Protein folding on the ribosome
Lisa D Cabrita\textsuperscript{1,2}, Christopher M Dobson\textsuperscript{1} and John Christodoulou\textsuperscript{2}

In living systems, polypeptide chains are synthesised on ribosomes, molecular machines composed of over 50 protein and nucleic acid molecules. As nascent chains emerge from the ribosomal exit tunnel and into the cellular environment, the majority must fold into specific structures in order to function. In this article we discuss recent approaches designed to reveal how such folding occurs and review our current knowledge of this complex self-assembly process.

Protein folding
It is well established that the activity of a protein molecule is inextricably linked to its three-dimensional fold, and that the information required for a protein to adopt its biologically active state is intrinsic to its amino acid sequence. The essential principles of the mechanism of protein folding are now known to be based on a stochastic search on a biased energy landscape (Figure 1) in which only a tiny fraction of all possible structures needs to be sampled [1]. Knowledge of how the folding process occurs, and how misfolding is avoided within the cellular environment, is central to our understanding of the nature of living systems. Moreover, protein misfolding can result in the degradation of newly synthesised polypeptide chains, a process that can give rise to medical conditions such as cystic fibrosis, or to their aggregation, a phenomenon associated with a wide range of devastating disorders that include neurodegenerative conditions such as Alzheimer’s and Parkinson’s diseases [2].

Our current understanding of the molecular basis of protein folding comes almost entirely from experimental investigations \textit{in vitro} of the renaturation of chemically or thermally denatured full-length proteins under a variety of solution conditions, coupled with theoretical studies and \textit{in silico} computer simulations. Such studies have been extended to examine the effects on the folding of auxiliary factors, such as molecular chaperones, and therefore serve, in addition, to identify at least some features of the protein folding process as it is likely to occur in a living system [3**]. In contrast to \textit{in vitro} experiments, where folding is initiated from a denatured, full-length protein, the starting point of this process \textit{in vivo} is the synthesis of the nascent polypeptide chain by the ribosome under physiological conditions. During its synthesis, the growing nascent chain threads through an ‘exit tunnel’ within the ribosomal particle (Figure 2) and into the cellular environment. In addition to the complexities that are an inherent feature of such an environment that is crowded with the wide range of molecules on which life depends, protein folding \textit{in vivo} is coupled with the progressive emergence of the nascent chain from the ribosome, which occurs in a vectorial manner from the N- to the C-terminus [3**]. The way in which structure develops in the growing nascent chain, both whilst it is tethered to the ribosome and following its release, represents a fascinating example of the complex interplay between structure and dynamics that is inherent in biological systems.

Studies of the renaturation of full-length molecules \textit{in vitro} indicate that small proteins at least can fold completely in seconds or less, and that the formation of secondary and tertiary contacts can occur on very much shorter timescales [1]. The rate of biosynthesis \textit{in vivo}, however, proceeds much more slowly, for example involving the incorporation of 2–4 amino acid residues per second into the growing chain in a typical eukaryotic systems [4]. The rate-limiting step in the folding of a nascent chain on the ribosome, therefore, could well be the production of the polypeptide chain itself. Whether a growing nascent chain can undergo a significant degree of folding during biosynthesis, or whether extrinsic folding effectively occurs only after release from the ribosome, has been the subject of considerable speculation as we discuss below.

The contemporary view of protein folding is that it can be represented as a process taking place on an energy surface or landscape in which the native structure corresponds, in the simplest case, to the lowest energy state (Figure 1). One can then visualise the process of the folding of an ensemble of denatured molecules as a set of trajectories that result from the fact that, on average, native-like interactions are more stable than non-native ones. Provided that the energy surface is appropriately biased towards the native state, therefore, a stochastic search
process can result in its discovery in a remarkably efficient manner. Generally however, and particularly for larger proteins, there are local minima on the energy surface representing relatively disordered ensembles of conformations that can act as kinetic traps for folding and may, in some cases, lead to misfolding and aggregation [5]. During biosynthesis, the incremental addition of each amino acid to the growing polypeptide chain progressively increases the total conformational space that is potentially accessible to the chain. If, however, the emerging nascent chain can acquire a significant degree of native-like structure, through the process of ‘co-translational folding’, the extent of conformational space that would otherwise be sampled during the search for the native fold may be substantially reduced. The fact that the nascent chain is tethered at its C-terminus to the peptidyl transferase centre (PTC) (Figure 2), and that it can interact both with the ribosome and with auxiliary factors such as molecular chaperones, is likely to exert an additional bias in the energy surface towards the native state. Any such additional bias is likely, in addition to aiding the conformational search for the native state, to decrease the probability of non-native structures being populated, facilitating efficient folding still further.

An indication of the effects of the progressive increase in the length of a protein chain is seen from biophysical studies of N-terminal polypeptide fragments of sperm whale apomyoglobin [6], an all α-helical single domain protein. Short fragments in solution can acquire a significant degree of β-sheet character, whilst α-helical structure develops as the chain becomes longer; indeed the native-like α-helical fold is substantially acquired when the chain length reaches 119 residues out of a total of 153. Moreover, the transition from predominantly β sheet to α helical content is concomitant with a decreased propensity for the fragments to self-associate [6]. The progressive formation of native-like interactions in vitro suggests that similar behaviour might occur in vivo, as the N-terminus of a protein emerges from the ribosome before the formation of the C-terminus. The way in which vectorial folding in vivo can differ from that of protein folding in vitro, where the interactions within the entire polypeptide chain can in principle be sampled at any
Protein folding on the ribosome

stage in the folding process, and how the fundamental principles of folding derived from in vitro studies are manifested in the complex environment of the cell, are important and challenging questions.

The structure of the ribosome and the nature of co-translational protein folding

The intact 70S ribosomal particle is a complex of over 50 proteins and three RNA molecules that form a large [50S] and a small [30S] subunit (Figure 2) in prokaryotes. The structure, assembly and function of the two subunits and of the intact ribosome have been studied in great detail by a multitude of biochemical [7–10] and biophysical approaches [11,12,13,14,15,16,17,18], the latter including cryoEM [19**] and X-ray crystallography [20**]. Indeed, studies using these latter techniques have captured the ribosomes engaged in various states of translation and revealed such phenomena as ratcheting, the rapid rotation of the ribosomal subunits that is required during protein elongation [20**,21**]. The latter studies in particular highlight the fact that, despite the overall complexity of the ribosome, it functions as a dynamic macromolecular machine.

Of the structural features of the 70S complex, the nature of the ribosomal exit tunnel, the channel that links the PTC where synthesis actually occurs to the cellular environment, and through which the nascent chain emerges, is of particular interest in the context of co-translational folding. The ribosomal tunnel is more than 80 Å in length, and varies between 10 Å and 20 Å in width. It is lined by segments of a large RNA molecule (the 23S rRNA) and of the ribosomal proteins L4 and L22 [22–24,25**] (Figure 2). In addition, the L23 protein is located at the exit interface and has been found to serve as a docking point for a range of ribosome-associated binding species, notably, trigger factor (TF) [26], as described later in this article (Figure 2). The ribosomal exit tunnel has been described as having ‘teflon-like’ qualities [23] as it is hydrophilic in character [24], a property that might allow the nascent chain to traverse the tunnel in a relatively secluded and unhindered manner.

Recent cryoEM studies [27**] at 6 Å resolution of ribosomes stalled during translation, have revealed clear density that shows that a nascent chain can make a number of interactions within the ribosomal exit tunnel, and provides further support for extensive biochemical evidence that certain stretches of amino acids, such as residues 150–166 of the SecM protein [28,29] that interact strongly with the L4 and L22 proteins, and also rare codons [30**] interfere with protein synthesis, as do many antibiotics [31**]. In addition, ‘molecular tape’ [32] measurements of elongation rates during protein synthesis suggest that positively charged residues such as arginine can generate a transient arrest of translation by altering the local electrostatic potentials within the tunnel, perhaps representing one means of communication between the site of synthesis on the ribosome and the nascent chain. Indeed, cryoEM studies [27**,33,34**] of translation-arrested ribosomes have provided no evidence for large scale conformational changes within the ribosomal subunits or the tunnel, also suggesting that signal transduction between the stalled nascent chain and the

COSTBI-754; NO OF PAGES 13

www.sciencedirect.com

Please cite this article in press as: Cabrita LD, et al. Protein folding on the ribosome. Curr Opin Struct Biol (2010). doi:10.1016/j.sbi.2010.01.005
PTC may be mediated by relatively subtle interactions. This conclusion is supported by simulation studies that suggest that a global rigidity is associated with the ribosomal tunnel [35\*] despite the presence of channels within the complex that could facilitate the flux of ions or water molecules [24]. Whether or not the ribosome is capable of more elaborate conformational changes, perhaps transiently under some conditions, is a question that will undoubtedly be answered more definitively as further studies are carried out.

Despite its apparent rigidity, it is increasingly clear, however, that the ribosomal exit tunnel is not a completely passive conduit, though whether or not it has the capacity to support, or even promote, a significant degree of folding by the nascent chain has been the subject of considerable speculation. Several studies using selective proteolysis [36,37] have indicated that the tunnel can accommodate 30–40 residues, a larger number than that expected for a fully extended polypeptide chain. A propensity for the nascent chain to adopt some preferred structural elements is supported by the recent cryoEM structures of translating ribosomes [27\*,34\*], which have revealed that the conformations of two nascent chains, each distinct in sequence remain superimposable within much of the tunnel. This finding may in part relate to the limited dimensions characteristic of much of the length of the tunnel, which may preclude higher orders of folding, although biophysical studies using chemical modification indicate that ‘folding zones’ exist within the tunnel [38] separated by constrictions generated by the ribosomal proteins L4 and L22.

The propensity for the nascent chain to adopt a significant degree of structure may also, however, be intrinsic to certain sequences; nascent chains of transmembrane-derived sequences, for example, are thought to be capable of adopting a compact, possibly α-helical structure in the tunnel [39]. The density observed in the cryoEM studies of nascent chains at the exit port (which is located approximately 80 Å from the PTC) (Figure 2), however, reveal that there is no longer a superposition of the nascent chain structure, but rather, each had adopted a preferred conformation. This finding suggests that some distinct, ‘low order’ structural preferences may be sampled by different nascent chains, a conclusion that is in agreement with recent cross-linking studies that have suggested that simple units of structure, such as α and β hairpins, can form in the exit port [40\*]. It is also consistent with previous cryoEM studies [41] that have suggested that various degrees of additional structure might form within certain regions of the tunnel.

The physical dimensions and apparent global rigidity of the ribosomal exit tunnel, in line with the experimental observations described above, suggest that major conformation resulting in the adoption of native-like structure by the nascent chain can only occur outside the ribosomal tunnel, though not necessarily before complete release from the associated auxiliary proteins, as discussed below. There is little information, however, about the structural preferences of ribosome-bound nascent chains (RNCs) that have passed through the tunnel. One approach that can uniquely provide high-resolution structural information on dynamic systems is NMR spectroscopy. Indeed, analysis of NMR spectra has provided unambiguous evidence for the existence of very highly dynamic regions of intact ribosomes and their subunits, notably the L7/L12 stalk region [42,43], and recent studies have revealed that RNCs can be similarly dynamic [44,45**]. This conclusion is consistent with fluorescence anisotropy data from studies of apomyoglobin [46\*] that indicate that, as the length of a nascent chains increases, the overall motional correlation time decreases as the probe emerges from the tunnel.

NMR data have also revealed that the emerging nascent chain has the propensity to adopt partially folded structures [45**] and, as a consequence of its incompletely folded nature, it will be prone to self-association in the crowded cellular environment. A recent cryo-electron tomography reconstruction of polysomes (Figure 3) has shown that initial conformation of the nascent chain emerging from the tunnel, which may be sampled by different nascent chains, is likely to be partially folded, consistent with recent cryoEM studies [27\*,34\*]. The emerging nascent chain is likely to adopt a partially folded structure, consistent with recent cryoEM studies [27\*,34\*]. The emerging nascent chain is likely to adopt a partially folded structure, consistent with recent cryoEM studies [27\*,34\*].

**Figure 3**

A model of one class of polysome organisation (top) as determined by cryo-electron tomography reconstruction techniques together with possible representative conformations of emerging nascent chains (green and red), where the ribosomes (30S in yellow, 50S in blue) are positioned in a pseudo-planar position and the exit tunnels face outwards, with each emerging nascent chain therefore being less prone to interacting with a neighbouring one. A tomograph of firefly luciferase (Luc550) (bottom left) was used to reconstruct the model of the ribosomes in a pseudo-planar position (bottom middle) and the putative mRNA pathway through the polysomes is also shown (bottom right). Data taken from [47**].
revealed, however, that multiple ribosomes bound to a single mRNA transcript are arranged in a staggered or pseudo-helical manner with the ribosomal exit tunnels facing outwards from the complex, which in turn maximises the distances between nascent chains limiting any unfavourable interactions between them [47**]. Further details of the conformational nature of the nascent chain outside the tunnel have, however, remained elusive to both X-ray crystallography and cryoEM. Engineering the nascent chain to incorporate highly stabilising motifs, or the binding of auxiliary factors may, however, introduce sufficient rigidity to enable RNCs to be observed by these techniques.

Characterising the nature and properties of RNCs in any detail presents a significant challenge for the methods of structural and cellular biology. Typically during protein synthesis, the existence of a stop codon within the mRNA sequence signals the end of translation and enables factors to be recruited that allow the nascent chain to be released from the ribosome. One strategy for defining the structural and dynamical properties of a nascent chain during co-translational folding involves the study of a series of constructs that are designed to mimic the progressive emergence of the chain through the tunnel, and hence to enable ‘snapshots’ to be taken of the elongation process. Generating RNCs for such studies requires translation to be arrested artificially; one method that is used to achieve this objective involves engineering truncated DNA constructs which lack a terminal stop codon. This approach enables the ribosome to retain the nascent chain of interest until the transfer messenger RNA (tRNA) surveillance mechanism initiates mRNA decay or stimulates the release of the nascent chain by trans-translation [48]. An attractive alternative strategy for RNC generation involves the incorporation of a motif derived from the secretion monitor protein (SecM) [28], a procedure that results in the nascent chain being retained on the ribosome as discussed above.

One approach to probing the relatively small numbers (tens to hundreds) of residues in a nascent chain in the presence of a total of some 7500 residues contained in the proteins that make up the ribosome itself, involves some type of selective labelling. Applications of this general approach have included the incorporation into the nascent chain of radioisotopes, non-natural amino acids and of stable isotopes for a range of biochemical and biophysical studies [39,46*,49]. The use of an in vitro transcription–translation (cell-free system) widely employed for the generation of RNCs, enables selective labelling to be carried out relatively easily; for example such a system can be supplied with $^{15}$C or $^{15}$N labelled amino acids that are then incorporated in the growing chain and enables its detection by NMR spectroscopy [44]. Cell-free extracts can be prepared with particular efficiency from E. coli, strains that are devoid of ribonucleases (e.g. MRE600); in addition, strains can be obtained that are, for example, enriched with rare codons (e.g. BL21-pRIPL) or lacking specific ribosomal or cellular components, enabling the effects of such components to be investigated.

More recent advances in the methods for the generation of RNCs have included the development of the PURE system [50], a reconstituted protein translation system, which contains the minimal set of components required for translation. It has been used to particular effect for investigating the study of the interactions between nascent chains and molecular chaperones, such as TF [51] and GroEL/ES [52,53], and has been coupled with the quartz-crystal microbalance technique (QCM) to examine protein synthesis through changes in the mass of the ribosome, enabling for example, the significance of the kinetics of the formation of the initiation complex in the overall rate of protein synthesis to be examined [54*]. The ability to remove components selectively from the PURE system (e.g. the ribosomal protein S1 [55]) not only facilitates the study of the role of such cellular factors, but also holds further promise for detailed selective labelling strategies for biophysical and structural studies.

The recent introduction of in vivo methods for the generation of RNCs [45**,56,57*] provides an exciting extension to studying protein folding in the natural cellular environment. In this approach, cells are stimulated to grow to high cell densities in an unlabelled medium and then the cells are transferred into an isotopically labelled medium where expression is induced and RNCs are generated using the SecM translation-arrest motif at the C-terminus of the nascent chain [45**]. Together, the dual strategy enables the production of selectively labelled nascent chains bound to isotopically silent ribosomes (Figure 4). When combined with advances in imaging [58**] and NMR spectroscopy [59–61], the in vivo generation of RNCs is also a significant step forward towards examining the folding process in real time.

Biochemical and biophysical studies of co-translational folding

Once appropriate samples of RNCs can be prepared, the key questions concern the ways in which their structural and dynamical properties can best be described. Many types of biochemical and biophysical techniques have been applied for this process and we discuss here a range of examples. For example, two-dimensional SDS-PAGE studies of RNCs of influenza haemagglutinin [62] have shown that the latter can form disulfide bonds and undergo modifications including N-linked glycosylation and glycan trimming whilst attached to the ribosome, features that are prerequisites for the RNCs to interact with the ER-associated molecular chaperone, calnexin. The formation of disulfide bonds in a co-translational manner has also been observed for nascent chains of the
HIV-1 glycoprotein, gp160 [63], and pulse-chase experiments have shown that manipulating the extent to which such disulfide bond formation takes place has a very significant impact on both the maturation and secretion of the protein.

Studies which have assessed the progressive acquisition of native-like structure and biological activity have provided particularly compelling evidence about the extent of folding that can occur co-translationally. Thus enzymatic studies of the RNCs of alphavirus capsid protein [64] reveal the co-translational formation of its N-terminal domain, together with the activation of the intrinsic autocatalytic activity required for the maturation of the full-length protein, when the C-terminus of the domain is separated by 43 amino acid residues from the PTC, just above the number of residues needed to span the exit tunnel. Consistent with these observations are studies that have exploited limited proteolysis to demonstrate the sequential folding of a two-domain synthetic fusion of human Ras protein and mouse dihydrofolate reductase (hRas-DHFR) [4]. Limited proteolysis has also been used, in concert with C-terminal truncations, to examine the five-domain cystic fibrosis transmembrane conductance regulator (CFTR) [65], and has revealed that each domain acquires native-like structure as it emerges sequentially from the ribosome, and that the complete folding of the protein occurs in a domain-by-domain, modular fashion.

Monoclonal antibodies have also been used to probe emerging nascent chains. In particular, in the cases of RNCs of rhodanese, chloramphenicol acetyltransferase (CA) and the MS2 viral coat protein [66]. It was found that 60, 85 and 44 residues, respectively, had to be synthesised in order to allow antibody recognition of an N-terminal fluorescence probe attached to nascent chains, suggesting that there is variability in the length of nascent chain required for emergence from the tunnel. With globin-RNCs, the synthesis of just the N-terminal 86 residues (out of the entire 140 amino acid sequence) has been found to be required for heme binding [67]. Meanwhile biological activity has been observed to develop in rhodanese [68], firefly luciferase [69] and the green fluorescence protein (GFP) [70], when the full-length nascent chains of these proteins were extended at their C-termini, by the addition of 23, 26 and 31 amino acids

Figure 4

The production of selectively isotopically labelled RNCs in vivo, involves a two-step process, where the E. coli cells are first stimulated to grow to high cell densities in an unlabelled medium (a), and then transferred into a isotopically labelled medium for expression (b). The purification of the RNCs from E. coli cells (c) following lysis involves the isolation of the RNCs from the E. coli lysate using affinity chromatography, which exploits the N-terminal affinity tag (typically a 6×His tag), followed by further purification using sucrose gradient ultracentrifugation. The RNCs can be detected by SDS-PAGE and immunoblotting (e.g. anti-6×His). Data taken from [45**].
respectively. Importantly, in these cases the additional amino acids just cover the length of the ribosomal exit tunnel, and although the minimum sequence requirements for biological activity have not yet been defined, the results show clearly that each of these proteins also has the capacity to fold in the near vicinity of the ribosome.

A variety of studies has therefore demonstrated unequivocally that nascent chains can indeed form structure and acquire activity whilst attached to the ribosome. By contrast, only a few studies have been carried out to probe specific differences between in vivo and in vitro folding [71,72,73,74], to discuss, for example, whether isolated polypeptide chains have a greater propensity to misfold in the absence of the structure that is formed co-translationally. Two studies provide the clearest evidence for the significance of co-translational folding in this respect and are therefore described here in some detail. Firstly, the homotrimERIC P22 tailspike protein (TS) has been found [72] to be able to acquire native-like structure whilst bound to the ribosome. More recent experiments have examined SecM-arrested C-terminal deletions of TS-RNCs [73] containing clusters of rare codons at naturally occurring domain boundaries corresponding to the N-terminal domain (the first 222 residues) and then to progressively increasing lengths of its C-terminal β-helix and interdigitated domains. A panel of conformation-dependent antibodies, that together are able to recognize the complete sequence of the C-terminal domains in each case, were shown to be sensitive probes of acquisition of folded structure; these were then found to recognize the RNCs indicating that the polypeptide chains had folded before their release. In addition, the number of antibodies recognised increased with the growing nascent chain, demonstrating that the TS structure is acquired in a progressive manner. This result is consistent with the sequential domain folding inferred in the studies of hRas-DHFR and CFTR (discussed above). The analogous truncated and isolated fragments were all shown to be aggregated in vitro and although some of these aggregates could be solubilised and renatured to a monomeric protein, none of the fragments recognized the equivalent antibodies. A similar trend has been seen in limited proteolysis experiments, which indicate that the nascent chains that had folded before release showed the presence of a protease-resistant domain but that the analogous truncated and isolated fragments that had been refolded in vitro did not. These data highlight the fact that different conformations can be sampled by nascent chains synthesised in vivo compared to the analogous polypeptides refolded in vitro, and that the vectorial emergence of a growing nascent chain appears to enable it to sample native like conformations that enhance further the probability of correct folding.

Further important information concerning the stepwise folding of multi-domain proteins has come from the study of RNCs of the 62 kDa protein firefly luciferase [71]. An N-terminal, 190 residue domain of this RNC was found by limited proteolysis methods to adopt native-like structure that is analogous to that observed in vitro in denaturation studies of the full-length protein. When isolated, luciferase was denatured and then renatured in a rabbit reticulocyte lysate containing molecular chaperones, this native-like domain was not observed. This result is consistent with that discussed above for P22 tailspike, where inherent differences appear to exist between the in vivo or in vitro folding of a nascent chain. The rate of refolding of luciferase is also accelerated by chaperones, as biophysical studies [74] indicate that luciferase synthesised in vivo acquires its biological activity within seconds, in contrast to in vitro refolded luciferase renatured in vitro that requires minutes to acquire its full native activity. Taken together, these studies suggest that the combination of vectorial emergence, together with the action of chaperones, leads to more favourable rates of folding of at least some polypeptide chains, presumably reducing the probability of the formation of long-lived partially folded intermediates that could be vulnerable to rapid degradation or aggregation.

A number of studies has determined clearly that there is an effect of translational speed on the folding of nascent chains. This phenomenon is associated with codon usage, which gives rise to a discontinuous translational rate that appears to be an important feature of in vivo folding [75,76–78]. The distribution of rare codons within a sequence can, therefore, moderate the speed of translation and, as a result, can play a significant role in the efficiency of folding; the nascent chain is presumably able to fold more efficiently if it has more time to develop native-like structure. This effect has been seen in studies of CA where the replacement of rare for synonymous codons was found to result in a 20% loss in the activity of the protein produced by recombinant methods [79]. A similar result has been found for the E. coli derived protein Sulf, where substitutions in two codons were found to result in formation of a relatively stable folding intermediate and to decrease the folding efficiency of the protein [80]. More striking is the case of the multidrug resistance 1 (MDR1) gene, where conformation-dependent antibodies and limited proteolysis indicate that rare codon substitutions give rise to differences in the detailed structure of translated proteins and to variation in the functional specificity [81].

Although such a wealth of biophysical and biochemical data has advanced greatly our general knowledge of the potential significance of co-translational folding, elucidating the molecular details of the conformational preferences populated by growing nascent chains is undoubtedly a key objective in order to gain detailed insight into the in vivo folding process. As mentioned above, recent evidence concerning the structures of nascent chains inside the exit tunnel has recently been obtained by cryoEM, although
such information has so far eluded X-ray crystallography, presumably because of the dynamic nature of the species involved. Outside the tunnel, regions of the nascent chain are clearly highly flexible; for example, rotational correlation times in the order of \( \text{ca.} 5 \text{ ns} \) have been determined for RNCs of apomyoglobin using fluorescence anisotropy measurements [46,8], and values of the same order of magnitude have been estimated for RNCs of ddFLN (ca. 15 ns) by NMR [44]. In each case these values are comparable to those of the folded protein free in aqueous solution. Such extensive motional freedom underpins the recent development of solution NMR spectroscopy for the study of RNCs [44,45,82,83] by reducing the spectral linewidths by several orders of magnitude relative to a rigid system the size of the ribosome. NMR methods represent powerful tools for the study of early folding events, because of their unique ability to report simultaneously on the structure and dynamical properties of proteins in terms of ensembles of conformations arising from highly disordered states [84,85].

Multi-dimensional NMR studies of RNCs have been carried out recently on a two-domain RNC construct (ddFLN) derived from the 120 kDa F-actin cross-linking gelation factor from *Dictyostelium discoideum*, a large family of proteins that organises filamentous actin in networks and stress fibres [86,87]. This construct contains a folding competent immunoglobulin domain (Dom5) fused to a

---

**Figure 5**

The progressive acquisition of structure during co-translational folding has been probed using NMR spectroscopy for the two-domain (Dom5 and 6) immunoglobulin-like protein, ddFLN. As Dom 5, corresponding to the 105 N-terminal residues of ddFLN begins to emerge from the ribosomal exit tunnel, it can adopt a partially folded state (a) as characterised by the presence of both overlapping and well-dispersed resonances in the \(^1\text{H}\) dimension in both the \(^{15}\text{N}–^1\text{H}\) (top) and \(^{13}\text{C}–^1\text{H}\) (lower) correlation spectra. When tethered to the PTC by the folding incompetent Dom6 (89 residues), the ribosome-bound Dom5 can adopt a native-like structure (b) equivalent in conformation to full-length, released Dom 5 after release from the ribosome (c). Data taken from [45**].
second but folding incompetent immunoglobulin domain (Dom6). The preparation of samples suitable for NMR (see above) was a major challenge in terms of the large quantities of ribosomes required to obtain spectra with sufficient signal-to-noise ratios, and the need for advanced methods of accumulating suitable NMR data. Nevertheless, high quality correlation spectra of selectively and uniformly isotopically $^{15}$N/$^1$C labelled ribosome-bound dDFLN were acquired and have revealed a wealth of residue-specific structural and dynamical detail. Analysis of the main chain $^{15}$N resonances from this RNC shows that Dom5 could acquire an overall fold, essentially identical to that of the native state of isolated Dom5. Analysis of the side chains of the RNC using $^1$C detection shows that these too are analogous to those of the native, isolated protein [82**]. There is, however, clear evidence for dynamical behaviour within the folded regions of the nascent chains that differs from that of isolated proteins and can be attributed to interactions with the ribosome. Interestingly, different types of perturbations to motional properties were found for the main chain [44] and side chain [82**] groups of the RNCs.

In order to initiate the study of the process of acquisition of structure during translation, a truncated variant of dDFLN, containing just the sequence of Dom5 along with the SecM arrest sequence was made and analysed by NMR. The results suggest [45**] that this RNC is able to adopt a partially folded state, with clear elements of native-like structure (Figure 5), before the complete emergence of the Dom5 sequence. The observation of such a co-translational folding intermediate provides a strong indication that it will be possible to observe the progressive development of structure as nascent chains emerge from the ribosomal tunnel. Furthermore, the use of newly developed NMR experiments, that have been shown to be capable of characterising large protein complexes such as the preosome [88,89], should enable the characterisation of even those regions of nascent chains whose dynamics are greatly limited by being constrained within the tunnel or within the vicinity of the exit port, as well by interactions with auxiliary factors. Of major importance in the context of all the NMR experiments are rapid developments in the use of NMR-derived structural restraints in molecular dynamics simulations that enable structural ensembles to be defined. Indeed, recent demonstrations that such ensembles can be determined from chemical shifts [90,91] without the need to carry out the extensive range of experiments needed in conventional structure determination procedures, has the potential to transform our understanding of the development of structure during folding.

The role of auxiliary factors in co-translational folding

As the nascent chain emerges from the ribosomal exit port, it has the opportunity to interact with auxiliary factors that either assist with folding, e.g. molecular chaperones [3], contribute to co-translational or post-translational modification, e.g. peptide deformylase [92*], or facilitate transport across the cell membrane, e.g. the signal recognition particle. The first chaperone that is thought to interact with the nascent chain is the 48 kDa TF, mentioned earlier in the article, which docks to the L23 protein located at the ribosomal exit tunnel (Figure 2). In the absence of a nascent chain, TF cycles on and off the ribosome with a mean residence time of 11–15 s [93*,94], but during translation the affinity of TF for the ribosome is increased ca. 30-fold [93*,94,95]. Both cryoEM [96**] and X-ray crystallographic [26] studies have revealed that TF undergoes a conformational change, forming a protective cavity (that has been described as a cradle [26]) for the folding of the nascent chain.

The existence of a cavity presumably enables enough of a polypeptide chain to emerge from the ribosome for at least a significant degree of folding to be completed. TF appears to mediate its function by scanning rapidly any exposed hydrophobic segments of the nascent chain, remaining bound to these regions even following its dissociation from the ribosome [51,94,96**]. In addition, TF is also able to reassociate with a given nascent chain under some conditions [93*,94], a characteristic that is likely to be important for multi-domain proteins; indeed, several such proteins have been found to fold more efficiently in the presence of TF, but at the expense of speed [97]. Experiments indicate that TF can stay associated with a released nascent chain for more than 30 s [94], and its eventual dissociation is then likely to occur when previously exposed hydrophobic patches become buried as the nascent chain folds. Following its detachment from TF, the partially folded nascent chain can then interact downstream with other chaperones, such as DnaK/J (Hsp70) and GroEL/GroES (E. coli) [4,97–99].

Another important facet of folding in vivo is the existence within cells of a potent sorting mechanism that enables the correct subcellular compartmentalization of any given polypeptide chain. In some instances, for example with secretory proteins destined for the ER, this translocation occurs in a co-translational manner and the sorting mechanism is made possible, in part, by an interplay at an early stage of nascent chain synthesis between TF and the signal recognition particle (SRP) [100–103]. The SRP, a multimeric protein complex, serves to shuttle the nascent chains of proteins destined for secretion, to the SRP receptor; the SRP binds strongly to the transmembrane sequences [101,104] and can readily displace TF. The SRP:RNC complex is then recognised by the SRP receptor and is subsequently transferred to the protein conducting channel (PCC) [105], is a heterotrimetric integral membrane protein complex, SecYEG (in eubacteria) and SecY/Sec61βγ (in eukaryotes), to
10 Folding and binding

enable co-translational translocation of the nascent chain to occur for downstream processing and folding [106].

Recent cryoEM studies of the SecY/Sec61 complex with a translating ribosome [34**], have revealed the tantalising evidence that the PCC behaves as an extension to the ribosomal exit tunnel for a protein destined for the ER [105,107–109] and as the dimensions of the PCC (10–20 Å [110]) are similar to those of the tunnel itself, extensive folding before the emergence of the protein into the outer membrane appears unlikely. The emergence of the nascent chain from the PCC in a co-translational manner raises an additional question of how and when folding takes place in this context. The process of co-translational translocation and folding thus represents an additional frontier in our efforts to understand the behaviour of newly synthesised protein molecules, in this case particularly for those systems requiring post-translational modifications to complete the folding process.

Concluding remarks

Studies of protein folding seek to establish the manner in which the acquisition of the native states of these ubiquitous molecules occurs within living systems. It is axiomatic that the underlying principles of folding in vitro and in vivo are the same; the probability that our understanding of how and when folding takes place in this context. The process of co-translational translocation and folding thus represents an additional frontier in our efforts to understand the behaviour of newly synthesised protein molecules, in this case particularly for those systems requiring post-translational modifications to complete the folding process.

In this article, we have discussed the current strands of activity designed to probe the details of folding in the cell, particularly from the standpoint of events that occur as the nascent chain is synthesised and emerges from the ribosome. Whilst much of great interest has been gained from a combination of biochemical and biophysical approaches, we believe that the opportunities that are beginning to be evident from the application of the methods of NMR spectroscopy in conjunction with the established techniques of EM and X-ray crystallography, are of exceptional interest. These approaches have, at least in principle, the capability of probing the structural and dynamical properties of RNCs at the level of individual residues, even in the presence of the extensive dynamical fluctuations that are involved in the transition of a disordered polypeptide chain into a fully structured protein. They therefore promise to provide unprecedented insight into one of the most fundamental steps in the translation of genetic information into biological activity, a process that must occur both efficiently and reliably in every living system.

Acknowledgments

The authors would like to thank members of their laboratories for useful discussions and Luke Goodsell for assistance with the figures. We acknowledge support from the Wellcome and Leverhulme Trusts (to C.M.D. and J.C); HFSP Young Investigators Award HFSP7/07 (J.C.); New Investigators Award, BB/SC/BRG0156511 (J.C.). L.D.G. is a NHRMRC C.J. Martin Research Fellow.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


An excellent review covering our current understanding of protein folding and misfolding highlighting the important role played by molecular chaperones.


An examination of the pre-translocation ribosome complex using single-molecule FRET, which shows the formation of tRNA hybrid states and the closure of L1 stalk.


A mutational analysis which shows that the decoding centre of the 3OS subunit participates in the pre-translocation ribosome complex, focussing on the interaction that is made with EF-G.


Using single-molecule fluorescence, this study follows the real-time dynamics of the L1 stalk region and its movements relative to the large subunit in the pre-translocation complex.


An excellent review of the structural data relating to the dynamics of the ribosomal machinery and draws together biochemical and biophysical information relevant to our current understanding of the decoding and translocation steps of protein elongation.


A remarkable cryoEM structure of a ribosome engaged in translocation and highlighting the phenomenon of ratcheting and the hybrid states of the tRNA.


An excellent review of our current understanding of decoding and elongation, which draws together the current information from cryoEM and X-ray crystallographic studies.


An excellent review focussing primarily on structural evidence concerning ribosome function.


A cryoEM study that provides an unprecedented insight into the structural nature of the nascent chain during translation-arrest.


A comprehensive review which examines the known inhibitors of protein synthesis and their interaction with the ribosome.


A computational study examining the dynamics of the ribosomal exit tunnel and the capacity for it to support co-translational folding.


An elegant study which illustrates that the ribosomal exit port has the capacity to support simple units of structure.


A study which makes use of fluorescence anisotropy to examine the dynamics of ribosome-bound nascent chains through the calculation of rotational correlation times.

12 Folding and binding

The use of 3D cryo-electron tomography to provide a high-resolution structure of a polysome and that suggests a potential mechanism for avoiding aggregation in vivo during protein synthesis.


59. A method that describes the use of single-molecule fluorescence to examine the kinetics of real-time folding of nascent chains of GFP as they emerge from the ribosome. FEBS J 2005, 262:281-291.

60. The use of single-molecule fluorescence to examine the kinetics of real-time folding of nascent chains of GFP as they emerge from the ribosome. FEBS J 2005, 262:281-291.


74. The use of optical tweezers to provide insight into the process of translation, showing three distinct rates for the translocation.


76. The impact of translational speed is explored using computational and experimental approaches to examine the E. coli protein, Suf I.


80. The impact of translational speed is explored using computational and experimental approaches to examine the E. coli protein, Suf I.

81. The impact of translational speed is explored using computational and experimental approaches to examine the E. coli protein, Suf I.
A study showing that $^{13}$C detection provides a highly sensitive means of deriving structural and dynamical information through NMR analysis of ribosome-bound nascent chains.

A detailed structural and dynamical characterization of ddFLN identifying the existence of a previously unknown in vitro protein folding intermediate.

A study which examines approaches to generating structures chemical dynamics and folding of an immunoglobulin domain using methyl NMR spectroscopy to examine high molecular weight proteins.


A cryoEM study of trigger factor bound to ribosomes demonstrating that trigger factor forms a protective environment for the nascent chain at the ribosome exit.


