

Conformations of Co-Translational Folding Intermediates

Michael S. Evans, Thomas F. Clarke IV and Patricia L. Clark*

Department of Chemistry and Biochemistry, University of Notre Dame, 251 Nieuwland Science Hall, Notre Dame, IN 46556 USA.

Abstract: While *in vitro* experiments have contributed much to our understanding of protein folding, we know much less about how proteins fold in the more complex environment of the cell. This review summarizes our current knowledge of the earliest *in vivo* folding intermediates: the conformations adopted by nascent polypeptides during synthesis by the ribosome. The challenges related to successful folding in the cellular environment, including off-pathway aggregation and macromolecular crowding, are also discussed.

Keywords: Protein, folding, co-translational, intermediate, misfolding, aggregation, ribosome, translation.

INTRODUCTION

Most of what is currently known about the process of protein folding has been learned from studying the folding behavior of purified proteins in dilute solution. For example, *in vitro* studies have been crucial for building our understanding of the presence of multiple pathways for folding chains, including the energy landscape concepts (i.e., 'folding funnels') developed from computational studies [1]. *In vitro* studies have also revealed much about the conformational preferences of denatured polypeptide chains. Long viewed as a statistical random coil (for example, [2]), several recent studies indicate that denatured chains have significant conformational preferences [3-5]. In addition, *in vitro* studies were instrumental in revealing the 'molten globule' as a common intermediate for folding [6], and the effect of contact order on folding rate [7, 8]. Yet *in vitro* studies of protein folding omit several key features present in the cellular folding environment. For example, what should we consider as the 'denatured ensemble' for newly synthesized polypeptide chains emerging vectorially from the ribosome? How many folding pathways are relevant under conditions in which co-translational folding can occur? Are there common intermediates for co-translational folding, something akin to the 'molten globules' observed *in vitro*? Although much remains to be learned with regard to these questions, below we focus on a variety of recent experimental and theoretical approaches exploring co-translational folding of nascent polypeptide chains.

PROTEIN FOLDING *IN VIVO*

Where examined, protein folding *in vivo* typically results in higher yields of native protein than protein folding *in vitro* [9, 10]. There are several possibilities to explain how proteins accomplish this feat in the seemingly more hostile

environment of the cell, relative to the more dilute, cooler *in vitro* environment, though our current understanding of these processes is far from comprehensive. One possibility is that protein stability is different *in vivo*. Yet studies using urea-treated *E. coli* to measure unfolding of a labeled protein have shown no change in stability versus measurements made *in vitro* [11, 12]. Given the potential for intermolecular interactions and/or destabilization of other cellular components, however, it is difficult to draw broad conclusions from these isolated reports. The folding process itself may be faster *in vivo* as well. For a relatively slow folding protein such as *Salmonella* phage P22 tailspike, the half-time for refolding *in vitro* is 24 hours [13], while the folding of newly synthesized chains *in vivo* occurs with a half time of five minutes [13, 14]. Another possibility is that proteins folding *in vivo* may avoid aggregation-prone conformations. This could be mediated by direct interactions with the ribosome or more peripheral parts of the translation apparatus, or perhaps vectorial appearance of the nascent chain alone.

Molecular chaperones are another component of the cellular folding environment. Yet studies indicate that only ~20% of newly synthesized chains require interactions with chaperones for proper folding [15, 16]. Indeed, for some of the more well-characterized chaperones such as DnaK, the cytoplasmic Hsp70 of *E. coli*, deletion of the chaperone results in no defects in cell growth or viability [17, 18]. Two excellent reviews on the details of chaperone interactions with nascent chains have been published recently [15, 16]. As a result, this review will not discuss the roles played by molecular chaperones in the folding of newly synthesized proteins. Likewise, this review will not discuss co-translational folding of proteins secreted into or through membranes to other cellular compartments; a comprehensive discussion of this topic can be found elsewhere [19]. Rather, this review will focus on the path of the elongating nascent chain as it exits the ribosome, and the conformations adopted by the chain as it interacts with increasing stretches of itself, the surface of the ribosome, and the crowded environment of the cytoplasm.

*Address correspondence to this author at the Department of Chemistry and Biochemistry, University of Notre Dame, 251 Nieuwland Science Hall, Notre Dame, IN 46556 USA; Tel: +1 574 631 8353; Fax: +1 574 631 6652; Email: pclark1@nd.edu

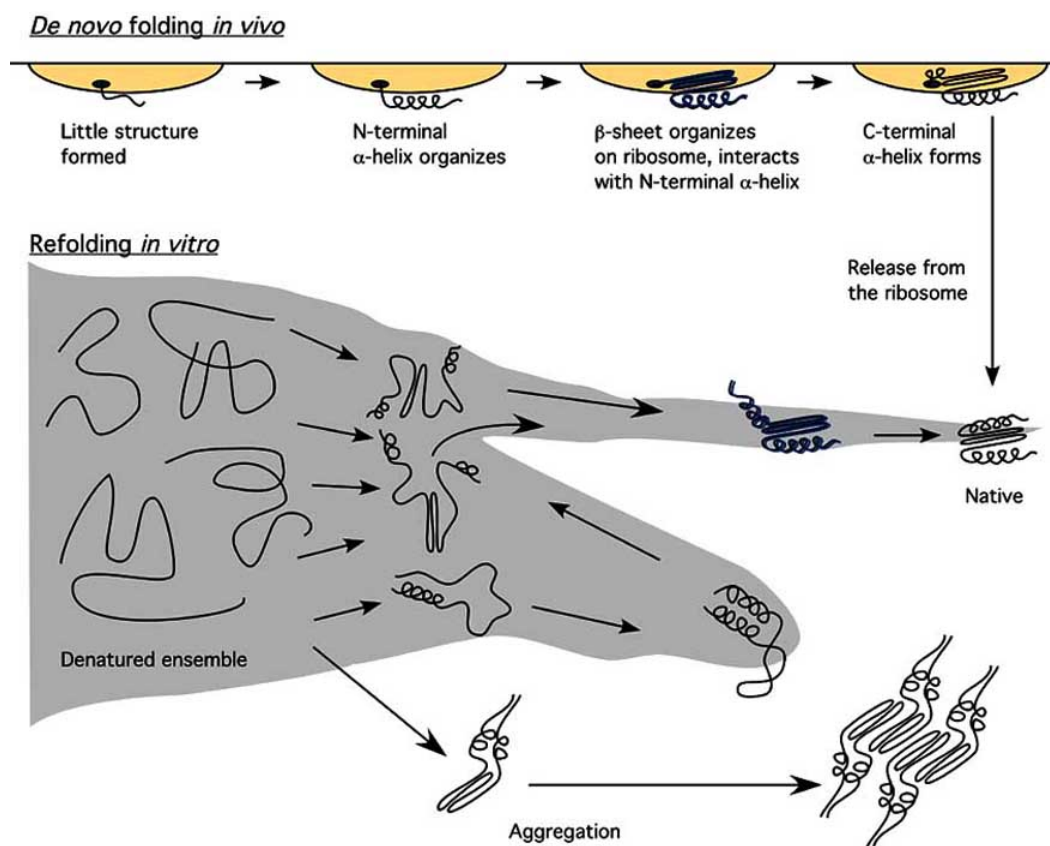


Figure 1. A conceptual model for *de novo* protein folding *in vivo* versus protein refolding *in vitro*. *In vivo*, the vectorial nature of protein synthesis necessarily limits the conformational space available to short nascent chains. As synthesis progresses, N- and C-terminal sequences are isolated spatially and temporally from each other, possibly preventing non-native interactions. During synthesis, the chain may dwell on the surface of the ribosome. In contrast, *in vitro* refolding begins from an ensemble of random conformations. In this case, N- and C-terminal interactions may form early during folding, leading to non-native interactions. Some non-native interactions may lead to aggregation, or to an intermediate that must unfold somewhat before accumulating more native interactions leading to the native state. In the figure, an on-pathway intermediate common to both *de novo* and *in vitro* folding is indicated by a blue shadow.

PARTITIONING BETWEEN FOLDING AND AGGREGATION

Aggregation mechanisms are not nearly as well understood as folding mechanisms, in part because of the difficulties of working with a non-equilibrium system and the lack of high-resolution structural information. Yet with the rapidly increasing number of human diseases now known to arise from aggregation, the mechanisms involved are now receiving more attention. Early studies using systems such as hen egg white lysozyme [20] and phage P22 tailspike [21] suggested the soluble, monomeric precursors of aggregation are misfolded intermediate conformations that form early during refolding *in vitro* (Figure 1). More recent studies indicate that mutations that promote aggregation correlate with substitutions that increase hydrophobicity and β -sheet secondary structural propensity, while decreasing net charge [22, 23]. These results suggest that co-translational folding intermediates, particularly those rich in β -sheet character and/or hydrophobic residues, may have significant aggregation propensity. Presumably, one important role of molecular chaperones is to suppress such aggregation. Yet experimental evidence demonstrating that the majority of

newly synthesized chains do not require the intervention of chaperones in order to fold correctly and efficiently [15, 16] suggests that alternative mechanisms, such as unique characteristics of vectorial folding intermediates, and/or interactions between the nascent chain and the ribosome itself (see below), may play a significant role in boosting the yield of protein folding *in vivo*.

MACROMOLECULAR CROWDING PROMOTES FOLDING AND AGGREGATION

It is impossible to consider the effect of co-translational intermediates on protein folding *in vivo* without considering the broader effects of the highly crowded conditions in the cell. Estimates for the total concentration of macromolecules in the *E. coli* cytoplasm range from 300-400 mg/mL, or about 30% of total volume [24]. When considering protein folding *in vivo*, this macromolecular crowding has two general effects [24]. First, crowding increases the thermodynamic activity of all species in solution, favoring the association of macromolecules, which can in turn enhance both folding and aggregation reactions. Second, crowding limits the diffusion of molecules. This limits the

conformational flexibility of polypeptides, adding complexity to folding and multimerization reactions. While the effects of macromolecular crowding can complicate folding studies, several recent experiments examining crowding *in vivo* and *in vitro* have made progress in understanding these complex effects.

Recent *in vitro* experiments indicate crowding agents enhance both the association of macromolecules with respect to folding and multimerization, and the adoption of a more compact state by otherwise randomly structured macromolecules. Van den Berg and coworkers used both Ficoll 70 and BSA as crowding agents to describe the kinetic effects on the refolding of reduced hen egg white lysozyme (HEWL) [25]. Following the formation of disulfide bonds, a population of native-like HEWL intermediates formed, and the conversion of these intermediates to the native structure was accelerated in the presence of crowding agents. The formation of an alternative intermediate, which perhaps must unfold to join the productive folding pathway, is slowed in the presence of crowding agents. In addition, the authors noted a general affect of lowered yields of native HEWL in the presence of crowding agents. These results demonstrate the association-enhancement property of macromolecular crowding, but also underscore a major difference between protein folding *in vivo* and *in vitro*: namely, proteins do fold efficiently in the crowded conditions of the cell; they are not plagued by the misfolding observed for HEWL under crowded conditions *in vitro*.

Macromolecular crowding, *via* excluded volume effects, has been shown to favor compact states over extended conformations. Sasahara and coworkers [26] demonstrated that acid denatured cytochrome *c* collapses to a molten globule conformation in the presence of crowding agents. This molten globule conformation is similar to a KCl-induced molten globule, where chloride ions interact with positively charged regions in the cytochrome *c* structure. Additional evidence for the favoring of a compact versus extended conformation under crowded conditions comes from the work of Qu and Bolen [27]. In this case, reduced and carboxyamided RNase T1 (TCAM), an intrinsically unfolded protein, was used. Using dextran 70 as the crowding agent, the authors showed a shift in the denatured-native equilibrium towards the native state of RNase T1, based on an enzymatic activity assay. Adding crowding agents thus resulted in a 5-7.5-fold increase in the equilibrium constant for folding. *In vivo*, the compactness stabilized by crowding may promote co-translational folding, protecting nascent chains from proteolytic digestion and/or inappropriate associations with other nearby nascent chains.

In vitro crowding experiments generally use an unreactive crowding agent. While such agents are chosen for their solubility properties, it is now clear that even inert crowding agents do not exert their effects in a uniform manner. Van den Berg and coworkers [25, 28] observed that BSA had a more deleterious effect on the refolding of HEWL than Ficoll or dextran. However, examination of the kinetics of normalized folding revealed that crowding with BSA accelerated the fast phase of folding relative to Ficoll. Qu and Bolen observed a greater amount of folded TCAM in the presence of dextran 70 than in the presence of either

Ficoll or BSA [27]. By examining the collapse of RNaseA and PEG in crowding agents of different size, Tokuriki and coworkers offered a possible explanation for the variety of results seen with different crowding agents [29]. Their results suggested that a crowding agent near in size to the compact state of the protein of interest would be most likely to favor collapse. In such a case, the volume available for the protein would be similar to its native state, but too small for extended, denatured conformations, favoring folding. These results emphasize the difficulty of approximating the cellular environment *in vitro*, with regard to both the size and chemical properties of macromolecules in solution.

While great progress has been made on understanding the effects of crowding on the folding of full-length proteins *in vitro*, there is little evidence to date on the effects of crowding *in vivo*. Recently, Dedmon and coworkers [30] have made progress in addressing this challenging problem. Their studies focused on FlgM, a *S. typhimurium* protein that is intrinsically disordered in dilute solution. Upon addition of its *in vivo* binding partner, the transcription factor ²⁸, the C-terminal portion of FlgM gains α -helical structure. By measuring HSQC NMR spectra of *E. coli* cells overexpressing FlgM, but not overexpressing the ²⁸ homologue, Dedmon and coworkers showed the *in vivo* environment induced C-terminal structure in FlgM, just as binding to its partner did *in vitro*. Additionally, C-terminal structure was induced by physiologically relevant concentrations of glucose and BSA. As progress is made on understanding the effects of crowding *in vivo*, it will be important to focus not only on full-length proteins, but also on polypeptide chains during synthesis. These will be challenging experiments, but the results described above with FlgM may provide a first approximation of the effect of crowding on nascent chains *in vivo*, particularly for short nascent chains classically assumed to be disordered in dilute solution.

VECTORIAL APPEARANCE OF NASCENT CHAIN: ENVIRONMENTS AND THE EFFECTS OF TRANSLATION RATE

Nascent chain synthesis occurs deep within the ribosome, at the junction between the two subunits. After leaving the P-site, the elongating chain progresses down a 100 Å long, 10-20 Å diameter tunnel through the 50S subunit, initially surrounded exclusively by 23S ribosomal RNA. Based on cross-linking data, relatively long (30 aa) peptides may fold back towards the peptidyl transferase center [31]. The ribosome exit tunnel is composed primarily of RNA, and does not display significant hydrophobicity. As a result, the majority of nascent chains are not expected to have significant chaperone-like interactions with the ribosomal tunnel [32]. This does not mean, however, that the tunnel is an inert component of co-translational folding. Approximately half-way through the length of the tunnel, ribosomal proteins L4 and L22 protrude into the tunnel, significantly reducing the conformational space available for the nascent chain [32]. Specific sequences have been shown to form stable interactions with this narrow portion of the tunnel, stalling translation [33]. After widening past this constriction, the width of the ribosomal tunnel may allow secondary structures to form within the tunnel itself [34].

The final ribosomal protein encountered by the nascent chain as it traverses the tunnel is L23, which serves as a docking site for ribosome-associated chaperones and factors involved in nascent chain translocation across membranes [35, 36].

What conformations are adopted by the nascent polypeptide chain once it exits the ribosome tunnel? As previously mentioned, only ~20% of the proteome requires chaperones to fold correctly *in vivo*, so our focus moves to the intermediates formed by the remaining ~80% of nascent chains. While the conceptual *in vitro* denatured state depicts a polypeptide chain sampling all conformations equally with no thermodynamic boundaries between them, the thermodynamic penalty for solvating exposed hydrophobic side chains means it is highly unlikely a nascent chain will assume an extended or 'random coil' conformation on the ribosome. The *in vivo* denatured state can be interpreted, therefore, as an absence of persistent structure, though some rough conformational preferences may be present. Simulations using polyglutamic acid and the "blob" model indicated that the transition to a stable fold could occur co-translationally, since the folding window required for initial contacts is only 32 aa [37]. This result agrees well with studies of stalled nascent chains tethered to ribosomes, which have demonstrated native-like contacts for peptides as short as 86 aa (including 46-66 aa outside of the ribosome tunnel) [38]. This suggests 32-66 aa residues must exit the ribosomal tunnel and begin sampling a broad conformational space before folding can commence; this process is limited by translation rate and not the nanosecond conformational sampling seen for isolated chains *in vitro* [39]. While it has previously been established that regions of secondary structure can form before a general global collapse of the polypeptide chain [40], the structural prerequisites for a chain bound to a ribosome have not been elucidated.

There is clearly not an obligate requirement for chain release before nascent chain folding: the addition of progressively longer linkers to the C-terminus can result in a fully functional enzyme while still bound to the ribosome (for example, [41]). The implication for *in vivo* nascent chain folding is clear: the nascent chain may obtain native secondary and tertiary structure while still bound to the ribosome, if enough distance exists between the ribosome and the folding peptide. In the study of Semliki Forest virus capsid protease domain, a linker of 45 aa was required [41]; however, depending on the topology of the native structure, this distance may vary. Furthermore, preliminary secondary structure, incomplete intermediate structure, and/or productive non-native interactions may form before this threshold is reached.

In vitro folding pathways do not require stepwise folding from N-terminus to C-terminus, yet this is precisely the limitation placed on co-translational folding of nascent chains. It is unlikely that *in vitro* folding intermediates that initiate from anywhere other than the N-terminus are significantly populated during synthesis *in vivo*, as the N-terminal regions of the nascent chain are unlikely to remain unstructured until interior portions emerge from the ribosome exit tunnel. For example, the folding of apomyoglobin *in vitro* proceeds through a well-characterized pair of intermediates with ordered structure in the C-terminal

-helix [42]. The two *in vitro* intermediates, H and AGH, both require the final helix to organize before additional folding occurs. However, studies of the conformational preferences of C-terminally truncated apomyoglobin peptides indicate that non-native structures may arise co-translationally, forming native-like structure only as the chain elongates further [43]. The most notable finding was the presence of regular structure in N-terminal apomyoglobin peptides as short as 36 aa; while this structure is non-native, it may resemble *in vivo* co-translational conformational preferences. Likewise, while vectorial appearance of the nascent chain may eliminate some intermediates observed *in vitro*, vectorial appearance may also favor the formation of certain on-pathway folding intermediates (Figure 1). For example, a study of firefly luciferase found that a co-translational proteinase-resistant folding intermediate, while seen in the unfolding pathway, is not populated in the chaperone-assisted *in vitro* refolding pathway [44]. Presumably, this on-pathway *in vivo* intermediate could be responsible for the increased speed and yield of luciferase folding *in vivo* compared to *in vitro*.

The significance of vectorial appearance of the nascent chain on folding is illustrated by a series of papers that have altered the kinetics of translation. Synonymous codon usage has been implicated in translational control as it relates to movement of the ribosomes and the presence of translational pauses. A number of factors, including tRNA availability [45], tRNA recharging [46], optimal codon/anti-codon pairing [47] and resemblance to conserved recognition sequences [48, 49], have optimized codon usage so that the ribosomal machinery progresses through some mRNA sequences quickly and efficiently while pausing at specific rare codons [50]. Unfortunately, few absolute measurements have been made for sequence-dependent translation rate (two notable exceptions are [51, 52]). Yet the effects can be significant: as few as eight rare codons can cause an increase in translation time of approximately three seconds [53]. The precise effects of this non-linear translation rate on nascent chain folding remain to be determined. Interestingly, however, a negative correlation between gene length and codon usage has been observed: longer genes tend to have more rare codons than very short genes [54]. Furthermore, -helical sequences tend to have fewer rare codons than -sheet sequences [55], and domain boundaries appear to be a common location for rare codons [56, 57]. These observations suggest a possible relationship between synonymous codon usage and structure formation: since folding rates correlate with contact order [7] and chain length [8], pauses after domains and -sheets could allow these structures to form co-translationally before the appearance of a C-terminal sequence that might contribute to an aggregation-prone conformation.

The effects of translation rate on folding have been examined for a few specific proteins. For example, eliminating three rare codons from a turn in a helix-turn-helix motif of EgFABP1 causes aggregation to increase dramatically [58]. Intriguingly, this motif is not essential for proper folding *in vitro* [59], suggesting the effect of the increased translation rate is not a local phenomenon, but instead allows a stable, off-pathway structure to form. Likewise, silent mutations in a series of codons

corresponding to amino acid residues in the active site of chloramphenicol acetyltransferase were found to decrease the activity of the protein. These silent mutations eliminated a specific pause observed during *in vitro* translation of the protein [60].

INTERACTIONS WITH THE RIBOSOME AND OTHER COMPONENTS OF THE TRANSLATION MACHINERY

Perhaps the most well studied example of specific interactions between a nascent chain and the translation machinery involves the N-terminal signal sequences of secretory proteins. The first 15-30 aa of these nascent chains interacts with the signal recognition particle (SRP). SRP also interacts with the ribosome, and this ternary complex halts translation of the nascent chain until the ribosome reaches the ER membrane [61].

And as introduced above, there are now several examples of specific interactions between nascent chains and the lining of the polypeptide chain exit tunnel in the large subunit of the ribosome. Many of these interactions result in translation arrest. For example, short (~20 aa) nascent chains from upstream open reading frames (uORFs) inhibit translation termination and chain release in *cis* in a sequence-dependant manner; two well-studied examples are the human cytomegalovirus (CMV) gpUL4 [62, 63], and *Neurospora crassa* arg-2 [64]. Presumably, specific interactions between the uORF nascent chains and the wall of the exit tunnel trigger translation arrest, though the precise mechanism has yet to be uncovered. In addition, two recent papers have identified sequence signatures in full length ORFs that, when present in the polypeptide exit tunnel, cause translation arrest [65, 66]. For one of these nascent chains (SecM), the sequence responsible for translation arrest is a long distance from the N-terminus (residues 155-166), and its effects can be suppressed by mutations in ribosomal protein L22 and nearby nucleotides of the 23S rRNA; these sites line a narrow portion in the middle of the polypeptide exit tunnel [65]. Perhaps most importantly, these cases represent direct control of translation by the nascent chain alone, and do not require a third player acting in a manner analogous to SRP. These examples illustrate that the ribosome itself has the ability to sense features of the nascent polypeptide chain, and modulate its polypeptide synthesis activity as a result [66, 67]. It remains to be seen what other roles exist for nascent chain control of polypeptide synthesis, and the resulting effects on protein folding.

What evidence exists for nascent chain:ribosome interactions as a mechanism for increasing intracellular folding yields? This is a largely unexplored area, but some insight may be gained from cells that exhibit global defects in polypeptide chain termination and release, such as the *[PSI⁺]* yeast strains that form prion-like structures that sequester the endogenous Sup35p/eRF3 nascent chain release factor. Intriguingly, Tuite and coworkers demonstrated that *[PSI⁺]* strains containing eRF3 prions are not only deficient in translation termination but also exhibit enhanced tolerance to heat and chemical stress, suggesting a possible connection between extended nascent chain:ribosome interaction and stabilization of folding

intermediates *in vivo* [68]. Supporting this hypothesis, these authors did not detect any differences in the up-regulation of stress response proteins (i.e. Hsp70, Hsp104p) in *[PSI⁺]* and *[psi⁻]* strains, and similar thermal and chemical tolerance was also observed for a yeast strain with a functional mutation in the eRF3 gene [68]. These results confirm that the improved thermal and chemical tolerance was not due to the action of molecular chaperones, or a secondary effect of the yeast prion itself, but may instead be a function of extended chain dwelling on the ribosome.

Likewise, newly synthesized P22 tailspike polypeptide chains dwell on *Salmonella* ribosomes far longer than required for chain termination and release [69]. One possibility is that tailspike chain dwelling and interactions with the ribosome may provide a mechanism for increasing productive intracellular folding yields. Indeed, tailspike is not unique in registering folding yields that are much higher *in vivo* than *in vitro* [10].

PROKARYOTES VERSUS EUKARYOTES: DIFFERENT CO-TRANSLATIONAL CONFORMATIONS?

A 1997 study by Netzer and Hartl suggested that nascent chains of multi-domain proteins may differ in the degree of co-translational folding depending on whether the polypeptide chains were produced in eukaryotes or prokaryotes: Eukaryotic translation appeared to allow N-terminal domains to fold to a native, enzymatically active state co-translationally, while prokaryotic translation did not [70]. However, subsequent studies have questioned the generality of this observation. For example, Helenius and coworkers demonstrated that an N-terminal chymotrypsin-like domain of a large viral preprotein can fold to its native state co-translationally in both eukaryotic and prokaryotic translation systems [71]. Likewise, Spirin and coworkers demonstrated that the eukaryotic multi-domain protein firefly luciferase was capable of co-translational folding to an enzymatically active state in a prokaryotic translation system, provided a C-terminal extension was added so that the entire luciferase sequence could thread out of the ribosome exit tunnel [72]. These results highlight the difficulties of uncovering general mechanisms for co-translational folding, given the relatively small data set for these studies.

CO-TRANSLATIONAL FOLDING INTERMEDIATES: CASE STUDIES

Despite the experimental challenges of studying co-translational folding intermediates, several studies have begun to elucidate *in vivo* mechanisms for successful, high yield protein folding to the native state. The unifying theme of these studies is that co-translational formation of native-like structure, combined with the vectorial nature of polypeptide biosynthesis, may protect nascent chains from kinetic traps encountered by denatured proteins refolding *in vitro*.

Demonstrating that co-translational folding is possible at all was an important prerequisite for more detailed structural studies. Kudlicki and coworkers showed that, when provided with C-terminal extensions, rhodanese tethered to ribosomes

can fold to an enzymatically active conformation [73]. Yet this study did not explicitly demonstrate co-translational folding of the wild type rhodanese sequence (i.e., folding may have occurred only after translation of the C-terminal extension). To directly demonstrate that γ -globin can fold co-translationally, Komar *et al.* programmed *in vitro* translation reactions with truncated mRNAs lacking stop codons in order to generate ribosome-bound γ -globin chains of discrete lengths [38]. Specific heme binding was detected on several truncated ribosome-bound nascent chains, including one containing only the first 86 amino acids of γ -globin, clearly showing that the polypeptide chain forms native-like structural contacts co-translationally.

Additional studies highlight the differences between *de novo* folding and *in vitro* refolding. Fedorov and Baldwin [10] investigated the formation of the enzymatically active bacterial luciferase heterodimer. Folded monomer was added to either refolding from denaturant or synthesized from an *in vitro* translation reaction; luciferase activity appeared earlier in reactions with the *in vitro* translated . When was translated *in vitro* using mRNA lacking a stop codon, and the polypeptide was subsequently released with puromycin, luciferase activity appeared more quickly when was added simultaneously, as opposed to after, puromycin treatment, suggesting released chains adopt a structured conformation that associates with more slowly than a co-translational conformation. Stopped-flow experiments investigating the kinetics of dimerization *in vitro* revealed the presence of two distinct phases, representing a distribution in the conformational ensemble of polypeptide chains. A small burst phase was attributed to the fast-associating co-translational-like , and a slow phase associated with the more stable, but less association-competent . The general model proposed by Fedorov and Baldwin to explain these results is that the co-translational intermediate is less stable, and closer in energy to the transition state for the / dimerization reaction. The derived from *in vitro* refolding was more stable, but this stability came at the cost of a slower association with . Perhaps most significantly, this study demonstrated that polypeptide chains folding co-translationally could avoid kinetic traps populated *in vitro*.

Experiments by Frydman and coworkers [44] have demonstrated monomeric firefly luciferase forms native structure co-translationally, and that this native structure is similar to native structure retained in a chemical unfolding intermediate. An N-terminal protease resistant fragment was identified from *in vitro* translation reactions at times prior to the appearance of full-length luciferase. This same domain, identified by mass spectrometry, was seen in proteolytic digestions of luciferase partially unfolded by guanidine HCl. Proteolytic digestion of luciferase refolding from denaturant in a translation lysate did not show a protease-resistant domain, indicating that the intermediate identified is unique to *de novo* synthesis of luciferase. Notably, refolding of luciferase *in vitro* is chaperone-dependent, and occurs more slowly than *de novo* folding, supporting the kinetic trap avoidance hypothesis of Fedorov and Baldwin. Both works demonstrate the difference between *de novo* folding and *in vitro* refolding of proteins, even when refolding occurs in an

environment similar to that used for *de novo* folding, as shown by the experiments with firefly luciferase.

Early formation of native structure on the ribosome may be important for the successful folding and multimerization of the trimeric tailspike attachment protein from *Salmonella* phage P22. Only about 20% of cold-trapped *in vivo* tailspike folding intermediates can fold productively *in vitro* at physiological temperatures. Those intermediates that can continue folding are trimeric, indicating that some proper folding in the monomer is critical for productive assembly [74]. More recent experiments by Clark and King have shown that *de novo* synthesized tailspike dwells on the 30S subunit and polysomes, as assayed by pulse-chase experiments and sucrose gradient fractionation of lysed cells [69]. When ribosomes were digested by RNase treatment, radiolabeled native tailspike increased, demonstrating that these ribosome-bound intermediates were on-pathway. Additionally, these ribosome-bound intermediates could be recognized by an -native monoclonal antibody, whereas early *in vitro* refolding intermediates were not recognized by the same antibody. And, following RNase digestion, there was an increase in antigenicity in 30S fractions, even though there were fewer 30S-tailspike complexes following digestion. Following digestion, either the formation of a more native-like tailspike structure, or a rearrangement of the 30S-tailspike complex, which made the epitope more accessible, could have been responsible for this increased recognition. Taken together, these experiments reinforce the difference between *in vitro* and *de novo* folding, and highlight the importance of the ribosome in productive folding. This work also offers tantalizing clues regarding the location on the ribosome where maturation of nascent chains can take place. While all these three case studies have focused attention on the importance of the ribosome and co-translational folding in partitioning nascent chains to productive folding pathways in the crowded environment of the cell, many details remain to be resolved.

PROSPECTS FOR FUTURE RESEARCH

Reports describing co-translational protein folding have provided glimpses of the earliest steps of protein folding *in vivo*. However, many questions remain. For instance, it will be necessary to characterize co-translational folding intermediates from many more proteins before conclusions will emerge about possible general features of co-translational intermediates. Also, some studies [69, 70] have implicated a role for the ribosome itself in co-translational folding, but how general is this effect? And while co-translational folding is conceptually intuitive for low contact order proteins, high contact order proteins would seem to present a formidable challenge for co-translational folding. How does the cellular folding machinery cope with this challenge? Codon usage may play a role here, introducing translational pause sites in order to allow domains to fold, but further experiments are needed to confirm this model. Certainly, with so many challenging yet fundamentally important questions still unanswered, it is likely that the characterization of *in vivo* folding mechanisms, and co-translational folding intermediates in particular, will receive increased attention in the coming years.

REFERENCES

- [1] Dill, K. A. and Chan, H. S. (1997) *Nat. Struct. Biol.*, 4, 10-19.
- [2] Levinthal, C. (1968) *Chim. Phys.*, 65, 44-45.
- [3] Crowhurst, K. A.; Tollinger, M. and Forman-Kay, J. D. (2002) *J. Mol. Biol.*, 322, 163-178.
- [4] Tcherkasskaya, O. and Uversky, V. N. (2003) *Protein Pept. Lett.*, 10, 239-245.
- [5] Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M. and Schwalbe, H. (2002) *Science*, 295, 1719-1722.
- [6] Arai, M. and Kuwajima, K. (2000) *Adv. Protein Chem.*, 53, 209-282.
- [7] Plaxco, K. W.; Simons, K. T. and Baker, D. (1998) *J. Mol. Biol.*, 277, 985-994.
- [8] Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D. and Finkelstein, A. V. (2003) *Protein Sci.*, 12, 2057-2062.
- [9] Seckler, R.; Fuchs, A.; King, J. and Jaenicke, R. (1989) *J. Biol. Chem.*, 264, 11750-11753.
- [10] Fedorov, A. N. and Baldwin, T. O. (1999) *J. Mol. Biol.*, 294, 579-586.
- [11] Ghaemmaghami, S. and Oas, T. G. (2001) *Nat. Struct. Biol.*, 8, 879-882.
- [12] Ignatova, Z. and Gierasch, L. M. (2004) *Proc. Natl Acad. Sci. USA*, 101, 523-8.
- [13] Fuchs, A.; Seiderer, C. and Seckler, R. (1991) *Biochemistry*, 30, 6598-6604.
- [14] Goldenberg, D.; Berget, P. and King, J. (1982) *J. Biol. Chem.*, 257, 7864-7871.
- [15] Frydman, J. (2001) *Ann. Rev. Biochem.*, 70, 603-647.
- [16] Hartl, F. U. and Hayer-Hartl, M. (2002) *Science*, 295, 1852-1858.
- [17] Deuerling, E.; Schulze-Specking, A.; Tomoyasu, T.; Mogk, A. and Bukau, B. (1999) *Nature*, 400, 693-696.
- [18] Teter, S. A.; Houry, W. A.; Ang, D.; Tradler, T.; Rockabrand, D.; Fischer, G.; Blum, P.; Georgopoulos, C. and Hartl, F. U. (1999) *Cell*, 97, 755-765.
- [19] Muller, M.; Koch, H. G.; Beck, K. and Schafer, U. (2001) *Prog. Nucleic Acid Res. Mol. Biol.*, 66, 107-157.
- [20] Goldberg, M. E.; Rudolph, R. and Jaenicke, R. (1991) *Biochemistry*, 30, 2790-2797.
- [21] King, J.; Haase-Pettingell, C.; Robinson, A. S.; Speed, M. and Mitragli, A. (1996) *FASEB J.*, 10, 57-66.
- [22] Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G. and Dobson, C. M. (2002) *Nat. Struct. Biol.*, 9, 137-143.
- [23] Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G. and Dobson, C. M. (2003) *Nature*, 424, 805-808.
- [24] Ellis, R. J. (2001) *Trends Biochem. Sci.*, 26, 597-604.
- [25] van den Berg, B.; Wain, R.; Dobson, C. M. and Ellis, R. J. (2000) *EMBO J.*, 19, 3870-3875.
- [26] Sasahara, K.; McPhie, P. and Minton, A. P. (2003) *J. Mol. Biol.*, 326, 1227-1237.
- [27] Qu, Y. and Bolen, D. W. (2002) *Biophys. Chem.*, 101-102, 155-165.
- [28] van den Berg, B.; Ellis, R. J. and Dobson, C. M. (1999) *EMBO J.*, 18, 6927-6933.
- [29] Tokuriki, N.; Kinjo, M.; Negi, S.; Hoshino, M.; Goto, Y.; Urabe, I. and Yomo, T. (2004) *Protein Sci.*, 13, 125-133.
- [30] Dedmon, M. M.; Patel, C. N.; Young, G. B. and Pielak, G. J. (2002) *Proc. Natl Acad. Sci. USA*, 99, 12681-12684.
- [31] Choi, K. M. and Brimacombe, R. (1998) *Nucl. Acids Res.*, 26, 887-895.
- [32] Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B. and Steitz, T. A. (2000) *Science*, 289, 920-930.
- [33] Nakatogawa, H. and Ito, K. (2004) *Chem. Biochem.*, 5, 48-51.
- [34] Hardesty, B. and Kramer, G. (2001) *Prog. Nucleic Acid Res. Mol. Biol.*, 66, 41-66.
- [35] Kramer, G.; Rauch, T.; Rist, W.; Vorderwulbecke, S.; Patzelt, H.; Schulze-Specking, A.; Ban, N.; Deuerling, E. and Bukau, B. (2002) *Nature*, 419, 171-174.
- [36] Gu, S. Q.; Peske, F.; Wieden, H. J.; Rodnina, M. V. and Wintermeyer, W. (2003) *RNA*, 9, 566-573.
- [37] Duhamel, J.; Kanagalingam, S.; O'Brien, T. J. and Ingratta, M. W. (2003) *J. Am. Chem. Soc.*, 125, 12810-12822.
- [38] Komar, A. A.; Kommer, A.; Krashennnikov, I. A. and Spirin, A. S. (1997) *J. Biol. Chem.*, 272, 10646-10651.
- [39] Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M. and Kiefhaber, T. (2003) *J. Mol. Biol.*, 332, 265-274.
- [40] Kippen, A.; Sancho, J. and Fersht, A. R. (1994) *Biochemistry*, 33, 3778-3786.
- [41] Kowarik, M.; Kung, S.; Martoglio, B. and Helenius, A. (2002) *Mol. Cell*, 10, 769-778.
- [42] Jennings, P. A. and Wright, P. E. (1993) *Science*, 262, 892-896.
- [43] Chow, C. C.; Chow, C.; Raghunathan, V.; Huppert, T. J.; Kimball, E. B. and Cavagnero, S. (2003) *Biochemistry*, 42, 7090-7099.
- [44] Frydman, J.; Erdjument-Bromage, H.; Tempst, P. and Hartl, F. U. (1999) *Nat. Struct. Biol.*, 6, 697-705.
- [45] Dong, H.; Nilsson, L. and Kurland, C. G. (1996) *J. Mol. Biol.*, 260, 649-663.
- [46] Elf, J.; Nilsson, D.; Tenson, T. and Ehrenberg, M. (2003) *Science*, 300, 1718-1722.
- [47] Akashi, H. (2003) *Genetics*, 164, 1291-1303.
- [48] Kane, J. F. (1995) *Curr. Op. Biotechnol.*, 6, 494-500.
- [49] Moszer, I.; Rocha, E. P. and Danchin, A. (1999) *Curr. Op. Microbiol.*, 2, 524-528.
- [50] Ramachandiran, V.; Kramer, G. and Hardesty, B. (2000) *FEBS Lett.*, 482, 185-188.
- [51] Curran, J. F. and Yarus, M. (1989) *J. Mol. Biol.*, 209, 65-77.
- [52] Sorensen, M. A. and Pedersen, S. (1991) *J. Mol. Biol.*, 222, 265-280.
- [53] Sorensen, M. A.; Kurland, C. G. and Pedersen, S. (1989) *J. Mol. Biol.*, 207, 365-377.
- [54] Duret, L. and Mouchiroud, D. (1999) *Proc. Natl. Acad. Sci. USA*, 96, 4482-4487.
- [55] Thanaraj, T. A. and Argos, P. (1996) *Protein Sci.*, 5, 1973-1983.
- [56] Thanaraj, T. A. and Argos, P. (1996) *Protein Sci.*, 5, 1594-1612.
- [57] Dessen, P. and Kepes, F. (2000) *Gene*, 244, 89-96.
- [58] Cortazzo, P.; Cervenansky, C.; Marin, M.; Reiss, C.; Ehrlich, R.; and Deana, A. (2002) *Biochem. Biophys. Res. Commun.*, 293, 537-41.
- [59] Kim, K.; Cistola, D. P. and Frieden, C. (1996) *Biochemistry*, 35, 7553-7558.
- [60] Komar, A. A.; Lesnik, T. and Reiss, C. (1999) *FEBS Lett.*, 462, 387-391.
- [61] Walter, P. and Johnson, A. E. (1994) *Ann. Rev. Cell Biol.*, 10, 87-119.
- [62] Cao, J. and Geballe, A. P. (1996) *Mol. Cell Biol.*, 16, 7109-7114.
- [63] Cao, J. and Geballe, A. P. (1998) *RNA*, 4, 181-188.
- [64] Wang, Z.; Fang, P. and Sachs, M. S. (1998) *Mol. Cell Biol.*, 18, 7528-7536.
- [65] Nakatogawa, H. and Ito, K. (2002) *Cell*, 108, 629-636.
- [66] Gong, F. and Yanofsky, C. (2002) *Science*, 297, 1864-1867.
- [67] Tenson, T. and Ehrenberg, M. (2002) *Cell*, 108, 591-594.
- [68] Eaglestone, S. S.; Cox, B. S. and Tuite, M. F. (1999) *EMBO J.*, 18, 1974-1981.
- [69] Clark, P. L. and King, J. (2001) *J. Biol. Chem.*, 276, 25411-25420.
- [70] Netzer, W. J. and Hartl, F. U. (1997) *Nature*, 388, 343-349.
- [71] Nicola, A. V.; Chen, W. and Helenius, A. (1999) *Nat. Cell Biol.*, 1, 341-345.
- [72] Kolb, V. A.; Makeyev, E. V. and Spirin, A. S. (2000) *J. Biol. Chem.*, 275, 16597-16601.
- [73] Kudlicki, W.; Chirgwin, J.; Kramer, G. and Hardesty, B. (1995) *Biochemistry*, 34, 14284-14287.
- [74] Goldenberg, D. and King, J. (1982) *Proc. Natl Acad. Sci. USA*, 79, 3403-3407.

Received on March 10, 2004, accepted on April 1, 2004.