three clusters, between 6.0 and 10.1 ns (19.1% in total), all show the helical sub-domain fluctuating in structure with the β-sheet sub-domain remaining more stable. Nearly all the structural changes are concentrated on the N-terminal helical section of the protein (Fig. 3b), with the β-sheet part remaining more native-like.

Clustering of the four 450 K GBSA trajectories gives seven major clusters (containing at least 1% of the data, using a distance cut-off of 250). The first two average structures (at 0.8 and 1.3 ns, containing 3.6 and 9.9% of data) along the trajectories show big changes in the helical sub-domain with some changes seen in the β-sheet part of the protein (Fig. 3c). These are followed by a structure where there is loss of most of the β-sheet (2.0 ns, 3.3%). The majority of the data is in the four clusters after 3.3 ns, which show fluctuating random coil conformations with no β-strand and only transient helical secondary structure.

Cluster analysis of the single 500 K GBSA (Fig. 3d) run shows three average conformations containing some native secondary structure: the near native structure (0.0 ns, 2.4% of data); a structure with a perturbed N-terminal (0.3 ns, 11.4%) similar to the intermediate seen in the 400 K simulations and a structure with a residual section of β-sheet (0.5 ns, 5.2%). The rest of the trajectory consists of a random coil ensemble which moves away from a globular structure as the trajectory continues.

Solvation Effects

Comparison of the radius of gyration from the explicit and implicit (Fig. 2b) unfolding simulations shows the multiple 400 and 450 K GBSA trajectories have little change in radius, presumably because there are no water molecules to push side chains out from the surface of the protein or to push the helical sub-domain away from the β-sheet. The 500 K GBSA trajectory shows extremely large spikes in the radius of gyration as the protein moves rapidly to a random coil ensemble, but still finishes with a small overall change. The two 500 K explicit solvent trajectories show similar patterns of change in the radius of gyration as have been seen before [22].

The different behaviour between explicit solvent and GBSA radius of gyration datasets is mirrored in the total solvent accessible surface area during the two kinds of simulations (Fig. 2d). The explicitly solvated 500 K simulations both show similar changes to Li and Daggett [22], with the 300 K control trajectory showing normal variation. The GBSA simulation at 300 K shows a small decrease in radius of gyration compared to the crystal structure. The 400 K GBSA trajectories show similar fluctuations to the 300 K explicit solvent control trajectory with no progressive change over time, the 450 K GBSA trajectories show about a ten percent increase during the simulation and the 500 K GBSA trajectory has extremely large spikes.

DISCUSSION

Comparison of the GBSA simulation data with the more established explicit solvent data shows that for barnase, using GBSA methods at 400 K, information can be gained concerning the first intermediate state, whilst with GBSA methods at 450 K, the change from an intermediate state to the final random coil ensemble can be explored.

The two separate explicit solvent trajectories at 500 K can be seen to be exploring subtly different unfolding pathways. The 400 K GBSA trajectories follow similar, but not identical, pathways to get to an intermediate state and these pathways are similar to the early part of the 500 K explicit solvent simulations. Of the 450 K GBSA trajectories, one follows a similar pathway to the first 500 K explicit solvent run, whilst the others have greater changes visible in the intermediate state than either of the 500 K explicit solvent trajectories. The 500 K GBSA trajectory unfolds too rapidly to allow characterisation of intermediate structures along its pathway.

The solvent accessible surface area shows the greatest difference between the GBSA and explicit solvent simulations. This is mainly due to side chains not being able to relax into solution in the GBSA simulations and later on in the higher temperature runs the lack of penetration of solvent into the hydrophobic cores between the helical and β-strand parts of the protein. The radius of gyration also shows little change in the 400 and 450 K GBSA simulations, compared to the increase in the explicit solvent trajectories, due to the same solvation effects. However, the 500 K GBSA simulation shows the protein exploding in a very short timescale and although there is evidence in the native contact and rmsd data of an intermediate similar to the other simulations, this structure has a short lifespan. Both of the 500 K explicit solvent simulations show similarities to the trajectory reported by Li and Daggett [22], but the two runs differ in their fine detail and appear to be following subtly different pathways between states. There are even larger differences between the 450 K GBSA trajectories, with the stable intermediate having markedly different properties in the four runs. This demonstrates an advantage of the more rapid GBSA method in providing statistically valid ensembles during the unfolding process.

The 400 K GBSA simulations appear to be following a pathway similar to the first 500 K explicit run, but with the only structural changes being loss of the β-strands a, b and 1 with changes in α-helices 2 and 3. The simulation appears to have only reached the first intermediate stage and the protein is much more compact that in the explicit solvent simulations. The 450 K GBSA simulations also follow the same pattern of rapid relaxation to a group of stable intermediate structures before the final push to the random coil ensemble.

The different behaviour of the protein at 500 K under both explicit solvent and GBSA simulations shows the importance of the solvent molecules in providing a restraining force on the protein. The 500 K GBSA simulation is no good for exploring intermediate structures as the protein almost immediately jumps to a random coil ensemble—however, dropping the temperature to 400 K gives rise to trajectories where the protein is trapped in the first intermediate state over the time scale of the simulation. For barnase, a temperature of 450 K appears to provide a suitable balance between the kinetic energy needed to unfold the protein and that which jumps the protein over any intermediate structure too quickly. Overall, the GBSA simulations appear to show that the transition from the native structure to the first intermediate is easier than the jump from that intermediate to any second intermediate.

CONCLUSIONS

Not only is the GBSA methodology about two and a half times faster than the explicit solvent simulations but the GBSA simulations change more rapidly than those using explicit solvation. The final random coil ensemble is reached in 4 ns at 450 K and in under 1 ns at 500 K with GBSA, compared to over 5 ns with 500 K explicit conditions. Thus there is a significant time advantage to using the GBSA methodology.

The different behaviour of the GBSA and the explicit solvent simulations appears to be confined mainly to the lack of water molecules penetrating into the protein with the GBSA methodology. There is a fine balance between the protein reaching a stable intermediate structure and the protein jumping immediately to a random coil ensemble. The temperatures used with barnase may not be applicable to the study of other proteins, such as the more stable γ-crystallin lens proteins, which with 500 K explicit solvent simulation show little change [41]. However, once conditions have been tuned for different systems, the method will allow the acquisition of molecular coordinates that capture features of early unfolding intermediates.

Acknowledgements

We would like to thank Drs Dave Houldershaw and Richard Westlake for computation support. We acknowledge the BBSRC for funding the Beowulf cluster. C. Slingsby and A.G. Purkiss are grateful for the financial support of the Medical Research Council, and J.T. Macdonald for a Medical Research Council studentship.

APPENDIX A: SIMULATION PARAMETERS

Both the explicit solvent and GBSA simulations were undertaken with the AMBER 7 [32,33] version of SANDER. The following parameters were used:

Parameter	Value during GBSA trajectories	Value during explicit solvent trajectories
Force evaluation (NTF)	1 (Complete)	1 (Complete)
Periodic boundary (NTB)	0 (None)	2 (Constant pressure)
Nonbonded cut-off (CUT)	18 Å	8 Å
Nonbonded pairlist update (NSNB)	10	10
Timestep (DT)	0.001 ps	0.001 ps
Temperature scaling (NTT)	1 (Constant temperature, weak coupling)	1 (Constant temperature, weak coupling)
Heat bath coupling constant (TAUTP)	1.0 ps	1.0 ps
Pressure scaling flag (NTP)	0 (none)	2 (anisotropic)
SHAKE constraints (NTC)	1 (Not performed)	1 (Not performed)
Internal dielectric (INTDIEL)	1	N/A
External dielectric (EXTDIEL)	78.5	N/A

Each trajectory was undertaken in 100 ps blocks, with velocities retained between blocks (IREST = 1). The GBSA trajectories were undertaken using the "standard" pairwise Born model [34] using the parameter IGB = 1. The explicit solvent trajectories used PME (USE_PME = 1) with the following parameters: ORDER = 4, NFFT1 = 64, NFFT2 = 48, NFFT3 = 60. All other parameters used default SANDER values.

References

[1] Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C. and Pepys, M.B. (1997) "Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis", *Nature* **385**, 787.

[2] Dobson, C.M. (2003) "Protein folding and disease: a view from the first Horizon Symposium", *Nat. Rev. Drug Discov.* **2**, 154.

[3] Dumoulin, M., Last, A.M., Desmyter, A., Decanniere, K., Canet, D., Larsson, G., Spencer, A., Archer, D.B., Sasse, J., Muyldermans, S., Wyns, L., Redfield, C., Matagne, A., Robinson, C.V. and Dobson, C.M. (2003) "A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme", *Nature* **424**, 783.
[4] Hammarstrom, P., Wiseman, R.L., Powers, E.T. and Kelly, J.W. (2003) "Prevention of transthyretin amyloid disease by changing protein misfolding energetics", *Science* **299**, 713.
[5] Dobson, C. (2002) "Protein misfolding and human disease", *Sci. World J.* **2**, 132.
[6] Neira, J.L. and Fersht, A.R. (1999) "Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments", *J. Mol. Biol.* **285**, 1309.
[7] Neira, J.L., Itzhaki, L.S., Ladurner, A.G., Davis, B., de Prat Gay, G. and Fersht, A.R. (1997) "Following co-operative formation of secondary and tertiary structure in a single protein module", *J. Mol. Biol.* **268**, 185.
[8] Fersht, A.R. and Daggett, V. (2002) "Protein folding and unfolding at atomic resolution", *Cell* **108**, 573.
[9] Torshin, I.Y. and Harrison, R.W. (2003) "Protein folding: search for basic physical models", *Sci. World J.* **3**, 623.
[10] Ivankov, D.N., Garbuzynskiy, S.O., Alm, E., Plaxco, K.W., Baker, D. and Finkelstein, A.V. (2003) "Contact order revisited: influence of protein size on the folding rate", *Protein Sci.* **12**, 2057.
[11] Makarov, D.E. and Plaxco, K.W. (2003) "The topomer search model: a simple, quantitative theory of two-state protein folding kinetics", *Protein Sci.* **12**, 17.
[12] Chen, Y.R. and Clark, A.C. (2003) "Equilibrium and kinetic folding of an alpha-helical Greek key protein domain: caspase recruitment domain (CARD) of RICK", *Biochemistry* **42**, 6310.
[13] Chiti, F., Stefani, M., Taddei, N., Ramponi, G. and Dobson, C.M. (2003) "Rationalization of the effects of mutations on peptide and protein aggregation rates", *Nature* **424**, 805.
[14] Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M. and Schwalbe, H. (2002) "Long-range interactions within a nonnative protein", *Science* **295**, 1719.
[15] Jaenicke, R. and Slingsby, C. (2001) "Lens crystallins and their microbial homologs: structure, stability, and function", *Crit. Rev. Biochem. Mol. Biol.* **36**, 435.
[16] Benedek, G.B. (1997) "Cataract as a protein condensation disease: the Proctor Lecture", *Investig. Ophthalmol. Vis. Sci.* **38**, 1911.
[17] Doyle, S.M., Anderson, E., Zhu, D., Braswell, E.H. and Teschke, C.M. (2003) "Rapid unfolding of a domain populates an aggregation-prone intermediate that can be recognized by GroEL", *J. Mol. Biol.* **332**, 937.
[18] Doruker, P., Atilgan, A.R. and Bahar, I. (2000) "Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor", *Proteins* **40**, 512.
[19] Zhou, R. (2003) "Free energy landscape of protein folding in water: explicit vs. implicit solvent", *Proteins* **53**, 148.
[20] Ferrara, P., Apostolakis, J. and Caflisch, A. (2002) "Evaluation of a fast implicit solvent model for molecular dynamics simulations", *Proteins* **46**, 24.
[21] Paci, E., Smith, L.J., Dobson, C.M. and Karplus, M. (2001) "Exploration of partially unfolded states of human alpha-lactalbumin by molecular dynamics simulation", *J. Mol. Biol.* **306**, 329.
[22] Li, A. and Daggett, V. (1998) "Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate", *J. Mol. Biol.* **275**, 677.
[23] Takei, J., Chu, R.A. and Bai, Y. (2000) "Absence of stable intermediates on the folding pathway of barnase", *Proc. Natl Acad. Sci. USA* **97**, 10796.
[24] Fersht, A.R. (2000) "A kinetically significant intermediate in the folding of barnase", *Proc. Natl Acad. Sci. USA* **97**, 14121.
[25] Arcus, V.L., Vuilleumier, S., Freund, S.M., Bycroft, M. and Fersht, A.R. (1995) "A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding", *J. Mol. Biol.* **254**, 305.
[26] Dalby, P.A., Clarke, J., Johnson, C.M. and Fersht, A.R. (1998) "Folding intermediates of wild-type and mutants of barnase. II. Correlation of changes in equilibrium amide exchange kinetics with the population of the folding intermediate", *J. Mol. Biol.* **276**, 647.
[27] Killick, T.R., Freund, S.M. and Fersht, A.R. (1998) "Real-time NMR studies on folding of mutants of barnase and chymotrypsin inhibitor 2", *FEBS Lett.* **423**, 110.
[28] Fersht, A.R., Matouschek, A. and Serrano, L. (1992) "The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding", *J. Mol. Biol.* **224**, 771.
[29] Matouschek, A., Serrano, L. and Fersht, A.R. (1992) "The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure", *J. Mol. Biol.* **224**, 819.
[30] Bond, C.J., Wong, K.B., Clarke, J., Fersht, A.R. and Daggett, V. (1997) "Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway", *Proc. Natl Acad. Sci. USA* **94**, 13409.
[31] Wang, T. and Wade, R.C. (2003) "Implicit solvent models for flexible protein-protein docking by molecular dynamics simulation", *Proteins* **50**, 158.
[32] Perlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, III, T.E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P.A. (1995) "AMBER, a package of computer programs for applying molecular mechancis, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules", *Comp. Phys. Commun.* **91**, 1.
[33] Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (2002) "AMBER 7", [www.scripps.edu].
[34] Tsui, V. and Case, D.A. (2000) "Theory and applications of the generalized Born solvation model in macromolecular simulations", *Biopolymers* **56**, 275.
[35] Bashford, D. and Case, D.A. (2000) "Generalized born models of macromolecular solvation effects", *Annu. Rev. Phys. Chem.* **51**, 129.
[36] Moraitakis, G., Purkiss, A.G. and Goodfellow, J.M. (2003) "Simulated dynamics and biological macromolecules", *Rep. Prog. Phys.* **66**, 383.
[37] Frishman, D. and Argos, P. (1995) "Knowledge-based protein secondary structure assignment", *Proteins* **23**, 566.
[38] Hubbard, S.J. and Thornton, J.M. (1993) "NACCESS" [www.http://wolf.bms.umist.ac.uk/naccess/].
[39] Purkiss, A.G. (2003) "CONTAX" [www http://people.cryst.bbk.ac.uk/(bpurk01/contax/index.html].
[40] Koradi, R., Billeter, M. and Wuthrich, K. (1996) "MOLMOL: a program for display and analysis of macromolecular structures", *J. Mol. Graph.* **14**, 51.
[41] Purkiss, A., Slingsby, C. and Goodfellow, J.M. (2000) "Simulation of the highly stable protein: bovine gammaB-crystallin at room and high temperature", *Protein Pept. Lett.* **7**, 211.