

MOLTEN GLOBULE AND PROTEIN FOLDING

By O. B. PTITSYN

Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia, and Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

I.	Introduction	83
II.	Equilibrium Molten Globule State	86
	A. General Properties	86
	B. Native-like Structural Organization	98
III.	Phase Transitions	112
	A. Denaturation	112
	B. Unfolding	127
	C. Biological Significance and Physical Explanation	136
IV.	Kinetics of Protein Folding	142
	A. Formation of Premolten Globule Intermediate	143
	B. Formation of Molten Globule Intermediate	154
	C. Formation of Native Tertiary Structure	168
V.	Mechanism of Protein Folding	171
	A. Folding Intermediates	171
	B. How Do Proteins Fold?	187
VI.	Physiological Role of Molten Globule	195
	A. Prediction	195
	B. Folding and Assembly	197
	C. Penetration into Membranes	205
	D. Miscellaneous	211
VII.	Conclusion	216
	References	217

I. INTRODUCTION

Shortly after Sanger determined the amino acid sequence of the first protein (Sanger, 1952), Anfinsen tried to find out whether protein sequence can spontaneously fold into the native three-dimensional (3D) structure with the corresponding function. The answer was "yes," as it has been shown that a protein with broken disulfide bonds unfolded by urea can spontaneously fold again, restoring its native disulfide bonds and full native activity (see Anfinsen, 1973, for a summary of these investigations).

Next the mechanism of protein folding was questioned. Does a protein fold by a complete search of all possible conformations looking for the structure with minimal energy or does it fall down into one of energy minima (not necessary the global one) by some mechanistic folding pathway coded in its sequence? The first possibility implies that protein

native 3D structures are under thermodynamic control, while the second possibility means that the native structure is the result of some definite pathway, i.e., it is under kinetic control.

The obvious difficulty with the first approach is that a very large number of protein conformations have to be searched (Levinthal, 1968, 1969), whereas the obvious difficulty with the second one is that proteins fold into the same structure in quite different situations *in vitro* and *in vivo*.

The more general approach proposed by the author 20 years ago (Ptitsyn, 1973) was based on the assumption that a protein folds through several intermediates and that each intermediate has an increasing number of native-like structural features which are not reconsidered but just fixed at subsequent stages of folding. This mechanism may be the result of a kinetic pathway if a protein chain cannot explore alternative possibilities as it folds. However, it may also be consistent with thermodynamic control, if "new" interactions emerging at each subsequent stage do not change the most favorable structure outlined by "old" interactions at preceding stages. It is quite possible that the contradiction between kinetic and thermodynamic approaches to protein folding exists only in nonnatural "proteins," i.e., in random sequences of different amino acid residues. Biological evolution may select sequences for which the shortest kinetic pathway leads directly to the global minimum of energy (Shakhnovich and Gutin, 1993a,b).

In a more concrete way this mechanism (see also Ptitsyn, 1987, 1991, 1992a) implied the existence of at least two kinetic intermediates (see Fig. 1). The first (early) intermediate has secondary structure regions

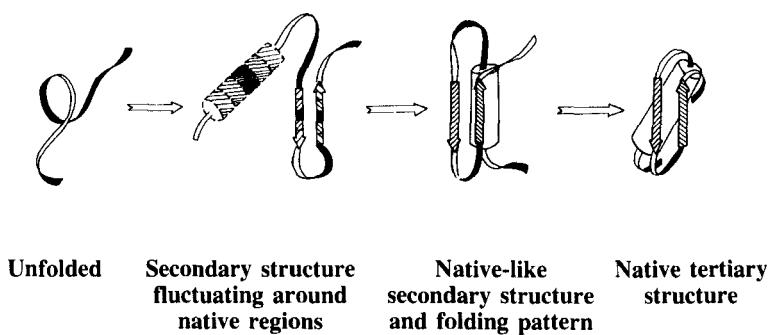


FIG. 1. Framework model of protein folding (Ptitsyn, 1973). The model has predicted three stages of protein folding: the formation of secondary structure regions fluctuating around their native positions, the merging of these regions into a compact state with the native-like folding pattern, and the transition of this intermediate state into the rigid native state.

fluctuating around their native positions, while the second (late) intermediate is compact and has native-like overall architecture but lacks the unique native tertiary structure (i.e., no rigid structure on atomic level). In other words, it has been assumed that a protein has three levels of its 3D structure—secondary structure, overall architecture, and rigid tertiary structure—and that these three structural levels are formed basically at three different stages of protein folding. It should be mentioned that the very existence of the second structural level (crude mutual positions of α and β regions) had been for the first time postulated as a part of this hypothesis before this structural level (now called folding pattern or tertiary fold) was identified from X-ray data on globular proteins (Levitt and Chothia, 1976; Richardson, 1977, 1981).

The difference between the folding pattern and the tertiary structure of proteins needs to be discussed. It is well known that the native proteins have rigid tertiary structure, which is supported by tight packing of side chains in the protein core, long-range hydrogen bonds, salt bridges, etc. This tertiary structure is dependent on all details of the amino acid sequence and therefore is specific for each given protein. However, in order to simplify this picture, we focus our attention on more global features, i.e., on the localization of α helices and β strands in protein chains and on their approximate arrangement in 3D space (Levitt and Chothia, 1976; Richardson, 1977, 1981; Ptitsyn and Finkelstein, 1980). This “crude” level of protein structure was initially considered just as a convenient mode for a simplified representation. However, almost immediately it was realized (Richardson, 1977, 1981) that this crude level is similar or even identical for different proteins. The analysis of these similarities led to the conclusion that it is due not to evolutionary divergence or functional convergence, but to the general laws of protein physics (Ptitsyn and Finkelstein, 1980; Finkelstein and Ptitsyn, 1987). This framework hypothesis has predicted that the folding pattern level of protein 3D structure can exist without support from rigid tertiary structure at the atomic level.

An important aspect of the framework hypothesis is that the secondary structure of the first intermediates was assumed to fluctuate not along the whole chain but rather around the native positions of α helices and β strands. In a similar way, it was assumed that the overall protein architecture in the second intermediate fluctuates around the native tertiary fold, i.e., around the native mutual positions of α helices and β strands. This first assumption was later called the framework model (Kim and Baldwin, 1982), while the second assumption is usually referred to as the molten globule model. However, these two assumptions are conceptually similar to each other and had been proposed simultane-

ously. Therefore, the term "framework model" is appropriate to apply to this hypothesis as a whole.

The main features of this scheme have been confirmed by direct experiments. In 1981 the second predicted intermediate was revealed (Dolgikh *et al.*, 1981) as the equilibrium state at mild denaturing conditions: the molten globule state. Then, a second intermediate was observed in experiments on folding kinetics (Dolgikh *et al.*, 1984; Semisotnov *et al.*, 1987) and it was shown that it is similar to the equilibrium intermediate (Baldwin, 1993) and has a native-like 3D structure (Fersht, 1993). The first intermediate has also been found in kinetic (Kuwajima *et al.*, 1987; Gilmanshin and Ptitsyn, 1987) and equilibrium (Uversky and Ptitsyn, 1994) experiments, although its properties have not been as clearly defined as those of the molten globule state. It differs from the predicted intermediate by being partially condensed.

In this chapter I review recent data on these intermediates (both equilibrium and kinetic), focusing attention on the molten globule state which has been more thoroughly studied than others. Moreover, it has been predicted (Bychkova *et al.*, 1988) and confirmed experimentally that the molten globule state can exist in a living cell and play an important role in a number of physiological processes (see Bychkova and Ptitsyn, 1993a, for review). This intriguing problem will also be discussed in this chapter (see Section VI). As the molten globule state has already been the subject of a number of reviews (Ptitsyn, 1987, 1992a; Kuwajima, 1989; Christensen and Pain, 1991; Dobson, 1991, 1992; Baldwin, 1991, 1993) I shall concentrate mainly on new data, mentioning the earlier data only as background material.

II. EQUILIBRIUM MOLTEN GLOBULE STATE

A. General Properties

1. Discovery

Probably the first clear evidence that protein molecules may have some properties intermediate between those of the rigid native structure and those of a completely unfolded random coil were obtained as early as 1967 in Tanford's and Brandts' groups (Aune *et al.*, 1967; Brandts and Hunt, 1967). These authors have shown that proteins denatured by acid or by temperature can possess another cooperative transition induced by urea or guanidinium chloride (GdmCl). It follows that not all denatured states are structurally equivalent—some of them are not completely unfolded. This very important idea that protein denaturation may be decoupled from protein unfolding has been convincingly argued

in the well-known review by Tanford (1968). It was established that protein molecules can be more or less completely unfolded mainly by large concentrations of urea or GdmCl, while other denaturing agents (temperature, pH, etc.) may transform protein molecules into different "partly unfolded" states (Tanford, 1968).

By following this approach, Tanford and others have shown the existence of at least two cooperative transitions on urea- or GdmCl-induced unfolding of carbonic anhydrase (Wong and Tanford, 1973), growth hormone (Holladay *et al.*, 1974), and β -lactamase (Robson and Pain, 1976a,b). Especially important were the investigations performed by Kuwajima and colleagues who established the existence of two well-resolved GdmCl-induced transitions in bovine (Kuwajima *et al.*, 1976) and human (Nozaka *et al.*, 1978) α -lactalbumins (see Fig. 2). The first transition can be monitored by a large decrease in the near-UV circular dichroism (CD) that reflects the loss of a rigid asymmetrical environment of aromatic side chains, i.e., the loss of a rigid tertiary structure. The second transition can be observed by a large decrease in the far-UV circular dichroism which corresponds to the melting of the secondary structure. Kuwajima (1977) realized that he was dealing with an interme-

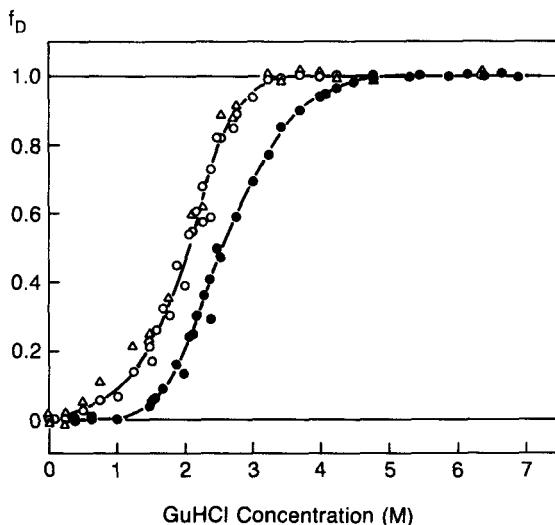


FIG. 2. Two stages of GdmCl-induced unfolding of bovine α -lactalbumin monitored by molar ellipticities $[\Theta]$ at 270 (\circ) and 296 nm (Δ) in aromatic region and at 222 nm (\bullet) in peptide region. Data are plotted in relative units $f_D = ([\Theta] - [\Theta]_N)/([\Theta]_D - [\Theta]_N)$, where $[\Theta]_N$ and $[\Theta]_D$ are the $[\Theta]$ values for the native and denatured states, respectively. Adapted with permission from Kuwajima *et al.* (1976).

diate state with a secondary structure, but without a rigid tertiary structure, and assumed that this state was an unfolded (noncompact) molecule with local secondary structure.

As the two GdmCl-induced transitions usually overlap, it is difficult to obtain this intermediate state without large admixtures of either native or unfolded states. Fortunately, it has been shown that carbonic anhydrase (Wong and Hamlin, 1974), bovine (Kuwajima *et al.*, 1976) and human (Nozaka *et al.*, 1978) α -lactalbumins, as well as some other proteins, can be transformed by low pH into a state with optical parameters similar to those of a GdmCl-induced intermediate. This opened the possibility of studying acid forms of these proteins to understand the physical properties of this intermediate state.

In 1981 we showed (Dolgikh *et al.*, 1981) that the acid forms of bovine and human α -lactalbumins are almost as compact as their native forms and have native-like secondary structure contents. On the other hand, they have no rigid tertiary structure, no cooperative temperature melting, and their large-scale intramolecular mobility is much larger than that in the native state. Therefore, this intermediate is not the unfolded chain with local secondary structure as was suggested by Kuwajima (1977), but a distinct physical state of protein molecules with an unusual combination of properties of native and unfolded proteins. This state, which was later named the "molten globule" (Ohgushi and Wada, 1983), has been the subject of very detailed experimental investigations (for review, see Ptitsyn, 1987, 1992a; Kuwajima, 1989, 1992; Christensen and Pain, 1991; Baldwin, 1991, 1993; Dobson, 1991, 1992; Bychkova and Ptitsyn, 1993a). In Section II,A I shall review the main properties of this state established in the early and middle 1980s mainly in my research group and discussed in previous reviews (Ptitsyn, 1987, 1992a). New important features of the molten globule state which have been revealed more recently by more sophisticated methods will be discussed in Section II,B.

2. Compactness

One of the most important properties of the molten globule state is that it is almost as compact as the native state. Figure 3 presents the first data demonstrating the compactness of the molten globule state. It shows that for both bovine and human α -lactalbumins the intrinsic viscosity in the pH-denatured state practically coincides with that of the native state. Almost the same value has been obtained for the temperature-denatured state of bovine α -lactalbumin. On the other hand, intrinsic viscosities of the GdmCl-unfolded states of both of these proteins (with intact S-S bonds) are about twofold larger than those in the native states.

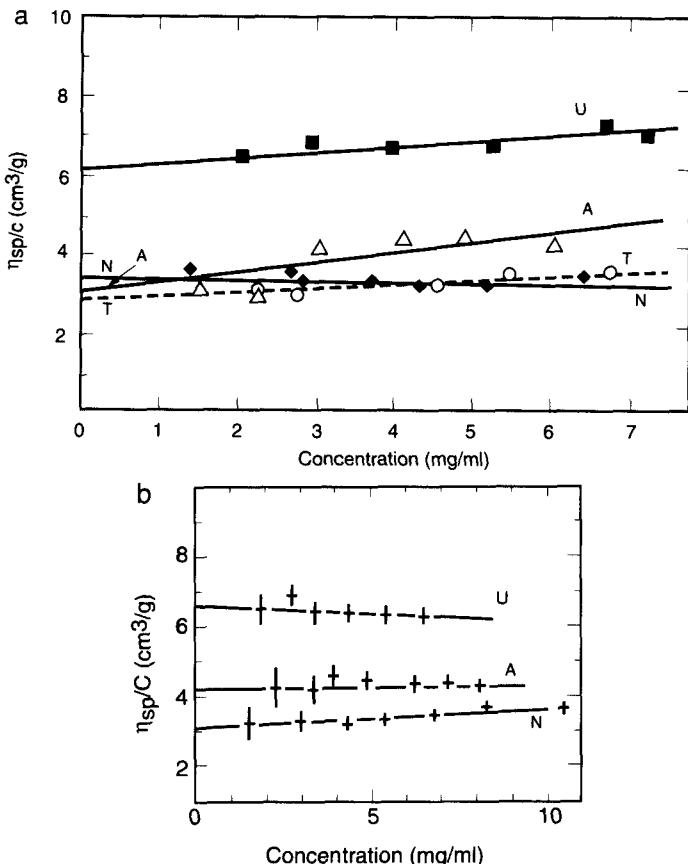


FIG. 3. First evidence of the compactness of the molten globule state. Intrinsic viscosities were estimated from the plots of reduced viscosity $\eta_{sp}/c = (\eta - \eta_0)/\eta_0 c$ versus c (η , viscosity of solution; η_0 , viscosity of solvent; c , concentration) for bovine (a) and human (b) α -lactalbumins. Plots are shown for the native states (N) (pH 7; 20°C), acid-denatured states (A) (pH 2; 20°C), and unfolded states (U) (6 M GdmCl; pH 7; 20°C) for both proteins. In addition, the plot for temperature-denatured state (T) of bovine α -lactalbumin (pH 7; 50°C in the absence of Ca^{2+} ions) is also shown (broken line). The experimental errors of η_{sp}/c are shown for human protein. Intrinsic viscosities [η] determined from these plots are: bovine α -lactalbumin, $[\eta]_N = 3.4$; $[\eta]_A = 3.1$, $[\eta]_T = 3.0$, $[\eta]_U = 6.1 \text{ cm}^3/\text{g}$; human α -lactalbumin, $[\eta]_N = 3.1$, $[\eta]_A = 4.2$, $[\eta]_U = 6.6 \text{ cm}^3/\text{g}$. Experimental errors in measurements of small intrinsic viscosities are $\pm 0.4 \text{ cm}^3/\text{g}$. Plots adopted with permission from Ph.D. thesis of D. A. Dolgikh (1983) and R. I. Gilmanishin (1985); intrinsic viscosities have been published by Dolgikh *et al.* (1981).

Similar data on intrinsic viscosities have been obtained for a number of other proteins. The average value of intrinsic viscosities of the six proteins for which experimental data were available in 1992 (Ptitsyn, 1992a; Bychkova and Ptitsyn, 1993b) is 4.2, which is $\sim 25\%$ larger than that for native proteins. This permits one to conclude that the molten globule state is compact, but it is insufficient support for quantitative conclusions as relative errors in small intrinsic viscosities can be rather large.

More quantitative estimates can be obtained from diffusion coefficients that can be precisely measured, either by quasielastic light scattering (Gast *et al.*, 1986) or by polarization interferometry (Bychkova *et al.*, 1990). Frequently, proteins tend to aggregate in the molten globule state, but this aggregation can be taken into account using their apparent molecular weights measured either by equilibrium sedimentation or by light scattering (Gast *et al.*, 1986). After this correction, the hydrodynamic radii measured by quasielastic light scattering and by polarization interferometry for pH- and temperature-induced molten globule states of human α -lactalbumin are surprisingly similar and exceed their native values by only 13 or 14% (Ptitsyn, 1992a). A similar value of 11% can also be obtained from the sedimentation coefficient of the pH-induced molten globule. (Gilmanshin *et al.*, 1982) after correction for a small aggregation (Ptitsyn, 1992a). Almost the same values (16 or 17%) have been obtained for bovine α -lactalbumin from diffusion (measured by polarization interferometer; see Bychkova *et al.*, 1990) and sedimentation (Gilmanshin *et al.*, 1982) data. (Note that without the correction for aggregation the ratio of hydrodynamic radii for the molten globule and the native state varies from 1.05 to 1.32.) The average value for the increase in the hydrodynamic radius of α -lactalbumins in the molten globule state compared with the native state is $14 \pm 2\%$ (Ptitsyn, 1992a) which corresponds to a volume increase of $50 \pm 8\%$. This should be contrasted with the increase in hydrodynamic radii of α -lactalbumin molecules unfolded by GdmCl (with intact S-S bonds). This increase is as large as $49 \pm 5\%$, which corresponds to a 3.3 ± 0.3 -fold increase in a molecular volume.

The increase in the molecular volume of α -lactalbumin in the molten globule state is surprisingly similar to the values obtained for quite different proteins, such as carbonic anhydrase and β -lactamase, by a different technique: size-exclusion chromatography [fast protein liquid chromatography (FPLC)]. Size-exclusion chromatography was applied (Uversky and Ptitsyn, 1994, 1995) to two proteins, carbonic anhydrase and β -lactamase, for which GdmCl-induced unfolding is known to take place through the molten globule state (Wong and Tanford, 1973; Rob-

son and Pain, 1976a,b). It was possible to measure elution volumes of both of these proteins in this state at 4°C and to compare them with the native state. The result was that the molten globule state of both proteins has a hydrodynamic radius 15 or 16% larger than the native state, which coincides almost exactly with the results for pH- and temperature-induced molten globule states of α -lactalbumins.

Diffuse X-ray scattering can be applied to obtain more information on the compactness of the molten globule state. The radius of gyration obtained from small-angle scattering (Dolgikh *et al.*, 1981, 1985) is similar to that of the native state (a more quantitative comparison is not reliable because small-angle X-ray scattering is very sensitive to aggregation and other intermolecular effects). More interesting is the middle-angle scattering which gives information on the distance distribution inside a protein. It was shown (Dolgikh *et al.*, 1985) that scattering curves for the native and the molten globule states of α -lactalbumin practically coincide, which gives very similar pair distribution functions. A more detailed study of cytochrome c (Kataoka *et al.*, 1993) used the so-called Kratky plot, i.e., the dependence of $I(Q) \times Q^2$ versus Q where $Q = 4\pi(\sin \Theta)/\lambda$ (2Θ scattering angle; λ , the wavelength of X-rays). This plot must monotonically increase and then reach a plateau for a random coil, while it goes through a rather sharp maximum for globules. Figure 4 shows that cytochrome c in the molten globule state (at pH 2 with electrostatic repulsion of positive charges reduced either by 0.5 M NaCl or by their partial acetylation) has a typical globular shape, while at low ionic strength, in the absence of acetylation, repulsion of positive charges transforms the molten globule into a random coil.

At large scattering angles, X-ray scattering begins to "see" smaller interatomic distances. In particular, all native globular proteins have a pronounced maximum corresponding to the Bragg distance of ~ 4.5 Å. This maximum is due mainly to a large number of long-range van der Waals contacts between atoms in a globular structure (Fedorov and Ptitsyn, 1977) and is absent or weak in unfolded or helical polypeptides (Damaschun *et al.*, 1986). It was shown that this maximum is also preserved in the molten globule state of α -lactalbumin induced by acid pH (Damaschun *et al.*, 1986) or by high temperature (Ptitsyn *et al.*, 1986; see also Ptitsyn, 1987). The only difference is that in both cases the maximum is shifted from 4.50 to 4.65 Å, i.e., by 3 or 4%, compared with that of the native state. It follows that the increase of distances between neighboring atoms in space is much smaller than the increase of the overall hydrodynamic radius (which is $\sim 15\%$, see above). This suggests that the protein preserves its "core" in the molten globule state and that the dimensions of this core increase even less than the increase of

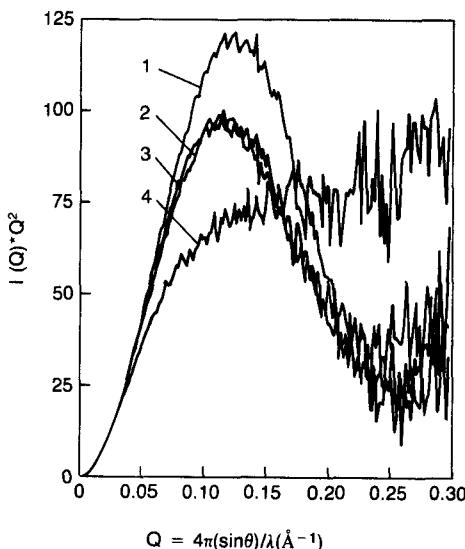


FIG. 4. Kratky plots for cytochrome c: 1, the native state at pH 7 (ionic strength 0.01 M); 2 and 3, the molten globule states at pH 2 and electrostatic repulsion reduced either by 0.5 M NaCl (2) or by partial acetylation of positively charged groups (3); 4, the unfolded state at pH 2 and low (0.02 M) ionic strength. Reproduced with permission from Kataoka *et al.* (1993).

"external" hydrodynamic dimensions. In other words, the expansion of a protein molecule in the molten globule state may be nonuniform (Ptitsyn *et al.*, 1986; Ptitsyn, 1992a): the core remains packed only a little more loosely than in the native state, while the "shell" is much more expanded.

3. Secondary Structure

Typically, far-UV CD spectra of proteins in the molten globule state are similar to or even more pronounced than those in the native state, which suggests a large amount of secondary structure (see Ptitsyn, 1987; Kuwajima, 1989, for reviews). This is illustrated by Fig. 5a, which shows far-UV circular dichroism spectra for bovine α -lactalbumin in the native state, two molten globule states (acid- and temperature-denatured), and the unfolded state. Circular dichroism spectra of both molten globule states clearly show the existence of a pronounced secondary structure (unlike the spectrum of protein in GdmCl). The difference between far-UV spectra in the native and the molten globule states does not necessarily mean a change in secondary structure. It can be explained as well by the disappearance of the contribution of aromatic side chains

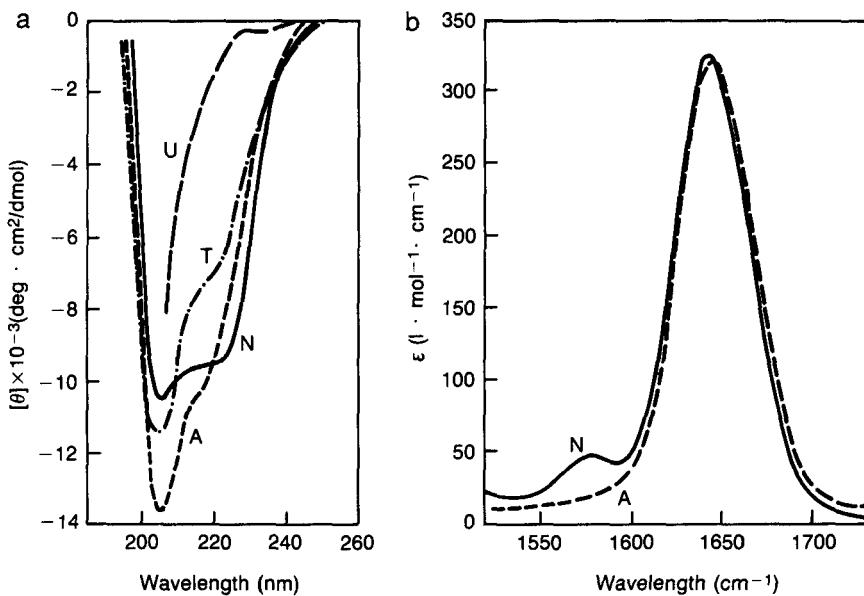


FIG. 5. (a) Far-UV circular dichroism spectra of bovine α -lactalbumin: N, native state (pH 7); A, acid state (pH 2); T, temperature-denatured state (pH 7 at 90°C); U, unfolded state (6 M GdmCl). CD spectra at 90°C and pH 2 is very similar to that at 90°C and pH 7. (b) Infrared spectra of bovine α -lactalbumin in the native (N) and acid (A) states. The difference between the spectra at 1550–1600 cm⁻¹ almost certainly is due to deionization of aspartic and glutamic acid residues at pH 2. Reproduced with permission from Dolgikh *et al.* (1985).

and/or disulfide bonds (Kronman *et al.*, 1966; Bolotina, 1987; Manning and Woody, 1989; Chaffotte *et al.*, 1992). This point of view is confirmed by the comparison of infrared spectra for the native and the molten globule (acid) states. Figure 5b presents infrared spectra of bovine α -lactalbumin in the amide I region (after the subtraction of side-chain contributions) and shows that these spectra are almost identical in the native and the molten globule states. In carbonic anhydrase (Dolgikh *et al.*, 1983; see also Ptitsyn, 1987) infrared spectra for the native and acid states differ substantially in the amide I region. However, the evaluation of the β -structural content gives almost the same values in both of these states, and the difference in infrared spectra can be ascribed to the broadening of the main β -structural infrared band that may suggest some disordering to the β structure.

More recently, a “pH-jump” method with two-dimensional nuclear magnetic resonance (2D NMR) has been applied for the identification

of stable secondary structure in the molten globule state and it was shown that a part of the stable α helices in α -lactalbumin, cytochrome c, and apomyoglobin occurs at nearly the same sequence positions as those in the native state (see below, Section II,B,1).

4. Internal Mobility

In contrast with far-UV CD, near-UV CD spectra of the molten globule state are dramatically reduced in comparison with the native ones, which suggests the virtual absence of a rigid tertiary structure (see Kuwajima, 1989, for a review). NMR spectra for the molten globule state are much simpler than those for native proteins and are more, or rather, similar to the spectra of unfolded chains (Dolgikh *et al.*, 1985; Kuwajima *et al.*, 1986; Rodionova *et al.*, 1989; Baum *et al.*, 1989; Dobson *et al.*, 1991). On the other hand, these spectra are still significantly perturbed, which suggests the existence of some traces of a rigid structure.

The very small ellipticities of near-UV CD spectra and the usually small perturbation of NMR spectra suggest strongly increased side-chain fluctuations in the molten globule state. In the case of carbonic anhydrase, this has been directly confirmed by spin echo measurements of the spin–spin relaxation times T_2 (Semisotnov *et al.*, 1989; see also Ptitsyn, 1992a). These measurements indicated that methyl groups in the molten globule state are much more mobile than those in the native state and nearly as mobile as those in the unfolded state. On the other hand, the mobility of aromatic side chains in the molten globule state is between those in the native and unfolded states. These data were also confirmed by the study of the two-stage urea-induced unfolding of carbonic anhydrase (Rodionova *et al.*, 1989; see also Ptitsyn, 1992a). It was shown that NMR signals of aliphatic groups change in parallel with the near-UV ellipticity, and changes in NMR signals of aromatic groups coincide with the far-UV ellipticity. This means that aliphatic groups begin to move in the molten globule intermediate, while the rotation of aromatic groups remains substantially hindered. It seems that the expansion of protein volume in the molten globule is sufficient to ensure relatively free movements of the relatively small and symmetric aliphatic groups but not enough to permit free movements of the larger and less symmetric aromatic groups. This important difference between the mobilities of aliphatic and aromatic side chains in the molten globule state was predicted by Shakhnovich and Finkelstein (1989). The large restrictions of mobility of aromatic side chains have also been confirmed (for tryptophan residues) by the polarization of fluorescence (Dolgikh *et al.*, 1981, 1985; Rodionova *et al.*, 1989).

Moreover, the molten globule state is characterized by a substantial increase in large-scale fluctuations of molecular structure which makes internal parts of a molecule accessible to solvent. It is reflected in a large increase in the rate of hydrogen exchange (Dilgikh *et al.*, 1981, 1983, 1985; Baum *et al.*, 1989) and a substantial increase of protein susceptibility to proteases (see, e.g., Merrill *et al.*, 1990).

5. *Hydrophobic Surface*

The semiflexible nature of the molten globule state permits some internal nonpolar groups to become exposed to water, thus making the surface of this state more hydrophobic than that for the native state. One of the clear manifestations of this phenomenon is that the molten globule state can bind nonpolar molecules from solution much more strongly than the native state. A typical example is 8-anilinonaphthalene 1-sulfonate (ANS) which binds to solvent-accessible clusters of nonpolar atoms in the native state (Stryer, 1965), but is bound more strongly by the molten globule state (Semisotnov *et al.*, 1987, 1991a). ANS provides a particularly convenient test for the molten globule state as its binding leads to a large increase in its fluorescence.

Another interesting example is protein labeling by hydrophobic photo-activatable reagents like 3-(trifluoromethyl)-3-(*m*-[¹²⁵I]iodophenyl)diazirine] (Dumont and Richards, 1988) or [³H]diazofluorene (DAF) (Lala and Kaul, 1992). These reagents also label the molten globule state of a protein much more strongly than the native and the unfolded states. This has been well established for fragment A of diphtheria toxin (Dumont and Richards, 1988), α -lactalbumin (Lala and Kaul, 1992), as well as for carbonic anhydrase and dihydrofolate reductase (V. E. Bychkova and D. E. Bochkarev, unpublished data, 1991). The degree of labeling of α -lactalbumin by DAF does not change when pH decreases from pH 7.4 to pH 4.0, but increases by almost an order of magnitude between pH 4 and pH 3, where α -lactalbumin transforms from the native into the molten globule state (Lala and Kaul, 1992).

The overall increase in hydrophobic surface in the molten globule state is also confirmed by more detailed structural information. For example, it has been shown that all four tryptophans of bovine α -lactalbumin are buried in the native state, but that two of them become exposed in the molten globule state (Lala and Kaul, 1992).

The increase in hydrophobic surface of the molten globule state leads to an increased affinity for membranes (Cavard *et al.*, 1988; Cabiaux *et al.*, 1989; Merrill *et al.*, 1990), liposomes (Zhao and London, 1988), or phospholipids (Kim and Kim, 1986; Lee and Kim, 1988). This effect

can be important for the penetration of proteins through membranes or for their insertion into membranes (see Section VI,C).

6. Model

To summarize early results on the physical properties of the molten globule state we concluded (Gilmanshin *et al.*, 1982; Shakhnovich and Finkelstein, 1982; Ptitsyn *et al.*, 1983; see also Ptitsyn, 1987) that this state preserves a protein core and the native-like secondary structure, but loses the tight packing of side chains, and has much more pronounced fluctuations than the native state. Moreover, it was initially suggested that the molten globule state must preserve not only the native-like secondary structure but also the native-like folding pattern (Ptitsyn, 1973, 1987).

This model was later confirmed by more detailed investigations using 2D NMR and site-specific mutagenesis. Side-chain resonances in the acid state of guinea pig α -lactalbumin manifest a small dispersion of chemical shifts and a substantial line broadening which suggest the absence of a rigid tertiary structure and the averaging of resonances over different local conformations (Alexandrescu *et al.*, 1993). On the other hand, side-chain resonances are generally shifted from their random coil values, and a large number of nuclear Overhauser effects (NOEs) between different nonpolar residues still exist in the acid state, suggesting the existence of some clusters of nonpolar groups (see also Chyan *et al.*, 1993). These results "are reasonably consistent with the conventional molten globule model" (Chyan *et al.*, 1993), which implies the presence of hydrophobic clusters but without rigid mutual positions of side chains.

Evidence for this point was obtained even earlier from site-specific mutations in apomyoglobin (Hughson and Baldwin, 1989; Hughson *et al.*, 1991). It was shown that mutations in the protein core that increase its hydrophobicity stabilize the molten globule state against unfolding, but destabilize the native state. This confirms the idea that the molten globule state is stabilized by liquid-like interactions of nonpolar groups in a water environment (i.e., by so-called "hydrophobic interactions") rather than by tight packing of these groups as is the case for native structure.

The most important feature of the molten globule state is that it does not just contain hydrophobic clusters but that its overall architecture (folding pattern) must be similar to that of the native protein. This prediction has been confirmed by 2D NMR data for modified insulin (Hua *et al.*, 1992, 1993) and by reconstruction of S-S bonds in the α subdomain of α -lactalbumin (Peng and Kim, 1994) (see Section II,B,2).

The model of the molten globule state (Ptitsyn, 1992a) presented in Fig. 6 summarizes all these data, illustrating the existence of native-like secondary structure and the native-like folding pattern with a loosely packed nonpolar core and without the rigid tertiary structure. It should be emphasized that this model implies that the protein core is expanded in the molten globule state, which permits at least a part of the side chains to jump from one rotamer into another (Shakhnovich and Finkelstein, 1989). Simultaneously, some nonpolar side chains from the loops and the ends of a protein chain may leave the core, which may lead to a partial unfolding of these chain regions and even a decrease in the lengths of secondary structure regions to some regions. As a result, this model portrays the molten globule as a "semiflexible" state of a protein molecule. Its relatively flexible loops and loosely packed core permit it to adapt itself to a wide variety of external conditions. On the other hand, it is likely that it has a native-like folding pattern that permits retention of the important features of its overall architecture despite these adaptations.

7. More a Rule Than an Exception

Until relatively recently, the equilibrium molten globule state had been considered to be a rare exception in the kingdom of proteins. However, it now becomes evident that this is not so. Fink, Goto, and collaborators have shown that a number of proteins can be transformed into the molten globule state either at low pH (see, e.g., Goto and Fink, 1989; Goto *et al.*, 1990a,b) or at intermediate concentrations of GdmCl (Hagihara *et al.*, 1993). A number of other authors have reported on states of protein molecules virtually identical with the molten globule.

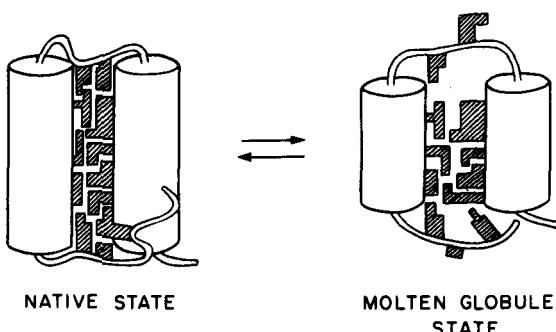


FIG. 6. Schematic model of the native and the molten globule states of protein molecule. Only two α helices are presented for simplicity. Nonpolar side chains are hatched. From Ptitsyn (1992a). In "Protein Folding" (T.E. Creighton, ed.). Copyright © 1992 by W. H. Freeman and Company. Reprinted with permission.

Data available to the beginning of 1992 (Bychkova and Ptitsyn, 1993b; see also Ptitsyn, 1992a) have shown that at least 20 proteins can pass into a compact state with pronounced secondary structure, but without rigid tertiary structure, which could therefore be identified with the molten globule state. This has led us to the conclusion that the molten globule "becomes now more the rule than the exception."

During the past few years the number of proteins reported to evidence the molten globule state continues to increase. Among new proteins with molten globule-like properties are apocytochrome b₅₆₂ (Feng *et al.*, 1991, 1994), heat-shock protein 73 (Palleros *et al.*, 1992), modified staphylococcal nuclease (James *et al.*, 1992), platelet factor 4 (Mayo *et al.*, 1992), reduced monomer of brain-derived neurotropic factor (Philo *et al.*, 1993; Narhi *et al.*, 1993), insulin with deleted pentapeptide (Hua *et al.*, 1993), and others. Some proteins active in a rigid dimer state can be observed as molten globule monomers under weakly denaturing conditions. These proteins include aspartate aminotransferase (Herold and Kirschner, 1990), platelet factor 4 (Mayo *et al.*, 1992), and brain-derived neurotropic factor (Philo *et al.*, 1993; Narhi *et al.*, 1993).

The wide occurrence of the equilibrium molten globule state, together with the observation that it serves as a universal kinetic intermediate in protein folding (see Sections IV and V) and is involved in a number of physiological processes (see Section VI), emphasizes its important role in biophysics and molecular biology.

B. Native-like Structural Organization

1. Native-like Secondary Structure

The overall properties of the molten globule state summarized previously lead to the understanding of many physical features of this state. However, these studies could not answer the central question: Does the molten globule have the important features of the native structural organization as was initially predicted (Ptitsyn, 1973), or is it just a relatively structureless compact state? The answer to this question is of primary importance (see, e.g., Baldwin, 1991), as only a native-like folding pattern of the molten globule state would permit straightforward consideration of the state as a necessary and important intermediate in protein folding.

There are two approaches for determining protein structures to atomic resolution: X-ray investigations of protein crystals (Blundell and Johnson, 1976; Perutz, 1992) and NMR investigations in solution (Wüthrich, 1986). Because no one has succeeded in crystallizing proteins in the

molten globule state, NMR is at present the only method that can be practically applied. For native proteins multidimensional NMR makes it possible to measure a large number of NOEs, which reflect the close proximities between different atoms in the well-ordered 3D structure. The main difficulty in applying this method to the molten globule state is that NMR spectra of this state usually exhibit small dispersions of chemical shifts and broad resonances, which make the detailed interpretation of its NMR spectra very difficult.

However, in the pioneering work of Dobson's group (Baum *et al.*, 1989), an approach has been proposed which partly overcomes this difficulty and permits one to at least obtain qualitative information on the stable secondary structures in the molten globule state. This approach involves the coupling of 2D NMR with hydrogen-deuterium exchange. In these experiments the protein is allowed to exchange in D₂O in the molten globule state (at acid pH) for various periods of time and then the protein is transformed into the native state by an increase in pH. Amide protons that are not well protected from the solvent in the molten globule state exchange during the incubation in D₂O, while slowly exchanging (well protected) protons remain protected and give resonances in the NMR spectrum at native conditions. Assignment of the resonances of a protein in the native state permits one to learn which protons are protected in the molten globule state.

Results from this pH-jump approach became more quantitative when Jeng *et al.* (1990) and Hughson *et al.* (1990) used the dependence of the number of exchanged protons on the time of incubation in D₂O to measure the kinetic exchange constants of each slowly exchanging proton. These experimental rate constants (k_{exp}) can be compared with the corresponding "intrinsic" values (k_{intr}) for each proton, calculated from the data of Englander's group (Englander and Poulsen, 1969; Molday *et al.*, 1972; Englander *et al.*, 1979; Englander and Mayne, 1992; Bai *et al.*, 1993).

The general results of this approach indicate that "protection factors" $P = k_{\text{intr}}/k_{\text{exp}}$ are much smaller in the molten globule state than in the native state (in which they often are 10⁸ or even larger). However, a number of protons are also substantially protected (with $P = 10^2$ or 10³) in the molten globule state. As a rule, these protons belong to NH groups which are involved in hydrogen bonds in a secondary structure. A short summary of the more detailed results follows.

a. *α-Lactalbumin.* NH protons of chain regions corresponding to native helices B and C are also substantially protected in the acidic (pH 2.0) molten globule state (Baum *et al.*, 1989; Dobson *et al.*, 1991;

Chyan *et al.*, 1993). Recent results are presented in Fig. 7. Almost all NH groups with protection factor $P > 10$ are located in chain regions 26–31 and 89–98, corresponding to native α helices B and C (see Acharya *et al.*, 1991). Outside the native helices B and C there are no protection factors exceeding those in denatured lysozyme, which is not in the molten globule state (Radford *et al.*, 1992b).

It is quite remarkable that all residues protected in the molten globule state are also strongly protected in the native state. This suggests that the protection in the molten globule state may be due to protein regions that preserve their native-like folding patterns. This conclusion is strongly supported by the fact that a number of residues involved in native contacts between helices B and C have especially large protection factors (>100) in the molten globule state, while the indole group of Trp-26 (also involved in these contacts) has $P > 900$. All these data support the conclusion (Chyan *et al.*, 1993) that the native nonpolar cluster between helices B and C can be preserved in the molten globule state and can participate in stabilizing its structure.

b. *Cytochrome c*. Similar results have been obtained for oxidized horse cytochrome c at pH 2.2 and high ionic strength (Jeng *et al.*, 1990). In this case, almost all NH groups with $P > 10$ are concentrated in chain regions 7–15, 64–70, and 91–101 residues, belonging to three main

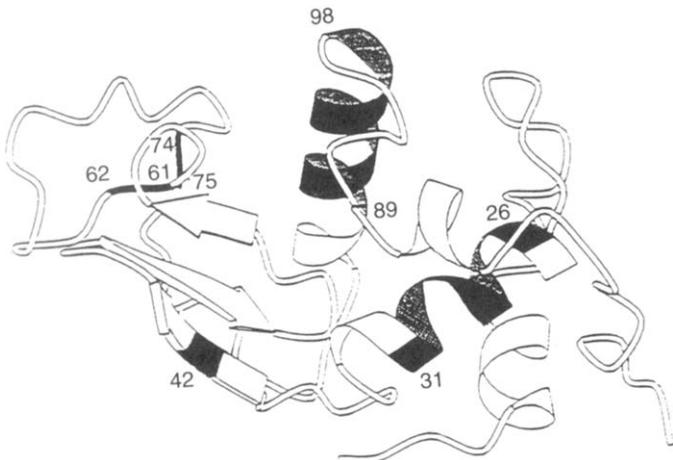


FIG. 7. Schematic presentation of the 3D structure of α -lactalbumin. Protection factors in the molten globule state have been measured for shaded segments. Almost all residues with $P > 10$ belong either to the native α -helix B (residues 26–31) or to the native α -helix C (residues 89–98). Adapted with permission from Chyan *et al.* (1993). Copyright 1993 American Chemical Society.

native α helices 3–14, 60–69, and 87–102 (Bushnell *et al.*, 1990) (see Fig. 8). In this case protection in the molten globule state is larger than in α -lactalbumin—almost all NH groups in these regions have protection factors $P > 100$ and some of them (Asp-15, Ile-96 and Leu-98, belonging to the N- and C-terminal native helices) even > 1000 .

Figure 9 compares the profiles of NH-exchange protection factors in the acid and the native states of cytochrome c. It shows that chain regions with slow exchange almost coincide in both states. Most of the strongly protected NH groups in the native state are clustered around the nonpolar core and the fact that these protons are strongly protected even in

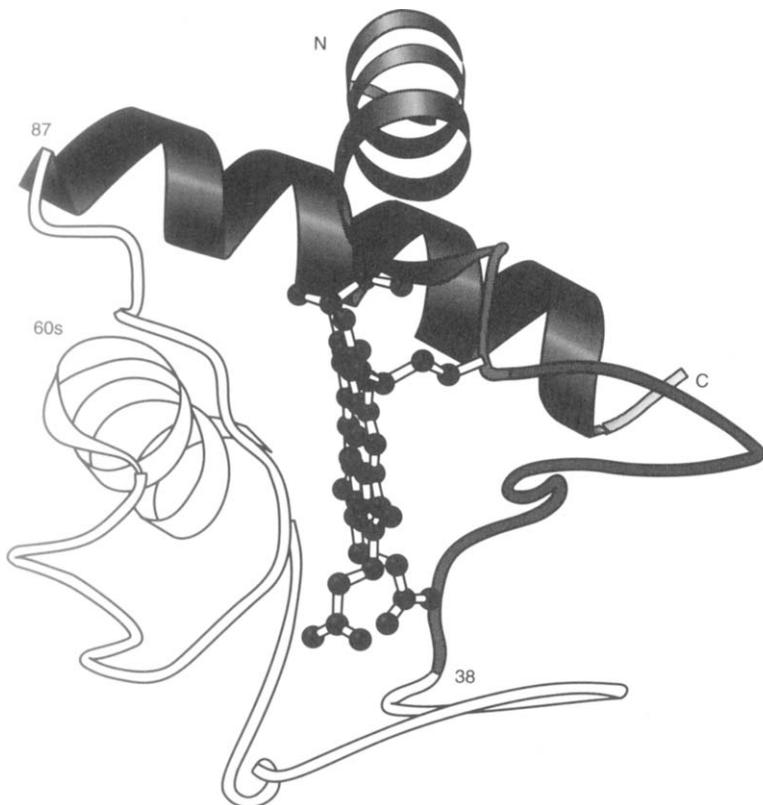


FIG. 8. Schematic presentation of 3D structure of cytochrome c illustrating the position of N- and C-terminal helices, 60's helix and heme. Shaded regions mark the locations of N-fragment (residues 1–38, including the heme), and C-fragment (residues 87–104) which form a complex in solution (see Section IV,B,1). Filled circles mark the heme. Adapted with permission from Wu *et al.* (1993). Copyright 1993 American Chemical Society.

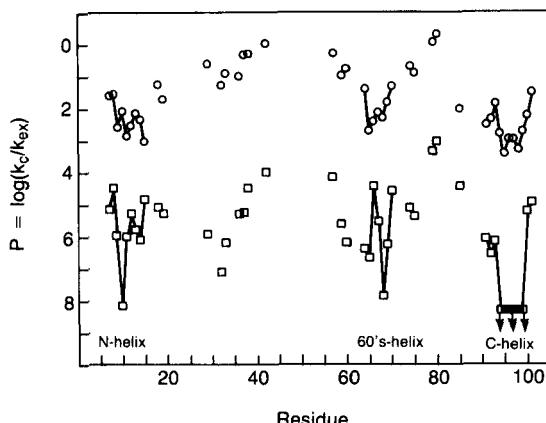


FIG. 9. Profile of NH exchange protection factors for oxidized horse cytochrome c is the native (\square) and acid, i.e., molten globule (\circ) states. Arrows show protection factors too large to be measured. Solid lines connect residues which form N-terminal helix (residues 7–15), 60's helix (residues 64–70), and C-terminal helix (residues 91–101) in the native protein. Reproduced with permission from Jeng *et al.* (1990). Copyright 1990 American Chemical Society.

the acid (molten globule) state is good evidence that the nonpolar core is at least partly formed in this state.

c. *Apomyoglobin*. At pH 4.2 apomyoglobin is in the molten globule state (Griko *et al.*, 1988); all its NH protons with $P > 10$ belong to native helices A, G, and H (Hughson *et al.*, 1990). Only one NH group (of residue Val-10 in helix A) has $P > 100$. Figure 10 compares protection factors in the acid and native states. It shows that helix A is protected relatively strongly ($< P > = 70$), helices G and H are protected more weakly ($< P > = 22\text{--}24$), while helix B has only marginal protection.

Figure 11 shows the mutual positions of three relatively well-protected helices in the native X-ray structure of myoglobin. It emphasizes that helices G and H form a hairpin covered by helix A. Since all these three helices are protected in the acid (molten globule) state, it is quite possible that their native mutual positions are preserved in this state (Hughson *et al.*, 1990).

The results obtained for these three proteins are remarkably similar. The acid (molten globule) state of these proteins preserves a substantial protection of amide protons in part of their native structures. It is interesting to note that in all three cases the protected parts involve α helices from both N and C termini (helices B and C in α -lactalbumin, N- and C-terminal helices as well as helix involving 60's residues in cytochrome

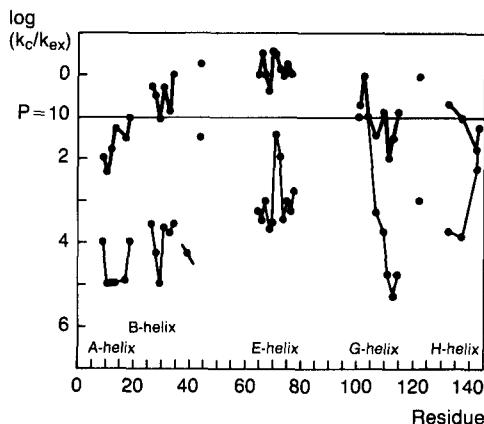


FIG. 10. Profiles of protection factors for NH proton exchange in sperm whale myoglobin in the acid (bold lines) and native (thin lines) states. Data are presented for 36 protons for which exchange can be quenched by the increase of pH. Lines connect residues belonging to five main helices of the native myoglobin. Adapted with permission from Baldwin and Roder (1991).

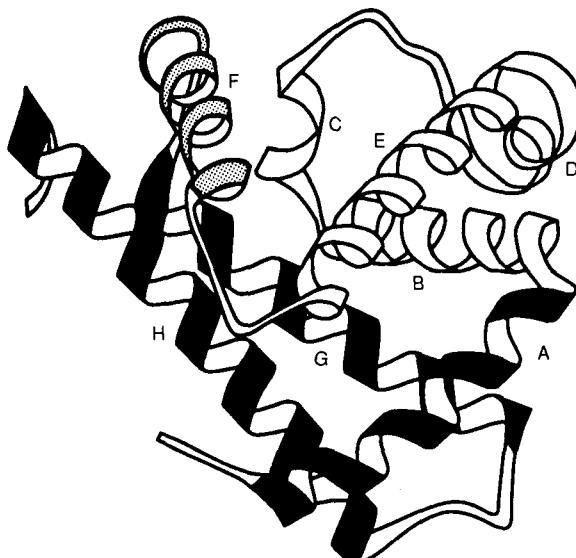


FIG. 11. Schematic presentation of the native structure of myoglobin. Darkly shaded regions indicate helices, A, G, and H which NH protons are protected from exchange in the acid (molten globule) state. Unshaded helices B-E are virtually not protected in this state. Helix F, which could not be probed by the pH-jump method, is lightly shaded. Reproduced with permission from Hughson *et al.* (1990). *Science* **249**, 1544-1548. Copyright 1990 by the AAAS.

c, helix A, and helices G and H in apomyoglobin). The remaining part of the native secondary structure in these proteins is practically unprotected in the molten globule state.

To interpret these data it is necessary to remember that NH protons can be protected from H→D exchange if they are involved in intramolecular hydrogen bonds and if these hydrogen bonds are relatively stable against thermal fluctuations. NH protons which are involved in strongly fluctuating hydrogen bonds would not be protected. This is well illustrated by the comparative studies of protein folding kinetics for lysozyme (Radford *et al.*, 1992a), cytochrome c (Elöve *et al.*, 1992), and interleukin-1 β (Varley *et al.*, 1993) using far-UV circular dichroism and H→D exchange (see Sections IV,A,1 and V,A,1). In both cases a very substantial part of the far-UV ellipticity is restored at the “burst” stage (within 2–4 msec), while substantial protection of NH protons starts much later (within 100 msec or even 1 sec).

The same can also be true under equilibrium conditions. Figure 12 shows the far-UV circular dichroism spectra for α -lactalbumin (Chyan *et al.*, 1993) and apomyoglobin (Hughson *et al.*, 1990) in the native and acid (molten globule) states. The circular dichroism spectra of α -lactalbumin are very similar in the native and molten globule states, indicating similar secondary structure contents. For apomyoglobin the difference is more pronounced but also not very large. The comparison of far-UV circular dichroism spectra of the molten globule state with protection factors shows that many unprotected NH groups can be involved in less stable and more fluctuating secondary structures. A similar observation has been made for temperature-denatured ribonuclease A: NH groups are practically nonprotected in this protein (Robertson and Baldwin, 1991) despite the existence of a considerable amount of secondary structure (Seshadri *et al.*, 1994).

We conclude that almost all protected NH protons belong to relatively stable α helices prevented by the rest of the protein from large fluctuations. However, the opposite is not necessarily true—unprotected chain regions may also be helical, but these regions can undergo relatively large fluctuations, exposing them to a solvent.

In summary, the pH-jump method has demonstrated the existence of a number of native-like stable α helices in the molten globule states of all three proteins studied. In all cases the surviving helices form clusters in the native structures which suggest that they can also form similar clusters (stabilized by hydrophobic interactions) in the molten globule state.

2. Native-like Folding Patterns

The data on protection of NH groups have shown that regions protected in the molten globule state form a subset of secondary structure

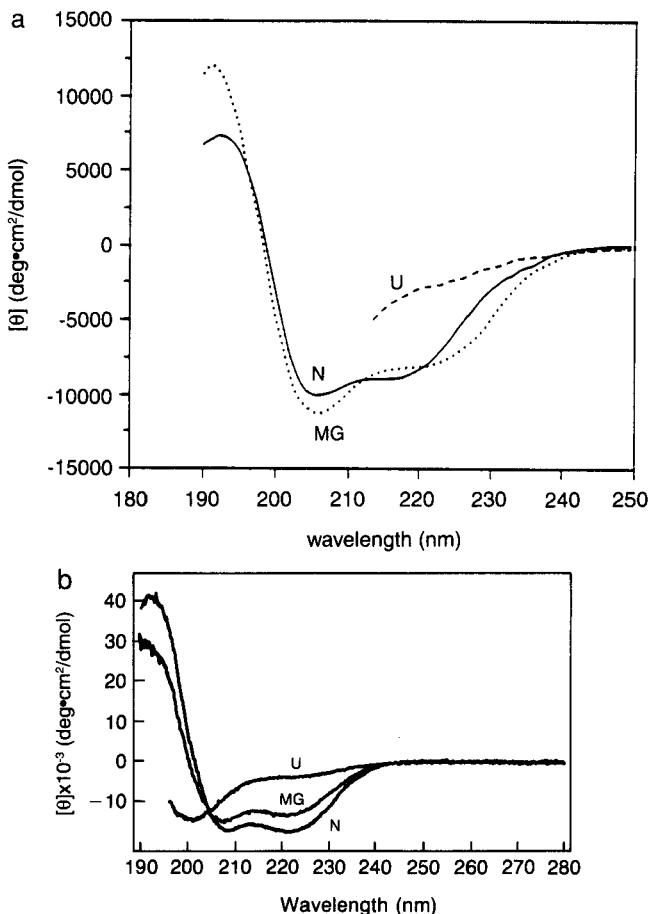


FIG. 12. Far-UV circular dichroism spectra of guinea pig α -lactalbumin (a) and sperm whale apomyoglobin (b) in the native (N), molten globule (MG), and unfolded (U) states. Conditions of experiments: α -lactalbumin, N (pH 7), MG (pH 2), U (9 M urea, pH 2); apomyoglobin, N (pH 6.1), MG (pH 4.3), U (pH 1.9). Adapted with permission from (a) Chyan *et al.* (1993). Copyright 1993 American Chemical Society; and (b) Hughson *et al.* (1990) *Science* **249**, 1544–1548. Copyright 1990 by the AAAS.

regions in the native state. Moreover, the secondary structure regions protected in the molten globule state are clustered together in the native state (like helices B and C in α -lactalbumin, N- and C-terminal helices in cytochrome c, or helices A, G, and H in apomyoglobin). These clusters include helices which are far apart from one another along the sequence but are close neighbors in 3D space in the native state. This strongly suggests that well-protected clusters of regions of secondary structure persist into the molten globule state.

A direct determination of the 3D structure of the molten globule in a water solution is very difficult due to the large broadening of resonances and the small dispersion of chemical shifts. More narrow resonances and sometimes even relatively large dispersion of chemical shifts can be observed for protein denatured in water-alcohol mixtures (see Section II,B,3), but it is unknown to what extent this denatured state is similar to the "normal" molten globule in aqueous solutions.

However, there is one example of a molten globule state with a relatively rich and informative 2D NMR spectrum in an aqueous solution: insulin with a deletion of residues 26 to 30 in the B chain. This "despentapeptide (B26-B30) insulin" (DPI) first described by Gattner (1975) and studied by Blundell's group (Pullen *et al.*, 1976) has a crystal structure similar to the corresponding portion of the intact protein (Bi *et al.*, 1984). 2D NMR spectra of DPI have been described in aqueous solution at pH 1.8 (Boelens *et al.*, 1990) and in a mixture of water with 20% (v/v) acetic acid (Hua and Weiss, 1990) and are found to be relatively informative. Later it was shown (Hua *et al.*, 1993) that DPI in both of these solutions has a number of molten globule-like properties. Its far-UV CD spectrum at pH 1.8 is similar to those of DPI at pH 7.0 and of native zinc-free insulin, but temperature-induced transitions of DPI both in water and in 20% acetic acid are much less cooperative compared with native proteins. These observations permit Hua *et al.* to suggest that DPI is in the molten globule state in both of these solutions.

2D NMR spectra of DPI in both of these conditions are similar to each other, but are different when compared with those of both native and unfolded proteins. They have practically the full set of NOEs at short and medium distances, which correspond to formation of α and β structure, respectively. However, only 87 long-range NOEs are observed compared with 343 predicted from crystal structure. The interpretation of these spectra (Hua *et al.*, 1992, 1993) implies that α helices and β strands remain practically intact at pH 1.8, but their mutual orientation being native-like fluctuates much more than in the native state. This is exactly what is expected from our model of the molten globule state described in Section II,A,6.

These results provide the best evidence obtained to date that the molten globule has a native-like structural organization differing from the native state mainly by the larger scale of its fluctuations. Is this the case only for this particular protein, or is it also true for other molten globule states which have broad NMR resonances and small dispersion of chemical shifts? The pH-jump technique shows that only part of the native secondary structure is protected in the molten globule states in α -lactalbumin, cytochrome c, and apomyoglobin (see Section II,B,1), and the question is

whether it means that other regions have no unique structural organization or that they have a native-like structure which fluctuates too much to protect the NH groups against hydrogen exchange.

An alternative approach for testing whether the molten globule contains a native-like folding pattern is to study the formation of the native S-S bonds in this state. If the molten globule does not have a native-like folding pattern, the reoxidation of SH groups will give a mixture of different S-S bonds. On the other hand, if the folding pattern of the molten globule state is native-like, native S-S bonds will dominate.

This technique was first applied (Ewbank and Creighton, 1991, 1993; Creighton and Ewbank, 1994) to human α -lactalbumin with one reduced S-S bond (Cys⁶-Cys¹²⁰) and three others (Cys²⁸-Cys¹¹¹, Cys⁶¹-Cys⁷⁷, Cys⁷³-Cys⁹¹, see Fig. 13) intact. It was shown that the Ca^{2+} -free form of this selectively reduced α -lactalbumin (with blocked Cys-6 and Cys-120) has far- and near-UV circular dichroism spectra similar to those of the acid-induced molten globule state. Moreover, this form is compact according to its electrophoretic mobility. The rearrangement of disulfide bonds in this selectively reduced α -lactalbumin gives many different species corresponding to many different sets of S-S bonds. This led Ewbank and Creighton (1991) to the conclusion that "there is no single preferred conformation for the molten globule state of α -lactalbumin in the absence of fixed disulfide bonds . . . indicating that the molten globule state of α -lactalbumin is much closer to an unfolded, but collapsed, form than to an expanded native conformation."

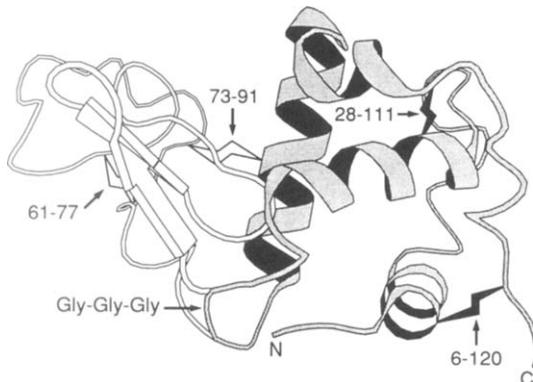


FIG. 13. Schematic representation of α -lactalbumin 3D structure (see Acharya *et al.*, 1991). Recombinant α -helical subdomain with linker of three glycines is shaded. Two disulfide bonds in α -helical subdomain are shown in black, while two others are shown in white. Residues in α subdomain are numbered according their numbering in the intact α -lactalbumin. Reproduced with permission from Peng and Kim (1994). Copyright 1994 American Chemical Society.

Quite different results have been obtained by Peng and Kim (1994) who have constructed a single-chain recombinant model of the α -helical domain of human α -lactalbumin and have shown that two S-S bonds in this domain can be reconstructed in its molten globule state.

Figure 13, which schematically presents the 3D structure of α -lactalbumin (see Acharya *et al.*, 1991), shows the α -helical domain of this protein (residues 1–37 and 86–123), which contains all four of its α helices, and the β -structural domain (residues 38–85), which contains a small antiparallel β sheet and loops. Of the four S-S bonds in α -lactalbumin two (6–120 and 28–111) are in the α -helical domain, one (61–77) is in the β domain, and one (73–91) connects the two domains. The recombinant domain constructed by Peng and Kim consists of residues 1–39 and 81–123 plus a short linker of three glycines. In addition, they have changed Cys-91 to Ala to avoid unwanted thiol-disulfide reactions.

Studies of different physical properties of the Ca^{2+} -free recombinant α domain (at pH 8.5 with 0.5 mM of EDTA) have shown that this domain has all the typical characteristics of the molten globule state. Far-UV circular dichroism spectrum is very pronounced, suggesting a substantial amount of α -helical secondary structure. The near-UV circular dichroism spectrum and ^1H NMR spectrum resemble those of the acid-induced molten globule form of the intact protein and show the absence of well-defined conformations of protein side chains. Dynamic light scattering shows that the recombinant α domain is rather compact. Temperature dependence of the far-UV circular dichroism signal demonstrates no cooperative temperature melting and the fluorescence maximum is intermediate between those for native and unfolded α -lactalbumin, suggesting the partial exposure of tryptophans in the recombinant model.

Figure 14 shows the results of reoxidation of reduced S-S bonds in the recombinant α -helical domain. When disulfide bond formation is allowed to occur in the molten globule state at pH 8.5 and at room temperature ("native conditions"), the native disulfide bonds (6–120; 28–111) strongly predominate: their ratio to two alternative sets (6–28; 111–120) and (6–111; 28–120) is equal to 90:6:4. As a control, when disulfide bonds are reconstructed in the unfolded state (6 M GdmCl) the ratio of these three species is equal to 8:85:7. This is exactly the same ratio as would be expected for an unfolded chain where the probability of the disulfide bond formation would be proportional to $n^{-3/2}$, where n is the number of residues between two cysteines. The dramatic difference between the observed 8% in the random polymer and 90% in the molten globule state clearly shows that the molten globule domain has a substantially native-like folding pattern. Discussing the discrepancy between their results and earlier results of Ewbank and Creighton (1991, 1993)

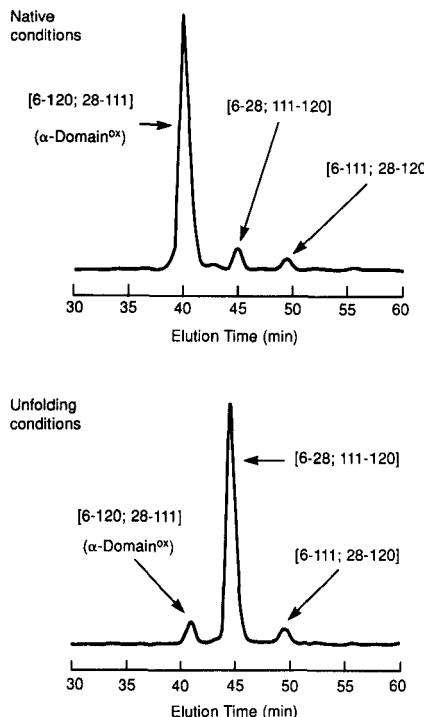


FIG. 14. Disulfide exchange in the recombinant α -helical domain of α -lactalbumin in the oxidized state analyzed by HPLC. Peaks refer to species with disulfide bonds (6–120; 28–111) (the native bonds in oxidized α domain), (6–28; 111–120) and (6–111; 28–220). Reproduced with permission from Peng and Kim (1994). Copyright 1994 American Chemical Society.

Peng and Kim (1994) suggest that the molten globule of α -lactalbumin may have an α -helical domain with a native-like fold but a relatively unstructured β -sheet domain. The authors emphasize that the existence of the native-like folding pattern in the molten globule state confirms my suggestion that the process of protein folding includes an early onset of the native tertiary fold and a late consolidation of side-chain packing.

Less strict but also rather convincing evidence has been obtained for another “classical” molten globule state, carbonic anhydrase, by chemical labeling of SH groups introduced by site-specific mutagenesis (Mårtensson *et al.*, 1993). It was shown that human carbonic anhydrase II is unfolded by GdmCl through the molten globule state and that the accessibility of SH groups introduced into different sites of a molecule correlates with their positions in the native protein. The main result is that the molten globule state has an ordered structure in the central

part of the native β sheet, while the peripheral part of this sheet is less ordered in the molten globule state. In particular, the SH group of cysteine introduced instead of valine-68 in the middle of a large native hydrophobic cluster remains protected from alkylation in the molten globule state and even at higher GdmCl concentrations.

3. Proteins Denatured by Water-Alcohol Mixtures

At the end of this chapter, it is worthwhile to mention some observations on the proteins denatured in water-alcohol mixtures. These proteins usually have NMR spectra with much narrower resonances than those in water solutions and sometimes even have a relatively large dispersion of chemical shifts, which makes these spectra more informative (Harding *et al.*, 1991; Buck *et al.*, 1993; Fan *et al.*, 1993; Stockman *et al.*, 1993; Alexandrescu *et al.*, 1994).

Especially interesting results have been obtained for ubiquitin (Harding *et al.*, 1991; Pan and Briggs, 1992; Stockman *et al.*, 1993). It was shown (Harding *et al.*, 1991) that ubiquitin at pH 2 and in 60% (v/v) methanol is in a "partly denatured" state with a very pronounced far-UV CD spectrum, i.e., a pronounced secondary structure (see also Wilkisson and Mayer, 1986), but with an NMR spectrum suggesting the absence of a rigid tertiary structure. Later it was also shown that this denatured state is nearly as compact as the native one and has no cooperative temperature melting (V. E. Bychkova, A. E. Dujsekina, and E. I. Tiktopulo, unpublished results, 1993).

The native 3D structure of ubiquitin (Vijay-Kumar *et al.*, 1987) includes a five-stranded β sheet covered on one side with an α helix (see Fig. 15a). Harding *et al.* (1991) and Pan and Briggs (1992) identified slowly exchanging NH groups in the alcohol-denatured state using the pH-jump method. These NH groups are distributed over the three main native β strands according to Harding *et al.* and over the whole native β sheet according to Pan and Briggs. In addition, slowly exchanged NH protons embrace the whole native α helix (Harding *et al.*, 1991; Pan and Briggs, 1992) and even the small β_{10} helix near the C end of the chain (Pan and Briggs, 1992).

Unusually narrow lines and relatively large dispersion of chemical shifts in NMR spectrum of ubiquitin denatured by pH 2 and in 60% methanol permit the partial assignments of the two-dimensional $^1\text{H}-^1\text{H}$ spectrum (Harding *et al.*, 1991), which provides more direct structural information. Partial assignment of resonances led to the determination of a number of NOEs between C^αH protons belonging to different β strands. Three such pairs, Gln-2/Gln-16, Phe-4/Thr-14, and Lys-6/Thr-12, presented good evidence that β strands 1 and 2 form a hairpin

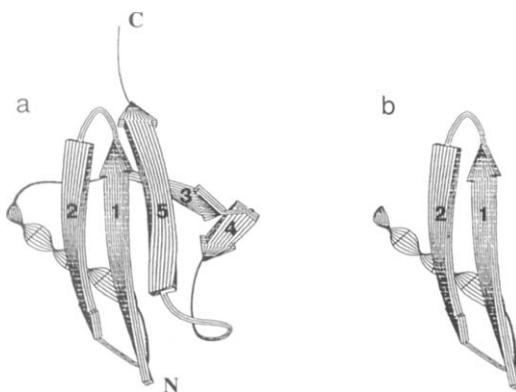


FIG. 15. (a) Schematic presentation of the native 3D structure of ubiquitin. The structure includes α helix (residues 23–34), five-stranded β sheet with β strands 1 (residue 1–7), 2 (11–17), 3 (41–45), 4 (48–50), 5 (64–70), and a number of reverse turns. (b) The part of the ubiquitin structure which is shown to be preserved in the alcohol-denatured state by 3D NMR data. Adapted with kind permission from Stockman *et al.* (1993). © 1993 ESCOM Science Publishers B.V.

in the alcohol-denatured state and even the mutual positions of residues in this hairpin are the same as those in the native state. In addition, cross-peaks between NH protons in the chain region corresponding to the native α helix (residues 23–34) show that this region also preserves its helical structure in the alcohol-denatured state (although in the denatured state the α helix is substantially more flexible).

Both of these conclusions were fully supported by subsequent 3D NMR studies on uniformly enriched [^{15}N]ubiquitin (Stockman *et al.*, 1993). Based on nearly complete assignment of ^1H and ^{15}N resonances, they observed a number of NOEs between ^{15}NH groups as well as between ^{15}NH and C^αH groups. These results also show the preservation of the native hairpin from β strands 1 and 2 (see Fig. 15b) as well as of the native α helix. In addition, a well-defined reverse turn 8–11 identical with that in the native state has been observed in the alcohol-denatured state. It follows that the whole N-terminal half of ubiquitin (residues 1–34) has a secondary structure which is very close to the native one. Although there is little or no evidence for the native-like mutual positions of the α helix and β sheet, it is likely that the N-terminal half of alcohol-denatured ubiquitin has a structure very similar to, though less stable than, that in the native state (see Fig. 15).

Little is known about the structure of the C-terminal half of ubiquitin in the alcohol-denatured state. 3D NMR (Stockman *et al.*, 1993) did not support the earlier suggestion (Harding *et al.*, 1991) that residues 64–70

forming a β strand in the native structure preserve this conformation in the denatured state. Moreover, the chemical shifts and sequential NOEs obtained for this region by 3D NMR in the denatured state are not typical for β structure and are more consistent with a very weak "transient" α helix. The same is true for other parts of the C-terminal half of the molecule. However, it is not obvious how these conclusions correlate with the close similarity between the profiles of protection factors for the native and the alcohol-denatured states (Pan and Briggs, 1992).

Some other proteins, lysozyme (Buck *et al.*, 1993) monellin (Fan *et al.*, 1993), and α -lactalbumins (Alexandrescu *et al.*, 1994), have also been studied by NMR in water-alcohol solvent mixtures. As in the case of ubiquitin, the lines of the NMR spectra are narrow, and sometimes the assignments of a number of resonances become possible. These assignments permit one to conclude that organic solvents can induce α -helical structure in some chain regions which were nonhelical in water solutions. For example, two α helices have so far been identified by 2D NMR analysis of α -lactalbumins; one of them (residues 86–90) virtually coincides with the native helix C, while the other one (residues 31–41) includes in the native state a linker between helix B (25–34) and the strand 3 (41–43) of its β sheet (Alexandrescu *et al.*, 1994). In a similar way the first strand of the β sheet in monellin (residues 17–30 of its A chain) transforms into a helical region in water-alcohol mixtures (Fan *et al.*, 1993).

Far-UV circular dichroism spectra in all these proteins are much more pronounced in water-alcohol mixtures than in the native state. The compactness of these proteins in water-alcohol mixtures has not yet been evaluated.

We conclude that, although NMR spectra of a number of proteins can be more informative in water-alcohol mixtures than in water solutions, the obtained structural information should be treated with caution as it is not known to what extent the 3D structure of proteins in these mixtures is similar to that of the "classical" molten globule state in water solutions.

III. PHASE TRANSITIONS

A. Denaturation

1. Temperature Denaturation

Even many years ago it was assumed (Anson, 1945) that protein denaturation is a highly cooperative process which embraces the protein

molecule as a whole (all-or-none transition). This point of view was supported by coincidental changes in different parameters for a number of proteins on denaturation, but concrete evidence was obtained much later. This evidence was based on a thermodynamic description of the denaturation process and on its comparison with what we expect for all-or-none transitions.

Thermodynamic description of all-or-none (two-state) transitions (Anson and Mirsky, 1934; Eisenberg and Schwert, 1951) is simple and straightforward. If Θ and $1 - \Theta$ are fractions of denatured and native molecules, respectively, the equilibrium constant between these two states is

$$K = \frac{\Theta}{1 - \Theta} = e^{-\Delta G/RT}, \quad (1)$$

where ΔG is free energy difference between denatured and native states, R is the gas constant, and T is the temperature. Denaturation induced by temperature, solvent composition, etc., can be described by the corresponding term for the energy, e.g., of temperature-induced denaturation

$$\frac{\partial \Theta}{\partial T} = \Theta(1 - \Theta) \Delta H/RT^2 \quad (2)$$

(ΔH being the enthalpy difference between denatured and native states). It follows from Eq. (2) that

$$\Delta H = 4RT_m^2 \left(\frac{\partial \Theta}{\partial T} \right)_{T_m}, \quad (3)$$

where T_m is the midpoint of a temperature transition, where $\Theta = (1 - \Theta) = 1/2$.

For all-or-none transition the fraction of denatured molecules can be determined from the change of any parameter X ,

$$\Theta = \frac{X - X_N}{X_D - X_N}, \quad (4)$$

where X_N and X_D are the values of X for native and denatured molecules, respectively. Equations (3) and (4) together permit the evaluation of the denaturation enthalpy ΔH from the slope of the temperature dependence of any parameter. This enthalpy has been referred to as the van't Hoff enthalpy of denaturation, which corresponds to the enthalpy of a "cooperative unit," i.e., the part of the material which undergoes a temperature transition as a whole (Privalov, 1979).

Of course, we cannot judge the all-or-none character of denaturation just by measuring a van't Hoff enthalpy (see, e.g., Lumry *et al.*, 1966). An important finding was that the answer can be obtained by a comparison of the van't Hoff enthalpy with the enthalpy measured directly by microcalorimetry of protein solutions (Privalov, 1979). It was shown (Privalov, 1979, 1992) that temperature denaturation of small proteins is accompanied by a single peak of heat absorption (see Fig. 16). The actual value of the denaturation enthalpy per protein molecule can easily be obtained from the area under this peak. For small proteins the van't Hoff values of denaturation enthalpy always coincide with their "real" (calorimetical) values. This means that temperature denaturation of small globular proteins is an all-or-none transition which embraces a molecule as a whole (Privalov, 1979; see also Freire and Biltonen, 1978).

The all-or-none character of the temperature melting of native proteins does not necessarily mean the absence of noncooperative pre- and post-transitional changes of proteins. These gradual transformations can change baselines of heat capacity (see Fig. 16), but may not be cooperative enough to alter the peak of the heat absorption.

An important question concerns the physical state of protein molecules at high temperatures. As early as 1967 it was shown that high temperatures do not lead to a complete unfolding of proteins. There are two

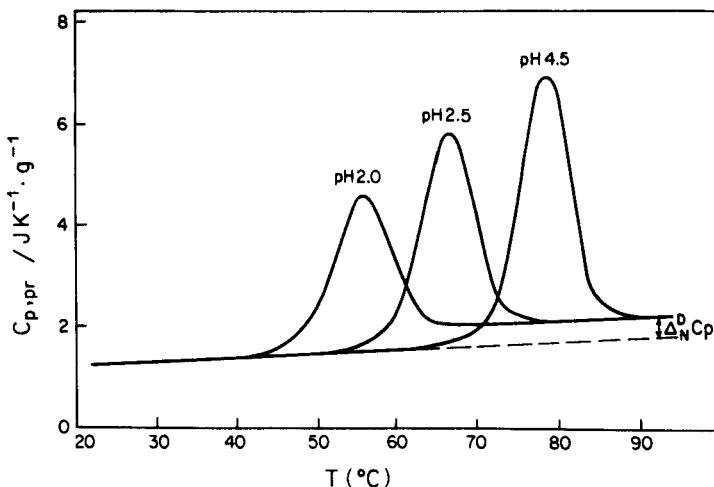


FIG. 16. Temperature dependence of the heat capacity of a typical small globular protein (hen egg white lysozyme) in aqueous solutions at different pH values. $\Delta_N^D Cp$ means a difference between the heat capacity of the temperature-denatured and native state. From Privalov (1992). In "Protein Folding" (T. E. Creighton, ed.). Copyright © 1992 by W. H. Freeman and Company. Reprinted with permission.

main arguments in favor of this conclusion: (1) temperature-denatured proteins can undergo an additional cooperative unfolding induced by GdmCl or urea (Aune *et al.*, 1967; Brandts and Hunt, 1967), and (2) intrinsic viscosities of proteins usually are substantially less at high temperatures than at high concentrations of urea or GdmCl (see, e.g., Hamaguchi and Sakai, 1965). These arguments have been considered in the well-known review by Tanford (1968) and have led him to the definite conclusion that the temperature-denatured state of protein molecules is far from being completely unfolded.

This conclusion has been questioned by Privalov (1979), but now it is supported by many other experiments. The most striking examples are temperature-denatured bovine and human α -lactalbumins. Temperature-denatured bovine α -lactalbumin has an intrinsic viscosity of $3.0 \text{ cm}^3/\text{g}$ (see Fig. 3a) and a very pronounced far-UV circular dichroism spectrum (Dolgikh *et al.*, 1981, 1985), i.e., properties typical for the molten globule state. Similar results have also been obtained for human α -lactalbumin. Figure 17a presents data on Stokes radii for this protein (in Ca^{2+} -free form) measured by quasielastic light scattering at different temperatures for neutral and acid pH. Figure 17a shows that the heating at neutral pH leads only to a very small expansion of protein similar to that obtained at acid pH at room temperature. Large-angle diffuse X-ray scattering shows that a protein core is preserved in the temperature-denatured state, being only slightly less tightly packed than that in the native proteins (Ptitsyn *et al.*, 1986; see also Ptitsyn, 1987). The far-UV circular dichroism spectrum of this protein is very pronounced even at 80°C , while its near-UV spectrum is practically absent at this temperature (see Fig. 17b).

The point is that the temperature denaturation of both bovine (Dolgikh *et al.*, 1981, 1985) and human (Pfeil *et al.*, 1986) α -lactalbumins is accompanied, as for all native proteins, by a cooperative heat absorption (see Fig. 17c), although in this case the thermal denaturation does not lead to a substantial unfolding of the protein chain or to a substantial decrease in its secondary structure.

Perhaps α -lactalbumins present an extreme case of temperature-denatured proteins with the properties almost indistinguishable from those of the acid-denatured molten globule state. However, the data available for other proteins also support the idea that these proteins are far from being completely unfolded in the temperature-denatured state. For example, the far-UV CD spectrum of temperature-denatured lysozyme (at 69°C) is more similar to the native than to the unfolded state (Kuwajima *et al.*, 1985) and its linear dimensions exceed the native state only by 16% (Hamaguchi and Sakai, 1965; Kugimiya and Bigelow, 1973)

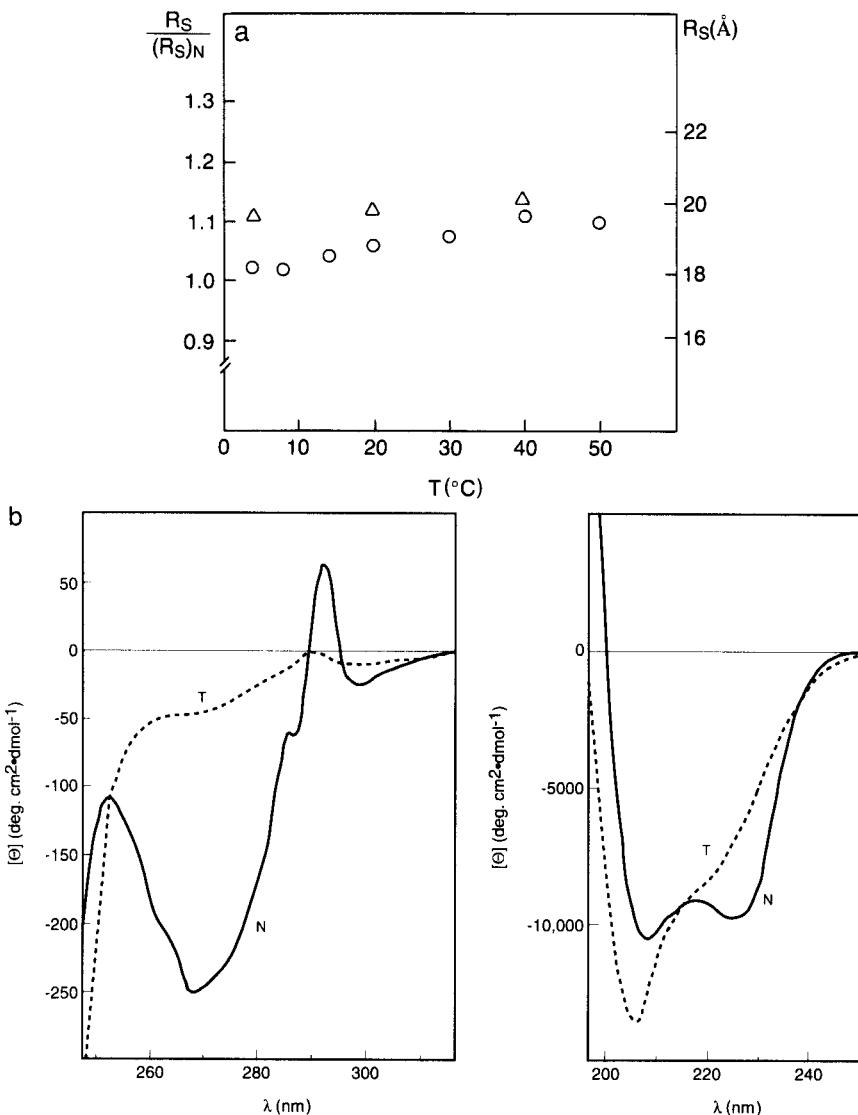


FIG. 17. (a) Temperature dependence of the Stokes radius, R_s , and its ratio to that of the native protein $R_s / (R_s)_N$ for human α -lactalbumin at pH 2.0 (Δ) and pH 7.5 (\circ) according to the data of Gast *et al.* (1986). Ca^{2+} -free form of the protein has been used to decrease its thermostability. Reproduced with permission from Ptitsyn (1987). (b) Near (left)- and far (right)-UV CD spectra of human α -lactalbumin at pH 7.0: N, the native state (0°C), T, the temperature-denatured state (80°C). Unpublished data of S. Yu. Venyaminov and V. E. Bychkova, 1983. (c) Temperature dependence of heat capacity $C_p(T)$ for the native (pH 7) state (solid line) and the molten globule state (pH 2) (dotted line) of human α -lactalbumin. Reproduced with permission from Pfeil *et al.* (1986).

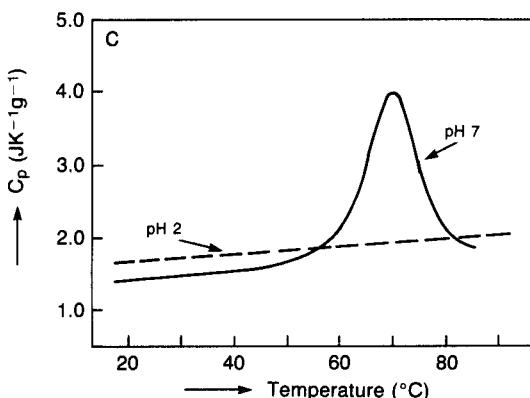


FIG. 17. (continued)

which coincides with a value typical for the molten globule state (see Section II,A,2).

In a similar way the linear dimensions of ribonuclease A in the temperature-denatured state obtained from intrinsic viscosity (Holcomb and van Holde, 1962), dynamic light scattering (Nicoli and Benedek, 1976), and size-exclusion chromatography (Corbett and Roche, 1984) exceed those of the native protein only by 7–20% (compared with about 50% in the unfolded state with intact S-S bonds). Raman (Chen and Lord, 1976) and far-UV CD (Labhardt, 1982; Privalov *et al.*, 1989; Robertson and Baldwin, 1991) spectra show the presence of a substantial amount of a secondary structure in the spectra of temperature-denatured ribonuclease A. An important observation that two of four histidines of temperature-denatured ribonuclease have nonrandom chemical shifts in NMR spectrum (Matthews and Westmoreland, 1975) emphasizes the existence of a residual structure in this protein.

More recently the temperature-denatured state of ribonuclease A has been studied by diffuse X-ray scattering and Fourier transform infrared spectroscopy (Sosnik and Trehewella, 1992). Sosnik and Trehewella (1992) have also found a limited increase in the radius of gyration in the temperature-denatured state (Fig. 18a). A pair distribution function of this state is qualitatively similar to that of the native state although it reflects the expansion of molecules that increases the longest distances in a molecule from 40–45 to 60–65 Å (Fig. 18b). Fourier transform infrared spectrum of the temperature-denatured state differs from that of the native state by a smaller intensity of the band at 1633 cm⁻¹ which is attributed to the β structure. However, the substantial residual intensity

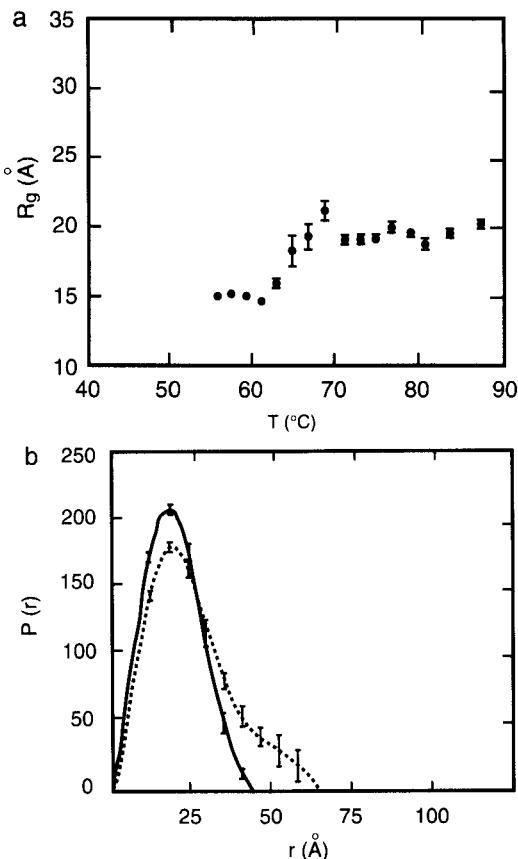


FIG. 18. (a) Temperature dependence of the radius of gyration, R_g , of ribonuclease A with intact S-S bonds at neutral pH and low ionic strength. (b) Pair-distribution function for ribonuclease A with intact S-S bonds at 57 (solid line) and 67°C (dotted line). (c) Fourier transform infrared spectra of ribonuclease A at 54 (solid line) and 66°C (dashed line). Reproduced with permission from Sosnick and Trewella (1992). Copyright 1992 American Chemical Society.

of this peak and the broadness of the rest of the spectrum (Fig. 18c) suggest that a large part of the secondary structure is preserved in the temperature-denatured state (see also Seshardi *et al.*, 1994).

All these data leave little doubt that the temperature-denatured proteins may be relatively compact and may possess a substantial secondary structure. Sometimes their properties are very similar to those of the molten globule state, in other cases they may be intermediate between the molten globule and the unfolded state, i.e., something like the partly folded or premolten globule state (Uversky and Ptitsyn, 1994, 1995; see Section V,A,2). On the other hand, it is possible that some proteins are

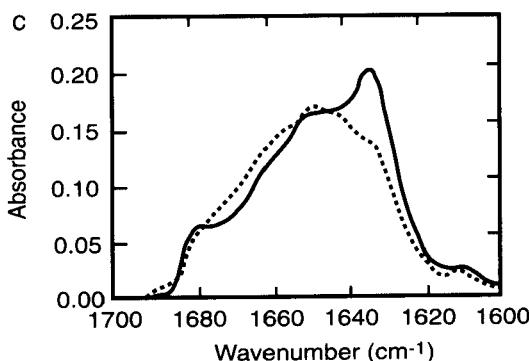


FIG. 18. (continued)

more unfolded at high temperature. For example, a Fourier transform infrared spectroscopy study of ribonuclease T1 leads to the conclusion that its thermally induced state is predominantly (though not completely) irregular (Fabian *et al.*, 1993).

2. Transition between Native and Unfolded States

It should be emphasized that the discussions on the physical state of temperature-denatured proteins do not diminish the importance of the discovery (Privalov, 1979) that temperature melting is an all-or-none transition, i.e., an intramolecular analog of a first-order phase transition. The fact that the native protein structure is destroyed according to the all-or-none scheme by at least one of the denaturing agents (temperature) suggests that it may also be the case for other denaturing agents and may not depend on the properties of a denatured state.

It is especially interesting to check this assumption for strong denaturing agents, like urea or GdmCl, which transform protein molecules into a more unfolded state compared with other types of denaturation. The all-or-none character of these transitions has been postulated by Tanford (1968), reasoned by Privalov (1979), and has recently been confirmed by direct experiments.

The answer to whether a given transition is all-or-none can in principle be obtained by an approach similar to that used for temperature-induced transitions. For the all-or-none transitions the dependence of the fraction of denatured molecules on the activity of a denaturing agent a is described by an equation equivalent to Eq. (2):

$$\frac{\partial \Theta}{\partial a} = \Theta (1-\Theta) \frac{\Delta \nu}{a}. \quad (5)$$

If follows from Eq. (5) that

$$\Delta v(a_m) = 4a_t \left(\frac{\partial \Theta}{\partial a} \right)_{a_m}, \quad (6)$$

where a_m is the activity of denaturant at the middle point of the solvent-induced transition where $\Theta = (1 - \Theta) = 1/2$. Here, Δv is the number of molecules of denaturing agent which must be added (or removed) to keep its activity constant following the all-or-none transition in one protein molecule (Ptitsyn and Uversky, 1994). For a simplified model in which all denaturant molecules are specified as "free" or "bound" to protein molecules (Aune and Tanford, 1969), Δv is the difference between the numbers of these molecules bound to one molecule of a protein in its two states. Equation (6), originally proposed for this model by Aune and Tanford (1969), has been shown (A. V. Finkelstein, unpublished results) to be valid for the general case at small protein concentrations. Δv determined by Eq. (6) is formally equivalent to the van't Hoff enthalpy of temperature-induced denaturation which is determined by Eq. (3), i.e., refer to the "cooperative unit" that undergoes denaturation as a whole. However, in the case of a urea- or GdmCl-induced transition it is practically impossible to measure the equivalent of calorimetric enthalpy, i.e., Δv per one protein molecule. Therefore, the approach proposed by Privalov (1979) for temperature-induced denaturation cannot be applied to this case.

Nevertheless, there are two direct lines of evidence that solvent-induced transitions have an all-or-none character. The first line of evidence (Ptitsyn and Uversky, 1994) is based on general principles of statistical physics (Hill, 1968), according to which the slope of an all-or-none transition in a small system must be proportional to the number of particles in this system. Applying this principle to all-or-none transitions in proteins (or other macromolecules) we conclude that the slopes of these transitions must be proportional to a protein molecular weight. On the other hand, if a transition embraces only a part of a macromolecule (a cooperative unit or domain), it will not "feel" the whole macromolecule and therefore its slope will not depend on molecular weight. It follows that Δv will increase with protein molecular weight only when the whole molecule is a cooperative unit, i.e., when urea- or GdmCl-induced transitions belong to the all-or-none type.

Figure 19 (Ptitsyn and Uversky, 1994) shows that Δv for the native-unfolded state transition is proportional to the molecular weight for all 60 proteins studied to date. The data have been collected from the

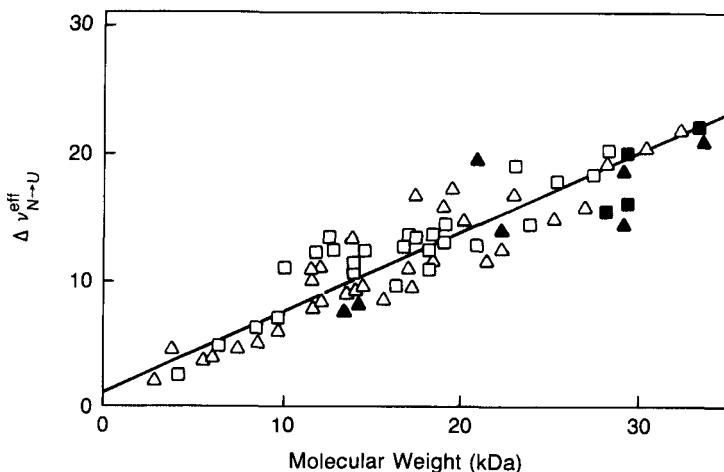


FIG. 19. Molecular weight dependence of the slopes of urea (\square , \blacksquare)- and GdmCl (\triangle , \blacktriangle)-induced transitions between the native and unfolded states in relatively small proteins. The open symbols refer to the proteins which undergo two-state solvent-induced transition. Filled symbols refer to proteins which undergo three-state transition (through the molten globule state). In the last case $\Delta v_{N \rightarrow U}$ were calculated as sums of Δv 's for $N \rightarrow MG$ and $MG \rightarrow U$ transitions ($\Delta v_{N \rightarrow U} = \Delta v_{N \rightarrow MG} + \Delta v_{MG \rightarrow U}$) (see Figs. 22 and 23). Adapted with permission from Ptitsyn and Uversky (1994).

literature for relatively small proteins (to avoid complications related to the domain structure of large proteins; see Section III,A,4) that have only one urea- or GdmCl-induced transition from the native to the unfolded state. The data can be approximated by

$$\Delta v_{N \rightarrow U} = 0.63 M + 1.59, \quad (7)$$

where M is a molecular mass (in kDa). It means that urea- and GdmCl-induced transitions between the native and unfolded states belong to the all-or-none type in all these proteins.

The second direct line of evidence for all-or-none transitions is the bimodal distribution of molecules in the transition region, which clearly shows the absence of molecules with intermediate properties. These data have been obtained using size-exclusion chromatography, which for slow all-or-none transitions permits the resolution of elution peaks for folded and unfolded molecules (Corbett and Roche, 1984; Uversky, 1993). This is illustrated by Fig. 20 which presents the data on urea-induced denaturation of myoglobin (Corbett and Roche, 1984). The striking feature of Fig. 20 is the bimodal distribution of elution volumes:

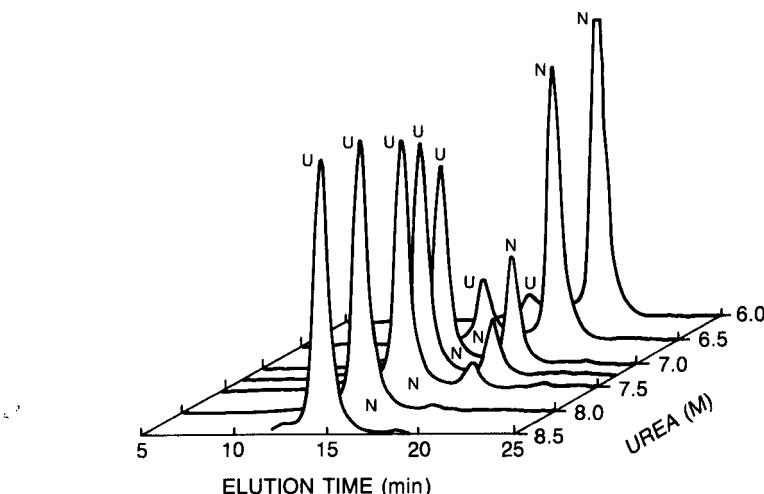


FIG. 20. Urea-induced unfolding of sperm whale myoglobin at 25°C studied by size-exclusion chromatography (HPLC). Peaks N and U refer to the native and unfolded states, respectively. Adapted with permission from Corbett and Roche (1984). Copyright 1984 American Chemical Society.

at small concentrations of urea only a peak for folded molecules is present, in the transition region a peak for unfolded molecules appears and increases at the expense of the first peak, and at large urea concentrations only the second peak remains. Similar results have also been obtained for bovine serum albumin (Withka *et al.*, 1987), staphylococcal nuclease (Shalongo *et al.*, 1992), and egg white lysozyme (Uversky, 1993).

All these data leave no doubt that solvent-induced denaturation of relatively small proteins encompasses the protein molecule as a whole without significant amounts of molecules which are composed individually of partly folded and unfolded regions.

Bimodal distributions even for an all-or-none transition could not be obtained if this transition were fast compared with the characteristic time of size-exclusion chromatography (15–30 min). For example, temperature-induced denaturation of ribonuclease does not show a bimodal distribution, but rather is described with a single elution peak shifted to larger molecular dimensions with the increase in temperature (Corbett and Roche, 1984). The reason for this behavior is the acceleration of this transition at high temperatures; as a result size-exclusion chromatography “sees” only elution volume averaged over two discrete states of a protein.

It is worthwhile to discuss to what extent proteins are unfolded at high concentrations of strong denaturing agents like urea or GdmCl.

Far-UV CD spectra of these proteins usually show no visible traces of secondary structure (see, e.g., Figs. 5a and 12) and are considered as reference spectra for unfolded chains. The data on intrinsic viscosities (Tanford, 1968) of these proteins at room temperature also do not show any visible deviations from the behavior typical for random coils. The Stokes radii of 15 proteins with molecular masses from 3 to 200 kDa in 6 M GdmCl are described by $R_s \sim M^{0.55}$, which is typical for random coils. Therefore, it is widely accepted that proteins are completely or almost completely unfolded at high concentrations of strong denaturants.

On the other hand, the dimensions of protein molecules in these solutions dramatically decrease with the increase of temperature. Ahmad and Salahuddin (1974) have studied this effect for eight proteins without S-S bonds or with reduced S-S bonds and have observed that their intrinsic viscosities in 6 M GdmCl decrease two- or three-fold if the temperature increases from 25 to 55°C. This decrease is much larger than the normal temperature dependence of molecular dimensions for random coils (Flory, 1953).

Even more important is the fact that the molecular weight dependence of protein dimensions is inconsistent with the random coil model at all temperatures except 25°C. The routine method for separating the influence of short- and long-range interactions on the dimensions of random coils is the plot of $[\eta]/M^{1/2}$ versus $M^{1/2}$ (Stockmayer and Fixman, 1963). It is known that the intrinsic viscosity unperturbed by long-range interaction is proportional to $M^{1/2}$ and the influence of long-range interactions can be approximated by another term which is roughly proportional to M . Therefore, the intercept of the plot of $[\eta]/M^{1/2}$ versus $M^{1/2}$ for random coil is equal to $[\eta]_0/M^{1/2}$ ($[\eta]_0$ is intrinsic viscosity at the Θ point) and depends only on short-range interactions, while the slope of this plot is proportional to a second virial coefficient and is a measure of long-range interactions.

Figure 21a presents these plots for eight proteins studied by Ahmad and Salahuddin (1974) in 6 M GdmCl at 25, 30, and 55°C. Only the plot at 25°C has an intercept which is expected for the random coils with a reasonable persistence length of 13 Å. Already at 30°C the intercept becomes almost five times smaller, which corresponds to a nonrealistic value of 5 Å for the persistence length, while above 45°C the intercept becomes negative. This behavior is absolutely inconsistent with that expected for random coil and is very strong evidence for the presence of some structure in proteins, even in 6 M GdmCl, at temperatures above 25°C. One can see (Fig. 21b) that the intercept of plots $[\eta]/M^{1/2}$ versus $M^{1/2}$ sharply decreases when temperature increases by only 5°C (from

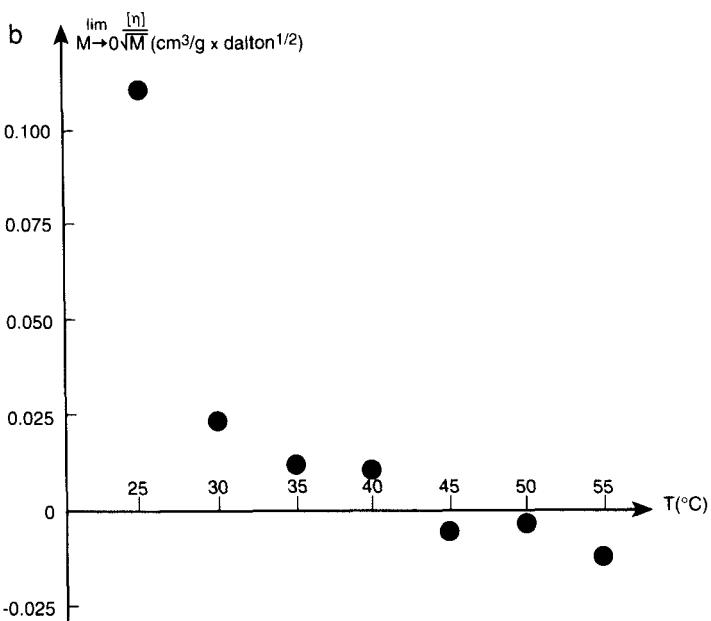
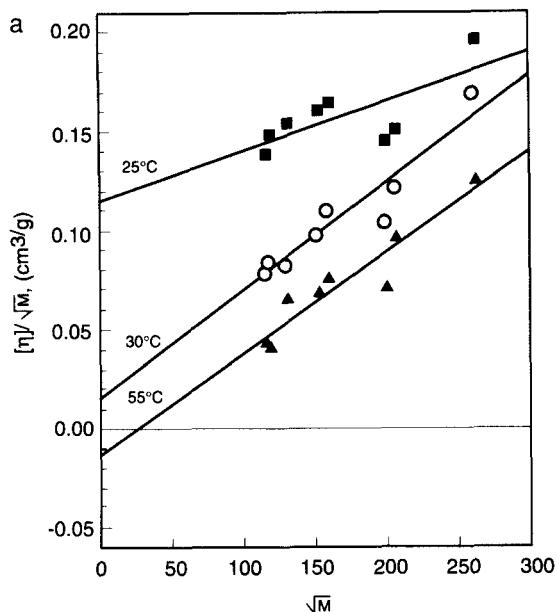


FIG. 21. (a) Plots of $[\eta]/M^{1/2}$ versus $M^{1/2}$ obtained from the data on intrinsic viscosities $[\eta]$ of eight proteins without disulfide bonds in 6 M GdmCl at 25 (■), 30 (○), and 55°C (▲) (Ahmad and Salahuddin, 1974). The plots for 35, 40, 45, and 50°C are between those for 30 and 50°C (I. N. Serdyuk and O. B. Ptitsyn, unpublished data). (b) Temperature dependence of the intercepts of plots of $[\eta]/M^{1/2}$ versus $M^{1/2}$. For random coils these intercepts are equal to $[\eta]_{\Theta}/M^{1/2}$ where $[\eta]_{\Theta}$ is intrinsic viscosity in Θ -solvent. Usually, plots of $[\eta]/M^{1/2}$ versus $M^{1/2}$ for random coils in solvents of different quality have different slopes but the same intercepts.

25 to 30°C) and then becomes much less dependent on temperature. It may mean a cooperative formation of some structure in proteins just below 30°C and it is not yet clear at what temperature this transition begins. The possibility cannot be ruled out that proteins may be even more unfolded at temperatures below 25°C.

NMR data provide direct evidence for residual structure in a protein denatured by a high concentration of urea. Wüthrich and collaborators have studied the amino-terminal 63-residue domain of the 434 repressor by 2D NMR in 7 M urea and have shown that segment 54–59 of this protein forms a hydrophobic cluster containing the side chains of Val-54, Val-56, Trp-58, and Trp-59 (Neri *et al.*, 1992). This cluster does not coincide with the native one, but is related to it by simple rearrangements of the residues 58–60. Other studies showed that a hydrophobic cluster in human carbonic anhydrase II exists at very high concentrations of GdmCl, probably up to 8 M (Mårtensson *et al.*, 1993).

All these data suggest that, although in the first approximation strong denaturants unfold proteins, the extent of this unfolding still requires further study.

3. Transition between Native and Molten Globule States

Protein denaturation traditionally was considered to be accompanied by substantial (Tanford, 1968) or even complete (Privalov, 1979) unfolding of protein chains. However, we have shown (Dolgikh *et al.*, 1981, 1985) that bovine and human α -lactalbumins can be transformed into a compact denatured state with pronounced secondary structure, i.e., the molten globule state (see Section II). It was shown that this transition can be produced not only by low pH or moderate concentrations of GdmCl, but also by high temperatures (see Section III,A,1). An important feature of the temperature denaturation of α -lactalbumins is that their van't Hoff enthalpies are equal to the calorimetric enthalpies, i.e., their temperature denaturation conforms to an all-or-none process. This was the first demonstration of the all-or-none character of the transition between the native and the molten globule state.

A more general case of native–molten globule state transitions is urea- or GdmCl-induced denaturation. In a number of proteins urea- or GdmCl-induced denaturation is decoupled from unfolding, i.e., protein unfolds through the intermediate molten globule state (see Sections II,A,1 and V,A,2). As in the case of native-unfolded state transitions, the all-or-none character of native–molten globule and molten globule–unfolded state transitions can be checked by the study of molecular weight dependences of their slopes. Figure 22 shows the plot of $\Delta\nu$ for $N \rightarrow MG$ (where MG is the molten globule state) transition ($\Delta\nu_{N \rightarrow MG}$)

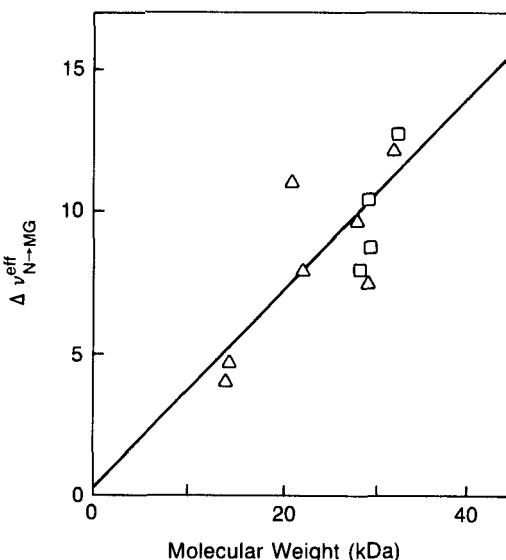


FIG. 22. Molecular weight dependence of the slopes of urea (□)- and GdmCl (Δ)-induced transitions between the native and the molten globule states. The data have been collected for proteins undergoing two solvent-induced stages of unfolding (from the native state to the molten globule state and from the molten globule state to the unfolded state). Transitions were monitored by near-UV circular dichroism, ^1H NMR and activity for bovine and human α -lactalbumin, bovine carbonic anhydrase B, human carbonic anhydrase II, bovine and reduced human growth hormones, *Staphylococcus aureus* and *Bacillus cereus* β -lactamases, and rhodanese. Adapted with permission from Ptitsyn and Uversky (1994).

against the molecular weight of proteins (Ptitsyn and Uversky, 1994). Despite the much smaller number of points compared with the $\text{N} \rightarrow \text{U}$ (where U is the unfolded state) transition (Fig. 19), one can see that $\Delta v_{\text{N} \rightarrow \text{MG}}$ again is roughly proportional to the protein molecular weight and it described well by

$$\Delta v_{\text{N} \rightarrow \text{MG}} = 0.35 M + 0.19. \quad (8)$$

Equation 8 supports the conclusion that the transition between the native and molten globule state in small proteins is of the all-or-none type.

4. Denaturation of Large Proteins

All data discussed above refer to small proteins in which denaturation is of the all-or-none character. On the other hand, temperature denaturation of large proteins does not follow the all-or-none scheme, as their van't Hoff enthalpy is half or even less than half of the enthalpy measured

calorimetrically (Privalov, 1982). It appears that these proteins consist of two or more cooperative units or melting domains and that the melting of these proteins goes through stage(s) in which each protein molecule consists of nonmelted and melted parts.

A similar picture was also observed for solvent-induced denaturation (O. B. Ptitsyn and V. N. Uversky, unpublished data, 1994). It was shown that at molecular masses larger than 30 kDa, $\Delta\nu$ for the N → U transition fails to be proportional to molecular weight and points begin scatter around $\Delta\nu$ values between 10 and 20. This shows the absence of the all-or-none transitions in large proteins, which appear to be melted as loosely associated molecular units or domains.

In many cases the all-or-none character of denaturation of large proteins reflects their domain structure, i.e., the fact that they consist of two or more structural "lobes" with much stronger interactions within each lobe than between them. In these cases heat capacity curves often have a complex profile suggesting the relatively independent melting of these lobes, which being somewhat isolated undergo all-or-none transitions. Fibrinogen (Privalov and Medved', 1982) and plasminogen (Novokhatny *et al.*, 1984) can serve as good examples of such behavior. However, some other proteins, e.g., papain (Tiktopulo and Privalov, 1978), pepsinogen (Mateo and Privalov, 1981), and retinol-binding protein (Bychkova *et al.*, 1992), melt by a single temperature transition with the profile typical for small proteins, but their van't Hoff enthalpies are half or even less than the calorimetric values. Formally, these proteins also can be considered to consist of two or more cooperative units or melting domains, but in this case these units or domains may have no clear structural meaning.

Moreover, we have described a similar behavior in homopolymers which undergo temperature-induced coil–globule transitions in water solutions (see, e.g., Meewes *et al.*, 1991; Anufrieva *et al.*, 1991). It has been shown (Tiktopulo *et al.*, 1994, 1995) that the van't Hoff enthalpy ΔH_{vH} of the globule–coil transition in these polymers does not depend on their molecular weight, while their calorimetric enthalpy, ΔH_{cal} , is proportional to the molecular weight. As a result, only for short chains is $\Delta H_{vH} \approx \Delta H_{cal}$, whereas for longer chains ΔH_{vH} is much smaller than ΔH_{cal} . Thus, the domain melting can also occur in homopolymers in which these domains cannot be considered as definite structural units.

B. Unfolding

1. Transition between Molten Globule and Unfolded States

The results presented in Section III,A show that denaturation of native proteins (monitored by the loss of their activity and their rigid tertiary

structure) is of the all-or-none character irrespective of the degree of molecular unfolding and the decay of its secondary structure. This very important feature of protein denaturation strongly contradicts former ideas connecting the all-or-none denaturation of proteins with the globule-coil transition in protein chains or with a cooperative loss of its secondary structure (see Chan and Dill, 1991; Karplus and Shakhnovich, 1992; Ptitsyn, 1992a, for recent reviews). The elegant theory of Shakhnovich and Finkelstein (1982, 1989; see also Karplus and Shakhnovich, 1992) explains this general feature of protein denaturation by the cooperative loss of tight packing of protein side chains (see Section III,C,3).

Quite unexpected was the direct evidence for the existence of a second all-or-none transition in globular proteins: the transition between two denatured states, the molten globule and the unfolded. The first argument in favor of this conclusion was that urea- and GdmCl-induced unfolding of carbonic anhydrases B shows the existence of only one intermediate, the molten globule state (Rodionova *et al.*, 1989; see also Ptitsyn, 1992a). In a similar way, acid- and urea-induced unfolding of sperm whale apomyoglobin (Barrick and Baldwin, 1993) and cytochrome c (Goto *et al.*, 1993) can be quantitatively described in terms of only three states: the native, the unfolded, and the intermediate (molten globule) states.

Strict evidence for the all-or-none molten globule-unfolded state transitions has been obtained by the same two approaches that were applied to native-unfolded state transitions—the molecular weight dependence of a cooperativity (Ptitsyn and Uversky, 1994) and the bimodal distribution in the transition region (Uversky *et al.*, 1992).

Figure 23 shows the molecular weight dependence of the slopes of molten globule-unfolded state transitions (Ptitsyn and Uversky, 1994). The data were collected from the literature both for proteins, which have two urea- or GdmCl-induced transitions from the native to the unfolded state (see above), and for proteins unfolded by urea or GdmCl from their acid-induced molten globule state. Figure 23 shows that all these data also fit linear molecular weight dependence:

$$\Delta\nu_{MG \rightarrow U} = 0.26 M + 1.05, \quad (9)$$

which demonstrates the all-or-none character of molten globule-unfolded state transitions in all proteins studied so far.

It is impossible to explain the all-or-none character of the $MG \rightleftharpoons U$ transition assuming that the molten globule consists of the native and the unstructured parts, one of them has been already melted in the $N \rightarrow MG$ transition, while another melts in the $MG \rightarrow U$ transition. In

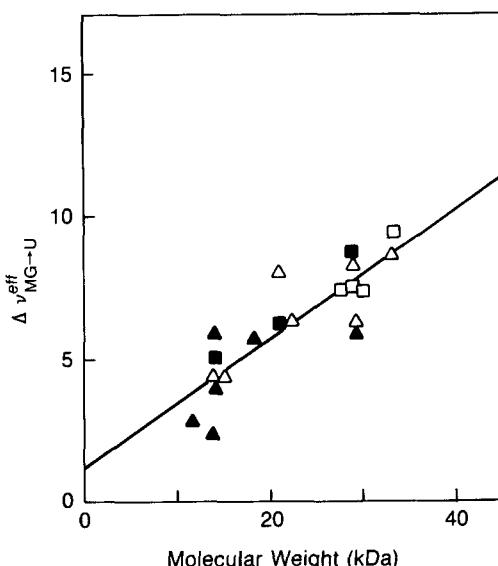


FIG. 23. Molecular weight dependence of the slopes of urea (□, ■)- and GdmCl (Δ, ▲)-induced transitions between the molten globule and the unfolded states. The data have been collected both for proteins undergoing two solvent-induced stages of unfolding (see legend to Fig. 22) and for proteins unfolded by urea or GdmCl from their acid-induced molten globule states (the latter proteins are marked by filled symbols). Transitions were monitored by far-UV circular dichroism, fluorescence, ¹H NMR, absorbance, chromatography and viscosity for bovine and human α -lactalbumins, bovine and human carbonic anhydrases B, bovine and reduced human growth hormones, *Staphylococcus aureus* and *Bacillus cereus* β -lactamases, rhodanese, carp parvalbumin, ribonuclease A, and β -lactoglobulin. Adapted with permission from Ptitsyn and Uversky (1994).

this case, all properties sensitive to tight packing of side chains would change in two stages, one in the $N \rightarrow MG$ and another in the $MG \rightarrow U$ transition. However, typical proteins which unfold through the molten globule state demonstrate a different behavior: their NMR and near-UV CD spectra change dramatically in the $N \rightarrow MG$ transition and do not change very much in the $MG \rightarrow U$ transition. This suggests that proteins in the molten globule state usually do not contain substantial regions with the tight-packed native structure.

More direct evidence for the all-or-none character of molten globule unfolding has been obtained by size-exclusion chromatography (Uversky *et al.*, 1992). Under native conditions, in which the molten globule is much more stable than an unfolded chain, a protein folds into the molten globule state within 1 sec (see Section IV,B,1). However, this process can be much slower (especially at low temperature) at those

GdmCl concentrations which correspond to the middle of equilibrium $MG \rightleftharpoons U$ transition where the molten globule state is much less stable.

In fact, Fig. 24 shows that a bimodal distribution of elution volumes is observed in the GdmCl-induced unfolding transitions of two proteins at 4°C. Thus, both these proteins lose their compact structure according to an all-or-none scheme in these transitions. The simplest explanation of this behavior is that in this case protein unfolding is coupled with protein denaturation which is known to be an all-or-none transition. However, it was shown long ago that both carbonic anhydrase (Wong and Tanford, 1973) and β -lactamase (Robson and Pain, 1976a,b) at room temperature have two GdmCl-induced transitions, monitored by near- and far-UV ellipticities, respectively—the denaturation at smaller concentrations of GdmCl and the unfolding at larger concentrations.

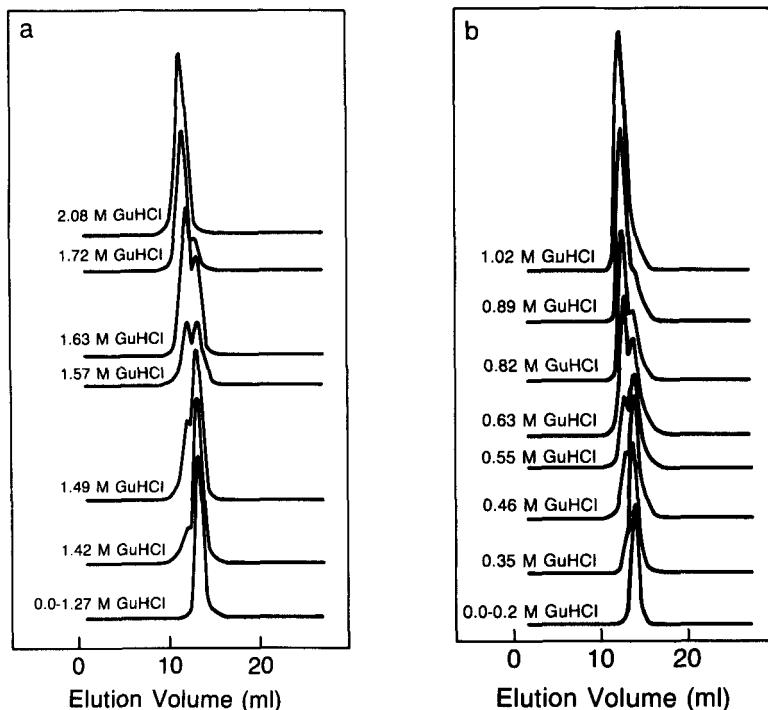


FIG. 24. Elution profiles of size-exclusion chromatography (FPLC) for bovine carbonic anhydrase (a) and *Staphylococcus aureus* β -lactamase (b) at 4°C. GdmCl concentrations are shown above the curves. Comparison of these data with GdmCl-induced denaturation of these proteins shows that a bimodal distribution is observed between two denatured species of protein molecules. Reproduced with permission from Uversky *et al.* (1992).

This is also the case at low temperature as denaturation monitored by activity and near-UV CD occurs at lower GdmCl concentrations than an all-or-none transition monitored by size-exclusion chromatography. In β -lactamase, denaturation occurs mainly between 0.1 and 0.6 M GdmCl, while all-or-none unfolding occurs between 0.35 and 0.9 M (see Fig. 44 in Section V,A,2). This difference is even larger in carbonic anhydrase which is denatured between 1.0 and 1.5 M GdmCl and unfolds between 1.3 and 1.8 M (Uversky *et al.*, 1992). Moreover, the maximum of ANS binding by carbonic anhydrase corresponds to 1.45 M GdmCl, which is near the end of denaturation but near the middle point of unfolding.

It is possible that the chromatographic columns shift the equilibrium between the native and denatured states in such a way that denaturation on a column coincides with unfolding. Special experiments have shown, however, that the activity of carbonic anhydrase, collected from an elution peak corresponding to compact protein molecules, follows the same GdmCl dependence as that measured without a column.

These data clearly show that the all-or-none transition evidenced by bimodal distribution of elution volumes has nothing to do with a protein denaturation. It presents the other all-or-none process: the transition between compact and less compact denatured molecules, i.e., the transition between the molten globule and a less compact state.

It is worthwhile to discuss here a point which is quite clear from a physical point of view but sometimes leads to misunderstanding—the strict definition of all-or-none (or first-order phase) transition. This definition implies only that the physical system (in this case, the protein molecule) undergoes this given transition as a whole. It does not mean that the final state of this transition cannot be subjected to another transition if the parameters of the environment continue to change. In fact, the melting of a crystal is called a phase transition despite that fact that at even higher temperatures a liquid undergoes another phase transition into a gas state. In a similar way the all-or-none transition of a protein molecule means only that the given transition embraces the molecule as a whole. This all-or-none transition can well be just one step in the whole process of the transformation of the native protein in a completely unfolded chain.

2. Molten Globule: A Third Thermodynamic State of Protein Molecules

The experiments described in Sections III,A,3 and III,B,1 leave no doubt that urea- or GdmCl-induced N \rightarrow MG and MG \rightarrow U transitions are of the all-or-none type. This means that the molten globule is not only structurally quite different from the unfolded and native protein states, but even represents a new third thermodynamic state in addition

to two previously known states, native and unfolded (see Uversky *et al.*, 1992; Ptitsyn and Uversky, 1994).

This result strongly contradicts the suggestion that all denatured states of proteins are thermodynamically and even structurally equivalent (Privalov, 1979, 1992). This idea, to a large extent, was based on the studies of temperature-, pH-, and GdmCl-induced denaturation of lysozyme (Pfeil and Privalov, 1976a,b,c). It was shown that the enthalpy of denaturation of lysozyme is a universal function of temperature for all three types of denaturation. The conclusion was that temperature- and pH-induced denaturation of lysozyme transforms it into a state which is thermodynamically similar to the unfolded state produced by GdmCl. It was also suggested that the thermodynamic similarity also means structural similarity and that both these conclusions can be extrapolated for all small proteins: "it follows indisputably that the denatured protein has no fixed residual structure" (Privalov, 1979).

It may be true that the temperature- and pH-denatured states of lysozyme have thermodynamic functions similar to those of the unfolded state (although small but significant changes may be invisible on the background of the strong dependencies of these functions on temperature, pH, etc). However, this conclusion by no means can be considered as a general case, since many proteins can be observed in the molten globule state at low or high pH, at intermediate concentrations of GdmCl or urea, and even at high temperature (see Ptitsyn, 1992a; Bychkova and Ptitsyn, 1993b, for reviews). Structural properties of the molten globule state are, of course, quite different from those of unfolded proteins because this state certainly has a number of native-like α helices and very likely even has many important features corresponding to a native-like folding pattern (see Section II,B).

In some papers (Privalov, 1979; Spolar *et al.*, 1989; Privalov and Makhatadze, 1990) heat capacity curves have been used as an argument for the assumption that all denatured states are thermodynamically or even structurally similar, as it was claimed that heat capacity of a protein does not depend on the type of denaturation. Of course, heat capacity is an extremely indirect structural characteristic of a protein (as contrasted with, e.g., 2D NMR), but even this characteristic has now been shown to be substantially different in the molten globule state and in the unfolded state. Kuwajima (1977) has calculated heat capacities of the intermediate (molten globule) and the unfolded states of bovine α -lactalbumin from thermodynamic analysis of denaturation and unfolding of this protein (see below). He found that C_p of the intermediate state is larger than that of the native state by about 1 kcal/degree · mol

which has later been confirmed by direct experiments (Dolgikh *et al.*, 1981, 1985; Pfeil *et al.*, 1986). On the other hand, the transition from the molten globule to the unfolded state leads to an additional increase in heat capacity of about 0.4 kcal/degree · mol. The fact that the heat capacity of the molten globule state is significantly larger than that of the native state has been confirmed by a number of authors (Xie *et al.*, 1991; Kuroda *et al.*, 1992; Haynie and Freire, 1993; Griko *et al.*, 1994; Griko and Privalov, 1994).

The main reason for the increase in heat capacity on protein denaturation is believed to be the exposure of nonpolar groups (Kauzmann, 1959; Privalov *et al.*, 1989). This suggests that nonpolar groups are exposed in the molten globule substantially less than in the unfolded state, although substantially more than in the native state.

Direct experimental data show the existence of many features of the native 3D structure in the molten globule state, demonstrating that it is something much more than just a "squeezed coil." However, in the absence of direct thermodynamic data, it was still possible to conclude that differences between proteins denatured in different ways are not qualitative from the thermodynamic point of view; and that "all these forms of protein should be assigned to the same macroscopic state" (Privalov, 1992).

Now there are two independent convincing arguments in favor of the opposite point of view. The first is that the slopes of the molten globule–unfolded state transitions induced by urea or GdmCl are proportional to a molecular weight of a protein (Ptitsyn and Uversky, 1994), which is the classical evidence for all-or-none transition (Hill, 1968). The second is that we can observe a bimodal distribution between the molten globule and a less compact state in the GdmCl-induced transition between these two states (Uversky *et al.*, 1992). These experiments probably will contribute to end the long discussion on the difference between various denatured states of protein molecules.

The existence of all-or-none transitions between the molten globule and unfolded states justifies the thermodynamic analysis of denaturation and unfolding of these proteins using the "three-state" model. Kuwajima was the first to apply this analysis and found that experimental data on far- and near-UV circular dichroism of α -lactalbumin can be well described by a three-state model with native (N), intermediate (MG), and unfolded (U) states (Kuwajima *et al.*, 1976; Nitta *et al.*, 1977; Kuwajima, 1977). For α -lactalbumin at neutral pH, at room temperature, and in H₂O, $\Delta G_{N \rightarrow MG} = 4.1$ kcal/mol, while $\Delta G_{MG \rightarrow U} = 2.5$ kcal/mol. A thermodynamic treatment of the far-UV circular dichroism spectra of apomyo-

globin (Barrick and Baldwin, 1993) has led to a similar conclusion ($\Delta G_{N \rightarrow MG} = 4.5$ kcal/mol, $\Delta G_{MG \rightarrow U} = 2.1$ kcal/mol). These estimates show that about two-thirds of $\Delta G_{N \rightarrow U}$ is due to the $N \rightleftharpoons MG$ transition leaving the rest to the $MG \rightleftharpoons U$ transition. This is an important conclusion as it suggests that the stability of native proteins is due mainly to the interactions which are absent in the molten globule state. In other words, these free energy differences "indicate that side-chain packing plays a dominant role in stabilizing the folded protein" (Barrick and Baldwin, 1993).

The "driving force" in the solvent-induced unfolding of the molten globule state (see Figs. 23 and 24) is the difference in the interactions of the protein in these two states with denaturant molecules. This effect is described by the term μn (μ is the chemical potential, n is the number of molecules) in Gibbs free energy. The question arises of how this unfolding changes the more usual thermodynamic functions (enthalpy or entropy). The first estimate was made by Kuwajima (1977) from the temperature dependence of far- and near-UV CD spectra of α -lactalbumin. The results at 25°C in water were $\Delta H_{N \rightleftharpoons MG} = 17.9$ kcal/mol and $\Delta H_{MG \rightleftharpoons U} = 10.8$ kcal/mol, showing that the enthalpy difference between the molten globule and the unfolded states represents only one-third of the total enthalpy of protein folding.

This explains why heating the molten globule state usually does not lead to a cooperative temperature-induced transition. A simulation of the heat capacity curve for a protein with ΔH expected for the $MG \rightleftharpoons U$ transition shows that the maximum for the $C_p(T)$ curve is so broad and so small that it can hardly be observed. Moreover, the temperature-denatured state of many proteins is far from being completely unfolded (see Section III,A,1). Therefore, ΔH for the temperature transition of the molten globule state can be either close to zero (if the temperature-denatured state also is the molten globule, as for the α -lactalbumins) or at least substantially smaller than that for the real $MG \rightarrow U$ transition. This explains why the heating of α -lactalbumins in the molten globule state does not lead to any measurable heat absorption (Dolgikh *et al.*, 1981, 1985; Pfeil *et al.*, 1986; Yutani *et al.*, 1992) as is the case for a number of other proteins for which these experiments have been performed. Cytochrome c, for which the heating of the molten globule state leads to a reasonably large heat absorption (Potekhin and Pfeil, 1989), is the only known exception, although small heat absorption has also been observed for retinol-binding protein (Bychkova *et al.*, 1992) and for despentapeptide insulin (Hua *et al.*, 1993).

Small ΔH for the $MG \rightleftharpoons U$ transition and its remarkable dependence on denaturant concentration also explain the absence of a measurable

heat absorption during solvent-induced unfolding of the molten globule state. Using isothermal microcalorimetry, Pfeil *et al.* (1986) have shown that cooperative heat absorption can be observed only in the GdmCl-induced $N \rightleftharpoons MG$ transition for α -lactalbumin, while the $MG \rightleftharpoons U$ transition does not lead to any measurable cooperative heat absorption. This result can also be explained by Kuwajima's estimates of $\Delta H_{N \rightleftharpoons MG}$ and $\Delta H_{MG \rightleftharpoons U}$ for this protein. In addition to the fact that $\Delta H_{MG \rightleftharpoons U}$ is half as large as $\Delta H_{N \rightleftharpoons MG}$ in water, both these values decrease markedly with increasing GdmCl concentration, and the $MG \rightleftharpoons U$ transition takes place at a much larger GdmCl concentration than the $N \rightleftharpoons MG$ transition. Figure 25 illustrates that $\Delta H_{MG \rightleftharpoons U}$ in the middle of the $MG \rightleftharpoons U$ transition can be much smaller than $\Delta H_{N \rightleftharpoons MG}$ in the middle of the $N \rightleftharpoons MG$ transition (2 and 14 kcal/mol, respectively, according to this estimate).

These arguments suggest that one usually cannot expect large changes in enthalpy (or entropy) of the molten globules at both high tempera-

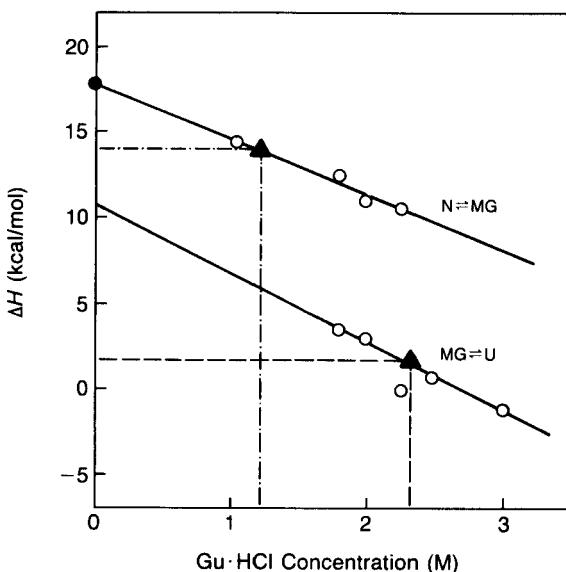


FIG. 25. GdmCl dependence of enthalpies for $N \rightleftharpoons MG$ and $MG \rightleftharpoons U$ transitions in bovine α -lactalbumin at 25°C (Kuwajima, 1977). The filled circle refers to ΔH between MG and N states in the absence of GdmCl. All ΔH values have been calculated from far- and near-UV circular dichroism data basing on the assumption of a three state ($N \rightleftharpoons MG \rightleftharpoons U$) transition. Adapted with permission from Kuwajima (1977). Filled triangles mark the middle points of $N \rightleftharpoons MG$ (1.2 M GdmCl) and $MG \rightleftharpoons U$ (2.3 M GdmCl) transition in human protein at 40°C (Pfeil *et al.*, 1986).

tures and high concentrations of strong denaturants. Despite this, their urea- or GdmCl-induced unfolding is a real all-or-none transition, because the slope of this transition is proportional to the protein's molecular weight and a bimodal distribution of protein molecules is observed in the transition region.

C. Biological Significance and Physical Explanation

1. Biological Significance

The high level of cooperativity of protein denaturation has an obvious biological significance. It prevents a protein from large-scale noncooperative thermal fluctuations which otherwise would damage its rigid active center and impair its function. Biological significance of the all-or-none temperature denaturation is that it makes large local fluctuations unlikely below the temperature at which the molecule becomes flexible as a whole (Ptitsyn, 1992a). Using more quantitative terms, native structures of small proteins or domains have about 10 kcal/mol of stability at "physiological" conditions, which gives only about 0.1 kcal/mol per one amino acid residue. Therefore, if amino acid residues could jump into a more flexible state independently from each other, about half of them would be in a flexible state. However, due to high cooperativity each residue can become flexible only together with the whole protein or domain which forces it to feel the full 10 kcal/mol and consequently only about 10^{-7} of them would be flexible (Karplus and Shakhnovich, 1992).

To understand the biological meaning of all-or-none destruction of the molten globule state it is necessary to remember the growing evidence that the molten globule state exists *in vivo* and plays an important role for a number of biological processes (Bychkova and Ptitsyn, 1993a; see also Section VI). In fact, the molten globule can be considered as another type of native protein in addition to proteins with rigid tertiary structure, and therefore must also be protected from large-scale thermal fluctuations. It follows that both all-or-none transitions in proteins can have important biological meaning, one of them, protecting rigid proteins and the other protecting the molten globules from large noncooperative thermal fluctuations.

2. Early Attempts at Physical Explanation

The physical explanation of the all-or-none character of protein denaturation has a rather long and controversial history, starting with attempts

to describe protein denaturation as the transition from a "liquid-like" compact globule to a "gas-like" statistical coil (Ptitsyn and Eizner, 1965; Ptitsyn *et al.*, 1968; De Gennes, 1975; Sanchez, 1970; Birshtein and Pryamitsyn, 1991; see Chan and Dill, 1991; Karplus and Shakhnovich, 1992, for reviews). The original idea was that the globule-coil transition may be a first-order phase transition as it is in the case of liquid-gas transitions. However, the fact that "molecules" (residues) of the polymer chain are linked with chemical bonds dramatically changes the situation. The point is that each gas molecule is free to move in all space available, while each residue in a coil can move only within a cone formed by chemical bonds. As a result, these theories (see Lifshitz *et al.*, 1978) lead to the conclusion that for long and flexible chains, the globule-coil transition is a gradual transition of the second order rather than of the first order. This conclusion fails to explain the all-or-none character of protein denaturation. Even in the case in which the theory predicts a weak all-or-none transition (for stiff and short chains) this transition is predicted between significantly swollen globule and a coil in such a way that only a small part of the change in protein density is associated with the globule-coil transition itself.

The attempt to take into account the heterogeneity of protein chains (Dill, 1985; Dill *et al.*, 1989; Stigter *et al.*, 1991) led to the conclusion that in this case globule-coil transition must be all-or-none. However, this conclusion is true only for a very specific and not very realistic pathway, namely, through a random collapse of a chain without any structuring or any segregation of polar and nonpolar groups (Karplus and Shakhnovich, 1992). A more rigorous treatment (Grosberg and Shakhnovich, 1986a,b) has shown that at least very long heteropolymers must undergo their globule-coil transitions as phase transitions of the second order, i.e., in a gradual way.

On the other hand, it has been shown theoretically that if a globule-coil transition in polymer molecule is accompanied by a substantial change of its local order (secondary structure), this coupling or two cooperative processes may make each of them an all-or-none process (Grosberg, 1984). There are no obvious experimental examples of this behavior.

Another approach (Ptitsyn, 1975) has suggested that the high cooperativity of protein denaturation may be caused just by the small dimensions of proteins or protein domains. If it is assumed that each block of a secondary structure can lose this structure only as a whole, the total number of intermediates between the native and the unfolded states of a small protein or domain must not be very large. Therefore, even a relatively small increase of free energies of intermediates, compared

with completely folded and completely unfolded states, may make the contribution of intermediates to the partition function negligible. However, more recent results (Ptitsyn and Uversky, 1994) have shown that the cooperativity of the denaturation increases rather than decreases with an increase in molecular weight (see Fig. 19) which is inconsistent with the idea that the cooperativity is the result of the small dimensions of protein molecules.

3. Theory of Protein Denaturation

The problem of the all-or-none protein denaturation completely changed after it was shown that the denaturation is all-or-none even when it neither leads to a large increase of protein volume nor to a substantial decrease of a secondary structure (Dolgikh *et al.*, 1981, 1985). This requires searching for other reasons for high cooperativity not associated with large changes of protein volume and/or its secondary structure.

A novel approach has been elaborated by Shakhnovich and Finkelstein (1982, 1989). Their theory treats protein denaturation not as the liquid—gas, but as the crystal—liquid type of transition. The central point of this theory is that the high cooperativity is due to cooperative destruction of tight packing of side chains that is present only in the native state but is absent in all denatured states, starting with the molten globule. The main idea of this theory is presented schematically in Fig. 26. In the native state side chains are tightly packed inside a protein core and therefore this core represents a minimum in both energy and entropy. The expansion of a molecule leads to a smooth increase of its energy as packing of side chains becomes less tight and their van der Waals attraction decreases. However, the increase of the entropy of a molecule is more complicated. Due to the existence of the intrinsic vibrational potential of side chains they have two different types of motion, librations inside a given rotational isomer and jumps from one rotational isomer into another. With the small expansions the amplitude of libration is limited by the intrinsic potential and does not depend on molecular volume. Therefore, at these degrees of expansion an increase in energy is not accompanied by an increase in entropy, which means the existence of a potential barrier. Only upon a further expansion do the free spaces close to each side chain become large enough to permit side chains to jump from one rotational isomer into another, which substantially increases the entropy. As a result, free energy of a protein has two minima, the native with small energy and small entropy and the denatured with large energy and large entropy. Intermediate values of protein volume correspond to large energy and small entropy and therefore to

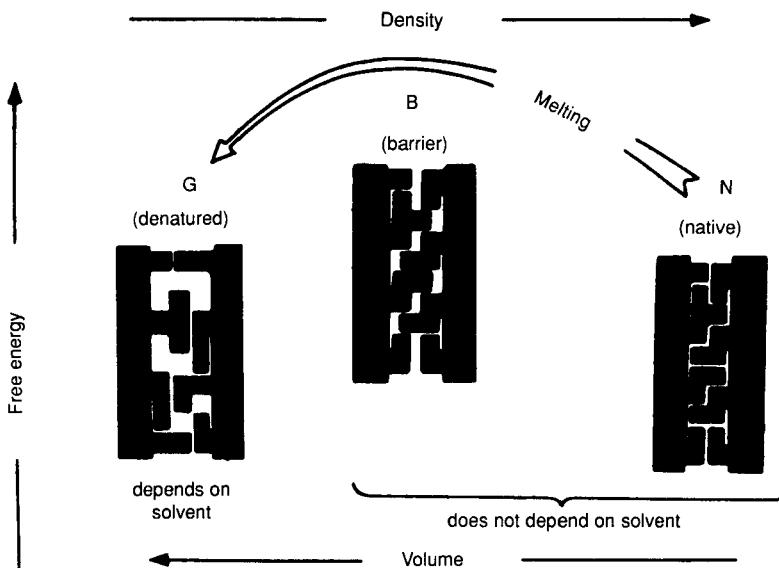


FIG. 26. Scheme of the melting of protein tertiary structure. From Shakhnovich, E. I., and Finkelstein, A. V. *Biopolymers* 28. Copyright © 1989 John Wiley & Sons. Reprinted by permission of John Wiley & Sons, Inc.

the maximum in free energy. Within the frame of this theory the exact nature of a denatured state is not of principal importance; it requires only that the volume of denatured state must be large enough to permit at least a substantial part of the side chains to rotate.

At least three predictions of this theory have been confirmed by direct experiments. First, the theory predicts that the increase of protein volume by 30% (i.e., linear dimensions by 10%) is enough to reach the "denatured" minimum of free energy, in which the increase of entropy will compensate for the increase of energy. In fact, numerous experimental estimates of the hydrodynamic radius of the molten globule state (see Section II,A,2) show that it is larger than that of the native state by 11–16%, which is very close to theoretical estimate. Second, the theory predicts that in this denatured state with a minimum volume only aliphatic side chains can rotate almost freely, while larger and more asymmetric aromatic side chains remain more hindered. In fact, NMR data (Semisotnov *et al.*, 1989; Rodionova *et al.*, 1989; see also Ptitsyn, 1992a) show that aliphatic side chains rotate almost freely in the molten globule state, while aromatic side chains are remarkably more hindered (see Section II,A,4). Thus, the movements of aromatic side chains and their immediate neighbors being more pronounced than those in the

native state (which is clearly shown from their near-UV CD and NMR spectra) are still substantially more restricted than the movements of aliphatic side chains. Third, the theory of Shakhnovich and Finkelstein implies that the transient state between the native and the denatured states corresponds to a protein density intermediate between these two states. This has been directly confirmed by Fersht's group using site-directed mutagenesis in a protein core. It was demonstrated that the core progressively consolidates on the way from a kinetic molten globule intermediate to a transient state (at the rate-limiting step barrier) and then to the completely folded state (Matouschek *et al.*, 1992a).

It follows that the all-or-none character of protein denaturation is well explained by this theory. It is interesting to note that the theory of Shakhnovich and Finkelstein compositionally does not take into account the heterogeneity of protein chains. Thus, their results can even be applied to a homopolymer if it possessed a tightly packed structure. Heterogeneity of protein chains may influence the results of this theory only in one aspect. If chains were quite homogeneous, the dimensions of a protein core could increase in an unlimited way with the increase of molecular weight. However, globular proteins consist of nonpolar and polar residues which are more or less randomly distributed along the chain. In these chains the dimensions of a nonpolar cores are limited, as the formation of large cores is possible only if a number of long chain regions are entirely nonpolar (Finkelstein and Ptitsyn, 1987). Therefore, the domain melting of large proteins (see Section II,A,4) is a direct effect of the quasirandom distribution of nonpolar and polar residues in typical globular proteins.

4. Biological Evolution and Phase Transition

It is much more difficult to explain the all-or-none character of the transition between the molten globule and the unfolded states, i.e., between compact and noncompact denatured states, each of which has no rigid tertiary structure. The all-or-none character of this transition cannot be explained using an analogy with the globule-coil transition in polymer chains, as the globule-coil transition has been theoretically predicted to be a smooth second-order phase transition rather than a first-order transition (see above). Moreover, it was shown experimentally that globule-coil transitions in synthetic random copolymers of polar and nonpolar residues definitely do not belong to an all-or-none type (Anufrieva *et al.*, 1975; Bychkova *et al.*, 1980; Semisotnov *et al.*, 1981). Rather, it goes through at least two equilibrium intermediates—a noncompact state with a pronounced secondary structure and a “half-compact” state, which is structured enough to hinder internal rotation

of fluorescent groups, but still does not achieve a compact globular structure.

Therefore, it is reasonable to assume (Bychkova and Ptitsyn, 1993a; Ptitsyn and Uversky, 1994; Ptitsyn, 1994) that an all-or-none molten globule–unfolded state transition is connected with a breakdown of the native-like folding pattern (which is almost certainly present in the molten globule state). If so, the whole picture of phase transitions in globular proteins becomes simple and beautiful: there are two levels of protein 3D structure—the crude folding pattern and the detailed tertiary structure—and each of them can be destroyed in a small protein or a domain only by an all-or-none transition, i.e., by a phase transition of first order. However, the physical mechanism of the all-or-none destruction of a folding pattern remains unclear.

An alternative approach to the explanation of the all-or-none denaturation of proteins was first proposed by Go *et al.* (Taketomi *et al.*, 1975; Go and Abe, 1981) and then elaborated by Bryngelson and Wolynes (1987, 1990, see also Skolnick *et al.*, 1990). The approach is based on the assumption that a folding mechanism essentially involves some “memory” on the protein native structure which is reflected into specific interactions between neighboring residues in this structure. It was assumed (Taketomi *et al.*, 1975; Go and Abe, 1981) that monomers which are in contact in the native state specifically attract each other in other conformations, while all other monomers do not interact. Bryngelson and Wolynes (1987, 1990) have even assumed that intrinsic energies of monomers are smaller in the native state than in all other states. Both Monte Carlo stimulation (Taketomi *et al.*, 1975; Go and Abe, 1981) and analytical theory (Shakhnovich and Gutin, 1989) have shown that for this model the native state is separated from the completely disordered denatured state by a first-order phase transition.

The assumption that interaction energies of a given pair of monomers (and, moreover, even their intrinsic energies) are quite different for the monomers which are and are not in contact in the native structure looks rather exotic and is at the first glance extremely difficult to justify from a physical point of view. In fact, interaction energy of a given pair of residues depends only on their distance and mutual orientation and cannot have memory, whether these residues have or have not interacted in the native structure. However, in a more general sense this approach may reflect the very important idea that the first-order transition needs a large energy gap between the native state and all nonnative states. As has been pointed out by Shakhnovich and Gutin (1993a,b), this gap can be created without any special assumptions regarding the parameters of intramolecular interactions. In fact, the large energetic gap between

the native state and all other states may be the result of the evolutionary selection of amino acid sequences with the aim of decreasing the energy of the "ground" (native) state without a large influence on the energies of other states (see Section V,B,2).

An attractive feature of this idea is that it can be applied both to protein denaturation ($N \rightleftharpoons U$ or $N \rightleftharpoons MG$ transition) and to protein unfolding ($MG \rightleftharpoons U$ transition). In fact, it is quite possible (see Section V,B,5) that evolutionary selection leads to a specific decrease in the energy of both the native tertiary structure and the native folding pattern. If so, it would mean that both the native tertiary structure and the native-like molten globule state can be separated from the more disordered states by the same mechanism. As a result, both these levels of protein structure can be destroyed according to an all-or-none scheme.

It should be emphasized that an all-or-none protein denaturation ($N \rightleftharpoons U$ or $N \rightleftharpoons MG$ transitions) can be alternatively explained by the theory of Shakhnovich and Finkelstein (1989) which can be true even for homopolymers and has nothing to do with the biological history of natural proteins. On the other hand, the only explanation proposed so far for an all-or-none protein unfolding ($MG \rightleftharpoons U$ transition) is based on the biological selection of protein sequences and therefore can refer only to natural (biologically selected) proteins. The question of whether this transition can be explained without reference to biological evolution requires future study.

IV. KINETICS OF PROTEIN FOLDING

Up to this point the molten globule has been described as an interesting phenomenon in protein physics without referring to its possible biological importance. However, it is now well known that the molten globule state plays a very important role in protein folding. Moreover, it is involved in many important physiological processes in a living cell. The role of the molten globule in protein folding will be described in this and subsequent sections, while its role in physiological processes will be discussed in the last section of this chapter.

There are many reviews on protein folding emphasizing different points of view on this process (Tanford, 1968, 1970; Jaenicke, 1981, 1987, 1988, 1991; Kim and Baldwin, 1982, 1990; Kuwajima and Schmid, 1984; Ptitsyn, 1987, 1992a; Kuwajima, 1989, 1992; Christensen and Pain, 1991; Matthews, 1991, 1993; Dobson, 1991; Schmid, 1991, 1992, 1993; Richards, 1992; Garel, 1992; Creighton, 1990, 1992; Baldwin, 1993; Fersht, 1993; Roder and Elöve, 1994).

A. Formation of Premolten Globule Intermediate

1. Formation of Fluctuating Secondary Structure

The first experimental evidence on kinetic intermediates was obtained in the late 1970s when it was shown that on folding β -lactamase (Robson and Pain, 1976b), ribonuclease A (Schmid and Baldwin, 1979; Kim and Baldwin, 1980), and carbonic anhydrase B (McCoy *et al.*, 1980), all go through an intermediate with pronounced secondary structure (monitored by far-UV circular dichroism or other methods) but without tertiary structure at the atomic level (monitored by near-UV circular dichroism). Almost simultaneously, Creighton (1980) showed the existence of a compact kinetic intermediate in protein folding. These first experiments (see Kim and Baldwin, 1982, 1990; Kuwajima, 1989, for reviews) established the existence of kinetic intermediates. However, questions about the number of these intermediates and the timescale of their formation remained unanswered.

Answers to these questions became possible only after the stopped-flow technique had been elaborated and combined with different physical methods, like circular dichroism or fluorescence, to provide information on different features of protein structure. A typical experiment consists of unfolding the protein by urea or GdmCl and then mixing this solution with a native buffer decreasing the concentration of denaturant well below the transition region. Recently, the dead time of mixing experiments has become as small as 2–4 msec (Elöve *et al.*, 1992; Radford *et al.*, 1992a). This permits separation of the “burst” stage of protein folding, which occurs within this dead time, from slower processes and to monitor these slower processes by the stopped-flow technique.

Even the first attempts to apply this technique to protein folding in a systematic way lead to two important observations. First, it was shown (Kuwajima *et al.*, 1987; Gilmanshin and Ptitsyn, 1987) that the far-UV ellipticity of proteins can change very significantly at the early stages of protein folding (within about 20 msec), which suggests the very fast restoration of a substantial part of the secondary structure. Second, Semisotnov *et al.* (1987) demonstrated the existence of another substantially slower phase of protein folding detected by fluorescence and electronic paramagnetic resonance methods with a half-time about 40 msec.

Figure 27 shows the first experimental evidence (Kuwajima *et al.*, 1987) of the very fast restoration of secondary structure upon a protein's refolding from the completely unfolded state. About 80% of the native far-UV ellipticity is restored within the dead time of experiment (about 20 msec), while the near-UV ellipticity remains unaltered within this

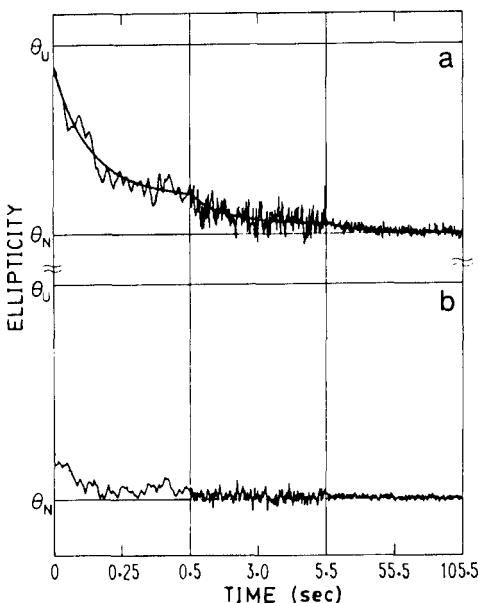


FIG. 27. Kinetics of refolding for cytochrome c (pH 6.8–7.0; 25°C) induced by concentration jump of GdmCl from 4 to 0.4 M. The curves are molar ellipticities in the near (a)- and far (b)-UV regions at 420 and 222.5 nm, respectively. Reproduced with permission from Kuwajima *et al.* (1987).

time and changes with a half-time of about 100 msec, i.e., at least an order of magnitude slower. This provides strong evidence that protein secondary structure can be substantially restored at the very early (burst) stage of protein folding.

Similar investigations have been performed by Kuwajima and co-workers for a number of other proteins and have been summarized by Kuwajima *et al.* (1993; see also Ptitsyn and Semisotnov, 1991). Figure 28 compares molar ellipticities in the far-UV region obtained just after the dead time of stopped-flow experiments (about 10 msec), Θ_i , and in the completely folded (native) state, Θ_N . For six proteins (cytochrome c, carbonic anhydrase, chymotrypsinogen, bovine and human α -lactalbumins, and lysozyme) $\Theta_i \approx \Theta_N$ (Θ_i is between 80 and 100% of Θ_N), for seven proteins (carp parvalbumin, β_2 subunit of tryptophan synthase, staphylococcal nuclease, dihydrofolate reductase, ribonuclease, β -lactamase, and phosphoglycerate kinase) $|\Theta_i| < |\Theta_N|$ (usually Θ_i is about 50 to 60% of Θ_N), and for one protein (β -lactoglobulin) $|\Theta_i| > |\Theta_N|$. This means that in all cases the far-UV ellipticity is substantially or even completely restored within the first 10 msec of protein

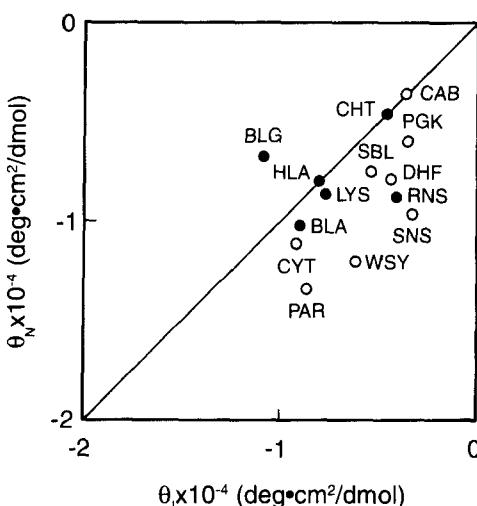


FIG. 28. Molar ellipticities at 220–225 nm of native proteins (Θ_N) versus those of the early intermediates of protein folding (Θ_I). Notations: LYS, hen egg lysozyme (Kuwajima *et al.*, 1987); CYT, horse ferricytochrome c (Kuwajima *et al.*, 1987); BLG, bovine β -lactoglobulin (Kuwajima *et al.*, 1987); PAR, carp parvalbumin III (Kuwajima *et al.*, 1988; WSY, β_2 subunit of *E. coli* tryptophan synthase (Goldberg *et al.*, 1990); SNS, *Staphylococcal* nuclease A (Sugawara *et al.*, 1991); DHF, *E. coli* dihydrofolate reductase (Kuwajima *et al.*, 1991); CAB, bovine carbonic anhydrase B (data by G. V. Semisotnov and K. Kuwajima, published in the review of Ptitsyn and Semisotnov, 1991); BLA, bovine α -lactalbumin (Ikeguchi *et al.*, 1992); RNS, bovine ribonuclease A; HLA, human α -lactalbumin; CHT, bovine chymotrypsinogen A; SBL, *Staphylococcal* β -lactamase; PGK, yeast phosphoglycerate kinase (unpublished data by G. V. Semisotnov and K. Kuwajima). Filled and unfilled circles refer to proteins with and without S–S bonds, respectively. Adapted with permission from Kuwajima *et al.* (1993).

folding which suggests the very fast (burst) partial or even complete restoration of secondary structure. The behavior of β -lactoglobulin in which $|\Theta_I|$ is larger than $|\Theta_N|$ was later observed also for egg white lysozyme (Radford *et al.*, 1992a; Chaffotte *et al.*, 1992) and was convincingly explained (Chaffotte *et al.*, 1992) by the contribution of S–S bonds and/or aromatic side chains in the far-UV CD spectrum of the native protein.

Since a substantial part of the far-UV ellipticity is restored within the dead time of experiments, values of ellipticities extrapolated from measured kinetic data to zero time are quite different from those for the unfolded state (see, e.g., Fig. 27b). If this extrapolation is performed for different wavelengths, the whole far-UV CD spectrum of the intermediate can be reconstructed (Kuwajima *et al.*, 1988, 1991; Sugawara *et al.*, 1991). Moreover, it is even possible to measure the far-UV CD spectrum

just after the dead time using the continuous-flow technique (Elöve *et al.*, 1992). The results confirm those obtained from molar ellipticity at one wavelength: the far-UV spectra of early intermediates are much more pronounced than those in the unfolded state but usually less pronounced than those in the native state.

Figure 29 (Elöve *et al.*, 1992) illustrates changes in the far-UV CD spectrum of cytochrome c upon protein folding. These experiments were performed at shorter dead time (4 msec) compared to the pioneering paper of Kuwajima *et al.* (1987). In addition, lower temperature and higher GdmCl concentration were used. This permits a more detailed picture compared with that presented in Fig. 27b. Under these conditions the far-UV spectrum at 4 msec is relatively close to the unfolded state and corresponds to the superposition of 40% native state and 60% unfolded state. At 100 msec the spectrum becomes substantially closer to the native state and corresponds to the superposition of 72% native state and 28% unfolded state, while at 5 sec the spectrum is indistinguishable from that of the native state.

An important next step was a comparison of these data with the protection of backbone NH groups against hydrogen exchange, obtained by the "pulse hydrogen exchange" technique (see below). This comparison has shown that NH groups of cytochrome c (Elöve *et al.*, 1992), lysozyme (Radford *et al.*, 1992a), and interleukin-1 β (Varley *et al.*, 1993)

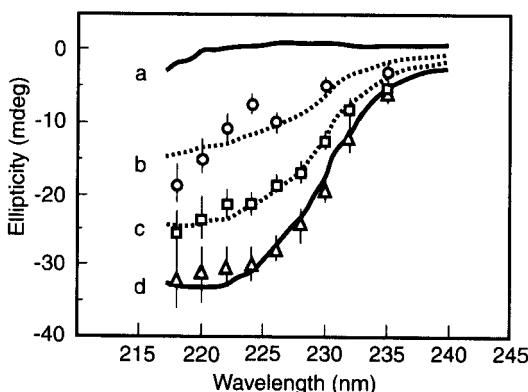


FIG. 29. Comparison of transient far-UV CD spectra of oxidized cytochrome c with equilibrium spectra for fully unfolded (a) and native (d) proteins. Transient spectra were obtained at 0.004 (○), 0.100 (□), and 5 (△) sec after the start of folding (at 10°C in 0.7 M GdmCl). Spectra of the unfolded (a) and native (d) states are shown by solid curves. The dashed curves indicate linear combinations of the far-UV CD spectra for the native (N) and the unfolded (U) states as follows: 0.40 N + 0.60 U at 4 msec (b) and 0.72 N + 0.28 U at 100 msec (c). Reproduced with permission from Elöve *et al.* (1992). Copyright 1992 American Chemical Society.

remain virtually nonprotected in the burst stage of folding, despite the high degree of restoration of the secondary structure (40% for cytochrome c and 90% for interleukin) measured by the far-UV CD spectra. The explanation is obvious—the secondary structure of early intermediates is not sufficiently stable to protect amide protons against exchange. In fact, to be protected from exchange, NH groups not only have to form hydrogen bonds but these bonds must be stable against thermal fluctuations. This provides good evidence that secondary structure is formed at the burst stage of protein folding mainly in a fluctuating state and becomes stable only at the subsequent stages.

The data on far-UV ellipticities of early transient intermediates can be used to evaluate the stability of the burst intermediate. Since the ellipticity extrapolated to zero time $\Theta(0)$ is proportional to the concentration of this state, its stability can be estimated from the dependence of $\Theta(0)$ on the final concentration of urea or GdmCl in refolding experiments (see, e.g., Ikeguchi *et al.*, 1986; Kuwajima *et al.*, 1986, 1988, 1991; Sugawara *et al.*, 1991; Jennings and Wright, 1993; Matthews, 1993). For example, Fig. 30 compares the stabilities of the burst-phase intermediate and native states for *Escherichia coli* dihydrofolate reductase (Kuwajima *et al.*, 1991). The figure shows that the burst-phase intermediate in this case is much less stable than the native protein and unfolds much less cooperatively.

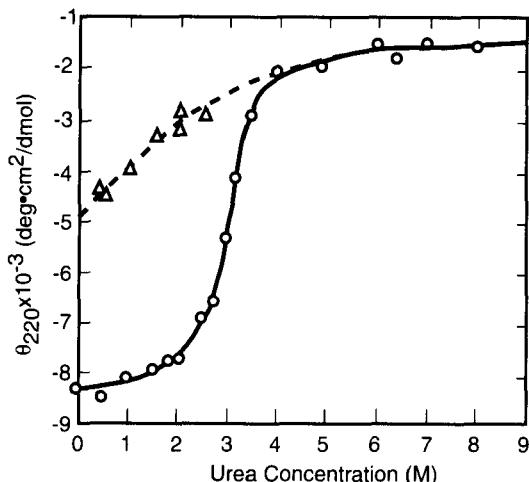


FIG. 30. Stability of *E. coli* dihydrofolate reductase against urea-induced unfolding measured by far-UV ellipticity [Θ]₂₂₀, at pH 7.8 and 15°C. ○, equilibrium curve; Δ, kinetic data extrapolated to zero time from $t > 18$ msec. Reproduced with permission from Kuwajima *et al.* (1991). Copyright 1991 American Chemical Society.

2. Partial Condensation

There are some indications that the formation of a fluctuating, unstable secondary structure in the burst state is accompanied by partial condensation of a protein chain. In at least two cases, partial condensation of a protein during the burst phase has been established by an energy transfer between a single tryptophan residue and another chromophore. Kawata and Hamaguchi (1991) have observed a very fast (within ~ 20 msec) increase of energy transfer between the single Trp-148 and a chromophore attached to the single cysteine residue in the constant fragment of the κ -immunoglobulin light chain. Elöve *et al.* (1992) have shown the existence of a burst phase (within 4 msec) in heme quenching of fluorescence of the single Trp-59 in cytochrome c. This burst phase leads to about 35% of the total decrease of Trp-59 fluorescence in this protein (see Fig. 41 in Section V,A,1).

Varley *et al.* (1993) have observed a very fast stage in the change of fluorescence of the single tryptophan residue in interleukin 1 β (35% of this change occurs with a half-time of 3.5 msec) which suggests the fast stage of partial burying of this tryptophan upon protein folding. Similar but more detailed results have been obtained by Khorasanizadeh *et al.* (1993) with the Phe-45 \rightarrow Trp mutant of ubiquitin which is active and structurally very similar to the wild-type protein. Using the engineered Trp-45 as a probe, Khorasanizadeh *et al.* (1993) observed a series of kinetic phases for the decrease of its fluorescence during refolding. If the final denaturant concentration is below 2 M GdmCl, up to 60% of this decrease occurs within the first 2 msec (the dead time of the experiment) at 25°C. The fluorescence extrapolated to zero time depends on GdmCl concentration according to a sigmoidal curve. This suggests the formation of a condensed state during the dead time in which Trp-45 is buried and its fluorescence is partly quenched. It is interesting that no partly condensed intermediate has been observed at low temperature (8°C). This suggests that a condensed state is stabilized by hydrophobic interactions which are known to increase with temperature (Kauzmann, 1959).

Another important event which takes place at the burst stage of protein folding is the partial binding of the hydrophobic fluorescent probe. It has been shown (Semisotnov *et al.*, 1987, 1991a; Rodionova *et al.*, 1989) that ANS is bound to the equilibrium molten globule much more strongly than to the native and the unfolded states. Since the binding of ANS is accompanied by a drastic increase in intensity in its fluorescence and by a blue shift of the fluorescence spectrum, the measurement of ANS fluorescence is a sensitive tool for identifying the molten globule state.

The application of this technique (in combination with stopped-flow) to the kinetics of protein folding (Semisotnov *et al.*, 1987, 1991a; Ptitsyn *et al.*, 1990; Goldberg *et al.*, 1990; Ptitsyn and Semisotnov, 1991; Varley *et al.*, 1993; Matthews, 1993) has shown the existence of at least two stages of ANS binding to a folding protein. The first stage is within the dead time of experiments (4–20 msec) and leads to at least 50% of the maximal ANS binding, while the second stage is much slower (several decades of msec, see Section IV,B,1). As an example, Fig. 31 shows kinetic data on the ANS binding to carbonic anhydrase and α -lactalbumin (Semisotnov *et al.*, 1991a). Figure 31 demonstrates that the intensity of ANS fluorescence observed just after the dead time is substantially larger than that for free ANS, which shows the existence of the burst stage of ANS binding.

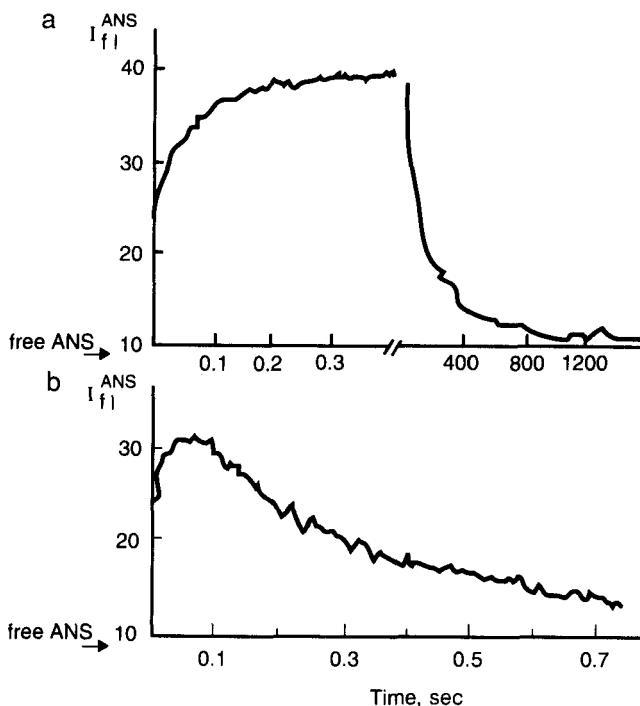


FIG. 31. Increase in binding of ANS upon folding of bovine carbonic anhydrase B (a) and of human α -lactalbumin (b). Both proteins were refolded from the unfolded state by their dilution from 8.5 to 1.4 M urea at 25°C. Decrease of ANS fluorescence intensity on a time scale of hundreds of seconds reflects the release of ANS from a protein at its transition to the native state (Semisotnov *et al.*, 1987, 1991a). From Semisotnov *et al.* (1991a) *Biopolymers*. Copyright © 1991 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

It is known (Stryer, 1965; Turner and Brand, 1968) that ANS binds to solvent-accessible nonpolar clusters in proteins. Therefore, the very fast stage of ANS binding can be explained (Semisotnov *et al.*, 1991a) by its binding to clusters of nonpolar groups which are formed at the surfaces of fluctuating α helices and especially of β structure. On the other hand, a very fast stage of ANS binding may also reflect the very fast formation of some elements of protein globular structure (Ptitsyn and Semisotnov, 1991).

Thus, early in protein folding a kinetic intermediate accumulates which is partly condensed, possesses fluctuating secondary structure, and has solvent-exposed clusters of nonpolar groups. We initially proposed (Uversky and Ptitsyn, 1994; Ptitsyn, 1994) to call this intermediate a partly folded state of protein molecules. However, the term premolten globule state (Uversky and Ptitsyn, 1995) seems to be better since it specifies the place of this state in the whole family of different partly folded states.

3. Secondary Structure in Unfolded Chains

Since substantial parts of both the native far-UV CD spectra and the native fluorescence quenching are restored within the dead time of rapid mixing experiments, it is difficult to determine the relative rates of these processes. To solve this problem it is necessary to penetrate inside the dead time of mixing experiments (few msec) which can be done if protein folding can be initiated by light. The first example of a light-triggered folding reaction was described by Roder's and Eaton's groups (Jones *et al.*, 1993). They have explored the observation that carbon monoxide (CO) binds to covalently attached heme of cytochrome c preferentially in the unfolded state. Therefore, photodissociation of CO triggers the folding reaction (see Fig. 32) and the kinetics of this process can be monitored by time-resolved absorption spectroscopy. In their first paper Jones *et al.* (1993) combined this approach with nanosecond laser pulses to monitor the transient absorption spectra at time delays from 10 nsec to 1 sec after the photodissociation of CO. Thus, they have identified the intramolecular transient binding of different methionine and histidine ligands to the heme. Light-triggered folding of cytochrome c opens the possibility of studying the folding of this protein without practically any limitation in time resolution. It becomes possible to extend kinetic measurements to extremely fast stages of the folding reaction, which may provide a new insight into the relationship between formation of secondary structure and condensation of a polypeptide chain.

In the absence of direct experimental data it is worthwhile to consider theoretical predictions concerning the relationship between the formation of secondary structure and the condensation of a protein chain.

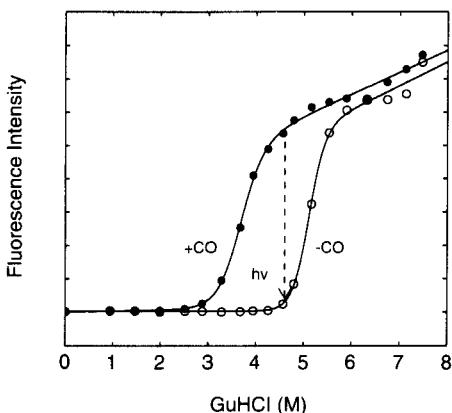


FIG. 32. GdmCl-induced unfolding of reduced horse cytochrome c in the presence (●) and absence (○) of CO (1 atm) at 40°C, monitored by the fluorescence of the single Trp-59. This fluorescence is almost completely quenched in the native protein by energy transfer to the heme and drastically increases in the unfolded state. The binding of CO stabilizes substantially the unfolded state, shifting the midpoint concentration of the N → U transition from 5.1 to 3.7 M GdmCl. Thus, in 4.6 M GdmCl the photodissociation of CO initiates the transition of cytochrome c from the unfolded state to the folded state. Adapted with permission from Jones *et al.* (1993).

Two extreme points of view have been proposed regarding this problem. According to the “framework model” (Ptitsyn, 1973, 1987, 1991) and especially the “diffusion–collision model” (Karplus and Weaver, 1976, 1994), the chain condenses by a “merging” (Ptitsyn, 1973) or “collision” (Karplus and Weaver, 1976) of a preexisting region with the secondary structure. On the other hand, Chan and Dill (1989, 1990a,b, 1993) argue that the condensed state of a protein chain, taken by itself, may shift the equilibrium toward the formation of α helices and β structure even in the absence of hydrogen bonds supporting these structures in an unfolded chain. Although the quantitative (but not qualitative) aspects of their calculations have been questioned (Gregoret and Cohen, 1991), the concept of Chan and Dill has become popular and has led to the belief that condensation of a protein chain must precede formation of its secondary structure.

There is another perhaps more general reason why condensation of a protein chain must stabilize its secondary structure. A chain cannot condense without the disruption of hydrogen bonds between peptide groups and water molecules. This disruption would lead to a large increase of free energy, unless these broken hydrogen bonds are compensated for by intramolecular hydrogen bonds. This compensation is possi-

ble only by the formation of a network of intramolecular hydrogen bonds between peptide groups, i.e., the formation of secondary structure (Finkelstein and Ptitsyn, 1987).

However, it is difficult to understand how proteins can condense without segregation of their nonpolar groups from polar groups, since otherwise polar groups would be included in a nonpolar core which would dramatically increase its free energy. This segregation is hardly possible in typical (quasirandom) sequences of natural proteins without a previous or simultaneous formation of secondary structure. The point is that α helices and β strands in all globular proteins have at least one nonpolar surface (Perutz *et al.*, 1965; Schiffer and Edmundson, 1967; Lim, 1974; Ptitsyn and Finkelstein, 1980) and therefore can merge with each other without involvement of polar side chains into a nonpolar core. This does not necessarily need evolutionary selection of protein sequences, since nonpolar surfaces can be formed in random sequences with the same average lengths and the same length distribution as those in natural proteins (Ptitsyn, 1983, 1984, 1985; Ptitsyn and Volkenstein, 1986). Therefore, it is interesting to study to what extent protein secondary structure can preexist in unfolded chains.

Numerous experimental data by Baldwin's, Wright's, Kallenbach's, Stellwagen's and other groups (see Dyson and Wright, 1991, 1993; Scholz and Baldwin, 1992; Ptitsyn, 1992b; Chakrabarty and Baldwin, 1993, for reviews) have shown that many short peptides in aqueous solution have far-UV CD spectra consistent with a pronounced α -helical structure. These data have demonstrated the important influence of charged groups on the stability of α helices (see, e.g., Shoemaker *et al.*, 1985, 1987; Lyn *et al.*, 1992a,b) predicted as early as 1977 (Finkelstein and Ptitsyn, 1977; Finkelstein *et al.*, 1977; Finkelstein, 1977).

The pronounced helicity of short peptides was initially considered to be surprisingly large and was attributed to strong and specific side chain-side chain interactions (see, e.g., Kim *et al.*, 1982). However, it was later shown (Finkelstein *et al.*, 1990) that a similar degree of helicity can be expected for short peptides from uncharged lysine or glutamic acid using experimental parameters for corresponding polypeptides and an extension of the Zimm-Bragg theory concerning polypeptides (Zimm and Bragg, 1959) to short molecules. Moreover, the helicities of short heteropeptides can be calculated in a quantitative way by the molecular theory of secondary structure for unfolded peptides (Ptitsyn and Finkelstein, 1980, 1983). These calculations generally correlate well with experimental CD data, explaining the dependence of helicity of short peptides on amino acid sequences, pH, and temperature (Finkelstein *et al.*, 1991).

It is especially interesting that NMR technique permits the measurement of NOEs between the residues which are near each other along the chain and thus to localize α helices (Wright *et al.*, 1990; Liff *et al.*, 1991; Dyson *et al.*, 1992a,b; Waltho *et al.*, 1993) and β turns (Wright *et al.*, 1988; Dyson *et al.*, 1988; Sumner *et al.*, 1990, 1992) in short peptides. It was shown, e.g., the presence of α -helical order in peptides corresponding to the A and D helices of myohemerythrin and the very weak ("nascent") helical order in peptides corresponding to two other long helices (B and C) of this protein (Dyson *et al.*, 1992a). On the other hand, no helical order has been observed in fragments of plastocyanin (Dyson *et al.*, 1992b) which has no helical structure in its native state. In a similar way, substantial helicity was found for the peptide corresponding to the H helix of myoglobin, whereas the G helix peptide exhibits only a very small amount of helical conformation (Waltho *et al.*, 1993).

Returning from small peptides to unfolded protein chains, it is necessary to emphasize that the same molecular theory predicts a pronounced secondary structure for unfolded protein chains, as illustrated in Figs. 33 and 34 which present the results of the calculations of Ptitsyn and Finkelstein (1980, 1983). Figures 33a and 34a show the results of the theory for unfolded protein chains. They demonstrate that α and β structures of proteins are roughly outlined even in their unfolded states, i.e., are determined mainly by local interactions in α helices or intermediate-range interactions in β structures. Figures 33b and 34b present the results of the theory of secondary structure for compact chains where the long-range interactions are modeled by the interaction of each chain region with a nonspecific hydrophobic template (Ptitsyn and Finkelstein, 1983). The comparison of the results of the two theories with each other and with X-ray data show that long-range interactions basically do not change the prediction of the theory, but make it substantially more definite: instead of smooth curves of α and β probabilities in unfolded chains we can see much higher "resolving power" in the compact molecules.

It follows that secondary structure (fluctuating around its native location) can exist even in unfolded chains, i.e., without support of long-range interactions. However, the stability of this structure is marginal, and taken by itself, this fluctuating structure can hardly substantially facilitate protein folding (Abkevich *et al.*, 1994). Thus, both protein condensation without a formation of secondary structure and the formation of secondary structure without condensation seem to be unfavorable processes and do not correspond to minima of free energy. It appears more plausible that the condensation of a chain and the formation of its more or less secondary structure is a single cooperative process,

i.e., both these events occur simultaneously (Ptitsyn, 1995). This single cooperative process may coincide with the formation of the kinetic premolten globule intermediate (see Sections IV,A,1, IV,A,2, and V,A,1) and its equilibrium analog (see Section V,A,2). The formation of the molten globule state may be a result of the further condensation of a protein molecule accompanied by the further increase in the stability of its secondary structure.

B. Formation of Molten Globule Intermediate

1. Early Evidence

The molten globule state was predicted (Ptitsyn, 1973) as a kinetic intermediate of protein folding and then was observed experimentally (Dolgikh *et al.*, 1981) as an equilibrium state of certain protein molecules at low pH or at moderate concentrations of GdmCl. Later it was shown that the molten globule accumulates in the folding of carbonic anhydrase, and thus can not only be an equilibrium but also a kinetic intermediate (Dolgikh *et al.*, 1984). Figure 35a presents this first evidence of the existence of a molten globule-like kinetic intermediate of protein folding. Figure 35a shows that the reduced viscosity and the far-UV ellipticity are almost restored to their native values within the dead time of these experiments which was ~ 2 min, as manual mixing had been used. In contrast, esterase activity of a protein and its near-UV ellipticity change by only 10–15% during the dead time and continue to change afterward with a half-time of ~ 25 min. This suggests the accumulation of an intermediate which is almost as compact as the native protein and has a similar amount of secondary structure, but has no rigid tertiary structure and no activity, i.e., it meets all the requirements formulated earlier for the equilibrium molten globule state (Dolgikh *et al.*, 1981). It is important to emphasize (see Fig. 35b) that a similar slow phase in folding kinetics has also been observed in the case in which folding is started from the equilibrium molten globule state (at ~ 2 M GdmCl) rather than from an unfolded state. This suggested that a kinetic intermediate observed in experiments on protein folding from an unfolded state may be similar to that of the equilibrium molten globule state. This conclusion has been confirmed by the comparison of protection of NH groups against deuterium exchange in these two states (Baldwin, 1993).

The use of stopped-flow techniques combined with CD spectroscopy, fluorescence, and electronic paramagnetic resonance permitted a “de-coupling” of the formation of a burst intermediate from the formation

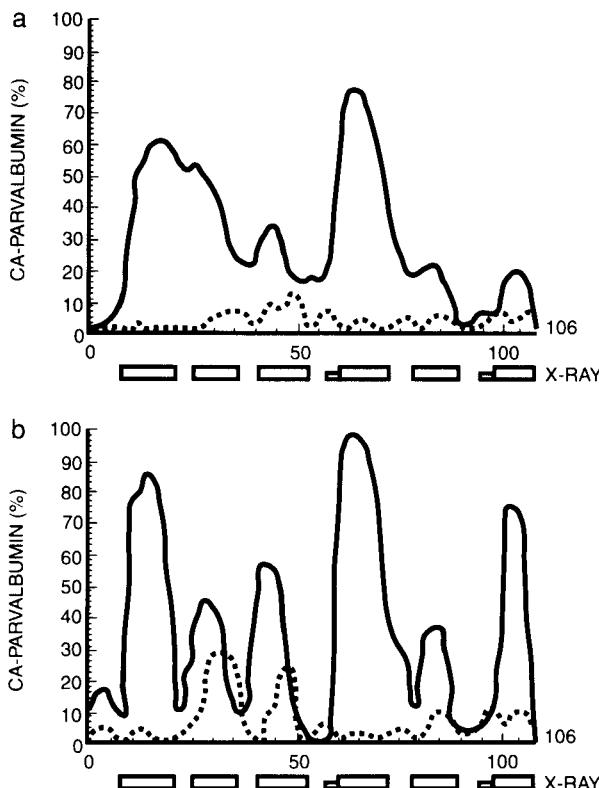


FIG. 33. Theoretically predicted secondary structure of carp parvalbumin for unfolded chain (a) and compact (globular) protein (b). Abscissa: number of residues from the N end of the chain; ordinate: probability of α (solid line) and β (dashed line) structure. X-ray positions of α helices (\blacksquare) and β strands (\square) are shown at the bottom of each figure. From Ptitsyn and Finkelstein (1983) *Biopolymers*. Copyright © 1983 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

of the kinetic molten globule. In fact, although a substantial part of protein secondary structure is restored within the first 20 msec of protein folding (see Section IV,A,1) Semisotnov *et al.* (1987) have shown that the molten globule-like kinetic intermediate in carbonic anhydrase is formed within 200 msec, i.e., much later than secondary structure which in this protein is also formed within the first 20 msec (Ptitsyn and Semisotnov, 1991; Kuwajima *et al.*, 1993).

Three different experimental techniques have been applied: (1) increase of fluorescence of randomly attached dansyl labels at the excitation of tryptophans which reflects the increase in energy transfer on condensation of the molecule; (2) decrease in the ESR label signal which

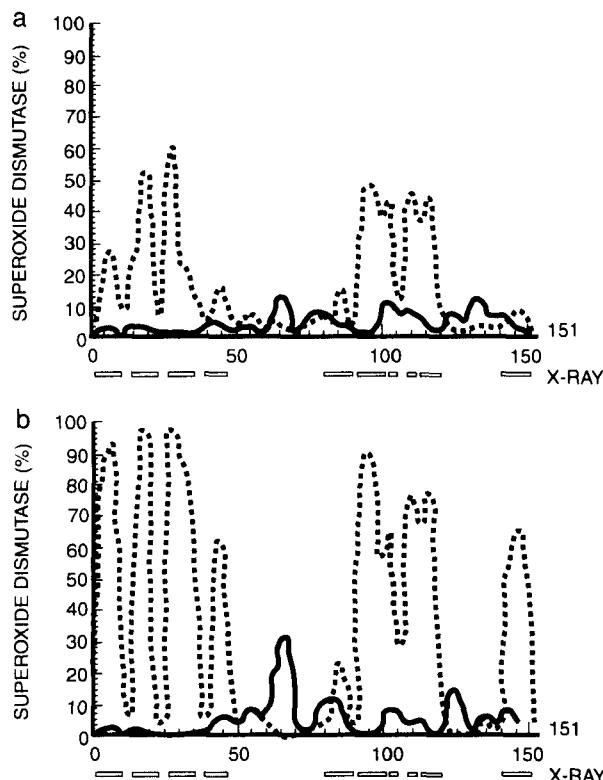


FIG. 34. Theoretically predicted secondary structure of superoxide dismutase for unfolded chain (a) and compact (globular) protein (b). The notation is the same as described in the legend to Fig. 33. From Ptitsyn and Finkelstein (1983) *Biopolymers*. Copyright © 1983 John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

reflects an immobilization of the spin label by the protein matrix; and (3) a further increase in fluorescence of a nonpolar probe (ANS), which reflects its binding to water-exposed hydrophobic clusters. Figure 36 shows that the first two approaches (see Fig. 36a) have similar kinetics with half-times ($t_{1/2}$) of 30 to 40 msec which coincide with the half-time of the slow phase of ANS binding (Fig. 36b). These half-times are much larger than those for the formation of secondary structure which are less than a few milliseconds. On the other hand, they are much smaller than the half-times for restoration of the native structure monitored by the increase of high-field resonances of ^1H NMR spectrum as well as by the decrease in ANS fluorescence on its release from a protein ($t_{1/2} = 140$ sec for both of these processes). Even slower ($t_{1/2} = 600$ sec) is the restoration of the rigid tertiary structure (monitored by ellipticity at 270 nm) and of the esterase activity of this protein (Semisotnov *et al.*, 1987).

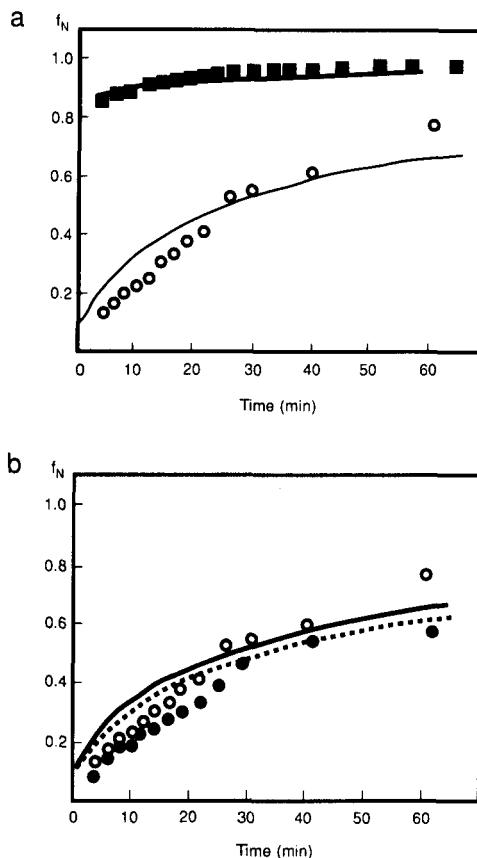


FIG. 35. First evidence of the molten globule kinetic intermediate in protein folding: (a) Kinetics of restoration of different parameters at the refolding of bovine carbonic anhydride B induced by its dissolution from 5.45 to 0.97 M GdmCl at 20°C. Bold curve, far-UV ellipticity ($[\Theta]_{220}$); ■, reduced viscosity ($\eta - \eta_0)/\eta_0 c$ (η_0 , viscosity of solvent; η , viscosity of solution; c -concentration); thin curve, near-UV ellipticity $[\Theta]_{270}$; ○, esterase activity. (b) Kinetics of restoration of $[\Theta]_{270}$ (dashed line) and esterase activity (●) upon the refolding of carbonic anhydride from the molten globule state induced by its dissolution from 1.97 to 0.97 M GdmCl at 20°C. The corresponding data for U → N transition (see a) are shown for comparison (solid line, $[\Theta]_{270}$; ○, esterase activity). Adapted with permission from Dolgikh *et al.*, (1984).

Of course, all three approaches used to detect the second kinetic intermediate are rather empirical. There are no clear reasons for the remarkable difference between energy transfer from tryptophans to randomly attached dansyl labels in carbonic anhydride (Semisotnov *et al.*, 1987, 1991a) or β -lactoglobulin (Ptitsyn and Semisotnov, 1991) and the energy transfer from a single tryptophan to a single fluorescent group

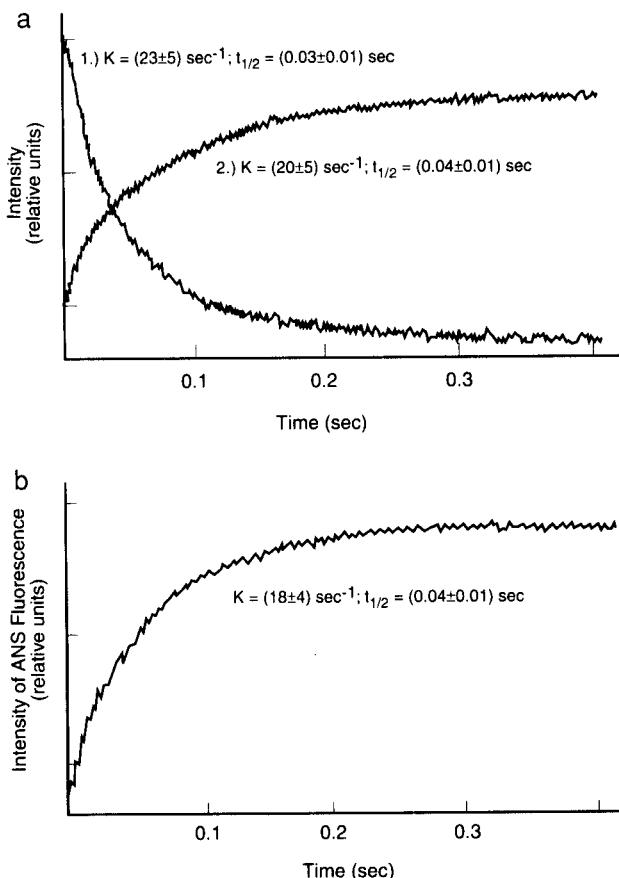


FIG. 36. Kinetics of folding bovine carbonic anhydrase B initiated by its dissolution from 8.5 to 4.2 M GdmCl at 23°C. (a) Decrease of intensity in the spin-label signal (1) and increase of intensity of dansyl label fluorescence at excitation of tryptophans (2). (b) Visible phase of increase in ANS fluorescence intensity (the ratio of molar concentrations of protein and ANS was 1:10). Rate constants (K) and half-time ($t_{1/2}$) of the processes are shown near the corresponding curves. Reproduced with permission from Semisotnov *et al.* (1987).

in an immunoglobulin fragment (Kawata and Hamaguchi, 1991) and cytochrome c (Elöve *et al.*, 1992) (see above, Section IV,A,2). However, the paper of Semisotnov *et al.* (1987), showed for the first time the existence of clearly visible processes of protein folding which occur within 0.1–0.2 sec, i.e., much slower than the formation of a substantial secondary structure but much faster than the formation of a tertiary structure. It is especially important that one of these processes is the binding of ANS

which has been shown to be a specific tool for monitoring the molten globule state (Semisotnov *et al.*, 1987, 1991a). This permits the suggestion that the second kinetic intermediate is similar to the molten globule state.

Since 1987, ANS binding has been used to monitor the formation of kinetic intermediates for a number of other proteins (see, e.g., Ptitsyn *et al.*, 1990; Goldberg *et al.*, 1990; Varley *et al.*, 1993; Matthews, 1993). The kinetics of ANS binding is qualitatively similar in all proteins studied: a burst stage of binding, followed by a further fast increase and then a slow release (see, e.g., Fig. 31 in Section IV,A,2). The maximal binding is achieved usually in 0.1–1 sec, while a subsequent release can require from a few seconds to more than 1000 sec. This behavior is explained (Semisotnov *et al.*, 1987, 1991a) by strong ANS binding to a kinetic molten globule intermediate, while a subsequent conversion of a molecule into the native state leads to the release of ANS. Based on these data it was concluded that the molten globule is a “general kinetic intermediate of protein folding” and that “secondary structure is formed before the molten globule state” (Ptitsyn *et al.*, 1990).

Three stages of protein folding obtained by optical methods—the burst formation of the premolten globule state (i.e., secondary structure formation and a partial condensation), the fast formation of the molten globule state and the slow formation of the native state—give a framework for the studies of structural features of kinetic intermediates. For example, it was shown that the formation of the first native-like antigenic determinant in the β_2 subunit of *E. coli* tryptophan synthase has a half-time of approximately 10 sec (Murry-Brelier and Goldberg, 1988; Blond-Elguindi and Goldberg, 1990), which is much slower than the formation of the molten globule state (~ 1 sec) but much faster than the formation of tertiary structure (~ 1000 sec) (see Goldberg *et al.*, 1990). It was concluded that the first epitope is formed within the molten globule state of this protein. In a similar way Matthews and collaborators (Touchette *et al.*, 1986; Kuwajima *et al.*, 1991; Matthews, 1993) have shown that the second stage of ANS binding is accompanied by the native-like packing of two tryptophan residues. As the protein cannot bind the inhibitor of its active site in this state, this state is not the native state but rather is a kinetic intermediate with some elements of the specific tertiary structure.

2. Partial Formation of Stable Folding Patterns

Systematic studies of the structure of kinetic intermediates become possible with the development of a pulsed hydrogen-exchange method by Udgaonkar and Baldwin (1988) and Roder *et al.*, (1988). This approach (see Fig. 37) consists of the following steps (Roder and Elöve, 1994): (1) the protein is unfolded in a mixture of D₂O with a strong

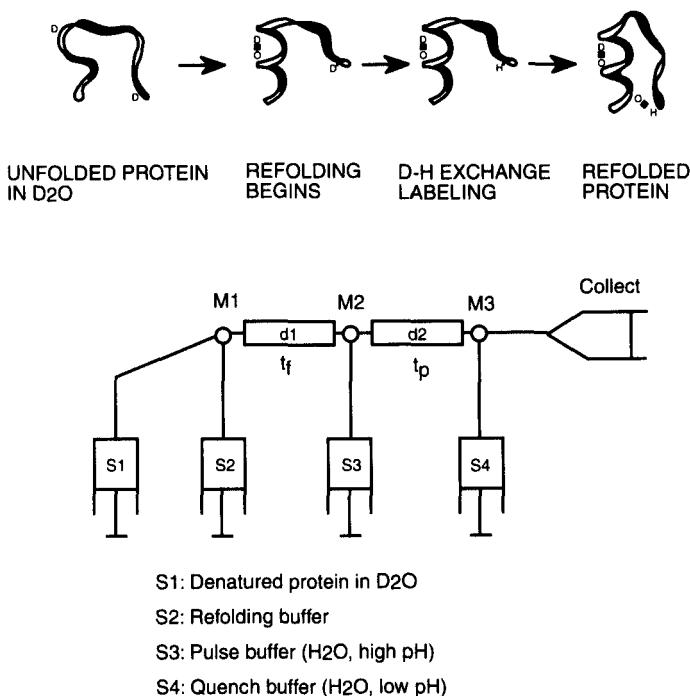


FIG. 37. Schematic illustration of the pulse hydrogen exchange method. *Top:* Basic principle of the method in which part of the D atoms are protected from the exchange by formation of α -helical hydrogen bonds and another part (which is involved in tertiary hydrogen bonds) remains unprotected at early stages of protein folding. *Bottom:* Scheme of the quenched-flow apparatus with three mixing stages (M1–M3) and two variable time delays (d₁, d₂). t_f is the refolding time, t_p is the duration of the labeling pulse. Reproduced with permission from Roder and Elöve (1994).

denaturing agent (urea or GdmCl), resulting in the replacement of all protons of NH groups by deuterium; (2) protein folding is initiated by rapid dilution of the denaturant by D₂O buffer; (3) after a given folding delay, t_f (which can be varied), the partially refolded protein is mixed with H₂O buffer at high pH. During this “labeling pulse” all exposed ND groups become protonated, i.e., transformed into NH groups, while those which are involved in stable hydrogen-bonded structures remain deuterated; (4) after a pulse time, t_p , the exchange is quenched by rapid lowering of the pH, while protein folding is allowed to go to completion; and (5) the localization of NH and ND groups trapped in the refolded protein is analyzed by 2D NMR spectroscopy. If these experiments are done at different times in protein folding, the time course of the forma-

tion of stable (relatively slowly fluctuating) hydrogen bonds in this process can be reconstructed.

This method has been applied to the folding of a number of different proteins including ribonuclease A (Udgaonkar and Baldwin, 1988, 1990), cytochrome c (Roder *et al.*, 1988; Elöve and Roder, 1991; Elöve *et al.*, 1994), barnase (Bycroft *et al.*, 1990; Matouschek *et al.*, 1992b), ubiquitin (Briggs and Roder, 1992), T4 lysozyme (Lu and Dahlquist, 1992), hen lysozyme (Radford *et al.*, 1992a), interleukin-1 β (Varley *et al.*, 1993), apomyoglobin (Jennings and Wright, 1993), ribonuclease T1 (Mullins *et al.*, 1993), and staphylococcal nuclease (Jacobs and Fox, 1994). The results of these studies have been summarized in a review by Baldwin (1993).

Usually this method reveals the existence of an early kinetic intermediate in which many (or even all) NH groups, involved in a secondary structure, are partly protected from hydrogen exchange, while NH groups, involved in tertiary hydrogen bonds, remain unprotected. The exceptions are ubiquitin (Briggs and Roder, 1992) and ribonuclease T1 (Mullins *et al.*, 1993) in which tertiary NH groups are protected almost simultaneously with NH groups involved in a secondary structure. The time for partial protection of NH groups in secondary structures may vary over at least two orders of magnitude—from less than 5 msec for apomyoglobin (Jennings and Wright, 1993) to \sim 1 sec for interleukin-1 β (Varley *et al.*, 1993). Usually at this early stage of protein folding NH groups are only partly protected. For example, for ribonuclease A (Udgaonkar and Baldwin, 1990) the intermediate formed within 50 msec has protection factors of about 10, but at 400 msec protection factors are too large to be measured (>1000). Protection factors measured at this stage of protein folding, e.g., for ribonuclease A (Udgaonkar and Baldwin, 1990) and T4 lysozyme (Lu and Dahquist, 1992), are between 10 and 100, i.e., comparable to those formed for equilibrium molten globules (see Section I,B,1).

In some cases, e.g., for ubiquitin (Briggs and Roder, 1992), almost all NH groups involved in secondary structure are protected more or less simultaneously. However, in other proteins, such as cytochrome c (Roder *et al.*, 1988; Elöve and Roder, 1991), lysozyme (Radford *et al.*, 1992a), interleukin-1 β (Varley *et al.*, 1993), and apomyoglobin (Jennings and Wright, 1993), some regions with secondary structure are protected earlier than others.

Figure 38 presents the data of Roder *et al.* (1988) on the protection kinetics of different NH groups in cytochrome c. One can see that NH groups belonging to the N- and C-terminal helices are protected much faster than those belonging to the other two helices and to the tertiary

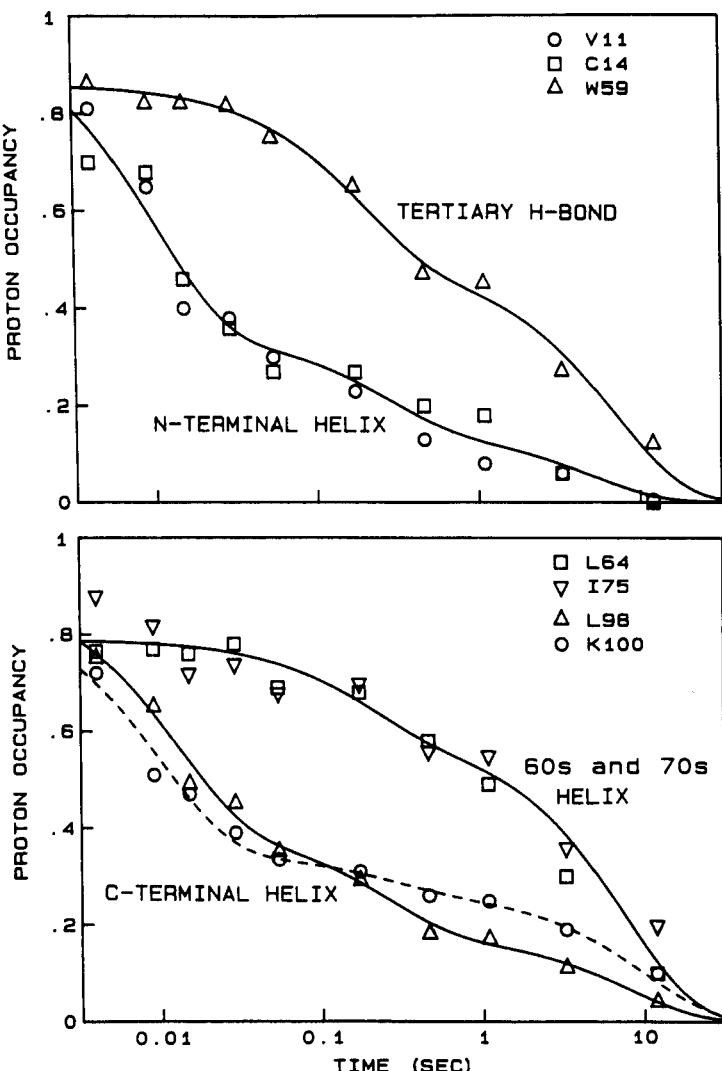


FIG. 38. Protection of NH groups for some residues of cytochrome c plotted as the dependence of proton occupancy on the refolding time (logarithmic scale). The NH group of Trp-59 is involved in a tertiary hydrogen bond between its indole group and the heme, other NH groups belong to the backbone of some residues selected to represent N-terminal α helix (Val-1, Cys-14), C-terminal helix (Leu-98, Lys-100), as well as two other helices involving 60's and 70's residues (Leu-64 and Ile-75, respectively). Reproduced with permission from *Nature*, Roder *et al.* (1988). Copyright 1988 Macmillan Magazines Limited.

hydrogen bond between the indole group of Trp-59 and the heme propionate side chain. N- and C-terminal helices form a complex in the native protein (see Fig. 8), and the close similarity between kinetics of their protection suggests that their docking occurs early in protein folding (Roder *et al.*, 1988). This conclusion has been confirmed by the identical pH dependence of NH protection in N- and C-terminal α helices (Elöve and Roder, 1991). Moreover, it was shown (Wu *et al.*, 1993) that a heme-containing N-terminal fragment (residues 1–38) forms a noncovalent complex with a synthetic peptide corresponding to the C-terminal helix (residues 87–104). The formation of this complex (see Fig. 8) is accompanied by a substantial increase in the amplitudes of the far-UV CD spectra, which suggests that helicity of these fragments increases on docking. Moreover, the formation of a complex leads to changes in the NMR spectrum, especially longitudinal relaxation times, which are attributed to the close distance between the C-terminal fragment and the heme (which is covalently attached to the N-terminal fragment).

It should be emphasized that both N- and C-terminal α helices are also protected from hydrogen exchange in the equilibrium molten globule state of cytochrome c at pH 2 and high ionic strength (Jeng *et al.*, 1990) (see Section III,A, Fig. 9). The only difference is that in the equilibrium molten globule state the 60's helix is protected as well; however, the failure to form this helix in the kinetic intermediates appears to be due to the presence of a nonnative histidine–heme ligand in the unfolded state that becomes trapped in a partially folded intermediate (Elöve and Roder, 1991; Elöve *et al.*, 1994).

In hen lysozyme (Radford *et al.*, 1992a) protection of NH groups occurs in two phases. The first phase, with a characteristic time $\tau = 7 \pm 4$ msec, is about the same in both α -helical and β -structural subdomains (see Fig. 7 for the 3D structure of α -lactalbumin which is very similar to that of lysozyme). About 40% of NH groups in the α -helical subdomain and about 25% of those in the β -structural subdomain are protected at this stage. Remaining NH groups are protected much more slowly and the average rates for their protection are four times faster for the α -helical domain than for the β -structural domain (e.g., for α helical A–D and the C-terminal 3_{10} helix involved in the helical subdomain the average time constant $<\tau> = 80$ msec, while for β sheet and 3_{10} helix included in the β -structural subdomain $<\tau> \approx 350$ msec). On the other hand, the near-UV ellipticity reappears in two phases with $\tau \approx 10$ and 300 msec, that are similar to the fast and slow stages of NH protection in the β -structural subdomain. This (as well as the behavior of some NH groups forming nonhelical hydrogen bonds) suggests that the α -helical

subdomain initially folds into the molten globule state and becomes native only together with the slowly folding β -structural domain. It is important to emphasize that helices B and C belonging to the α -helical subdomain are protected in the equilibrium molten globule state of α -lactalbumin, which is homologous to lysozyme (Baum *et al.*, 1989; Dobson, 1991; Chyan *et al.*, 1993) (see Section II,B,1).

Another interesting example is interleukin-1 β (Varley *et al.*, 1993). Folding of this protein is much slower than that of cytochrome c; initial protection is observed only within 1 sec and full protection requires well over 25 sec. Figure 39 shows where residues protected at different stages of folding are located in the 3D structure of this protein, demonstrating (see Fig. 39a) that NH groups protected within the first few seconds (with $t_{1/2} \approx 1$ sec) belong mainly to four β strands (6–9) that are adjacent along the chain. Two of them (strands 8 and 9) are included in the antiparallel β -barrel which is the main “framework” of this protein, while two others (6 and 7) form one of three β hairpins covering the “bottom” of this β barrel. It appears as if the formation of the stable structure of interleukin begins from two virtually independent β hairpins located in the same region of 3D structure (see Fig. 39b). It is interesting to note that almost all residues, not protected within the first 1 sec, are protected only much more slowly than 25 sec, probably together with restoration of protein tertiary structure whose half-time is ~ 20 min.

The application of the pulsed hydrogen-labeling approach to apomyoglobin (Jennings and Wright, 1993) was very effective (Fig. 40). Unlike interleukin, this protein folds very fast: a substantial part of the NH groups is protected at the burst stage, i.e., within the first 6 msec (the dead time of the experiment), and the protein is completely refolded within 4 sec. Among 38 NH protons that can be used as probes, 20 are protected within the first 6 msec. All of them belong to three helices, A, G, and H, and to a part of helix B, i.e., to the same part of the protein, which is protected in the equilibrium molten globule state (Hughson *et al.*, 1990). The rest of helix B is protected much more slowly (within ~ 1 sec) and helices C and E, as well as loop CD, even more slowly (within ~ 2 sec). A close similarity in protection patterns of the kinetic intermediate and the equilibrium molten globule state provides very convincing evidence that the structures of these two intermediates are similar.

The far-UV CD data on this protein show the restoration of about 65% of its ellipticity at 222 nm in the first 4 msec of folding, while the other 35% is restored with a relaxation time of ~ 1 sec. The comparison of stabilities of the burst intermediate and the native protein against urea shows that the burst intermediate is substantially less stable; its

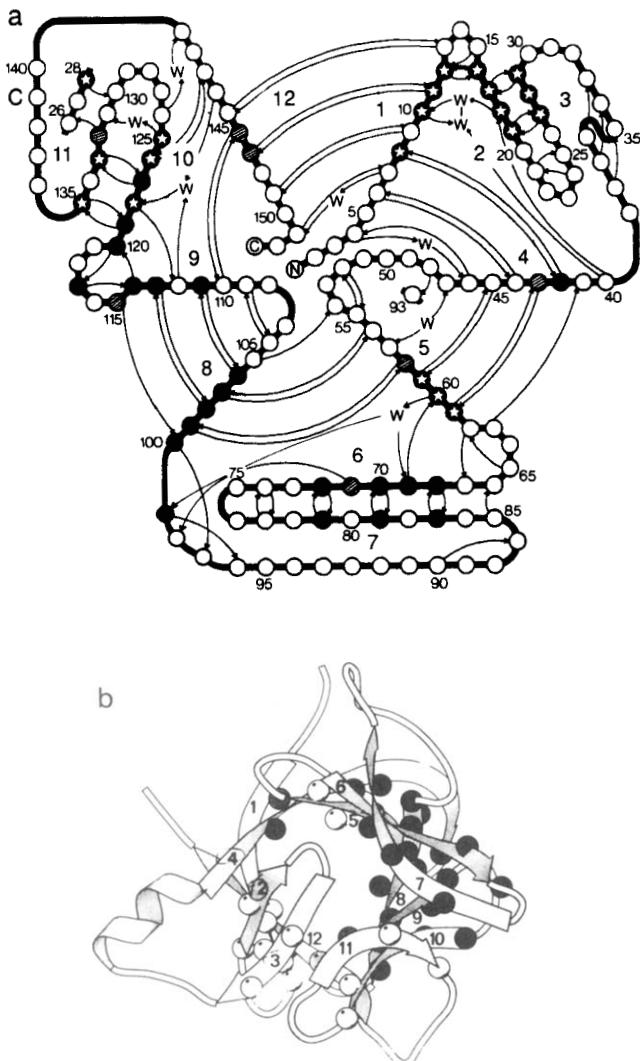


FIG. 39. (a) Schematic presentation of hydrogen bonds and the topology of interleukin-1 β illustrating sequential formation of a stable secondary structure. Groups protected with half-times of 0.7–1.5, 15–25, and ≥ 25 sec are shown by solid circles, hatched circles, and solid circles with white stars, respectively. Hydrogen bonds are marked by arrows from the donor NH groups to the acceptor C = O group or water (W). (b) Schematic presentation of the 3D structure of interleukin-1 β . The residues protected with $t_{1/2} = 0.7\text{--}1.5, 15\text{--}25$, and ≥ 25 are shown as black, gray, and white circles, respectively. β strands 1, 4, 5, 8, 9, and 12 form a six-stranded β barrel that is closed off at one end by three β hairpins 2–3, 6–7, and 10–11. The open face of the barrel is located on the back of the structure. Reproduced with permission from Varley *et al.* (1993). *Science* **260**, 1110–1113. Copyright 1993 by the AAAS.

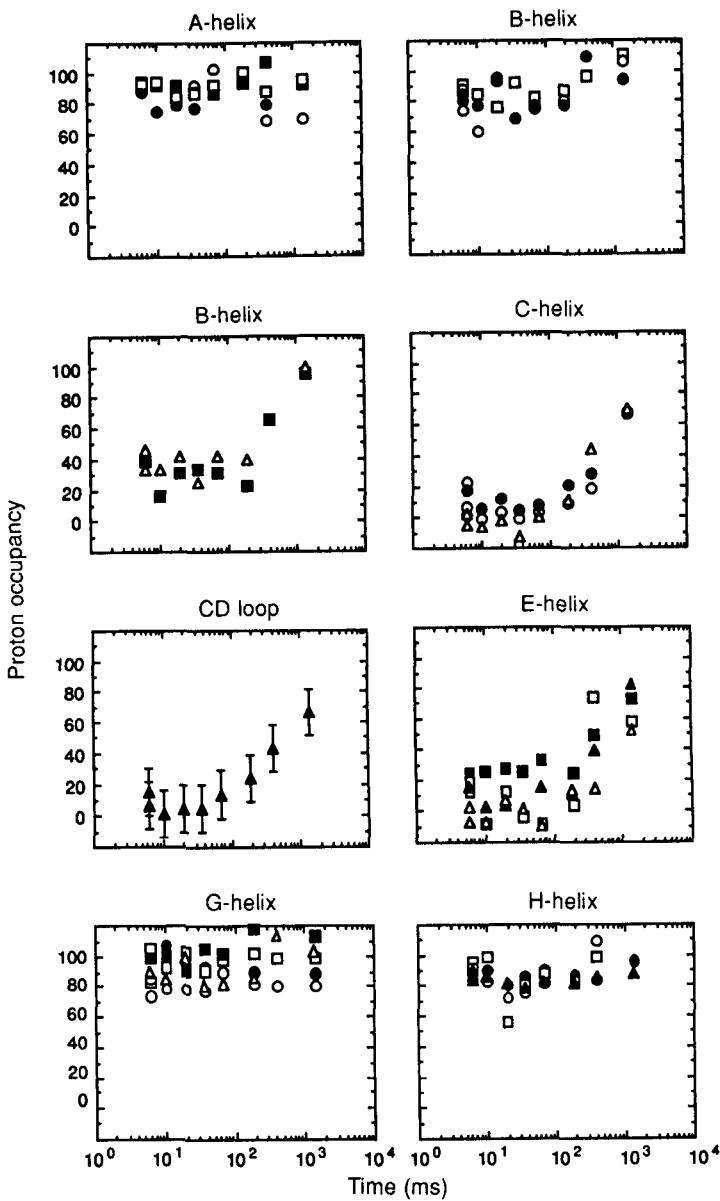


FIG. 40. Representative data on the kinetics of NH proton protection in the apomyoglobin backbone from exchange with solvent deuterium. Representative experimental errors are shown for the CD loop. The figure shows the very fast protection of helices A, G, H, and a part of helix B, a slower protection of another part of helix B, and an even slower protection of helices C and E and loop CD. Reproduced with permission from Jennings and Wright (1993). *Science* **262**, 892–896. Copyright 1993 by the AAAS.

estimated stability compared to the unfolded state is as small as 2.5 kcal/mol.

In summary, we conclude that the pulsed hydrogen exchange method demonstrates that some regions of secondary structure are protected at early stages of protein folding, in some cases even within a few milliseconds. The extent of this protection is similar to that observed in the equilibrium molten globule state. Early protected α helices and β strands belong to the native helical complexes or β structure, which confirms the formation of a native-like structural organization at the early stages of protein folding. Moreover, in apomyoglobin, regions protected in a kinetic intermediate coincide exactly with those protected in the equilibrium molten globule state. In cytochrome c at least two of the three regions protected in the equilibrium molten globule state are also protected in a kinetic intermediate, and two of the four helices protected in a kinetic intermediate of lysozyme are also protected in the equilibrium molten globule state of its close homolog α -lactalbumin. This suggests that the kinetic intermediate with partly protected secondary structure may be similar to the equilibrium molten globule state.

The sequence of events leading to the formation of native protein structure is, however, more complicated than described above. The protection from hydrogen exchange of at least some of the amides is independent of pH and the length of labeling pulse. This is inconsistent with a sequential folding mechanism which would imply that each NH group is increasingly protected upon protein folding. Therefore, these observations have been considered as evidence for parallel folding pathways, with some leading to a more rapid protection of NH groups than others. If it takes place over a long time, it usually (but not always; see Section IV,C) can be explained by proline *cis-trans* isomerization (see Schmid, 1992, 1993, for reviews). However, the "splitting" of folding pathways has also been observed on a short time scale (seconds or less), for example, for ribonuclease A (Udgaonkar and Baldwin, 1990), cytochrome c (Elöve and Roder, 1991; Elöve *et al.*, 1992), lysozyme (Radford *et al.*, 1992a), and ribonuclease T1 (Mullins *et al.*, 1993). The existence of parallel folding pathways has also been proposed for dihydrofolate reductase based on inhibitor-binding studies (Matthews, 1993).

Three explanations of these multiple folding pathways have been proposed. The first is the assumption that proline *cis-trans* isomerization which typically occurs in about 20 sec, can also contribute to the splitting of folding pathways on a time scale of seconds (Elöve and Roder, 1991; Briggs and Roder, 1992; Roder and Elöve, 1994). In the case of cytochrome c, trapped nonnative heme ligands have also been shown to cause heterogeneous folding behavior (Elöve and Roder, 1991; Elöve *et*

et al., 1994; Roder and Elöve, 1994). The second explanation is the existence of *cis* peptide bonds in nonproline residues (Udgaonkar and Baldwin, 1990), whose lifetimes have been determined as several seconds (Portnova *et al.*, 1970). The third, and most interesting, explanation is that the fast hydrophobic collapse of protein chains may produce a mixture of structured molten globules with the native-like folding pattern and unstructured molten globules without well-defined folding patterns, and that unstructured molten globules need more time to achieve their native-like folding patterns (Radford *et al.*, 1992a; Baldwin, 1993). This situation may result in kinetic "traps" and parallel folding pathways.

C. Formation of Native Tertiary Structure

The last stage of protein folding is the formation of the rigid native tertiary structure. This stage can be monitored by the restoration of native NMR and near-UV CD spectra, by the restoration of native protection of all NH groups, by the release of ANS, etc. The process of the formation of rigid tertiary structure is accompanied by the restoration of protein enzymatic activity which can also be used to monitor this process. The time scale of the slowest (rate-limiting) step of protein folding may vary from about a few seconds (see, e.g., Elöve *et al.*, 1992; Radford *et al.*, 1992a) to thousands of seconds (see, e.g., McCoy *et al.*, 1980; Semisotnov *et al.*, 1987; Goldberg *et al.*, 1990; Ptitsyn and Semisotnov, 1991).

The reason for this "superslow" folding usually is *cis-trans* isomerization of proline residues (Schmid, 1992, 1993). In native proteins each proline is in either the *trans* or the *cis* conformation and these conformations in a number of cases are essential for the formation of the native 3D structure. In unfolded chains prolines are no longer fixed in one of these conformations by long-range interactions and therefore can isomerize to reach a statistical equilibrium between the two states of each proline. This equilibrium can be achieved only on a rather long time scale (usually ~ 20 sec) as the energy barrier between the *cis* and *trans* states of proline is as large as ~ 20 kcal/mol (see, e.g., Schmid, 1992). The formation of the native 3D structure from an unfolded chain is dependent on the slow isomerization of all prolines, for which unique *cis* or *trans* conformations are essential in the 3D structure. This usually is the slowest (rate-limiting) step of protein folding.

For some proteins, e.g., for carbonic anhydrase (Semisotnov *et al.*, 1987, 1990) and phosphoglycerate kinase (Betton *et al.*, 1985; Semisotnov *et al.*, 1991b), the formation of the native 3D structure requires thousands of seconds, which is much slower than usual proline *cis-trans* isomerization. The near-UV CD spectrum of bovine carbonic anhydrase B

is completely restored only after 2500 sec with a half-time of about 600 sec (Semisotnov *et al.*, 1987). This can be explained (Semisotnov *et al.*, 1990) by the correlated *cis-trans* isomerization of two or more proline residues. Alternatively, nonnative proline isomers can be trapped in stable intermediates, resulting in very high barriers to isomerization (see, e.g., Schmid, 1992).

It is obvious that the proline-dependent barrier for protein folding is not the intrinsic feature of this process. This is confirmed by a number of observations that proteins with "essential" prolines usually fold as a mixture of two or more sorts of molecules, one of them folds much faster than others. The explanation is that the fast-folding molecules have native conformations of essential prolines also in the unfolded state, while the others have one or more prolines in nonnative conformations (Kim and Baldwin, 1982, 1990; Schmid, 1992). The removal of prolines either by using a protein from another species or by site-directed mutagenesis can lead to the elimination of slowly folding molecules (see Kim and Baldwin, 1990; Schmid, 1992, for references).

The nonintrinsic character of the barrier between the kinetic molten globule and the native state has been clearly demonstrated in an interesting paper by Englander's group (Sosnik *et al.*, 1994). It is known that cytochrome c folds at neutral pH within 10 sec with the accumulation of the kinetic molten globule state having partly protected NH groups (see Fig. 41). Sosnik *et al.* have shown, however, that this kinetic intermediate does not accumulate at low pH, where cytochrome c folds with a time constant of ~ 15 msec simultaneously with protein condensation in a single kinetically unresolved step. The reason is that the intermediate accumulated at neutral pH is due to the nonnative heme liganding of the histidines. This is impossible at low pH because histidines are ionized. The dramatic decrease in the barrier between the molten globule and native states at low pH is emphasized by the observation that cytochrome c folds from the molten globule state within the dead time of the instrument (3 msec).

These results clearly show that the packing barrier between the molten globule and the native states can be strongly reduced if the correct packing is not inhibited by artificial barriers like nonnative proline conformations or nonnative liganding.

An important contrasting observation was that in horse (Betton *et al.*, 1985), as well as in pig and yeast (Semisotnov *et al.*, 1991b), phosphoglycerate kinases the slow-folding phase is not connected with proline isomerization. A convincing experiment supporting this conclusion is based on the "double-jump" technique in which a native protein is jumped into unfolding conditions for different time intervals and then jumped

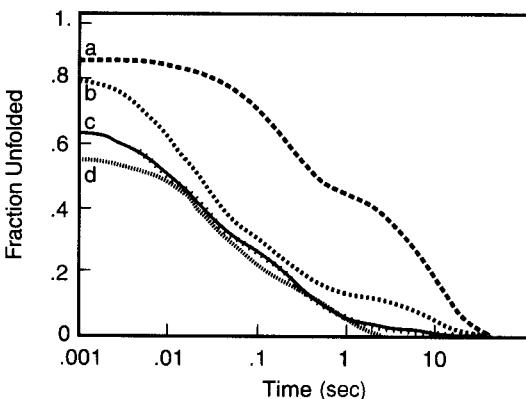


FIG. 41. Kinetics of refolding for oxidized horse cytochrome c following a jump from 4.2 to 0.7 M GdmCl at pH 6.2 and 10°C. (a) NH protection for the helices encompassing 60's and 70's residues; (b) NH protection for the N- and C-terminal helices; (c) intensity of fluorescence of the single Trp; (d) ellipticity at 222 nm. The kinetic curve for the near-UV ellipticity (at 289 nm), not shown, is quite close to that of NH protection for 60's and 70's helices (Roder and Elöve, 1994). Reproduced with permission from Elöve *et al.* (1992). Copyright 1992 American Chemical Society.

back into the native conditions. If the reason for the slow phase is some slow isomerization in unfolded chains, the amplitude of this phase will increase with the increase in the protein's incubation time under unfolding conditions, as is the case for a number of proteins, e.g., for carbonic anhydrase (Semisotnov *et al.*, 1990). However, in phosphoglycerate kinases the amplitude of the slow phase is independent of the time of its incubation under unfolding conditions which rules out the possibility that the slow phase is due to any slow isomerization in unfolded chains including proline *cis-trans* isomerization.

An intriguing possibility is that this slow but proline-independent formation of tertiary structure may be connected with traps in protein folding (see Section IV,B,2). In fact, since proteins usually unfold quickly at high concentrations of strong denaturants, all protein molecules should be equally unfolded after the first jump, whereas at the second jump a part of them can fold quickly while another part can fall into deep traps and need longer times for escape.

In larger proteins consisting of two or more "folding units" the rate-limiting step of folding can involve association or repacking of these units as it has been demonstrated for the α subunit of tryptophan synthase (Beasty *et al.*, 1986; Hurle *et al.*, 1986; Matthews, 1993) and for dihydrofolate reductase (Perry *et al.*, 1989; Garvey and Matthews, 1989; Matthews, 1993).

V. MECHANISM OF PROTEIN FOLDING

A. *Folding Intermediates*

1. *Kinetic Intermediates*

The optical studies of protein folding summarized in Section IV clearly show the existence of at least two kinetic intermediates: the premolten globule state with substantial secondary structure and partial condensation of the chain, and the molten globule state with a much more stable secondary structure and much greater compactization of the chain. Recent data in which optical methods have been compared with pulse hydrogen exchange techniques strongly support this conclusion.

Figure 41 presents the data on the folding kinetics for cytochrome c (Elöve *et al.*, 1992) obtained by three different methods: far-UV ellipticity $[\Theta]_{222}$, intensity of fluorescence of the single Trp (quenched by the heme in the native state), and protection of NH groups involved in helical structure. Curves a, b, and c (or d) show the existence of at least three main stages of folding and consequently at least two kinetic intermediates.

The first intermediate forms within the dead time of experiments (4 msec). The changes of its far-UV ellipticity and Trp fluorescence are ~ 45 and $\sim 35\%$ of their total changes during protein folding, respectively. This suggests that the first intermediate has a pronounced secondary structure and is partly compact. However, its NH groups are not significantly protected, suggesting that it has no stable features of 3D structure. This partly compact intermediate with fluctuating secondary structure can be called the kinetic premolten globule (or partly folded) state of a protein molecule.

The second intermediate forms within ~ 30 msec and is characterized mainly by a substantial protection of NH groups belonging to the N- and C-terminal α helices (the curves of protection of these groups become similar to those of the far-UV ellipticity and the fluorescence). There is a good evidence (see Section IV,B,2) that N- and C-terminal helices dock to each other at this stage, which indicates the restoration of at least a part of the native folding pattern. However, two other helices (which embrace 60's and 70's residues) remain practically unprotected, and the protein still has no rigid tertiary structure. This intermediate with at least a part of the native folding pattern but without a native tertiary structure can be called the kinetic molten globule state.

The full restoration of the native tertiary structure of cytochrome c, monitored by the protection of other NH groups (not belonging to N-

and C-terminal helices) and by restoration of near-UV CD spectrum, takes about 10 sec.

A qualitatively similar picture has also been observed for hen egg lysozyme (Radford *et al.*, 1992a). Figure 42 presents the comparison of different kinetic curves for this protein. Again, at least three stages of protein folding and, consequently, at least two kinetic intermediates, can be observed. Approximately 65% of the change in the far-UV ellipticity takes place within the dead time of the experiment (2 msec), which is not accompanied by any visible protection of NH groups or any changes of near-UV ellipticity. It is the first intermediate with a pronounced but fluctuating secondary structure and without a tertiary structure (there are no data on the degree of its compactness). The second intermediate forms within 70 msec when the far-UV ellipticity reaches its maximum value. As shown by Chaffotte *et al.* (1992), the maximum of the far-UV ellipticity in lysozyme reflects formation of the intermediate in which full secondary structure is achieved. The subsequent decrease in ellipticity is due to contributions of S-S bonds or aromatic groups which become effective after the formation of the rigid tertiary structure. NH groups of the α -helical subdomain are almost entirely protected at this stage, while the protection of these groups in the β -structural subdomain is substantially lowered. However, this does not mean that the α -helical subdomain is completely folded, as the near-UV ellipticity at this stage is far from its native value, although most tryptophans in lysozyme belong to its α -helical subdomain (Radford *et al.*, 1992a). The further increase in the near-UV ellipticity occurs in parallel with the protection of NH

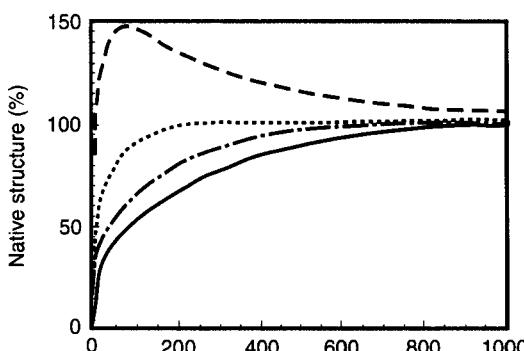


FIG. 42. Kinetics of refolding for hen egg lysozyme following a jump from 6 to 0.6 M GdmCl at pH 5.2 and 20°C. Abscissa axis, time in milliseconds. (---) ellipticity at 225 nm, (· · ·) NH protection for α -helical subdomain, (· · ·) NH protection for β structural subdomain, (—) ellipticity at 289 nm. Reproduced with permission from *Nature*, Radford *et al.* (1992a). Copyright 1992 Macmillan Magazines Limited.

groups belonging to the β subdomain, which suggests that the α subdomain "waits" in the molten globule state until the β subdomain becomes folded before transforming together to the native state.

Again, we observe the first intermediate with a fluctuating secondary structure and the second intermediate in which at least a substantial part of the molecule forms the molten globule state. Note that in both cases the first intermediate formed at least an order of magnitude faster than the second intermediate (a few milliseconds compared to about 80 msec), while the second intermediate is formed at least an order of magnitude faster than the native state (which forms in up to 10 sec). This confirms that we are dealing with two intermediates which sequentially accumulate on protein folding.

Even more different are the kinetic curves for various characteristics of the third protein, interleukin-1 β (Varley *et al.*, 1993). About 90% of its native far-UV ellipticity is restored within the first 10 msec of protein folding, which is accompanied by a substantial increase in the intramolecular quenching of Trp fluorescence and by a substantial ANS binding. Then, with a half-time of \sim 1 sec, NH groups of two β hairpins are protected, and much later (up to 10^3 sec) the protein folds into its native state.

Apomyoglobin appears to be an exception since \sim 65% of far-UV ellipticity and NH protection for helices A, G, and H are restored within the dead time of the experiment (Jennings and Wright, 1993). However, this does not necessarily mean that NH protection is simultaneous with a partial restoration of far-UV CD spectrum. It is quite possible that in this case NH groups are protected after the formation of a fluctuating secondary structure; both of these processes are simply very fast and take place within the dead time of experiments. Other related examples are T4 lysozyme (Lu and Dahlquist, 1992) and staphylococcal nuclease (Jacobs and Fox, 1994).

It follows that the prediction of the framework model (Ptitsyn, 1973) and early experimental data on the existence of at least two kinetic intermediates (Semisotnov *et al.*, 1987; Goldberg *et al.*, 1990) are generally confirmed by more recent experiments using the pulsed hydrogen exchange method. We conclude that two kinetic intermediates, the pre-molten globule state and the molten globule state, are typical for protein folding.

2. Equilibrium Intermediates

It is interesting to compare these kinetic intermediates with intermediates which may exist at different stages of equilibrium folding. In order to mimic the conditions of kinetic experiments it is worthwhile to study

equilibrium protein folding upon the dilution of solutions from high concentrations of urea or GdmCl. This point is very important as other denaturing conditions like high temperature and low or high pH may create protein conformations which are very far from being unfolded (see Sections II,A and III,A,1). In fact, even at high concentrations of urea or GdmCl proteins may not be completely unfolded (see Section III,A,2) but at least the degree to which they are unfolded is the same in kinetic and equilibrium experiments. It is unclear whether residual structures which may survive or form at high concentrations of urea or GdmCl can influence protein folding.

New possibilities for studying protein folding from concentrated solutions of urea or GdmCl are opened by size-exclusion chromatography [high-performance liquid chromatography (HPLC) or FPLC] which is especially useful for the transitions between two states differing substantially in their compactness (Corbett and Roche, 1984). If these states interconvert more slowly than the characteristic time of an experiment (15 to 30 min), size-exclusion chromatography can separate the two states in the transition region as two peaks with different exclusion volumes and can monitor what happens to each peak separately. Elution volumes can be calibrated to calculate Stokes radii; as a result these radii can be determined by size-exclusion chromatography with an accuracy quite comparable to that of the usual hydrodynamic methods (Uversky, 1993). Therefore, for slow transitions between two states with substantially different compactnesses, size-exclusion chromatography provides evidence on three independent processes. The first process is the conversion of protein molecules from a more compact state to a less compact one, the second process is the change in hydrodynamic volume of more compact molecules, and the third process is the change in hydrodynamic volume of less compact molecules. Figure 43 illustrates these three processes for the GdmCl-induced denaturation of Lysozyme (Uversky, 1993). Figure 43a shows the conversion of molecules from a more compact to a less compact state, while the insert in Fig. 43b demonstrates the GdmCl dependence of elution volumes for more compact and less compact molecules separately.

A number of proteins (e.g., myoglobin and lysozyme) have only one sigmoidal curve for urea- or GdmCl-induced transitions, which together with chromatography data shows the absence of any intermediates between the native and unfolded states in these proteins. For example, Fig. 43b shows that the denaturation of lysozyme (monitored by the decrease of circular dichroism in the aromatic region) is coupled in this case with an all-or-none transition between compact and less compact states (monitored by the bimodal distribution of elution volumes), with

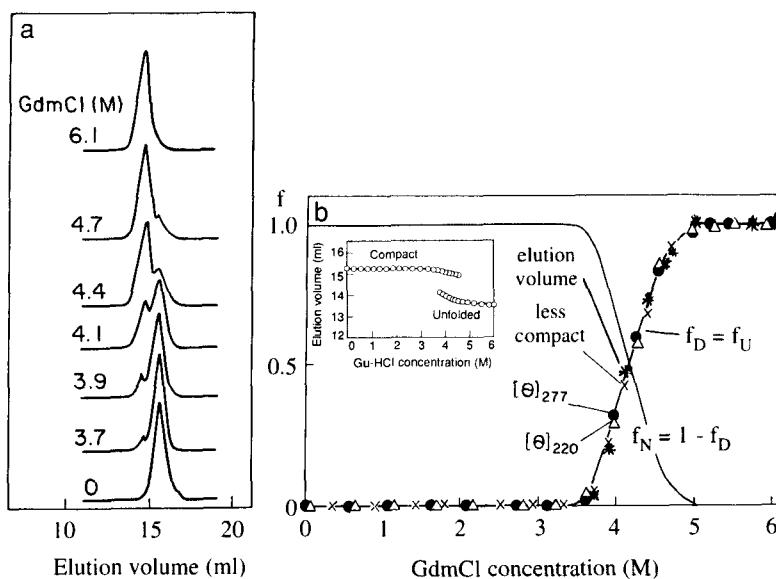


FIG. 43. (a) FPLC elution profiles of hen egg white lysozyme at 4°C and pH 6.8 for different concentrations of GdmCl. (b) Two-state unfolding of lysozyme by GdmCl: (●) the fraction of denatured molecules obtained from $[\Theta]_{277}$; (×) the fraction of less compact molecules obtained from the relative areas under the two peaks in (a); (*) fraction of unfolded molecules obtained from an average elution volume; (Δ) fraction of molecules without secondary structure obtained from $[\Theta]_{220}$. Insert: GdmCl dependences of elution volumes of compact and unfolded states. Adapted with permission from Uversky and Ptitsyn (1994). Copyright 1994 American Chemical Society.

the unfolding (monitored by a change of elution volume averaged over both compact and less compact states), and with the loss of secondary structure (monitored by the changes of circular dichroism in the peptide region). It means that only two states—the native and completely unfolded—exist in this protein and that the fraction of unfolded molecules (calculated from the area of the corresponding FPLC peak, from the average elution volume, or from the far-UV ellipticity) increases at the expense of the fraction of native molecules (calculated from the near-UV ellipticity).

However, in a number of proteins it is possible to decouple these processes and to see one or even two equilibrium intermediates between the native and the completely unfolded states. For example, Fig. 44 (Uversky and Ptitsyn, 1994) illustrates GdmCl-induced unfolding of β -lactamase at room temperature. Figure 44a shows a large decrease in the elution volume (i.e., an increase of hydrodynamic volume) with an

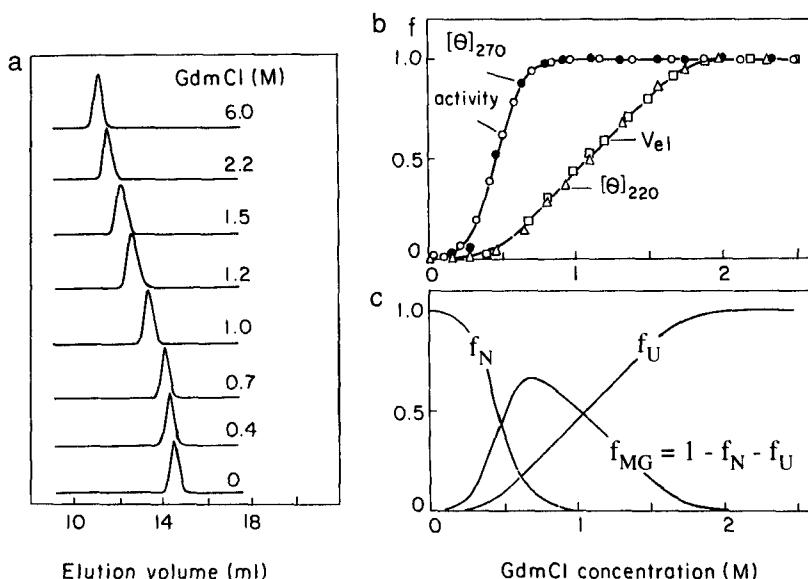


FIG. 44. (a) FPLC elution profiles of staphylococcal β -lactamase at 23°C and pH 6.8 for different concentrations of GdmCl. (b) Three-state unfolding of β -lactamase at 23°C: (○, ●) fractions of denatured molecules obtained from activity and $[\Theta]_{270}$, respectively; (□) fraction of unfolded molecules obtained from average elution volume V_{e1} ; (Δ) fraction of molecules without secondary structure obtained from $[\Theta]_{220}$. (c) GdmCl dependences of fractions of native (N), molten globule (MG), and unfolded (U) molecules calculated from the data from Fig. 44b. Adapted with permission from Uversky and Ptitsyn (1994). Copyright 1994 American Chemical Society.

increase in the GdmCl concentration. This decrease of elution volume is much larger than expected (and measured) for unfolded proteins (see, e.g., Fig. 43a and the insert in Fig. 43b) and clearly reflects a fast exchange between more compact and less compact states. Figure 44b shows that in this case two unfolding curves are observed: the first corresponding to protein denaturation (i.e., to the loss of its activity and its rigid tertiary structure), and the second corresponding to protein unfolding (i.e., the change of average elution volume) and the loss of its secondary structure (i.e., the change of far-UV ellipticity). The compact intermediate with a native-like content of secondary structure but without tertiary structure and activity meets all the usual requirements for the molten globule state and in fact is very similar to the classical molten globule observed in many proteins at low pH. Similar GdmCl- or urea-induced equilibrium intermediates have been observed for a number of proteins, but only the use of FPLC has permitted demonstration that these intermediates really are nearly as compact as the native state.

Figure 44c illustrates how the fraction of native molecules (monitored by the activity and near-UV CD) decreases with the increase in GdmCl concentration, being replaced by a compact nonnative intermediate having a native-like secondary structure (i.e., by the molten globule), which at higher GdmCl concentration is in its turn replaced by unfolded molecules.

Moreover, it was shown (Uversky and Ptitsyn, 1994) that at low temperature β -lactamase has three different curves of GdmCl-induced unfolding and therefore unfolds through at least two equilibrium intermediates. This is illustrated in Fig. 45. Figure 45a demonstrates the bimodal distribution of elution volumes for β -lactamase at 4°C (see the analogous curve for carbonic anhydrase in Fig. 24a). Figure 45b shows that this all-or-none transition between a compact and a less compact state occurs at higher GdmCl concentration than protein denaturation as monitored by its activity or near-UV ellipticity. Moreover, Fig. 45b also demonstrates that the elution volume of a less compact state continues to change at much higher concentration of GdmCl where the all-or-none transition

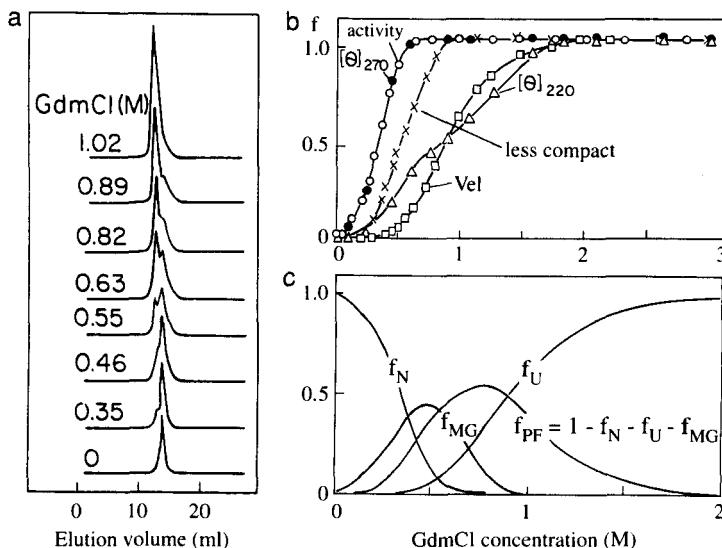


FIG. 45. (a) FPLC elution profiles of staphylococcal β -lactamase at 4°C and pH 6.8 for different concentrations of GdmCl. (b) Four-state unfolding of β -lactamase at 4°C: (x) fraction of less compact molecules obtained from the relative area of the two peaks in Fig. 45a; other notations are as described in the legend to Fig. 44. (c) GdmCl dependences of fractions of native (N), molten globule (MG), premolten globule (or partly folded, PF), and unfolded (U) molecules calculated from the data from Fig. 45b. Adapted with permission from Uversky and Ptitsyn (1994). Copyright 1994 American Chemical Society.

between compact and less compact states is already finished. This decrease of elution volume is much larger than normal swelling of unfolded molecules and strongly suggests the existence of a fast equilibrium between at least two different types of less compact molecules with very different degrees of compactness. The fourth curve in Fig. 45b shows the change in the far-UV ellipticity which looks like a superposition of two transitions—the increase of less compact molecules and their further GdmCl-induced expansion. Very similar results have also been obtained for bovine carbonic anhydrase B (Uversky and Ptitsyn, 1995).

The existence of three transitions indicates the presence of at least two intermediates between the native and the completely unfolded states that differ in their compactness. Similar intermediates have been observed for β -lactamase by Creighton and Pain (1980) by urea-gradient electrophoresis, but it remained uncertain whether both of these intermediates are at equilibrium or whether one of them corresponds to a transient kinetic state. The more compact intermediate revealed in our experiments is almost as compact as the native state and has a native-like content of secondary structure. Moreover, analogous studies of another protein, carbonic anhydrase (Uversky *et al.*, 1992; Uversky and Ptitsyn, 1995), have shown that this equilibrium intermediate strongly binds a fluorescent hydrophobic probe (ANS). This means that it fulfills all the criteria for the molten globule state. The less compact intermediate was observed for the first time in these experiments, and we named it the equilibrium partly folded or premolten globule state (Uversky and Ptitsyn, 1994, 1995). The evaluation of hydrodynamic volumes for both β -lactamase and carbonic anhydrase shows that it is no more than twice as large as the volume of the native state and no more than 1.5 times larger than that of the molten globule state (Uversky and Ptitsyn, 1994, 1995). By contrast, the hydrodynamic volumes of unfolded molecules are in both cases 11 times larger than that of the native state. Therefore, this new equilibrium state, being substantially more expanded than the native and the molten globule states, still belongs to the “family of compact states” of protein molecules. Figure 45b also shows that this intermediate has a substantial portion (about 50%) of the native far-UV ellipticity, i.e., a pronounced secondary structure. The experiments on ANS binding by carbonic anhydrase on its GdmCl-induced unfolding show that the curve of ANS binding has a nonsymmetric maximum. This maximum can be presented as a superposition of two maxima, corresponding to ANS binding by the premolten globule and the molten globule states, respectively (Uversky and Ptitsyn, 1995). This suggests that the premolten globule state can also bind ANS, although it does so more weakly than for the molten globule state.

Figure 45c compares the fraction of native molecules (f_N) of β -lactamase (with activity and rigid tertiary structure), the fraction of molten globules (i.e., compact but nonnative molecules) (f_{MG}), the fraction of fully unfolded chains (with a large hydrodynamic volume and without secondary structure) (f_U), and the fraction of molecules in the premolten globule or partly folded state (f_{PF}) which can be calculated as $f_{PF} = 1 - f_N - f_{MG} - f_U$. One can see how the native molecules are replaced first by the molten globule intermediates (with the maximum at 0.45 M GdmCl equal to 45% of all molecules), then by partly folded or premolten globule molecules (with the maximum at 0.80 M GdmCl equal to 55% of all molecules), and finally at GdmCl concentrations larger than 2M by unfolded molecules.

The properties of the new premolten globule state (partial compactness, partial secondary structure, relatively weak ANS binding) look similar to the properties of the first (burst) kinetic intermediate in protein folding (see Sections IV,A,1, IV,A,2, and V,A,1). In addition, the properties of the equilibrium molten globule state were previously recognized as being very similar to those of the second (early) kinetic intermediate ("kinetic molten globule state"). Therefore, it appears that we can mimic the kinetic process of the protein folding (at the fast dilution of its solution in high urea or GdmCl concentration) by the equilibrium process of its folding (at step-by-step dilution of these solutions). If so, it opens new possibilities for learning more about the properties of the first kinetic intermediate by studying its equilibrium analog.

There is an even deeper analogy between kinetic and equilibrium intermediates in protein folding. The two kinetic intermediates (the premolten globule and the molten globule) accumulate upon protein folding, i.e., are separated from each other and from the native state by high potential barriers. On the other hand, we have shown that the equilibrium molten globule state is separated from both the unfolded and the native states by all-or-none transitions, i.e., it is also separated from both of them by high potential barriers (see Sections III,A,3 and III,B,1). Moreover, in the case in which the premolten globule state can be observed, the all-or-none transition seems to occur between the partly folded and the molten globule states (Uversky and Ptitsyn, 1995) (see Fig. 45). This correlates with the accumulation of the premolten globule state before its conversion into the molten globule state in the kinetics of protein folding. It seems likely that the premolten globule and the molten globule states are two general intermediates in protein folding which can be observed in both kinetic and equilibrium experiments.

3. 3D Structures of Molten Globule Intermediates and the Transient State

A series of very interesting papers by Fersht and his collaborators using site-detected mutagenesis (see Fersht, 1993, for a review) provided deep insight into the structure of the molten globule kinetic intermediate. Moreover, this approach also permits the description of the structure of a transient state which corresponds to the potential barrier between the molten globule kinetic intermediate and the native state.

Fersht extended to protein folding (Matouschek *et al.*, 1989, 1990; Fersht *et al.*, 1992) the same approach which he had previously elaborated to study the role of interaction energies between substrates and enzymes during enzymatic catalysis (Wells and Fersht, 1986; Fersht, 1987). The idea is simple and elegant. To check whether the given interaction (known from the X-ray protein 3D structure) is already present at a given stage of protein folding, we have to remove this interaction with site-specific mutagenesis to see how its removal will influence free energies of different kinetic states in protein folding. These free energies can be evaluated by the usual kinetic and equilibrium measurements, e.g., by fluorescence in combination with stopped-flow technique. To make the change of interactions as addressable as possible it is worthwhile to replace just a part of a side chain. For example, removing the OH group from serine ($\text{Ser} \rightarrow \text{Ala}$ mutation) permits study of the influence of its hydrogen bond(s) on free energies of different kinetic states and thus to reveal the stage of protein folding at which this particular hydrogen bond is formed. After many such mutations have been created and studied, one can obtain the 3D structure of a kinetic state almost at atomic resolution.

In principle there are no limitations for using this method at each stage of protein folding, although its application to earlier stages may face serious technical problems. To date this approach has been applied to the most interesting structures: the structure of a kinetic intermediate, which accumulates just before a protein transforms into its native state, and the structure of a transient state, which corresponds to the potential barrier between this intermediate and the native state.

This approach is illustrated in Fig. 46, which presents the free energy profile of the unfolded (U), intermediate (I), transient (#), and folded (F) states of a wild-type protein and its mutant. The free energy differences between mutant and wild-type proteins in intermediate, transient, and folded states will indicate at which of these states the specific part of the protein structure is formed.

An important extension of this method was the exploration of double mutants that permits isolating interactions between a given pair of side

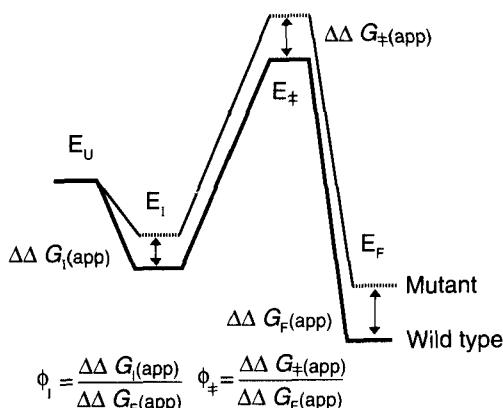


FIG. 46. Free energy profile for the folding of wild-type and mutant proteins according to the scheme $U \rightleftharpoons F$ via transition state. The free energies of unfolded states of both proteins are shown to be the same for illustrative purposes. Reproduced with permission from Fersht (1993).

chains from all other interactions. If two residues, X and Y, are mutated separately and then together, the free energy of $X \cdots Y$ interactions can be obtained as a difference between free energies $\Delta G_{X \rightarrow WT}$ and $\Delta G_{XY \rightarrow Y}$, where WT, X, Y, and XY stand for the wild type of protein and for a protein with mutated X, Y, and both X and Y residues, respectively (Serrano *et al.*, 1990, 1991; Horovitz *et al.*, 1990). In an analogous manner even triple mutants can be analyzed (Fersht, 1993). The relative influence of the given mutations on intermediate, transient, and folded states can be characterized by the parameters Φ introduced in Fig. 46.

Of course, this method is based on some assumptions discussed in detail by Fersht *et al.* (1992). The main assumption is that mutations do not alter the folding pathway, i.e., that the wild-type protein and the mutant fold through the same intermediate and transient states. However, the self-consistency of the results obtained with many mutations make all the principal results very convincing.

The position of this intermediate state on the folding pathway (just before its rate-limiting step; see Fig. 46), as well as its properties revealed in this study leave little doubt that it is a kinetic molten globule intermediate (Matouschek *et al.*, 1990). Therefore, this approach can be considered as complementary to and probably even more powerful than the pulsed hydrogen exchange technique for the study of the kinetic molten globule state.

This approach has been applied to the very detailed study of the 3D structure of barnase (bacterial ribonuclease), for which crystal and

solution 3D structure have been determined by high-resolution X-ray analysis (Mauguen *et al.*, 1982; Baudet and Janin, 1991) and 2D NMR (Bycroft *et al.*, 1991), respectively. This structure (Fig. 47) includes three α helices in the N-half of the chain and five antiparallel β strands in its C-half. The major hydrophobic core is formed by the packing of the main α_1 helix against the β_2 and β_3 strands. The results have been published in a series of papers in one issue of *Journal of Molecular Biology* (Serrano *et al.*, 1992a,b,c; Matouschek *et al.*, 1992a,b) and have been extensively reviewed by Fersht (1993). A short summary of these results follows.

a. α_1 Helix. The main (first) α -helical region in barnase (residues 6–18) is schematically shown in the top part of Fig. 48. Thr-6 at the very N-terminal of this helix stabilizes it by a hydrogen bond of its OH group with the backbone NH group of Gly-9 and by weaker hydrogen bonds of this group with residues 7 and 8. In addition, the side chain of Thr-6 packs against the region before α_1 helix and against one of the loops. Mutations Thr-6 \rightarrow Ala, Gly do not significantly change the free energies of intermediate and transient states (see the bottom part of Fig. 48).

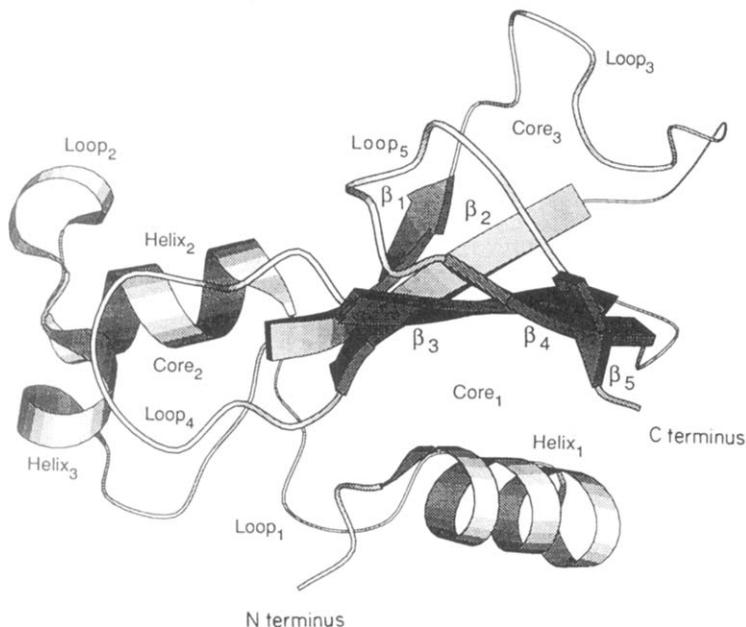


FIG. 47. 3D structure of barnase. Reproduced with permission from Serrano *et al.* (1992a).

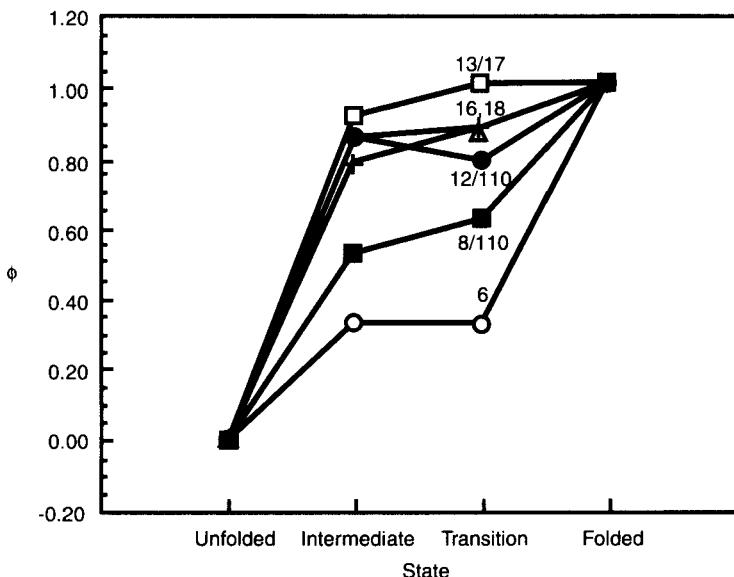
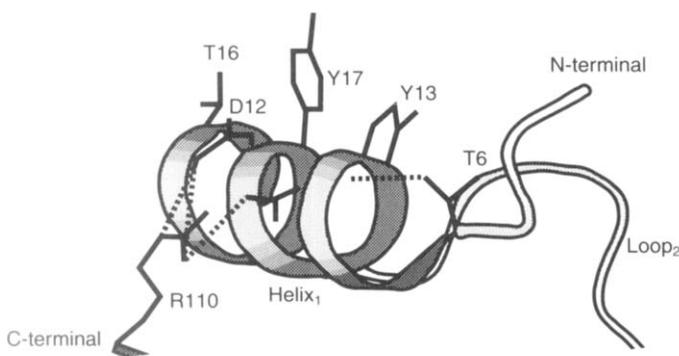


FIG. 48. Structure of the α helix in barnase (top) and the influence of its mutations on the free energies of intermediate, transient, and folded states (bottom). Parameter ϕ has been introduced in Fig. 46. Mutations Thr-16 \rightarrow Ser (+) and His-18 \rightarrow Gln (Δ), as well as double mutations of Tyr-13/Tyr-17 (\square) and Asp-12/Arg-110 (\bullet) lead to significant changes in the free energy of the intermediate and transient states, while double mutations of Asp-8/Arg-110 (\blacksquare) and mutation Thr-6 \rightarrow Ala (\circ) have smaller influence on the free energies of these states. Reproduced with permission from Fersht (1993).

This means that these interactions are formed only in the last step of protein folding. On the other hand, the mutants $\text{Thr-16} \rightarrow \text{Ser}$ and $\text{His-18} \rightarrow \text{Glu}$ change the free energies of intermediate and transient states approximately to the same extent as the folded state. This means that the interactions which involve the CH_3 group in the Thr-16 side chain and the ring of His-18 come to the scene as early as those in the intermediate state.

This is strong evidence that the C terminus of α_1 helix is already restored at the stage of the formation of the intermediate, while its N terminus becomes fixed only in the completely folded state. This conclusion is confirmed by a number of double mutants which show that interactions of Tyr-13 or Thr-16 with Tyr-17, as well as salt bridge between Asp-12 and the C-terminal Arg-110 are already present in the intermediate state, while the salt bridge between Asp-8 with the same Arg is fully present only in the folded state. This permits localizing the C-terminal part of the α_1 helix, residues 13–18, which is already formed in the intermediate state, and to conclude that the rest (N-terminal part) of this helix is formed only in the completely folded state.

The β sheet consisting of five antiparallel β strands is shown in Fig. 47. Different mutations in β_1 (residues 50–55), β_2 (70–76), β_3 (85–99), β_4 (94–99) and β_5 strands (106–108) present a nice picture of propagation of the structure from the central β_3 strand first to "half-edge" strands β_2 and β_4 , and then to edge strands β_1 and β_5 (Fig. 49). In fact, replacements in the central strand β_3 , decreasing its interactions with strands β_2 and β_4 , increase the free energy of the transient state as much as that of the folded state ($\Phi = 1$) and also significantly influence free energy of the intermediate ($\Phi \approx 0.6$). The interactions between β strands nearer to the β sheet's edges (β_1 with β_2 and β_5 with β_4) have smaller influence on free energies of both transient and intermediate states. It appears that the β sheet grows from the center to the edges, when the protein proceeds from the kinetic molten globule intermediate to the folded state.

A progressive growth of structure from the center to edges is also seen in the main nonpolar core of barnase. Figure 50 shows that the mutation of Ile-88 in the center of a main nonpolar core significantly ($\Phi \approx 0.6$) alters the free energy of the intermediate state and changes the free energy of the transient state almost as much as that of the folded state. Mutations of Ile-109 and Ile-96 also significantly alter the free energies of the intermediate and transient states ($\Phi \approx 0.6$), while mutations of Ile-76, which is on the edge of a nonpolar core, influence only the free energy of the folded state.

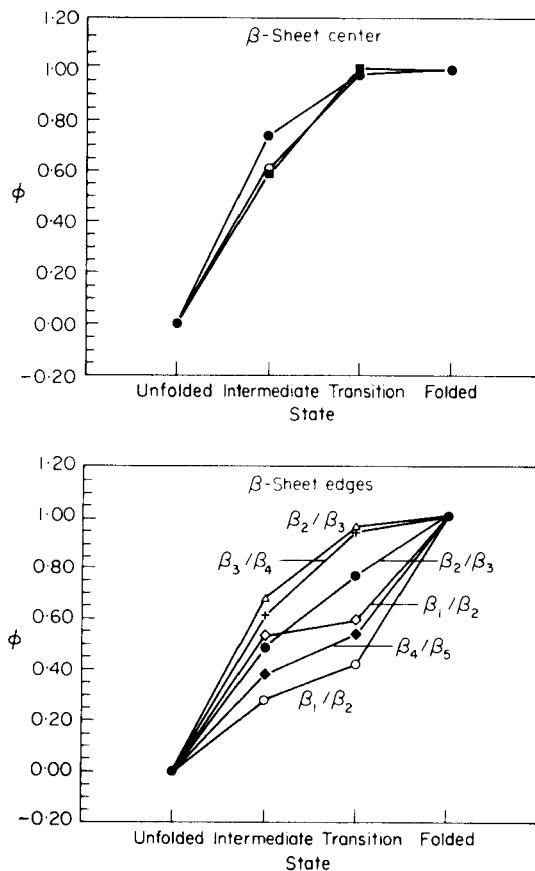


FIG. 49. Influence of mutations in the β sheet of barnase on the free energies of its intermediate, transient, and folded states. *Top:* Mutations Val-88 → Ala (○), Val-89 → Thr (■), and Val-96 → Ala (●) showing that the center of the β sheet is already substantially formed in the intermediate state and completely formed in the transient state. *Bottom:* Mutations Ile-55 → Thr (○), Ile-55 → Ala (◇), Val-76 → Ala (●), Thr-105 → Val (◆), Ser-91 → Ala (+), and Ser-92 → Ala (Δ) showing the progressive growth of β structure from the center of the β sheet to its edges in intermediate and transient states. Reproduced with permission from Matouschek *et al.* (1992a).

Of special importance are the data showing the existence of a number of long-range interactions in the intermediate and transient states. For example, interactions of Val-10 and Leu-14 (α_1 helix) with β sheet and interaction of Ile-88 (β_3 strand), Ile-96 (β_4 strand), and Ile-109 (C-terminal) with α_1 helix are partly realized in the intermediate and transient states. This demonstrates that the position of α_1 helix relative to

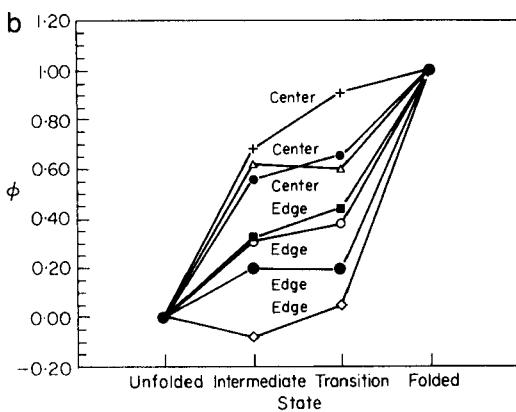
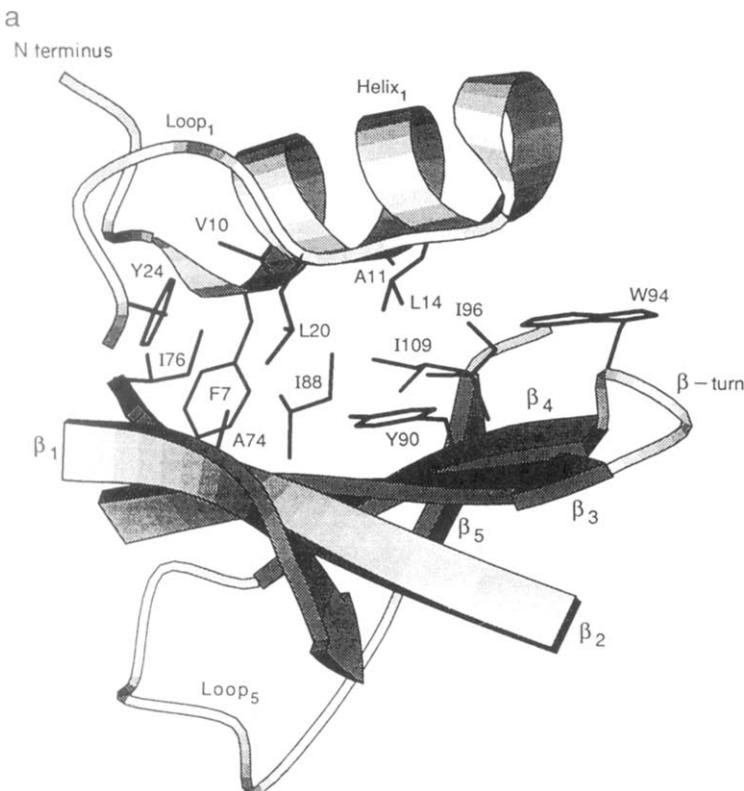


FIG. 50. Structure of the nonpolar core of barnase (a) and the influence of its mutations on free energies of the intermediate, transient, and folded states (b). Mutations Ile-88 → Val (+), Ile-109 → Val (●), Ile-96 → Val (△), Val-10 → Ala (○), Val-10 → Thr (■), Leu-14 → Ala (●), and Ile-76 → Val (◇) show the progressive growth of a nonpolar core from its center to its edges in intermediate and transient states. Residues 10 and 14 belong to the α helix, while other mutated residues are in the β sheet. Reproduced with permission: (a) from Serrano *et al.* (1992a) and (b) from Matouschek *et al.* (1992a).

β sheet (as well as the mutual positions of different β strands inside this sheet: see above) is already present to some extent in the intermediate state. This strongly confirms the idea (Ptitsyn, 1973, 1991) that the kinetic molten globule intermediate has a native-like folding pattern. Some evidence in favor of this idea has also been obtained by pulsed hydrogen exchange of a number of proteins including barnase (see above). However, unlike site-directed mutagenesis, pulsed hydrogen exchange sees mainly hydrogen bonds inside secondary structure and gives no direct evidence on the mutual positions of the regions of secondary structure which are not connected by hydrogen bonds.

In barnase, the pulsed hydrogen exchange approach (Bycroft *et al.*, 1991; Matouschek *et al.*, 1992b) has confirmed the most important features obtained by site-directed mutagenesis (like the early formation of the C-terminal part of α_1 helix and of β sheet). In addition, the pulse-labeling technique reveals some other structural features of the kinetic intermediate, including early formation of the C-terminal part of α_2 helix, of α_3 helix, and of several β turns (see Fersht, 1993).

The study of kinetic intermediates by site-directed mutagenesis is in its infancy. This approach can be applied to the earliest stages of protein folding and can be used to answer questions about the existence of multiple folding pathways (Fersht, 1993). It can analyze 3D structures of different kinetic intermediates even for very short-lived states if these states can be observed in the kinetics of protein folding.

B. How Do Proteins Fold?

1. Paradox of Protein Folding

The framework model of protein folding and its experimental confirmation give a conceptional framework for some general aspects of protein folding. However, it does not solve the mystery of protein folding and does not pretend to do so. The framework model does not answer the main question: how can a protein find its unique 3D structure among the large number of alternatives? What this model has really done is that it has localized this mystery showing the stage of protein folding which is faced with this difficulty.

The paradox of protein folding was first formulated by Levinthal (1968, 1969) in the following way. A protein chain consisting of n residues has a total number of conformations equal to $\Omega = \gamma^n$, where γ is the number of conformations per residue. As each residue (including its side chain) has about 10 different conformations, the total number of protein conformations is astronomically large, and a protein cannot search all of them to achieve its global energy minimum.

It was the first but probably not the best formulation of the main difficulty of protein folding. In fact, there are many examples in which the finding of a global minimum among astronomically large numbers of conformations is possible in a reasonable time. Examples include the formation of the double helix in DNA, of helical or β structures in polypeptides, and other cases, in which a search of conformations can be performed more or less independently in the different regions of a macromolecule. It is important to emphasize that heteropolymers (like DNA or heteropolypeptides) can achieve their equilibrium state as easily and as fast as homopolymers.

The real paradox has been formulated only relatively recently as a result of studies of this problem using the rigorous terms of theoretical physics (Shakhnovich and Gutin, 1989, 1990, 1993a,b; Bryngelson and Wolynes, 1989).

Shakhnovich and Gutin (1990) showed that a heteropolymer with random sequence (i.e., without any preferences in the distribution of the various monomers along a chain) dramatically decreases the number of its conformations from an “exponentially large” number (i.e., exponentially increasing with increase in the number of monomers) to just a few conformations. This dramatic change occurs at the temperature

$$T_c = B(\rho)^{1/2} / 2k(\ln \gamma)^{1/2}, \quad (10)$$

where B is the mean square deviation of pairwise monomer interaction energies from their average value (a measure of chain heterogeneity), ρ is the density of monomers (a measure of chain compactness), γ is the partition function per monomer (a measure of chain rigidity), and k is the Boltzmann constant. Moreover, at temperatures $T < T_c$ a significant fraction of heteropolymers will have only one conformation. Thus, it appears that the existence of a unique 3D structure does not create any problem for a protein chain from a thermodynamic point of view, as a reasonably large fraction of random sequences can have this structure under normal conditions. This result confirms the suggestion that a protein is no more than an “edited” random copolymer (Ptitsyn, 1983, 1984, 1985; Ptitsyn and Volkenstein, 1986) and seems to be the solution of the thermodynamic problem of why a protein can have a unique 3D structure.

The real difficulty is that at low temperatures ($T < T_c$) folding rates dramatically slow down and the global energy minimum of a heteropolymer becomes inaccessible for kinetic reasons (Bryngelson and Wolynes, 1989; Sali *et al.*, 1994). In addition to the global minimum a chain also has a number of local minima, separated from each other by

high potential barriers. A folding chain can easily reach one of these "wrong" local minima and be trapped in it, which may increase the time to achieve its global minimum to values which make this practically impossible.

The framework model of protein folding has introduced three main stages of this process—the formation of the premolten globule state, of the molten globule state, and of the native state—and indicates at which of these stages a protein meets with this difficulty. The first stage does not create any serious problems, since the structure of the premolten globule state is highly fluctuating and can be formed or destroyed almost independently in different regions of a chain. These processes are not connected with large potential barriers and therefore are fast and efficient. The equilibrium state over corresponding degrees of freedom can be reached on a millisecond time scale.

The real problems start with the second stage, the formation of a kinetic molten globule intermediate. Each protein chain (even if its secondary structure is specified) can possess a number of folding patterns, differing by the mutual position of α and/or β regions. Each of these folding patterns corresponds to a local minimum of chain energy. These local minima are separated from each other and from the global minimum by high potential barriers, which can practically prevent a system from reaching its equilibrium.

2. Biological Evolution and Protein Folding

If proteins cannot easily escape the traps, i.e., local energy minima with nonnative folding patterns, the only logical way out is to avoid them. Therefore, a mechanism must exist which leads protein molecules directly to their global energy minimum without searching local ones. This idea was strongly supported by computer simulation of a folding of small model heteropolymers (Shakhnovich *et al.*, 1991; Sali *et al.*, 1994); Monte Carlo simulations were performed for the folding of a short (27-mer) random heteropolymer chain on a cubic lattice. In this simple case all compact conformations without self-overlapping of a chain can be exactly enumerated, which permits locating the conformation with the minimal energy for each sequence. Thus, in this case we know the native conformation for each sequence and the problem is whether or not a chain can reach this conformation within a reasonable time.

Monte Carlo simulation of the folding of these simple model chains leads to a very interesting result (Shakhnovich *et al.*, 1991; Sali *et al.*, 1994): only a small part of the random sequences can fold in a reasonable time. The sequences which can do so have one remarkable difference

from other sequences: the “gap” between the state of the lowest energy and the next state is on the average substantially larger than that in other sequences. Thus, an interesting empirical rule has been established for these model systems—the large gap between the global and the next higher energy minima is necessary for fast folding.

The most important idea was that this large gap between the native and alternative conformations could have been created in natural proteins by biological evolution (Shakhnovich and Gutin, 1993a,b). This implies that biological evolution has a chance to ensure both fast folding and the relatively large stability of a native protein by adjusting a single parameter—the gap between the energy of the ground state of a heteropolymer and the energy of its next state. This is illustrated in Fig. 51, which compares a typical energy spectrum of a random heteropolymer with the energy spectrum of a sequence selected to increase this gap. The possibility for a substantial decrease in the energy of one protein conformation without altering the energies of the others needs some discussion. It is due to the property of heteropolymers which has been postulated by Bryngelson and Wolynes (1987) and strictly shown by Shakhnovich and Gutin (1990), i.e., that heteropolymers meet all requirements of the so-called random energy model introduced by Derrida (1981) in the theory of spin glasses. The important property of this model is that its lowest energy levels correspond to structurally quite different conformations. This property is mostly nontrivial since to date a heteropolymer is the only physical system which is shown to have it. For example, simple low-molecular weight crystals have quite opposite properties—there are N states with one defect, $N(N-1)/2$ states with two defects, etc., for a crystal of N molecules, and both energies and the physical structures of all these states are very similar. Chemical links between monomers in a polymer chain make almost all its conformations self-overlapping; a relatively small part of the remaining conformations happen to be quite different from each other from the structural point of view. This is the reason why mutations can strongly stabilize one of these energy minima, but almost always do not influence the others. We can, for example, stabilize the ground state of a random heteropolymer by engineering a salt bridge between monomers A and B which is close in its 3D structure. Due to the main theorem of the random energy model, alternative structures must be quite different from the ground state and therefore sites A and B in general are far from each other in these structures. Therefore, these mutations will not affect their interactions and will not stabilize alternative structures.

Shakhnovich and Gutin have shown that the evolutionary “design” of a sequence leading to the large decrease in the energy of its given 3D

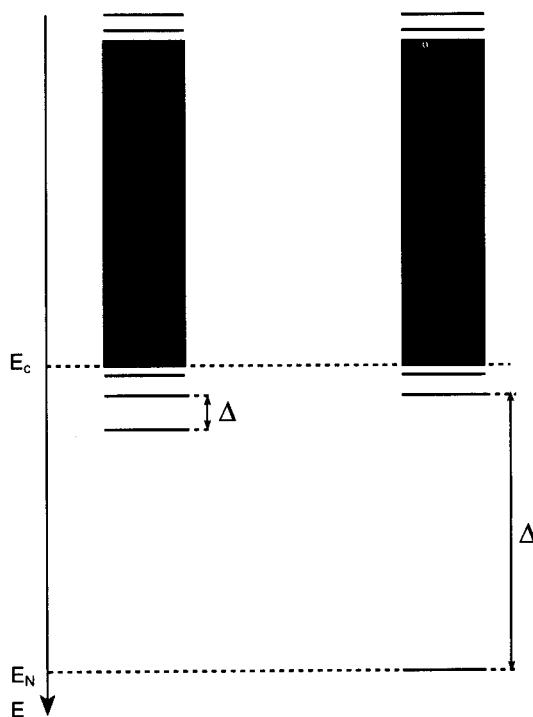


FIG. 51. Scheme of energy levels for random (left) and evolutionarily selected (right) protein sequences. Black regions correspond to a continuous spectrum of energy. Discrete levels represent the "native" (ground) state and some of the "wrong folded" (alternative) states. They correspond to quite different 3D structures of protein chains (though each of them can be considered as a number of sublevels reflecting small variations in structure). In the evolutionary selected sequences (right) the energy of one of the lower minima becomes much lower than that of all others. E_c , lower boundary of continuous energy spectrum; E_N , energy of the ground (native) state; Δ , energy gap between the ground and the next state.

structure is not such a difficult task. They have elaborated a general algorithm for the design of sequences which will "fit" any given 3D structure (Shakhnovich and Gutin, 1993a,b). However, the most important and quite unexpected result was that the heteropolymers designed to have a stable 3D structure are folded much faster than random sequences (Shakhnovich and Gutin, 1993b). This is illustrated in Fig. 52 which compares the efficiency of folding for selected and random sequences for short (27-mers) heteropolymers on cubic lattice. The efficiency of folding is measured by the "foldicity" of a chain, i.e., by a fraction of Monte Carlo runs which leads to achieving the ground state for a given sequence.

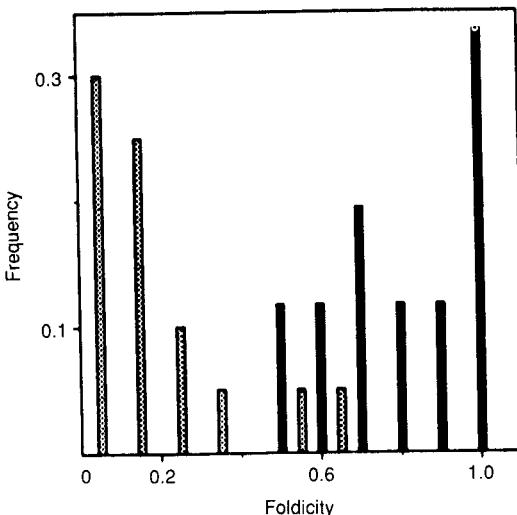


FIG. 52. Distribution of "foldicities" for 17 selected sequences (filled bar) and 20 random sequences (dotted bar) with nondegenerate ground states. For each sequence 10 runs of 20,000,000 Monte Carlo steps were made. The "foldicity" of a sequence is specified as the fraction of runs for which a unique ground state was found. Reproduced with permission from Shakhnovich and Gutin (1993b).

The next important step was the transition from model short chains to more realistic long chains. In fact, if we can design a sequence which makes the given 3D structure much more stable than the others, we will know the 3D structure which corresponds to the ground state without enumerating all compact states and, therefore, can consider chains of realistic lengths. The computer experiments performed by Shakhnovich (1994) for "evolutionarily designed" chains up to 100 monomers have shown that these more realistic chains can also fold quickly into their native (i.e., ground) states. Thus, it was established that the existence of a large energy gap between the lowest minimum and the others is not only necessary but sufficient for fast folding.

3. Possible Solution of Paradox

The idea of the evolutionarily designed gap between the native and all other states of protein molecules may solve the main paradox of protein folding. This is illustrated in Fig. 51 which shows that all states of protein molecules can be divided into three types: (1) a very large number of substantially unfolded states with high energy (continuous part of the spectrum), (2) a few wrong folded states with a discrete set of energies, and (3) a correctly folded state with deep global minimum

of energy. Shakhnovich and Gutin (1993b) have shown that at all temperatures only the states of the first and the third types contribute to the partition function of molecules. The states of the first type contribute because they are very numerous and, therefore, their total entropy is very high, while the state of the third type does so because its energy is much lower than those of all other states. The states of the second type have relatively high energies and are not numerous enough to have high entropy and, therefore, do not substantially contribute to a partition function at any temperature.

It follows that evolutionary selected sequences avoid wrong folded alternative structures and jump directly from the unfolded state to the global minimum. This transition takes place at temperatures which are higher than T_c and, therefore, are high enough to overcome kinetic difficulties. It is a real first-order phase transition, as it is accompanied by a sharp change of energy from that of the unfolded state to that of the native state, and a change of entropy from that of the unfolded state to zero. This transition has nothing to do with the transition in random heteropolymers. A random heterochain moves from the lower boundary of its continuous energy spectrum, E_c , to one of the traps, i.e., local minima which correspond to alternative (nonnative) 3D structures. In contrast, the evolutionary selected sequence jumps directly from the middle of the continuous energy spectrum (i.e., from an unfolded state) to the global minimum of energy, ignoring all alternatives.

It is likely that this idea, presented by Shakhnovich and Gutin (1993b), solves the main paradox of protein folding, as it permits a protein to avoid the choice between a thermodynamically impossible transition to the native state at high temperatures and a kinetically impossible transition to this state at low temperatures. This choice is inevitable for all statistical systems but is avoided in systems which have been specially selected by nature to avoid it.

4. Nucleation of Protein Folding

If we accept that wrong folded protein structures are thermodynamically unfavorable, the question arises as to how a protein can avoid them at a kinetic pathway of its folding. The answer is probably related to the two-state (phase) character of protein folding. It is known that first-order phase transitions involve a slow nucleation stage with a fast subsequent growth of a nucleus. A similar nucleation-growth mechanism has been proposed for protein folding based on studies of folding-unfolding kinetics (Tsong *et al.*, 1972) and on the analysis of native protein structures (Wetlaufer, 1973). This idea has been strongly supported by Monte Carlo calculations of a kinetics of protein folding for model lattice

proteins (Abkevich *et al.*, 1994). These model proteins do not contain side chains and, therefore, folding is stopped at the molten globule state.

According to these calculations (performed for sequences designed to fit the given tertiary fold; see above) folding of small lattice chains consists of two stages. The first stage is a slow formation of a spatially localized substructure involving a set of native contacts ("folding nucleus"), while the second stage is the fast growth of this nucleus to the native structure embracing the whole molecule.

The important result of these calculations is that fluctuations of the native structure of model proteins can be divided into "locally unfolded" (with preserved nucleus) and "globally unfolded" (in which the nucleus is destroyed). This is consistent with experimental data on the heterogeneity of protection factors in the molten globule state (see Section II,B,1 and II,B,2) and on the similarity of protection patterns in the equilibrium and the kinetic molten globules (see Section IV,B,2). We cannot exclude the possibility that the nucleus is protected before the rest of a chain in folding kinetics and is protected more than the rest of a chain in the equilibrium molten globule state. The examples of these hypothetical nuclei may be the complex of N- and C-terminal helices in cytochrome c, the α -helical subdomain in α -lactalbumin or lysozyme, and the complex of helices A, B, and H in apomyoglobin (see Sections II,B and IV,B).

The "nucleation barrier" of protein folding is intrinsic for the folding process as each protein first has to form its nucleus. Usually, this barrier is much smaller than the barrier between the molten globule and the native states, but the second barrier can be dramatically decreased (see Section IV,C) and in these cases the nucleation barrier represents a rate-limiting step of protein folding.

5. Transition to Native Structure

There is one additional problem connected with the formation of the native protein from the molten globule. Van der Waals attraction, tertiary hydrogen bonds, and other more specific interactions which are absent or substantially reduced in the molten globule but present in the native protein can be larger than interactions which stabilize the molten globule state (see Section III,B,2). It may happen that the folding pattern with minimal energy will lead to the native structure with energy larger than that achievable from one of the alternatives. In this case the final folding of the protein again would require a search of all possible folding patterns which would make this process hopeless. Therefore, the coincidence of the folding pattern of the kinetic intermediate with that of the final

stage is a necessary condition for successful folding (Ptitsyn, 1973, 1987).

In trying to find a solution of this second paradox of protein folding, it was suggested that all folding patterns decrease their energies in a similar way when they jump into the tightly packed native structure (Ptitsyn, 1991, 1994). This assumption appears reasonable as at least van der Waals energy of packing may be similar in all tightly packed structures (although it is almost certainly not the case for more specific interactions, like tertiary hydrogen bonds or salt bridges). However, a more simple and more general explanation can be suggested. Biological evolution designs sequences for which the native folding pattern corresponds to a very deep global energy minimum (see Fig. 51). This folding pattern will automatically survive in the native state if additional interactions in rigid tertiary structures cause changes smaller than the gap between the native and the next folding pattern. In fact, if energy differences between tertiary structure and folding patterns are substantially smaller than the gap between right and wrong folding patterns, they can alter the mutual ranks of the nonnative folding pattern but cannot deprive the native-like folding pattern of its first rank.

Additional design of protein sequences may be necessary to ensure that the native-like folding pattern can achieve the global minimum of tertiary structure with respect to the corresponding degrees of freedom (rotation of side chains, etc.). Here we can also think about the continuous energy spectrum of the molten globule state, another “glass transition” to a number of minima, and solve this paradox by a specific decrease in the global minimum energy in the course of biological evolution.

The recent studies of the protein folding problem by the rigorous methods of theoretical physics are very exciting. They provide good examples of how physics can do much more for biology than just provide a powerful arsenal of structural and other methods and sophisticated computational approaches. More importantly, it can introduce to biology the strict physical mode of thinking, leave less room for plausible speculations, and replace speculations with a more careful quantitative analysis.

VI. PHYSIOLOGICAL ROLE OF MOLTEN GLOBULE

A. *Prediction*

The properties of the molten globule state described above suggest the following model of this state (see Fig. 6). This model is a “semirigid”

molecule which preserves the main features of the native folding pattern, but has much easier internal motions including rotational isomerization of many internal side chains and flexibility of the loops. Therefore, it is a state with a rather mobile shell but with a less flexible core. It can expose a number of its nonpolar groups, thus increasing its hydrophobic surface, but it can also screen its nonpolar groups even more than the native state. It allows much more pronounced small-scale fluctuations than the native state, but its unfolding is a phase transition of the first order, which protects it from occasional loss of its folding pattern by large-scale thermal fluctuations.

These properties appear almost ideal for a protein in a living cell which must adjust itself to a large set of different conditions—in the cytoplasm, near or inside membranes, in cell organelles, etc. Some of these conditions are denaturing ones, e.g., inside lysosomes where the pH can be as low as pH 4.5–5.0, or near membranes in which the immediate pH can be substantially lower than the overall pH.

These concepts have permitted us to suggest that the molten globule state may play an important role in a number of physiological processes (Bychkova *et al.*, 1988). Although this article mainly focuses on the possible role of molten globule in protein translocation across membranes, we have also assumed that proteins may be in the molten globule state in their nascent form and that this state can be trapped by chaperones. We have assumed that protein degradation at acid pH or in the ubiquitin system may also demand the presence of this state.

Many important aspects of this hypothesis have been confirmed since 1988 by direct experiments, and in 1993 we reviewed the first evidence for the existence and role of the molten globule state *in vivo* (Bychkova and Ptitsyn, 1993a). An excellent review on protein folding *in vivo*, including many related problems, has been presented by Freedman (1992). Other reviews (Fischer and Schmid, 1990; Jaenicke, 1991; Nilsson and Anderson, 1991; Gething and Sambrook, 1992) can also be recommended. Here, I shall repeat the most important data obtained up to 1992 and shall concentrate mainly on the most recent results.

There is another kinetic and equilibrium intermediate (premolten globule state) which is more expanded than the molten globule state (see Sections IV,A and V). It is quite possible that this state can also play a role in some physiological processes. For example, there is some evidence that chaperone GroEL (see below) recognizes α -lactalbumin in a state which is slightly more expanded than the molten globule state (Okazaki *et al.*, 1994; Robinson *et al.*, 1994).

B. Folding and Assembly

1. Folding and Chaperones

The biosynthesis of a small protein or protein domain needs about 10 sec (Alberts *et al.*, 1983), which is substantially longer than the formation time of the molten globule state (\sim 1 sec), but much shorter than the formation time of the native state in many proteins (up to thousands of seconds). Therefore, it was natural to assume that many nascent protein chains can be in the molten globule state just after their biosynthesis (Bychkova *et al.*, 1988; Bychkova and Ptitsyn, 1993a). A cell might need to maintain this state in order to use it for self-assembly, transmembrane transport, and other processes which need protein molecules in a semiflexible, rather than in a rigid, state.

It is very likely that the molten globule state of nascent protein chains is preserved by chaperones which bind newly synthesized protein chains but cannot bind completely folded proteins. It was shown that GroEL, a chaperone from *E. coli*, facilitates folding of many proteins; it recognizes and stabilizes kinetic intermediates of protein folding and prevents them from aggregating (Bochkareva *et al.*, 1988; Buchner *et al.*, 1991; Martin *et al.*, 1991; van der Vies *et al.*, 1992). The important point is that chaperones facilitate protein folding by inhibition rather than by acceleration (Weiss *et al.*, 1988; Laminet *et al.*, 1990; Martin *et al.*, 1991; van der Vies *et al.*, 1992, etc.). Proteins begin to fold rapidly after addition of Mg^{2+} ATP to their complexes with chaperones (Laminet *et al.*, 1990; Martin *et al.*, 1991), and the rate of their folding in the presence of GroEL and Mg^{2+} ATP is not much different from that in the absence of both factors (Laminet *et al.*, 1990). This suggests that GroEL traps proteins in a state before the rate-limiting step (Laminet *et al.*, 1990), which is consistent with our prediction that chaperones recognize proteins in the molten globule state.

Direct structural evidence confirming that proteins are in their molten globule state in complexes with chaperones has been obtained by Hartl and colleagues (Martin *et al.*, 1991, 1992) and Lorimer, Jaenicke, and co-workers (van der Vies *et al.*, 1992). Martin *et al.* (1991, 1992) have shown that dihydrofolate reductase and rhodanese in complexes with GroEL have a maxima for tryptophan fluorescence (see Fig. 53a) and its acrylamide quenching intermediate between those of the native and unfolded states; proteins in these complexes strongly bind ANS (see Fig. 53b), show an efficient energy transfer from tryptophan residues to ANS, and are much more sensitive to proteolysis than the native proteins. In other words, these proteins in their complex with GroEL

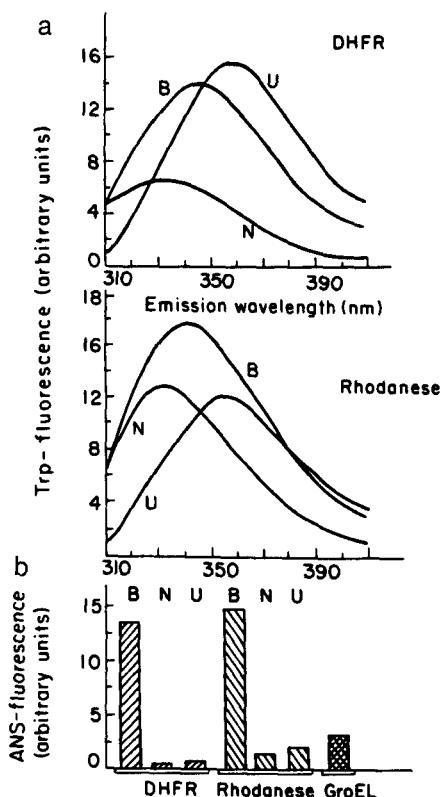


FIG. 53. (a) Fluorescence spectra of tryptophans in dihydrofolate reductase (DHFR) and rhodanase in the native (N) and unfolded (U) states as well as in the state bound (B) to GroEL. (b) Binding of ANS to DHFR and rhodanase in the N, U, and B states (see above). The cross-hatched rectangle presents ANS binding by the native GroEL in the absence of bound proteins. Adapted with permission from Martin *et al.* (1991).

fulfill all the usual requirements of the molten globule state. In a similar way, van der Vies *et al.* (1992) have shown that ribulose-bisphosphate carboxylase (Rubisco) forms a stable binary complex with GroEL in a molten globule-like state which accumulates upon protein folding from GdmCl- and acid-denatured states.

Very interesting data showing that chaperones can recognize the non-native states of many different proteins have been obtained by Lorimer's group (Viitanen *et al.*, 1992). They have shown that ~50% of the total protein fraction of *E. coli* forms complexes with GroEL (which is an oligomer of 14 identical subunits). However, they do so only in a partly unfolded state, but not in the native state. Figure 54 shows that a mixture

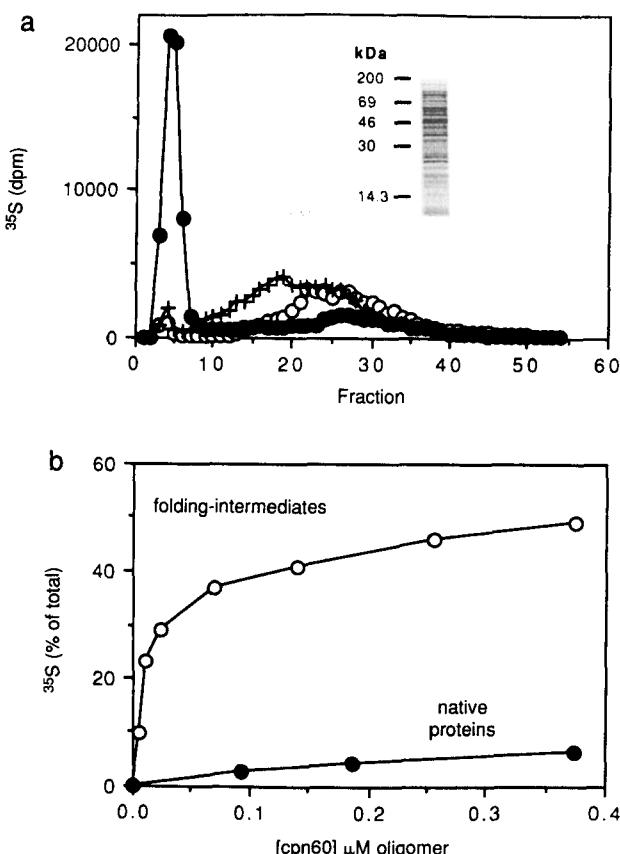


FIG. 54. (a) Binding of a total ^{35}S -labeled protein mixture from *E. coli* with GroEL. Radioactivity profiles of protein mixtures are shown after its incubation with 5 M GdmCl and then dissolution in native buffer with (●) or without (○) GroEL. (+) Radioactivity profile for the mixture of native proteins (not incubated with 5 M GdmCl). Inset: Fluorograph of the protein mixture used in this study. (b) Titration of ^{35}S -labeled *E. coli* proteins with GroEL (chaperonin 60) after the incubation in 5 M GdmCl and return to the native conditions (○) and in the native states (●). From Vitanen *et al.* (1992). Copyright 1992 Cambridge University Press. Reprinted with the permission of Cambridge University Press.

of native proteins containing [^{35}S]methionine has a smooth distribution of elution volumes expected for a mixture of proteins of different molecular weights. The same mixture, after its incubation in 5 M GdmCl and following dilution to return to native conditions, has a similar distribution (slightly shifted to smaller molecular weights presumably due to the disruption of native oligomeric structures). However, in the presence of GroEL this mixture (after incubation in 5 M GdmCl) demonstrates

a strong (~50%) binding to this chaperonin. Since the mixture after incubation in 5 M GdmCl may differ from the native mixture only by the presence of a fraction of proteins which have not yet succeeded in folding, it means that GroEL binds about 50% of total proteins of *E. coli* in the form of folding intermediates. Titrations by GroEL of the mixture of *E. coli* proteins before and after incubation in GdmCl (see Fig. 54b) show large differences demonstrating a large affinity of folding intermediates for GroEL. The physiological significance of this observation has been illustrated by showing that the complex mixture of *E. coli* proteins interacts with only one or a few sites on the GroEL oligomer.

The strong affinity of folding intermediates to GroEL is confirmed in Fig. 55, which shows what occurred when the mixture of proteins was unfolded by GdmCl and then was incubated in a native buffer before GroEL was added. If the time of this incubation increases, the fraction of proteins bound to GroEL strongly decreases and the fraction of unbound proteins increases (see Fig. 55b), demonstrating that the material trapped by GroEL is not a stable population of molecules, but a mixture of transient intermediates, as it must be if this material consists of molecules which have not finished their folding. Figure 55a illustrates the elution profile of the protein mixture before and after 9 hr of incubation.

This important work emphasizes that the binding of folding intermediates by chaperonin 60 is a rule rather than an exception and, thus, is not based on some specific amino acid sequences (Lorimer, 1992). All these data taken together leave little doubt that GroEL recognizes proteins in the molten globule-like state. It is possible that the molten globule can be recognized by a hole in oligomers, and it can adjust itself to the dimensions of this hole.

2. *Catalysis of Protein Folding*

There are two other proteins which facilitate protein folding, protein disulfide-isomerase (PDI) (Goldberger *et al.*, 1963; Venetianer and Straub, 1963) and peptidylprolyl isomerase (PPI) (Fischer *et al.*, 1984). Unlike GroEL or related chaperones, these proteins possess real catalytic functions: PDI catalyzes disulfide bond formation and PPI catalyzes *cis-trans* isomerization of proline residues (see Schmid, 1991, 1992; Lorimer, 1992; Freedman, 1992, for reviews). Some evidence has accumulated suggesting that both these folding enzymes combine their catalytical function with the ability to recognize folding intermediates, i.e., with chaperone-like function.

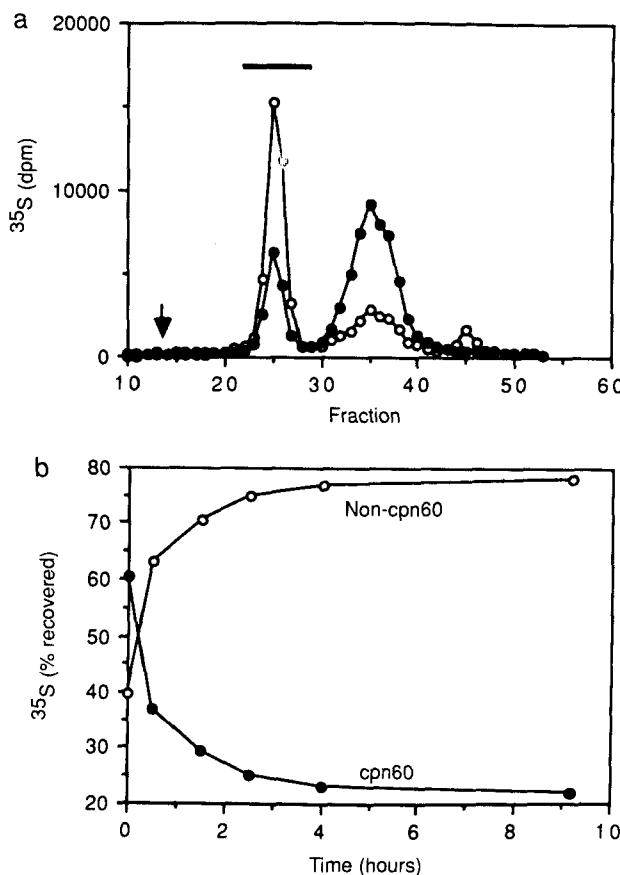


FIG. 55. (a) Radioactive profiles of ^{35}S -labeled protein mixture from *E. coli* which has been unfolded by 5 M GdmCl and returned to native conditions without its subsequent incubation (○) or after its incubation for 9 hr (●). The horizontal bar shows the elution volume of 14-mer of GroEL (cpn 60), while the arrow shows the void volume. (b) Decrease of fraction bound to cpn 60 and the increase of fraction unbound to cpn 60 as function of incubation time of ^{35}S -labeled protein mixture from *E. coli*. From Viitanen *et al.* (1992). Copyright 1992 Cambridge University Press. Reprinted with the permission of Cambridge University Press.

The first data suggesting that PDI may have chaperone-like properties were the observations that it binds to nascent chains of immunoglobulin M, preventing them for nonspecific aggregation (Roth and Pierce, 1987). Later it was also shown that PDI interacts with unstructured peptides (Noiva *et al.*, 1991; Morjana and Gilbert, 1991) in a manner similar to hsp-70 chaperones (Flynn *et al.*, 1989).

More recently, very convincing evidence was obtained that PDI catalyzes disulfide bond formation in bovine pancreatic trypsin inhibitor (BPTI) mainly in its kinetically trapped folding intermediates (Weissman and Kim, 1993). The native BPTI has three disulfide bonds, two of them, (5–55) and (30–51), are on one side of a molecule, while the third, (14–38), is on the other side. Two intermediates accumulate during reoxidation and refolding of the reduced and unfolded protein; both of them contain two native disulfide bonds, (14–38), and one of the pairs on the opposite side of the molecule, (5–55) or (30–51): the intermediates N^* (5–55, 14–38) and N' (30–51, 14–38) (see Creighton, 1977, 1992). The third intermediate N_{SH}^{SH} (5–55, 30–51) with two native disulfide bonds on the same side of the molecule rapidly reoxidizes to native BPTI (N), while two other intermediates, N^* and N' , do not (Fig. 56). It is very likely that the reason for this difference is that the reoxidation of the last S–S bonds in N^* and N' is hindered for steric reasons.

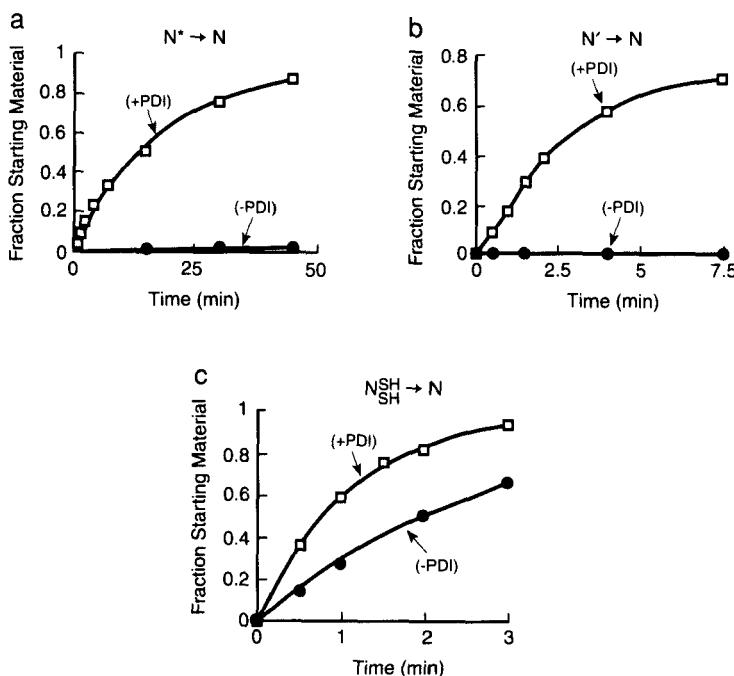


FIG. 56. Effects of PDI on the productive folding in the native states of three intermediates of BPTI with two native disulfide bonds: N^* (5–55, 14–38), N' (30–51, 14–38), and N_{SH}^{SH} (30–51, 5–55). The molar concentrations of BPTI and PDI are in the ratio of 5:1. Reproduced with permission from *Nature*, Weissman and Kim (1993). Copyright 1993 Macmillan Magazines Limited.

The interesting observation was that this very large difference between the behavior of N_{SH}^{SH} and N^* or N' is practically removed by PDI. This enzyme leads to only a three-fold increase in the oxidation rate of N_{SH}^{SH} , which in this case is reasonably high even in the absence of PDI. By contrast, it can accelerate much smaller rates of oxidation of N^* and N' several thousand times, making them quite comparable with that for N_{SH}^{SH} (see Fig. 56). It should be noted that PDI has only moderate influence on the rate of disulfide bond formation in the fully reduced protein and in single-disulfide intermediates (30–51) and (5–55). It is clear that the large effects of PDI are specific to the kinetically trapped folding intermediates.

It was also shown that PDI catalyzes the rearrangement of N' and N^* and N_{SH}^{SH} even in the absence of redox reagents, which demonstrates that PDI acts mainly by increasing the rate of intramolecular rearrangement. This may be compared with the influence of high concentrations of denaturants (6 M urea), which accelerates the rate of rearrangement of both N^* (Creighton and Goldenberg, 1984) and N' (Weissman and Kim, 1991), suggesting that this rearrangement needs substantial loss of a rigid protein structure (see Weissman and Kim, 1992). It is quite possible that protein disulfide-isomerase transforms BPTI into a semi-flexible molten globule state making possible the rearrangements of its molecules which almost completely hindered in the native state.

Another interesting publication (La Mantia and Lennarz, 1993) has demonstrated that a mutant of PDI, that has lost the disulfide-isomerase activity, still influences the growth of a yeast strain. These authors suggested that the function of PDI may be mainly related not to its catalytic function but to its ability to bind newly synthesized polypeptides in the lumen of the endoplasmic reticulum and to assist their folding. It is possible that PDI has a peptide-binding domain which may act in a chaperone-like manner.

Another enzyme involved in protein folding, PPI, also has a chaperone function. The class of heat-shock proteins with molecular weight of 56–59 kDa (called cyclophilins) is identical to PPI (Schreiber, 1991). It was shown that PPI not only accelerates the folding of carbonic anhydrase but, when added early in the folding process, prevents it from aggregating, thus ensuring 100% restoration of enzyme activity (Freskgård *et al.*, 1992). Therefore, it is possible that PPI can act as a chaperone at the early stage of protein folding, when the molten globule state is formed, and as a catalyst at its late stage, when *cis-trans* isomerization of proline residues should be accelerated. It is reasonable to assume that PDI and PPI bind their protein substrates in the molten globule state, preventing aggregation, and then accelerate the last stages of their folding.

3. Assembly

It has been observed that some proteins, e.g., aspartate aminotransferase (Herold and Kirschner, 1990) and brain-derived neutrotropic factor (Philo *et al.*, 1993), function in a dimer state, but are molten globules in a monomer form. It was shown that their monomeric molten globules can be recognized by chaperones, which help them to form homo- or heterodimers. For example, α subunits of bacterial luciferase are compact under physiological conditions, have a pronounced secondary structure, but have no rigid tertiary structure and no activity, i.e., are in the molten globule state (Flynn *et al.*, 1993). GroEL binds both α and β subunits of this protein (releasing them after addition of Mg²⁺ATP) but does not bind to active dimers. It is likely that the α subunit of bacterial luciferase is expressed as the molten globule and then waits in a complex with GroEL for β subunits (which are expressed later) to form native active dimers.

Much more complicated processes, such as the assembly of the cytoskeleton, including intermediate filaments, microtubules, and microfilaments, also can in some cases involve transitions from molten globule monomers to rigid dimers. There is a specific cytosolic complex, called TCP1, which assists in protein folding and is a mammalian equivalent of GroEL (Gao *et al.*, 1992; Frydman *et al.*, 1992). This complex is a heterooligomeric toroidal particle, which includes heat-shock proteins and has a total molecular mass of about 900 kDa. Later it was shown (Sternlicht *et al.*, 1993) that newly synthesized α -tubulin, β -tubulin, and actin enter a large complex (with a molecular mass of about 900 kDa), which coelutes with TCP1 and precipitates with TCP1 antibody. Tubulin monomers can be released from this complex in a state competent to form heterodimers. The mitotic spindle assembly is also destroyed in TCP1 mutants (Ursic and Culberston, 1991). These data support the suggestion that TCP1 complex is a cytosolic chaperone which is also involved in actin and tubulin folding *in vivo*. The β subunit of tubulin in the TCP1 complex may be structurally different from those in free states and in heterodimers, as they have different epitopes. This suggests that tubulin monomers are in a nonnative state before dimerization, and the molten globule is a good candidate for this state. Actin monomers released from the TCP1 complex are stable in solution and can be polymerized only in the presence of a number of other proteins, including hsp-70 (Haus *et al.*, 1993).

It has been proposed that ubiquitin, which is a heat-shock protein, is involved not only in protein degradation, but also in protein folding (Varshavsky *et al.*, 1988) and in the assembly of nucleosomes, ribosomes,

and other cell organelles (see Section VI,D,3). New data appear which demonstrate that such a classical chaperone as *E. coli* DnaK has similar properties. Alix and Guérin (1993) have shown that thermosensitive mutants of DnaK lead to accumulation at high temperature (45°C) of wrong ribosomal particles with sedimentation constants 45, 35, and 25 S, in addition to normal 30 and 50 S. These abnormal particles do not form at 30°C. This suggests that the thermosensitivity of these mutants is due to the wrong ribosome assembly and that the product of gene Dnak participates in ribosome biogenesis at elevated temperatures.

These data suggest that some ribosomal proteins may be in a nonnative state and require chaperones to be included in ribosomes. The assumption that some ribosomal proteins do not have rigid structures before ribosome assembly is consistent with the observation that a number of these proteins (including those with tryptophan residues) have pronounced far-UV CD spectra typical for native proteins, but that their near-UV CD spectra are more typical of these unfolded proteins (Veniaminov *et al.*, 1981). It is likely that these proteins are in the molten globule state in solution and need assistance to be included into ribosomes.

C. Penetration into Membranes

1. Nonnative Translocation-Competent State

The first evidence of physiological process in which a nonnative protein state is involved comes from the studies of transmembrane protein transport performed mainly by Schatz and collaborators (Eilers and Schatz, 1986, 1988; Vestweber and Schatz, 1988; Glick and Schatz, 1991; see also Randall and Hardy, 1986). These studies led to the definite conclusion that the translocation-competent state of proteins and their precursors is not the native state. In fact, it was shown (see Bychkova and Ptitsyn, 1993a, for references) that: (1) native proteins stabilized by specific ligands or disulfide bonds are not translocated; (2) transient intermediates trapped on both sides of the membrane surface are sensitive to proteases and thus are nonnative; and (3) newly synthesized polypeptide chains or proteins renatured from high concentrations of strong denaturants are competent for translocation, but lose this competence during incubation without membranes.

Most important were the experiments of Schatz's group (Eilers *et al.*, 1988), which showed that proteins renatured from urea or GdmCl are translocated faster and more efficiently than native proteins and retain this ability even at low temperature, but lose it after incubation without membranes. At least four conclusions have been drawn from these very

important observations: (1) the translocation-competent state is non-native, (2) this state is a folding intermediate, (3) the translocation-competent state is separated from the native state by a high potential barrier, and (4) this state exists for a relatively long time after the start of renaturation. Only the molten globule state meets all of these requirements and therefore we have suggested (Bychkova *et al.*, 1988) that the translocation-competent state of protein molecules is the molten globule state.

The same paper of Schatz's group also presented evidence that even native proteins being bound to the membrane surface are transformed into the translocation-competent state similar to that of folding intermediates. They conclude that "a cell organelle can partly unfold on its surface precursors which must be imported into this organelle and that it is unlikely that this reaction is physiologically irrelevant" (Eilers *et al.*, 1988). In subsequent papers by Schatz's group (Endo and Schatz, 1988; Eilers *et al.*, 1989; Endo *et al.*, 1989) it was shown that the partial unfolding of native proteins near membranes is due to negative charges on the membrane surface.

In fact, as early as 1979 it was realized (Eisenberg *et al.*, 1979) that a membrane surface with a strong negative potential can attract protons leading to a local decrease in pH of about 2 units at 5–15 Å from the surface in accord with simple electrostatic theory. Therefore, it looks as if "membranes not only need the molten globule state to translocate proteins, but also induce the transition to this state by their own negative potential" (Bychkova and Ptitsyn, 1993a). Of course, an even more effective source of a local decrease in pH can be a permanent efflux of protons through membranes and pumps (see, e.g., Skulachev, 1988).

It is also possible that the decrease of effective dielectric constant near the membrane surface can strengthen electrostatic interactions and enable a local decrease in pH to transform native proteins into the molten globule state (Bychkova and Ptitsyn, 1993a).

Moreover, it was shown that a membrane surface cannot only partly unfold native proteins but it can partly fold unfolded protein chains. A good example was presented by De Jongh *et al.* (1992) who demonstrated that apo- and holocytochrome c, being structurally quite different in the absence of membrane (apo is a virtually unfolded chain, whereas holo is a typical rigid protein), are in the presence of detergent micelles in virtually the same state that has many features of the molten globule.

It is reasonable to assume that in a typical case nascent polypeptides are transferred from ribosomes to membranes (see Fig. 57, center) by chaperones which prevent them both from folding too fast and from aggregating. However, even if a protein comes to a membrane in the

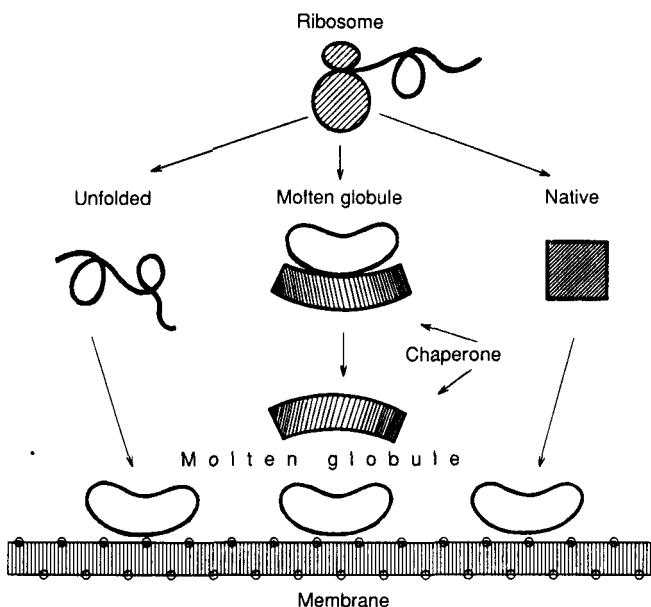


FIG. 57. A scheme illustrating that proteins must be in the molten globule state near a membrane surface. Usually they are transferred in this state to membranes directly from ribosomes by chaperones, but even if they are unfolded (like apocytochrome c) or folded into rigid structures they can be transformed into the molten globule state by the influence of a membrane. From Bychkova and Ptitsyn (1993a). Copyright © 1993 by Data Trace Chemistry Publishers, Inc. Published in *Chemtracts*, Volume 4, Number 3, pp. 133–163 and reproduced here with permission.

native state (Fig. 57, right) or in the unfolded state, like apocytochrome c (Fig. 57, left), the membrane surface transforms it into the molten globule state.

2. Toxin Penetration into Membranes

The first experimental evidence that a protein near a membrane is in the molten globule state was obtained for a precursor of a protein OmpA from the outer membrane of *E. coli* (Lecker *et al.*, 1990), but the most convincing data come from a study of toxin penetration into membranes. These data were obtained for the colicins, a group of toxins which kill cells via the loss of K⁺ ions (Bourdineaud *et al.*, 1990). Their action consists of three steps: colicin binds to a receptor of the outer membrane, translocates through this membrane, and finally permeabilizes the inner membrane. This protein consists of three domains, each of which is responsible for one of these functions: translocation, receptor binding, and pore formation.

A water-soluble form of the C-terminal (pore-forming) domain of colicin A consists of 10 α helices, 2 of which form a nonpolar hairpin inside the molecule, while 8 others, being amphiphilic, surround the nonpolar hairpin, screening it from water (Parker *et al.*, 1989). The affinity of this domain to membranes is much larger at low pH and the rate of colicin insertion into membranes sharply increases between pH 5.5 and 4.2 (van der Goot *et al.*, 1991). On the other hand, the pore-forming domain of colicin A transforms to the molten globule state at low pH, and the striking point is that the local pH at the membrane surface, needed for the penetration of this domain, is the same as the overall pH required for the native–molten globule state transition in aqueous solution (van der Goot *et al.*, 1991).

This interesting observation led van der Goot *et al.* (1992) to the idea that local pH at the membrane surface denatures the pore-forming domain to the molten globule state, which makes this domain flexible and permits its hydrophobic hairpin to penetrate into the membrane (Lakey *et al.*, 1992). Alternatively, van der Goot *et al.* (1992) suggest that the molten globule state of the pore-forming domain may be triggered not by low pH but by the unfolding of this domain by colicin A receptor and/or by other parts of the translocation machinery. In both cases colicin A reaches the inner membrane in the molten globule state.

It is likely that pore-forming domains of other toxins have a similar behavior. For example, the pore-forming domain of colicin E1 also transforms into the molten globule state at low pH (Merrill *et al.*, 1990), which suggests a similar mechanism for its penetration into membranes. δ Toxin from *Bacillus thuringiensis* also has a pore-forming domain, which includes seven α helices which enclose a single long hydrophobic helix inside its structure (Li *et al.*, 1991).

The role of the molten globule state in penetrating into a membrane can also be suggested for human complement C9. This protein is converted from a water-soluble state to integral membrane oligomers during the assembly of the membrane–attack complex. This transformation is coupled with a partial denaturation of this protein, as one of its temperature-induced transitions occurs already at 36°C; it means that this protein is partially denatured at body temperature (Lohner and Esser, 1991).

Acid-induced transitions from the native state to the molten globule state are very likely relevant to the translocation of endocytosed toxins which enter into the low pH interior of endocytic vesicles prior to transmembrane transfer from these vesicles into the cytosol. As the pH within these vesicles is as low as pH 4.5, these toxins are under denaturing conditions before being translocated. The best studied among these

toxins is the diphtheria toxin (see London, 1992, and references therein). This toxin consists of two disulfide-linked fragments, A and B, with molecular masses of 27 and 37 kDa, respectively. The catalytic A fragment releases into the cytoplasm and kills the cell by ribosylating of elongation factor 2, which is involved in protein biosynthesis.

The X-ray structure of this toxin (Choe *et al.*, 1992) shows that the catalytic fragment A is similar to catalytic domains of some other toxins. Fragment B consists of two domains: the receptor-binding domain, which is a flattened β barrel, and the membrane translocation domain, which has a structure similar to pore-forming domains of colicin A (Parker *et al.*, 1989) and δ toxin (Li *et al.*, 1991). It consists of nine α helices; two of them are unusually hydrophobic and are buried inside the protein.

There is evidence that both fragments A and B can be transformed into the molten globule state by low pH or high temperature (Dumont and Richards, 1988; Zhao and London, 1988; Cabiaux *et al.*, 1989; Ramsay *et al.*, 1989; Ramsey and Freire, 1990; Jiang *et al.*, 1991; London, 1992; van der Goot *et al.*, 1992) and these transitions correlate with the degree of membrane insertion (Papini *et al.*, 1987; Dumont and Richards, 1988; Zhao and London, 1988; Cabiaux *et al.*, 1989). Combining all existing data, van der Goot *et al.* (1992) came to the general scheme for translocations of diphtheria toxin shown in Figs. 58.

After toxin has entered the endosome, fragment B transforms into the molten globule state and inserts into the lipid layer. Then, upon further acidification of endosomes, fragment A also transforms into the molten globule state which triggers its insertion and translocation through the membrane. Unlike fragment B, the native–molten globule state transition of fragment A is reversible, which permits it to refold in the cytoplasm back to its native structure and to restore its biological activity.

Pseudomonas exotoxin A also consists of three domains (Allured *et al.*, 1986), and its translocation domain is α helical. It also undergoes a transition to the molten globule state at acid pH (Farahbakhsh and Wisnieski, 1989; Jiang and London, 1990). It is reasonable to suggest that the mechanism of translocation of this toxin shares some important features with that for diphtheria toxin.

3. Cell Fusion

Another example of pH-induced penetration of nonpolar protein segments into a membrane is the entry of viruses (e.g., influenza virus) into a cell. Often this entry takes place via endocytic vesicles, and it is believed to be triggered by a conformational transition in a so-called fusion protein influenced by low pH inside these vesicles. This transition may lead to an exposure of a nonpolar fragment of a fusion protein

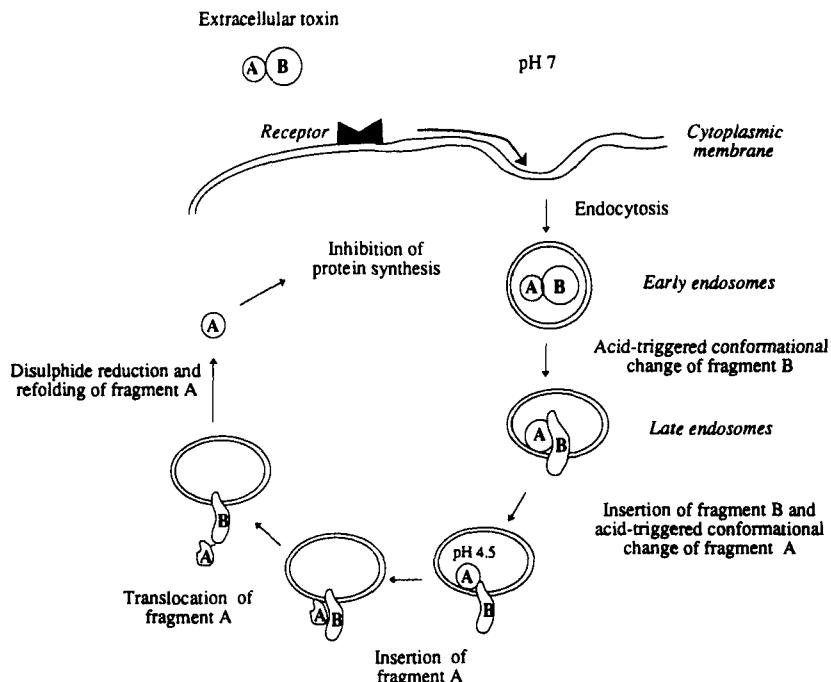


FIG. 58. Scheme of the steps in the translocation of diphtheria toxin. Reproduced with permission from van der Goot *et al.* (1992).

("fusion peptide") and its penetration into membrane of the target cell (Li *et al.*, 1991; White, 1992; Tsûrûdome *et al.*, 1992).

The best-studied example is hemagglutinin, the protein which participates in the fusion of influenza virus with the target cell by the mechanism described above (Tsûrûdome *et al.*, 1992). Recently, Blumenthal's group has studied the acid-induced conformational transition of this protein and has shown that this transition transforms hemagglutinin into the molten globule state (Krumbiegel *et al.*, 1995). They have shown that upon the transition of hemagglutinin from the native state (pH 7.4; 37°C) to "fusogenic conditions" (pH 4.9; 37°C) its near-UV CD spectrum practically vanishes. The transition monitored by the near-UV CD spectrum is rather sharp (within less than one pH unit). Temperature and the enthalpy of temperature melting of hemagglutinin decrease on changing the pH from 7.4 to 5.3, while at pH 4.8 no temperature melting could be observed. On the other hand, the far-UV CD spectrum of hemagglutinin and anisotropy of its tryptophan fluorescence remain practically unaltered upon this transition, which suggests that the protein preserves

its secondary structure and its compactness at fusogenic pH. In addition, hemagglutinin binds a nonpolar hydrophobic fluorescent probe (Nile Red) substantially more strongly at fusogenic pH than at neutral pH. Thus, hemagglutinin is in a typical molten globule state at fusogenic conditions. This result strongly supports the idea that the exposure of fusion peptide and its penetration into the target membrane are triggered by the native–molten globule state transition in hemagglutinin.

4. Translocation

It is natural to assume that the molten globule state is a necessary prerequisite not only for protein insertion into membranes but also for its translocation through membranes (Bychkova *et al.*, 1988; van der Goot *et al.*, 1992; Bychkova and Ptitsyn, 1993a). In many cases nascent protein chains may be recognized by chaperones just after their biosynthesis and transported to membranes in the “salvage pathway” for proteins which have “missed their chance of cotranslational targeting” (Rapoport, 1990).

The state in which protein molecules exist during the transit across membranes depends, of course, on the mechanism of this transit. The transit of a protein chain in its extended conformation, which has been observed for translocation into mitochondria (Rassow *et al.*, 1990), is energetically unfavorable if transit takes place directly through the hydrophobic phospholipid bilayer. In this case the molten globule would be a better candidate to the translocating state (Bychkova *et al.*, 1988; van der Goot *et al.*, 1992). However, if proteins are translocated through hydrophilic tunnels formed by transmembrane proteins (Rapoport, 1990; Simon and Blobel, 1991; High *et al.*, 1991), the molten globule state may be required just to facilitate the local unfolding of a chain to an extended conformation, which is used for translocation (van der Goot *et al.*, 1992).

D. Miscellaneous

1. Transfer of Nonpolar Ligands

Proteins can release nonpolar ligands (retinol, fatty acids, bilin, etc.) near a membrane surface transferring them into the target cells. The best studied among these proteins is plasma retinol-binding protein (RBP) which delivers retinol to specific receptors on the cell surface, but may also do it with liposomes (which do not contain any receptors). The rate of this release strongly increases at low pH (Fex and Johannesson, 1987). Bychkova *et al.* (1992) have shown that retinol can be completely

released from RBP at low pH even in the absence of membranes and that this release is accompanied by the transition of RBP into the molten globule state. The evidence of this conclusion is that RBP at $\text{pH} \leq 4$ is relatively compact and has a pronounced secondary structure but has no rigid tertiary structure and almost no cooperative temperature melting, and, in addition, strongly binds ANS.

Figure 59 illustrates an analogy between the insertion into membranes of nonpolar regions of toxins and other proteins and the targeted release of nonpolar ligands from their carriers to membranes. Both nonpolar protein regions and specific nonpolar ligands are usually deeply buried in a protein moiety and tightly packed within the rest of the protein molecule. Therefore, the release of both of them needs to overcome a high potential barrier. This barrier is dramatically reduced in the molten globule state which has no tight packing between the released part of

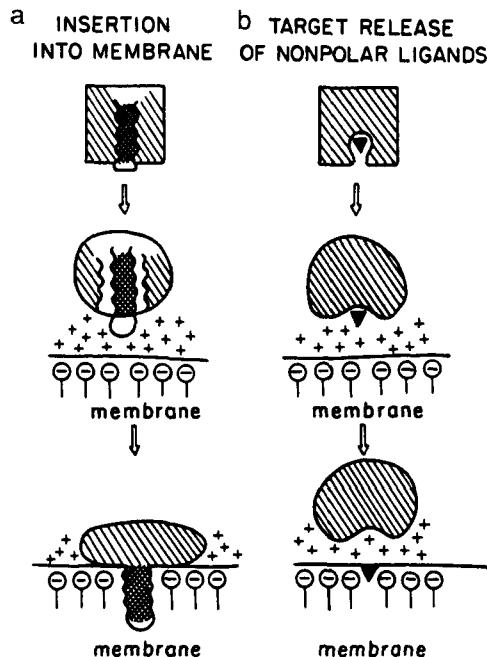


FIG. 59. The possible common mechanism of protein insertion into membrane (a) and of target release of nonpolar ligands (b). Both of these processes may need the molten globule state as a necessary prerequisite. The transition of proteins into the molten globule state may be induced by the strong negative electrostatic potential of a membrane surface (which decreases pH near a membrane) and probably also by the local decrease of effective dielectric constant near a membrane surface. From Bychkova and Ptitsyn (1993a). Copyright © 1993 by Data Trace Chemistry Publishers, Inc. Published in *Chemtracts*, Volume 4, Number 3, pp. 133–163 and reproduced here with permission.

a protein and the rest of the molecule. Therefore, the molten globule state may be a necessary prerequisite not only for protein penetration into a membrane but also for a targeted release of nonpolar protein ligands (Bychkova and Ptitsyn, 1993a).

2. *Protein-Receptor Interactions*

The key-lock mechanism of protein-receptor interactions probably is not always correct, since these interactions may need some semiflexible structure of both these components to facilitate their mutual adjustments. For example, it was shown that engineering of the insertion of Gly into insulin molecules which increases its internal flexibility can lead to a similar or even higher potency for interaction with insulin receptor, while too much flexibility decreases its receptor-binding potency (Nakagawa and Tager, 1993). This suggests that insulin binds to its receptor in a folded but not tightly packed conformation. The nature of this conformation was revealed by Hua *et al.* (1992, 1993) who presented good evidence that a DPI, analogous to monomeric insulin, at low pH or in 20% acetic acid is in the molten globule state (see Section II,B,2). It was also shown that a fully active engineered insulin monomer at physiological conditions has tertiary NOEs similar to those of DPI (Weiss *et al.*, 1991). This leads to the suggestion that the functional state of insulin in the bloodstream is the molten globule state (Hua *et al.*, 1993).

On the other hand, there is evidence that receptors also may need some flexibility to form complexes with their protein ligands. Apo forms of steroid receptors are found in cytosol in large complexes with different types of heat-shock protein, hsp-90, hsp-70, and hsp-56-59 (Tai *et al.*, 1992, 1993). It was suggested (Tai *et al.*, 1993) that these heat-shock proteins form a core structure of a universal chaperone which may be responsible for recognition, folding, assembly, and disassembly of steroid receptors or other regulatory proteins at or near the end of translocation. Steroid receptors can bind their ligands only when associated with heat-shock proteins and have no binding activity without these proteins (Camaña *et al.*, 1993). Holosteroid receptors (with bound ligands) lose their affinities to heat-shock proteins and acquire their DNA-binding and transcriptional activities.

Apo forms of steroid receptors are more sensitive to digestion than their holo forms (Allan *et al.*, 1992), and their proteolytic fragments are recognized by different antibodies (Smith *et al.*, 1992). It was also shown that the ligand-binding domain of steroid receptors is directly involved in the interaction with hsp-90 (Cadepond *et al.*, 1993). Thus, there is a state of steroid receptors competent for the binding of ligand, which

these receptors acquire only in a complex with heat-shock proteins. The molten globule state is a good candidate for this state.

3. Ubiquitin-Dependent Processes

We have suggested (Bychkova *et al.*, 1988) that not only the lysosomal pathway of protein degradation (at pH 4.5–5.0), but also such degradation in the ubiquitin-dependent system may involve a preceding transition of a protein into the molten globule state. There is some evidence supporting this suggestion. First, it was shown that a necessary prerequisite of protein degradation often is a side-chain modification of proteins (see Hershko and Ciechanover, 1982, 1992; Rechsteiner, 1987, and references therein). These modifications can change the conformational state of a protein, which has been demonstrated for the metal-catalyzed oxidation of glucose-6-phosphate dehydrogenase (Fritgues *et al.*, 1993). In this case the properties of the modified protein are consistent with the assumption that the protein is in the molten globule state. Second, there is evidence that the degradation-competent states of proteins may be related to the state recognized by chaperones, as chaperones can also enhance protein degradation (Chiang *et al.*, 1989; Sherman and Goldberg, 1992).

This suggests (Bychkova *et al.*, 1988; Bychkova and Ptitsyn, 1993a) that covalent attachment of ubiquitin to a protein may transform it into the molten globule state. This assumption is supported indirectly by some evidence that ubiquitin may be involved in the insertion of proteins into membranes (Zhuang and McCauley, 1989; Zhuang *et al.*, 1992; Finley and Chau, 1991).

In general, ubiquitin is a heat-shock protein whose role in the life of a cell is by no means limited to protein degradation (Rechsteiner, 1988; Jentsch *et al.*, 1990). It is involved in the assembly of ribosomes (Finley *et al.*, 1989), nucleosomes (Bonner *et al.*, 1988), microfilaments (Ball *et al.*, 1987), microtubules (Murti *et al.*, 1988), etc. This suggests that ubiquitin can destabilize a protein's structure in such a way that the protein becomes competent for incorporation into different cell structures and that this "incorporation competent" protein state is the molten globule state (Bychkova and Ptitsyn, 1993a).

4. Diseases

An intriguing possibility is that the molten globule may be involved in mechanisms of some diseases connected with point mutations in proteins, which lead to their mislocation in a cell. An example of these diseases is cystic fibrosis, which is caused by a deletion of Phe-508 from cystic fibrosis transmembrane conductance regulator (CFTR). It was

shown (Yang *et al.*, 1993) that this mutant protein was retained in the endoplasmic reticulum instead of being delivered to the plasma membrane. Newly synthesized CFTR enters into a complex with chaperone hsp-70 and the wild-type protein dissociates from hsp-70 before its transport to the Golgi and subsequent traffic to plasma membrane. By contrast, the mutant protein remains in the complex with hsp-70 in the endoplasmic reticulum and is rapidly degraded in a pre-Golgi nonlysosomal compartment.

Thus, hsp-70 can discriminate between the normal and mutant forms of CFTR, which suggests that this chaperone performs quality control during the biosynthesis of plasma membrane proteins. Mutants which can not be released from chaperones remain in the endoplasmic reticulum where they are degraded (Machamer *et al.*, 1990; Accili *et al.*, 1992).

It is likely that a similar mechanism also operates in the pathogenesis of some other diseases including, for example, hypercholesterolemia, connected with a mutation in low-density lipoprotein receptor (Lehrman *et al.*, 1987), and emphysema, connected with mutation in α -1-antitrypsin (Lomas *et al.*, 1992). In both cases mutant proteins may also be in complexes with chapterones and therefore are retained in the endoplasmic reticulum.

There are convincing data showing that point mutations can hinder the last stage of protein folding and trap a protein in a folding intermediate under physiological conditions. Probably the first evidence was obtained by King and coauthors (Goldenberg and King, 1981; Smith and King, 1981). These authors have shown that some temperature-sensitive mutants of phage P22 tailspike protein cannot fold and assemble into active proteins. Some of these mutants can fold at lower temperatures which suggests that those mutations happen to be consistent with native structure (Goldenberg *et al.*, 1983a,b).

After the molten globule state was discovered, Craig *et al.* (1985) showed that two mutants of β -lactamase (replacements of Thr-40 by Ile and Asp-146 by Asn) are in a nonactive form which is slightly less compact than the native state and has a native content of secondary structure, but has very substantially reduced near-UV CD spectra and is susceptible to proteolysis.

More recently, Lim *et al.* (1992) studied point mutations in a nonpolar core of the N-terminal domain of λ repressor. Replacements of one nonpolar residue by another nonpolar residue were basically consistent with the native structure, while replacements of nonpolar residues by polar ones (Leu-18 \rightarrow Asn, Val-36 \rightarrow Asn, Val-36 \rightarrow Asp) dramatically change the properties of the protein. These mutants become inactive *in vivo*, have 100- to 1000-fold smaller affinities to antibodies against

the native protein, no cooperative temperature melting, and strongly reduced cooperativity of GdmCl-induced unfolding. Their NMRs become broad and much less dispersed than those in the native protein. On the other hand, these mutants restore 50 to 80% of their native far-UV CD ellipticities and their NMR spectra still have resonances substantially shifted from their expected random coil positions. The authors concluded that these mutants are blocked in the molten globule state.

Keeping these examples in mind, we cannot exclude the possibility that mutations, which cause some diseases by changing the intracellular pathway of proteins, also inhibit the last stage of protein folding and trap a protein in the molten globule state (Bychkova and Ptitsyn, 1995). This does not permit a mutant protein to use normal trafficking in the cell and leads to its mislocation and degradation.

VII. CONCLUSION

This review is an attempt to describe the present state of the studies of the molten globule state and its role in protein folding and physiological processes. Several points should be emphasized in conclusion: (1) the molten globule very likely has a native-like overall architecture (folding pattern) without including the rigid packing of side chains (Hua *et al.*, 1992, 1993; Peng and Kim, 1994); (2) the molten globule is separated by first-order phase transitions from both the native and the unfolded states and, therefore, represents a third thermodynamic state of protein molecules (Uversky *et al.*, 1992; Ptitsyn and Uversky, 1994); (3) the molten globule is a universal kinetic intermediate in protein folding (Ptitsyn *et al.*, 1990); the structure of this intermediate is similar to the structure of the equilibrium molten globule (Baldwin, 1993) and has many native-like features (Fersht, 1993); (4) successful protein folding may be the result of evolutionary selection of sequences, which makes the native structure much more stable than each of the alternatives and permits the protein to fold by a directed process (Shakhnovich and Gutin, 1993a,b); and (5) the molten globule is involved in a number of physiological processes in the living cell (Martin *et al.*, 1991; van der Goot *et al.*, 1991, 1992; Bychkova *et al.*, 1992; Flynn *et al.*, 1993; Krumbiegel *et al.*, 1995, and others; see Bychkova and Ptitsyn, 1993a, for review).

The last point is probably the most important since it demonstrates the direct bridge between the physics of protein and cell biology and suggests that there are in fact two native states of protein molecules, the rigid state and the molten globule state.

ACKNOWLEDGMENTS

I gratefully acknowledge the kind help of Irina Ptitsyna in preparing the manuscript. I thank the Laboratory of Mathematical Biology of the National Cancer Institute, NIH,

for support and hospitality. The properties of the molten globule state were the topics of numerous discussions with Robert Baldwin, Valentina Bychkova, Tom Creighton, Ken Dill, Dmitry Dolgikh, Christopher Dobson, Vladimir Filimonov, Alexey Finkelstein, Rudolph Gilmanishin, Michel Goldberg, Yuri Griko, Alexander Gutin, Neville Kallenbach, Martin Karplus, Peter Kim, Kunihiro Kuwajima, Roger Pain, Peter Privalov, Heinrich Roder, Gian-Luigi Rossi, Gottfried Schatz, Gennady Semisotnov, Vladimir Uversky, Sergey Venyaminov, Peter Wright, and other colleagues. I am especially thankful to Eugene Shakhnovich for cooperation and numerous useful discussions which were of special importance for this work. Valentina Bychkova has written part of Section VI. Valentina Bychkova, Robert Jernigan, Frederic Richards, Heinrich Roder, and Eugene Shakhnovich have read the manuscript and have made many important comments. R. L. Blumenthal, C. M. Dobson, P. S. Kim, P. L. Privalov, H. Roder, and P. Wright provided me with preprints of their papers before publication. Alexandra Dujsekina and Tatjana Salnikova have checked the references. I am thankful to all of them for their generous help. The work on this review was partly supported by a grant from Human Frontier Science Program Organization.

REFERENCES

- Abkevich, V. I., Gutin, A. M., and Shakhnovich, E. I. (1994). *Biochemistry* **33**, 10026–10036.
- Accili, D., Kadowaki, T., Kadowaki, H., Mosthaf, L., Ullrich, A., and Taylor, S. I. (1992). *J. Biol. Chem.* **267**, 586–590.
- Acharya, K. R., Ren, J. S., Stuart, D. I., Phillips, D. C., and Fenna, R. E. (1991). *J. Mol. Biol.* **221**, 571–581.
- Ahmad, F., and Salahuddin, A. (1974). *Biochemistry* **13**, 245–249.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., eds. (1983). “Molecular Biology of the Cell,” Chapters 3, 6, 7. Garland Publishing, New York.
- Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., and Dobson, C. M. (1993). *Biochemistry* **32**, 1707–1718.
- Alexandrescu, A. T., Ng, Y.-L., and Dobson, C. M. (1994). *J. Mol. Biol.* **235**, 587–599.
- Alix, J.-H., and Guérin, M.-F. (1993). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9725–9729.
- Allan, G. F., Leng, X. H., Tsai, S. Y., Weigel, N. L., Edwards D. P., Tsai, M. J., and O’Malley, B. W. (1992). *J. Biol. Chem.* **267**, 19513–19520.
- Allured, V. S., Collier, R. J., Carroll, S. F., and McKay, D. B. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1320–1324.
- Anfinsen, C. B. (1973). *Science* **181**, 223–230.
- Anson, M. L. (1945). *Adv. Protein Chem.* **2**, 361–386.
- Anson, M. L., and Mirsky, A. E. (1934). *J. Gen. Physiol.* **17**, 393–398.
- Anufrieva, E. V., Bychkova, V. E., Krakovyack, M. G., Pautov, V. D., and Ptitsyn, O. B. (1975). *FEBS Lett.* **55**, 46–49.
- Anufrieva, E. V., Krakovyack, M. G., Gromova, R. A., Lushchik, V. B., Ananyeva, T. D., and Sheveleva, T. V. (1991). *Dokl. Akad. Nauk SSSR* **319**, 895–898.
- Aune, H. C., and Tanford, C. (1969). *Biochemistry* **8**, 4586–4590.
- Aune, H. C., Salahuddin, A., Zarlenyo, M. H., and Tanford, C. (1967). *J. Biol. Chem.* **242**, 4486–4489.
- Bai, Y., Milne, J. S., and Englander, S. W. (1993). *Proteins: Struct., Funct. Genet.* **17**, 75–86.
- Baldwin, R. L. (1991). *Chemtracts: Biochem. Mol. Biol.* **2**, 379–389.
- Baldwin, R. L. (1993). *Curr. Opin. Struct. Biol.* **3**, 84–91.
- Baldwin, R. L., and Roder, H. (1991). *Curr. Opin. Struct. Biol.* **1**, 218–220.
- Ball, E., Karlik, C. C., Beal, C. J., Saville, D. L., Sparrow, J. C., Bullard, B., and Fyrberg, E. A. (1987). *Cell (Cambridge, Mass.)* **51**, 221–228.

- Berrick, D., and Baldwin, R. L. (1993). *Biochemistry* **32**, 3790–3796.
- Baudet, S., and Janin, J. (1991). *J. Mol. Biol.* **219**, 123–132.
- Baum, J., Dobson, C. M., Evans, P. A., and Hanly, C. (1989). *Biochemistry* **28**, 7–13.
- Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., Onuffer, J. J., and Matthews, C. R. (1986). *Biochemistry* **25**, 2965–2974.
- Betton, J.-M., Desmadril, M., Mitraki, A., and Yon, J. M. (1985). *Biochemistry* **24**, 4570–4577.
- Bi, R. C., Dauter, Z., Dodson, E., Dodson, G., Giordano, F., and Reynolds, C. (1984). *Biopolymers* **23**, 391–395.
- Birshtein, T. M., and Pryamitsyn, V. A. (1991). *Macromolecules* **24**, 1554–1560.
- Blond-Elguindi, S., and Goldberg, M. E. (1990). *Biochemistry* **29**, 2409–2417.
- Blundell, T. B., and Johnson, L. N. (1976). "Protein Crystallography." Academic Press, New York.
- Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. (1988). *Nature (London)* **336**, 254–257.
- Boelens, R., Ganadu, M. L., Verheeden, P., and Kaptein, R. (1990). *Eur. J. Biochem.* **191**, 147–153.
- Bolotina, I. A. (1987). *Mol. Biol. (Moscow)* **21**, 1625–1635.
- Bonner, W. M., Hatch, C. L., and Wu, R. S. (1988). In "Ubiquitin" (M. Rechsteiner, ed.), pp. 157–172. Plenum, New York.
- Bourdineaud, J. P., Boulanger, P., Lazdunski, C., and Letellier, L. (1990). *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1037–1041.
- Brandts, J. E., and Hunt, I. (1967). *J. Am. Chem. Soc.* **89**, 4826–4838.
- Briggs, M. S., and Roder, H. (1992). *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2017–2021.
- Bryngelson, J. D., and Wolynes, P. G. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524–7528.
- Bryngelson, J. D., and Wolynes, P. G. (1989). *J. Phys. Chem.* **93**, 6902–6915.
- Bryngelson, J. D., and Wolynes, P. G. (1990). *Biopolymers* **30**, 177–188.
- Buchner, J., Schmid, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kieffhaber, T. (1991). *Biochemistry* **30**, 1586–1591.
- Buck, M., Radford, S. E., and Dobson, C. M. (1993). *Biochemistry* **32**, 669–678.
- Bushnell, G. W., Louie, G. V., and Brayer, G. D. (1990). *J. Molec. Biol.* **214**, 585–595.
- Bychkova, V. E., and Ptitsyn, O. B. (1993a). *Chemtracts: Biochem. Mol. Biol.* **4**, 133–163.
- Bychkova, V. E., and Ptitsyn, O. B. (1993b). *Biofizika* **38**, 58–66.
- Bychkova, V. E., and Ptitsyn, O. B. (1995). *FEBS Lett.* **359**, 6–8.
- Bychkova, V. E., Semisotnov, G. V., Ptitsyn, O. B., Gudkova, O. V., Mitin, Yu. V., and Anufrieva, E. V. (1980). *Mol. Biol. (Moscow)* **14**, 178–186.
- Bychkova, V. E., Pain, R. H., and Ptitsyn, O. B. (1988). *FEBS Lett.* **238**, 231–234.
- Bychkova, V. E., Bartoshevich, S. F., and Klenin, S. I. (1990). *Biofizika* **35**, 242–248.
- Bychkova, V. E., Berni, R., Rossi, G.-I., Kutyshenko, V. P., and Ptitsyn, O. B. (1992). *Biochemistry* **31**, 7566–7571.
- Bycroft, M., Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1990). *Nature (London)* **346**, 488–490.
- Bycroft, M., Ludvigsen, S., Fersht, A. R., and Poulsen, F. M. (1991). *Biochemistry* **30**, 8697–8701.
- Caamaño, C. A., Morano, M. I., Patel, P. D., Watson, S. J., and Akil, H. (1993). *Biochemistry* **32**, 8589–8595.
- Cabiaux, V., Brasseur, R., Wattiez, R., Falmagne, P., Ruivschert, J.-M., and Goormaghtigh, E. (1989). *J. Biol. Chem.* **264**, 4928–4938.
- Cadepond, F., Binart, N., Chambraud, B., Jibard, N., Schweizer-Groyer, G., Segard-Maurel, J., and Baulieu, E. E. (1993). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10434–10438.

- Cavard, P., Sauve, P., Heitz, F., Pattus, F., Martinez, C., Dijkman, R., and Lazdunski, C. (1988). *Eur. J. Biochem.* **172**, 507–512.
- Chaffotte, A. F., Guillon, Y., and Goldberg, M. E. (1992). *Biochemistry* **31**, 9694–9702.
- Chakrabarty, A., and Baldwin, R. L. (1993). *ACS Symp. Ser.* **526**, 166–177.
- Chan, H. S., and Dill, K. A. (1989). *Macromolecules* **22**, 4559–4573.
- Chan, H. S., and Dill, K. A. (1990a). *J. Chem. Phys.* **92**, 3118–3135.
- Chan, H. S., and Dill, K. A. (1990b). *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6388–6392.
- Chan, H. S., and Dill, K. A. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 447–490.
- Chan, H. S., and Dill, K. A. (1993). *Physics Today*, 24–32.
- Chen, M. C., and Lord, R. C. (1976). *Biochemistry* **15**, 1889–1897.
- Chiang, H.-L., Terlecky, S. R., Plant, C. P., and Dice, J. F. (1989). *Science* **246**, 382–385.
- Choe, S., Bennetti, M. J., Fujii, G., Curmi, P. M. G., Kantardjeff, R., Collier, J., and Eisenberg, D. (1992). *Nature (London)* **357**, 216–222.
- Christensen, H., and Pain, R. H. (1991). *Eur. Biophys. J.* **19**, 221–229.
- Chyan, C.-L., Wormald, C., Dobson, C. M., Evans, P. A., and Baum, J. (1993). *Biochemistry* **32**, 5681–5691.
- Corbett, R. J. T., and Roche, R. S. (1984). *Biochemistry* **23**, 1888–1894.
- Craig, S., Hollecker, M., Creighton, T. E., and Pain, R. H. (1985). *J. Mol. Biol.* **185**, 681–687.
- Creighton, T. E. (1977). *J. Mol. Biol.* **113**, 275–293.
- Creighton, T. E. (1980). *J. Mol. Biol.* **137**, 61–80.
- Creighton, T. E. (1990). *Biochem. J.* **270**, 1–16.
- Creighton, T. E. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 301–351. Freeman, New York.
- Creighton, T. E., and Ewbank, J. J. (1994). *Biochemistry* **33**, 1534–1538.
- Creighton, T. E., and Goldenberg, D. P. (1984). *J. Mol. Biol.* **179**, 497–526.
- Creighton, T. E., and Pain, R. H. (1980). *J. Mol. Biol.* **137**, 431–436.
- Damaschun, G., Gernat, C., Damaschun, H., Bychkova, V. E., and Ptitsyn, O. B. (1986). *Int. J. Biol. Macromol.* **8**, 226–230.
- De Gennes, P. G. (1975). *J. Phys. Lett.* **36**, L 55–L 57.
- De Jongh, H. H. J., Killian, J. A., and de Kruijff, B. (1992). *Biochemistry* **31**, 1636–1643.
- Derrida, B. (1981). *Phys. Rev. B* **24**, 2613–2626.
- Dill, K. A. (1985). *Biochemistry* **24**, 1501–1509.
- Dill, K. A., Alonso, D. O. V., and Hutchinson, K. (1989). *Biochemistry* **28**, 5439–5449.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995). *Protein Sci.* **4**, 561–602.
- Dobson, C. M. (1991). *Curr. Opin. Struct. Biol.* **1**, 22–27.
- Dobson, C. M. (1992). *Curr. Opin. Struct. Biol.* **2**, 6–12.
- Dobson, C. M., Hanley, C., Radford, S. E., Baum, J. A., and Evans, P. A. (1991). In "Conformations and Forces in Protein Folding" (B. T. Nall and K. A. Dill, eds.), pp. 175–181. Am. Assoc. Adv. Sci., Washington, D.C.
- Dolgikh, D. A. (1983). Ph.D. Thesis, Institute of Protein Research, Academy of Sciences of the USSR, Pushchino.
- Dolgikh, D. A., Gilmanishin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Yu., and Ptitsyn, O. B. (1981). *FEBS Lett.* **136**, 311–315.
- Dolgikh, D. A., Abaturov, L. V., Brazhnikov, E. V., Lebedev, Yu. O., Chirgadze, Yu. N., and Ptitsyn, O. B. (1983). *Dokl. Akad. Nauk SSSR* **272**, 1481–1484.
- Dolgikh, D. A., Kolomiets, A. P., Bolotina, I. A., and Ptitsyn, O. B. (1984). *FEBS Lett.* **165**, 88–92.
- Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Bushuev, V. N., Gilmanishin, R. I., Lebedev, Yu. O., Semisotnov, G. V., Tiktropulo, E. I., and Ptitsyn, O. B. (1985). *Eur. Biophys. J.* **13**, 109–121.

- Dumont, M. E., and Richards, F. M. (1988). *J. Biol. Chem.* **263**, 2087–2097.
- Dyson, H. J., and Wright, P. E. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 519–538.
- Dyson, H. J., and Wright, P. E. (1993). *Curr. Opin. Struct. Biol.* **3**, 60–65.
- Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988). *J. Mol. Biol.* **201**, 161–200.
- Dyson, H. J., Merutka, G., Walther, J. P., Lerner, R. A., and Wright, P. E. (1992a). *J. Mol. Biol.* **226**, 795–817.
- Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., and Wright, P. E. (1992b). *J. Mol. Biol.* **226**, 819–835.
- Eilers, M., and Schatz, G. (1986). *Nature (London)* **322**, 228–232.
- Eilers, M., and Schatz, G. (1988). *Cell (Cambridge, Mass.)* **52**, 481–483.
- Eilers, M., Hwang, S., and Schatz, G. (1988). *EMBO J.* **7**, 1139–1145.
- Eilers, M., Endo, T., and Schatz, G. (1989). *J. Biol. Chem.* **264**, 2945–2950.
- Eisenberg, M. A., and Schwert, G. W. (1951). *J. Gen. Physiol.* **34**, 583–606.
- Eisenberg, M. A., Gresalfi, T., Riccio, T., and McLaughlin, S. (1979). *Biochemistry* **18**, 5213–5223.
- Elöve, G. A., and Roder, H. (1991). *ACS Symp. Ser.* **470**, 50–63.
- Elöve, G. A., Chaffotte, A. F., Roder, H., and Goldberg, M. E. (1992). *Biochemistry* **31**, 6876–6883.
- Elöve, G. A., Bhuyan, A. K., and Roder, H. (1994). *Biochemistry* **33**, 6925–6935.
- Endo, T., and Schatz, G. (1988). *EMBO J.* **7**, 1153–1158.
- Endo, T., Eilers, M., and Schatz, G. (1989). *J. Biol. Chem.* **264**, 2951–2956.
- Englander, J. J., Calhoun, D. B., and Englander, S. W. (1979). *Anal. Biochem.* **92**, 517–524.
- Englander, S. W., and Mayne, L. (1992). *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243–265.
- Englander, S. W., and Poulsen, A. (1969). *Biopolymers* **7**, 379–393.
- Ewbank, J. J., and Creighton, T. E. (1991). *Nature (London)* **350**, 518–520.
- Ewbank, J. J., and Creighton, T. E. (1993). *Biochemistry* **32**, 3677–3693.
- Fabian, H., Schulz, C., Naumann, D., Landt, O., Haft, U., and Saenger, W. (1993). *J. Mol. Biol.* **232**, 967–981.
- Fan, P., Bracken, C., and Baum, J. (1993). *Biochemistry* **32**, 1573–1582.
- Farahbakhsh, Z. T., and Wisnieski, B. J. (1989). *Biochemistry* **28**, 580–585.
- Fedorov, B. A., and Ptitsyn, O. B. (1977). *Dokl. Akad. Nauk SSR* **233**, 716–718.
- Feng, Y., Wand, A. J., and Sligar, S. G. (1991). *Biochemistry* **30**, 7711–7717.
- Feng, Y., Sligar, S. G., and Wand, A. J. (1994). *Nat. Struct. Biol.* **1**, 30–35.
- Fersht, A. R. (1987). *Biochemistry* **26**, 8031–8037.
- Fersht, A. R. (1993). *FEBS Lett.* **325**, 5–16.
- Fersht, A. R., Matouschek, A., and Serrano, L. (1992). *J. Mol. Biol.* **224**, 771–782.
- Fex, G., and Johannesson, G. (1987). *Biochim. Biophys. Acta* **901**, 255–264.
- Finkelstein, A. V. (1977). *Mol. Biol. (Moscow)* **11**, 811–819.
- Finkelstein, A. V., and Ptitsyn, O. B. (1977). *Biopolymers* **16**, 469–495.
- Finkelstein, A. V., and Ptitsyn, O. B. (1987). *Prog. Biophys. Mol. Biol.* **50**, 171–190.
- Finkelstein, A. V., Ptitsyn, O. B., and Kositsyn, S. A. (1977). *Biopolymers* **16**, 497–524.
- Finkelstein, A. V., Badretdinov, A. Y., and Ptitsyn, O. B. (1990). *Nature (London)* **345**, 300.
- Finkelstein, A. V., Badretdinov, A. Y., and Ptitsyn, O. B. (1991). *Proteins: Struct., Funct., Gene.* **10**, 287–299.
- Finley, D., and Chau, V. (1991). *Annu. Rev. Cell Biol.* **7**, 25–69.
- Finley, D., Bartel, B., and Varshavsky, A. (1989). *Nature (London)* **338**, 394–401.
- Fischer, G., and Schmid, F. X. (1990). *Biochemistry* **29**, 2205–2212.
- Fischer, G., Bang, H., and Mech, C. (1984). *Biomed. Biophys. Acta* **43**, 1101–1111.
- Flory, P. J. (1953). "Principles of Polymer Chemistry." Cornell Univ. Press, Ithaca, NY.

- Flynn, G. C., Chappell, T. G., and Rothman, J. E. (1989). *Science* **245**, 385–390.
- Flynn, G. C., Becker, C. J. M., Baase, W. A., and Dahlquist, F. W. (1993). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10826–10830.
- Freedman, R. B. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 455–539. Freeman, New York.
- Freire, E., and Biltonen, R. L. (1978). *Biopolymers* **17**, 463–479.
- Freskgård, P.-O., Bergenhem, N., Jonsson, B.-H., Svensson, M., and Carlsson, U. (1992). *Science* **258**, 466–468.
- Fritgut, B., Szweda, L. I., and Stadtman, E. R. (1993). American Society of Biochemistry & Molecular Biology and the Division of Biological Chemistry. Thesis, American Chemical Society, Joint Meeting, San Diego, CA.
- Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P., and Hartl, F.-U. (1992). *EMBO J.* **11**, 4767–4778.
- Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H., and Cowan, N. J. (1992). *Cell (Cambridge, Mass.)* **69**, 1043–1050.
- Garel, J.-R. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 405–454. Freeman, New York.
- Garvey, E. P., and Matthews, C. R. (1989). *Biochemistry* **28**, 2083–2093.
- Gast, K., Zirwer, D., Welfle, H., Bychkova, V. E., and Ptitsyn, O. B. (1986). *Int. J. Biol. Macromol.* **8**, 231–236.
- Gattner, H. G. (1975). *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1397–1404.
- Gething, M.-J., and Sambrook, J. (1992). *Nature (London)* **355**, 33–45.
- Gilmanshin, R. I. (1985). Ph.D. Thesis, Institute of Protein Research, Academy of Sciences of the USSR, Pushchino.
- Gilmanshin, R. I., and Ptitsyn, O. B. (1987). *FEBS Lett.* **223**, 327–329.
- Gilmanshin, R. I., Dolgikh, D. A., Ptitsyn, O. B., Finkelstein, A. V., and Shakhnovich, E. I. (1982). *Biofizika* **27**, 1005–1016.
- Glick, G., and Schatz, G. (1991). *Annu. Rev. Genet.* **25**, 21–44.
- Gō, N., and Abe, H. (1981). *Biopolymers* **20**, 991–1011.
- Goldberg, M. E., Semisotnov, G. V., Fritgut, B., Kuwajima, K., Ptitsyn, O. B., and Sugai, S. (1990). *FEBS Lett.* **263**, 51–56.
- Goldberger, R. F., Epstein, C. J., and Anfinsen, C. B. (1963). *J. Biol. Chem.* **238**, 628–635.
- Goldenberg, D. P., and King, J. (1981). *J. Mol. Biol.* **145**, 633–651.
- Goldenberg, D. P., Smith, D. H., and King, J. (1983a). *Biopolymers* **22**, 125–129.
- Goldenberg, D. P., Smith, D. H., and King, J. (1983b). *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7060–7064.
- Goto, Y., and Fink, A. L. (1989). *Biochemistry* **28**, 945–952.
- Goto, Y., Calciano, L. J., and Fink, A. L. (1990a). *Proc. Natl. Acad. Sci. U.S.A.* **87**, 573–577.
- Goto, Y., Takahashi, N., and Fink, A. L. (1990b). *Biochemistry* **29**, 3480–3488.
- Goto, Y., Hagihara, Y., Hamada, D., Hoshino, M., and Nishii, I. (1993). *Biochemistry* **32**, 11878–11885.
- Gregoret, L., and Cohen, F. (1991). *J. Mol. Biol.* **219**, 109–122.
- Griko, Yu. V., and Privalov, P. L. (1994). *J. Mol. Biol.* **235**, 1318–1325.
- Griko, Yu. V., Privalov, P. L., Venyaminov, S. Yu., and Kutyshenko, V. P. (1988). *J. Mol. Biol.* **202**, 127–138.
- Griko, Yu. V., Freire, E., and Privalov, P. L. (1994). *Biochemistry* **33**, 1889–1899.
- Grosberg, A. Yu. (1984). *Biofizika* **29**, 563–573.
- Grosberg, A. Yu., and Shakhnovich, E. I. (1986a). *Zh. Eksp. Teor. Fiz.* **91**, 837–850.
- Grosberg, A. Yu., and Shakhnovich, E. I. (1986b). *Biofizika* **31**, 1054–1057.
- Hagihara, Y., Aimoto, S., Fink, A. L., and Goto, Y. (1993). *J. Mol. Biol.* **231**, 180–184.

- Hamaguchi, K., and Sakai, H. (1965). *J. Biochem. (Tokyo)* **57**, 103–114.
- Harding, M. M., Williams, D. H., and Woolfson, D. N. (1991). *Biochemistry* **30**, 3120–3128.
- Haus, U., Trommler, P., Fisher, P. R., Hartmann, H., Lottspeich, F., Noegel, A. A., and Schleicher, M. (1993). *EMBO J.* **12**, 3763–3771.
- Haynie, D. T., and Freire, E. (1993). *Proteins: Struct., Funct., Genet.* **16**, 115–143.
- Herold, M., and Kirschner, K. (1990). *Biochemistry* **29**, 1907–1913.
- Hershko, A., and Ciechanover, A. (1982). *Annu. Rev. Biochem.* **51**, 335–364.
- Hershko, A., and Ciechanover, A. (1992). *Annu. Rev. Biochem.* **61**, 761–807.
- High, S., Gorlich, D., Wiederun, M., Rapoport, T. A., and Dobberstein, B. (1991). *J. Cell Biol.* **113**, 35–44.
- Hill, T. L. (1968). In "Thermodynamics of Small Systems" (W. A. Benjamin, ed.). Wiley, New York.
- Holcomb, D. N., and van Holde, K. E. (1962). *J. Phys. Chem.* **66**, 1999–2006.
- Holladay, L. A., Hammonds, R. G., Jr., and Puett, D. (1974). *Biochemistry* **13**, 1653–1661.
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990). *J. Mol. Biol.* **216**, 1031–1044.
- Hua, Q.-X., and Weiss, M. A. (1990). *Biochemistry* **29**, 10545–10555.
- Hua, Q.-X., Kochyan, M., and Weiss, M. A. (1992). *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2379–2383.
- Hua, Q.-X., Ladbury, J. E., and Weiss, M. A. (1993). *Biochemistry* **32**, 1433–1442.
- Hughson, F. M., and Baldwin, R. L. (1989). *Biochemistry* **28**, 4415–4422.
- Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990). *Science* **249**, 1544–1548.
- Hughson, F. M., Barrick, D., and Baldwin, R. L. (1991). *Biochemistry* **30**, 4113–4118.
- Hurle, M. R., Tweedy, N. B., and Matthews, C. R. (1986). *Biochemistry* **25**, 6356–6360.
- Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. (1986). *Biochemistry* **25**, 6965–6972.
- Ikeguchi, M., Sugai, S., Fujiro, M., Sugawara, T., and Kuwajima, K. (1992). *Biochemistry* **31**, 12695–12700.
- Jacobs, M. D., and Fox, R. O. (1994). *Proc. Natl. Acad. Sci. U.S.A.* **91**, 449–453.
- Jaenicke, R. (1981). *Annu. Rev. Biophys. Bioeng.* **10**, 1–67.
- Jaenicke, R. (1987). *Prog. Biophys. Mol. Biol.* **49**, 117–237.
- Jaenicke, R. (1988). In "Protein Structure and Protein Folding" (R. Huber and E. L. Winnacker, eds.), Colloq. Mosbach, Vol. 39, pp. 16–36. Springer Verlag, Berlin and New York.
- Jaenicke, R. (1991). *Biochemistry* **30**, 3147–3161.
- James, E., Wu, P. G., Stites, W., and Brand, L. (1992). *Biochemistry* **31**, 10217–10225.
- Jeng, M. F., Englander, S. W., Elöve, G. A., Wang, A. I., and Roder, H. (1990). *Biochemistry* **29**, 10433–10437.
- Jennings, P. A., and Wright, P. E. (1993). *Science* **262**, 892–896.
- Jentsch, S., Seufert, W., Sommer, T., and Reins, H.-A. (1990). *Trends Biochem. Sci.* **15**, 195–198.
- Jiang, J. X., and London, E. (1990). *J. Biol. Chem.* **265**, 8636–8641.
- Jiang, J. X., Abrams, F. S., and London, E. (1991). *Biochemistry* **30**, 3857–3864.
- Jones, C. M., Henry, E. R., Hu, Y., Chan, C.-K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, I., and Eaton, W. A. (1993). *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11860–11864.
- Karplus, M., and Shakhnovich, E. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 127–195. Freeman, New York.
- Karplus, M., and Weaver, D. L. (1976). *Nature (London)* **260**, 404–406.
- Karplus, M., and Weaver, D. L. (1994). *Protein Sci.* **3**, 650–688.
- Kataoka, M., Haghara, Y., Mihara, K., and Goto, Y. (1993). *J. Mol. Biol.* **229**, 591–596.
- Kauzmann, W. (1959). *Adv. Protein Chem.* **14**, 1–63.

- Kawata, Y., and Hamaguchi, K. (1991). *Biochemistry* **30**, 4367–4373.
- Khorasanizadeh, S., Peters, I. D., Butt, T. K., and Roder, H. (1993). *Biochemistry* **32**, 7054–7063.
- Kim, J., and Kim, H. (1986). *Biochemistry* **25**, 7867–7874.
- Kim, P. S., and Baldwin, R. L. (1980). *Biochemistry* **19**, 6124–6129.
- Kim, P. S., and Baldwin, R. L. (1982). *Annu. Rev. Biochem.* **51**, 459–489.
- Kim, P. S., and Baldwin, R. L. (1990). *Annu. Rev. Biochem.* **59**, 631–660.
- Kim, P. S., Bierzynski, A., and Baldwin, R. L. (1982). *J. Mol. Biol.* **162**, 187–199.
- Kronman, M. J., Blum, R., and Holmes, L. G. (1966). *Biochemistry* **4**, 917–944.
- Krumbiegel, M., Remeta, D. P., Booy, F. P., Kunton, J. R., Ginsburg, A., and Blumenthal, R. (1995). Submitted for publication.
- Kugimiya, H., and Bigelow, C. C. (1973). *Can. J. Biochem.* **51**, 581–585.
- Kuroda, Y., Kidokoro, S.-I., and Wada, A. (1992). *J. Mol. Biol.* **223**, 1139–1153.
- Kuwajima, K. (1977). *J. Mol. Biol.* **114**, 241–258.
- Kuwajima, K. (1989). *Proteins: Struct., Funct., Genet.* **6**, 87–103.
- Kuwajima, K. (1992). *Curr. Opin. Biotechnol.* **3**, 462–467.
- Kuwajima, K., and Schmid, F. X. (1984). *Adv. Biophys.* **18**, 43–74.
- Kuwajima, K., Nitta, K., Yoneyama, M., and Sugai, S. (1976). *J. Mol. Biol.* **106**, 359–373.
- Kuwajima, K., Hiroka, Y., Ikeguchi, M., and Sugai, S. (1985). *Biochemistry* **24**, 874–881.
- Kuwajima, K., Harushima, Y., and Sugai, S. (1986). *Int. J. Pept. Protein Res.* **27**, 18–27.
- Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., and Nagamura, T. (1987). *FEBS Lett.* **221**, 115–118.
- Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M., and Sugai, S. (1988). *Biochemistry* **27**, 7419–7428.
- Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R., and Sugai, S. (1991). *Biochemistry* **30**, 7693–7703.
- Kuwajima, K., Semisotnov, G. V., Finkelstein, A. V., Sugai, S., and Ptitsyn, O. B. (1993). *FEBS Lett.* **334**, 265–268.
- Labhardt, A. M. (1982). *J. Mol. Biol.* **157**, 331–335.
- Lakey, J. H., González-Manas, J. M., van der Goot, F. G., and Pattus, F. (1992). *FEBS Lett.* **307**, 26–29.
- Lala, A. K., and Kaul, P. (1992). *J. Biol. Chem.* **267**, 19914–19918.
- La Mantia, M. L., and Lennarz, W. J. (1993). *Cell (Cambridge, Mass.)* **74**, 899–908.
- Laminet, A. A., Ziegelhoffer, J., Georgopoulos, C., and Plückthun, A. (1990). *EMBO J.* **9**, 2315–2319.
- Lecker, S. H., Driessens, A. J. M., and Wickner, W. (1990). *EMBO J.* **9**, 2309–2314.
- Lee, J. W., and Kim, H. (1988). *FEBS Lett.* **268**, 408–414.
- Lehrman, M. A., Schneider, W. J., Brown, M. S., Davis, C. G., Elhammer, A., Russell, D. W., and Goldstein, J. L. (1987). *J. Biol. Chem.* **262**, 401–410.
- Levinthal, C. (1968). *J. Chem. Phys.* **65**, 44–45.
- Levinthal, C. (1969). In "Mossbauer Spectroscopy of Biological Systems" (P. Debrunner, L.-C. Tsibris, and E. Munck, eds.), pp. 22–24. Univ. of Illinois Press, Urbana.
- Levitt, M., and Chothia, C. (1976). *Nature (London)* **261**, 552–558.
- Li, J. D., Carroll, J., and Ellar, D. J. (1991). *Nature (London)* **353**, 815–821.
- Liff, M. I., Lyn, P. C., and Kallenbach, N. R. (1991). *J. Am. Chem. Soc.* **113**, 1014–1019.
- Lifshitz, I. M., Grosberg, A. Yu., and Khokhlov, A. R. (1978). *Rev. Mod. Phys.* **50**, 683–713.
- Lim, V. I. (1974). *J. Mol. Biol.* **88**, 857–872.
- Lim, W. A., Farrugio, D. C., and Sauer, R. T. (1992). *Biochemistry* **31**, 71–77.
- Lohner, K., and Esser, A. F. (1991). *Biochemistry* **30**, 6620–6625.

- Lomas, D. A., Evans, D. L., Finch, J. T., and Carrell, R. W. (1992). *Nature (London)* **357**, 605–607.
- London, E. (1992). *Biochim. Biophys. Acta* **1113**, 25–51.
- Lorimer, G. H. (1992). *Curr. Opin. Struct. Biol.* **2**, 26–34.
- Lu, J., and Dahlquist, F. W. (1992). *Biochemistry* **31**, 4749–4756.
- Lumry, R., Biltonen, R. L., and Brandts, J. F. (1966). *Biopolymers* **4**, 917–944.
- Lyn, P. C. C., Zhou, Jelveh, N., Wemmer, D. E., and Kallenbach, N. R. (1992a). *J. Am. Chem. Soc.* **114**, 6560–6562.
- Lyn, P. C. C., Gans, P. J., and Kallenbach, N. R. (1992b). *J. Mol. Biol.* **223**, 343–350.
- Machamer, C. E., Doms, R. W., Bole, D. G., Helenius, A., and Rose, J. K. (1990). *J. Biol. Chem.* **265**, 6879–6883.
- Manning, M. C., and Woody, R. W. (1989). *Biochemistry* **28**, 8609–8613.
- Mårtensson, L.-G., Jonsson, B.-H., Fresgård, P.-O., Kihlgren, A., Svensson, M., and Carlsson, V. (1993). *Biochemistry* **32**, 224–231.
- Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F.-U. (1991). *Nature (London)* **352**, 36–42.
- Martin, J., Horwich, A. L., and Hartl, F.-U. (1992). *Science* **258**, 995–998.
- Mateo, P. L., and Privalov, P. L. (1981). *FEBS Lett.* **123**, 189–192.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989). *Nature (London)* **342**, 122–126.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990). *Nature (London)* **346**, 440–445.
- Matouschek, A., Serrano, L., and Fersht, A. R. (1992a). *J. Mol. Biol.* **224**, 819–835.
- Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M., and Fersht, A. R. (1992b). *J. Mol. Biol.* **224**, 837–845.
- Matthews, C. R. (1991). *Curr. Opin. Struct. Biol.* **1**, 28–35.
- Matthews, C. R. (1993). *Annu. Rev. Biochem.* **62**, 653–683.
- Matthews, C. R., and Westmoreland, D. G. (1975). *Biochemistry* **14**, 4532–4538.
- Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C., and Jack, A. (1982). *Nature (London)* **297**, 162–164.
- Mayo, K. H., Barker, S., Kuranda, M. J., Hunt, A., Myers, J. A., and Maione, T. E. (1992). *Biochemistry* **31**, 12255–12265.
- McCoy, L. F., Rone, E. S., and Wong, U.-P. (1980). *Biochemistry* **19**, 4738–4743.
- Meewes M., Ricka, J., de Silva, M., Nyffenegger, R., and Binkert, T. (1991). *Macromolecules* **24**, 5811–5816.
- Merrill, A. R., Cohen, F. S., and Cramer, W. A. (1990). *Biochemistry* **29**, 5829–5836.
- Molday, R. S., Englander, S. W., and Kallen, R. G. (1972). *Biochemistry* **11**, 150–159.
- Morjana, N. A., and Gilbert, H. F. (1991). *Biochemistry* **30**, 4985–4990.
- Mullins, L. S., Pace, C. N., and Ranshel, F. M. (1993). *Biochemistry* **32**, 6152–6156.
- Murry-Brelier, A., and Goldberg, M. E. (1988). *Biochemistry* **27**, 7633–7640.
- Murti, K. G., Smith, H. T., and Fried, V. A. (1988). *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3019–3023.
- Nakagawa, S. H., and Tager, H. S. (1993). *Biochemistry* **32**, 7237–7243.
- Narhi, L. O., Rosenfeld, R., Wen, J., Arakawa, T., Prestrelski, S. J., and Philo, J. S. (1993). *Biochemistry* **32**, 10819–10825.
- Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992). *Science* **257**, 1559–1563.
- Nicoli, D. F., and Benedek, G. B. (1976). *Biopolymers* **15**, 2421–2437.
- Nilsson, B., and Anderson, S. (1991). *Annu. Rev. Microbiol.* **45**, 607–635.
- Nitta, K., Kitz, N., Kuwajima, K., and Sugai, S. (1977). *Biochim. Biophys. Acta* **490**, 200–208.
- Noiva, R., Kimura, H., Roos, J., and Lennarz, W. J. (1991). *J. Biol. Chem.* **266**, 19645–19649.

- Novokhatny, V. V., Kudinov, S. A., and Privalov, P. L. (1984). *J. Mol. Biol.* **179**, 215–232.
- Nozaka, M., Kuwajima, K., Nitta, K., and Sugai, S. (1978). *Biochemistry* **17**, 3753–3758.
- Ohgushi, M., and Wada, A. (1983). *FEBS Lett.* **164**, 21–24.
- Okazaki, A., Ikura, T., Nikaido, K., and Hiraoka, K., (1994). *Nat. Struct. Biol.* **1**, 439–445.
- Palleros, D. R., Reid, K. L., McCarty, J. S., Walker, G. C., and Fink, A. L. (1992). *J. Biol. Chem.* **267**, 5279–5285.
- Pan, Y., and Briggs, M. S. (1992). *Biochemistry* **31**, 11405–11412.
- Papini, E., Schiavo, G., Tomasi, M., Colombatti, M., Rappuoli, R., and Montecucco, C. (1987). *Eur. J. Biochem.* **169**, 637–644.
- Parker, M. W., Pattus, F., Tucker, A. D., and Tsernoglou, D. (1989). *Nature (London)* **337**, 93–96.
- Peng, Z., and Kim, P. S. (1994). *Biochemistry* **33**, 2136–2141.
- Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., and Matthews, C. R. (1989). *Biochemistry* **28**, 7961–7968.
- Perutz, M. F. (1992). "Protein Structure: New Approaches to Disease and Therapy." Freeman, New York.
- Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–678.
- Pfeil, W., and Privalov, P. L. (1976a). *Biophys. Chem.* **4**, 23–32.
- Pfeil, W., and Privalov, P. L. (1976b). *Biophys. Chem.* **4**, 33–40.
- Pfeil, W., and Privalov, P. L. (1976c). *Biophys. Chem.* **4**, 41–50.
- Pfeil, W., Bychkova, V. E., and Ptitsyn, O. B. (1986). *FEBS Lett.* **198**, 287–291.
- Philo, J. S., Rosenfeld, R., Arakawa, T., Wen, J., and Narhi, L. O. (1993). *Biochemistry* **32**, 10812–10818.
- Portnova, S. L., Bystrov, V. F., Balashova, T. A., Ivanov, V. T., and Ovchinnikov, Yu. A. (1970). *Izv. Akad. Nauk SSSR, Ser. Kim.*, pp. 825–830.
- Potekhin, S. A., and Pfeil, W. (1989). *Biophys. Chem.* **34**, 55–62.
- Privalov, P. L. (1979). *Adv. Protein Chem.* **33**, 167–241.
- Privalov, P. L. (1982). *Adv. Protein Chem.* **35**, 1–104.
- Privalov, P. L. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 83–126. Freeman, New York.
- Privalov, P. L., and Makhatadze, G. I. (1990). *J. Mol. Biol.* **213**, 385–391.
- Privalov, P. L., and Medved', L. V. (1982). *J. Mol. Biol.* **159**, 665–683.
- Privalov, P. L., Tiktupulo, E. I., Venyaminov, S. Yu., Griko, Yu. V., Makhatadze, G. I., and Khechinashvili, N. N. (1989). *J. Mol. Biol.* **205**, 737–750.
- Ptitsyn, O. B. (1973). *Dokl. Akad. Nauk SSSR* **210**, 1213–1215.
- Ptitsyn, O. B. (1975). *Dokl. Akad. Nauk SSSR* **213**, 1253–1255.
- Ptitsyn, O. B. (1983). In "Conformations in Biology" (R. Srinivasan and R. H. Sarma, eds.), pp. 49–58. Adenine Press, New York.
- Ptitsyn, O. B. (1984). *Mol. Biol. (Moscow)* **18**, 574–590.
- Ptitsyn, O. B. (1985). *J. Mol. Struct. (Theochem.)*, Elsevier Sci. Publ. D. V., pp. 45–65, Amsterdam.
- Ptitsyn, O. B. (1987). *J. Protein Chem.* **6**, 273–293.
- Ptitsyn, O. B. (1991). *FEBS Lett.* **285**, 176–181.
- Ptitsyn, O. B. (1992a). In "Protein Folding" (T. E. Creighton, ed.), pp. 243–300. Freeman, New York.
- Ptitsyn, O. B. (1992b). *Curr. Opin. Struct. Biol.* **2**, 13–20.
- Ptitsyn, O. B. (1994). *Protein Eng.* **7**, 593–596.
- Ptitsyn, O. B. (1995). *Curr. Opin. Struct. Biol.* **5**, 74–78.
- Ptitsyn, O. B., and Eizner, Yu. Ye. (1965). *Biofizika* **10**, 3–6.
- Ptitsyn, O. B., and Finkelstein, A. V. (1980). *Q. Rev. Biophys.* **13**, 339–386.

- Ptitsyn, O. B., and Finkelstein, A. V. (1983). *Biopolymers* **22**, 15–25.
- Ptitsyn, O. B., and Semisotnov, G. V. (1991). In "Conformations and Forces in Protein Folding" (B. T. Nall and K. A. Dill, eds.), pp. 155–168. Am. Assoc. Adv. Sci., Washington, D. C.
- Ptitsyn, O. B., and Uversky, V. N. (1994). *FEBS Lett.* **341**, 15–18.
- Ptitsyn, O. B., and Volkenstein, M. V. (1986). *Biomol. Struct. Dyn.* **4**, 137–156.
- Ptitsyn, O. B., Kron, A. K., and Eizner, Yu. Ye. (1968). *J. Polym. Sci., Part C* **16**, 3509–3517.
- Ptitsyn, O. B., Dolgikh, D. A., Gilmanishin, R. I., Shakhnovich, E. I., and Finkelstein, A. V. (1983). *Mol. Biol. (Moscow)* **17**, 569–576.
- Ptitsyn, O. B., Damaschun, G., Gernat, C., Damaschun, H., and Bychkova, V. E. (1986). *Stud. Biophys.* **112**, 207–221.
- Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerownik, E., and Razgulyaev, O. I. (1990). *FEBS Lett.* **262**, 20–24.
- Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, A., Brandenburg, D., Zahn, H., Gliemann, J., and Cammelfoft, S. (1976). *Nature (London)* **259**, 369–373.
- Radford, S. E., Dobson, C. M., and Evans, P. A. (1992a). *Nature (London)* **358**, 302–307.
- Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M., and Evans, P. A. (1992b). *Proteins: Struct., Funct., Genet.* **14**, 237–248.
- Ramsay, G., and Freire, E. (1990). *Biochemistry* **29**, 8677–8683.
- Ramsay, G., Montgomery, D., Berger, D., and Freire, E. (1989). *Biochemistry* **28**, 529–533.
- Randall, L. L., and Hardy, S. J. S. (1986). *Cell (Cambridge, Mass.)* **46**, 921–928.
- Rapoport, T. A. (1990). *Trends Biochem. Sci.* **15**, 355–358.
- Rassow, J., Hartl, F. U., Guiard, B., Pfanner, N., and Neupert, W. (1990). *FEBS Lett.* **275**, 190–194.
- Rechsteiner, M. (1987). *Annu. Rev. Cell Biol.* **3**, 1–30.
- Rechsteiner, M., ed. (1988). "Ubiquitin." Plenum, New York.
- Richards, F. M. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 1–58. Freeman, New York.
- Richardson, J. S. (1977). *Nature (London)* **268**, 495–500.
- Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
- Robertson, A. D., and Baldwin, R. L. (1991). *Biochemistry* **30**, 9907–9914.
- Robinson, C. V., Gros, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M., and Radford, S. E. (1994). *Nature (London)* **372**, 646–651.
- Robson, B., and Pain, R. H. (1976a). *Biochem. J.* **155**, 325–330.
- Robson, B., and Pain, R. H. (1976b). *Biochem. J.* **155**, 331–334.
- Roder, H., and Elöve, G. A. (1994). In "Frontiers in Molecular Biology" (R. H. Pain, ed.), pp. 26–59. Oxford Univ. Press, Oxford.
- Roder, H., Elöve, G. A., and Englander, S. W. (1988). *Nature (London)* **335**, 700–704.
- Rodionova, N. A., Semisotnov, G. V., Kutyshenko, V. P., Uversky, V. N., Bolotina, I. A., Bychkova, V. E., and Ptitsyn, O. B. (1989). *Mol. Biol. (Moscow)* **23**, 683–692.
- Roth, R. A., and Pierce, S. B. (1987). *Biochemistry* **26**, 4179–4182.
- Sali, A., Shakhnovich, E. I., and Karplus, M. (1994). *J. Mol. Biol.* **235**, 1614–1636.
- Sanchez, I. C. (1970). *Macromolecules* **12**, 980–988.
- Sanger, F. (1952). *Adv. Protein Chem.* **7**, 1–67.
- Schiffer, M., and Edmundson, A. B. (1967). *Biophys. J.* **7**, 121–135.
- Schmid, F. X. (1991). *Curr. Opin. Struct. Biol.* **1**, 36–41.
- Schmid, F. X. (1992). In "Protein Folding" (T. E. Creighton, ed.), pp. 197–241. Freeman, New York.
- Schmid, F. X. (1993). *Annu. Rev. Biophys. Biomol. Struct.* **22**, 123–143.

- Schmid, F. X., and Baldwin, R. L. (1979). *J. Mol. Biol.* **135**, 199–215.
- Scholz, J. M., and Baldwin, R. L. (1992). *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.
- Schreiber, S. L. (1991). *Science* **251**, 283–287.
- Semisotnov, G. V., Zikherman, K. Kh., Kasatkin, S. B., Ptitsyn, O. B., and Anufrieva, E. V. (1981). *Biopolymers* **20**, 2287–2309.
- Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blank, J., and Ptitsyn, O. B. (1987). *FEBS Lett.* **224**, 9–13.
- Semisotnov, G. V., Kutyshenko, V. P., and Ptitsyn, O. B. (1989). *Mol. Biol. (Moscow)* **23**, 808–815.
- Semisotnov, G. V., Uversky, V. N., Sokolovsky, I. V., Gutin, A. M., Razgulyaev, O. I., and Rodionova, N. A. (1990). *J. Mol. Biol.* **213**, 561–568.
- Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas', A. F., and Gil'manshin, R. I. (1991a). *Biopolymers* **31**, 119–128.
- Semisotnov, G. V., Vas, M., Chemeris, V. V., Kashparova, N. I., Kotova, N. N., Razgulyaev, O. I., and Sinev, M. M. (1991b). *Eur. J. Biochem.* **202**, 1083–1089.
- Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990). *Biochemistry* **29**, 9349–9352.
- Serrano, L., Bycroft, M., and Fersht, A. R. (1991). *J. Mol. Biol.* **218**, 465–475.
- Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., and Fersht, A. R. (1992a). *J. Mol. Biol.* **224**, 783–804.
- Serrano, L., Matouschek, A., and Fersht, A. R. (1992b). *J. Mol. Biol.* **224**, 805–818.
- Serrano, L., Matouschek, A., and Fersht, A. R. (1992c). *J. Mol. Biol.* **224**, 847–859.
- Seshadri, S., Öberg, K. A., and Fink, A. L. (1994). *Biochemistry* **33**, 1351–1355.
- Shakhnovich, E. I. (1994). *Phys. Rev. Lett.* **72**, 3907–3910.
- Shakhnovich, E. I., and Finkelstein, A. V. (1982). *Dokl. Akad. Nauk SSSR* **267**, 1247–1250.
- Shakhnovich, E. I., and Finkelstein, A. V. (1989). *Biopolymers* **28**, 1667–1680.
- Shakhnovich, E. I., and Gutin, A. M. (1989). *Biophys. Chem.* **34**, 187–199.
- Shakhnovich, E. I., and Gutin, A. M. (1990). *Nature (London)* **346**, 773–775.
- Shakhnovich, E. I., and Gutin, A. M. (1993a). *Protein Eng.* **6**, 793–800.
- Shakhnovich, E. I., and Gutin, A. M. (1993b). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7195–7199.
- Shakhnovich, E. I., Farztdinov, G. M., Gutin, A. M., and Karplus, M. (1991). *Phys. Rev. Lett.* **67**, 1665–1667.
- Shalongo, W., Jagannadham, M. V., Heid, P., and Stellwagen, E. (1992). *Biochemistry* **31**, 11390–11396.
- Sherman, M. Y., and Goldberg, A. L. (1992). *EMBO J.* **11**, 71–77.
- Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. I., Chaiken, L. M., Stewart, J. M., and Baldwin, R. L. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2349–2353.
- Shoemaker, K. R., Kim, P. S., York, E. I., Chaiken, L. M., Stewart, J. M., and Baldwin, R. L. (1987). *Nature (London)* **326**, 563–567.
- Simon, S. M., and Blobel, G. (1991). *Cell (Cambridge, Mass.)* **65**, 371–380.
- Skolnick, J., Kolinski, A., and Yaris, R. (1990). *Comments Mol. Cell. Biophys.* **6**, 223–248.
- Skulachev, V. P. (1988). "Membrane Energetics." Springer-Verlag, Berlin.
- Smith, D. F., Stensgard, B. A., Welch, W., and Toft, D. O. (1992). *J. Biol. Chem.* **267**, 1350–1356.
- Smith, D. H., and King, J. (1981). *J. Mol. Biol.* **145**, 653–676.
- Sosnik, T. R., and Trehewella, J. (1992). *Biochemistry* **31**, 8329–8335.
- Sosnik, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994). *Nat. Struct. Biol.* **1**, 149–156.
- Spolar, R. S., Ha, J. H., and Record, M. T., Jr. (1989). *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8382–8385.

- Sternlicht, H., Farr, G. W., Sternlicht, M. L., Driscoll, J. K., Willison, K., and Yaffe, M. B. (1993). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9422–9426.
- Stigter, P., Alonso, D. O. V., and Dill, K. A. (1991). *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4176–4180.
- Stockman, B. J., Euvrard, A., and Scabill, T. A. (1993). *J. Biomol. NMR* **3**, 285–296.
- Stockmayer, W. H., and Fixman, M. (1963). *J. Polymer Part Sci. C* **1**, 137–141.
- Stryer, L. (1965). *J. Mol. Biol.* **13**, 482–495.
- Sugawara, T., Kuwajima, K., and Sugai, S. (1991). *Biochemistry* **30**, 2698–2706.
- Sumner, S. C. J., Gallagher, K. S., Davis, D. G., Covell, D. G., Jernigan, R. L., and Ferretti, J. A. (1990). *J. Biomol. Struct. Dyn.* **8**, 687–707.
- Sumner, S. C. J., Jiang, S.-P., Jernigan, R. L., and Ferretti, J. A. (1992). *J. Biomol. Struct. Dyn.* **10**, 429–439.
- Tai, P.-K. K., Albers, M. W., Chang, H., Faber, L. E., and Schreiber, S. L. (1992). *Science* **256**, 1315–1318.
- Tai, P.-K. K., Chang, H., Albers, M. W., Schreiber, S. L., Toft, D. O., and Faber, L. E. (1993). *Biochemistry* **32**, 8842–8847.
- Taketomi, H., Ueda, Y., and Gö, N. (1975). *Int. J. Pept. Protein Res.* **7**, 445–459.
- Tanford, C. (1968). *Adv. Protein Chem.* **23**, 121–282.
- Tanford, C. (1970). *Adv. Protein Chem.* **24**, 1–95.
- Tiktopulo, E. I., and Privalov, P. L. (1978). *FEBS Lett.* **91**, 57–58.
- Tiktopulo, E. I., Bychkova, V. E., Rička, J., and Ptitsyn, O. B. (1994). *Macromolecules* **27**, 2879–2882.
- Tiktopulo, E. I., Uversky, V. N., Lushchik, V. B., Klenin, S. I., Bychkova, V. E., and Ptitsyn, O. B. (1995). *Macromolecules*. In press.
- Tsong, T. Y., Baldwin, R. L., and McPhie, P. (1972). *J. Mol. Biol.* **63**, 453–469.
- Tsürûdome, M., Gluck, R., Graf, R., Falchetto, R., Schaller, U., and Brunner, J. (1992). *J. Biol. Chem.* **267**, 20225–20232.
- Touchette, N. A., Perry, K. M., and Matthews, C. R. (1986). *Biochemistry* **25**, 5445–5452.
- Turner, D. C., and Brand, L. (1968). *Biochemistry* **7**, 3381–3390.
- Udgaoonkar, J. B., and Baldwin, R. L. (1988). *Nature (London)* **335**, 694–699.
- Udgaoonkar, J. B., and Baldwin, R. L. (1990). *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8197–8201.
- Ursic, D., and Culbertson, M. R. (1991). *Mol. Cell. Biol.* **11**, 2629–2640.
- Uversky, V. N. (1993). *Biochemistry* **32**, 13288–13298.
- Uversky, V. N., and Ptitsyn, O. B. (1994). *Biochemistry* **33**, 2782–2791.
- Uversky, V. N., and Ptitsyn, O. B. (1995). *J. Mol. Biol.*
- Uversky, V. N., Semisotnov, G. V., Pain, R. H., and Ptitsyn, O. B. (1992). *FEBS Lett.* **314**, 89–92.
- van der Goot, F. G., Gonzalez-Mañas, J. M., Lakey, J. H., and Pattus, F. (1991). *Nature (London)* **354**, 408–410.
- van der Goot, F. G., Lakey, J. H., and Pattus, F. (1992). *Trends Cell Biol.* **2**, 343–348.
- van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H., and Jaenicke, R. (1992). *Biochemistry* **31**, 3635–3644.
- Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., and Clore, G. M. (1993). *Science* **260**, 1110–1113.
- Varshavsky, A., Bachmair, A., Finley, D., Gonda, D., and Wünning, I. (1988). In "Ubiquitin" (M. Rechsteiner, ed.), pp. 287–324. Plenum, New York.
- Venetianer, P., and Straub, F. B. (1963). *Biochim. Biophys. Acta* **67**, 166–168.
- Venyaminov, S. Yu., Gudkov, A. T., Gogia, Z. V., and Tumanova, L. G. (1981). "Absorption and Circular Dichroism Spectra of Individual Proteins from *E. Coli* Ribosomes." Communication of Biological Research Center, Institute of Protein Research, Academy of Sciences of the USSR, Pushchino.

- Vestweber, D., and Schatz, G. (1988). *EMBO J.* **7**, 1147–1151.
- Viitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992). *Protein Sci.* **1**, 361–369.
- Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., and Cook, W. J. (1987). *J. Mol. Biol.* **194**, 531–534.
- Walther, J. P., Feher, V. A., Merutka, G., Dyson, H. J., and Wright, P. E. (1993). *Biochemistry* **32**, 6337–6355.
- Weiss, J. B., Ray, P. H., and Bassford, P. J., Jr. (1988). *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8978–8992.
- Weiss, M. A., Hua, Q.-X., Lynch, C. L., Frank, B. H., and Shoelson, S. E. (1991). *Biochemistry* **30**, 7379–7389.
- Weissman, J. S., and Kim, P. S. (1991). *Science* **253**, 1386–1393.
- Weissman, J. S., and Kim, P. S. (1992). *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9900–9904.
- Weissman, J. S., and Kim, P. S. (1993). *Nature (London)* **365**, 185–188.
- Wells, T. N. C., and Fersht, A. R. (1986). *Biochemistry* **25**, 1881–1886.
- Wetlaufer, D. B. (1973). *Proc. Natl. Acad. Sci. U.S.A.* **70**, 697–701.
- White, J. M. (1992). *Science* **258**, 917–924.
- Wilkinson, K. D., and Mayer, A. N. (1986). *Arch. Biochem. Biophys.* **250**, 390–399.
- Withka, J., Moncuse, R., Baziotis, A., and Maskiewicz, R. (1987). *J. Chromatogr.* **398**, 175–202.
- Wong, K.-P., and Hamlin, L. M. (1974). *Biochemistry* **13**, 2678–2683.
- Wong, K.-P., and Tanford, C. (1973). *J. Biol. Chem.* **248**, 8518–8523.
- Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988). *Biochemistry* **27**, 7167–7175.
- Wright, P. E., Dyson, H. J., Feher, V. A., Tennant, L. L., Walther, J. P., Lerner, R. A., and Case, D. A. (1990). *UCLA Symp. Mol. Biol. New Ser.* **109**, 1–13.
- Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J., and Roder, H. (1993). *Biochemistry* **32**, 10271–10276.
- Wüthrich, K. (1986). "NMR of Proteins and Nucleic Acids." Wiley, New York.
- Xie, D., Bhakuni, V., and Freire, E. (1991). *Biochemistry* **30**, 10673–10678.
- Yang, Y., Janich, S., Cohn, J. A., and Wilson, I. M. (1993). *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9480–9484.
- Yutani, K., Ogasahara, K., and Kuwajima, K. (1992). *J. Mol. Biol.* **228**, 347–350.
- Zhao, J.-M., and London, E. (1988). *J. Biol. Chem.* **263**, 15369–15377.
- Zhuang, Z., and McCauley, R. B. (1989). *J. Biol. Chem.* **264**, 14594–14596.
- Zhuang, Z., Marks, B., and McCauley, R. B. (1992). *J. Biol. Chem.* **267**, 591–596.
- Zimm, B. H., and Bragg, J. R. (1959). *J. Chem. Phys.* **31**, 526–535.