NMR investigations of co-translational protein folding on ribosomal particles

Thesis submitted by
Hélène M.M. Launay

For the degree of
Doctor of Philosophy in Biochemistry and Structural Biology

Research Department of Structural and Molecular Biology University College London
Gower Street, London, WC1E 6BT, UK

November 28, 2011

Declaration

I, **Hélène M.M. Launay**, declare that all the work presented in this thesis is the result of my work only. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The work herein was carried out while I was a graduate student at the University College London, Research Department of Structural and Molecular Biology, under the supervision of Dr John Christodoulou.

Abstract

The 2.4 MDa ribosome complex is responsible for protein synthesis in all domains of life. During its biosynthesis, the nascent polypeptide chain (NC) threads through the ribosomal exit tunnel and into the cellular milieu. There is much evidence to indicate that during translation and whilst tethered to the ribosome, the NC has its first opportunity to acquire structure, which assists in its folding to the active biological state, in a process known as co-translational folding. The studies of the structural and molecular determinants of this process present a challenge due to the NC's intrinsic conformational heterogeneity.

NMR spectroscopy has the unique ability to report on both protein structure and dynamic at a residue-specific level and this thesis describes the development of NMR methodologies to allow monitoring the progressive folding of an immunoglobulin domain (ddFLN-dom5) NC as it emerges from the ribosome. Snapshots of the emergence of ddFLN-dom5 from the ribosome were generated using different lengths of *in vivo* translated, homogeneously stalled and selectively labelled ribosome-bound NCs (RNC) which are then extracted from *E. coli* cells for NMR analysis. A strategy is described that allows monitoring *in situ* the integrity of the RNC samples, and the attachment of the NC to its parent ribosome.

Despite the high-molecular-weight of the ribosomal complexes, their instability and the low achievable sample concentrations, a range of useful NMR tools are being developed. Importantly, comparisons of ^{1}H-^{13}C methyl-TROSY HMQC and ^{1}H-^{15}N SOFAST-HMQC NMR spectra of the ddFLN-dom5-RNCs with the isolated domain in both native and denatured conditions allows the detailed analysis of the folding equilibrium of the RNC. A robust data analysis methodology was designed to optimise the significance of low signal to noise spectra. These NMR data reveal clear evidence for co-translational folding when the C-terminal end of the ddFLN-dom5 is at lengths $\geq$47 residues from the peptidyl transferase centre (PTC). At this and longer linking lengths, the chemical shifts observed for ddFLN-dom5-RNC are identical to those of the isolated native domain. The RNC resonances show heterogeneous linewidth indicative of conformational exchange between native and non-native states on the order of the chemical-shift timescale (ms).

Overall, this study sets the stage for future opportunities for investigations of the structural and dynamical properties of RNCs at a residue-specific level.

Results shown in this thesis were presented in the following papers:

S.-T. D. Hsu, P. Fucini, L. D. Cabrita, H. Launay, C. M. Dobson, J. Christodoulou, **Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy**, *Proc Natl Acad Sci USA*, **104**(42) pp 16516 (2007)

L. D. Cabrita, S.-T. D. Hsu, H. Launay, C. M. Dobson, J. Christodoulou, **Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy**, *Proc Natl Acad Sci USA*, **106**(52) pp 22239 (2009)

F. Blombach, H. Launay, V. Zorraquino, D. C. Swarts, L. D. Cabrita, D. Benelli, J. Christodoulou, P. Londei, J. van der Oost, **An HflX-type GTPase from Sulfolobus solfataricus binds to the 50S ribosomal subunit in all nucleotide-bound states**, *J Bacteriol*, **193** (11) pp 2861

Results shown in this thesis were presented in the following meetings:

Advance NMR Practical EMBO course, Il Ciocco, Italy (Aug 2008) poster presentation

NMR GpP meeting, Paris, France (13 Feb 2009) Short oral presentation

Advance NMR Practical EMBO course, Munich, Germany (Aug 2009) poster presentation

Keystone symposium, Frontiers of NMR in Biology, Montana, USA (Jan 2011) Short oral presentation in plenary session & poster presentation

Acknowledgements

The work presented in this thesis was carried out in the Department of Structural and Molecular Biology, University College London, London UK. I am indebted to the Countess of Lisbourne and the head of Department of Structural and Molecular Biology Gabriel Waksman for the generous financial support that allowed me to undertake my PhD studies.

First, I would like to express my sincere gratitude to Dr John Christodoulou for his supervision, continuous support and advice during the course of my thesis. He has been very understanding, while always keeping high expectation for my work, which made my PhD study a very inspiring time. John is a stimulating scientist whom I would like to keep working with. I would also like to thank Prof Paul Driscoll and Prof Christine Orengo for sitting in my PhD committee meeting.

I am grateful to my thesis examinors Dr Paul Barker and Dr Flemming Hansen who have taken the time to edit my thesis with great details, and with whom I had a great scientific discussion during my viva.

I would like to thank all members of my group as well as Dr John Kirkpatrick. John has always been there to provide the well appreciated help to learn and perform NMR. Dr Lisa Cabrita has been of tremendous help for the biochemistry, and this project relies on her inspiring biochemical expertise. I also consider as a privilege to work with Dr Christopher Waudby, without whom a lot of the NMR data presented in this study would not have reached such a level of analysis. I would like to thank the other PhD students of our group, Maria-Evangelia Karyadi and Xiaolin Wang with whom I worked closely, as well as Luke Goodsell, Annika Weise, Géraldine Levy, Anne Wentink, and TOshitaka TAjima. I am also grateful to the visiting students I worked with Piotr Gierszewski and Johanna Reul whom I am convinced will decome first class scientists. I am also grateful to the collaborators I worked with: Fabian Blombach and Daniel Sohmen, without whom the work with archaeal ribosome would have not be possible.

Last but not least I would like to thank Goodenough college where I met inspiring people, my friends from UCL, my parents, my familly and Sidi Souvi for their patience and support at all time.

Contents

List of abbreviations

AFM	Atomic force microscopy
ApoMb	apomyglobin
ATP	Adenosine triphosphate
BEST	Band-selective excitation short-transient
CFTR	Cystic fibrosis transmembrane conductance regulator
CPMG	Carr-Purcell- Meiboom-Gill
CRINEPT	Cross-correlated relaxation-enhanced polarization transfer
CRIPT	Cross- correlated relaxation-induced polarization transfer
cryo-EM	Cryo-electron microscopy
CSA	Chemical shift anisotropy
ddFLN	Filamin gelation factor from *Dictyostelium discoideum*
ddFLN-dom5	domain 5 from ddFLN
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EF-G	Elongation factor G
EF-Tu	Elongation factor thermo unstable
FID	Free induction decay
fnIII	Fibronectin type III domains
FRET	Förster resonance energy transfer
GTP	Guanosine-5′-triphosphate
H-D	Hydrogen/Deuterium
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HMQC	Heteronuclear multiple quantum coherence
HRP	Horseradish Peroxidase
HSQC	Heteronuclear single quantum coherence
INEPT	Insensitive nuclei enhanced by polarization transfer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Daltons
$\Delta\nu_{^1\mathrm{H}}$	^{1}H linewidth
MALDI	Matrix-assisted laser desorption/ionization

MALDI-TOF	MALDI time-of-flight
MD	Molecular dynamic
MDa	Mega Daltons
MDG	Phosphate (M), L-aspartic acid (D) and glucose (G) based medium
mRNA	messenger RNA
MW	Molecular weight
NC	Nascent chain
Ni-IDA	Nickel charged beads (chelating ligand: iminodiacetic acid)
Ni-NTA	Nickel charged beads (chelating ligand: nitrilotriacetic acid)
NMR	Nuclear Magnetic Resonance
NS	Number of scans
NUS	Non uniform sampling
OD	Optical density
P. aero	*Pyrobaculum aerophilum*
PFG	Pulsed-field gradient
ppm	part per million
PRE	Paramagnetic relaxation enhancement
PTC	peptidyl transferase center
RDC	Residual dipolar coupling
r_{h}	hydrodynamic radius
RNA	Ribonucleic acid
RNC	Ribosome-bound nascent chain
ROI	Region of interest
rpm	Rotations per minute
rprotein	Ribosomal protein
rRNA	Ribosomal RNA
S (S_{LZ})	Order parameter
SecM	Secretion monitor protein
S_{entropy}	Conformational entropy
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOFAST-HMQC	Band-selective optimized-flip-angle short-transient HMQC
SN	Signal to noise ration
SSF	*Sulfolobus Solfataricus*
STE	Stimulated echo

τ_c	Rotational correlation time
TI	Titin domain
tRNA	transfer RNA
wf	window function
w/v	Weight per volume
zf	Zero filling
XSTE	Heteronuclear STE

Chapter 1

Introduction

1.1 Introduction to co-translational folding

It is well known that the biological activity of a protein is linked to its three dimensional structure, and the information required for folding is contained within the amino-acid sequence of the polypeptide chain [1–3]. The ribosome, which is the cellular apparatus responsible for protein biosynthesis within all living systems, produces and eventually releases the linear polypeptide chains into the cellular environment. It is known that the ribosome participates in the folding process of new, emerging polypeptides [4–7], but the details of how a fledgling nascent chain (NC) can adopt its biologically active fold and avoids alternative misfolded conformations while on the ribosome is not entirely clear. This thesis describes studies using nuclear magnetic resonance (NMR) spectroscopy to investigate NCs on their parent ribosome.

In this Chapter, a general background to the ribosome, protein folding and *co-translational protein folding* is described. The description of co-translational folding includes direct structural evidence already obtained experimentally about the emerging nascent polypeptide via different techniques, including cryo-electron microscopy (cryo-EM) and fluorescent anisotropy decay studies, as well as predictions derived from computational simulation approaches. Since NMR spectroscopy is the primary tool used in this thesis, the NMR methods that have been used during this study are also

introduced.

1.1.1 Ribosome structure

Over the past decade, a wealth of ribosome structures have been determined via investigations using X-ray crystallography [8–15] and cryo-EM [16–22]. These have revealed the inner workings of the ribosome at high resolution at different stages of translation [23, 24] (Figure 1.1). The *initiation* step is characterised by the recognition of the translation initiation site of the mRNA by the small ribosomal subunit (termed 30S in bacteria and archaea), and the recruitment of the large ribosomal subunit (termed 50S in bacteria and archaea) with the help of translation initiation factors. During *elongation*, the 30S subunit recruits the anti-codon acylated transfer RNA (tRNA) that complements the mRNA coding sequence, with the help of elongation factor EF-Tu. The recognition of the cognate tRNA triggers GTP hydrolysis by EF-Tu, which provides the energy to accommodate the acylated tRNA at the A-site (Figure 1.1C) [27, 28]. This is followed by peptide bond formation between the incoming amino-acid loaded on the P-site tRNA and the growing nascent chain at the peptidyl transferase centre (PTC, Figure 1.1B) located at the heart of the 50S subunit. The elongation process results in the synthesis of the nascent polypeptide from the N-terminal to the C-terminal end, which then traverses through the 50S subunit via a constricted 100Å-long tunnel, in which the last 20Å is thought to be wider and is termed the vestibule or "exit port" (Figure 1.1B). During *translocation*, which describes the progression of the ribosome complex along the mRNA from the 5′ end to the 3′ end, the A-site tRNA is first transferred to the P-site where the peptidyl-transfer reaction occurs, and the empty tRNA is then translocated to the E-site to be expelled from the ribosome [24] – with the help of the elongation factor EF-G and GTP hydrolysis (illustrated in Figure 1.1C). *Termination* marks the release of the NC

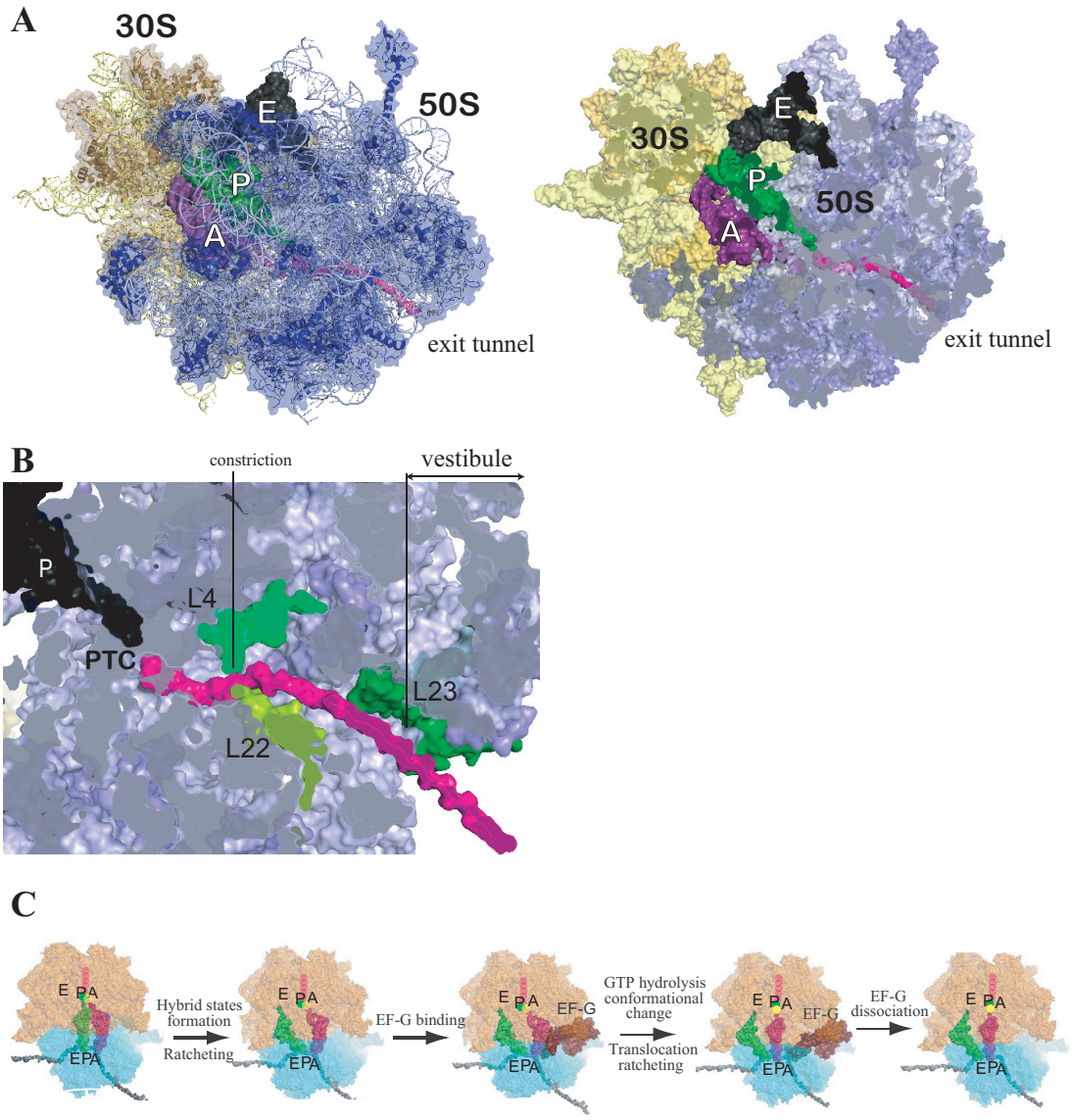


Figure 1.1: **A:** Representation of the 70S *E. coli* ribosomal complex, with the 50S (blue, pdb file *2AVY.pdb*) and the 30S (yellow, pdb file *2AW4.pdb*) subunits [25]. The A-, P- and E-site tRNAs are represented in purple, green and black respectively (adapted from the cryo-EM structure of the archaeal ribosome, pdb file *2J00.pdb* [20]). To highlight the location of the ribosomal exit tunnel, an outline of the nascent polypeptide chain is shown in magenta (aligned with the cryo-EM density of the TnaC stalling sequence in stalled bacterial ribosomes [26]). **B:** Cross-section of the ribosomal exit tunnel, with the three ribosomal proteins that line the ribosome tunnel wall L4, L22 and L23 (the latter being also the universal docking site for auxiliary factor binding [6]). The two main structural features of the tunnel, the constriction and the exit vestibule are depicted. **C:** The tRNA translocation from the A- to P-site and P- to E-site is shown, adapted from Schmeing *et al* [24], with the 50S subunit in gold and the 30S in blue. The nascent polypeptide is shown in magenta, with the amino-acid *i* in green and *i+1* in yellow.

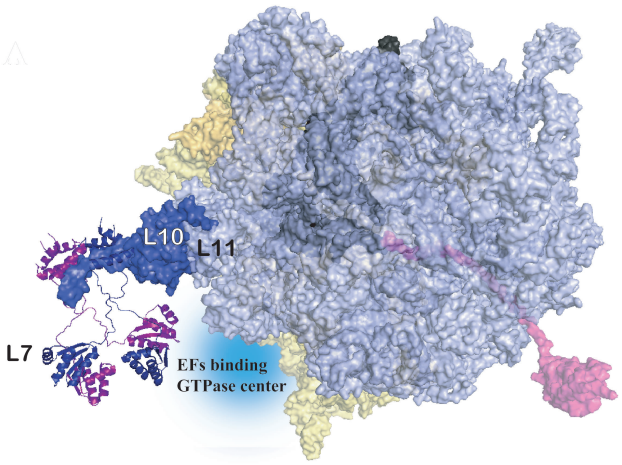


Figure 1.2: A model of the 70S particle highlighting the L7/L12 stalk region. Two L7 dimers are represented as ribbons, from pdb *1RQU.pdb* [31], with L10 represented as the blue surface, from pdb *1ZAV.pdb* [32]. The 50S, 30S, tRNA and model nascent polypeptide are identical to Figure 1.1. Docking of the L10 structure to the 70S ribosome was achieved by using structural alignment with pdb file *2WRJ.pdb* [14].

from the PTC with assistance from release factors and the recycling of the individual ribosomal subunits.

1.1.1.1 Dynamics of elongation

Protein translation is a dynamic process [29] in particular during the stages of tRNA recognition and translocation [30]. The tRNA selection begins with the binding of EF-Tu to the flexible L7/L12 stalk region of the 50S subunit [30]. The stalk region is intrinsically flexible, and consists of multiple copies of L7 (four in bacteria, six in eukaryotes and archaea), whose C-terminal domains are linked to the ribosome-bound N-terminal domain via a 20-residue-long disordered loop (Figure 1.2). The L7/L12 stalk region has not been resolved completely in the numerous cryo-EM and X-ray structures of ribosome complexes due to its inherent dynamic nature, whereby it acts as a “fishing

line" for the recruitment of EF-Tu. Cryo-EM relies on the averaging of images from an ensemble of homogeneous molecules, and hence the variety in conformations of the stalk region prevents good image averaging. NMR is therefore a suitable tool to investigate the dynamics of the stalk region [33, 34] and its interaction with EF-Tu. The C-terminal domains of the L7/L12 proteins were found to mediate binding with EF-Tu [35] at the 30S-50S interface (Figure 1.2).

Recognition of cognate tRNA is also a highly dynamic process that involves allosteric communication between the decoding centre of the 30S subunit within the A-site, and the GTP hydrolysis centre of EF-Tu. Structural snapshots of the GTP hydrolysis obtained through the use of non-hydrolysable GTP analogues provided evidence of the allosteric mechanism, which is mediated by large conformational changes in the tRNA [30, 35]. Prior to GTP hydrolysis, differential kinetics for transient binding of non-cognate or cognate tRNA at the decoding centre of the A-site were observed by Förster resonance energy transfer (FRET) studies, and the kinetic selection of these transient interactions was found to be critical for tRNA selection [36].

Following tRNA recognition, the *ratcheting* mechanism that induces translocation is also known to be a highly dynamic process. The transfer of the acylated tRNA from the A-site to the P-site and of the deacylated tRNA from the P-site to the E-site after peptidyl transfer to the growing NC, involves large conformational changes divided into two stages: first there is the counterclockwise ratchet-like rotation of the 30S subunit with respect to the 50S subunit (Figure 1.1C). Then, secondly, binding of EF-G and GTP hydrolysis induces the relaxation of the ratcheted state (Figure 1.1C), which completes translocation. The oscillations at room temperature of the ribosome structure between the "ratcheted" and "unracheted" states, and the hybrids structures between these states (Figure 1.1C) were observed via a combination of FRET [37], cryo-EM [38] and crystal

structures [35], and suggests that the ribosome is a Brownian motor: i.e. the translation machinery oscillates between different conformations separated by low energy barriers, and only the binding of EF-G and GTP hydrolysis is able to drive the oscillations towards effective translocation [30].

Finally, one aspect of the translation machinery to be considered, and which is perhaps the most interesting from the point of view of the growing nascent chain, is the flexibility of the PTC centre and the ribosomal exit tunnel. Only recently, have cryo-EM studies of ribosome-bound NC (RNC) shed light into the intimate set of interactions between P-site tRNA, the ribosome exit tunnel and the NC [26, 39–41]. In particular, structures of SecM-stalled RNCs suggest that a set of interactions between the ribosome exit tunnel and the RNC induces conformational changes of the A-site tRNA that prevents translocation [40], again revealing allostery between the ribosome tunnel and the tRNAs. Also, the cryo-EM studies showed that the structure of the ribosomal RNAs that form the PTC centre was altered in cases where the stalling sequence of a TnaC leader peptide formed specific contacts with the ribosome exit tunnel.

Computational studies are often used to shed light into a range of biological processes such as molecular mechanisms, protein-protein interactions and protein folding. However, because of the size of the ribosome complex, computational studies of the entire translation machinery are only just beginning to emerge; these have enabled insights into the mechanism of translocation and allowed models of the dynamic pathway of the ribosome cycle to be generated [42, 43]. A number of computational studies have also sought to examine structural aspects of the ribosome, such as the exit tunnel. These studies suggest a teflon-like property, presumably to allow the NC to traverse the tunnel unhindered [44, 45]. Moreover, these studies are being extended to the NC itself [46–48], and are aimed at modelling the conformational space a NC can

begin to sample upon emergence from the ribosome exit tunnel.

1.1.2 Protein folding

Over the last 50 years, a large number of studies have attempted to understand the fundamental principles underlying the folding of a polypeptide chain. In 1973, Anfinsen proposed the hypothesis of protein folding [1] in answer to Levinthal's paradox [49], which stated that the number of possible conformations accessible to a polypeptide chain (10^{300} for a 150-residue protein [49]) is far greater than the number that could possibly be sampled in a reasonable time. Anfinsen showed that the amino acid sequence encodes the necessary information to drive the conformational sampling into the unique low-energy state, i.e. the sequence drives protein folding [1]. A vast range of protein folding studies have followed, which include the characterisation of stable folds, the analysis of folding rates under varying conditions [50], the investigation of transition-state ensembles [51] and of potential stable, obligatory or non-obligatory intermediate states [52,53]. The majority of studies of protein folding *in vitro* have used isolated proteins and have proposed that polypeptide chains explore only a small number of all the possible conformations within the energetic folding funnel [50, 52–54] (Figure 1.3A) to acquire the biologically active structure, which is usually the most energetically preferred state, and which is often populated within 1 ms (for simple folds) [3, 7]. It is well-established from a range of studies (for recent reviews see [55]) that, during the folding process, an isolated polypeptide chain does not follow a simple folding route, but that multiple folding pathways coexist, all leading to the same native state. Detailed investigation of native states have led to the view that different conformations can exist in equilibrium under native conditions, separated by energy barriers that correspond to transition state ensembles [51]. Classically, mutational analysis and comparison of folding rates have

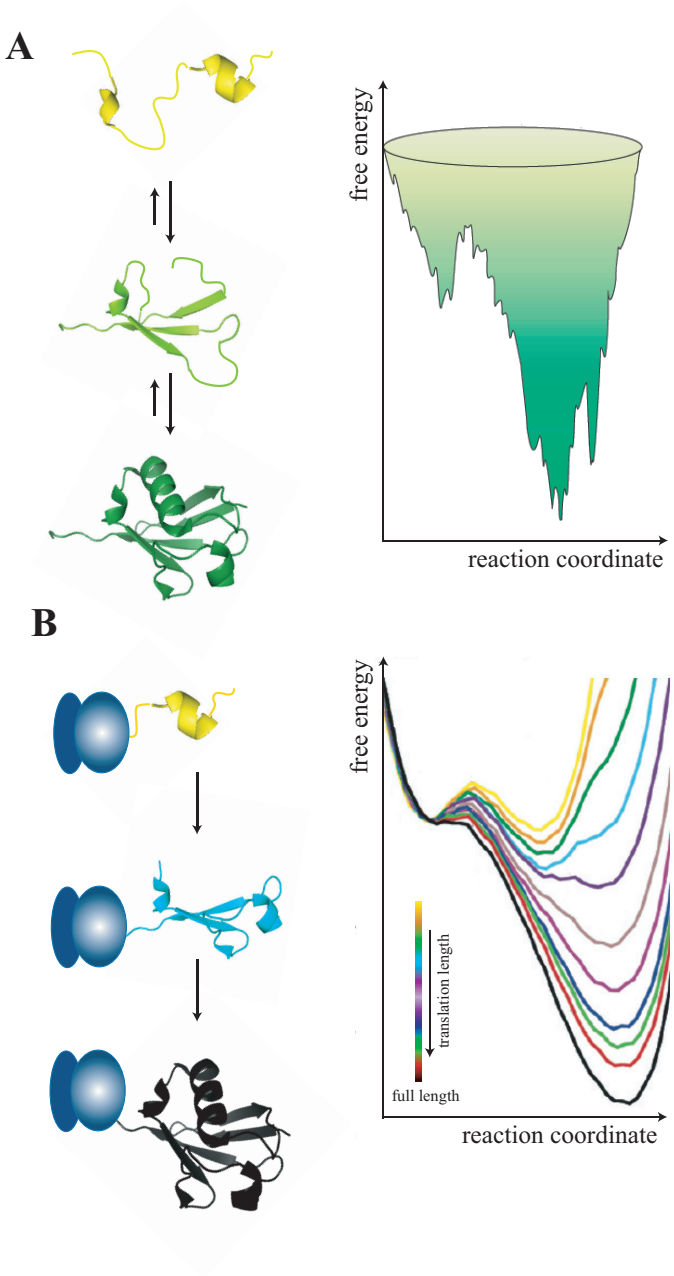


Figure 1.3: **A:** Schematic of protein conformational search during *in vitro* re-folding (left), with the corresponding free-energy landscape (right), adapted from [7]. **B:** In contrast, during biosynthesis (the ribosome is represented by the blue spheres), the free-energy landscape of the nascent polypeptide during synthesis is constantly modified (right, adapted from [48]).

revealed the nature of transition state ensembles [50], which have often been found to make native-like contacts [56]. However, a recent detailed study suggested that non-native contacts can also be formed transiently [57]. Folding of small domains has been observed to be highly cooperative [54], but experimental results, notably from hydrogen-exchange studies [52,53], have suggested that populations of native-like partially folded intermediate states can typically assist in the acquisition of the native structure by driving the polypeptide forwards along the most efficient downhill path of the folding funnel. These intermediate states are often referred to as collapsed states, i.e. compact heterogeneous structural ensembles without persistent secondary or tertiary structure [4]. Intermediate states, and in particular molten globules [58] also have the propensity to misfold and lead to the formation of kinetically trapped aberrant species, which can result in aggregation or in some instances reorganise to thermodynamically stable ordered structures such as amyloid fibres [59]. Numerous studies have investigated the relationship between protein misfolding and human diseases [60], as well as how protein misfolding is avoided in the cellular milieu [2].

Experimental investigations of protein folding often require trapping thermodynamically unstable states such as disordered states using a chemical denaturant (guanidinium chloride or urea). Biophysical strategies such as measuring the thermal stability or hydrogen-deuterium exchange rates are established tools for determining the thermodynamic properties of the conformations adopted by the polypeptide chain under these conditions. High-resolution spectroscopic methods are increasingly being used to extract further details of protein folding. For example, atomic force microscpy (AFM) has been used to understand two-stage folding of single domains through mechanically-induced unfolding [61]. Also, NMR spectroscopy has been widely used to characterise the ensemble of conformations adopted by a polypeptide chain during protein folding

[62], under conditions close to native conditions, including low-populated metastable states [63,64] and transition state ensembles [51].

1.1.3 Co-translational protein folding

Studies of protein folding have largely examined isolated proteins, where the entire amino-acid sequence is available throughout the folding process. In living systems, the paradigm of protein folding is different, in that the NC emerges from the ribosomal exit tunnel one amino-acid at a time, into the cellular milieu. This emerging N-terminal region of a nascent polypeptide has the opportunity to explore conformational space before the entire sequence is available for folding [5] (Figure 1.3B).

1.1.3.1 Comparison of folding rate and elongation rate

The maximum rate of folding into the native state is limited by the rate at which the nascent polypeptide chain emerges from the exit tunnel, which is in turn dictated by the rate of peptide synthesis at the PTC, *ca.* 43-137 peptide-bonds/min in *E. coli* [65,66] (120-180 peptide-bonds/min within a eukaryotic system [67]). There are very few studies reporting the *de novo* folding rate [67], but a notable example is the measurement of the folding timescale of the multidomain protein CFTR as the NC emerges from the eukaryotic ribosomal exit tunnel (upon cleavage of the peptidyl-tRNA bond) in real time by FRET. The folding time was found to under 2 min, implying that the folding rate was close to the synthesis rate of this 111-residue domain [67]. Protein translation rates are known to be highly regulated and are found to depend on the codon composition of the transcript [65, 66]. The clustering of codons that pair to low abundance tRNA appears to have an increased probability of occurring at *ca.* 19-20 residues downstream of the C-terminal end of individual domains within multi-domain proteins, resulting

in slow-translating stretches at domain boundaries [66]. The attenuation of translation rates downstream of an individual domain is thought to be be necessary for the correct folding of full-length Suf1 protein, as shown by limited proteolysis experiments [65]. As the ribosomal tunnel can hold 24-40 residues [41, 47], rare codon clusters 20 residues downstream of the C terminal end of domains would appear to allow the newly synthesised polypeptide chain to explore conformational space and to reach a transient thermodynamic equilibrium prior to the synthesis of the rest of the protein. In the majority of studies of co-translational folding detailed below (as well as in this PhD study), translation is arrested in order to achieve thermodynamic equilibrium (Section 2.1.1), mimicking situations where the elongation rate is slow, and the folding rate rapid. This transient equilibrium (i.e. reshaping the energy landscape as in Figure 1.3 B) might not be attained in cases of fast elongation rates and slow folding, as is perhaps the case within domains. While the structural studies of stalled nascent chains are necessary to probe the folding process and provide parallel information to kinetics studies, ultimately, equilibrium studies will need to be combined with kinetic studies.

1.1.3.2 Biochemical evidence for co-translational protein folding

The nascent polypeptide within the ribosomal exit tunnel is highly confined (Figure 1.1B, [26,40,41]), and it is only when the N-terminal region of the NC reaches the exit vestibule (the last 20 Å of the ribosome exit tunnel, Figures 1.1B and 1.4) that it begins to have increased conformational freedom [68]. Indeed, there is increasing evidence to suggest that during synthesis and prior to its completion, the emerging NC can begin to sample conformations that assist its folding to its biologically active state in a process known as *co-translational folding* [5, 68]. An understanding of the structure and dynamics of the NC as it exits the ribosome will provide a detailed picture how protein folding takes

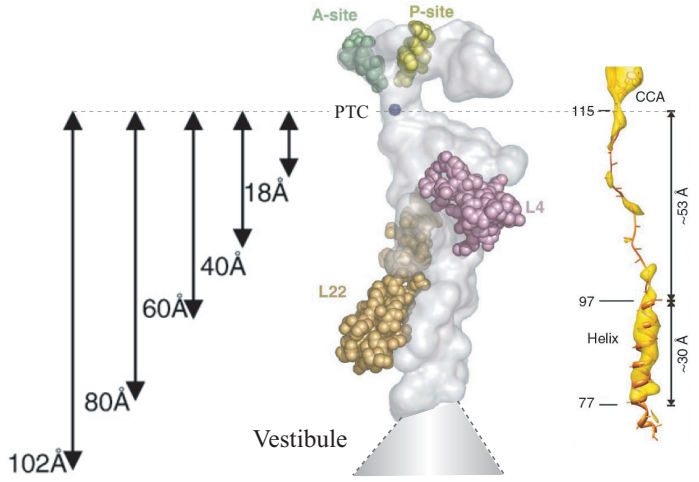


Figure 1.4: Surface of the inside of the ribosome tunnel (from [68,69]). As in [69], an extra 20 Å is added to account for the longer tunnel as reported in the crystal structures of the ribosome [8,70]. The electron density observed for a nascent chain of dipeptidylaminopeptidase B [41] is shown on the right at the same scale.

place within the cellular environment, and importantly, if and how this process differs to that observed *in vitro* (Section 1.1.2). At present, very little structural information is available for the NC during synthesis, although biochemical studies have indicated that the NC has the capacity to orchestrate certain cellular events that mediate downstream processes. Prior to the completion of synthesis for example, the appearance of a signal peptide sequence within the ribosomal exit tunnel can invoke the recruitment of the signal recognition particle (SRP), which, together with the Sec machinery, co-ordinates an intricate translocation mechanism, a process which can occur co-translationally [39]. Much biochemical and biophysical evidence has shown that, whilst still tethered to the ribosome, the NC can acquire catalytic activity [71–73] (Figure 1.5B), form disulphide bonds [74,75] and also recognise conformation-dependent antibodies [76] (Figure 1.5A). Limited proteolysis studies (Figure 1.5D) have revealed that for the multi-domain protein

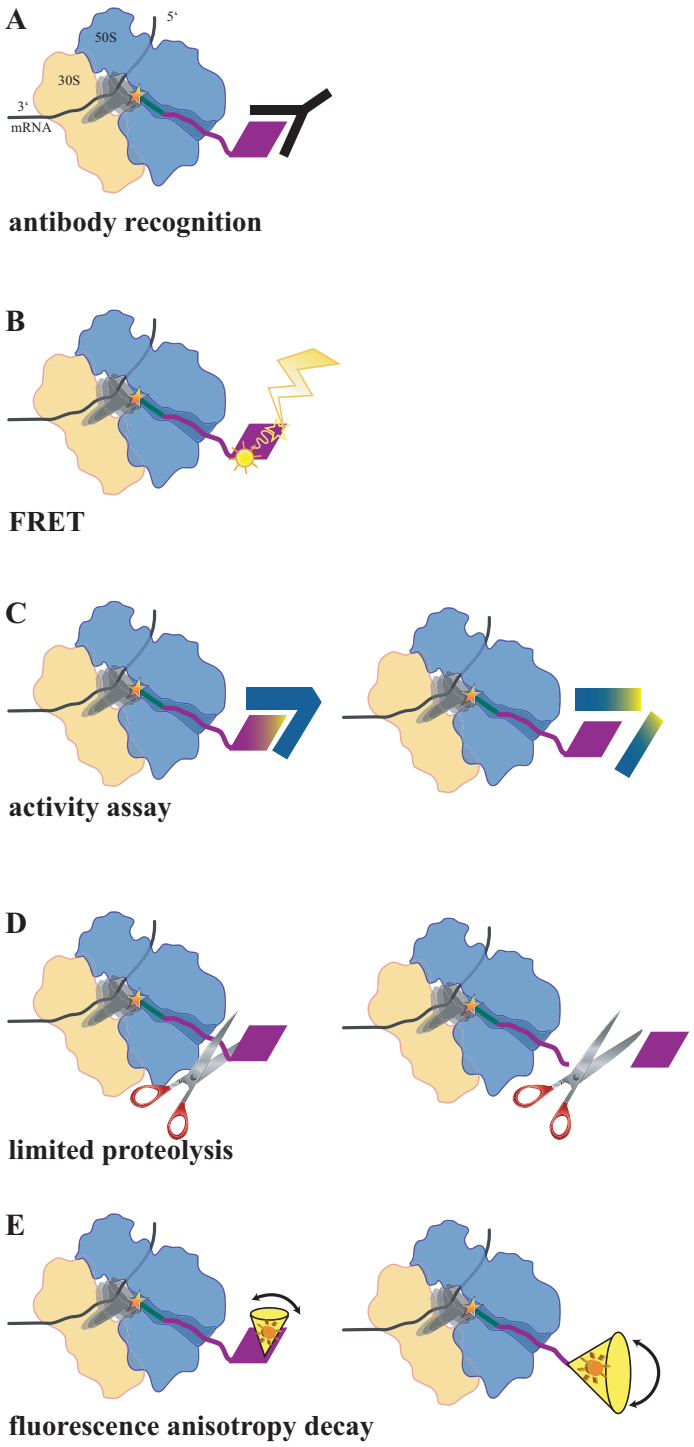


Figure 1.5: Evidence for co-translational protein folding includes indications from: **A:** conformation-specific antibody recognition [76,77], **B:** FRET [67,78], **C:** activity assays [71–73,79], **D:** limited proteolysis [80,81], **E:** fluorescence anisotropy decay [82,83].

CFTR, folding of the NC can occur in a sequential domain-by-domain manner [80] and FRET studies have shown that it requires ATP binding [67]. Together these studies, which describe largely functional attributes of the NC, provide compelling evidence that native-like structure can be acquired in a co-translational manner.

1.1.3.3 Co-translational protein folding and in vitro re-folding

Perhaps the most well-studied system comparing aspects of co-translational *de novo* protein folding and *in vitro* re-folding has been the P22 tailspike protein [77]. In this study, conformation-dependent antibodies raised against the N-terminal domain of the native protein showed a marked binding preference for conformations on the ribosome in the case of stalled RNCs, compared to the equivalent truncated polypeptide chain refolded *in vitro*, thus suggesting that different conformations are sampled on the ribosome. Similarly, a stalled RNC of the first 190 N-terminal residues of firefly luciferase were found by limited proteolysis to populate native-like structures. These native-like intermediate structures were not observed when the released protein was denatured and refolded *in vitro* [81], which indicates that the native-like intermediates were specific to vectorial emergence of the NC during co-translational folding. In addition, it was found in this study that the action of molecular chaperones present *in vivo* is likely to influence the manner in which these intermediate structures are acquired [81] (Section 1.1.3.5). The difference between co-translational folding and *in vitro* refolding is supported further by biophysical studies which have revealed that luciferase acquires its activity within seconds during biosynthesis, but in contrast, the acquisition of activity upon refolding from a chemically denatured state is considerably slower, on the order of minutes [79].

One key difference between *in vitro* refolding and *in vivo de novo* folding lies in the vectorial nature of protein synthesis (and co-translational folding). In an interesting

study, AFM coupled to modelling via molecular dynamics has been used in an attempt to model vectorial folding by limiting the conformational space of the C-terminal region of a polypeptide chain using the AFM tips. For a multidomain ankyrin repeat protein, NI6C [84], studied in this way, it was found that the vectorial re-folding pathway appears to involve a nucleation of the N-terminal repeats and condensation of the C-terminal end of the protein.

Overall, the details of how refolding under steric constraints that model ribosome tethering and the passage through the ribosomal exit tunnel (Section 1.1.1) resembles *de novo* co-translational folding remains to be seen, and important factors such as the effects of the geometry of the ribosome exit tunnel, interaction with cellular chaperones (Section 1.1.3.5) and molecular crowding (Section 1.1.3.4) are likely to be key determinants. Indeed, it seems likely that the folding processes of proteins have evolved such that the formation of folding intermediates during biosynthesis prevents misfolded conformations, as opposed to what has been often seen within *in vitro* refolding experiments (Section 1.1.2); with the vectorial emergence of the NC, the potential interactions the NC establishes with the ribosome inside and outside the ribosomal exit tunnel, and other factors mentioned (molecular chaperones, cellular crowding) all exerting an influence on the co-translational folding process.

1.1.3.4 Molecular crowding and protein biosynthesis

The dense cellular milieu contains up to 400 $g.L^{-1}$ of biomaterial in *E. coli* [85]. While the consequences of cellular crowding on protein folding [86,87] have been investigated, the effect of crowding on co-translational folding are poorly understood. The presence of crowding agents was found to accelerate not only refolding rates, but also to hasten protein aggregation [86]. The latter observation is consistent with the prediction that

macromolecular crowding increases the association constant [88]. A factor likely to assist the NC in avoiding misfolding within the crowded cellular environment would therefore be the arrangement of polysomes (the clustering of multiple ribosomes on a single mRNA transcript). Indeed, the orientations of individual ribosomes within the polysome [89] appears to limit the extent to which each emerging NC can interact with its partially folded neighbour. The increased binding constants that result from crowding conditions [85] are also likely to increase chaperone activity that can assist protein folding [86].

Computational studies have also suggested that the excluded volume effect due to macromolecular crowding is likely to alter the entropy of the nascent polypeptide chain [87]. As a consequence, it alters the thermodynamics and kinetics of protein folding *in vivo*, by destabilising disordered states which are generally more expanded, in favour of more compacted states.

1.1.3.5 Molecular chaperones, the trigger factor

The molecular chaperone trigger factor (TF, Figure 1.6) is the first chaperone that can interact with the nascent chain, and has a docking site on the ribosome in the vicinity of the exit port of the exit tunnel (ribosomal protein L23). TF shows a strong affinity for untranslating ribosomes (K_D=1.1 μM [90][1]) and its docking site is close to the exit port of the ribosomal tunnel, so that TF forms a molecular cradle for the emerging NC [93]. TF has an even higher affinity for translating ribosome (K_D=50 nM [82,94]), where it can make initial contact with *ca.* 40% of the RNC [6,82,95], and is known to cycle on and off the ribosome; it is therefore likely to alter the nature of the intermediate species that are formed during translation.

[1]The ribosome concentration in *E. coli* is ca. 5-20 μM [91], and the association of TF to ribosome is competing with the dimerisation of TF, which constant is on the order of 18 μM [92].

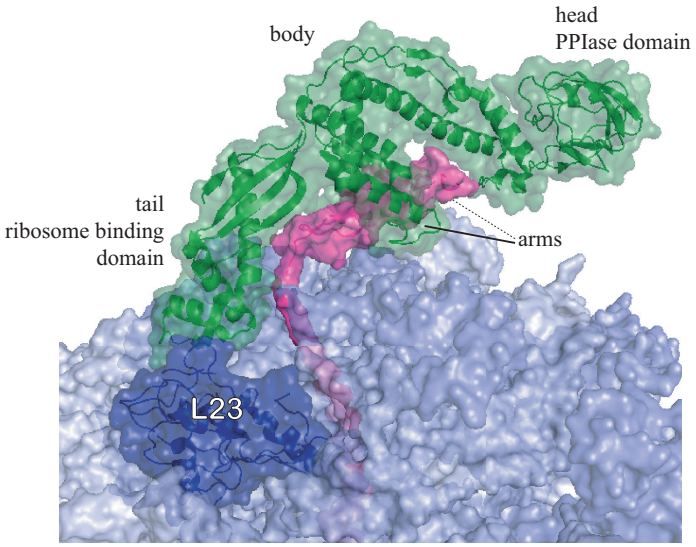


Figure 1.6: 47 kDa Trigger factor in green (pdb files *1W26.pdb* [96] and *2VRH.pdb* [97]) interacting with the 50S subunit in blue (pdb file *2AW4.pdb* [25]) at the universal docking site (protein L23). For clarity, a nascent polypeptide chain has been modelled in magenta as in Figure 1.1.

1.1.4 Structural evidence for co-translational folding

The structure and function of the intact 70S ribosome and its component 50S and 30S subunits has been described in an array of exquisite crystallographic (reviewed in [24]) and cryo-EM structures (reviewed in [23], Section 1.1.1) and although several biochemical and biophysical studies have demonstrated co-translational folding, little structural evidence exists for the conformational preferences of RNCs, and where within the ribosomal environment and at what stage during synthesis co-translational folding begins to occur. Within the ribosome, the ribosome exit tunnel is of considerable interest in the context of co-translational folding. As mentioned, the tunnel is located within the 50S subunit, and is lined with parts of the 23S rRNA and the ribosomal proteins L4 and L22 that are likely to mediate the exit of the NC [68, 70]. Modelling studies

have suggested that it is a water-filled, yet rigid structure [42, 44] (Figure 1.4A). With dimensions of 20 Å at its widest points [8] the tunnel has been the subject of a central question — whether or not it has the capacity to support any degree of folding or structure of the NC?

Recent cryo-EM studies of translation-arrested ribosomes have demonstrated that a pair of NCs, unrelated in sequence, assume an unfolded conformation that extends throughout the length of the tunnel [26, 39] and can indeed interact with the inner wall of the tunnel. When, however, a stabilising helical motif was introduced to the NC sequence at positions that reflect different locations within the exit tunnel [41], the NC readily formed helical structure. The formation of such structure occurred as little as 53 Å away from the PTC (Figure 1.4) and this is in strong support of previous FRET studies [78], which also indicated formation of secondary structure.

Perhaps more intriguing is the recent proposal for the "exit port" (Figure 1.4). Mapping of the conformational space accessible to the voltage-gated potassium channel protein Kv1.3 RNC by "molecular tape measurements" using cysteine-modifying reagents have suggested that the ribosome tunnel widens at the exit port [69]. Moreover, the absence of electron density for the NC at the exit port in cryo-EM maps of translating ribosomes holding TnaC or Sec61/SecY [26,39], while a well defined density is observed closer to the PTC, suggests a heterogeneity in conformation at the exit port, and that the vestibule might offer an entropic window for folding of the NC [98].

Computational modelling has suggested that some native contacts can be achieved by barnase-RNC and CI-2 RNC when the nascent polypeptide is only 25 residues long [46]. At these lengths, the NC can only extend to the exit port. Molecular dynamics simulations of four domains with different folds [47] have predicted that a coil-to-globule transition can occur at distances >65 Å from the PTC (the globular state is defined as

a compact heterogeneous structural ensemble without persistent secondary or tertiary structure), which corresponds to the distance spanned by a *ca.* 24-residues extended peptide.

The expectation is that there is an increase in conformational space sampled by the NC once it reaches the exit port, which is likely to be accompanied by a greater degree of freedom. This has been examined for apomyglobin-RNC (ApoMb-RNC) using fluorescence depolarisation measurements. The motions of ApoMb-RNC were modelled using the model-free approach proposed by Lipari and Szabo [82, 99]. The model-free approach makes use of two parameters to describe internal motions: the timescale of the motions is expressed in terms of a local correlation time ($\tau_{c,NC}$), and the amplitude of the motions is expressed in terms of an order parameter (S^2_{NC}: the greater the amplitude of the motion, the lower the order parameter). Fluorescence depolarisation measurements of ApoMb-RNC [82,83,100] showed a correlation between NC length and both the amplitude and timescale of these motions. For NC lengths of <57 residues, the NC appears to be largely immobilised, and its motions are too restricted (in other words, S^2_{NC} is too high) to allow independent tumbling of the NC. On the contrary, at longer translation lengths (>57 residues) the RNCs were found to be flexible, and their motions were found to have a correlation time of 3-11 ns, which reflects the greater degree of freedom upon emergence from the ribosome exit tunnel. This correlation time corresponds to the rotational diffusion of a small domain, and indicates that the small globular RNC moves independently of the ribosome [83]. In addition, the timescale of the motions appears to decrease as the length of translated construct increases, from $\tau_{c,NC} = 9 \pm 2$ ns at a NC length of 57 residues down to $\tau_{c,NC} = 5 \pm 1$ ns at a NC length of 153 residues, reflecting the progressive gain in flexibility as the fraction of the NC that has emerged from the ribosomal exit tunnel grows.

The timescale of the motion is also a function of the size of the domain, with large domains having longer correlation times [99]. The correlation time of ApoMb-RNC (3-11 ns) is much faster than the rotational correlation time found for the released 150 residues domain (41 ns[2]), which can perhaps be rationalised by a either a smaller size of the potential co-translational folding intermediates [82, 83], or the absence of defined structure. These experimental studies together with simulation studies [46–48] indicate that the NCs are not only able to sample a myriad of conformations (ms timescale), but are also highly dynamic on a ns timescale, and renders the characterisation of their structures inaccessible by either X-ray crystallography or cryo-EM. Thus, a parallel with the L7/L12 stalk region mentioned earlier (Section 1.1.1.1) can be drawn, which leads to the idea that NMR can be a suitable tool to investigate both the structure and dynamics of the emerging RNC.

In this thesis, the use of NMR for examining RNCs is considered, specifically its ability to provide detailed quantitative data about their dynamics. The remainder of this Chapter surveys the technique, and a synopsis of the key aspects of NMR used in biological systems is presented in Section 1.2. As the work described in Chapters 2-4 requires a detailed understanding of rapid-acquisition techniques, diffusion NMR and transverse relaxation optimised spectroscopy, these more advanced concepts are introduced in Section 1.3.

1.2 NMR spectroscopy in studying protein folding

An array of NMR experiments has been used extensively to study protein folding and unfolding pathways, as well as the different conformations the polypeptide chain

[2]This value is based on fluorescence anisotropy decay measurements on the released nascent chain by Ellis *et al* [83], and might suggest an oligomerisation of the protein. The expected τ_c for a 150 residues domain would be *ca.* 10 ns assuming a globular domain based on the Stoke-Einstein-Debye equation 1.18.

populates during folding/unfolding. These are briefly summarised below.

1.2.1 NMR studies of protein unfolding and folding pathways

NMR has been found to provide detailed information on the structural and dynamical properties of intermediate states occupied during unfolding. NMR studies using isolated full-length proteins have enabled the characterisation of persistent native-like conformations in unfolding intermediate states within highly-denaturing conditions for proteins such as GFP [101] and the ribosomal protein L9 [102]. Details of a molten globule unfolding intermediate of alpha-lactalbumin [103] have been revealed, highlighting the intrinsic dynamic properties of such intermediate states. Unfolding intermediates that are easily stabilised under denaturing conditions are often used to gain insight into the analogous folding intermediates, but refolding experiments provide more direct information on the folding pathways. For instance, the refolding pathway of β_2-microglobulin after chemically-induced unfolding, characterised using real-time NMR acquisition [104] showed the existence of multiple folding pathways: i) a two-stage folding, ii) a three-stage folding with a long-lived native-like intermediate, and iii) a three-stage folding pathway with a short-lived intermediate state.

The pioneering development of relaxation dispersion spectroscopy (described in Section 1.3.1) has also been used to examine the conformations of sparsely populated (1-2%) metastable states present under native conditions for the G48M SH3-fyn [105], the KIX domain [106] and the villin headpiece HP67 [107] amongst others. These metastable states are thought to be the analogues of folding intermediates, and are seen to possess residual native-structural elements, with large disordered regions. The structure and dynamics of these metastable states are of great interest to the understanding of the function of small signalling domains, which is often related to

their native conformational sampling. The ps-to-ns dynamics of the low populated state of G48M SH3-fyn have been characterised using the carbon chemical-shift dispersion of Leu and Ile sidechains to model the timescale of methyl rotamerisation [108, 109]. More than just allowing the characterisation of the structural and dynamic properties of the metastable states, relaxation dispersion spectroscopy also allows examination of the chemical exchange kinetics between native and intermediate states as well as the thermodynamic features of the folding process [63, 110]. Further insight into the kinetics of exchange and the properties of the transition states related to the energy barriers separating the native and metastable states were provided by a study of the folding pathway of a nitrogen regulatory protein using relaxation dispersion spectroscopy and mutagenesis. This study showed that transient non-native hydrogen bonds reduce the free energy of the transition state [57], thus increasing the rate of folding.

1.2.2 NMR studies of disordered proteins

Over recent years, NMR methods have allowed the determination of structural ensembles that describe the conformational space adopted by disordered states. For example, the unfolded ensemble of urea-denatured ubiquitin has been characterised using paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) NMR measurements [111, 112]. Both approaches, in combination with molecular dynamics (MD) modelling, have shown that the disordered ensemble contains a number of compact conformations which retain some residual structure. For instance, PRE restraints combined with MD simulation have shown that for α-synuclein (but not for β-synuclein [113]) compact conformations are stabilised by a number of long-range tertiary interactions [114], despite the absence of defined structural elements.

The development of restrained MD modelling to enable the use of ensemble NMR

data as inputs into structural determination is an area of rapid development over recent years [115], in particular in attempts to describe unfolded and highly dynamic states [116]. Chemical shifts contain detailed structural information, although their potential is only recently beginning to be exploited [117–119]. Not only are chemical-shift-restrained simulations able to provide structures of globular states [117–119], but they have also been able to reproduce the conformational sampling of disordered proteins such as sendai virus nucleoprotein [120].

1.2.3 NMR studies of high-molecular-weight complexes

NMR spectroscopy has typically been considered as being size-limited, in part because of relaxation processes, mediated by dipolar interactions and chemical-shift anisotropy (discussed further in Section 1.3.1), whose rates increase as the size of the molecule increases, and the rate of rotational diffusion decreases. However, developments of new labelling strategies and TROSY spectroscopy techniques (described in detail in Section 1.3.2) have allowed the NMR investigation of the functional states of a number of high-molecular-weight complexes [121]. For example, the NMR study of the 20S proteasome revealed the mechanism that enables substrate localisation within the catalytic chamber [122, 123], which is facilitated by the dynamics of the surface residues of the entrance pore and the catalytic chamber, as well as the network of substrate-enzyme interactions inside the catalytic chamber [124]. Similarly, the highly dynamic substrate (hDHFR) of the 900kDa GroEL-ES complex was characterised, with the network of chaperone-substrate interactions identified using CRIPT and CRINEPT NMR experiments [125] (TROSY-based experiments described further in Section 1.3.2). The resulting set of conformations of the GroEL-ES bound hDHFR was described as a dynamic ensemble of randomly distributed structures.

1.2.4 NMR studies of ribosome complexes

As was mentioned in Section 1.1.1.1, analysis of ^{1}H-^{15}N spectra of 70S ribosomes [33,34] revealed resonances from the L7/L12 stalk, which are observed as the result of resonance narrowing through the independent motion afforded by the 20 residue-long flexible linker that tethers the C-terminal domain to the ribosome body. The NMR observation of flexible regions of high molecular weight complexes is not a new discovery in NMR spectroscopy and dynamic regions within large amyloid structures have also been observed to give rise to resonances using conventional NMR methods [**?**,126–128].

The similarity of the chemical shifts of ribosome-bound L7/L12 to those obtained from purified L7 shows that the structure of the ribosome stalk is closely resembling that of the isolated proteins [31,33]. The observation of sharp resonances from the ribosomally associated C-terminal domain of L7 is indicative of rapid tumbling and high flexibility. The effective correlation time of the ribosomally associated L7/L12 stalk determined via ^{1}H-^{15}N spin relaxation studies is much shorter than that expected for the ribosome complex overall ($\tau_{c_{L7/L12}}$ = 14 ns, [33], compared to $\tau_{c_{ribo}} \sim$ 2500 ns [129] at 25°C).

1.2.5 Studying structural and dynamical properties of RNCs by NMR spectroscopy

By analogy to the ribosome-bound stalk region, the possibility of using NMR to study RNCs was first demonstrated on a pair of tandem repeat immunoglobulin-like domains derived from the 120kDa F-actin cross-linking gelation factor from *Dictyostelium discoideum* (domain 5 and domain 6, abbreviated as ddFLN-dom5+6) [130]. The RNC construct of this initial study comprised ddFLN-dom5 anchored to the ribosome via ddFLN-dom6, which behaves as an extended linker (Figure 1.7). The analysis of the resonances derived from ^{1}H-^{15}N correlation spectra using SOFAST-HMQC experiments

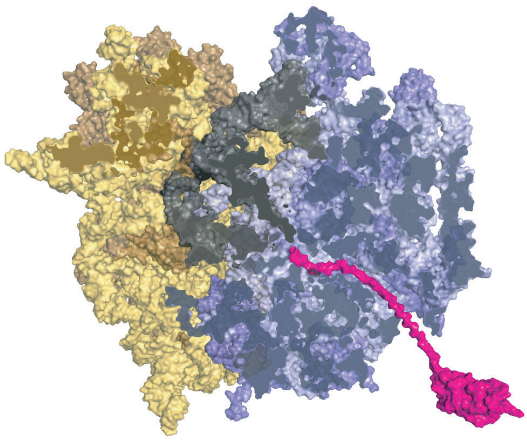


Figure 1.7: Schematic representation of ddFLN-dom5+6 RNC as in [130], by a cross-section of the translating 70S ribosome (pdb file 2J00 and 2J01). The 30S subunit is shown in yellow, the 50S subunit in blue, and mRNA in black. The three tRNAs are represented with dark grey, grey and light grey for the A-site tRNA, P-site tRNA and E-site tRNA, respectively. The domain of interest, ddFLN-dom5, is shown in magenta, with its native-fold, while the unfolded ddFLN-dom6 linker resides principally inside the ribosome exit tunnel.

(described in detail in Section 1.3.3) revealed that the ribosomally attached ddFLN-dom5-RNC adopted a fold that resembles closely the native fold observed in the corresponding isolated domain [130]. Analysis of ^{1}H-^{13}C correlation spectra of ddFLN-dom5+6-RNC using HMQC experiments provided further insight into the RNC folded state, afforded by the increased sensitivity resulting from both the increased flexibility of the sidechains and the three equivalent protons of methyl groups [131]. The inherent dynamic behaviour of the NC has been confirmed from the observation of narrow linewidths associated with the NC resonances (*ca.* 25-30 Hz) compared to the linewidths expected for the ribosome alone (>1000 Hz, [33]). Similarly, the NMR study of an SH3-RNC via ^{1}H-^{15}N CRINEPT experiments (Section 1.3.2) have revealed that resonances from the NC have linewidths amenable to NMR observation (36 Hz [132]). The ^{1}H-^{15}N spectrum was characteristic of a disordered SH3 domain despite the C-terminal end of the domain

being 50 residues away from the PTC [132].

1.3 NMR spectroscopic methods for the study of ribosome complexes

This section introduces the key concepts for the NMR spectroscopy of RNCs. The challenges related to NMR studies of RNC complexes arise from the size of the complex, which results in broadened linewidths, the limited maximum achievable concentration, which leads to low sensitivity, and the overall limited stability of the complex, which prevents accumulation of data over more than two to four days at most. Following a a brief introduction to NMR relaxation, specific NMR experiments which were used and optimised to overcome these challenges will be described in detail. For instance, the transverse relaxation-optimised spectroscopy (TROSY) NMR technique allows narrowing of broad linewidths, and longitudinal relaxation-optimised spectroscopy (SOFAST-HMQC) allows the rapid acquisition of heteronuclear spectra, compensating for the low concentrations and short lifetimes of RNCs samples. Pulsed-field gradient diffusion experiments were used to monitor the attachment of the observed polypeptide to the ribosome complex. Finally, the choice of nucleus labelling that allows the appropriate NMR experiments to probe for the unfolded and folded states of the RNC, in order to gain information to the folding equilibrium of the RNC, will be discussed.

1.3.1 Introduction to spin dynamics

The key concept which underlies NMR data accumulation is relaxation, which is the return to the equilibrium magnetic state, proceeding via transitions between the energy levels of the nuclear spin system. These transitions are caused by field fluctuations that have the same frequency as that corresponding to the difference in energy between

two spin states. The equilibrium state corresponds to a net longitudinal magnetisation (referred to as I_z^0) and a zero transverse magnetisation. The spin-spin relaxation rate (R_2) describes the return to equilibrium of the transverse magnetisation (during acquisition), such that a fast R_2 rate results in early decay of the recorded magnetisation (10-50 ms or shorter for proteins), broad signals and decreased peak height. The spin-lattice relaxation rate (R_1) describes the return to equilibrium of the longitudinal magnetisation, and a relatively fast R_1 is necessary for the recovery of magnetisation between scans to allow a high repetition rate of data accumulation within a reasonable time frame (typically seconds). Both R_2 and R_1 depend on the signal intensity of the frequencies generated by the fluctuating fields that match the frequencies of the relevant NMR transitions. The probability of finding the frequency ω within the thermal motions of a molecule depends on the rotational diffusion of the molecule, or the correlation time τ_c (assuming that rotational diffusion is the only significant motion process) [133, 134], is termed $J(\omega)$ and is given by:

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \tag{1.1}$$

The fluctuating fields of interest in bio-molecular NMR have their origins predominantly in dipolar interactions with other NMR-active nuclei, and in the chemical shift anisotropy. Selective labelling can reduce the number of NMR active nuclei within a molecule and therefore limit the dipolar contributions to the relaxation [136]. During a sequence of pulses, it is possible to avoid spin states that are associate with fast-relaxing transitions [137], for example in TROSY pulse sequences (described in Section 1.3.2).

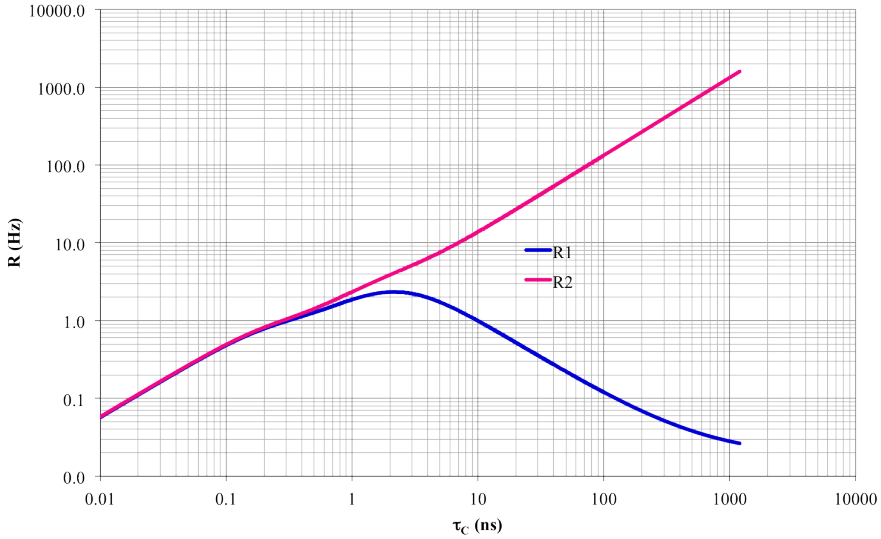


Figure 1.8: Heteronuclear ^{15}N relaxation rates R_2 $(1/T_2)$ and $R_1(1/T_1)$ as function of molecule correlation time τ_c, calculated for a ^{1}H frequency of 700MHz. A small protein domain (20kDa) has an estimated correlation time, τ_c of *ca.* 5-10ns, whereas a spheric particle of the size of the ribosome (2.4 MDa) has a τ_c of *ca.* 2500ns [129]. This figure was generated with the help of Dr. John Kirkpatrick, UCL, using the expression for the relaxation rates in term of spectral density from [135].

1.3.1.1 Effects of internal motions

The transverse and longitudinal spin relaxation results in the decay of the recorded magnetisation, and recording the relaxation rates allows the spectral density $(J(\omega))$ function to be partially mapped. The model-free method proposed by Lipari and Szabo [99] expresses $J(\omega)$ in terms of the overall rotational correlation time, and the amplitude and internal correlation time of the local motion of the main vector of the interactions leading to relaxation (*e.g.* the NH bound for ^{15}N relaxation). Indeed, regions of a protein or complex often have local ps-ns dynamics, which result in apparent faster tumbling than the rest of the molecule. To differentiate at a residue-specific level, the

contribution to $J(\omega)$ arising from the overall rotational diffusion (correlation time τ_c) compared to that arising from the local ps-ns flexibility (correlation time τ_e), the model-free method makes use of the order parameter S, which is a measure of the amplitude of the internal ps-ns motions for each residue.

$$J(\omega) = \left[\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2)\tau'}{1 + \omega^2 \tau'^2} \right] \tag{1.2}$$

where $\tau' = \left(\tau_c^{-1} + \tau_e^{-1}\right)^{-1}$.

S describes the width of the distribution of orientations of the bond vector (NH for example), and can be interpreted as a function of the cone semi-angle θ that limits the ps-ns motions of the residues (Figure 1.9):

$$S = \frac{1}{2}\cos(\theta)(1 + \cos(\theta)) \tag{1.3}$$

Conformational entropy is also a function of the distribution of the bond vector orientations. A simple relation between the difference of model-free S^2 between two conformations (A & B) and the difference in conformational entropy, S_{entropy}, between these two conformations has been proposed [138], which assumes that the motions are restricted within a cone (the model-free order parameter is written S_{LZ} to highlight the distinction from S_{entropy}).

$$S_{\text{entropy, B}} - S_{\text{entropy, A}} = \ln\left(\frac{3 - \sqrt{1 + 8S_{\text{LZ, B}}}}{3 - \sqrt{1 + 8S_{\text{LZ, A}}}} \right) \tag{1.4}$$

In the case of the ribosome, a range of correlation times needs to be considered to model the distinct motions occurring within the complex, including the overall rotational diffusion time $\tau_{c,\text{ribo}}$, the correlation time of motions of flexible domains such as the C-

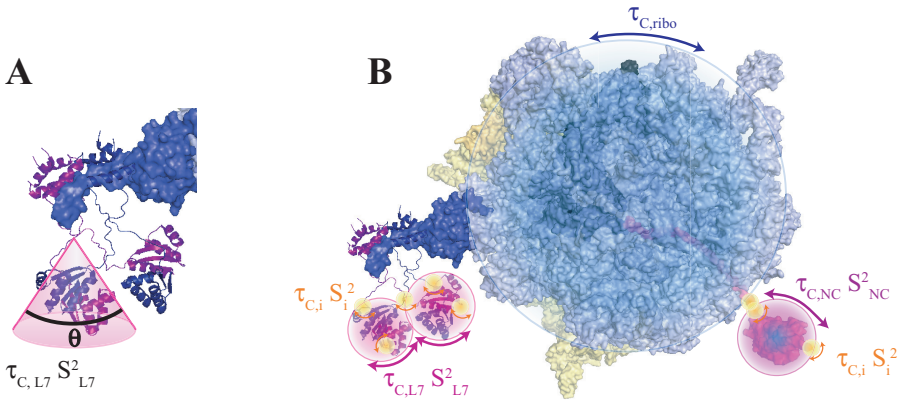


Figure 1.9: **A:** Amplitude of the motions described by the model-free approach, using the S^2 parameter, shown on the ribosome for the dynamic L7 stalk. **B:** Schematic representation of the motions associated with different regions of the ribosome complex: the overall rotational correlation $\tau_{c,ribo}$, the motions of flexible domains $\tau_{c,L7}$, $\tau_{c,NC}$ (with their order parameter S^2_{L7}, S^2_{NC}, Section 1.3.1), and the internal motions of sidechains or flexible loops $\tau_{c,i}$ (with their order parameter S^2_i).

terminal domain of the L7 stalk ($\tau_{c,L7}$, S^2_{L7}), and the fast internal motions associated with dynamic loops or side-chains ($\tau_{c,i}$, S^2_i) (Figure 1.9B).

1.3.1.2 Effects of chemical exchange

Processes of chemical exchange can also contribute to the relaxation rate R_2, and this modulation of R_2 by chemical exchange is distinguishable from anisotropic rotational diffusion [139]. Depending on the timescale, chemical exchange results in de-phasing of the magnetisation that can be partially refocused by applying a train of 180° pulses (CPMG pulse train) with a frequency that is close to that of the chemical exchange process in relaxation dispersion experiments [140]. By varying the frequency of the CPMG pulse train, the magnetisation de-phasing can be quantified allowing

characterisation of the kinetics of the chemical exchange process, and extraction of parameters describing the participating states, such as their chemical shift [63], or residual dipolar coupling [141].

1.3.2 TROSY NMR spectroscopy

High-molecular-weight complexes are associated with long rotational correlation times, which result in rapid transverse relaxation (as can be seen in Figure 1.8). The early decay of the recorded transverse magnetisation results in line-broadening and losses in NMR signal sensitivity. Recent developments in NMR spectroscopy, such as ^{1}H-^{15}N TROSY [142, 143], ^{1}H-^{15}N CRIPT & CRINEPT [144] and methyl-TROSY HMQC [145, 146] have facilitated the study of large complexes such as the 900kDa GroEL-ES complex [147] and the 20S proteasome [122], and suggest that studies of the RNCs by similar strategies are possible. Each of these TROSY experiments rely on carefully constructed magnetisation pathways that avoid conversion of slow-relaxing spin states with fast-relaxing spin states, and guide the magnetisation through spin-states whose relaxation is slowed by the effects of interference between different relaxation mechanisms [137] (Figure 1.10).

^{1}H-^{15}N TROSY relies on the destructive interference of the ^{1}H-^{15}N dipolar interactions and chemical shift anisotropy (CSA), which is predicted to be optimal at a ^{1}H frequency of $\sim$ 1 GHz and reduces significantly the rate of two of the four transitions (Figure 1.10A, transitions shown in blue). The pulse sequences of ^{1}H-^{15}N TROSY experiments selectively transfers the magnetisation from the slow-relaxing ^{15}N transition ($R^{s}_{2,N}$) during indirect evolution to the slow relaxing ^{1}H transition ($R^{s}_{2,H}$) for detection, and minimises the time spent in fast-relaxing spin-states during magnetisation transfer [143]. ^{1}H-^{15}N CRIPT and CRINEPT spectroscopy further exploits the cross-correlated relaxation in the ^{1}H-^{15}N spin system of large proteins to improve the efficiency of the

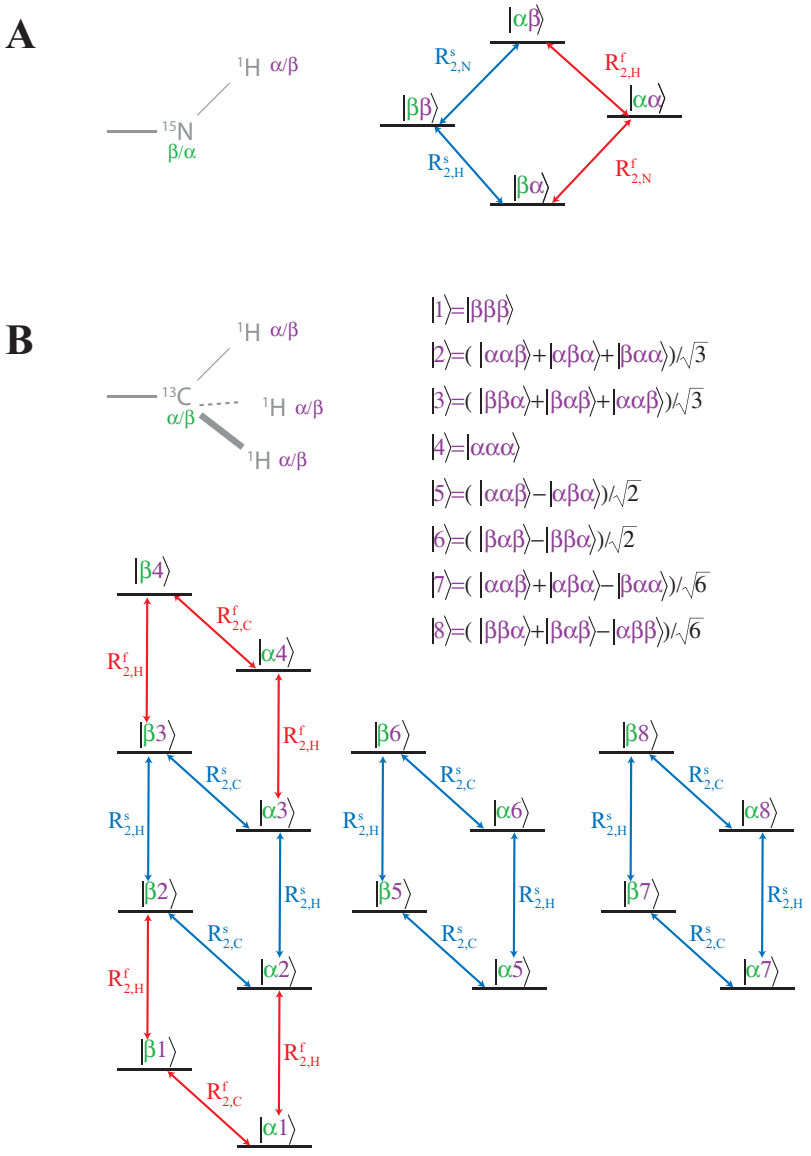


Figure 1.10: Transitions between magnetic states in **A:** ^{1}H-^{15}N spin pair or **B**: ^{1}H-^{13}C CH_3 spin system. The proton magnetisation states are shown in purple and the heteronuclear states is shown in green. **A:** The transitions whose rates are reduced by the destructive interference between the dipolar interactions and CSA are shown in blue, and are labelled slow (s). **B:** The transitions which do not involve ^{1}H-^{1}H and ^{1}H-^{13}C dipolar interactions are shown in blue and are labelled slow (s). **A** is adapted from [148] and **B** is adapted from [145].

magnetisation transfers over that achieved using classic Insensitive Nuclei Enhanced Polarization Transfer (INEPT) elements [144]. Despite the remarkable sharpening of resonances observed in the case of GroEL [147], the spectral improvements achieved through this destructive interference pathway depend critically on the tumbling rate of the molecule and on the magnetic field.

The examination of the transitions occurring in a CH_3 group isolated from external protons shows that the multiple ^{1}H-^{1}H and ^{1}H-^{13}C dipole-dipole interactions cancel each other for all transitions involving eight of the possible 16 spin-states (shown on the right in Figure 1.10B), in a field-independent manner. A variety of pulse sequences were examined before it was discovered that in the HMQC scheme, the recorded magnetisation results from pathways that populate exclusively the eight slow-relaxing spin-states (and pathways that involve exclusively the fast-relaxing spin-states), making the HMQC a TROSY scheme [145] in the sense that slow- and fast-relaxing spin-states are not mixed. ^{1}H-^{1}H dipole-dipole interactions with external protons constitute the primary remaining relaxation process, and can produce spin-flips, which result in the mixing of the slow- and fast-relaxing spin-states; this source of signal loss can be significantly attenuated using high deuteration levels (>90%). Any ^{13}C-^{13}C coupling also contributes to rapid relaxation, but over the past decade, it has been clearly demonstrated [121] that this effect can be reduced using selective labelling, such as uniform-[^{12}C,^{2}H], Ileδ1-[$^{13}C^{1}H_3$] labelling. To achieve this highly selective Ileδ1 labelling, methyl-^{13}C, 3,3 ^{2}H, α-ketobutyric acid is introduced into the growth medium, together with ^{2}H, ^{12}C glucose or acetate as the principal carbon source [149,150]. Similarly, the combination of selective labelling U-[^{12}C,^{2}H], Leuδ1-,Valγ1-[$^{13}C^{1}H_3$] can also be achieved using the biosynthetic precursor of Val and Leu, 3-methyl-^{13}C, 3,4,4,4 ^{2}H α-isovalerate. ^{1}H-^{13}C and ^{2}H-^{1}H dipolar interaction arising during the pulse sequence results in dipole-dipole cross-

relaxation of the multiple-quantum spin-system during the ^{13}C evolution time [145]. The overall sensitivity of the experiment is thus dependent on the rotation correlation time ($\tau_{c,ribo}$) and the order parameter of the methyl axis (S^2_{axis}, Figure 1.11). It is assumed that motion around methyl axis is so rapid that it is independent of the slower rotational motion of the entire molecule.

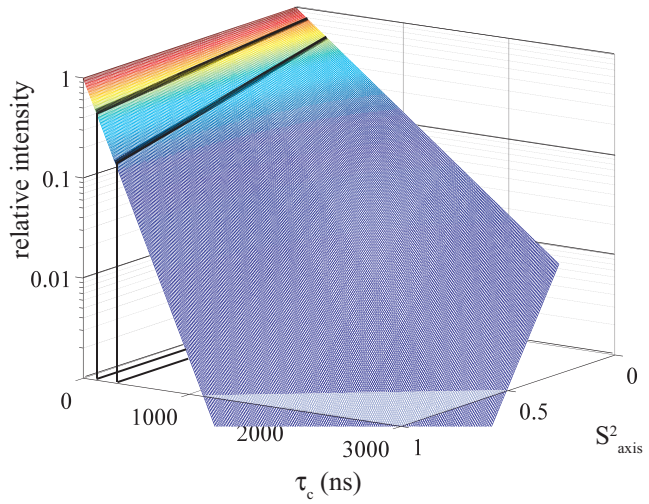


Figure 1.11: Predicted maximal intensity in methyl-TROSY HMQC of RNC, as a function of $\tau_{c,ribo}$ and the order parameter describing the reorientation of the methyl 3-fold axis (S^2_{axis}), using equation 1.10 and 1.5 from Ollerenshaw *et al* [145]. $\sum_i \frac{1}{r^6_{HH,i}}$ and $\sum_i \frac{1}{r^6_{HD,i}}$ are defined as in Ollerenshaw *et al* [145], and such that $\left(\sum_i r^{-6}_{HH,i}\right)^{-\frac{1}{6}}$ = 5.5 Å and $\left(\sum_i r^{-6}_{HD,i}\right)^{-\frac{1}{6}}$ = 1.8 Å. The black lines show the tumbling time of the proteasome as studied by [122], 300 ns, and that of malate synthase G, 118 ns [145].

The performance of the methyl-TROSY HMQC experiment was demonstrated by the NMR investigation of high MW complexes (often using high temperatures to decrease the rotational correlation time of the large complexes), such as the 670 kDa 20S archaeal proteasome [122] (60°C), the 468 kDa archaeal oligomeric TET2 protease [151] (37°C), the 230 kDa nucleosomal protein 2-nucleosome complex [152] (45°C), and the 204 kDa

ATPase motor of the Sec translocase [153] (25°C). The overall rotational correlation time of the complex approached values of 300 ns in the first two cases.

Ollerenshaw *et al* [145] have shown that the proton transverse relaxation rate in the methyl-TROSY HMQC has contributions from the methyl proton - carbon dipolar coupling (first term in equation 1.5), the dipolar coupling of the methyl proton with all the external deuterium (second term of equation 1.5) and the dipolar coupling of the methyl protons with the remaining protons (last term in equation 1.5), considering that only the slow-relaxing pathways contribute to the signals, which is likely to be the case for broad resonances arising from ribosome-attached domains:

$$\begin{aligned} R_{2,\mathrm{H}} &= \frac{1}{45}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{S^2_{\mathrm{axis}}\gamma^2_{\mathrm{H}}\gamma^2_{\mathrm{C}}\hbar^2\tau_{\mathrm{c,ribo}}}{r^6_{\mathrm{HC}}} + \frac{8}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \sum_i \frac{\gamma^2_{\mathrm{H}}\gamma^2_{\mathrm{D}}\hbar^2\tau_{\mathrm{c,ribo}}}{r^6_{\mathrm{HD,i}}} \\ &\quad + \frac{1}{4}\left(\frac{\mu_0}{4\pi}\right)^2 \sum_i \frac{\gamma^4_{\mathrm{H}}\hbar^2\tau_{\mathrm{c,ribo}}}{r^6_{\mathrm{HH,i}}} \end{aligned} \tag{1.5}$$

Those three relaxation terms are described as functions of the relevant spectral density function ($J_{\mathrm{CH}}(\omega)$, $J_{\mathrm{HD}}(\omega)$ and $J_{\mathrm{HH}}(\omega)$) with a single overall correlation time $\tau_{\mathrm{c,ribo}}$ in the case of ribosome complexes (Section 1.3.1) ($J_{\mathrm{CH}}(\omega)$ is weighted by the order parameter of the side chain motions S^2_{axis}). The spectral density functions $J(\omega)$ are approximated to $J(0)$ only since $(\omega_{\mathrm{C}}\tau_{\mathrm{c}})^2$ is much larger than 1 for τ_{c}>20ns.

At 25°C, the rigid ribosomal complex was shown to have a rotational correlation time of $\tau_{\mathrm{c,ribo}} \sim$ 2500ns [129], which is significantly greater than in any methyl-TROSY work yet reported. Nonetheless, the ribosome is known to contain flexible regions such as the L7/L12 stalk which has a relatively short correlation time ($\tau_{\mathrm{c,L7}} \approx$14 ns, [33]) as described in Figure 1.9. Indeed, in high MW complexes, internal motions of long extensions such as the stalk region or in the case of an RNC the nascent chain, result in an apparent reduction of their tumbling time (Figure 1.9). As described in Section

1.3.1.1, the model-free approach introduces an internal correlation time that reflects these motions ($\tau_{c,L7}$,$\tau_{c,NC}$), and their coupling to the overall rotational correlation time ($\tau_{c,ribo}$). The model-free method incorporates the effects of internal motion into the expression for $J(\omega)$, and modifies $J_{CH}(0) = S^2_{axis}\tau_{c,ribo}$ to

$$J_{CH}(0) = S^2_{axis}\left(S^2_e\tau_{c,ribo} + (1 - S^2_e)\tau'\right) \tag{1.6}$$

and $J_{HH}(0) = J_{HD}(0) = \tau_{c,ribo}$ to

$$J_{HH,HD}(0) = \left(S^2_e\tau_{c,ribo} + (1 - S^2_e)\tau'\right) \tag{1.7}$$

Thus, the model-free method applied to equation 1.5 gives rise to a new expression of the relaxation rates that includes τ' and S^2_e:

$$\begin{aligned} R_{2,H} \quad = \quad & \left[\frac{1}{45}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{S^2_{axis}\gamma^2_H\gamma^2_C\hbar^2}{r^6_{HC}} + \frac{8}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \sum_i \frac{\gamma^2_H\gamma^2_D\hbar^2}{r^6_{HD,i}} + \frac{1}{4}\left(\frac{\mu_0}{4\pi}\right)^2 \sum_i \frac{\gamma^4_H\hbar^2}{r^6_{HH,i}}\right] \\ & \times \left(S^2_e\tau_{c,ribo} + (1 - S^2_e)\tau'\right) \end{aligned} \tag{1.8}$$

The latter expression is therefore more appropriate for the analysis of the ^{1}H relaxation rates of flexible regions of the ribosome complex.

Local anisotropy within the flexible domain (such as dynamics of individual loops or secondary elements) results in additional terms in the expression of $J_{HH,HD}(0)$ ($J_{CH}(0)$ is again simply scaled by S^2_{axis}) [154,155]:

$$J_{HH,HD}(0) = S^2_e S^2_i\tau_{c,ribo} + S^2_e(1 - S^2_i)\tau' + (1 - S^2_i)\tau'' \tag{1.9}$$

with $\tau'' = \left(\tau^{-1}_{c,ribo} + \tau^{-1}_{c,e} + \tau^{-1}_{c,i}\right)^{-1}$, in which the local motions are characterised by $\tau_{c,i}$ and S^2_i. Again, this can complicate further the expression of $R_{2,H}$. Similarly, chemical

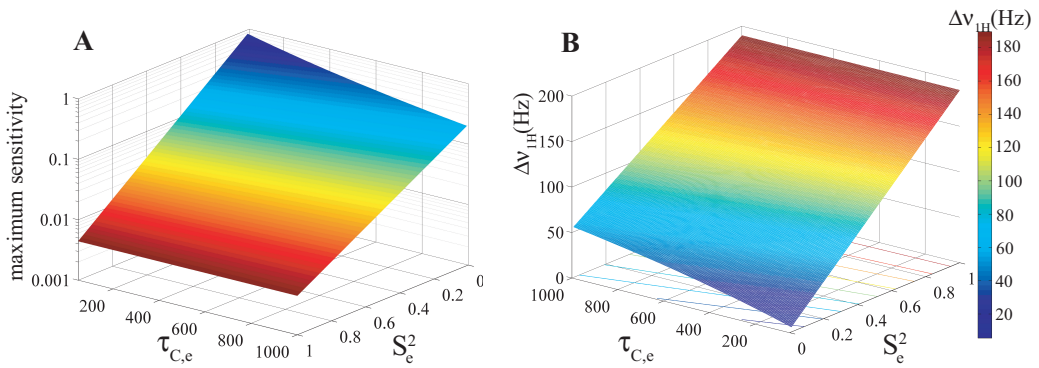


Figure 1.12: A: Predicted maximal intensity in methyl-TROSY HMQC spectra of resonances from a flexible domain attached to the ribosome, as a function of the internal correlation time $\tau_{c,e}$ and the order parameter of the internal motions (S_e^2) (refer to main text for details), assuming S^2_{axis}=0.5 and $\tau_{c,ribo}$=2500 ns. **B:** Predicted proton linewidth for the same resonances, calculated from $R_{2,H}$ using equation 1.11 (refer to main text for details).

exchange on the order of the chemical shift timescale (100 ns-1 ms) as mentioned in Section 1.3.1.2 would require an additional term $R_{exchange}$ to $R_{2,H}$.

The maximum intensity of the methyl-TROSY HMQC, defined as the amplitude of the first point of the FID (i.e. at the beginning of the acquisition time, without any ^{13}C chemical shift evolution), depends solely on the proton relaxation rate:

$$I_{HMQC} = \exp(-4\tau R^s_{2,H}) \tag{1.10}$$

where τ is equal to $1/4J_{CH}$. Figure 1.12A shows the expected I_{HMQC} of ribosome-attached flexible domains as a function of $\tau_{c,e}$ and S_e^2, with $\tau_{c,ribo}$=2500 ns, and S^2_{axis}=0.5, and $\sum_i \frac{1}{r^6_{HH,i}}$ and $\sum_i \frac{1}{r^6_{HD,i}}$ defined as in Ollerenshaw *et al* [145]. For $\tau_{c,e}$ close to $\tau_{c,ribo}$ (>1000 ns), as the model of internal motion is not strictly valid, the overall tumbling of the molecule $\tau_{c,ribo}$ is not independent of $\tau_{c,e}$. The extended model-free approach to analyse slow interdomain motion can be applied in such cases [156].

R_{2H} is an NMR parameter that is readily measured in heteronuclear experiments from the ^{1}H linewidth ($\Delta\nu_{^1H}$):

$$R_{2H} = \pi\Delta\nu_{^1H} \tag{1.11}$$

Conversely, the estimation of the internal motion parameters of the observed domain can be derived from the analysis of the proton linewidth (Figure 1.12B). Empirical measurements show that the uncertainty in linewidth is inversely proportional to the signal-to-noise [157]:

$$\sigma_{LW} = \frac{17.2}{SN} \tag{1.12}$$

The ^{1}H linewidths of RNC resonances are presented in Section 3.2.2.1.

1.3.3 Fast acquisition NMR spectroscopy

The NMR study of co-translational folding requires the preparation of relatively large amounts of homogeneous material (>10^3 times more than what is required for fluorescence spectroscopy). Despite this feat of preparative biochemistry [158], the maximal achievable concentration of the ribosomal material is on the order of 25 mg/ml in order to avoid ribosome aggregation and significant increases in the viscosity of the sample [129], and at *ca.* 10 μM, represents a molar concentration that is 10- to 100-fold lower than that typically used for NMR spectroscopy. Consequently, accumulating sufficient data to overcome the inherent problems associated with poor signal-to-noise ratios within the limited lifetime of the sample (typically <24 hours) is a significant challenge.

Central to the study of RNCs has been the application of NMR pulse sequences that allow rapid acquisition, in particular the SOFAST-HMQC experiment [159], which has been used in our studies of ddFLN-RNCs [130] and also in the study of barnase-

RNCs [160].

The SOFAST-HMQC experiment relies upon the use of selective excitation so that unperturbed nuclei act as a reservoir of cool spins, allowing the more rapid R_1 relaxation of the excited spins (via enhanced longitudinal cross-relaxation). The return to equilibrium of the longitudinal magnetisation can be written as:

$$\frac{d\Delta I_{iz}(t)}{dt} = -\rho_i \Delta I_{iz}(t) - \sum_{j \neq i} \sigma_{ij} \Delta I_{jz} \tag{1.13}$$

where ρ_i is the auto-relaxation rate of the *i*th spin (which depends on ^1H_i-^1H_j dipolar couplings as well as small contributions from the proton chemical shift anisotropy and the ^{1}H-^{15}N dipolar coupling) and σ_{ij} is the cross-relaxation rate between the *i*th and *j*th spins; in the case of uniform excitation and at time t=0, $R_1 = \rho_i + \sum_{j \neq i} \sigma_{ij}$. However, after selective excitation, a large number of *j*th spins are unperturbed ($\Delta I_{jz} \sim 0$), which reduces the contribution to the cross-relaxation rate of the *i*th spins significantly. Because σ_{ij} has large negative values for large molecules with long rotational tumbling times (Figure 1.13), the longitudinal relaxation rate is increased for selective excitation. This is demonstrated in Figure 1.13 using the simple model of an isolated two-spin system.

In addition to the use of selective excitation, a decrease in the excitation angle from 90° to 60° (the excitation angle is 120° in practice, but after the 180° ^{1}H pulse the final excitation angle is 60°), according to the optimised Ernst angle excitation, permits the reduction of the inter-scan delay from $1.25 \times T_1$ to $0.6 \times T_1$. The optimal excitation angle (α_e) is a function of the inter-scan delay (T) and R_1: $\cos(\alpha_e) = \exp(-R_1.T)$ [161]. The blue curve in Figure 1.14 shows α_e as a function of the interscan delay (T) for R_1=2Hz. The sensitivity per unit time depends on the interscan delay (T), α_e, R_1 and the time for

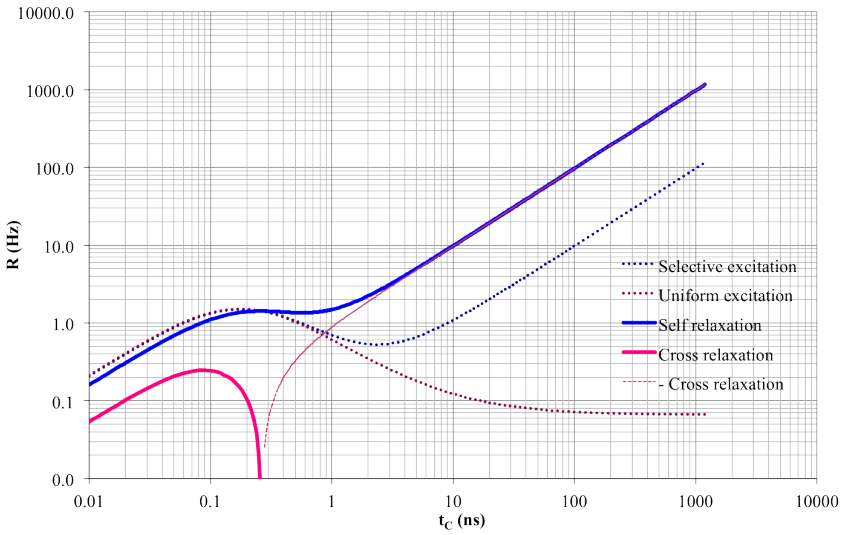


Figure 1.13: ^{1}H amide spin-lattice relaxation rates R_1 for a two-spin system as a function of the rotational correlation time, τ_c, calculated at a ^{1}H frequency of 700 MHz. Refer to the main text for the description of the proton auto-relaxation rate ρ_H and longitudinal cross-relaxation rate σ_{HH}. For uniform excitation, $R_1 = \rho_H + \sigma_{HH}$. For a selective excitation, an approximation of R_1 is shown assuming an isolated two-spin system, in which the contribution from cross-relaxation is down-weighted by a factor of 0.9 (refer to main text). A small protein domain has a correlation time of *ca.* 5-10 ns, whereas a spherical particle of the size of the ribosome has a correlation time of *ca.* 2500 ns [129]. This figure was generated using the expression for the relaxation rates in terms of the spectral density from [135].

one scan (T_{scan}) [159,161]:

$$I \propto \frac{1 - \exp(-R_1.T)\ \sin(\alpha_e)}{1 - \exp(-R_1.T)\ \cos(\alpha_e)\ \sqrt{(T_{scan})}} \tag{1.14}$$

and is shown in Figure 1.14 in red, right axis. The experiment therefore benefits from fast recycling yielding greater sensitivity per unit time. The result of multiple quantum acquisition in the indirect dimension is broader ^{15}N linewidths compared to single quantum acquisition (typically from 7 Hz to 15 Hz for the 16.7kDa protein

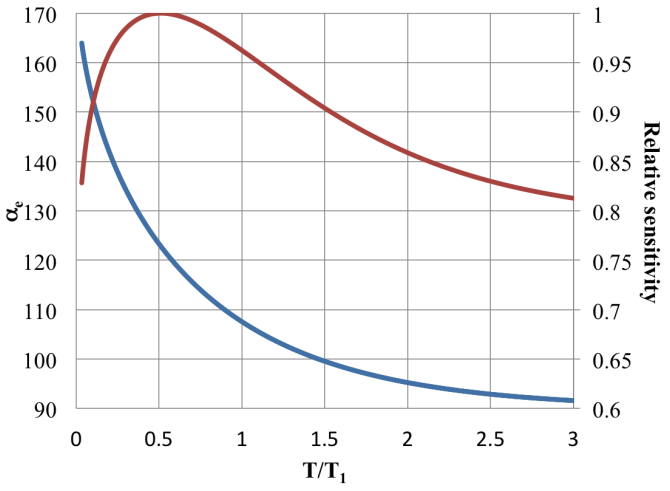


Figure 1.14: Optimised excitation angle α_e (left axis, blue curve) and relative sensitivity per unit time I (right axis, red curve) as a function of the recycling delay T noted as T/T_1, for T_1=500 ms.

calmodulin) [162]. In an attempt to overcome this issue, a BEST-HSQC scheme was developed using selective excitation or inversion for all proton pulses, thereby combining the advantages of both the SOFAST-HMQC and HSQC experiments [163]. The BEST-HSQC sequence was tested and was found not to be as sensitive as the SOFAST-HMQC. This difference is probably due to the increased number of proton pulses in the HSQC sequence, which must all be included as selective shaped pulses in the BEST-HQSC sequence. The SOFAST-HMQC therefore offers the best compromise to study high-MW complexes such as the ribosome.

1.3.4 Measurements of diffusion coefficients by NMR spectroscopy

As will be described in Chapter 2, Section 2.3.3 the lifetime of the ribosome complexes is limited, and monitoring of the sample stability proves to be of major importance in

gauging the length of time during which the integrity of the ribosome is maintained, and the ribosomal proteins or nascent chain are attached to the ribosome complex. As the release of the observable L7/L12 stalk (Section 1.2.3) does not result in a significant linewidth decrease of the L7/L12 resonances (because the stalk region tumbles independently of the ribosome core, Section 1.2.3), analysis of the 2D spectrum is not sufficient to detect release. The attachment of the dynamic ribosomal proteins (and ribosome-bound nascent polypeptide) can therefore be best monitored by NMR via the introduction of translational diffusion measurements [33]. Pulsed-field gradient (PFG) diffusion NMR spectroscopy typically allows the deconvolution of the translational diffusion coefficients of different species in solution that have distinct chemical shifts [164, 165]. The introduction of a PFG of length δ and strength G dephases the magnetisation as a function of the z-position of the particles (Figure 1.15A, arrow b). The magnetisation is then rephased by the introduction of a second PFG with opposite phase after a time Δ (Figure 1.15A, arrow c). By recording the progressive loss of signal intensity that results from increasing the strength of the PFG, one can extract the translational diffusion coefficient (D) from the Stejskal-Tanner equation [164]:

$$\frac{I}{I_0} = \exp\left[D\left(\Delta - \frac{\delta}{3}\right).(\delta.G.\gamma_{\mathrm{H}})^2\right] \tag{1.15}$$

The above equation assumes gradient pulses are applied with square shapes. In practice, gradient pulse are usually shaped to reduce eddy currents at the start and end of the pulse. Therefore, the product $\delta.G$, which represents the integrated gradient, is modified to include a shape factor, F_{shape}, that depends on the shape of the gradient pulses.

The accurate determination of small translational diffusion coefficients such as those associated with the ribosome (on the order of 10^{-11} $\mathrm{m^2 s^{-1}}$) requires a strong signal attenuation for the strongest gradient, which is achieved by using a long diffusion

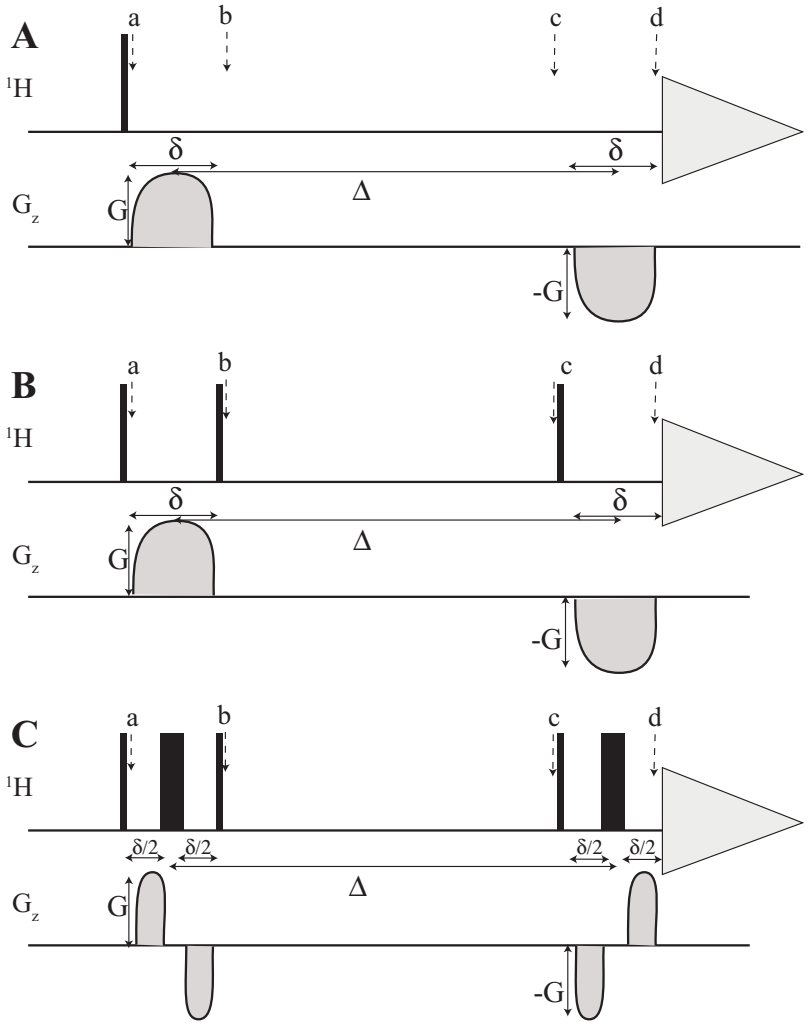


Figure 1.15: Schematic of three different diffusion pulse sequences. Narrow and wide bars show 90° and 180° pulses respectively. Refer to main text for definition of Δ, δ and G. **A:** PFG echo sequence. At the time indicated by the arrow a, the magnetisation is I_x. At time b, the magnetisation is dephased by the PFG, modulated along the z axis: $I_x.\cos(2.\gamma.G.\delta.z) + I_y.\sin(2.\gamma.G.\delta.z)$. At time c, the magnetisation is still phase modulated along the z axis, and the molecules are displaced by longitudinal diffusion. At time c, the magnetisation is rephased by the PFG, but the molecules have been displaced by diffusion, so that the magnetisation is modulated by equation 1.15. **B** STE (stimulated echo) sequence. **C** STE with bipolar gradient pulses.

delay and long gradient pulses. The choice of these parameters is typically dictated by limitations of the probe and by longitudinal and transverse relaxation occurring during delays in the sequence, as discussed below. To avoid the rapid return to equilibrium via transverse relaxation during the diffusion delay Δ in the PFG echo sequence (Figure 1.15A), the use of a stimulated echo (STE, Figure 1.15B) converts the magnetisation into I_z during Δ, which is associated with slower longitudinal relaxation. However, STE experiments are associated with a loss of 50% of the magnetisation that is not returned to the z-axis after the second 90° pulse (arrow b in Figure 1.15B) due to the dephasing of the magnetisation in the x-y plane. A loss of signal is thus unavoidable, but nonetheless, the fast transverse relaxation associated with biomolecules renders STE diffusion experiments the most sensitive. STE experiments allow the use of longer diffusion delays, Δ, although the choice of Δ remains limited by the longitudinal relaxation ($\Delta < T_1$) of the observed nucleus, and as mentioned in Section 1.3.3, T_1 can be as short as hundreds of ms to seconds in the case of ^{1}H longitudinal magnetisation. The choices for the other experimental parameters are also dictated by practical limitations: for example, the length of the gradient pulses, δ, is limited by the transverse relaxation, and the gradient strength, G, is limited by the probe capacity and the need to avoid extensive sample heating. The use of STE with bi-polar gradient pulses reduces eddy current effects (Figure 1.15C), and were used throughout this thesis.

In the case of selective isotopic labelling within a mixture or a complex (as for RNCs, as described further throughout this thesis), an attractive feature of translational-diffusion measurements by NMR is that the experiments can be nucleus-selective. The recent introduction of heteronuclear STE (XSTE) diffusion experiments [166] permits recording of only the magnetisation of ^{1}H nuclei that are coupled to ^{15}N or ^{13}C nuclei. The magnetisation during Δ is purely ^{15}N (^{13}C) longitudinal magnetisation as a result of

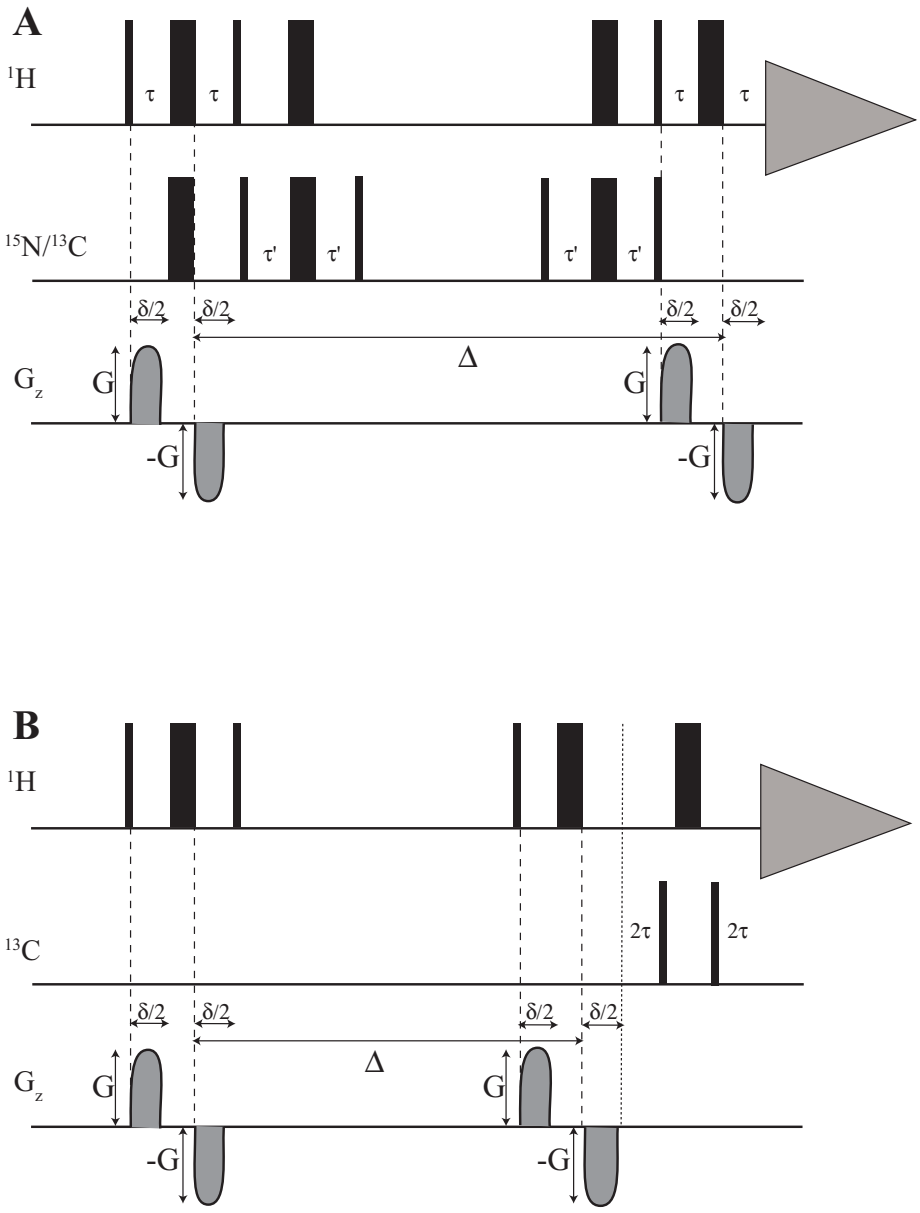


Figure 1.16: Schematic of heteronuclear PFG diffusion pulse sequences. Narrow and wide bars represent 90° and 180° pulses. $\tau = 1/4J_{XH}$. Refer to main text for definition of Δ, δ and G.**A** ^{15}N (^{13}C) XSTE diffusion experiment. $\tau' = 1/4J_{NH}$, and $\tau' = 1/10J_{CH}$ (refer to main text). **B** ^{13}C-edited STE-HMQC diffusion experiment.

the refocused-INEPT element that precedes Δ (Figure 1.16A). Note that in the case of the ^{13}C XSTE diffusion sequence, the delay of the second spin-echo in the first refocused-INEPT and the first spin-echo of the last refocused-INEPT is optimized (labelled τ' in Figure 1.16A) for a $^{13}CH_3$ group, i.e. $1/10J_{CH}$ instead of $1/4J_{CH}$. Nonetheless, the sensitivity of ^{15}N (^{13}C) XSTE diffusion experiments is limited due to the extensive transverse relaxation occurring during both refocused-INEPT elements of the pulse sequence (as well as the loss due to the STE mentioned earlier). Recently, ^{13}C nucleus selection has been introduced through the use of bi-polar STE prior to a conventional methyl-TROSY HMQC sequence (Figure 1.16B), which reduces the otherwise extensive line-broadening due to transverse relaxation [167]. In this case, transverse relaxation pathways are still active during the gradient pulses, but at least some of the methyl-TROSY coherences are preserved throughout the STE part of the sequence.

Several useful characteristics of the observed species can be extrapolated from the diffusion coefficient, such as the hydrodynamic radius, the rotational correlation time and an estimation of the molecular weight. The hydrodynamic radius (r_h) of the particles that gives rise to the STE diffusion spectra can be estimated from the experimentally measured diffusion coefficient using the Stokes-Einstein relation:

$$r_h = \frac{k_B.T}{6\pi.D.\eta} \tag{1.16}$$

where k_B is the Boltzmann constant (1.38×10^{-23}kg m^2 s^{-2} K^{-1}), $\eta(T)$ is the viscosity of water (in kg m^{-1} s^{-1}, at temperature T) with [168]:

$$\eta(T) = 2.41 \times 10^{\frac{247.8}{T-140}-5} \tag{1.17}$$

Assuming a spherical geometry, the rotational correlation time can be estimated from

the hydrodynamic radius using the Stokes-Einstein-Debye relation:

$$\tau_c = \frac{4\pi.\eta_T.r_h^3}{3.k_B.T} \tag{1.18}$$

Lastly, assuming a spherical particle, the molecular weight of the particle that has a given r_h can be estimated by the following equation:

$$MW = \frac{4\pi.N_A.(r_h - r_w)^3}{3.\nu_0} \tag{1.19}$$

where N_A is the Avogadro number (6.02×10^{23} mol^{-1}), r_w is the thickness of the water layer around the particle (estimated to be around 0.2nm [134]), ν_0 is the specific volume for the particle (estimated to be 7.3×10^{-7} m^3.g^{-1} for proteins and 5.5×10^{-7} m^3.g^{-1} for nucleic acids) [169].

1.3.5 Nuclei labelling

The aim of this work is to analyse the folding equilibrium of nascent chains as they emerge from the ribosome, and therefore, both the folded states and unfolded states are investigated in parallel by heteronuclear NMR spectroscopy. Typically, folded states are often associated with wide chemical shift dispersion of the amide ^{1}H, ^{15}N and ^{13}C nuclei [134], and therefore overlap of resonances is reduced using ^{1}H-^{15}N and ^{1}H-^{13}C 2D NMR spectroscopy. Moreover, the methyl ^{1}H chemical shift can undergo high-field shifting due to the ring current effect from nearby aromatic residues [170], and are therefore sensitive probes for structural interactions. Unfolded states, however, are often associated with narrow chemical shift dispersion of the ^{1}H and ^{13}C nuclei (of side chains), and only the ^{15}N nucleus (as well as the ^{13}C carbonyl) remains dispersed due to the strong influence of the residue type on its chemical shift [134]. As a consequence, ^{15}N labelling is

required for the study of the unfolded states by ^{1}H-^{15}N SOFAST-HMQC spectra (Section 1.3.3) and ^{1}H-^{13}C labelling is preferred to monitor the folded state due to the higher sensitivity of the methyl group over the backbone amide group (which results from the three equivalent protons and typically low order parameter of the sidechain methyl groups).

Moreover, unfolded states are highly flexible and are therefore associated with a very fast effective correlation time. A persistence length on the order of seven residues for unfolded states has been reported [171], so that the overall size of the protein has very little influence on the effective correlation times of individual residues in an unfolded peptide. Folded states, on the other hand, are globular structures with longer effective correlation times, that strongly depend on the size of the molecule (as equation 1.19 reports, *ca.* 10 ns for a 20 kDa domain). Therefore, the study of folded RNCs requires strategies to reduce linewidths, such as the methyl-TROSY HMQC experiment, together with a high deuteration level [121].

Lastly, the effect of temperature is also to be considered. Increasing the sample temperature during NMR acquisition is often used as a means of decreasing the rotational correlation time of high MW complexes, however, the stability of the ribosome complex is significantly reduced with increasing temperature [172]. Moreover, at high temperatures, the amide protons of unfolded states typically undergo faster exchange with the protons of the solvent, which results in significant line-broadening [173]. In light of these conflicting considerations, a temperature of 25°C was chosen as a compromise in studies of both the ribosome and the RNCs, as described in this thesis.

1.4 Conclusion

This introduction has summarised on how the structures of the ribosome complex from X-ray and cryo-EM have allowed an improved understanding of the mechanism of translation. However, these structural studies have provided little information on the conformations of the nascent polypeptide chains. NMR spectroscopy, which has been demonstrated to be a suitable tool for the study of protein folding, can report on different states occupied in the folding pathway as it occurs for isolated protein: the unfolded ensemble, the intermediate states, and even recently reporting on the transition states. There is little structural information on co-translational folding pathways, and the aim of this work is to use NMR spectroscopy to study the folding equilibrium of RNCs at different stages of translation. The study relies on the careful optimisation of a series of advanced NMR experiments suitable for RNCs.

Chapter 2

A strategy for the production and NMR analysis of ribosome-bound nascent chains complexes

2.1 Introduction

The central aim of this thesis is to study co-translational protein folding via NMR spectroscopy by creating snapshots of the emergence of a domain from the ribosomal exit tunnel via stalled RNCs. Generating homogeneous and stable lengths of newly synthesised nascent chains at a concentration amenable for NMR study, and with the appropriate isotopic labelling, is a significant challenge, and here an *in vivo* approach to produce these RNC samples has been explored. Additional challenges for the NMR study of RNCs reside in the low sensitivity due to the low maximal molar concentration (10 μM), the slower tumbling of the large RNC, and the limited sample lifetime. This chapter presents the development of a set of methodological tools (both spectroscopic and biochemical) to facilitate the detailed NMR study of RNCs and co-translational folding. Chapter 3 will subsequently describe the NMR investigation of

co-translational folding of increasing lengths of an ddFLN-immunoglobulin RNC using this methodology.

2.1.1 Production of RNCs for NMR studies

Upon completion of synthesis, the translation process is terminated by auxiliary factors which interact with the ribosome to expel the synthesised NC. Preparative methods to homogeneously arrest translation and keep the newly synthesised polypeptide chain on the PTC have been developed over the recent years [5] and allowed the generation of RNCs for biochemical and biophysical studies. These are either *in vitro* transcription-translation [174–176] or *in vivo* methods [158, 160, 177, 178], and have resulted typically in very small quantities of RNCs (1-10 μL of 1-5μM). Using linearised DNA or mRNA transcripts, that lack a stop codon to terminate translation [130, 179], together with *in vitro* transcription-translation reactions [175, 176] results into synchronised translational stalling. For studies by NMR spectroscopy, the production of high quantities of selectively isotopically labelled NC bound to isotopically-silent ribosomes is needed. Indeed, the ribosome is not NMR silent as previous studies have shown that the L7/L12 stalks gives rise to well dispersed signals in ^{1}H-^{15}N HSQC spectra [33], which can interfere with the analysis of the RNC. With unlabelled ribosomes in the *in vitro* transcription-translation reactions, the use of isotopically labelled amino-acid allows the selective labelling of the NC alone [130]. The pure cell-free production method allows, in principle, control of all aspects of translation, including the co-factors being present during translation [174–176]. This method, however, is very costly in terms of routine usage and has therefore only been used for the production of small amounts of RNCs. The conversion to *in vivo* production methods to generate RNCs in a cost effective way is not permitted by the use of linearised DNA or mRNA transcripts, thus recently

introduced stalling motifs are used *in vivo* [178].

The application of translation-arrest motifs, such as those derived from the secretion monitoring protein (SecM) [180] and the tryptophan operon (tnaC) [26] have opened the door to *in vivo* production of RNC [132, 158, 160, 178]. In the case of SecM, a 17 amino acid stalling region is recognised by the ribosomal exit tunnel, and translation is stalled in the pre-translocation step prior to peptide bond formation with the Pro-tRNA^{166} at the A-site and Gly-tRNA^{165} at the P-site [40]. Cryo-EM studies showed electron density revealing a shift of the carbonyl-carbon of the Gly-tRNA^{165} away from the Pro-tRNA^{166} (Figure 2.1A & B), and suggest that this move disables translocation. Contacts of the SecM residues with regions of the 23S rRNA that line the ribosomal exit tunnel are thought to mediate the conformational changes of the rRNA network that accommodate the P-site 3′ aminoacylated A nucleotide (Figure 2.1A).

An *in vivo* approach is not only cost-effective, but it also offers the exciting future possibility of examining protein folding directly within the cellular environment. The *in vivo* approach that we introduced uses *E. coli* to generate large quantities of SecM-stalled RNC for NMR studies [132,158,160]. NMR study of RNCs produced *in vivo* requires the manipulation of the growth and expression media to enable selective isotopic labelling to take place for the NC alone. In this study, the *in vivo* method has been developed in collaboration with Dr Lisa Cabrita (UCL) and is described in Section 2.3.1.

2.1.2 NMR characterisation of RNCs complexes

As NMR spectroscopy has a significantly greater sensitivity for fast tumbling molecules (τ_c <50 ns) compared to slower tumbling high molecular weight complexes, the release of the NC from the ribosome can potentially give rise to additional or higher intensity cross-peaks in (heteronuclear) spectra. A conservative philosophy to analyse NMR data

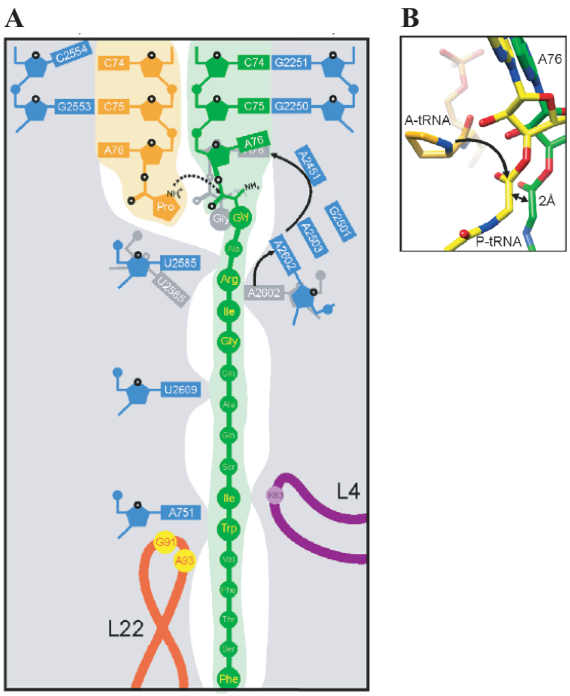


Figure 2.1: **A:** Schematic of the conformational changes of the ribosomal rRNA and the P-site tRNA induced by the SecM sequence. In particular, the P-site A76 (3′ aminoacylated) is shifted away from the A-site tRNA (dotted arrow) and prevents translocation **B:** Overlay of the P-site 3′ aminoacylated A nucleotide models from cryo-EM maps of SecM stalled RNC in green and TnaC stalled RNC in yellow that highlights the moves of the P-site tRNA 3′ ends induced by the SecM sequence. Both A and B are adapted from [40]

from RNC is therefore adopted: i.e. the null hypothesis is that signals are not from RNC, and collection of strong evidence is required to accept the alternative hypothesis that the signals indeed arise from RNCs. It is essential to monitor the stability, integrity and attachment of the NC to the ribosome complex before and during NMR data acquisition. An array of NMR and biochemical tools for the characterisation of the RNC have been developed in this study and are presented in the latter section of this Chapter (Section 2.3.3).

2D ^{15}N-^{1}H SOFAST-HMQC and ^{13}C-^{1}H HMQC spectra are recorded in order to

assess the structure and dynamics of the RNCs, but given the low concentrations present, signals may be weak even after several days of acquisition. The analysis of such spectra can be non-trivial, therefore we first discuss the development of appropriate statistical methods for the analysis of very low intensity spectra (Section 2.2). The application of such methodology to analyse the sensitivity of the ^{15}N-^{1}H SOFAST-HMQC of RNCs is presented in Section 2.3.2.3 and the sensitivity of the ^{13}C-^{1}H HMQC of RNCs is presented in Sections 2.3.2.5 & 2.3.2.7.

Together, this Chapter describes the set of tools that were used in the detailed study of co-translational folding of a ddFLN-dom5 that is presented in the next Chapter 3.

2.2 Theory: Development of a statistical analysis methodology to quantify signal observability

The sensitivity of NMR spectra recorded on ribosomes and particularly on RNC remains low, partially because of the low maximum concentration of the ribosome complex (10μM) and because of the size of the complex. A method has been developed that allows the definition of a threshold for the observability of resonances based on the noise level of the spectrum, which has a gaussian distribution with a standard deviation σ_{noise}. Loosely, a resonance is observable when its intensity is significantly higher than the noise. At low signal to noise however, this significance must be quantified more formally, using standard statistical methods. By normalising the signal intensity by the noise, we define the z-score (or signal to noise ratio, SN):

$$z = \text{SN} = \frac{\text{signal intensity} - \mu_{\text{noise}}}{\sigma_{\text{noise}}} \tag{2.1}$$

where μ_{noise} is the mean of the noise of the spectrum and σ_{noise} is the standard deviation of the noise. For a properly baselined spectrum, $\mu_{\text{noise}} = 0$. The z-score is sampled from the standard normal distribution of the noise, thus, a z-score of >1.64 indicates that the signal intensity has <5% chance of arising from the noise distribution. This significance level is used throughout the thesis.

To explore in practical terms the effect of sensitivity and processing in signal observability, a series of spectra from a purified isolated immunoglobulin domain (9μM) were recorded, with increasing number of scans (NS) from 4 to 1024. Peaks were picked at fixed positions from the most sensitive spectrum, and noise data were picked within the same chemical shift range as the signals. The z-scores of noise and signal intensities of these spectra are shown in Figure 2.2A. As expected by definition, the distribution of the z-scores of the noise is centred at 0, and has a standard deviation of 1. Spectra recorded with NS=1024, 64 and 16 scans are shown in Figures 2.2B, C & D, and most of the resonances that were observed in the highest sensitive spectrum are not observed "by eye" in the least sensitive spectrum (NS $\leq$ 64). This is reflected in the distribution of the z-score of the signal intensity in the spectra recorded with a low number of scans, in which most of the signals were associated with a z-score lower than 1.64 (dotted line in Figure 2.2). Nonetheless, despite the fact that only few resonances were observable in these spectra, the centre of the signal intensities distribution is higher than the centre of the noise distribution, clearly demonstrating that the analysis of the entire distribution rather than single signals, allows us to confirm the presence of the spectra even when individual resonances are not observed.

The distribution of signal intensities narrows when the overall intensity of the spectrum is reduced. This is reflected in Figure 2.3, which shows the standard deviation of the signal distribution (σ_{signals}) as a function of the number of scans. The standard

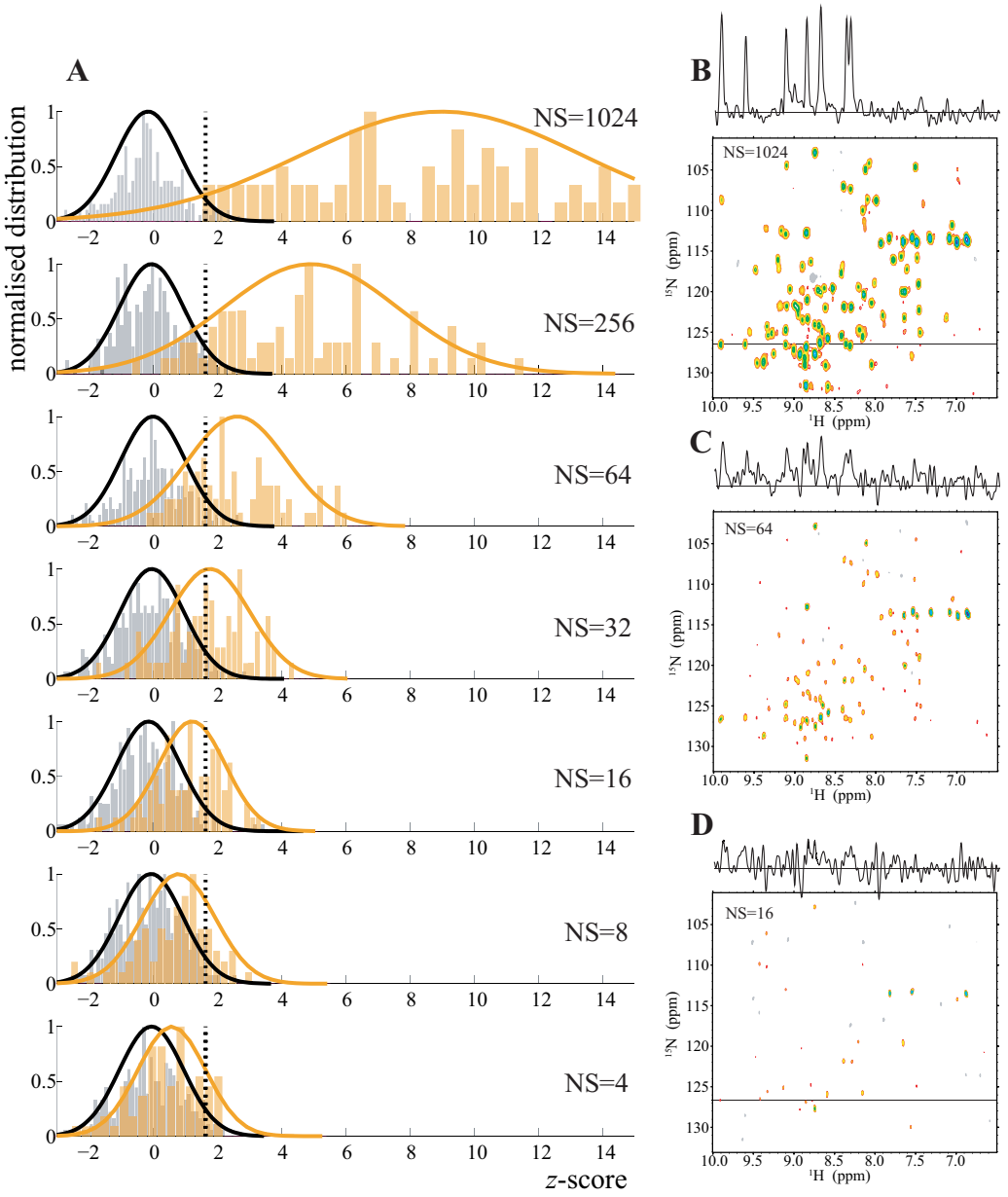


Figure 2.2: **A:** Signal distribution (orange) and noise distribution (grey) for immunoglobulin domain ddFLN-dom5 spectra recorded with decreasing number of scans. The spectra are processed with no zero filling nor window function to avoid bias (refer to main text). The distributions are normal gaussian distributions. **B, C & D:** Spectra of isolated immunoglobulin domain (ddFLN-dom5) recorded with decreasing number of scans, from NS=1024 to NS=16. A slice at ^{15}N frequency = 126.7ppm is shown on the top. The grey signals are negative.

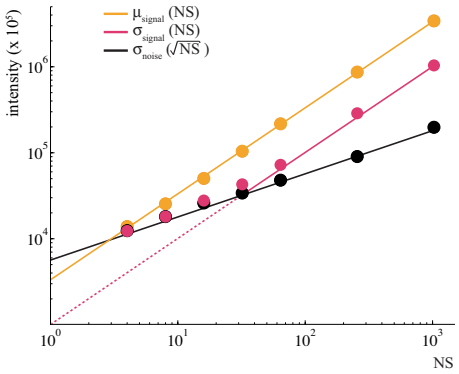


Figure 2.3: Plot of the centre of the peak intensity distribution (μ_{signals}, orange), the standard deviation of the peak intensity distribution (σ_{signals}, magenta) and the standard deviation of the noise distribution (σ_{noise}, grey), as a function of the number of scans used to record the spectra, on a log-log scale.

deviation increases linearly with NS, as does the intensity (Figure 2.3). However, as the overall intensity is lowered, the width of the signal intensities distribution modelled by σ_{signals} deviates from the theoretical $\sigma_{\text{signals}}<\sigma_{\text{noise}}$ (magenta versus grey lines), and in reality, $\sigma_{\text{signals}} = \sigma_{\text{noise}}$. The consequence of this observation is that for low SN data, the observed resonances are not necessarily the ones that have intrinsically the sharper linewidth and higher intensity, but rather are randomly distributed due to the gaussian noise distribution.

The effect of applying window functions and zero-filling to the z-scores of the observed signals is shown in Figure 2.4. The mean z-score is improved by applying both window function which effectively is averaging the noise because the multiplication becomes a convolution in the Fourier domain (green curve in Figure 2.4), or zero-filling to extend the FID domain which results in increased resolution in the Fourier domain, and allows more accurate peak picking (orange curve in Figure 2.4).

This basic analysis is highly dependent on the peak-picking, and how the signal

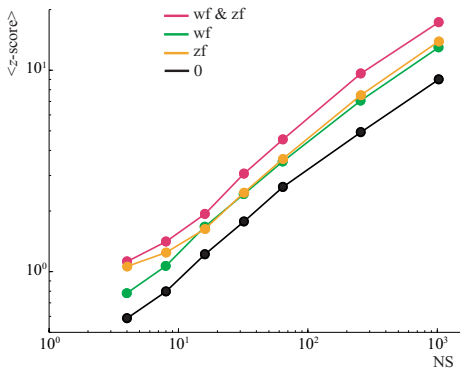


Figure 2.4: Plot of the centre of the mean of the signal intensities z-scores as a function of the number of scans used to record the spectra on a log-log scale, when using different processing modes: no window function nor zero filling (black), a sine-bell window function and no zero filling (green), no window function and zero filling to double the size of the real data (orange), and both window function and zero filling (magenta).

intensity is defined. It is therefore important to introduce the uncertainty of the signal position into the analysis. This is rarely considered in spectra with high SN, but at low SN can become an important factor as shown in Figure 2.5.

The uncertainty in the signal position is directly proportional to its linewidth, as depicted in equation 2.2 and inversely proportional to its SN [181,182].

$$\sigma_{\mathrm{Hz}} = \frac{1}{\sqrt{2}} * \frac{\mathrm{LW}}{\mathrm{SN}} \tag{2.2}$$

For a typical RNC NMR measurement, SN is close to 2, and the linewidths are >40Hz in the ^{1}H dimension and >100Hz ^{15}N dimension. Based on Equation 2.2, the uncertainty of the resonance frequency is therefore 0.02ppm in the ^{1}H dimension and 0.5ppm in the ^{15}N dimension. The spectra typically have a digital resolution of 0.014ppm/0.2ppm (^{1}H/^{15}N), where the limited resolution is due to the short acquisition times that are needed to benefit from the fast recycling rate of the SOFAST-HMQC

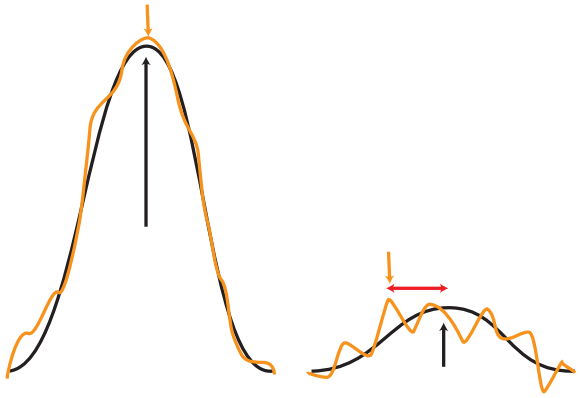


Figure 2.5: Schematic of a resonance with high SN (left) and low SN (right). The black line shows the theoretical signal and the orange line shows the addition of the noise to the signal. The black arrows shows the theoretical resonances chemical shift, the orange arrows show the maximum of the experimental spectra, and the red arrow show the chemical shift difference between the experimentally determined peak position and the theoretical peak position for low SN signals.

(Section 1.3.3), and the fast R_2 during acquisition. The intrinsic uncertainty on the signal positions (0.02ppm/0.05ppm), is therefore on the order of the digital resolution (0.014ppm/0.2ppm). As a consequence, a signal cannot be localised to a single data point, but only to a region of interest (ROI) of at least 3×3 data points (or 9 variables). The signal intensity is then defined as the maximum within this 3×3 ROI. This maximum cannot be compared to the normal noise distribution represented by a gaussian distribution and its standard deviation, but needs to be compared to the distribution of maxima of random 3×3 ROIs distributed within the noise region of the spectrum. The nine-variate normal distribution $f(x_1, x_2, ...x_9)$ is the extension of the noise distribution from a single variable to a 3×3 ROI, as such it is a function of the correlation between neighbouring points (ρ) and the gaussian noise standard deviation (σ). The correlation between neighbouring peaks (ρ) is introduced during processing by the window function and zero filling. The spectra recorded on the immunoglobulin

1.0	0.05	-0.01
-0.01	0.00	0.1
0.02	0.00	-0.01

1.0	0.6	0.05
0.4	0.2	-0.08
-0.06	-0.07	-0.04

Table 2.1: Example of a correlation matrix between neighbouring points, on the left for a spectrum processed without window function nor zero-filling, and the right for a spectrum processed with a sine-bell window function and a zero filling to double the size of the Real data. The shaded points of the matrix shows the cases where the correlation is strong (with the i+1, j+1 datapoints, where i is the ^{15}N position and j is the ^{1}H position).

domain were used to optimise the processing such that the SN is maximised and the correlation is minimised. An acceptable correlation does not spread beyond the 3×3 ROI (Table 2.1).

For the ROI to be an observable signal, the probability that the maximum of the ROI (X) is significantly higher than the noise distribution of 9 variables $x_1, x_2, ...x_9$ needs to be close to 1:

$$F(X) = P(\text{all } x_i \leq X) = F(X) = P(x_1 \leqslant X \cap x_2 \leqslant X \cap ... \cap x_9 \leqslant X) \tag{2.3}$$

Because of the correlation between neighbouring points ρ, $F(X)$ can be rewritten as:

$$F(X) = P(x_1 \leqslant X) * P(x_2 \leqslant X|x_1) * ... * P(x_9 \leqslant X|x_1, x_2, x_3, ..., x_8) \tag{2.4}$$

This probability is the multivariate normal cumulative density function (mvncdf) at X, it is the integral of $f(x, x, x, ...)$ from X to $+\infty$ in each nine dimensions:

$$F(X, X, X...) = \int_{-\infty}^{X} \int_{-\infty}^{X} ... \int_{-\infty}^{X} f(x_1, x_2, ..., x_9) dx_1 dx_2 ... dx_9 \tag{2.5}$$

If $F(X, X, X, ...)$ is higher than 95%, the maximum in the 3×3 ROI has less than 5% chance to be generated by the noise, we therefore conclude that the observation is a genuine resonance. This 95% threshold is the equivalent of a z-score of 1.64 for a normal gaussian distribution.

Figure 2.6A shows the histograms of the z-score of signal intensity as defined by the maximum within a 3×3 ROI, as well as the nine-variate normal distribution of the z-score of the noise at varying sensitivity (dotted line). Note that now the distribution of the z-score of the noise are not centred at 0. The vertical black line shows the value at which the mvncdf=95% (a signal which intensity is above this line has less than 5% chance of being generated by the noise). At a high sensitivity (NS=1024), the intensity of each resonance is higher than the noise threshold. However, with a lower sensitivity (NS$\leq$64), the distribution of signal intensity overlaps with the noise multivariate distribution, and a random set of signals are not observable. As the search radius is increased, statistical significance demands a greater signal intensity. This is reflected by the increased z-score of the 5% threshold, from 1.64 for a single point, to *ca.* 2.5 when using a 3×3 ROI, to 2.8 when using a 5×5 ROI (Figure 2.7). The consequence of this is that when the chemical shift is allowed to vary, the required intensity to determine the significance of a resonance is higher. New resonances that are different from either the assigned native state or disordered states will be identified only if they have a strong intensity (SN >3).

Throughout this work, the analysis of resonances intensities is performed using the maximum of the 3×3 ROI centred at either the resonances assigned to either the native state of the ddFLN-dom5, or the 8M urea denatured state. Both the centre of the peak intensity distribution and individual resonances are compared to the noise. When the centre of the peak intensity distribution is shifted from the noise, a spectrum is present, even if individual resonances do not have a statistically significant SN. However, in

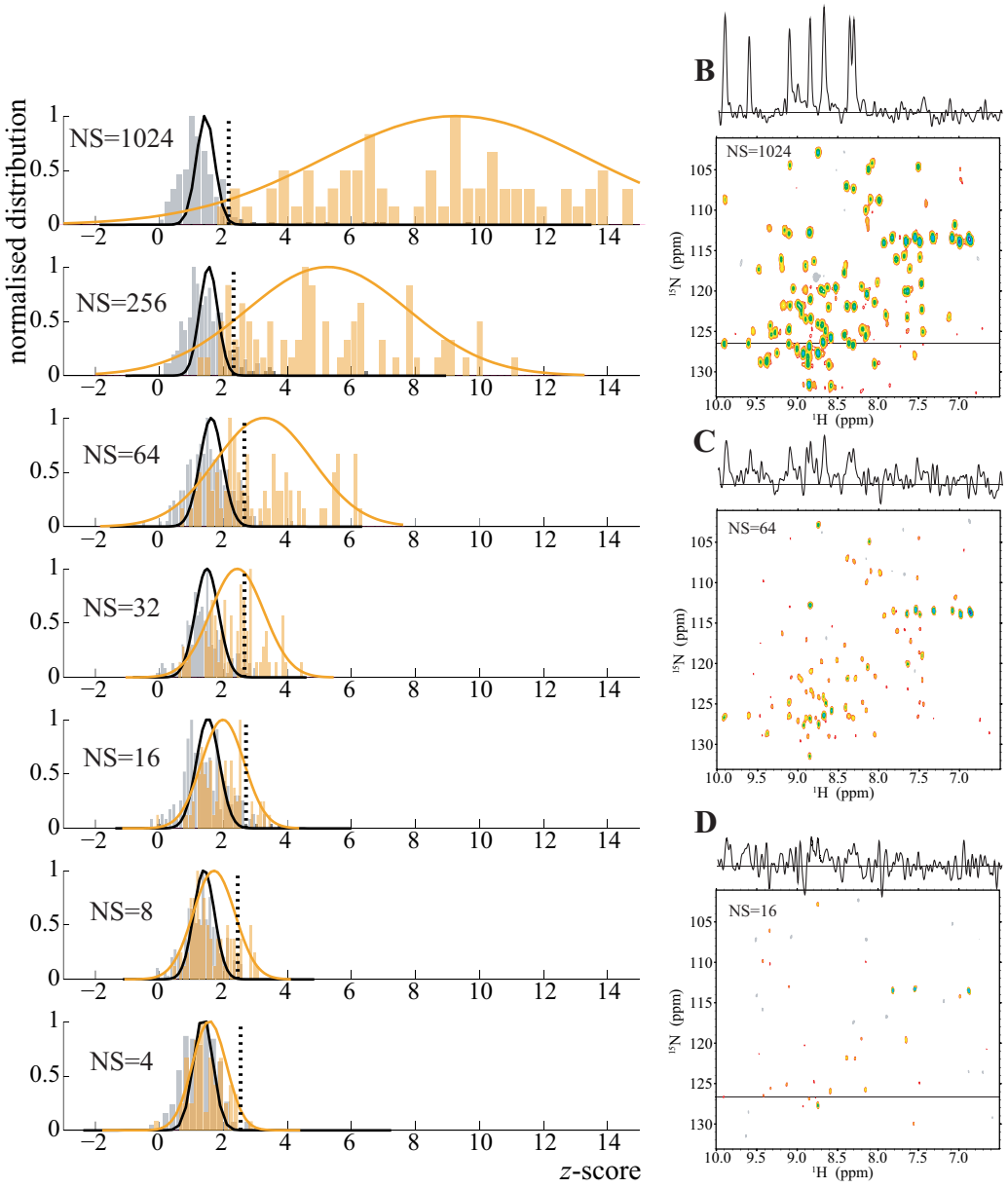


Figure 2.6: **A:** Distribution of maxima of 3×3 ROIs centred on the noise region (grey) and at signal positions (orange) in the same spectra as shown in Figure 2.2. The same scale as Figure 2.2 is used, i.e. normalised with the true standard deviation of the noise. **B, C & D:** Spectra of isolated immunoglobulin domain (ddFLN-dom5) as in Figure 2.2.

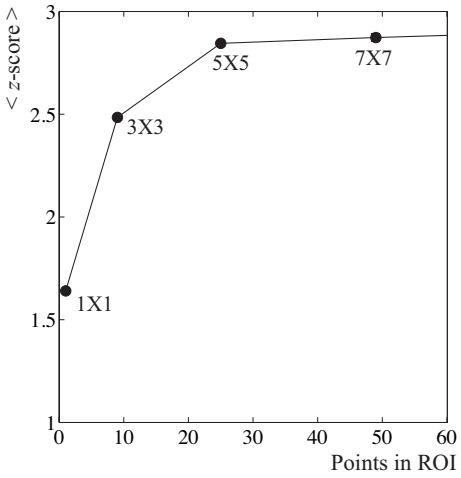


Figure 2.7: Plot of the *z*-score of the 5% threshold as a function of the uncertainty in the peak position, experimentally determined from noise regions of ^{1}H-^{15}N HSQC spectra.

this case, it is not possible to compare individual resonance intensities as the width of the peak intensity distribution depends uniquely on the noise. Also, for a non-native resonance to be identified, its SN needs to be higher than 2.8, i.e. it is higher than the noise allowing an extremely wide ROI.

2.3 Results and Discussion

2.3.1 *In vivo* production of SecM-stalled RNCs

2.3.1.1 SecM-RNC constructs

This Chapter describes the methodology that was developed to study co-translational protein folding by NMR, with a model domain used as an example: domain 5 from the gelation factor (ABP-120) from *Dictyostelium discoideum* (ddFLN-dom5). Translation of varying lengths of the subsequent domain ddFLN-dom6 in addition to ddFLN-dom5 offers the opportunity to observe the emergence of ddFLN-dom5 from the exit

tunnel. The conformational states occupied by ddFLN-dom5-RNC at different stages of translations of the ddFLN-dom6 linker were monitored by NMR spectroscopy, making use of large quantities of homogeneously stalled ddFLN-dom5+6-RNC. Previous NMR studies of the ddFLN domains have shown that the truncation of the sequence coding for the final G-strand of ddFLN-dom6 renders the domain folding incompetent [183] (Figure 2.8B) such that it acts as a flexible linker tethering the folding-competent ddFLN-dom5 to the PTC. To generate snapshots of the emergence of ddFLN-dom5 from the ribosome exit tunnel, a series of RNC constructs were designed which have increasing lengths of ddFLN-dom6 (Figure 2.8C). At the C-terminus of the construct, 17 amino-acids from SecM allow translational pausing via the binding of the NC to the ribosome tunnel (Figure 2.8A), while at the N-terminus there is a 6×His affinity tag to facilitate affinity purification (Figure 2.8C).

While in this study, the linker is the disordered ddFLN-dom6 truncation, the nature of the linker can be modified, and only requires being disordered to provide sufficient flexibility to the RNCs to allow NMR observation.

2.3.1.2 In vivo production of selectively labelled RNCs

The *in vivo* method to generate selectively-labelled ddFLN-dom5-RNCs is described in the Material and Methods, Chapter 6 [158]. Briefly, *E. coli* cells were grown in unlabelled media to an OD_{600} ∼4, and then transferred into isotopically-enriched (with $^{15}NH_4Cl$, ^{13}C-glucose) media for expression (Figure 2.9A). Typically, 60 to 80 nmoles of ribosomes were recovered from 500ml of cell culture after cell lysis and sucrose cushion (Figure 2.9C & D). 6 nmoles of NC-occupied ribosomes were typically isolated after Ni-IDA affinity chromatography (Figure 2.9D). Finally, 3 nmoles of pure ddFLN-dom5-RNCs were typically recovered from the 10% to 35% sucrose gradient (Figure 2.9D). The silver-

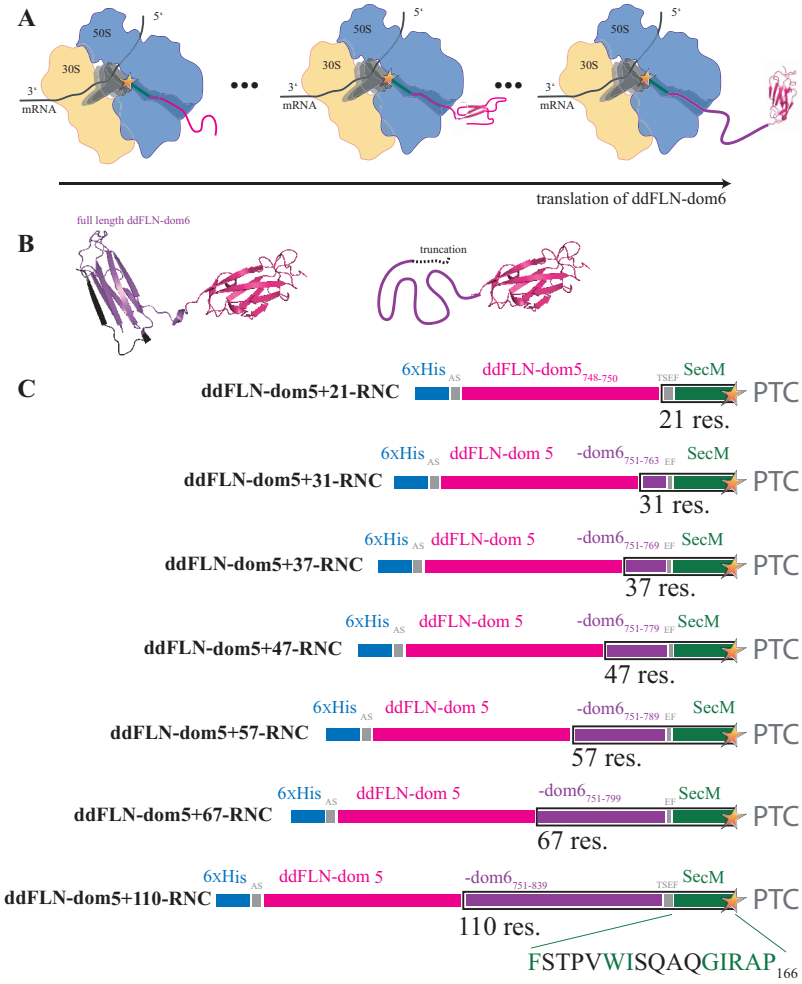


Figure 2.8: A: Schematic of snapshots of the emergence of ddFLN-dom5-RNC from the ribosomal exit tunnel at different lengths of translation of ddFLN-dom6 (purple). The SecM stalling sequence is shown in green, and ddFLN-dom5 is shown in magenta. **B:** Left: crystal structure of ddFLN (pdb ID: *1QFH.pdb*) where ddFLN-dom5 is shown in pink and ddLFN-dom6 in magenta. Right: schematic of the truncation of the C-terminal strand of ddFLN-dom6 (shown in black) rendering the domain folding incompetent. **C:** Simplified map of the ddFLN-dom5-NC constructs used in this study. Varying lengths of ddFLN-dom6 are used for different ddFLN-dom5-RNCs, from ddFLN-dom5+21-RNC which incorporate none of the ddFLN-dom6 sequence, to ddFLN-dom6$_{751-840}$ to produce ddFLN-dom5+110-RNC. The numbering in the name of the RNCs corresponds to the number of amino-acids from the C-terminal end of ddFLN-dom5 to the PTC centre (black rectangle).

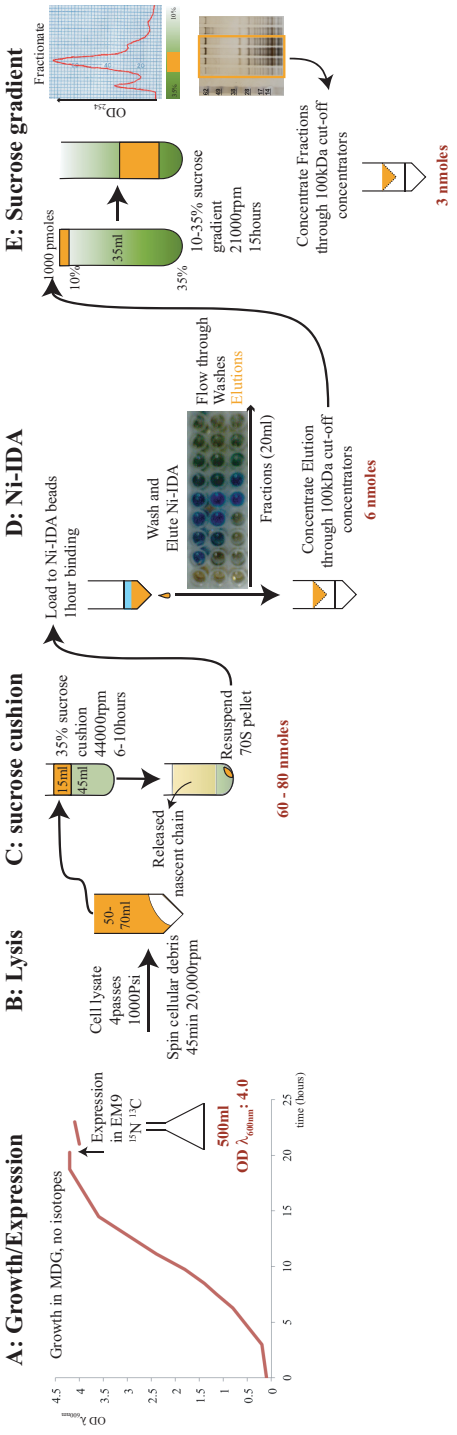


Figure 2.9: Schematic representation of the ddFLN-dom5-RNC preparation. Refer to main text for details.

stained SDS-PAGE of these purified RNCs showed the typical pattern of the 54 ribosomal proteins (Figure 2.9D), and in addition, the absorbance ratio OD_{260}/OD_{280} gave a value 1.9 ± 0.1, which was indicative of a homogeneous ribosome sample [184]. A lower ratio is indicative of an excess of proteins and a higher ratio indicates loss of ribosomal proteins.

Another important aspect of the RNC preparation is the NC occupancy level. The occupancy of the nascent chain in the ribosomes throughout the purification was monitored by 6×His western blot, where 10pmoles of 70S complexes after each purification steps are loaded along side 10, 5 and 2pmoles of purified ddFLN-dom5 with a 6×His tag (Figure 2.10A). The level of immunofluorescence from the 10pmoles of 70S is quantified and compared to the immunofluorescence from the known quantity of isolated ddFLN-dom5 controls to give an estimate of the NC quantity (Figure 2.10B). The immunofluorescence of 10pmoles of the ribosome pellet after the first purification step indicates that <20% of the ribosomes are occupied with a nascent chain (Figure 2.10A & B blue squares). The subsequent affinity purification improved the overall NC-occupancy in the sample, yielding 75-80% of RNCs (Figure 2.10A & B orange diamonds). The high occupancy is maintained after sucrose gradient purification (Figure 2.10A & B red diamonds), suggesting that the NC remains attached during this final purification step.

Moreover, the purified RNCs were also assessed for the presence of co-purified trigger factor (TF), using an anti-TF antibody (Figure 2.10C). It was found that ddFLN-dom5-RNCs typically contained <5% TF (Figure 2.10D), which is a very small amount as TF is typically present in a equimolar ratio with ribosomes in the cellular environment [94], and is unlikely to interfere in any significant way with the NMR observations of the NC.

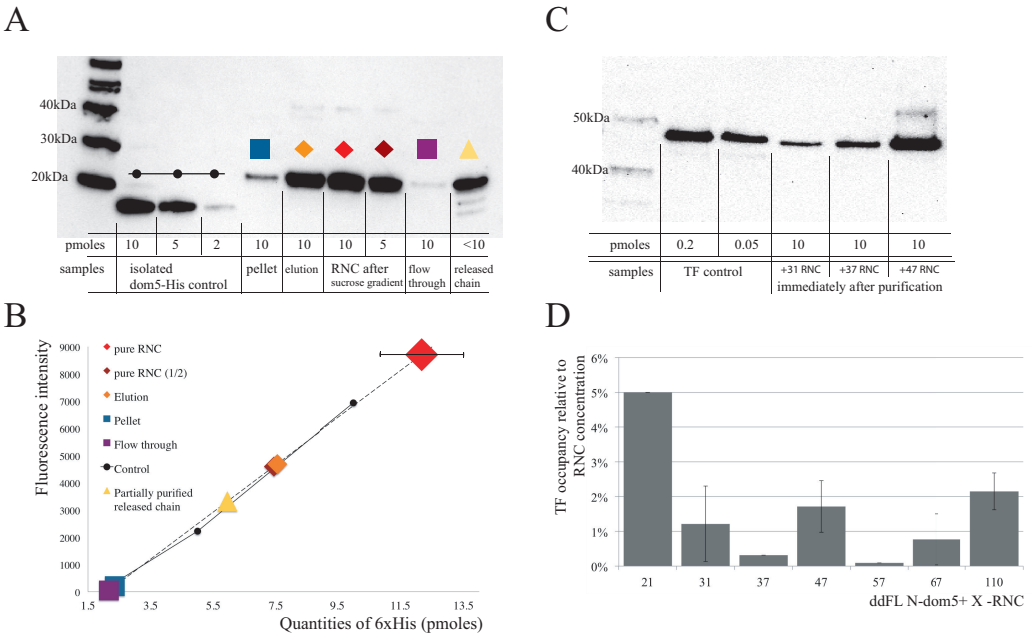


Figure 2.10: **A:** Anti-His western blot of samples collected at different stages of the ddFLN-dom5+47-RNC purification. The quantities in pmoles indicated are based on quantification of 70S ribosome in each sample from OD_{260} measurement. **B:** Quantification of the ddFLN-dom5+47-RNC sample based on the signal intensity of the western. The dotted line is the linear fitting from the three control samples (isolated ddFLN-dom5). The error bar for purified ddFLN-dom5-RNC corresponds to the difference in occupancy found for the 10 and 5 pmoles samples (brown and red diamonds). **C:** Anti-TF western for ddFLN-dom5+31-RNC, ddFLN-dom5+37-RNC and ddFLN-dom5+47-RNC. **D:** Quantification of TF in the ddFLN-dom5-RNC samples based on the signal intensity of the western. The error bars correspond to standard deviation of the TF occupancy found in two to three independent preparations.

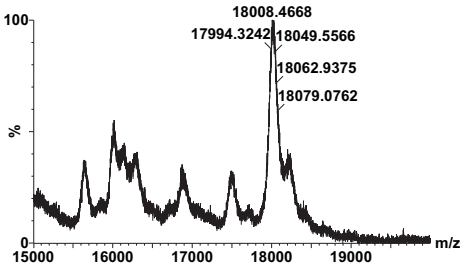


Figure 2.11: MALDI mass spectrum of partially purified and released ddFLN-dom5+47-NC. The expected molecular weight of ^{1}H-^{12}C-^{14}N ddFLN-dom5+47-NC is 17,115Da. The expected molecular weight of ^{1}H-^{13}C-^{15}N ddFLN-dom5+47-NC is 18081Da. Recorded by Dr Lisa Harris, UCL.

2.3.1.3 Quantification of the selective labelling of RNCs

The acquisition of heteronuclear correlation NMR spectra requires homogeneous isotope enrichment of the nascent chain with ^{13}C and/or ^{15}N. To achieve this during *in vivo* production of the RNC, the growth media contains ^{12}C glucose and $^{14}NH_4Cl$ to allow production of NMR-silent ribosomes, while the expression media is isotopically enriched (^{13}C glucose or $^{15}NH_4Cl$) to allow for the production of ribosome-bound nascent chains with NMR-active heteronuclei [158]. The homogeneous isotopic labelling of the NC was monitored by performing MALDI-TOF mass spectrometry of the purified release chain from the cells expressing ddFLN-dom5-RNCs. The expected mass increase as a result of isotopic labelling is reflected in the spectrum (Figure 2.11), where the NC has of a mass of 18.0 kDa compared to the expected value of 17.1 kDa for unlabelled protein. The 966 Da difference in mass showed that the released ddFLN-dom5-NC is 99% labelled.

During the expression period of the RNCs, i.e. during the 60 mins of induction, the cells produced some further ribosomes. While the expression protocol keeps the expression period to a minimum, the ribosomes produced during this period are isotopically labelled. The extent of ribosome labelling was quantified by recording

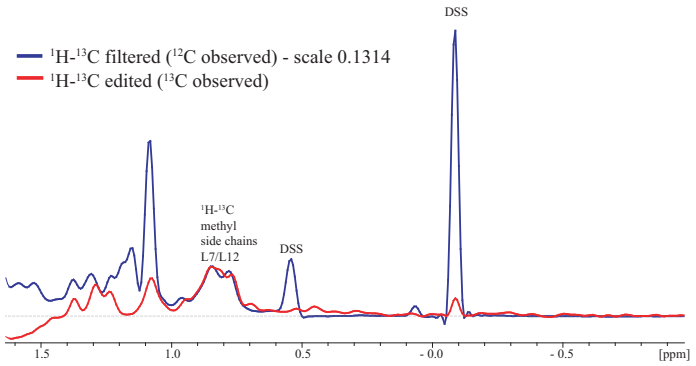


Figure 2.12: ^{1}H spectra of empty ribosomes purified from the ddFLN-dom5+110-RNC preparation. The first increment of ^{1}H-^{13}C-edited HMQC (red) and ^{1}H-^{13}C filtered HMQC (blue) via phase cycling (Section 6.2.7). The ^{1}H-^{13}C filtered spectrum (blue) is scaled such that the L7/L12 signals have the same intensity as in the ^{1}H-^{13}C-edited spectrum (red), i.e. to 13% of its initial intensity.

^{13}C/^{15}N-filtered and ^{13}C/^{15}N-edited proton spectra, which were recorded using opposite phase cycling in an HMQC-based experiment (Figure 2.12) [157]. The flexible L7/L12 stalk region which gives rise to intense NMR signals [33] is used as a probe for the labelling of the entire ribosomal complex. ^{13}C/^{15}N-filtered and ^{13}C/^{15}N-edited proton spectra (Material and Method, Section 6.2.7) of empty ribosomes isolated from the ddFLN-dom5-RNCs preparation showed that the intensities of the L7/L12 methyl signals in the ^{13}C-edited spectrum typically corresponded to 10-15% of those in the ^{13}C-filtered spectrum (Figure 2.12), suggesting that around 10-15% of the L7/L12 stalk are isotopically enriched and gives rise to signals. (Of course, this means that the entire ribosome is labelled to approximately the same extent, but these do not give rise to observable resonances.) A similar ratio for the amide signals of L7/L12 in the ^{15}N-edited spectrum compared to that in the ^{15}N-filtered spectrum was typically observed. The extent to which the ribosomes were isotopically labelled in the purified ddFLN-dom5-RNC samples was also determined based on the comparison of the intensity of the

L7/L12 resonances to that expected from a completely isotopically labelled 70S sample of the same concentration; this was found to vary from <5% to 15%.

2.3.1.4 Control spectrum: interaction of isolated NC with empty ribosomes

The ddFLN-dom5 used in this study have been chosen such that the native proteins do not interact significantly with the 70S complex [130, 179], as this could restrict the tumbling of the nascent chain by providing a second anchor point to the MDa complex (the first one being the tethering at the PTC centre), and thus broaden resonances potentially beyond detection. In order to confirm that the ribosome does not specifically interact with the ddFLN-dom5 NC, the interaction of the purified release chain with empty 70S was examined by NMR spectroscopy. The ^{1}H-^{15}N HMQC of purified ddFLN-dom5 in the presence of empty ribosomes is shown in Figure 2.13B, which shows that the observed resonances do not present extensive broadening compared to those of ddFLN-dom5 isolated nascent chain (black spectrum in Figure 2.13B), with linewidths of the amide protons varying from 20 Hz to 40Hz (Figure 2.13B). A similar experiment performed with an isolated and purified released nascent chain of ddFLN-dom5+37 which, in addition to the isolated domain contains the hydrophobic SecM sequence and the linker sequence (and thus is more representative of the ribosomally tethered sequence), is shown in Figure 2.13C. Resonances attributed to SecM residues were typically extensively broadened in the presence of the ribosome, however, the ddFLN-dom5 resonances are not broadened, suggesting that this domain does not interact with the ribosome.

In order to investigate further any potential interaction of ddFLN-dom5+37 with empty ribosomes, ^{1}H STE spectra were recorded to discriminate the translational diffusion coefficient of ribosome-bound or isolated domains. The relative intensity

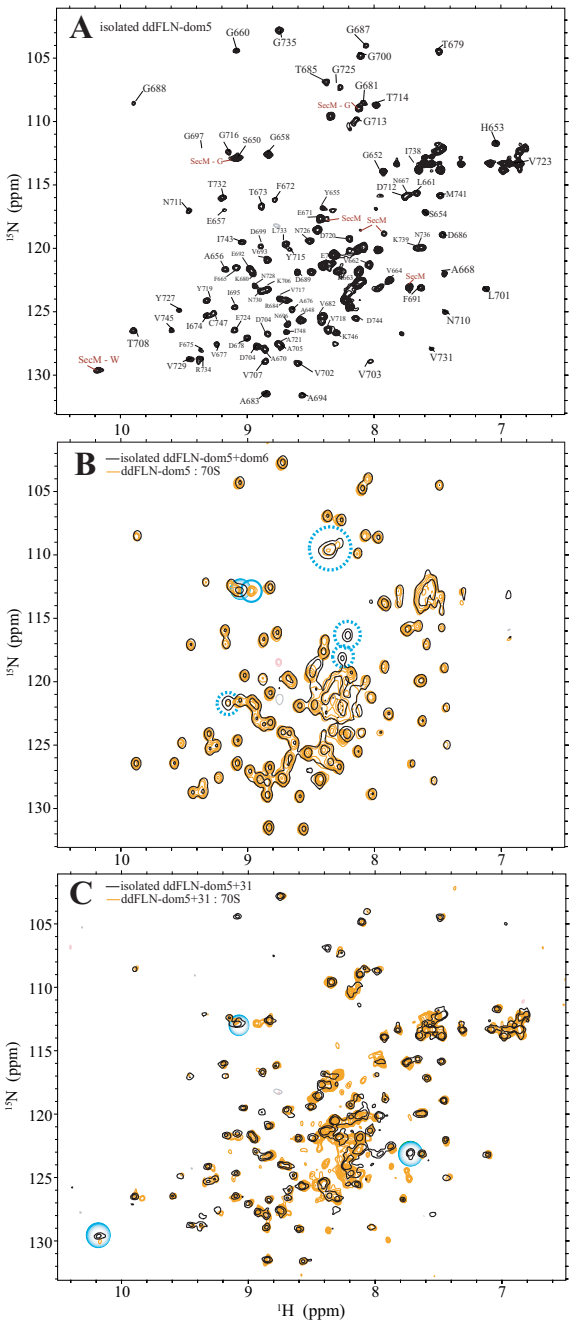


Figure 2.13: A: ^{1}H-^{15}N HSQC spectrum of the isolated released ddFLN-dom5+37-NC labelled with the assignment of the ddFLN-dom5 resonances and SecM resonances recorded at 25°C. **B:** Overlay of the ^{1}H-^{15}N SOFAST-HMQC of the isolated ddFLN-dom5+dom6 alone (10µM, black) and with empty 70S ribosomes (10µM, orange). The intensity losses for six ddFLN-dom6 resonances (dotted blue circles) can be attributed to the truncation of ddFLN-dom6 as well as the resonances shift highlighted for a ddFLN-dom6 resonances (plain blue circles) **C:** Overlay of ^{1}H-^{15}N SOFAST-HMQC of the purified released ddFLN-dom5+37-NC alone (10µM, black) and with empty 70S ribosomes (10µM, orange). The position of the discrete SecM resonances that do not overlay with ddFLN-dom5 or linker resonances are highlighted with the blue circles.

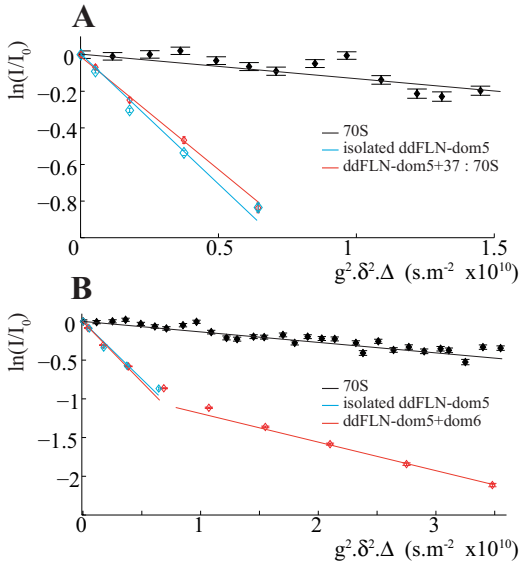


Figure 2.14: A: Plot o f the relative signal intensity in the ^{1}H STE diffusion spectra (log I/I_0 versus g^2), for isolated ddFLN-dom5 (blue), empty 70S ribosomes (black), and the intensities of the isolated ddFLN-dom5+37-NC in the presence of ribosomes observed via ^{13}C XSTE experiments (red). **B:** Same as A, with ddFLN-dom5+dom6 in red.

of the ddFLN-dom5 signals in ^{1}H STE spectra recorded with increasing gradient strengths is shown in Figure 2.14. Translational diffusion coefficients were calculated from these data using Equation 1.15. Isolated ddFLN-dom5 was found to have a translational diffusion coefficient D=1.40$\pm$0.05$\times10^{-10}$ m^2s^{-1} and the resonances from isolated ribosome complex were associated with D=1.7$\pm$0.4$\times10^{-11}$ m^2s^{-1} (Figure 2.14A). In the presence of 70S ribosomes, the diffusion coefficient of ^{15}N-^{13}C ddFLN-dom5+37 was selectively measured using a ^{1}H-^{13}C XSTE experiment [166] to filter out the resonances arising from the unlabelled 70S complexes, and gave a diffusion coefficient of 1.30$\pm$0.05$\times10^{-10}$ m^2s^{-1}, a value identical (within error) to that found for ddFLN-dom5 in isolation. This high value for the diffusion of the purified chain in the presence of empty ribosomes indicates that it does not interact significantly with the 70S complex

on the timescale of the diffusion experiment (in which the diffusion delay Δ=100ms, Material and Methods Section 6.2.4), however the broadening of the SecM resonances in the SOFAST-HMQC suggested a transient interaction (timescale $\ll$100 ms) that did not involve the folded ddFLN-dom5 domain.

For isolated ddFLN-dom5+110, which contains a longer hydrophobic ddFLN-dom6 linker sequence, the STE experiment showed a bi-exponential decay of the intensity with increasing gradient strength, indicative of heterogeneity (Figure 2.14B). Two diffusion coefficients can be measured from this curve, one extrapolated from the intensities in the spectra recorded with the four lowest gradient strengths, that is on the order of the isolated ddFLN-dom5 (D=1.40$\pm$0.05$\times10^{-10}$ m^2s^{-1}, Figure 2.14B), and one extrapolated from the intensities in the four spectra recorded with the highest gradient strengths, that is 5$\pm$0.5$\times10^{-11}$ m^2s^{-1}, a value that is 2-fold lower than that of ddFLN-dom5. This may be a result of the susceptibility to proteolysis cleavage of the ddFLN-dom6 disordered linker, and the bi-exponential decay of intensity in the STE diffusion experiment might be reflecting a truncation of the unfolded ddFLN-dom6 linker.

2.3.2 NMR spectroscopy of RNCs

A set of NMR experiments were assessed and a subset seemed appropriate for the study of ddFLN-dom5-RNC, and are presented below. First ^{1}H 1D and ^{1}H STE diffusion spectra of ddFLN-dom5-RNC are described. Then, a detailed analysis of the unfolded region of ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5-RNC is presented, which can be assigned to either unfolded ddFLN-dom5 or the disordered ddFLN-dom6 linker, depending on the linker length (ddFLN-dom5+21-RNC or ddFLN-dom5+110-RNC, respectively). Then, the ability of the ^{1}H-^{15}N SOFAST-HMQC to report on the folded state of ddFLN-dom5+110-RNC is discussed. The analysis of the dispersed

folded resonances in the ^{1}H-^{13}C HMQC and methyl-TROSY HMQC spectra of ddFLN-dom5+110-RNC are presented next. Finally, the analysis of ^{15}N XSTE and ^{13}C-edited STE-HMQC diffusion experiments are described (Section 2.3.2.8), that allows the confirmation of the attachment of the NC to the ribosome complex.

2.3.2.1 Assessment of ribosome purity via ^{1}H 1D NMR

^{1}H 1D spectra recorded of a 10 μM ddFLN-dom5-RNC sample (>80% occupancy) were closely similar to those recorded of empty 70S ribosomes under the same conditions (Figure 2.15A). The absence of obvious signals from ddFLN-dom5 in these proton spectra is surprising (Figure 2.15A). Based on signal intensities of the amide region, the ratio of the ddLFN-dom5 resonances to those of L7/L12 stalk is thus estimated to be <0.1%, although the occupancy of the ribosome complex was confirmed to be >80% as seen by an anti-His western western blot (Figure 2.10). Two copies of the L7/L12 stalks are thought to be visible for each ribosome [33], and only one nascent chain per ribosome. Therefore, the higher abundance of NMR-observable L7/L12 cannot explain the very significantly reduced sensitivity of the resonances of the nascent chain (<0.1%). It is possible that differences in dynamics may contribute to the absence of visible nascent chain resonances: in these spectra, restricted tumbling of the NC compared to that of the L7/L12 stalk may result in broader linewidths for the NC resonances, and thus lower sensitivity for those resonances. Irrespective of the exact reasons, it is clear that the observed 1D ^{1}H spectra is that of the ribosome alone, therefore, they were used to assess the purity and concentration of the ribosome complexes.

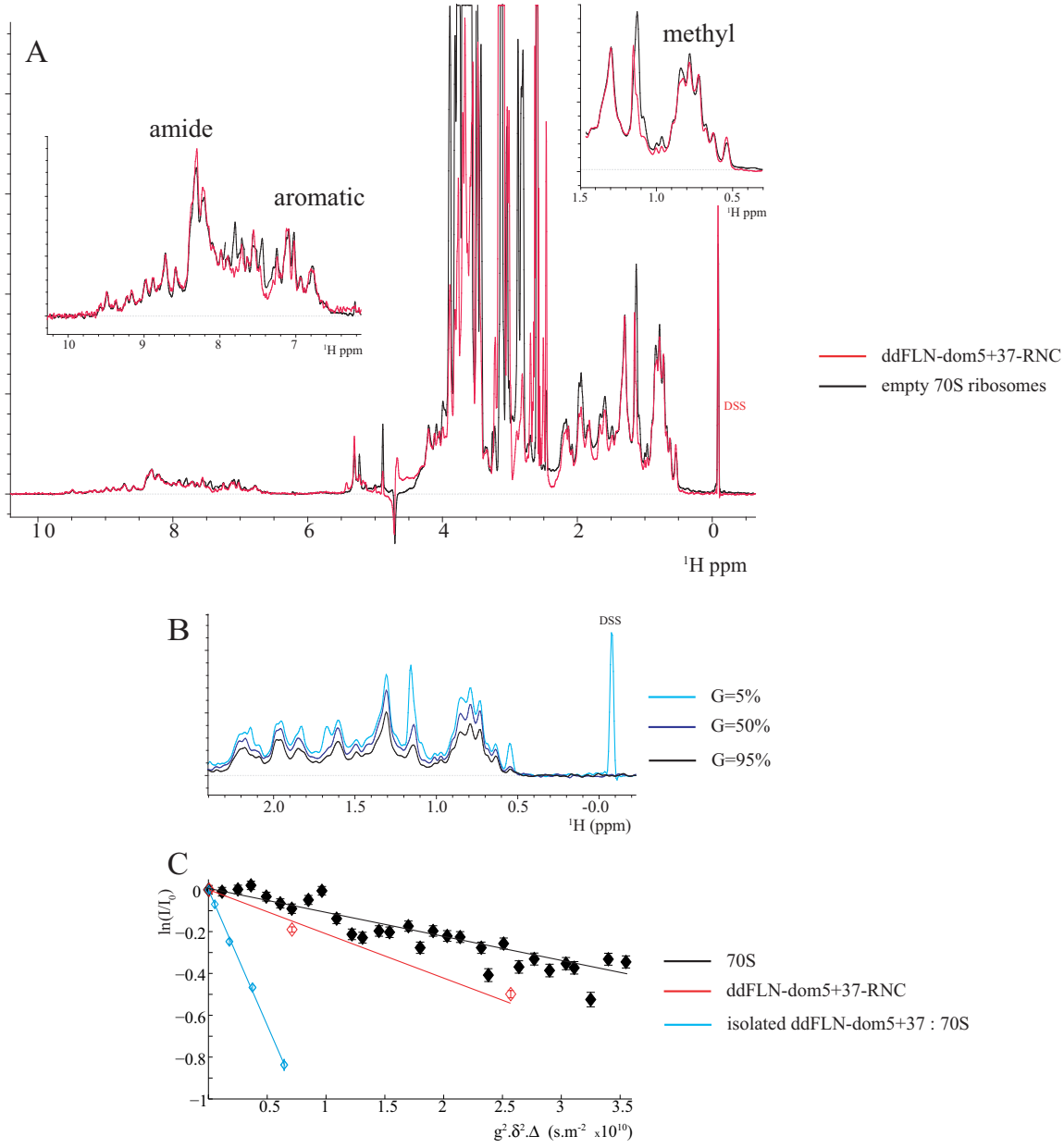


Figure 2.15: **A:** Overlay of ^{1}H 1D spectra of empty 70S ribosomes (black) with that of ddLFN-dom5+37-RNC (red), recorded at 25°C on a Bruker 700MHz spectrometer. Ribosome concentrations have been normalised and the occupancy of the RNC was >80%. **B:**^{1}H STE diffusion experiment of ddLFN-dom5+37-RNC with gradient strengths of g=2.5, 25 and 47.5 G.cm^{-1}, a diffusion delay Δ=100ms, and gradient lengths δ=4ms, in blue, purple and red respectively. **C:** Plot of the relative signal intensity of the STE diffusion spectra, for ddFLN-dom5-RNC (red), empty 70S ribosomes (black) and the isolated ddFLN-dom5+37-NC in the presence of ribosomes (blue).

2.3.2.2 *Assessment of the ribosome integrity by ^{1}H STE NMR*

The observation of resonances in the methyl and amide regions (inserts of Figure 2.15A) from ribosomal complexes which have been assigned to the flexible L7/L12 stalk region ([33] and Chapter 4) offers one possible probe for the integrity of the 70S complexes. The method employed to monitor the ribosome integrity was the ^{1}H STE diffusion NMR experiment, observing the L7/L12 methyl signals (the amide undergo exchange with H_2O) to report on the translational diffusion of the entire ribosome complex. ^{1}H STE diffusion spectra of ddFLN-dom5-RNC are shown in Figure 2.15B. The observed decay of intensity in the STE spectra with increasing gradient strength (Figure 2.15C) indicates that the observed signals arise from slow diffusing species, with a calculated translational diffusion coefficient of $2.0\pm0.5\times10^{-11}$ m^2s^{-1} (Figure 2.15B). This value is equal, within the experimental uncertainty, to the value obtained for the 70S complex (D= $1.7\pm0.4\times10^{-11}$ m^2s^{-1}, Figure 2.15B in line with previous measurements from light-scattering spectroscopy [172]), and indicates that the purified RNCs are intact ribosome complexes.

2.3.2.3 *Analysis of the unfolded region of ^{1}H-^{15}N SOFAST-HMQC spectra of RNCs*

A ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+21-RNC is shown in Figure 2.16A & B. The spectrum shows a cluster of signals in the central region of the spectrum (between 7.8-8.5 in the ^{1}H dimension). In the case of longer linker length, well-dispersed resonances are also observed with very low intensity and are analysed in details in the next Section 2.3.2.4. The analysis of the central region of the ^{1}H-^{15}N SOFAST-HMQC spectrum of RNC is described first.

The cluster of unfolded signals in the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+21-RNC overlays reasonably well with the ^{1}H-^{15}N HSQC spectrum of 8M urea

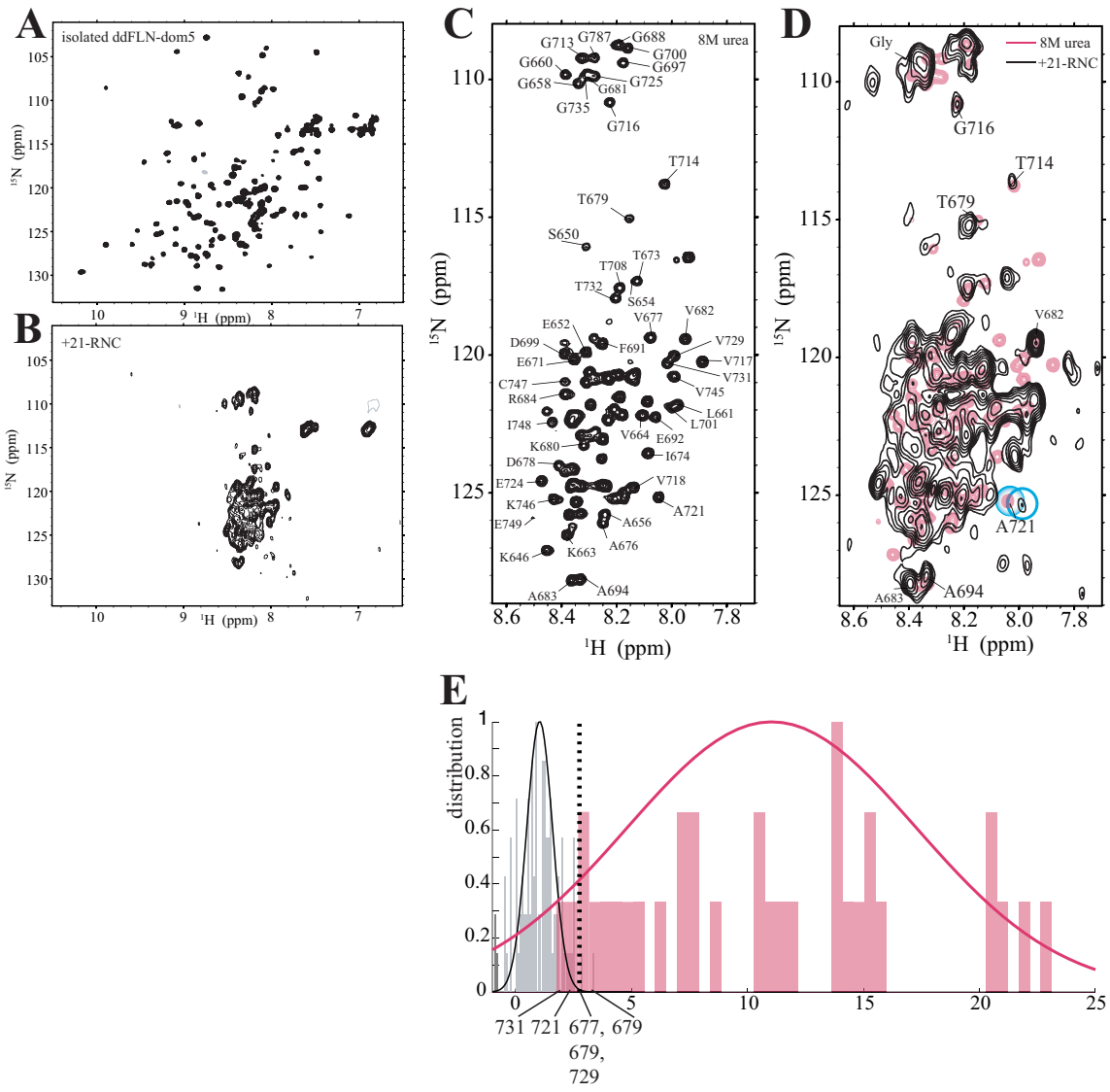


Figure 2.16: **A:** ^{1}H-^{15}N HSQC spectra of the isolated ddFLN-dom5+21 as a released NC. **B:** ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+21-RNC. **C:** ^{1}H-^{15}N HSQC spectra of the assigned ddFLN-dom5 in 8M urea [185]. **D:** ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+21-RNC (black), overlaid with spectra of the urea unfolded ddFLN-dom5 (magenta). The depicted assignment is extrapolated from the 8M urea assignment of the isolated ddFLN-dom5. The blue circled resonances highlight possible chemical shift changes between the urea denaturated states and the RNC resonance for A721. **E:** Distribution of the signal intensity for the 32 resonances that do not overlap in spectrum B. The signal intensity is defined as the maximum within a 3×3 ROI centred at the urea assigned position. The nine-variate distribution of the noise is shown in grey. Except for few resonances mentioned in the main text (V677, T679, T685, A721, V729, V731), all resonance intensities are significantly higher than the noise distribution. The black line shows the noise threshold as defined in Section 2.2, above which the signals have less than 5% chance of being a result of random noise. All spectra were recorded at 25°C.

denaturated ddFLN-dom5 (shown in Figure 2.16C with assignments [185]). 32 of the observed and resolved ddFLN-dom5+21-RNC resonances could be attributed to unfolded ddFLN-dom5 based on the assignment (Figure 2.16D), with most peaks within 0.05ppm in the ^{1}H dimension and 0.5ppm in the ^{15}N dimension. These resonances can be probes to report on the unfolded states of the RNC.

The RNC spectrum (Figure 2.16D) shows a significant broadening of the disordered ddFLN-dom5 resonances which results in a significantly greater overlap in the central region of the RNC spectrum compared to the urea-denatured spectrum of the isolated protein. The distribution of the *z*-score of the intensities of the 32 unfolded ddFLN-dom5 resonances is shown in Figure 2.16E (using the methodology described in Section 2.2, where the signal intensity is the maximum of a 3×3 ROI centred at the resonances in the urea denatured ddFLN-dom5 spectra), and most of the resonances have a statistically significant intensity. Figure 2.17 shows the *z*-score for each of the residues which had an unambiguous assignment in the ddFLN-dom5+21-RNC spectrum (plot A), or in the 8M urea denatured spectrum of the isolated protein (plot B). The intensities of the individual residues do not show a clear correlation along the sequence, and the two plots show broadly the same variation along the sequence. The attachment of the RNC does not appear to result in significant broadening of the C-terminal residues compared to the N-terminal ones, in line with the high flexibility observed along unfolded chains [171].

However, some unfolded resonances in the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+21-RNC, such as those corresponding to V677, T679, T685, A721, V729, V731 appear to have a particularly low intensity (Figure 2.16E & 2.17C), but might also have altered chemical-shift values compared to those in the urea-denaturated spectra (Figure 2.16D). This difference in chemical shift between the unfolded ddFLN-dom5 and the RNC spectra might be explained by a difference in the conformational-sampling

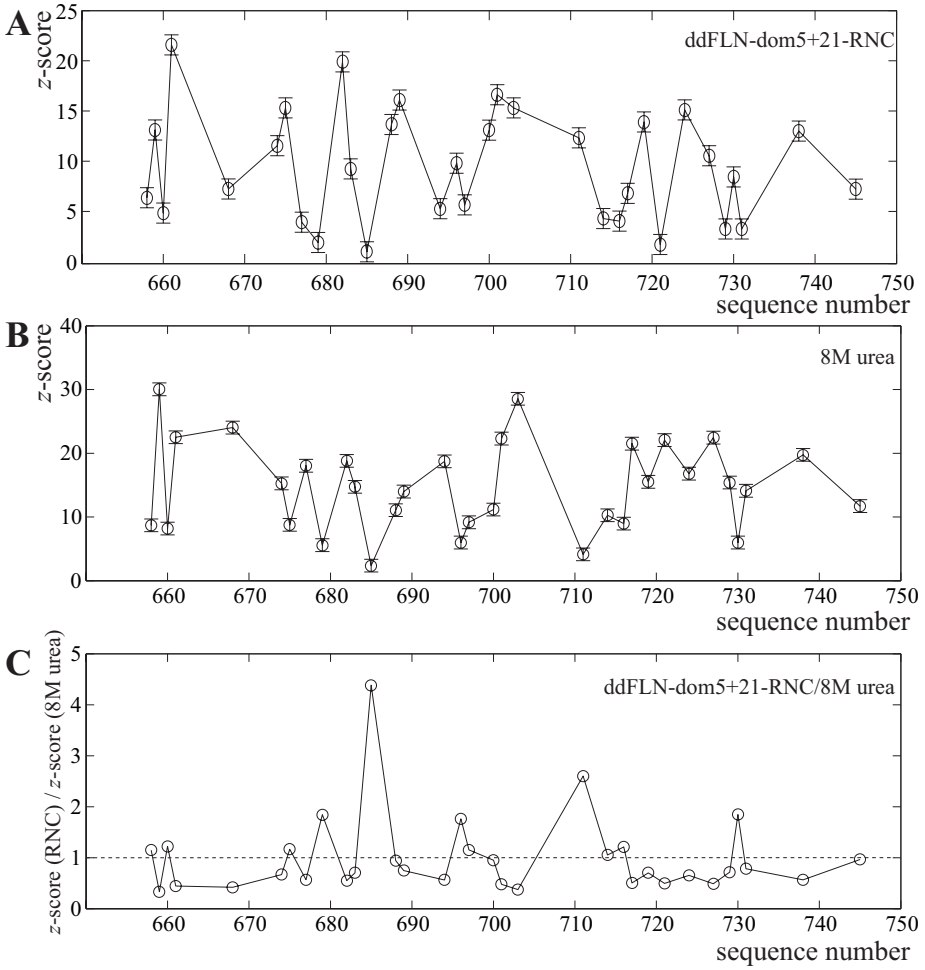


Figure 2.17: A: z-score of the unfolded resonances in ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+21-RNC plotted against the residue number. **B:** z-score of the unfolded resonances ^{1}H-^{15}N HSQC spectrum of the 8M urea denatured ddFLN-dom5 plotted against the residue number. **C:** Plot of the intensities of ddFLN-dom5+21-RNC resonances divided by those of the 8M urea denatured ddFLN-dom5, normalised so that the mean of the ratio equals one. The variation of the ratio varies from 0.5 to 1.5 times the mean, except for residues T685, N711 and N730.

the domains can perform in urea compared to when attached to the ribosome via its C-terminal end.

Resonances corresponding to an unfolded chain are also observed in spectra of ddFLN-dom5-RNCs with long linker lengths. Figure 2.18A shows the disordered region of the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC. The large cluster of overlapping resonances in the centre of the spectrum cannot be assigned to disordered ddFLN-dom5: discrete ddFLN-dom5 resonances were not observable, with their intensity being lower than the noise threshold (Figure 2.18B & D). Instead, the RNC resonances seem to overlay well with the spectrum of truncated ddFLN-dom6 linker (Figure 2.18C). The distribution of the z-score of ddFLN-dom6 resonances intensities (Figure 2.18D, cyan) shows that the sensitivity of the disordered ddFLN-dom6 signals is sufficiently higher than the noise ($SN = 2$ to 8) to allow good accuracy for determination of resonance frequency and sensitivity. This observation showed that the overlapping central region of the ^{1}H-^{15}N SOFAST-HMQC spectra of RNCs provides distinct probes for the unfolded ddFLN-dom6 linker and for unfolded ddFLN-dom5.

We have shown here that the analysis of a set of discrete resonances attributed to unfolded ddFLN-dom5, that do not overlay with resonances of the ddFLN-dom6 linker, nor to background L7 resonances, may be used to investigate the disordered state of ddFLN-dom5-RNCs. Further results using this approach will be presented in Chapter 3.

2.3.2.4 Analysis of the "well-dispersed" resonances in ^{1}H-^{15}N SOFAST-HMQC spectra of RNCs.

In addition to intense resonances from disordered linker regions, discussed above, the ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+110-RNC exhibited dispersed signals of low intensity (Figure 2.19A & C). These dispersed resonances overlay well with folded

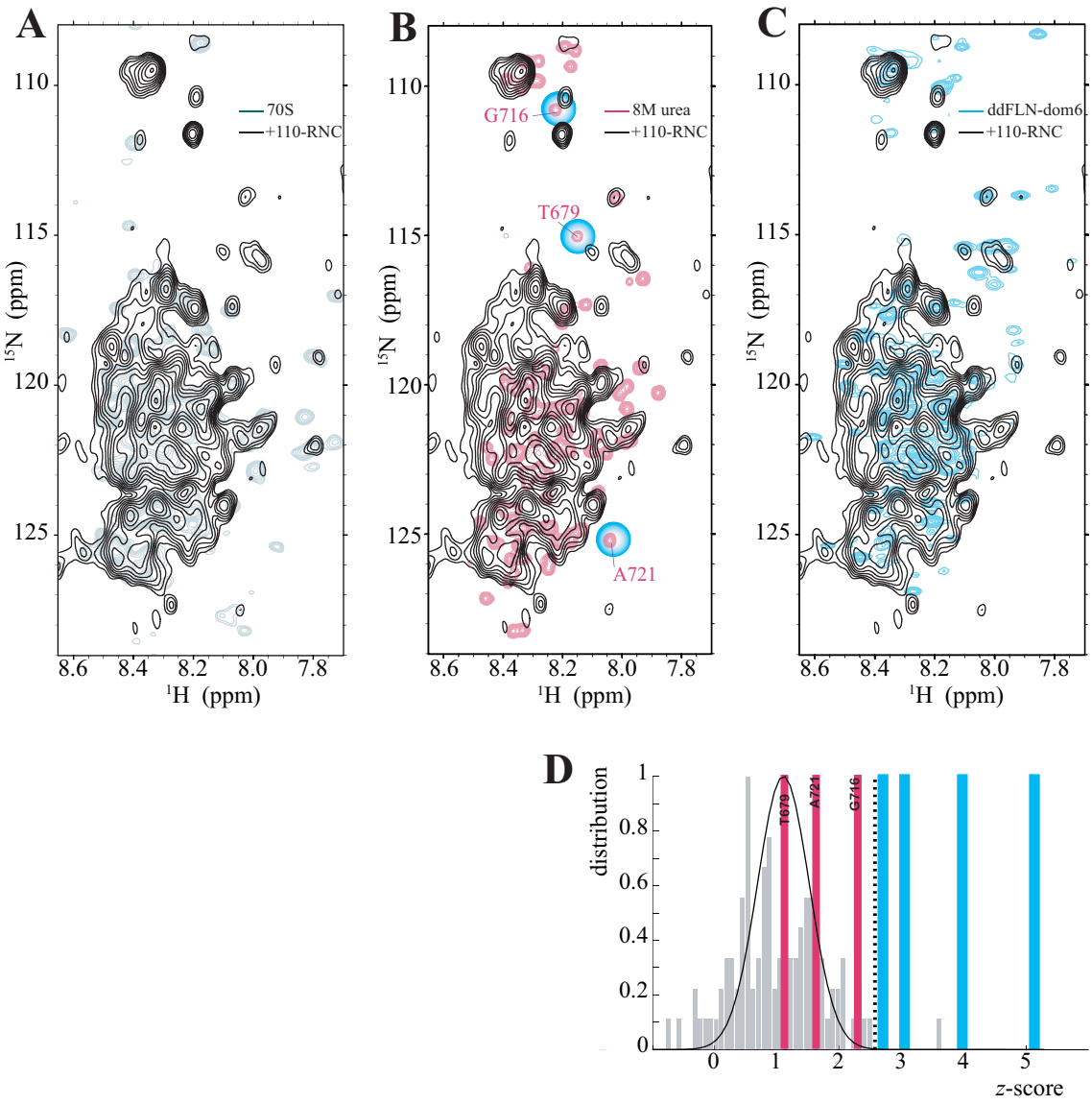


Figure 2.18: A: ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC (black). The ^{1}H-^{15}N SOFAST-HMQC spectrum of labelled 70S ribosomes is shown in the background in dark green. **B:** Overlay of ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC (black) and isolated 8M urea unfolded ddFLN-dom5 (magenta) [185]. The blue spheres indicate the resonances that do not overlay with resonances arising from L7/L12 or ddFLN-dom6 linker. **C:** Overlay of the ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+110-RNC (black) and purified released ddFLN-dom5+110 construct (cyan). **D:** Distribution of the *z*-score of the intensities of unfolded ddFLN-dom5 resonances in magenta and ddFLN-dom6 resonances in cyan in the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC. The 5% threshold as defined in Section 2.2 is shown by the dotted black line.

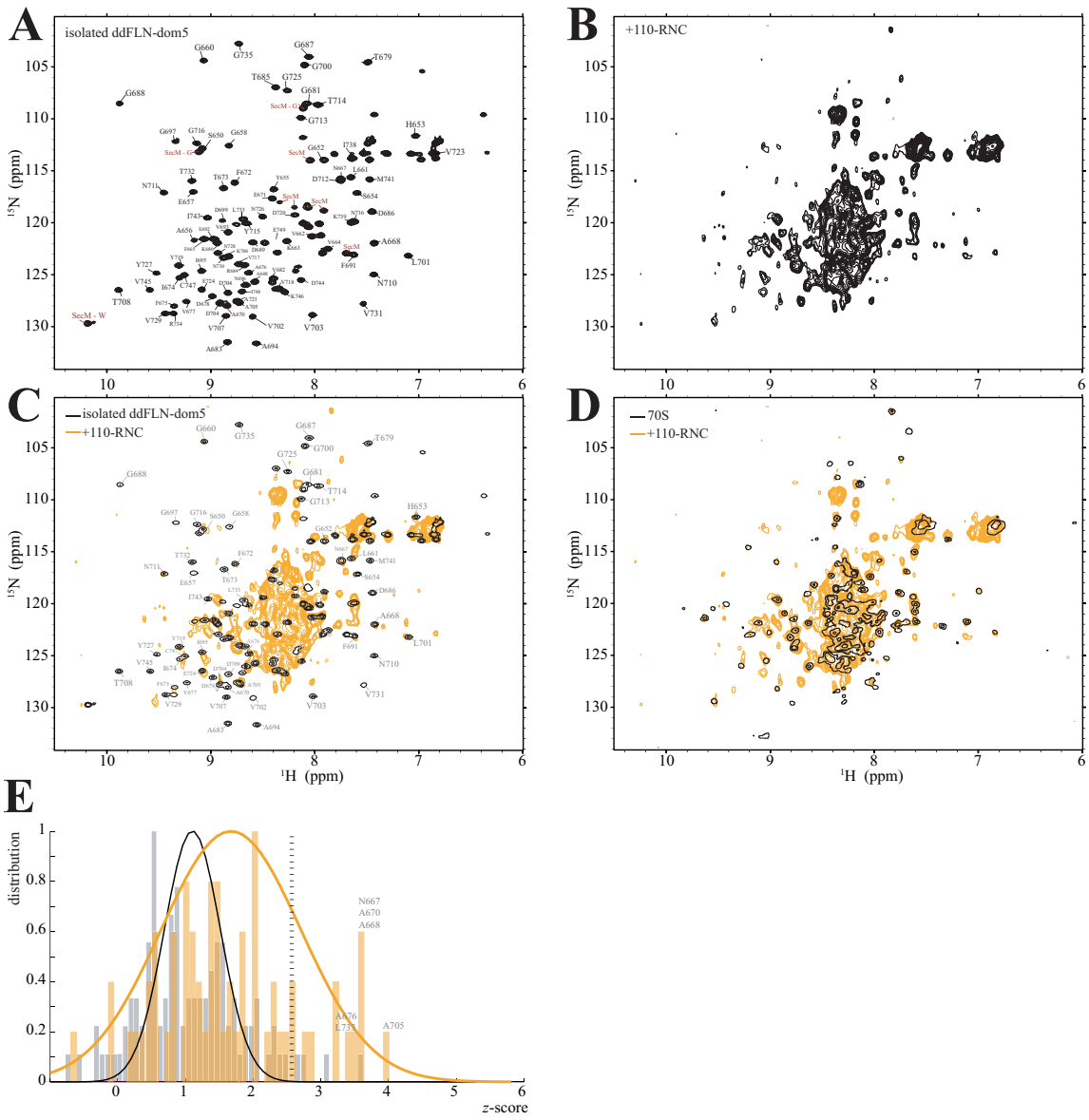


Figure 2.19: A: ^{1}H-^{15}N HSQC spectrum of isolated ddFLN-dom5+21 as a released NC, with peak assignments from [185]. The SecM resonances are also indicated with dark red labels. **B:** ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC. **C:** Overlay of ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+110-RNC (orange) and that of purified released ddFLN-dom5+21(black). **D:** Overlay of ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+110-RNC (orange), and that of uniformly ^{15}N labelled 70S ribosomes (black). **E:** Intensity distribution of the 62 resonances for folded ddFLN-dom5 in the RNC spectrum C, with the noise distribution in grey. The resonances with the highest sensitivity are labelled above the intensity bar.

ddFLN-dom5 resonances by comparison with a spectrum of isolated ddFLN-dom5+21 (Figure 2.19B). In the RNC spectrum, 9 well-dispersed ddFLN-dom5 resonances out of 64 resonances analysed (labelled in Figure 2.19C) had an intensity that was significantly higher than the noise threshold as defined in Section 2.2 (Figure 2.19E). These resonances hinted at the presence of folded ddFLN-dom5-RNC. Moreover, the centre of distribution of the *z*-score of the 64 dispersed ddFLN-dom5 resonances intensity is shifted compared to the noise distribution (Figure 2.19E), indicating resonances corresponding to folded ddFLN-dom5 within the RNC spectrum are present but not observable, as it was the case in the distribution observed for a model spectrum of isolated ddFLN-dom5 recorded with very low number of scans, and with very low sensitivity, shown in Figure 2.6 (NS=16). Together, this indicates that the folded state of the RNC is populated at this linker length, but individual "folded" resonances may not be observed. As has been described in the case of the model spectra, at low SN, the distribution of resonance intensity is only dependent on the noise, not on their intrinsic linewidth, and observation of a particular resonance was found to be highly sample-dependent – i.e. no significance should be attached to the observation of particular individual resonances. This low intensity significantly hinders more detailed analysis of the folded conformation of ddFLN-dom5-RNCs via ^{1}H-^{15}N SOFAST-HMQC (e.g. diffusion measurements or linewidth analysis).

This analysis of the ^{1}H-^{15}N SOFAST-HMQC spectra of RNCs suggests that, while the amide resonances were found to be reliable probes of the disordered regions of RNC, allowing the discrimination between unfolded ddFLN-dom5 and linker ddFLN-dom6, alternative methods are required for the detailed analysis of weak folded resonances. This is pursued in the next Section using other NMR probes, and in particular the ^{13}C-based observations of residue sidechains.

2.3.2.5 Analysis of the well-dispersed resonances in ^{13}C-^{1}H HMQC of RNCs.

Resonances from sidechains are often associated with increased sensitivity relative to the backbone resonances due to their low order parameter as compared to the backbone [186], and further in the case of methyl groups (Leu, Val, Ile, Ala, Thr and Met), for which the three equivalent protons also provide enhanced sensitivity [187]. ^{1}H-^{13}C HMQC spectra of ddFLN-dom5+110-RNC (Figure 2.20B) show several dispersed methyl resonances that overlay closely with native-ddFLN-dom5 resonances (Figure 2.20A and 2.20D orange spectrum). The intensity distribution of 16 high-field shifted methyl resonances (<0.6 ^{1}H ppm) of ddFLN-dom5 in ddLFN-dom5+110-RNC is shown in Figure 2.20E. As shown by the insert of Figure 2.20D, five resonances were found to overlap in RNC spectra (resonances I695 γ_2 and a L7 resonance, V729 γ_1 and V664 γ_2, V729 γ_2 and A670 β). Compared to the spectrum of the isolated, released chain, the extent to which those resonances are overlapped has increased, and this relates to their broader linewidth induced by the ribosome tethering. These five resonances are therefore analysed separately from the other 11 resonances, and were found to have statistically significant intensities. The intensity distribution of the remaining resonances was found to overlap with the noise distribution, and only two of these discrete resonances have intensities that were found to be higher than the noise threshold. As was the case for the ^{1}H-^{15}N SOFAST-HMQC resonances (Section 2.3.2.4), the discrete methyl resonances in the ^{1}H-^{13}C HMQC have a distribution that is shifted from the noise, indicating that the folded state is being populated, but individual resonances remain hard to observe. In order to increase the sensitivity of the dispersed methyl resonances, the methyl-TROSY HMQC method was tested with high level of deuteration and selective labelling of the Ile $\delta 1$.

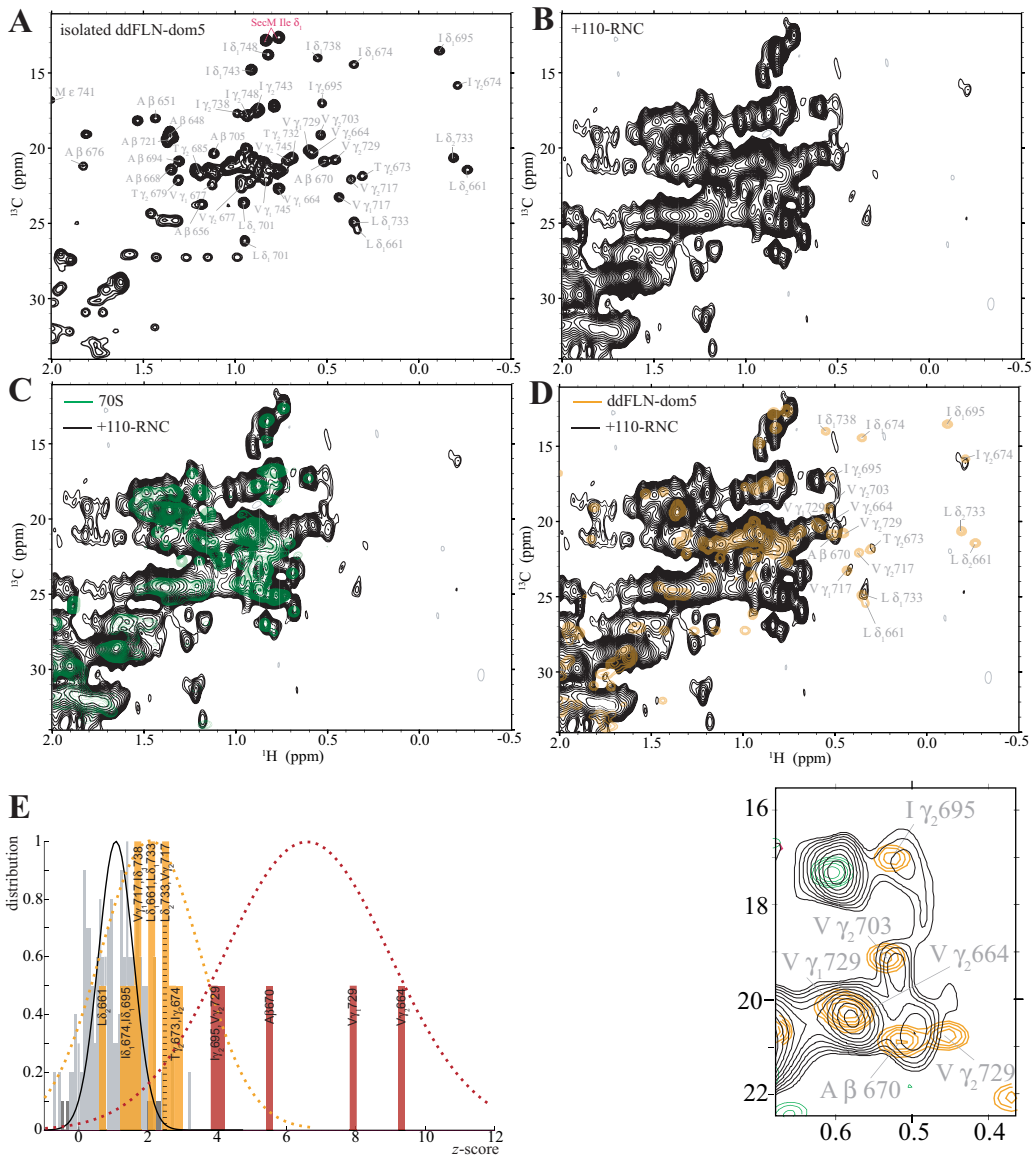


Figure 2.20: A: ^{1}H-^{13}C HMQC spectrum of isolated ddFLN-dom5+21 purified as a released NC, with the assignment of the ddFLN-dom5 methyl resonances from [185]. Two isoleucines assigned to SecM are also shown, labelled in red. **B:** ^{1}H-^{13}C HMQC spectrum of ddFLN-dom5+110-RNC. **D:** Overlay of ^{1}H-^{13}C HMQC spectra of ddFLN-dom5+110-RNC (black) and that of ^{13}C labelled 70S ribosome (green)(the resonances in the latter arise from L7L12 stalk). **D:** Overlay of ^{1}H-^{13}C HMQC spectra of ddFLN-dom5+110-RNC (black) and isolated ddFLN-dom5+110-NC (magenta). An insert of the spectra is shown below, highlighting overlapped resonances in the RNC spectrum. **E:** Intensity distribution of the 16 methyl resonances assigned in B in the ^{1}H-^{13}C HMQC spectra of ddFLN-dom5+110-RNC. The intensities of the discrete signals are shown in orange, and the intensities of the overlapped signals are shown in dark red. The noise distribution is shown in grey, and the noise threshold as defined in Section 2.2 is shown with the dotted black line.

2.3.2.6 Analysis of the dispersed resonances in ^{1}H-^{13}C methyl-TROSY HMQC of RNCs

The combined use of deuteration and selective $^{13}CH_3$ labelling currently presents optimal conditions to perform methyl-TROSY NMR spectroscopy in the ^{1}H-^{13}C HMQC acquisition (Section 1.3.2), reducing T_2 relaxation resulting in narrower, more intense resonances. U-[^{12}C,^{2} H], Ileδ1-[$^{13}C^{1}H_3$] labelled ddFLN-dom5-RNC samples were produced, and the methyl-TROSY HMQC of ddFLN-dom5+110-RNC is shown in Figure 2.21. Signals arising from the high-field shifted (i.e. folded) isoleucine side-chains of ddFLN-dom5 were all observed with a statistically significant intensity, and a SN ranging from 4.3 to 12.5 (Figure 2.21B & D). Compared to SN of *ca.* 1-3 in previous spectra, and this increased intensity allows further characterisation of the folded state of the RNC.

The observation of these intense resonances allows determining their proton linewidths ($\Delta\nu_{^{1}H}$): the ^{1}H linewidths of the deuterated ddFLN-dom5-RNC resonances were measured by fitting the signals to a lorentzian lineshape, whilst no window function was applied in the ^{1}H dimension (Figure 2.21), and the uncertainty in linewidth was estimated from equation 1.12 [157]. ^{1}H linewidths of the RNCs ranged from 14-28 Hz with uncertainty of 1-4 Hz. These linewidths were significantly narrower than the linewidths expected for resonances arising from the ribosomal particle ($\gg$100 Hz, Figure 1.12), indicating that the nascent chain has a high degree of motional freedom, and an internal tumbling time $\tau_{c,NC}$ (Section 1.3.2) that is significantly shorter than that of the entire ribosome complex ($\tau_{c,ribo} \sim$ 2500 ns, [129]). These linewidths of the RNC resonances were however significantly greater than these of the corresponding isolated nascent chain resonances (*ca.* $\leq$ 5-6 Hz), which could not be measured accurately in this case because the ^{1}H acquisition time, otherwise limited by the probe capacity and the ^{1}H-^{13}C decoupling pulse, was too short to avoid truncation in the FID. These results indicate that the NC has restricted mobility due to the attachment to the ribosomal complex.

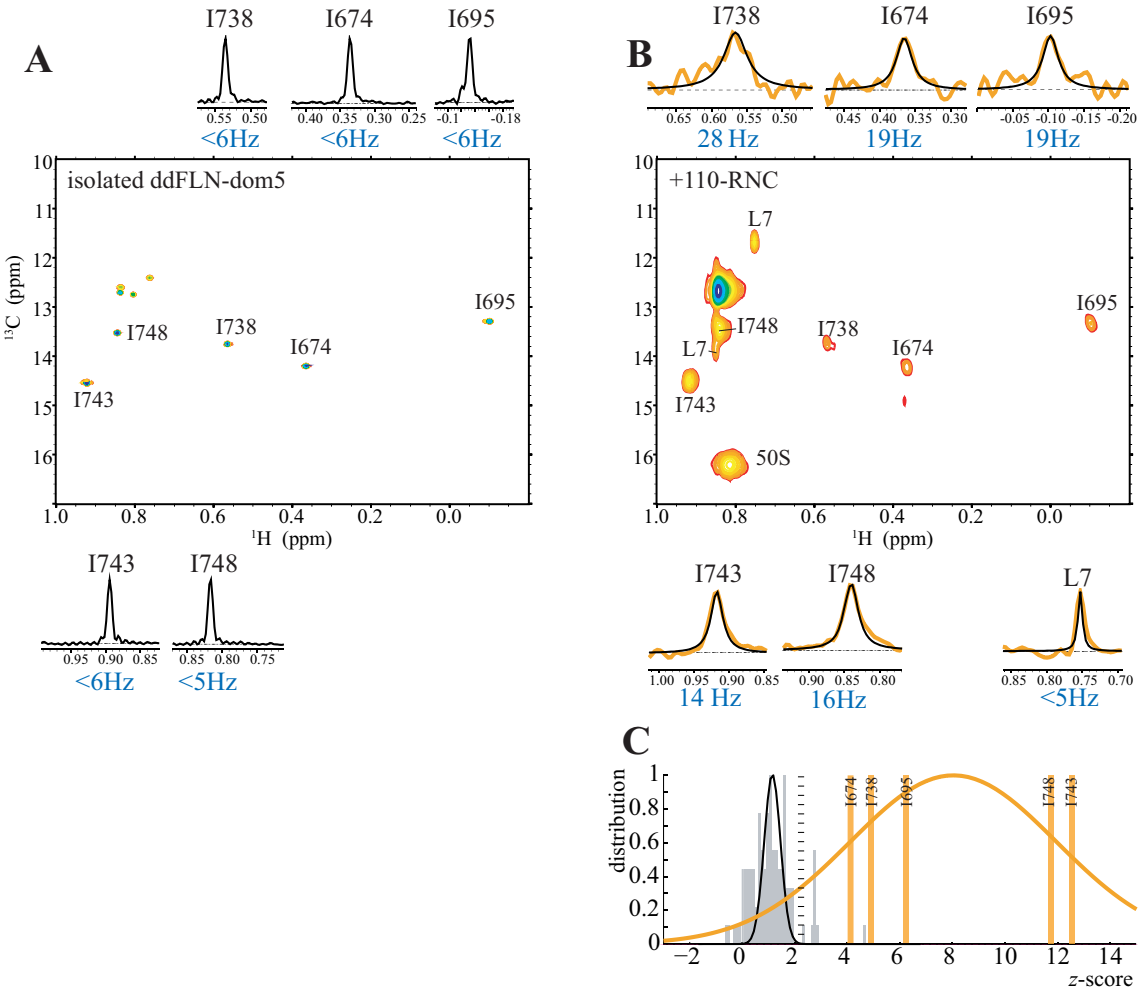


Figure 2.21: ^{1}H-^{13}C HMQC spectra of: **A:** purified released U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+47-NC, 100 μM recorded for 1 hour, and **B:** U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+110-RNC, 10 μM recorded for 30 hours. The corresponding proton slice of the signals are shown next to the assigned resonances, with their SN ratio in green and ^{1}H linewidth in black. The intensity distribution for the five ddFLN-dom5 Isoleucine resonances are shown in **D**. The signal intensity is the maximum in a 3×3 ROI, and the 5% threshold for the noise is shown with the black dotted line.

The ^{13}C acquisition time was kept deliberately short ($t_{1\ max}$=20 ms) to allow a maximise signal accumulation, but consequently this did not allow the accurate determination of the ^{13}C linewidths.

Thus, we have shown in this section that the analysis of the resonances observed in the methyl-TROSY HMQC (and in particular the increased sensitivity allowing measurement of the ^{1}H linewidths) of ddFLN-dom5-RNCs provides a wealth of information on the folding states of the RNC, and this strategy was utilized for the detailed investigation of the folded states of ddFLN-dom5-RNC with varying ddFLN-dom6 linker lengths presented in Chapter 3.

2.3.2.7 Analysis of the aromatic resonances of RNCs

Aromatic resonances are also good reporters of the formation of globular structure because of their hydrophobicity and their propensity to cluster within the hydrophobic core of simple protein folds [188]. Moreover, there is no aromatic residue in the C-terminal end of the L7 stalk, which simplifies the analysis of the RNC spectrum since the intense resonance of L7 in the case of even small amounts of mixed labelling in the *in vivo* derived RNC would not be present in this region. Figure 2.22A shows the ^{1}H-^{13}C aromatic spectrum of isolated released ddFLN-dom5+21, in which 10 "dispersed" resonances can be assigned to the Tyr and the Phe residues within ddFLN-dom5 [185]. The resonances corresponding to a Trp residue also observed (labelled in red in Figure 2.22A) can be readily assigned to the single Trp in the sequence, present in the SecM stalling sequence. The ^{1}H-^{13}C aromatic spectrum of ddFLN-dom5+110-RNC is shown in Figure 2.22B. The spectrum shows a set of overlapping residues that have chemical shifts close to the random-coil chemical shift of Phe and Tyr (^{13}C: 127-132 ppm / ^{1}H: 6.9-7.3 ppm & ^{13}C:115-120 ppm / ^{1}H: 6.6-6.9 ppm, respectively) which overlap with

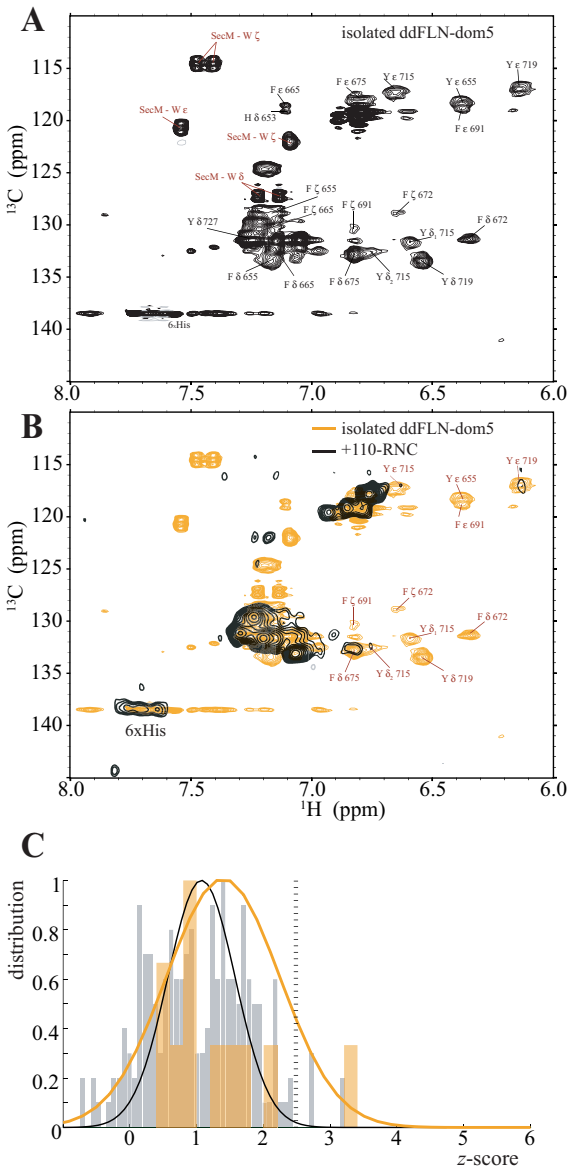


Figure 2.22: **A:** ^{1}H-^{13}C HMQC spectrum of isolated ddFLN-dom5+21 released NC, with the ddFLN-dom5 peak assignment from [185] and the SecM Trp resonances in red. The Trp chemical shift are as follows from the BioMagResBank: C$\delta_{1,2}$:127 ppm, Hδ_1:7.2 ppm, Cϵ_3:120 ppm, Hϵ_3:7.5 ppm, Cζ_2:114.5 ppm, Hζ_2:7.4 ppm, Cζ_3:122 ppm, Hζ_3:7.0 ppm. **B:** ^{1}H-^{13}C HMQC spectrum of ddFLN-dom5+110-RNC (black) overlaid with that of the purified released chain in magenta (same as in A). **C:** Signal intensity distribution for the ddFLN-dom5 resonances in the RNC spectra shown in B. The signal intensity is the maximum of a 3×3 ROI centred at the resonances frequency in the released assignment, and the noise is shown in grey.

resonances from the isolated domain. Resonances assigned to folded ddFLN-dom5 aromatic residues were not observable: their intensities distribution overlaps with the noise distribution (Figure 2.22C). This is in line with the lower intensity of aromatic signals than the methyl resonances where even in the latter case some of the individual resonances were at the noise threshold (Section 2.3.2.5). The ratio of intensities of the "unfolded" aromatic resonances to that of the dispersed "folded" ones (which are at the noise level) was found to be 40 ± 10. This differential intensity was increased compared to the spectrum of the isolated domain, where the unfolded resonances have an intensity that is 10 ± 5 times higher than the dispersed ones. This might reflect the broadening of the folded resonances due to the restricted motion imposed by the ribosome tethering, while the aromatic sidechains of unfolded linker residues have a higher degree of freedom.

Notably, the absence of visible Trp resonances indicates that the C-terminal end of the nascent chain is not observable, probably due to the restricted tumbling due to the likely occlusion within the ribosomal tunnel. Indeed, the Trp within the SecM sequence is known to make stable contacts within the constriction of the ribosome exit tunnel (Figure 2.1) [40], therefore it is expected to have a high order parameter and also to tumble with the τ_c of the ribosomal complex. In addition, despite the restricted tumbling of the C-terminal end of the RNC, a set of His sidechain resonances arising from the $6\times$His tag are observable, reporting on the faster tumbling of the N-terminal end of the RNC.

Together, the intensity of the aromatic resonances are typically too weak to probe for the presence of folded ddFLN-dom5-RNC, however, they appear to be a reliable means of assessing the integrity of the RNC by reporting on the relative sensitivity of the His resonances at the N-terminus versus the Trp resonances at the C-terminus.

2.3.2.8 Attachment of RNCs as followed by heteronuclear-edited STE diffusion

The direct monitoring of the ribosomal attachment of the species giving rise to the observed resonances can be performed by measuring the translational diffusion coefficient associated with the RNC resonances and comparing this to the translational diffusion coefficient of ribosome-bound L7/L12 (D = $1.7\pm0.4\times10^{-11}$ m^2s^{-1}, Section 2.3.2.1) or that of released ddLFN-dom5 in the presence of one molar equivalence of ribosome (D = $1.30\pm0.05\times10^{-10}$ m^2s^{-1}, Section 2.3.1.4). With values for the parameters of the STE experiments appropriate for the determination of the diffusion coefficient of the ribosome (i.e. a diffusion delay, Δ=100 ms, gradient lengths δ=4 ms, Section 1.3.4), the intensity of signals arising from released NCs is expected to decay more than 95% from the spectrum recorded at low gradient strength (2.5 $G.cm^{-1}$) to that recorded at high gradient strength (47.5 $G.cm^{-1}$). A signal arising from ribosome-bound species, however, is expected to have an attenuation of 30%. This indicates that the diffusion coefficients of the ribosome and that of the released proteins have values that are significantly different to assess ribosome-attachment based on the STE diffusion measurement on the NC resonances.

Heteronuclear STE experiments [166, 167] were used to determine the translational diffusion coefficient of the NC specifically, excluding most of the resonances from the L7/L12 stalk. As mentioned in Section 2.3.2.3, the resonances characteristic for the unfolded regions of the RNC are observed in the ^{1}H-^{15}N SOFAST-HMQC spectra, and this suggests that the ^{15}N XSTE diffusion measurement [166] can report on the attachment of the NC giving rise to these unfolded resonances. On the other hand, the ^{1}H-^{13}C-edited STE diffusion experiment [167] can report on the attachment of the NC giving rise to folded ddFLN-dom5 resonances observed in the ^{1}H-^{13}C HMQC spectra.

Figure 2.23 shows the ^{15}N XSTE and ^{13}C-edited STE-HMQC spectra of ddFLN-

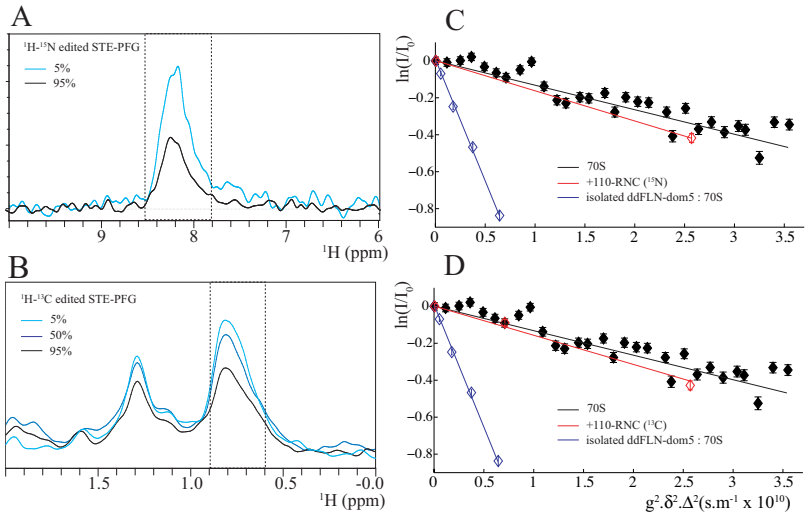


Figure 2.23: **A:** ^{1}H-^{15}N edited XSTE spectra of ddFLN-dom5+110-RNC recorded with g=2.5 (cyan), g=47.5 G.cm^{-1} (black). **B:** ^{13}C-edited STE-HMQC spectra of ddFLN-dom5+110-RNC recorded with g=2.5 (cyan), g=25 (blue), g=47.5 G.cm^{-1} (black). **C:** Plot of the relative signal integral delimited by the dotted lines in A in the different spectra of the diffusion experiments in a log scale, versus the square of the gradient strength, in the ^{1}H-^{15}N edited XSTE spectra of ddFLN-dom5+110-RNC in red and for the purified released in the presence of ribosome in blue (Section 2.3.1.4). The intensity of the ^{1}H STE of empty 70S is shown in black. **D:** Same as C for the intensity of the ^{13}C-edited STE-HMQC spectra.

dom5+110-RNC. The ^{15}N XSTE spectra of ddFLN-dom5+110-RNC showed a signal characteristic of unfolded resonances with a low ^{1}H chemical shift dispersion (Figure 2.23A). This signal showed a 30% decay from the ^{15}N XSTE recorded with a low gradient strength (2.5 G.cm^{-1}) to that recorded with a high gradient strength (47.5 G.cm^{-1}). The integral of the signal at 7.8-8.5 ppm was analysed to determine the diffusion coefficient of the unfolded RNC (ranging from 2.3 to 1.7$\pm$0.5$\times$10^{-11} m^{2}s^{-1}, Figure 2.23C), which showed that those signals arise from a ribosome-bound NC.

The ^{13}C-edited STE-HMQC spectra recorded at gradient strengths g=2.5, 25 and 47.5 G.cm^{-1} are shown in Figure 2.23B, which also showed a similar attenuation to

that observed in the ^{15}N XSTE spectra (Figure 2.23D). If we assume that the extent of labelling of the ribosomal-bound L7/L12 within RNC samples is 10%, as typically found to be visible in the heteronuclear spectra of RNC, the signals within the ^{13}C-edited STE-HMQC spectra thus represents the sum of the RNC signal plus those from L7/L12. The attribution of the resonances in the ^{13}C-edited STE-HMQC diffusion spectra was done by analysing the ^{1}H-^{13}C HMQC spectrum, and the intensity of the signals arising from 10% labelled L7/L12 is compared to the intensity of the signals arising from the NC which was dependent on the length of ddFLN-dom6 linker (analysed in details in Chapter 3). The L7/L12 contribution to the signals observed in the ^{13}C-edited STE-HMQC spectra was therefore found to vary from 20% to 70%, depending on the ddFLN-dom6 linker length. In the case of intermediate linker lengths (ddFLN-dom5+37 & ddFLN-dom5+47), the signals in the ^{13}C-edited STE-HMQC spectra arise mainly from L7/L12 resonances, suggesting that the diffusion experiments report mainly on the attachment of the stalk region. In cases of shorter and longer RNC constructs ($\leq$ ddFLN-dom5+31 & $\geq$ ddFLN-dom5+67) the intensity of either the disordered ddFLN-dom5 or the ddFLN-dom6 linker (Section 2.3.2.3) were high enough to allow attribution of the observed signals in the ^{13}C-edited STE-HMQC spectra to the NC, thus allowing reporting on the attachment of the NC to the ribosome complex.

The absence of high field shifted methyl resonances in the ^{1}H-^{13}C spectra of 70S ribosomes offers the exciting potential of monitoring the translational diffusion coefficient for the "folded" ddFLN-dom5-RNC resonance in a specific manner. For protonated RNCs, ^{13}C-edited STE-HMQC spectra did not reveal any dispersed ddFLN-dom5 resonances, despite very high data accumulation, which is likely to be a result of transverse relaxation during the spin-echo that hosts the 2ms gradients. However, ^{13}C-edited STE-HMQC spectra of U-[^{12}C,^{2} H], Ileδ1-[$^{13}C^{1}H_3$] labeled ddFLN-dom5+110-

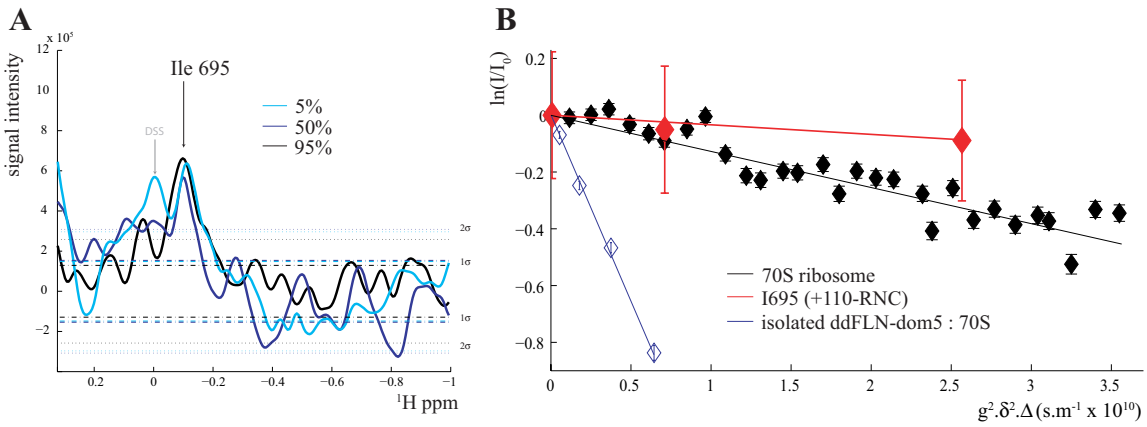


Figure 2.24: A: ^{13}C-edited STE-HMQC spectra of U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+110-RNC recorded with g=2.5 (cyan), g=25 (blue), g=47.5 G.cm^{-1} (black), δ=1.5ms and Δ=100ms. The dotted lines indicates one and two standard deviation of the noise. **B:** Plot of the relative signal intensity in the different spectra of the diffusion experiments in a log scale, versus the square of the gradient strength, for the I695 resonances in the ^{13}C-edited STE-HMQC spectra of ddFLN-dom5+110-RNC in red and for the purified released in the presence of ribosome in blue (Section 2.3.1.4). The intensity of the ^{1}H STE of empty 70S is shown in black.

RNC displayed an NMR signal that corresponded to the resonances of "folded" Ileδ$_1$ 695 (Figure 2.24), which do not overlay with any 70S ribosomal resonances. Despite the low SN (*ca.* 3.5), the absence of signal decay from 2.5 G.cm^{-1} to 47.5 G.cm^{-1} indicates slow translational diffusion similar to that of the 70S complex, thus enabling the demonstration that the NC is attached to the ribosome unambiguously (Figure 2.24B).

This result showed the ability to monitor the attachment of the nascent chain *in-situ* during the typically long acquisition time required for RNCs. RNC are typically unstable and highly transient complexes in nature; the release of NC from the ribosome would significantly compromise the analysis of RNC spectra as even small amounts of release would result in the appearance of sharp resonances for folded ddFLN-dom5 that can overlay with RNC resonances. The ribosome-bound ddFLN-dom5-NC giving rise to the "folded" ^{1}H-^{13}C methyl-TROSY HMQC spectrum or "unfolded" ^{1}H-^{15}N SOFAST-HMQC

spectrum was directly monitored by interleaving the acquisition of the heteronuclear two-dimensional spectra with ^{1}H STE and either ^{13}C-edited STE-HMQC or ^{15}N XSTE diffusion experiments, respectively.

2.3.3 A complete strategy to monitor the integrity of RNCs

NMR data on RNCs are typically recorded over several days, and it is thus critically important to monitor and to understand the integrity of the RNCs over this time-frame. This section presents the overall strategies developed to monitor the stability of the RNC, using a combination of NMR experiments and biochemical assays, and attempts therefore to draw together an assessment of its behaviour over time.

The integrity of the RNC complex can be characterised by four main features (Figure 2.25): firstly the general integrity of the sample, which can suffer for example from bacterial growth, and secondly the stability of the 70S ribosome itself. Both of these aspects can be (and are) monitored using ^{1}H NMR experiments (described further in Section 2.3.3.1). The attachment of the NC to the ribosomal complex and in particular the release of the SecM stalling sequence may be monitored by the analysis of its Trp resonances which will be described in Section 2.3.3.3. Lastly, and perhaps most importantly, the integrity of the NC itself may be monitored by the analysis of the ddFLN-dom5 resonance intensities over time, the measurement of the NC diffusion coefficient over time via ^{13}C-edited STE-HMQC diffusion experiments and also using biochemical methods, in particular with the emphasis on the use of an anti-His western blot timecourse analysis, described in Section 2.3.3.4).

Over 30 RNC samples have been produced and studied during the course of this PhD. The data presented here have been chosen to best illustrate the various aspects of the stability of the RNCs samples over time, and have consequently been drawn together

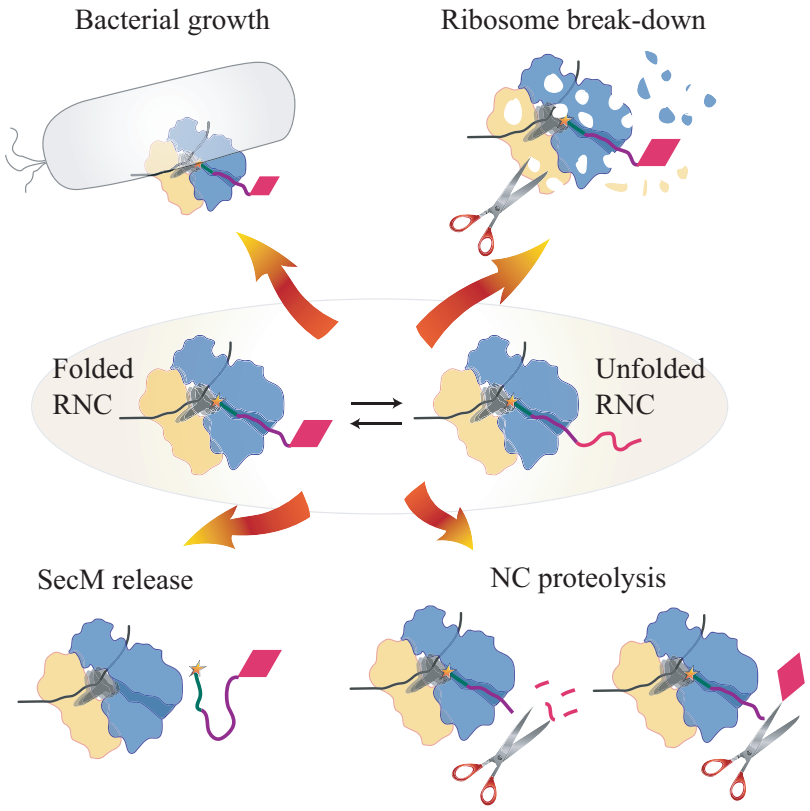


Figure 2.25: Schematic of the different processes by which the RNCs complexes can break-down. Refer to main text Section 2.3.3 for details.

from many different samples.

2.3.3.1 Stability of the 70S monitored by changes in the proton spectra over time

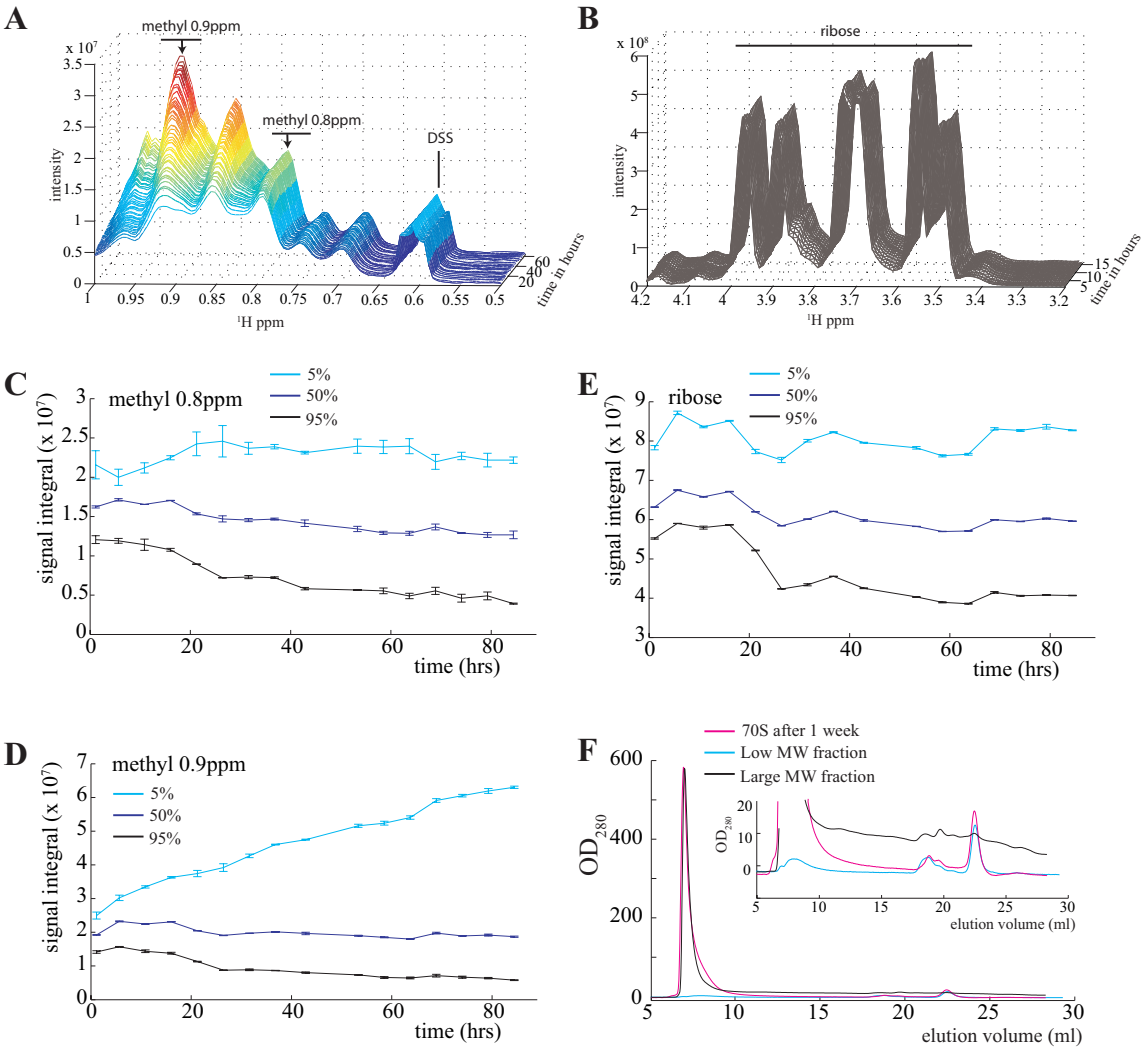
^{1}H NMR spectra can be used to report on the integrity of 70S ribosomes within the RNC over the time-course of the NMR acquisition, in particular when recorded in an interleaved manner with heteronuclear spectra and diffusion experiments. The very short acquisition time of a ^{1}H NMR spectrum (2-5 mins) renders it particularly favourable for the analysis of the sample integrity without compromising the time

available for acquisition of NMR data relevant for the co-translational folding analysis (e.g. heteronuclear spectra). Over the course of the sample lifetime, ^{1}H spectra of RNCs typically exhibit a number of changes, particularly in the methyl region (Figure 2.26).

Upon closer inspection, three resonances were able to be assigned to the L7/L12 methyls by comparison with ^{1}H-^{13}C HMQC of ^{13}C labelled ribosome (Figure 2.26 and Chapter 4). Among them, a signal at 0.8 ppm typically did not change over time (Figure 2.26A & C cyan curve), however, the signal at 0.9 ppm was found to double in intensity over 20-30 hours of sample lifetime in ^{1}H spectra of each of the RNC sample (Figure 2.26A & D, cyan curve).

The properties of these signals with increasing intensity over the three days of NMR acquisition can be examined by measuring the translational diffusion over the same time-course, using interleaved ^{1}H STE diffusion experiments. Figure 2.26C & D shows the intensity of the signals centred around 0.8 ppm and 0.9 ppm in the ^{1}H STE diffusion experiment with gradient strengths of g=2.5 (cyan), g=25 (blue), g=47.5 G.cm^{-1} (black). The increase in intensity of the signal at 0.9 ppm was typically

Figure 2.26 ***(following page)*****: A:** Stack of ^{1}H 1D spectra of a ddFLN-dom5+110-RNC sample recorded at hourly intervals during the NMR acquisition time (0-85 hours). Two sections of the methyl region are highlighted, the methyl resonance centred at 0.8 ppm assigned to the L7/L12 stalk and a methyl resonance centred at 0.9 ppm, which is positioned at a random coil methyl chemical shift. **: B:** Same as A, but the ribose region of the spectra is shown. **C, D & E:** Plots of the variation of intensity of the ^{1}H STE diffusion experiment with the NMR experimental time. Low gradient strenght spectra (cyan, g=2.5 G.cm^{-1}), middle gradient strength (blue, g=25 G.cm^{-1}) and high gradient strength (black, g=47.5 G.cm^{-1}) are shown. Plot B shows the signal intensity vs time for the resonance centred at 0.7 ppm. Resonances centred at 0.9 ppm is shown in plot C, and the ribose signal is shown in plot E. Refer to the main text for the analysis of the variations over time. **F:** Size exclusion profile on an analytical Superdex 200 of a ribosome incubated at 25°C for one week. The profile of the high molecular weight fraction is shown in magenta and that of the low molecular weight fraction is shown in cyan. The lower graph shows the same profiles with a zoom on the low absorbance values.


A
methyl 0.9ppm
methyl 0.8ppm
DSS
x 10^7
intensity
time in hours
¹H ppm
B
ribose
x 10^8
intensity
time in hours
¹H ppm
C
methyl 0.8ppm
5%
50%
95%
signal integral (x 10^7)
time (hrs)
E
ribose
5%
50%
95%
signal integral (x 10^7)
time (hrs)
D
methyl 0.9ppm
5%
50%
95%
signal integral (x 10^7)
time (hrs)
F
70S after 1 week
Low MW fraction
Large MW fraction
OD280
elution volume (ml)

associated with an increase in the diffusion coefficient from $2.0\pm0.5\times10^{-11}$ m^2s^{-1} to $10.0\pm0.5\times10^{-11}$ m^2s^{-1}. The spectra recorded with a high gradient strength (black line in figure 2.26D, g=47.5 $G.cm^{-1}$) showed a much slower decay, indicating that the ribosome-bound resonances disappeared slowly. Together, this suggests that the increase in intensity corresponds to the appearance of fast diffusing released ribosomal proteins that contribute to the random-coil signal at 0.9ppm (doubling time typically 20-30 hours), whilst the ribosomally-associated L7/L12 became released in a much longer timescale (half time typically 3-6 days).

In order to quantify the amount of released material observed in the ^1H 1D spectra, a ribosome sample was analysed biochemically after the ^1H 1D spectra recorded from that sample over time had showed the same pattern as in Figure 2.26A. After seven days at 25°C, the ^1H STE diffusion spectra of the 70S ribosome sample exhibited a similar increase in translational diffusion (D=$10.0\pm0.5\times10^{-11}$ m^2s^{-1}). The sample was then analysed biochemically by analytical sucrose gradient sedimentation and size-exclusion chromatography (Figure 2.26F). In the size exclusion chromatography profile, the 70S ribosomal fractions eluted in the void volume (∼7ml), and the absorbance at OD_{280} of the latter eluting species accounted for less than 2% of the initial ribosome content (OD_{280}=10 versus OD_{280}=580 for the 70S ribosomal fraction), indicating that only a small proportion of the ribosomes were disassembled. The ratio OD_{260}/OD_{280} of the low MW fraction were found to be similar to that of an intact ribosome sample (2.0), indicating that the ratio of ribosomal RNAs and ribosomal proteins in the released fractions was that of the intact ribosome, thus suggesting that the entire ribosome was being disassembled (as opposed to specific proteins being released).

In an intact ribosome, only two copies of L7/L12 are observable in the ^1H 1D spectrum. However, the disassembling of the 70S ribosome results in the appearance

of 53 additional proteins in the proton spectrum, that were not observed initially. Consequently 2% of released ribosomal proteins (as quantified biochemically) resulted in the appearance of new signals in the ^{1}H 1D and ^{1}H STE spectra that were sharp, intense, and associated with a fast translational diffusion. Therefore, the acquisition of ^{1}H 1D and ^{1}H STE spectra can be used to quantify the disassembly of the entire ribosomal complex in the RNC samples.

Stability of the ribosomal RNA monitored in ^{1}H 1D and ^{1}H STE diffusion spectra

Aside from changes in the methyl region of the ^{1}H 1D spectra, examination of other regions of ^{1}H spectra revealed ribose signals that were found to have a diffusion coefficient identical to that of ribosome-bound L7/L12 methyl resonances (Figure 2.26E). The analysis of these ribose signals over the time-course of a typical NMR acquisition period of an RNC can be used to investigate their attachment to the ribosome complex. The intensity of these ribosome signals in the high gradient strength ^{1}H STE spectra (black line in Figure 2.26E, g=47.5 G.cm^{-1}) were found to decrease over time, indicating the loss of ribosome-bound ribose, in line with the previous biochemical observation that both ribosomal protein and RNA were being released when the ribosome complex is being degraded. However, the intensity of the ribose resonances in the ^{1}H 1D spectra remained stable, suggesting that the released ribosomal RNA observed via OD_{260} measurements was not observed in the conventional ^{1}H spectra, and only the ribosomally-attached ribose are observed in the ^{1}H 1D and ^{1}H STE diffusion spectra.

2.3.3.2 Microbial contamination observed via changes in the ^{1}H 1D spectra

The typical release of ribosomal proteins described above follows simple exponential kinetics. In contrast, a second set of resonances in the proton spectra were observed

to changes with sigmoidal kinetics (Figure 2.27). The decrease in the intensity of signals (Figure 2.27 A & B) which can be assigned to sucrose, was concomitant with the appearance of intense peaks (Figure 2.27C & D). These peaks have been tentatively assigned to a product of the metabolism of sucrose. Changes in the HEPES buffer ^{1}H chemical shifts (Figure 2.27E&F) occurred on the same timescale, and were found to correspond to a decrease in the pH of the sample from 7.4 to 6.8 (measured electrochemically). These simultaneous changes can be attributed to the growth of a microbial contaminant digesting sucrose. In RNC samples that exhibited these changes in the ^{1}H 1D, the lag phase of the sigmoidal growth kinetics varied between 10-30 hours and the log phase occurred between 12-38 hours. The careful removal of sucrose via buffer exchange in a 100kDa cut-off concentrator significantly increases the lag phase of the growth, and microbial contamination can be prevented during the entire NMR acquisition time (3-4 days). Nonetheless, because these changes seems to have an effect that is not reflected in the ribosomal resonances, in other words because the kinetics of the changes of the ribosome signals is significantly different than the growth kinetics discussed here, it is likely that it has little effect on the ribosome stability if the amount of sucrose is low enough.

Together these results demonstrate that the stability and integrity of the ribosomal complex can be monitored via ^{1}H 1D spectra recorded in an interleaved manner with the heteronuclear spectra. In the majority of samples, only a small percentage of the ribosomal proteins are released, and the bulk of the ribosomes remains intact over the course of a week at 25°C.

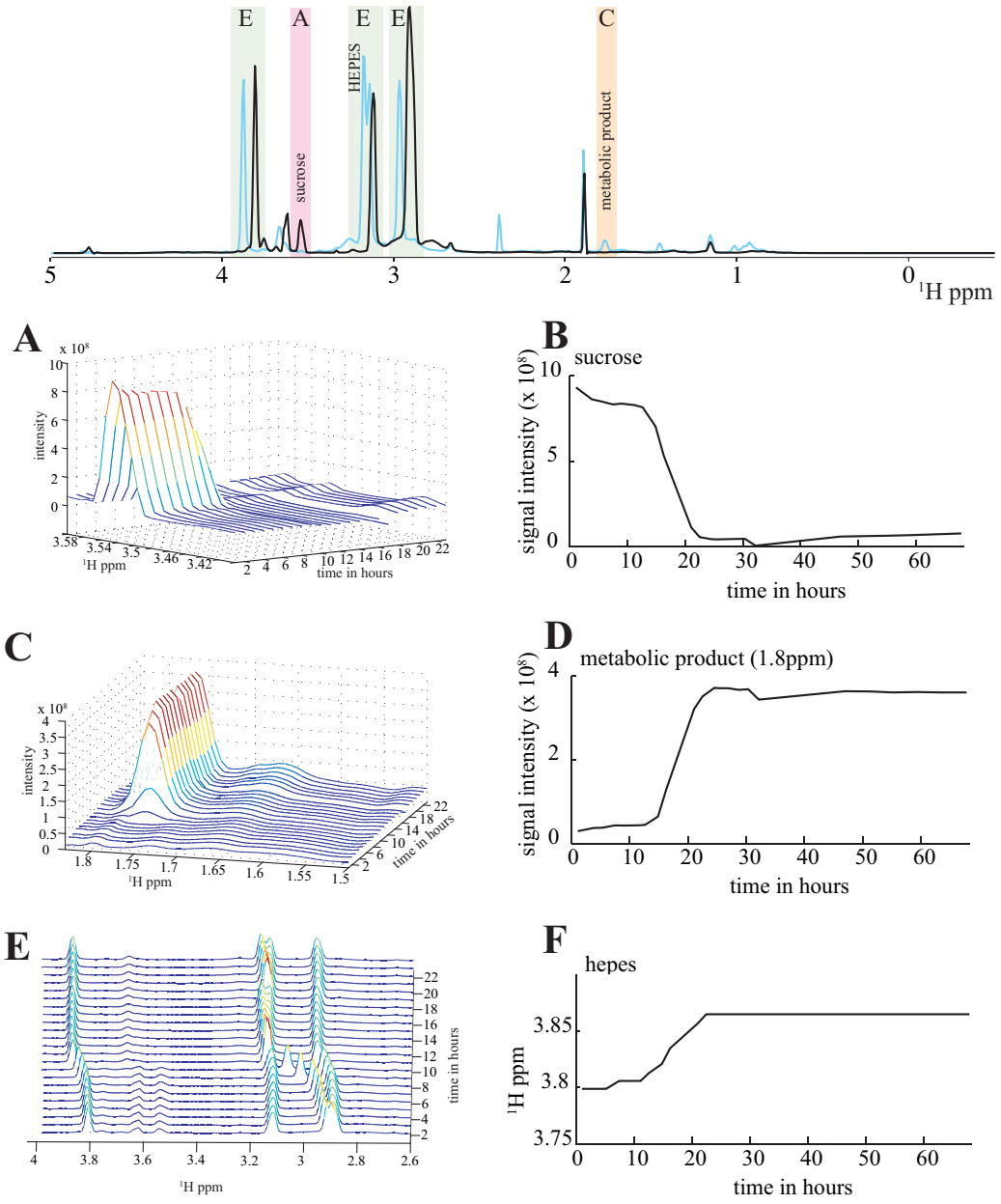


Figure 2.27: Top: ^{1}H 1D spectrum of a ddFLN-dom5+110-RNC sample recorded at time=0 (black, i.e. immediately after exchange in a 100kDa cut-off concentrator) and at time=70 hours (blue). The HEPES, sucrose and metabolic products chemical shifts are highlighted in green, magenta and orange respectively. **A, C & E:** Stack of ^{1}H spectra of ddFLN-dom5+110-RNC recorded after increasing acquisition time at 25°C, zoomed to show the ppm region corresponding to the sucrose chemical shift (A), a resonances that increased over time (C) and the HEPES chemical shift (E). The relative intensity of the sucrose peak at 3.4 ppm versus acquisition time is shown in **B**, that of the sharp peak at 1.8ppm is shown in **D** and the chemical shift of the the HEPES peak over time is shown in **F**.

2.3.3.3 Monitoring SecM-based release

To understand the process by which the NC is released from the ribosomal complex, and in particular to test whether the SecM stalling sequence is being released, the appearance of signals from the sidechain of the Trp residue within SecM can be analysed over the time-course of the NMR acquisition (Section 2.3.2.7). Only on two occasions (samples of ddFLN-dom5+31 & +37RNCs) did an increase in intensity for the "folded" well-dispersed ddFLN-dom5 resonances correlate with the appearance of a Trp signal with a statistically significant intensity (Figure 2.28I), which is suggestive of the SecM region being released from the ribosomal tunnel. Nonetheless, for the majority (10 out of 30) of analysed RNC samples, the absence of observation of the Trp signal whilst the ddFLN-dom5 resonances showed a significant increase in intensity indicated that the release of the NC was not SecM mediated (Figure 2.28K). Together, these observations suggest that whilst SecM release is not the primary cause of NC release, it can be monitored by analysing the intensity of the sidechain Trp resonances over the time.

2.3.3.4 Monitoring the integrity of the NC itself

The combined analysis of ^{1}H 1D, ^{1}H STE spectra and the Trp sidechain resonance over time suggests that most RNCs do not release via 70S ribosome degradation or SecM mediated released. Nonetheless, an increase in intensity for the "folded" ddFLN-dom5 resonances was typically observed in the ^{1}H-^{15}N SOFAST-HMQC (Figure 2.28K) and ^{1}H-^{13}C HMQC (Figure 2.29B) spectra. For example, after 20-30 hours of acquisition at 25°C, the sensitivity for dispersed ddFLN-dom5 resonances increased significantly (Figure 2.29B compared to A).

In order to report whether the resonances associated with such increase in intensity in the heteronuclear spectra arise from released NC, ^{13}C-edited STE-HMQC diffusion

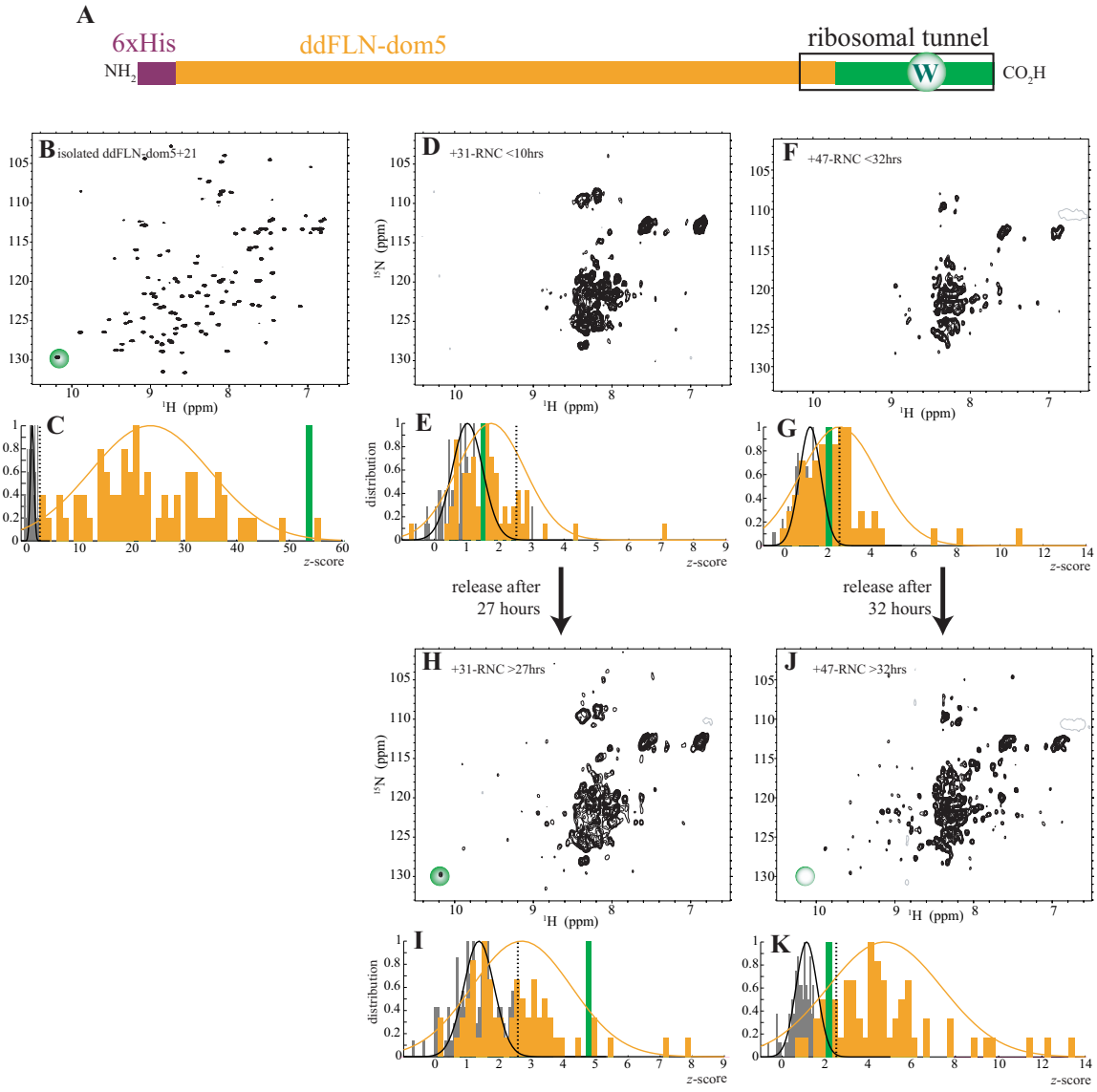


Figure 2.28: **A:** Schematic of the ddFLN-dom5-RNC constructs, highlighting the Trp residue within SecM that interacts with the ribosomal exit tunnel. **B:** ^{1}H-^{15}N HSQC spectrum of purified released ddFLN-dom5+31-RNC. The Trp sidechain resonance is circled in green. **C:** Normalised histogram of the *z*-score of the intensity of the ^{1}H-^{15}N resonances of folded ddFLN-dom5 shown in spectrum B in orange, and Trp sidechain in green. **D:** ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+31-RNC, recorded within the first 10 hours of acquisition. **H:** ^{1}H-^{15}N SOFAST-HMQC spectrum of the same sample recorded within 27-37 hours of sample life-time. The sidechain tryptophan is highlighted in green, and overlay with the resonances of the SecM tryptophan seen in spectrum B. **F:** ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+47-RNC, recorded within the first 32 hours of acquisition. **G:** ^{1}H-^{15}N SOFAST-HMQC spectrum of the same sample recorded within 32-66 hours of sample lifetime. Although the ddFLN-dom5 resonances have increased in intensity, the absence of the Trp sidechain resonances is highlighted with the hollow circle. **: E, G, I, K:** Same as C for spectra shown in D, F, H & J respectively.

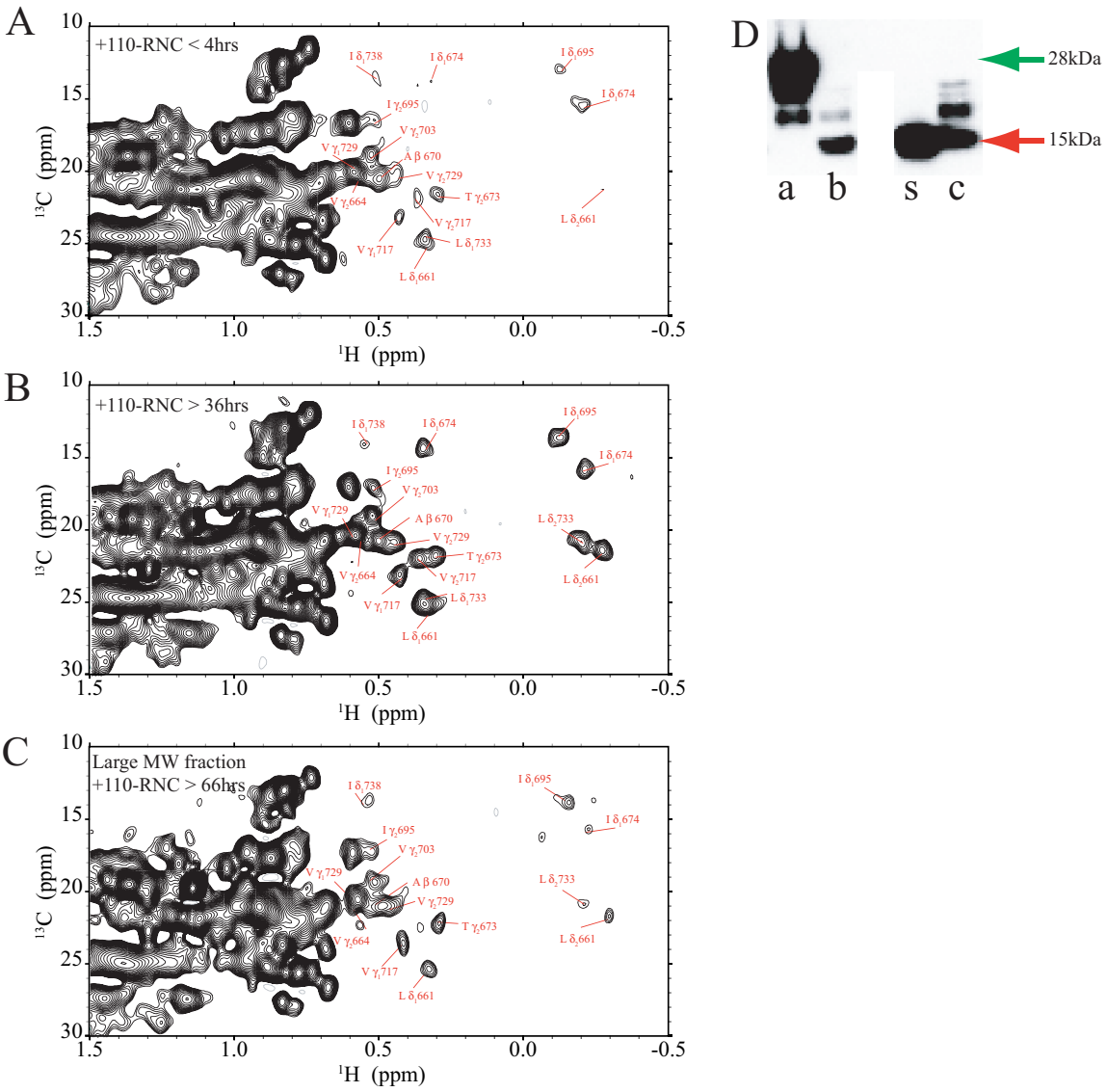


Figure 2.29: ^{1}H-^{13}C HMQC spectra ddFLN-dom5+110-RNC. **A:** Spectrum acquired within 7.3 hours at 25°C, with a total acquisition time of 3.3 hours. The 6×His western band of the sample before the acquisition of A is shown in **D** lane **a**. **B:** Spectrum acquired from 36 to 47 hours at 25°C. The 6×His western blot of the sample after the acquisition of B is shown in **D** lane **b**. After B is recorded, the low molecular weight fractions are separated from the ribosome fraction via a 100kDa concentrator. The 6×His western band of the low molecular weight fractions separated from the sample after B is recorded is shown in **D** lane **s**. The ribosomal fraction is shown in **D** lane **c**. The spectrum of sample **c** is shown in **C**. All spectra are recorded with the same total acquisition time (3.3 hours) and displayed with the same threshold.

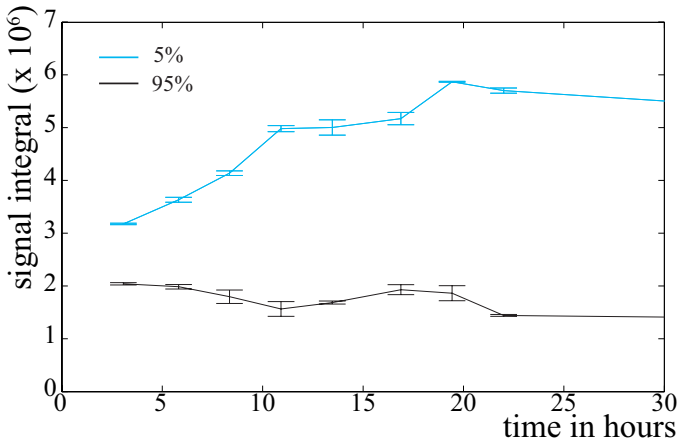


Figure 2.30: Signal integral from the random-coil methyl signals in the ^{13}C-edited STE-HMQC spectra of ddFLN-dom5+110-RNC (0.6-0.8 ^{1}H ppm) recorded with g=2.5 (black) and g=47.5 G.cm^{-1} (blue) as a function of sample acquisition time at 25°C.

spectra can be recorded throughout the NMR acquisition period, as was described in Section 2.3.2.8. The diffusion data typically show a similar increase in signal intensity in the low gradient strength (g=2.5 G.cm^{-1}) diffusion spectrum within the first 24 hours of acquisition (Figure 2.30 shows an example from the data recorded on ddFLN-dom5+110-RNC). The translational diffusion coefficient calculated from these data, using Equation 1.15, increases over time from $2\pm0.5\times10^{-11}$ m^{2}s^{-1} initially to $10\pm0.5\times10^{-11}$ m^{2}s^{-1} after 24 hours of acquisition, which reveals the appearance of low-molecular weight labelled proteins. After 24 hours of acquisition at 25°C, the measured translational diffusion coefficient reached a plateau. The timescale of the increase in diffusion was found to be similar to the timescale of increase in intensity of the "folded" ddFLN-dom5 resonances in the ^{1}H-^{13}C HMQC (Figure 2.31A), suggesting that those "folded" ddFLN-dom5 resonances arise from released NCs.

To further confirm the observation that the "folded" ddFLN-dom5 resonances arise from released NCs, the sample was subjected to a 100 kDa cut-off concentrator after 30

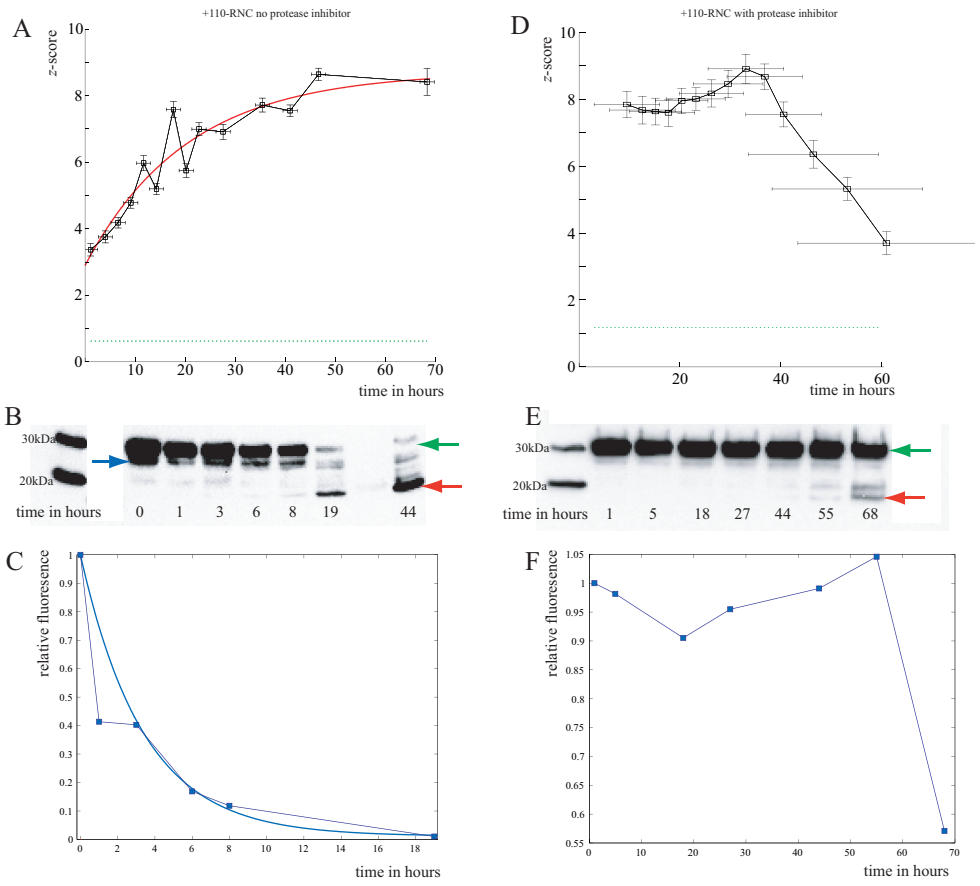


Figure 2.31: **A:** Mean z-scores of 16 high field shifted resonances assigned to folded ddFLN-dom5 residues (Section 2.3.2.5) in ddFLN-dom5+110-RNC (Figure 2.29). Vertical error-bars show the standard error of the mean, and horizontal error-bars correspond to the time of acquisition for each spectrum. The green dotted line shows the 5% threshold corresponding to the limit of observability of the mean of 16 resonances (Section 2.2). Data are fitted to Equation 2.7, k=20 $\pm$ 13 hours^{-1} (r^2=0.94) **B:** Anti-His western blot of ddFLN-dom5+110-RNC during the timecourse. The blue arrow shows a lower molecular weight band below full-length ddFLN-dom5+110-NC at time 0. The red arrow shows the 14kDa product of the proteolysis and the green arrow shows the full-length 28kDa NC. **C:** Intensity of the chemiluminescence signal in the 6$\times$His western blot shown in B for full-length ddFLN-dom5. Data are fitted to Equation 2.8, k=3$\pm$3 hours^{-1} (r^2 0.99). **D**, **E** and **F**: same as A, B and C but for U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+110-RNC. In this spectra, only the five dispersed Ile defined in Section 2.3.2.5 are analysed.

hours of NMR experiments, and the ribosomal fractions were analysed by NMR, and it was found that the folded ddFLN-dom5 resonances were significantly reduced compared to before sample processing (Figure2.29C compared to B).

To understand the origin of the release (which was not SecM mediated), the sample was analysed by anti-His western blot before the NMR acquisition (Figure 2.29D, lane a) and after 30 hours of acquisition (Figure 2.29D, lane b). It was found that the 6×His band was truncated to *ca.* 15 kDa (red arrow) from 28 kDa (the size of the full length, green arrow). This indicates that the 6×His ddFLN-dom5+110-NC was being cleaved and released from the ribosome, thus giving rise to intense "folded" signals in the heteronuclear spectrum. To analyse further the size of the truncated product, the low molecular weight fractions of the sample upon filtering through a 100 kDa cut-off concentrator was analysed by MALDI-TOF mass-spectrometry. The MALDI mass-spectrum showed that the released chain corresponded to the NC with a cleavage of ddFLN-dom6 at a position eleven amino acids downstream from ddFLN-dom5, and 99 amino acids upstream from the PTC centre (Figure 2.32B).

These observations from ddFLN-dom5+110-RNC suggest that the analysis of the intensity of the "folded" ddFLN-dom5 resonances in ^{1}H-^{13}C HMQC spectra over the NMR acquisition period, together with the analysis of the ^{13}C-edited STE-HMQC diffusion spectra and the western blot analysis of samples collected at the same time-points can be used to analyse the timescale of the cleavage in RNC samples.

The increase in intensity of the ddFLN-dom5 resonances was typically found to follow single exponential kinetics (Figure 2.31A) as did the decrease in the intensity of the band corresponding to the full-length construct in the western blot (Figure 2.31B). The two events therefore appear to be correlated and to follow similar kinetics which resembles proteolysis, and where the size of the proteolysis product typically

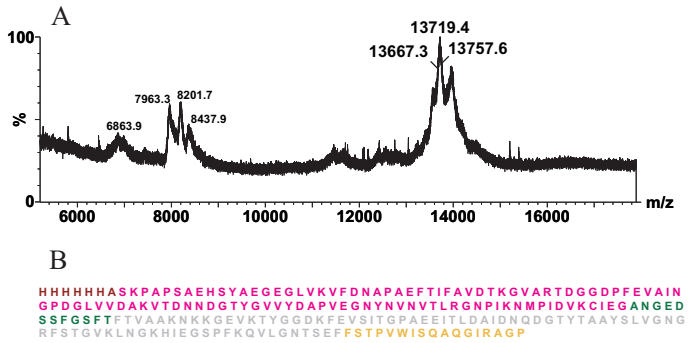


Figure 2.32: **A:** MALDI mass-spectrum of the isolated released chain from ddFLN-dom5+110-RNC sample (via filtering through a 100 kDa cut-off concentrator as described in the text) after 30 hours of NMR acquisition, corresponding to sample "**s**" in the 6×His western shown in Figure 2.29D. **B:** Sequence of ddFLN-dom5-NC from N to C-terminii, with in dark red the 6×His tag, in magenta ddFLN-dom5, in green the length of the linker that would correspond to a size of 13.8kDa, the rest of the linker is shown in grey before the SecM sequence in yellow.

corresponds to the size of the ddFLN-dom5 folded domain observed by NMR.

The proteolysis process is a first order reaction:

$$\mathrm{RNC_{FL}} \xrightarrow{\mathrm{k}} \mathrm{NC_{trunc}} \tag{2.6}$$

with the intact, full length RNC ($\mathrm{RNC_{FL}}$) concentration decreasing as:

$$[\mathrm{RNC_{FL}}]_t = [\mathrm{RNC_{FL}}]_0 . \exp\left(\frac{-t}{k}\right) \tag{2.7}$$

and the truncated and released NC ($\mathrm{NC_{trunc}}$) increasing as:

$$[\mathrm{NC_{trunc}}]_t = [\mathrm{NC_{trunc}}]_\infty . \left[1 - \exp\left(\frac{-t}{k}\right)\right] \tag{2.8}$$

In Figure 2.31A, the increase in the mean intensity of the released high field shifted NC resonances observed from a ddFLN-dom5+110-RNC sample was fitted to Equation 2.8

(red line in Figure 2.31A), which gave a time constant for proteolysis of 20 $\pm$ 5 hours^{-1} (r^2=0.94). The same equation was then used to fit the increase in intensity observed in low gradient strength ^{1}H-^{13}C edited diffusion spectrum of Figure 2.30 (g=2.5 G.cm^{-1}), which gave a time constant for proteolysis of 13 $\pm$ 4 hours^{-1} (r^2 0.96). The decrease in band intensity of the full-length RNC observed in western blots (Figure 2.31B & C) was fitted to Equation 2.7 and gave a time constant of 3$\pm$1 hours^{-1} (r^2 0.99). The truncation of other ddFLN-dom5-RNC samples followed similar kinetics, where the typical time-constant of the proteolysis reaction varied from 10 to 20 hours. Interestingly, the nature of the proteolysis was found to vary depending on the ddFLN-dom6 linker length, and is described in Chapter 3 (Section 3.2.3).

The observed proteolysis was also found to be significantly attenuated by the presence of protease inhibitors. In particular the Complete inhibitor tablet from Roche supplemented with 1μM of the inhibitor pepstatin (pepsin protease inhibitor), and 5mM EDTA was determined to be particularly effective. As an example, Figure 2.31D shows that the signal intensity of the high field shifted ddFLN-dom5-RNC resonances in the ^{1}H-^{13}C methyl-TROSY HMQC spectrum mentioned in Section 2.3.2.5 did not increase over time, and the western blot analysis of the samples performed in parallel did not show any evidence for proteolysis over two days (Figure 2.31E). The decay in intensity beyond two days might be related to precipitation of the ribosome sample that was observed visually.

2.3.3.5 Assessing the RNC integrity post NMR

The attachment of the NC to the 70S ribosome after NMR can be assessed by pelleting the ribosomes of the RNC samples post-NMR acquisition through a 17% sucrose cushion, and monitoring the presence of the NC in the ribosome pellet via anti-His western blot

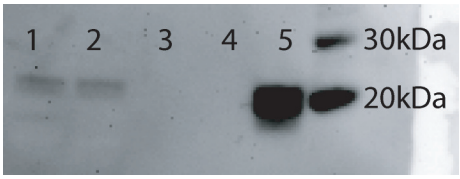


Figure 2.33: Western blot analysis of a sample of ddFLN-dom5+47-RNC sample after two days of NMR acquisition at 25°C. Lane 1: 10pmoles of post NMR sample, no treatment. Lane 2: 10pmoles of pellet through a 17% sucrose cushion (spin 30min at 120,000rpm). Lane 3 & 4 supernatant of the cushion. Lane 5: 10pmoles of purified ddFLN-dom5+47-RNC.

detection. Figure 2.33 shows and example of such analysis, the western blot signal was found to be nearly identical in the NMR sample of ddFLN-dom5+47-RNC prior pelleting and in the pellet, indicating that the NC remains attached.

2.4 Concluding remarks

We have shown here that the *in vivo* production method of ddFLN-dom5-RNC can generate homogeneously stalled, selectively labelled RNC samples amenable to NMR study. This chapter represents a description of the developed methods for the analysis of these RNC samples, highlighting the need for cautious continuous analysis of the integrity of the RNC samples via both biochemical and NMR methods to ensure that data recorded arise from RNC samples and not from degraded products.

Analysis of the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+21-RNC revealed intense resonances of the disordered state of ddFLN-dom5, indicating firstly that the NC populates the unfolded state, and secondly that the analysis of ^{1}H-^{15}N SOFAST-HMQC spectra of RNC with different linker lengths can report on the unfolded state of the RNC at increasing translation lengths (Figure 2.34A). Analysis of the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110-RNC revealed resonances from the disordered ddFLN-

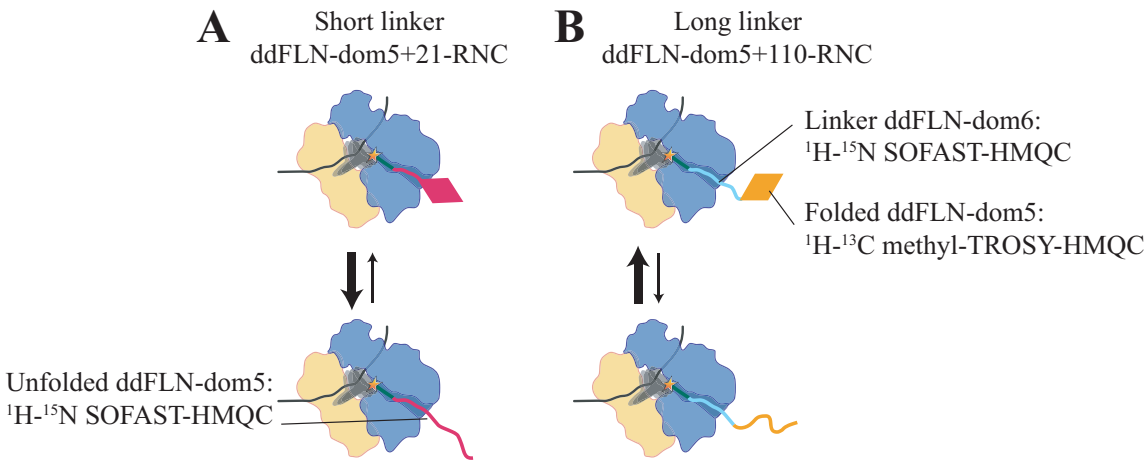


Figure 2.34: Schematic of the conclusion drawn from the analysis of the ^{1}H-^{15}N SOFAST-HMQC of ddFLN-dom5+21-RNC (**A**), and ^{1}H-^{15}N SOFAST-HMQC & ^{1}H-^{13}C methyl-TROSY HMQC of ddFLN-dom5+110-RNC (**B**). Refer to main text for details.

dom6 linker, confirming that the linker is intrinsically disordered at these lengths, and no distinct resonances that arose from a disordered ddFLN-dom5 (Figure 2.34B). The ^{1}H-^{13}C methyl-TROSY HMQC spectra of the deuterated, selectively Ile δ_1 labelled ddFLN-dom5+110-RNC sample showed resonances similar to the natively folded state of ddFLN-dom5 with statistically significant intensities, indicating that firstly ddFLN-dom5 populates the native folded state at this length, and secondly that ^{1}H-^{13}C methyl-TROSY HMQC spectra are suitable to provide structural and dynamical informations on the folded states of ddFLN-dom5-RNCs (Figure 2.34B). The combined use of ^{1}H-^{15}N SOFAST-HMQC and ^{1}H-^{13}C methyl-TROSY HMQC provide dual probes for examining the unfolded and folded characteristics, respectively, of RNCs, and were therefore used to examine the folding equilibrium of ddFLN-dom5-RNCs at different stages of emergence of this domain from the ribosomal exit tunnel, as is discussed in Chapter 3.

As a result of the developments described in this Chapter, the attachment of the RNC can now be routinely confirmed by the use of ^{13}C-edited STE-HMQC diffusion

experiments. The comparison of the linewidths of RNC methyl groups in ^{1}H-^{13}C methyl-TROSY HMQC spectra of ddFLN-dom5+110-RNC further confirms the RNC attachment to the ribosome. The attachment of the RNC rendered the C-terminal residues of the RNC constructs non-observable in conventional ^{1}H-^{15}N or ^{1}H-^{13}C HMQC acquisition, and in particular the resonances from the single Trp residue within the SecM stalling sequence, and this particular feature can be used to discriminate ribosome-bound from SecM-released nascent chain.

While the low intensity of the "folded" ddFLN-dom5-RNC resonances would suggest increasing the acquisition time (or indeed the temperature used for NMR acquisition), this was restricted due to the short lifetime of the complex. From the combined NMR and biochemical data, it appears that the RNCs are indeed susceptible to release, however, the nature of this process varies. Figure 2.35 shows the different ways by which the integrity of the RNCs was found to be degraded. When the timescale of release was close to the timescale of NMR acquisition, the signals from the release chain rendered the analysis of the RNC resonances challenging due to their significantly sharper linewidth compared to the ribosome-attached domain (Figure 2.21). It was found to be important to monitor the timescale of release by the methods developed in Section 2.3.3, which included the highly interleaved acquisition of:

1. ^{1}H 1D spectra from which any microbial contamination can be detected,
2. ^{1}H STE spectra to determine the translational diffusion associated with the proton resonances that can report on the integrity of the ribosomes (Figure 2.35A),
3. ^{13}C-edited STE-HMQC spectra which can report on the presence of released NC (Figure 2.35B);
4. ^{1}H-^{13}C HMQC and/or ^{1}H-^{15}N SOFAST-HMQC spectra, from which the intensity

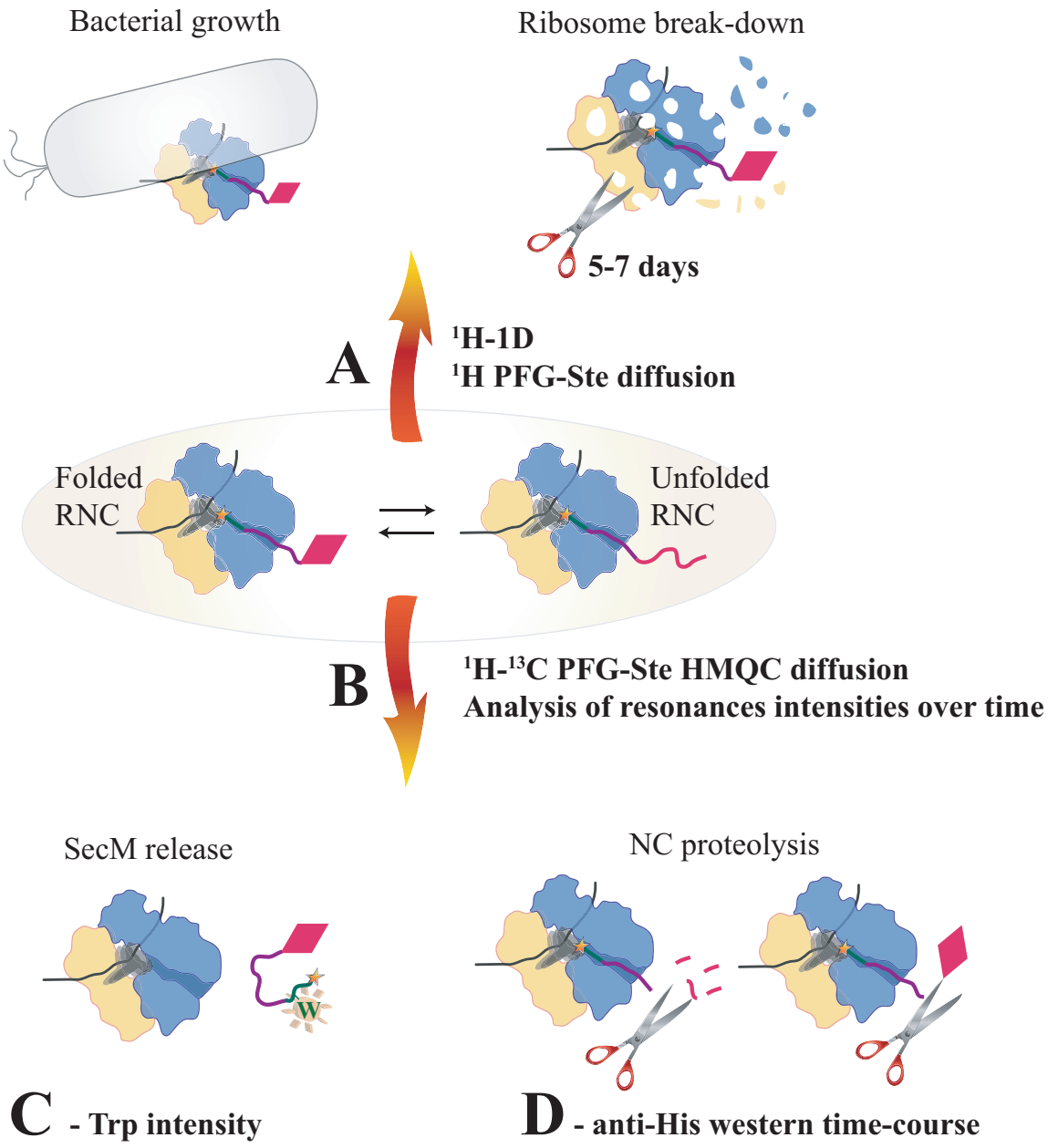


Figure 2.35: Schematic of the different processes by which the RNCs complexes can be degraded, as well as the methods developed to monitor this break-down.

of the dispersed ddFLN-dom5 resonances can be analysed over the time-course, increase in which are indicative of NC release (Figure 2.35B).

5. The observation of Trp sidechain indole resonance in ^{1}H-^{15}N SOFAST-HMQC spectra can identify SecM based released (Figure 2.35C), although this was found to be rare (two out of 30 samples).

6. In parallel to these NMR experiments, anti-His western blot of samples collected over time-course of NMR acquisition can reveal proteolysis at the ddFLN-dom6 linker (Figure 2.35D).

7. Finally, the integrity of the RNC sample after NMR experiment can monitor the presence of the NC within the ribosomal pellet.

In conclusion, this chapter has presented a general methodology for the NMR study of RNC, including a set of NMR experiments reporting on the folded and unfolded state of the RNCs, a methodology for the analysis of low intensity spectra, and a series of NMR and biochemical controls to monitor the attachment of the NC to the ribosome complex. This methodology forms the basis for a detailed analysis of the folding equilibrium of ddFLN-dom5-RNC as it emerges from the ribosome exit tunnel, described in the following Chapter 3.

Chapter 3

Snapshots of the emergence of ddFLN-dom5 from the ribosome exit tunnel followed by NMR spectroscopy

3.1 Introduction

3.1.1 ddFLN-dom5 has an immunoglobulin fold

The filamin protein gelation factor (ABP-120) from *Dictyostelium discoideum* (ddFLN) is an immunoglobulin-repeat protein consisting of six immunoglobulin domains followed by an actin-binding domain.The six immunoglobulin domains share a high sequence homology (average 40%, [189]), (Figure 3.1A). The native state structure of ddFLN-dom5 has been solved by X-ray crystallography [190] (figure 3.1A), and reveals a very similar structure to the NMR structure of ddFLN-dom4 [191] and the crystal structure of ddFLN-dom6 [190]. The structure consists of an antiparallel β-sheet sandwich, where each of the seven β-strands is labelled from A to G (N-term to C-term) according to the standard nomenclature for immunoglobulin folds [190] (Figure 3.1B). Within the structure, the N-terminal strand A closes the back β-sheet, and strand A' closes the front β-sheet, and makes contacts with the C-terminal strand G via hydrogen bounds (Figure 3.1C). The

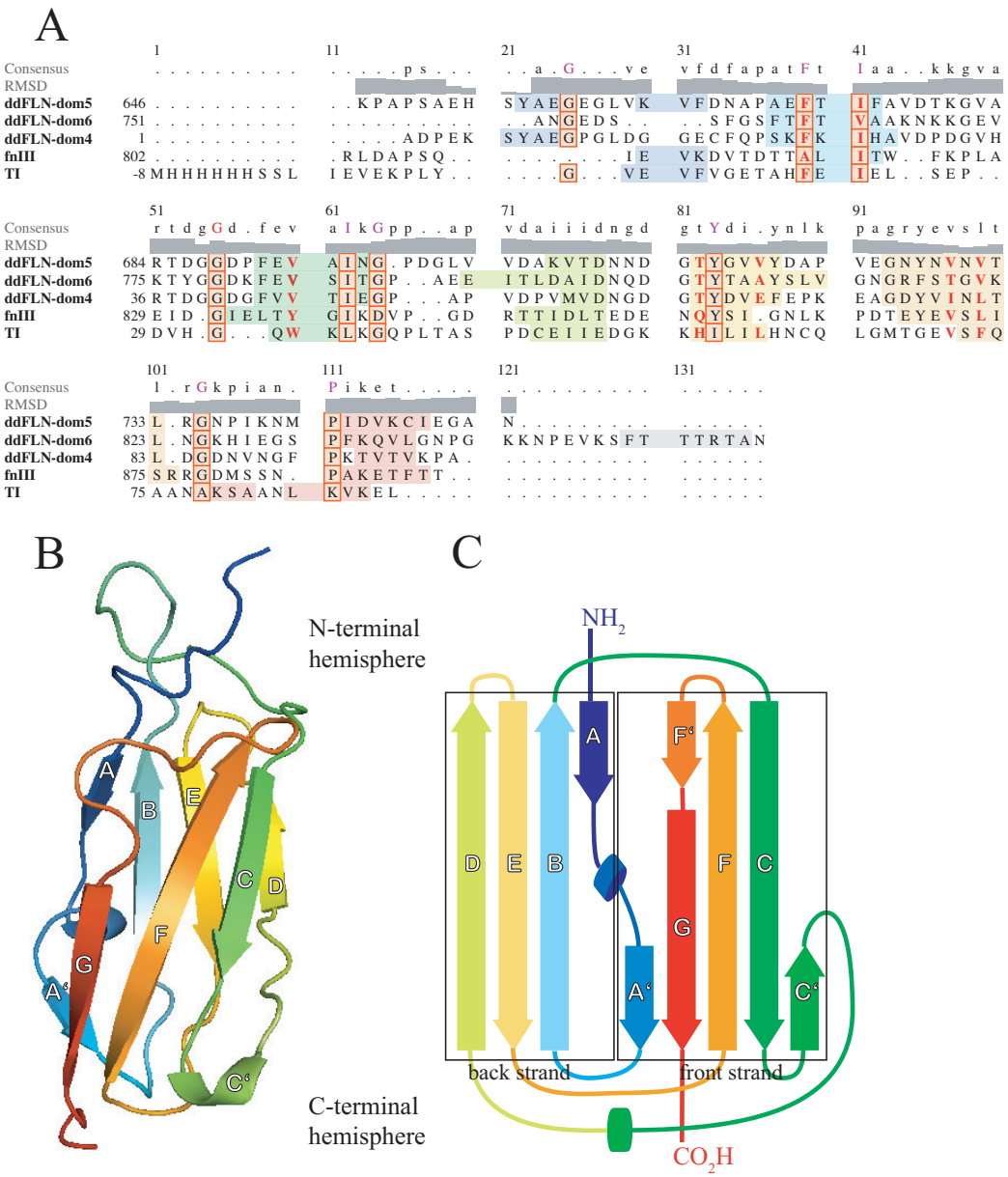


Figure 3.1: **A:** Sequence alignment of ddFLN-dom5, ddFLN-dom6, ddFLN-dom4, fnIII and TI (pdb entries *1QFH.pdb* [190], *1KSR.pdb* [191], *1TEN.pdb* [192] and *1TIU.pdb* [193] respectively), the secondary elements are shown with the same color coding as in A&B, the conserved residues are in the red rectangle, and the residues that have been shown to form the folding nucleus [56] are shown in red bold letters. **B** The X-ray crystal structure of ddFLN-dom5 with the six β-strands labelled A to G (pdb entry *1QFH.pdb*, [190]). **C:** A topology map of ddFLN-dom5 adapted from [190].

dynamics of the native state of the domain has been studied by NMR, and the overall high order parameter (S^2 >0.95) of the β-sandwich indicates a very stable β-sheet [183], which is also consistent with low b-factor values derived from the crystal structure [190]. Within the structure, the N-terminal hemisphere (loops BC, DE and FG, Figure 3.1) has been shown to be ordered (S^2>0.95), whereas the C-terminal hemisphere (loops A'B, C'D and EF, Figure 3.1) is more dynamic (and in particular the loop from strands C' to D) and contains residues with lower order parameters ($S^2 \approx 0.7$) and high b-factor values from the crystal structure [183].

The immunoglobulin domains of ddFLN share an overall homology in structure and in sequence with other immunoglobulin domains from other organisms (Figure 3.1C). The *in vitro* folding of immunoglobulin-like domains has been extensively studied over the last 15 years, with detailed studies of the folding of ddFLN-dom4 [61], of fibronectin type III domains (fnIII) [194, 195] and of titin domain (TI) [196, 196], using a range of biophysical and structural techniques. One of the major biophysical techniques that has been used to study the conformations of immunoglobulin domains is hydrogen/deuterium (H-D) exchange, which in combination with denaturation/renaturation experiments, has been used to probe structural features during the folding of the proteins [52, 197].

3.1.2 Introduction to hydrogen exchange studies of protein folding

Hydrogen/deuterium (H-D) exchange reports on both the structure and dynamics of a given system, and it is often used in conjunction with mass spectrometry and/or NMR. For example, H-D exchange rates measured by NMR at a residue specific level were used to define the folding characteristics of lysozyme [198], and to characterise the formation of intermediate ensembles populated during the folding process of a number of proteins

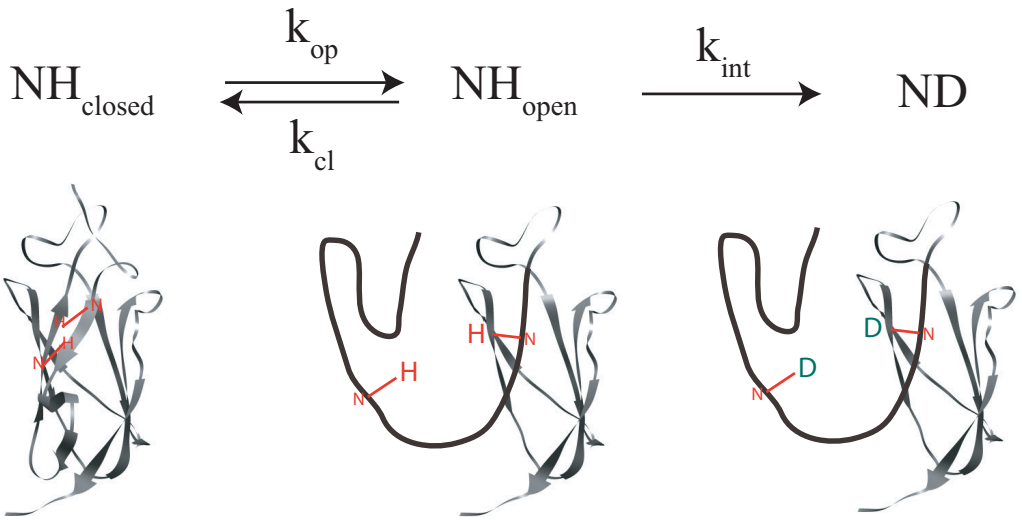


Figure 3.2: Linderstrøm-Lang model of the hydrogen exchange reaction [200]. Refer to main text for details on the analysis of the different regimes for the H-D exchange reaction.

[52,53], among them fnIII [194,199]. The use of NMR to record H-D exchange at a residue specific level relies on the fact that deuterium is not detected within standard ^{1}H NMR experiments. The experiment therefore consists of the rapid dilution of a protonated sample of protein in D_2O, followed immediately by the recording of the decay in the intensity of ^{1}H-^{15}N amide resonances in real time as the amide protons exchange with the deuterium within the solvent.

The Linderstrøm-Lang model of H-D exchange proposes that the rate of H-D exchange (k_{ex}) can be related to the open/closing of structural segments of the protein [200] (Figure 3.2). Briefly, in the "closed" conformation, the amides are protected from exchange with the deuterium of the solvent, and in the "open" conformation, the amide protons exchange with an intrinsic rate that only relates to the peptide bond and the protection offered from the two surrounding sidechains (or in other words, the exchange

rate of the amide proton in the case where no protection is offered by the structure of the protein). Above pH of *ca.* 3, the intrinsic exchange rate k_{int} is base-catalysed and therefore is directly proportional to the pH, and has been measured on dipeptides of each of 400 amino-acid combinations [201]. Two regimes can be analysed depending on the relative values of k_{int} and the closing rate of the structural element k_{cl}. In the case of high pH (EX1 regime), k_{op} is the rate limiting step, $k_{\text{int}} \gg k_{\text{cl}}$, and the rate of conformation exchange is simply $k_{\text{op}} = k_{\text{ex}}$. In the case of low pH (EX2 regime), the rate limiting step is the intrinsic H-D exchange, $k_{\text{int}} \ll k_{\text{cl}}$, and the exchange rate k_{ex} probes for the thermodynamic equilibrium for the opening-closing of the structure (quantified by the protection factor P):

$$P = k_{\text{cl}}/k_{\text{op}} = k_{\text{int}}/k_{\text{ex}} \tag{3.1}$$

P relates to the difference in free energy of the observed residue,

$$\Delta G_{\text{op}-\text{cl}} = RT \ln(P)$$

between the "closed" state and the "open", exchange-competent state [202]. By measuring the protection factors of a different folding equilibrium than the native equilibrium (using chemical denaturation for example), the folding free energy landscape can be mapped at a residue specific level [52, 53]. Additionally, the protection factors of the native state can also be used to map the stability of the structure and these studies have been performed on several immunoglobulin-like fold proteins [194, 199].

3.1.3 The *in vitro* folding pathways of immunoglobulin domains

As has been mentioned before, the aim of this chapter is to study the co-translational folding pathway of ddFLN-dom5, and to compare this to the *in vitro* folding pathways

that have been proposed for immunoglobulin domains. The information available on the *in vitro* folding pathways derives from different experiments that probed the thermodynamic equilibrium (H-D exchange [52, 197]), kinetics of folding (ϕ-value analysis [203]), and structural details along the folding pathways (AFM [204]). Particular questions are relevant for the comparison of the *in vitro* and co-translational folding: is the folding cooperative or are their any intermediate states being occupied? Are these intermediate states obligatory? Are the transition states native like? What are the stability of the different states?

To the best of our knowledge, there is no structural details of the (re)folding pathway of ddFLN-dom5, but the urea unfolding of ddFLN-dom5 monitored by NMR, UV CD and tyrosine fluorescence was seen to be mainly cooperative, with the disordered and folded states in slow interconversion on the NMR timescale (*ca.* 100ms) [183]. From the folding mid-point (4.2M urea), the free energy of unfolding was found to be $\Delta G_{\mathrm{F-U}} = 7.8 \pm 1.9$ kcal.mol^{-1}. Nonetheless, on the basis of NMR intensities, a transient population of an invisible intermediate state was inferred and at most 20% of the domains were observed to populate this unfolding intermediate at folding mid-point. In Section 3.2.1 of this current chapter, the stability of the native state of ddFLN-dom5 is investigated using H-D experiments, and compared to the free energy of unfolding.

Additional information on the folding pathway of immunoglobulin domains were derived from several other immunoglobulin domains which structural homology with ddFLN-dom5 is shown in Figure 3.1A.

Hydrogen-deuterium exchange experiments on the immunoglobulin domain fnIII have shown that β-strands B, C and F are resistant to hydrogen exchange [194, 199] ($\Delta G_{\mathrm{op-cl}} > 5$ kcal.mol^{-1}), whereas residues in the A' and G strands are somewhat less stable [194] ($\Delta G_{\mathrm{op-cl}} \sim$ 3-4 kcal.mol^{-1}), and can sample a wider conformational range

at room temperature. The increased stability observed for the inner stands B, C and F as measured by H-D exchange is likely to be related to their involvement in the folding core of the protein as determined by ϕ-value analysis [194]. ϕ-value analysis consists of quantifying the effect of the mutation of a set of residues one by one to the chemical folding and unfolding kinetics. A residue with a high ϕ-value indicates that the residue is structured in the folding transition state; a residue with a low ϕ-value indicates that the residue only becomes structured late in the folding pathway, after the transition state. Residues in strands B, C and F of fnIII were found to be typically associated with high ϕ-values [194, 195], which suggest that these are structured in the transition state ensemble.

The folding landscape of titin (TI) has been also studied using ϕ-value analysis and computational modelling. β-strands B, C, E and F were found to be associated with high ϕ-values [196], which indicates that they make native-like contacts in the transition state ensemble [56], and residues in the A' and G strands are associated with lower ϕ-values in TI [196] and might be relatively disordered in the transition state ensemble. A comprehensive computational study that incorporates these experimental results showed that the transition state ensemble of TI is native like, with strands B, C, E and F fully formed, and strands A, A' and G quasi-unstructured [196], and that seven residues of strands B, C, E and F form a folding nucleus around which the rest of the structure condenses at the early folding stages (shown in red letters in Figure 3.1C).

The folding pathway of ddFLN-dom4 was also investigated using atomic force microscopy (AFM) force-unfolding measurements [61]. This study showed the population of an obligatory intermediate state during re-folding that followed the mechanical unfolding of a ddFLN-dom1-5 chain using the force provided by the AFM tips to stretch the protein [61]. The authors show that the folding of ddFLN-dom4

followed a three-state folding pathway, whereas the other immunoglobulin domains of ddFLN followed a slower, two-state folding pathway. The structural properties of the folding intermediate states were found to be similar to that of the unfolding intermediate state, otherwise characterised [205], in which strands C-G adopt a native-like fold and strands A & B are disordered.

Together, these experimental data on three homologous immunoglobulin domains suggest that most immunoglobulin domains fold cooperatively (with the exception of ddFLN-dom4), with a transition state ensemble in which either strands B, C and F (in the case of fnIII), B, C, E and F (in the case of TI) make native like contacts. ddFLN-dom4 is a fast folding domain that populates an intermediate state during unfolding and refolding. Interestingly, it is thought that the presence of a stable intermediate state fastens the folding rate of ddFLN-dom4 *ca.* 10 times [61] compared to other ddFLN domains which do not populate apparent intermediates. However, the observation of a non-obligatory, "invisible" intermediate state, found to be populated at mid-point folding by *ca.* 20% of ddFLN-dom5 suggests that a there are probably multiple folding pathways that remain to be characterised.

3.1.4 Co-translational folding intermediates

So far, there is no experimental evidence to confirm that *in vitro* folding pathways are sampled during *de novo* co-translational folding. On the contrary, evidence for step-wise folding was observed in the case of the P22 tailspike protein [206], in which co-translational folding intermediate states were recognised by conformation-specific antibody of the N-terminal domain at translation lengths that were aggregating *in vitro*. Similarly, the folding of the N-terminal 190 residues of firely luciferase was found to precede and accelerate the folding of the entire sequence co-translationally [81]. This

step-wise folding was not observed *in vitro* in the presence of chaperones, suggesting that the vectorial emergence of the N-terminal domain is required for the formation of the intermediate state. Early step-wise folding of CFTR-RNC was also revealed by FRET studies, for which a three-state co-translational folding was observed despite its complex topological organisation and high contact number [67]. This experimental evidence suggests that the co-translational folding pathway (and transition state ensemble) is different to the *in vitro* refolding pathway, as presented in Figure 1.3B, and may be associated with the population of intermediate states that are not typically accessible in the refolding pathway.

Computational modelling of the emergence of small protein domains such as protein G or the monomeric λ-repressor has also suggested that, despite the *in vitro* folding being cooperative, transient intermediates that are specific to co-translational folding, may be populated before the emergence of the entire domains from the ribosomal tunnel [48] (Figure 3.3).

To date, there are very few structural studies of co-translational folding. The co-translational folding of ddFLN-dom5 is investigated in this chapter, using the methodology described in Chapter 2.

3.2 Results and Discussion

3.2.1 Hydrogen exchange study of *in vitro* ddFLN-dom5 folding

To determine whether ddFLN-dom5 domain shares the same native thermodynamic equilibrium properties as the immunoglobulin family, the protection factors from ddFLN-dom5 were determined by H-D exchange experiments as followed by NMR. Lyophilised samples of ddFLN-dom5 were dissolved in D_2O, and a series of SOFAST-

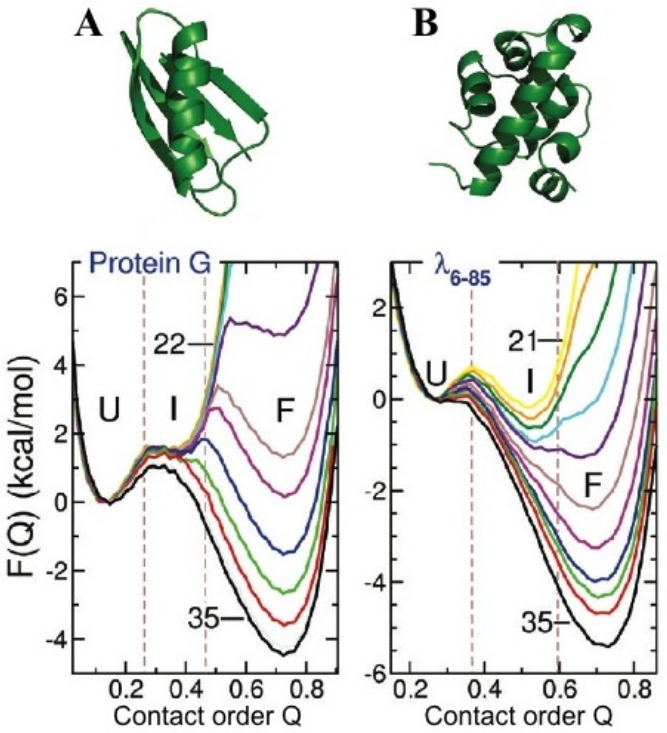


Figure 3.3: Folding free energy landscape with the contact order as the reaction coordinate for two RNCs: Protein G (**A**) and monomeric λ-repressor (**B**). The free energy landscape of the RNC at different stage of translation are shown with different colour: when the domain is attached to the PTC with a linker length of 22 residues in yellow, to a linker length of 35 residues in black. Protein G samples a metastable intermediate, while monomeric λ-repressor populates a stable intermediate at linker lengths of 22-24 residues. Figure adapted from [48].

HMQC were recorded with 5min intervals (2min dead time). 40 residues out of the 69 being analysed were found to have an H-D exchange rate that followed a linear dependancy with pH at pD* 6.15, 6.6, 7.3 and 7.7 (Supplementary Figure A.1), thus confirming that these follow the EX2 regime (Section 3.1.2). Overall, the free energy of opening/closing of the structure was found to be similar within error to the folding free energy obtained from the urea-denaturation of ddFLN-dom5 ($\Delta G_{\mathrm{F-U}} = 7.8 \pm 1.9$

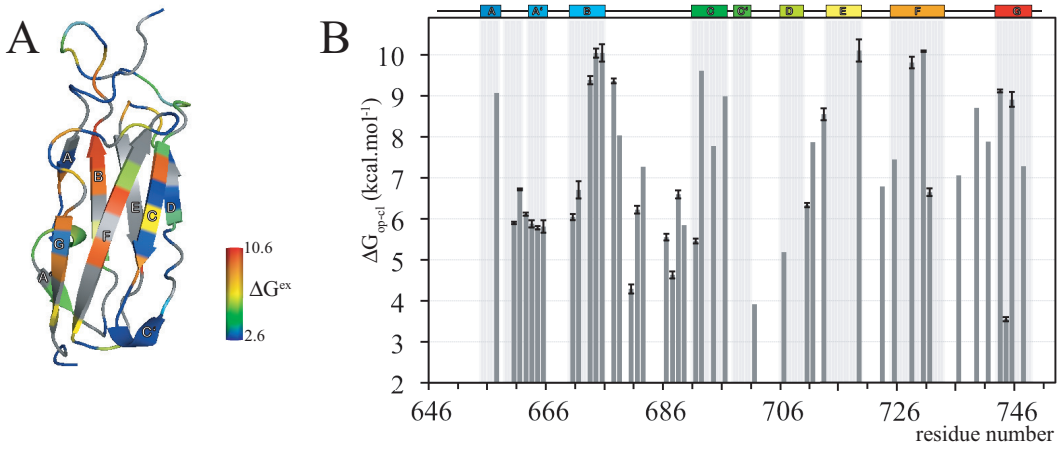


Figure 3.4: A: Structural mapping of the ΔG_{op-cl} of ddFLN-dom5. Residues for which the amide proton exchange within the experiment dead time (3min) have been assigned to a $\Delta G_{op-cl} \sim$ 2.6 kcal.mol^{-1}. Residues that follow the EX1 regime as well as prolines are coloured in grey. **B:** ΔG_{op-cl} of ddFLN-dom5 as a function of residue numbers (the secondary elements are indicated on the top). Only those residues determined to be in the EX2 exchange regime are included.

kcal.mol^{-1}, [183]). Residues of the A, A′, C and G strands were found to have a lower average protection factor ($\langle \Delta G_{op-cl}^{A,A',C,D \text{ and } G} \rangle$=6.0 kcal.mol^{-1}, Figure 3.4) than those of the B & F strands ($\langle \Delta G_{op-cl}^{B\&F} \rangle$=8.8 kcal.mol^{-1}, Figure 3.4), which is consistent with observation from other immunoglobulin-like proteins. Most of the residues of strand E were found to follow the EX1 regime, suggesting a relatively slow "closing" rate ($k_{cl} < k_{int} \approx 0.5 s^{-1}$). Their "opening" rates were also found to be slow ($\langle k_{op} \rangle$=3×10^{-7} s^{-1}), which also suggest a very stable "close" conformation for that central strand, also in line with previous studies, and a high energy barrier between the "open" and "close" conformation. This slow closing rate and slow opening rates was also observed in residues of strand F and one residue of the BC loop. The stability of strand C was, however, altered in ddFLN-dom5 ($\langle \Delta G_{op-cl}^{C,\ ddFLN-dom5} \rangle$< 6.1 kcal.mol^{-1}) compared that of fnIII ($\langle \Delta G_{op-cl}^{C,\ fnIII} \rangle$=7.3 kcal.mol^{-1} [199]).

Together, the protection factors measured for ddFLN-dom5 confirm a very stable β-sandwich, in which strands B, E and F are relatively rigid compared to strands A,C,D & G. The stability of the loops were also found to display interesting dynamical properties; the loops of the C-terminal hemisphere were particularly dynamic ($\langle \Delta G_{\mathrm{op-cl}} \rangle < 3.4$ kcal.mol^{-1}), compared to the loops of the N-terminal hemisphere which were found to be of a higher stability ($\langle \Delta G_{\mathrm{op-cl}}^{\mathrm{BC\ loop}} \rangle > 5.7$ kcal.mol^{-1}). The relative dynamic differences between the N & C terminal hemispheres is of interest in the context of co-translational folding. During vectorial emergence, the N-terminus emerges first and would in principle have the capacity to form stable structure prior to the emergence of the C-terminal hemisphere.

3.2.2 Characterising the co-translational folding properties of ddFLN-dom5 using NMR spectroscopy

The folding equilibrium of the ddFLN-dom5-RNC at different stages of emergence from the ribosomal exit tunnel was investigated using NMR. Samples of stalled RNCs of ddFLN-dom5 with increasing lengths of ddFLN-dom6 were produced as in Figure 3.5. The various lengths of ddFLN-dom6 were designed to behave as a disordered linker that positions the preceding domain at the exit port of the ribosomal exit tunnel, thus providing snapshots for the emergence of ddFLN-dom5 from the exit tunnel (Section 2.3.1.1). As mentioned in Section 1.1.4 the ribosomal exit tunnel has the capacity to hold between 24 and 40 residues of the emerging NC. RNCs containing linker lengths >47 residues are thought to have the immunoglobulin domain fully exposed and available to fold, while linker lengths of 21-47 may display different capacities to adopt structures due to their proximity to the ribosome exit tunnel. ddFLN-dom5+21-RNC is thought to have the last G-strand of ddFLN-dom5 partially sequestered within the ribosome tunnel,

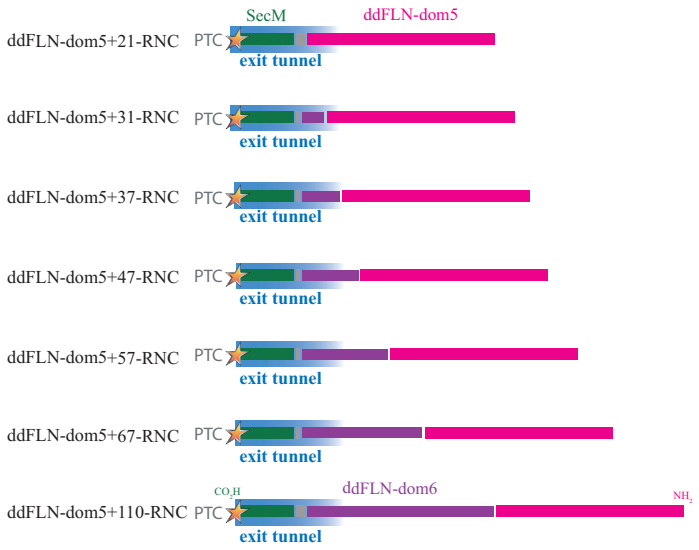


Figure 3.5: Schematic of the ddFLN-dom5-RNC constructs as in Figure 2.8, with the residues that are expected to span through the 100Å long exit tunnel shaded in blue.

and constructs ddFLN-dom5+31 & +37-RNC have the immunoglobulin domain either fully exposed or partially sequestered. To study the progressive emergence and folding of ddFLN-dom5, isotopically labelled samples of each RNC were prepared and analysed by NMR using the methodology as described in Chapter 2, Section 2.3.2.

3.2.2.1 Analysis of ^{1}H-^{13}C methyl-TROSY HMQC of RNC

As described in Chapter 2, Section 2.3.2.6, high levels of perdeuteration within the ribosomal complex, and selective ^{13}C labelling of the NC Ileδ1 methyl groups labelling combined with methyl-TROSY HMQC methods provides a sensitive means of assessing folded structure in RNCs. Samples of uniform-[^{12}C,^{2}H], Ileδ1-[$^{13}C^{1}H_3$] labelled ddFLN-dom5+110, +47 and +37-RNCs were therefore produced. Ile residues were chosen because of the high dispersion of their chemical shift, their extended sidechains that provide additional dynamics to the methyl group, and their distribution throughout the

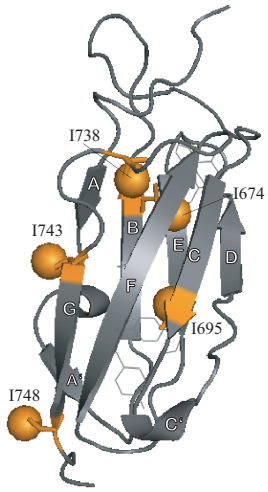


Figure 3.6: Methyl groups of the Ile mapped on the structure of ddFLN-dom5.

structure (Figure 3.6) with Ile residues found within the hydrophobic core of the protein (I674, I695 & I738), and also at the surface (I743 & I748); covering strands B (I674), C (I695) and G (I743 & I748). Ile residues are present in the N-terminal hemisphere (I738) and in the C-terminal hemisphere (I748). Together, these Ile residues are a very sensitive set of reporters of native structure.

The ^{1}H-^{13}C methyl-TROSY HMQC spectrum of isolated ddFLN-dom5+47-NC (Figure 3.7A) revealed three high-field shifted resonances assigned to I695, I674 and I738 [185], while the two solvent-exposed Ile of strand G (I743 & I748) were found to have a lower ^{1}H chemical shift dispersion. Four additional resonances were observed at chemical shifts characteristic of random coil conformations and these were assigned to the two Ile of the ddFLN-dom6 truncation and the two Ile of the SecM sequence. The ^{1}H-^{13}C methyl-TROSY HMQC spectrum of the isolated ddFLN-dom5 was used to assess spectra obtained for each of the RNC lengths.

The ^{1}H-^{13}C methyl-TROSY HMQC spectrum of ddFLN-dom5+110-RNC contains

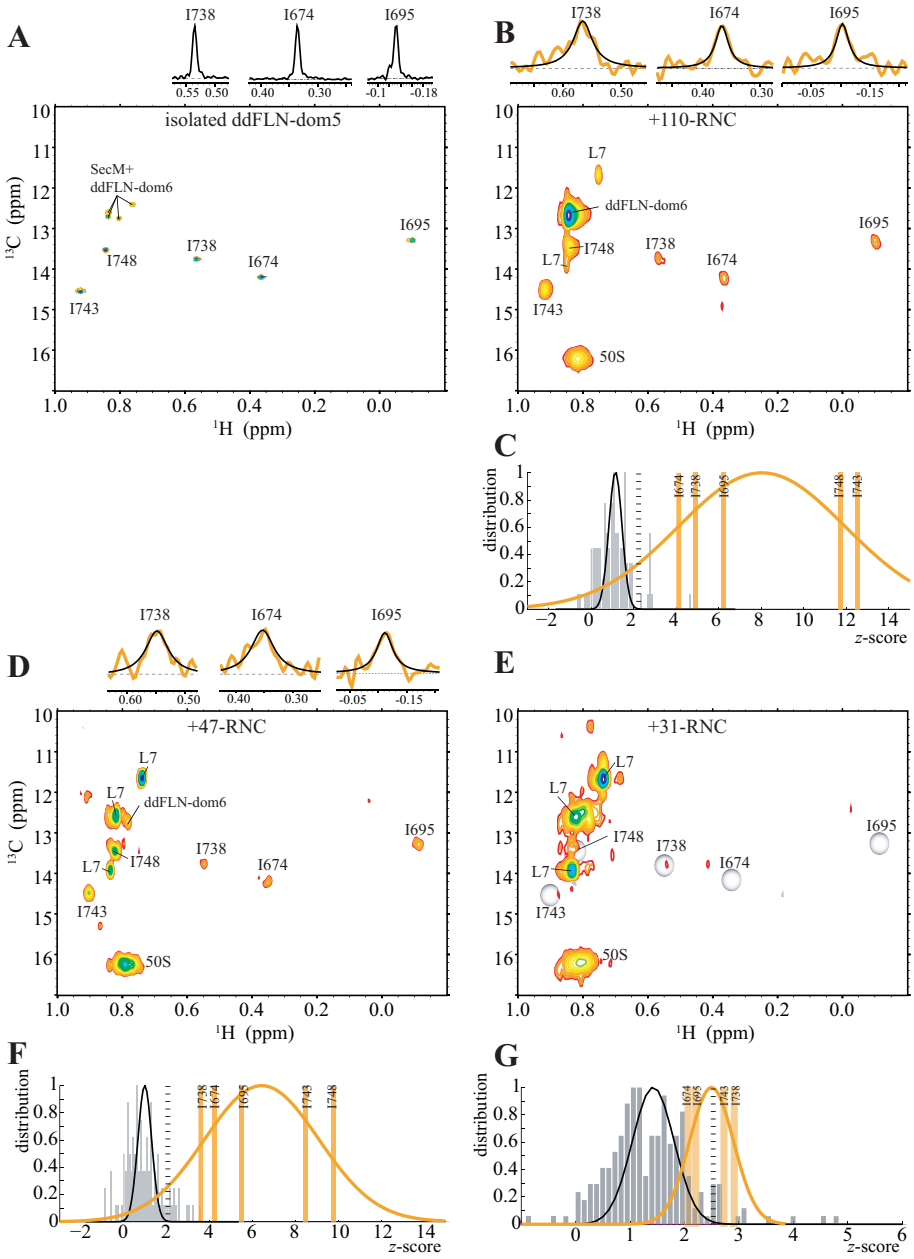


Figure 3.7: ^{1}H-^{13}C HMQC spectra of: **A:** purified released U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+47-NC, **B:** U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+110-RNC (A & B are the same spectra as shown in Figure 2.21), **D:** U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+47-RNC and **E:** U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+37-RNC. Proton cross-sections of the signals are shown next to the assigned resonances. The intensity distribution for the five ddFLN-dom5 Isoleucine resonances are shown in **C, F & G** for ddFLN-dom5+110, +47 and +37-RNC respectively.

signals from each of the five ddFLN-dom5 Ile native resonances (Figure 3.7B), with a signal to noise ratio (SN or *z*-score, Section 2.2) that allows for their observation with confidence (Figure 3.7C), indicating that the nascent chain is folded. The chemical shift of the observed ddFLN-dom5 resonances were identical to that of the released chain, within 0.01 ^{1}H ppm and 0.08 ^{13}C ppm. Additional resonances were also observed, that can be assigned to the ddFLN-dom6 linker, L7/L12 stalk and resonances from the 50S subunit.

The dispersed ddFLN-dom5 Ile resonances were also observed in the ^{1}H-^{13}C methyl-TROSY HMQC spectrum of ddFLN-dom5+47-RNC (Figure 3.7D), with identical chemical shift within 0.01 ^{1}H ppm and 0.1 ^{13}C ppm, and a sufficient SN to confirm their significance (Figure 3.7F). Similar to ddFLN-dom5+110-RNC spectrum, resonances from the L7/L12 stalk and the 50S subunit were also observed, but the cluster of resonances assigned to the ddFLN-dom6 linker in the spectrum of ddFLN-dom5+110-RNC was significantly reduced in the case of ddFLN-dom5+47-RNC, which is consistent with the shorter ddFLN-dom6 sequence associated with the shorter RNC.

The spectrum of sample ddFLN-dom5+37-RNC contained intense resonances from the L7/L12 stalk and 50S subunit, and the dispersed folded ddFLN-dom5 were associated with a significantly lower SN (Figure 3.7E & G). Only two of the five resonances were associated with a SN that was significantly higher than the noise (resonances for I738 could not be analysed due to its overlap with L7/L12 resonances). The distribution of the ddFLN-dom5 resonances intensities is nevertheless higher than the noise distribution. Any potential new resonances had a SN that was lower than the noise threshold allowing a search in the entire spectrum (*z*-score=2.8, Section 2.2). Before analysing the characteristic of each residues of the folded state of ddFLN-dom5-RNCs as observed in these spectra, the attachment of the NC over the time of acquisition of the

^{1}H-^{13}C methyl-TROSY HMQC was analysed.

Figure 3.8A, B & C shows the intensity of the five Ile resonances of ddFLN-dom5 in the ^{1}H-^{13}C HMQC spectra recorded over a time-course of 50 to 100 hours. Each spectrum is recorded for 3 hours, followed by the acquisition of ^{1}H 1D spectrum and ^{1}H STE diffusion spectra. The intensities of the ddFLN-dom5 resonances remained stable for 40 hours in the case of ddFLN-dom5+110-RNC, for 60 hours in the case of ddFLN-dom5+47-RNC, but increased rapidly after only 6 hours in the case of ddFLN-dom5+37-RNC. Figure 3.9 shows the spectrum recorded within the first 12 hours of acquisition of ddFLN-dom5+37-RNC, compared to that recorded from 38-48 hours, and the spectrum recorded after 38 hours reveals the appearance of sharp ddFLN-dom5 Ile resonances. These have linewidths which are significantly sharper than the linewidths of the ddFLN-dom5+110-RNC resonances (>13Hz, Section 2.3.2.6), and are indeed similar to those observed for released protein (<10Hz), indicating that these resonances do not arise from ribosome-attached NC but rather from released protein.

The attachment of the NC was also examined by the ^{13}C-edited STE-HMQC diffusion experiment (Section 2.3.2.8), recorded in an interleaved manner with the heteronuclear spectra. Three gradients strength were used: 2.5, 25 and 47.5 G.cm^{-1}, and the decay of intensity from the spectrum recorded with a low gradient strength compared to that recored with a high gradient strength gave an estimation of the translational diffusion coefficient of the NC (cyan vs black curves in Figure 3.8D, E & F). A decay of *ca.* 30% is indicative that the NC has a diffusion coefficient equal to that of the ribosome, whereas greater attenuations are associated with the release of the NC (Section 2.3.2.8). The decay of intensity associated with the unfolded methyl signals (^{1}H ppm 0.6-1.0) of sample ddFLN-dom5+110-RNC is similar to what is expected for a ribosomal particle, and remains stable for *ca.* 40 hours. Moreover, the decay of intensity associated with

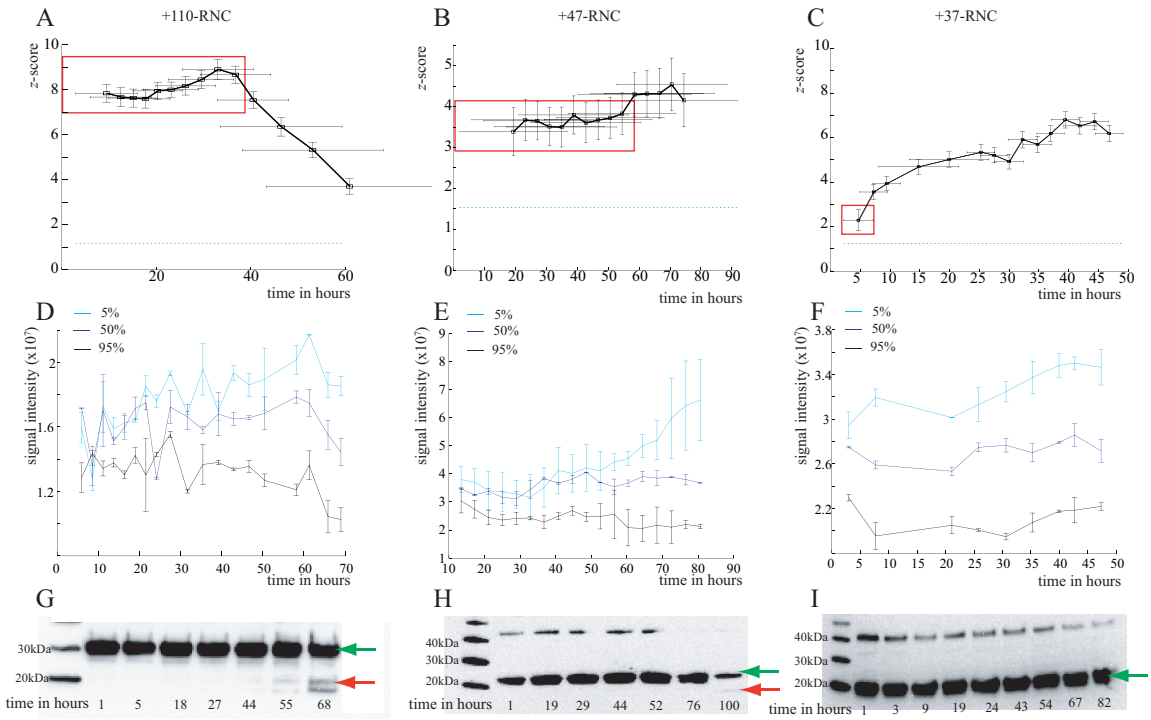


Figure 3.8: A, B & C: Mean z-score of the intensity of five ddFLN-dom5 Ile resonances (Section 2.3.2.5) in ddFLN-dom5+110, +47 and +37-RNC respectively, followed over-time (see also Figure 2.31). The intensities shown here are intensity measured on the sum of five spectra in the case of ddFLN-dom5+110-RNC, ten spectra in the case of ddFLN-dom5+47-RNC and 3 spectra in the case of ddFLN-dom5+37-RNC. The vertical error-bars show the standard error of the mean, and the horizontal error-bars correspond to the time of acquisition for each spectrum. The green dotted lines show the 5% threshold corresponding to the limit of observability for the mean of five resonances (Section 2.2). The red rectangle shows the data that are summed for the intensity analysis and linewidth analysis shown in Section 3.2.2.1. **D, E & F:** Signal intensity of the ^{13}C-edited STE-HMQC diffusion spectra, recorded with gradient strengths of g=2.5 (cyan), 25 (blue) and 47.5 (black) G.cm^{-1}, for ddFLN-dom5+110, +47 and +37-RNC respectively. **G, H & I** Anti-His western of ddFLN-dom5+110, +47 and +37-RNC respectively, during the timecourse that covers the NMR acquisition. The red arrow shows the 14kDa product of the proteolysis and the green arrow shows the full-length NC.

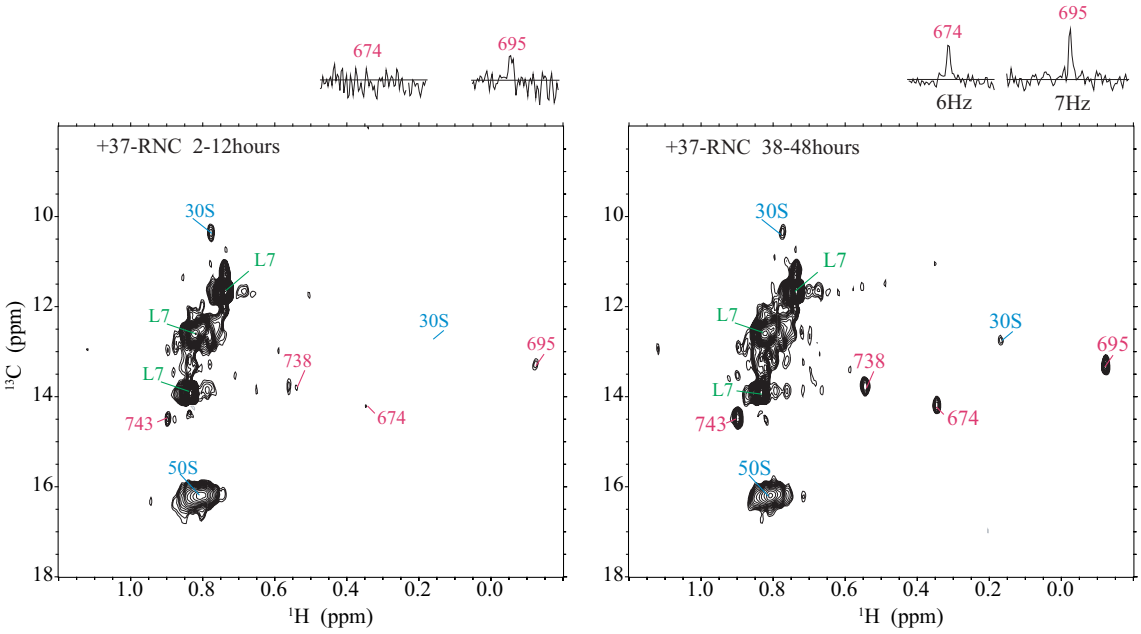


Figure 3.9: Ile-methyl-TROSY HMQC spectra of U-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] ddFLN-dom5+37-RNC. Left: spectrum recorded within the first 12 hours of acquisition. Right: spectrum recorded after 38 hours of acqusition. Both spectra are recorded with the same number of scans, and the 1D ^{1}H traces at the ^{13}C chemical shift of Ile 674 and Ile 695 are shown above the spectra. The linewidth of the appearing ddFLN-dom5 resonances are on the order of 6-7Hz, similar to the linewidth of the isolated protein signals.

the dispersed resonance of I695 could be measured directly, and was found to be on the order of that expected from a ribosome particle (data shown in Section 2.3.2.8). The intensities of unfolded methyl signals of both samples ddFLN-dom5+47 & +37-RNCs in the ^{13}C-edited STE-HMQC diffusion spectra also indicated that the diffusion coefficient of those resonances are on the order of that of a ribosome particle, and remained stable for *ca.* 55 hours (ddFLN-dom5+47-RNC) and 40 hours (ddFLN-dom5+110-RNC). In the heteronuclear spectrum of ddFLN-dom5+37-RNC however, most of the signals observed could be attributed to the L7/L12 stalk as the overall signal for the NC was weak in the NMR data accumulated over the period of the time analysed (Figure 3.7E). The ^{13}C-

edited STE-HMQC diffusion experiment therefore mainly reported on the attachment of those ribosome signals, and not the attachment of the NC.

The anti-His western blotting of samples taken over the time of NMR acquisition did not show significant levels of proteolysis, indicating this was not the main factors in limiting the sample lifetime. However, due to the high deuteration level, it was not possible to probe for a tryptophan resonance which would be used to monitor a SecM base release (Section 2.3.3.3).

The conservative approach to RNC analysis detailed in Chapter 2 stipulates that the null hypothesis is that the observed signals arise from released NC unless strong evidence indicates the attachment of the NC to the ribosome. In the case of ddFLN-dom5+110 & +47-RNCs, the observation of RNC signals in the ^{1}H-^{13}C methyl-TROSY HMQC spectra allowed the attribution of the slow diffusing signals in the ^{1}H-^{13}C-edited STE diffusion spectra to the folded NC, which indicated that the NC remains attached for 40-60 hours respectively. However, data from ddFLN-dom5+37-RNC were more ambiguous. The weak intensity of the folded NC resonances in the ^{1}H-^{13}C methyl-TROSY HMQC meant that these were not observed in the ^{1}H-^{13}C-edited STE diffusion spectra, and there is therefore no evidence that the signals observed within the first six hours of acquisition are those of the attached NC – particularly as subsequent spectra showed the appearance of released NC. Nevertheless, it is clear that data recorded from ddFLN-dom5+110 & +47-RNCs within 40-60 hours are from attached NC, and the spectra are therefore suitable for further analysis.

In particular, the ^{1}H linewidths ($\Delta\nu_{^{1}\mathrm{H}}$) of each resonances could be measured, which relate to the dynamics of the residues. A Lorentzian function was used to fit each signal in spectra of isolated ddFLN-dom5, ddFLN-dom+110 & +47-RNCs, and both the fitted lineshape and the raw spectrum are shown above the spectra in Figure 3.7.

Resonance	SN in +110-RNC	SN in +47-RNC	$\Delta\nu_{^1H}$ (Hz) in +110-RNC	$\Delta\nu_{^1H}$ (Hz) in +47-RNC	$\Delta\nu_{^1H}$(Hz) in isolated ddFLN-dom5
748	11.8	9.8	16±1	19±2	<5
743	12.5	8.7	14±1	15±2	<6
738	4.9	3.8	28±4	31±5	<6
695	6.4	5.5	19±3	24±3	<6
674	4.3	4.3	19±4	34±4	<6
L7	15.3	37	5.1±1.1	4.7±0.5	

Table 3.1: Signal to noise and linewidth of the dispersed, native-like ddLFN-dom5 Ileδ1 in ddFLN-dom5+110 and ddFLN-dom5+47-RNC. The linewidth uncertainty are calculated with equation 1.12.

The linewidths of isolated ddFLN-dom5 are sharper than 6Hz (Table 3.1) – limited by the proton acquisition time. Similar values for the linewidths of the L7 resonances were found in the spectra of ddFLN-dom+110 & +47-RNCs (5.1±1.1 & 4.7±0.5 Hz), indicating that the dynamics of the ribosome-bound L7 is similar to that of an isolated domain (in line with previous NMR studies of the L7/L12 [33]). The linewidths of the folded ddLFN-dom5 Ileδ1 resonances in ^{1}H-^{13}C methyl-TROSY HMQC spectra of ddFLN-dom5+110 & +47-RNCs were found to be broader than those of L7: *ca.* 20Hz on average, with a range of 14-28Hz (Table 3.1), indicating that the RNCs are less mobile than the L7 stalk. These linewidths were, however, sharper than the linewidth expected from the ribosome complex, estimated to be *ca.* 190Hz (Section 1.3.2, assuming S^2_{axis}=0.5), indicating that the NC does experience independent motions from the ribosome core, attributable to the flexible ddFLN-dom6 linker.

The linewidths of each of the dispersed resonances of ddFLN-dom5+110-RNC are mapped on the ddFLN-dom5 structure in Figure 3.10A. Most were nearly identical within error (16-19 Hz with uncertainties of 1-3Hz), with the exception of I738 (24±4Hz), which is in the F-G loop in the N-terminal hemisphere and the sidechain of which is

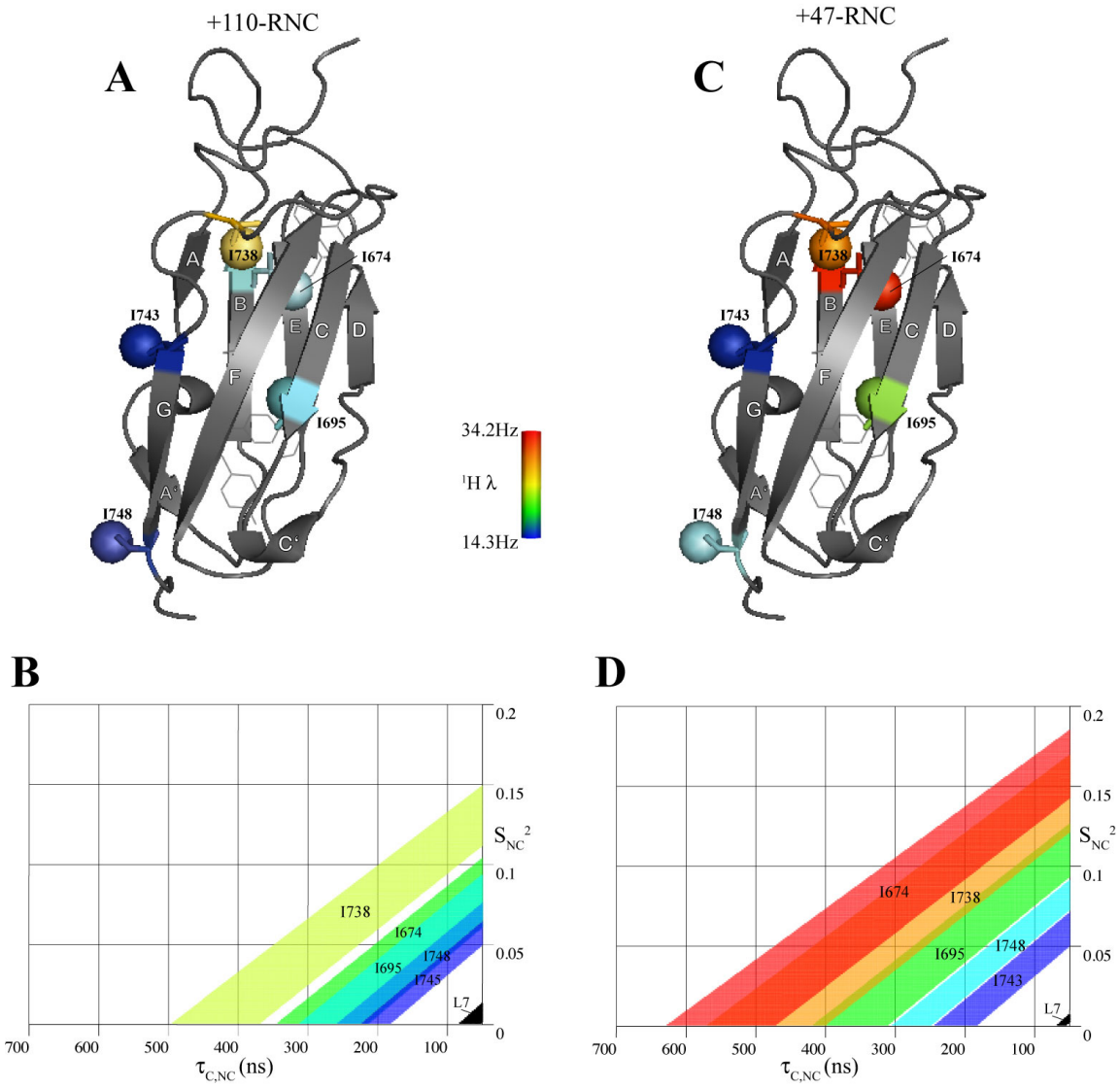


Figure 3.10: Structural mapping of the Ile $\Delta\nu_{1H}$ of ddFLN-dom5+110-RNC in **A** and ddFLN-dom5+47-RNC in **C**. **B & D:** Map of the allowed $\tau_{c,NC}$ and S^2_{NC} that satisfy the experimental linewidth (shown in the same colour coding as in A and C respectively) considering $\tau_{c,ribo}$=2500ns and S^2_{axis}=0.5 for all Ile residues. The regions shaded indicates the quoted uncertainties in Table 3.1. The L7 linewidth are shown in black. Refer to Chapter 1 Section 1.3.2 for details about the plot.

within the hydrophobic core of the protein facing strand B. As the linewidth is directly correlated to the amplitude of the rotation of the methyl group (S^2_{axis}, Section 1.3.2), restricted rotation of the methyl group could explain the broader linewidth of I738 in the ddFLN-dom5+110-RNC spectrum, possibly due to the buried position with which I738 is present within the structure.

To model the overall flexibility of the RNC provided by the flexible linker, the model-free parameters that characterise the motion of the NC relatively to the ribosome (its correlation time $\tau_{c,NC}$ and order parameter S^2_{NC}) were derived from the linewidths using equations 1.11 & 1.8 (Chapter 1, Section 1.3.2). The allowed values of $\tau_{c,NC}$ and S^2_{NC}, given the observed linewidths and $\tau_{c,ribo}$=2500 ns are shown in figure 3.10B (see also Figure 1.12B). The correlation time associated with the motions of the NC was at most 300 ns, assuming a very low order parameter (i.e. the amplitude of the motion not limited). If the order parameter of the NC is larger (at most S^2_{NC}=0.1), the correlation time of the motion will be lower ($\tau_{c,NC} \sim 50$ ns) to satisfy the observed linewidth. By comparison, the motion associated to the L7 linewidth were found to have a correlation time $\tau_{c,L7} \leq 60$ ns and $S^2_{NC} \leq 0.01$ (black lines in Figure 3.10B & D). The dynamics of L7 have also been measured by ^{1}H-^{15}N relaxation, which showed a correlation time $\tau_{c,L7} \sim$ 14ns [33], suggesting that motions of the C-terminal end of L7 are beyond the validity of the model used here which was derived for $\tau_{c,L7} \geq 50$ns (Section 1.3.2).

The linewidths of the RNC resonances indicate that, overall, the natively folded ddFLN-dom5 in ddFLN-dom5+110-RNC tumbles independently of the ribosome, although with a longer tumbling time as compared to the C-terminal domain of the L7/L12 stalk. This is also supported by the low maximum order parameter values (allowed if $\tau_{c,NC}$ is close to 50 ns) which indicates that the amplitude of the motions experienced by the nascent chain is large, and the NC is not in a confined space.

In order to understand the overall motion of the folded RNC in ddFLN-dom5+47-RNC, the linewidths associated with the folded resonances of ddFLN-dom5+47-RNC ($\Delta\nu_{^1\mathrm{H}}$=15.0-34.2Hz) were analysed by comparison with their linewidths in the ddFLN-dom5+110-RNC spectrum. Resonances I743 and I748 were found to have the same linewidths in both the ddFLN+47-RNC spectrum (15$\pm$2 & 19$\pm$2 Hz) and ddFLN-dom5+110-RNC spectrum (14$\pm$1 & 16$\pm$1 Hz), while other resonances were found to be broader in ddFLN-dom5+47-RNC spectrum ($\Delta\nu_{^1\mathrm{H}}$=14-28Hz). This suggests that the broad linewidths for residues I695, I738 and I674 are not due to the limited tumbling due to the shorter linker. The persistence length of a disordered peptide is known to be on the order of seven residues [171], i.e. within a disordered linker, the $i+8^{\mathrm{th}}$ residue tumbles independently of the i^{th} residue. The C-terminal end of ddFLN-dom5 is expected to be between 7 to 23 residues away from the exit port of the ribosome exit tunnel in ddFLN-dom5+47-RNC (assuming that 24-40 residues occupy the ribosomal exit tunnel, Section 3.2.2), and given that the truncated ddFLN-dom6 sequence results in a disordered linker, and a persistence length of seven residues, it is expected that this linker length provides enough flexibility for the RNC to tumble independently of the ribosome. This independent tumbling was confirmed by the linewidth of the folded I743 and I748 resonances in the ddFLN-dom5+47-RNC spectrum, these linewidths were associated with an overall motion of the NC that has a maximum correlation time of $\tau_{c,\mathrm{NC}} = 300$ ns, and a maximum order parameter of $S^2_{\mathrm{NC}} = 0.1$ (Figure 3.10D), and indicates that the overall motion experienced by the NC in ddFLN-dom5+47-RNC is of similar timescale and amplitude as in ddFLN-dom5+110-RNC. Together, these results suggest that the ddFLN-dom6 linker provides the required flexibility to allow ddFLN-dom5 to sample a large volume in both RNCs, and provide the NC with a large conformational entropy.

Nonetheless, the analysis of the linewidths of the folded ddFLN-dom5 resonances

in ddFLN-dom5+47-RNC were found to be very heterogeneous, from 34±4 Hz (I674 and I738) to 15±2 Hz (I748 and I743). This heterogeneity is significantly greater than the experimental uncertainty, and might indicate internal dynamics in addition to the overall motion of the entire domain (analysed above). This differential linewidth seems to be greater than that observed for ddFLN-dom5+110-RNC, which suggest that this is beyond the differential dynamics of the side chains (i.e. differential S^2_{axis}). The solvent exposed Ile methyl groups (Ile 748 and Ile 743) were found to have a sharper linewidth (Figure 3.10) compared to those packed in the hydrophobic core of the domain (Ile 674, Ile 674 and Ile 738). In addition, the Ile methyl groups of the N-terminal hemisphere (Ile 674, Ile 738) were found to be the most broadened, in particular residues I674 of strand B, which was found to be very stable in the native state of the protein by H-D exchange. This preliminary observation of differential dynamics may be indicative of chemical exchange with a non-native intermediate state, or there may be additional anisotropic dynamics associated with individual loops or side-chains (Section 1.3.2). A knowledge of proton linewidths alone does not permit the accurate determination of local anisotropy or chemical exchange at this intermediate linker length, and further experimental data will be required to provide additional information. Another possibility for the differential linewidths observed within ddFLN-dom5+47-RNC may be a transient interaction of ddFLN-dom5 with the ribosome [130]. This would result in the broadening of the residues of the N-terminal hemisphere as in a chemical exchange situation, and would also reduce significantly the tumbling of the NC (i.e. increase $\tau_{c,NC}$ and S^2_{NC}).

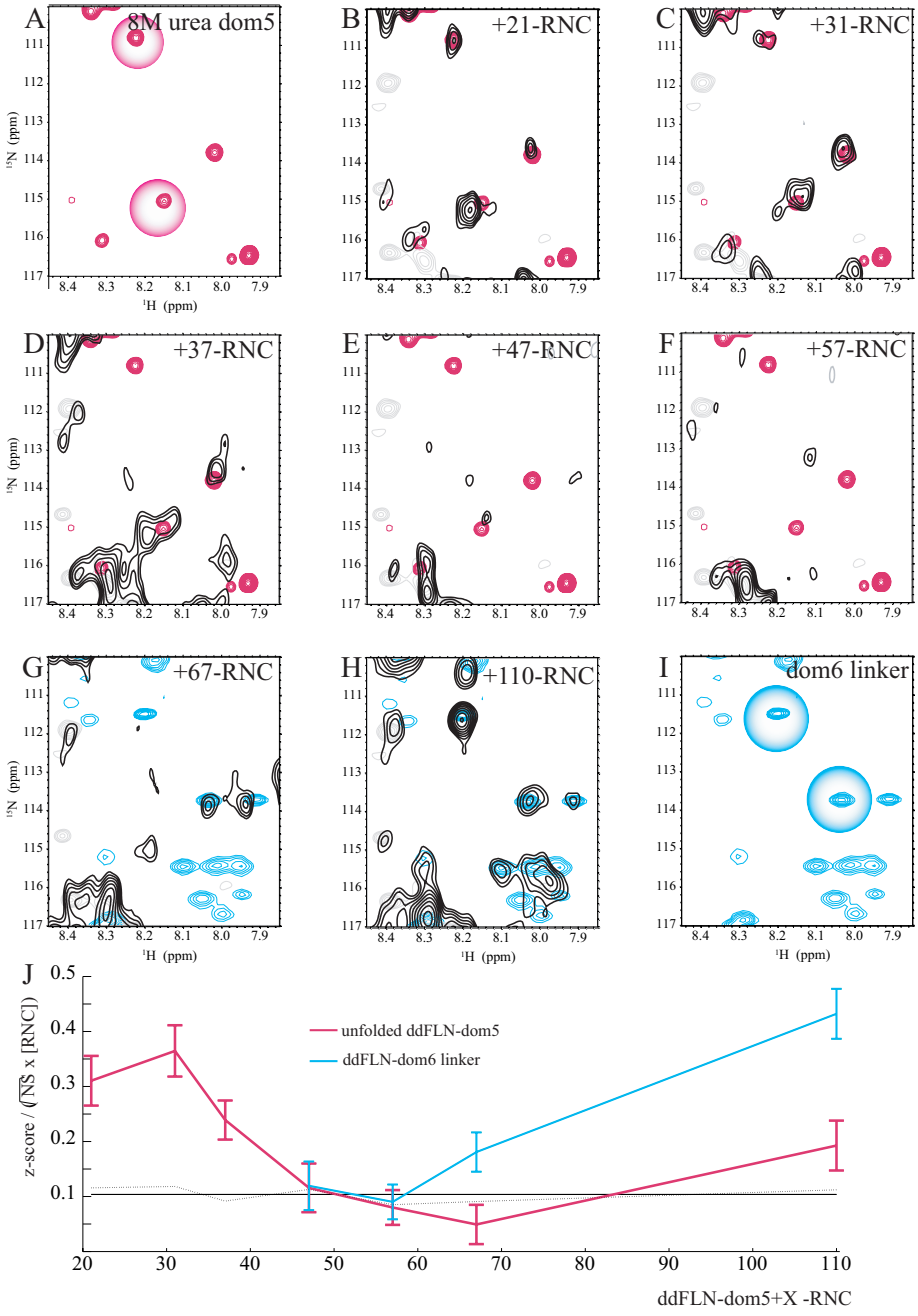
In summary, while it is clear for the ddFLN-dom5+47-RNC that the immunoglobulin domain is fully emerged and folded, the proton linewidths reveal the existence of non-native dynamics. This may indicate conformational exchange or transient interactions with the ribosome surface. These observations provide a strong foundation for future

work to analyse these effects further.

3.2.2.2 *Analysis of the disordered regions of ddFLN-dom5-RNCs in the ^{1}H-^{15}N SOFAST-HMQC*

In order to analyse further the folding equilibrium of ddFLN-dom5 as a function of ddFLN-dom6 translation length, the unfolded states of ddFLN-dom5 were examided using the ^{1}H-^{15}N SOFAST-HMQC of each sample as described in Section 2.3.2.3 and compared to the ^{1}H-^{15}N HSQC spectrum of 8M urea denatured ddFLN-dom5. The ^{1}H-^{15}N HSQC spectrum of unfolded ddFLN-dom5 contains several discrete resonances – in particular, resonances G716 and T679, which do not overlay with either L7/L12, folded ddFLN-dom5 or disordered ddFLN-dom6 resonances. These resonances were therefore used to monitor the unfolded state of ddFLN-dom5 at each linker length (Figure 3.11, magenta circles). Unlike the analysis of the folded resonances, the assessment of whether

Figure 3.11 ***(following page)*****:** ^{1}H-^{15}N SOFAST-HMQC spectra of disordered ddFLN-dom5 in magenta, truncation of ddFLN-dom6 in cyan, 70S in grey, and ddFLN-dom5-RNCs in black. **A**: ^{1}H-^{15}N SOFAST-HMQC spectra of 8 M urea denatured ddFLN-dom5 [185]. **B**: ddFLN-dom5+21-RNC, **C**: ddFLN-dom5+31-RNC, **D**: ddFLN-dom5+37-RNC, **E**: ddFLN-dom5+47-RNC, **F**: ddFLN-dom5+57-RNC, **G**: ddFLN-dom5+67-RNC and **H**: ddFLN-dom5+110-RNC, **I**: Truncated ddFLN-dom6$_{751-840}$. The magenta circle in A shows resonances G716 and T714 of unfolded ddFLN-dom5 and the cyan circles in G shows two resonances of disordered ddLFN-dom6. **J**: Normalised *z*-score (SN) of the mean intensity for resonances G716 and T714 of unfolded ddFLN-dom5 in magenta and disordered ddFLN-dom6 in cyan as a function of linker length. The *z*-score is normalised for the 70S concentration and the square root of the number of scans: each spectra was recorded with 352 scans, and the intensities shown here correspond to the intensity in the sum of 5 spectra for ddFLN-dom5+21-RNC (7 μM), 10 spectra for ddFLN-dom5+31-RNC (4.8 μM), 11 spectra for ddFLN-dom5+37-RNC (7 μM), 4 spectra for ddFLN-dom5+47-RNC (7 μM), 5 spectra for ddFLN-dom5+57-RNC (10 μM), 8 spectra for ddFLN-dom5+67-RNC (7 μM) and 3 spectra for ddFLN-dom5+110-RNC (9 μM). The error bars are the standard deviation of the noise of each spectrum divided by $\sqrt{2} \times \sqrt{\text{NS}} \times$[RNC]. The grey doted line shows the 5% threshold of the noise of each spectrum, and the black lines shows the average threshold.


A
8M urea dom5
B
+21-RNC
C
+31-RNC
D
+37-RNC
E
+47-RNC
F
+57-RNC
G
+67-RNC
H
+110-RNC
I
dom6 linker
^{15}N (ppm)
^{1}H (ppm)
J
unfolded ddFLN-dom5
ddFLN-dom6 linker
z-score / (√NS x [RNC])
ddFLN-dom5+X -RNC

the unfolded resonances arose from attached RNC was easier to monitor. This was because the unfolded resonances are distinct from the folded resonances observed from the released chain that was found to be able to fold as seen in Section 2.3.3.3; and the high intensity of the unfolded resonances that allows recording the ^{1}H-^{15}N SOFAST-HMQC in a shorter timescale (compared to ^{1}H-^{13}C methyl-TROSY HMQC, *ca.* 6-12 hours); as well as less ambiguity in the ^{15}N XSTE diffusion experiment in which the dispersed L7/L12 amide resonances are very distinct from the unfolded ddFLN-dom5 resonances.

Resonances from disordered ddFLN-dom5 could be observed in the ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+21, +31 and +37-RNC (Figure 3.11A, B and C respectively), indicating that at these linker lengths ddFLN-dom5 populates the disordered state. The intensity of the disordered resonances remained constant (within error) from linker lengths of +21 to +31 residues. The ratio of intensities of unfolded ddFLN-dom5 resonances in the ddFLN-dom5+31-RNC compared to those in the +21-RNC spectra is approximately constant throughout the ddFLN-dom5 sequence (Figure 3.12), with the exception of five resonances that had a higher intensity in the ddFLN-dom5+31-RNC spectrum, indicating that the unfolded state in both lengths sample similar conformations. The intensities of the unfolded resonances decrease however as the linker length increases from +31 to +47 residues. At a linker length of +37 residues, the intensities of the unfolded resonances were significantly lower than at linker lengths of +21 or +31 residues, despite the longer linker providing additional flexibility to the domain. This lower intensity was not associated with a lower occupancy of the NC as controlled by the anti-His western, but instead relates to the intrinsic properties of the ddFLN-dom5+37-RNC. As mentioned in Section 3.2.2, at linker length of +37, ddFLN-dom5+37-RNC is expected to be either fully exposed or at most with 3 residues within the exit tunnel. Although resonances from the folded state could

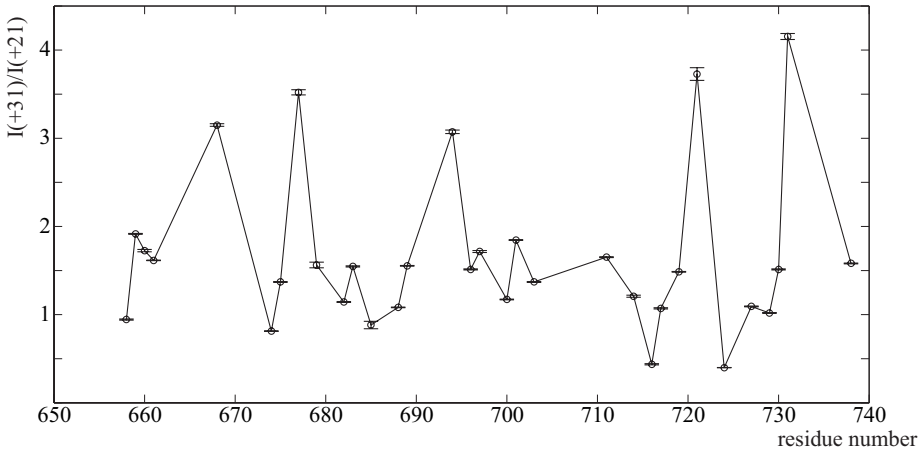


Figure 3.12: Intensities for the unfolded resonances in ddFLN-dom5+31-RNC spectrum divided by that in ddFLN-dom5+21-RNC spectra, normalised for the number of scans and RNC concentration, as plotted versus the residue number.

not be analysed in details in the ^{1}H-^{13}C methyl-TROSY HMQC of ddFLN-dom5+37-RNC (due to early release of the NC), one cannot exclude that a low population of a folded state is populated, giving rise to weak signals in the ^{1}H-^{13}C spectrum (Figure 3.7G). The decrease in the intensity of the unfolded ddFLN-dom5 resonances together with the absence of discernible "folded" resonances suggest that perhaps the currently observed unfolded state is in a slow exchange with a "NMR-invisible" folded state, and therefore only a subset of the domains populates the unfolded state; this would explain the lower intensity for those resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+31-RNC. The presence of a folded state (with perhaps non-native structure or dynamics) that is NMR-invisible is also supported by the analysis of ddFLN-dom5+47-RNC spectra. At a linker length of +47 residues, "folded" but severely broadened resonances were observed in the ^{1}H-^{13}C methyl-TROSY HMQC. It is reasonable to expect such broadening to be even more significant at shorter linker lengths, thus rationalising the absence of observed signals.

In ^{1}H-^{15}N SOFAST-HMQC spectra of ddFLN-dom5+57 to +67-RNCs, resonances from disordered ddFLN-dom5 were not observed, despite the linker providing enough flexibility, indicating that at these linker lengths the unfolded state was not populated (as was the case for ddFLN-dom5+47-RNC). Moreover, the appearance of weak resonances assigned to unfolded ddFLN-dom6 (Figure 3.11G, H & J) at linker length $\geq$ 67 residues indicates that at these lengths parts of the unfolded linker have emerged from the ribosome exit tunnel with enough flexibility to allow NMR observation. Moreover, in the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+110 resonances assigned to disordered ddFLN-dom6 were observed with higher intensities compared to ddFLN-dom5+67-RNC, which is consistent with the greater exposure of the long linker. This also indicates that the ddFLN-dom6 linker remains disordered at all lengths of this study. Close inspection of weak unfolded ddFLN-dom5 resonances in ddFLN-dom5+110-RNC spectrum revealed that these most likely reflected an artefact of the intense ddFLN-dom6 resonances that partially overlap the ddFLN-dom5 resonances; this is also supported by the analysis of the ^{1}H-^{13}C HMQC of the same sample that revealed that ddFLN-dom5 adopts the native conformation, signifying the essentially complete folding of the ddFLN-dom5 domain.

Together, the analysis of the unfolded resonances in the ^{1}H-^{15}N SOFAST-HMQC spectra of RNCs with increasing linker lengths indicates that the unfolded state is populated for linker length $\leq$ 37 residues, and not populated at linker length $\geq$ 47 residues. At linker length of +37 residues, the low intensity of the unfolded resonances might also be indicative of a equilibrium between an observed unfolded state and a yet-to-be-analysed partially folded state, in slow exchange compared to the NMR timescale (*ca.* 100ms). One can not exclude such an equilibrium for shorter lengths, although the greater intensities for the unfolded resonances suggests that at shorter lengths, the

equilibrium is favouring the unfolded state.

3.2.3 Analysis of the differential resistance to proteolysis and the implication for co-translational folding

As described in Chapter 2, Section 2.3.3, a number of NMR and biochemical strategies were developed to assess the integrity of the RNC over time. Section 2.3.3.4 presented evidence that for ddFLN-dom5+110-RNC, proteolysis can occur at the ddFLN-dom5 linker, resulting in the release of the folded immunoglobulin domain and giving rise to sharp resonances in the heteronuclear spectra (Figure 2.29 showed the spectrum before proteolysis and the spectrum after proteolysis has occurred within 24 hours). A similar increase in the intensity of native-like resonances over time in ^{1}H-^{13}C HMQC spectra of ddFLN-dom5+47-RNC (Figure 3.13D) was also attributed to the release of the native domain via truncation of the linker, as well as in the case of ddFLN-dom5+57 and +67 (Supplementary figures, Figure A.3). The timescale of proteolysis as observed by this increase in intensity in the heteronuclear spectra varied: 20$\pm$13hrs for ddFLN-dom5+110-RNC, 16$\pm$4 hrs for ddFLN-dom5+47-RNC, 66$\pm$33 hrs for ddFLN-dom5+57-RNC and 65$\pm$35 hrs for ddFLN-dom5+67-RNC. The folded domain however appears to be resistant to proteolysis (as seen by the heteronuclear spectra recorded after proteolysis, MS and anti-His western, Section 2.3.3.4) and thus proteolytic cleavage was targeted at the disordered linker downstream the C-terminal end of the domain. The resistance of folded domains to limited proteolysis has previously been used to probe for native structure [207] and is often used to identify protein domains [208]. In this regard, it is interesting to note that in RNCs with linker length of 47 residues and longer, ddFLN-dom5 is resistant to the proteolysis accounted here, under the influence of trace amounts of proteases.

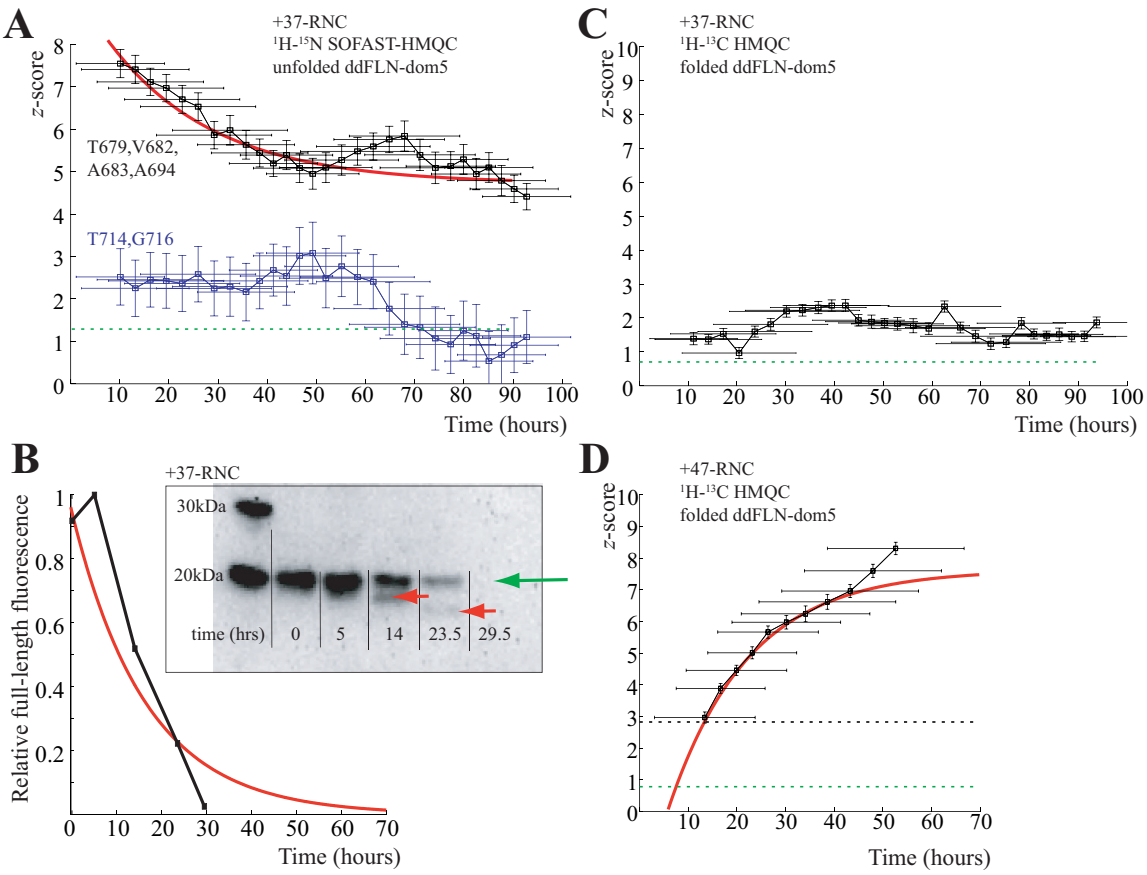


Figure 3.13: A: Decrease of the mean mean z-scores of the intensities of ^{1}H-^{15}N unfolded ddFLN-dom5 resonances over time in ddFLN-dom5+37-RNC. The vertical error-bars show the standard error of the mean, and the horizontal error-bars show the time of acquisition of 8 spectra. The black line shows the mean z-scores of the unfolded resonances T679, V682, A683 and A694. The blue line shows the mean z-scores of unfolded resonances T714 and G716. The dotted green line shows the 5% threshold corresponding the limit of observability for the mean of five resonances. Data are fitted to equation 2.8 (red curve), k=21.5 $\pm$ 7 hours (r^2=0.92). **B**: Decrease in the mean z-scores of the intensities of the signal in the 6×His western blot corresponding to full-length ddFLN-dom5+37-NC. The western blot is shown in the insert. The green arrow shows the size of the full-length RNC, and the red arrows highlight fragments that have a size similar to that of ddFLN-dom5. In the plot, the y axis is the normalised signal intensity from the western fluorescent signal. The x axis is the time in hours. Data are fitted to equation 2.8 (red curve), k=11.5 $\pm$ 8 hours (r^2=0.93). **C**: Mean z-score of the intensities of folded methyl resonances ddFLN-dom5+37-RNC, in ^{1}H-^{13}C HMQC. The axis are the same as in A, as well as the green line. **D**: Same as C for ddFLN-dom5+47-RNC. The red curve shows the fit of equation 2.7, k=16 $\pm$ 4 hours (r^2 0.99).

In contrast to RNC with a long linker length, RNC constructs with short linker lengths were not subject to the equivalent proteolytic behaviour over time, and were also seen to populate the unfolded state as seen in the previous Section 3.2.2.2. Samples of ddFLN-dom5+21-RNC and ddFLN-dom5+37-RNC were however subject to proteolysis over time in the absence of protease inhibitors, as shown by the disappearance of the full-length band in the anti-His western time-course (Figure 3.13B) with a time constant of 12 $\pm$ 8 hours. This proteolysis, however, was not correlated with the appearance of native ddFLN-dom5 resonances in the ^{1}H-^{13}C HMQC spectra, the intensities of which remained below detection during the entire time-course (Figure 3.13C). On the contrary, the intensities of isolated "unfolded" resonances in the ^{1}H-^{15}N SOFAST-HMQC (ddFLN-dom5 residues T679, V682, A683 and A694) appeared to decrease over time (Figure 3.13A, black curve) with a time constant of 21.5 $\pm$ 7 hours. This result suggests that with linker lengths of 21 and 37 residues, the entire immunoglobulin domain has the propensity to sample conformation that renders it sensitive to proteolysis, in which the domain can be truncated within 4-29 hours. It it interesting to note that the most intense resonances in the ^{1}H-^{15}N SOFAST-HMQC of ddFLN-dom5+37-RNC also correspond to resonances of the N-terminal end of the domain. Unfolded resonances from the C-terminal end of the domain (T714 and G716, Figure 3.13A, blue curve) were typically associated with a lower SN in the ^{1}H-^{15}N SOFAST-HMQC, and were also observed to have decreasing intensities from 70 hours, at which time the intensities of the N-terminal resonances reached a plateau. This might indicate that the proteolysis targets the disordered NC in a N- to C-terminal manner.

Although this result is not equivalent to a controlled limited proteolysis experiment, it is worth noting the correlation between the susceptibility of RNCs to proteolysis and NMR observations. RNCs of ddFLN-dom5 with a linker length $\geq$ 47 residues

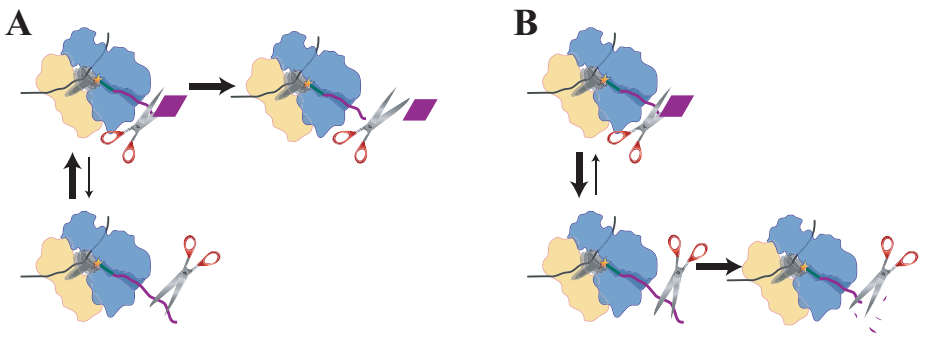


Figure 3.14: Schematic of the relation between resistance to proteolysis reaction and the folding equilibrium. **A:** In the case where the RNC populates the folded state, the folded domain is resistant to proteolysis. **B:** In the case where the RNC populates unfolded or partially folded states, the domain is susceptible to proteolysis. Refer to the main text.

gave rise to dispersed resonances in the ^{1}H-^{13}C methyl-TROSY HMQC, and proteolysis typically occurred downstream in the linker region, thus releasing the folded domain from the ribosome (Figure 3.14A). ddFLN-dom5-RNC with linker length $\leq$ 37 residues, on the other hand, gave rise to resonances assigned to the disordered state in the ^{1}H-^{15}N SOFAST-HMQC, and proteolysis caused the progressive truncation of the unfolded immunoglobulin from the N-terminus to the C-terminus (Figure 3.14B).

The disappearance of the full-length RNC band in the anti-His western blot also correlated with the appearance of a lower MW bands whose sizes were similar to that of the ddFLN-dom5 domain, as it was observed in the case of ddFLN-dom5+110+RNC (Figure 2.31B) and ddFLN-dom5+47+RNC (Figure 3.8E). Segments of a protein that are resistant to proteolytic cleavage in limited proteolysis experiments were attributed to residual structural elements in intermediate states such in as molten globule ensemble, for example in the case of bovine alpha-lactalbumin [209, 210]. The observation of an anti-His band of the size of the ddFLN-dom5 domain suggests that the domain was sampling a conformation that protects it from proteolysis in ddFLN-dom5+37-RNC, and

this observation is in line with the potential population of folded but "invisible" state ddFLN-dom5+37-RNC that was inferred from the lower intensity of unfolded resonances in the ^{1}H-^{15}N SOFAST-HMQC spectrum of ddFLN-dom5+37-RNC compared to that of shorter RNCs, and the shift of the distribution of the intensities of folded ddFLN-dom5 resonances compared to the noise distribution in the ^{1}H-^{13}C methyl-TROSY HMQC.

A controlled limited proteolysis experiment might assist in determining differential rates of proteolysis as a function of linker lengths, and probe for a low population of folded state that is resistant to limited proteolysis. Such an experiment would provide complementary information to the NMR results, and in particular provide additional information to the transient population of NMR invisible folded states in short RNCs.

3.3 Further discussion and concluding remarks

NMR spectroscopy was used to map the co-translational folding properties of ddFLN-dom5 as it emerges from the ribosomal exit tunnel. At a linker length of 21 residues, the C-terminal end of ddFLN-dom5 has not emerged from the ribosome exit tunnel, and is not available for the entire domain to fold. In the isolated native state, strands B, E & F were seen to be the most stable, but strand G also makes stable H-bounds with strand F (ΔG^{I743}_{op-cl}=9.1kcal.mol^{-1}, ΔG^{V745}_{op-cl}=8.9kcal.mol^{-1}). The observation of unfolded ddFLN-dom5 at short linker lengths suggest that the availability of strand G is required for complete folding to occur, and perhaps these H-bonds are necessary to stabilise the folded state. The unfolded state was also observed to be populated at linker lengths up to 37 residues. This length exceeds the number of residues needed to pass through the ribosome tunnel, assuming a fully expanded conformation (24 residues [48]). This indicates therefore, that either the ddFLN-dom6 linker has the propensity to adopt a compact structure in the tunnel (Figure 3.15A), or that the ribosome itself is playing

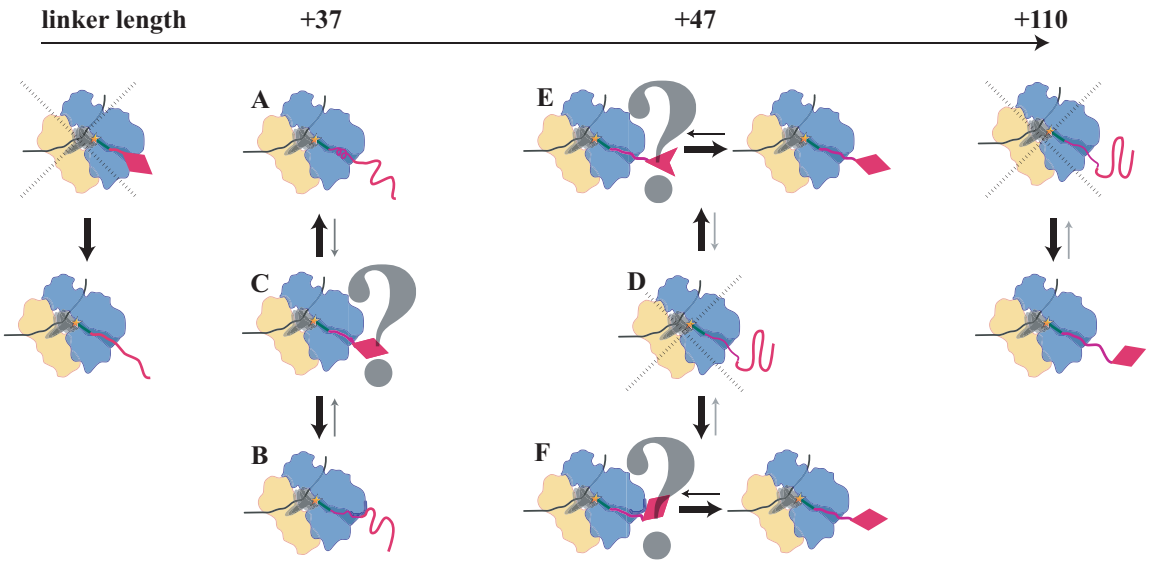


Figure 3.15: Schematic of the folding equilibrium at increasing linker lengths. The question marks indicate the "invisible by NMR" states, on which the current set of NMR and biochemical data does not provide any information. Refer to the main text.

an active role that prevents co-translational folding at short translation lengths (Figure 3.15B). There is no evidence that the N-terminal end of ddFLN-dom6 has a propensity to form an α-helix *in vitro*, although one can not exclude the formation of such helix within the confining environment of the ribosome exit tunnel. Indeed, the formation of α-helices in the last 50Å of the tunnel have been observed for a sequence that is prone to helix formation [41], and SecM was seen to adopt a slightly compact conformation near the exit port of the tunnel [40]. The need for a longer linker to observe complete folding could also be due to transient interactions of the ribosome with the nascent chain, and in particular with the linker sequence. If the linker itself interacts with the ribosome, its ability to provide enough flexibility for the C-terminal end of ddFLN-dom5 to sample the conformational space required to fold might be altered (Figure 3.15B).

The NMR data alluded to the presence of an invisible folded state at linker length of 37 residues as observed in the decreased intensity of the unfolded resonances compared to those of ddFLN-dom5+21-RNC in the ^{1}H-^{15}N SOFAST-HMQC. The very low intensity of the folded ddFLN-dom5+37-RNC resonances in the ^{1}H-^{13}C methyl-TROSY HMQC could be the result of the very low population of this intermediate state (Figure 3.15C). In addition, the linker length might not be long enough to provide the require flexibility for the NC to tumble independently of the ribosome, thus preventing NMR observation. Indeed, the persistence length of a disordered protein has been shown to be around seven residues [171], and at a linker length of 37 residues, at most 13 of these residues are present outside of the ribosomal tunnel, however, this might not be a linker length which ensures independent tumbling. This restricted tumbling is surprising with respect to computational studies which have suggested that the exit port of the ribosome exit tunnel present an entropic window for the nascent chain to sample a larger conformational space (detailed in the introduction Section 1.1.4). Indeed, previous fluorescent anisotropy studies of partially folded ApoMb-RNC which has not emerged fully from the ribosome were also shown to undergo an independent tumbling with a correlation time as short as 3-11ns [83]. This suggests therefore that either the intermediates observed for ApoMb-RNC have a more compact structure that the invisible folded state populated by ddFLN-dom5+37-RNC (and thus are tumbling faster), or that the linker sequences used here behave differently to that of the C-terminal disordered extension of Apo-Mb-RNC.

An interaction of ddFLN-dom5 itself with the ribosome would also limit further its flexibility and render its resonances broadened beyond detection. At lengths of 47 residues and longer, however, the observed folded RNC resonances were observed to tumble independently of the ribosome complex via the analysis of the proton linewidths in the ^{1}H-^{13}C methyl-TROSY HMQC, therefore any potential interaction with

the ribosome is likely to be partially disrupted at longer linker lengths, in which the domain has the capacity to fold completely (Figure 3.15F). The use of a combination of NMR and biophysical/biochemical assays can shed light to any potential NMR invisible intermediates state populated at these lengths. For example, the data presented here from uncontrolled limited proteolysis hindered the resistance to proteolysis of a low population of folded domain in ddFLN-dom5+37-RNC.

With a longer linker length of 47 residues, the population of a disordered state was not observed in the ^{1}H-^{15}N SOFAST-HMQC (Figure 3.15D). Native-like resonances were observed in the ^{1}H-^{13}C methyl-TROSY HMQC, indicating that the immunoglobulin domain is folded. The dynamics of the ddFLN-dom5 at this stage of co-translational folding differs, however, from that of the fully exposed domain seen with a linker length of 110 residues. These dynamics might reflect a chemical exchange with non-native states that remains to be characterised (Figure 3.15E). This is interesting in light of the increased dynamics of the N-terminal hemisphere that was observed (with selective broadening of residues I738 and I674), in comparison to the overall stability of the N-terminal hemisphere as observed in the native state by H-D experiment on isolated ddFLN-dom5. In particular, residue I674 were seen to be very stable in the native state ($\Delta G^{\mathrm{I674}}_{\mathrm{op-cl}}$=10kcal.mol^{-1}), however, it was selectively broadened in the ddFLN-dom5+47-RNC, indicating that this residue is subject to chemical exchange or local anisotropic effects. In contrast, resonances of the G-strand had similar linewidths in the ^{1}H-^{13}C methyl-TROSY HMQC spectra of ddFLN-dom5+47 and +110 RNC, which indicates that these have similar dynamics and that the overall motion of ddFLN-dom5 is not strongly affected by the increased linker length. Although the linewidth analysis excluded a strong interaction with the ribosome, transient interactions that could shift the folding equilibrium cannot be excluded (Figure 3.15F).

In summary, the current data provides for the first time a structural and dynamical understanding of the different conformations sampled during co-translational protein folding. The newly described methods to generate RNCs and analyse their properties by NMR will enable future experiments of this kind to be explored. To complete this understanding, the main challenge resides in the limited NMR observability of RNC with restricted tumbling and of rigid regions of the ribosome complex, aside from the challenges related to the variability in the sample preparation and stability. The nascent chain spans through the large ribosomal subunits only; it would be therefore interesting to analyse the effect of reducing the size of the complex to the individual subunits on the NMR observability. The use of higher temperatures can also permit the NMR observation of slower tumbling regions, this has been tested on stable archaeal ribosomes, and is the subject of the next chapter.

Chapter 4

NMR investigations of intact ribosomes

4.1 Introduction

NMR studies of high molecular weight biological systems have increased in number recently, as was discussed in Section 1.2.3. Notable examples are studies of the 20S proteasome, of GroEL-ES and of the SecA machinery. The NMR study of the 20S proteasome revealed the N-terminal residues as forming a primitive gate at the entrance pore [122,123], that regulate protein degradation, with exchange rate between the closed and opened conformations of these N-terminal residues was found to be on the order of a second. Also, the dynamics of the surface residues, the entrance pore and the catalytic chamber were found to assist substrate localisation [122]. The network of substrate-enzyme interactions inside the catalytic chamber was also investigated by NMR, revealing that these interactions shift the folding equilibrium of the substrate by destabilising the folded states [124], thus facilitating protein degradation. The NMR investigation of a substrate inside the GroEL-ES chaperonin chamber revealed an opposite effect of confinement to that of the proteasome which actually destabilised the disordered state of the substrate [125]. In the NMR studies of the translocation machinery [153] the recognition mode of hundreds of different signal sequences by the cognate receptor SecA that mediate translocation through the Sec pathway was shown by NMR to be mediated by hydrophobic and electrostatic interactions with a single groove

on SecA [153]. The same NMR study revealed an oscillation between two conformations of SecA: an open conformation where the groove is accessible and a closed conformation where the C-terminal tail is occluding the groove. This equilibrium was proposed to be an autoinhibitory regulation mechanism, which is relieved by interaction with the SecB chaperone.

These studies offered the hope to the possible use of NMR to investigate aspects of the ribosome, in particular the dynamical features of translation, such as the network of transient interactions between the NC and the ribosome exit tunnel, the allostery between the exit port of the exit tunnel, the universal docking site, and the PTC and tRNA binding sites (Section 1.1.1.1). As seen from the investigations of RNCs presented in Chapters 2 & 3 of this thesis, the NMR study of the ribosome and its dynamics is challenging. Its size exceed the typical maximum size for traditional NMR and computational studies. While initial NMR studies of ribosome were described as early as the 1980s with the observation of the L7/L12 stalk [33,211–213], the detailed NMR investigation of the entire ribosomes has not been presented to date. This Chapter describes some preliminary NMR investigations of entire ribosome complexes.

4.2 Results and Discussion

4.2.1 NMR of the *E. coli* ribosome

4.2.1.1 Production of purified 70S, 50S and 30S particles.

Isotopically labelled ribosome complexes suitable for NMR were purified from *E. coli* cells grown in $^{15}N/^{13}C$ enriched media as described in Section 6.1.7, or in media supplemented with methyl-^{13}C, 3,3-^{2}H, α-ketobutyric acid in a perdeuterated background to produce selectively Ileδ_1 labelled complex (Section 1.3.2). Following

purification, the ribosomal complexes were assessed for purity using agarose gels and SDS-PAGE (Figure 4.1). In 1% w/v agarose gels, two RNA bands were typically observed, which represented the migration of the 23S and the 16S ribosomal RNAs (Figure 4.1A). The very small 5S rRNA migrates beyond the ethidium bromide line. In the SDS-PAGE of a typical 70S sample (Figure 4.1B), a series of bands were observed, which correspond well to the expected distribution of sizes of the ribosomal proteins.

This purification strategy typically yielded up to 36±6 nmoles of pure 70S (>95%) per litre of culture. To purify the component subunits, the purification strategy was modified by decreasing the magnesium concentration to 1mM to allow splitting of the subunits, without them dissociating into their component proteins. The optical density at 254nm (OD_{254}) profile of the fractionation of the sucrose gradient (Figure 4.1C) shows good separation of the individual subunits, which was further confirmed by the distinct pattern of the proteins within the 50S and 30S complexes revealed by SDS-PAGE analysis of the fractions (Figure 4.1D). The recovery of isolated subunits from the 70S ribosomes loaded was as high as 94% for the large subunit, but only *ca.* 40% for the small subunit, based on the concentration measured by OD_{260}.

4.2.1.2 Translational diffusion of the 70S particle and the 50S & 30S subunits monitored by NMR spectroscopy

Both the 70S ribosome complex and the individual subunits were characterised by NMR. As described in Section 2.3.2.1 (Figure 2.15), the ribosome gave rise to well-defined resonances in proton spectra, as also seen for the isolated 50S and 30S subunits (Figure 4.2A&B). The methyl resonances at <0.7ppm that arise from the L7/L12 stalk, which gave rise the majority of the signals in the 70S spectrum, could easily be identified in the proton spectrum of the 50S subunit (Figure 4.2B). The spectrum of the 30S subunit

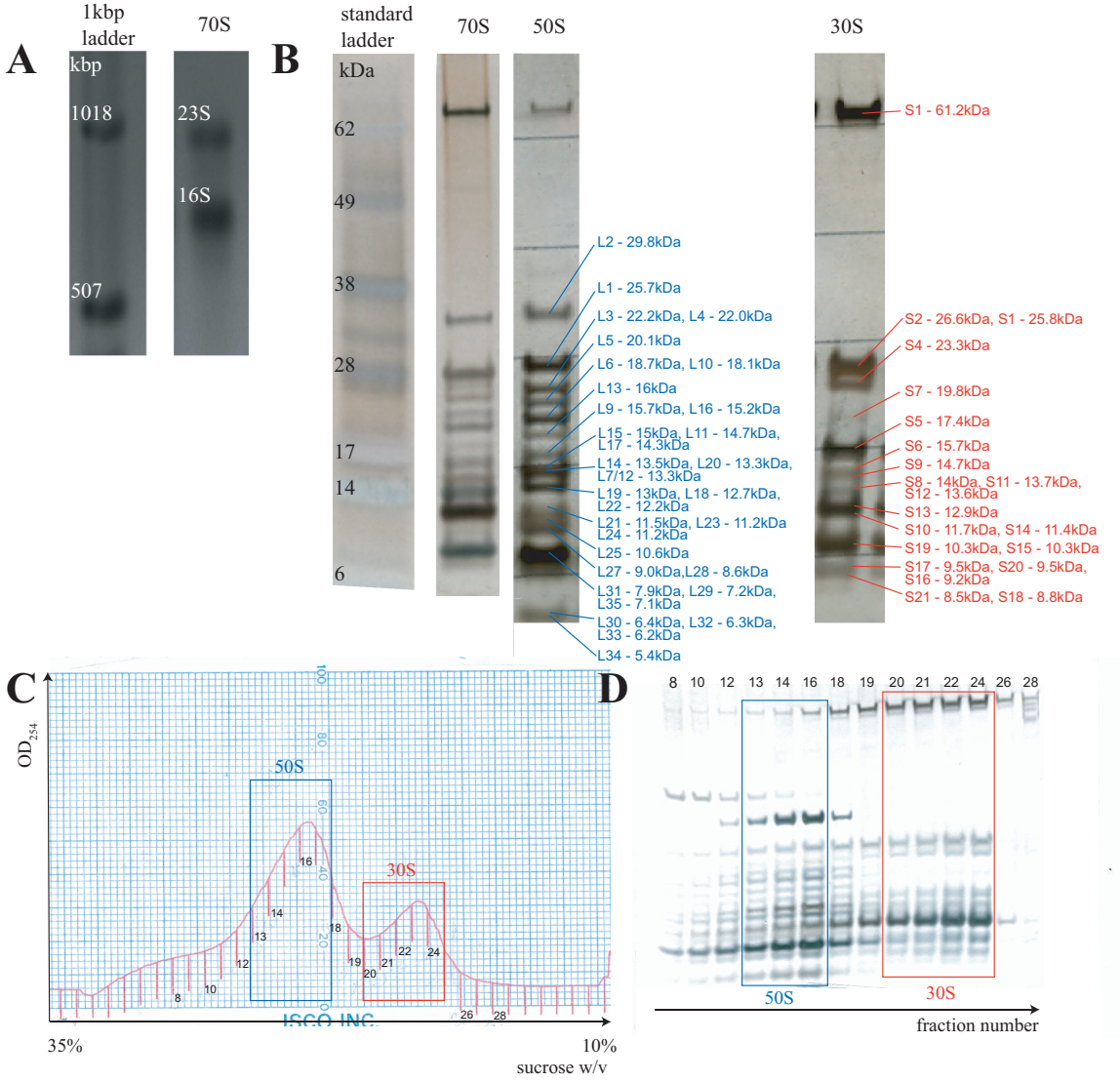


Figure 4.1: **A:** 1% w/v agarose gel of 70S ribosome, with a 1k DNA ladder on the left. The attribution of the RNA bands is shown in white. **B:** SDS-PAGE of ribosome complexes. From left to right: protein ladder, 70S, 50S and 30S. The tentative attribution of the protein bands to each ribosomal protein is indicated in blue for the large subunit and in red for the small subunit. **C:** OD_{254} profile of the 17%-26% w/v sucrose gradient of a 70S ribosomes sample at low Mg concentration (1mM). SDS-PAGE of the numbered fractions is shown in **D**. The fractions that contained pure 50S subunits are in the blue rectangle, and those that contained pure 30S subunits are in the red rectangle.

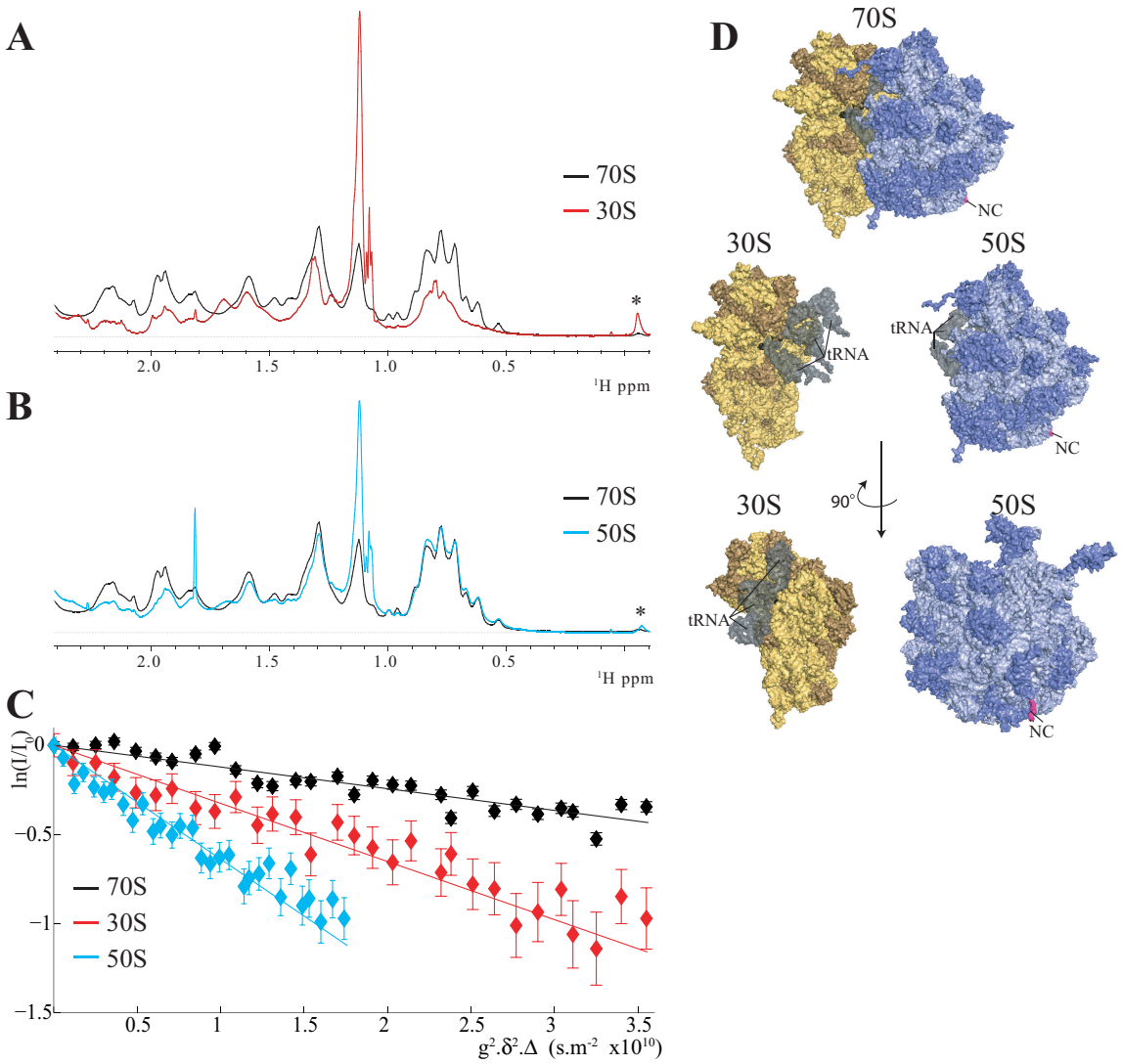


Figure 4.2: **A:** Proton spectrum of the intact ribosome complex (black), overlaid with that of the 30S subunit (red). Resonances observed in the methyl region of the 70S spectrum are likely to arise from the flexible L7/L12 stalk. **B:** Proton spectrum of the intact ribosome complex (black), overlaid with that of the 50S subunit (blue). **C:** Plot of the relative signal intensity in spectra of the ^{1}H STE diffusion experiment (on a log scale) versus the square of the gradient strength on the x axis, for 70S in black, 50S in blue and 30S in red. The signal labelled with * indicates a signal that is associated with a fast diffusion coefficient, and does not arise from the ribosome complex. **D:** View of the 70S ribosome and the 30S & 50S subunits. The A,P and E tRNA are shown in grey, as well as the nascent chain in the magenta.

showed mainly resonances indicative of unfolded regions, with a narrow chemical shift dispersion of the methyl resonances (none at less than 0.7ppm) in agreement with previous studies [33]. These resonances have been tentatively assigned to a disordered extension of the S1 protein as the S1 extension is associated with a high flexibility from X-ray and cryo-EM studies of the 30S subunit. To understand the hydrodynamic behaviour of the ribosome and its subunits, and to ensure the ribosomal attachment of the proteins giving rise to the observed resonances, ^{1}H STE diffusion spectra were recorded, and used to estimate the translational diffusion coefficients of the complexes. As shown previously (Section 2.3.2.1, Figure 2.15), the resonances from the 70S complex show a diffusion coefficient of $1.7\pm0.4\times10^{-11}$ m^2s^{-1} at 25°C (4.2C), a value identical to the diffusion coefficient measured by light-scattering spectroscopy [172]. This value equates to a hydrodynamic radius (r_h) of 13±0.5nm (using the Stokes-Einstein relation, Equation 1.16), which is consistent with the modelled hydrodynamic radius [214] from the crystal structures of *E. coli* ribosomes [25] (r_h=13nm). This r_h relates to an overall correlation time for the ribosome complexe of $\tau_{c,ribo}$=1250ns, (using the Stokes-Einstein-Debye relation, Equation 1.18). Previous measurement of $\tau_{c,ribo}$ using phosphorescence anisotropy decay experiments found a $\tau_{c,ribo}$=3300ns at 15°C [129], corresponding to $\tau_{c,ribo}$=2500ns at 25°C. The discrepancy between the rotational correlation time measure by phosphorescence anisotropy decay and that calculated using the hydrodynamic radius derived from NMR and light scattering experiments may be due to the aspherical shape of the ribosome [129].

The diffusion coefficients of the isolated subunits were measured using a similar ^{1}H STE diffusion experiments and were found to be $2.7\pm0.4\times10^{-11}$ m^2s^{-1} for the 30S subunit and $5.1\pm0.4\times10^{-11}$ m^2s^{-1} for the 50S subunit (Figure 4.2C), which correspond to hydrodynamic radii of 12nm and 5nm, respectively. Diffusion coefficients of ribosomal

subunits measured by light-scattering spectroscopy were found to be 2.1 and 1.9×10^{-11} m^2s^{-1} for the 30S and 50S prepared under identical conditions [172]. The NMR measurement of the diffusion coefficient of the 30S subunits is in good agreement with the value obtained from light-scattering spectroscopy (within uncertainty). It was surprising that the NMR determined translational diffusion coefficient of the 50S subunit was found to be greater than that of the 30S subunits, and indeed under the NMR conditions, the ratio of their diffusion coefficients is opposite to the ratio of their sedimentation coefficients (50 & 30 Svedberg units, respectively), which were determined from sucrose gradient sedimentation. The diffusion coefficient measured for the 50S subunit relates to a hydrodynamic radius that is smaller than the radius modelled from the X-ray crystal structure (r_h=9.5-10nm [25, 214]), and may suggest additional conformational changes upon splitting. The hydrodynamic radius of the 30S subunits extrapolated from its NMR-determined diffusion coefficient (r_h=12nm) is larger than the hydrodynamic radius modelled from the X-ray crystal structure (r_h=7-8nm [25, 214]). The slow NMR-determined diffusion coefficient may suggest that the 30S subunit can dimerise under the NMR conditions used here. It is known that the inactive form of the 30S has the propensity to dimerise [215].

4.2.1.3 Heteronuclear NMR spectra of 70S ribosomes and the 50S & 30S subunits

The ^{1}H-^{15}N heteronuclear spectra of the component subunits are shown in Figure 4.3. The spectrum of the 30S subunit contains *ca.* 35 cross-peaks, all characteristic of resonances arising from disordered polypeptide (Figure 4.3C) and which were assigned to the disordered extension of S1 [33]. Resonances that overlay well with the 77 residues of the C-terminal domain of isolated L7 [31] were observed with narrow linewidth ($\Delta\nu_{^1H}$ ~40Hz in the ^{1}H-^{15}N SOFAST-HMQC) in the spectrum of 70S ribosomes as

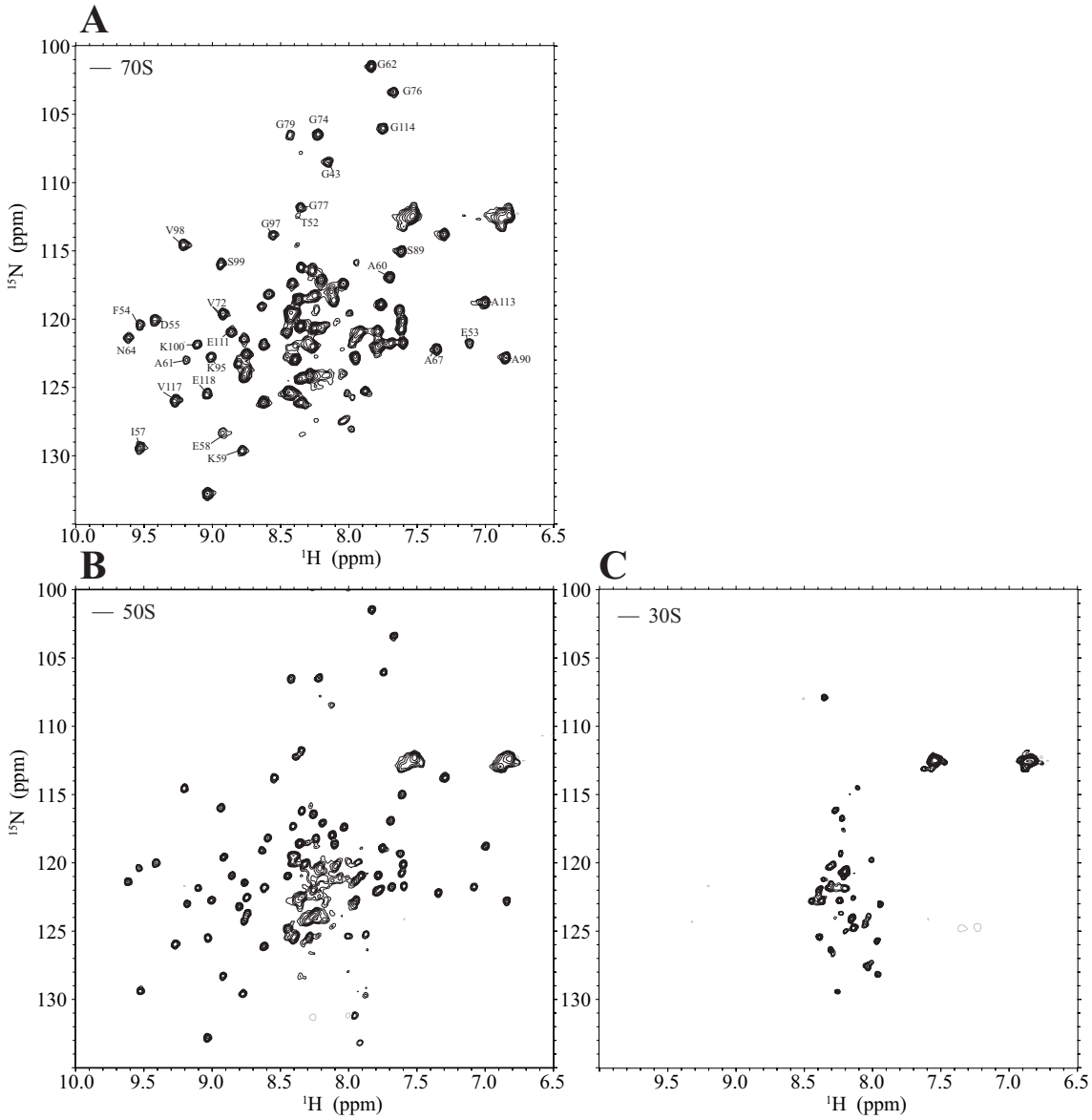


Figure 4.3: ^{1}H-^{15}N SOFAST-HMQC spectra of 70S ribosomes (**A**), 50S subunit (**B**) and 30S subunit (**C**). Partial assignment of the residues of the C-terminal domain of L7/L12 are shown in the 70S ribosome spectrum, the complete assignment is shown in Figure 4.4.

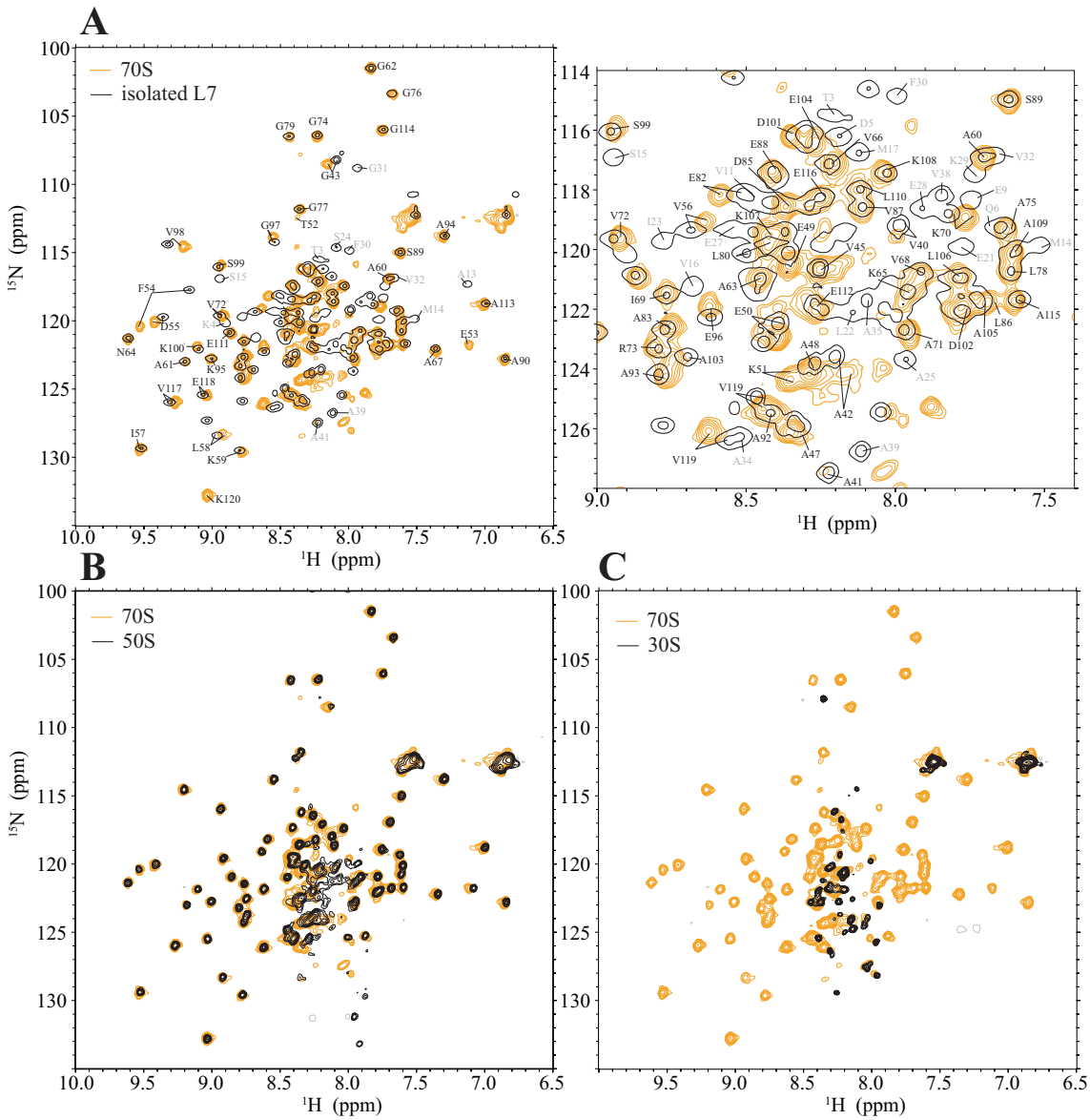


Figure 4.4: **A:** Left: ^{1}H-^{15}N SOFAST-HMQC spectrum of 70S ribosomes (orange), overlaid with that of isolated L7 (black). Resonances assigned to the C-terminal domain are labelled in black, and those assigned to the N-terminal domain are labelled in grey. Right: Magnified insert. **B:** ^{1}H-^{15}N SOFAST-HMQC spectrum of 50S subunits (black), overlaid with that of 70S ribosomes (orange). **C:** ^{1}H-^{15}N SOFAST-HMQC spectrum of 30S subunits (black), overlaid with that of 70S ribosomes (orange).

well as in that of the 50S subunit (Figure 4.4A & B). Resonances from the N-terminal domain of L7/L12 (residues 2-30) were apparently broadened beyond detection, as were 10 N-terminal resonances of the linker between the N-terminal and C-terminal domains (residues 31-41). Resonances from the remainder of the linker (residues 42-51) could be observed as well as those from the C-terminal domain.

The detection of the linker resonances more than ten residues from the L7/L12 N-terminal domain is pertinent as regards the observability of the nascent chains as a function of the distance from the ribosome, that was described in Chapters 2 & 3. The N-terminal domain of the L7 stalk sits on L10, itself attached to the rigid ribosome via interaction with L11 (Figure 1.2). L11 is associated with a well defined electron density in the cryo-EM and X-ray crystal structures of the *E. coli* ribosome [11, 16], and is thus considered to be part of the rigid core of the ribosome. While L10 and its interacting partners the N-terminal domains of L7/L12 appear to tumble too slowly to be observed in the ^{1}H-^{15}N SOFAST-HMQC spectrum, the disordered linker of L7/L12 appears to provide sufficient flexibility for the tenth linker residue and beyond to be observed, in line with the persistence length of an intrinsically disordered peptide, which is on the order of seven residues [171]. If this dynamical behaviour is also characteristic of the ddFLN-dom6 linker, one would expect to observe resonances that are no less than 10 residues away from the exit port of the ribosomal exit tunnel.

The ^{1}H-^{13}C HMQC spectra of the ribosomal subunits are shown in Figure 4.5. The spectra of 50S subunits and 70S ribosomes show a number of methyl resonances (>40) that overlap with typical V, A, T and L chemical shifts as shown in the boxes regions of Figure 4.5 [216]. Notably three distinct Ileδ_1 resonances can be identified (Figure 4.5A & B). The C-terminal domain of L7/L12 contains two Ile residues (I57 & I69); the additional Ileδ_1 resonances observed in these ^{1}H-^{13}C HMQC spectra of the 50S and 70S particles

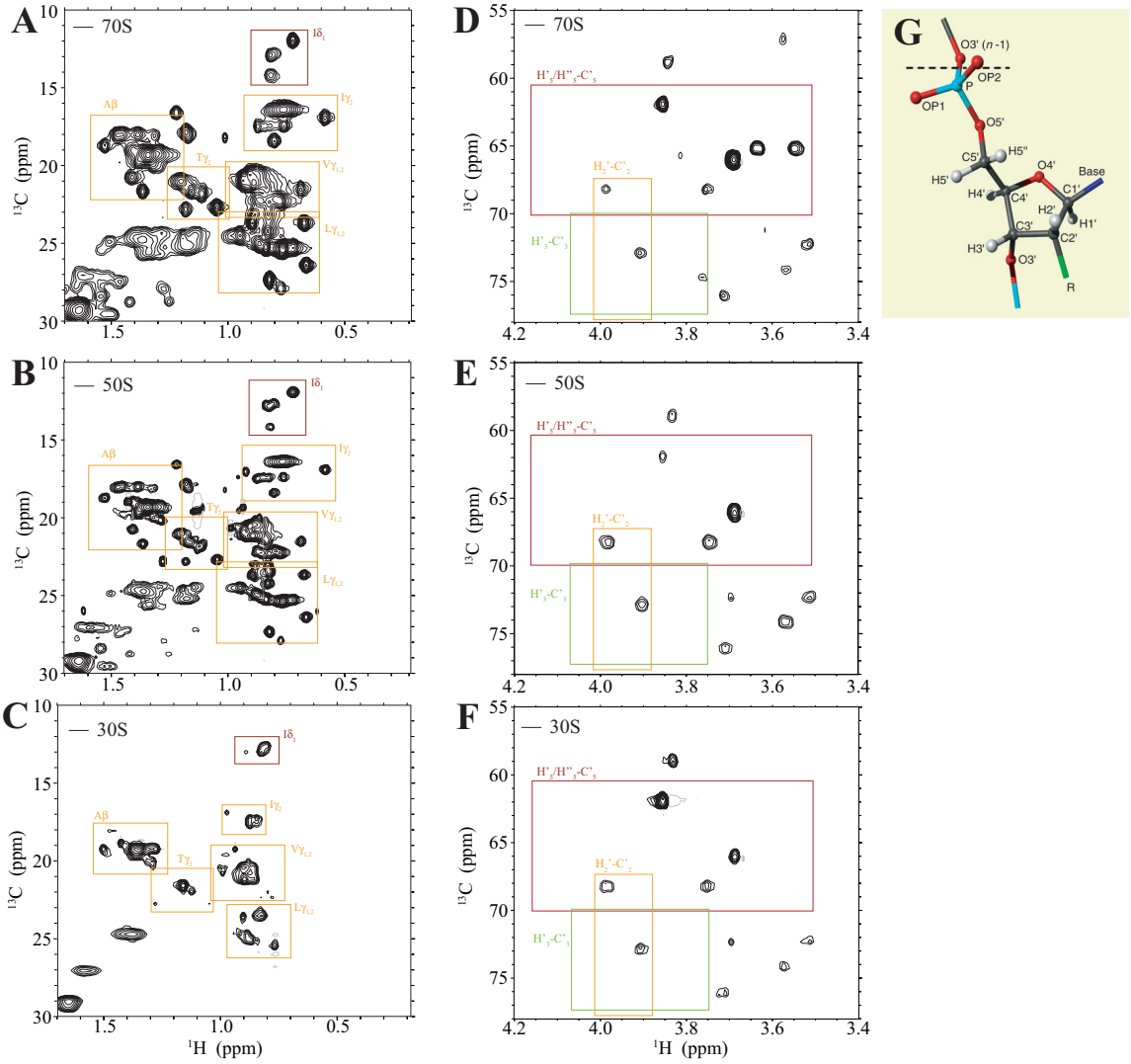


Figure 4.5: **A, B & C:** ^{1}H-^{13}C HMQC spectra of 70S ribosomes, and 50S & 30S subunits, respectively, showing the methyl region of the spectra. The typical chemical shift of methyl groups are indicated with the squares, from the BioMagResBank [216]. **D, E & F:** Ribose region of the ^{1}H-^{13}C HMQC spectra of 70S ribosomes, and 50S & 30S subunits respectively. The typical chemical shift of ribose H-C resonances are indicated with the squares, from the BioMagResBank [216]. **G:** Structure of the RNA ribose showing the nuclei nomenclature.

does not appear to have been observed in the corresponding ^{1}H-^{15}N SOFAST-HMQC spectra. The ^{1}H-^{13}C HMQC spectrum of 30S subunit contains resonances that overlap with the typical chemical shift of methyl group of unfolded I, V, L, A and T residues ([216], Figure 4.5C). No high-field-shifted methyl resonances (^{1}H ppm <0.5ppm) were observed in the spectra of either 70S ribosomes, or 50S or 30S subunits (Figure 4.5A&B). This is in agreement with both the disordered properties of S1 in the 30S subunit, and the absence of aromatic residues within the C-terminal domain of L7 in the 70S ribosomes and 50S subunit, which means the methyl resonances do experience ring-current shifts.

Interestingly, a notable feature of these ^{1}H-^{13}C HMQC spectra are resonances from each of the subunits and from that of the 70S ribosomes that can be attributed to ribose groups (Figure 4.5D, E & F). These resonances were associated with a diffusion coefficient of $1.6\pm0.5\times10^{-11}$ m^2s^{-1} in the ^{1}H STE spectra of 70S ribosome, which is very similar to the diffusion coefficient of the intact 70S complex, and strongly indicates that these resonances arise from ribosomally asscociated RNA. However, the corresponding resonances from the RNA bases are not observed, rendering the assignment of the ribose resonances challenging.

In order to investigate the utility of NMR to further probe the core of the ribosome complex, ^{1}H-^{13}C methyl-TROSY HMQC spectra were recorded from perdeuterated ribosome samples that were selectively labelled at the Ileδ1 methyl groups: U-[^{12}C,^{2} H], Ileδ1-[$^{13}C^{1}H_3$], a strategy that has been used successfully to probe other large systems, and the RNCs (Chapter 3). In conventional ^{1}H-^{13}C HSQC spectra, the three Ile previously observed from the 50S and 70S particle (among which two resonances were assigned to the L7/L12 C-terminal domain, Figure 4.6A) gave rise to intense and narrow resonances. The presence of these slow-relaxing signals prevented the acquisition time being decreased to values that would optimise the signal-to-noise for much broader

signals (ideally as short as *ca.* 3ms for the ^{13}C acquisition time), because truncation of the slow decaying FIDs of the flexible Ile resonances results in sinc side bands that overlap with and obscure low intensity resonances. When the ^{1}H-^{13}C methyl-TROSY HMQC spectrum of the intact 70S particle was recorded, *ca.* 10 broad cross-peaks were observed (Figure 4.6B), with signal-to-noise ratio (SN) ranging from 5.1 to 9.5, in addition to the three intense Ile resonances mentioned previously. This SN was higher than the noise threshold (2.8) for new resonances to be detected, allowing a search within the entire spectrum (Section 2.2). Four of these resonances were observed in the ^{1}H-^{13}C methyl-TROSY HMQC spectrum of the 30S subunit (Figure 4.6C), and notably these included the only high-field-shifted resonance (peak B, 0.18ppm). The ^{1}H-^{13}C methyl-TROSY HMQC spectrum of the 50S subunit revealed a cluster of overlapping broad signals at random coil chemical shifts (^{13}C ppm: 11-14, ^{1}H ppm: 0.65-0.9ppm, Figure 4.6D), together with a cluster of Ile resonances that is low-field-shifted in the ^{13}C dimension, with ^{13}C chemical shift higher than 16ppm, which are indicative of a trans conformation of the Ile side-chains [217].

The number of observed resonances is clearly significantly lower than the number of Ile residues in the ribosomal complexes (334 in the 70S complex, 189 in the 50S subunit and 145 in the 30S subunit). Nonetheless, in the ^{1}H-^{13}C methyl-TROSY HMQC spectra of the isolated subunits and of the intact ribosome, overlapping resonances are observed that were broadened beyond detection in the ^{1}H-^{13}C HSQC spectra. In the 50S subunit, these resonances exceed the number of Ile residues in L7 domains (two in the C-terminal domain and four in the N-terminal domain), and in the 30S subunit, the observation of a high-field shifted resonances suggests that a folded protein is observed in addition to the S1 disordered extension. This indicates that residues from more rigid regions of the subunits give rise to observable resonances. The combination of broad linewidth and

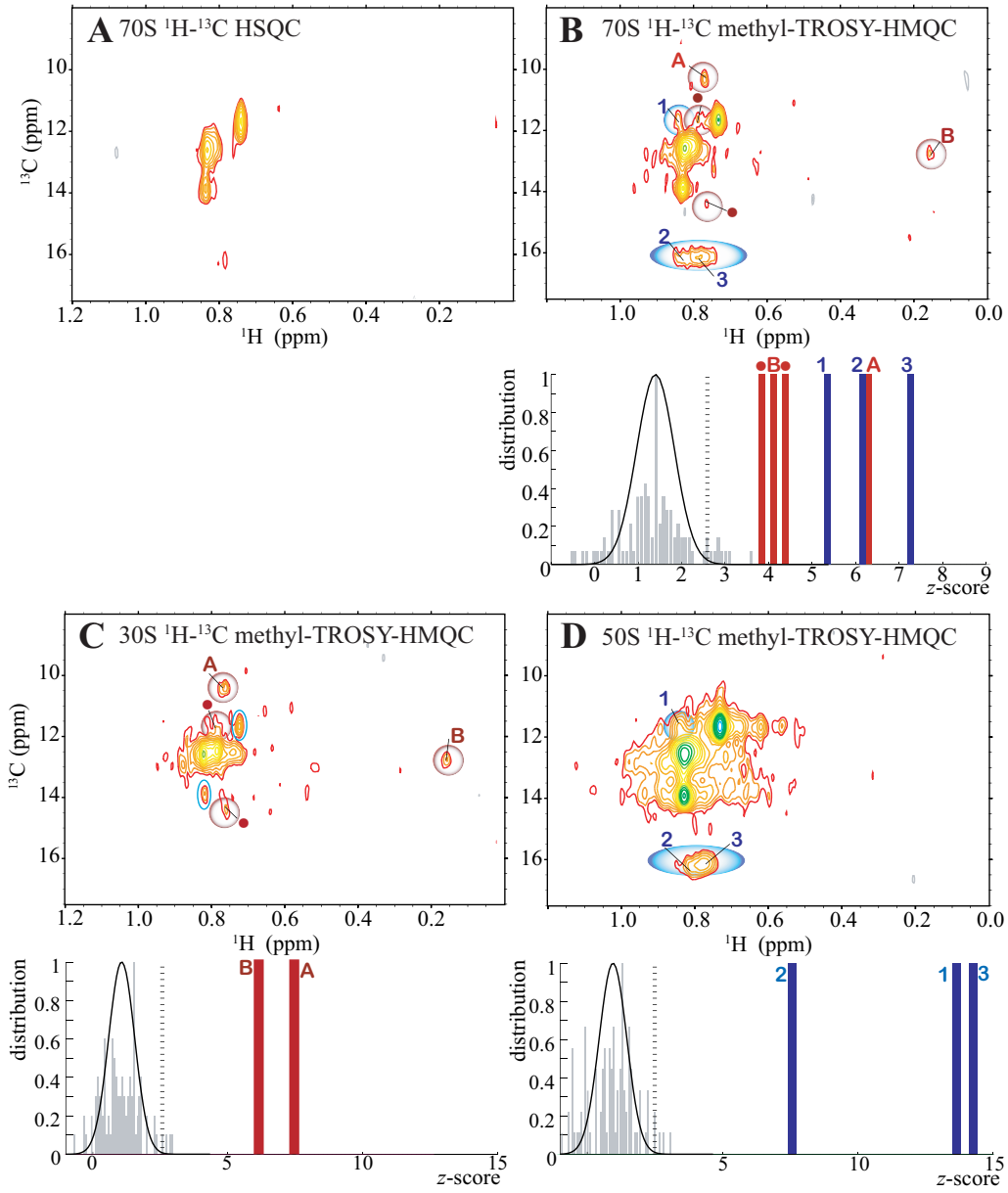


Figure 4.6: **A:** ^{1}H-^{13}C HSQC spectrum of 6 μM U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] 70S ribosomes (recorded for 1 hour). **B:** ^{1}H-^{13}C methyl-TROSY HMQC spectrum of 6μM U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] 70S ribosomes (recorded for 3.5 hour). **C:** ^{1}H-^{13}C methyl-TROSY HMQC spectrum of 3μM U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] 30S subunits (recorded for 17.5 hour). Resonances arising from the L7/L12 of the small amount of non-separated 50S are shown in the blue circles in the centre of the spectrum. **D:** ^{1}H-^{13}C methyl-TROSY HMQC spectrum of 6μM U-[^{12}C,^{2}H], Ileδ1-[^{13}C^{1}H$_3$] 50S subunits (recorded for 17.5hour). The signal intensity distributions are shown below the spectra. The signals assigned to the 30S subunits are indicated with a letter or a red dot and are shown in red in the intensity distributions, and those assigned to the 50S subunits are indicated with a number and are shown in blue in the intensity distributions.

absence of chemical shift dispersion renders the observed resonances largely overlapped, and complicates their NMR analysis.

One commonly used strategy to attempt to decrease the linewidths of broadened resonances is to shorten the rotational correlation time by increasing the temperature. To examine whether ribosomes were amenable to elevated temperatures, their stability was first investigated at ambient temperature (25°C) using the established NMR methodology presented in Chapter 2, specifically the use of ^{1}H 1D and ^{1}H STE diffusion experiments. ^{1}H STE diffusion experiments indicated the appearance of protein signals associated with a high diffusion coefficient after 24 hours in both 70S ribosome samples (D=3.0$\pm$0.3$\times10^{-11}$ m^{2}s^{-1} compared to D=1.7$\pm$0.3$\times10^{-11}$ m^{2}s^{-1} at time zero) as well as in samples of the individual subunits, with the spectra recorded with a low gradient strength undergoing an increase in intensity after 24 hours (Figure 4.7C). This coincided with the appearance of additional resonances in the ^{1}H-^{13}C HSQC spectrum (Figure 4.7A to B). Hence, both the ^{1}H STE diffusion and the ^{1}H-^{13}C HSQC experiments indicated the release of ribosomal proteins, which could be related to the break-down of entire ribosome, as indicated by analytical sucrose gradient runs of identical ribosome samples. Indeed, the SDS-PAGE of the low-molecular-weight fractions of the analytical sucrose gradient (black rectangles in Figure 4.8B & E) presented a similar pattern to that of the ribosomal fractions (orange rectangle in Figure 4.8). This results suggests that the break-down of the ribosomes is perhaps mediated by dissociation of the ribosomal proteins from the rRNA scaffolding, leading to the disassembly of the complex.

Although the integrity of the ribosomal complex could be easily monitored *in situ*, the limited stability of the ribosome samples suggested that elevated temperatures were likely to result in increased break-down. At 30°C, visible precipitates were detected after 2-3 hours of NMR acquisition. The instability of the ribosomes at temperatures $\geq$

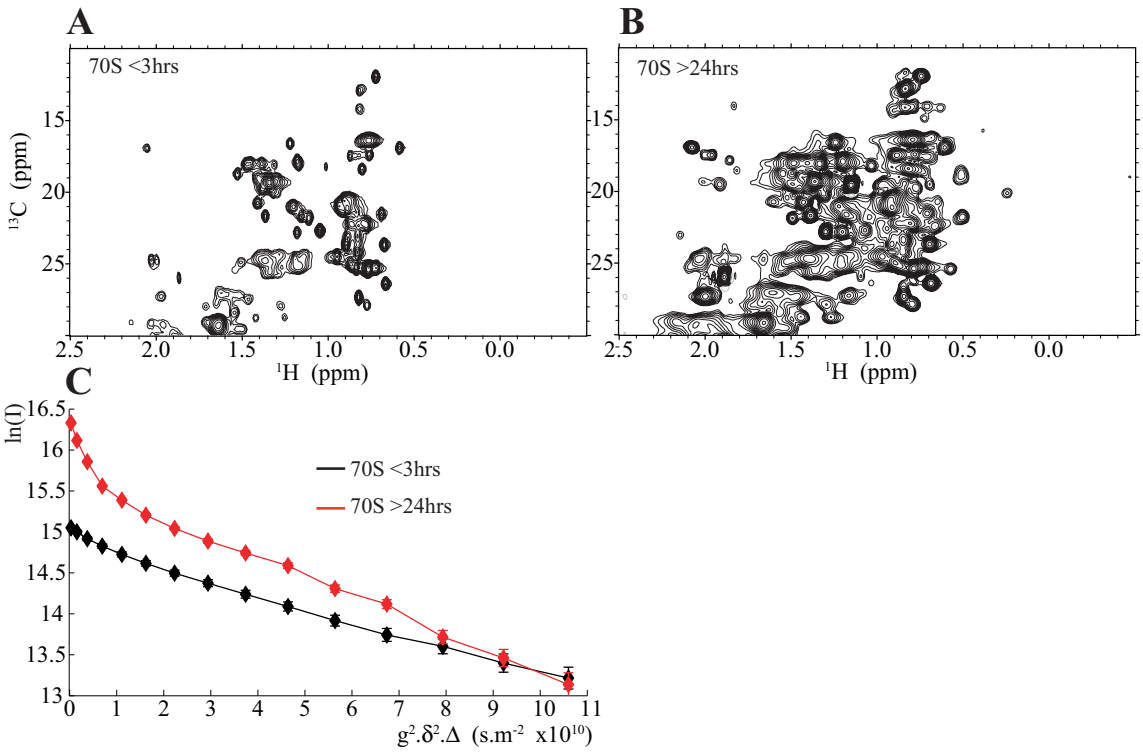


Figure 4.7: A: Methyl region of the ^{1}H-^{13}C HSQC spectrum of 70S ribosomes that were associated with a diffusion coefficient of 1.64×10^{-11} m^2s^{-1}. **B:** Methyl region of ^{1}H-^{13}C HSQC spectrum of 70S ribosomes that were left for 24 hours at 25°C. **C:** Plot of the signal intensity in the different spectra of the ^{1}H STE diffusion experiment on a log scale versus the square of the gradient strengths in the x axis, for fresh 70S ribosomes in black, and after 24 hours at 25°C in red. In the latter case, the intensity decay was characteristic of a heterogeneous sample, in which more than two species with different diffusion coefficients are being observed. The diffusion coefficient calculated from the four lowest gradient strengths is $D=1.0\pm0.1\times10^{-10}$ m^2s^{-1}.

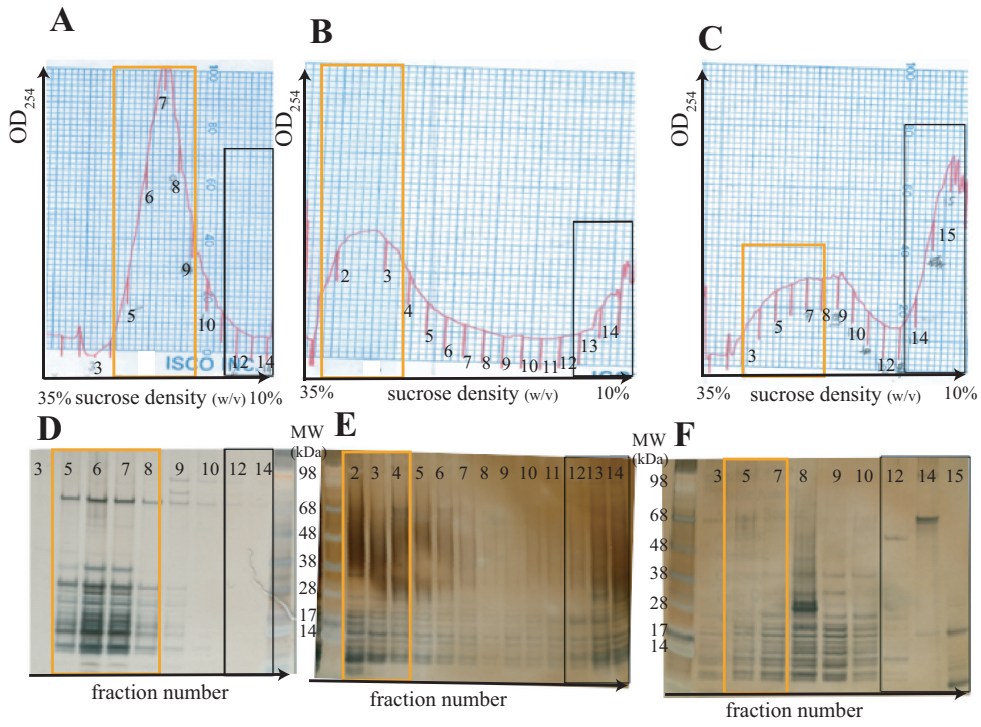


Figure 4.8: Analytical sucrose gradient profile of (**A**) an intact 70S ribosome sample, with a diffusion coefficient measured by ^{1}H STE diffusion experiments of 1.7×10^{-11} m^2.s^{-1}, (**B**) a sample that showed an increase in its diffusion coefficient after 24 hours in the spectrometer, see main text and (**C**) a sample that showed a diffusion coefficient of 1×10^{-10} m^2.s^{-1}, after being stored for one week at 4°C. **D,E & F** show the corresponding SDS-PAGE of the fractions throughout the sucrose gradient profiles shown in A, B and C, respectively. The black rectangles contain the low molecular weight fractions and the orange rectangles contain the 70S fractions.

30°C has been previously reported, with notably the precipitation and break-down of the complex related to an increase in the diffusion coefficient [172].

4.2.2 NMR studies of ribosomes from thermophilic archaea

In order to explore the possibility of using elevated temperatures to probe ribosome structure, ribosomes from two archaeal sources were investigated. Unlike the well-characterised *E. coli* ribosomes, the purification of archaeal ribosomes is not as well established [218]. Sucrose gradient sedimentation of *Sulfolobus Solfataricus* (*SSF*) or *Pyrobaculum aerophilum* (*P. aero*) ribosomes is sufficient to split the individual subunits (Figure 4.9), and their purity was correlated to the salt concentration used during purification: increasing the salt concentration to 500 mM NH_4Cl was found to be insufficient for removing all extrinsic factors (e.g. elongation factors, EFs, or trigger factor, TF) from the ribosomal subunits, but higher concentrations of NH_4Cl are reported to lead to partial disintegration of the subunits [219, 220]. The separated subunits were quantified by the absorbance at OD_{260} (1 OD unit at 1 cm pathlength corresponds to 60 pmoles/L of 50S subunits and 70 pmoles/L of 30S subunits [218]). 2.5 nmoles of 50S subunits and 4.6 nmoles of 30S subunits were purified from 3.24 g of *SSF* cells using the protocol described in Section 6.1.10.

The proton spectrum of 30S *SSF* subunits resembles that of the 30S *E. coli* subunits (Figure 4.10A), despite there being no archaeal protein homologous to the bacterial S1 protein; therefore the disordered region of another protein must be observable in the 30S *SSF* subunit, which remains to be characterised. ^{1}H-^{15}N SOFAST-HMQC spectrum of 30S *SSF* subunits showed *ca.* 16 resonances, which were clustered at chemical-shifts characteristic of disordered peptide, further indicating that the observed protein in the 30S *SSF* subunit shares the unfolded characteristic with the disordered extension of S1

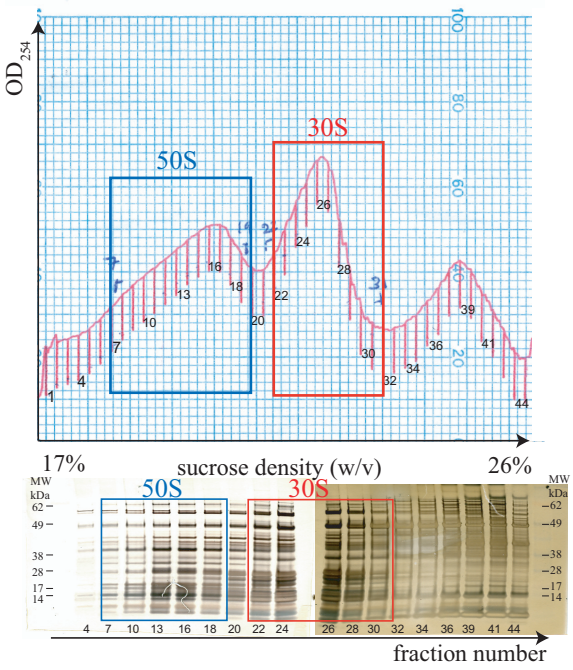


Figure 4.9: OD_{254} profile of the 17%-26% w/v sucrose gradient of *SSF* ribosomes. SDS-PAGE of the numbered fractions is shown below. The fractions that contained pure 50S subunits are in the blue rectangle, and those that contained the 30S subunits are in the red rectangle.

protein in *E. coli* 30S subunit (Figure 4.11D).

The ^{1}H 1D spectrum of 50S *SSF* or *P. aero* subunits presented a cluster of signals with narrow proton chemical shift dispersion in the methyl region (>0.7ppm, Figure 4.10A&C). Archaeal ribosomes also possess a stalk region composed of six copies of L7/L12, which are expected to be observed in the spectra of 50S subunits [221]. The comparison of the methyl resonances of *SSF* or *P. aero* 50S subunits reveals resonances with a reduced chemical shift dispersion than the resonances of *E. coli* L7/L12 (Figure 4.10B & C). Also, the intensity of the resonances (normalised to the concentration of the sample) appears to be lower in the archaeal ribosome spectra than in the corresponding *E. coli* spectra. The ^{1}H-^{15}N SOFAST-HMQC spectrum of 50S *SSF* subunits showed

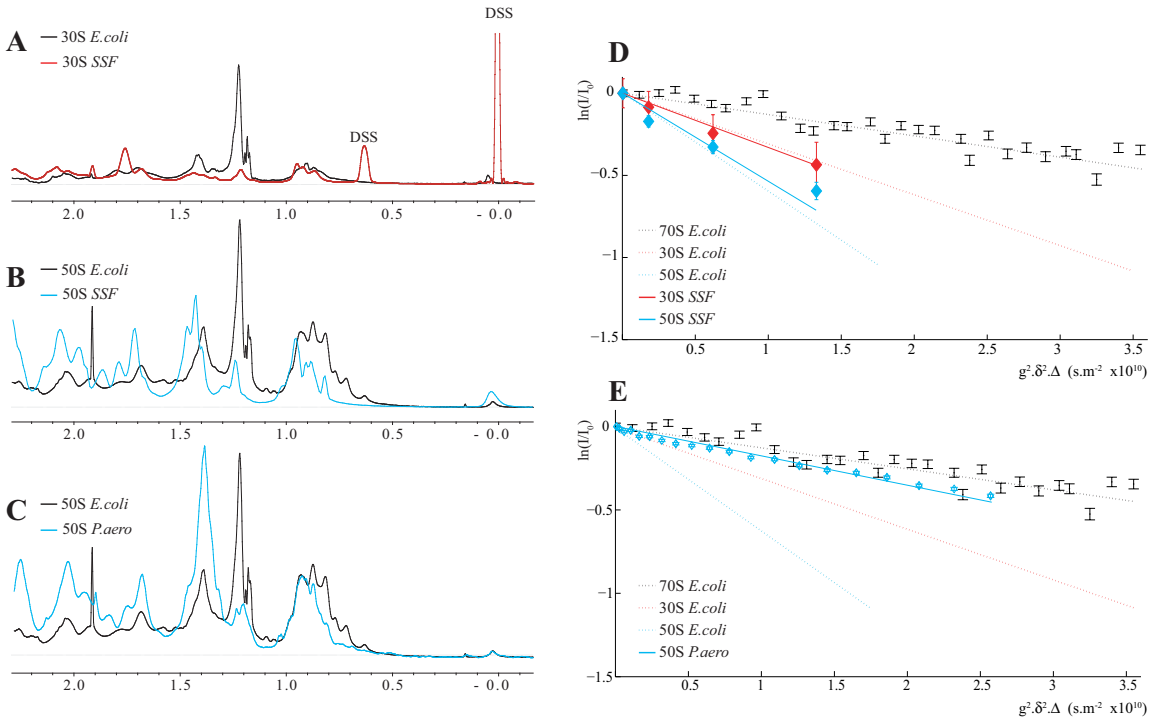


Figure 4.10: Overlays of ^{1}H 1D spectra from different ribosomal subunits. **A:** *E. coli* (black) and *SSF* (red) 30S subunits, **B:** *E. coli* (black) and *SSF* 50S (blue) subunits and **C:** *E. coli* (black) and *P. aero* (blue) 50S subunits. **D:** Plot of the relative signal intensity in ^{1}H STE spectra on a log scale versus the square of the gradient strengths on the x axis, for 70S in black, *E. coli* 50S and 30S in the dotted blue and red respectively, and *SSF* 50S and 30S with the blue and red filled diamonds respectively. **E:** Same as shown in D, for *P. aero* 50S subunit.

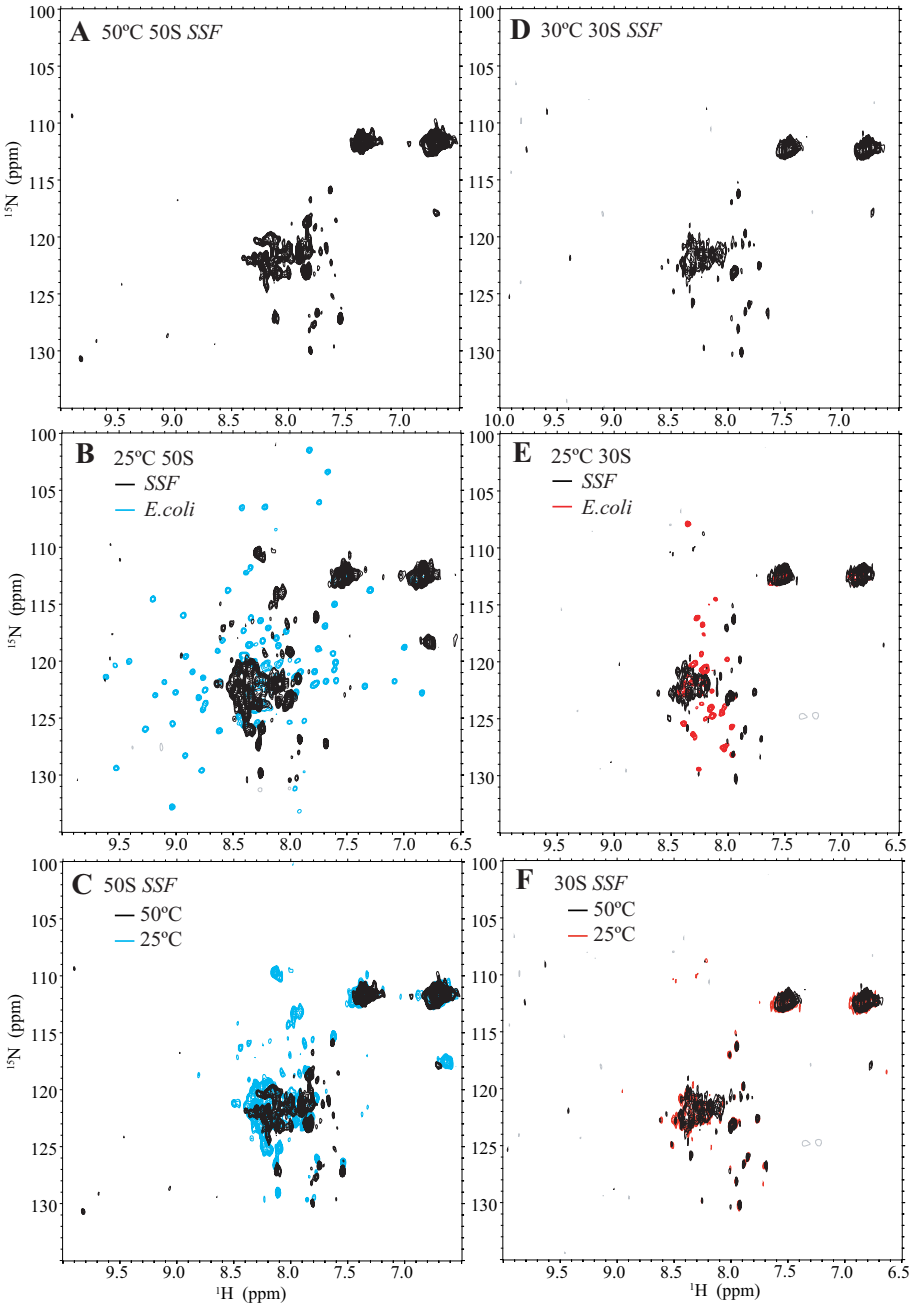


Figure 4.11: **A:** ^{1}H-^{15}N spectrum of *SSF* 50S subunits recorded at 50°C. **B:** ^{1}H-^{15}N spectrum of *SSF* 50S subunits (black), overlaid with that of *E. coli* 50S subunits (blue), both recorded at 25°C. **C:** Overlay of ^{1}H-^{15}N spectra of *SSF* 50S subunits recorded at 25°C (blue) and 50°C (black). **D:** ^{1}H-^{15}N spectrum of *SSF* 30S subunits recorded at 30°C. **E:** ^{1}H-^{15}N spectrum of *SSF* 30S subunits (black), overlaid with that of *E. coli* 30S subunits (red), both recorded at 25°C. **F:** Overlay of ^{1}H-^{15}N spectrum of *SSF* 30S subunits recorded at 25°C (red) and 30°C (black).

a narrow amide ^{1}H chemical shift dispersion with all resonances found between 7.5 - 8.5 ppm (Figure 4.11B), and the amide resonances were found to be broadened at increased temperature indicating that the amide protons undergo exchange with the solvent (Figure 4.11C). Under the conditions that were optimal for the NMR acquisition of the L7/L12 spectrum from 50S *E. coli* ribosomes (25°C), only a few resonances are observed from the 50S *SSF* subunits (<30), and the large differences in chemical shift do not allow the attribution of these observed resonances to the C-terminal domain of L7/L12 (Figure 4.11B).

With six L7 copies per ribosome [221], it is perhaps surprising that the NMR spectra of archaeal 50S subunits did not present the well-resolved L7 resonances observed in the NMR spectrum of *E. coli* ribosomes. However, the limited current knowledge of the biochemistry of the archaeal organism [218] does not preclude the presence of translation factors, such as elongation factors (EFs) [219,220] interacting with the L7/L12 C-terminal domain. In the present work, these could not be excluded from the SDS-PAGE of the 50S or 30S archaeal subunits (Figure 4.9). The presence of EFs can immobilise L7 and result in extensive resonance broadening [19,32,33].

The methyl region of the spectra of the archaeal subunits were used in ^{1}H STE diffusion experiments (Figure 4.10D&E); the resultant diffusion coefficients are summarised in Table 4.1. The 50S *E. coli* and *SSF* subunits were found to have identical diffusion coefficients within experimental error ($5.1\pm0.4\times10^{-11}$ m^2s^{-1} and $4.2\pm0.2\times10^{-11}$ m^2s^{-1}), indicating that the subunits from both species share similar size and conformation. The diffusion of the 50S *P. aero* subunit was slower ($1.7\pm0.2\times10^{-11}$ m^2s^{-1}), closer to the diffusion coefficient reported for the 50S *E. coli* subunits from light-scattering (1.9×10^{-11} m^2s^{-1} [172], Section 4.2.1.2). The diffusion coefficient of the 30S *E. coli* and *SSF* subunits were also identical within error, indicating that ribosome subunits

.

Experiment	**Species**	**Diffusion coefficient** $\times 10^{-11}$ m^2s^{-1}
Light scattering	70S *E. coli*	1.7
NMR	70S *E. coli*	1.7±0.3
Light scattering	50S *E. coli*	1.9
NMR	50S *E. coli*	5.1±0.4
NMR	50S *SSF*	4.2±0.2
NMR	50S *P. aero*	1.7±0.2
Light scattering	30S *E. coli*	2.1
NMR	30S *E. coli*	2.7±0.4
NMR	30S *SSF*	3.2±0.7

Table 4.1: Translation diffusion coefficients determined by ^{1}H STE NMR diffusion or light scattering experiments at 25° for the small and large ribosomal subunits of different organisms. *SSF*: *Sulfolobus solfataricus*. *P. aero*: *Pyrobaculum aerophilum*. Light scattering values are from [172].

from both species share similar hydrodynamic properties.

The stability of the *P. aero* 50S subunits at elevated temperatures (50°C) was assessed using ^{1}H 1D spectra and ^{1}H STE diffusion experiments. After several hours of NMR acquisition (24 hours), no changes in either the proton NMR spectrum or the diffusion coefficient were observed, indicating increased stability of the archaeal ribosomes compared to the *E. coli* ribosomes.

In an attempt to observe slower tumbling regions of the ribosomal complex, ^{1}H 1D spectra were recorded at elevated temperature (50°C) on the 50S *P. aero* subunit. These showed an increase in intensity for the broad resonances, in particular in the amide region of the spectra (Figure 4.12A). ^{1}H 1D spectra for samples in H_2O were generally recorded using water pre-saturation for water suppression, with a typical pre-saturation power corresponding to a B_1 field of 50-100 Hz. Recording ^{1}H 1D spectra on a *P. aero* 50S subunit sample prepared in D_2O allowed examination of the effect of pre-saturation on the intensity of the broad 50S resonances. Comparison of the spectrum recorded without pre-saturation to that recorded with pre-saturation at a power employed for H_2O samples

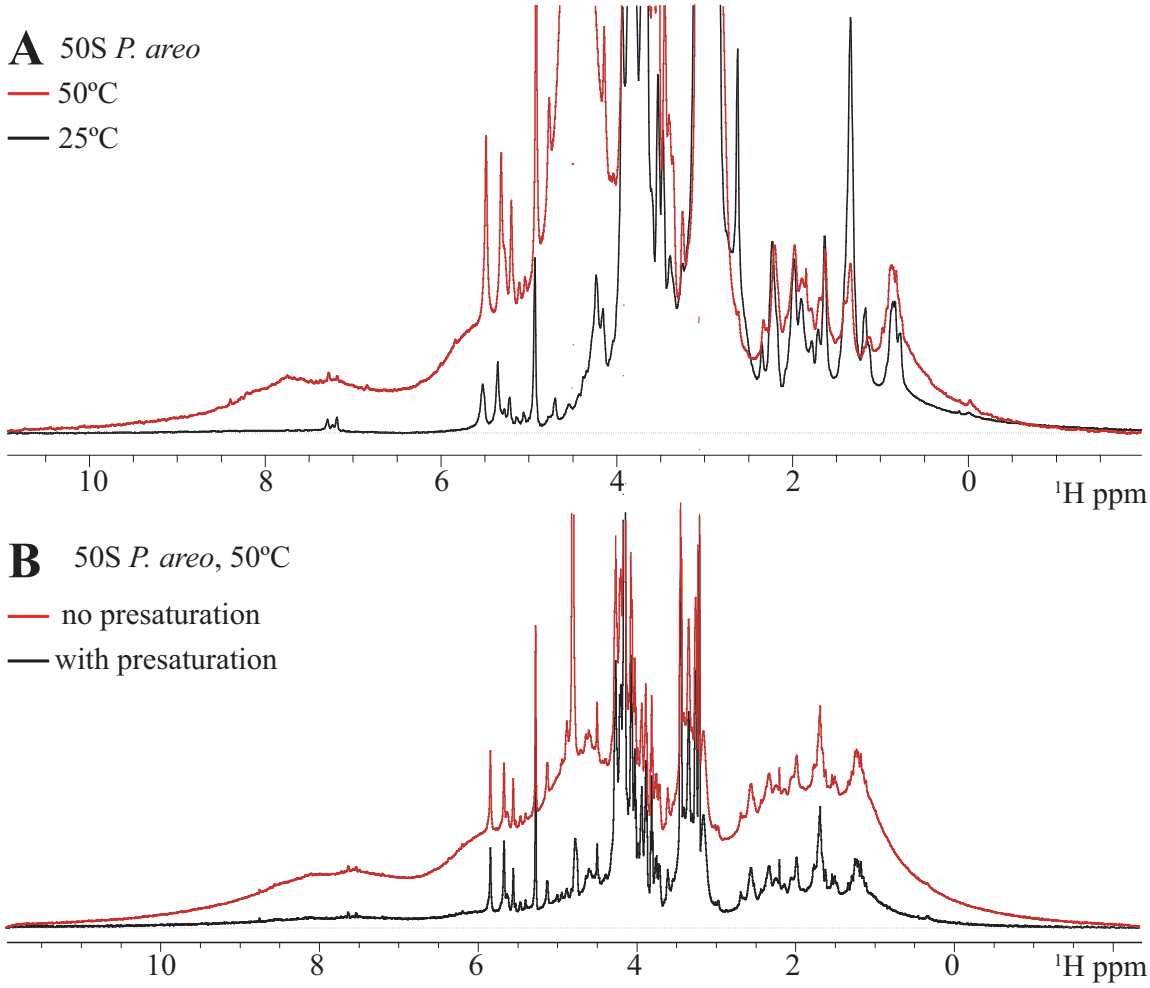


Figure 4.12: A: ^{1}H 1D spectra of *P. aero* 50S subunits recorded at 50°C (red) and 25°C (black). **B:** ^{1}H 1D spectra of *P. aero* 50S subunits recorded at 50°C with and without pre-saturation for one second (black and red, respectively).

indicated that pre-saturation significantly attenuated the intensity of the very broad 50S resonances across the entire spectrum (Figure 4.12B). This indicated that the broad resonances were highly sensitive to cross-relaxation with Hα nuclei, whose resonance frequency lies close to the frequency of water, and are thus saturated during pre-saturation. As was demonstrated in Section 1.3.3, cross-relaxation is dependent on the rotational correlation time, and the cross-relaxation rate is fast for large molecular weight species. The observed cross-relaxation suggests that the broad resonances observed in the ^{1}H 1D spectra arise from rigid regions of the ribosome.

4.3 Concluding remarks

The preliminary NMR investigations of intact ribosomes described in this Chapter revealed the following observations: the L7/L12 C-terminal domain that is observed in conventional ^{1}H-^{15}N SOFAST-HMQC spectra of *E. coli* 70S ribosomes, as well as 10 residues from the linker region between the C-terminal and N-terminal domains are tumbling independently from the ribosome complex. The other proteins that form the stalk region (L10 and parts of L11) as well as the N-terminal domain of L7 and the first 10 residues of the linker appear to be insufficiently flexible to be observable in conventional NMR spectra. The stalk region thus provides a good model for the independent tumbling of a flexible region on the ribosome, a feature which was exploited in the NMR study of RNCs presented in Chapter 3.

In the ^{1}H-^{13}C HSQC spectrum of intact 70S ribosomes, resonances from the C-terminal domain of L7 are observed, with at least one additional Ile resonance that originates from a more rigid region of the 50S subunit. In the ^{1}H-^{13}C methyl-TROSY HMQC spectrum of deuterated Ileδ1 labelled 50S subunits, a number of overlapping resonances are observed, which also arise from rigid regions of the subunit. In the

^{1}H-^{13}C methyl-TROSY HMQC spectrum of deuterated Ileδ1 labelled 30S subunits, at least two resonances that have disperse chemical shifts (Figure 4.6C, peaks labelled with letters A & B) indicate that a folded protein is observable. Although the number of observed resonances does not correspond to anywhere even close to the number of Ile residues in the ribosome complex, the observation of such resonances differentially between conventional and methyl-TROSY based spectroscopy indicates that a range of dynamics must be considered for the study of ribosome complexes: ı) some regions such as the C-terminal domain of L7, tumble independently of the ribosome, ıı) the core of the ribosome complex that contains most of the ribosomal proteins has a rotational correlation time that renders their resonances broadened beyond detection even in ^{1}H-^{13}C methyl-TROSY HMQC spectra and ııı) this study reveals the presence of regions undergoing internal motions that reduces the linewidth of their resonances as modelled in Figure 1.12, Section 1.3.2, such that they can be observed in ^{1}H-^{13}C methyl-TROSY HMQC spectra, although these are not as flexible as the stalk region. The assignment of broad resonances, and the characterisation of these dynamical regions is challenging. A combination of factor-binding can be used together with selective labelling of specific ribosomal protein [222].

The use of archaeal ribosomes permits using increased temperatures to decrease the overall rotation correlation time of the ribosome complex. At these elevated temperatures, broad resonances (^{1}H linewidth on the order of one ppm) are observed, whose intensities were beyond detection in the spectra recorded at 25°C.

Ribose resonances were also observed in ^{1}H-^{13}C HSQC spectra of the 70S ribosome and the corresponding subunits, indicating that some ribosomal RNAs are sufficiently flexible enough to give rise to sharp signals in conventional NMR experiments. The bases of these nucleotides are not observed, suggesting that they might undergo chemical

exchange on the order of ms, or are simply more rigid than the ribose rings. The assignment of such ribose resonances can be performed via selective RNA labelling via an *in vitro* transcription production [223] and RNA-TROSY-based experiments [224]. The development of strategies to generate ribosomal complexes *in vitro* [225–228] will facilitate the production of archaeal ribosome samples [218] with the optimal labelling for NMR of proteins [121] and rRNA [223].

The measurement of translational diffusion of *E. coli* and archaeal ribosome provides insight into conformational changes of the subunits upon splitting. The 50S *E. coli* and *SSF* subunits appear have a reduced hydrodynamic radius compared to that expected from the X-ray structure of the subunits, and that was not the case for the 50S *P. aero* subunits. The 30S *E. coli* and *SSF* subunits may dimerise under the conditions used for the NMR study, despite the fact that indentical conditions were not seen to produce inactive subunits previously [215].

The NMR characterisation of archaeal ribosomes will allow the investigation of ribosome binding factors [219], and demonstrates that NMR can be a useful tool to drive the preliminary studies of the archaeal translational machinery, which has not been investigated comprehensively to date.

Chapter 5

Concluding remarks

This thesis presents a strategy for the use of NMR spectroscopy to study, at a residue specific level, the process of biosynthesis on the ribosome. Central to such a study are the preparative methods for RNCs: the *in vivo* method recently introduced [158] and optimised during the period of this study yielded preparations in which the attached NC was homogeneously stalled during synthesis, selectively isotopically labelled, and at the large quantities required for NMR studies. This allowed the production of all of the RNC complexes described in this thesis. The low levels of background labelling (observed in NMR spectra via the presence of isotopically labeled L7) of the ribosome complex had only a small effect on the analysis of the NMR signals of ddFLN-dom5-RNCs. Emerging strategies for more controlled selective RNC labelling strategies include the manipulation of endoribonucleases to prevent translation of any ribosomal protein during expression [229].

The maximum working ribosome concentration (*ca.* 10 μM) is low compared the concentrations typically used for solution state NMR. Moreover, the sample lifetime is limited to a few days (two to three days) at best (as described in Chapter 2). As shown, this leads to weak NMR signals with achievable values for the signal-to-noise ratio ranging from two to five. This provided the impetus for the development of the methodology for the analysis of low signal-to-noise data of RNCs that was presented in Chapter 2. In this work, the presence of "folded" resonances in the RNC spectra -

where the individual peaks corresponding to native-like ddFLN-dom5 were not always clearly visible "by eye" across different spectra - was assessed not by the observation of individual resonances, but by comparing the distribution of intensities of a number of candidate ddFLN-dom5 resonances to the distribution of the noise. In summary, the knowledge of the peak positions allowed the identification of low signal-to-noise resonances at a statistically significant confidence level, and thus even low populations of the folded or unfolded conformations could be detected. This methodology allowed the detailed analysis of RNC spectra and their comparison with those of the isolated folded ddFLN-dom5, or ddFLN-dom5 denaturated in urea. The comparison with the latter proved useful although differences were observed (Chapter 2, Section 2.3.2.3) and that suggests spectra of C-terminal truncations of the ddFLN-dom5 domain could provide more meaningful comparisons. Preliminary studies have shown that truncations of the last G strand of ddFLN-dom5 renders the domain folding incompetent (as was observed for ddFLN-dom6 in the +110 RNC (Chapter 2)).

This thesis presented various experiments to analyse the folding state of the nascent polypeptide. From the work presented here, the use of rapid acquisition methods, e.g. the ^{1}H-^{15}N SOFAST-HMQC experiment, proved particularly important in probing for disordered states of the RNC. It is best combined with the ^{1}H-^{13}C methyl-TROSY HMQC experiment, which was found to be the most suitable experiment to probe for more compact conformations. These appear to undergo conformational heterogeneity, including possible interactions of the attached nascent polypeptide with the ribosome. The essential function of diffusion NMR experiments in RNC studies is also very clear from this work and these enabled the close correlation of changes in NMR spectra over time with observations using biochemical methods.

The NMR study of ddFLN-dom5 co-translational protein folding (Chapter 3) revealed

that the NMR observable folding mid-point appears to be when the domain is *ca.* 37 residues from the PTC, at which length the low intensities of unfolded ddFLN-dom5 resonances compared to their intensities in RNCs of shorter linker lengths (21 & 31 residues) suggested the presence of "NMR invisible" states. At a linker length of 47 residues, the domain appears to sample an observable native folded state, but with non-native dynamics that suggest exchange with a non-native state a ms timescale. At a linker length of 110 residues, ddFLN-dom5 seems to have near-native-like properties. Heretofore, NMR analysis of the dynamics of RNCs relies on the analysis of the linewidths of the methyl resonances, although spin relaxation experiments [108, 230] currently under investigation could provide more detailed descriptions of the dynamics of RNCs. At the 110 linker length, the resonances observed for ddFLN-dom5 are broader than those for the (potentially analogous) C-terminal domain of L7 suggesting that either the RNC linker sequence (ddFLN-dom6) is less dynamic than the L7 linker, or that ddFLN-dom5 undergoes significant conformational exchange when on the ribosome. Parallel investigations on an entirely disordered nascent chain of alpha-synuclein [231] suggest that interactions with the ribosome are causing resonance broadening. These will need to be further investigated.

Resonance assignment of RNCs was undertaken in this work by comparison with isolated polypeptides. The parallel NMR study of the entire 70S ribosome has revealed resonances that arise from less flexible regions of the ribosome (Chapter 4), and suggests that new resonances can be observed, despite their chemical shift not being known *a priori*. Such an observation indicates that the "NMR invisible" states of the RNC can be observed using a similar strategy. For such resonances (for which the chemical shift is not known *a priori*) assignment will need to be undertaken with the RNCs themselves. For this purpose, selected labelling strategies are possible e.g. site-specific labelling

of proteins through incorporation of position-specific isotopically labelled amino acids [232]. Also likely to be useful for sequential assignment of RNCs is the development of three-dimensional experiments within the *BEST-NMR* suite (band-selective excitation short-transient-NMR experiments) [233], which allow reduced inter-scan delays in a manner analogous to the ^{1}H-^{15}N SOFAST-HMQC experiment. Non-uniform sampling (NUS) [234] also allow reductions in the acquisition time of indirect dimensions without compromising sensitivity or resolution.

The exploration of the ribosome itself using NMR (Chapter 4) revealed not only that a range of dynamics ought to be considered: the dynamics of side chains of the residues observed by NMR, the flexibility of regions of the ribosomes complex (such as the L7/L12 stalk that tumble independently from the ribosome, or regions that undergo internal motions) and the correlation time of the entire ribosome complex; but also that some ribosomal-RNAs (rRNA) have a high flexibility. The observation of both rRNA and proteins from more rigid regions of the ribosome confirms that the ribosome is a very dynamic complex, in which motions on different timescales are to be investigated. The preliminary solution-state NMR observations of slow tumbling regions of the ribosome presented in Chapter 4 does not preclude the potential benefit of solid-state NMR acquisition on sedimented ribosomes [235] in which more rigid regions of the mega-dalton complex may be observed. The benefits of studying archaeal ribosomes have been demonstrated in Chapter 4. *In vitro* reconstitution of complete, functional ribosomes has been established [225–228], and offers the opportunity to use of selective labelling specific ribosomal proteins and rRNAs.

Overall, the success of the NMR strategy developed for obtaining unique residue-specific structural information on the co-translational folding process reveals a wealth of possibilities for exploring the biosynthesis of a range of different topologies e.g: beta

sheets, mixed alpha/beta (e.g. alpha-antitrypsin), multi-domain (e.g. SufI) and knotted (e.g. ubiquitin carboxyl-terminal esterase L1, UCH-L1), and to compare results obtained from NMR with those obtained from biophysical methods. Furthermore, we can begin to explore the influence of the ribosome on the emerging nascent chain, to understand the influence of chaperones such as trigger factor on folding, and to attempt to understand co-translational misfolding. NMR can provide detailed structural information on highly dynamic systems such as RNCs, such as proton distances and relative bond orientations [62] from PRE and RDC measurements, respectively. The data shown in Chapters 2 and 3 show that the spectra of the ddFLN-dom5-RNC are highly reproducible and therefore, it should be possible to record RDCs from RNC samples (and also PREs using suitable labelling strategies). In the case of RDCs, preliminary investigations [236] show that alignment strategies for RDC measurements are likely to be successful. Future work will also investigate H-D exchange methods of the ddFLN-RNCs to provide information on the folding equilibria at progressive lengths. The "basic" NMR parameter, obtained in the studies described in this thesis, the chemical shift, also contains detailed structural information. Indeed, recent developments in molecular dynamics simulations show that accurate structural ensembles can be determined from chemical shifts alone [117–119]. These methods are being investigated using the NMR data obtained in this work for the ddFLN-dom5-RNC. This has the potential to transform our understanding of the development of structure during folding and it has been very exciting being part of the development of this work.

Chapter 6

Materials and methods

The following chapter describes the methodology used in the studies presented in this thesis. It covers the preparation from *E. coli* of selectively isotopically labelled ribosomes, and of ribosome-bound nascent chain complexes (RNCs) that were used in the NMR studies, as well as biochemical analyses used to assess the integrity of the RNCs and the NMR-based strategies that were used to describe the structural and dynamical properties of the samples studied.

All biochemicals and reagents were purchased from Sigma (unless otherwise specified) and were of the highest grade possible.

6.1 Molecular biology and biochemical methods

6.1.1 DNA constructs

Each of the ddFLN-RNCs constructs used (ddFLN-dom5+21, +31, +37, +47, +67, +110-NC) were in the pLDC vector [158] and cloned by Dr Lisa Cabrita. The vector contains ampicillin resistance for selection and a T7 promoter which enables protein expression to be induced using IPTG. These DNA constructs also have an N-terminal hexa-histidine tag to facilitate protein purification and detection and a C-terminal SecM motif which is used to enable stalling of the nascent chain during translation. The ddFLN5 DNA construct used for the preparation of the isolated protein was also in the pLDC vector,

but contained a stop codon at the C-terminus instead of the SecM motif. This clone was generated by Xiaolin Wang.

6.1.2 Bacterial strains

The bacterial strain DH5α (Invitrogen) was used plasmid maintenance. For the expression of isolated proteins, RNCs and for the isolation of ribosomes, BL21-GOLD (DE3) (Stratagene), was used. Both of these bacterial strains were purchased as competent cells.

6.1.3 Growth media composition

The media used for routine bacterial growth was LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl). For plates, the media was supplemented with 7.5 g/L agar, which was added prior to autoclaving.

For media with ampicillin, 100 μg/mL of ampicillin was added after sterilisation, when the temperature was $< 50^{\circ}$C.

For the expression of isotopically-labelled isolated proteins a M9 medium was used. The medium comprised of M9 salts (7.1 g/L Na_2HPO_4, 3.4 g/L KH_2PO_4, 5.84 g/L NaCl, pH 7.4) which was supplemented with 2 mM $MgSO_4$, 100 mM $CaCl_2$, 0.8 g/L yeast nitrogen base without amino acids (Difco), 1 g/L ^{15}N NH_4Cl and 2 g/L ^{13}C glucose.

MDG media without L-aspartic acid ([237], Table 6.1) was used for the preparation of isotopically labelled ribosomes. In the case of RNC preparation, MDG media (Table 6.2) was used for unlabelled growth and enhanced M9 media (Table 6.3) was used for expression of the isotopically labelled NC.

For selectively methyl labelled RNCs, perdeuteration was used, where all media components were prepared in 99% deteurium oxide and deuterated carbon sources were

Compound	Final concentration
Glucose (D-glucose-1,2,3,4,5,6,6-d_7)	2 g/L
$MgSO_4$	2 mM
Trace metals*	0.02%* (w/v)
Na_2HPO_4	25 mM
KH_2HPO_4	25mM
NH_4Cl	50 mM (1 g/L**)
Na_2SO_4	5mM

Table 6.1: MDG media [237] for the production of isotopically labelled ribosomes. In brackets are the modifications applied when producing deuterated ribosomes. * 100% trace metals: 2mM $CoCl_2$, 2mM $CuCl_2$, 2mM $NiCl_2$, 2mM Na_2SeO_3, 2mM $CoCl_2$, 2mM Na_2MoO_4, 2mM H_3BO_3, 10mM $ZnSO_4$, 10mM $MnCl_2$, 20mM $CaCl_2$, 50mM $FeCl_3$. ** For ^{15}N labelling 1 g/L $^{15}NH_4Cl$ was used.

Compound	Final concentration
L-aspartic acid	0.2% w/v
Glucose (d_4 acetic acid)	2 g/L (3g/L)
$MgSO_4$	2 mM
Trace metals	0.02% (w/v)
Ampicillin	0.1 mg/ml
Na_2HPO_4	25 mM
KH_2HPO_4	25mM
NH_4Cl	50 mM
Na_2SO_4	5 mM

Table 6.2: MDG media used for the generation of unlabelled ribosomes for RNC expression. In brackets are the modifications applied when producing deuterated RNCs. Refer to table 6.1 for the composition of 100% trace metal.

Compound	Final concentration
Na_2HPO_4	7 g/L
KH_2PO_4	3.4 g/L
KH_2PO_4	5.84 g/L
Glucose (d_4 acetic acid)	2 g/L (3g/L)
NH_4Cl	1 g/L
$MgSO_4$	2.5 mM
$CaCl_2$	0.2 mM
Trace metal	0.0125% (w/v)
Ampicillin	0.1 mg/mL
BME vitamins (Sigma) (yeast nitrogen base (Difco))	2.5 mL (0.8 g/L)

Table 6.3: Enhanced M9 media for the growth of isotopically labelled ribosomes. In brackets are the modifications applied when producing deuterated RNCs. Refer to table 6.1 for the composition of 100% trace metal.

used. For the preparation of uniform-[^{12}C,^{2}H] , Ileδ1-[$^{13}C^{1}H_3$] RNCs within uniform-[^{12}C,^{2}H] ribosomes, d_4 acetic acid (3 g/L) was used instead of ^{13}C glucose. For the preparation of uniform-[^{12}C,^{2}H], Ileδ1-[$^{13}C^{1}H_3$] ribosomes, D-glucose-1,2,3,4,5,6,6-d_7 (2 g/L) was used instead of ^{13}C glucose. The BME vitamins within the enhanced M9 media was replaced with 0.8 g/L yeast nitrogen base (Difco). The isoleucine precursor (2-ketobutyric acid-4-^{13}C,3,3-d_2 sodium salt hydrate) was used at a concentration of 100 mg/L.

6.1.4 Transformation

To 25 μL of competent *E.coli* cells, 1 μL of a plasmid was added. The cells were incubated on ice for 30 min, followed by a heat shock at 42°C for 45 s. They were then incubated on ice for 30 s and incubated for 1 hour in 900 μL of LB media at 37°C, 200 rpm to allow for expression of the antibiotic resistance gene. The cells were plated on LB-Amp plates. The plates were incubated at 37°C for 12-16 hours, until the colonies were 2-3 mm in size.

6.1.5 Expression of selectively isotopically labeled ribosome-bound nascent chain complexes

3 mL of LB media were inoculated with a single colony, and left to grow overnight at 37°C, 200 rpm. Two flasks of 500 mL of MDG media were inoculated with the overnight pre-culture and left to grow overnight at 37°C, 180 rpm. The optical density at 600 nm (OD_{600}) was measured before the cells were pelleted at 3500 rpm for 15 min. The pellet from 1 L of culture was resuspended in 500 mL of enhanced M9 salts and pelleted again at 3500 rpm for 15 min. The resuspension/pelleting was repeated twice to remove the residual MDG media. The cells were then finally resuspended in 500 mL of enhanced M9 media without isotopes ($^{15}NH_4Cl$ or/and ^{13}Cglucose). The absorbance of the cells at OD_{600} was monitored before the addition of $^{15}NH_4Cl$ and/or ^{13}C glucose isotopes (for uniform isotopic strategy). For RNC expression, IPTG was added to a final concentration of 1 mM and the cells were incubated at 37°C and at 180 rpm. After 10 min of incubation, rifampicin was added to the medium at a final concentration of 0.15 mg/mL. After 1 hour of incubation, the cells were pelleted at 3500 rpm for 20 min and transferred into a 50 mL falcon tube and flash-frozen and stored at -20°C for later purification.

6.1.6 Expression of uniformally deuterated ribosome, Ileδ1 selectively labelled ribosome-bound nascent chain complexes

3 mL of LB media were inoculated with a single colony, and left to grow overnight at 37°C and at 200 rpm. 3 mL of deuterated (80%) MDG media in which the carbon source was protonated glucose, were inoculated with the overnight pre-culture and left to grow for 10 hours at 37°C, 180 rpm. 3 mL of deuterated (100%) MDG media in which the two carbon sources were 1 g/L of protonated glucose and 1.5 g/L of d_4 acetic acid were inoculated with the culture and left to grow for 12 hours at 37°C, 180 rpm. 3 mL of

deuterated (100%) MDG media in which the carbon source was 3 g/L of d_4 acetic acid were inoculated with the culture and left to grow for 10 hours at 37°C, 180 rpm. 100 mL of deuterated (100%) MDG media in which the carbon source was 3 g/L of d_4 acetic acid were inoculated with the culture and left to grow for 12 hours at 30°C, 180 rpm. Two flasks of 500 mL of deuterated (100%) MDG media in which the carbon source is 3 g/L of d_4 acetic acid were inoculated with the culture and left to grow for 20 hours at 30°C, 180 rpm. The absorbance of the cells at OD_{600} was monitored to ensure that the end of the log phase was reached (OD_{600}=2-2.5). The cells were then pelleted at 3500 rpm for 15 min. The pellet from 1 L of culture was resuspended in 500 ml of deuterated (100%) enhanced M9 media with 3 g/L of d_4 acetic acid for RNC expression. The absorbance of the cells at OD_{600} was monitored before the addition of 2-ketobutyric acid-4-^{13}C,3,3-d_2 sodium salt hydrate to a final concentration of 100 mg/L. After 60 min of incubation at 30°C and at 180 rpm, a time that allowed the cell to metabolise the precursor, RNC expression was induced with the addition of IPTG to a final concentration of 1 mM and the cells were incubated at 30°C, 180 rpm. After 10 min of incubation, rifampicin was added to the medium at a final concentration of 0.15 mg/mL. After 2 hours of incubation, the cells were pelleted at 3500 rpm for 20 min and transferred into a 50 mL falcon tube and flash-frozen and stored at -20°C for later purification.

6.1.7 Production of isotopically labelled 70S ribosomes

BL21 *E. coli* cells were streaked onto LB-agar plates, and a single colony was used to inoculate 3 mL of LB media without antibiotic. The pre-culture was incubated overnight at 37°C, 180rpm. Two flasks of 500 mL of MDG without L-aspartic acid and enriched with $^{15}NH_4Cl$ and/or ^{13}Cglucose isotopes were inoculated with the overnight pre-culture and left to incubate overnight at 37°C, 180rpm. The OD_{600} was measured before the cells

were pelleted at 3500 rpm for 15 min. The cells were then resuspended in M9 media enriched with $^{15}NH_4Cl$ and/or ^{13}Cglucose. After 1 hour of incubation, the cells were pelleted at 3500 rpm for 20 min and transferred into a 50 mL falcon tube, flash-frozen and stored at -20°C for later purification.

6.1.8 Purification of RNC

6.1.8.1 Buffer composition

- Buffer B: 500 mM KOAc, 50 mM HEPES, 6 mM MgOAc, 2 mM BME, pH 7.4.
- Lysis buffer: 1 M KOAc, 50 mM HEPES, 12 mM MgOAc, 5 mM EDTA, 5 mM ATP, 2 mM BME, 2 pellet/100mL protease inhibitor (Roche), 1 mM pepstatin, 1 mg/mL lysozyme, traces of DNAse 1, pH 7.4.
- 35% sucrose cushion: 35% w/v sucrose, 1 M KOAc, 50 mM HEPES, 12 mM MgOAc, 5 mM EDTA, 5 mM ATP, 2 mM BME, 0.3 pellet/100mL protease inhibitor (Roche), 1 mM pepstatin, pH 7.4.
- Resuspension buffer: 500 mM KOAc, 50 mM HEPES, 6 mM MgOAc, 5 mM ATP, 2 mM BME, 0.1 pellet/100mL protease inhibitor (Roche), 1 mM pepstatin, pH 7.4.
- RNC elution buffer: 500 mM KOAc, 50mM HEPES, 6 mM MgOAc, 150 mM Imidazole, 2 mM BME, 0.1 pellet/100mL protease inhibitor (Roche), 1 mM pepstatin, pH 7.4.
- Sucrose gradient: Two stock solutions of 35% and 10% w / v sucrose were used to prepare the linear sucrose gradients and both were prepared in the following buffer: 1 M KOAc, 50 mM HEPES, 12 mM MgOAc, 5 mM EDTA, 2 mM BME, 0.1 pellet/100mL protease inhibitor (Roche), 1 mM pepstatin, pH 7.4.

- NMR tico buffer: 30 mM NH_4Cl, 10 mM HEPES, 12 mM $MgCl_2$, 5 mM EDTA, 1 mM BME, 0.1 pellet/100mL protease inhibitor (Roche), 1 μM pepstatin, 1 mM BME, pH 7.4.

6.1.8.2 RNC purification

The cells were resuspended in the lysis buffer, and lysis was performed using four passes in a French Press at 1000 psi. The cellular debris was then pelleted by centrifugation at 18000 rpm for 45 min. After cell lysis the cell lysate was loaded onto a 35% w/v sucrose cushion with volume ratio of 1:3 (cell lysate : sucrose), and subjected to centrifugation for 12 hours at 40 000 rpm (Type 45Ti rotor). The resulting ribosome pellet was disolved in resuspension buffer and released NCs were purified from the supernatant of the sucrose cushion by Ni NTA chromatography. The ribosome suspension was loaded onto 5 mL of Ni IDA beads (Protino) for 1 hour at 4°C. The unbound ribosomes and other impurities were washed from the beads with the resuspension buffer. The RNC were then eluted from the beads using the elution buffer. Following concentration using a filtration device with a 100 kDa cut-off, 1 nmole of RNCs were loaded onto 35 mL 10-35% w/v sucrose gradients formed in SW28 tubes using a manual gradient maker, and sedimented for 15 hours at 21 000 rpm (SW28 rotor). The sucrose gradient was fractionated and the OD_{254} of each fractions of the gradients was monitored using the Foxy Junior system (Presearch). The factions were assessed using SDS-PAGE. The ribosomal fractions are concentrated and buffer-exchanged in tico buffer through a 100 kDa cut-off concentrator to remove sucrose, and any residual low MW co-factors and released ribosomal proteins. The RNCs were flash-frozen and stored at -80°C. Upon defrosting and prior to NMR analysis, the RNC samples were buffer-exchanged with tico buffer using a 100 kDa cut-off concentrator to remove any released proteins.

Marilia Karyadi and Lisa Cabrita helped with the production and purification of ddFLN-dom5+21-RNC and ddFLN-dom5+31-RNC, and Johanna Reul helped with the production of ddFLN-dom5+37-RNCs.

6.1.9 Purification of ribosome samples

The ribosome purification was identical to the RNC purification with the omission of the Ni IDA chromatography step.

Piotr Gierszewski prepared uniform-[^{12}C,^{2} H], Ileδ1-[^{13}C^{1}H$_3$] *E. coli* 70S ribosomes.

6.1.9.1 Purification of SSF ribosomal subunits

SSF cells were grown in ^{15}N isotopically enriched media to produce ^{15}N labelled *SSF* ribosomes (prepared by Fabian Blombach, Germany). The purification of the *SSF* ribosomes followed the same protocol as used for *E. coli* ribosomes, except that a 16-26% w/v sucrose gradient with 500 mM NaCl was used for the sucrose gradient purification step.

Daniel Sohmen produced and purified *P. aero* 50S subunits.

6.1.10 Purification of isolated subunits

To produce the isolated 30S and 50S subunits, purified 70S ribosomes were incubated for 1 hour in tico buffer with 1 mM $MgCl_2$, and loaded on a 16-26% w/v sucrose gradient and sedimented for 15 hours at 21 000rpm.

6.1.11 Sodium Dodecyl Sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

For SDS-PAGE, the protein samples were mixed with LDS loading dye (Invitrogen) and boiled for 5 min. The protein samples were run on 4-12% (w/v) Bis-Tris gels. The gels probing for RNCs or ribosomes were typically silver stained using the SilverQuest silver staining kit (Invitrogen), following the manufacturer's instructions. Gels of isolated proteins were stained with Coomassie Stain (0.025% w/v coomassie brilliant blue, 40% v/v ethanol, 10% v/v glacial acetic acid) and then destained using Destain (40 % v/v ethanol, 10% v/v glacial acetic acid).

6.1.12 Quantification of ribosomes and RNCs concentration using optical density measurements

The concentration of ribosomes and RNC samples was determined using UV spectroscopy at OD_{260}, where 1 OD_{260} in a 1cm pathlength corresponded to 24 nM of 70S ribosomes [184]. An OD_{260}/OD_{280} ratio of two was used to indicated a pure 70S preparation, as this ratio reflects both the rRNA and protein content present within the ribosomal complex [184]. For the component subunits, $1OD_{260}$ was equivalent to either 36 nM of 50S subunits or 72 nM of 30S.

6.1.13 Immunodetection – Western Blot

- Blotting buffer: 6.05 g/L Tris, 14.4 g/L glycine, 20% (v/v) methanol, 0.01%(v/v) SDS. (stored at 4°C)
- Tris Buffered Saline (TBS) : 2.42 g/L Tris, 11.69 g/L NaCl, pH 7.4 (stored at 4°C)
- TBS-Tween: 0.1%(v/v) Tween-20 in TBS

- Blocking buffer:

  Histidine-Tag detection: 20% (w/v) casein in TBS-Tween.

  anti-TF: 0.5% (w/v) skim milk powder in TBS-Tween.

- Antibody dilution:

  Histidine-Tag detection: Penta-His HRP conjugate (Qiagen) (1:5000)

  anti-TF primary antibody: TF rabbit polyclonal antibody (1:2500) (Genscript)

  anti-TF secondary antibody: Horse-radish conjugated anti-rabbit antibody (raised in goat) (1:1000) (New England Biolabs)

- chemiluminescence detection (ECL): SuperSignal West Pico chemiluminescence substrate (Pierce)

After the SDS-PAGE, the gel along with 2 sponges, nitrocellulose membrane and filter paper were soaked in cold blotting buffer. The following were stacked onto the Western apparatus (negative face): sponge, filter paper, gel, nitrocellulose membrane, filter paper, sponge. The "sandwich" was pushed down to release trapped air bubbles. The apparatus was clamped up within the gel running tank and the inner chamber was filled with blotting buffer and the outside chamber with water. The transfer took place for 2 hours at 250 mA on ice. After the transfer, the membrane was placed face-up into a container and allowed to block for one hour with blocking buffer with shaking. The membrane was incubated with the primary antibody for more than 1 hour (typically overnight), then washed four times for 10 min with TBS-Tween. For anti-histidine detection, the membrane was then treated with ECL reagents for 5 min, wrapped in plastic-wrap and then the chemiluminescence was detected with Fujifilm LAS-1000 scanner using exposure times of 30 s to 10 min. For anti-TF detection, the membrane was incubated

with the secondary antibody for 1 hour at room temperature, followed by an additional wash step (four intervals of 10 min) and ECL treatment and detection as described above for anti-histidine detection.

The density of the western blot bands were quantified by using the software ImageK 1.44o, Wayne Rasband, NIH.

6.1.14 Nascent chain and trigger factor occupancy within RNCs

To measure the NC occupancy, 5 and 10 pmoles of RNCs together with a series of purified ddFLN-dom5 standards ranging in amount (2 to 10 pmoles) were run on SDS-PAGE, transferred onto the western blot membrane and probed for the his-tag. Densitometry analysis of the resulting bands enabled a standard curve based upon the ddFLN-dom5 standards to be generated and the NC occupancy was determined from this. The quantification of TF occupancy was performed in a similar manner, except that the standards were purified TF, and these were loaded at amounts ranging from 0.2 to 2 pmoles.

6.1.15 RNC proteolysis time-course

To assess the stability of the RNCs over time, a sample was incubated at 25°C for 3-7 days, during which 10 pmoles samples were collected each 3-6 hours, and run on SDS-PAGE and the his-tag was probed using anti-his western blot (Section 6.1.13).

6.1.16 Expression and purification of isolated ddFLN-dom5

6.1.16.1 Expression of isolated ddFLN-dom5

A single colony was used to inoculate 3mL of LB media. The pre-culture was incubated overnight at 37°C, 200 rpm. 500 mL of M9 enriched with $^{15}NH_4Cl$ and/or ^{13}Cglucose

isotopes was inoculated with the pre-culture and incubated for 6 hours at 37°C, 180 rpm. When the OD_{600} reached 0.6, IPTG was added to the media to a final concentration of 1 mM. After 4 hours of incubation, the cells were pelleted at 3500 rpm for 20 min and transferred into a 50 mL falcon tube, flash-frozen and stored at -20°C.

6.1.16.2 Buffer composition

- Phosphate buffer: 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4.
- Phosphate washing buffer: 25 mM Imidazole, 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4.
- Phosphate elution buffer: 150 mM Imidazole, 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4.

6.1.16.3 Purification of isolated ddFLN-dom5

The cells were resuspended in phosphate buffer supplemented with DNAase and protease inhibitors, and lysis was performed using sonication, performed on ice with interleaved periods of 10s of sonication and 30s of rest (10 cycles). The cellular debris was then pelleted by centrifugation at 18000 rpm for 45 min. The cell lysate was loaded on Ni NTA beads, and the binding was performed for 1 hour at 4°C. The beads were washed with phosphate washing buffer and the protein was eluted with phosphate elution buffer. The sample was then buffer exchanged in phosphate buffer and loaded on a 5 mL Hitrap Q-sepharose column (GE healthcare) using the Akta system. The Q-sepharose column was washed with phosphate buffer and the protein was eluted with a 60 mL gradient of salt from 0 M to 1 M NaCl and 1.5 mL fractions were collected. The sample was then concentrated with a 10 kDa cut-off concentrator and loaded on a Superdex 75 16/60 column (GE healthcare) in tico buffer and 1 mL fractions of the proteins were collected.

Pure fractions of ddFLN-dom5 (as assessed by SDS-PAGE) were pooled and concentrated using a 10 kDa cut-off concentrator. The final concentration of protein is determined using the OD_{280}, and the absorbance coefficient (for ddFLN-dom5, ϵ=5960 $L.mol^{-1}cm^{-1}$) was calculated from the EXPASY website using the sequence .

Released NC extracted from the purification of RNC are purified using the same protocol for Ni-NTA chromatography.

6.2 NMR experiments

All NMR experiments on RNCs were recorded on a 700 MHz Bruker Avance III spectrometer at 25°C, using Topspin 2.1. The water frequency was calibrated by placing the water on-resonance in a simple presaturation-1D. The calibrated water frequency was always the carrier frequency of the proton channel. The pulse length for the 90° hard pulse was calibrated by finding the 360° null at the maximum power level. 2D heteronuclear correlation spectra of RNCs were recorded in an interleaved fashion with 1D ^{1}H, ^{1}H STE, ^{15}N, ^{13}C XSTE or ^{13}C-edited STE diffusion experiments.

6.2.1 ^{1}H 1D

^{1}H 1D experiments were run with excitation sculpting for water suppression, with an acquisition time of 50 ms and a spectral width of 21 ppm. The number of scans varies from 128 to 512.

6.2.2 ^{1}H STE diffusion

^{1}H STE diffusion experiments were run using bi-polar gradients. Smoothed-square gradient pulses of 2 ms were used in the bi-polar encoding & decoding spin echoes. A diffusion delay of 100 ms separated the encoding & decoding gradient pairs. Three

gradient strengths were recorded: 5%, 50% and 95% of the maximum gradient strength (0.563 $T.m^{-1}$). The acquisition time was set to 50 ms, and the spectral width to 25 ppm. The number of scans varied from 128 to 512.

6.2.3 ^{15}N XSTE diffusion

^{15}N XSTE diffusion experiments were run using bipolar encoding gradients (2 ms smoothed-square gradient pulses) during the first spin-echo of the the first refocused-INEPT and decoding gradients during the second spin-echo of the last refocused-INEPT as published by Ferrage *et al* [166]. Additional water presaturation pulses were added during the recycling delay and during the diffusion delay (after the first refocused-INEPT). A diffusion delay of 100 ms separated the bipolar encoding/decoding gradient pairs. Three gradient strengths were recorded: 5%, 50% and 95% of the maximum gradient strength. The ^{1}H acquisition time was set to 50 ms during which ^{15}N decoupling was applied using the garp4 decoupling sequence, and the spectral width was set to 25 ppm. Data were accumulated for one to two hours.

6.2.4 ^{13}C STE diffusion

6.2.4.1 ^{13}C XSTE diffusion

^{1}H-^{13}C XSTE diffusion experiments were run identically to the ^{1}H-^{15}N XSTE diffusion experiment, but the delay of the second spin-echo in the first refocused-INEPT and the first spin-echo of the last refocused-INEPT was optimised for $^{13}CH_3$ groups, i.e. $1/10J_{CH}$ instead of $1/4J_{CH}$. The reason for adjusting this delay is to optimise the J-coupling magnetisation transfer for a CH_3 group.

6.2.4.2 ^{13}C-edited STE-HMQC diffusion

^{13}C-edited STE-HMQC experiments were run with the STE elements prior to a ^{1}H-^{13}C HMQC sequence, as in Didenko *et al* [167]. Bipolar 2 ms smoothed-square gradient pulses were used, with a 100 ms diffusion delay. The experiment was run in a pseudo 2D fashion (with no ^{13}C chemical shift evolution), with three gradient strengths: 5%, 50% and 95% of the maximum gradient strength. During ^{1}H acquisition (52 ms) ^{13}C decoupling was applied using the garp4 decoupling sequence. Data were accumulated for 15 min to one hour.

6.2.5 ^{1}H-^{15}N SOFAST-HMQC

^{1}H-^{15}N heteronuclear experiments were run in the SOFAST-HMQC fashion [159]. PC9 (1408 ms) and REBURP (857 ms) selective pulses were used for 120° excitation and inversion respectively, centred at 8.5 ppm. Low power ^{15}N decoupling was used during the 46 ms ^{1}H acquisition (garp4 decoupling sequence with the 90° pulse lengthened from 107 μs to 350 μs). The recycling delay was shortened to 50 ms. The ^{15}N acquisition time was limited to 14.1 ms during which the ^{13}C nuclei were decoupled using a ^{13}C adiabatic broadband CHIRP inversion pulse. The ^{15}N spectral width was 32 ppm (64 increments), and the ^{1}H spectral width was 16 ppm. Data were accumulated for 45 min to three hours.

6.2.6 ^{1}H-^{13}C HMQC & ^{1}H-^{13}C methyl-TROSY HMQC

^{1}H-^{13}C HMQC experiments were run using a standard HMQC sequence with additional water presaturation during the 1 s recycling delay, a clean-up gradient-pair around the ^{1}H 180 refocusing pulse, and removal of the ^{13}C equilibrium magnetisation by an additional ^{13}C 90° pulse followed by a PFG [145, 146]. The ^{13}C acquisition time (t_1)

was set to 3.3 ms (10 ms in the case of ^{1}H-^{13}C methyl-TROSY HMQC), with a 110 ppm spectral width (40 ppm in the case of ^{1}H-^{13}C methyl-TROSY HMQC). No ^{15}N decoupling was performed. The ^{1}H acquisition time is set to 52 ms during which ^{13}C decoupling was applied using the garp4 decoupling sequence. The ^{1}H spectral width was 14 ppm. Data were accumulated for two to four hours.

6.2.7 Labelling experiments

"Labelling experiments" to assess the ^{13}C (^{15}N) isotopic labelling were performed using a standard ^{1}H-^{13}C HMQC experiment (^{1}H-^{15}N SOFAST-HMQC) for which the phase cycling of the receiver was modified to either select or reject ^{1}H-^{13}C (^{1}H-^{15}N) magnetisation. Water suppression was performed using a presaturation pulse during the 1.3 s recycling delay. Signal intensities from both experiments (^{13}C/^{15}N-^{1}H edited or filtered) were directly comparable, and proportional to the relative population of isotopically labelled and unlabelled proteins.

6.2.8 Hydrogen-deuterium exchange

For hydrogen-deuterium exchange (Section 3.2.1), the sample was dialysed in 100 mM bi-carbonate and lyophilised for 4-12 hours. The lyophilised sample was disolved in deuterated tico buffer with varying pH* from 6.15 to 7.7, and immediately inserted in the 700 MHz NMR spectrometer to record a series of 5 min ^{1}H-^{15}N SOFAST-HMQC spectra. The spectra were processed using nmrPipe (package version 2009), using a lorentz-to-gaussian window function (GMB in nmrPipe) in both dimensions. The resonances were fitted with a gaussian lineshape and the simulated resonances were used to measure the signal intensity. The decay of signal intensity over time was fitted with a single exponential to extrapolate the hydrogen-deuterium exchange rate k_{ex}. Refer to Section

3.1.2 for details.

6.3 Processing NMR data

NMR experiments were processed using nmrPipe (package version 2009) [238]. In some cases, Topspin 2.1 (Bruker) was used. When stated, the data was imported in Matlab (MathWorks, Matlab release R2010b).

6.3.1 ^{1}H 1D

^{1}H 1D were converted from Bruker to nmrPipe format, a cosine-bell window function was applied to the data before zero-filling to double the size of the data. After Fourier transformation, the imaginary data were discarded. A zero-order phase correction was then applied, followed by a linear baseline correction. The nmrPipe script is shown in Supplementary Material Script 1. The data was then converted from nmrPipe format to text format, and imported into Matlab. The time at which each spectrum was recorded was also imported in Matlab.

- The signal integral from 0.75 ppm to 0.85 ppm was calculated for each spectrum and plotted against the time at which those spectra were recorded. The uncertainties of these integrals were calculated as the standard deviation over the same bandwidth at -0.5ppm and shown with error-bars. This methyl region of the spectrum is referred to as the “ribosome bound signal”.

- Similarly, the signal integral from 0.94 ppm to 1.0 ppm was calculated and this methyl region of the spectrum is referred to as the “ribosome-released signal”.

- Similarly, the signal integral from 3.7 ppm to 3.9 ppm was calculated and this region of the spectrum is referred to as the “rRNA signal”.

- Maximum signal intensity (corresponding to a sharp signals of a small metabolite molecule) at 1.8 ppm (as well as 2.85 ppm) was plotted over time and is referred to "metabolite signal".
- Maximum signal intensity at 5.4 ppm was plotted over time and is referred to as "residual sucrose".
- The chemical shifts of the HEPES signal at 3.85 ppm-3.95 ppm was plotted over time and is referred to as "HEPES chemical shift".

6.3.2 ^{1}H STE diffusion spectra

First, the diffusion coefficient for the high-molecular-weight particles giving rise ^{1}H signals at 0.8 ppm was estimated using the ratio of their intensities in the 50% and 95% gradient strength spectra, and the diffusion coefficient for the low-molecular-weight particles giving rise to ^{1}H signals at 1ppm was estimated using the ratio of their intensities in the 5% and 50% gradient strength spectra. This was done in Topspin 2.1. The diffusion coefficients was calculated based on the following equation:

$$D = \frac{ln(I_2/I_1)}{((G_1\gamma_{\mathrm{H}})^2 - (G_2\gamma_{\mathrm{H}})^2) \times (\Delta - \frac{\delta}{3})} \tag{6.1}$$

where γ_{H} is the ^{1}H gyromagnetic ratio (2.67522128$\times 10^8$ rad.s^{-1}.T^{-1}), I_1 and I_2 are the intensities at the two different gradient strengths, Δ is the diffusion delay (100 ms), δ is the length of the gradient pair (4 ms), and G_1 and G_2 are the gradient strengths calculated according to:

$$g_{\mathrm{i}} = g_{\%}.G_{\mathrm{max}}.F_{\mathrm{shape}}.\delta \tag{6.2}$$

where $g_{\%}$ is the % of maximum gradient strength, G_{max} is the maximum gradient strength (0.563 $T.m^{-1}$), F_{shape} is the shape factor related to the integral of the gradient used (F_{shape}=0.63 for a sine gradient and F_{shape}=0.9 for a smoothed-square gradient).

^{1}H STE spectra were converted from Bruker to nmrPipe format, and process in an identical manner as the ^{1}H 1D spectra, and imported into Matlab.

- For each of the gradient strength increments, the signal integral from 0.75 ppm to 0.85 ppm was calculated and plotted against the time at which the spectra were recorded. The uncertainties of these integrals were calculated as the standard deviation over the same bandwidth at -0.5 ppm and shown with error-bars. This methyl region of the spectrum is referred to as the "methyl 0.8 ppm" signal.
- Similarly, the signal integral from 0.94 ppm to 1.0 ppm was calculated and this methyl region of the spectrum is referred to as the "methyl 0.9 ppm" signal.
- Similarly, the signal integral from 3.7 ppm to 3.9 ppm was calculated and this methyl region of the spectrum is referred to as the "ribose" signal.

6.3.3 ^{15}N XSTE diffusion spectra

Data were processed in a similar fashion to the ^{1}H STE spectra in Topspin 2.1. Data were also converted to nmrPipe format and processed in a similar fashion to the ^{1}H STE spectra. Once exported to Matlab, the the bandwidth of the signal integral was from 8.0 to 8.5 ppm.

6.3.4 ^{13}C-edited STE-HMQC diffusion spectra

Data were processed in a similar fashion as the ^{15}N XSTE diffusion spectra. Once exported to Matlab, the bandwidth of the signal integral was from 0.75 to 1.0 ppm.

6.3.5 ^{1}H-^{15}N SOFAST-HMQC spectra

^{1}H-^{15}N SOFAST-HMQC spectra were exported to nmrPipe. A simple cosine-bell window function was applied to the ^{1}H dimension before zero-filling to double the size of the data. After Fourier transformation, the imaginary data were discarded. A zero-order phase correction was applied and the data were then transposed. A simple cosine-bell window function was applied to the ^{15}N dimension before zero-filling to double the size of the data. After Fourier transformation, the imaginary data was discarded. A zero-order phase correction of -90° and a first-order phase correction of 180° were applied followed by a linear baseline correction in the ^{15}N dimension. The data was then transposed again and a linear baseline correction in the ^{1}H dimension was applied. The nmrPipe script is shown in Supplementary Material script 2. Figures were produced with spectra processed using linear prediction in the indirect dimension, and zero-filling to 4096 points in the direct dimension and 1024 points in the indirect dimension.

Three to five consecutive ^{1}H-^{15}N SOFAST-HMQC spectra were added in a moving average fashion using the nmrPipe library script addNMR. The moving averages were then exported to text files and opened in Matlab. The correlation between neighbouring points was determined using a region devoid of resonances of the spectrum for ^{1}H chemical shift of 10.5 to 11 ppm (refer to Section 2.2 for details, and the Matlab script is shown in Supplementary Material Script 3). The noise levels of the spectra were determined using the standard deviation over 100 points chosen randomly between 6.5 and 10.5 ppm (^{1}H chemical shift), which do not overlap with resonances of ddFLN-dom5 or L7/L12. The distribution of maxima within regions of interest (ROI) of 3×3 datapoints (Section 2.2) of the noise and signals were used to analyse the signal intensities (refer to Section 2.2 for details, and the Matlab script is shown in Supplementary Material Script 4).

6.3.6 ^{1}H-^{13}C HMQC

^{1}H-^{13}C HMQC spectra were processed in a similar fashion as the ^{1}H-^{15}N SOFAST-HMQC spectra.

One to ten consecutive ^{1}H-^{13}C HMQC were summed up, exported to text files and opened in Matlab. The correlation between neighbouring points was determined using a empty region of the spectrum for ^{1}H chemical shift of -0.4 to -0.2 ppm (refer to Section 2.2 for details). The noise of the spectra are determined using the standard deviation between 100 points chosen randomly between -0.2 and 1.0 ppm (^{1}H chemical shift), which do not overlap with resonances of ddFLN-dom5 or L7/L12.

6.4 Other structural analyses

6.4.0.1 Structural alignment of immunoglobulin domains

The sequence alignment of ddFLN-dom5, ddFLN-dom6, ddFLN-dom4, fnIII and TI was performed in Chimera UCSF, with a restraint from the structural alignment of 30%.

6.4.0.2 Model of hydrodynamic radii from structure

The hydrodynamic radii were modelled from the pdb structures using Hydropro [214], using a minimum bead size of 5 and a maximum bead size of 10.

Appendix A

Supplementary Material

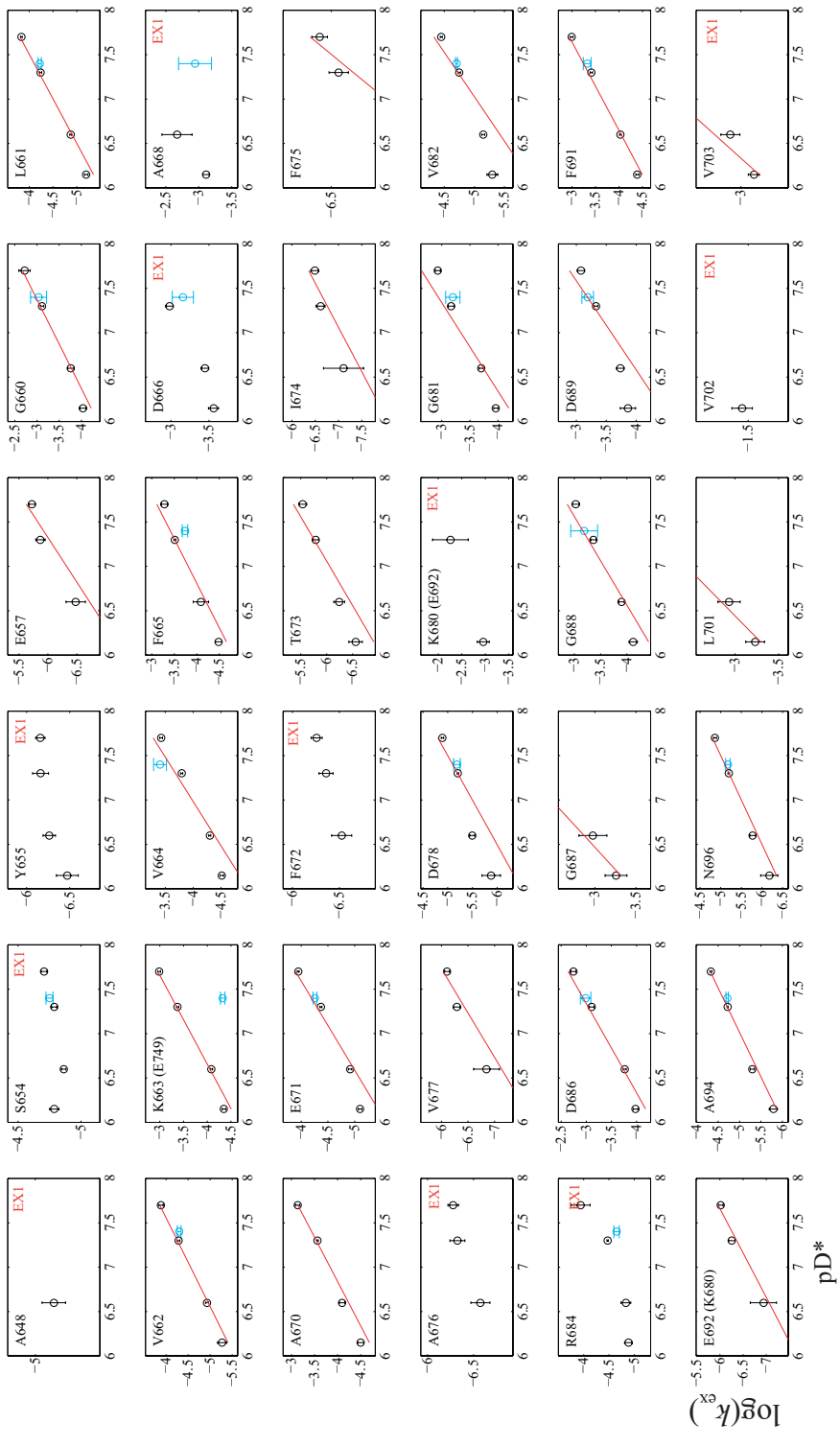


Figure A.1: Plot of the H-D exchange rate as a function of the pH measured in D_2O for each residues of purified isolated ddFLN-dom5 (black circles) are shown together with those measured from ddFLN-dom5+6 (blue circles). The y axis is the $\log_{10}$ of k_{ex}. When the exchange was found to be in the EX2 regime, the red line shows $\log_{10}(k_{int})$-$\log_{10}(P)$. This figure is showing the data for the first 36 residues. The last 40 residues are shown in Figure A.2.

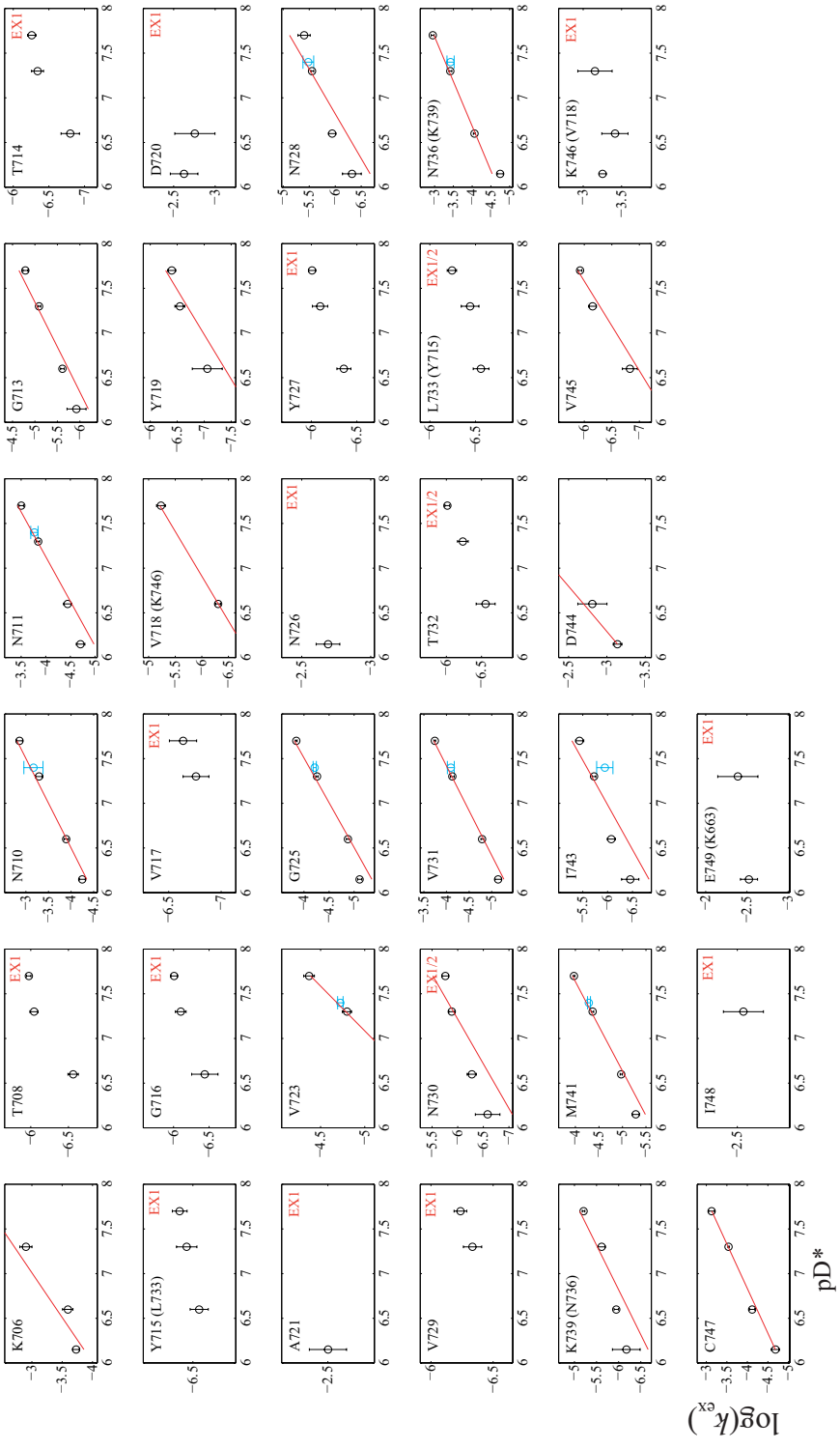


Figure A.2: Refer to Figure A.1.

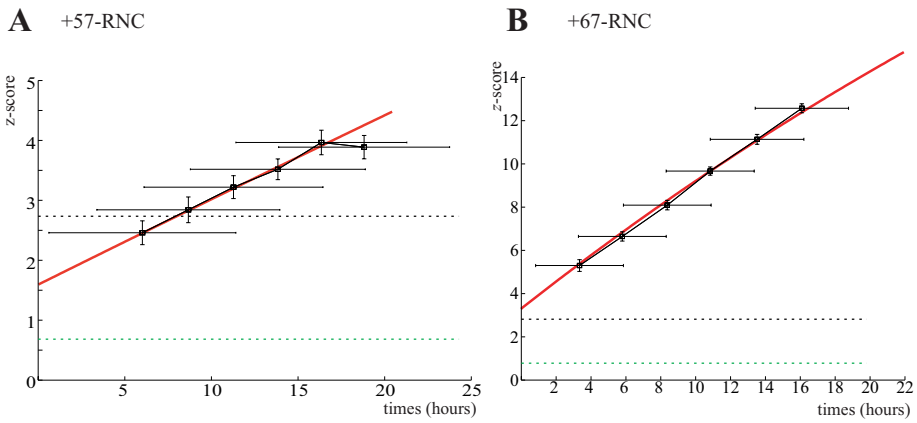


Figure A.3: z-score of ^{1}H-^{13}C high field shifted methyl resonances for folded ddFLN-dom5 in ddFLN-dom5+57-RNC in **A** and ddFLN-dom5+67-RNC in **B**. The y axis is the mean z-score of the 16 signal intensity (maximum within a 3×3 region of the spectra centred at the resonances). The x axis is the time of acquisition in hours. The y error bars are the standard deviation of the random noise of the individual spectra divided by $\sqrt{16}$. The x error bars shows the time of acquisition. The black dotted line shows the 5% threshold of the nine-variate distribution of the random noise. The green green line shows the 5% threshold divided by $\sqrt{16}$. The red curve is the fit of equation 2.8, with a time constant of 66 $\pm$ 33 hours, R^2 0.90 for A and 65 $\pm$ 35 hours, R^2 0.999 for B.

```
bruk2pipe -in ./fid \
-bad 0.0 -aswap -AMX -decim 1418.66666666667 -dspfvs 20 -grpdly 67.9896545410156 \
-xN 2048 \
-xT 1024 \
-xMODE DQD \
-xSW 14097.744 \
-xOBS 700.133 \
-xCAR 4.773 \
-xLAB 1H \
-ndim 1 \
-out ./test.fid -verb -ov
nmrPipe -in test.fid \
| nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2.0 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 $phase -p1 0.0 -di -verb \
| nmrPipe -fn EXT -x1 -4ppm -xn 13ppm -sw \
| nmrPipe -fn BASE -nw 20 -nl -1.5ppm -0.5ppm 10.5ppm 11.5ppm \
-ov -out test.ft1
pipe2txt.tcl -index PPM test.ft1 >1D.txt
```

Script 1: nmrPipe script for the conversion and processing of 1D ^{1}H spectrum. **$phase** was estimated manually, using the visualisation software nmrDraw.

```
bruk2pipe -in ./ser \
-bad 0.0 -aswap -DMX -decim 1792 -dspfvs 20 -grpdly 67.9841766357422 \
-xN 1024 -yN 64 \
-xT 512 -yT 32 \
-xMODE DQD -yMODE States-TPPI \
-xSW 11160.714 -ySW 2270.405 \
-xOBS 700.133 -yOBS 70.952 \
-xCAR 4.773 -yCAR 117.066 \
-xLAB 1H -yLAB 15N \
-ndim 2 -aq2D States \
-out ./test.fid -verb -ov
nmrPipe -in test.fid \
| nmrPipe -fn SP -off 0.45 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 $phase -p1 0.00 -di -verb \
| nmrPipe -fn EXT -left -sw \
| nmrPipe -fn TP \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 -90.00 -p1 180.00 -di -verb \
| nmrPipe -fn BASE -nw 6 -nl 103ppm 105ppm 131ppm 133ppm \
| nmrPipe -fn TP \
| nmrPipe -fn BASE -nw 20 -nl 11ppm 10.8ppm 6.2ppm 5.5ppm \
-ov -out test.ft2
pipe2txt.tcl -index PPM test.ft2 >2D.txt
```

Script 2: nmrPipe script for the conversion and processing of 2D ^{1}H-^{15}N SOFAST-HQMC (similar scripts were used for ^{1}H-^{13}C HMQC spectrum).

```
s=0;
for k=1:9
for l=1:9
s=s+1;
row(s,1)=k;
row(s,2)=l;
for i=1:(noise_size_row-l)
for j=1:(noise_size_row-k)
neighbours_data((N15size-k+1)*i-(N15size-k+1)+j,1)=specNoise_data(i,j);
neighbours_data((N15size-k+1)*i-(N15size-k+1)+j,s)=specNoise_data((i+l-1),(j+k-1));
end
end
[rho, pval, rho_l, rho_u] = corrcoef(neighbours_data(:,1);neighbours_data(:,s));
rho_data((k),(l))=rho_temp(1,2);
pval_data((k),(l))=pval_temp(1,2);
rho_low_data((k),(l))=rho_low_temp(1,2);
rho_up_data((k),(l))=rho_up_temp(1,2);
end
end
for i=1:9
sigma_corr_temp(i,i)=1;
end

sigma_corr_temp(1,2)=rho_data(1,2);
sigma_corr_temp(2,1)=rho_data(1,2);
sigma_corr_temp(3,2)=rho_data(1,2);
sigma_corr_temp(2,3)=rho_data(1,2);
sigma_corr_temp(5,4)=rho_data(1,2);
sigma_corr_temp(4,5)=rho_data(1,2);
sigma_corr_temp(5,6)=rho_data(1,2);
sigma_corr_temp(6,5)=rho_data(1,2);
sigma_corr_temp(7,8)=rho_data(1,2);
sigma_corr_temp(8,7)=rho_data(1,2);
sigma_corr_temp(9,8)=rho_data(1,2);
sigma_corr_temp(8,9)=rho_data(1,2);
sigma_corr_temp(1,4)=rho_data(2,1);
sigma_corr_temp(4,1)=rho_data(2,1);
sigma_corr_temp(5,2)=rho_data(2,1);
sigma_corr_temp(2,5)=rho_data(2,1);
sigma_corr_temp(6,3)=rho_data(2,1);
sigma_corr_temp(3,6)=rho_data(2,1);
sigma_corr_temp(4,7)=rho_data(2,1);
sigma_corr_temp(7,4)=rho_data(2,1);
sigma_corr_temp(1,4)=rho_data(2,1);
sigma_corr_temp(4,1)=rho_data(2,1);
sigma_corr_temp(5,8)=rho_data(2,1);
sigma_corr_temp(8,5)=rho_data(2,1);
sigma_corr_temp(9,6)=rho_data(2,1);
sigma_corr_temp(6,9)=rho_data(2,1);
sigma_corr_temp(1,4)=rho_data(2,1);
sigma_corr_temp(4,1)=rho_data(2,1);
sigma_corr_temp(1,5)=rho_data(2,2);
sigma_corr_temp(5,1)=rho_data(2,2);
sigma_corr_temp(1,5)=rho_data(2,2);
sigma_corr_temp(5,1)=rho_data(2,2);
sigma_corr_temp(2,4)=rho_data(2,2);
sigma_corr_temp(4,2)=rho_data(2,2);
sigma_corr_temp(2,6)=rho_data(2,2);
sigma_corr_temp(6,2)=rho_data(2,2);
sigma_corr_temp(3,5)=rho_data(2,2);
sigma_corr_temp(5,3)=rho_data(2,2);
sigma_corr_temp(4,8)=rho_data(2,2);
sigma_corr_temp(8,4)=rho_data(2,2);
sigma_corr_temp(7,5)=rho_data(2,2);
sigma_corr_temp(5,7)=rho_data(2,2);
sigma_corr_temp(5,9)=rho_data(2,2);
sigma_corr_temp(9,5)=rho_data(2,2);
sigma_corr_temp(6,8)=rho_data(2,2);
sigma_corr_temp(8,6)=rho_data(2,2);
sigma_corr_temp(1,3)=rho_data(1,3);
sigma_corr_temp(3,1)=rho_data(1,3);
sigma_corr_temp(4,6)=rho_data(1,3);
sigma_corr_temp(6,4)=rho_data(1,3);
sigma_corr_temp(9,7)=rho_data(1,3);
sigma_corr_temp(7,9)=rho_data(1,3);
sigma_corr_temp(1,6)=rho_data(2,3);
sigma_corr_temp(6,1)=rho_data(2,3);
sigma_corr_temp(4,9)=rho_data(2,3);
sigma_corr_temp(9,4)=rho_data(2,3);

sigma_noise=std(specNoise_data(:));
sigma(:,:)=sigma_corr_temp(:,:)*sigma_noise*sigma_noise;
```

Script 3: Matlab commands used to calculate the correlation between neighbouring points (**rho_data**). **specNoise_data** is a region of the spectrum where there is only noise (10-10.5ppm in the ^{1}H dimension and 110-130ppm in the ^{15}H). The correlation matrix (**sigma**) is build for use in script 4.

```
for i=1:243
noise_intensity_data(1,i)=spec_data(noise_position(i,1),noise_position(i,2));
end
std_noise_data=std(noise_intensity_data);

for j=1:243
noise_intensity_data(1,j)=max(reshape(spec_data(noise_position(j,1)-1:noise_position(j,1)+1,noise_position(j,2)-1:noise_position(j,2)+1),1,9));
end

for i=1:73
peak_intensity_data(1,i)=max(reshape(spec_data(peak_position(i,1)-1:peak_position(i,1)+1,peak_position(i,2)-1:peak_position(i,2)+1),1,9));
end

step_x=(max(peak_intensity_data(1,:))-min(noise_intensity_data(1,:)))/50;
step_y=(max(noise_intensity_data(1,:))-min(noise_intensity_data(1,:)))/50;
x=(min(noise_intensity_data(1,:)-10*step_x):step_x:max(peak_intensity_data(1,:))+10*step_x);
y=(min(noise_intensity_data(1,:)-10*step_y):step_y:max(noise_intensity_data(1,:))+20*step_y);
[noise_distribution_data]=histc(noise_intensity_data,y);
[peak_distribution_data]=histc(peak_intensity_data,x);

size_n=size(y);
for i=1:9
mu(1,i)=mean(noise_intensity_data(1,:));
end
for i=1:size_n(1,2)
for k=1:9
y_mvncdf(i,k)=[y(1,i)];
end
Mvn_pdf(:,i)=mvnpdf(y_mvncdf(i,:),mu(1,9),sigma(1:9,1:9));
F_data(1,i)=mvncdf(y_mvncdf(i,:),[000000000],sigma(1:9,1:9));
end
threshold=find(F_data >= 0.95,1,'first');

size_p=size(x);
mu_peak=mean(peak_intensity_data);
std_peak=std(peak_intensity_data);
for i=1:size_p(1,2)
gaus_p(1,i)=exp(-(x(1,i)-mu_peak)^ 2/(2*std_peak^ 2));
end

hold on
bar(y/std_noise_data,noise_distribution_data/max(noise_distribution_data),'FaceColor',[0.5,0.5,0.5],'EdgeColor','none')
bar(x/std_noise_data,peak_distribution_data/max(peak_distribution_data),0.85,'m','EdgeColor','none')
plot(y/std_noise_data,Mvn_pdf(1,:)/max(Mvn_pdf(1,:)),'k')
plot(x/std_noise_data,Mvn_pdf(1,:)/max(Mvn_pdf(1,:)),'m')
plot([y(threshold)/std_noise_data y(threshold_1)/std_noise_data],[0 1],':k')
```

Script 4: Matlab commands used to generate the intensity distribution plots shown in Chapter 2. **spec_data** is the entire 2D spectrum. **noise_position** contains the positions within the spectrum that does not correspond to RNC or L7/L12 resonances. **peak_position** contains the position of the RNC resonances. **sigma** is defined in Script 3.

Bibliography

[1] Anfinsen, C. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223–30.

[2] Gershenson, A. and Gierasch, L. M. (2011) Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol*, **21**, 32–41.

[3] Dobson, C. M. (2003) Protein folding and misfolding. *Nature*, **426**, 884–90.

[4] Cavagnero, S. (2011) Protein folding at the exit tunnel. *Annu Rev Biophys*, **40**, 337–359.

[5] Cabrita, L. D., Dobson, C. M., and Christodoulou, J. (2010) Protein folding on the ribosome. *Curr Opin Struct Biol*, **20**, 33–45.

[6] Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol*, **16**, 589–97.

[7] Clark, P. L. (2004) Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci*, **29**, 527–534.

[8] Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 a resolution. *Science*, **289**, 905–20.

[9] Naganuma, T., Nomura, N., Yao, M., Mochizuki, M., Uchiumi, T., and Tanaka, I. (2010) Structural basis for translation factor recruitment to the eukaryotic/archaeal ribosomes. *J Biol Chem*, **285**, 4747–56.

[10] Dunkle, J. A. and Cate, J. H. D. (2010) Ribosome structure and dynamics during translocation and termination. *Annu Rev Biophys*, **39**, 227–244.

[11] Dunkle, J. A., Xiong, L., Mankin, A. S., and Cate, J. H. D. (2010) Structures of the escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci USA*, **107**, 17152–7.

[12] Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010) Crystal structure of the eukaryotic ribosome. *Science*, **330**, 1203–9.

[13] Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., and Ban, N. (2011) Crystal structure of the eukaryotic 40s ribosomal subunit in complex with initiation factor 1. *Science*, **331**, 730–6.

[14] Gao, Y.-G., Selmer, M., Dunham, C. M., Weixlbaumer, A., Kelley, A. C., and Ramakrishnan, V. (2009) The structure of the ribosome with elongation factor g trapped in the posttranslocational state. *Science*, **326**, 694–9.

[15] Schmeing, T. M., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-trna. *Nature*, **438**, 520–4.

[16] Gao, H., et al. (2003) Study of the structural dynamics of the e coli 70s ribosome using real-space refinement. *Cell*, **113**, 789–801.

[17] Julián, P., Milon, P., Agirrezabala, X., Lasso, G., Gil, D., Rodnina, M. V., and Valle, M. (2011) The cryo-em structure of a complete 30s translation initiation complex from escherichia coli. *PLoS Biology*, **9**, e1001095.

[18] Villa, E., et al. (2009) Ribosome-induced changes in elongation factor tu conformation control gtp hydrolysis. *Proc Natl Acad Sci USA*, **106**, 1063–8.

[19] Schuette, J.-C., et al. (2009) Gtpase activation of elongation factor ef-tu by the ribosome during decoding. *EMBO J*, **28**, 755–65.

[20] Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70s ribosome complexed with mrna and trna. *Science*, **313**, 1935–42.

[21] Armache, J.-P., et al. (2010) Cryo-em structure and rrna model of a translating eukaryotic 80s ribosome at 5.5-a resolution. *Proc Natl Acad Sci USA*, **107**, 19748–53.

[22] Becker, T., Armache, J.-P., Jarasch, A., Anger, A. M., Villa, E., Sieber, H., Motaal, B. A., Mielke, T., Berninghausen, O., and Beckmann, R. (2011) Structure of the no-go mrna decay complex dom34-hbs1 bound to a stalled 80s ribosome. *Nat Struct Mol Biol*, **18**, 715–20.

[23] Mitra, K. and Frank, J. (2006) Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu Rev Biophys Biomol Struct*, **35**, 299–317.

[24] Schmeing, T. M. and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–42.

[25] Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. D. (2005) Structures of the bacterial ribosome at 3.5 a resolution. *Science*, **310**, 827–34.

[26] Seidelt, B., et al. (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, **326**, 1412–5.

[27] Rodnina, M. V., Gromadski, K. B., Kothe, U., and Wieden, H.-J. (2005) Recognition and selection of trna in translation. *FEBS Lett*, **579**, 938–42.

[28] Daviter, T., Gromadski, K. B., and Rodnina, M. V. (2006) The ribosome's response to codon-anticodon mismatches. *Biochimie*, **88**, 1001–11.

[29] Petrov, A., Kornberg, G., O'Leary, S., Tsai, A., Uemura, S., and Puglisi, J. D. (2011) Dynamics of the translational machinery. *Curr Opin Struct Biol*, **21**, 137–45.

[30] Frank, J. and Gonzalez, R. L. (2010) Structure and dynamics of a processive brownian motor: the translating ribosome. *Annu Rev Biochem*, **79**, 381–412.

[31] Bocharov, E. V., Sobol, A. G., Pavlov, K. V., Korzhnev, D. M., Jaravine, V. A., Gudkov, A. T., and Arseniev, A. S. (2004) From structure and dynamics of protein l7/l12 to molecular switching in ribosome. *J Biol Chem*, **279**, 17697–706.

[32] Diaconu, M., Kothe, U., Schlünzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G., Stark, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural basis for the function of the ribosomal l7/12 stalk in factor binding and gtpase activation. *Cell*, **121**, 991–1004.

[33] Christodoulou, J., et al. (2004) Heteronuclear nmr investigations of dynamic regions of intact escherichia coli ribosomes. *Proc Natl Acad Sci USA*, **101**, 10949–54.

[34] Mulder, F. A. A., Bouakaz, L., Lundell, A., Venkataramana, M., Liljas, A., Akke, M., and Sanyal, S. (2004) Conformation and dynamics of ribosomal stalk protein l12 in solution and on the ribosome. *Biochemistry*, **43**, 5930–6.

[35] Voorhees, R. M., Schmeing, T. M., Kelley, A. C., and Ramakrishnan, V. (2010) The mechanism for activation of gtp hydrolysis on the ribosome. *Science*, **330**, 835–8.

[36] Lee, T.-H., Blanchard, S. C., Kim, H. D., Puglisi, J. D., and Chu, S. (2007) The role of fluctuations in trna selection by the ribosome. *Proc Natl Acad Sci USA*, **104**, 13661–5.

[37] Fei, J., Bronson, J. E., Hofman, J. M., Srinivas, R. L., Wiggins, C. H., and Gonzalez, R. L. (2009) Allosteric collaboration between elongation factor g and the ribosomal l1 stalk directs trna movements during translation. *Proc Natl Acad Sci USA*, **106**, 15702–7.

[38] Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003) Locking and unlocking of ribosomal motions. *Cell*, **114**, 123–34.

[39] Becker, T., et al. (2009) Structure of monomeric yeast and mammalian sec61 complexes interacting with the translating ribosome. *Science*, **326**, 1369–73.

[40] Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D. N., and Beckmann, R. (2011) Secm-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biology*, **9**, e1000581.

[41] Bhushan, S., Gartmann, M., Halic, M., Armache, J.-P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D. N., and Beckmann, R. (2010) alpha-helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol*, **17**, 313–317.

[42] Munro, J. B., Sanbonmatsu, K. Y., Spahn, C. M. T., and Blanchard, S. C. (2009) Navigating the ribosome's metastable energy landscape. *Trends Biochem Sci*, **34**, 390–400.

[43] Whitford, P. C., Geggier, P., Altman, R. B., Blanchard, S. C., Onuchic, J. N., and Sanbonmatsu, K. Y. (2010) Accommodation of aminoacyl-trna into the ribosome involves reversible excursions along multiple pathways. *RNA*, **16**, 1196–204.

[44] Fulle, S. and Gohlke, H. (2009) Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. *J Mol Biol*, **387**, 502–17.

[45] Petrone, P. M., Snow, C. D., Lucent, D., and Pande, V. S. (2008) Side-chain recognition and gating in the ribosome exit tunnel. *Proc Natl Acad Sci USA*, **105**, 16549–54.

[46] Elcock, A. H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput Biol*, **2**, e98.

[47] O'Brien, E. P., Hsu, S.-T. D., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2010) Transient tertiary structure formation within the ribosome exit port. *J Am Chem Soc*, **132**, 16928–37.

[48] O'Brien, E. P., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2011) New scenarios of protein folding can occur on the ribosome. *J Am Chem Soc*, **133**, 513–526.

[49] Levinthal, C. (1968) Are there pathways for protein folding? *Journal of Medical Physics*, **65**, 44–45.

[50] Fersht, A. (1995) Characterizing transition states in protein folding: an essential step in the puzzle. *Curr Opin Struct Biol*, **5**, 79–84.

[51] Villali, J. and Kern, D. (2010) Choreographing an enzyme's dance. *Curr opin Chem Biol*, **14**, 636–643.

[52] Englander, S. and Mayne, L. (1992) Protein folding studied using hydrogen-exchange labeling and two-dimensional nmr. *Annu Rev of Biophys Biomol Struct*, **21**, 243–265.

[53] Baldwin, R. (1993) Pulsed h/d-exchange studies of folding intermediates. *Curr Opin Struct Biol*, **3**, 84–91.

[54] Dobson, C., Evans, P., and Radford, S. (1994) Understanding how proteins fold: The lysozyme story so far. *Trends Biochem Sci*, **19**, 31–37.

[55] Sosnick, T. R. and Barrick, D. (2010) The folding of single domain proteins - have we reached a consensus? *Curr Opin Struct Biol*, **21**, 12–24.

[56] Geierhaas, C. D., Paci, E., Vendruscolo, M., and Clarke, J. (2004) Comparison of the transition states for folding of two ig-like proteins from different superfamilies. *J Mol Biol*, **343**, 1111–23.

[57] Gardino, A. K., et al. (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell*, **139**, 1109–18.

[58] Dyson, H. and Wright, P. (2006) Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. *J Mol Biol*, **355**, 139–156.

[59] Chiti, F. and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol*, **5**, 15–22.

[60] Chiti, F. and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, **75**, 333–66.

[61] Schwaiger, I., Schleicher, M., Noegel, A. A., and Rief, M. (2005) The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep*, **6**, 46–51.

[62] Grzesiek, S. and Sass, H.-J. (2009) From biomolecular structure to functional understanding: new nmr developments narrow the gap. *Curr Opin Struct Biol*, **19**, 585–95.

[63] Baldwin, A. J. and Kay, L. E. (2009) Nmr spectroscopy brings invisible protein states into focus. *Nat Chem Biol*, **5**, 808–14.

[64] Bouvignies, G., Vallurupalli, P., Hansen, D. F., Correia, B. E., Lange, O., Bah, A., Vernon, R. M., Dahlquist, F. W., Baker, D., and Kay, L. E. (2011) Solution structure of a minor and transiently formed state of a t4 lysozyme mutant. *Nature*, **477**, 111–4.

[65] Zhang, G., Hubalewska, M., and Ignatova, Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol*, **16**, 274–80.

[66] Zhang, G. and Ignatova, Z. (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PloS one*, **4**, e5036.

[67] Khushoo, A., Yang, Z., Johnson, A. E., and Skach, W. R. (2011) Ligand-driven vectorial folding of ribosome-bound human cftr nbd1. *Mol Cell*, **41**, 682–92.

[68] Voss, N., Gerstein, M., Steitz, T., and Moore, P. (2006) The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol*, **360**, 893–906.

[69] Lu, J., Kobertz, W. R., and Deutsch, C. (2007) Mapping the electrostatic potential within the ribosomal exit tunnel. *J Mol Biol*, **371**, 1378–91.

[70] Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–30.

[71] Makeyev, E. V., Kolb, V. A., and Spirin, A. S. (1996) Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Lett*, **378**, 166–170.

[72] Kudlicki, W., Chirgwin, J., Kramer, G., and Hardesty, B. (1995) Folding of an enzyme into an active conformation while bound as peptidyl-trna to the ribosome. *Biochemistry*, **34**, 14284–14287.

[73] Nicola, A. V., Chen, W., and Helenius, A. (1999) Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol*, **1**, 341–5.

[74] Land, A., Zonneveld, D., and Braakman, I. (2003) Folding of hiv-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J*, **17**, 1058–67.

[75] Chen, W., Helenius, J., Braakman, I., and Helenius, A. (1995) Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc Natl Acad Sci USA*, **92**, 6229–33.

[76] Tsalkova, T., Odom, O. W., Kramer, G., and Hardesty, B. (1998) Different conformations of nascent peptides on ribosomes. *J Mol Biol*, **278**, 713–23.

[77] Clark, P. L. and King, J. (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem*, **276**, 25411–20.

[78] Woolhead, C. A., McCormick, P. J., and Johnson, A. E. (2004) Nascent membrane and secretory proteins differ in fret-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, **116**, 725–36.

[79] Kolb, V. A., Makeyev, E. V., and Spirin, A. S. (1994) Folding of firefly luciferase during translation in a cell-free system. *EMBO J*, **13**, 3631–7.

[80] Kleizen, B., van Vlijmen, T., de Jonge, H. R., and Braakman, I. (2005) Folding of cftr is predominantly cotranslational. *Mol Cell*, **20**, 277–287.

[81] Frydman, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1999) Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*, **6**, 697–705.

[82] Ellis, J. P., Culviner, P. H., and Cavagnero, S. (2009) Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Sci*, **18**, 2003–15.

[83] Ellis, J. P., Bakke, C. K., Kirchdoerfer, R. N., Jungbauer, L. M., and Cavagnero, S. (2008) Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem Biol*, **3**, 555–66.

[84] Lee, W., Zeng, X., Zhou, H.-X., Bennett, V., Yang, W., and Marszalek, P. E. (2010) Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations. *J Biol Chem*, **285**, 38167–72.

[85] Ellis, R. J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci*, **26**, 597–604.

[86] van den Berg, B., Ellis, R. J., and Dobson, C. M. (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J*, **18**, 6927–33.

[87] Mittal, J. and Best, R. B. (2010) Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J*, **98**, 315–320.

[88] Elcock, A. H. (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struct Biol*, **20**, 196–206.

[89] Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., and Baumeister, W. (2009) The native 3d organization of bacterial polysomes. *Cell*, **136**, 261–71.

[90] Maier, R., Eckert, B., Scholz, C., Lilie, H., and Schmid, F.-X. (2003) Interaction of trigger factor with the ribosome. *J Mol Biol*, **326**, 585–92.

[91] Liveris, D., Klotsky, R. A., and Schwartz, I. (1991) Growth rate regulation of translation initiation factor if3 biosynthesis in escherichia coli. *J Bacteriol*, **173**, 3888–93.

[92] Patzelt, H., Kramer, G., Rauch, T., Schönfeld, H.-J., Bukau, B., and Deuerling, E. (2002) Three-state equilibrium of escherichia coli trigger factor. *Biol Chem*, **383**, 1611–9.

[93] Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, **431**, 590–6.

[94] Rutkowska, A., Mayer, M. P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E., and Bukau, B. (2008) Dynamics of trigger factor interaction with translating ribosomes. *J Biol Chem*, **283**, 4124–32.

[95] Hartl, F. U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, **295**, 1852–8.

[96] Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., and Deuerling..., E. (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, **431**, 590–596.

[97] Merz, F., et al. (2008) Molecular mechanism and structure of trigger factor bound to the translating ribosome. *EMBO J*, **27**, 1622–1632.

[98] Kosolapov, A. and Deutsch, C. (2009) Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol*, **16**, 405–11.

[99] Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *J Am Chem Soc*, **104**, 4546–4559.

[100] Weinreis, S. A., Ellis, J. P., and Cavagnero, S. (2010) Dynamic fluorescence depolarization: A powerful tool to explore protein folding on the ribosome. *Methods*, **52**, 57–73.

[101] rong Huang, J., Craggs, T. D., Christodoulou, J., and Jackson, S. E. (2007) Stable intermediate states and high energy barriers in the unfolding of gfp. *J Mol Biol*, **370**, 356–71.

[102] Shan, B., McClendon, S., Rospigliosi, C., Eliezer, D., and Raleigh, D. P. (2010) The cold denatured state of the c-terminal domain of protein l9 is compact and contains both native and non-native structure. *J Am Chem Soc*, **132**, 4669–77.

[103] Higman, V. A., Rösner, H. I., Ugolini, R., Greene, L. H., Redfield, C., and Smith, L. J. (2009) Probing the urea dependence of residual structure in denatured human alpha-lactalbumin. *J Biomol NMR*, **45**, 121–31.

[104] Corazza, A., et al. (2010) Native-unlike long-lived intermediates along the folding pathway of the amyloidogenic protein beta2-microglobulin revealed by real-time two-dimensional nmr. *J Biol Chem*, **285**, 5827–35.

[105] Vallurupalli, P., Hansen, D. F., and Kay, L. E. (2008) Structures of invisible, excited protein states by relaxation dispersion nmr spectroscopy. *Proc Natl Acad Sci USA*, **105**, 11766–71.

[106] Schanda, P., Brutscher, B., Konrat, R., and Tollinger, M. (2008) Folding of the kix domain: characterization of the equilibrium analog of a folding intermediate using 15n/13c relaxation dispersion and fast 1h/2h amide exchange nmr spectroscopy. *J Mol Biol*, **380**, 726–41.

[107] O'Connell, N. E., Grey, M. J., Tang, Y., Kosuri, P., Miloushev, V. Z., Raleigh, D. P., and Palmer, A. G. (2009) Partially folded equilibrium intermediate of the villin headpiece hp67 defined by 13c relaxation dispersion. *J Biomol NMR*, **45**, 85–98.

[108] Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2009) Measurement of methyl group motional parameters of invisible, excited protein states by nmr spectroscopy. *J Am Chem Soc*, **131**, 12745–54.

[109] Hansen, D. F., Neudecker, P., Vallurupalli, P., Mulder, F. A. A., and Kay, L. E. (2010) Determination of leu side-chain conformations in excited protein states by nmr relaxation dispersion. *J Am Chem Soc*, **132**, 42–3.

[110] Neudecker, P., Lundström, P., and Kay, L. E. (2009) Relaxation dispersion nmr spectroscopy as a tool for detailed studies of protein folding. *Biophys J*, **96**, 2045–54.

[111] rong Huang, J. and Grzesiek, S. (2010) Ensemble calculations of unstructured proteins constrained by rdc and pre data: a case study of urea-denatured ubiquitin. *J Am Chem Soc*, **132**, 694–705.

[112] Esteban-Martín, S., Fenwick, R. B., and Salvatella, X. (2010) Refinement of ensembles describing unstructured proteins using nmr residual dipolar couplings. *J Am Chem Soc*, **132**, 4626–32.

[113] Bertoncini, C. W., Rasia, R. M., Lamberto, G. R., Binolfi, A., Zweckstetter, M., Griesinger, C., and Fernandez, C. O. (2007) Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. *J Mol Biol*, **372**, 708–22.

[114] Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005) Mapping long-range interactions in alpha-synuclein using spin-label nmr and ensemble molecular dynamics simulations. *J Am Chem Soc*, **127**, 476–7.

[115] Mittag, T. and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol*, **17**, 3–14.

[116] Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, **5**, 789–96.

[117] Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007) Protein structure determination from nmr chemical shifts. *Proc Natl Acad Sci USA*, **104**, 9615–20.

[118] Shen, Y., Bryan, P. N., He, Y., Orban, J., Baker, D., and Bax, A. (2010) De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci*, **19**, 349–56.

[119] Wishart, D. S., Arndt, D., Berjanskii, M., Tang, P., Zhou, J., and Lin, G. (2008) Cs23d: a web server for rapid protein structure generation using nmr chemical shifts and sequence data. *Nucleic Acids Res*, **36**, W496–W502.

[120] Jensen, M. R., Salmon, L., Nodet, G., and Blackledge, M. (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc*, **132**, 1270–2.

[121] Ruschak, A. M. and Kay, L. E. (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR*, **46**, 75–87.

[122] Sprangers, R. and Kay, L. E. (2007) Quantitative dynamics and binding studies of the 20s proteasome by nmr. *Nature*, **445**, 618–22.

[123] Religa, T. L., Sprangers, R., and Kay, L. E. (2010) Dynamic regulation of archaeal proteasome gate opening as studied by trosy nmr. *Science*, **328**, 98–102.

[124] Ruschak, A. M., Religa, T. L., Breuer, S., Witt, S., and Kay, L. E. (2010) The proteasome antechamber maintains substrates in an unfolded state. *Nature*, **467**, 868–71.

[125] Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L., and Wüthrich, K. (2005) Direct nmr observation of a substrate protein bound to the chaperonin groel. *Proc Natl Acad Sci USA*, **102**, 12748–53.

[126] Baldwin, A. J., Anthony-Cahill, S. J., Knowles, T. P. J., Lippens, G., Christodoulou, J., Barker, P. D., and Dobson, C. M. (2008) Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution nmr diffusion measurements. *Angew Chem Int Ed Engl*, **47**, 3385–7.

[127] Sillen, A., Leroy, A., Wieruszeski, J.-M., Loyens, A., Beauvillain, J.-C., Buée, L., Landrieu, I., and Lippens, G. (2005) Regions of tau implicated in the paired helical fragment core as defined by nmr. *Chembiochem*, **6**, 1849–56.

[128] Landrieu, I., Leroy, A., Smet-Nocca, C., Huvent, I., Amniai, L., Hamdane, M., Sibille, N., Buée, L., Wieruszeski, J.-M., and Lippens, G. (2010) Nmr spectroscopy of the neuronal tau protein: normal function and implication in alzheimer's disease. *Biochem Soc Trans*, **38**, 1006–11.

[129] Lavalette, D., Amand, B., and Pochon, F. (1977) Rotational relaxation of 70s ribosomes by a depolarization method using triplet probes. *Proc Natl Acad Sci USA*, **74**, 1407–11.

[130] Hsu, S.-T. D., Fucini, P., Cabrita, L. D., Launay, H., Dobson, C. M., and Christodoulou, J. (2007) Structure and dynamics of a ribosome-bound nascent chain by nmr spectroscopy. *Proc Natl Acad Sci USA*, **104**, 16516–21.

[131] Hsu, S.-T. D., Cabrita, L. D., Fucini, P., Christodoulou, J., and Dobson, C. M. (2009) Probing side-chain dynamics of a ribosome-bound nascent chain using methyl nmr spectroscopy. *J Am Chem Soc*, **131**, 8366–7.

[132] Eichmann, C., Preissler, S., Riek, R., and Deuerling, E. (2010) Cotranslational structure acquisition of nascent polypeptides monitored by nmr spectroscopy. *Proc Natl Acad Sci USA*, **107**, 9111–6.

[133] Grzesiek, S. (2008) Embo notes. *Embo course*.

[134] Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., and Skelton, N. J. (2007) Protein nmr spectroscopy: Principles and practice 2nd edition. *Academic press*.

[135] Peng, J. W. and Wagner, G. (1994) Investigation of protein motions via relaxation measurements. *Method Enzymol*, **239**, 563–596.

[136] Skrisovska, L., Schubert, M., and Allain, F. H.-T. (2010) Recent advances in segmental isotope labeling of proteins: Nmr applications to large proteins and glycoproteins. *J Biomol NMR*, **46**, 51–65.

[137] Fernández, C. and Wider, G. (2003) Trosy in nmr studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol*, **13**, 570–80.

[138] Yang, D. and Kay, L. E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from nmr-derived order parameters: application to protein folding. *J Mol Biol*, **263**, 369–82.

[139] Kneller, J. M., Lu, M., and Bracken, C. (2002) An effective method for the discrimination of motional anisotropy and chemical exchange. *J Am Chem Soc*, **124**, 1852–1853.

[140] Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2008) Using relaxation dispersion nmr spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR*, **41**, 113–20.

[141] Baldwin, A. J., Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2009) Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion nmr spectroscopy. *J Am Chem Soc*, **131**, 11939–48.

[142] Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) Attenuated t2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to nmr structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA*, **94**, 12366–71.

[143] Nietlispach, D. (2005) Suppression of anti-trosy lines in a sensitivity enhanced gradient selection trosy scheme. *J Biomol NMR*, **31**, 161–166.

[144] Riek, R., Wider, G., Pervushin, K., and Wüthrich, K. (1999) Polarization transfer by cross-correlated relaxation in solution nmr with very large molecules. *Proc Natl Acad Sci USA*, **96**, 4918–4923.

[145] Ollerenshaw, J. E., Tugarinov, V., and Kay, L. E. (2003) Methyl trosy: explanation and experimental verification. *Magn Res Chem*, **41**, 843–852.

[146] Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E., and Kay, L. E. (2003) Cross-correlated relaxation enhanced 1h[bond]13c nmr spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc*, **125**, 10420–8.

[147] Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002) Nmr analysis of a 900k groel–groes complex. *Nature*, **418**, 207–211.

[148] Greisinger, C. (2009) Embo course notes. *Embo course*.

[149] Goto, N. K., Gardner, K. H., Mueller, G., Willis, R. C., and Kay, L. E. (1999) A robust and cost-effective method for the production of val, leu, ile (Ît'1) methyl-protonated 15n-, 13c-, 2h-labeled proteins. *J Biomol NMR*, **13**, 369–374.

[150] Goto, N. K. and Kay, L. E. (2000) New developments in isotope labeling strategies for protein solution nmr spectroscopy. *Curr Opin Struct Biol*, **10**, 585–92.

[151] Schanda, P. and Brutscher, B. (2005) Very fast two-dimensional nmr spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc*, **127**, 8014–8015.

[152] Kato, H., van Ingen, H., Zhou, B.-R., Feng, H., Bustin, M., Kay, L. E., and Bai, Y. (2011) From the cover: Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based nmr. *Proc Natl Acad Sci USA*, **108**, 12283–8.

[153] Gelis, I., Bonvin, A. M. J. J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C. G. (2007) Structural basis for signal-sequence recognition by the translocase motor seca as determined by nmr. *Cell*, **131**, 756–69.

[154] Clore, G. M., Szabo, A., Bax, A., Kay, L., Driscoll, P., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc*, **112**, 4989–4991.

[155] Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. (1992) Analysis of the backbone dynamics of the ribonuclease h domain of the human immunodeficiency virus reverse transcriptase using 15n relaxation measurements. *Biochemistry*, **31**, 9150–7.

[156] Baber, J. L., Szabo, A., and Tjandra, N. (2001) Analysis of slow interdomain motion of macromolecules using nmr relaxation data. *J Am Chem Soc*, **123**, 3953–9.

[157] Kirkpatrick, J. (2010) Personal communication.

[158] Cabrita, L. D., Hsu, S.-T. D., Launay, H., Dobson, C. M., and Christodoulou, J. (2009) Probing ribosome-nascent chain complexes produced in vivo by nmr spectroscopy. *Proc Natl Acad Sci USA*, **106**, 22239–44.

[159] Schanda, P., Kupce, E., and Brutscher, B. (2005) Sofast-hmqc experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR*, **33**, 199–211.

[160] Rutkowska, A., Beerbaum, M., Rajagopalan, N., Fiaux, J., Schmieder, P., Kramer, G., Oschkinat, H., and Bukau, B. (2009) Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett*, **583**, 2407–13.

[161] Brutscher, B. (2009) Embo course notes. *Embo course*.

[162] Bax, A., Ikura, M., Kay, L. E., and Torchia, D. A. (1990) Comparison of different modes of two-dimensional reverse-correlation nmr for the study of proteins. *J Magn Res*, **86**, 304–318.

[163] Lescop, E., Kern, T., and Brutscher, B. (2009) Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear nmr: Example of longitudinal-relaxation-enhanced best-type (1)h-(15)n correlation experiments. *J Magn Res*.

[164] Stejskal, E. and Tanner, J. (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*, **42**, 288–293.

[165] Price, K. E., Lucas, L. H., and Larive, C. K. (2004) Analytical applications of nmr diffusion measurements. *Anal Bioanal Chem*, **378**, 1405–7.

[166] Ferrage, F., Zoonens, M., Warschawski, D. E., Popot, J.-L., and Bodenhausen, G. (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient nmr method. *J Am Chem Soc*, **125**, 2541–5.

[167] Didenko, T., Boelens, R., and Rudiger, S. G. D. (2011) 3d dosy–trosy to determine the translational diffusion coefficient of large protein complexes. *Protein Eng Design Selection*, **24**, 9–103.

[168] Seeton, C. (2006) Viscosity–temperature correlation for liquids. *Tribology Letters*, **22**, 67–78.

[169] Ortega, A. and de la Torre, J. G. (2005) Efficient, accurate calculation of rotational diffusion and nmr relaxation of globular proteins from atomic-level structures and approximate hydrodynamic calculations. *J Am Chem Soc*, **127**, 12764–5.

[170] Blanchard, L., Hunter, C. N., and Williamson, M. P. (1997) The effect of ring currents on carbon chemical shifts in cytochromes. *J Biomol NMR*, **9**, 389–95.

[171] Dyson, H. J. and Wright, P. E. (2004) Unfolded proteins and protein folding studied by nmr. *Chem Rev*, **104**, 3607–22.

[172] Gabler, R., Westhead, E. W., and Ford, N. C. (1974) Studies of ribosomal diffusion coefficients using laser light-scattering spectroscopy. *Biophys J*, **14**, 528–45.

[173] Croke, R. L., Sallum, C. O., Watson, E., Watt, E. D., and Alexandrescu, A. T. (2008) Hydrogen exchange of monomeric alpha-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in escherichia coli. *Protein Sci*, **17**, 1434–45.

[174] Takahashi, S., Iida, M., Furusawa, H., Shimizu, Y., Ueda, T., and Okahata, Y. (2009) Real-time monitoring of cell-free translation on a quartz-crystal microbalance. *J Am Chem Soc*, **131**, 9326–32.

[175] Matsuura, T., Yanagida, H., Ushioda, J., Urabe, I., and Yomo, T. (2007) Nascent chain, mrna, and ribosome complexes generated by a pure translation system. *Biochem Biophys Res Commun*, **352**, 372–7.

[176] Lakshmipathy, S. K., Tomic, S., Kaiser, C. M., Chang, H.-C., Genevaux, P., Georgopoulos, C., Barral, J. M., Johnson, A. E., Hartl, F. U., and Etchells, S. A. (2007) Identification of nascent chain interaction sites on trigger factor. *J Biol Chem*, **282**, 12186–93.

[177] Schaffitzel, C. and Ban, N. (2007) Reprint of "generation of ribosome nascent chain complexes for structural and functional studies" [j. struct. biol. 158 (2007) 463-471]. *J Struct Biol*, **159**, 302–10.

[178] Evans, M. S., Ugrinov, K. G., Frese, M.-A., and Clark, P. L. (2005) Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods*, **2**, 757–62.

[179] Gilbert, R. J. C., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., Robinson, C. V., Dobson, C. M., and Stuart, D. I. (2004) Three-dimensional structures of translating ribosomes by cryo-em. *Mol Cell*, **14**, 57–66.

[180] Nakatogawa, H. and Ito, K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell*, **108**, 629–36.

[181] Bax, A., Kontaxis, G., and Tjandra, N. (2001) Dipolar couplings in macromolecular structure determination. *Method Enzymol*, **339**, 127–174.

[182] Kontaxis, G., Clore, G. M., and Bax, A. (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J Magn Res*, **143**, 184–196.

[183] Hsu, S.-T. D., Cabrita, L. D., Fucini, P., Dobson, C. M., and Christodoulou, J. (2009) Structure, dynamics and folding of an immunoglobulin domain of the gelation factor (abp-120) from dictyostelium discoideum. *J Mol Biol*, **388**, 865–79.

[184] Hill, W. E., Rossetti, G. P., and Holde, K. E. V. (1969) Physical studies of ribosomes from escherichia coli. *J Mol Biol*, **44**, 263–77.

[185] Hsu, S.-T. D., Cabrita, L. D., Christodoulou, J., and Dobson, C. M. (2009) 1h, 15n and 13c assignments of domain 5 of dictyostelium discoideum gelation factor (abp-120) in its native and 8m urea-denatured states. *Biomol NMR Assign*, **3**, 29–31.

[186] Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A., and Torchia, D. A. (1992) Dynamics of methyl groups in proteins as studied by proton-detected 13c nmr spectroscopy. application to the leucine residues of staphylococcal nuclease. *Biochemistry*, **31**, 5253–63.

[187] Tugarinov, V. and Kay, L. E. (2005) Methyl groups as probes of structure and dynamics in nmr studies of high-molecular-weight proteins. *Chembiochem*, **6**, 1567–77.

[188] Xiao, S., Bi, Y., Shan, B., and Raleigh, D. P. (2009) Analysis of core packing in a cooperatively folded miniature protein: the ultrafast folding villin headpiece helical subdomain. *Biochemistry*, **48**, 4607–16.

[189] Noegel, A. A., Rapp, S., Lottspeich, F., Schleicher, M., and Stewart, M. (1989) The dictyostelium gelation factor shares a putative actin binding site with alpha-actinins and dystrophin and also has a rod domain containing six 100-residue motifs that appear to have a cross-beta conformation. *J Cell Biol*, **109**, 607–18.

[190] McCoy, A. J., Fucini, P., Noegel, A. A., and Stewart, M. (1999) Structural basis for dimerization of the dictyostelium gelation factor (abp120) rod. *Nat Struct Biol*, **6**, 836–41.

[191] Fucini, P., Renner, C., Herberhold, C., Noegel, A. A., and Holak, T. A. (1997) The repeating segments of the f-actin cross-linking gelation factor (abp-120) have an immunoglobulin-like fold. *Nat Struct Biol*, **4**, 223–30.

[192] Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992) Structure of a fibronectin type iii domain from tenascin phased by mad analysis of the selenomethionyl protein. *Science*, **258**, 987–91.

[193] Improta, S., Politou, A. S., and Pastore, A. (1996) Immunoglobulin-like modules from titin i-band: extensible components of muscle elasticity. *Structure*, **4**, 323–37.

[194] Billings, K., Best, R., Rutherford, T., and Clarke, J. (2008) Crosstalk between the protein surface and hydrophobic core in a core-swapped fibronectin type iii domain. *J Mol Biol*, **375**, 560–571.

[195] Cota, E., Steward, A., Fowler, S., and Clarke, J. (2001) The folding nucleus of a fibronectin type iii domain is composed of core residues of the immunoglobulin-like fold1. *J Mol Biol*, **305**, 1185–1194.

[196] Fowler, S. and Clarke, J. (2001) Mapping the folding pathway of an immunoglobulin domain:: Structural detail from phi value analysis and movement of the transition state. *Structure*.

[197] Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) Hydrogen exchange: the modern legacy of linderstrøm-lang. *Protein Sci*, **6**, 1101–9.

[198] Morozova, L. A., Haynie, D. T., Arico-Muendel, C., Dael, H. V., and Dobson, C. M. (1995) Structural basis of the stability of a lysozyme molten globule. *Nat Struct Biol*, **2**, 871–5.

[199] Cota, E., Hamill, S. J., Fowler, S. B., and Clarke, J. (2000) Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *J Mol Biol*, **302**, 713–25.

[200] Linderstrøm-Lang, K. U. (1958) Deuterium exchange and protein structure. *In: Neuberger A, ed. Symposium on protein structure*.

[201] Zhang, Y.-Z. (1995) Protein and peptide structure and interactions studied by hydrogen exchange and nmr. *Ph.D. Thesis, Structural Biology and Molecular Biophysics, University of Pennsylvania, PA, USA*.

[202] Best, R. B. and Vendruscolo, M. (2006) Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of ci2. *Structure*, **14**, 97–106.

[203] Fersht, A. (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. *San Francisco: W.H. Freeman*.

[204] Li, P.-C. and Makarov, D. E. (2004) Simulation of the mechanical unfolding of ubiquitin: probing different unfolding reaction coordinates by changing the pulling geometry. *J Chem Phys*, **121**, 4826–32.

[205] Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A., and Rief, M. (2004) A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat Struct Mol Biol*, **11**, 81–5.

[206] Evans, M. S., Sander, I. M., and Clark, P. L. (2008) Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J Mol Biol*, **383**, 683–92.

[207] Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta Biochim Pol*, **51**, 299–321.

[208] Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim Biophys Acta*, **1382**, 191–206.

[209] de Laureto, P. P., Filippis, V. D., Bello, M. D., Zambonin, M., and Fontana, A. (1995) Probing the molten globule state of alpha-lactalbumin by limited proteolysis. *Biochemistry*, **34**, 12596–604.

[210] de Laureto, P. P., Frare, E., Gottardo, R., and Fontana, A. (2002) Molten globule of bovine alpha-lactalbumin at neutral ph induced by heat, trifluoroethanol, and oleic acid: a comparative analysis by circular dichroism spectroscopy and limited proteolysis. *Proteins*, **49**, 385–97.

[211] Tritton, T. R. (1979) Ribosome-chloramphenicol interactions: a nuclear magnetic resonance study. *Arch Biochem Biophys*, **197**, 10–7.

[212] Tritton, T. R. (1980) Proton nmr observation of the escherichia coli ribosome. *FEBS Lett*, **120**, 141–4.

[213] Bushuev, V. N. and Gudkov, A. T. (1988) [9] nuclear magnetic resonance techniques for studying structure and function of ribosomes. *Method Enzymol*, **164**, 148–58.

[214] Torre, J. G. D. L., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J*, **78**, 719–30.

[215] Guérin, M. F. and Hayes, D. H. (1987) Comparison of active and inactive forms of the e. coli 30s ribosomal subunits. *Biochimie*, **69**, 965–74.

[216] Ulrich, E. L., et al. (2008) Biomagresbank. *Nucleic Acids Res*, **36**, D402–8.

[217] Hansen, D. F., Neudecker, P., and Kay, L. E. (2010) Determination of isoleucine side-chain conformations in ground and excited states of proteins from chemical shifts. *J Am Chem Soc*, **132**, 7589–91.

[218] Benelli, D. and Londei, P. (2007) In vitro studies of archaeal translational initiation. *Method Enzymol*, **430**, 79–109.

[219] Blombach, F., Launay, H., Zorraquino, V., Swarts, D. C., Cabrita, L. D., Benelli, D., Christodoulou, J., Londei, P., and van der Oost, J. (2011) An hflx-type gtpase from sulfolobus solfataricus binds to the 50s ribosomal subunit in all nucleotide-bound states. *J Bacteriol*, **193**, 2861–7.

[220] Londei, P., Teichner, A., Cammarano, P., Rosa, M. D., and Gambacorta, A. (1983) Particle weights and protein composition of the ribosomal subunits of the extremely thermoacidophilic archaebacterium caldariella acidophila. *Biochem J*, **209**, 461–70.

[221] Ilag, L. L., Videler, H., McKay, A. R., Sobott, F., Fucini, P., Nierhaus, K. H., and Robinson, C. V. (2005) Heptameric (l12)6/l10 rather than canonical pentameric complexes are found by tandem ms of intact ribosomes from thermophilic bacteria. *Proc Natl Acad Sci USA*, **102**, 8192–7.

[222] Bonincontro, A., Cametti, C., Nierhaus, K. H., Ortore, M. G., and Risuleo, G. (2005) Ribosomes deprived of select proteins show similar structural alterations induced by thermal treatment of native particles. *Cell Biochem Biophys*, **42**, 55–60.

[223] Thakur, C. S., Sama, J. N., Jackson, M. E., Chen, B., and Dayie, T. K. (2010) Selective 13c labeling of nucleotides for large rna nmr spectroscopy using an e. coli strain disabled in the tca cycle. *J Biomol NMR*, **48**, 179–92.

[224] Simon, B., Zanier, K., and Sattler, M. (2001) A trosy relayed hcch-cosy experiment for correlating adenine h2/h8 resonances in uniformly 13c-labeled rna molecules. *J Biomol NMR*, **20**, 173–6.

[225] Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011) Assembly of bacterial ribosomes. *Annu Rev Biochem*, **80**, 501–26.

[226] Sykes, M. T. and Williamson, J. R. (2009) A complex assembly landscape for the 30s ribosomal subunit. *Annu Rev Biophys*, **38**, 197–215.

[227] Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat Biotechnol*, **19**, 751–5.

[228] Nierhaus, K. H. and Dohme, F. (1974) Total reconstitution of functionally active 50s ribosomal subunits from escherichia coli. *Proc Natl Acad Sci USA*, **71**, 4713–7.

[229] Suzuki, M., Mao, L., and Inouye, M. (2007) Single protein production (spp) system in escherichia coli. *Nat Protoc*, **2**, 1802.

[230] Tugarinov, V., Sprangers, R., and Kay, L. E. (2007) Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer nmr spectroscopy. *J Am Chem Soc*, **129**, 1743–50.

[231] Waudby, C., Cabrita, L. D., Christodoulou, J., and et al (2011) Personal communication.

[232] Jones, D. H., Cellitti, S. E., Hao, X., Zhang, Q., Jahnz, M., Summerer, D., Schultz, P. G., Uno, T., and Geierstanger, B. H. (2010) Site-specific labeling of proteins with nmr-active unnatural amino acids. *J Biomol NMR*, **46**, 89–100.

[233] Lescop, E., Schanda, P., and Brutscher, B. (2007) A set of best triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Res*, **187**, 163–9.

[234] Kupce, E. and Freeman, R. (2003) Frequency-domain hadamard spectroscopy. *J Magn Res*, **162**, 158–65.

[235] Bertini, I., Luchinat, C., Parigi, G., Ravera, E., Reif, B., and Turano, P. (2011) Solid-state nmr of proteins sedimented by ultracentrifugation. *Proc Natl Acad Sci USA*, **108**, 10396–9.

[236] Wang, X., Kirkpatrick, J., Christodoulou, J., and et al (2011) Personal communication.

[237] Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207–34.

[238] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) Nmrpipe: a multidimensional spectral processing system based on unix pipes. *J Biomol NMR*, **6**, 277–93.