NMR snapshots of nascent chains emerging from the ribosome during biosynthesis

Sammy H.S. Chan[1], Christopher A. Waudby[1‡], and John Christodoulou[1‡]

[1]Institute of Structural & Molecular Biology, Department of Structural & Molecular Biology, University College London, Gower Street, London WC1E 6BT, United Kingdom and Department of Biological Sciences, Birkbeck College, Malet Street, London WC1E 7HX, United Kingdom

[‡]To whom correspondence should be addressed (c.waudby@ucl.ac.uk & j.christodoulou@ucl.ac.uk)

NMR spectroscopy is a powerful tool to study the dynamic process of how proteins acquire their biological fold at near-atomic resolution. Advances in spectroscopic and biochemical approaches have now expanded the ability of NMR methods to probe increasingly larger biological machinery including the 2.4-MDa 70S E. coli ribosome, where polypeptide chains are assembled. Detailed structures of the ribosome have been determined by X-ray crystallography and cryo-electron microscopy studies that describe the rigid ribosomal core and exit tunnel that the nascent chain must pass during its biosynthesis. However, most folding occurs when the nascent chain has emerged beyond the exit tunnel and it is here where NMR spectroscopy is uniquely able to provide high-resolution structural insights into on its co-translational folding. In this review, we will discuss recent progress in NMR studies of ribosome-bound nascent chain complexes, highlighting the emerging role of the ribosome and how these results are shaping our understanding of co-translational protein folding in the cell.

1. Introduction

The ribosome is the molecular machinery responsible for the production of all protein molecules. Beyond this basic function as the site of peptide synthesis though, it is increasingly clear that the ribosome also plays an active role in the regulation of protein synthesis and in the protein quality control network both in prokaryotic and eukaryotic cells.[1,2] The conformation of the ribosome is under continuous change throughout the different steps of translation[3] during which a multitude of translation factors can interact. Given the large molecular weight of the ribosome (approximately 2.4 MDa for 70S ribosomes and 4.3 MDa for the 80S human ribosome), residues within the core ribosomal particle are too slowly tumbling to be observable by solution-state NMR spectroscopy. Instead, the power of NMR spectroscopy is its unique ability to probe flexible regions on the surface of the particle, and in particular, to observe resonances from the emerging nascent polypeptide chain (NC), the ribosome's biosynthetic product. Such observations are highly complementary to other structural biology techniques in which flexible regions cannot be resolved. NMR studies of ribosomal particles have therefore enabled structural and dynamical insights into polypeptide chains on the ribosome during their earliest stages of biosynthesis.

1.1 De novo protein folding in the cell

The covalent linkage of amino acids to form a nascent polypeptide chain occurs at peptidyl transferase centre (PTC) at the core of the ribosome particle. Here the NC remains covalently tethered to tRNA molecules throughout the biosynthetic process until its released by termination factors. The order in which amino acids are added, i.e. its primary sequence, is the only information required for NCs to adopt their biologically active three-dimensional structure.[4] Folding therefore occurs spontaneously and, for small domains, can occur very rapidly, on the timescale of milliseconds to seconds.[5] However, for more complex proteins such as multi-domain proteins, accounting for over 70% of the eukaryotic proteome,[6] the incremental addition of each amino acid to the growing NC creates an exponential increase in conformational space that must be sampled by the NC (Figure 1) and presents a significant challenge for the folding process. There is therefore a high risk of forming non-native inter-domain contacts that could potentially lead to misfolded states[7] (magenta trajectory in Figure 1), implicated in a whole host of human diseases.[8] Co-translational folding is a means by

which the very large number of available structural conformations is restricted through sequential folding of (sub)domains to bias folding towards the native state and away from such hazardous processes. Folding can occur before complete synthesis and release from the ribosome because elongation proceeds at a more modest rate of approximately 20 amino acids per second in prokaryotes.[9] Indeed, the relative rate of elongation has evolved to provide kinetic fine-tuning of the folding process (Figure 1), evidenced by a generally slower translation rate in eukaryotes (approximately 6 amino acids per second [9]), and non-uniform translation rates generated by a variety of different cellular factors such as mRNA structure and tRNA concentrations.[10] Cells also possess a complex network of molecular chaperones, including the ribosome-asssociated trigger factor,[11] to promote efficient folding and ensure protein homeostasis.[11]

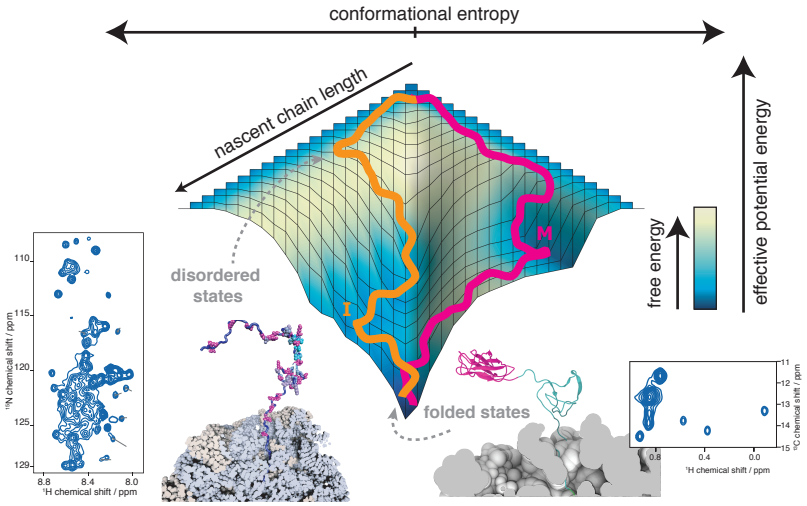

Figure 1. Co-translational folding free energy landscape. The conformational entropy of the NC (represented by the width of the surface) and its effective potential energy (depth of the landscape) continuously changes as it emerges from the ribosome with increasing polypeptide length during biosynthesis. The balance between favourable conformational entropy and potential energy is reflected by the total free energy at each point. The route taken along the energy landscape by the NC is dependent on the relative rates of folding and translation. Two trajectories are shown: when folding is rapid relative to translation (orange), folding proceeds via an intermediate free energy minimum (I), while slower folding relative to translation (magenta) results in a kinetically trapped, misfolded intermediate (M). NMR spectroscopy is able to probe the structure and dynamics of the NC, in both its disordered (such as the α-synuclein RNC,[34] bottom left) and folded (such as the FLN5+110 RNC,[33] bottom right) state, during co-translational folding. Figure adapted from reference.[1]

In vivo NCs do not fold in the dilute buffered solutions typically used for protein folding studies but rather in the densely crowded cytosolic environment where there may be many specific and non-specific interactions between the NC and other cellular factors.[12] These include ribosome-associated factors recruited to NCs at different stages of its biosynthesis for a variety of processes including protein quality control,[13] targeting and translocation,[14,15] and modifications.[16] Within this highly heterogeneous environment, other proteins, nucleic acids and biomolecules contribute to a dense macromolecular concentration of up to 400 g/L,[17] imposing crowding effects and physically constraining the structure and mobility of proteins.[18,19] Due to these processes, the energy landscape of a NC is continuously remodelled during its biosynthesis. Only by determining how each of these factors impact the emerging NC will we improve our understanding of *de novo* folding in the cell and the mechanisms by which misfolding can be avoided.

1.2 Studying co-translational protein folding by NMR spectroscopy

Ongoing developments in physical methods and preparative biochemistry have enabled increasingly detailed structures of the ribosome to be determined during different stages of translation using both X-ray crystallography and cryo-electron microscopy studies,[3,20,21] the latter undergoing a recent 'resolution revolution' stemming from technological advances in direct electron detectors and image-processing

software.[22] These methods have successfully reported on the structure of NCs with secondary structure[23-25] and simple folds[26] within the exit tunnel and vestibule. However, emergence of the NC from the confined environment of the ribosomal exit tunnel (having a diameter between 10 Å at its narrowest and 20 Å at its widest points[27]) is required to enable the acquisition of complex tertiary structure. Because of the large gain in flexibility beyond the exit tunnel, these important states have been unable to be resolved by these structural techniques. In contrast, NMR spectroscopy is uniquely positioned to characterise these dynamic NCs in solution with near-atomic resolution, having also undergone recent improvements in sensitivity and resolution by, for example, availability of high-field spectrometers with cryogenic probes and advanced isotopic labelling schemes combined with pulse sequences that have pushed the size limitation of NMR studies to increasingly large biological systems.[28-30]

Initial studies of the flexible bL12 stalk region of the 70S E. coli ribosome,[31] a key region required for the recruitment of elongation factors, demonstrated the capacity of NMR spectroscopy to study selectively and in detail flexible regions of the ribosome, despite the overall slow tumbling of the 2.4 MD 70S particle. These experiments paved the way for a series of NMR studies of translation-arrested ribosome-nascent chain complexes (RNCs)[32-38]: complexes in which translation has been arrested at a defined point in the NC sequence, allowing acquisition of an NMR 'snapshot' of folding at a particular NC length. Homogenous samples of RNCs can now be generated in vitro or in vivo with a variety of isotopic labelling schemes and purified at sufficiently high yields to produce high-resolution NMR spectra of exclusively the NC.[32,35] Among other applications, this has enabled chemical shifts to be determined for a pair of immunoglobulin-like filamin domains (FLN5 and FLN6) during translation, which were subsequently used as restraints in molecular dynamics (MD) simulations to determine the first structural ensemble of an RNC.[33] These models have provided key insights into the process of polypeptide structure acquisition and ways in which the ribosome might influence co-translational folding pathways. In parallel, NMR is being developed as a tool to investigate in vivo samples.[39-41] Such in-cell NMR studies are now beginning to permit an exploration into protein structure, dynamics and folding under increasingly more biologically relevant conditions.[42-46]

In this chapter, we will review the key developments in our understanding of co-translational protein folding that have arisen from recent NMR studies of RNCs and ribosomal particles, and provide an account of new developments in biochemical and spectroscopic methods that were crucial to permitting such technologically challenging experiments. We offer a perspective on future NMR studies of RNCs and its use in increasingly complex and biologically relevant analyses that are required to gain a comprehensive understanding of co-translational folding as it occurs within living cells.

2. NMR observations of the folding of isolated proteins and co-translational folding intermediate analogues

The complexity of co-translational folding within living cells necessitates a reductionist approach to delineate the many factors affecting this complex process. Isolated proteins studied in the absence of the ribosome provide an important experimental baseline from which the perturbing effect of the ribosome can be deduced in subsequent RNC studies. Importantly, isolated proteins generally provide the opportunity to acquire more complex and information-rich NMR experiments, unconstrained by the technical challenges of NMR of RNCs (see section 4).

2.1 Probing transiently populated intermediate states in folding pathways

A powerful aspect of NMR spectroscopy lies in its ability to study a wide range of states across the folding energy landscape that may interchange or exchange on timescales from picoseconds to days.[47] In particular, relaxation dispersion NMR experiments can report on sparsely populated "excited" states (populated in some cases by <1%[48]) in chemical exchange with a ground state on micro- to millisecond timescale.[49] Detailed kinetic and thermodynamic descriptions can be obtained through such measurements without the need for denaturants that may alter the protein structures[48]. Similarly, saturation transfer methods such as chemical and dark-state exchange saturation transfer (CEST and DEST, respectively) may be used to characterise excited

states in slower exchange, relying on differences in the chemical shift or transverse relaxation rate between the major and minor states.[50] Such measurements are now enabling atomic resolution structures of transient folding intermediates[51] and intermediates associated with aggregation[52,53] to be determined.

2.2 Investigating potential co-translational folding intermediates using a C-terminal truncation strategy

A key distinction between studies of the reversible folding of full-length proteins that have underpinned much of our understanding of folding mechanisms so far and de novo folding in the cell is the gradual, vectoral emergence of the NC from the ribosomal exit tunnel that enables folding to take place in the absence of its complete amino acid sequence. The co-translational folding energy landscape must therefore be considered as a series of nested landscapes corresponding to increasing chain length and larger conformational space, in which different structural states may be favoured relative to in vitro denatured, full-length protein[1,54] (Figure 1). The implications of the progressive increase in chain length during biosynthesis have been investigated experimentally using N-terminal polypeptide fragments to mimic partially emerged and synthesised NCs. Initial studies of the small single-domain proteins barnase (110 amino acids)[55] and chymotrypsin inhibitor-2 (CI2) (64 amino acids)[56] found using ^{1}H NMR, fluorescence and circular dichroism that short fragments were largely disordered, while secondary structure formed in parallel with tertiary structure at longer chain lengths. However, stable, native-like structure formed only when >90% of all residues were present.[56] While such examples highlight the ability of NCs to make significant native-like intramolecular contacts in the absence of C-terminal residues, the ribosomal exit tunnel sequesters approximately 30 amino acids.[27,57] Therefore, the late onset of folding observed here indicates that these proteins are in vivo more likely to fold post-translationally rather than co-translationally.

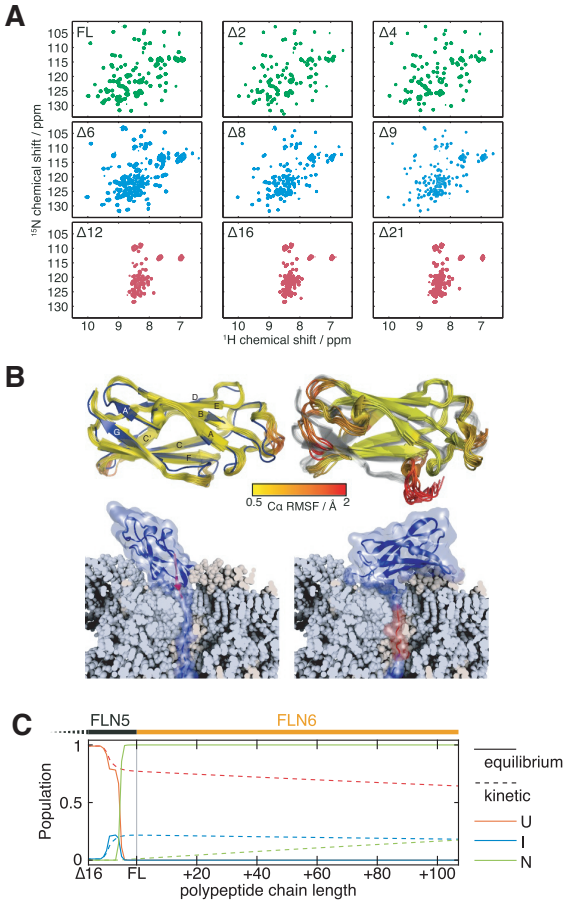

Figure 2. Mimicking co-translational folding of FLN5 using isolated C-terminal truncations. (A) ^{1}H,^{15}N HSQC spectra of progressively truncated FLN5 constructs (denoted FLN5Δ*x*, where *x* represents the extent of truncation) reveals native-like structure in full-length (FL), Δ2, Δ4 and Δ6, intermediate and unfolded structure in Δ6, Δ8 and Δ9, and only unfolded structure in Δ12, Δ16 and Δ21. (B) Ensemble structures determined by chemical shift-restrained simulations of full-length FLN5 (top left) and the Δ6 intermediate (top right), the latter retaining a native-core with a disordered C-terminal G strand. Structure-based models show these ensembles are structurally compatible with tethering to the ribosome with a 20- and 14-amino acid linker for the full-length (bottom left) and Δ6 intermediate conformations (bottom right), respectively. (C) Populations of native, unfolded and intermediate states at equilibrium and predicted under non-equilibrium conditions with a translation rate of 5 amino acids per second. Figures taken from reference.[51]

More recently, we have used a similar C-terminal truncation approach to model co-translational folding of the FLN5 filamin domain.[51] This study arguably provided a stronger basis on which to study co-translational folding effects as FLN5 is part of the larger tandem-repeat protein also known as gelation factor from *Dictyostelium discoideum*. We found that the length-dependent folding pathway of FLN5 proceeded via a stable intermediate associated with the isomerisation of a native cis-proline, with a native-like core but with the final G-strand detached and in a disordered state (Figure 2A). Backbone chemical shifts from several FLN5 fragments were used as restraints in molecular dynamics (MD) simulations to determine a structure of this putative co-translational folding intermediate (Figure 2B). Experiments using ^{15}N-relaxation dispersion NMR also identified an additional sparsely populated intermediate (0.3-1% population) associated with rapid folding of the core B-F strands. The MD results are significant in the context of co-translational folding, as accessibility and disorder of the C-terminus suggests that the intermediate is structurally compatible with tethering of the NC to the ribosome, and modelling of the closest possible approach of the intermediate tethered to the ribosome showed that only a 14-18 amino acid linker (comprising the subsequent FLN6 domain) is sterically required.[51]

2.3 Integrating co-translational folding kinetics with equilibrium measurements

The structural and thermodynamic analysis of the FLN5 folding pathway generated using the C-terminal truncation approach described above enabled an understanding of which conformational states were likely to be accessible by the NC. However, co-translational folding is a fundamentally kinetic, non-equilibrium process that is directly coupled to the rate of translation at the PTC.[10] The extent of co-translational folding therefore depends on the rate of folding relative to translation.[1] When the folding rate is relatively slow ($k_{folding} << k_{translation}$), most folding will occur only after the full-length NC has been synthesised and released from the ribosome, resulting in an essentially identical folding pathway to that observed in in vitro refolding studies. Equally, for relatively fast folding rates ($k_{folding} >> k_{translation}$), exchange between folded and unfolded states occurs rapidly and repeatedly during translation such that any co-translational effects are ultimately negated. Therefore, the intermediate regime is where folding may be most strongly modulated by co-translational effects. Through considering the rate of translation, the states along the folding pathway that are kinetically accessed (i.e. the route along the energy landscape) can thereby be identified.

To address this, the structural and thermodynamic data of FLN5 truncations were complemented by kinetic data. In particular, 2D N_z-exchange measurements[58] and real-time NMR temperature-jump measurements were used to measured folding and unfolding rates and proline isomerisation rates (of seven native-state trans prolines and one cis proline) respectively.[51] The data were finally combined with an approximate eukaryotic translation rate to construct a Markov model of the non-equilibrium co-translational folding pathway (Figure 2C). This provided a crucial way to use measurements made under equilibrium conditions by NMR and predict whether the states observed are kinetically accessed during co-translational folding. Specifically, this model predicted that only 17% of FLN5 NCs would be fully folded once the subsequent FLN6 domain has been translated, and thus posed a high risk of forming non-native inter-domain contacts. The action of PPIases was predicted to accelerate this step such that the risk of misfolding (calculated as the probability of populating adjacent unfolded domains during translation) can essentially be eliminated.[51] These experiments highlighted the need to consider the coupling of translation to folding to understand the biological implications of NMR observations made under equilibrium conditions, and also the role of auxiliary factors in the co-translational folding of NCs in the cell (discussed further in section 5.3).

It is known that within living cells, the rate of translation is non-uniform due to a host of various in vivo factors, including tRNA abundance[59], rare and synonomous codon usage,[60,61] mRNA secondary structure,[62] and electrostatic interactions within the exit tunnel.[63] The modulation of the local[64] and global[9] elongation rates has been shown to have evolved to optimise NC folding[65] and can therefore ultimately affect the final protein structure[66] , stability[67] and biological function.[68] Moreover, folding and unfolding rates may also vary during translation, being modulated by chain length and proximity to the ribosomal surface,[69,70] while competing cellular processes under kinetic control such as misfolding can further complicate the co-translational folding pathway.[71] A powerful and ambitious avenue in the future would therefore be to combine

high-resolution, equilibrium structural measurements by NMR with experimentally derived kinetic measurements of translation, folding, misfolding and binding events in real-time.[72]

3. NMR investigations of the ribosome

The ability to study dynamic regions of the ribosome using high-resolution heteronuclear NMR spectroscopy was first reported in experiments with uniformly ^{15}N-labelled 70S E. coli ribosomes.[31,73] Despite extremely slow tumbling of the 2.4-MDa ribosome particle, approximately 100 well-resolved and relatively sharp resonances could be observed in ^{1}H-^{15}N HSQC spectra (Figure 3A).[31] These signals were shown to arise from the hinge and two of the four copies of the C-terminal domains (CTD) of the mobile bL12 stalk proteins from the GTPase-associated region involved in the recruitment of translation factors. The chemical shifts of bL12 on the ribosome were found to be identical to those of free bL12, indicating that its structure was not affected by tethering to the ribosome. The ^{15}N-spin relaxation measurements showed that rotational correlation times were much shorter (13.6 ns) than expected for rigidly ribosome-bound proteins although still longer than those of free bL12, indicating that at least two of the four bL12 CTDs tumble relatively independently from the core ribosome. The flexibility of the CTD has been proposed to assist in recruiting cytosolic translation factors,[74] consistent with the observed loss of NMR signal when elongation factor G (EF-G) was titrated into the sample. This mobility has prohibited their observation in cryo-EM and crystallographic structures,[3] and therefore provides an excellent example of the complementarity of NMR with such techniques, and the unique ability of NMR to probe inherently dynamic regions of macromolecular machinery.

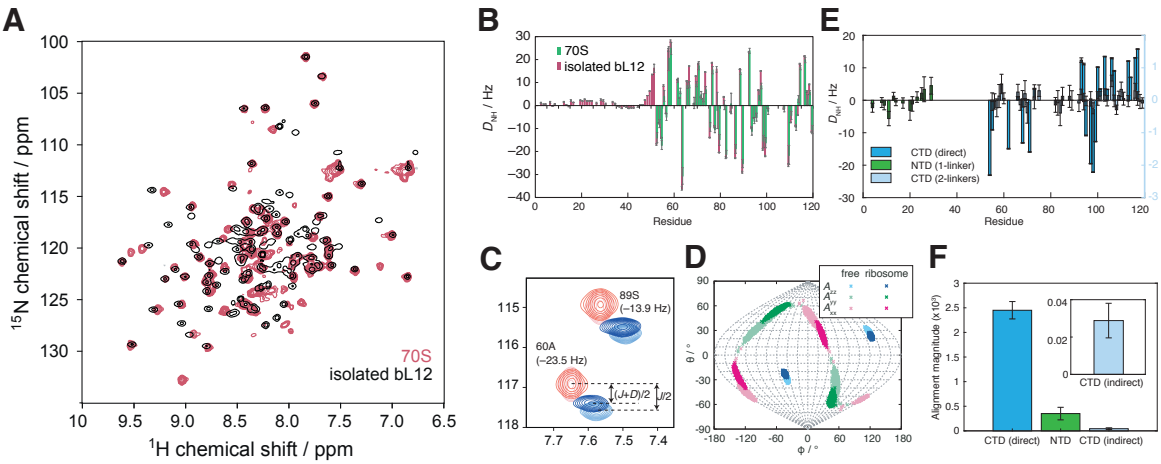

Figure 3. NMR spectroscopy of dynamic regions of the 70S E. coli ribosome. (A) Overlay of ^{1}H,^{15}N HSQC spectra of 70S ribosomes and isolated bL12 protein. Figure adapted from [31]. (B) Amide RDC measurements of ribosome-bound and isolated bL12 aligned in bacteriophage. (C) Excerpt from ^{1}H,^{15}N HSQC and TROSY spectra of 70S ribosomes in isotropic and aligned conditions. (D) Sanson-Flamsteed projection of the principal axes of the alignment tensor show differences between those calculated for ribosome-bound and isolated bL12. (E) Amide RDC measurements of isolated bL12 domains directly and indirectly aligned by mixed ^{15}N/lanthanide binding tag labelling. (F) Magnitude of fitted alignment tensors of the directly and indirectly aligned bL12 domains reveals alignment is propagated through the hinge region. Figures B-F taken from reference (Wang et al, Scientific Reports 2019, in press).

The apparent rotational correlation times measured for ribosome-bound showed large variation between different regions of the CTD,[31] suggesting that the tumbling of bL12 was anisotropic. From these data, estimations of diffusion tensors could be determined and showed a discrepancy between free and ribosome-bound CTD, implying that the ribosome imposes a preferential orientation on the CTD of bL12. These results have been corroborated in recent work (Wang et al, Scientific Reports 2019, in press), in which residual dipolar coupling (RDC) measurements of 70S ribosomes aligned in Pf1 bacteriophage were used to refine a 3D structure of the CTD (Figure 3). The refined alignment tensor was found to differ slightly from that of free bL12 (Figure 3B-D). Furthermore, alignment by a paramagnetic lanthanide-binding tag on the CTD of isolated bL12 showed propagation of alignment through the hinge region to the other bL12 domains (Figure 3E-F). This

provided direct experimental validation of a previously described structural ensemble refined using SAXS and NMR relaxation measurements[75] and suggest that EF-G is efficiently recruited to the ribosome through an anti-correlated extension and retraction mechanism of action of the two CTDs. Importantly, the ability to measure RDCs of intact ribosomes under weakly anisotropic solution conditions is an exciting technical achievement that paves the way for more detailed structural analyses of RNCs.

4. Studying ribosome-nascent chain complexes by NMR spectroscopy

The preparation of ribosomes and RNC samples for NMR studies requires a rigorous biochemical approach to obtain high yields of homogeneous complexes.[76] In the case of RNCs, ribosomes are must be uniformly programmed to arrest the translation of NCs at a specific position along its sequence and produced, typically in vivo, to maximise isotopically labelling of the NC while minimising labelling of its parent ribosome remaining unlabelled. Furthermore, the stability of RNC samples is limited, with typical lifetimes of several hours to days before the NC is released from the ribosome or the ribosomal complex itself is degraded.[77] The integrity of the complex therefore requires continuous monitoring by both spectroscopic and biochemical means to ensure that NMR signals derive exclusively from intact RNCs. In addition to this limited stability, RNC samples are also characterised by their high molecular weight (2.4 MDa) and low maximum concentrations (approximately 10 μM, equivalent to 24 mg/mL) resulting in very weak NMR signals. However, despite these spectroscopic challenges, NMR offers the unique ability to provide rich structural and dynamic information on the ribosome and NC at a residue-specific level.[33,34,36,78,79]

4.1 Biochemical preparations of RNC samples for NMR analysis

4.1.1 Translational arrest of RNCs by stalling sequences

The capacity to uniformly and selectively programme ribosomes to arrest translation of NCs relies on the use of short peptidyl motifs that interact with specific sites in the ribosomal exit tunnel during biosynthesis. A variety of stalling motifs have been identified in both prokaryotic and eukaryotic ribosomes,[80] and the recent revolution in cryo-electron microscopy has been instrumental in elucidating the mechanisms of their activity.[81-84] From these studies, a number of common features may be identified. The N-termini of stalling motifs commonly interact with the uL4 and uL22 ribosomal proteins near the constriction site (where the tunnel narrows) while additional interaction typically occur between the C-terminus and nucleotides surrounding the PTC. Specific aminoacyl-tRNAs that have slow peptide bond formation rates may also be encoded at the C-terminus to increase stalling activity.[85] These three features are all used in the translational arrest mechanism of the 17-residue motif deriving from the E coli secretion monitor protein SecM (FxxxxWIxxxxGIRAGP).[81] As with other known stalling motifs, SecM functions as a regulatory switch; in this case, tuning downstream gene expression of translocation factors through stalling-induced changes in mRNA secondary structure induced [86]. Other stalling sequences, derived from both prokaryotic proteins such as TnaC [3] and eukaryotic proteins such as XBP1u,[84] have also been used to produce RNCs that function as equilibrium snapshots of growing NCs. The use of stalled RNCs has been instrumental in biochemical,[57,87] cryo-EM,[81-84] and NMR studies[32-38,79] that have advanced our understanding of early NC structure acquisition.

The experimental stability of purified RNCs is highly dependent on the stalling efficiency of the arrest peptide, i.e. the strength of their interactions with the ribosomal tunnel. Recent studies have therefore determined the residues and contacts with ribosomal proteins that are crucial to its translational arrest and mutant versions of arrest peptides have been developed with significantly improved stalling [84,88] that have been exploited to produce RNCs sufficiently stable for cryo-EM analysis.[26] The increase in stability and sample lifetimes of RNCs with arrest-enhanced peptide sequences is likely to be extremely valuable in future NMR studies, particularly given the long acquisition times required.

4.1.2 Selective isotopic labelling of RNCs

The preparation of RNC samples for NMR analysis requires efficient (i.e. cost effective) and selective isotopic labelling. Because of the need for high yields, to limit the cost of production, and the desire to produce RNCs within their native, cellular environment, the in vitro transcription-translation systems used in early studies[79] have been largely replaced with in vivo methods. Using the well-established host E. coli, for which genetic manipulation is easy and of which cultivation times are short, culturing conditions have been developed based on auto-induction methods so that high cell density growths can be sustained to produce a large pool of unlabelled (thus NMR background-free) ribosomes.[35,76] Isotopically labelled precursors or media may then be introduced during the short induction time of protein expression and together with the high cell density conditions, this effectively limits the incorporation of isotopes to the NC only. Several rounds of purification follow cell lysis, including ultracentrifugation to select for ribosomes and affinity chromatography to an N-terminal tag (such as a hexahistidine or Strep tag) to select for only those ribosomes harbouring a NC (Figure 4).[76]

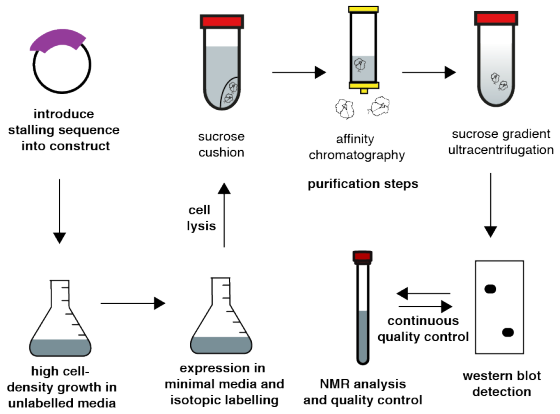

Figure 4. In vivo production of RNC samples for NMR analysis. After mutation of the stop codon to a stalling sequence, e.g. SecM, RNCs are expressed in a two-step process: firstly, bacterial cells are grown to high cell densities in unlabelled media producing a large pool of unlabelled ribosomes; then, expression in minimal media containing appropriate isotopes is induced to express and selectively label the NC. Cell lysis and purification steps follow, typically involving ultracentrifugation of sucrose cushions to pellet ribosomal particles, affinity chromatography to select for only ribosomes harbouring a NC, and a final sucrose density ultracentrifugation to select for 70S ribosomes. Detection by western blot analysis occurs both before and during NMR data acquisition as crucial means to monitor sample integrity[76] (see Figure 6 and main text).

The choice of isotopes should be carefully considered in the context of the aims of the experiment and to ensure the greatest level of observability of RNC resonances. Uniform ^{15}N-labelling provides the easiest and most cost-effective isotopic incorporation into RNCs through the use of minimal media with ^{15}N-ammonium chloride as the sole nitrogen source. It has been most successfully used in studies of RNCs with disordered conformations, such as α-synclein RNCs,[34] where the flexibility provides good sensitivity and large chemical shift dispersion in the ^{15}N dimension allows a large number of residues to be resolved. The increased cost of fractional deuteration or perdeuteration (uniform ^{2}H-labelling) combined with ^{15}N-labelling may be a factor in why this has not been utilised in RNC studies so far, but the potential reduction in transverse relaxation rates, particularly when combined with TROSY-optimised experiments[89] offers an attractive prospect to study RNCs of higher molecular weight and with greater spectroscopic sensitivity and resolution.[90]

Alternative isotopic labelling schemes may be required for the observation of folded states on the ribosome because of unfavourable relaxation properties resulting from reduced flexibility. While early studies utilised only ^{15}N or ^{13}C labelling,[31,35,36,91] in more recent studies[33] we have favoured perdeuteration combined with selective ^{13}C-labelling of sidechain methyl groups.[92] This has greatly increased the experimental sensitivity because of the increase in proton concentration (three times that of amide groups), rapid rotation about thee three-fold symmetry axis, and the use of methyl-TROSY pulse sequences that avoid mixing slow and fast relaxing coherences.[93] While there is a significant cost to producing such samples, future studies may extend $^{13}CH_3$-labelling to other sidechain methyl groups such as leucine, valine, alanine, threonine and methionine,[93] to increase the number of observable probes for folded RNCs at relatively little extra expense.

The use of other heteronuclei is yet to be explored in NMR studies of RNCs. In particular, ^{19}F-labelling is attractive because of its high sensitivity (with a gyromagnetic ratio of 83% that of ^{1}H) and sensitivity to the chemical environment due to its wide chemical shift range (approximately 400 ppm), and high chemical shift anisotropy (CSA).[94] Because ^{19}F may be incorporated into specific amino acids, and because there is no biological background, small numbers of ^{19}F resonances can be analysed by simple 1D NMR experiments,

eliminating the need for multi-dimensional experiments (and thus the √2 reduction in sensitivity with each spectral dimension). Due to these properties, ^{19}F NMR has been used to examine protein folding and unfolding in real-time,[95] large biomolecular complexes such as membrane proteins,[96] and in-cell NMR.[97] Incorporation of ^{19}F in proteins can be achieved through the use of auxotrophic strains of E. coli which lack the ability to synthesize specific amino acids and instead enable the introduction of non-natural fluorinated amino acids.[98,99] Amber codon suppression using orthogonal tRNA and its cognate aminoacyl-tRNA synthetase evolved to recognise ^{19}F-labelled (and other non-natural) amino acids offer an alternative and more selective approach[100] that may be more suited to RNC studies.

4.2 Practical notes on NMR of RNCs

A defining feature of NMR studies of ribosomes and RNCs is the low sensitivity arising from the large molecular weight of the ribosome, limited concentrations (approximately 10 μM which is an order of magnitude less than typical concentrations although this corresponds to 24 mg/mL) and short sample lifetimes (section 4.1.1).[77] The limited stability of RNCs in particular means that increasing the number of scans to compensate for weak signals is not generally an option. Indeed, continual evaluation of the sample stability is essential and this can also require a substantial fraction of the total acquisition time. We have previously discussed the NMR sensitivity limitations of RNC samples,[77] and so we focus here on summarising the main points and detailing new applications and developments made since.

4.2.1 Optimising NMR sensitivity of RNCs

Given the low sensitivity of RNC NMR discussed above, the continued refinement of NMR methods and strategies has been central to progress in this field. The increasing availability of high- [101] and ultrahigh (>1 GHz) field spectrometers,[102] and cryogenic probes[103] have certainly alleviated some of the inherent insensitivity of NMR of large biological systems. However, for the greater sensitivity improvements crucial for increasingly sophisticated observations of RNCs, further optimisation is required through sample preparation, by the choice of isotopic labelling scheme, the choice and set-up of the experiments, or potentially enhancing the stalling efficiency of the arrest peptide sequence and thus stability of the RNC (discussed in section 4.1.1), effectively increasing the acquisition time available for each sample.

The majority of RNC studies published so far have employed the ^{1}H,^{15}N SOFAST-HMQC experiment to observe disordered conformations of NCs.[33,34,37,79] This experiment employs selective Ernst-angle excitation of amide resonances to accelerate the longitudinal relaxation of the observed spins, and when coupled with an optimised inter-scan recovery delay, enables a high number of scans to be rapidly acquired.[104] This is particularly effective for disordered proteins as there is additional transfer of magnetisation from the solvent via rapid amide proton exchange.[105] Similar longitudinal relaxation-optimised experiments such as the BEST-family of 3D and TROSY experiments [106] may yet prove useful in future RNC studies by NMR.

The longitudinal relaxation rate can be further enhanced to reduce the relatively long inter-scan recovery times, in an orthogonal approach, by the addition of soluble paramagnetic compounds into NMR samples (Figure 5). The paramagnetic relaxation enhancement (PRE) effect arises from long-range dipolar interactions between unpaired electrons in the paramagnetic centre and a nucleus, resulting in an additional paramagnetic contribution to both the longitudinal and transverse relaxation rates. For paramagnetic metal ions with electronic relaxation rates much shorter than their diffusional correlation time, such as Fe(III) and Ni(II),[107] this results in the paramagnetic longitudinal relaxation enhancement (PLRE) effect, where the longitudinal relaxation is strongly accelerated with a limited increase in the transverse relaxation.[108-110] For samples of 70S ribosomes, addition of the PLRE agent NiDO2A increased SOFAST-HMQC sensitivity by almost two-fold, with only marginal increases in linewidths associated with the increase in transverse relaxation [110]. Crucially, no specific interactions were detected between NiDO2A and any of the proteins studied and incubation of RNCs with NiDO2A showed no observable difference in sample lifetime (determined by NMR diffusion measurements and western blot analysis, see section 4.2.2), indicating that sample integrity was not compromised. The increase in SOFAST-HMQC sensitivity in NiDO2A-doped RNC samples, within the same

experimental time, enabled observation of additional resonances that could unambiguously be assigned to the NC (Figure 5A). The sensitivity gains obtained by PLRE are particularly advantageous to time-limited studies of unstable complexes such as RNCs, as the increase in longitudinal relaxation rates can therefore be usefully employed to either obtain spectra of higher signal-to-noise or resolution, or to use the additional time to acquire for acquisition of other (and perhaps more complex) experiments. Indeed, sensitivity improvements by PLRE agents are not limited to SOFAST-HMQC experiments, but have been shown to greatly improve sensitivity of NMR diffusion measurements of RNCs[110] (see section 4.2.2) and also direct ^{13}C-detected experiments.[109]

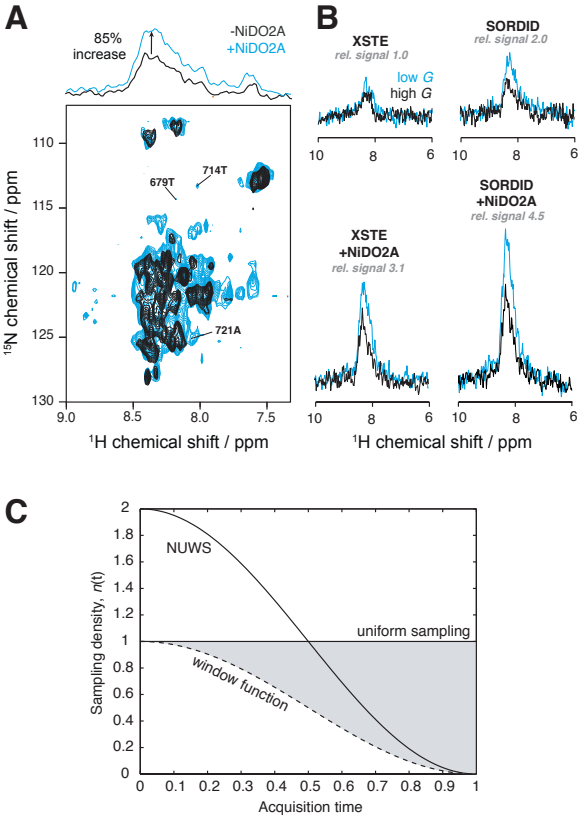

Figure 5. Strategies to optimise NMR sensitivity of ribosomal particles. (A) Addition of the PLRE agent NiDO2A to a disordered FLN5 RNC enhances ^{1}H,^{15}N SOFAST-HMQC sensitivity by approximately 85% (1D spectrum shown on top), enabling observation of additional NC resonances (labelled) undetectable in the absence of NiDO2A within the same experimental time. (B) Longitudinal relaxation-optimised SORDID experiments improve sensitivity of diffusion NMR measurements of RNC samples by 2-fold, relative to XSTE experiments. Further enhancements of up to 4.5-fold are made on addition of NiDO2A. (C) Uniform and non-uniform weighted sampling schemes for acquisition of the indirect dimension of multi-dimensional spectra. The NUWS approach provides 10-20% increase in sensitivity for low concentration samples. Figures adapted from references.[110,111]

The sensitivity of multi-dimensional NMR experiments can be further increased by sampling more efficiently on the Nyquist grid. Non-uniform weighted sampling (NUWS) has been shown to improve signal-to-noise by 10-20% by sampling every point in the Nyquist grid with a variable number of scans according to a weighting function (Figure 5C).[111] This technique has been successfully employed in SOFAST-HMQC experiments of α-synuclein RNCs (section 5.2)[34] and may have further applications to higher dimensional experiments in the future.[112]

Finally, we note that because of the limited sensitivity, the NMR observations of RNCs to date used 2D correlation spectra. While the chemical shifts, line widths and signal intensities provided by such spectra are a rich source of structure and dynamics (as we describe below), we expect that future studies will employ the sensitivity enhancements outlined above to increase the dimensionality and complexity of NMR experiments suitable for RNC studies.

4.2.2 Sample quality control using optimised NMR experiments

The inherent instability of ribosomal particles necessitates continual monitoring of NMR samples to ensure that the observed signals derive exclusively from intact ribosomal complexes rather than released or degraded material. Such species can be effectively distinguished using NMR measurements of translational diffusion, as originally applied to NMR studies of the ribosome (see section 3),[31] and which have become an integral part

of all RNC studies to date. The translational diffusion coefficients measured by these experiments are inversely proportional to the hydrodynamic radius according to the Stokes-Einstein equation, and their application in studies of ribosomal complexes therefore exploits the difference in the hydrodynamic radius of small proteins in free solution (typically 2-3 nm) and the 70S ribosome (12.6 nm).[31]

When applied to RNCs, continual measurements of the diffusion coefficient of the NC specifically are essential to ensure a complete assessment of sample integrity (Figure 6B), as release of the NC typically occurs before degradation of the ribosomal particle.[76,77] For ^{15}N-labelled samples of RNCs, this was originally achieved through ^{15}N-XSTE experiments[113] in which the diffusion of spins between a pair of encoding and decoding pulse field gradients (PFG) results in reduced signal intensity, dependent on the gradient strength and length of the diffusion delay according to the Stejskal-Tanner equation.[114] In the case of ^{15}N-XSTE experiments (Figure 6B, inset), storage of the spatially encoded magnetisation on ^{15}N spins serves both to edit isotopes and to optimise sensitivity by reducing longitudinal relaxation losses during diffusion delays (since ^{15}N longitudinal relaxation rates are slower than that of ^{1}H).[113] Further sensitivity enhancements can be made through the use of the longitudinal relaxation-optimised SORDID experiment,[115] which allows recovery of ^{1}H magnetisation (typically performed during long inter-scan delays) to occur simultaneously with the diffusion delay, thereby reducing the overall experimental time by approximately half (Figure 5B). Moreover, by exploiting the dependence of the PLRE effect (section 4.2.1) on the squared gyromagnetic ratio to accelerate recovery of ^{1}H magnetisation without adversely affecting storage on N_z during diffusion delays, the sensitivity of NMR diffusion measurements of RNCs can be increased even further, by up to 4.5-fold relative to XSTE experiments in the absence of NiDO2A (Figure 5B), corresponding to a reduction of over 20-fold in experimental time required to obtain the same spectra.[110]

For ^{13}C-labelled RNCs, translational diffusion is typically assessed by the ^{1}H-STE-^{1}H,^{13}C-HMQC experiment instead (Figure 6C),[33,76] which as far as possible is optimised to take advantage of the methyl-TROSY effect by selecting for slow relaxing coherences [93]. The effect of PLRE agents on the sensitivity of ^{13}C-based experiments have not yet been investigated, but given their use in other ^{13}C-based experiments,[109] may be of potential benefit.

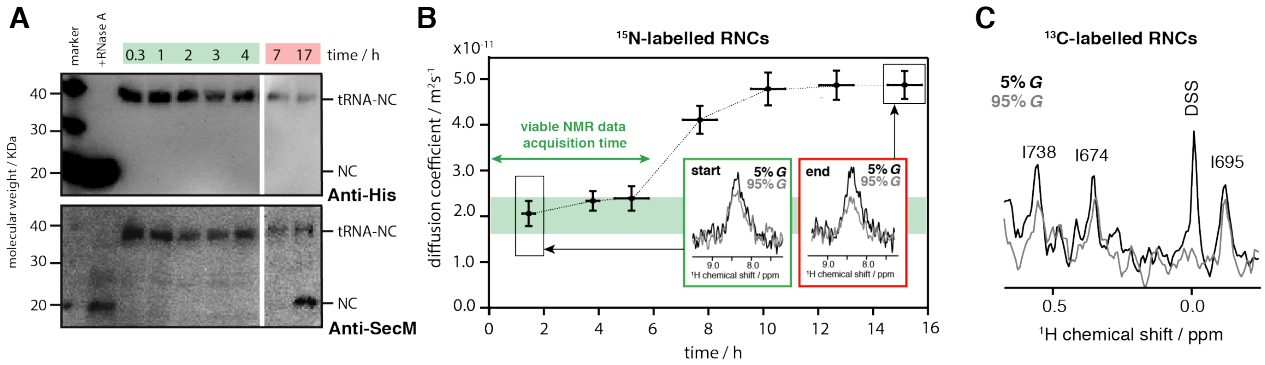

Figure 6. Biochemical and NMR quality control of RNC samples. (A) Time-course of RNC integrity assessed by western blot analysis detected at the N- and C-terminus of NCs using anti-His and anti-SecM antibodies respectively. Upshifted bands correspond to the tRNA-bound form of the RNC with release of the NC monitored at a band approximately 17 kDa below. (B) In parallel, the diffusion coefficient of RNC samples is continuously monitored using STE experiments interleaved during NMR data acquisition, in which ~1.8 x 10^{-11} m^2s^{-1} corresponds to fully ribosome-bound species. Insets show the amide envelopes of ^{1}H,^{15}N SORDID spectra used to calculate the diffusion coefficients. (C) ^{1}H,^{13}C-STE experiments enable similar measurements for selective methyl labelled RNCs. Figures adapted from references.[32,110]

Finally, biochemical analysis of identical samples incubated in parallel with NMR samples remains a critical orthogonal means of assessing the integrity of RNCs (Figure 6A). SDS-PAGE analysis followed by western blotting against NC-specific antibodies is used to detect ribosome-bound species. These migrate with an additional approximately 17-KDa band-shift relative to the NC molecular mass caused by the presence of the P-site peptidyl tRNA covalently linked to the NC by an ester bond that is retained by the low pH conditions.[76] Together with NMR measurements detailed above, these analyses ensure confidence that reported NMR resonances originate exclusively from ribosome-bound species.

5. Studies of co-translational folding studies using NMR spectroscopy

In this final section, we now focus on reviewing recent NMR studies of co-translational folding in RNCs. There have been studies of the co-translational folding of barnase[38] and the SH3 domain from α-spectrin[37] on the ribosome, and we have previously summarised their findings.[77] Here, we focus on the most recent studies, which have used one of two model NC systems. The most comprehensively studied RNCs comprise varying lengths of FLN5 and FLN6, the fifth and sixth filamin domains of the *Dictyostelium discoideum* gelation factor (isolated FLN5 proteins discussed in sections 2.2-2.3).[33,35,36] These have yielded residue-level insights into early structure acquisition on the ribosome of multi-domain protein which represent the majority of eukaryotic proteins[116] and for which co-translational folding is likely to play a major role during biosynthesis. Conversely, the intrinsically disordered protein (IDP) α-synuclein has also been studied as an RNC by NMR,[34] serving as a model of the unfolded state without the complications of folding and enabling us now in retrospect to draw comparisons with the disordered states of folding-competent FLN5 and SH3 RNCs.

5.1 Co-translational folding of an immunoglobulin domain

5.1.1 Monitoring length-dependent folding on the ribosome

The first observations of ribosome-bound NCs by NMR spectroscopy were made studying polypeptide chains of the FLN5 domain, with a linker comprising fragments of the following FLN6 domain.[33,36,79] The initial studies found that both disordered and native-like folded conformations of RNCs were sufficiently dynamic to be observable by NMR spectroscopy, and thus tumbled relatively independent from the ribosomal particle, paving the way for future RNC studies. The use of ^{1}H,^{15}N and ^{1}H,^{13}C SOFAST-HMQC experiments was crucial to obtain spectra with sufficiently high sensitivity for analysis as was the development of robust sample preparations of RNCs, using both in vitro and in vivo methods, at the very large scales required for NMR studies (several orders of magnitude higher than any previous biophysical, biochemical and cryo-EM work on RNCs, see section 4.1). Preliminary insights into the co-translational folding of FLN5 were obtained by the spectroscopic analysis of two SecM-stalled RNC constructs, either the single FLN5 domain or the combined FLN5-6 domain, in which FLN5 was 21 (FLN5+21) and 110 residues (FLN5+110) from the PTC respectively (Figure 7A). The NMR spectra of FLN5-6 RNC showed dispersed ^{1}H,^{13}C-HMQC resonances of side-chains with no significant chemical shift changes to that of isolated FLN5, providing the first direct evidence of native folded structure on the ribosome [36]. In contrast, at the shorter linker length, the narrow dispersion of ^{1}H,^{15}N-HMQC amide backbone cross-peaks, which overlapped closely with spectra of isolated FLN5 denatured in 8 M urea, indicated the presence of disordered conformations.[35] The lack of folding in this RNC was attributed to the occlusion of approximately 30 C-terminal residues in the ribosomal exit tunnel.

These studies raised the question of when during translation folding of FLN5 occurs on the ribosome, and whether this process may be modulated by the ribosome. This was addressed in a subsequent study, in which RNCs with variable lengths of the FLN6 linker (plus the SecM arrest peptide) were systematically investigated (Figure 7A), to create a series of 'snapshots' of the length-dependent co-translational folding pathway of FLN5.[33] When the FLN5 domain was between 21 and 45 residues away from the PTC, resonances were observed in ^{1}H-^{15}N SOFAST-HMQC spectra (Figure 7C) that were narrowly dispersed in the proton dimension and all overlaid extremely closely with those from isolated unfolded FLN5 variants (unfolded by either the destabilising Y719E mutation or by C-terminal truncation of 12 residues). Three well-resolved amide resonances were selected that had comparable ^{1}H linewidths across all RNC lengths. Their intensity was then used to determine the population of the unfolded NC state (having ruled out any changes in their relaxation properties). This analysis revealed a substantial in the population of the unfolded state between NC linker lengths of 42 to 45 residues, consistent with a transition towards the folded state. In contrast, ^{1}H,^{15}N-correlation spectra of FLN5+67 and FLN5+110 RNCs (Figure 7B) showed resonances that could not be assigned to unfolded FLN5, but rather to the unfolded FLN6 linker, providing further evidence of FLN5 folding at longer lengths.

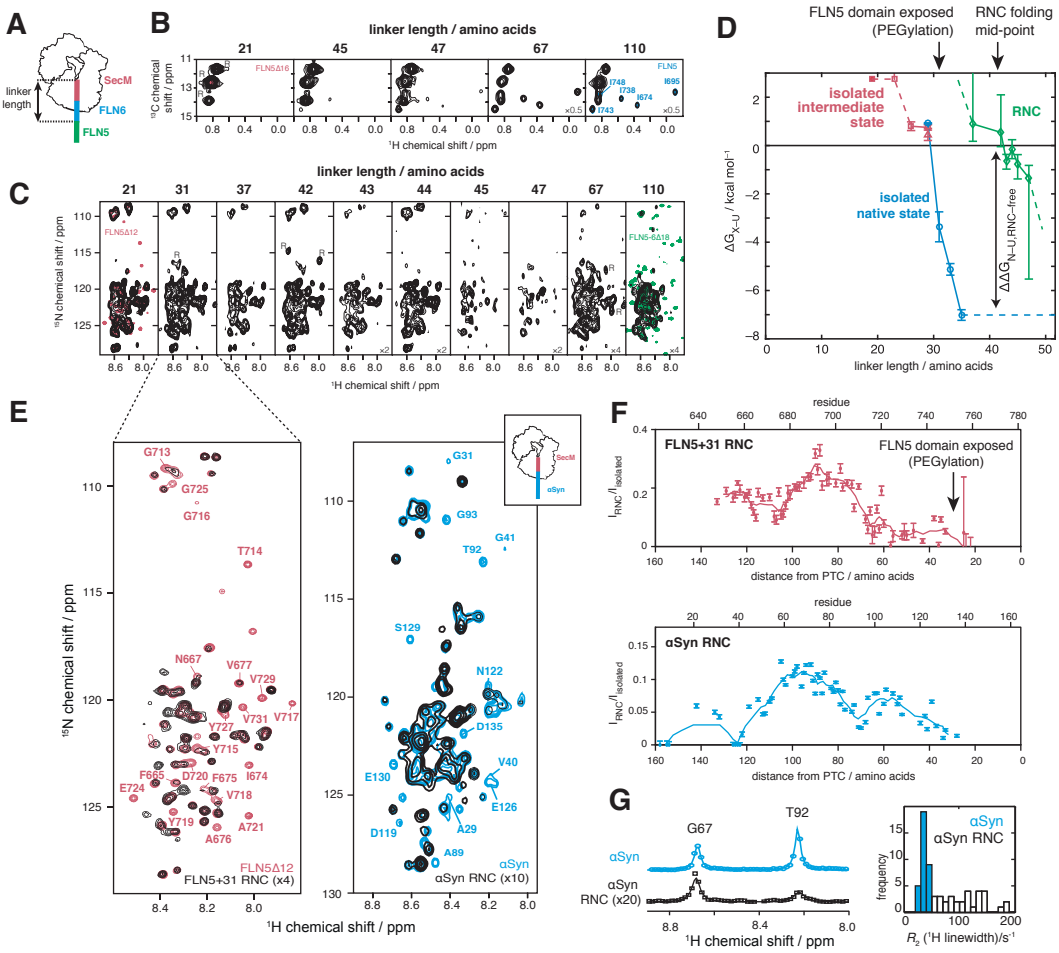

Figure 7. Co-translational folding of FLN5 on the ribosome by NMR spectroscopy and comparisons of its disordered state with α-synuclein RNCs. (A) To create equilibrium snapshots of co-translational folding, a series of RNC constructs were made in which the FLN5 domain is tethered to the PTC by a linker comprising SecM and a variable number of FLN6 residues. (B) ^{1}H,^{13}C HMQC spectra of ^{2}H,[^{13}CH$_3$-ile] labelled FLN5 RNCs, showing a progressive increase in the signal intensity of resonances of the five native isoleucine residues with longer linker lengths. (C) ^{1}H,^{15}N SOFAST-HMQC spectra of uniform ^{15}N-labelled FLN5 RNCs showing a progressive decrease in the signal intensity of disordered FLN5 cross-peaks with longer linker lengths, followed by observation of signals from the FLN6 linker residue in the FLN5+110 RNC spectrum. (D) Thermodynamic characterisation of ribosome-bound FLN5, measured by SOFAST-HMQC signal intensity analysis, and isolated FLN5 truncations. The length at which the FLN5 domain emerges from the exit tunnel and its native C747 residue is exposed, determined by PEGylation, is offset from the observed folding mid-point on the ribosome. (E) ^{1}H,^{15}N SOFAST-HMQC spectra of FLN5+31 RNC (left) and α-synuclein RNC (right), overlaid with the isolated protein spectrum. (F) Non-uniform signal line broadening observed for FLN5+31 (top) and α-synuclein (bottom) RNCs. Significant broadening is observed for FLN5+31 until 58 residues from the PTC despite becoming solvent accessible at 34 residues, whereas α-synuclein RNC resonances are broadened only at 28 residues from the PTC. (G) Linewidth analysis of α-synuclein RNC using ^{1}H cross-sections through ^{1}H,^{15}N SOFAST-HMQC spectra (left) used to determine and compare the ^{1}H R_2 rates of ribosome-bound and isolated α-synuclein (right). Figures adapted from references.[33,34,51]

Notwithstanding our early observations,[36,78] we have found that folded FLN5 NCs are very challenging to detect via ^{15}N-labelling. Separate NMR samples were therefore produced in which isoleucine residues were selectively ^{13}C-methyl labelled against a perdeuterated background to reduce relaxation and increase NMR sensitivity (discussed in sections 4.1.2 and 4.2.1). The methyl-TROSY effect was exploited by acquisition of ^{1}H-^{13}C HMQC spectra (Figure 7B), which revealed the appearance of five native-like resonances only when FLN5 was at least 45 residues away from the PTC. The signal intensities of the methyl cross-peaks increased with length, which may be attributable both to an increase in the population of the folded state but also to

increased mobility away from the slowly tumbling ribosomal particle. These observations supported the folding transition at 42-45 residues from the PTC observed by uniform ^{15}N-labelling.

5.1.2 Delayed folding on the ribosome

To explore the folding transition in more detail, two further experiments were undertaken. Firstly, the solvent accessibility of the emerging FLN5 domain was studied by PEGylation: the chemical modification by a 5-KDa PEG-maleimide of a cysteine residue C747 at the inter-domain FLN5-FLN6 boundary. The resulting 5-KDa mass difference could be detected by SDS-PAGE following PEGylation of RNCs (rendered folding-incompetent by the destabilising mutation Y719E) and this was used as a measure of NC emergence from the exit tunnel. Complete PEGylation was found when the FLN5 domain was ≥31 residues away from the PTC. Secondly, comparisons were made with C-terminal truncations of the isolated FLN5 domain (section 2.2) which showed that deletion of the four terminal residues could be tolerated whilst retaining a fully natively folded conformation.[51] Collectively, these data showed that acquisition of native structure would be expected, on the basis of steric accessibility, at a linker length of 31 residues and greater. In a later study,[51] structure-based models also showed that the native FLN5 state was sterically accessible on the ribosome with only a 28-amino acid linker (in a relaxed conformation). However, contrary to these expectations, NMR measurements required 42-45 residues of the FLN6 linker before folding commences. Moreover, in contrast to isolated FLN5, which was observed to populate an intermediate state when 6-16 C-terminal residues were truncated, no such intermediate was detected on the ribosome, despite it being sterically compatible with an approximately 18-amino acid residue linker.[51] Both the intermediate and native state are therefore destabilised on the ribosome, the latter destabilised by $\Delta\Delta G \sim 7$ kcal mol^{-1} comparable to the free energy of folding for FLN5, based on measurements of the unfolded state populations by NMR (Figure 7D).[33,51]

To understand the origin of this remarkable destabilisation of folded FLN5 NCs and/or stabilisation of disordered states by the ribosome, the FLN6 residues linking the FLN5 domain to the ribosome were replaced with a poly(glycine-serine) linker.[33] However, the offset between the emergence from the exit tunnel and folding remained, suggesting that inter-domain interactions alone were insufficient to cause destabilisation of the domain on the ribosome. Next, high-resolution ^{1}H-^{15}N SOFAST-HMQC experiments of FLN5 RNCS of linker lengths 21-42 (using both the 'natural' FLN6 linker and the poly(glycine-serine) linker) acquired at high fields (950 MHz) were analysed at a residue-specific level (Figure 7E, left).[33] These showed significant variations in line broadening across the sequence, particularly in the regions F655 to V667 where 70% of signal intensity was lost (relative to free FLN5) and the C-terminal region G713 to G750 which was completely undetectable (Figure 7F). These data indicate a loss of mobility in segments of the NC due to interactions with the ribosomal surface, resulting in an increased effective rotational correlation times. The interaction with the ribosome surface was significantly greater than that measured for isolated disordered FLN5 in the presence of 70S ribosomes and was therefore ascribed to the increased effective ribosome that results when the NC is tethered to the ribosome via the PTC. Indeed, this effect would be most pronounced at the shortest linker lengths and at the C-terminus. It is noteworthy that the six most C-terminal residues of FLN5 (V745 to G750) were also found, in a later study, to contribute almost the entire stability of the domain;[51] their participation in interactions with the ribosome in addition to folding may therefore begin to explain the inhibition of acquisition of stable folded structure by the ribosome. Thus, the results of this study begin to show how the ribosome might modulate co-translational folding processes, in this case through a chaperone-like holdase activity. If this is found to be a general mechanism, it may help to minimise formation of misfolded protein by stabilising unfolded states, similar to the action of chaperones such as trigger factor (section 5.3) - a mechanism that is certainly important in the biosynthesis of multi-domain proteins.

Finally, the possibility that disordered FLN5 conformations may differ on and off the ribosome cannot be entirely excluded from our considerations. Coarse-grained MD simulations have found that the steric effect of the ribosome vestibule and surface can shift unfolded NC ensembles towards more anisotropic shapes and partially folded structure.[70] While the observed backbone amide chemical shift differences between disordered FLN5 NCs and isolated unfolded FLN5 variants were small (<0.1 ppm), other chemical shifts (such as H^{α}, C^{α} and C^{β}) are more sensitive to changes in secondary structure,[117,118] and may therefore give additional

insights into early structure acquisition. Measurements of tertiary structure or compaction of the NC, by, for example, NOE [119] or PRE methods,[120] may thus also be of particular interest, although on a practical level, such experiments will be highly challenging given the low sensitivity available. However, RDC measurements of ribosome-bound bL12 (section 3) offer hope to achieving such measurements on RNCs.

5.2 Comparisons of the disordered FLN5+31 RNC with α-synuclein RNCs

5.2.1 Probing ribosome-nascent chain interactions with NMR line broadening analysis

The NMR studies of FLN5 discussed above show that the ribosome provides a unique environment for protein folding to take place, where interactions between the NC and nearby ribosomal proteins and rRNA may compete with NC intramolecular contacts to modulate the co-translational folding process. These effects have also been observed for other RNC systems using other biophysical methods.[69,121] To investigate these interactions further, we used α-synuclein as a convenient model as any interactions can be examined without the complication of protein folding effects.[34] The relatively flat (i.e. weakly funnelled) free energy landscapes of disordered proteins also results in high sensitivity to such external perturbations. Using a similar approach to FLN5 RNCs,[33] α-synuclein RNCs were translationally arrested using the SecM motif and prepared by expression in vivo with uniform ^{15}N-labelling, before several purification steps. ^{1}H-^{15}N SOFAST-HMQC experiments were acquired using non-uniform weighted sampling to enhance sensitivity (section 4.2.1). High quality spectra were thus obtained of the α-synuclein RNC (Figure 7E, right). Chemical shifts of observed resonances closely matched those of isolated α-synuclein, indicating that similar ensembles of disordered state conformations were present on and off the ribosome. However, large reductions in signal intensities were observed relative to isolated α-synuclein (>87%), and complete broadening of the C-terminus was observed beyond residue D135 (Figure 7F). The loss of resonances of C-terminal residues was attributed to occlusion in the exit tunnel and hence severely restricted tumbling (effectively a thousand-fold increase in the correlation time). The detection of D135 resonance indicated the minimum distance, 28 residues from the PTC, at which the NC possessed sufficient mobility to become observable. This is consistent with ensemble structures determined using coarse-grained MD simulations (section 5.2.2) in the same study found that D135 was the first NC residue to be emerge from the exit tunnel into the vestibule. In contrast, the C-terminal residues G713 to G750 and the FLN6 linker residues were all broadened beyond detection in ^{1}H-^{15}N SOFAST-HMQC spectra of the disordered FLN5+31 RNC,[33] indicating that sufficient mobility was obtained only beyond 58 residues from the PTC (Figure 7F). The comparison of the two systems clearly shows that the broadening of FLN5 resonances cannot solely be attributed to occlusion in the exit tunnel. Instead, these observations provide further evidence of FLN5 NC interactions with nearby ribosomal RNA or proteins. Unfortunately, due to these line broadening effects, the ability of NMR to study more strongly interacting segments of the NC is limited, so ultimately we expect that an integrative approach using other techniques such as cryo-EM, which has been successfully employed to examine NCs within the exit tunnel,[24,26,81] and molecular dynamics (see section 5.2.2), will be key to understanding these effects comprehensively.

Beyond the α-synuclein C-terminus, and similarly to FLN5 RNCs, large variations in line broadenings (i.e. cross-peak intensity) were observed across the α-synuclein.[34] An overall increase in signal was observed from the C-terminus to residue K58, consistent with increasing mobility further from the ribosome as predicted by the coarse-grained simulations (section 5.2.2). However, an overall decrease in intensity was then observed from residue K58 to the N-terminus. An analysis of the charge distribution across the sequence shows a high density of positive charges at the N-terminus, and negative charges around the C-terminus. Given the negatively charged ribosomal surface,[122] it is thought that electrostatic interactions may contribute to NC-ribosome contacts.[69,123] Consistent with this finding. isolated α-synuclein with additional positive or negative charges introduced at the N-terminus showed increased and decreased line broadening respectively.

Substantial signal reductions in the α-synuclein RNC spectrum were also observed for resonances assigned to residues centred around M5-S9, V37-V40 and, to a lesser extent, around T92/G93. These regions coincide with the aromatic residues present in the sequence (F4, Y39, and F94). On the other hand, the region of highest hydrophobicity, the NAC region (residues 61-91), which contains many leucine and alanine residues, was also the region of greatest intensity. This suggests that the ribosome surface does not interact stronger

with hydrophobic aliphatic residues, but may have a particular propensity to interact with aromatic residues. However, whether the properties of single amino acids dictate ribosomal interactions more generally across other NC systems is yet to be systematically examined.

To quantify the extent of NC interactions, ^{1}H linewidths of RNC resonances were measured. A three-fold increase in R_2 was observed from residues G67 to T92, from 37 ± 7 s^{-1} (mean ± SD) for the isolated protein to 109 ± 38 s^{-1} in the RNC. In contrast, amide proton relaxation rates of a NC tightly bound to the ribosomal surface was estimated to be 16.1 ± 4.4 x 10^{-3} s^{-1}, by averaging ^{1}H-^{1}H dipolar interactions across an ensemble of all-atom models of the α-synuclein RNC (section 5.2.2). Assuming fast exchange between the free and ribosome-associated states (typically found for weak interactions), it was estimated that 0.45 ± 0.24% of this NC segment was bound (rigidly) to the ribosome surface. A similar analysis of linewidths also concluded that disordered SH3 RNCs were less than 1% bound to the ribosome surface at a given time.[37] Although this interaction is weak (demonstrating also the sensitivity of NMR measurements), residues that were unobservable by NMR are likely to have stronger interactions with the ribosome. This is likely to be the case for the C-terminal residues of FLN5 RNCs that show significant broadenings in NMR spectra[33] but also contribute almost the entire stability of the domain [51] (section 5.1.2), and thus potentially result in delayed folding. A quantitative assessment of binding parameters and thermodynamics of ribosome-NC interactions together with systematic changes to the NC sequence will give further insights into how such effects contribute to modulating folding on the ribosome.

5.2.2 Combining NMR with MD simulations to probe NC interactions

Based on the results above, NC interactions with the ribosome are thought to derive at least in part from the highly negative electrostatic potential of the ribosomal surface, which in turn arises predominantly from the phosphodiester backbone of the ribosomal RNA,[124] although there may be some contribution from counter ions on the surface.[123] However, because charged and aromatic residues are also found to interact with other components in the cell,[42,43] and hydrophobic residues may also promote the interaction with chaperones such as trigger factor (TF) (section 5.3), there are likely to be other ways in which specific interactions with the ribosome are coordinated to tune protein folding. The ribosomal proteins accessible to emerging NCs were identified in the coarse-grained simulations of α-synuclein RNCs (Figure 8A) initially discussed in the previous section. The CamTube force field was used, which presents the NC as a flexible tube and thus efficiently samples protein conformations with limited computational cost [125], crucial in the study of such large biological systems. Ribosomal proteins close to the exit tunnel and within 10 Å of the emerging nascent polypeptide included uL22, uL23, uL24 and uL29, which are implicated in the binding of several factors including signal-recognition particle (SRP) and TF [126], and bL17, bL32, and the 23S rRNA (Figure 8A).

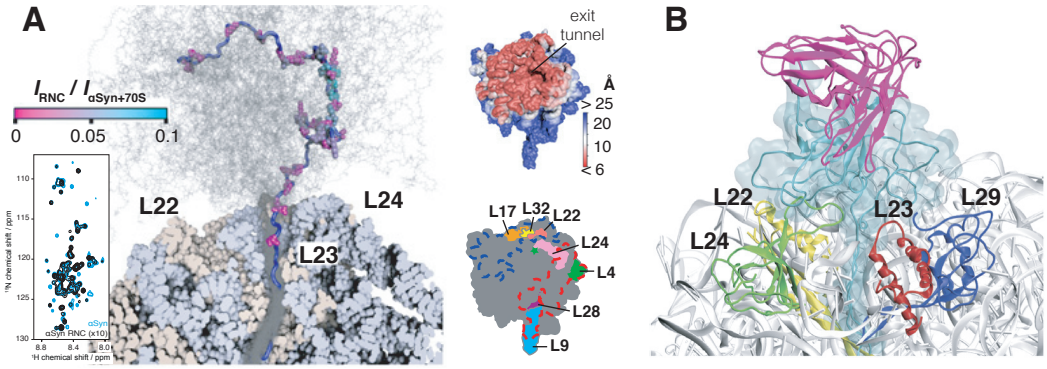

Figure 8. Ensemble RNC structures determined by MD simulations and NMR measurements. (A) Structural modelling of the α-synuclein RNC using coarse-grained simulations. For reference, NC residues are coloured by their observed NMR resonance intensities relative to isolated α-synuclein (inset shows spectrum as in Figure 7E). Distances of closest approach for the NC are coloured on the ribosome surface (top right) with a schematic diagram of ribosomal proteins in close proximity to the exit tunnel (bottom right). (B) Structural ensemble of FLN5+110 RNC using chemical shift restraints. Transient interactions are made between the disordered FLN6 linker and ribosomal proteins lining the exit tunnel. Figures taken from references[33,34].

To investigate whether NCs indeed do interact with these sterically accessible proteins during biosynthesis, NMR chemical shift measurements were used to restrain MD simulations and so determine a structural ensemble of an RNC. The FLN5+110 RNC provided probes for both the folded FLN5 domain and the disordered FLN6 linker, through $^{1}H,^{13}C$-methyl and $^{1}H,^{15}N$-amide backbone chemical shifts respectively, and these were therefore applied as structural restraints in replica-averaged metadynamics calculations.[33] The calculated ensemble (Figure 8B) showed that the folded FLN5 domain was tethered to the PTC by a disordered yet compact FLN6 linker with a propensity to form native-like secondary structure (approximately 20% population on average) and intramolecular contacts. The folded domain was sufficiently dynamic to access large regions of the ribosome surface, making transient contacts with the 23S rRNA and uL29. Consistent with results from coarse-grained simulations of the α-synuclein RNC, the disordered FLN6 linker also made contacts with the 23S RNA and the ribosomal proteins that line the exit port, uL22, uL29, and most frequently, uL23 (30%) and uL24 (55%). Transient interactions with uL23 and uL24 have been also detected for other RNCs by cross-linking experiments,[127] while biophysical studies using loop-deletion strains have shown that obstruction of the exit tunnel and vestibule by the protruding uL23 and uL24 β-hairpin loops can sterically modulate the co-translational folding of NCs by reducing the available space for folding.[128] It is not known whether such specific interactions are responsible for the folding delay of FLN5 on the ribosome, or whether uL23 and uL24 universally interact with NCs as chaperones, but in the future, the use of modified ribosomes combined with NMR analysis may help understand the role these ribosomal proteins play during co-translational folding.

MD simulations have become an important and effective tool in the study of large, complex biological systems,[129] aided by the exponential growth in computational processing power,[130] and are beginning to be applied to the complete ribosome over increasing time scales.[131] Continued improvements in, for example, sampling methods and force fields to combine protein and RNA simulations, are likely to be crucial to progress in their application to ribosomal particles. Nevertheless, it is clear that MD simulations not only enable structural determination of RNCs restrained by experimental NMR data (and will certainly benefit from future structural NMR measurements), but can also play an important role in guiding future experimental work.

5.3 Interactions with trigger factor by NMR

As the newly synthesised NC emerges from the exit tunnel, a host of auxiliary factors within the densely crowded cellular milieu compete to bind and interact. In the context of protein folding, the multitude of molecular chaperones that work in a protein quality control network to maintain protein homeostasis are Of particular interest.[11] TF is the only known ribosome-associated chaperone in prokaryotics, but shares similar functionality with Hsp70-based chaperones in eukaryotes.[132,133] TF is a homodimer that is constitutively expressed and is highly in the cytosol, occurring in two- to three-fold molar excess to ribosomes.[134] TF interacts with the ribosome as a monomer, docking in an ATP-independent manner at the intersection of ribosomal proteins uL23 and uL29 and domain III of the 23S rRNA, enabling direct access to emerging nascent chains.[135,136] The interaction with ribosomes is rapid and reversible to scan for substrate NCs. In the presence of a NC, its dissociation is inhibited and thus multiple copies of TF can remain bound to substrate NCs, even following dissociation of TF from the ribosome.[137] This 'holdase' activity shields hydrophobic sequence elements within NCs to supress misfolding and aggregation.[138,139] In the case of multi-domain proteins, where NCs are particularly susceptible to inter-domain misfolding during co-translational folding, TF can protect against denaturing effects of unfolded polypeptides on adjacent folded domains.[140] Furthermore, force spectroscopy measurements have shown that TF can also act as a foldase by accelerating the folding reaction under moderate mechanical force.[141]

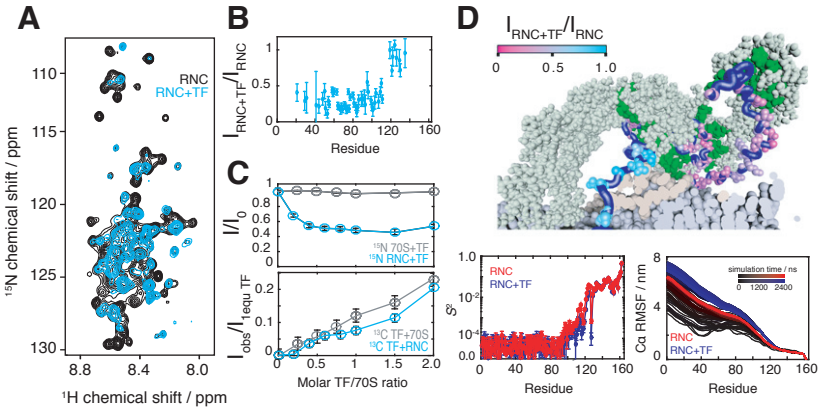

Figure 9. RNC interactions with trigger factor by NMR spectroscopy. (A) ^{1}H,^{15}N SOFAST-HMQC spectra of α-synuclein RNC in the presence and absence of equimolar trigger factor. (B) Analysis of line broadenings observed for the SOFAST-HMQC resonances on addition of trigger factor. (C) Changes in the integral of the 1D ^{1}H,^{15}N SOFAST-HMQC amide envelope of 70S ribosomes and α-synuclein RNC upon addition of increasing trigger factor (top), and in the integrals of 1D ^{13}C-HMQC methyl resonances of trigger factor in the presence of 70S ribosomes and α-synuclein RNC. (D) Structural modelling of α-synuclein RNC with trigger factor (top). Amide S2 order parameters (bottom left) and Cα RMDF (bottom right) determined from simulations of α-synuclein RNC with and without trigger factor. Figures taken from references.[34]

The dynamic interplay of dimerisation, ribosome binding and NC interactions by TF complicate its study, but there have been successes by NMR spectroscopy in determining its dimeric structure[142,143] and in structurally understanding its substrate binding activity, at least in the absence of the ribosome.[138] Key to these studies has been the selective methyl labelling and exploitation of methyl-TROSY to overcome the large molecular weight of TF (48 KDa). The interaction of TF with a translating NC has also been examined by NMR, in the studies of the α-synuclein RNCs introduced above (section 5.2).[34] Although α-synuclein RNCs are is a weak substrate of TF, its binding could nevertheless be detected; moreover, stronger binding would cause significant line broadening resulting in unobservable NMR resonances. Chemical shift perturbations were minimal on addition of TF to ^{15}N-labelled α-synuclein RNCs were (Figure 9A), indicating no significant structural changes occurred within the NCs. However, the ^{1}H-^{15}N SOFAST-HMQC signal intensities of RNC residues K21-G111 were substantially reduced (by approximately 70%), while the C-terminal region remained unperturbed (Figure 9B-C). These data were consistent with coarse-grained modelling of the TF-bound RNC in the same study (Figure 7D), which showed that only the NC residues in the broadened region were able to contact the substrate binding sites of TF previously proposed.[138] A potentially interesting interplay between NC interactions with TF and the ribosome surface was also observed by the simulations, which found that binding of TF altered the region of ribosomal surface and ribosomal proteins accessible to the NC.[34] It is interesting to speculate whether the NC interactions between ribosomal proteins and RNA and with TF demonstrate a 'hand-over' of contacts of disordered segments of NC, in a similar fashion to its cooperation with DnaK,[144] to minimise the formation of misfolded or aggregated states during biosynthesis. In any case, the ability for TF to alter the dynamics of RNCs of α-synuclein, a weak TF substrate, demonstrate the complexity of the co-translational folding landscape of an emerging NC during the early stages of structure acquisition.

5.4 In-cell NMR spectroscopy

As the nascent polypeptide emerges from the exit tunnel, it enters a densely crowded cellular environment where, in addition to the interactions with the ribosomal surface (sections 5.2.1 and 5.2.2) and auxiliary factors such as TF (section 5.3), it is exposed to high concentrations of proteins, nucleic acids and other biomolecules that collectively can exceed 300 gL^{-1}.[17] Aside from biological events that occur in the cell, the crowded environment therefore also imposes physical effects on the NC structure and dynamics. Experimental[145] and computational work[146] have shown that there are several competing effects in vivo: the excluded volume effect, arising from the very high concentration of molecules in solution, promotes compaction of proteins by sterically disfavouring expanded states, whereas so-called 'quinary structure' arises from non-specific

interactions including electrostatics and hydrogen bonding, and can be either stabilising or destabilising.[147,148] Because of its non-invasive nature, the ability to probe dynamics in solution, and the capacity for selective labelling, NMR spectroscopy is emerging as a key technique to study protein structure and folding in living cells. The use of in-cell NMR has enabled, for example, characterisation of α-synuclein structure in both E. coli[43] and mammalian cells,[42] de novo in-cell protein structure determination,[39,41] in vivo analysis of protein-protein[149,150] and protein-nucleic acid interactions,[151] observation of post-translational modifications,[152] and measurements of protein stability[44,153-155]. A key challenge associated with in-cell NMR is the lack of observable resonances obtainable for certain proteins for reasons that are poorly understood. While this can be somewhat mitigated with the use of more sensitive labelling schemes such as selective methyl labelling and ^{19}F-labelling (section 4.1.2),[156] it is crucial to ensure that observed signals derive from intracellular species rather than extracellular ones and this can be achieved non-invasively by diffusion NMR.[157] With increasing advances in NMR sensitivity through both preparative and spectroscopic means, the enticing prospect of studying RNCs in living cells by NMR spectroscopy is becoming increasingly within reach.

Concluding remarks

In this review, we have discussed recent progress in NMR spectroscopy to probe the emerging NC during biosynthesis on the ribosome. These studies have shown that NMR is uniquely able to study at high-resolution the dynamic NC beyond the exit tunnel where most structure acquisition occurs, and have begun to identify and characterise the distinct scenarios in which protein folding occurs in vivo. In future studies, we expect that NMR spectroscopy will be used to study RNCs under increasingly biologically relevant conditions, and will also form an important part of an integrative approach combining complementary techniques, such as cryo-electron microscopy and MD simulations, to enable us to gain a more comprehensive understanding of de novo protein folding in the cell.

Acknowledgements

We would like to thank all members of the Christodoulou group for helpful discussions during the preparation of this review, and acknowledge funding by the Wellcome Trust.

References

1 C. A. Waudby, C. M. Dobson and J. Christodoulou, *Trends in Biochemical Sciences*, 2019.
2 C. M. Kaiser and K. Liu, *Journal of Molecular Biology*, 2018, **430**, 4580–4591.
3 V. Ramakrishnan, *Cell*, 2002, **108**, 557–572.
4 C. B. Anfinsen, *Science*, 1973, **181**, 223–230.
5 S. E. Jackson, *Fold Des*, 1998, **3**, R81–91.
6 J.-H. Han, S. Batey, A. A. Nickson, S. A. Teichmann and J. Clarke, *Nat Rev Mol Cell Biol*, 2007, **8**, 319–330.
7 A. Borgia, K. R. Kemplen, M. B. Borgia, A. Soranno, S. Shammas, B. Wunderlich, D. Nettels, R. B. Best, J. Clarke and B. Schuler, *Nature Communications*, 2015, **6**, 8861.
8 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2006, **75**, 333–366.
9 E. Siller, D. C. DeZwaan, J. F. Anderson, B. C. Freeman and J. M. Barral, *Journal of Molecular Biology*, 2010, **396**, 1310–1318.
10 F. Gloge, A. H. Becker, G. Kramer and B. Bukau, *Current Opinion in Structural Biology*, 2014, **24**, 24–33.
11 F. U. Hartl, A. Bracher and M. Hayer-Hartl, *Nature*, 2011, **475**, 324–332.
12 A. Gershenson and L. M. Gierasch, *Current Opinion in Structural Biology*, 2011, **21**, 32–41.
13 B. Chen, M. Retzlaff, T. Roos and J. Frydman, *Cold Spring Harb Perspect Biol*, 2011, **3**, a004374.
14 Y. Nyathi, B. M. Wilkinson and M. R. Pool, *Biochim. Biophys. Acta*, 2013, **1833**, 2392–2402.

15 D. Akopian, K. Shen, X. Zhang and S.-O. Shan, *Annu. Rev. Biochem.*, 2013, **82**, 693–721.
16 S. Varland, C. Osberg and T. Arnesen, *Proteomics*, 2015, **15**, 2385–2401.
17 S. B. Zimmerman and S. O. Trach, *Journal of Molecular Biology*, 1991, **222**, 599–620.
18 A. Dhar, A. Samiotakis, S. Ebbinghaus, L. Nienhaus, D. Homouz, M. Gruebele and M. S. Cheung, *Proc Natl Acad Sci USA*, 2010, **107**, 17586–17591.
19 Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois and G. J. Pielak, *J. Am. Chem. Soc.*, 2012, **134**, 16614–16618.
20 T. A. Steitz, *Nat Rev Mol Cell Biol*, 2008, **9**, 242–253.
21 A. Javed, J. Christodoulou, L. D. Cabrita and E. V. Orlova, *Acta Crystallogr D Struct Biol*, 2017, **73**, 509–521.
22 E. H. Egelman, *Biophysical Journal*, 2016, **110**, 1008–1012.
23 B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J. P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz and R. Beckmann, *Science*, 2009, **326**, 1412–1415.
24 S. Bhushan, M. Gartmann, M. Halic, J.-P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D. N. Wilson and R. Beckmann, *Nat Struct Mol Biol*, 2010, **17**, 313–317.
25 S. Bhushan, H. Meyer, A. L. Starosta, T. Becker, T. Mielke, O. Berninghausen, M. Sattler, D. N. Wilson and R. Beckmann, *Molecular Cell*, 2010, **40**, 138–146.
26 O. B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, J. D. Puglisi, E. P. O'Brien, R. Beckmann and G. von Heijne, *CellReports*, 2015, **12**, 1533–1540.
27 N. R. Voss, M. Gerstein, T. A. Steitz and P. B. Moore, *Journal of Molecular Biology*, 2006, **360**, 893–906.
28 D. P. Frueh, A. C. Goodrich, S. H. Mishra and S. R. Nichols, *Current Opinion in Structural Biology*, 2013, **23**, 734–739.
29 R. Sprangers and L. E. Kay, *Nature*, 2007, **445**, 618–622.
30 J. Fiaux, E. B. Bertelsen, A. L. Horwich and K. Wüthrich, *Nature*, 2002, **418**, 207–211.
31 J. Christodoulou, G. Larsson, P. Fucini, S. R. Connell, T. A. Pertinhez, C. L. Hanson, C. Redfield, K. H. Nierhaus, C. V. Robinson, J. Schleucher and C. M. Dobson, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 10949–10954.
32 A. M. E. Cassaignau, H. M. M. Launay, M.-E. Karyadi, X. Wang, C. A. Waudby, A. Deckert, A. L. Robertson, J. Christodoulou and L. D. Cabrita, *Nature Protocols*, 2016, **11**, 1492–1507.
33 L. D. Cabrita, A. M. E. Cassaignau, H. M. M. Launay, C. A. Waudby, T. Wlodarski, C. Camilloni, M.-E. Karyadi, A. L. Robertson, X. Wang, A. S. Wentink, L. S. Goodsell, C. A. Woolhead, M. Vendruscolo, C. M. Dobson and J. Christodoulou, *Nat Struct Mol Biol*, 2016, **23**, 278–285.
34 A. Deckert, C. A. Waudby, T. Wlodarski, A. S. Wentink, X. Wang, J. P. Kirkpatrick, J. F. S. Paton, C. Camilloni, P. Kukic, C. M. Dobson, M. Vendruscolo, L. D. Cabrita and J. Christodoulou, *Proc Natl Acad Sci USA*, 2016, **113**, 5012–5017.
35 L. D. Cabrita, S. T. D. Hsu, H. Launay, C. M. Dobson and J. Christodoulou, *Proc Natl Acad Sci USA*, 2009, **106**, 22239–22244.
36 S.-T. D. Hsu, L. D. Cabrita, P. Fucini, J. Christodoulou and C. M. Dobson, *J. Am. Chem. Soc.*, 2009, **131**, 8366–8367.
37 C. Eichmann, S. Preissler, R. Riek and E. Deuerling, *Proc Natl Acad Sci USA*, 2010, **107**, 9111–9116.
38 A. Rutkowska, M. Beerbaum, N. Rajagopalan, J. Fiaux, P. Schmieder, G. Kramer, H. Oschkinat and B. Bukau, *FEBS Letters*, 2009, **583**, 2407–2413.
39 D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nature*, 2009, **457**, 102–105.
40 E. Luchinat and L. Banci, *IUCrJ (2017). M4, 108-118 [doi:10.1107/S2052252516020625]*, 2017, 1–11.
41 T. Müntener, D. Häussinger, P. Selenko and F.-X. Theillet, *J Phys Chem Lett*, 2016, **7**, 2821–2825.
42 F.-X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, *Nature*, 2016, **530**, 45–50.
43 C. A. Waudby, C. Camilloni, A. W. P. Fitzpatrick, L. D. Cabrita, C. M. Dobson, M. Vendruscolo and J. Christodoulou, *PLoS ONE*, 2013, **8**, e72286.

44 W. B. Monteith, R. D. Cohen, A. E. Smith, E. Guzman-Cisneros and G. J. Pielak, *Proc Natl Acad Sci USA*, 2015, **112**, 1739–1742.
45 A. P. Schlesinger, Y. Wang, X. Tadeo, O. Millet and G. J. Pielak, *J. Am. Chem. Soc.*, 2011, **133**, 8082–8085.
46 P. Selenko, Z. Serber, B. Gadea, J. Ruderman and G. Wagner, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 11904–11909.
47 I. R. Kleckner and M. P. Foster, *BBA - Proteins and Proteomics*, 2011, **1814**, 942–968.
48 P. Neudecker, A. Zarrine-Afsar, W.-Y. Choy, D. R. Muhandiram, A. R. Davidson and L. E. Kay, *Journal of Molecular Biology*, 2006, **363**, 958–976.
49 P. Neudecker, P. Lundström and L. E. Kay, *Biophysical Journal*, 2009, **96**, 2045–2054.
50 N. J. Anthis and G. M. Clore, *Q. Rev. Biophys.*, 2015, **48**, 35–116.
51 C. A. Waudby, T. Wlodarski, M.-E. Karyadi, A. M. E. Cassaignau, S. H. S. Chan, A. S. Wentink, J. M. Schmidt-Engler, C. Camilloni, M. Vendruscolo, L. D. Cabrita and J. Christodoulou, *Proc Natl Acad Sci USA*, 2018, **115**, 9744–9749.
52 T. K. Karamanos, A. P. Kalverda, G. S. Thompson and S. E. Radford, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2015, **88-89**, 86–104.
53 A. Sekhar, J. A. O. Rumfeldt, H. R. Broom, C. M. Doyle, R. E. Sobering, E. M. Meiering and L. E. Kay, *Proc Natl Acad Sci USA*, 2016, **113**, E6939–E6945.
54 P. CLARK, *Trends in Biochemical Sciences*, 2004, **29**, 527–534.
55 A. D. Kippen, J. Sancho and A. R. Fersht, *Biochemistry*, 1994, **33**, 3778–3786.
56 J. L. Neira and A. R. Fersht, *Journal of Molecular Biology*, 1999, **285**, 1309–1333.
57 J. Lu and C. Deutsch, *Nat Struct Mol Biol*, 2005, **12**, 1123–1129.
58 S. A. Robson, R. Peterson, L.-S. Bouchard, V. A. Villareal and R. T. Clubb, *J. Am. Chem. Soc.*, 2010, **132**, 9522–9523.
59 S. Varenne, J. Buc, R. Lloubes and C. Lazdunski, *Journal of Molecular Biology*, 1984, **180**, 549–576.
60 H. Akashi, *Genetics*, 1994, **136**, 927–935.
61 G. Rosenblum, C. Chen, J. Kaur, X. Cui, H. Zhang, H. Asahara, S. Chong, Z. Smilansky, Y. E. Goldman and B. S. Cooperman, *J. Am. Chem. Soc.*, 2013, **135**, 11322–11329.
62 X. Qu, J.-D. Wen, L. Lancaster, H. F. Noller, C. Bustamante and I. Tinoco, *Nature*, 2011, **475**, 118–121.
63 J. Lu and C. Deutsch, *Journal of Molecular Biology*, 2008, **384**, 73–86.
64 G. Zhang, M. Hubalewska and Z. Ignatova, *Nat Struct Mol Biol*, 2009, **16**, 274–280.
65 S. J. Kim, J. S. Yoon, H. Shishido, Z. Yang, L. A. Rooney, J. M. Barral and W. R. Skach, *Science*, 2015, **348**, 444–448.
66 F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M. V. Rodnina and A. A. Komar, *Molecular Cell*, 2016, **61**, 341–351.
67 M. Zhou, J. Guo, J. Cha, M. Chae, S. Chen, J. M. Barral, M. S. Sachs and Y. Liu, *Nature*, 2013, **495**, 111–115.
68 C. Kimchi-Sarfaty, J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar and M. M. Gottesman, *Science*, 2007, **315**, 525–528.
69 C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco and C. Bustamante, *Science*, 2011, **334**, 1723–1727.
70 E. P. O'Brien, J. Christodoulou, M. Vendruscolo and C. M. Dobson, *J. Am. Chem. Soc.*, 2011, **133**, 513–526.
71 L. M. Alexander, D. H. Goldman, L. M. Wee and C. Bustamante, *Nature Communications*, 2019, **10**, 2709.
72 M. ZEEB, *Methods*, 2004, **34**, 65–74.
73 F. A. A. Mulder, L. Bouakaz, A. Lundell, M. Venkataramana, A. Liljas, M. Akke and S. Sanyal, *Biochemistry*, 2004, **43**, 5930–5936.
74 M. C. Wahl and W. Möller, *Curr. Protein Pept. Sci.*, 2002, **3**, 93–106.
75 P. Bernadó, K. Modig, P. Grela, D. I. Svergun, M. Tchorzewski, M. Pons and M. Akke, *Biophysical Journal*, 2010, **98**, 2374–2382.

76 A. M. E. Cassaignau, H. M. M. Launay, M.-E. Karyadi, X. Wang, C. A. Waudby, A. Deckert, A. L. Robertson, J. Christodoulou and L. D. Cabrita, *Nature Protocols*, 2016, **11**, 1492–1507.
77 C. A. Waudby, H. Launay, L. D. Cabrita and J. Christodoulou, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2013, **74**, 57–75.
78 L. D. Cabrita, S.-T. D. Hsu, H. Launay, C. M. Dobson and J. Christodoulou, *Proc Natl Acad Sci USA*, 2009, **106**, 22239–22244.
79 S.-T. D. Hsu, P. Fucini, L. D. Cabrita, H. Launay, C. M. Dobson and J. Christodoulou, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 16516–16521.
80 A. R. Buskirk and R. Green, *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 2017, **372**.
81 S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D. N. Wilson and R. Beckmann, *PLoS Biol*, 2011, **9**, e1000581.
82 J. Zhang, X. Pan, K. Yan, S. Sun, N. Gao and S.-F. Sui, *eLife*, 2015, **4**.
83 B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J.-P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz and R. Beckmann, *Science*, 2009, **326**, 1412–1415.
84 V. Shanmuganathan, N. Schiller, A. Magoulopoulou, J. Cheng, K. Braunger, F. Cymer, O. Berninghausen, B. Beatrix, K. Kohno, G. von Heijne and R. Beckmann, *eLife*, 2019, **8**.
85 S. J. Hersch, S. Elgamal, A. Katz, M. Ibba and W. W. Navarre, *J. Biol. Chem.*, 2014, **289**, 28160–28171.
86 J. G. E. S. D. W. R. B. K. Schulten, E. Schreiner, D. N. Wilson, R. Beckmann and K. Schulten, *Biophysical Journal*, 2012, **103**, 331–341.
87 J. Lu and C. Deutsch, *Biochemistry*, 2005, **44**, 8230–8243.
88 F. Cymer, R. Hedman, N. Ismail and G. von Heijne, *J. Biol. Chem.*, 2015, **290**, 10208–10215.
89 K. Pervushin, R. Riek, G. Wider and K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, **94**, 12366–12371.
90 C. FERNANDEZ, *Current Opinion in Structural Biology*, 2003, **13**, 570–580.
91 C. Eichmann, S. Preissler, R. Riek and E. Deuerling, *Proc Natl Acad Sci USA*, 2010, **107**, 9111–9116.
92 N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis and L. E. Kay, *J. Biomol. NMR*, 1999, **13**, 369–374.
93 L. E. Kay, *J. Magn. Reson.*, 2011, **210**, 159–170.
94 M. A. Danielson and J. J. Falke, *Annu. Rev. Biophys. Biomol. Struct.*, 1996, **25**, 163–195.
95 J. G. Bann, J. Pinkner, S. J. Hultgren and C. Frieden, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 709–714.
96 D. Rose-Sperling, M. A. Tran, L. M. Lauth, B. Goretzki and U. A. Hellmich, *Biol. Chem.*, 2019, **400**, 1277–1288.
97 C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl and G. J. Pielak, *J. Am. Chem. Soc.*, 2010, **132**, 321–327.
98 J. Völler, H. Biava, B. Koksch, P. Hildebrandt and N. Budisa, *ChemBioChem*, 2015, **16**, 742–745.
99 J. Broos, E. Gabellieri, E. Biemans-Oldehinkel and G. B. Strambini, *Protein Sci.*, 2003, **12**, 1991–2000.
100 J.-S. Völler, T. M. Thi To, H. Biava, B. Koksch and N. Budisa, *Enzyme Microb. Technol.*, 2017, **106**, 55–59.
101 D. Rovnyak, J. C. Hoch, A. S. Stern and G. Wagner, *J. Biomol. NMR*, 2004, **30**, 1–10.
102 C. M. Quinn, M. Wang and T. Polenova, *Methods Mol. Biol.*, 2018, **1688**, 1–35.
103 H. Kovacs, D. Moskau and M. Spraul, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2005, **46**, 131–155.
104 P. Schanda, Ē. Kupče and B. Brutscher, *J. Biomol. NMR*, 2005, **33**, 199–211.
105 S. Gil, T. Hošek, Z. Solyom, R. Kümmerle, B. Brutscher, R. Pierattelli and I. C. Felli, *Angew. Chem. Int. Ed.*, 2013, **52**, 11808–11812.
106 Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold and B. Brutscher, *J. Biomol. NMR*, 2013, **55**, 311–321.
107 J. Rantaharju, J. Mareš and J. Vaara, *The Journal of Chemical Physics*, 2014, **141**, 014109.
108 S. Cai, C. Seu, Z. Kovacs, A. D. Sherry and Y. Chen, *J. Am. Chem. Soc.*, 2006, **128**, 13474–13478.
109 F.-X. Theillet, A. Binolfi, S. Liokatis, S. Verzini and P. Selenko, *J. Biomol. NMR*, 2011, **51**, 487–495.
110 S. H. S. Chan, C. A. Waudby, A. M. E. Cassaignau, L. D. Cabrita and J. Christodoulou, *J. Biomol. NMR*, 2015, **63**, 151–163.
111 C. A. Waudby and J. Christodoulou, *Journal of Magnetic Resonance*, 2012, **219**, 46–52.

112 B. Simon and H. Köstler, *J. Biomol. NMR*, 2019, **73**, 155–165.
113 R. Augustyniak, F. Ferrage, R. Paquin, O. Lequin and G. Bodenhausen, *J. Biomol. NMR*, 2011, **50**, 209–218.
114 J. Hrabe, G. Kaur and D. N. Guilfoyle, *J Med Phys*, 2007, **32**, 34–42.
115 R. Augustyniak, F. Ferrage, C. Damblon, G. Bodenhausen and P. Pelupessy, *Chem. Commun.*, 2012, **48**, 5307.
116 D. Ekman, A. K. Björklund, J. Frey-Skött and A. Elofsson, *Journal of Molecular Biology*, 2005, **348**, 231–243.
117 D. C. Dalgarno, B. A. Levine and R. J. Williams, *Biosci. Rep.*, 1983, **3**, 443–452.
118 D. S. Wishart and B. D. Sykes, *Meth. Enzymol.*, 1994, **239**, 363–392.
119 R. Horst, G. Wider, J. Fiaux, E. B. Bertelsen, A. L. Horwich and K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 15445–15450.
120 G. M. Clore and J. Iwahara, *Chem. Rev.*, 2009, **109**, 4108–4139.
121 A. M. Knight, P. H. Culviner, N. Kurt-Yilmaz, T. Zou, S. B. Ozkan and S. Cavagnero, *ACS Chem. Biol.*, 2013, **8**, 1195–1204.
122 D. V. Fedyukina, T. S. Jennaro and S. Cavagnero, *J. Biol. Chem.*, 2014, **289**, 6740–6750.
123 P. E. Schavemaker, W. M. Śmigiel and B. Poolman, *eLife*, 2017, **6**.
124 N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 905–920.
125 P. Kukic, A. Kannan, M. J. J. Dijkstra, S. Abeln, C. Camilloni and M. Vendruscolo, *PLoS Comput. Biol.*, 2015, **11**, e1004435.
126 I. Buskiewicz, E. Deuerling, S.-Q. Gu, J. Jöckel, M. V. Rodnina, B. Bukau and W. Wintermeyer, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 7902–7906.
127 J. H. Peterson, C. A. Woolhead and H. D. Bernstein, *Molecular Microbiology*, 2010, **30**, no–no.
128 R. Kudva, P. Tian, F. Pardo-Avila, M. Carroni, R. B. Best, H. D. Bernstein and G. von Heijne, *eLife*, 2018, **7**.
129 J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu and K. Schulten, *Current Opinion in Structural Biology*, 2015, **31**, 64–74.
130 M. Vendruscolo and C. M. Dobson, *Curr. Biol.*, 2011, **21**, R68–70.
131 K. Y. Sanbonmatsu, *Current Opinion in Structural Biology*, 2012, **22**, 168–174.
132 T. Rauch, H. A. Hundley, C. Pfund, R. D. Wegrzyn, W. Walter, G. Kramer, S.-Y. Kim, E. A. Craig and E. Deuerling, *Molecular Microbiology*, 2005, **57**, 357–365.
133 K. Ito, *Molecular Microbiology*, 2005, **57**, 313–317.
134 S. K. Lakshmipathy, R. Gupta, S. Pinkert, S. A. Etchells and F. U. Hartl, *J. Biol. Chem.*, 2010, **285**, 27911–27923.
135 A. Hoffmann, B. Bukau and G. Kramer, *BBA - Molecular Cell Research*, 2010, **1803**, 650–661.
136 L. Ferbitz, T. Maier, H. Patzelt, B. Bukau, E. Deuerling and N. Ban, *Nature*, 2004, **431**, 590–596.
137 C. M. Kaiser, H.-C. Chang, V. R. Agashe, S. K. Lakshmipathy, S. A. Etchells, M. Hayer-Hartl, F. U. Hartl and J. M. Barral, *Nature*, 2006, **444**, 455–460.
138 T. Saio, X. Guan, P. Rossi, A. Economou and C. G. Kalodimos, *Science*, 2014, **344**, 1250494–1250494.
139 R. Maier, B. Eckert, C. Scholz, H. Lilie and F.-X. Schmid, *Journal of Molecular Biology*, 2003, **326**, 585–592.
140 K. Liu, K. Maciuba and C. M. Kaiser, *Molecular Cell*, 2019, **74**, 310–319.e7.
141 S. Haldar, R. Tapia-Rojo, E. C. Eckels, J. Valle-Orero and J. M. Fernandez, *Nature Communications*, 2017, **8**, 668.
142 L. Morgado, B. M. Burmann, T. Sharpe, A. Mazur and S. Hiller, *Nature Communications*, 2017, **8**, 1992.
143 T. Saio, S. Kawagoe, K. Ishimori and C. G. Kalodimos, *eLife*, 2018, **7**.
144 V. R. Agashe, S. Guha, H.-C. Chang, P. Genevaux, M. Hayer-Hartl, M. Stemp, C. Georgopoulos, F. U. Hartl and J. M. Barral, *Cell*, 2004, **117**, 199–209.
145 M. Sarkar, C. Li and G. J. Pielak, *Biophys Rev*, 2013, **5**, 187–194.
146 M. Feig, I. Yu, P.-H. Wang, G. Nawrocki and Y. Sugita, *J. Phys. Chem. B*, 2017, **121**, 8009–8025.
147 A. Christiansen, Q. Wang, M. S. Cheung and P. Wittung-Stafshede, *Biophys Rev*, 2013, **5**, 137–145.

148 W. B. Monteith, R. D. Cohen, A. E. Smith, E. Guzman-Cisneros and G. J. Pielak, *Proc Natl Acad Sci USA*, 2015, **112**, 1739–1742.
149 S. Majumder, J. Xue, C. M. DeMott, S. Reverdatto, D. S. Burz and A. Shekhtman, *Biochemistry*, 2015, **54**, 2727–2738.
150 D. S. Burz, K. Dutta, D. Cowburn and A. Shekhtman, *Nature Protocols*, 2006, **1**, 146–152.
151 R. Hänsel, L. M. Luh, I. Corbeski, L. Trantirek and V. Dötsch, *Angew. Chem. Int. Ed.*, 2014, **53**, 10300–10314.
152 P. Selenko, D. P. Frueh, S. J. Elsaesser, W. Haas, S. P. Gygi and G. Wagner, *Nat Struct Mol Biol*, 2008, **15**, 321–329.
153 J. Danielsson, X. Mu, L. Lang, H. Wang, A. Binolfi, F.-X. Theillet, B. Bekei, D. T. Logan, P. Selenko, H. Wennerström and M. Oliveberg, *Proc Natl Acad Sci USA*, 2015, **112**, 12402–12407.
154 W. B. Monteith and G. J. Pielak, *Proc Natl Acad Sci USA*, 2014, **111**, 11335–11340.
155 M. P. Latham and L. E. Kay, *Journal of Molecular Biology*, 2014, **426**, 3214–3220.
156 G. Xu, Y. Ye, X. Liu, S. Cao, Q. Wu, K. Cheng, M. Liu, G. J. Pielak and C. Li, *Biochemistry*, 2014, **53**, 1971–1981.
157 C. A. Waudby, M. D. Mantle, L. D. Cabrita, L. F. Gladden, C. M. Dobson and J. Christodoulou, *J. Am. Chem. Soc.*, 2012, **134**, 11312–11315.