Kinetics of Folding of Proteins and RNA

D. THIRUMALAI* AND S. A. WOODSON

Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received February 20, 1996

I. Introduction

The assembly of biological molecules, most notably globular proteins[1] and RNA,[2,3] into unique three-dimensional structures with well-defined topology is a complex and fascinating phenomenon in molecular biology. There are two aspects to the problem of folding of proteins and RNA. The first is the prediction of the three-dimensional structure of the folded state from the one-dimensional primary sequence of amino acids (for proteins) and nucleotides (for RNA). The second question is concerned with the kinetics of approach to the essentially unique folded state (also referred to as the native state) starting from an initial ensemble of disordered structures. In this Account we describe recent advances in our understanding of the kinetics of *in vitro* folding of globular proteins in terms of the underlying energy landscape. We further show that similar considerations can be usefully applied to describe the general features of the folding of RNA molecules.

The pioneering experiments of Anfinsen[4,5] and subsequent studies established that in many cases protein folding is a self-assembly process; i.e., the information needed for obtaining the three-dimensional structure is encoded in the primary sequence. These experiments did not provide the mechanisms of folding to the native conformation. The intellectual impetus to understand the kinetic mechanisms of protein folding came from Levinthal[6] who wondered how a protein molecule searches the astronomically large number of conformations to reach the native state on a biologically relevant time scale. It has been proposed recently, through statistical mechanical studies of several classes of minimal models, that the key to resolving the Levinthal paradox lies in elucidating the ways in which proteins explore the energy landscape.[7−11] In the minimal models only those aspects of a polypeptide chain which are thought to be crucial for describing the folding process are retained. These include chain connectivity, approximate representation of hydrophobic interactions, and self-avoidance between the various residues. Theoretical studies using the minimal models have shown that the free energy surface of typical proteins is rugged; i.e., there are many minima besides the one corresponding to the native state, which are separated by free energy barriers of varying heights. An examination of the dynamics in such a complex energy hypersurface leads to general kinetic scenarios for protein folding which are just beginning to be confirmed experimentally.

More recently, interest in the problem of RNA folding has been renewed by the discovery of catalytic RNA.[12,13] Pioneering work on the structure of transfer RNA first established that RNAs could form complex structures.[14−16] The expectation that catalytic RNAs should also fold into well-defined, compact structures has been borne out by a battery of biochemical, spectroscopic, and crystallographic experiments designed to probe RNA structure.[3,17,18]

From the energy landscape perspective it is natural to suggest that the considerations that lead to the theoretical developments of protein folding should also apply to RNA folding. In general terms, the requirements for RNA folding are analogous to those of protein folding. As is true for polypeptides, the number of conformations in the fully denatured state (the Levinthal limit) is large. For RNA sequences, the kinetic problem consists of forming the correct secondary structure, that is, Watson−Crick base pairs between complementary sequences, and achieving the correct three-dimensional organization of the structural elements.

In this context the challenge is to understand how the interplay of interactions among polynucleotides establishes an energy surface such that the native state is activated in a biologically meaningful time.

D. Thirumalai received his undergraduate education at the Indian Institute of Technology, Kanpur. In 1982 he received his Ph.D. degree from the University of Minnesota. He has been on the faculty of the University of Maryland since 1985. His current research interests lie in the areas of polymer science and biomolecular folding.

S. A. Woodson received her Ph.D. from Yale University in 1987 and did postdoctoral work at the University of Colorado, Boulder. She is currently an associate professor in the Department of Chemistry and Biochemistry at the University of Maryland, College Park. Her research interests are RNA folding and the splicing and evolution of group I introns.

(1) *Protein Folding*; Creighton, T. E., Ed.; Freeman: New York, 1992.
(2) Gesteland, R. F.; Atkins, J. F. *The RNA World*; Cold Spring Harbor Press: New York, 1993.
(3) Pyle, A. M.; Green, J. B. *Curr. Opin. Struct. Biol.* **1995**, *5*, 303.
(4) Anfinsen, C. A. *Science* **1973**, *181*, 223.
(5) Anfinsen, C. A.; Scheraga, H. A. *Adv. Protein Chem.* **1975**, *29*, 205.
(6) Levinthal, C. In *Mossbauer Spectroscopy in Biological Systems*; Debrunner, P., Tsibris, J. C. M., Monck, E., Eds.; University of Illinois Press: Urbana, 1968.
(7) Karplus, M.; Shakhnovich, E. I. In *Protein Folding*; Creighton, J. E., Ed.; W. H. Freeman: New York, 1992; p 127.
(8) (a) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561. (b) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1994**, *100*, 9238.
(9) (a) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524. (b) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct. Genet.* **1995**, *21*, 167. (c) Leopold, P.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *89*, 8721.
(10) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1169.
(11) Zwanzig, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2801.
(12) Cech, T. R.; Zang, A.; Grabowski, P. J. *Cell* **1981**, *27*, 487.
(13) Guerrier-Takada, C.; Altman, S. *Science* **1984**, *223*, 285.
(14) Kim, S. H.; Quigley, G. J.; Suddath, F. L.; McPherson, F. L.; Sneden, D.; Kim, J. J.; Weinzierl, J.; Rich, A. *Science* **1973**, *179*, 285.
(15) Moran, D.; Camarmond, M. B.; Fischer, J.; Weiss, R.; Thierry, J. C.; Ebel, J. P.; Giegé, R. *Nature* **1980**, *288*, 669.
(16) Woo, N. H.; Roe, B. A.; Rich, A. *Nature* **1980**, *286*, 346.
(17) (a) Jaeger, J. A.; Santa Lucia, J. Jr.; Tinoco, I. Jr. *Annu. Rev. Biochem.* **1993**, *62*, 255. (b) Moore, P. B. *Curr. Opin. Struct. Biol.* **1993**, *3*, 340.
(18) (a) Pley, H.; Flaherty, U. M.; McKay, D. B. *Nature* **1994**, *372*, 68. (b) Scott, W. G.; Finck, J. T.; Klug, A. *Cell* **1995**, *81*, 991.

S0001-4842(95)00093-8 CCC: $12.00

In both instances the problem boils down to describing the origins of the complex energy landscape and the ways of navigating the folding routes. Briefly our goals in this Account are to (i) obtain, in qualitative terms, the kinetic partitioning mechanism (KPM) from the energy landscape perspective,[19,21] and use this as a unifying concept to describe folding of proteins and RNA, (ii) discuss the existing experimental evidence in support of this picture, and (iii) argue how the KPM when applied to RNA folding suggests that RNA chaperonins should exist. These RNA cofactors or chaperones (which are only beginning to be identified) would rescue misfolded RNA structures and perhaps lead to an enhancement in the rates of folding.

II. Kinetic Partitioning Mechanism (KPM)

It is known from the study of several models of disordered systems that whenever one has several competing interactions in a system then the free energy surface could become rugged,[9,10] which means that there are several minima and a distribution of barrier heights.[21] Such systems are considered "frustrated" in the sense that all favorable interactions for a given particle cannot be simultaneously satisfied. Examples of such systems are spin glasses in which both ferromagnetic (attractive) and antiferromagnetic (repulsive) interactions are simultaneously present. As a result not all spins can have the most preferred interactions with every neighboring spin. This leads to energetic frustration. The kind of frustration that manifests itself in biological molecules is a bit more subtle than that found in spin glasses. In proteins and RNA, hydrophobic groups would prefer to be buried, creating a very compact globule, whereas hydrophilic residues (moieties that are highly polar) would be better accommodated by more extended structures. Since the hydrophobic species are dispersed throughout the primary sequence, it is clear that on any length scale (less than that of the entire molecule) there would be tendency for the hydrophobic residues to be in proximity. The resulting structures, although locally favorable, would be incompatible with the global fold. This conflict between local requirements and global considerations leads to "topological frustration".[22] A relatively unique ordered structure results when this frustration is minimized[22b,23] (or eliminated) on the scale of the size of the molecule. It is obvious that there are many ways of forming incorrect structures. Some of these can have many aspects in common with the native structure.[21] The presence of these low-energy native-like misfolded structures could serve as natural kinetic traps in the folding process.

Since RNA and proteins are "designed" in the process of evolution, it is likely that topological frustration is minimized. As a result these systems do not exhibit kinetic behavior similar to that of spin glasses which are frustrated (largely energetic) on all length scales. If we assume that natural biopolymers minimize topological frustration, it follows that generic

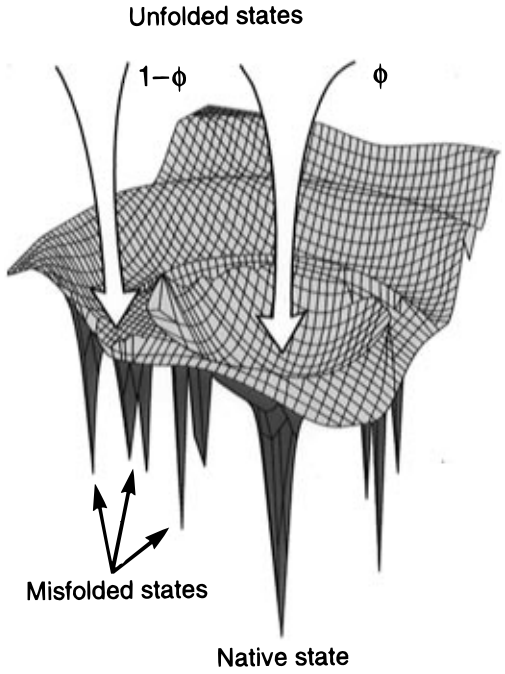

Figure 1. A schematic diagram of the free energy landscape for biomolecules. There are many minima that are separated by barriers of varying heights. For clarity we do not display the barriers between the various states. The deepest free energy minimum corresponds to the native conformation. In addition there are other deep minima in which the biomolecule adopts a misfolded conformation. A certain fraction, Φ, of initial population of unfolded molecules reach the native state rapidly without being trapped in any intermediate while the remaining fraction, $1 - \Phi$, become trapped as misfolded structures. The transition from the misfolded states to the native state involves overcoming a free energy barrier. This process, which slows the rate of folding, necessarily involves unraveling of (at least partially) the chain. The partition factor, Φ, is sensitive to mutations and external conditions.

random sequences of amino acids and nucleotides will not be able to fold to unique structures on any significant time scale. Thus, although for a given value of N (number of amino acids or nucleotides) one can form astronomically large numbers of possible sequences, only a very small percentage of such sequences would qualify to be biologically competent. It is also likely that among these only a small subset can fold on biologically relevant time scales. As a consequence of this minimal topological frustration in natural biopolymers there must exist sufficient free energy bias toward the native structure.[9,10,22,23] In other words there is a preferred basin of attraction (or funnel)[9b,c] in the free energy hypersurface corresponding to the native conformation. However, because of topological frustration, there are other deep minima corresponding to misfolded structures which play an important role in the kinetics of folding. The tendency to form misfolded structures becomes more prominent in high molecular weight proteins and RNA. For these larger systems under normal conditions, direct folding to the native conformation becomes rare. Folding of some of these larger biomolecules may require the presence of other cofactors.

A schematic sketch of the free energy surface is shown in Figure 1. From Figure 1 one can obtain, in a straightforward manner, the basic notions of the kinetic partitioning mechanism,[19,21] which we will argue is the unifying feature that allows us to analyze

(19) Thirumalai, D.; Guo, Z. *Biopolymers* **1995**, *35*, 137.
(20) Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
(21) Todd, M.; Lorimer, G.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4030.
(22) (a) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci.* **1990**, *87*, 3526. (b) Go, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.
(23) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6092.

in broad terms the results of folding kinetics of both RNA and proteins. Imagine the process by which an ensemble of unfolded structures (denatured) begin to navigate in this rough free energy landscape in search of the native conformation. There is a fraction of molecules, Φ, whose conformations map directly onto the native structure. These molecules would reach the native conformation extremely rapidly without encountering any discernible intermediates. The remaining fraction, $1 - \Phi$, would necessarily land in one of the deep metastable minima. Since subsequent rearrangement requires activation over a free energy barrier, the folding of this set of molecules would be slow. Due to the multivalley structure of the free energy surface, the ensemble of initially denatured molecules partition into fast folders (Φ being the fraction of such molecules) and slow folders. Hence, the mechanism that emerges from this consideration has been termed the kinetic partitioning mechanism, KPM.[19−21]

The partition factor Φ determines whether the folding process rapidly produces a high yield of the native conformation. Since the kinetic partition factor is determined by the structure of the free energy surface, it is not surprising that Φ depends on intrinsic factors like the sequence itself, as well as extrinsic factors such as pH, temperature, ionic strength, etc. It follows that Φ will vary in response to mutations of the primary sequence, a feature that accords well with laboratory experience.

The mechanisms by which the fast folders and slow folders reach the native conformation have been elucidated theoretically using the minimal model representation of proteins.[8] Guo and Thirumalai, using off-lattice models, have argued that the fast process corresponds to a native conformation nucleation collapse mechanism (NCNC).[19,20] Lattice model simulations[24] have been used to suggest that for a given sequence there are predetermined residues that trigger the nucleation collapse process. According to NCNC, once a critical number of residues form native-like contacts, a mobile structure in which the topology (including loops) is close to that of the native structure is formed. This mobile structure very rapidly reaches the native conformation. Thus, for the fraction of molecules that fold without forming detectable intermediates the *collapse process and the acquisition of native structure are almost synchronous*.

The remaining fraction, $1 - \Phi$, of molecules reach the native conformation by a complex three-stage multipathway mechanism (TSMM).[20,25] According to this mechanism the chain collapses into a compact conformation in the initial stage. This collapse process is nonspecific, and the structures that are obtained are largely determined by considerations such as loop formation probability and local favorable interactions. In the second stage the biopolymer diffusively searches among the set of compact structures, and as this process continues energetic forces drive the system into lower energy states. At the end of this stage the molecule contains several native-like features. However, since the initial collapse is nonspecific, the structures found at the end of this process are misfolded. Studies of minimal models suggest that these misfolded structures have between 50% and 80% tertiary contacts in common with the native conformation.[20,25−27] The final stage of TSMM corresponds to activated transitions from the misfolded structures to the native conformation.

A few comments concerning the general aspects of the kinetic partition mechanism as they pertain to the folding of biomolecules are worth making: First, it should be emphasized that both the direct process involving a nucleation mechanism and the three-stage multipathway mechanism may operate simultaneously under physiological conditions *in vitro* and *in vivo*. Second, a direct consequence of the KPM is that after a transient time the fraction of molecules that remain unfolded follows biexponential kinetics.[20] The biexponential kinetics does not occur because all the molecules go through a single similar major intermediate. The origin of the biexponential kinetics is due to the parallel occurrence of the direct nucleation process and the TSMM. For the fraction of molecules that follow the TSMM there is a partial unraveling of the chain in the transition state which occurs close to the native conformation. Third, the various time scales involved in the TSMM have been estimated for biomolecules with an arbitrary value of N, the number of residues in the protein, using scaling arguments.[28] The estimates show that the initial nonspecific collapse in the three-stage kinetics is very rapid and occurs on a time scale of about $10-100\ \mu$s for $N \approx 100$. The relative rates of the second and the third stages for a given value of the viscosity, or average surface tension between the residues and water, depend crucially on N. For $N \lesssim 30$ the second stage appears to be slower and could act as the rate-determining step, whereas for $N \gtrsim 30$ the rate-limiting step involves activated transitions from the set of misfolded structures to the native conformation. The reason for this is that the average free energy barrier separating the misfolded structures and the native conformation varies only as $N^{1/2}$,[28] which is significantly smaller than N. For example, for a protein with $N \approx 150$ the average free energy barrier height is $\sim 12 k_B T$ or 7.4 kcal/mol at $T \simeq 25$ °C. The time scale for overcoming this barrier is on the order of about 2 s. The reduction in barrier from the naive expectation of $N k_B T$ is due to the presence of disorder in biomolecules. The disorder arises due to the presence of conflicting interactions. This leads to the presence of a large number of structures that can access the native state. The entropy associated with these structures leads to an effective reduction in the free energy barrier separating the native conformation and the misfolded structures. Finally, if the partition factor Φ is large, then the time required to reach native state scales algebraically with N, i.e., a folding time of $\sim N^{\omega}$ where $3.8 \leq \omega \leq 4.2$.[28] For $N = 100$ this estimate yields time scales on the order of a few milliseconds. In general only for small proteins are the majority of molecules expected to reach the native conformation ($\Phi \approx 1$) via the nucleation collapse mechanism.[29] In this case there is only one dominant basin of attraction or funnel in the free energy landscape.[10,27]

(24) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Biochemistry* **1994**, *33*, 10026.
(25) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369.
(26) Socci, N.; Onuchic, J. N.; Wolynes, P. G. Preprint, October, 1995.
(27) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626.
(28) Thirumalai, D. *J. Phys. I* **1995**, *5*, 1457.
(29) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869.

There are features of the KPM that are in harmony with other approaches to the protein folding kinetics. This is especially the case when the partition factor Φ tends to unity. In this case the folding is rapid and is uninhibited by trapping in misfolded structures. In terms of the energy landscape perspective these fast folding processes are best described in terms of a dominant folding funnel or basin of attraction. A quantitative analysis of folding of small α-helical proteins in the $\Phi = 1$ regime has been recently attempted.[27]

Although most of the formulation of the KPM has been arrived at by studying models of proteins, we suggest that, due to the expected similarities in the free energy landscapes of RNA and proteins, roughly similar scenarios should be observable in the kinetics of RNA folding. We substantiate this further by analyzing the recent experiments of Emerick and Woodson[30] within this framework.

III. Protein Folding Experiments

In the last few years refolding kinetic experiments using pulsed hydrogen exchange in conjunction with NMR[31,32] on a number of proteins have been reported which can be readily interpreted in terms of the kinetic partitioning mechanism. Rather than describe each of these, we have selected two examples, one of which follows the KPM with a fractional value of Φ while the other appears to have $\Phi \approx 1$. We analyze these two experiments in some detail within the framework of the kinetic partitioning mechanism.

(1) Folding of a Cold-Shock Protein (CspB). In a recent experiment Schindler et al.[33] have studied the folding kinetics of a relatively small protein cold-shock protein, CspB, from *Bacillus subtilis*. The native structure of CspB, a protein with 67 amino acid residues, consists of a single five-stranded β-barrel. It is found that urea-induced unfolding is well described by a simple two-state mechanism.[33] The refolding of CspB in 0.6 M urea showed that the native state is reached extremely rapidly (in about 1.5 ms) without the protein being trapped in any intermediates. The probes monitoring the folding kinetics follow monoexponential kinetics. This is an example that, in the picture of KPM, would imply the partition factor Φ is nearly unity. It has been shown[20] that if $\Phi \approx 1$ then the folding kinetics are described by a single exponential. Theoretical studies indicate that for $\Phi \approx 1$ the rapid folding to the native conformation is dominated by the nucleation collapse mechanism.[19,20] For this process it has been estimated that the folding time scales as N^{ω} with ω varying between 3.8 and 4.2.[28] Using standard estimates of viscosity and average surface tension, the theoretical estimate for folding times for the specific nucleation collapse process with $N = 67$ is about 10^{-4} s. This estimate is in approximate agreement with the measurements of Schindler et al.[33]

(2) Refolding of Hen-Egg Lysozyme. In retrospect the earliest experiment that revealed clearly the salient aspects of KPM is the refolding of hen-egg lysozyme.[34] In experiments using a pulse labeling approach in conjunction with two-dimensional NMR, Radford et al. showed that the protection of amide protons (against hydrogen exchange) at several sites follows biexponential kinetics.[34] It can be shown that the kinetics of protection against hydrogen exchange offer a measure of the fraction of unfolded molecules as a function of time.[20] Radford et al. observed that folding of lysozyme consists of a fast process (less than about 10 ms) and a slow process that takes place on the order of 300 ms. They also showed that roughly 25% of the protein molecules reached a native-like conformation within 10 ms. These initial observations have been supplemented with further experiments that have been used to detect folding intermediates in the hydrogen exchange process by electrospray ionization mass spectrometry.[35] On the basis of these experiments Radford and Dobson have provided a rather detailed sketch of the folding pathways in hen-egg lysozyme.[36]

The broad features of the refolding kinetics described above are in accord with the theoretical picture based on the kinetic partitioning mechanism. The partition factor Φ for the condition of their experiments is 0.25. The fast process that leads directly to the formation of the native-like state corresponds, in our picture, to a nucleation collapse mechanism. The slow process corresponds to an initial collapse of the protein on a relatively short time scale to one of the misfolded structures and the subsequent rearrangement of this structure to the native conformation. If the slow process is dominant, as is the case for hen-egg lysozyme ($1 - \Phi \approx 0.75$), then it is clear that the rate-determining step would occur late in the folding process. The theoretical interpretation of the fast and slow processes can be tested by studying the temperature dependence of the refolding process.[19] Scaling arguments predict that the slow process would occur on a time scale on the order of $\tau_0 \exp(N^{1/2})$ (assuming the barrier height scales as $N^{1/2}$).[28] For hen-egg lysozyme this gives about 90 ms assuming τ_0 is on the order of about 10^{-6} s. This estimate is in rough accord with experiments.

Although we have only discussed two specific experiments for proteins, we have argued that the kinetic partitioning mechanism can account for refolding kinetics of various proteins that have been reported so far.[37−40] Furthermore, by a suitable generalization this basic idea has been used to propose a theory for chaperonin-assisted folding of proteins.[21] Thus, KPM unifies models of both *in vitro* and *in vivo* protein folding.

IV. Folding of RNA

(1) Folding of tRNA. Recent experiments[30] suggest that RNA folding can also be described by the

(30) (a) Emerick, V. L.; Woodson, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9675. (b) Emerick, V. L.; Pan, J.; Woodson, S. A. *Biochemistry* **1996**, in press.
(31) Roder, H.; Elove, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700.
(32) Udgaonkar, J. B.; Baldwin, R. *Nature* **1988**, *335*, 694.
(33) Schindler, T.; Herrlev, M.; Marahiel, M. A.; Schmid, F. X. *Nature Struct. Biol.* **1995**, *2*, 663.
(34) Radford, S.; Dobson, C. M.; Evan, P. A. *Nature* **1992**, *358*, 302.
(35) Minankar, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Nature* **1993**, *262*, 896.
(36) Radford, S. E.; Dobson, C. M. *Philos. Trans. R. Soc. London, B* **1995**, *348*, 17.
(37) Otzen, D. E.; Itzhaki, L. S.; Elmasry, N. F.; Jackson, S. E.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10422.
(38) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
(39) Kuszewski, J.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1994**, *3*, 1945.
(40) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. *Nature Struct. Biol.* **1994**, *1*, 149.

kinetic partition mechanism. We will first consider certain RNA sequences for which the partition factor Φ is near unity and then discuss the tendency of larger RNAs to misfold ($\Phi < 1$).

Examples of sequences that fold readily under favorable *in vitro* conditions include tRNAs, as well as certain small group I self-splicing introns.[41,42] Early NMR and temperature-jump experiments on $tRNA^{fMet}$ demonstrated that individual stem loops form on the microsecond time scale at physiological temperatures[43,44] in agreement with rates of helix formation for oligonucleotides.[45,46] Overall, folding of tRNA was estimated to proceed with a time constant of about 0.1−1 s.[43] The unfolding transition is highly cooperative[47] in the presence of Mg^{2+}, less so in its absence.[43] These time scales would be consistent with the picture that these small RNA molecules reach their folded state without being trapped in misfolded structures. The existence of direct pathways to the native state leads to Φ on the order of unity.

(2) Folding of *Tetrahymena* Ribozyme. The KPM described in section II suggests that the degree of topological frustration should increase as the length of the RNA increases. This would imply that, for larger RNA molecules, Φ is expected to be considerably less than unity and the occurrence of misfolded structures becomes more likely. Overall, the folding of such RNAs would appear slower due to the activated transitions from the misfolded structures to the native state. Fluorescence experiments on the *Tetrahymena* group I ribozyme by Bevilacqua and coworkers provide an estimate for folding rates in larger molecules.[48] Base pairing of a small substrate RNA with the ribozyme at 15 °C (4×10^6 M^{-1} s^{-1}) was approximately 10-fold slower than association of analogous oligonucleotides. Subsequent docking of the substrate helix in the active site of the ribozyme was much slower (2.5 s^{-1}), suggesting that formation of tertiary interactions may be rate determining for the folding of large RNAs.[48] These experiments indicate that folding in these larger RNAs does not occur via the nucleation mechanism.

The most direct evidence for the general validity of the KPM for RNA folding comes from the refolding experiments on precursor RNA containing the *Tetrahymena* ribozyme (Figure 2). Using self-splicing kinetics and gel electrophoresis, Emerick and Woodson showed that a mixed population of active and native inactive conformers are in slow exchange (0.1 min^{-1}) at 30 °C.[30] An advantage of gel electrophoresis is that values for Φ can be readily determined. Approximately 70−90% of wild-type *Tetrahymena* precursor RNA is misfolded after transcription at 30 °C.[49] Reannealing of the RNA heated to 75 °C followed by cooling to 30 °C reduces the proportion of inactive molecules to 20−30%. This suggests that the RNA is

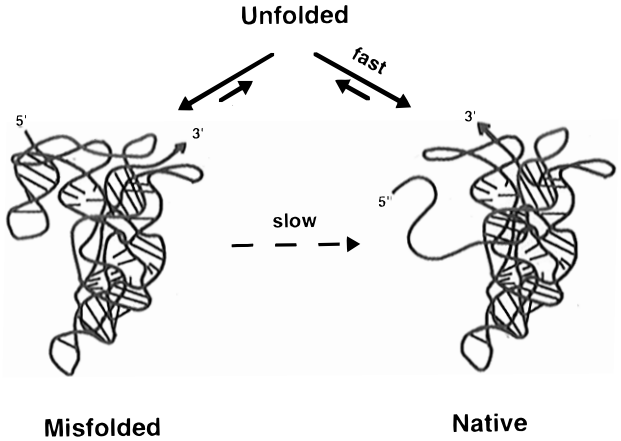

Figure 2. Model for folding of the *Tetrahymena* group I precursor RNA. At physiological temperatures (30 °C), a fraction of the RNA ($\Phi = 0.2$) rapidly achieves the native state (right) while the remainder is partitioned into misfolded structures (left). Each pathway (denoted by double arrows) is reversible and involves multiple steps from an ensemble of unfolded structures. The native state contains base pairs between the 5′ exon (green, right) and the 5′ end of the intron (red). Misfolded precursor RNAs typically contain alternative base pairs within the 5′ exon (green, left). To reach the native state, these molecules must partially unfold, resulting in slow folding kinetics. The core ribozyme structure is represented in blue and is adapted from Michel and Westhof.[77]

trapped in an inactive conformation, but can be driven into the native state by subsequent denaturation and renaturation.[30,49] *Tetrahymena* precursor RNA sequences that do not achieve the native state may be trapped in specific misfolded structures that contain an alternative set of base pairs near the 5′ splice site.[50] This alternative secondary structure is known to inhibit self-splicing activity.[51]

The predicted difference in thermodynamic stability of the native and misfolded structures is relatively small (a few kilocalories per mole). Accordingly, the fraction of RNA molecules that fold correctly during transcription is sensitive to changes in sequence; even single-point mutations can increase or decrease the probability of reaching the native state by several-fold.[42,49] As predicted for folding of relatively large proteins,[28] slow transitions late in the folding pathway arise from the need to escape from local minima in the energy landscape.

The various observations of Emerick and Woodson[30] are generally consistent with the basic KPM. According to this model the RNA quickly becomes trapped in misfolded structures that rearrange on a longer time scale. The partition factor for this large RNA at $T = 30$ °C is indeed quite small ($\Phi \approx 0.2$),[30] which means that the majority of the molecules have to overcome the free energy barriers separating the misfolded structures and the native state. The kinetic barriers can be overcome either by heating and recooling, i.e., by an annealing mechanism, or on very long time scales.

An unusual feature of the *Tetrahymena* pre-rRNA is that while the energetic barrier between native and non-native structures is high, the temperature dependence of the rate of isomerization is shallow. Conversion of inactive to active forms displays Arrhenius

(41) Privalov, P. L.; Filominov, V. V. *J. Mol. Biol.* **1978**, *122*, 447.
(42) Jaeger, L.; Westhoff, E.; Michel, F. *J. Mol. Biol.* **1993**, *234*, 331.
(43) Crothers, D. M.; Cole, P. E.; Hilben, C. W.; Schulman, R. G. *J. Mol. Biol.* **1974**, *87*, 63.
(44) Riesner, M. T.; Lohman, T. M.; de Hasseth, P. *J. Mol. Biol.* **1976**, *107*, 145.
(45) Porschke, D.; Eigen, M. *J. Mol. Biol.* **1971**, *62*, 361.
(46) Craig, M. E.; Crothers, D. M.; Doty, P. *J. Mol. Biol.* **1971**, *62*, 383.
(47) Kearns, D. R.; Patel, D. J.; Schulman, R. G.; Yamane, T. *J. Mol. Biol.* **1971**, *61*, 265.
(48) Bevilacqua, P. C.; Kierzek, R.; Johnson, K. A.; Turner, D. H. *Science* **1992**, *258*, 1355.
(49) Emerick, V. L.; Woodson, S. A. *Biochemistry* **1993**, *32*, 14062.
(50) Woodson, S. A.; Emerick, V. L. *Mol. Cell. Biol.* **1993**, *13*, 1137.
(51) Woodson, S. A.; Cech, T. R. *Biochemistry* **1991**, *30*, 2042.

behavior, with an activation energy of 10−15 kcal/mol.[30] The estimate based on scaling arguments gives the free energy of activation as $N^{1/2}K_BT$, which for *Tetrahymena* pre-RNA ($N = 650$) turns out to be around 15 kcal/mol.[28] This value is much lower than enthalpies typically seen for opening of RNA helices.[43,52] The rate of this transition may be limited by disruption of non-base-pairing interactions, for which the entropic term is significant. Exchange of only a few base pairs at a time, as occurs during branch migration, also results in a low activation barrier (25 kcal/mol).[53,54]

Biphasic folding kinetics in RNAs have also been observed by single-strand specific chemical modification.[42,55a] Zarrinkar and Williamson[55b] have probed the folding kinetics of *Tetrahymena* ribozyme by trapping single-stranded regions with an excess of complementary oligonucleotides. They have found evidence for multiple kinetic intermediates, in which the formation of a long-range double helix was rate determining and surprisingly slow (0.74 min^{-1}).[55b] On the basis their experiments they have suggested that the kinetic assembly of RNA proceeds by a hierarchical mechanism, that is, sequential folding of subdomains. In contrast, the KPM points to the existence of multiple parallel paths to the native state and supposes that under nonfavorable conditions ($T = 30$ °C and no annealing) there is a small fraction of molecules that rapidly reach the native state. These findings may appear to be in conflict with those of Emerick and Woodson[30] and with the KPM mechanism.[19,20] However, this apparent contradiction can be understood by noting that the shortest time scale in the experiments of Zarrinkar and Williamson is approximately 1 min. Theoretical arguments[28] suggest that for large enough RNAs the early kinetic events in RNA folding would occur on the time scale of milliseconds or less depending on the sequence and on external conditions. It is generally acknowledged[55,56] that better experimental methods that can monitor RNA conformational changes on shorter time scales are needed to resolve these questions.

V. RNA Folding *in Vivo*

In order to achieve efficient folding of RNAs in the cell, kinetic barriers to the native structure such as misfolded intermediates must be overcome. It is therefore likely that *in vivo* RNA folding may be an assisted self-assembly process. The need for "RNA chaperonins" or a family of cofactors becomes noticeable if we consider the low partition factor for *Tetrahymena* precursor RNA ($\Phi \approx 0.2$ at $T = 30$ °C), which means the majority of molecules are misfolded. In protein folding, the chaperonin-assisted assembly is needed most whenever Φ is small, such as typically happens for large proteins.[57] In this case the complex formed between the chaperonin and the substrate protein utilizes the energy upon ATP hydrolysis to overcome the barriers separating the misfolded structures and the native states.[21,58]

There is also evidence for a similar "chaperone-like" mechanism for the folding of large RNAs. For the *Tetrahymena* ribozyme, splicing is still 20−50 times more rapid *in vivo* (20 min^{-1}) than *in vitro* (0.6 min^{-1}),[59,60] providing a lower limit for the rate at which the active structure is achieved in the cell. Splicing is facilitated to the same extent in bacteria as it is in *Tetrahymena*, demonstrating that a species-specific protein is not required.[61] Instead, general cellular components or proteins that bind the ribosomal RNA sequences that flank the ribozyme appear to be sufficient to accelerate formation of the native structure. Natural RNA chaperonins have not been identified yet, but likely candidates include abundant ribosomal proteins or the protein hnRNP A1.[62] Nonspecific RNA binding proteins have been shown to accelerate the rate of RNA-catalyzed reactions *in vitro* by promoting dissociation of misfolded structures.[63,64] Finally, RNA molecules may act as RNA chaperones. For example, small nucleolar RNAs are complementary to sequences throughout ribosomal RNA, suggesting that they play a role in assembly of active ribosomes.[65] It appears likely that cells will be shown to possess a class of proteins or even other RNA molecules that function as chaperones for RNA transcripts, just as chaperones are required to fold polypeptides.

VII. Differences between Proteins and RNA

Despite the obvious analogies between the folding of proteins and RNA, key differences remain to be explored. An important factor in the folding of RNA is that condensation of a polyanion is necessarily sensitive to differences in ionic strength, as the electrostatic repulsion along the backbone provides a strong opposing force to the collapse of hydrophobic bases.[66] Nearly all complex RNAs require Mg^{2+} for biological activity, for both catalysis and stabilization of the native structure.[67,68] Mg^{2+} binding is cooperative and coincides with the formation of tertiary structure.[69,70]

Although it is evident that divalent ions are specifically coordinated in the native state of RNA, their role in the folding process is not yet understood. Two extreme positions may be considered. On the one hand, metal coordination may only occur late in the folding process, after formation of the binding site. In this case, the presence of Mg^{2+} during the early stages of folding should have only a small influence on partitioning between misfolded and native states. On

(52) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 9373.
(53) LeCuyer, K. A.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3373.
(54) Panyutin, J. G.; Hsieh, P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2021.
(55) (a) Banerjee, A. R.; Turner, D. H. *Biochemistry* **1995**, *34*, 6504. (b) Zarrinkar, P. P.; Williamson, J. R. *Science* **1994**, *265*, 918.
(56) Herschlag, D. *J. Biol. Chem.* **1995**, *270*, 2071.
(57) Todd, M.; Vitanen, P. V.; Lorimer, G. *Science* **1994**, *265*, 659.
(58) Thirumalai, D. In *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions*; Doniach, S., Ed.; Plenum Press: New York, 1994; p 115.
(59) Brehm, S. F.; Cech, T. R. *Biochemistry* **1983**, *22*, 2390.
(60) Bass, B. L.; Cech, T. R. *Nature* **1984**, *308*, 820.
(61) Zhang, F.; Ramsay, E. S.; Woodson, S. A. *RNA* **1995**, *1*, 284.
(62) Dreyfuss, G.; Maturis, M. J.; Pinol-Roman, S.; Bond, C. G. *Annu. Rev. Biol. Chem.* **1993**, *62*, 289.
(63) Tsuchihashi, Z.; Khosla, M.; Herschlag, D. *Science* **1993**, *262*, 99.
(64) Coetzee, T.; Herschlag, D.; Belfort, M. *Genes Dev.* **1994**, *8*, 1575.
(65) Steitz, J. A.; Tycowski, K. *Science* **1995**, *270*, 1626.
(66) Record, M. T.; Lohman, T. M.; de Hasseth, P. *J. Mol. Biol.* **1976**, *107*, 145.
(67) Grosshans, C. A.; Cech, T. R. *Biochemistry* **1989**, *28*, 6888.
(68) Pyle, A. M. *Science* **1993**, *261*, 709.
(69) Celander, D. W.; Cech, T. R. *Science* **1991**, *251*, 401.
(70) Lynch, D. C.; Schimmel, P. R. *Biochemistry* **1974**, *13*, 1841.

the other hand, rapid association of divalent ions with the unfolded chain could serve as a nucleation event that is perhaps similar to heme ligation in cytochrome *c*.[40] Here, the KPM model would predict Φ to be strongly Mg^{2+}-dependent. In fact, the true situation may be somewhat between these two, in that binding of magnesium to partially folded molecules induces further conformational change.[55,70]

Another difference between proteins and RNA is that nucleic acid secondary structure is energetically stable in the absence of tertiary contacts.[52] On the basis of thermodynamic considerations, it would appear that formation of secondary and tertiary contacts can be decoupled. Two broad transitions are typically observed during thermal denaturation of tRNA[71] and group I robozymes,[42] of which the low-temperature transition corresponds to disruption of tertiary interactions.[42,72] This has led to the suggestion that the formation of secondary structure precedes that of native tertiary contacts.[47,55,72] As discussed above, tertiary interactions appear to form more slowly than Watson−Crick base pairs. On the other hand, tertiary interactions contribute to the relative stabilities of alternative structures and as such contribute to the overall energy landscape.[49,73] Additional experiments are clearly required to determine to what extent thermodynamic intermediates observed during denaturation determine the kinetics of folding.

VIII. Concluding Remarks

It appears to us that we have only begun to appreciate the need to understand folding kinetics in both RNA and proteins from a unified point of view. The energy landscape perspective provides a clear way to rationalize and anticipate general scenarios of folding of biomolecules. The kinetic partitioning mechanism should be viewed as a tentative proposal to account for folding kinetics of biomolecules. Our analysis also suggests that in both cases it is clear that there is a need for experiments on short enough time scales so that the early events can be monitored. Such experiments have been initiated recently in the protein folding community.[74] From a computational point of view it is clear that minimal model and realistic representations of RNA sequences would be very desirable.[75,76] These studies have already suggested that drastic differences should exist in the refolding kinetics of small and large RNA molecules. The kinetics and thermodynamics of such models may provide additional insights into RNA folding. Theoretical work is needed to extend the results of minimal model studies to understand folding in real proteins. Efforts along these lines have just begun.[27,28]

This work was supported in part by grants from the National Science Foundation, Air Force Office of Scientific Research, National Institutes of Health, American Cancer Society, Pew Charitable Trust, and Camille and Henry Dreyfus Foundation. We are grateful to our collaborators C. J. Camacho, V. Emerick, Z. Guo, H. Orland, and G. Lorimer for numerous discussions. D.T. acknowledges several illuminating discussions on protein folding over the years with J. D. Bryngelson, H. S. Chan, K. A. Dill, W. Eaton, J. Onuchic, J. Straub, P. Wolynes, and R. Zwanzig.

AR9500933

(71) Stein, A.; Crothers, D. M. *Biochemistry* **1976**, *15*, 160.
(72) Bannerjee, A.; Jaeger, J. A.; Turner, D. H. *Biochemistry* **1993**, *32*, 153.
(73) Gluick, T. C.; Draper, D. E. *J. Mol. Biol.* **1994**, *241*, 246.
(74) Jones, C. M.; Henry, E. R.; Hu, Y.; Chan, C. K.; Luck, S. D.; Bhuyan, A.; Roder, H.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11860.
(75) Higgs, P. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 2531.
(76) Higgs, P. *J. Phys. I* **1993**, *3*, 43.
(77) Michel, F.; Westhoff, E. *J. Mol. Biol.* **1990**, *216*, 585.