

## Stabilization of $\alpha$ -helical secondary structure during high-temperature molecular-dynamics simulations of $\alpha$ -lactalbumin

Pei Fan, Dorothea Kominos, Douglas B. Kitchen, Ronald M. Levy and Jean Baum

*Department of Chemistry, Rutgers University, Piscataway, NJ 08855, USA*

Received 8 July 1991

Molecular-dynamics simulations of the partial unfolding of  $\alpha$ -lactalbumin are described. The effects of introducing various intramolecular constraints on the unfolding are compared. In one set of simulations, intrahelical constraints designed to force the preservation of the helical secondary structure were imposed. In another set of simulations, the hydrophobic core was loosely constrained by introducing distance constraints between side chains of residues in the hydrophobic core of the protein. Our primary result, which is the subject of this communication, is the observation that the introduction of loose geometric constraints within the hydrophobic core of the protein stabilizes the  $\alpha$ -helical secondary structure against thermal disruption at high temperature.

### 1. Introduction

One of the main difficulties in studying protein folding is the transient nature of the folding intermediates. Proteins have been identified, which under certain conditions of pH, temperature, or solvent composition, exist in a stable state which is intermediate between a folded and unfolded state [1]. These intermediate states, called “molten globule” states [2], possess a considerable amount of secondary structure, appear to be compact, but have no defined tertiary structure [3]. Studies of molten globule proteins, such as  $\alpha$ -lactalbumin, cytochrome c and apomyoglobin are important in understanding the subset of interactions that stabilize native proteins. In addition, it has been suggested that molten globule states may be found on the protein folding pathway [1,4]. Recently, NMR hydrogen-exchange trapping experiments have been used to study kinetic intermediates on the folding pathway [5–7].

$\alpha$ -lactalbumin has been studied extensively over the past 20 years. Early studies by Kuwajima and co-workers showed the presence of stable intermediates during the denaturant-induced, or acid-induced unfolding of  $\alpha$ -lactalbumin [1]. Ptitsyn and coworkers, based on a comparison of CD measurements at neu-

tral and low pH [3], suggested that  $\alpha$ -lactalbumin exhibited molten-globule properties. Recent NMR results on  $\alpha$ -lactalbumin have indicated that there are elements of native-like secondary structure in the molten-globule state and that there exist two stable helices in this partially structured state [8–10]. The amide protons that are protected from hydrogen exchange in the molten-globule state arise predominantly from two  $\alpha$ -helical regions of the native protein; these are the B and C helices which include residues 23–34 and 86–99, respectively, in the native state [11]. Similar results have been obtained for the molten-globule states of cytochrome c and apomyoglobin [12,13]. The experimental results have motivated us to use molecular-dynamics simulations to investigate the origin of the stability of the helices in the molten-globule state of  $\alpha$ -lactalbumin.

In an effort to understand the nature of the interactions in the molten globule state, a series of molecular dynamics calculations are being carried out on  $\alpha$ -lactalbumin under a variety of conditions. The basic question we asked in this work is whether the presence of a hydrophobic core of residues stabilizes the  $\alpha$ -helical secondary structure against thermal disruption at high temperature; and the converse question, does the presence of  $\alpha$ -helices stabilize the hy-

drophobic core? Several related questions arise including: to what extent does the protein unfold at elevated temperatures in the times accessible to molecular-dynamics simulations; how stable is the secondary structure and do tertiary hydrophobic interactions stabilize it? The strategy is to generate partially unfolded structures starting from the native state of the protein by high-temperature molecular dynamics. The partially unfolded structures are obtained by introducing various kinds of intramolecular constraints on the high-temperature dynamics of  $\alpha$ -lactalbumin; these calculations are compared with the results of simulations without any constraints.

In contrast to the very large number of detailed molecular-dynamics studies of the native state of proteins that have appeared [14,15], there have been few attempts to simulate protein folding or unfolding [16–19]. This is in part due to the fact that the protein dynamics time scales which can be simulated are on the order of tens to hundreds of picoseconds, whereas the time scales for complete unfolding at ambient temperatures are many orders of magnitude slower. In initial simulations of  $\alpha$ -lactalbumin, we have employed two simplifications of the model in order to speed-up both the unfolding process and the calculations – solvent is not explicitly included in the simulations described in this communication and they are performed at very high temperatures (500 and 1000 K for the work reported here; a control simulation at 298 K was also carried out). Our primary result is the observation that the introduction of loose geometric constraints within the hydrophobic core of the protein stabilizes the  $\alpha$ -helical secondary structure against thermal disruption at high temperature.

## 2. Methods

Molecular dynamics simulations of  $\alpha$ -lactalbumin were performed using the IMPACT program package [20] and the all-atom force field of Weiner et al. [21]. A 7.5 Å atom based cut-off was used to construct the nonbonded interaction list which was updated every 10 steps (5 fs). The rattle algorithm [22] was applied to all bonds in the protein, and the temperature of the system was controlled by coupling to a heat bath [23]. Since the NMR studies were performed on guinea-pig  $\alpha$ -lactalbumin, and X-ray coordinates are

not available for this protein, the coordinates were model built from the X-ray structure of the homologous baboon  $\alpha$ -lactalbumin [11]. The resultant structure was minimized using 4000 steps of conjugate gradient minimization. As part of the minimization procedure, the theoretical NMR ring-current shifts were calculated for the guinea-pig structures using the program VNMR [provided by J. Hoch] and then compared with the experimental NMR spectrum to optimize the orientation of the aromatic residues.

As part of our study of the molten-globule properties of  $\alpha$ -lactalbumin, we are carrying out some simulations of the protein without solvent and others including the solvent explicitly. The all-atom representation of guinea-pig  $\alpha$ -lactalbumin contains 1982 atoms. Four thousand eight hundred water molecules (14400 atoms) are needed to fully solvate the protein. A 1 ps trajectory (a 0.5 fs time step is used at high temperature) of the fully solvated protein takes  $\sim 19$  cpu hours on a Silicon Graphics 210 GTX as compared with  $\sim 2.7$  cpu hours per ps for the simulations without explicit solvent reported here. In order to partially account for solvent effects, a distance dependent dielectric function was used to mimic the electrostatic screening effects of water [24]. The screening function  $\epsilon$  employed for the Coulomb interaction between two atoms separated by a distance  $R$ , was  $\epsilon = 4R$ . In tests on the highly refined crystal structure of crambin, this screening function was found to reproduce best the implicit effects of solvent [25]. The relative speed of the vacuum simulations enabled us to generate and analyze several sets of simulations at different temperatures and with different constraint sets. We compare the results of simulations at 298, 500 and 1000 K.

In one set of simulations no constraints were imposed on the protein and molecular-dynamics simulations were done for 20 ps at 1000 K, 50 ps at 500 K and 50 ps at 298 K. In a second set of simulations, intrahelical constraints designed to force the preservation of the helical secondary structure were imposed. The intrahelical constraints were chosen between the  $\text{NH}(i)\text{--NH}(i+1)$ ,  $\text{O}(i)\text{--NH}(i+4)$  and  $\text{HA}(i)\text{--NH}(i+3)$  pairs of atoms within the B and C helices resulting in a total of 52 constraints in this set. The upper and lower bounds for all pairs of constrained atoms were chosen by adding or subtracting

0.2 Å from the measured distances in the native structure. A harmonic constraint function  $K(r-r_0)^2$  was applied with  $K=100$  kcal/(mol Å<sup>2</sup>) where  $r_0$  corresponds to the lower or upper bound. Simulations with this constraint set were carried out at 1000 K for 20 ps. In a third set of simulations, the hydrophobic core was loosely constrained by introducing twenty-six distance constraints between side chains of residues in the hydrophobic core (hydrophobic constraints). The hydrophobic box region of the protein, shown in fig. 1, is composed of Ile 95, Trp 104, Tyr 103, Phe 60; the B and C helices border on the hydrophobic box region. For each pair of hydrophobic residues that are within 4.5 Å of each other, a maximum of three distance constraints were imposed between them to try to maintain the hydrophobic interactions. Upper and lower bounds ranged from within 1 to 2 Å. Pairs of interactions were chosen, either between the two helices or between the residues on one helix and the hydrophobic core; no intrahelical side chain to side chain constraints were chosen. The hydrophobic residues in the helices are pointing towards each other or towards the hydrophobic core as shown in fig. 1, whereas the hydrophilic residues are pointing away from the core of the protein as well as away from the region between the

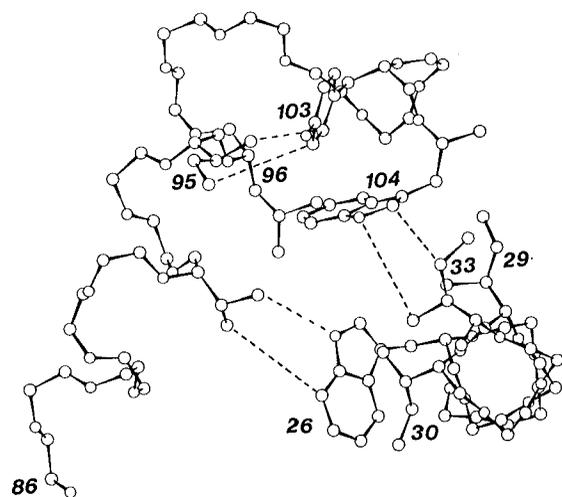


Fig. 1. The B and C helices and the hydrophobic core region of guinea-pig  $\alpha$ -lactalbumin. Some of the constrained side chains are explicitly shown. The broken lines represent examples of atoms connected by distance constraints imposed between pairs of residues in the hydrophobic constraint set.

two helices. Fig. 1 shows the constraints that were chosen for the pairs Trp 26 (B helix) and Val 92 (C helix), Ile 33 (B helix) and Trp 104, and Ile 95 (C helix) and Tyr 103. Two simulations were run with hydrophobic constraints: 20 ps at 1000 K, and 50 ps at 500 K.

### 3. Results and discussion

Fig. 2a shows a ribbon diagram of the native structure of guinea-pig  $\alpha$ -lactalbumin. The native structure is composed of 3  $\alpha$ -helices, helices A, B and C; a fourth helix, the D helix, found in baboon  $\alpha$ -lactalbumin is, according to model building, broken into two short 3–10 helices in guinea-pig  $\alpha$ -lactalbumin because of the substitution of a Pro at position 109.  $\alpha$ -lactalbumin also contains a short antiparallel  $\beta$ -sheet region [11]. The B and C helices, labelled in fig. 2a, are found by hydrogen-exchange NMR studies to be stable in the molten-globule state [8–10]. The structures obtained from the unconstrained dynamics and the dynamics with intrahelical and hydrophobic constraints after 10 ps at 1000 K are shown in figs. 2b–d.

In the first set of simulations, the native protein was heated without constraints; this serves as a control to determine the extent to which the native protein unfolds at high temperatures. From the sche-

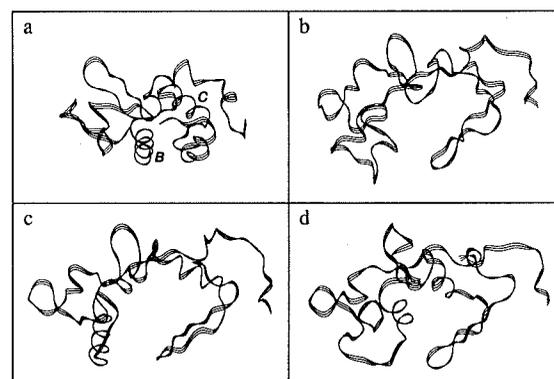


Fig. 2. Ribbon diagrams of  $\alpha$ -lactalbumin (a) in the native state, the B and C helices are shown; (b) after 10 ps of molecular-dynamics simulations at 1000 K without any constraint; (c) after 10 ps of molecular-dynamics simulations with the intrahelical constraints at 1000 K; (d) after 10 ps molecular-dynamics simulations with the hydrophobic constraints at 1000 K.

matic drawing shown in fig. 2b, it appears that the protein loses all of its secondary structure and has expanded in this unconstrained simulation, although the general globular shape is maintained. During this dynamics simulation, the C helix starts to fray at the amino terminus and the N-terminus makes a transition to a 3-10 helix before it falls apart. A similar transition through the 3-10 helix was observed by Tirado-Rives and Jorgensen in their simulations of the  $\alpha$ -helical ribonuclease S peptide [26]. Because NMR studies indicate the existence of stable regions of secondary structure in the B and C helices in the molten-globule state, we carried out two sets of simulations with interatomic constraints in order to examine the interrelationship between the integrity of the hydrophobic core and the integrity of the  $\alpha$ -helical secondary structure.

Fig. 2c shows the structure after 10 ps of molecular dynamics at 1000 K with intrahelical constraints, in which the B and C helices were prevented from breaking up. It can be seen that the two helices are moving apart from one another, leaving a cavity in the middle of the protein and thus exposing a large number of hydrophobic residues. This structure, although consistent with the experimental hydrogen exchange data for the molten-globule state, is not favourable from a thermodynamic point of view (see below). In the vacuum simulations we have attempted to model hydrophobic attractions by loosely constraining hydrophobic residues to remain near one another. Fig. 2d shows a ribbon diagram of the structure after 10 ps of molecular dynamics at 1000 K with hydrophobic constraints. The results of the simulation are quite interesting. Not only did the residues that were constrained stay together, but the B and C helices persist although they are not explicitly constrained as helices. These results are in sharp contrast to those obtained when the protein is heated with no constraints, as shown in fig. 2b.

To investigate the structures and the time course of the helical unfolding more closely, we selected a criterion to estimate fractional helicity which we monitored over the course of the simulations. A residue was considered to be helical if at least five consecutive  $N(i)-N(i+1)$  distances were greater than 1.8 Å and less than 3.6 Å [27]. If the  $N(i, i+1)$  distance criterion was not met in one position, but the  $\alpha N(i, i+3)$  and  $\alpha\beta(i, i+3)$  distances were less than

5 Å at that position, then the residue was also considered helical. This criterion worked best in describing what was structurally present by visual examination of the protein. Figs. 3a and b show the % helix content versus time for two sets of simulations. Fig. 3a indicates the % helicity of the B and C helices during the course of unconstrained molecular dynamics at 1000 K. Without constraints, the B and C helices fall apart rapidly as a function of time. At the end of 10 ps, both helices have been completely disrupted. In

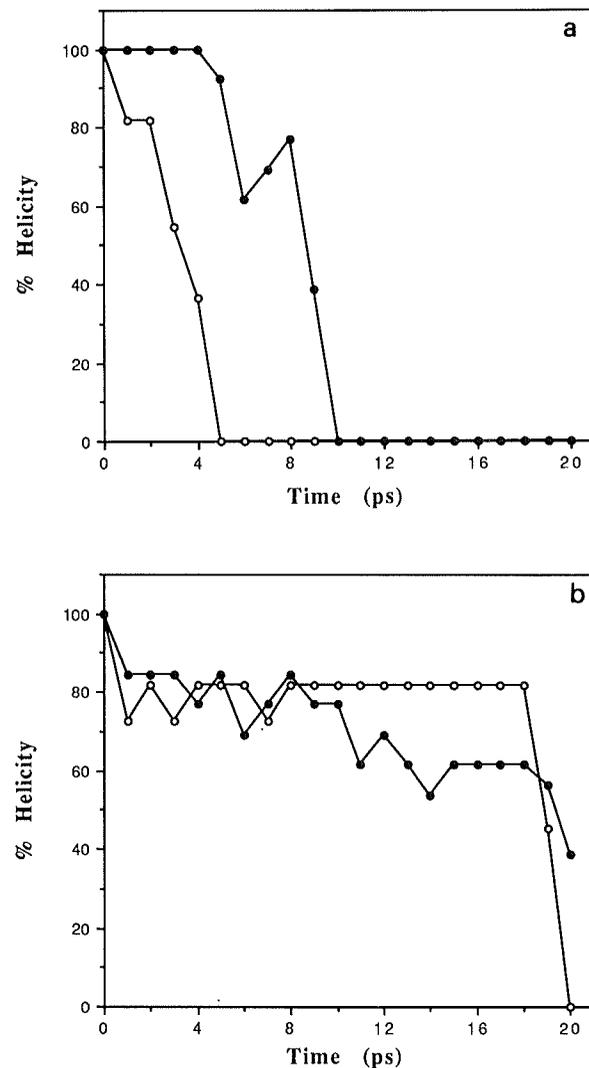


Fig. 3. % helicity of the (○) B and (●) C helices during the dynamics course at 1000 K, (a) with no constraint; (b) with hydrophobic constraints.

contrast, as shown in fig. 3b, with hydrophobic constraints after 10 ps of molecular dynamics at 1000 K, 82% of the residues in the B helix remain helical and 77% of the residues in the C helix remain helical. Part of the B and C helices remain even after 15 ps of dynamics. At 20 ps, one of the helices has unfolded completely. Adding the hydrophobic constraints to the simulations alters the time it takes for the helical unfolding at 1000 K. At 20 ps, the distance constraint set between hydrophobic residues is still fulfilled, while the % helicity ultimately reaches 0%. This fact indicates that the helical units of secondary structure are not forced to remain intact by the constraint set. Rather, the interactions between the side chains of the helical residues and the hydrophobic core slows down the rate of thermal disruption of the helical secondary structure. This suggests that the tertiary hydrophobic interactions stabilize the secondary structure. To further examine this point, we have repeated the simulations of the unconstrained protein and the protein with hydrophobic constraints for 50 ps each at 500 K. The results are shown in figs. 4a and b. The trends are the same as observed at higher temperature but the helical unfolding is slowed down in both simulations. In the unconstrained simulation at 500 K, the C helix is  $\sim 50\%$  disrupted after 20 ps and all helix content has been lost within 50 ps. About 50% of the B helix remains intact after the 50 ps simulation at 500 K. We observed additional protection of the helices in the simulations with hydrophobic constraints. About 60% of the C and 75% of the B helix remain after 50 ps of dynamics at 500 K when the hydrophobic core is constrained. Upon further lowering the temperature to 298 K; we observed in an unconstrained simulation of  $\alpha$ -lactalbumin at room temperature (results not shown) that residues 23–34 and 86–99 remain 100% helical during a 50 ps simulation.

We have used an empirical free-energy scale based on accessible surface areas [28] to estimate the free-energy changes associated with the unfolding observed in our high-temperature simulations (1000 K) of  $\alpha$ -lactalbumin. The results are shown in fig. 5. For the simulation without constraints, the hydration free energy increases as time increases, indicating that a larger number of hydrophobic residues have been exposed to the surrounding solvent; while for the simulation using hydrophobic constraints, the hydration

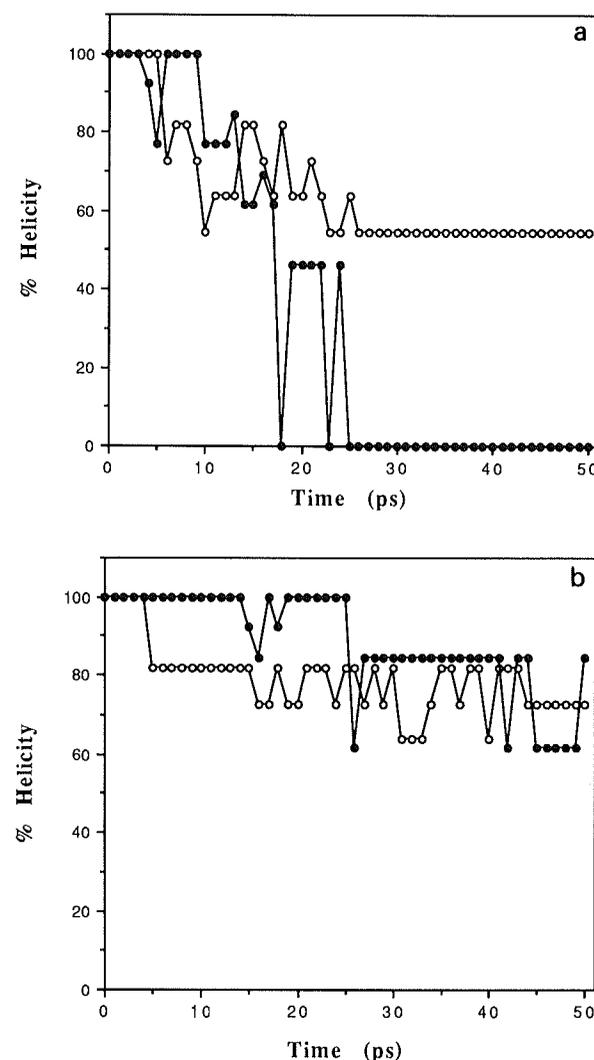


Fig. 4. % helicity of the (○) B and (●) helices during the dynamics course at 500 K, (a) with no constraint; (b) with hydrophobic constraints.

free energies fluctuate around  $-70$  kcal/mol during the course of the simulation. The hydration free energies of the structures obtained with the helix constraints are always higher than those obtained with hydrophobic constraints during the dynamics simulations. This is consistent with the pictures of these states (fig. 2) in which the interior of the protein appears much more exposed in the simulation with the helix constraints as compared with the simulation with hydrophobic constraints. At the end of 15 ps,

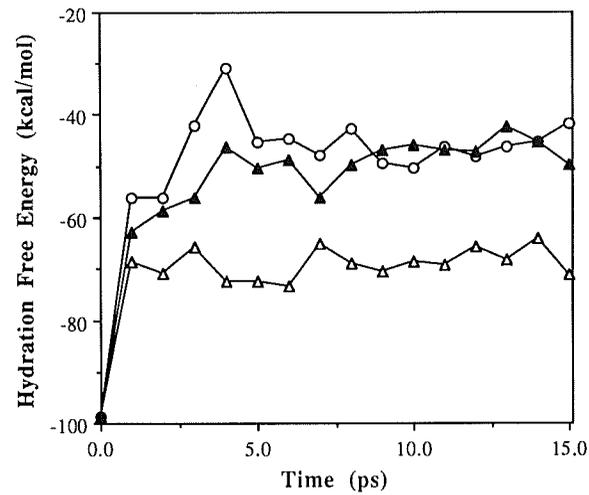


Fig. 5. Calculated hydration free energies as a function of simulation time for the molecular dynamics simulations at 1000 K, (○) with no constraint; (▲) with intrahelical constraints; (△) with hydrophobic constraints.

the structure generated from the simulation with hydrophobic constraints is about 20 kcal/mol more stable than the one obtained using helix constraints. As expected, structures in which the core of the protein is more compact have a lower hydrophobic free energy as estimated by the empirical surface accessibility criteria.

#### 4. Conclusions

Considerable interest has developed in using NMR, CD and molecular-dynamics simulations to study the stability of  $\alpha$ -helices in small peptide fragments [26,29–34]. However, simulations and experimental studies of peptide fragment  $\alpha$ -helices alone do not address the role played by the remainder of the protein in stabilizing the  $\alpha$ -helix. It is known from the analysis of solved structures that the sequence-dependent preference for  $\alpha$ -helical structure is weak, and that, therefore, nonlocal interactions must play an important role in stabilizing the  $\alpha$ -helical secondary structure [35]. For example, experimental studies on BPTI peptide fragments showed that the isolated C-terminal  $\alpha$ -helix in aqueous solution is only marginally stable [36], but that the interaction of this helix with the central  $\beta$ -sheet region increased the stability

substantially [37]. The studies of  $\alpha$ -lactalbumin reported here constitute an initial effort to use simulations to study the role of nonlocal interactions in stabilizing the  $\alpha$ -helix. When the protein is unfolded without constraints it loses all the secondary structure within 10 ps at high temperature. When the B and C helix are restrained, they remain intact but move apart from each other. In contrast, when the hydrophobic core of the protein is loosely constrained, the  $\alpha$ -helical secondary structure is stabilized against thermal disruption. It is interesting to note that the propagation of the constraints on the stability of the protein is unidirectional; constraining the helices alone does not stabilize the hydrophobic core of the protein, whereas constraining the hydrophobic core does stabilize the  $\alpha$ -helices. These results suggest that hydrophobic tertiary interactions are important in stabilizing the  $\alpha$ -helical secondary structure. Further questions that are being investigated via molecular-dynamics simulations are: is it possible to identify an ensemble of structures which have features associated with protein molten globules; what are the dynamics of the side chains in molten-globule states; what role does the solvent play in the early steps of unfolding and does solvent penetrate the protein interior? Characterization of the interactions that stabilize molten-globule proteins and developing molecular models of these states will provide insight into protein folding mechanisms.

#### Acknowledgement

We thank Dr. Phil Evans and Dr. Chris Dobson for helpful discussions. This work has been supported by NIH grant GM-30580. J. Baum is a Searle Scholar. We acknowledge a grant of supercomputer time from the National Cancer Institute of the National Institutes of Health.

#### References

- [1] K. Kuwajima, *Proteins* 6 (1989) 87.
- [2] M. Ohgushi and A. Wada, *FEBS Letters* 164 (1983) 21.
- [3] O.B. Ptitsyn, *J. Protein Chem.* 6 (1987) 273.
- [4] O.B. Ptitsyn, R.H. Pain, G.V. Semisotnov, E. Zerovnik and O.I. Razgulyaev, *FEBS Letters* 262 (1990) 20.
- [5] J.B. Udgaonkar and R.L. Baldwin, *Nature* 335 (1988) 694.

- [6] H. Roder, G.A. Elöve and S.W. Englander, *Nature* 335 (1988) 700.
- [7] A. Miranker, S.E. Radford, M. Karplus and C.M. Dobson, *Nature* 349 (1991) 633.
- [8] J. Baum, C.M. Dobson, P.A. Evans and C. Hanley, *Biochemistry* 28 (1989) 7.
- [9] C.L. Chyan, C. Hanley, P.A. Evans, C.M. Dobson, and J. Baum, in preparation.
- [10] C.M. Dobson, C. Hanley, S.E. Radford, J. Baum and P.A. Evans, in: *Conformations and Forces in Protein Folding*, eds. B.T. Nall and K.A. Dill (1991) 175.
- [11] K.R. Acharya, D.I. Straut, N.P.C. Walker, M. Lewis and D.C. Phillips, *J. Mol. Biol.* 208 (1989) 99.
- [12] M. Jeng, S.W. Englander, G.A. Elöve, A.J. Wand and H. Roder, *Biochemistry* 29 (1990) 10433.
- [13] F.M. Hughson, P.E. Wright and R.L. Baldwin, *Science* 249 (1990) 1544.
- [14] J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ. Press, Cambridge, 1987).
- [15] C.L. Brooks III, M. Karplus and B.M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (Wiley, New York, 1988).
- [16] M. Levitt, *J. Mol. Biol.* 170 (1983) 723.
- [17] B.R. Brooks, *Chem. Scripta* 29A (1989) 165.
- [18] H.S. Chan and K.A. Dill, *Proc. Natl. Acad. Sci. USA* 87 (1990) 6388.
- [19] J. Skolnick and A. Kolinski, *Science* 250 (1990) 1121.
- [20] D.B. Kitchen, F. Hirata, J.D. Westbrook, R.M. Levy, D. Kofke and M. Yarmush, *J. Comput. Chem.* 11 (1990) 1169.
- [21] S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, *J. Comput. Chem.* 7 (1986) 230.
- [22] H.C. Andersen, *J. Comput. Phys.* 52 (1983) 24.
- [23] H.J.C. Berendsen, J.P. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak, *J. Chem. Phys.* 81 (1984) 3684.
- [24] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.* 4 (2) (1983) 187.
- [25] M. Whitlow and M. Tector, *J. Am. Chem. Soc.* 108 (1986) 7163.
- [26] J. Tirado-Rives and W. Jorgensen, *Biochemistry* 30 (1991) 3864.
- [27] K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
- [28] D. Eisenberg and A.D. McLachlan, *Nature* 319 (1986) 199.
- [29] P.S. Kim and R.L. Baldwin, *Nature* 307 (1984) 329.
- [30] H.J. Dyson, M. Rance, R.A. Houghten, P.E. Wright and R.A. Lerner, *J. Mol. Biol.* 201 (1988) 201.
- [31] F.M. DiCapua, S. Swaminathan and D.L. Beveridge, *J. Am. Chem. Soc.* 112 (1990) 6768.
- [32] D.A. Case, *Biophys. J.* 397A (1991) U. 59.
- [33] V. Daggett, P.A. Kollman and I.D. Kuntz, *Biopolymers*, in press.
- [34] R.M. Levy, D. Perahia and M. Karplus, *Proc. Natl. Acad. Sci. USA* 79 (1982) 1346.
- [35] W. Kabsch and C. Sander, *Proc. Natl. Acad. Sci. USA* 81 (1984) 1075.
- [36] E.M. Goodman and P.S. Kim, *Biochemistry* 28 (1989) 4343.
- [37] T.G. Oas and P.S. Kim, *Nature* 336 (1988) 42.

6  
1

4

2  
3