Available online at www.sciencedirect.com
ScienceDirect
BioSystems 93 (2008) 78–89
Bio Systems
www.elsevier.com/locate/biosystems

# Hydrophobic condensation and modular assembly model of protein folding

Tian-Yow Tsong [a,b,*], Chin-Kun Hu [c,d], Ming-Chya Wu [c,e,f]

[a] College of Biological Sciences, University of Minnesota, Minneapolis, MN 55455, USA
[b] National Chiao-Tung University, Hsinchu 30010, Taiwan
[c] Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
[d] Center for Nonlinear and Complex Systems, and Department of Physics, Chung Yuan Christian University, Chungli 32023, Taiwan
[e] Research Center for Adaptive Data Analysis, National Central University, Chungli 32001, Taiwan
[f] Department of Physics, National Central University, Chungli 32001, Taiwan

Received 28 January 2008; received in revised form 1 April 2008; accepted 7 April 2008

**Abstract**

Despite several decades of intense study, protein folding problem remains elusive. In this paper, we review current knowledge and the prevailing thinking in the field, and summarize our work on the *in vitro* folding of a typical small globular protein, staphylococcal nuclease (SNase). Various thermodynamic and kinetic methods have been employed to determine the energetic and construct the energy landscape of folding. Data presented include, but not limit to, the identification of intermediate states, time courses of their spread and convergence on the landscape, and finally the often ignored step, the refinement of the overall conformation and hence the activation of the enzyme. Our goal is to have a complete perspective of the folding process starting from its initial unfolded state to the fully active native state. Analysis leads to these findings: the folding starts with the condensation of the hydrophobic side chains in different locales of the peptide chain. The newly forged hydrophobic environment facilitates formation of helix- and sheet-like frameworks at different domains. Consolidation and inter-docking of these frameworks or domains then stabilizes the overall conformation and refines the structure to activate the enzyme. Based on these observations we favor folding-by-parts and propose a modular assembly model for the *in vitro* folding of SNase.


*PACS:* 87.14.Ee; 87.15.Cc; 87.15. −v

*Keywords:* Hydrophobic condensation; Modular assembly model; Staphylococcal nuclease; Protein folding

## 1. Introduction

Proteins are biopolymers displaying soft characteristics; their conformations fluctuate in solution and spread in a wide range of conformational space. From chemical point of view, a protein is a linear, heterogeneous polymer assembled from 20 different amino acid residues linked by covalent peptide bonds into the polypeptide chain that adopts unique native three-dimensional structures allowing it to carry out its intricate biological functions.

Proteins in cell are synthesized in ribosomes from the genetic information encoded in the cellular DNA. Protein folding *in vivo* is in some cases co-translational. It is initiated before the completion of protein synthesis while the nascent chain is still attached to the ribosome. Other proteins undergo the major part of their folding in the cytoplasm after their release from the ribosome. Others fold in specific compartments, such as mitochondria or the endoplasmic reticulum after translocation through membranes. Details of the folding process depend on the particular environment in which folding takes place. The fundamental principles of folding are presumably universal.

According to the thermodynamic hypothesis of Anfinsen (1972), protein folding is a process driven by the free energy of stabilization when a peptide chain in a random configuration folds into a compact three-dimensional structure. In other words, the structure of a protein is thermodynamically determined (Anifnsen, 1973; Tanford, 1968; Jaenicke, 1980). A typical protein of molecular weight 20 kDa (Dalton) consists of over thousand of atoms, and the intra-molecular interaction energy for the hydrogen bonding, electrostatic interaction, hydrophobic effects, van der Waals' interaction, etc. may amount to hundreds

* Corresponding author at: College of Biological Sciences, University of Minnesota, Minneapolis, MN 55455, USA.
*E-mail addresses:* tsong001@umn.edu (T.-Y. Tsong), huck@phys.sinica.edu.tw (C.-K. Hu), mcwu@ncu.edu.tw, mcwu@phys.sinica.edu.tw (M.-C. Wu).

0303-2647/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2008.04.007

of $k_BT$ for each category. Proteins are in their native states in aqueous solvent near neutral pH at 20–40 °C, which is the typical cellular environment, and they denature or unfold and lose their functionality by exposure to certain chemical agents or to high or low temperatures. To fulfill specific functions, proteins must have a well-defined three-dimensional structure. The stability must typically exceed the energy of thermal motion, $RT \approx 2.5\,\mathrm{kJ\,mol^{-1}}$ with the gas constant $R$ and the absolute temperature $T$. This level of stability can be achieved only by integration of a large number of elements into a domain, which is defined as the single cooperative structural unit. This requirement determines the lower size of a folding domain to about 5 kDa, and the upper limit to about 30 kDa (Privalov, 1979, 1982). Consequently, protein molecules may have one or more functional domains.

In this paper, we review current knowledge and the prevailing thinking in the field, and summarize our work on the *in vitro* folding of a typical small globular protein, staphylococcal nuclease (SNase). Various thermodynamic and kinetic methods have been employed to determine the energetic and construct the energy landscape of folding.

This paper is organized as follow. In next section, we shortly review dominant forces in protein folding, the hydrophobic interaction and conformation entropy. The measurement of the conformational energy is presented in Section 3. The effect of local hydrophobicity is discussed in Section 4, and the hydrophobic condensation and modular assembly model is presented with the discussions of other contested protein folding model in Section 5. We consider a case study of SNase in Section 6, in which a series experimental studies and theoretical investigations are summarized, and from which a folding model of the protein is proposed. Finally, we conclude briefly in Section 7.

## 2. Dominant Forces in Protein Folding

### 2.1. Hydrophobic Interaction

The protein folding problem is to understand and predict the compact native structures from knowledge of amino acid sequence. One step to tackle this problem has been the unravelling of the various forces involved in the folding. Go and Taketomi (1978a) showed, by simulations on lattice model, that lattice protein molecule is stabilized by specific long-range (non-local) and short-range (local) interactions. The former is essential for highly cooperative stabilization of the native conformation and latter accelerate the folding and unfolding transition. Local interactions, such as hydrogen bonding, have been considered as essential forces in folding. It is the major component for the formation of the secondary structures. However, the information needed to specify the conformation of an $n$-atom segment is proportional to $n \times (x_i, y_i, z_i)$, where $x_i, y_i, z_i$ are coordinates of the $i$th atom ($i \in n$). Secondary-structure propensities cannot encode this much information. This implies that non-local interactions must be involved. Kauzmann (1959) argues that a strong force for folding proteins is the tendency of non-polar amino acids to associate in water, instead of hydrogen bonding. Today, a common view is that there is no single dominant force in folding (Dill, 1990). However, it is generally accepted that the principle forces of protein folding are hydrophobicity (HP) and conformational entropy (Chan and Dill, 1990). HP is an important driving force in the process of polymer folding and protein stabilization (Dill, 1990; Privalov and Gill, 1988; Southhall et al., 2002). It has been used to investigate protein stability (Kanehisa and Tsong, 1980; Kiefhaber and Baldwin, 1995), and the HP plot (Kyte and Doolittle, 1982) with a fixed window size is useful for the identification of protein surface-exposed regions. It can also be used to predict chain turns in globular proteins (Rose, 1978).

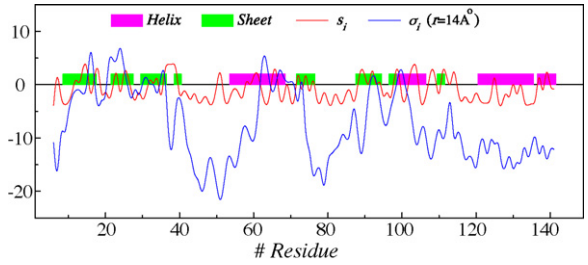


Fig. 1. HP ($s_i$) and local hydrophobicity (LHP, $\sigma_i$) as a function of residues for SNase (1EY0) estimated by the Kyte–Doolittle scale (Kyte and Doolittle, 1982). Adapted from Ref. Wu et al. (preprint). See Fig. 3 for the definition of LHP.

HP is defined as the free energy of transfer from water to a non-polar liquid (Kyte and Doolittle, 1982; Creighton, 1993). The ratio of the saturated concentrations $[X]$ of the molecule in the non-polar liquid and in the water at equilibrium gives the partition coefficient $K$ (Creighton, 1993):

$$K = \frac{[X]_{\text{water}}}{[X]_{\text{non-polar liquid}}}. \tag{1}$$

The free energy transferred from the water to the non-polar liquid is given by $\Delta G = -RT \ln K$, which is a measure of HP of a molecule.[1] The more hydrophobic molecules have the more positive HPs. Fig. 1 shows HP and local hydrophobicity (LHP) to be defined below as a function of residues for SNase estimated by the Kyte–Doolittle scale (Kyte and Doolittle, 1982).

In the simplified H–P model, in which amino acids are classified as hydrophobic monomers (denoted by H) and polar or charged monomers (denoted by P), the folding code is binary and delocalized throughout the amino acid sequence. This simple model can account for the properties that characterize protein folding: two-state cooperativity, secondary and tertiary structures, and multistage folding kinetics (Dill et al., 1995). Sun et al. (1995) showed, by inverse folding algorithms with the aim of designing amino acid sequences which fold to desired target structures, that a designed sequence can fold to have good hydrophobic cores and generate unique folders in the lattice model tests. Under the influence of hydrophobic interaction, the resulting compromise allows hydrophilic side-chains access to the aqueous solvent while at the same time minimizing contact between hydrophobic side-chains and water. From this aspect, one can consider that water triggers the folding process.

[1] Hydrophobicity by definition has a unit of kcal/mol. For simplicity, we neglect the unit throughout the paper.

It is remarkable to note that HP characterized by the free energy alone masks the underlying physics, but it fails to provide predictive power for more complex situations, such as size, shape, and multiple-body effects (Silberstein et al., 1998). HP is seen as "nonspecific", and hydrogen bonding and helical propensities are seen as the "specific" components of the folding code that directs a protein to fold to its unique native structure. A possible process to combine two components in a folding model is the hydrophobic collapse to be discussed below.

### 2.2. Conformational Entropy

While protein stability is sensitive to the details of the inter-atomic interactions, folding mechanisms appear to depend more on the low resolution geometrical properties of the native state (Baker, 2000). Recent investigations indicate that protein-folding rates and mechanisms are largely determined by a protein's topology represented by contact order, rather than by its inter-atomic interactions (Baker, 2000; Alm and Baker, 1999; Fersht, 1999). Thus the structures of folding transition states are similar in proteins with similar native structures.

In folding process, a protein must cross over a higher-free-energy barrier of a transition state. In the unfolded state the protein can take up any one of many conformations, whereas in the native state it has only one or a few distinct conformations. The flexibility of residue side chains is reduced upon helix formation, because the bulky helix backbone is sterically incompatible with some side-chain conformers. Such restriction is energetically disfavored and is expressed as a loss in conformational entropy (Baldwin and Rose, 1999). Therefore, unfolding a structure should result in an increase of entropy.

Understanding of the enthalpic and entropic driving forces of this process requires measurement of the difference between a protein's native state and a fully unfolded, solvent-exposed reference state. Understanding the mechanism or pathway of folding, however, requires knowledge of transient kinetic intermediates, which in some cases may be the same as intermediate states present at equilibrium. Differential scanning calorimetry (DSC) can be used directly to measure the thermodynamic characteristics of such equilibrium intermediate states (Carra and Privalov, 1996).

## 3. Conformational Energy of a Protein

Conformational transition of a protein can be characterized by its free energy change and varying degree of cooperativity. The upper limit of cooperativity corresponds to the purely two-state process $A \rightleftharpoons B$, where the cooperative unit coincides with the entire protein (Lumry et al., 1966).

A very useful criterion for the two-state behavior is afforded by calorimetric measurements. Tsong et al. (1970) proposed that the value for the enthalpy change per mole evaluated by the van't Hoff equation, $\Delta H_{\mathrm{vH}}$, under a given set of conditions is equal to the calorimetric value, $\Delta H_{\mathrm{cal}}$, observed under the same conditions if the transition is of the two-state type; otherwise the inequality $\Delta H_{\mathrm{vH}} < \Delta H_{\mathrm{cal}}$ holds. For a transition with significantly populated and equally spaced intermediate states between the initial and the final states, the general relation reads as (Tsong et al., 1970)

$$\Delta H_{\mathrm{vH}} = \frac{n+1}{3}\Delta H_{\mathrm{cal}}, \tag{2}$$

where $n$ is the number of intermediate states. Here, the equilibrium data required for the estimation of $\Delta H_{\mathrm{vH}}$ are best obtained from the calorimetric observations themselves (Tsong et al., 1970).

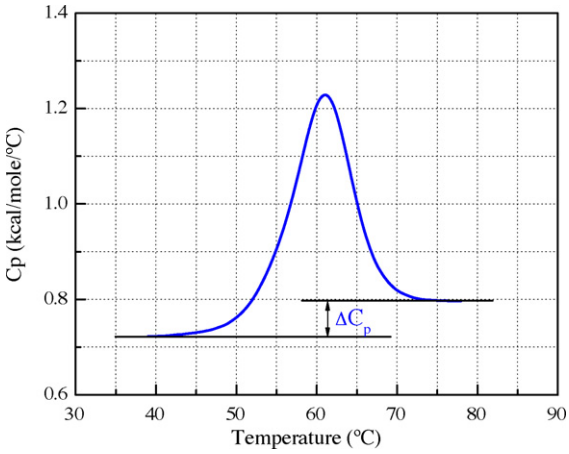


Fig. 2. Typical calorimetric measurement of thermal unfolding of a protein.

The enthalpy $\Delta H_{\mathrm{cal}}$ is calculated from heat capacity $C_{\mathrm{p}}$, which is the energy required to raise the temperature at constant pressure. Fig. 2 shows a typical $C_{\mathrm{p}}$ curve of a protein in solution. The contribution of the protein, its partial $C_{\mathrm{p}}$, is determined by subtracting the corresponding measurements of the aqueous solvent. The partial $C_{\mathrm{p}}$ of the folded protein initially changes slightly as the temperature is increased. Unfolding is apparent as a peak in the heat capacity curve, indicating a large absorption of heat. When unfolding is complete, at higher temperatures, the partial $C_{\mathrm{p}}$ of the protein levels off and is greater than its original value. According to the definition, the partial heat capacity $C_{\mathrm{p}}$ is a measure of the temperature dependence of both the enthalpy and the entropy of the protein in solution, that is,

$$C_{\mathrm{p}} = \frac{\partial H}{\partial T} = \frac{T\partial S}{\partial T}. \tag{3}$$

The partial heat capacities of the pre- and post-transition regions are those of the folded and unfolded states in water, respectively, and give the temperature dependence of their enthalpies and entropies. The area under the transition peak gives the enthalpy change upon unfolding.

However, the heat capacity increment $\Delta C_{\mathrm{p}}$, defined as the difference of $C_{\mathrm{p}}$ before and after transition temperature, cannot be explained solely by an increase of conformational freedom of a polypeptide chain upon unfolding. Calculation shows that such a conformational heat capacity effect of unfolding should be much lower than that observed (Sturtevant, 1977). Transfer of non-polar solutes into water is known to give rise to a significant increase in the heat capacity (Edsall, 1935; Gill et al., 1976). The exposure of internal groups which are essentially non-polar

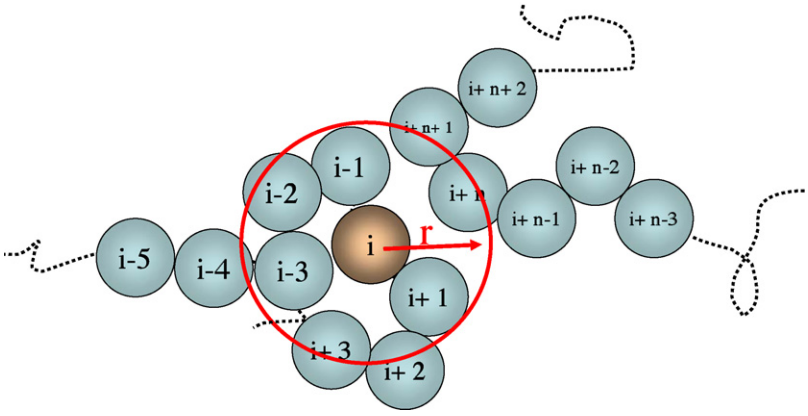


Fig. 3. Schematic drawing for the definition of LHP in this work. One sphere represents one amino acid residue. After Ref. Wu et al. (preprint).

molecules to water in protein unfolding is thus attributed to the increase of the heat capacity. On the other hand, transfer of non-polar molecules into water is associated with a significant decrease of entropy. Hydration effects can then be considered as being responsible for the low solubility of non-polar compounds in water, for their HP (Privalov, 1997).

## 4. Local Hydrophobicity Stabilizes Secondary Structures in Proteins

Kanehisa and Tsong (1980) have defined LHP as the sum of HPs contributed from two nearest neighboring residues on both sides of the linear chain. They have shown that there is correlation between amino acid sequence and segments of secondary structures based on the statistics of 47 globular proteins. The probability of occurrence of helix and β-sheet residues increases as a function of LHP. Thus, LHP stabilizes secondary structures in proteins. This definition of LHP, however, does not consider the HPs from other residues closed to the *i*th residue in its tertiary structure.

We have revised the definition of LHP to include three-dimensional structure information of protein molecules, as shown in Fig. 3 (Wu et al., preprint). The LHP $\sigma_i$ around the *i*th residue is defined as the sum of HP values of residues surrounding the residue within a specified distance $r$, i.e.,

$$\sigma_i(r) = \sum_{\langle i,j \rangle} s_j \, \mathrm{e}^{-r_{ij}/r}, \tag{4}$$

where $\langle i, j \rangle$ indicates that the distance $r_{ij}$ between pair residues $i$ and $j$ is less than $r$ which defines the effective region of LHP. The revised LHP is calculated from protein structure files in the Protein Data Bank (PDB). A typical LHP curve is shown in Fig. 1, in which the Kyte–Doolittle scale (Kyte and Doolittle, 1982) has been used. Following the work of Kanehisa and Tsong (1980) and the hydrophobic scale (Trp, 3.4; Phe, 2.5; Tyr, 2.3; Leu, 1.8; Ile, 1.8; Met, 1.3; Cys, 1.0; His, 0.5; Ala, 0.5; Thr, 0.4; and 0 for the rest of amino acids) used by them, the probability density of occurrence of helix and β-sheet residues as functions of LHP can be estimated by an ensemble statistics of PDB protein structures. Fig. 4 shows that the statistics based on 655 PDB entries selected by the protein sequence culling server (Wang and Dunbrack, 2003). The probability of occurrence of 20 amino acids in proteins is inhomogeneous, while statistics of LHP of a residue in helix and sheet is the normal distribution. Gly, Ala and Glu have high probabilities in helices, Val and Ile have higher probabilities in sheets, whereas Leu has a high probability in both structures. From the HP point of view, residues with higher HP, such as Leu, Val, and Ile, in higher LHP environment are more likely in sheet. On the other hand, the non-polar residue, Ala, in negative LHP is likely in helices. By dividing LHP into 6 class (I: 0–4, II: 4–8, III: 8–12, IV: 12–16, V: 16–20, and VI: $>$ 20) and perform the statistics on the occurrence of 20 amino acid residues with LHP values in these class, we plot the relative probability as function of LHP in Fig. 4(c). The curves show that amino acid residues in helices and sheets are likely to have higher LHP, consistent with the results of Kanehisa and Tsong (1980).

Simple statistics reveals the correlation between LHP and HP of amino acid residues in specific secondary structures. This implies that LHP is an important factor for the stabilization of the secondary structures in proteins. Nevertheless, it is also apparent that such statistics on a global scale alone is insufficient to predict secondary structures from the amino acid sequence. In contrast to this global property, it is possible to establish correlation between LHP and HP and secondary structures for a specific protein. By considering an amino acid sequence as a time series, Wu et al. (preprint) compared the decomposition of HP and LHP in different scales with locations of specified secondary structures, and found that folding information of HP is imprinted in amino acid sequences and propagated through LHP in folding processes. This finding suggests that if a polypeptide chain folds to a compact dimension in an early stage, local environments shall lead to the formation of local structures. Subsequent folding processes do not significantly alter the environments. This picture is consistent with the hydrophobic collapse model. Furthermore, LHP has been used to predict secondary structures contents in mutants from wild-type (WT) proteins (Wu et al., preprint). We will discuss this in our later case study of SNase.

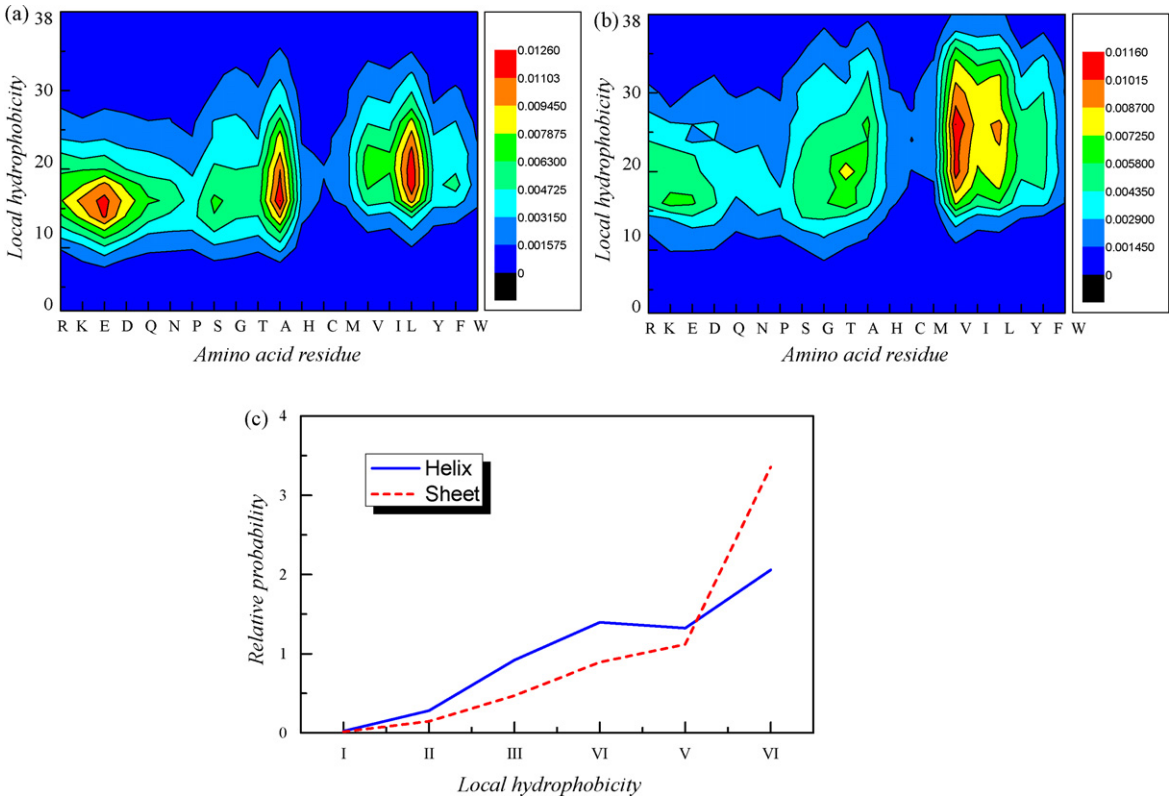


Fig. 4. The probability density of occurrence of (a) helix and (b) β-sheet residues as functions of LHP defined by Eq. (4). (c) The average of 20 amino acids for the helix and sheet residue probabilities.

## 5. Hydrophobic Condensation and Modular Assembly in Protein Folding

Fig. 5 shows schematic representation of some contested protein folding models, the framework model (Kim and Baldwin, 1982), the hydrophobic collapse model (Kanehisa and Tsong, 1978), and the hybrid model (Ptitsyn, 1987).

In the framework model, folding is driven by local interactions (Kim and Baldwin, 1982), and implies that isolated helices form early in the protein folding pathway and then assemble into the native tertiary structure (Dill et al., 1995).

The hydrophobic collapse model also known as the cluster model (Kanehisa and Tsong, 1978) postulates that folding is driven by the collapse of the local hydrophobic side chains which concurrently facilitate the formation of folding units or structural domains. The subsequent folding process is the modular assembly of these structural domains. The cluster model is energetically consistent with the two-state model (Kanehisa and Tsong, 1978). However, the folding kinetics is hierarchic (Baldwin and Rose, 1999). Hierarchic folding is defined as a process in which folding begins with structures that are local in sequence and marginal in stability; these local structures interact to produce intermediates of ever-increasing complexity and grow, ultimately into the native conformation (Baldwin and Rose, 1999). Multiple folding routes may co-exist in hierarchic folding process, and the stabilities of the intermediates and their combinatorial associations will determine the dominant pathways (Baldwin and Rose, 1999; Rose, 1979).

In the hybrid model, partially folded intermediates are formed in the early folding stage. It is in compact globular shape with a high degree of secondary structure and few specific tertiary interactions (for review, see Refs. (Kuwajima, 1989; Dobson, 1992; Ptitsyn, 1992)). The main difference between the hydrophobic collapse model and the hybrid model is that in the former secondary structures form after the hydrophobic collapse, while in the latter both hydrophobic clusters and secondary structures form simultaneously. These folding intermediates are reminiscent of the molten globules postulated by many authors.

Our work puts special emphasis on the hydrophobic collapse model. Historically, Go (Go, 1975; Go and Taketomi, 1978b,c) has first approximated the protein structure as one globular part with one or more random coil parts and discussed the overall feature of the protein denaturation. Kanehisa and Tsong have extended the concept to construct the cluster model (Kanehisa and Tsong, 1978, 1979); local structures or clusters are formed at an early stage in different regions of the polypeptide chain. The model implies that non-local interactions driving collapse processes in heteropolymers can give rise to protein structure, stability, and folding kinetics. It also implies that hydrophobic collapse facilitates secondary structure formation, rather than the reverse. Proteins are special among polymers not primarily because of the 20 types of their monomers, but because the amino acids in proteins are linked in specific sequences. Folding code resides mainly in global patterns of contact interactions, which are non-local, and arise from the arrangements of polar and non-polar monomers in the sequence (Dill et al., 1995).

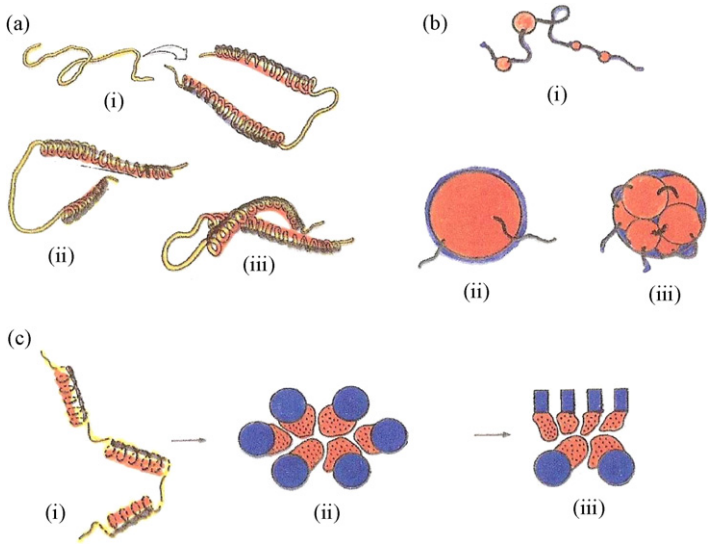


Fig. 5. Schematic representation of folding models: (a) the framework model (adapted from Ref. Kim and Baldwin (1982)); (b) the hydrophobic collapse model (adapted from Ref. Kanehisa and Tsong (1978)); (c) the hybrid model (Ptitsyn, 1987). Experimental test of these models is straightforward. In model (a) formation of secondary structures should precede formation of hydrophobic clusters. In model (b) the reverse is true. And in model (c) the two processes should parallel each other.

Rose and Roy (1980) showed that linear chain regions rich in hydrophobic residues serve as small clusters that fold against each other, with concomitant or even later fixation of secondary structure. A helix or strand would arise in this process as one of a few energetically favorable alternatives for a give cluster, followed by a shift in the equilibrium between secondary structure conformers upon cluster association (Rose and Roy, 1980).

## 6. A Case Study: Staphylococcal Nuclease Folding

In this section, we present a case study of SNase. We will review a series of our equilibrium and kinetic studies of SNase and its mutants in the past decade and from which infer the folding model of the protein.

### 6.1. Stepwise Folding of SNase

SNase can be denatured by acid and denaturants, such as guanidine HCl (GdmCl) (Chen et al., 1991, 1992a, b; Su et al., 1996), or by elevating temperature beyond 65 °C (Su et al., 1996). Acid denatured protein retains some sheet-like circular dichroism (CD) spectrum. The complete loss of secondary structure is afforded by exposing the protein to 1.5 M or higher, of GdmCl (Su et al., 1996). Analysis of the equilibrium unfolding indicates that $\Delta H_{vH} = \Delta H_{cal}$, suggesting that SNase is a two-state folder and has an unit cooperativity. No presence of any folding intermediates is detectable. However, when kinetics of folding from these denatured and unfolded states is monitored, complex time courses are detected. The complex time courses means the existence of folding intermediates. These seemingly contradictory observations require energetic analysis of the stepwise folding reactions. We used folding from GdmCl unfolded protein as an example to describe the complete folding/unfolding reactions.

### 6.2. Nucleation Events

Fig. 6 gives kinetic records of GdmCl induced folding/unfolding of SNase. Fig. 6(a) illustrates the complete signal change, measured with the CD changes at 230 nm. The stopped-flow instrument employed has a mixing time of 0.2 ms. Reactions faster than 0.2 ms cannot be resolved by this equipment. Nevertheless, the missing signal can easily be estimated by looking at the total signal changes. Fig. 6(b) indicates that 10–15% of the protein secondary structures is lost within 0.2 ms. Conversely, Fig. 6(c) indicates 10–15% gain of the secondary structure in the unfolding in the first 0.2 ms. Since the helical content of SNase is only 25%, the reactions faster than 0.2 ms account for only 2–3% of the entire peptide, or the equivalent of a few amino acid residues in SNase. This fast initial phase was interpreted as the *nucleation events* in the folding and the backbone adjustment in the unfolding.

### 6.3. Hydrophobic Condensation and the Subsequent Growth of Secondary Structures

Rapid formation of hydrophobic cluster is monitored by the appearance of anilinyl-naphthalene-sulfonate (ANS) binding sites. ANS is a fluorescence probe for the polarity of its binding sites. When binding site is non-polar the fluorescence intensity of bound ANS increases by 50 folds, as compared to its binding to a polar binding site. Fig. 7 shows that with this probe we have found that during SNase folding, ANS binding sites accumulate within 20 ms in different locales of the peptide chain. Surprisingly, the bound ANS is subsequently sequestered from these binding sites in a complex reaction in 1–5 s. This latter reaction is also accompanied by a parallel accumulation of the protein secondary structures (Fig. 6(b)). During this kinetic phase the newly forged secondary structures consolidate and the internal structures of protein start to refine and larger structure domains

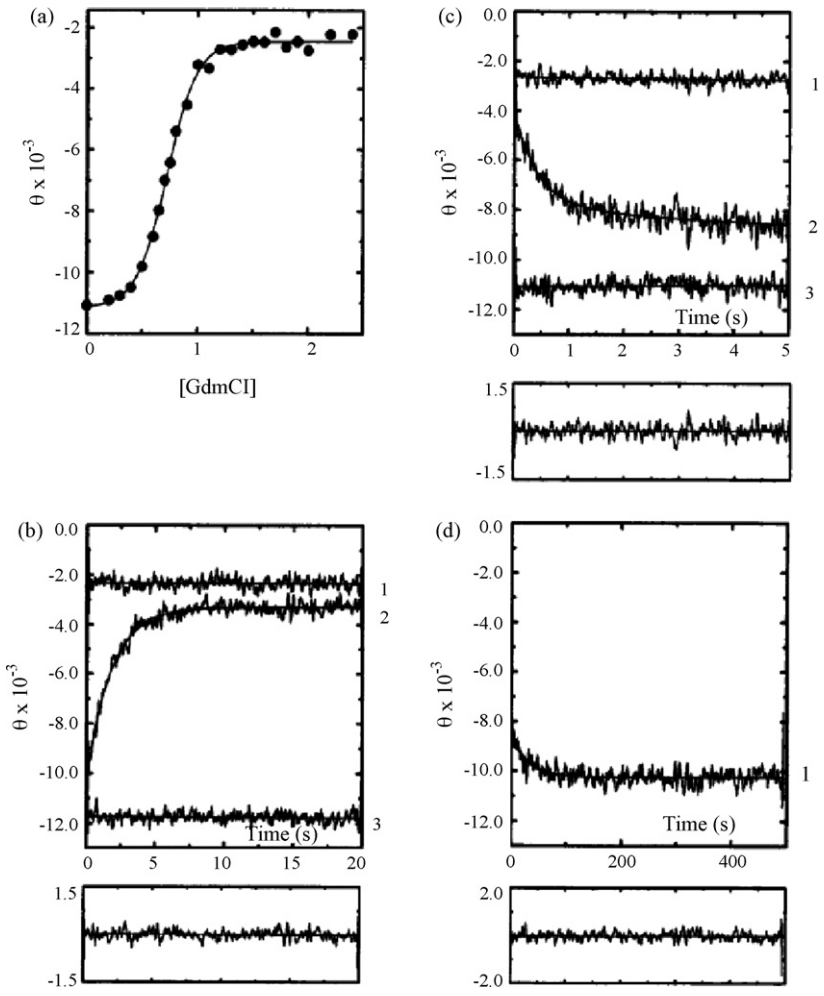


Fig. 6. Kinetic records of GdmCl-induced folding/unfolding transition. (a) Equilibrium unfolding curve measured directly inside the observation chamber of the stopped-flow instrument. The midpoint of transition, Cm, was 0.75 M. (b) Time course of the unfolding of SNase at 2.0 M GdmCl. Curve 1 is the equilibrium position of the unfolded state, curve 2 the time-resolved signals, and curve 3 the equilibrium position of the initial state, $N_0$. Curve 2 was fit into a single exponential decay process, with a relaxation time of 1.77 s. The line drawn through a kinetic record is the non-linear least-squares fit of the experimental curve. Lower panel shows residues of the fit. (c) Time course for folding to 0.34 M GdmCl, kinetic measurements in the range 20 ms to 5 s were fit into a two-relaxation event (429 ms and 6.1 s). (d) An additional slow-folding reaction (relaxation time of 30.6 s) was time-resolved. [$\theta$], Mean residue ellipticity at 222 nm, expressed in deg cm$^2$ dmol$^{-1}$. After Ref. Su et al. (1996).

start to emerge. These domains enriched in secondary structure form the frameworks or the modules for further assembly.

### 6.4. Kinetic Intermediates Versus Thermodynamic Intermediates

As mentioned above, the complex kinetics of folding of SNase points to the presence of several folding intermediates. For the time-resolved kinetics shown in Fig. 6, two folding intermediates can be inferred. As shown, the time course of folding is composed of three kinetic phases, whereas the time course of unfolding is monophasic. After a rigorous analysis of the time courses and amplitudes of kinetic phases, by both the single-jump and the double jump experiments, we proposed a kinetic scheme consistent with all the experimental data (Chen et al., 1991; Su et al., 1996):

$$U_3 \rightleftharpoons U_2 \rightleftharpoons U_1 \rightleftharpoons N_0. \quad (5)$$

According to this scheme the folding is triphasic if the $U_3$ to $U_2$ reaction is slower than the $U_2$ to $U_1$ reaction and this latter reaction is slower than the $U_1$ to $N_0$ reaction. The unfolding on the other hand, is monophasic. It detects only the $N_0$ to $U_1$ reaction. The slower conversions among $U_i$'s are not visible by the optical signal as their CD or fluorescence properties are similar. The kinetic scheme is also equivalent to the $U_i \rightleftharpoons N_0$ reaction. Note that the stabilities of all three $U_i$'s are similar and indistinguishable by thermodynamic analysis.

### 6.5. Compact Dimension of Denature States of SNase

Denatured states describe a major change of a protein from its original native structure. GdmCl-unfolded state is a specific subset of denatured states whose conformations are highly open and solvent-exposed and with little or no residual structures. In contrast the acid-denatured states of globular proteins are usually compact and exhibit secondary structure like CD spectrum.

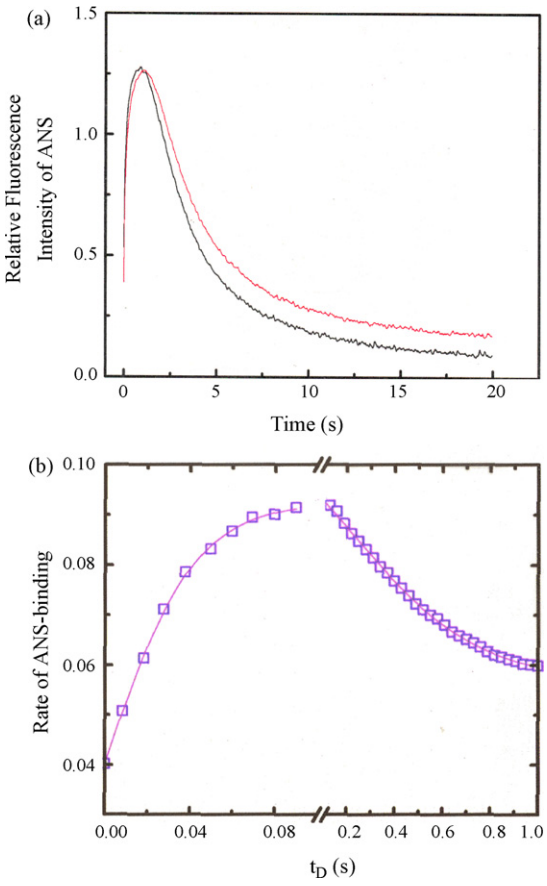


Fig. 7. The typical kinetic traces or time courses of ANS-binding during SNase folding with GdmCl concentration jump at 25 °C. (a) two separate experiments are shown. (b) Rate of ANS-binding as a function of time after protein folding was initiated. Note that when [ANS] is kept constant the rate of ANS binding is proportional to the concentration of ANS binding sites. Both experiments indicate rapid accumulation of ANS binding sites in 20 ms followed by a sequestering of bound ANS and disappearance of the binding sites in 2–5 s. The latter reaction parallels the appearance of the secondary structures shown in Fig. 6.

Flanagan et al. (1992) studied the *truncated* SNase and found it to be compact but disordered, which suggested that extensive solvent exclusion generates a compact polypeptide chain prior to the development of persistent secondary structural features. We used the technique of fluorescence energy transfer to study the folding kinetics of SNase and determine the dimension of the denatured states of SNase by measuring the distance between K45C (energy acceptor) and W140 (donor) (Chow et al., 2008). We have found that the denatured states of SNase are quite compact. Fig. 8 shows the folding/unfolding kinetics of SNase by acid (upper series) and by GdmCl (lower series). The denatured substates have similar enthalpy and tryptophan fluorescence, and their equilibrium cannot be shifted by temperature changes. However, they are kinetically distinctive. The range of distance changes between two probes is between 25.6 and 25.4 Å while it is 20.4 Å for the native state. These distances are within the sphere shown in Fig. 9, indicating that the denatured states of SNase are highly compact regardless of which denatured states (pH-induced or GdmCl-induced) are induced. For the native state of K45, the end-to-end distance between K45 and W140 is 21.46 Å and the mean radius of gyration $\langle R_G \rangle = 8.81$ Å, which is smaller than a random coil conformation ($\langle R_G \rangle = 34$ Å) or a sphere-like structure ($\langle R \rangle = 11$ Å). For the denatured states of K45C, the largest end-to-end distance between K45C and W140 is $r = 25.6$ Å and $\langle R_G \rangle = 10.51$ Å, which is still much smaller than a random coil conformation ($\langle R_G \rangle = 34$ Å) and is similar to a sphere-like structure ($\langle R \rangle = 11$ Å). These results suggest that the dimension of $D_i$ and $U_i$ are compact resembling the dimensions of the native state (Chow et al., 2008).

The compact dimension of SNase implies that hydrophobic collapse of the molecule takes place in the early stage of folding. This conjecture was verified by considering the effects of LHP in the formation of secondary structures (Wu et al., preprint). We have established the correlation coefficient between LHP and secondary structure at each residue from the WT protein structure to estimate the percentage contents of secondary structures from changes of LHPs in mutants (Wu et al., preprint). We have tested WT SNase and a number of mutants with

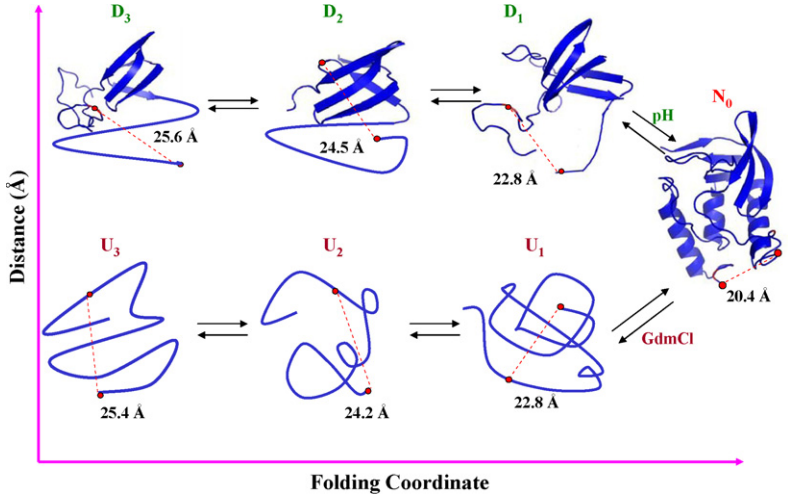


Fig. 8. Folding kinetics of SNase. $N_0$ is the native state, $D_i$'s are pH denatured states, and $U_i$'s are GdmCl denatured states. The dashed line indicates the distance between K45C and W140, determined by the technique of fluorescence energy transfer. After Ref. Chow et al. (2008).

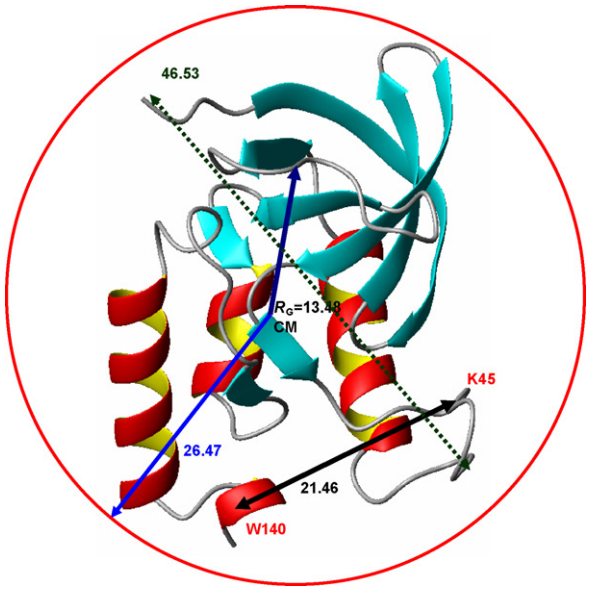


Fig. 9. Some dimensions of the crystal structure of SNase. The distances are in unit of Å.

one-point mutation, W140A (Tryptophan at 140 is replaced by Alanine) and E75G, and two-point mutations, F61W/W140A and Y93W/W140A, and the estimations are qualitatively and quantitatively (within 8% deviation) consistent with those from CD measurements (Wu et al., preprint). The successful estimation from changes of LHP in the mutants provides further evidence that effects of LHP enter into the folding process in early stage of folding and plays an essential role in the formation of secondary structures throughout the whole folding process.

### 6.6. The Docking of Frameworks and Enzyme Activation

SNase is generally described as a single domain globular protein. However, a careful analysis of the packing density reveals that the overall structure of SNase is consisted of two independently folded domains, an N-terminal β-core and a C-terminal domain (Anfinsen, 1972; Griko et al., 1994). Molecular dynamic simulation (Lamy and Smith, 1996) and an NMR study (Shortle and Abeygunawardana, 1993) showed that Δ137–149 retains the N-terminal β-core, but it is largely disordered and fluctuates rapidly between folded and unfolded structure. The formation of the C-terminal cluster also stabilizes the N-terminal domain, while the combination of two domains is an assembly process. The three intermediate states in Eq. (5) in the folding of SNase can be considered as states consisting of the N-terminal cluster and C-terminal cluster forming in the collapsed space.

After folding for 2–5 s, the N-terminal and the C-terminal domains are largely formed. However, these domains lack conformational energy. 95% of the conformational energy of the SNase is acquired at a step where the two domains, or frameworks properly dock to each other in a 20 s reaction. The long-range interaction in the folding of SNase is generated by the C-terminal cluster. It is critical for SNase to fold into its native conformation, while it is considered as a rate-limiting step in SNase folding. Hirano et al. (2005) showed that the information encoded in W140 of SNase is hierarchic and intricate. W140 plays a central role in formation of the C-terminal cluster (Hirano et al., 2005; Hu et al., preprint). The side-chain information encoded by residue 140 is essential to maintain a stable native structure and this residue must be an aromatic side chain.

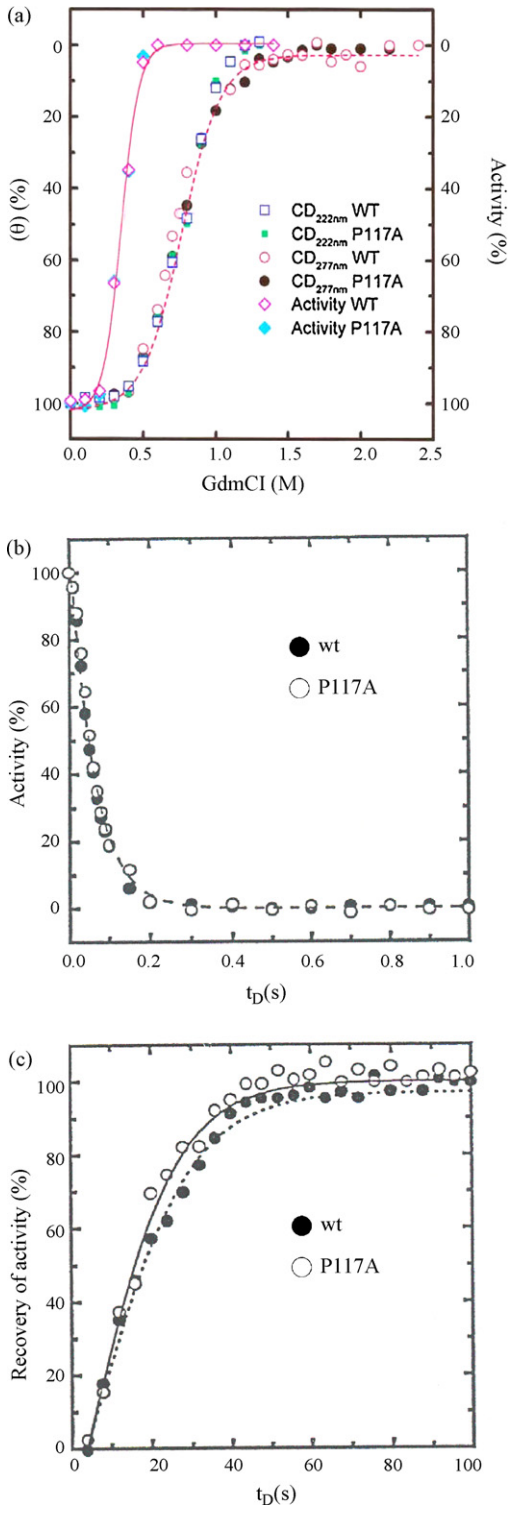


Fig. 10. Enzyme activation/inactivation with its nuclease activity in folding/unfolding transition. (a) Activity and circular dichroism monitoring of the GdmCl induced folding/unfolding of WT and P117A mutant. (b) Unfolding kinetic of active site. (c) Folding kinetics of active sites.

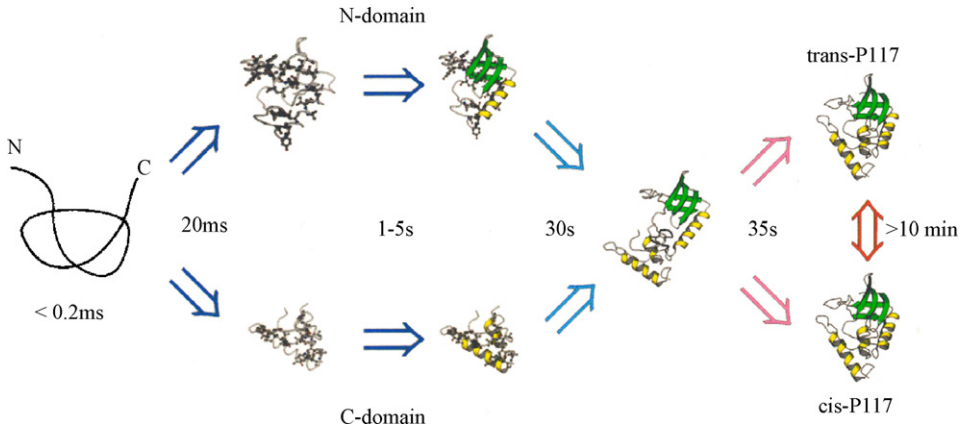


Fig. 11. Folding of SNase by modular assembly. After Ref. Tsong and Chang (2003).

Furthermore, the five-membered nitrogen-containing ring of the indole is thought to bear the essential information for tertiary structure (Hirano et al., 2005).

We further note that the SNase folding is completed once the full activity of the enzyme is acquired. Both (*trans* and *cis*) forms of the enzyme exhibit full nuclease activity. Our data indicates that in GdmCl enzyme activity is lost at 0.4 M while at this concentration of GdmCl 95% of the secondary structures and the enthalpy of folding remain unaffected. Data also suggests that the *cis–trans* isomerization of Pro-117 plays little role in the enzyme activation. Enzyme is fully activated either Pro-117 is in the *cis*- or in the *trans*-isomeric form. The stability of enzyme is approximately $8k_BT$. Almost the entire $8k_BT$ of the conformational stability of the native state is acquired at the last step (30 s) in folding. However, these kinetic steps have activation barriers ranging from $25k_BT$ to $30k_BT$ (Su et al., 1996). Experiments show that the least activation path determines the pathway of protein folding (Su et al., 1996). Multiple pathways will result if the activation barriers of different pathways have similar heights (Tsong and Chang, 2003). These results of enzyme activation are summarized in Fig. 10.

### 6.7. The Modular Assembly Model

Based on above thermodynamic and kinetic studies we are proposing a modular assembly model to account for the stepwise folding pathways of SNase as shown in Fig. 11 (Tsong and Chang, 2003). The folding of SNase starts from presumably an ensemble of conformations with no intramolecular interaction energy. These reactions, taking place between ns and μs, represents events of small activation processes. They give rise to less than 10% of the CD signal. The events that follow are the formation of the hydrophobic clusters, semi-independently, in the N-domain and the C-domain. These events take place in 10–50 ms. The accumulation of helices and sheets inside the newly formed hydrophobic clusters is slow, in 50 ms to 5 s, presumably because of the crowded environment. The interlocking of the N-domain and C-domain then takes place in 30 s. The formation of the two active forms (*trans*-P117 and *cis*-P117) requires further structural fine-tuning, and its takes place in 35 s. The *trans* to *cis* isomerization (Brandts et al., 1975) of proline 117 is a slow equilibrium reaction for SNase in solution (Kaneshia and Tsong, 1979; Tsong and Su, 1999). Finally, the active form assumes the conformation shown in the lower rightmost structure (*cis*-P117).

### 6.8. Rugged Landscape of SNase Folding

The analysis based on the transition-state theory reveals much insight concerning the nature of the folding intermediates. Fig. 12(a) illustrates the three activation barriers that the peptide chain must cross over during folding of GdmCl-unfolded SNase at 25 °C and pH 7.0. Enthalpic and entropic contributions to the activation energy are indicated. Here the activation energy, $\Delta G^{\ddagger}$, and the activation entropy $\Delta S^{\ddagger}$ are calculated according to (Su et al., 1996)

$$k_{obs} = \left(\frac{k_BT}{h}\right) \exp\left[-\frac{\Delta G^{\ddagger}}{RT}\right], \quad (6)$$

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}, \quad (7)$$

where $k_{obs}$ is the observed folding rate, $k_B$ and $h$ are respectively the Boltzmann constant and the Planck constant ($6.25 \times 10^{12}$ $s^{-1}$). The first transition state for the $U_3 \rightarrow U_2$ reaction involves mainly the activation enthalpy, while the second and the third ones encounter mainly the entropic barriers. The large negative activation entropies ($\approx -40$ e.u.) for the $U_2 \rightarrow U_1$ and $U_1 \rightarrow N_0$ reactions point to highly structured transition states for these two reactions. Data for the renaturation of the acid-denatured protein gives a similar result. Since not all of the vast landscape of conformational energy is accessible to a peptide chain (Fig. 12(b)), the peptide does not behave like an ideal random-flight, statistical polymer chain. Random search may occur within the confine of each substate; however, random search amongst "microscopic states" does not define a macroscopic pathway for folding. Furthermore, comparison of the activation processes in folding of SNase from acid and folding from GdmCl indicates that sheet-like chain conformation in the $D_i$'s did not reduce the activation barriers of folding. The energy landscape derived from this study resembles the landscape of the cluster converging kinetics of Kanehisa and Tsong (1979).

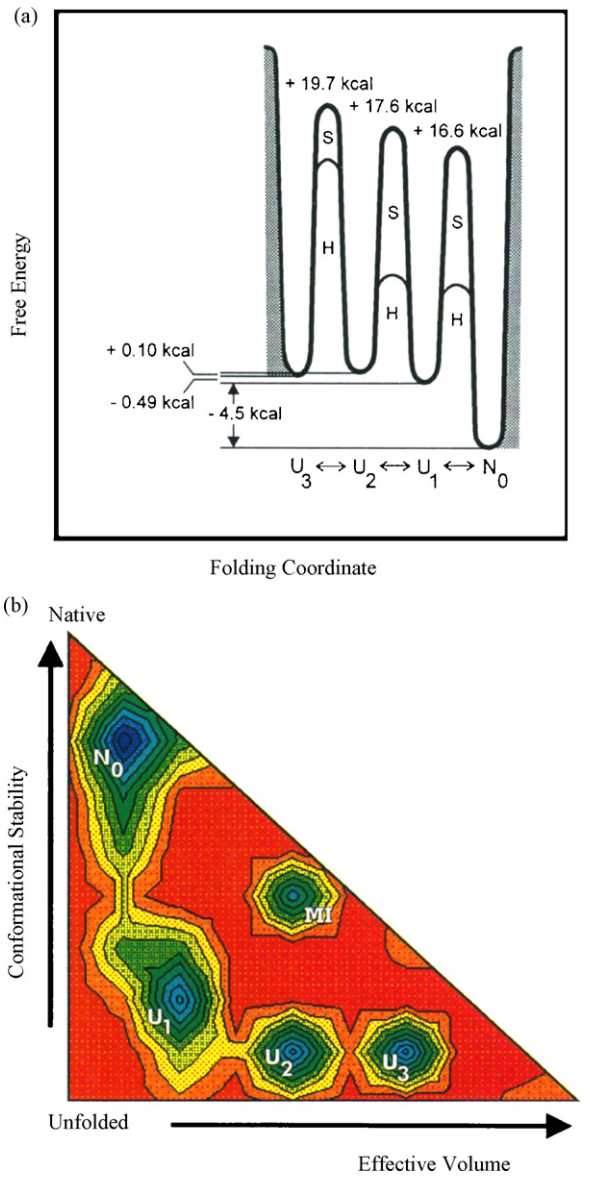


Fig. 12. (a) Activation barriers for the three folding steps of SNase. Activation energy is separated into two contributions, one enthalpic and the other entropic. The scale for free energy is proportional to the numerical values given. (b) Hypothetical energy landscape for kinetic activation during the folding of SNase at 0.34 M GdmCl and 25 °C. Pseudo-colors with the convention of a landscape map were adopted to indicate levels of activation energy. MI means a "metastable intermediate", or a kinetic trap. When a solvent condition permits reversible folding, the peptide chain descends to the global free-energy minimum of the native state via the path that requires the least activation. The *x*-axis denotes the effective volume, or the hydrodynamic volume, and the *y*-axis denotes the conformational stability of the peptide chain. Scales for the *x*- and the *y*-axes are qualitative. After Ref. Su et al. (1996).

### *6.9. Microscopic States Within Each Macroscopic State*

While the experimental data discussed above support the kinetic scheme of Eq. (5), our rigorous analysis of other kinetic experiments (not discussed here) suggests the existence of microscopic states (Chen et al., 1992b). In other words, each of the four states ($N_0$, $U_1$, $U_2$ and $U_3$) is composed of many microscopic states in rapid equilibrium. This observation is consistent with the softmatter characteristic of proteins.

## 7. Conclusions and Perspective

As with the models for protein-folding mechanisms, most of the successful methods attempt to ignore the complex details of the inter-atomic interactions and instead focus on the coarse-grained features of sequence–structure relations. Problems in which the full atomic detail of interactions in the native state is important will almost certainly require considerably more detailed models.

Recently, disordered proteins have attracted much attention. One reason is its importance for the understanding the properties of the misfolded proteins (Mittag and Forman-Kay, 2007). It has been apparent that at the biochemistry level, the stringent quality-control systems will come into play if the folding process fails, ensuring that the misfolded products are targeted for degradation before they may cause irrevocable harm. Those that escape this cellular surveillance are prone to forming aggregate that can damage or kill cells through mechanisms that are just beginning to be understood (Smith, 2003; Dobson, 2003; Sitia and Braakamn, 2003; Goldberg, 2003; Selkoe, 2003; Cohen and Kelly, 2003; Bates, 2006). A huge variety of previously unrelated diseases, such as prion diseases, diabetes and cancer, Huntington's disease, Parkinson's disease, Alzheimer's disease, and amyotrophic lateral sclerosis, share the pathological feature of aggregated misfolded protein deposits. This suggests the exciting possibility that these protein-misfolding diseases are linked by common principles, and may therefore present common targets for therapeutic intervention (Smith, 2003). From the aspect of hydrophobic collapse model, we believe that one important factor of protein misfolding is due to multiple folding pathways in the system where subtle changes in solvent conditions may direct folding reaction to a metastable state or to an astray pathway that triggers a protein filament formation. Studying folding pathways in misfolded proteins with stepwise folding process in our model should be instructive. Our future study will compare folding/unfolding reactions of SNase under quasi-equilibrium conditions in solution with that of non-equilibrium, mechanical stretching by atomic force microscopy (Carrion-Vazquez et al., 1999).

Computer simulation methods have been used widely to explore structures, folding, and unfolding of proteins (Eisenmenger et al., 2001; Hayryan et al., 2001, 2005; Lin et al., 2003; Nguyen et al., 2005; Kouza et al., 2005, 2006, 2008; Busa et al., 2005; Li et al., 2006; Eisenmenger et al., 2006; Li et al., 2007; Ghulghazaryan et al., 2007). It is of interest to apply such methods to study some aspects of hydrophobic condensation and modular assembly model of protein folding presented in the present paper.

## Acknowledgements

This work was supported by the National Science Council of the Republic of China (Taiwan) under Grant Nos. NSC 96-2911-M001-003-MY3 (C.-K. Hu), and NSC 96-2112-M-008-

021-MY3 (M.-C. Wu), Ministry of Education ATU Program (T.-Y. Tsong), National Center for Theoretical Sciences in Taiwan, and by Academia Sinica (Taiwan) under Grant No. AS 95-TP-A07.

## References


Alm, E., Baker, D., 1999. Curr. Opin. Struct. Biol. 9, 189–196.
Anfinsen, C.B., 1972. Biochem. J. 128, 737–749.
Anifnsen, C.B., 1973. Science 181, 223–230.
Baker, D., 2000. Nature 405, 39–42.
Bates, G.P., 2006. Science 311, 1385–1386.
Baldwin, R.L., Rose, G.D., 1999. Trends Biochem. Sci. 24, 26–33.
Brandts, J.R., Halvorson, H.R., Brennan, M., 1975. Biochemistry 14, 4953–5963.
Busa, J., Dzurina, J., Hayryan, E., Hayryan, S., Hu, C.-K., Plavka, J., Pokorny, I., Skrivanek, J., Wu, M.-C., 2005. Comp. Phys. Commun. 165, 59–96.
Carra, J.H., Privalov, P.L., 1996. FASEB J. 10, 67–74.
Carrion-Vazquez, M., Marszalek, P.E., Oberhauser, A.F., Fernandez, J.M., 1999. Proc. Natl. Acad. Sci. U.S.A. 96, 11288–11292.
Chan, H.-S., Dill, K.A., 1990. Proc. Natl. Acad. Sci. U.S.A. 87, 6388–6392.
Chen, H.-M., You, J.-L., Markin, V.S., Tsong, T.-Y., 1991. J. Mol. Biol. 220, 771–778.
Chen, H.-M., Markin, V.S., Tsong, T.-Y., 1992a. Biochemistry 31, 1483–1491.
Chen, H.-M., Markin, V.S., Tsong, T.-Y., 1992b. Biochemistry 31, 12369–12375.
Chow, C.-Y., Wu, M.-C., Fang, H.-J., Hu, C.-K., Chen, H.-M., Tsong, T.-Y., 2008. Compact dimension of denatured states of staphylococcal nuclease. Proteins: Struct. Funct. Bioinf. 72, 901–909.
Cohen, F.E., Kelly, J.W., 2003. Nature 426, 905–909.
Creighton, T.E., 1993. Proteins: Structures and Molecular Properties, 2nd ed. W.H. Freeman and Company, New York, pp. 153–160.
Dill, K.A., 1990. Biochemistry 29, 7133–7155.
Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.-P., Thomas, P.D., Chan, H.-S., 1995. Protein Sci. 4, 561–602.
Dobson, C.M., 2003. Nautre 426, 884–890.
Dobson, C.M., 1992. Curr. Opin. Struct. Biol. 2, 6–12.
Edsall, J.T., 1935. J. Am. Chem. Soc. 57, 1506–1507.
Eisenmenger, F., Hansmann, U.H.E., Hayryan, S., Hu, C.-K., 2001. Comput. Phys. Commun. 138, 192–212.
Eisenmenger, F., Hansmann, U.H.E., Hayryan, S., Hu, C.-K., 2006. Comp. Phys. Commun. 174, 422–429.
Fersht, A., 1999. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. Freeman, New York.
Flanagan, J.M., Kataoka, M., Shortle, D., Engelman, D.M., 1992. Proc. Natl. Acad. Sci. U.S.A. 89, 748–752.
Ghulghazaryan, R.G., Hayryan, S., Hu, C.-K., 2007. J. Comp. Chem. 28, 715–726.
Gill, S.J., Nichols, N.F., Wadso, I., 1976. J. Chem. Thermodyn. 8, 445–452.
Go, N., 1975. Int. J. Peptide Protein Res. 7, 445–459.
Go, N., Taketomi, H., 1978a. Proc. Natl. Acad. Sci. U.S.A. 75, 559–563.
Go, N., Taketomi, H., 1978b. Int. J. Peptide Protein Res. 13, 235–252.
Go, N., Taketomi, H., 1978c. Int. J. Peptide Protein Res. 13, 447–461.
Goldberg, A.L., 2003. Nature 426, 895–899.
Griko, Y.V., Gittis, A., Lattman, E.E., Privalov, P.L., 1994. J. Mol. Biol. 243, 93–99.
Hayryan, S., Hu, C.-K., Hu, S.-Y., Shang, R.-J., 2001. J. Comput. Chem. 22, 1287–1296.
Hayryan, S., Hu, C.-K., Skrivanek, J., Hayryan, E., Pokorny, I., 2005. J. Comp. Chem. 26, 334–343.
Hirano, S., Kamikubo, H., Yamazaki, Y., Kataoka, M., 2005. Proteins: Struct. Funct. Bioinf. 58, 271–277.
Hu, H.-Y., Wu, M.-C., Fang, H.-J., Forrest, M.D., Hu, C.-K., Tsong, T.-Y., Chen, H.-M. The role of tryptophan in staphylococcal nuclease stability, preprint.
Jaenicke, R., 1980. Protein Folding. Elsevier/North-Holland Biomedical Press, Amsterdam.
Kauzmann, W., 1959. Adv. Protein Chem. 14, 1–63.
Kanehisa, M.I., Tsong, T.-Y., 1978. J. Mol. Biol. 124, 177–194.
Kanehisa, M.I., Tsong, T.-Y., 1979. Biopolymers 18, 1375–1388.
Kaneshia, M.I., Tsong, T.-Y., 1979. J. Mol. Biol. 133, 279–283.
Kanehisa, M.I., Tsong, T.-Y., 1980. Biopolymers 19, 1617–1628.
Kim, P.S., Baldwin, R.L., 1982. Ann. Rev. Biochem. 51, 459–489.
Kiefhaber, T., Baldwin, R.L., 1995. J. Mol. Biol. 252, 122–132.
Kouza, M., Chang, C.-F., Hayryan, S., Yu, T.-H., Li, M.S., Huang, T.-H., Hu, C.-K., 2005. Biophys. J. 89, 3353–3361.
Kouza, M., Li, M.S., O'Brien, E.P., Hu, C.-K., Thirumalai Jr., D., 2006. J. Phys. Chem. A 110, 671–676.
Kouza, M., Hu, C.-K., Li, M.S., 2008. J. Chem. Phys. 128, 045103.
Kuwajima, K., 1989. Protein: Struct. Funct. Genet. 6, 87–103.
Kyte, J., Doolittle, R.F., 1982. J. Mol. Biol. 157, 105–132.
Lamy, A., Smith, J.C., 1996. J. Am. Chem. Soc. 118, 7326–7328.
Lin, C.-Y., Hu, C.-K., Hansmann, U.H.E., 2003. Proteins: Struct. Funct. Genet. 52, 436–445.
Li, M.-S., Hu, C.-K., Klimov, D.K., Thirumalai, D., 2006. Proc. Natl. Acad. Sci. U.S.A. 103, 93–98.
Li, M.S., Kouza, M., Hu, C.-K., 2007. Biophys. J. 92, 547–561.
Lumry, R., Biltonen, R., Brandts, J.F., 1966. Biopolymers 4, 917.
Mittag, T., Forman-Kay, J.D., 2007. Curr. Opin. Struct. Biol. 17, 3–14.
Nguyen, P.H., Stock, G., Mittag, E., Hu, C.-K., 2005. Proteins: Struct. Funct. Bioinf. 61, 795–808.
Privalov, P.L., 1979. Adv. Protein Chem. 33, 167–241.
Privalov, P.L., 1982. Adv. Protein Chem. 35, 1–104.
Privalov, P.L., Gill, S.J., 1988. Adv. Protein Chem. 39, 191.
Privalov, P.L., 1997. J. Chem. Thermodyn. 29, 447–474.
Ptitsyn, O.B., 1987. J. Protein Chem. 6, 273–293.
Ptitsyn, O.B., 1992. The molten globule state. In: Creighton, T.E. (Ed.), Protein Folding. Freeman and Co., New York, pp. 243–300.
Rose, G.D., 1978. Nature 272, 586–590.
Rose, G.D., 1979. J. Mol. Biol. 134, 447–470.
Rose, R.D., Roy, S., 1980. Proc. Natl. Acad. Sci. U.S.A. 77, 4643–4647.
Selkoe, D.J., 2003. Nature 426, 900–904.
Shortle, D., Abeygunawardana, C., 1993. Structure 1, 121–134.
Silberstein, K.A.T., Haymet, A.D.J., Dill, K.A., 1998. J. Am. Chem. Soc. 120, 3166–3175.
Sitia, R., Braakamn, I., 2003. Nature 426, 891–894.
Smith, A., 2003. Nature 426, 883.
Southhall, N.T., Dill, K.A., Haymet, A.D.J., 2002. J. Phys. Chem. B 106, 521.
Sturtevant, J.M., 1977. Proc. Natl. Acad. Sci. U.S.A. 74, 2236–2240.
Sun, S., Brem, R., Chan, H.-S., Dill, K.A., 1995. Protein Eng. 8, 1205–1213.
Su, Z.-D., Arooz, M.T., Chen, H.-M., Gross, C.J., Tsong, T.-Y., 1996. Proc. Natl. Acad. Sci. U.S.A. 93, 2539–2544.
Tanford, C., 1968. Adv. Protein Chem. 23, 121–282.
Tsong, T.-Y., Hearn, R.P., Wrathall, D.P., Sturtevant, J.M., 1970. Biochemistry 9, 2666–2677.
Tsong, T.-Y., Su, Z.-D., 1999. AIP Conf. Proc. 487, 37–53.
Tsong, T.-Y., Chang, C.-H., 2003. AAPPS Bull. 23, 12–18.
Wang, G., Dunbrack Jr., R.L., 2003. Bioinformatics 19, 1589, (The website of the protein sequence culling server is http://dunbrack.fccc.edu/pisces/. Here we used version cullpdb_pc20_res1.6_R0.25_d040724_chains743.).
Wu, M.-C., Hu, C.-K., Chen, H.-M., Tsong, T.-Y. Local hydrophobicity and protein secondary structure formation: theoretical model and experimental test, preprint.