7

Molecular Dynamics Studies of Protein and Peptide Folding and Unfolding

Amedeo Caflisch and Martin Karplus

1. Introduction

Proteins are fascinating. As objects in three-dimensional space, they are sometimes elegant and always complex molecules, yet they consist of only 20 different amino acid building blocks. The function of proteins is determined by their three-dimensional structure, and the majority of biological processes involve one or more protein molecules. The mechanism of the evolutionary development of specific proteins is one of the unsolved problems of biology. Most proteins are very sensitive to their environment; small temperature or pH changes can alter both their stability and their ability to function. The native structure of a protein is determined by the amino acid sequence (Anfinsen, 1972). In solution, many proteins have been shown to refold by themselves under conditions that lead to a stable native state, but *in vivo* the folding process can be very complicated and often involves other proteins, such as chaperones (Gething and Sambrook, 1992).

The time scale for protein folding in solution ranges from nanoseconds to hours. This is much too long for a steepest descent to a minimum of the potential energy surface and much too short for an exhaustive search of the conformational space. It is clear from minimization studies of folding (Nemethy and Scheraga, 1977; Levitt, 1983) and from experimental (Austin et al., 1975) and theoretical studies (Elber and Karplus, 1987; Noguti and

The Protein Folding Problem and Tertiary Structure Prediction
K. Merz, Jr. and S. Le Grand, Editors

Gō, 1989a–e; Caflisch et al., 1992) of the protein conformation space that it is highly complex, with a multiminimum character that prevents folding from being the analogue of a simple, monotonic descent into a potential well. This means that barriers must be overcome in the folding process and the finding of the minimum has analogies to the complexities arising in glassy systems. Since the seminal analysis of Levinthal (1969) it has been realized that alternatives to a random search must be operative in protein folding. Although the conformation space of the denatured state is vast, there would be no search problem if each of the amino acid residues could find its correct conformation independent of the others, or if only interactions with nearest neighbors were involved. Rapid folding would be expected, as in the helix–coil transition of polymers. Protein folding has been compared to crystallization (Harrison and Durbin, 1985), but it is important to realize that it is significantly different. In the former, once a nucleus is formed, there is no problem; i.e., there are many similar sites at which subsequent molecules can coalesce independently. In protein folding, even if there were a stable nucleus, it is not clear that the protein could continue to fold by the independent condensation of residues. The fundamental distinction between the helix–coil transition or crystallization and protein folding is that long-range interactions play an essential role in determining the native state. In the limit, this implies that the conformational energy of each amino acid residue depends on all of the others in the polypeptide chain. It is this aspect of the search of the full configuration space, with its vast numbers of conformers, that leads to the Levinthal paradox.

A recent study (Zwanzig et al., 1992) argues that the Levinthal paradox can be resolved simply by introducing local main-chain propensities. They examined a bead model for the polypeptide chain that has only local interactions and found that a small bias of the main-chain propensities results in a very large reduction in the apparent first passage time to a folded state. Such an analysis misses the essential point, outlined in the previous paragraph, concerning the difference between the helix–coil transition, which corresponds to the case studied by Zwanzig et al., and the cooperative transition of a protein.

Naively, the best way to find out how a protein folds is to start with a model for the denatured state of a protein and to simulate the transition to the native state. The simulation itself would solve the Levinthal paradox if the model used were sufficiently detailed and accurate. An approach that should work, in principle, is to use an atom-based model for the potential energy function and to follow a molecular dynamics trajectory from one or more denatured conformers to the native state in the presence of the

appropriate solvent. With the available methodology and computing power, such a simulation would require approximately 10^{11} hours (or 10^7 years) for a 100-residue protein where the experimental transition to the folded state takes place in about 1 sec. With the teraflop supercomputers that are on the horizon (Deng et al., 1990), such simulations would still take 10^3 years, although it would be possible to examine the faster portions of the folding transition by such a direct method. Clearly, when such simulations become possible they should be done, even though it is likely that a very large fraction of the trajectory would be uninteresting and that a very large amount of human time would be required to interpret the results.

The practical difficulties in doing such brute force simulations has led to several types of theoretical approaches to the dynamics of protein folding. One approach is based on the complete atomic model, but biases the system to speed up the folding transition by several orders of magnitude (as in simulated annealing with NMR distance constraints), or instead studies unfolding under special conditions (e.g., use of a forcing potential, high temperature) that reduce the time involved from microseconds to picoseconds. Related studies deal with small protein fragments (e.g., α-helices, β-turns) for which the conformation space is sufficiently small so that full searches can be accomplished and/or transitions of interest occur on a manageable time scale. Alternatively, the entire polypeptide is considered, but a simplified description is introduced. One approach is to approximate each amino acid residue by one or two quasi particles and to use the C_α pseudo-dihedral angles as the only variables. The problem can be further reduced by restricting the quasi particles to move on a lattice (e.g., a cubic lattice) and using Monte Carlo procedures to follow the folding process. Such lattice calculations have the advantage over the physically more correct ("off-lattice") atomic models in that the total number of conformations is reduced, and a large fraction of all the conformations can be examined in a given time (Shakhnovich and Gutin, 1990; Shakhnovich et al., 1991).

In addition to methods that are based on following the dynamics of a polypeptide chain with varying levels of detail, there exist a range of models for protein folding that are phenomenological in character (Karplus and Weaver, 1976, 1979). These models provide a qualitative description of the folding process, often including a conceptual approach to the solution of the Levinthal paradox. In this review we concern ourselves primarily with molecular dynamics simulations of peptides and proteins that are being used to analyze elements of protein folding and unfolding. Such simulations constitute an approach to the protein folding problem that has been made possible relatively recently by improved simulation technology and the increase in speed of the available computers. A review of the theory of

protein folding that considers other aspects of the problems is given by Karplus and Shakhnovich (1992).

2. Fragment Studies

An approach that can be used to obtain information concerning the detailed kinetics of protein folding is to study peptide fragments that can be treated by detailed simulations. A number of such studies have been made, and more are in progress. To illustrate the possibilities, we shall describe several of them here and outline the more important results.

Czerminski and Elber (1989, 1990) did vacuum simulations of a blocked alanine tripeptide, which is the shortest unit that can form an i to $i + 4$ helical hydrogen bond. The simplicity of this system and the absence of solvent permitted them to make a full analysis of the multiminimum nature of the potential surface. Minima were found with the three peptide units in the neighborhood of the local minima for (ϕ, ψ) equal to $(-60, -60;$ α-helix), $(-120, 120;$ β-sheet), $(-60, 60)$ and $(60, -60)$. A total of 138 minima were found in the energy range up to 6.3 kcal/mol, relative to the lowest minimum. This is to be compared with $4^{3.5}$, or 128, combinations of the most probable (ϕ, ψ) values. The 6,216 reaction paths between the minima were investigated. For direct paths connecting two minima without an intermediate, the barriers were found to be in the range of 0.5 to 5.5 kcal/mol, with the higher energies most common. Most of the direct paths involved changes in only one dihedral angle, with two thirds of the "flips" involving ψ and one third involving ϕ. Whether indirect paths led to significantly lower barriers between pairs of minima was not discussed. A master equation approach was used to study the dynamics, with the kinetic constants estimated from transition state theory. This led to lifetimes for the various minima in the picosecond to nanosecond range at 400 K.

The "folding/unfolding" transition of a blocked alanine dipeptide was investigated recently (Lazaridis et al., 1991). The reaction path between the "folded" configuration $(\phi_1,\psi_1;\ \phi_2,\psi_2)$ equal to $(-72, -57;\ -81, -31)$ and an "extended" conformation $(-83, 129;\ -83, 129)$ was determined *in vacuo* by a free-energy simulation method with a dielectric constant of 50 to mimic the effect of an aqueous solvent. Two paths were found that are essentially mirror images of each other with ψ_1 and ψ_2 undergoing successive changes (ϕ_1 changes very little); an intermediate was found where either ψ_1 or ψ_2 had undergone a transition. The calculated barriers were in the range of 1.5 to 2 kcal/mol. In a free-energy simulation including an explicit model for the water molecules, similar results were obtained.

The major difference in the latter is that the extended configuration with hydrogen bonds to water is about 3.4 kcal/mol more stable than the "turn," which is similar to an earlier value (Tobias et al., 1990); in the model with a dielectric constant of 50, the extended conformer was less stable by about 1 kcal/mol. The form of the potential of mean force in the presence of water is similar to that in the absence of water with an intermediate and with barriers of 2.3 and 2.8 kcal/mol.

Case et al. (personal communication) have studied the dynamics of tetrapeptides for which there are NMR data indicating that the turn conformation makes a significant contribution to the equilibrium population in aqueous solution. In a 5 ns simulation of the peptide Ala-Pro-Gly-Asp, they found transitions between "turns" with 1–3 and 1–4 bifurcated C=O···H-N hydrogen bonds and essentially extended conformations on a time scale of about 500 ps. It was noted that the transitions had diffusive elements, suggesting that a simple transition state description might not be valid. This corresponds to the results of stochastic dynamics simulations of the β-strand to coil transition (Yapa et al., 1992).

A study (Robert and Karplus, unpublished results) of the folding of the peptide $(Ala)_2(Asp)(Ala)_6$ from an extended chain to an α-helix *in vacuo* (dielectric constant of 1) and in aqueous solution suggest similar folding mechanisms, but significantly different time scales. *In vacuo*, a structure formed rapidly (3 ps), in which the carboxyl group of the Asp interacted with three NH groups. This remained stable for about 80 ps, when a fourth NH group moved in and this initiated rapid (in a few ps) helix formation. The simulation in aqueous solution required about 200 ps to form the three NH/Asp configurations, and helix initiation required on the order of 600 ps.

Tirado-Rives and Jorgensen (1991) used a molecular dynamics simulation to study a 15-residue ribonuclease A S-peptide analogue at 5° C and 75° C in the presence of an explicit water model. They found that the peptide was stable for 300 ps at 5° C, while it unfolded in less than 500 ps at 75° C. Although experiments show that the peptide is more stable at low temperature (in agreement with the calculation), it is clear that equilibrium was not reached in the dynamics, since the observed helix–coil equilibrium constant is near unity at 5° C. In the two higher-temperature simulations, unfolding occurred; in one, the transition took about 100 ps and in the other about 350 ps. Examination of the simulation suggested that an important element of the unfolding transition is the replacement of an α-helical (i to $i + 4$) hydrogen bond by water hydrogen bonds through an intermediate involving a 3_{10} (i to $i + 3$) or reverse turn hydrogen bond. This is in accord with the x-ray studies of Sundaralingam and Sekharudu (1989), who observed that when a water molecule inserts into an α-helix,

a reverse turn is frequently formed; see also Karle et al. (1990). Details of the dihedral angle transitions are not given in the paper, so no conclusion can be drawn about their rates. Also, it is important to recognize that in a simulation of helix unfolding, nothing is learned about the rate-determining helix-initiation step.

In related work, Soman et al. (1991) have simulated the isolated helix H (residues 132 to 149) of myoglobin for 1 ns at 300 K. They found that the helix unfolded progressively from the C to the N terminus. Tight (i to $i+3$) turns (3_{10} helices) were intermediates in the unfolding and often appeared prior to aqueous solvation. Insertion of H_2O occured in some cases, but not all. Often a single water molecule was inserted between the C=O and NH, but sometimes several water molecules were involved in transient intermediates. The transition were found to have diffusive elements with a time scale on the order of 100 ps. Considerable detail about the unfolding process is given in the paper.

A 100 ps molecular dynamics simulation of hydrated decaalanine (initially in a canonical right-handed conformation) at 300 K was performed by DiCapua et al. (1990). The distances between carbonyl oxygen and the amide nitrogen in the i to $i+4$ hydrogen bonds were monitored. During the trajectory the helix was stable, except for the 5 NH–OC 1 hydrogen bond, which separated at about 15 ps (for 10 ps) and at ca. 70 ps for the remainder of the simulation. Analysis of the orientation of the water revealed that a bridge structure (i.e., water acting as donor for the CO group of residue 1 and acceptor for the NH of residue 5) was formed after about 70 ps but was not present during the transient destabilization, although a water was inserted and acted mainly as a hydrogen-bond donor. The helix destabilization coupled to water insertion is in agreement with simulations described above and the experimental results of Sundaralingam and Sekharudu (1989).

The helix denaturation of a polyalanine peptide (13 residues) was simulated both *in vacuo* and in aqueous solution with an explicit water model by Daggett and Levitt (1992a). They performed several simulations, lasting 200 ps each, at temperatures ranging from 298 to 473 K. The helical state was stable *in vacuo* at all temperatures, whereas in solution the helix denatured at higher temperatures. The role of water in destabilizing and denaturing the helix was investigated. Daggett and Levitt concluded that the type of water insertion proposed by Sundaralingam and Sekharudu (1989) was not the mechanism of helix unfolding. Instead they found that the helices tended to open up due to high-temperature fluctuation, after which the polar C=O and NH groups of the helix interacted with water molecules.

Although the simulations of Daggett and Levitt (1992a) were too short to reach equilibrium, the free energy was estimated from the average waiting

time for the helix→coil and coil→helix transition for each residue (a fast event treated as if it were at equilibrium) and then averaged over all residues during the last 100 ps of each trajectory. It was found that in solution the free energy change for the helix→coil transition decreased with increasing temperature and transitions to coil conformations were unfavorable below 423 K. *In vacuo*, such transitions were unfavorable at all temperatures and the free energy was nearly temperature independent. Moreover, the free energy of activation for the single residue transition was found to increase with increasing temperature in all cases except for the helix→coil transition in solution.

Another type of study of peptide fragments makes use of free energy simulation methods for estimating the change of stability when amino acids are altered. An example is provided by molecular dynamics simulations, which were used to estimate the change in the conformational stability of proline-replaced polyalanine α-helices (Yun et al., 1991). An unfavorable effect was found for proline replacements in the middle of the helix, whereas a substitution at the first position of the helical N-terminus yielded a favorable contribution to the free energy of folding. In a related study, good agreement with experimental results was obtained for the differences in stability of oligoalanine peptides with single amino acid replacements in the middle of the chain (Hermans et al., 1992). Although this type of work is not concerned with the dynamics of folding, it is nevertheless of interest in the present context because it may be useful in interpreting the stability of intermediates or protein fragments involved in folding mechanisms, such as the diffusion collision model.

A number of simulations of protein fragments have reached longer time scales than those possible in the all-atom plus solvent simulations described above. The method uses a potential of mean force surface that implicitly includes the solvent and employs stochastic dynamics to introduce the dynamic effects of the solvent. To further simplify the problem, each residue is represented by a single interaction center ("atom") located at the centroid of the corresponding side chain, and the centers are linked by "virtual" bonds (Levitt, 1976; McCammon et al., 1980). For the potential energy of interaction between the residues, assumed to be Val in an α-helical simulation (McCammon et al., 1980), a set of energy parameters obtained by averaging over the side-chain orientations was used (Levitt, 1976). Terms that approximate solvation and the stabilization energy of helix formation were included. The diffusive motion of the chain "atoms" expected in water was simulated by using a stochastic dynamics algorithm based upon Brownian dynamics. Starting from an all-helical conformation, the dynamics of several residues at the end of a 15-residue chain was monitored in a number

of independent 12.5 ns simulations at 298 K. The mobility of the terminal residues was quite large, with an approximate rate constant of 10^9 per sec for the transition between coil and helix states. This mobility decreased for residues further into the chain. Unwinding of an interior residue required simultaneous displacements of several residues in the coil, so larger solvent frictional forces were involved. The coil region did not move as rigid body. Instead, the torsional motions of the chain were correlated so as to minimize dissipative effects. Such concerted behavior is not consistent with the conventional idea that successive transitions occur independently. Analysis of the chain diffusion tensor showed that the frequent occurrence of correlated transitions resulted from the relatively small frictional forces associated with such motions (Pear et al., 1981). Further, the correlated nature of the torsional transition suggests that unwinding occurs in a relatively localized fashion and that a limiting value of about 10^7 per sec would be reached for the interior of the helix.

A similar model has been used recently to study β-sheet to coil transitions in a β-hairpin (Yapa et al., 1992). From an analysis of ten 90 ns simulations, it was shown that rate constants for the transition between the coil and strand state are on the order of 10^{10} to 10^{11} per sec. This high rate occurs even though the adiabatic potential has barriers on the order of 3 kcal/mol. Unlike the α-helix results, the transition rate constant decreases only slowly as one goes from the end of the strand toward the interior. Also, nonterminal residues are sometimes found in the coil state, while the end residues form a regular sheet. This behavior may be related to the occurrence of β-bulges.

A corresponding approach has been applied to the behavior of structural motifs of proteins. An example is a simulation of two α-helices connected by a coil segment (Lee et al., 1987). This simulation served to examine a possible elementary step in protein folding, as described by the diffusion-collision model (Karplus and Weaver, 1976), i.e., the coalescence of a pair of helices. The system considered in the simulations is a 24-residue peptide in which the first and last eight residues formed an α-helix and the intervening eight residues are initially in a random-coil conformation. Twelve trajectories were generated with a total time of 820 ns. The exact lengths of the individual trajectories were not important because the parameters were chosen so that a stable, coalesced structure does not form. Instead, the system folds and unfolds many times during the simulation, and rate constants for the coalescence and dissociation reactions could be determined. The values obtained are on the order of 10^8 per sec. For the unfolded system, there is a strong bias in the connecting loop towards shorter distances, with a maximum in the radial distribution function near 27 Å, even though the

fully extended conformation has a length of 45 Å. This is due primarily to the entropic contribution, since the intervening chain has many more allowed conformations at intermediate lengths. Such a trend has been observed experimentally in a study of the end-to-end distances in a series of oligopeptides (Haas et al., 1978). This model calculation of coalescence used stable helices to reduce the time required. In a real system, helix to coil transitions would be coupled to the collisional association events. Given the estimates of the helix to coil transition rates, such a study should be possible by combining a stochastic dynamics simulation with a kinetic model for the helix–coil transition.

3. Simplified Protein Simulations

Simulation of folding for a complete protein represent a much more complex problem than the fragment studies just described. Early work in this area (Levitt and Warshel, 1975) used a C_α-type model for the protein and coupled energy minimization with thermalization *in vacuo* in an attempt to fold bovine pancreatic trypsin inhibitor (BPTI), a 58-residue protein. Conceptually, such an approach is of interest; if the potential surface, or better, potential of mean force surface of a protein, were such that a relatively simple procedure can find the minimum, there would be no Levinthal paradox. The limiting case would be a single potential well connecting the extended coil conformations to the native structure. Clearly that is not what is found for real proteins, although an empirical potential can be modified to obtain a relatively well behaved surface that leads to a folded structure. In the potential used for the BPTI folding, for example, biases for turns and extended strands were introduced in the regions that have such conformations in the native structure (Levitt and Warshel, 1975; Hagler and Honig, 1978). With such a potential, one out of five runs that began with the terminal α-helix present and involved 600 cycles of minimization and thermalization had features of the native form, although the root mean square (RMS) difference was about 6 Å, as compared with 3.4 Å, the best value possible for the simplified model.

Another way of modifying the potential to achieve rapid folding is to introduce distance constraints. These have been used to simplify the energy minimization problem and, more recently, as a way of employing molecular dynamics with a simulated annealing protocol for structure determination by NMR (Brünger and Karplus, 1991). In a trial study of the protein crambin with NOE interproton distance constraints (Brünger et al., 1986), it was found that the native structure is achieved from an extended conformation in a few picoseconds of high-temperature molecular dynamics, a

time on the order of 10^{12} times faster than that observed in solution. This decrease in time appears to involve two aspects of the constraints. The first is that stable secondary structural elements are formed in 1 psec, instead of approximately 10 msec, and the second aspect is that the need to search a large portion of the conformation space is eliminated by the long-range constraints (in terms of distance along the chain rather than physical distance). It was shown that if the secondary structure is not formed before collapse, misfolded structures that are local minima can result. This is consistent with the diffusion–collision and related phenomenological description of protein folding (Karplus and Weaver, 1979; Kim and Baldwin, 1990).

A BPTI fragment that contains the 30,51 S–S bond and has a large portion of the β-sheet and the C-terminal α-helix has been studied experimentally by Oas and Kim (1988); they have shown that the system is marginally stable with a structure close to that of native BPTI. When this construct was simulated at 500 K (Robert and Karplus, unpublished results), rapid denaturation occurred. A corresponding simulation at 500 K with the explicit inclusion of solvent showed the beginnings of denaturation over a time scale of 200 psec. The C-terminal α-helix began to unfold and water molecules replaced the terminal hydrogen bonds of the β-sheet. The behavior corresponds to the fragment studies cited previously, in term of the slowing of the time scale relative to vacuum simulations and the explicit role of water–protein interactions at a certain stage of denaturation.

4. High-Temperature Unfolding Simulations of Proteins

Recently, a number of high-temperature molecular dynamics simulations of protein denaturation have appeared, and more are in progress. The simulation of protein unfolding is of interest for several reasons. First, the utilization of the x-ray structure of the native protein as the starting point provides initial conditions corresponding to a known, physically stable state. The choice of an unfolded or denaturated structure for the simulation of protein folding would be more difficult because of the lack of experimental information and the multiplicity of possible starting structures. Second, it is known that protein folding is cooperative in that the transition between the denaturated and native state is of the all-or-none type at equilibrium (Karplus and Shakhnovich, 1992). There is evidence that the rate-limiting step (i.e., the main free-energy barrier to folding and unfolding) is near the native state in many proteins (Creighton, 1988). Thus, the investigation of the initial phases of protein unfolding can help to elucidate the cooperative nature of protein folding, as well as of unfolding. Third, a number

of detailed experimental studies of the protein unfolding transition have been made; an example is given in the barnase results described below. Finally, it has been shown by experiments that several proteins have stable, partially folded intermediates under mild denaturing conditions (Dolgikh et al., 1981; Ptitsyn et al., 1990; for reviews, see Kuwajima, 1989; Ptitsyn, 1992). These intermediates are characterized by native-like secondary structure and a native-like tertiary fold. However, their tertiary structure is less rigid than in the native state, probably due to the absence of the tight packing of side chains that occurs in the protein interior (Ptitsyn, 1992). The term "molten globule" has been introduced for such compact states (Ohgushi and Wada, 1983), which appear to be intermediates on the folding pathway in some cases. There is evidence that the main energy barrier in protein folding is encountered after the formation of the molten globule state, i.e., between the molten globule and the native state. It is likely that this part of the folding process is mirrored by the presently available simulations of the unfolding process.

Vacuum molecular dynamics simulations with a polar hydrogen model for reduced BPTI were performed by heating the system from 300 to 1000 K over a time of 450 psec (Brady et al., unpublished results). Such a simulation of denaturation is in accord with the thermodynamic analysis of Shakhnovich and Finkelstein (1989) in which the first step of denaturation is regarded as a "vacuum" process and the protein expands to a molten, globule-like state with a transition temperature of the order of 500 K. A relatively sharp transition in the RMS deviation from the native structure was found at about 500 K; the RMS increased from about 2 Å to 4 Å. After this, the RMS increased slowly until there was a second transition close to 1000 K, in which the RMS reached about 7 Å. The first transition corresponds to a loss of tertiary structure with preservation of secondary structure; i.e., the β-sheet and α-helix are present but reoriented relative to each other. In the second transition all the secondary structure is lost, with the α-helix disappearing before the β-sheet. The loss of the secondary structures is accompanied by the formation of new hydrogen bonds, including "C5 ring"-like structures, a local minimum on the vacuum dipeptide map. The final system is a random, coil-like globule, whose radius of gyration is only slightly larger that that of native BPTI.

Fan et al. (1991) have utilized high-temperature vacuum molecular dynamics to study the molten globule state of α-lactalbumin. Some native-like secondary structure was recently shown to be present in the molten-globule state of guinea pig α-lactalbumin by NMR hydrogen exchange experiments (Baum et al., 1989; Dobson et al., 1991). A number of main-chain NH groups were observed to be highly protected from solvent exchange in the

molten globule state; these were found to occur mainly in segments that are helical in the native structure. In particular, two major helical regions, including residues 23–34 (helix B) and residues 89–96 (helix C), were identified as stable by these experiments. To investigate the stability of α-helices B and C in α-lactalbumin, Fan et al. performed two sets of simulations (with a dielectric constant of $4r$, r being the distance in Å between interacting atoms); one was at a temperature of 500 K (lasting 50 ps) and the other at 1000 K (lasting 20 ps). Starting coordinates were model-built from the x-ray structure of the homologous baboon α-lactalbumin (Acharya et al., 1989) since no x-ray coordinates are available for the guinea pig protein. Fan et al. imposed harmonic constraints on the helices in one simulation and on the hydrophobic core in another, for each value of the temperature. Upper and lower bounds for the distances of all pairs of constrained atoms were introduced by adding or subtracting 0.2 Å from the distances in the native structure. They found that constraining the helices does not stabilize the hydrophobic core, whereas constraining the hydrophobic core does stabilize the α-helices. The hydrophobic contribution to the free energy of unfolding was estimated from the change in the solvent accessible surface area during the simulation. It was found that this contribution was more favorable for the structures obtained with hydrophobic constraints than those obtained with the helix constraints. As expected, conformations characterized by a more compact hydrophobic core have a better hydrophobic free energy. Fan et al. do not give quantitative description of the behavior of the radius of gyration and of RMS deviation from the x-ray structure during the simulation. It is difficult to determine, therefore, whether what they observe corresponds to the molten globule state. Nevertheless, their results on the coupling of helix stability with the existence of a hydrophobic core is of considerable interest.

Daggett and Levitt (1992b) investigated the unfolding of reduced bovine pancreatic trypsin inhibitor (BPTI) by high-temperature molecular dynamics. BPTI was solvated in a box of water molecules and periodic boundary conditions were applied. Simulations were performed at constant temperature by coupling to an external bath (Berendsen et al., 1984). The system was brought to the target temperature by scaling the velocities. Five simulations (lasting up to 550 ps) were performed: native BPTI at 298 K and 423 K and fully reduced BPTI at 298 K, 423 K, and 498 K; in the latter, all three S–S bonds were not present. Properties of BPTI as a function of temperature and the state of the disulfide bonds were averaged over the entire simulation. Since no detailed description of the behavior of such properties as a function of time is given in the paper, it is difficult to interpret the results, e.g., to determine whether the simulation reached a stable unfolding inter-

mediate. The main conclusion presented by Daggett and Levitt (1992b) is that reduced BPTI at high temperature was relatively compact but had a larger radius of gyration than native BPTI (a volume increase of 5 to 25%), while there was little change at 423 K in the oxidized form; presumably a higher temperature would have been required to denature the native protein on the simulation time scale. Reduced BPTI at high temperature showed increased main-chain and side-chain mobility over that of the native protein. Since the average number of water molecules inside reduced BPTI was observed to be independent of structure, the expansion was not caused by solvent penetration. Thus, the present results are rather similar to those observed in the vacuum simulation of Brady et al. (unpublished) described above. However, it is important to note that the authors used a reduced density in their constant volume, high-temperature simulations. This is likely to be the reason for the lack of water penetration, which is in disagreement with experiment (Ptitsyn, 1992) and with the barnase simulations described below. Moreover, when a low-temperature (298 K) simulation with normal water density was performed for the unfolded reduced structure, there was a significant increase of protein–water interactions over that in the native state. Because of the lack of experimental results on BPTI kinetic intermediates, the authors compared the BPTI simulation with data for other proteins.

Mark and van Gunsteren (1992) have reported a simulation of the thermal denaturation of hen egg white lysozyme. As in Daggett and Levitt (1992b), the enzyme was solvated in a box of water molecules with periodic boundary conditions and the temperature was kept constant by coupling to an external heat bath. They performed a control run of 550 ps at 300 K and a run at 500 K, which was branched after 85 ps by cooling the system to 320 K over 5 ps. The simulation at 320 K was continued for 190 ps, while the 500 K simulation was stopped after 180 ps. They monitored the radius of gyration, the RMS deviation of the Cα atoms from the starting x-ray structure, and changes in the main-chain hydrogen bonding pattern as a function of simulation time. At 500 K, the radius of gyration increased steadily for about 90 ps and then remained effectively constant for 40 ps before rising again more rapidly than before. A similar behavior was observed for the C_α RMS deviation from the x-ray structure. The simulation branched from the intermediate state showed an approximately constant radius of gyration at 320 K, a marked increase in solvent penetration (the authors report a 50% increase in solvent accessible surface area but do not discuss the penetration of the water molecules) and a high degree of native-like secondary structure. Hydrogen exchange NMR experiments (Miranker et al., 1991) have shown that lysozyme has two distinct folding domains, which corre-

spond to the two structural domains (McCammon et al., 1976; Janin and Wodak, 1983). No difference between the two domains was observed in the simulation. Van Gunsteren (private communication) has argued that this result is not inconsistent with the hydrogen exchange experiments, i.e., that rapid fluctuations in accessibility can explain the NMR results. However, this argument is incorrect because it is based on a misunderstanding of the mechanism of hydrogen exchange, which, for the lysozyme, measures the equilibrium between the unfolded (exposed NH) and folded (protected NH) structures, rather than fluctuations in the latter. From the number of helical main-chain hydrogen bonds as function of time, Mark and van Gunsteren (1992) concluded that the formation of secondary structure occurs after the collapse of the peptide chain. This is in sharp contrast with experimental results based on far-UV ellipticity measurements, which demonstrate that there is formation of secondary structure at an early stage of folding (Gilmanshin and Ptitsyn, 1987; Kuwajima et al., 1987; Goldberg et al., 1990).

Recently, molecular dynamics simulations lasting 500 ps or more of apomyoglobin in aqueous solution were carried out by Brooks (1992) at 312 K, and by Tirado-Rives and Jorgensen (1993) at 298 K and at 358 K (pH values of 6 and 4 were used at the high temperature). The studies used different force fields; the CHARMM polar hydrogen model (Brooks et al., 1983) was used by Brooks (1992) and the AMBER-OPLS all-atom model by Tirado-Rives and Jorgensen (1993); the same water model, TIP3P (Jorgensen et al., 1983), was used by the two groups. The two room-temperature simulations (298 K and 312 K) address the native structure of apomyoglobin. In both simulations helices A, E, G, and H were stable. However, for the remaining helices, there are differences. Brooks (1992) found that the other helices also existed during the entire simulation, with B as stable as A, E, G, and H, while C, D, and F were more mobile and moved into the heme cavity. By contrast, Tirado-Rives and Jorgensen (1992) found that helix B unfolded in part and helices C, D, and F unfolded completely in a relatively short time. Both results are consistent with the experimental findings of Hougson et al. (1990, 1991) in that helices A, G, and H constitute a stable subdomain that retains substantial helicity upon partial unfolding in environments of decreasing pH; as to the other helices, the experiments do not clearly distinguish between the two simulations. Brooks (1992) found that the all-atom RMS deviation (RMSD) from the crystal structure reached a plateau value after about 60 ps. Different behavior of the RMSD for the main-chain heavy atoms relative to the crystal structure were found at 358 K by Tirado-Rives and Jorgensen (1993); in the 358 K run at pH = 4, the RMSD increased continuously until the end of the simulation (500 ps),

whereas in the 358 K run at pH = 6 it reached a plateau value after the first 350 ps. It is surprising that in the 298 K control run (pH value of 6.0) the RMSD did not reach a plateau value. Tirado-Rives and Jorgensen state that the helical content averaged over the last 50 ps of both trajectories at 358 K was in agreement with CD estimates (Houghson et al., 1990). No discussion of the role of water is given in either paper.

High-temperature molecular dynamics simulations have been performed to study the unfolding of barnase, a ribonuclease excreted from *Bacillus amyloliquefaciens* for which a large body of experimental work on folding and unfolding is available from Fersht and co-workers. Barnase is a monomeric enzyme (110 residues) that is of particular interest as a folding model, as suggested by Fersht (for a review, see Fersht, 1993). Both its crystal structures (Mauguen et al., 1982; Baudet and Janin, 1991) and its structure in solution are known (Bycroft et al., 1991); there are no disulfide linkages to constrain the unfolded state and the three prolines in barnase are all trans. The structure of barnase consists of three α-helices and a five-stranded β-sheet that are stabilized by three hydrophobic cores in the native structure (Figure 7-1). Transition states and pathways of barnase folding have been investigated in detail by protein engineering (Matouschek et al., 1989; Fersht et al., 1992a; Matouschek et al., 1992a; Meiering et al., 1992; Serrano et al., 1992a, 1992b) and NMR hydrogen-exchange trapping experiments (Bycroft et al., 1990; Matouschek et al., 1992b). The two techniques, which yield complementary results (Fersht et al., 1992b; Serrano et al., 1992c), have shown that the rate-determining step for both folding and unfolding involves the crossing of a free-energy barrier near the native state.

Solvated barnase denaturation simulations started with the x-ray structure and were performed with a deformable boundary potential (Brooks and Karplus, 1983, 1989); the system consisted of 9009 solvent atoms and 1091 protein atoms. Two denaturation simulation were performed at 600 K (A600, 120 ps; R600, 150 ps) and a 300 K control trajectory was run for 250 ps. The 600 K temperature was used to speed up the unfolding transition by at least six orders of magnitude relative to the experimental denaturation temperature of approximately 327 K; the activation energy for unfolding is 20 kcal/mol (Kellis et al., 1989; Matouschek et al., 1990). The A600 simulation was started after 4 ps of simulation at 300 K, while the R600 run was initiated after 100 ps of simulation at 300 K. The two different initial structures were chosen to evaluate the effect of the initial conditions on the overall unfolding behavior. The A600 simulation was branched after 90 ps dynamics, by cooling the system to 300 K; it was continued for 160 ps; the name B300 is used to designate this simulation. The

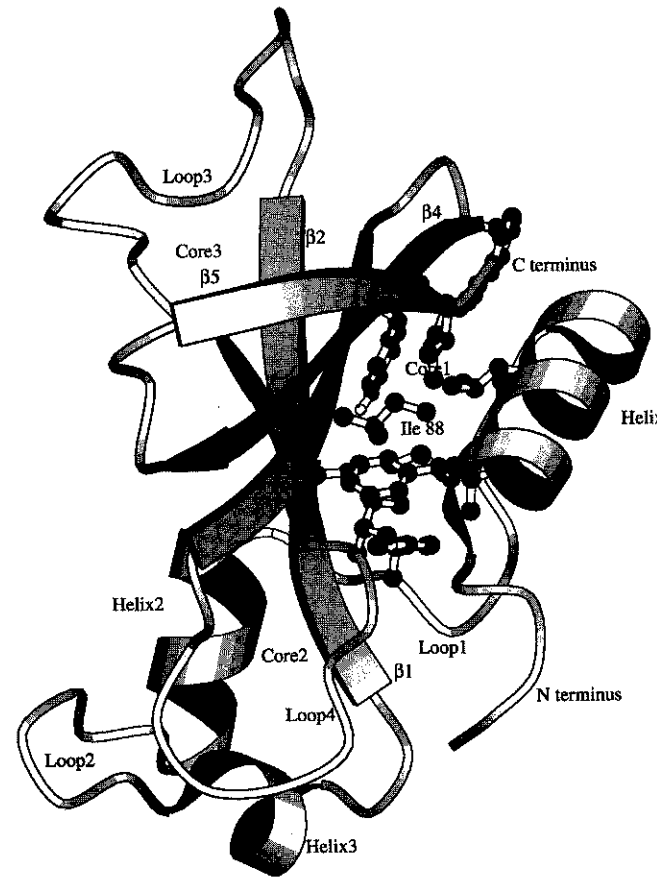

Figure 7-1. Schematic picture of the backbone of barnase, emphasizing the secondary structure elements. Side chains of hydrophobic $core_1$ are plotted in a ball-and-stick representation. The secondary structural elements include the following residues: N-terminus (1–5), $helix_1$ (6–18), $loop_1$ (19–25), $helix_2$ (26–34), $loop_2$ (35–40), $helix_3$ (41–45), type II β-turn (46–49), $strand_1$ (50–55), $loop_3$ (56–69), $strand_2$ (70–75), $loop_4$ (76–84), $strand_3$ (85–90), type I β-turn (91–94), $strand_4$ (95–100), type III' β-turn (101–104), $strand_5$ (105–108), C-terminus (109–110). The atomic coordinates of the protein were kindly provided by Dr. A. Cameron and Professor G. Dodson of the University of York. Figure made with the Molscript program (Kraulis, 1991).

temperature of the system was controlled by weak coupling to an external bath (Berendsen et al., 1984).

In addition, two molecular dynamics runs starting with the x-ray structure and lasting 250 ps were performed at 600 K (V600) and 800 K (V800)

without explicit solvent molecules (*in vacuo*) starting from the x-ray structure; a distance dependent dielectric constant was used, as in the simulation of Fan et al. (1991). During the V600 *in vacuo* run, the radius of gyration (R_g) decreased to about 13.0 Å in the first 50 ps (R_g of the minimized x-ray structure is 13.6 Å), and was approximately constant during the rest of the simulation. The heavy atom RMS positional deviation (RMSD) from the x-ray structure increased to about 4.5 Å in the first 60 ps and did not change significantly after that. In V800, R_g increased to about 14.3 Å during the first 10 ps; it then decreased to a value that oscillated between 13.5 and 14.0 Å in the last 50 ps, while the heavy atom RMSD from the x-ray structure reached a plateau value of 9 Å after about 170 ps. The structures obtained at the end of the V600 and V800 runs are characterized by a tight packing of the loops and a large amount of distortion in the secondary structure elements. At the end of V600 the major α-helix ($helix_1$) of barnase preserves about 40% of its helical content and packs against the stable β-sheet. In V800, $helix_1$ unfolded during the first 50 ps; most of the β-sheet was disrupted in the first 50 ps (apart from strands 3 and 4, which lasted until 120 ps). The *in vacuo* simulations are briefly compared with the simulations that included explicit water molecules at the end of this section.

Plots of R_g and RMSD as a function of time for the simulations with explicit solvent are given in Figures 7-2a and 7-2b. In the control simulation at 300 K, R_g showed a very small increase (the average R_g is 13.7 Å), which was mainly due to the rearrangements of surface hydrophilic side chains, and the heavy atom RMSD from the x-ray structure is 1.9 Å (the main-chain atom RMSD is 1.5 Å) during the last 50 ps. The overall conformation, hydrophobic core compactness (Figure 7-5a–c), and secondary structure elements are found to be stable at room temperature. Furthermore, an average of 0, 2, and 1 water molecules were present in $core_1$ (within 7 Å from the core center), $core_2$ (within 6 Å), and $core_3$ (within 6 Å), respectively. Experimental evidence indicates that $core_2$ contains three water molecules, whereas $core_1$ and $core_3$ are fully inaccessible to solvent in the x-ray structure (A. Cameron and K. Henrick, unpublished results; Serrano et al., 1992c).

In A600 and R600, R_g started to increase only after 30 ps, while the RMSD increased immediately. This suggests that the protein was undergoing a local conformational search to find convenient pathway(s) for denaturation. In the A600 simulation, R_g increased between 30 and 45 ps to about 14.2 Å; this was followed by the formation of an intermediate state, while in R600, R_g continuously increased during the first 70 ps. In A600, the intermediate state lasted for about 18 ps; the average R_g of the 45–63 ps

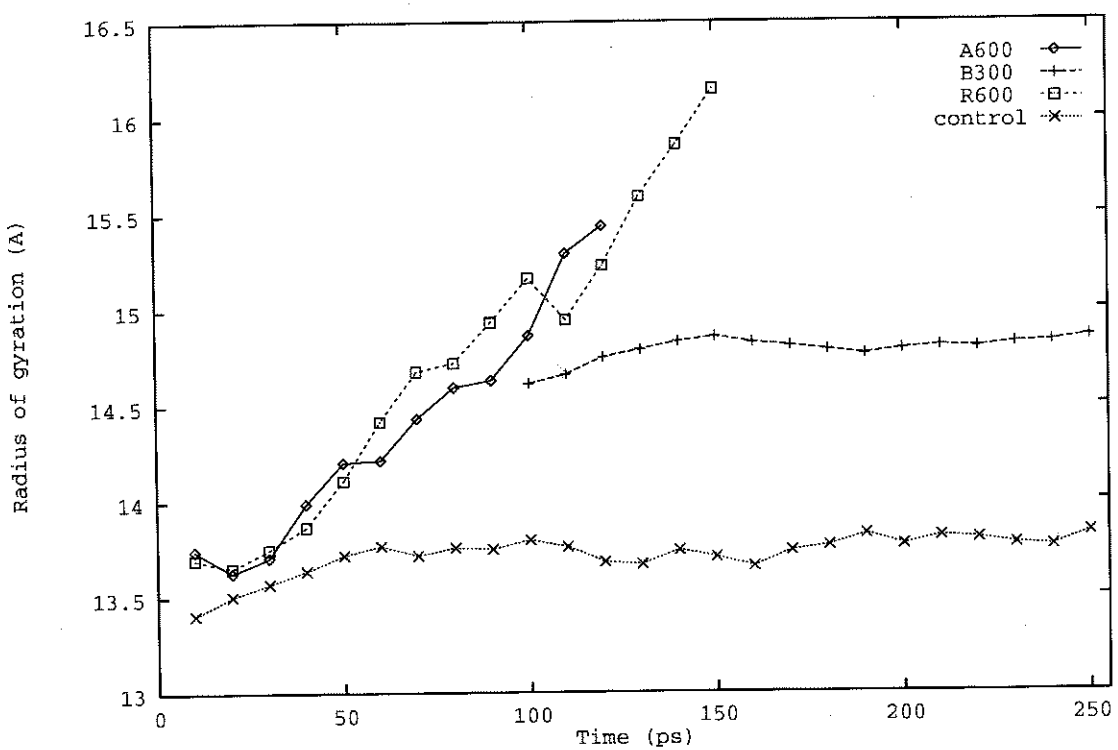

Figure 7-2a. Radius of gyration as a function of simulation time averaged over 10 ps intervals: A600 (*solid line*), B300 (*dashed line*), R600 (*broken line*), control run at 300 K (*dotted line*).

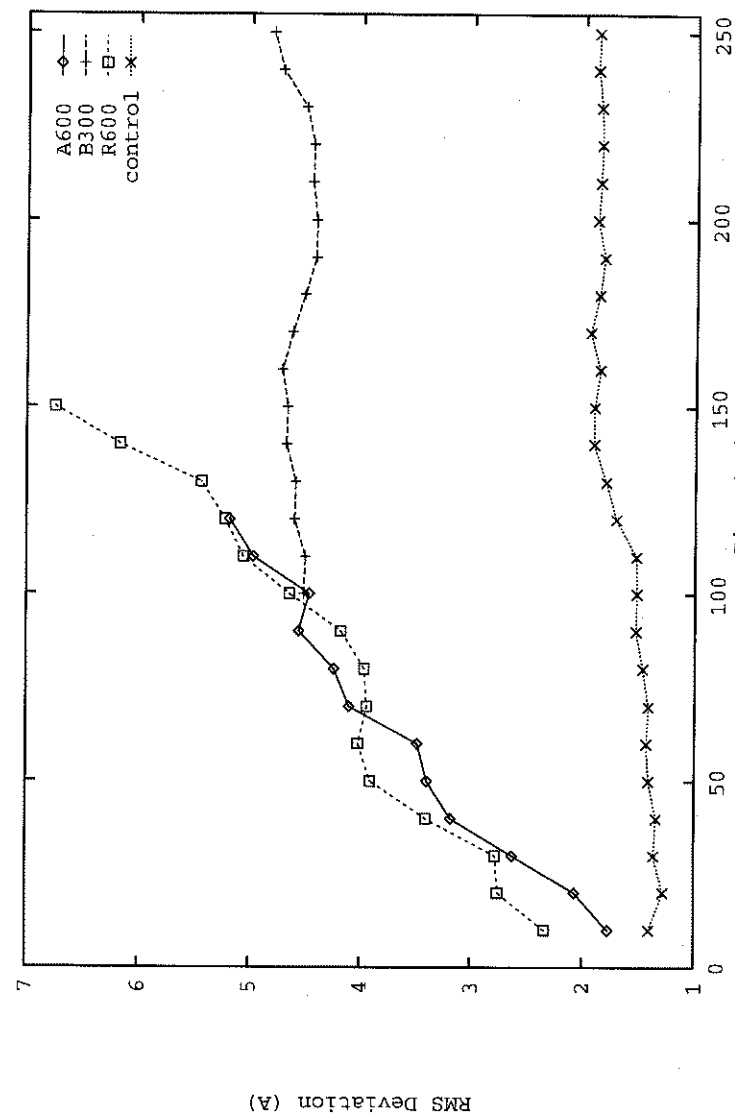

Figure 7-2b. Heavy atom root mean square deviation from the x-ray structure as a function of simulation time averaged over 10 ps intervals: A600 (*solid line*), B300 (*dashed line*), R600 (*broken line*), control run at 300 K (*dotted line*).

structures was 14.2 Å. This intermediate was characterized by a structure similar to the native one and a partially solvated $core_2$ (solvation of the cores is discussed in the next section). In A600 and R600, another intermediate state was formed after about 80 and 70 ps, respectively. It was characterized by a relatively small distortion in the secondary structural elements, a fully solvated $core_2$, and partially solvated $core_1$ and $core_3$ (see Figure 7-1 for notation). During the B300 simulation (started after 90 ps of A600 dynamics), R_g remained approximately constant at an average value of 14.8 Å; assuming a spherical geometry, the volume increase from the minimized x-ray structure amounts to 29%. Fersht and co-workers have shown that there exists an intermediate on both the folding and unfolding pathway of barnase (Matouschek et al., 1992a). It has some of the properties of B300 (e.g., unfolded N-terminus, $loop_1$, $loop_2$, and $loop_4$, distorted secondary structure elements, solvated $core_2$, and weakened hydrophobic interactions in $core_1$ and $core_3$) but no R_g measurement is available. Measurements of the α-lactalbumin molten globule by ultracentrifugation (Kronman et al., 1967) and by quasielastic light scattering (Gast et al., 1986) show a volume increase of about 30%.

In the high-temperature simulations, there were many structural changes (most of which are common to A600 and R600), but the overall fold was conserved (Figures 7-3a and 7-3b), as it appears to be in the experimentally observed intermediate (Matouschek et al., 1992a). In A600, the N-terminus, $loop_1$, and $loop_2$ began to unfold during the first 30 ps. This was followed by partial denaturation of the hydrophobic cores; $core_2$ denatured relatively rapidly, followed by $core_1$, $core_3$, and $loop_3$. The same sequential order for the hydrophobic core solvation was seen in R600. Solvation of hydrophobic $core_1$ was coupled with a large distortion of both $helix_1$ and the edge strands of the β-sheet. $Helix_1$ and $helix_2$ lost about 50% of the native α-helical hydrogen bonds in the A600 and R600 simulations; $helix_3$ unfolded after about 20 ps. In the β-sheet, about half of the native interstrand hydrogen bonds had disappeared after 100 ps in both trajectories; in R600 the β-sheet was almost fully solvated in the last 30 ps (120 to 150 ps), while in the last 20 ps of A600 (100 to 120 ps), there were about 50% of the native interstrand hydrogen bonds. In the B300 simulation, the percentages of native interstrand hydrogen bonds at the edges of the sheet (strands 1–2 and 4–5), and at the center of the sheet (strands 2–3 and 3–4), stabilized at about 50% and 70%, respectively. Interactions with water molecules replaced many of the helical and β-sheet hydrogen bonds (see Section 5).

The accessibility of $core_1$ to water was coupled with the relative motion of $helix_1$ and the β-sheet (see Figures 7-3a and 7-3b); in A600 they began to move apart at about 30 ps and their separation was continuous during the

30–75 ps period; between 75 and 100 ps there was a small closing movement in accord with a decrease of water in the core, followed by expansion for the remainder of the simulation (see $strand_3$ residues 85–90 and $helix_1$ residues 6–18 in Figure 7-3a). During the R600 simulation an essentially continuous opening motion was seen (Figure 7-3b). In B300, the relative orientation of $helix_1$ and the β-sheet was stable, the steady state of partial solvation of $core_1$ was coupled with a nearly constant number of hydrogen bonds in $helix_1$ and the central part of the β-sheet (strands 2–3 and 3–4). The structure at the end of B300 was similar to the conformation of A600 after 90 ps. The all-atom (backbone, side chain) RMSD after optimal superposition between these two structures was 2.32 Å (1.89 Å, 2.68 Å). This is to be compared with the deviations from the native structure; they were 4.55 Å (4.15 Å, 4.91 Å) after 90 ps of A600 and 4.94 Å (4.25 Å, 5.36 Å) at the end of B300.

5. Solvent Role in the Unfolding Transition

The results from the simulations described in the previous section provide considerable insights into the nature of the unfolding transition of globular proteins. An important point is that water penetration plays an essential role in the disruption of secondary structure and of hydrophobic clusters. This is in disagreement with theoretical arguments, which have suggested that water is not involved in the transition from the native to the molten globule state (Shakhnovich and Finkelstein, 1989). It also disagrees with the results of Daggett and Levitt (1992b), who used a low water density in their high-temperature simulations. However, it is in accord with measurements of the partial specific volume of the molten globule and the heat capacity change in the transition from the native to the molten globule state (Ptitsyn, 1992). It has been concluded that there are several hundred water molecules in the interior of the molten globule state of α-lactalbumin. In what follows, we summarize the available results. Much of the information (all of it for hydrophobic cores and β-sheets) comes from the barnase simulation outlined above.

The picture that emerges is that of a stepwise but partly cooperative unfolding phenomenon from the native to a compact globule state. By this we mean that the initial break-up of secondary structure is simultaneous with and coupled to the denaturation of the hydrophobic cores that are involved in the stabilization of the tertiary structure. During the transition there is a limited expansion of the protein and the overall fold is preserved,

Figure 7-3a. Stereo drawing of the barnase C_α atoms to illustrate the relative helix$_1$/β-sheet motion during A600. 1 ps (*thin line and labels*), 70 ps (*thicker line*), 90 ps (*medium line*), last ps (*thick lines and labels*). (In all stereofigures, the view is cross-eyed for the left pair of images and wall-eyed for the right pair).

Figure 7-3b. Stereo drawing of the barnase C_α atoms to illustrate the relative helix$_1$/β-sheet motion during R600. 1 ps (*thin line and labels*), 70 ps (*medium line*), 90 ps (*thicker line*), last ps (*thick lines and labels*). (In all stereofigures, the view is cross-eyed for the left pair of images and wall-eyed for the right pair).

though the loops and secondary structural elements are present in a looser and significantly distorted form. In all of this, entropic effects due to the elevated temperature may play a role, i.e., at high temperature a flexible polypeptide chain with its polar groups participating in hydrogen bonds to water has a lower free energy than a more rigid helix with enthalpically stronger intrahelical hydrogen bonds. This effect may be amplified by the use of the room-temperature water density.

For α-helices and β-sheets the denaturation process shows similar characteristics. The essential factor is that water molecules compete with the existing hydrogen bonds of the secondary structure. Simulations of isolated helices (see Section 2 on fragment studies) and helices as parts of proteins (apomyoglobin and barnase) have shown that the water molecules attack primarily the more accessible carbonyl groups. A water molecule that has formed a hydrogen bond to a carbonyl oxygen can then insert into the helix and replace the helical C=O···NH hydrogen bond. Of course not every water molecule that makes a hydrogen bond to a carbonyl group inserts directly. Moreover, some helical hydrogen bonds are broken without simultaneous water insertion. For β-sheets, the process involves waters that also interact primarily with carbonyl oxygens but now the waters insert to replace interstrand hydrogen bonds.

As a detailed example, we consider the A600 simulation results for helix_1 of barnase. The C-terminal hydrogen bond was lost after about 30 ps and did not reform. The three N-terminal hydrogen bonds were broken after about 50 ps; they reformed in the last part of the simulation. Reformation of the N-terminal hydrogen bond in helix_1 (at 90 ps and 110 ps) occurred through a 3_{10} hydrogen bond; this phenomenon was observed during molecular dynamics simulations of an α-helical analogue of the ribonuclease A S-peptide (Tirado-Rives and Jorgensen, 1991) and is in accord with the x-ray studies of Sundaralingam and Sekharudu (1989). The helical hydrogen bonds between 15NH···CO11 and 16NH···CO12 were broken after about 70 ps; they were converted into a 3_{10} form in the last 20 ps of A600. A water molecule accepted from the NH group of residue 15 and donated to the CO of residue 11 during the entire B300 simulation. The NH group of residues 12 and 13 converted from the 3_{10}- to the α-type in the first half of the B300 simulation, while the NH of residue 14 underwent the opposite conversion after about 80 ps. The 11NH···CO7 and 17NH···CO13 helical hydrogen bonds and the 16NH···CO13 forming a 3_{10} hydrogen bond were stable during B300.

The barnase β-sheet is composed of five antiparallel β-strands in consecutive order (Figure 7-1) and provides a clear example of the role of water in its high-temperature denaturation. In its native state, the sheet has

a regular structure, apart from a β-bulge in strand$_1$ between residue 53 and 54; it has a twist of about 90 degrees (from strand$_1$ to strand$_5$; Figure 7-1). Solvent insertion began at the edges of the β-sheet (see Figures 7-4a and 7-4b). The C-terminal part of strand$_1$ (near the β-bulge) was rapidly solvated in both A600 and R600 (Figure 7-4a); this is likely to originate from the instability of this irregular structure. At 70 ps several solvent molecules were located between strands and participate in hydrogen bonds with the main-chain NH and CO groups as donor, acceptor or both. In R600, hydrogen bonds between strands 1 and 2 came apart at about 70 ps (in A600 they separated after about 100 ps). This was followed by separation (accompanied by water insertion) of strands 3 and 4 at 120 ps, strands 2 and 3 at 130 ps, and strands 4 and 5 at 150 ps; in A600, except for strands 1 and 2, one or more interstrand hydrogen bonds remained until the end of the simulation.

To focus on the nature of the motions of water molecules interacting with the β-sheet during R600, six waters that spent more than 30 ps within 3 Å of any main-chain atom of the β-sheet are labeled (from A to F) in Figures 7-4a and 7-4b. From 60 to 110 ps, water A belonged to a cluster of solvent molecules interacting with the β-bulge part of strand$_1$. At 70 ps (bottom stereopicture of Figure 7-4a), it donated a bifurcated hydrogen bond to the CO group of strand$_1$ residues 52 and 53; at 90 ps (top stereopicture of Figure 7-4b), it acted as acceptor for another water, which donated a bifurcated hydrogen bond to the same CO groups. From 50 to 110 ps, water B was inserted between strands 1 and 2; it donated to the CO group of residue 73 in strand$_2$. At 110 ps, it also accepted from the NH groups of residue 53 (strand$_1$) and residue 73 (strand$_2$), acting as an interstrand bridging element. It then moved to the C-terminal part of strand$_3$, where it donated to the CO group of residue 90 at 150 ps. Water C interacted with the N-terminal part of strand$_1$ from 50 to 100 ps; it then moved away from the β-sheet and came closer again at 130 ps for about 5 ps. Water D solvated the N-terminal part of strand$_4$ during the 90–130 ps period. Waters E and F inserted between the C-terminal part of strand$_4$ and the N-terminal part of strand$_5$ during the 70–130 ps period.

The role of water in disruption of hydrophobic cores is more complex. Water molecules penetrate in a number of ways, all of which tend to involve the preservation of their hydrogen bonds. This may be achieved by waters hydrogen bonding to polar groups (e.g., OH of Tyr residues) and to charged groups that enter the hydrophobic core. Alternatively, the water molecules form hydrogen-bonded chains or clusters around hydrophobic residues. Also, some waters may also be involved in breaking up the secondary structural elements that are held together by the hydrophobic cores. Any

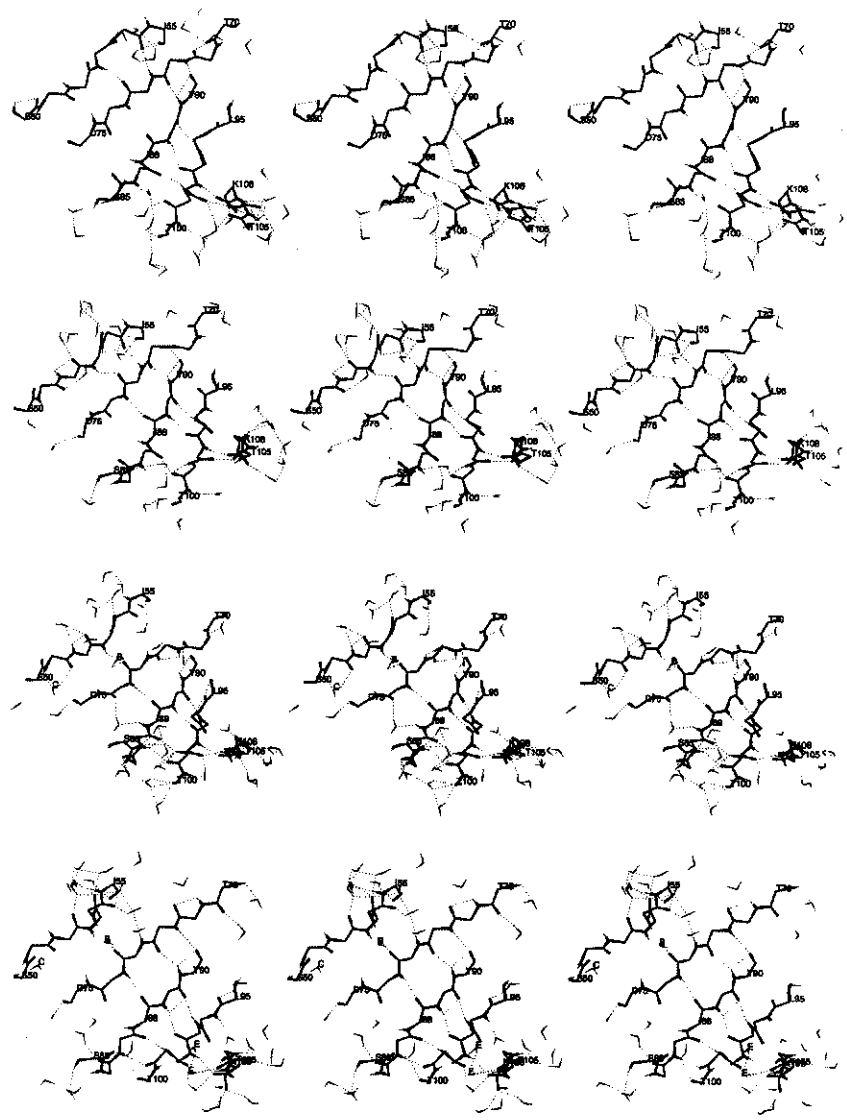

Figure 7-4a. Stereoview of water penetration into the β-sheet during R600. Main-chain N and O atoms are thick, hydrogens are thin, and hydrogen bonds are dotted; water molecules within 3 Å of any main-chain atom of the β-sheet are shown. Water molecules labeled from A to F are discussed in the text. From top to bottom: 10, 30, 50, 70 ps.

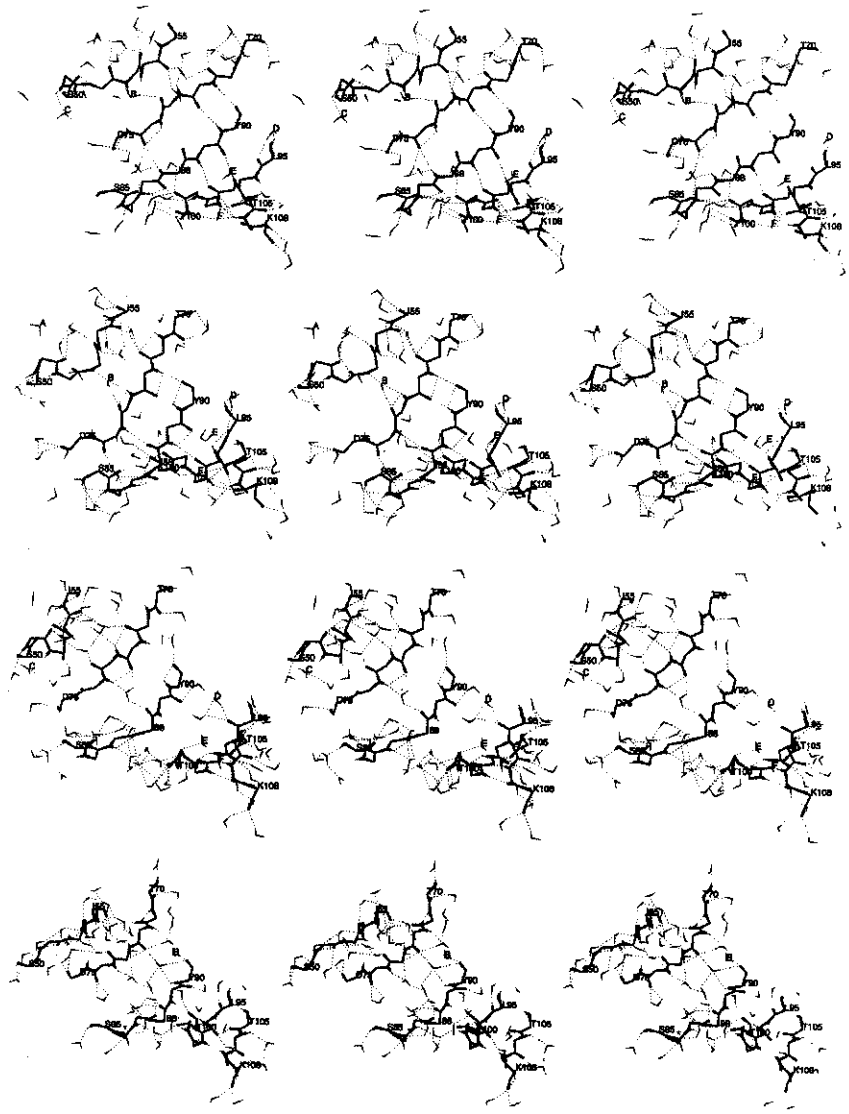

Figure 7-4b. Stereoview of water penetration into the β-sheet during R600. Main-chain N and O atoms are thick, hydrogens are thin, and hydrogen bonds are dotted; water molecules within 3 Å of any main-chain atom of the β-sheet are shown. Water molecules labeled from A to F are discussed in the text. From top to bottom: 90, 110, 130, 150 ps.

such water penetration is coupled to opening up of the hydrophobic core by relative motion of the secondary structural elements. In barnase, for example, the solvent-exposed surface area of the core increases as more water molecules penetrate into the core. Figure 7-5a–c shows the time dependence of the solvent-accessible surface area of the side chains of the principal hydrophobic core of barnase. Increase in $core_1$ surface area (Figure 7-5a) and water influx (not shown) are nearly simultaneous and begin at about 35 ps in A600 and 25 ps in R600. A maximum of 16 water molecules has penetrated at 82 and 84 ps in A600 (average of 12 for the 67–86 ps period); most of them are expelled again so that at 90 and 98 ps only 7 water molecules are located within 7 Å of its center (average of 10 for the 89–98 ps period). In the last 10 ps of A600 (from 111 to 120 ps) the number of water molecules in $core_1$ increases again to an average of 17. The water molecules penetrate mainly in the two directions parallel to the $helix_1$ axis. Furthermore, hydrophobic residues at the edges of the core (Leu 20, Tyr 24, Tyr 90, Trp 94, and Ile 109) were solvated first, whereas residues in the center of the core (Leu 14, Ala 74, Ile 88, and Ile 96) were more compact, in agreement with protein engineering results (Serrano et al., 1992c). The solvation of the main hydrophobic core was coupled to both the relative motion of $helix_1$ and the β-sheet and the distortion of the secondary structure elements. During the B300 simulation, the solvent accessible surface area of $core_1$ was nearly constant (average of 431 $Å^2$; 14 water molecules within 7 Å). This steady state of solvation of the core was correlated with the nearly constant number of hydrogen bonds in the secondary structure elements.

Solvation of $core_2$ began after about 10 ps in A600 and R600 (Figure 7-5b). It was coupled to the rearrangement of the subdomain consisting of $helix_2$, $loop_2$, and $helix_3$ which unfolded along with the β-turn formed by residues 46–49 (Figures 7-3a and 7-3b). Apart from a small closing movement of $core_2$ during the 20–50 ps period in R600, the number of water molecules increased continuously during the first 100 ps; a plateau value was reached after about 100 ps. $Core_3$ was more stable than either $core_2$ or $core_1$. It began to unfold after about 60 ps and 70 ps in the R600 and A600 trajectories, respectively. The late unfolding of $core_3$ and $loop_3$, relative to $core_2$, is consistent with experimental results (Serrano et al., 1992c).

During the *in vacuo* simulation V600 the solvent-accessible surface area of the cores showed an oscillating behavior around average values of 461 $Å^2$, 390 $Å^2$, and 162 $Å^2$, for $core_1$, $core_2$, and $core_3$, respectively. In V800, opening of the cores began in the first 10 ps ($core_2$ and $core_1$) or 100 ps ($core_3$); the solvent-accessible surface area reached a plateau value

Figure 7-5a. Solvent-accessible surface area (Å^2) of nonpolar side-chain atoms of hydrophobic core$_1$ as a function of time (ps). The Lee and Richards (1971) algorithm (CHARMM implementation) and a probe sphere of 1.4 Å radius were utilized; A600 (*solid line*), B300 (*dashed line*), R600 (*broken line*), control run at 300 K (*dotted line*).

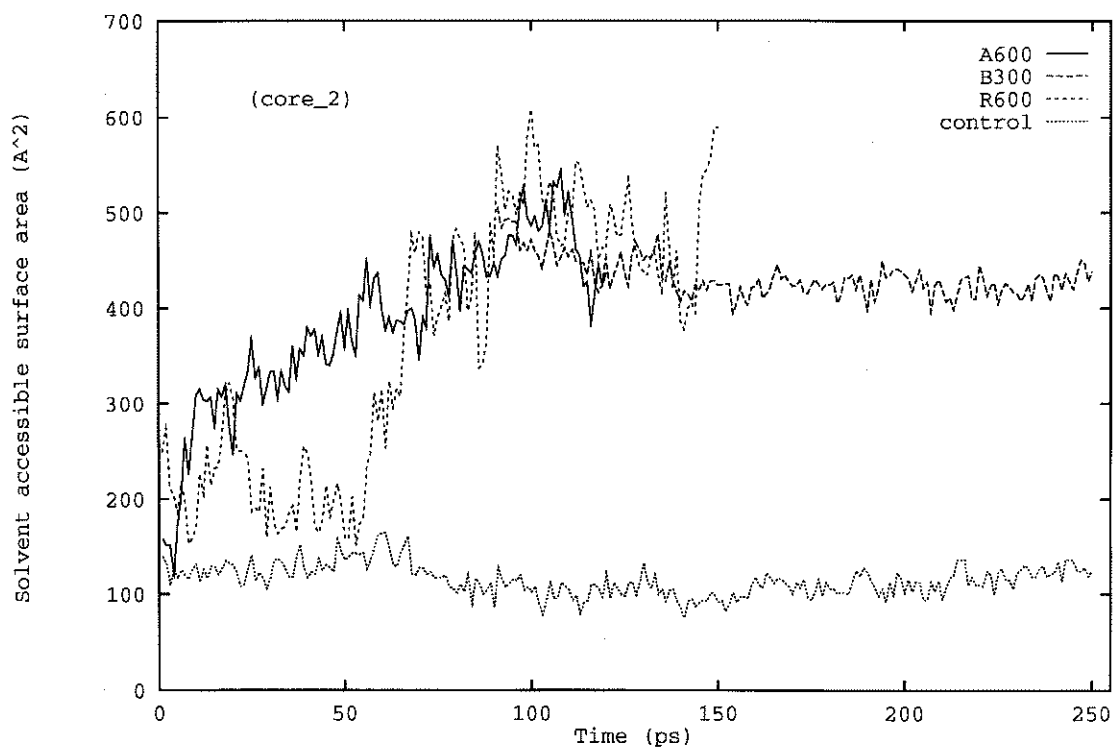

Figure 7-5b. Solvent-accessible surface area (Å^2) of nonpolar side-chain atoms of hydrophobic core$_2$ as a function of time (ps). The Lee and Richards (1971) algorithm (CHARMM implementation) and a probe sphere of 1.4 Å radius were utilized; A600 (*solid line*), B300 (*dashed line*), R600 (*broken line*), control run at 300 K (*dotted line*).

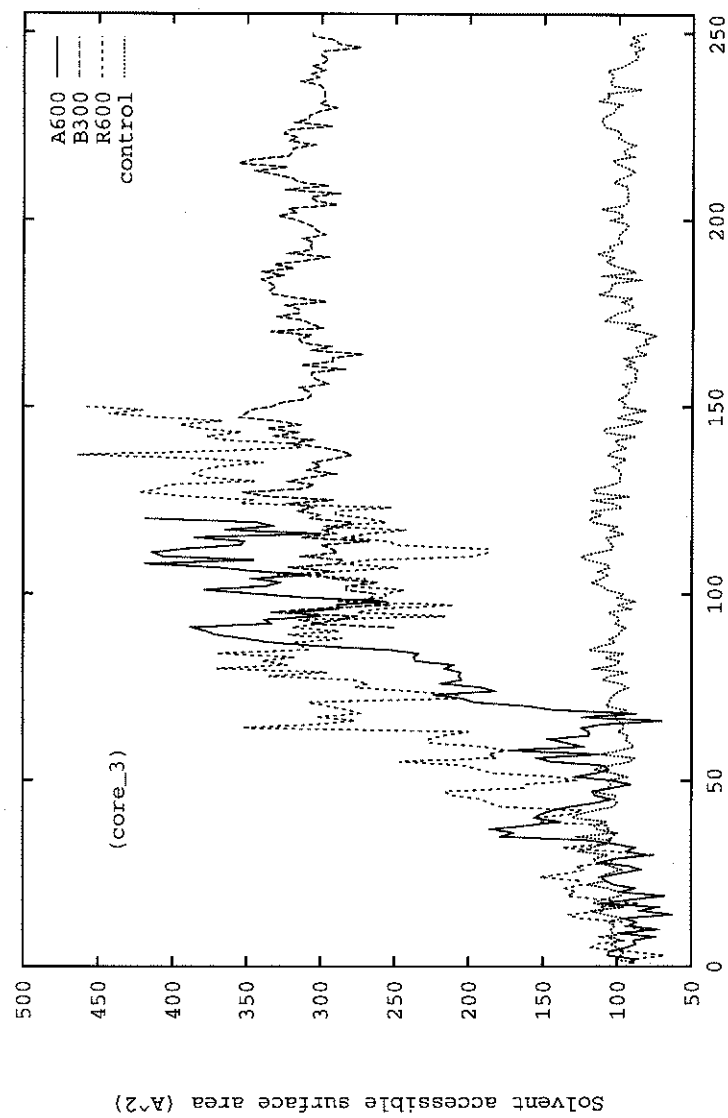

Figure 7-5c. Solvent-accessible surface area (Å^2) of nonpolar side-chain atoms of hydrophobic core$_3$ as a function of time (ps). The Lee and Richards (1971) algorithm (CHARMM implementation) and a probe sphere of 1.4 Å radius were utilized; A600 (*solid line*), B300 (*dashed line*), R600 (*broken line*), control run at 300 K (*dotted line*).

after 40 ps, 50 ps, and 150 ps, for $core_2$, $core_1$, and $core_3$, respectively; the sequential order was the same as in A600 and R600.

6. Conclusions

At the present stage of computational hardware, folding studies are limited to small fragments or to simplified models of proteins. The results that have become available are particularly useful as a source of parameters for phenomenological descriptions of protein folding, such as the diffusion–collision model. Much more can be done at an all-atom level of detail to simulate the dynamics of the early stages of unfolding from the native state to a compact organized globule. Here the results of a number of recent simulations focus on the important role in the denaturation process of explicit water molecules, rather than the general dielectric effect of aqueous solvent. Both for secondary structural elements (α-helices and β-sheets), and for hydrophobic cores, the entrance of water molecules that are hydrogen bonded to each other or to the protein polar groups are an essential part of the denaturation process. Some detailed examples, particularly from simulations of barnase, have been presented to illustrate the dynamic role of water molecules in protein unfolding.

Much remains to be done to obtain a full understanding of protein unfolding and to extend that knowledge to the folding process. All simulations of unfolding of entire proteins, in contrast to fragments, were performed at unrealistically high temperatures to increase the rate so that it falls within the nanosecond time scale of present day simulations. Most of these high-temperature simulations used explicit models of water molecules at a density corresponding to that appropriate for room temperature; it is likely that this also contributes to increasing the unfolding rate. These special conditions will have to be investigated further to determine whether significant artifacts are introduced. Also, alternative approaches to the denaturation process should be examined. Finally, the question of how best to use the information obtained from unfolding studies for analyzing the folding process has yet to be determined.

Acknowledgments. The work was supported in part by a grant from the National Institutes of Health and a gift from Molecular Simulations, Inc. A.C. was supported by the Schweizerischer Nationalfonds (Swiss National Science Foundation).

REFERENCES

Acharya KR, Straut DI, Walker NPC, Lewis M, Phillips DC (1989): Refined structure of baboon α-lactalbumin at 1.7 Å resolution. Comparison with c-type lysozyme. *J Mol Biol* 208:99–127

Anfinsen CB (1972): The formation and stabilization of protein structure. *Biochem J* 128:737–749

Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975): Dynamics of ligand binding to myoglobin. *Biochemistry* 14:5355–5373

Baudet S, Janin J (1991): Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. *J Mol Biol* 219:123–132

Baum J, Dobson CM, Evans PA, Hanly C (1989): Characterization of a partly folded protein by nmr methods: Studies on the molten globule state of guinea pig α-lactalbumin. *Biochemistry* 28:7–13

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984): Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983): CHARMM : A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* 4:187–217

Brooks CL III, Karplus M (1983): Deformable stochastic boundaries in molecular dynamics. *J Chem Phys* 79:6312–6325

Brooks CL III, Karplus M (1989): Solvent effects on protein motion and protein effects on solvent motion. *J Mol Biol* 208:159–181

Brooks CL III (1992): Characterization of "native" apomyoglobin by molecular dynamics simulation. *J Mol Biol* 227:375–380

Brünger A, Clore GM, Gronenborn AM, Karplus M (1986): Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to Crambin. *Proc Natl Acad Sci USA* 83:3801–3805

Brünger A, Karplus M (1991): Molecular dynamics simulations with experimental restraints. *Acc Chem Res* 24:54–61

Bycroft M, Matouschek A, Kellis JT, Serrano L, Fersht AR (1990): Detection and characterization of a folding intermediate in barnase by NMR. *Nature* 346:488–490

Bycroft M, Ludvigsen S, Fersht AR, Poulsen FM (1991): Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry* 30:8697–8701

Caflisch A, Niederer P, Anliker M (1992): Monte Carlo minimization with thermalization for global optimization of polypeptide conformations in Cartesian coordinate space. *Proteins: Structure, Function and Genetics* 14:102–109

Creighton TE (1988): Toward a better understanding of protein folding pathways. *Proc Natl Acad Sci USA* 85:5082–5086

Czerminski R, Elber R (1989): Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proc Natl Acad Sci USA* 86:6963–6967

Czerminski R, Elber R (1990): Reaction path study of conformational transitions in flexible systems: Applications to peptides. *J Chem Phys* 92:5580–5601

Daggett V, Levitt M (1992a): Molecular dynamics simulations of helix denaturation. *J Mol Biol* 223:1121–1138

Daggett V, Levitt M (1992b): A model of the molten globule state from molecular dynamics simulations. *Proc Natl Acad Sci USA* 89:5142–5146

Deng Y, Glimm J, Sharp DH (1990): Los Alamos Report, pages LA-UR-90–4340

DiCapua FM, Swaminathan S, Beveridge DL (1990): Theoretical evidence for destabilization of an α-helix by water insertation: Molecular dynamics of hydrated decaalanine. *J Am Chem Soc* 112:6768–6771

Dobson CM, Hanley C, Radford SE, Baum JA, Evans PA (1991): In *Conformations and Forces in Protein Folding*. Nall BT, Dill KA, eds. pages 175–181

Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov SY, Ptitsyn OB (1981): A-Lactalbumin: Compact state with fluctuating tertiary structure. *FEBS Letters* 136:311–315

Elber R, Karplus M (1987): Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science* 235:318–321

Fan P, Kominos D, Kitchen DB, Levy RM, Baum J (1991): Stabilization of α-helical secondary structure during high-temperature molecular-dynamics simulations of α-lactalbumin. *Chemical Physics* 158:295–301

Fersht AR, Matouschek A, Serrano L (1992a): The folding of an enzyme. I: Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782

Fersht AR, Matouschek A, Sancho J, Serrano L, Vuilleumier S (1992b): Pathway of protein folding. *Faraday Discuss* 93:183–193

Fersht AR (1993): Protein folding and stability: The pathway of folding of barnase. *FEBS letters* 325:5–16

Gast K, Zirwer D, Welfle H, Bychkova VE, Ptitsyn OB (1986): Quasielastic light scattering from human α-lactalbumin: Comparison of molecular dimensions in native and 'molten globule' state. *Int J Biol Macromol* 8:231–236

Gething M-J, Sambrook J (1992): Protein folding in the cell. *Nature* 355:33–45

Gilmanshin RI, Ptitsyn OB (1987): An early intermediate of refolding α-lactalbumin forms within 20 ms. *FEBS Letters* 223:327–329

Goldberg ME, Semisotnov GV, Friguet B, Kuwajima K, Ptitsyn OB, Sugai S (1990): An early immunoreactive folding intermediate of the tryptophan synthase β_2 subunit is a 'molten globule.' *FEBS Letters* 263:51–56

Haas E, Katchalski-Katzir E, Steinberg IZ (1978): Brownian motion of the ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers* 17:11–31

Hagler AT, Honig B (1978): On the formation of protein tertiary structure on a computer. *Proc Natl Acad Sci USA* 75:554–558

Harrison S, Durbin R (1985): Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA* 82:4028–4030

Hermans J, Anderson A, Yun RH (1992): Differential helix propensity of small apolar sidechains studied by molecular dynamics simulations. *Biochemistry* 31:5646–5653

Houghson FM, Wright PE, Baldwin RL (1990): Structural characterization of a partly folded apolyoglobin intermediate. *Science* 249:1544–1548

Houghson FM, Barrick D, Baldwin RL (1991): Probing the stability of partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry* 30:4113–4118

Janin J, Wodak S (1983): Structural domains in proteins and their role in the dynamics of protein function. *Prog Biophys Mol Biol* 42:21–78

Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML (1983): Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935

Karle IL, Flippen-Anderson JL, Uma K, Balaram P (1990): Apolar peptide models for conformational heterogeneity, hydration, and packing of polypeptide helices: Crystal structure of hepta- and octapeptides containing α-aminoisobutyric acid. *Proteins: Structure, Function and Genetics* 7:62–73

Karplus M, Weaver DL (1976): Protein folding dynamics. *Nature* 260:404–406

Karplus M, Weaver DL (1979): Diffusion-collision model for protein folding. *Biopolymers* 18:1421–1437

Karplus M, Shakhnovich E (1992): Protein folding: Theoretical studies of thermodynamics and dynamics. In *Protein Folding*. Creighton TE, ed. New York: WH Freeman

Kellis JT Jr, Nyber K, Fersht AR (1989): Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28:4914–4922

Kim PS, Baldwin RL (1990): Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry* 59:631–660

Kraulis P (1991): Molscript, a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950

Kronman MJ, Holmes LG, Robbins FM (1967): Inter and intramolecular interactions of α-lactalbumin. VIII The alkaline conformational change. *Biochim Biophys Acta* 133:46–55

Kuwajima K, Yamaya H, Miwa S, Sugai S, Nagamura T (1987): Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. *FEBS Letters* 221:115–118

Kuwajima K (1989): The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Structure, Function and Genetics* 6:87–103

Lazaridis T, Tobias DJ, Brooks CL III, Paulaitis ME (1991): Reaction path and free energy profiles for conformational transitions: An internal coordinate approach. *J Chem Phys* 95:7612–7625

Lee S, Karplus M, Bashford D, Weaver DL (1987): Brownian dynamics simulation of protein folding: A study of the diffusion-collision model. *Biopolymers* 26:481–506

Lee B, Richards FM (1971): The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400

Levinthal C (1969): In *Mössbauer Spectroscopy in Biological Systems*, DeGennes P et al., eds. Urbana, IL: University of Illinois Press. Proceedings of a meeting held at Allerton House, Monticello, IL

Levitt M, Warshel A (1975): Computer simulation of protein folding. *Nature* 253: 694–698

Levitt M (1976): A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* 104:59–107

Levitt M (1983): Protein folding by restrained energy minimization and molecular dynamics. *J Mol Biol* 170:723–764

Mark AE, van Gunsteren WF (1992): Simulation of the thermal denaturation of hen egg white lysozyme: Trapping the molten globule state. *Biochemistry* 31:7745–7748

Matouschek A, Kellis JT Jr, Serrano L, Fersht AR (1989): Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340:122–126

Matouschek A, Kellis JT Jr, Serrano L, Bycroft M, Fersht AR (1990): Transient folding intermediates characterized by protein engineering. *Nature* 346:440–445

Matouschek A, Serrano L, Fersht AR (1992a): The folding of an enzyme. IV: Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224:819–835

Matouschek A, Serrano L, Meiering EM, Bycroft M, Fersht AR (1992b): The folding of an enzyme. VH/H exchange—nuclear magnetic resonance studies on the folding pathway of barnase: Complementarity to and agreement with protein engineering studies. *J Mol Biol* 224:837–845

Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A (1982): Molecular structure of a new family of ribonuclease. *Nature* 297:162–164

McCammon JA, Gelin BR, Karplus M, Wolynes PG (1976): The hinge-bending motion in lysozyme. *Nature* 262:325–326

McCammon JA, Northrup SH, Karplus M, Levy RM (1980): Helix-coil transitions in a simple polypeptide model. *Biopolymers* 19:2033–2045

Meiering EM, Serrano L, Fersht AR (1992): Effect of active site residues in barnase on activity and stability. *J Mol Biol* 225:585–589

Miranker A, Radford SE, Karplus M, Dobson CM (1991): Demonstration by NMR of folding domains in lysozyme. *Nature* 349:633–636

Nemethy G, Scheraga HA (1977): Protein folding. *Quart Rev Biophys* 10:239–352

Noguti T, Gō N (1989a): Structural basis of hierarchical multiple substates of a protein. I: Introduction. *Proteins: Structure, Function and Genetics* 5:97–103

Noguti T, Gō N (1989b): Structural basis of hierarchical multiple substates of a protein. II: Monte Carlo simulation of native thermal fluctuations and energy minimization. *Proteins: Structure, Function and Genetics* 5:104–112

Noguti T, Gō N (1989c): Structural basis of hierarchical multiple substates of a protein. III: Side chain and main chain local conformations. *Proteins: Structure, Function and Genetics* 5:113–124

Noguti T, Gō N (1989d): Structural basis of hierarchical multiple substates of a protein. IV: Rearrangements in atom packing and local deformations. *Proteins: Structure, Function and Genetics* 5:125–131

Noguti T, Gō N (1989e): Structural basis of hierarchical multiple substates of a protein. V: Nonlocal deformations. *Proteins: Structure, Function and Genetics* 5:132–138

Oas TG, Kim PS (1988): A peptide model of a protein folding intermediate. *Nature* 336:42–48

Ohgushi M, Wada A (1983): 'Molten globul state': A compact form of globular proteins with mobile side-chains. *FEBS Letters* 614:21–24

Pear MR, Northrup SH, McCammon JA, Karplus M, Levy RM (1981): Correlated helix–coil transitions in polypeptides. *Biopolymers* 20:629–632

Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI (1990): Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters* 262:20–24

Ptitsyn OB (1992): The molten globule state. In *Protein Folding*, Creighton TE, ed. New York: WH Freeman, pages 243–300

Serrano L, Kellis JT Jr, Cann P, Matouschek A, Fersht AR (1992a): The folding of an enzyme. II Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* 224:783–804

Serrano L, Matouschek A, Fersht AR (1992b): The folding of an enzyme. III: Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224:805–818

Serrano L, Matouschek A, Fersht AR (1992c): The folding of an enzyme. VI: The folding pathway of barnase: Comparison with theoretical models. *J Mol Biol* 224:847–859

Shakhnovich EI, Finkelstein AV (1989): Theory of cooperative transitions in protein molecules. I. Why denaturation of a globular protein is a first order phase transition. *Biopolymers* 28:1667–1680

Shakhnovich EI, Gutin A (1990): Enumeration of all compact conformations of copolymers with random sequence of links. *J Chem Phys* 93:5967–5971

Shakhnovich EI, Farztdinov G, Gutin A, Karplus M (1991): Protein folding bottlenecks: A lattice Monte Carlo simulation. *Physical Review Letters* 67:1665–1668

Soman KU, Karimi A, Case DA (1991): Unfolding of an α-helix in water. *Biopolymers* 31:1351–1361

Sundaralingam M, Sekharudu YC (1989): Water-inserted α-helical segments implicate reverse turns as folding intermediates. *Science* 244:1333–1337

Tirado-Rives J, Jorgensen WL (1991): Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. *Biochemistry* 30:3864–3871

Tirado-Rives J, Jorgensen WL (1993): Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* 32:4175–4184

Tobias DJ, Sneddon SF, Brooks CL III (1990): Reverse turns in blocked dipeptides are intrinsically unstable in water. *J Mol Biol* 216:783–796

Yapa K, Weaver DL, Karplus M (1992): β-sheet coil transitions in a simple polypeptide model. *Proteins: Structure, Function and Genetics* 12:237–265

Yun RH, Anderson A, Hermans A (1991): Proline in α-helix: Stability and conformation studied by dynamics simulation. *Proteins: Structure, Function and Genetics* 10:219–228

Zwanzig R, Szabo A, Bagchi B (1992): Levinthal's paradox. *Proc Natl Acad Sci USA* 89:20–22