

***In vitro* and in-cell NMR approaches
to the study of
ribosome-nascent chain complexes**

Sammy HS Chan

A dissertation submitted for the degree of
Doctor of Philosophy in Structural and Molecular Biology

**Research Department of Structural and Molecular Biology
University College London**

2017

Declaration

I, **Sammy HS Chan**, 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 Research Department of Structural and Molecular Biology, University College London, under the supervision of Prof John Christodoulou.

Abstract

The ubiquitous ribosome is responsible for the synthesis of proteins, which generally require correct three-dimensional folds for their biologically active function. As the nascent polypeptide progressively emerges from the ribosomal exit tunnel, it is able to acquire structure concurrently with its biosynthesis while it is exposed to the densely packed and crowded cellular environment. NMR spectroscopy is uniquely able to probe such dynamic structures and folding processes at near-atomic resolution, and its application to translationally stalled ribosome-nascent chain complexes (RNCs) has allowed us to begin to understand co-translational protein folding in the cell. However, these studies are frequently restricted by weak signals arising from high molecular weight of ~ 2.4 MDa, the resulting low maximum achievable concentrations, and limited sample lifetimes of typically just a few hours, which present a major barrier to progress in this field.

To overcome the substantial spectroscopic challenges of NMR studies of RNCs, we have employed the use of the paramagnetic longitudinal relaxation enhancement (PLRE) agent NiDO2A in combination with longitudinal relaxation-optimised experiments, and report improvements of up to 4.8-fold in the sensitivities of two-dimensional ^1H - ^{15}N NMR correlation spectra and NMR diffusion measurements that are crucial in assessing the integrity of unstable RNCs. We describe how the PLRE approach can be exploited in more complex NMR experiments to extract direct structural information on RNCs of an immunoglobulin-like domain using three-dimensional NMR and residual dipolar coupling measurements, which have previously been unattainable due to low spectroscopic sensitivity.

We further describe a strategy to translate current *in vitro* RNC studies into NMR investigations within living cells, and address how ribosome-bound species can be distinguished from those isolated from the ribosome *in vivo*. Experiments in reconstituted cytosol revealed that macromolecular crowding effects perturb the structure and dynamics of FLN5. We anticipate our studies will significantly advance the application of NMR to large macromolecular assemblies, and to studies of protein folding under increasingly more physiologically relevant conditions.

Publications

Work carried out during my graduate studies, including results shown in this thesis, have been presented in the following publications:

Chan, S. H. S., Waudby, C. A., Cassaignau, A. M. E., Cabrita, L. D. & Christodoulou, J. Increasing the sensitivity of NMR diffusion measurements by paramagnetic longitudinal relaxation enhancement, with application to ribosome-nascent chain complexes. *Journal of Biomolecular NMR* 63, 151–163 (2015).

Waudby, C.A., Karyadi, M.E., Wlodarski, T., Chan, S.H.S., Schmidt-Engler, J., Cassaignau, A.M.E., Wentink, A., Carmilloni, C., Cabrita, L.D., Vendruscolo, M. & Christodoulou, J. The experimental free energy landscape associated with the co-translational folding of a filamin domain. *Submitted*

Acknowledgements

This dissertation is the culmination of the four years of my PhD studies, which have been immensely rewarding, and of countless times when I have needed help, support, guidance and reassurance from others. I thank everyone who has been a part of it.

I will always be indebted to my supervisor Prof John Christodoulou, for the opportunity of working in his group and for the confidence he had in me from the very beginning. His guidance and advice in science (and in life) have been invaluable.

I am grateful to everyone in the research group who has helped me in innumerable ways during the past few years. In particular, I wish to thank Dr Lisa Cabrita for her help throughout my PhD and guiding me through all the molecular biology and biochemistry that have been crucial to this work. I am hugely thankful to Dr Chris Waudby, whom I owe all my knowledge of NMR spectroscopy to, and for his always insightful, thoughtful comments and discussions on my work. Thank you to Anaïs Cassaignau, whom I worked closely with in this work, for her helpful thoughts and equally helpful distractions, and to Dr Xiaolin Wang for her insights and technical help with the RDC and other NMR experiments. I would of course like to thank my remaining fellow colleagues, past and present, all of whom have made the experience all the more enjoyable. I am grateful to Dr Angelo Mendes Pinto De Figu and Dr John Kirkpatrick for their valuable technical assistance in NMR, and staff at the NMR facilities of the Francis Crick Institute in London and the University of Birmingham.

My PhD studies have been funded by the Biotechnology and Biological Sciences Research Council (BBSRC) through the London Interdisciplinary Doctoral Programme.

Last but definitely not least, my family and friends. Thank you, Michelle Chan, Olivia Chan and Harry Ng, you have always been there for me. And to my parents, thank you for giving me the opportunities you never had, I hope I've made you proud.

Contents

Chapter 1: Introduction	21
1.1 Protein folding and structure	21
1.1.1 Mechanisms of protein folding from in vitro studies	21
1.2 Protein folding in the cell	23
1.2.1 Protein biosynthesis by the ribosome	24
1.2.2 Co-translational protein folding	25
1.2.3 Interactions and protein folding within the exit tunnel	26
1.2.4 Protein folding outside the exit tunnel	28
1.2.5 Translation rates and protein folding	30
1.2.6 The cellular environment	31
1.2.7 Molecular chaperones and other auxiliary factors	32
1.2.8 Summary	34
1.3 Applications of biomolecular NMR spectroscopy	36
1.3.1 NMR spectroscopy of large biological systems and in living cells	37
1.3.2 NMR spectroscopy of ribosomal particles	37
1.4 Introduction to NMR theory	40
1.4.1 The physical origin of NMR	40
1.4.2 The basic pulse-acquire experiment	40
1.4.3 Multi-dimensional NMR	41
1.4.4 Relaxation and NMR sensitivity	42
1.4.5 Transverse relaxation-optimised spectroscopy (TROSY)	44
1.5 Perspectives	45
Chapter 2: Increasing the sensitivity of NMR diffusion measurements by paramagnetic longitudinal relaxation enhancement	47
2.1 Introduction	48
2.1.1 NMR diffusion measurements	48
2.1.2 Improving the sensitivity of NMR experiments	49
2.2 Results	54
2.2.1 PLRE studies on isolated proteins	54
2.2.2 PLRE studies on 70S ribosomes	59
2.2.3 PLRE studies on ribosome-nascent chain complexes	62
2.3 Discussion	64

Chapter 3: Structural studies of ribosome-nascent chain complexes by NMR spectroscopy	69
3.1 Introduction	69
3.1.1 Co-translational folding of an immunoglobulin domain	69
3.1.2 Structural studies of RNCs by NMR spectroscopy	72
3.2 Theory and applications of 3D NMR	73
3.2.1 Triple-resonance NMR experiments	73
3.2.2 Sensitivity of 3D NMR experiments	74
3.2.3 Using chemical shifts to predict secondary structure	76
3.3 Theory and applications of residual dipolar couplings	77
3.3.1 The physical origin of residual dipolar couplings	77
3.3.2 Alignment methods	78
3.3.3 Measuring residual dipolar couplings	79
3.3.4 Applications of residual dipolar couplings	79
3.4 Results	80
3.4.1 Production of ¹³ C/ ¹⁵ N-labelled RNCs	80
3.4.2 3D NMR of FLN5+31 A ₃ A ₃ RNC	82
3.4.3 PLRE studies on BEST-HNCO and BEST-HNCA experiments of FLN5Δ12	87
3.4.4 3D NMR of FLN5+31 RNC	89
3.4.5 Towards residual dipolar coupling measurements of FLN5+31 RNC	92
3.5 Discussion	95
Chapter 4: <i>In vivo</i> observations of ribosome-nascent chain complexes by in-cell NMR spectroscopy	97
4.1 Introduction	97
4.1.1 <i>In vivo</i> approaches to study protein structure and folding	98
4.2 Applications of in-cell NMR spectroscopy	99
4.3 Technical aspects of in-cell NMR spectroscopy	100
4.3.1 In-cell sample preparation by overexpression	101
4.3.2 In-cell sample preparation by delivering purified proteins into cells	101
4.3.3 Selective isotopic labelling strategies	102
4.3.4 Further technical aspects of in-cell NMR experiments	104
4.3.5 From <i>in vitro</i> to in-cell NMR studies of RNCs	104
4.4 Results	105
4.4.1 Developing a strategy to observe RNCs in living cells by in-cell NMR	105
4.4.2 Structural insights of FLN5Δ12 in living cells, reconstituted cytosol and crowding agents	114

4.5 Discussion	116
4.5.1 FLN5+31 RNC remains disordered in vivo despite crowding effects	116
4.5.2 Approaches to identifying ribosome-bound and released species	117
4.5.3 Future perspectives	117
Chapter 5: Concluding remarks	119
Chapter 6: Materials and methods	123
6.1 Molecular biology and biochemistry	123
6.1.1 Plasmids and bacterial strains	123
6.1.2 Growth media and antibiotic	123
6.1.3 Production of E. coli competent cells	125
6.1.4 Site-directed mutagenesis	126
6.1.5 Transformation	127
6.1.6 DNA purification and quantification	127
6.1.7 Agarose gel electrophoresis	127
6.1.8 Sodium dodecyl sulphate polyacrylamide gel electrophoresis	128
6.1.9 Immunodetection	128
6.1.10 Protein expression and purification	129
6.1.11 Ribosome and ribosome-nascent chain expression and purification	132
6.1.12 Preparation of in-cell NMR samples	136
6.1.13 Preparation of lysed in-cell NMR samples	137
6.1.14 Preparation of reconstituted cytosol	137
6.1.15 Preparation of NiDO2A	138
6.1.16 Preparation of residual dipolar coupling measurement samples	138
6.2 NMR spectroscopy	138
6.2.1 NMR experimental conditions	138
6.2.2 NMR data acquisition, processing and analysis	139
A.1 pLDC vector	145
A.2 DNA and protein sequences	145
A.2.1 α -Synuclein	145
A.2.2 FLN5	146
A.2.3 FLN5 Δ 12	146
A.2.4 FLN5 A ₃ A ₃	146
A.2.5 FLN5+31 RNC	147
A.2.6 FLN5+31 A ₃ A ₃ RNC	147
A.2.7 FLN5+31 NCA1	148

A.2.8 Primers for SecM Δ 1 mutation	148
A.3 FLN5 crystal structure	148
A. 4 NMR processing scripts	148
A.4.1 Script for conversion and processing of 1D ^1H NMR experiments	148
A.4.2 Script for conversion and processing of pseudo-2D NMR experiments	149
A.4.3 Script for conversion, addition and processing of several pseudo-2D ^1H - ^{15}N SORDID diffusion experiments	149
A.4.4 Script for conversion and processing of 2D experiments	150
A.4.5 Script for conversion, addition and processing of several 2D experiments to a single data set	151
A.4.6 Script for conversion and processing of 3D NMR experiments	152
A.4.7 Script for conversion, addition and processing of several 3D NMR experiments into a single data set	153
A.5 ^1H - ^{15}N NMR assignments used in this study	154
A.5.1 Isolated bL12	154
A.5.2 α -Synuclein	155
A.5.3 FLN5	156
A.5.4 FLN5 Δ 12	157
A.5.5 FLN5 A ₃ A ₃	158
References	159

List of figures

1.1	Energy landscape model for protein folding	23
1.2	The structure of the prokaryotic ribosome	24
1.3	Protein folding on the ribosome	27
1.4	Progression of strategies used to study <i>de novo</i> protein folding	35
1.5	Measuring protein dynamics by NMR spectroscopy	36
1.6	The NMR experiment	41
1.7	The dependence of relaxation times on the rotational correlation time	43
1.8	Transverse relaxation-optimised spectroscopy	45
2.1	Pulse sequences for ¹⁵N heteronuclear diffusion experiments	51
2.2	The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of NMR measurements of isolated proteins	55
2.3	The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of NMR measurements of <i>E. coli</i> 70S ribosomes	61
2.4	The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of SOFAST-HMQC experiments of FLN5+31 RNC	63
2.5	Comparison of NMR diffusion measurements of a FLN5 RNC, and monitoring its integrity using both biochemical and NMR analysis	64
3.1	2D NMR spectroscopy of FLN5+31 WT and A₃A₃ RNC mutant	71
3.2	Residual dipolar couplings	78
3.3	Monitoring of FLN5+31 A₃A₃ and WT RNC sample stability during NMR data acquisition	81
3.4	3D NMR spectroscopy of FLN5+31 A₃A₃ RNC using BEST-HNCO and BEST-HNCA experiments	83
3.5	Analysis of HNCO and HNCA experiments of FLN5+31 A₃A₃ RNC.	85
3.6	Comparison of secondary structure populations between isolated and ribosome-bound FLN5+31 A₃A₃, calculated with the δ2D method	87
3.7	The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of HNCO and HNCA experiments of FLN5Δ12.	88
3.8	BEST-HNCO experiments of FLN5+31 RNC using PLRE	90
3.9	Analysis of the HNCO experiment of FLN5+31 RNC	92
3.10	Preliminary investigations of residual dipolar coupling measurements of FLN+31 RNC using PLRE	94

4.1	General scheme of in-cell NMR experiments	106
4.2	In-cell NMR of isolated FLN5 constructs	107
4.3	NMR characterisation of FLN5+31 NCA1	109
4.4	In-cell NMR of Ile-δ1-$^{13}\text{CH}_3$-labelled FLN5+31 NCA1	110
4.5	Deuteration of uniformly ^{15}N-labelled samples expressed in <i>E. coli</i> at high densities	112
4.6	Identification of in-cell FLN5+31 RNC NMR resonances	113
4.7	Residue-specific analysis of NMR signal intensities of FLN5Δ12 in cells and crowded environments	115

List of tables

2.1 Longitudinal relaxation rates of FLN5 and α -synuclein as a function of NiDO2A concentration	57
2.2 Measured ^1H and ^{15}N transverse relaxation rates of α -synuclein, FLN5, and <i>E. coli</i> 70S ribosomes as a function of NiDO2A concentration.	57
2.3 Longitudinal relaxation rates of <i>E. coli</i> 70S ribosomes as a function of NiDO2A concentration	62
3.1 Summary of selected triple-resonance experiments to obtain chemical information on C' , C^α and C^β	75
3.2 Measured ^1H and ^{15}N longitudinal and transverse relaxation rates of FLN5 Δ 12 as a function of NiDO2A concentration	89
5.1 Composition of M9 media	124
5.2 Composition of MDG growth media	125
5.3 Composition of EM9 expression media	125
5.4 Composition of transformation buffers 1 and 2 used for the production of competent cells	126
5.5 Typical composition of PCR mixture used	126
5.6 Typical PCR temperature cycle used	127
5.7 Composition of an SDS-PAGE gel	128
5.8 Composition of SDS-PAGE loading dye	128
5.9 Buffer compositions for western blot immunodetection	129
5.10 Buffer compositions for the purification of isolated FLN5 constructs	130
5.11 Buffer compositions for the purification of isolated α Syn constructs	131
5.12 Buffer compositions for the purification of ribosomes	132
5.13 Buffer compositions for the purification of RNCs	134

Abbreviations

1D, 2D, 3D	One-, two-, three-dimensional
BEST	Band-selective short transient
BME	β -Mercaptoethanol
BSA	Bovine serum albumin
CFTR	Cystic fibrosis transmembrane conductance regulator
Cryo-EM	Cryo-electron microscopy
CSP	Chemical shift perturbation
DEAE	Diethyl-aminoethyl
DNA	Deoxyribonucleic acid
DSS	4,4-Dimethyl-4-silapentane-1-sulfonic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediamine-tetraacetic acid
EM9	Enhanced minimal media
FID	Free induction decay
FLN5, FLN6	5 th , 6 th immunoglobulin-like domain of the filamin gelation factor
FRET	Förster resonance energy transfer
HMQC	Heteronuclear multiple-quantum coherence
HSQC	Heteronuclear single-quantum coherence
IDP	Intrinsically disordered protein
IPTG	isopropyl- β -D-1-thiogalactopyranoside
LB	Lysogeny broth media
M9	Minimal media
MD	Molecular dynamics
MDG	Phosphate- (M), L-aspartic acid- (D) and glucose-based (G) media
mRNA	Messenger ribonucleic acid
MW	Molecular weight
MWCO	Molecular weight cut-off
NC	Nascent chain
NiDO2A	Nickel 1,4,7,10-tetraazacyclododecane-1,7-bis(acetic acid)
Ni-IDA	Nickel-iminodiacetic acid
Ni-NTA	Nickel-nitrilotriacetic acid

NMR	Nuclear magnetic resonance
OD	Optical density
PCR	Polymerase chain reaction
PLRE	Paramagnetic longitudinal relaxation enhancement
PRE	Paramagnetic relaxation enhancement
PTC	Peptidyl transferase centre
$R_{1(p/d)} [R_1^{H/N}]$	(Paramagnetic/diamagnetic) longitudinal relaxation rate [of $^1\text{H}/^{15}\text{N}$]
$R_{2(p/d)} [R_2^{H/N}]$	(Paramagnetic/diamagnetic) transverse relaxation rate [of $^1\text{H}/^{15}\text{N}$]
RNA	Ribonucleic acid
RNC	Ribosome-nascent chain complex
rRNA	Ribosomal ribonucleic acid
s.d.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SecM	Secretion monitor
$\text{SNR}_{(t)}$	Signal-to-noise ratio (per unit time $^{1/2}$)
SOFAST	Band-selective optimized-flip-angle short-transient
SORDID	Signal optimisation with recovery in diffusion delays
SRP	Signal recognition particle
STE	Stimulated echo
T_1	Longitudinal relaxation
T_2	Transverse relaxation
TBS(T)	Tris-buffered saline (with Tween-20)
T_{rec}	Recovery time
tRNA	Transfer ribonucleic acid
TROSY	Transverse relaxation-optimised spectroscopy
XSTE	Heteronuclear stimulated echo

Chapter 1

Introduction

1.1 Protein folding and structure

Protein folding is the physical process by which protein molecules acquire their native, three-dimensional structure that is essential to their biological activity and function. In Anfinsen's classic experiment, ribonuclease A was chemically denatured and then observed to spontaneously refold back into its native structure; it was therefore concluded that the amino acid sequence contains all the information required for proteins to adopt their three-dimensional fold (Anfinsen 1973).

Rather than sampling all the conformations possible from the very large number of degrees of freedom permitted by the numerous peptide bonds, the formation of native protein structure is driven by thermodynamics, kinetics, and the reduction of the Gibb's free energy of the whole system (Dill 1990). Over the last half century, *in vitro* protein folding studies have found that in general the predominant driving force for stabilising protein structure is the gain in water entropy upon burial of hydrophobic residues, termed the 'hydrophobic effect'. Protein structures may be further stabilised through a combination of hydrogen bond formation, electrostatic interactions (such as salt bridges) and van der Waal's interactions. These stabilising forces collectively result in an increase in enthalpy ΔH , countering the reduction in conformational entropy ΔS upon folding and compaction of the polypeptide chain. On balance, these opposing forces result in only a small reduction in Gibb's free energy of the system ΔG of 5-10 kcal/mol, where $\Delta G = \Delta H - T\Delta S$ and T is the absolute temperature. The native state is therefore only marginally more stable than its disordered, random coil conformation, and this may allow conformational changes crucial to its biological function to occur (Dill et al. 2008).

1.1.1 Mechanisms of protein folding from *in vitro* studies

Much of our understanding of protein folding mechanisms is based on Anfinsen's original experiment and the ability to chemically or thermally denature protein structures into unfolded conformations, and subsequently observe the refolding process in aqueous solutions. From this idea, several step-wise models of protein folding pathways have been proposed, essentially differing on the succession of secondary and tertiary structure formation. In the 'diffusion-collision' model, only secondary structure elements are

formed within a short-lived intermediate state, before native tertiary contacts are made (Karplus and Weaver 1994). As an example, kinetic data measured by NMR spectroscopy suggest that barnase folds via an intermediate state containing fully formed helices and β -sheets without significant native tertiary structure (Sanz and Fersht 1993). In contrast, in the 'hydrophobic collapse' model (Dill et al. 2008), observed in the folding of lysozyme (Kiefhaber 1995), the hydrophobic effect causes the rapid collapse of the protein into a 'molten globule' state, followed by structural rearrangement and formation of secondary and tertiary structure. Between these two extreme cases, the 'nucleation' model proposes a single rate-limiting step where a nucleus of small, localised structural units is formed and from which native secondary and tertiary structure rapidly propagate (Wetlaufer 1973). This model has been used to explain the two-step folding process of chymotrypsin inhibitor 2 (CI2) where only the unfolded and native states are significantly populated, and between which the transition state consists of a single α -helix within a diffuse, extended nucleus (Jackson and Fersht 1991). These classic models of protein folding have enabled descriptions of structure acquisition of relatively small proteins (less than 100 amino acids), and also alluded to the requirement of larger systems to proceed through several folding intermediates. It is becoming increasingly clear that for the most proteins, more involved events occur during the folding process that cannot be described by a single pathway.

The contemporary view is that for the majority of proteins, the mechanism by which structure is acquired is not simply a distinct, sequential assembly, but instead a complex series of parallel routes defined within an 'energy landscape' (Bryngelson et al, 1995) (Figure 1.1a). Through a stochastic search of accessible conformations within a rugged, funnel-shaped energy landscape, a large ensemble of high-energy unfolded polypeptide conformations are directed to the low energy native state, often the lowest energy. As native contacts are on average more stable than non-native ones, the energy surface is biased towards the native state, and this allows the transition from a random coil to the correct fold to occur by sampling only a small subset of conformations, enabling folding to occur very rapidly (Onuchic and Wolynes 2004). Experimentally it has been found that in some cases, this can occur within microseconds (Lindorff-Larsen et al. 2011).

Low energy minima within the folding landscape are also populated by ensembles of disordered states that may be lower in free energy than the native state. These occur where an imbalance of stabilising enthalpic and disruptive entropic forces arises, acting as kinetic traps, and can be exacerbated by perturbations in, for example, pH, temperature, and protein concentration, but can also be caused by mutations in the

protein sequence. In solution, this results in the formation of misfolded or aggregated protein. In cells, these events are deleterious, being implicating in medical conditions such as cystic fibrosis and neurodegenerative diseases such as Alzheimer's and Parkinson's (Chiti and Dobson 2006). An understanding of the mechanisms of protein folding and how off-pathway misfolding is avoided in the cell is therefore of both significant biological and medical interest.

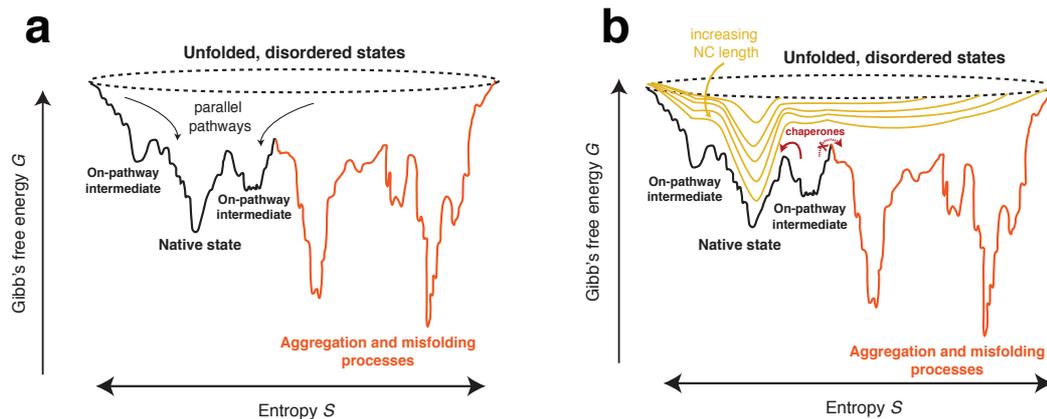


Figure 1.1 Energy landscape model for protein folding of (a) a full-length, denatured polypeptide *in vitro* and (b) of an actively translated polypeptide on the ribosome. (a) In the folding energy landscape, unfolded, disordered conformations populate high-energy states and follow parallel pathways towards the lower energy native state via intermediate conformations. Low-energy minima are also populated by ensembles of disordered states that lead to stable, misfolded or aggregated protein (orange). (b) *In vivo*, during translation, the energy landscape continually changes (yellow) as more conformations become accessible with the incremental addition of each new residue. Constraints within the ribosomal tunnel and auxiliary factors are likely to reduce the width of the accessible ensemble of structures. In addition, molecular chaperones (red) assist in reducing energy barriers between intermediate and native states, and increase activation energies between intermediate to misfolded or aggregated states.

1.2 Protein folding in the cell

While the basis of our knowledge in how proteins acquire structure have been underpinned by refolding studies in aqueous solutions, the experimental conditions used in this reductionist approach deviate significantly from how protein folding occurs in the cell. Indeed there is experimental evidence that folding in the cell is significantly more efficient than refolding *in vitro* (Hartl and Hayer-Hartl 2009). This discrepancy in folding efficiency is likely the result of a combination of features unique to *in vivo* folding and these are discussed below.

1.2.1 Protein biosynthesis by the ribosome

Across all domains of life, the ribosome is the site of protein biosynthesis during which the genetic information encoded within the genome sequence is translated into protein molecules. The 2.3-MDa, 70S *E. coli* ribosome is a complex macromolecular assembly of 54 different proteins and three large RNA molecules (Figure 1.2). The small 30S subunit is involved in the decoding of the mRNA. The large 50S subunit harbours both the peptidyltransferase centre (PTC) that forms the active site of peptide bond formation. Detailed structural studies (Dunkle and Cate 2010; Frank and Gonzalez 2010) have shown that through a series of concerted actions, the mRNA transcript is decoded by specific tRNA-amino acid complexes recruited to the PTC, where anti-codons of the tRNA molecules form base pairs with complementary codons in the mRNA. Peptide bond formation occurs between amino acids of two adjacently aligned tRNA molecules at the P- and A-sites. Displacement of the initial tRNA and recruitment of the tRNA complementary to the next codon in sequence occurs in parallel with a conformational change in the ribosome by ratchet-like rotations of the small subunit relative to the large subunit, as found by cryo-electron microscopy analysis of the 70S *E. coli* ribosome (Frank and Agrawal 2000). Following these events, further amino acids are added to a growing nascent chain in a step-wise, vectorial fashion (Schmeing and Ramakrishnan 2009).

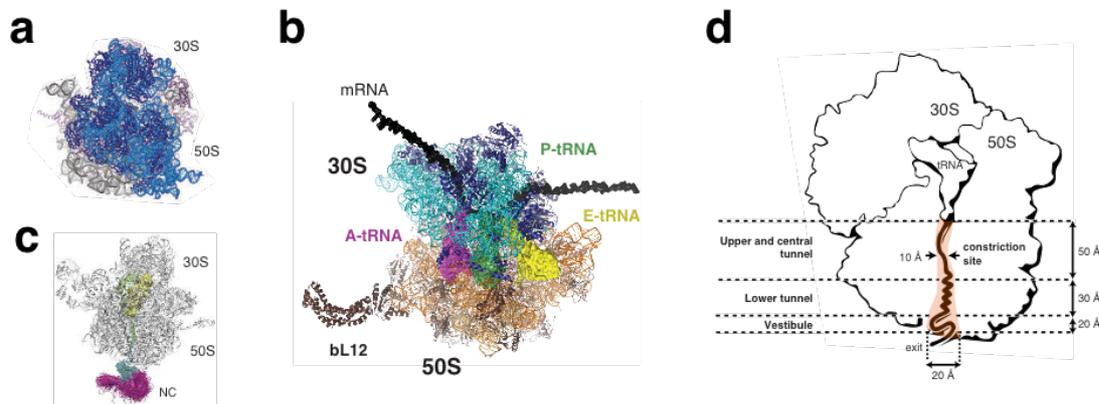


Figure 1.2 The structure of the prokaryotic ribosome. (a) Crystal structure of the intact 70S ribosome from *E. coli* with a resolution of 3.5 Å from a landmark study that paved the way for many other structural studies (Schuwirth et al. 2005). (b) 70S ribosome in complex with A-, P- and E-site tRNAs, and mRNA, which binds in a cleft between the 'head' and the 'body' of the 30S subunit. The mobile bL12 stalk region on the 50S was solved in isolation and fit onto the 70S ribosome (Schmeing and Ramakrishnan 2009). (c) Structural ensemble of RNCs of a pair of immunoglobulin-like domains (FLN5-6) calculated using NMR chemical shift restraints (Cabrita et al. 2016). (d) Schematic representation of the 70S ribosome showing the 30S and 50S subunits with the ribosomal exit tunnel traversing the large subunit.

The nascent polypeptide elongates through the ribosomal exit tunnel, which traverses the 50S subunit. The exit tunnel is ~ 100 Å long and has a diameter between 10 Å at its narrowest and 20 Å at its widest points, allowing it to accommodate between approximately 30-60 residues depending on whether the nascent chain adopts either an extended or more compact conformation respectively (Lu and Deutsch 2005a; Voss et al. 2006). The tunnel wall is predominantly formed by 23S rRNA and segments of uL4, uL22 and uL23 ribosomal proteins. At approximately 30 Å from the PTC, the exit tunnel is most constricted and is where the conserved β -hairpin loop of uL22 lies closest to the loop of uL4. At its furthest end, approximately 80 Å from the PTC, the tunnel widens up at the exit vestibule and is lined with various ribosomal proteins, such as the ubiquitously conserved uL22, uL23, uL24 and uL29 proteins, that form interaction sites of external ribosome-associated factors for the processing, targeting and folding of nascent chains (Kramer et al. 2009). Indeed, while much of the ribosomal exit tunnel remains rigid, the lengths between the PTC and constriction site, and the vestibule, are sufficiently flexible to accommodate elongation regulation (section 1.2.7) and co-translational protein folding (section 1.2.2) (Fulle and Gohlke 2009).

1.2.2 Co-translational protein folding

As an actively translated nascent polypeptide chain emerges from the ribosome in a vectorial manner from its N- to C-terminus, the total conformational space sampled by the nascent chain progressively increases with the addition of each new residue. However, the generally slow process of translation [15-20 and 3-4 amino acids per second in prokaryotes and eukaryotes respectively (Siller et al. 2010)] relative to the folding process means that the nascent polypeptide can start to sample conformations before complete synthesis. The acquisition of native-like protein structure in parallel with its translation, through the process of 'co-translational folding', can therefore significantly reduce sampling of the total conformational space, exerting a bias in the energy landscape towards the native state (Figure 1.1b).

Exploring the length-dependence of folding, biophysical studies using C-terminal truncations of proteins (de Prat Gay et al. 1995; Neira and Fersht 1999; Chow et al. 2003) have demonstrated how native-like (and non-native) folding can occur in the absence of the entire primary sequence. Most strikingly, short chain fragments of sperm whale apomyoglobin were found to fold into non-native β -sheets and self-assemble into amyloid-like species, while increasing chain length resulted in a transition to more native-like α -helices (Chow et al. 2003). While isolated from the ribosome, such studies have shown how folding can begin in the absence of the entire amino acid sequence.

However, tethering of the C-terminus of the nascent chain to the ribosome is likely to bias the folding pathway by, for example, constraining the extent of conformations accessible.

Studies of the 60-kDa multi-domain firefly luciferase have demonstrated how folding can also occur on the ribosome (Frydman et al. 1999). Limited proteolysis studies revealed the formation of an intermediate containing a folded N-terminal subdomain with a native-like, protease-resistant fold. When luciferase was denatured and refolded in isolation from the ribosome, the intermediate was not detected, even in the presence of molecular chaperones (section 1.2.7). This study therefore showed how *de novo* protein folding on the ribosome could follow a co-translational folding pathway with conformational intermediates distinct from those in renaturation experiments. The conclusion that nascent chains can adopt native-like structures when still ribosome-bound has been further corroborated in studies where RNCs were recognized by highly specific conformational antibodies (Makeyev et al. 1996), and in experiments that showed the formation of disulfide bonds in nascent chains enabling the acquisition of enzymatic activity (Land et al. 2003). In addition to such biochemical studies, structural evidence is now emerging of co-translational protein folding occurring both within and outside the ribosomal exit tunnel, as discussed below.

1.2.3 Interactions and protein folding within the exit tunnel

While the ribosomal exit tunnel was once thought to be merely a passive conduit, there is growing evidence that interactions can occur with the nascent polypeptide, allowing regulation of protein synthesis and assisting association of factors (section 1.2.7). For example, short peptide sequences such as SecM (Gumbart et al. 2012) and TnaC (Seidelt et al. 2009) interact with the ribosomal exit tunnel during biosynthesis leading to translation arrest. The C-terminal stalling sequence of SecM (¹⁵⁰FxxxxWIxxxxGIRAGP¹⁶⁶) interacts with the 23S rRNA bases A2058, A2062 and C2-methylated A2503 and these specific ribosomal interactions enable it to work as an independent translation-arrest element, even when engineered within an unrelated amino acid sequence (Huber et al. 2011; Gumbart et al. 2012).. In its biological setting, SecM is involved in the regulation of the cellular protein secretion activity in *E. coli*. Translational stalling by SecM briefly disrupts a hairpin in the *secM-secA* mRNA to expose the *secA* Shine-Dalgarno sequence, enabling *secA* translation by other ribosomes. In parallel, the nascent SecM polypeptide is co-translationally targeted by SRP (section 1.2.7) to the Sec translocase. SecA is therefore synthesized in the vicinity of the membrane, where SecA drives the release and export of SecM across the SecYEG transport channel into the periplasm

{Yap:2011ix}{Nakatogawa:2004ug}. Translational stalling by TnaC similarly leads to regulation of downstream genes (Gong and Yanofsky 2002).

Cryo-electron microscopy (cryo-EM) structures of RNCs (Becker et al. 2009; Seidelt et al. 2009) have shown that nascent chains can adopt specific conformations through interactions with the exit tunnel well beyond the PTC and before they fully emerge from the exit vestibule. While these analyses have shown that nascent chains remain mostly extended in the exit tunnel, there is evidence that at least some degree of intramolecular secondary and even simple tertiary interactions are possible for certain protein sequences (Figure 1.3), despite the narrow dimensions of the exit tunnel .

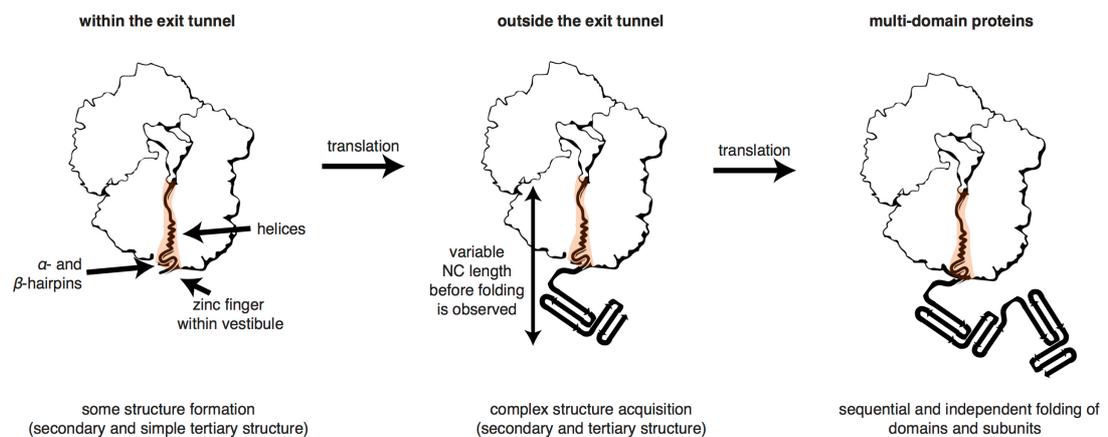


Figure 1.3 Protein folding on the ribosome. *De novo* protein folding is innately coupled to translation and therefore also the emergence of the nascent chain. Within the ribosomal exit tunnel (left), the constrained dimensions enable only some secondary structure elements to be formed, although recent studies have observed a simple tertiary structure motif within the vestibule, section 1.2.3. Outside of the exit tunnel (centre), the increase in conformational space enables tertiary and long-range contacts to be made, and more complex structure to be acquired. However, the extent to which the nascent chain has to emerge for native-like folding to occur is variable between proteins, section 1.2.4. In the folding of multi-domain proteins or proteins with multiple subunits (right), domains and subunits typically fold independently and sequentially from one another to reduce the probability of deleterious, stable inter-domain/unit contacts, which can lead to misfolded or aggregated states, section 1.2.4.1.

To examine the extent of folding in the tunnel, Förster resonance energy transfer (FRET) studies of a transmembrane segment have revealed that distances measured between the pair of fluorophores were consistent with compaction into helical conformations that were also observed when inserted into lipid bilayers (Woolhead et al. 2004). Moreover, in studies using an all-extended peptide in which its helical content was varied and where the accessibility of cysteine residues was probed by addition of PEG

resulting in a shift in its molecular mass (pegylation), it was found that distinct regions within the ribosomal exit tunnel play an active role in stabilising compact structures (Lu and Deutsch 2005a; Lu and Deutsch 2005b). Coarse-grain molecular dynamics (MD) simulations of a model helix in a cylindrical cavity have shown that such α -helices can be entropically stabilized simply by confinement in the cylindrical tunnel, and this effect is most pronounced with longer nascent chains (Ziv et al. 2005).

Closer toward the exit port of the tunnel, in the last 20 Å, the wider diameter permits greater flexibility of the nascent polypeptide and thereby the formation of intramolecular tertiary interactions to sequentially fold small subdomains, such as α - and β -hairpins (Kosolapov and Deutsch 2009). In addition, the recent observation of a simple tertiary motif, specifically a zinc-finger domain, by cryo-EM (Nilsson et al. 2015) suggests that the exit tunnel vestibule further provides a sufficiently wide yet sheltered folding environment akin to molecular chaperones (section 1.2.7). However, the limited dimensions of the exit tunnel have also been shown to restrict the accessible conformations of the emerging nascent chain. In recent studies of the N-terminal domain of HemK monitored in real-time by FRET, compact, non-native nascent chain conformations were observed that were likely located within the vestibule. However, only after the full domain sequence emerged out of the ribosomal exit tunnel did rearrangement into the native-like structure occur. It was postulated that this was a mechanism to drive the folding landscape away from misfolded states at shorter nascent chain lengths (Holtkamp et al. 2015).

1.2.4 Protein folding outside the exit tunnel

The studies described above collectively showed that while some preliminary protein folding can occur within the exit tunnel, the acquisition of more complex structures occurs outside of the ribosomal exit tunnel where the conformational space available is significantly expanded, enabling intra-molecular long-range tertiary contacts to occur (Figure 1.3).

However, the extent to which the length of nascent chain is required to emerge from the ribosome and for native-like folding to occur appears to vary between proteins. For example, studies have shown that RNCs of bovine rhodanese, bacterial chloramphenicol acetyltransferase and MS2 coat protein required 60, 72 and 44 N-terminal residues for antibody recognition, respectively, (Tsalkova et al. 1998) while the complete protein sequence with at least 26 additional residues were required for enzymatic activity of firefly luciferase to be observed (Makeyev et al. 1996). While there is clear variability in the length of nascent chain required for folding to occur outside of

the exit tunnel, likely in part by the extent of interactions and structure formation inside the exit tunnel (section 1.2.3), these studies demonstrated the ability for proteins to fold in the vicinity of the ribosome.

Several studies have also shown that co-translational folding enables the formation of on-pathway folding intermediates distinct from those *in vitro* to increase the probability of correct folding. For example, fluorescence studies in combination with limited protease digestion of fluorescent β -barrel proteins revealed that sequestration of 10-15 residues within the exit tunnel prevented stable barrel formation; folding only occurred once the C-terminus extended beyond the ribosome exit tunnel (Kelkar et al. 2012). However, RNCs with 10 of the 11 β -strands outside of the ribosomal exit tunnel formed stable, non-native but on-pathway intermediates that folded efficiently upon ribosome release (Kelkar et al. 2012). Moreover, in antibody-binding and limited proteolysis studies of P22 tailspike protein (TS) RNCs, progressive acquisition of structure was found concomitant with chain length (Clark and King 2001; Evans et al. 2008). Stable, compact conformations of the tailspike β -helix domain were populated on the ribosome prior to complete synthesis that were distinct from aggregation-prone, released and refolded truncations (Evans et al. 2008).

1.2.4.1 Co-translational folding of multi-domain proteins

The vast majority of the proteome consists of multi-domain proteins (Han et al. 2007), where sequence similarities between domains pose a unique challenge in folding. During the folding of such proteins, the risk of inter-domain, non-native interactions are significant (Borgia et al. 2011) and these can result in adverse misfolding and aggregation processes (Wright et al. 2005). Co-translational folding studies of multi-domain proteins have revealed that such deleterious, stable contacts are minimized through several, different means (Figure 1.3).

In the folding of cystic fibrosis transmembrane conductance regulator (CFTR) (Kleizen et al. 2005), Ras-DHFR fusion protein (Netzer and Hartl 1997), firefly luciferase (Frydman et al. 1999), and the SH3 domain of α -spectrin (Eichmann et al. 2010), (sub)domains acquire native structure independently and sequentially upon emergence from the ribosome, during which succeeding, C-terminal units remain occluded within the ribosomal exit tunnel.

In contrast, in NMR studies of another multi-domain protein (FLN), RNCs of the fifth immunoglobulin-like domain (FLN5) were shown to require a further 11-14 residues of the subsequent domain in addition to the full FLN5 sequence for persistent, native-like folding to appear (Cabrita et al. 2009; Cabrita et al. 2016). The restriction in

the folding of the domain was attributed, in part, to interactions with the ribosomal surface that is thought to reduce inter-domain contacts – the co-translational folding of FLN5 RNCs is discussed in more detail in Chapter 3. It therefore appears that several strategies have evolved to regulate native multi-domain structure acquisition, but each involves the formation of folding intermediates distinct to those found in the absence of the ribosome.

1.2.5 Translation rates and protein folding

Protein folding is innately coupled to translation, and indeed decreases in the speed of translation elongation are sufficient to affect the folding efficiency (Tsai et al. 2008). The significant differences between the abundances of different tRNA have been found to cause variation in the rate of translation of each codon. This results in a discontinuous translation rate and a distribution of codons, where codons with highly abundant cognate tRNAs are translated more rapidly and are also found more frequently. Conversely, rare codons with low abundance of cognate tRNAs have slower translation rates (Komar 2009).

In a study on the folding of SufI in *E. coli*, the translation rate was enhanced by addition of a large excess of tRNA and synonymous substitution of slow-translating codons with codons pairing to highly abundant tRNAs. Using these two strategies, slow-translating codons were identified that clustered together, allowing temporary separation of individually folding segments to kinetically stabilize each unit and progressively build the native state (Zhang et al. 2009). Similarly, it has been found that rare codons are often clustered in regions encoding for domain boundaries and β -strands elements, which can fold rapidly but require long-range stabilising interactions (Thanaraj and Argos 1996). By controlling the rate of translation and increasing the likelihood of co-translational protein folding to occur, off-pathway kinetic traps leading to misfolded states can therefore be avoided.

A more recent example of this is seen in FRET experiments, in which largely the translation elongation rates thorough codon usage appeared to be tailored to the requirements of particular regions within the first nucleotide-binding domain from the CFTR (Kim et al. 2015). Folding occurred in sequential, kinetically-coupled steps to allow the formation of the β -sheet core by delaying the folding of the α -subdomain until the C-terminal strands were available, facilitating correct folding of the entire domain (Kim et al. 2015).

Synonymous codon usage has also been found to direct the co-translational folding of some proteins towards different conformations. For example, a synonymous

single-nucleotide polymorphism (SNP), in which the codon-usage pattern is altered to incorporate clusters of rare codons without any amino acid substitutions, in the Multidrug Resistance 1 (MDR1) gene results in a change in the timing of the co-translational folding of P-gp protein, and thereby altering its substrate and inhibitor interaction site conformations (Kimchi-Sarfaty et al. 2007). In a more recent study, Gamma-B crystallins were produced from mRNAs with altered codon bias but the same amino acid sequence. By monitoring the co-translational folding in real time by FRET and fluorescence intensity changes, it was found that modulation of the translation rates manifested into different protein conformations associated with different cysteine oxidation states. It therefore appears that synonymous codons provide a secondary code for *in vivo* protein folding (Buhr et al. 2016).

The translation rate can also be altered by other mechanisms to modulate the folding of proteins. For instance, sequences with positively-charged lysines and arginine can lead to charge-specific interactions within the exit tunnel, significantly reducing elongation rates during protein synthesis (Lu and Deutsch 2008). Alternatively, secondary structure elements within mRNA molecules affect both the local stability of the mRNA but also the speed and fidelity of translation. Indeed, particularly stable RNA structures are thought to synchronise with the rates of co-translational folding of certain structural elements within proteins (Faure et al. 2016).

1.2.6 The cellular environment

Once the nascent polypeptide has emerged from the ribosomal exit vestibule, it is exposed to the cellular environment, a densely packed milieu with up to 300 g/L of macromolecules, such as proteins, nucleic acids and carbohydrates, occupying ~30% of the cellular volume (Zimmerman and Trach 1991). This results in the macromolecular crowding effect: a combination of excluded volume effects and increased intermolecular, non-specific interactions, such as electrostatic or hydrophobic interactions (Gershenson and Gierasch 2011). The presence of crowders reduces the available space to proteins through repulsive forces. This entropic, excluded volume effect has been shown to favour compaction of unfolded and disordered states, shifting folding equilibria to more stable, less extended forms, and this can be further reinforced by non-specific interactions that are repulsive. In contrast, attractive, non-specific interactions with the protein backbone favour greater surface exposure; they are therefore destabilising to folded states and may increase their propensity to unfold (Wang et al. 2012). The overall result of macromolecular crowding is dependent on the balance of these opposing effects.

For example, in a study of an unfolded variant of the immunoglobulin domain G binding domain of protein L (ProtL), monitored by NMR spectroscopy and circular dichroism (CD), the folding equilibrium was shifted to the stabilised folded state on addition of dextran, an artificial crowder introducing excluded volume forces (Ådén and Wittung-Stafshede 2014). However in a separate study, NMR experiments of the same protein within living *E. coli* cells (in-cell NMR) showed that folding did not occur because destabilising non-specific attractive interactions (negligible with dextran) were not compensated by the favourable excluded volume effect (Schlesinger et al. 2011).

Although crowding effects can shift the folding equilibrium to more compact and stable states, these can also promote the formation of misfolded states or aggregates (van den Berg et al. 1999). The extent of such off-pathway folding may, however, be limited as molecular chaperones (section 1.2.7) can display increased rates of activity in crowded conditions, such as protein disulfide isomerase, which has been found to show enhanced chaperone activity under such conditions, sufficient to prevent the aggregation of lysozyme (van den Berg et al. 1999).

Finally, macromolecular crowding also results in an increase in viscosity, retarding both molecular translational and rotational diffusion and, depending on the extent of the forces described above, can substantially reduce the rates of cellular processes, including protein folding (Dhar et al. 2010).

It is therefore clear that the cellular environment perturbs the structure and folding of proteins, in ways that are often difficult to replicate *in vitro*, as illustrated in the study of ProtL described above. In-cell NMR spectroscopy has therefore emerged as a high-resolution technique to observe biological macromolecules *in situ*. In-cell NMR is the focus of Chapter 4, where more detailed analysis of in-cell studies and the effects of macromolecular crowding can also be found.

1.2.7 Molecular chaperones and other auxiliary factors

As the nascent chain is synthesised on the ribosome and emerges from the exit tunnel, a host of auxiliary factors within the cellular milieu compete to bind and interact. Of most interest in the context of protein folding are molecular chaperones, a group of proteins that maintain protein homeostasis (proteostasis) in the cell and ensure quality control over protein folding across all kingdoms of life, despite the evolution of distinct chaperone systems (Kramer et al. 2009). In addition to assisting in *de novo* folding and refolding of aggregated proteins (Figure 1.1b), the multitude of chaperones within the cell enables their versatile functions in other cellular processes such as assembly of oligomeric proteins, protein transport, and protein degradation (Hartl and Bracher

2011). Whereas some chaperones are constitutively expressed, others are stress-induced, such as heat shock proteins (HSP) (Rauch et al. 2005).

The most detailed studies of ribosome-associated chaperone systems are of trigger factor, found in prokaryotes but which shares functional similarities with the two ribosome-associated systems in eukaryotic cells: the highly conserved nascent chain-associated complex (NAC); and the Hsp70/Hsp40-based chaperone system consisting of Ssz and its co-chaperone Zuo, which form the ribosome-associated complex (RAC), and work together with the Ssb in yeast to form the 'ribosomal chaperone triad' (Wegrzyn and Deuerling 2005). While little is known on the mechanisms of action and structures of the eukaryotic ribosome-associated chaperones, it has been shown that trigger factor can restore at least some of their functions when expressed in *Saccharomyces cerevisiae* yeast cells lacking the chaperone triad (Ito 2005), suggesting functional conservation of ribosome-associated chaperones despite structural differences.

In the absence of translation, trigger factor cycles on and off the ribosome with a mean residence time of 11-15 s (Maier et al. 2003; Kaiser et al. 2006), but in the presence of nascent chains, the affinity of trigger factor to the ribosome is increased up to 30-fold (Rutkowska et al. 2008). Trigger factor binds near the exit tunnel of the 70S ribosome in a 1:1 stoichiometry through electrostatic interaction, resulting in a conformational change into a "cradle" that exposes a hydrophobic surface towards within the ribosomal exit vestibule (Hoffmann et al. 2010), confirmed by the X-ray crystal structure (Schlunzen et al. 2005) and a subsequent solid-state NMR study (Barbet-Massin et al. 2015) of trigger factor bound to the large ribosomal subunit. The mechanism of action therefore involves binding to exposed hydrophobic patches of nascent chains; FRET studies have shown that it can remain bound to these regions even when dissociated from the ribosome (Kaiser et al. 2006).

These findings have been corroborated in a recent NMR study of trigger factor, which revealed a multivalent binding, holdase mechanism between the chaperone and a protein substrate, the 47-kDa alkaline phosphatase (PhoA) (isolated from the ribosome), to keep PhoA in an extended, unfolded conformation (Saio et al. 2014). Multiple sites in trigger factor, located in two different domains that were found to bind transiently in several, dispersed regions of PhoA, mediated predominantly by hydrophobic contacts and conformational rearrangements within the trigger factor binding sites, rather than through recognition of specific substrate primary structure. The rapid cycling of trigger factor between nascent chains (lifetime of ~ 20ms) may be particularly important in protecting the folding of multi-domain proteins, further evidenced by other studies with β -galactosidase and luciferase as substrates (Agashe et al. 2004). In the same NMR study,

promiscuous recognition of trigger factor was also found to enable its unfoldase activity with the aggregation-prone maltose-binding protein (MBP) by shielding exposed hydrophobic regions yet retaining the substrate in its disordered conformation (Saio et al. 2014). Following dissociation from trigger factor, nascent chains can interact downstream with other chaperones, including DnaK/J and GroEL/GroES (Hartl and Hayer-Hartl 2009).

The binding of trigger factor to ribosomes competes with recruitment of the signal recognition particle (SRP), which mediates the transport of secretory and membrane proteins to the plasma membrane (in prokaryotes) or the endoplasmic reticulum (in eukaryotes), demonstrating the interplay of different cellular factors on the ribosome (Akopian et al. 2013). In addition to N-terminal signal sequences on nascent polypeptides, interactions between the nascent chain and ribosomal exit tunnel can also recruit the SRP: in *E. coli*, translating ribosomes that harbor short nascent chains (~30 residues) in particular have an approximately 100-fold increase in affinity for SRP (relative to vacant ribosomes), despite such nascent chains being too short to reach the end of the exit tunnel. It is thought that interactions with the intra-tunnel loop of uL23, presumably non-specific, induce a conformational change of uL23 to actively recruit SRP (Bornemann et al. 2008; Berndt et al. 2009). Protein folding in the cell therefore occurs in a dynamic, complex environment, where both folding and nascent chain processing can occur in a co-translational fashion.

1.2.8 Summary

It is clear that the unique aspects of *de novo* protein folding in the cell described above each modulate the energy landscape (Figure 1.1a) significantly from its *in vitro* analogue (Figure 1.1b). While early work in protein folding relied on renaturation of full-length proteins, higher levels of complexity through the use of C-terminal truncations, translationally stalled RNCs, manipulations to the translation rates and the recent development of in-cell NMR (although yet to be applied to RNCs), have helped narrow the gap between observations made *in vitro* and their relevance *in vivo* (Figure 1.4).

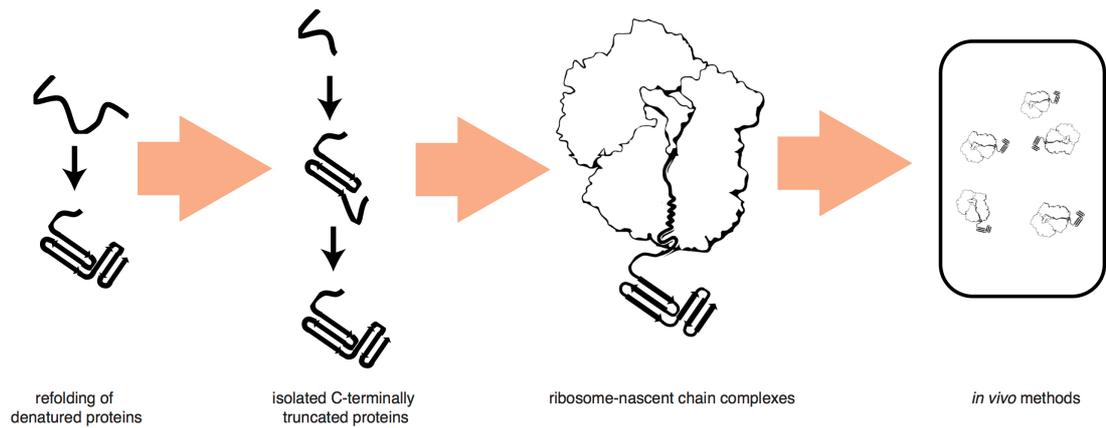


Figure 1.4 Progression of strategies used to study *de novo* protein folding. Classical studies of protein folding exploited the ability of full-length proteins to adopt disordered conformations in denatured states before dilution in renaturing buffer conditions to observe spontaneous refolding to the native state (far-left). The vectorial emergence of the nascent chain was investigated initially using isolated proteins that were progressively truncated at the C-terminus (left). More recent protein folding studies involved the use of ribosome-nascent chain complexes, where often a C-terminal stalling motif is introduced to induce translational arrest (right), and in which perturbations to its folding can be induced by, for instance, manipulations in translation rates or the length of the nascent chain. Studies are now emerging of RNCs co-translationally folding observed in real-time, but detailed structural studies of ribosome-bound nascent chains are relatively sparse. *In vivo* methods (far-right) are being increasingly used in the study protein structure and dynamics, but have yet to applied to RNCs or extensive protein folding studies (see Chapter 4).

However, there is still little understanding on the structural information on the emerging nascent chain on translating ribosomes, particularly on folding outside of the exit tunnel, where more complex protein structure acquisition is thought to occur. The inherently dynamic nature of the ribosome-bound nascent chain has so far eluded high-resolution techniques such as X-ray crystallography and cryo-EM, but is highly amenable to their study by NMR spectroscopy. Recent RNC studies by NMR spectroscopy (Cabrita et al. 2016; Deckert et al. 2016) have begun to provide insights, at a residue-specific level, the role of the ribosome in the folding of nascent polypeptide chains and interactions between the nascent chain and ribosomal surface (section 1.3.2). However, there is a notable absence of direct structural information on ribosome-bound nascent chains, and also how applicable the conclusions of *in vitro* studies of RNCs are to the *in vivo* environment. In this work, we describe the development of strategies using both *in vitro* and in-cell NMR spectroscopy in studies of ribosomal particles, which we anticipate will assist in addressing these questions (see Perspectives, section 1.5).

1.3 Applications of biomolecular NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for the study of biological macromolecules. By non-invasively probing the local magnetic environment of individual nuclei, high-resolution characterisation of molecular structure, dynamics, and interactions is achieved with near-atomic resolution.

At equilibrium, protein molecules are characterised by an ensemble of conformations that undergo continual exchange within a defined time-scale and so the observed state is a time- and ensemble-averaged conformation. The versatility of NMR spectroscopy has uniquely enabled the ability to obtain high-resolution structural information on dynamic molecules using a variety of methods that report on the different conformational changes and processes (Mittermaier and Kay 2009). This has provided detailed, atomistic insights into the functions and mechanisms of action of important biological molecules and processes across very wide range of times, from ps to s, including protein folding, which occurs within this range but can also occur at longer timescales (hours) (Markwick et al. 2008) and is therefore ideally suited to their study by NMR (Figure 1.5).

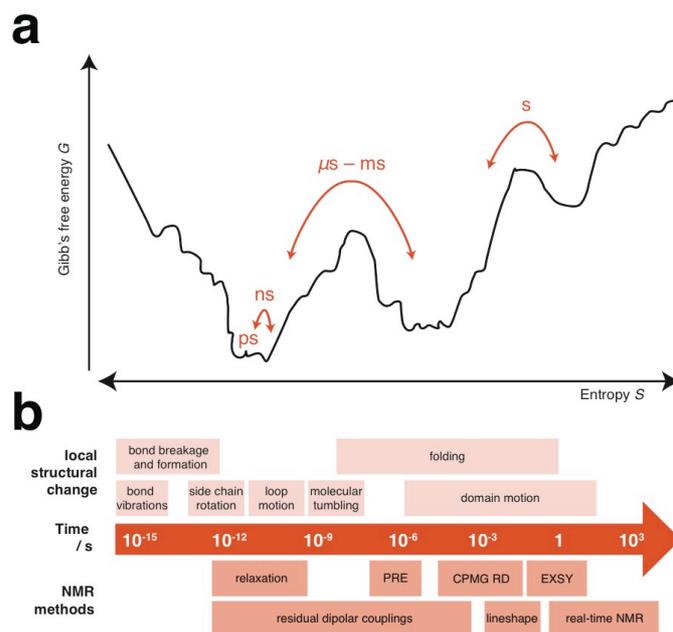


Figure 1.5 Measuring protein dynamics by NMR spectroscopy. (a) Schematic energy landscape showing thermodynamically accessible protein conformations within defined time-scales. (b) Protein conformational changes over a range of time-scales (upper) can be studied with a variety of different NMR methods (lower): nuclear spin relaxation measurements, residual dipolar couplings, paramagnetic relaxation enhancement (PRE), Carr-Purcell Meiboom-Gill Relaxation Dispersion (CPMG RD), lineshape analysis, exchange spectroscopy (EXSY), and real-time NMR. Figure adapted from reference (Kleckner and Foster 2011), where a more detailed discussion of these techniques can be found.

With the capability to acquire such rich information, molecular dynamic (MD) simulations have emerged as an increasingly important tool in structural determination, in which NMR data can be used as restraints in MD calculations to either bias force fields towards the biological structure or validate simulation results (Robustelli et al. 2010).

1.3.1 NMR spectroscopy of large biological systems and in living cells

The study of large macromolecular assemblies and protein complexes has previously been the preserve of structural techniques such as X-ray crystallography and cryo-EM. However, the advent of cryogenic probes and increased availability of high-field spectrometers (Kovacs et al. 2005) (section 1.4.4), alongside developments in NMR sample preparations and experimental schemes (section 1.4.5), within the last half century heralded a shift in the use of NMR spectroscopy beyond its classical size limitation (< 25 KDa). These advances have collectively increased the sensitivity and resolution of NMR experiments, and have alleviated the two main problems of NMR studies of large systems: the significant signal overlap arising from the increased number of probes present in larger complexes, and fast relaxation of the NMR signal resulting in low spectroscopic sensitivity (section 1.4.4). These advances, such as transverse relaxation-optimised spectroscopy (TROSY, section 1.4.5) and the development of isotopic labelling strategies (section 4.3.3), have recently been demonstrated in landmark studies of the 900-KDa GroEL-GroES complex (Fiaux et al. 2002) and the 670-KDa 20S proteasome (Sprangers and Kay 2007). In particular, the unique ability of NMR spectroscopy to probe dynamic regions of structure have enabled the study of ribosomal particles by NMR spectroscopy to study co-translational protein folding (Christodoulou et al. 2004; Cabrita et al. 2016; Deckert et al. 2016; Cassaignau et al. 2016), and this is discussed in more detail in section 1.3.2.

Alongside the application of NMR to such large macromolecular assemblies, the same technological advances in NMR have also enabled its use in the observation of biomolecules within living cells, termed in-cell NMR (Freedberg and Selenko 2014). NMR spectroscopy is well poised to study *in vivo* structure and dynamics because of its non-invasive nature coupled with its near-atomic resolution. We discuss applications and technical aspects of in-cell NMR in Chapter 4.

1.3.2 NMR spectroscopy of ribosomal particles

The ability of NMR spectroscopy to study dynamic structures was exploited in studies of *E. coli* 70S ribosomes (Christodoulou et al. 2004). Despite the high molecular weight of the complex (~ 2.3 MDa), approximately 100, well-resolved ^1H , ^{15}N -heteronuclear single-quantum correlation (HSQC) resonances were observed, attributable to the hinge and C-

terminal domains of the mobile bL12 stalk proteins. This demonstrated that the dynamic bL12 proteins tumbled relatively independently from the ribosomal structure. This was corroborated by ^{15}N -spin relaxation measurements, which showed that the apparent rotational correlation time (13.6 ns) was longer than that of free bL12, but shorter than expected for a rigidly ribosome-bound protein. The chemical shifts of ribosome-bound bL12 were not significantly perturbed relative to those of the isolated protein, indicating that its structure remained largely unchanged despite its attachment to the ribosome (Christodoulou et al. 2004).

These results demonstrated the utility and feasibility of NMR spectroscopy to probe the mobile regions of the ribosome, and thereby formed the basis of subsequent investigations of ribosome-bound nascent chains by NMR. NMR studies of RNCs were first reported for a pair of immunoglobulin-like domains derived from the gelation factor of *Dictyostelium discoideum* (FLN5-6) (Hsu et al. 2007). NMR spectra of uniformly ^{15}N - and ^{13}C -labelled FLN5-6 RNC showed that the FLN5 domain adopted a native-like fold while the FLN6 linker remained largely disordered. In more recent studies of the same RNC system (Cabrita et al. 2016), the FLN6 linker was progressively truncated towards the C-terminus generating a series of co-translational folding intermediate mimics, which was subsequently observed by NMR. The results indicated that interactions between the nascent chain and the ribosomal surface prevented the folding of the FLN5 domain at particular linker lengths (longer than those necessary for folding in isolation from the ribosome), as a possible mechanism to reduce the probability of inter-domain contacts and subsequent misfolding processes. This study is discussed in more detail in Chapter 3. The length-dependence folding on the ribosome has also been investigated for the SH3 domain from α -spectrin, which utilized the longitudinal relaxation-optimised ^1H - ^{15}N -CRINEPT-HMQC- ^1H -TROSY experiment (sections 1.4.4 and 1.4.5).

Nascent chain interactions with the ribosomal surface have similarly been observed in NMR studies of ribosome-bound α -synuclein, an intrinsically disordered protein (IDP) (Deckert et al. 2016). While the nascent chain remained highly disordered on the ribosome, analysis of the residue-specific signal intensities indicated interactions of the N-terminal region with the ribosome, which have been suggested to arise from both electrostatic interactions and interactions with aromatic residues (A. Deckert, *et al.*, in preparation; A.M.E. Cassaignau, *et al.*, in preparation), and are likely to have biological implications, particularly in the context of co-translational protein folding.

1.3.2.1 Technical aspects of NMR investigations of RNCs

The preparation of homogenous, stable and isotopically enriched RNC samples is critical to their study by NMR. Initial studies generated RNC samples using an *in vitro* coupled transcription-translation system (Hsu et al. 2007). By using a linearized DNA template lacking a stop codon, release factors cannot bind, leaving the C-terminus of the nascent chain covalently bound to the P-site tRNA by the aminoacyl ester bond. While this method generates isotopically labelled nascent chains on background-free ribosomes, the preparation is extremely costly.

An alternative approach is the generation of RNCs *in vivo* within their native environment. Translational arrest of nascent chains is achieved by incorporating a C-terminal stalling motif that interacts directly with the ribosomal exit tunnel, such as SecM or TnaC (section 1.2.3). To generate selectively labelled nascent chains in high yields, *E. coli* cells are initially grown in minimal media that can sustain high-cell densities. Expression is then induced together with incorporation of isotopic labelling, such that only the nascent chain is labelled. RNCs are then purified by sucrose cushions, followed by metal affinity purification to select for only ribosomes harbouring a nascent chain (Cassaignau et al. 2016).

The features of NMR studies of ribosomal particles are the high molecular weight and the resulting low maximum achievable sample concentrations of $\sim 10 \mu\text{M}$, which are 1-2 orders of magnitude smaller than typically used in NMR experiments. In addition, purified RNCs are typically stable for only a few hours in buffered solutions, which collectively means that sensitivity is at a premium and that continuous assessment of the sample is essential. These significant spectroscopic challenges mean that improvements in sensitivity are crucial to RNC studies by NMR. We outline the current strategies to overcome low sensitivity in this work.

Critical to all of the NMR studies of ribosomal particles are the use of NMR diffusion measurements, which enable calculation of the diffusion coefficient of the observed species to ensure that the observed NMR resonances derive exclusively from ribosome-bound rather than free species. NMR diffusion measurements and their application to RNCs are the focus of Chapter 2, where a more comprehensive overview of this technique can be found.

1.4 Introduction to NMR theory

A general overview of the theory of NMR as applied to biological molecules is given below. Details of specific NMR experiments and strategies used in this study have been described in the Introductions of the corresponding Chapters.

1.4.1 The physical origin of NMR

The basis for nuclear magnetic resonance is the observation that certain atomic nuclei have intrinsic magnetic moments. These NMR-active nuclei possess a nuclear spin quantum number of $I > 0$, and include ^1H , ^{31}P , ^{13}C and ^{15}N . The latter two isotopes occur with natural abundances of only 1.11% and 0.37% respectively, but biological molecules can be enriched by labelling biomolecules with ^{13}C - and ^{15}N -isotopes through biosynthetic pathways (section 4.3.3).

In the absence of an external magnetic field, the magnetic moments of the nuclei are randomly orientated resulting in a net zero magnetization of the sample. When the sample is placed into the NMR spectrometer, an external magnetic field B_0 is applied, and the distribution of nuclear spin orientations becomes biased towards lower energy states. These states are aligned with the magnetic field along the z -axis, producing equilibrium z -magnetization. The energy gap ΔE between the two spin states defines the 'Larmor frequency' of the spin (section 1.4.2), and is proportional to the strength of the external magnetic field: $\Delta E = \hbar\gamma B_0$, where \hbar is the reduced Planck's constant and the gyromagnetic ratio γ is a constant relating to a nucleus' sensitivity to the magnetic field; ^1H has the highest gyromagnetic ratio, while ^{13}C and ^{15}N are approximately $1/4$ and $1/10$ as sensitive as ^1H respectively. Populations between energy levels are manipulated by radio frequency pulses during NMR experiments and differences in population directly relate to the magnetization of the sample. However, the population differences, given by the Boltzman distribution, that are detected are inherently small ($\Delta E / RT \sim 10^{-4}$) limiting the sensitivity of NMR spectroscopy.

1.4.2 The basic pulse-acquire experiment

In the most basic one-dimensional (1D) experiment, pulse-acquire (Figure 1.6a), equilibrium z -magnetization is rotated onto the $-y$ -axis (perpendicular to the external frame) by a brief radio frequency pulse. Magnetization then precesses in the transverse plane at the Larmor frequency $\omega = -\gamma B_0$ which is detected as the free induction decay (FID) by a radio frequency receiver coil. The FID follows an exponential decay as the bias in the transverse plane reduces (relaxes) and spins return to the equilibrium magnetization. The FID is converted from a time-domain signal to frequency-domain

signal by Fourier transformation which gives rise to a resonance of a particular frequency within an interpretable spectrum, in this case, a 1D NMR spectrum

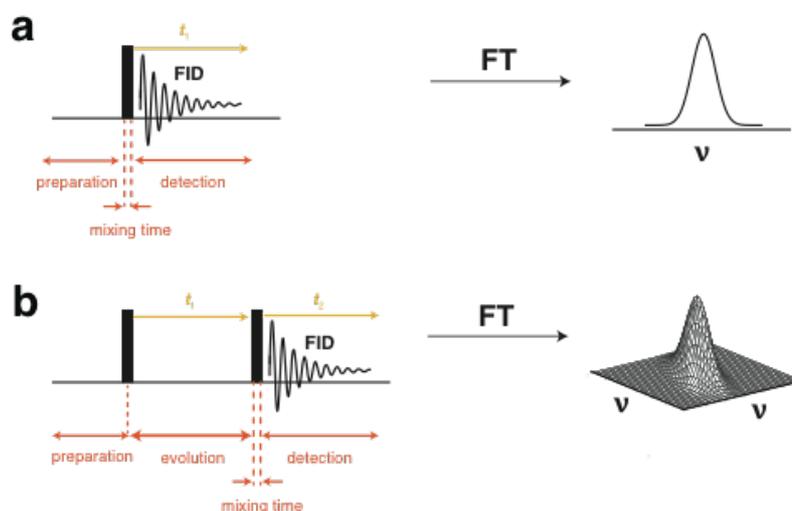


Figure 1.6 The NMR experiment. (a) The basic pulse-acquire experiment is composed of a preparation time, mixing time during which a radio-frequency pulse is applied, and a detection time when the free induction decay (FID) is detected. Fourier transform is then applied to the FID to produce a frequency-domain signal that is more readily interpretable. (b) In multi-dimensional NMR, a further evolution time is inserted after the preparation time for each additional dimension (see main text).

The observed resonance frequencies in NMR spectra are directly related to the local chemical environments of each nucleus. The difference between the signal frequency and the reference frequency from a standard molecule is known as the chemical shift δ and is independent of the applied field. Chemical shifts arise from motions of electrons, induced by the external magnetic field, which generate secondary fields. The net magnetic field around each nucleus is therefore affected by both the external static field and the local secondary field. In general, the effect of the local secondary field, termed nuclear shielding, is to oppose the external magnetic field, effectively reducing its strength. Perturbations in chemical shifts therefore report on a change in chemical environment and structure of the molecule, and can thus, for example, determine the site of interactions with other proteins or ligands.

1.4.3 Multi-dimensional NMR

NMR spectra are a rich source of information on the electronic shielding and thus the chemical environment of individual nuclei. All this information can in principle be extracted from 1D experiments, although for larger molecules, the resulting complexity renders this practically impossible. Resolving the spectral information across multiple

orthogonal dimensions can minimize spectral crowding of NMR signals so that more detailed interpretations become accessible.

In two-dimensional (2D) NMR, an indirect evolution time t_1 is inserted after the preparation time (i.e. the first set of radio frequency pulses), during which magnetization is encoded with information on the second frequency dimension (Figure 1.6b). A mixing period follows, where magnetization is transferred into an observable state, which is then detected during the direct evolution time t_2 . A 2D spectrum is therefore acquired as sets of 1D spectra, each with increasing indirect evolution periods. The experimental scheme for three-dimensional (3D) NMR is analogous to that for 2D NMR, with the addition of a second evolution period followed by a mixing period. In this case, the second evolution period gives rise to information on the third frequency dimension.

The acquisition of multi-dimensional NMR becomes increasingly more time-consuming with a reduction of $\sqrt{2}$ in sensitivity for each additional dimension, as each additional evolution period is independently incremented and both the imaginary and real components of the signal must be sampled. Increasingly more scans of the experiment are required to sample the indirect-domain time axis in step-wise increments, but reducing the number of experimental scans NS necessitates a trade-off between spectral width SW and/or resolution Δf in the indirectly detected dimension as $SW/\Delta f = NS$ (where the resolution in turn is proportional to the maximum evolution time in the indirect dimension $t_{1,max}$) (Zhang et al. 2014). The spectral width is therefore typically set to the minimum range to improve resolution and/or sensitivity (see also section 1.4.4)

To further reduce spectral overlap, particularly for large biological molecules, the shifts of different nuclei, such as ^1H and ^{13}C or ^{15}N , can be correlated using heteronuclear correlation experiments by transfer of magnetization through relatively large one-bond heteronuclear scalar couplings (magnetic moments of the electrons in the bonding orbitals). For example, the heteronuclear single-quantum correlation (HSQC) experiment is widely used for recording one-bond correlations between ^{15}N and ^1H of the amide backbone in protein NMR. The resonances can be readily assigned using triple-resonance experiments (where ^1H , ^{15}N and ^{13}C spins are correlated together), and this is discussed in Chapter 3.

1.4.4 Relaxation and NMR sensitivity

Relaxation is the process that results in the bulk magnetization returning to its equilibrium during an NMR experiment. Longitudinal (T_1) relaxation corresponds to the z-magnetisation returning to its non-zero equilibrium value and alignment with the

external magnetic field after a radio frequency pulse (section 1.4.). Transverse (T_2) relaxation relates to the dephasing of x - and y -magnetisation to its equilibrium value of zero due, in part, to a distribution of frequencies of local fields.

The rates of both longitudinal and transverse relaxation (the inverse of the relaxation times, $R = 1/T$) depends on the overall rate of rotational tumbling of the molecule in solution, characterised by the rotational correlation time, and chemical exchange processes. The rotational correlation time τ_c can be estimated using the Stokes-Einstein-Debye equation (Ortega and García de la Torre 2005):

$$\tau_c = \frac{4\pi\eta r_h^3}{3k_B T} \quad (1.1)$$

where k_B is the Boltzmann constant, T is the temperature, η is the viscosity and r_h is the hydrodynamic radius of the molecule. The relaxation times as a function of the rotational correlation time are shown in Figure 1.7.

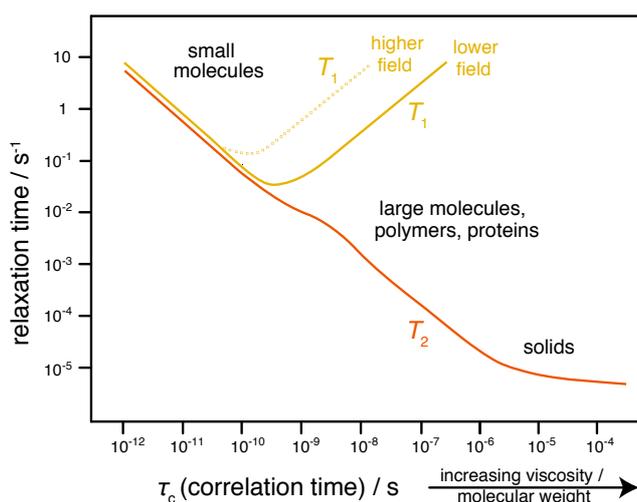


Figure 1.7 The dependence of the longitudinal (T_1) and transverse (T_2) relaxation times on the rotational correlation time τ_c , which in turn is proportional to the viscosity and molecular weight. Adapted from (Bloembergen et al. 1948).

The relaxation rates of molecules can be perturbed by the presence of unpaired electrons from paramagnetic reagents. As the magnetic moment of electrons are much greater than that of protons, local fields generated by electrons, which contribute to relaxation, are also much greater, such that the presence of paramagnetic species significantly enhances the relaxation of the molecule under investigation (Clare and Iwahara 2009). We demonstrate that this effect can be usefully employed to significantly enhance the rate of longitudinal relaxation, thereby increase the sensitivity of NMR experiments in Chapters 2 and 3, where other strategies to optimize longitudinal relaxation are also discussed.

The linewidths $\Delta\nu$, the full-width at half-height of NMR resonance line shapes, are related to the rate of transverse relaxation R_2 by $\Delta\nu = R_2/\pi$ in s^{-1} , and this affects the resolution and signal-to-noise ratio (SNR) of NMR spectra (Cavanagh et al. 2010). The signal intensity of each NMR peak (peak height) is optimised by complete longitudinal relaxation between each experiment. The integral of the area or volume under the peak is constant to the number of nuclei resonating at the same frequency (thus, the concentration of the sample), and therefore the signal intensity is attenuated (exponentially) by continuous transverse relaxation during the experiment. The result of the latter is that large molecules such as proteins, which have a long rotational correlation time and thus have rapid transverse relaxation (Figure 1.7), suffer from increased broadening of resonances and low spectroscopic sensitivity. While a greater number of experimental scans can be used to increase the signal, noise is simultaneously acquired during the same time period meaning that the overall sensitivity is proportional to only \sqrt{NS} .

At the hardware level, the NMR sensitivity can be improved by increases in spectrometer field strength ($SNR \propto B_0^{3/2}$) (Rovnyak et al. 2004), and the use of cryogenic probes, which reduce the resistance and thermal noise in the coil and preamplifier to enhance the sensitivity by 3 to 4-fold compared to conventional room-temperature probes (Kovacs et al. 2005). These technological advances have alleviated some of the insensitivity of NMR as applied to large biological systems in particular, but have been complemented with developments in experimental schemes as well.

1.4.5 Transverse relaxation-optimised spectroscopy (TROSY)

As previously discussed, the fast transverse relaxation modulated by the rotational tumbling of molecules in solution severely restricts the size of molecules that can be studied by NMR spectroscopy. The development of TROSY (Pervushin et al. 1997), which reduces the effective transverse relaxation rate, has enabled a dramatic increase in the size limit of NMR (previously < 25 kDa), permitting the study of systems in the 100s kDa to MDa range, such as the 900-kDa GroEL-GroES complex (Fiaux et al. 2002).

The TROSY effect originates from the interference of multiple pathways of relaxation, which can have an additive or subtractive influence on the effective relaxation rate. An example is found in 1H - ^{15}N bonds, where scalar couplings between 1H and ^{15}N nuclei result in a splitting of the 2D resonance into four components, each with significantly different linewidths which directly reflect the relaxation interference (Figure 1.8b). In conventional 2D NMR experiments, for example, the HSQC experiment, the four components are typically collapsed by ‘decoupling’ into a single resonance with

an averaged relaxation rate (Figure 1.8a). The TROSY technique exclusively selects for the slowly relaxing and narrow resonance, eliminating the remaining faster, broader cross-peaks (Figure 1.8c) (Pervushin et al. 1997). The TROSY effect therefore discards 75% of the signal, but are generally outweighed by the reductions in transverse relaxation effects for large molecules (> 25 kDa) (Fernández and Wider 2003).

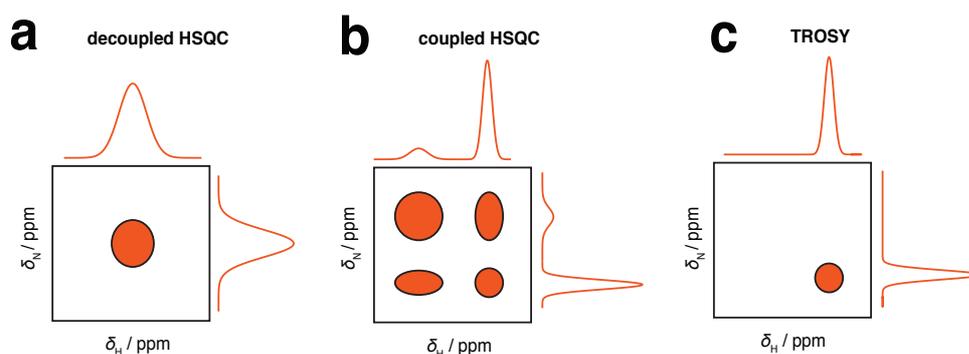


Figure 1.8 Transverse relaxation-optimised spectroscopy, TROSY. (a) In a typical 2D HSQC experiment, decoupling pulses are applied resulting in a single resonance in which relaxation rate and thus linewidth is averaged. (b) Without decoupling, the HSQC signal is split into the four components, each with different relaxation rates that result in different linewidths. (c) In TROSY, only the slowly relaxation resonance line is selected, eliminating the faster relaxing signal, resulting in a single resonance with a narrower linewidth.

Reductions in the effective relaxation rates can also be achieved by labelling strategies through complete (perdeuteration) or partial replacement of hydrogen atoms with deuterons. The theoretical and technical aspects of perdeuteration, and selective ^1H , ^{13}C -labelling of methyl groups within a deuterated background in combination with the ‘methyl-TROSY’ effect, are discussed in section 4.3.3.

1.5 Perspectives

Advances in NMR spectroscopy in all aspects, from sample preparation to experimental schemes and availability of more powerful hardware, have enabled their application to large macromolecular assemblies, including ribosome-nascent chain complexes. These studies are beginning to bring insights, at a residue-specific level, into how the ribosome modulates the co-translational folding of nascent chains (Christodoulou et al. 2004; Cabrita et al. 2016; Deckert et al. 2016). As previously discussed, there are significant spectroscopic challenges in the NMR study of RNCs, and continuous improvements in sensitivity are crucial to progress in these investigations. We therefore developed a strategy, described in Chapter 2, to improve the sensitivity of NMR experiments that are

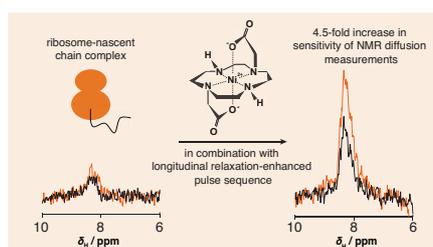
typically used in the study of RNCs, NMR diffusion measurements to monitor their stabilities over time, and 2D $^1\text{H},^{15}\text{N}$ -correlation experiments, using paramagnetic longitudinal relaxation enhancement (PLRE).

To demonstrate how such improvements in sensitivity can be usefully employed, we applied the PLRE effect in NMR experiments that have hitherto been unattainable in sensitivity. The acquisition of 3D NMR experiments in parallel with residual dipolar coupling (RDC) measurements of RNCs is described in Chapter 3, where we also discuss how these data can be used to gain high-resolution and direct structural information on ribosome-bound nascent chains, which has so far eluded most studies. We chose the FLN5 RNC as a model system in our investigations as we were motivated to understand whether the ‘delay’ in folding on the ribosome (section 1.3.2) resulted in intermediate structure acquisition during co-translational folding.

Finally, in Chapter 4, we discuss the development of a strategy to translate our understanding of RNCs and co-translational folding from *in vitro* studies, into investigations within living cells. We employed different labelling strategies to observe ribosome-bound and free species, enabling distinction of RNCs *in vivo*. In addition, to being to explore how the cellular environment affects the structure and dynamics of nascent chains during protein folding, we used reconstituted cytosol in experiments of disordered FLN5 in isolation, and found that the excluded volume effect and non-specific interactions with cytosolic components perturbed its structure from its *in vitro* conformation; we propose that these macromolecular crowding effects (also observed in-cell) are likely to have structural and biological implications on FLN5 on the ribosome. We anticipate these studies will help in narrowing the gap between *in vitro* observations and their relevance *in vivo*, and together with structural information from *in vitro* studies, enable a more comprehensive understanding of how protein folding occurs in the cell.

Chapter 2

Increasing the sensitivity of NMR diffusion measurements by paramagnetic longitudinal relaxation enhancement



Translational diffusion of macromolecules within NMR samples can be examined by non-invasive means using stimulated echo (STE) experiments, enabling accurate measurements of their molecular sizes. These measurements can be important probes of intermolecular interactions and processes such as protein folding and unfolding, and are crucial in monitoring the integrity of large macromolecular assemblies such as RNCs. However, NMR studies of such ribosomal complexes can be severely constrained by their slow tumbling, low solubility (with maximum concentrations of up to $10 \mu\text{M}$), and short lifetimes resulting in weak signal, and therefore continuing improvements in spectroscopic sensitivity are crucial. In this Chapter, we explore the use of the paramagnetic longitudinal relaxation enhancement (PLRE) agent NiDO2A on the sensitivity of heteronuclear XSTE and SORDID STE experiments, which can be used to monitor the integrity of such unstable complexes. We exploit the dependence of the PLRE effect on the nuclear gyromagnetic ratio and the electronic relaxation time to accelerate recovery of ^1H magnetization without adversely affecting storage on N_z during diffusion delays, nor introducing significant line broadening associated with transverse relaxation. By applying the longitudinal relaxation-optimized SORDID pulse sequence together with NiDO2A, we obtain up to a 4.8- and 1.7-fold increase in sensitivity for disordered and globular proteins, respectively. When applied to 70S *Escherichia coli* ribosomes and RNCs, NMR diffusion signal enhancements of up to 4.5-fold are achievable, alongside ~ 1.9 -fold improvements in two-dimensional NMR sensitivity. We anticipate these results will significantly advance the use of NMR spectroscopy to probe dynamic regions of ribosomes and other such large, unstable macromolecular assemblies.

2.1 Introduction

2.1.1 NMR diffusion measurements

NMR diffusion measurements are a powerful probe of biological molecular structure and dynamics in which the translational properties of molecules can be examined non-invasively, using a very wide variety of gradient echo NMR experiments (Johnson 1999; Sprangers and Kay 2007); a summary of the theory of these experiments is outlined below. These measurements can be utilized to determine diffusion coefficients, which in turn can be related to hydrodynamic radii and hence molecular structure by the Stokes-Einstein equation. The development of NMR stimulated echo methods has thereby advanced studies in a wide range of biological areas, such as the analysis of peptide aggregation and amyloid formation (Baldwin et al. 2008; Eichmann et al. 2010); macromolecular crowding effects (Li et al. 2009; Hyberts et al. 2012); protein-ligand binding events (Lucas and Larive 2004; Waudby and Christodoulou 2012); measuring hydrodynamic radii within different conformational ensembles including partially folded and molten globule states (Wilkins et al. 1999; Pervushin et al. 2002); and in-cell NMR to distinguish between intra- and extracellular proteins (Waudby et al. 2012), the latter of which is further discussed in Chapter 4. With increasing applications of NMR spectroscopy in understanding the biology of complex systems, NMR diffusion measurements are likely to develop growing prominence.

NMR diffusion measurements have also played an important role in NMR studies of large macromolecular assemblies, including investigations of ribosomes and RNCs (Christodoulou et al. 2004; Schanda et al. 2005; Hsu et al. 2007; Cabrita et al. 2009; Eichmann et al. 2010; Cabrita et al. 2016; Deckert et al. 2016). The study of such ribosomal particles is of major biological interest, but the high molecular weight and the resulting low maximum achievable concentrations, most often combined with limited sample lifetimes (hours to days), commonly results in very low signals that present significant spectroscopic challenges (Gil et al. 2013; Waudby et al. 2013b). For example, RNC sample lifetimes are limited primarily by release of the nascent chain from the ribosome before degradation of the ribosome itself (Ferrage et al. 2003; Waudby et al. 2013b; Cabrita et al. 2016). The monitoring of translational diffusion throughout the NMR experimental time, typically in an interleaved manner, therefore ensures that the observed resonances arise from an intact complex. In particular, isotope-edited diffusion experiments, such as the heteronuclear stimulated-echo (XSTE) experiment (Ferrage et al. 2003) have been critical in allowing the attachment of the isotopically-labelled nascent chain to the unlabelled ribosome to be monitored specifically (Hsu et al. 2007;

Cabrita et al. 2009; Eichmann et al. 2010; Waudby et al. 2013b; Cabrita et al. 2016; Deckert et al. 2016),; this approach is similar to one first used to study the dynamic bL12 stalk regions of non-translating ribosomes (Christodoulou et al. 2004).

2.1.1.1 Theory of NMR diffusion measurements

Diffusion NMR measurements use magnetic field gradients, during which the field becomes spatially dependent. The field strength is typically varied linearly along the direction of the external magnetic field, and defined by the rate of change of the field (i.e. in $T\ m^{-1}$). In a stimulated echo sequence, the first gradient 'encodes' a phase on a nuclear spin that is dependent on its position in the NMR sample; the second gradient 'decodes' the phase. If the spin has not been displaced during the diffusion delay between these two gradients, the phases cancel out. However, if the spin moves during the delay, it acquires a net phase depending on the distance displaced. The total spins in the sample therefore collectively acquire a phase error, leading to attenuation of the NMR signal that is dependent on both the gradient strength and the diffusion coefficient (Johnson 1999). By repeating the experiment across different gradient strengths, the diffusion coefficient of the sample can be calculated using the Stokes-Einstein equation.

2.1.2 Improving the sensitivity of NMR experiments

Continual improvements in experimental sensitivity are central to progress in studies of ribosomal particles, given such constraints their study by NMR. Recent gains in spectroscopic sensitivity and resolution have been made through the availability of high-field spectrometers (Rovnyak et al. 2004) and cryogenic probes (Kovacs et al. 2005). Transverse relaxation optimized spectroscopy (TROSY) (Pervushin et al. 1997; Fernández and Wider 2003) and in particular, methyl-TROSY (Tugarinov et al. 2003) in combination with advanced isotopic labeling schemes (Tugarinov et al. 2006), have revolutionised the study of large systems by NMR spectroscopy, such as the 900-kDa GroEL-GroES complex (Fiaux et al. 2002) and 670-kDa 20S proteasome (Sprangers and Kay 2007). These methods are now beginning to be applied to the study of RNCs (Eichmann et al. 2010). Furthermore, by sampling more efficiently on the Nyquist grid, through non-uniform (sparse) sampling (Hyberts et al. 2012) or non-uniform weighted sampling (Waudby and Christodoulou 2012), sensitivity improvements in multi-dimensional NMR can be achieved in parallel.

2.1.3 Longitudinal relaxation-optimised experiments

In typical NMR measurements, the majority of spectrometer time (>90%) is devoted to the recovery delay between scans, during which no new data is actively acquired but in which the sample magnetization slowly recovers towards Boltzmann equilibrium through the process of longitudinal relaxation (see Chapter 1). The impact of this delay on experimental sensitivity has been known since the earliest days of Fourier transform NMR, when optimal signal recovery was found to be intrinsically linked to the length of the recovery delay, as well as the excitation flip angle (Ernst and Anderson 1966). More recently, to accelerate the longitudinal relaxation of observed spins, longitudinal cross-relaxation following selective excitation of a subset of spins has been exploited. This optimization of longitudinal relaxation was first applied in LTROSY experiments (Pervushin et al. 2002), and was later coupled with selective Ernst angle excitation and fast repetition rates in the SOFAST-HMQC experiment (Schanda et al. 2005). In these experiments, since saturation of water protons is avoided, the latter experiment has proven to be particularly advantageous for IDPs as there is an additional transfer of 'cold' water protons onto the protein via rapid amide proton exchange (Gil et al. 2013).

Longitudinal relaxation-optimized experiments have also been developed for NMR diffusion measurements. The XSTE scheme was the first pulse sequence to exploit the disparities in longitudinal relaxation times between different nuclei (as discussed in Chapter 1), enabling storage of the spatially-encoded magnetization on ^{15}N spins rather than on ^1H (such as in homonuclear stimulated echoes) to reduce the loss of signal during the long diffusion delay (Ferrage et al. 2003) (Figure 2.1a). This was then combined with the selective excitation of amide protons (BEST-XSTE), and Ernst angle excitation with shorter recovery times (SOFAST-XSTE) (Augustyniak et al. 2011) to increase experimental sensitivity. In this Chapter, we focus on the more recently introduced SORDID experiment, which provides further signal gains by enabling the recovery of ^1H magnetization to occur simultaneously with the diffusion of spatially-encoded ^{15}N longitudinal magnetization, thereby reducing the overall experimental time to approximately 50% that of conventional XSTE experiments (Augustyniak et al. 2012a) (Figure 2.1b).

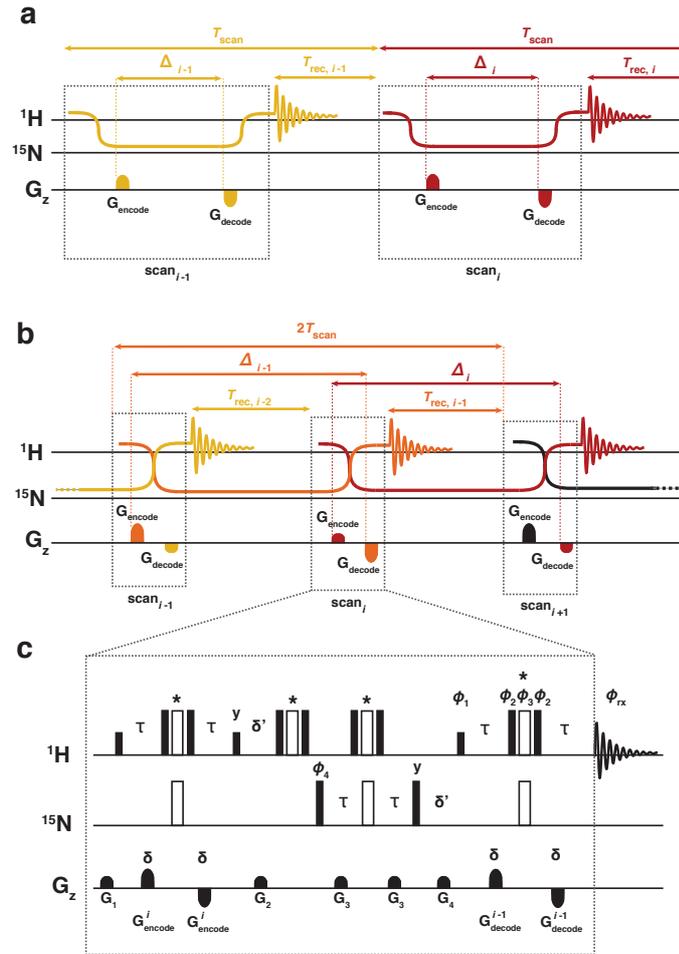


Figure 2.1 Schematic (a, b) and detailed (c) pulse sequences for ^{15}N heteronuclear diffusion experiments, adapted from reference (Augustyniak et al. 2011). (a) In XSTE experiments, after a recovery period T_{rec} , proton magnetization is transferred onto N_z where it is encoded during the diffusion delay Δ and, after decoding, is detected on ^1H in one scan. (b) In SORDID experiments, the diffusion delay extends over two scans, allowing the recovery time of the previous scan to occur during the diffusion delay of the following scan. (c) In the detailed pulse sequence of SORDID, filled and open rectangles represent 90° and 180° pulses, respectively. Hard composite pulses 75° , 285° , 75° are indicated by stars, and amide-selective 90° squared pulses are represented by lower filled rectangles. Unless otherwise mentioned, all pulses are applied along the x-axis of the rotating frame. The phase cycling employed was: $\varphi_1 = (-x)_8, (-y)_8, x_8, y_8$; $\varphi_2 = (34^\circ)_4(124^\circ)_4(214^\circ)_4(304^\circ)_4$; $\varphi_3 = (144^\circ)_4(234^\circ)_4(324^\circ)_4(324^\circ)_4$. In addition, with careful consideration of the interleaving scans, the following phase cycling was employed to select heteronuclear coherences: $\varphi_4 = x, -x, x, x$; noting that the receiver phase shift $\Delta\varphi_{\text{rx}}^{(i)} = \Delta\varphi_4^{(i)}\Delta\varphi_4^{(i-1)}$, the receiver phase was $\varphi_{\text{rx}} = x, -x, -x, x$. The receiver phase encompassing the two phase cycles was therefore $\varphi_{\text{rx}} = x, -x, -x, x, -x, x, x, -x, -y, y, y, -y, y, -y, -y, y, -x, x, x, -x, x, -x, -x, x, y, -y, -y, y, -y, y, y, -y$. The INEPT delays were $\tau = -1/4J_{\text{NH}} = 2.72$ ms ($J_{\text{NH}} \approx -92$ Hz). The encoding and decoding gradients ($G_{\text{encode}}^i = G_{\text{decode}}^i$) had variable strengths but constant duration $\delta = 1$ ms.

2.1.3.1 Paramagnetic longitudinal relaxation enhancement

Another, orthogonal, approach to increase the rate longitudinal relaxation is the introduction of soluble paramagnetic compounds into NMR samples (Eletsky et al. 2003; Hiller et al. 2005; Cai et al. 2006; Wickramasinghe et al. 2007). The solvent paramagnetic relaxation enhancement (PRE) effect arises from long-range magnetic dipolar interactions between unpaired electrons from the paramagnetic center and a nucleus, and results in an increase in both longitudinal and transverse relaxation rates (Otting 2010):

$$R_1 = R_{1d} + R_{1p} \quad (2.1)$$

$$R_2 = R_{2d} + R_{2p} \quad (2.2)$$

where R_{1d} and R_{2d} are the diamagnetic relaxation rates in the absence of a PRE agent, and R_{1p} and R_{2p} are the additional contributions to relaxation induced by the paramagnet. Without specific interactions between the PRE agent and the protein, the 'outer sphere' or diffusional model, paramagnetic relaxation is due to fluctuations in dipolar interactions between nuclei and electrons that arises from a combination of intrinsic electron spin relaxation and the translational diffusion of the PRE agent relative to the nucleus (Bertini et al. 2001; Bernini et al. 2009). In this case, the paramagnetic components of the relaxation rates are given by (Helm 2006):

$$R_{1p} = \frac{32\pi}{405} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{1000 N_A [M] \gamma_n^2 \gamma_e^2 \hbar^2 S(S+1)}{d (D_M + D_P)} [7 J_2(\omega_e) + 3 J_1(\omega_n)] \quad (2.3)$$

$$R_{2p} = \frac{16\pi}{405} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{1000 N_A [M] \gamma_n^2 \gamma_e^2 \hbar^2 S(S+1)}{d (D_M + D_P)} [4 J_1(0) + 13 J_2(\omega_e) + 3 J_1(\omega_n)] \quad (2.4)$$

where μ_0 is the permeability of free space, N_A is Avogadro's constant, $[M]$ is the concentration of the paramagnetic species, γ_n and γ_e are the nuclear and electronic gyromagnetic ratios, ω_n or ω_e are the nuclear and electronic Larmor frequencies, \hbar is the reduced Planck constant, S is the electron spin quantum number, D_M and D_P are the diffusion coefficients of the paramagnetic species and the protein, and d is the distance of closest approach between the paramagnetic center and nuclear spin. The spectral density functions $J_i(\omega)$ (with $n = 1, 2$) are:

$$J_n(\omega) = \frac{1 + \frac{1}{4} \left(n\omega\tau_D + \frac{\tau_D}{T_{ne}} \right)^{1/2}}{1 + \left(n\omega\tau_D + \frac{\tau_D}{T_{ne}} \right)^{1/2} + \frac{4}{9} \left(n\omega\tau_D + \frac{\tau_D}{T_{ne}} \right) + \frac{1}{9} \left(n\omega\tau_D + \frac{\tau_D}{T_{ne}} \right)^{3/2}} \quad (2.5)$$

where T_{1e} and T_{2e} are the longitudinal and transverse electron relaxation times, and the diffusional correlation time $\tau_D = d^2/(D_M+D_P)$ (Bertini et al. 2001). For small paramagnetic compounds, τ_D is on the order of 2 ns (when based on a closest approach distance of 1 nm and a 5-Å hydrodynamic radius). In contrast, electronic relaxation times can vary over several orders of magnitude, from ps to μ s (Bertini et al. 2001), and this can therefore strongly influence the relative magnitudes of the paramagnetic contributions R_{1p} and R_{2p} .

In this respect, a variety of paramagnetic relaxation agents have been employed to enhance the sensitivity of NMR experiments. For example, in NMR in the solid state, where T_1 -relaxation is particularly slow, sensitivity enhancements of up to 2.9-fold were obtained in the presence of the paramagnetic Cu(II)-EDTA complex (Wickramasinghe et al. 2007). In solution-state NMR studies, doping samples with 1 mM Gd(III)-DO2A provided sensitivity increases of up to 1.3-fold in TROSY experiments of the 800-kDa chaperone GroEL (Hiller et al. 2005). However, the acceleration of ^1H longitudinal relaxation was accompanied in both cases by increased line broadening effects associated with significantly large increases in the transverse relaxation rates. This can be attributed to the long electron spin relaxation times of Cu(II) and Gd(III) (10^{-9} - 10^{-8} s) (Bertini et al. 2001) which are comparable to the diffusional correlation time. While such transverse PRE effects can be useful to structurally probe the solvent accessibility of proteins (Clore and Iwahara 2009; Madl et al. 2011), in general these effects are detrimental for optimizing the experimental sensitivity.

In contrast, paramagnetic metal ions such as Fe(III) and Ni(II), which have electronic relaxation times much shorter than their diffusional correlation time (e.g. $\sim 10^{-11}$ s for Ni(II)) (Rantaharju et al. 2014) can be used to shorten recycle times by increasing the rate of proton longitudinal relaxation rates with only very minimal effect on transverse relaxation (Cai et al. 2006; Theillet et al. 2011). In this case, Equations 2.3 and 2.4 (above) reach a limit in which the relaxation rate $R_{2p} \approx 1.2 R_{1p}$. This limit is not sensitive to the details of the relaxation mechanism: an 'inner sphere' model in which the paramagnetic agent binds transiently to the protein was previously analyzed and found to give similar results (Cai et al. 2006). Since longitudinal relaxation in biological molecules is typically much slower than transverse relaxation ($R_{1d} < R_{2d}$), R_1 can therefore be increased significantly with little relative effect on R_2 (Equations 2.1-2.2); we term this the paramagnetic longitudinal relaxation enhancement (PLRE) effect.

In this Chapter, we exploit the dependence of the PLRE effect of NiDO2A (a neutral and hydrophilic chelate of Ni(II)) on the square of the gyromagnetic ratio (Equations 2.3-2.4) to improve the sensitivity of heteronuclear NMR diffusion

measurements, in which magnetization spatially encoded on ^1H is stored longitudinally on ^{15}N spins, which has a smaller gyromagnetic ratio (Ferrage et al. 2003). We show that sensitivity enhancements can be achieved by using longitudinal relaxation-optimized SORDID experiments to exploit reduce recycle times in NiDO2A-doped samples of both α -synuclein, an IDP and FLN5, a globular immunoglobulin domain from *Dictyostelium discoideum* (FLN) that we have previously used in co-translational folding studies (Hsu et al. 2009; Cabrita et al. 2009). When applied to isotopically labeled *E. coli* 70S ribosomes, and to an RNC we show that such large gains in sensitivity can be also achieved without compromising sample integrity (as experimentally observed). We therefore expect these enhancements will greatly facilitate future NMR investigations of such large, dilute, and unstable macromolecular machines; examples of which are discussed in Chapter 3.

2.2 Results

2.2.1 PLRE studies on isolated proteins

The effect of NiDO2A was first explored on the sensitivity of the 2D SOFAST-HMQC correlation experiment, and XSTE and SORDID NMR diffusion measurements of two well-characterized, structurally different isolated proteins (i.e. off the ribosome). Two-dimensional ^1H - ^{15}N SOFAST-HMQC experiments were acquired on uniformly ^{15}N -labeled samples of α -synuclein (Figure 2.2a) and FLN5 (Figure 2.2d), in both the absence and presence of NiDO2A PLRE agent at a concentration of 40 mM. We found that NiDO2A did not induce chemical shift changes and only marginal line broadening was observed; this is consistent with previous studies using NiDO2A (Cai et al. 2006; Theillet et al. 2011). This spectroscopically observed absence of interaction has been attributed to the very tight affinity of Ni^{2+} to DO2A^{2-} ligand with sub-femtomolar K_d (Chang et al. 1999), which outcompetes the interaction of Ni^{2+} with the hexahistidine purification tag present in FLN5. Moreover, the overall neutral charge of the resulting complex is thought to minimize electrostatic interactions with nucleic acids and proteins (Cai et al. 2006).

To quantify the NiDO2A-induced SOFAST-HMQC sensitivity enhancements, we acquired a series of 1D ^1H - ^{15}N SOFAST-HMQC experiments and varied the recovery delay between 50 ms and 1 s. The SNR of the resulting 1D spectra was calculated by integration of the amide envelope (see Chapter 6). The sensitivity improved up to 1.4-fold for FLN5 as the concentration of NiDO2A was increased to 40 mM NiDO2A (Figure 2.2b), while 2D SOFAST-HMQC experiments showed residue-specific sensitivity enhancements of 1.45 ± 0.17 (s.d.), which were distributed uniformly across the entire protein sequence. But in the presence of 60 mM NiDO2A, the highest concentration tested, the improvement in sensitivity observed was smaller, with a 1D SOFAST-HMQC sensitivity increase of 1.3-

fold and mean residue-specific increase of 1.25 ± 0.17 , due to increasing ^1H transverse relaxation rates. However, in all cases, line widths did not broaden such that resonances became overlapped; indeed for NiDO2A concentrations up to 60 mM, we could still acquire well-resolved spectral resonances.

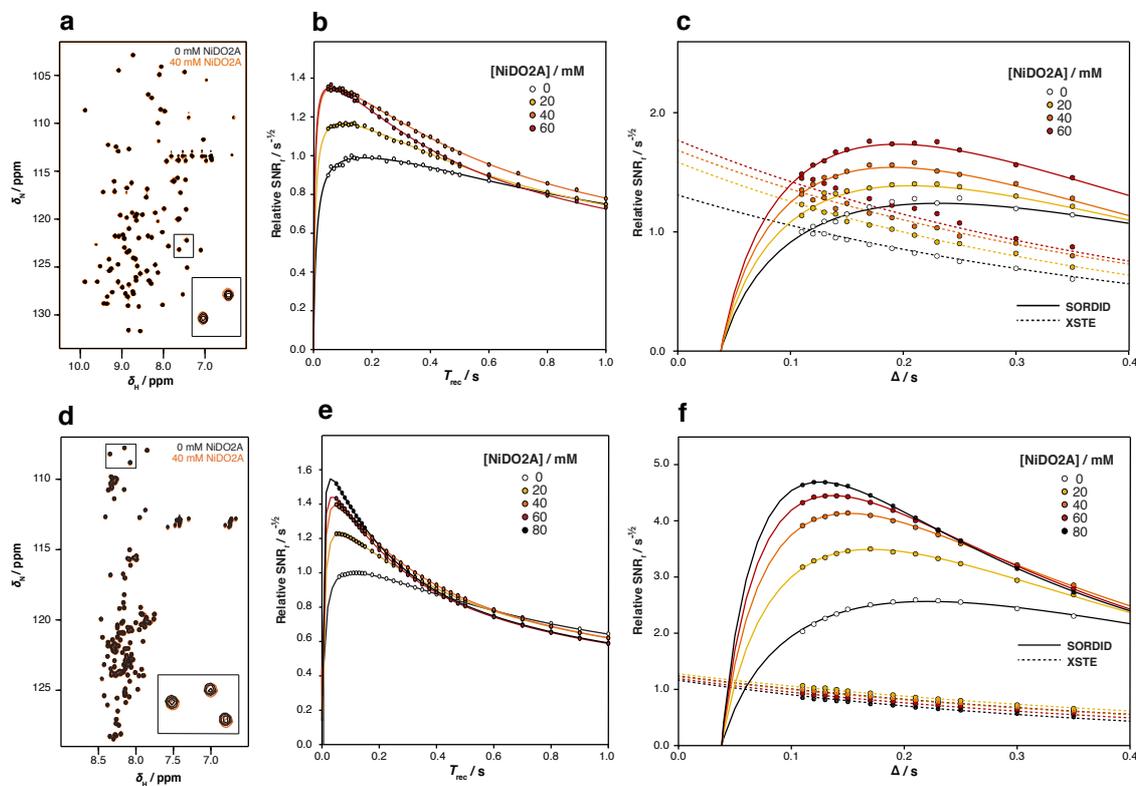


Figure 2.2 The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of NMR measurements of isolated proteins, (a-c) FLN5 and (d-f) α -synuclein. Overlay of 2D ^1H - ^{15}N SOFAST-HMQC spectra (a, d) in the presence and absence of 40mM NiDO2; insets show magnified views of highlighted resonances. The relative sensitivity of (b, e) 1D ^1H - ^{15}N SOFAST-HMQC experiments as a function of the recovery delay T_{rec} , and of (c, f) XSTE and SORDID diffusion experiments as a function of the diffusion delay Δ . Data are globally fitted to their theoretical sensitivity expressions (Equations 6.3–6.4)

We next explored the use of the PLRE agent with FLN5 in NMR diffusion experiments (Figure 2.2c). The sensitivity of the XSTE experiment (Ferrage et al. 2003) was compared with the longitudinal relaxation-optimized SORDID experiment (Augustyniak et al. 2012b); we adjusted G_{max} to obtain constant echo attenuation (I/I_0) across a range of diffusion delays between 110 and 350 ms. The value of the SNR acquired from each diffusion delay (and normalized for the square root of total experimental time) was fitted globally using a constant scaling factor and experimentally determined transverse relaxation rates to theoretical expressions for SNR_t , (Equations 6.3-6.4), as obtained by theoretical analysis of the trajectory of magnetization during the

XSTE and SORDID pulse sequences (Table 2.2). We found good agreement between the fitted ^{15}N and ^1H longitudinal relaxation rates and those measured directly using inversion-recovery experiments in general (Table 2.1). Where differences arise, this may at least in part be due to the differential weighting of different residues in the 1D amide envelope observed by diffusion NMR and relaxation measurements. Only small increases in R_1^{N} were observed as expected (Equation 2.3) such that loss of sensitivity is minimized during storage of magnetization on ^{15}N nuclei, and this is therefore particularly advantageous to heteronuclear diffusion NMR measurements.

[NiDO2A] / mM	$R_1 (^{15}\text{N}) / \text{s}^{-1}$			$R_1 (^1\text{H}) / \text{s}^{-1}$			
	observed	XSTE fit	SORDID fit	observed hard	observed amide- selective	XSTE fit	SORDID fit
0	2.00 ± 0.26	1.73 ± 0.18	2.31 ± 0.13	0.77 ± 0.03	5.49 ± 1.31	0.85 ± 0.27	1.26 ± 0.19
20	1.98 ± 0.26	1.91 ± 0.19	2.62 ± 0.14	1.24 ± 0.06	6.15 ± 1.10	1.76 ± 2.81	1.62 ± 0.25
40	1.99 ± 0.28	1.70 ± 0.15	2.82 ± 0.15	1.75 ± 0.05	6.88 ± 1.12	2.26 ± 3.15	1.98 ± 0.31
60	2.02 ± 0.26	1.60 ± 0.28	2.61 ± 0.16	2.57 ± 0.08	7.93 ± 0.95	8.69 ± 9.27	2.31 ± 0.36

FLN5. Scaling coefficients: $A_{\text{XSTE}} = 3.52 \pm 0.76$ (Equation 5.3); $A_{\text{SORDID}} = 4.96 \pm 0.69$ (Equation 5.4).

[NiDO2A] / mM	$R_1 (^{15}\text{N}) / \text{s}^{-1}$			$R_1 (^1\text{H}) / \text{s}^{-1}$			
	observed	XSTE fit	SORDID fit	observed hard	observed amide- selective	XSTE fit	SORDID fit
0	2.57 ± 0.29	1.59 ± 0.17	1.73 ± 0.07	1.28 ± 0.09	4.43 ± 0.24	2.49 ± 0.51	3.03 ± 0.15
20	2.41 ± 0.30	1.44 ± 0.17	1.94 ± 0.10	3.70 ± 0.07	6.85 ± 0.22	7.82 ± 2.83	5.75 ± 0.32
40	2.40 ± 0.31	1.62 ± 0.14	1.77 ± 0.14	5.77 ± 0.09	8.90 ± 0.24	8.57 ± 5.17	8.44 ± 0.51
60	2.33 ± 0.30	1.79 ± 0.12	1.68 ± 0.18	7.79 ± 0.08	10.89 ± 0.26	10.40 ± 7.74	11.04 ± 0.71
80	2.32 ± 0.31	2.00 ± 0.16	1.50 ± 0.23	9.74 ± 0.14	13.00 ± 0.33	14.20 ± 8.14	14.11 ± 1.00

α -Synuclein. Scaling coefficients: $A_{\text{XSTE}} = 1.54 \pm 0.27$ (Equation 5.3); $A_{\text{SORDID}} = 3.55 \pm 0.18$ (Equation 5.4).

Table 2.1 Longitudinal relaxation rates of FLN5 (upper) and α -synuclein (lower) as a function of NiDO2A concentration, from direct measurements using inversion-recovery experiments with non-selective or band-selective pulses, and determined from fitting of XSTE and SORDID experimental data.

[NiDO2A] / mM	$R_2 (^{15}\text{N}) / \text{s}^{-1}$			$R_2 (^1\text{H}) / \text{s}^{-1}$		
	α -synuclein	FLN5	70S ribosomes	α -synuclein	FLN5	70S ribosomes
0	3.62 ± 0.11	10.19 ± 0.19	6.92 ± 2.06	11.01 ± 0.16	24.61 ± 2.55	32.24 ± 6.24
20	3.74 ± 0.37	10.27 ± 0.20	-	14.91 ± 0.44	27.46 ± 1.30	-
40	3.81 ± 0.07	10.45 ± 0.11	8.03 ± 1.77	18.43 ± 1.51	29.16 ± 1.99	45.71 ± 5.86
60	3.89 ± 0.05	10.59 ± 0.41	-	21.43 ± 0.53	34.14 ± 2.20	-
80	3.94 ± 0.05	-	-	23.91 ± 0.66	-	-

Table 2.2 Measured ^1H and ^{15}N transverse relaxation rates of α -synuclein, FLN5, and *E. coli* 70S ribosomes as a function of NiDO2A concentration.

The sensitivity was increased by between 1.0 and 1.9-fold, depending on the diffusion delay Δ , when using SORDID instead of the XSTE experiment, and in the absence of NiDO2A (Figure 2.2c). The most significant enhancements were observed for the longer diffusion delays, and this may be attributed to the slow ^1H longitudinal relaxation rate observed for FLN5 (Table 2.1). This therefore favours the long recovery time of 1 s

within XSTE experiments, but not in SORDID experiments where recovery times are shorter than (and coupled to) the diffusion delay. As sensitivity is limited in experiments with long diffusion delays mainly by slow R_1^H , the use of PLRE agents is clearly advantageous. We observed sensitivity increases of ~50% for both XSTE and SORDID experiments in the presence of 20–60 mM NiDO2A.

Our fitted and observed rates (Table 2.1) showed that 1H longitudinal relaxation accelerates with increasing NiDO2A concentrations. However, as the XSTE employs a fixed recovery delay, then once a sufficiently fast R_1^H is achieved (i.e. where magnetization completely returns back to equilibrium within the recovery delay) then the signal is attenuated by fast transverse relaxation rates instead. This is reflected in the large uncertainties of the fitted R_1^H rates for XSTE (Table 2.1). On the other hand, those fitted to SORDID measurements were found to be lower systematically than those measured experimentally following a soft inversion pulse. This observation may reflect the several additional hard pulses in the SORDID pulse sequence during longitudinal relaxation recovery, so that the overall excitation of spins is expected to be less selective than just the single selective inversion pulse used to measure R_1^H directly. Overall, by combining the use of SORDID and 60 mM NiDO2A for FLN5, and optimizing Δ , increases of ~1.7-fold in the sensitivity of NMR diffusion experiments were found, compared with the conventional XSTE experiment.

We investigated the effect of protein structure on PLRE-sensitivity enhancements by repeating the above experiments with the IDP α -synuclein. We found that SOFAST-HMQC sensitivity enhancements of 1.5-fold were achieved for α -synuclein (in 80 mM NiDO2A) (Figure 2.2e), with mean residue-specific increases of 1.36 ± 0.40 in 40 mM NiDO2A. These data indicate that the magnitude of the solvent PRE effect is slightly greater to that observed for the globular FLN5 (Figure 2.2b), which we suggest is due to the greater solvent accessibility of residues in the IDP, enabling a closer approach of the paramagnetic agent (Equations 2.3-2.4)

When 20 mM NiDO2A was added to NMR samples of α -synuclei, a small increase of 10% in XSTE sensitivity was observed. However, at higher NiDO2A concentrations this improvement was found to gradually decrease (Figure 2.2f). We attributed this to the unnecessarily long recovery time of 1 s used in XSTE experiments, which, given that faster 1H longitudinal relaxation rates are observed for α -synuclein (Table 2.1), results in sensitivity reductions due to accelerated transverse relaxation. In contrast, in the absence of NiDO2A, greater sensitivity of 2.6-fold was observed using SORDID compared with XSTE, and this increased to 4.8-fold at optimal diffusion delays when 80 mM PLRE agent was added. Once again the shorter recovery times in SORDID experiments, which

are coupled with the diffusion delay, are highly advantageous, so faster R_1^H induced by greater NiDO2A concentrations require shorter T_{rec} , and so result in higher sensitivity. Although the selectivity of excitation is reduced by the additional hard pulses in the SORDID experiment during R_1^H recovery (as previously discussed), highly labile, solvent-exposed amide protons enable additional gains in 1H longitudinal relaxation, and thus sensitivity, due to increased amide/water proton chemical exchange. The combined effect of these two opposing effects therefore results in fitted R_1^H rates that agree well with those measured by a non-selective pulse for globular FLN5, but faster fitted R_1^H rates than those observed with a hard pulse for disordered α -synuclein. Instead, this results in excellent agreement with those measured directly using selective inversion pulses for α -synuclein (Table 2.1). Such effects are also reflected in the fitted global scaling factors for the XSTE and SORDID experiments, where good agreement was found for FLN5 between these two scaling values, but a larger scaling factor (and therefore additional sensitivity) was found for SORDID than XSTE experiments of α -synuclein.

2.2.2 PLRE studies on 70S ribosomes

Because such large improvements in the sensitivity of NMR diffusion measurements were readily achievable for isolated proteins using NiDO2A, we next explored its use in experiments of macromolecular complexes where increases in sensitivity are a more essential requirement because of limits on time, concentration and molecular size (see Introduction). We selected the 70S ribosome as a test case because despite its large molecular weight of ~ 2.4 -MDa, sharp resonances can still be observed for the flexible bL12 stalk complex (Mulder et al. 2004; Christodoulou et al. 2004). We chose to study the magnitude of the PLRE effect using a NiDO2A concentration of 40 mM as, at this concentration, we observed the optimum sensitivity enhancement for both diffusion and 2D NMR measurements of the globular FLN5 protein (and thus the optimal balance of increase in longitudinal and transverse relaxation rates).

We obtained well-resolved resonances deriving mainly from the bL12 ribosomal stalk region by recording 1H - ^{15}N SOFAST-HMQC spectra of purified ribosomes (Figure 2.3a) (Christodoulou et al. 2004). On addition of 40 mM NiDO2A, no chemical shift changes were observed, which demonstrated that no significant interaction of the ribosome with NiDO2A was observed or measurable. However, we did observe an increase in the SOFAST-HMQC intensity of the 1D amide envelope of 20% (Figure 2.3c), and in 2D spectra, we found residue-specific sensitivity increases of 1.44 ± 0.25 (s.d.), which therefore enabled us to better discern some very weak resonances (Figure 2.3b, c, d). While some additional line broadening was induced by acceleration of transverse

relaxation on addition of 40 mM NiDO2A (Table 2.1), the resonances observed in the 2D spectra generally remained very well resolved.

We investigated the sensitivity of XSTE and SORDID diffusion measurements in a similar manner as for isolated α -synuclein and FLN5 proteins, by measuring the improvements in sensitivity across different diffusion delays in the absence and presence of 40 mM NiDO2A (Figure 2.3f). The combined use of SORDID and NiDO2A provided a 2-fold increase in sensitivity. This was significantly greater than found by either method alone, which showed sensitivity gains of approximately 10% and 20% respectively. Upon fitting of the diffusion data, the directly measured relaxation rates agreed well with those that were fitted (Tables 2.2-2.3). Crucially, no differences in the integrity or stability of ribosomes arose from the addition of NiDO2A, as determined by measurements of diffusion coefficient of the ribosomes over time. Indeed the sample lifetimes observed were between 29-47 h (Figure 2.3g), typical of ribosome preparations.

[NiDO2A] /mM	$R_1(^{15}\text{N}) / \text{s}^{-1}$			$R_1(^1\text{H}) / \text{s}^{-1}$			
	observed	XSTE fit	SORDID fit	observed non-select.	observed SORDID- select.	XSTE fit	SORDID fit
0	2.69 ± 0.44	1.87 ± 0.16	3.55 ± 0.22	1.09 ± 0.39	11.02 ± 3.32	1.63 ± 0.26	1.82 ± 0.45
40	2.40 ± 0.71	1.99 ± 0.13	3.78 ± 0.30	1.70 ± 0.39	13.34 ± 3.88	1.77 ± 0.31	2.92 ± 0.76

Scaling coefficients: $A_{\text{XSTE}} = 2.48 \pm 0.24$ (Eq. 8); $A_{\text{SORDID}} = 4.78 \pm 0.72$ (Eq. 9).

Table 2.3 Longitudinal relaxation rates of *E. coli* 70S ribosomes as a function of NiDO2A concentration, from direct measurements using inversion-recovery experiments with non-selective or band-selective pulses, and determined from fitting of XSTE and SORDID experimental data.

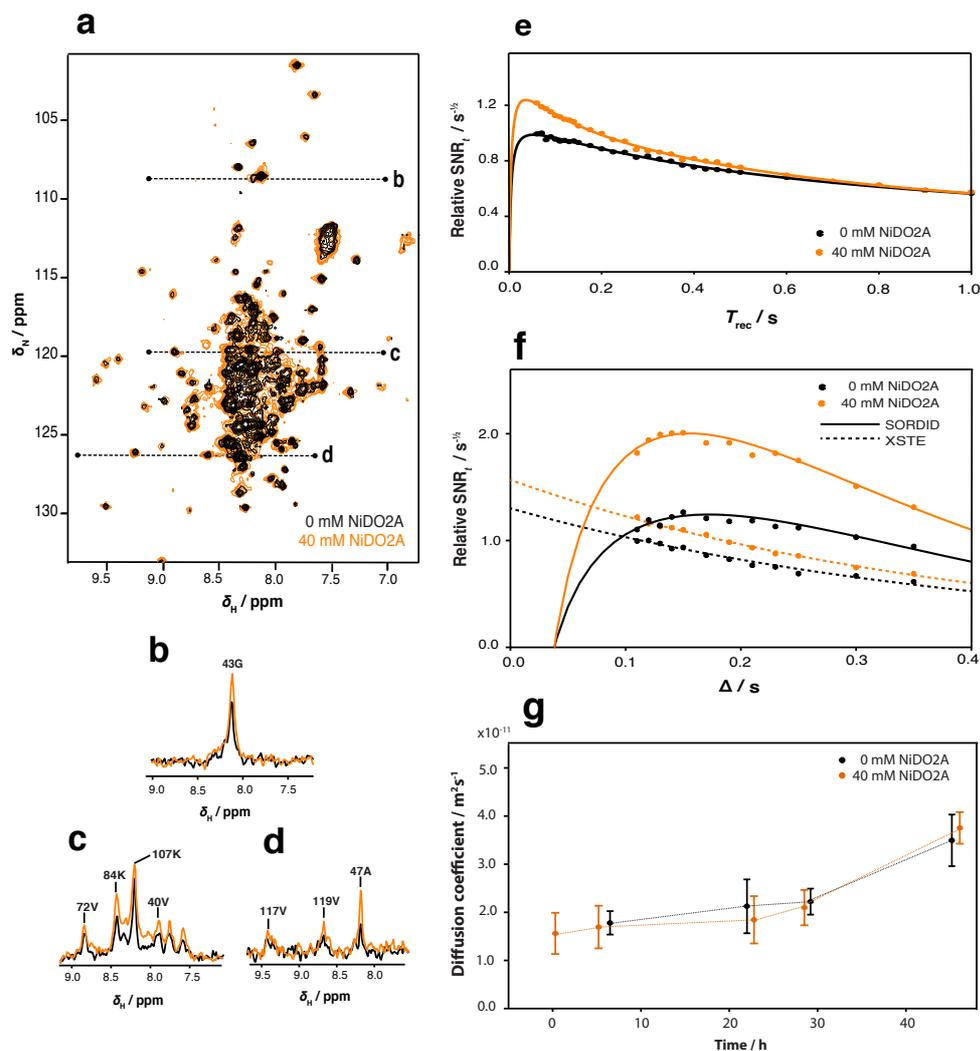


Figure 2.3 The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of NMR measurements of *E. coli* 70S ribosomes. (a) Overlay of 2D ^1H - ^{15}N SOFAST-HMQC spectra in the presence and absence of 40 mM NiDO2A PLRE agent. Dashed lines indicate locations of one-dimensional slices as shown in (b, c, d) and labeled with known assignments from the bL12 stalk complex. The relative sensitivity of (e) 1D ^1H - ^{15}N SOFAST-HMQC experiments as a function of the recovery delay T_{rec} , and of (f) XSTE and SORDID diffusion experiments as a function of the diffusion delay Δ . Data are globally fitted to their theoretical sensitivity expressions (Equations 6.3-6.4). To assess the integrity of 70S ribosomes during NMR acquisition, and the effect of NiDO2A on the sample stability, its diffusion coefficient was continually monitored (g) using XSTE and SORDID measurements in the presence and absence of 40 mM NiDO2A.

2.2.3 PLRE studies on ribosome-nascent chain complexes

We were motivated to explore the benefits of paramagnetic longitudinal relaxation-induced sensitivity gains as applied to RNCs because, as previously discussed, there are significant challenges in their study by NMR spectroscopy, due primarily to low concentrations and short lifetimes of purified RNC samples. The experiments were extended to a ribosome-bound nascent chain of FLN5, linked with a 31-residue sequence derived from FLN6 and the SecM motif to cause translational stalling of the ribosome, termed FLN5+31 RNC, and in which the selectively, uniformly ^{15}N -labelled nascent chain adopts a disordered conformation (Cabrita et al. 2016). One-dimensional ^1H - ^{15}N SOFAST-HMQC correlation spectra (Figure 2.4a-b) were acquired, and an enhancement in the amide proton signals of $\sim 85\%$ was observed when 40 mM NiDO2A was added to the sample (Figure 2.4a), an increase in sensitivity comparable to that obtained for α -synuclein (Figure 2.2), which is consistent with the disordered nature of the nascent chain. As a result of such large sensitivity enhancements, a greater number of well-resolved SOFAST-HMQC resonances become observable within the same experimental time of 30 minutes. At least three of these cross-peaks could be unambiguously assigned to the FLN5 nascent chain, while additional resonances were attributed to background labeling of the ribosome (specifically the bL12 stalk region) (Figure 2.4b). Importantly, nascent chain resonances did not show any changes in the chemical shift on addition of 40 mM NiDO2A, and analysis of one-dimensional slices showed that the signal enhancements far outweigh the smaller increases in line broadening (Figure 2.4c, d, e).

XSTE and SORDID diffusion spectra of the RNC were acquired using identical experimental times of 30 min (Figure 2.5a). Longitudinal relaxation optimization, achieved by using SORDID or introducing 40 mM NiDO2A into the sample, resulted in 2- and 3.1-fold enhancements in sensitivity respectively, compared to XSTE experiments in the absence of PRE agent (Figure 2.5b). However, even greater sensitivity gains were achieved when both methods were combined, and 4.5-fold enhancement in sensitivity was observed. Importantly, all four measurements yielded identical diffusion coefficients (Figure 2.5c), corresponding to a diffusion coefficient of $(2.22 \pm 0.13) \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ and therefore indicating that the observed resonances were indeed ribosome associated.

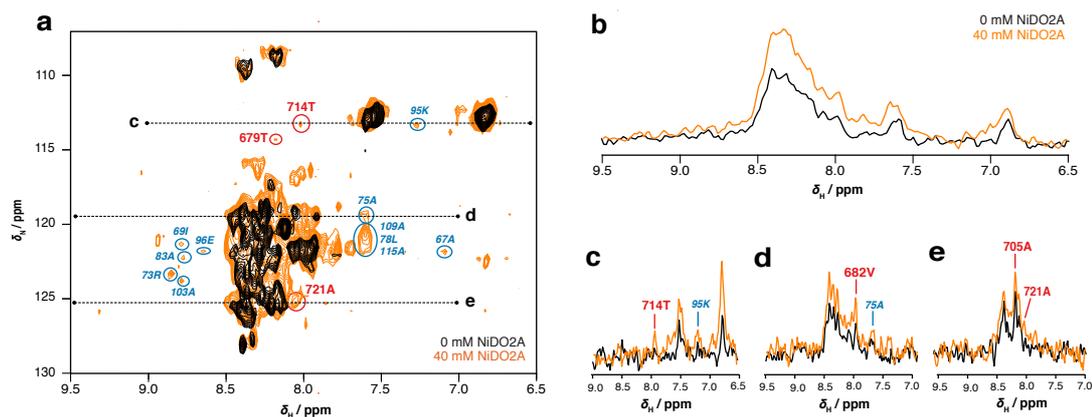


Figure 2.4 The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of SOFAST-HMQC experiments of FLN5+31 RNC. Overlay of (a) 2D ^1H - ^{15}N SOFAST-HMQC spectra (of 30 min total experimental time each) and of (b) 1D ^1H - ^{15}N SOFAST-HMQC spectra in the presence and absence of 40 mM NiDO2A. Additional resonances are observable in the PLRE-induced spectrum and all can be assigned to the nascent chain or the bL12 stalk region of background-labeled ribosomes (see main text). Those that can be assigned unambiguously to either region are circled in red or blue respectively, and labeled with their known assignments. Dashed lines indicate locations of one-dimensional slices as shown in (c, d, e) and labeled with known and unambiguous assignments of disordered FLN5 or bL12.

We assessed the integrity of the RNC sample both before and after the addition of NiDO2A, and then also over the total experimental time, by repeatedly acquiring sets of XSTE and SORDID experiments in an interleaved manner (Figure 2.5d). These experiments both showed a constant diffusion coefficient over the first approximately five hours of data acquisition (during which the previously discussed experiments were completed), before the diffusion coefficient measured increased, after approximately seven hours, which is indicative degradation of the sample by release of the FLN5 nascent chain from the ribosome (resulting in an effectively smaller species). These NMR diffusion measurements were complemented by western blot analysis of identical samples in parallel with the NMR acquisition period and incubated with and without 40 mM NiDO2A. Aliquots of these samples were taken at intervals, and analysed by western blotting to assess the integrity of the nascent polypeptide. Both anti-his and anti-SecM western blots showed a reduction in band intensity of the tRNA-bound form of the NC (~40 kDa) and, at ~ 7 h, the subsequent appearance of a ~23kDa species, corresponding to the released nascent chain. We observed no difference in stability between samples incubated with and without NiDO2A (Figure 2.5e). These observations are in agreement with the time-course of SORDID and XSTE measurements. Taken together with the SOFAST-HMQC data, we therefore conclude that there is no measurable evidence of

interaction of NiDO2A with ribosomes or RNCs, nor is there any indication of any effect on the nascent chain or ribosomal particle stability, consistent with our previous experiments using isolated protein and ribosomes.

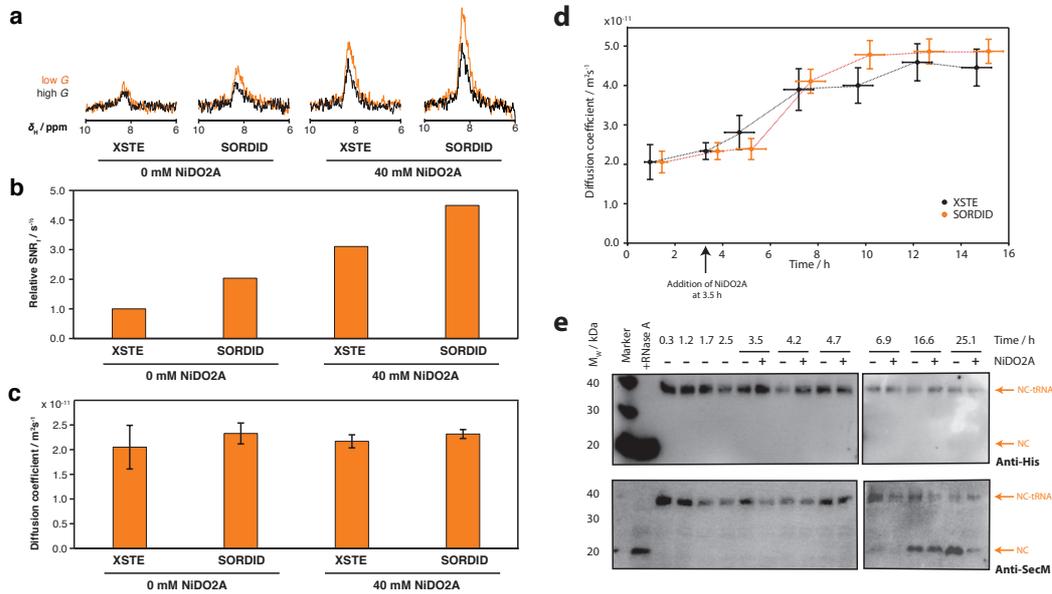


Figure 2.5 Comparison of NMR diffusion measurements of FLN5+31 RNC, and monitoring its integrity by both biochemical and NMR analysis. (a) XSTE ($\Delta = 100$ ms; $G = 5\%$, $95\% G_{\max}$) and SORDID ($\Delta = 190$ ms; $G = 10.4\%$, $69.5\% G_{\max}$) spectra acquired within 30 min each, in the presence and absence of 40 mM NiDO2A. (b) Relative sensitivity for each diffusion experiment. (c) Diffusion coefficients measured using each diffusion experiment. (d) Diffusion coefficients of the RNC sample measured using sets of XSTE and SORDID experiments to examine the integrity and stability of the RNC. (e) Western blot analysis to assess the effect of 40 mM NiDO2A on the RNC integrity, detected using anti-his (upper) and anti-SecM (lower) antibodies. Upshifted bands (~ 40 kDa) correspond to the tRNA-bound form of the RNC in which the nascent chain is attached to the ribosome, with release of the nascent chain (and tRNA) monitored by the band at a lower molecular weight of ~ 23 kDa.

2.3 Discussion

The recent advances in the improvement of the sensitivity and resolution of NMR spectroscopy, for example, the development of TROSY sequences, advanced isotopic labelling strategies, and the greater availability of higher field spectrometers equipped with cryo-probes, have enabled the application of NMR spectroscopy to increasingly larger and more complex biological systems. The study of these large and often unstable macromolecular assemblies requires both continuing improvements in NMR

spectroscopic sensitivity with a non-invasive, rapid means to monitor the sample integrity. In this Chapter, we have explored the use of the paramagnetic longitudinal relaxation enhancement (PLRE) effect to improve the sensitivity of heteronuclear NMR diffusion measurements that are essential in probing the sample stability of ribosomal particles such as RNCs. We observed significant improvements by using the PLRE agent NiDO2A (Cai et al. 2006), particularly when combined with the longitudinal relaxation-optimised SORDID experiment (Augustyniak et al. 2012a), in which we observed up to 1.7- and 4.8-fold enhancements in sensitivity for globular and disordered proteins respectively, relative to conventional XSTE experiments in the absence of NiDO2A (Ferrage et al. 2003), without adverse effect on the storage of magnetization on N_z during the diffusion period and with only negligible line broadening of resonances. Similarly, for FLN5+31, in which the nascent chain adopts a disordered conformation, a 4.5-fold increase in NMR diffusion sensitivity was achieved. These sensitivity gains are in addition to improvements in 2D SOFAST-HMQC sensitivity of 20-85%, as previously reported for α -synuclein (Theillet et al. 2011).

The PLRE effect occurs to different extents both between globular and disordered proteins and throughout protein sequences. This may be rationalized by a combination of two effects. Firstly, the greater solvent accessibility of amides in disordered and less structured regions increases the distance between the PLRE agent and nuclei, increasing the strength of dipolar interactions (distance of closest approach in Equations 2.3-2.4). Secondly, the PLRE agent will also accelerate the longitudinal relaxation of water protons, and their exchange with amide protons in IDPs will therefore also lead to an increase in the effective R_1^H rates (Cai et al. 2006; Gil et al. 2013).

So far we have focused our discussion on the generally favourable effects of PLRE. However, increases in longitudinal relaxation rates by PLRE are inevitably also associated with the acceleration in transverse relaxation (Equations 2.3-2.4) and while choices of fast relaxing paramagnetic species such as Ni(II) or Fe(III) can mitigate this effect significantly (see section 2.1.3.1) some additional line broadening is unavoidable. For instance, the two-fold increase in proton line widths resulted from the doubling of the average proton R_2 rates observed for α -synuclein in 80 mM NiDO2A. However, the more than four-fold increases in both longitudinal relaxation rates and sensitivity enhancements considerably outweigh these effects, and indeed SOFAST-HMQC spectra acquired in the presence of PLRE agent continued to show well-resolved resonances with few, if any, additional overlapped peaks. Clearly, the optimum concentration of PLRE agent must be judged on a case-by-case basis for each protein, based on the acceptable compromise between spectroscopic resolution and sensitivity. We note that the addition

of soluble PLRE agents also increases the line width of water ^1H nuclei, which can reduce the effectiveness of water suppression, but found that this can be largely alleviated by the introduction of phase cycling to select heteronuclear coherence transfer pathways within the SORDID experiment (Figure 2.1), and by processing spectra using standard solvent suppression filters and baseline correction post-acquisition (see Materials and Methods).

In this Chapter, we have shown that the gains in sensitivity obtained by paramagnetic longitudinal relaxation enhancement are particularly advantageous to time-limited investigations of large, complex macromolecular assemblies such as RNCs. The additional sensitivity can therefore be exploited to reduce the uncertainty in diffusion coefficients measured using NMR diffusion experiments of the same experimental time (Figure 4c). However, given the short life time of a stable RNC, the additional sensitivity can instead be used to substantially decrease the required measurement time and therefore more rapidly assess the sample integrity. Given the 4.5-fold increase in sensitivity we observe here, the same quality of spectra acquired by XSTE experiments in 30 min may be obtained in less than 1.5 min by combined use of NiDO2A and SORDID. As sample integrity and lifetimes are not compromised on addition of the PLRE agent, the additional available time may therefore be used for longer acquisition of other, more useful NMR experiments. This enables higher resolution spectra to be obtained, as previously demonstrated (Theillet et al. 2011), or can be used to acquire spectroscopically more complex experiments, as discussed in the following Chapter. Alternatively, longer experimental times can be used to enhance the sensitivity of existing experiments. For instance, in NMR experiments of RNC samples, we have typically acquired 2D and diffusion measurements with equal time periods. However, by increasing the time allocated to 2D measurements (instead of NMR diffusion measurements which are essentially control experiments), we estimate the net sensitivity increase when combined with PLRE-induced gains to be approximately 2.6-fold. We expect this will greatly facilitate future measurements of RNCs and similar challenging biological systems.

An extension of the effect of PRE agents on the sensitivity of ^{13}C -SORDID or ^{13}C -XSTE to study diffusion of ^{13}C -labeled samples has not yet been investigated, and its potential benefits remain unclear at this stage. This is in part due to the larger $^1\text{J}_{\text{CH}}$ coupling constant that exists and in which the INEPT transfer period is significantly reduced from ^{15}N experiment, thus limiting the maximum gradient pulse length that may be applied. As a result, longer diffusion delays are therefore needed to determine slow-diffusing molecules, therefore reducing overall sensitivity because of increased ^{13}C longitudinal relaxation, together with undesirable PRE-induced increases in R_1^{C} rates. In

fact, increases in the latter will be greater than those of observed for ^{15}N , because of the larger gyromagnetic ratio of ^{13}C . Whether these effects are outweighed by increases in ^1H longitudinal relaxation by PRE remains uncertain. An alternative is the use of the ^1H -STE- $^1\text{H},^{13}\text{C}$ -HMQC (Didenko et al. 2011), which may offer greater sensitivity as it is optimized for the methyl-TROSY effect in which mixing of slow and fast relaxing coherences is minimized. However, the implications of accelerated ^1H longitudinal relaxation to this process on the experimental sensitivity will need further investigation. Nevertheless, with future prospects in alternative isotopic-labeling strategies, such as specific methyl labeling (Goto et al. 1999) and fractional or perdeuteration (Sattler and Fesik 1996) to further increase sensitivity of NMR measurements of RNCs and other macromolecular assemblies, the development of potential increases in ^{13}C diffusion experiments are highly attractive.

Chapter 3

Structural investigations of ribosome-nascent chain complexes by NMR spectroscopy

3.1 Introduction

While detailed descriptions of the ribosome structure are emerging in parallel with structural studies of protein folding inside the ribosomal exit tunnel, comparatively little is known about co-translational protein folding when the nascent chain has emerged from the exit vestibule and where more extensive and complex structure acquisition is thought to occur. The inherently dynamic nascent chain is highly amenable to their study by NMR spectroscopy, although such experiments are restricted by the intrinsically low sensitivity. In Chapter 3, we have shown how PLRE effects can be exploited to increase the sensitivity of 2D correlation and diffusion NMR experiments of RNCs, and in this Chapter, we discuss how this strategy can be extended to extract more detailed, structural information on the nascent chain using 3D NMR spectroscopy and the measurement of residual dipolar couplings (RDCs), which have eluded most large biological systems so far. We believe these approaches will enable unprecedented insights into the structure of co-translational folding intermediates and how the ribosome modulates the dynamics and structure of nascent chains.

3.1.1 Co-translational folding of an immunoglobulin domain

Recent NMR studies of a pair of immunoglobulin-like domains (FLN5-6) on the ribosome have provided detailed, structural descriptions of co-translational protein folding outside of the exit tunnel (Cabrita et al. 2016). In this study, a series of SecM-translationally stalled, ribosome-bound FLN5 nascent chains were produced in which its FLN6 linker was progressively truncated at the C-terminus to mimic co-translational intermediates. A dual isotopic labelling strategy, through selective ^{13}C -methyl labelling of isoleucine residues and uniform ^{15}N -labelling, enabled the observation of an increasing population in folded nascent chains and a concomitant reduction in disordered states, respectively, as FLN5-6 emerged from the ribosome (further details on NMR labelling strategies are described in section 4.3.3).

Analysis of NMR signal intensities revealed a midpoint of transition from disordered to folded FLN5 when the domain was between 42-45 residues from the peptidyl transferase centre. This was in contrast to a pegylation assay that showed complete emergence of FLN5 from the ribosomal tunnel at 31 residues from the PTC, and also a parallel study of FLN5 in isolation from the ribosome, which found only native-like states were populated even when 4 residues were truncated from its C-terminus. These data suggested that the RNC of FLN5 linked to the PTC by 31 residues (termed RNC FLN5+31) would in principle be able to acquire native structure, but NMR data suggested an additional 11-14 residues were required when tethered to the ribosome.

Replacement of FLN6 residues with a poly(glycine-serine) linker showed that FLN5 persisted in a disordered conformation, suggesting that FLN5 folds independently and that inter-domain interactions are not the predominant contributor to the effects observed. The combined experimental evidence therefore showed that the ribosome modulates the folding energy landscape of FLN5 nascent chains.

Examination of residue-specific 2D ^1H - ^{15}N SOFAST-HMQC spectra of the RNCs found substantial broadening within two stretches of residues in particular, F665 to V667 and G713 to G750 (Figure 3.1a-c), that was attributed to the binding of disordered FLN5 to the ribosome. Indeed, in the same study, MD simulations using replica-averaged chemical shift restraints showed an ensemble of FLN5-6 RNCs in which both folded and disordered FLN nascent chains interacted with the ribosomal surface, which were likely to be analogous to the experimental NMR studies. Whereas the folded domain made substantial contacts with the 23S rRNA and uL29 protein, the most frequent interactions made by disordered FLN6 occurred with the loop of uL24 near the exit vestibule. The increased nascent chain interactions (relative to isolated protein) with the ribosome were attributed to a combination of higher effective local ribosomal concentration, steric occlusion effects, and inherent sequence determinants within FLN5-6 (further discussed below).

The nascent chain interactions with the ribosomal surface, also observed more generally in cryo-EM and biochemical studies, therefore demonstrate that the ribosome can perturb the structural and dynamic processes of complex secondary and particularly tertiary structure acquisition as the nascent chain emerges vectorially from the tunnel. In the case of FLN RNCs, the ribosome serves to stabilise the unfolded state of FLN5, reducing the possibility of misfolding and aggregation processes, and has therefore been suggested to be a common mechanism of multi-domain proteins with high sequence identities among domains.

These studies raised several interesting questions into how the ribosome influences the *de novo* folding of FLN5 and this prompted further study. Firstly, the likely sequence determinants of FLN5-6 of ribosomal interactions were investigated. As the poly(GS) linker did not significantly influence the folding of FLN5 (Cabrita et al. 2016) (described above), the linking residues between FLN5 and the PTC were likely to be independent of the folding process, and we therefore studied the amino acid sequence of FLN5 in more detail. The two clusters of residues with significantly broadened NMR resonances in the RNC spectra (Figure 3.1a-c) seemed to be associated with aromatic residues (F665, F672 and F675 in one cluster, Y715, Y719 and Y727 in the other, more C-terminal cluster); indeed, patches of aromatic side chains seem to be common nascent chain interaction sites with ribosomes as similar observations have been made in NMR investigations of ribosome-bound α -synuclein (Deckert et al. 2016).

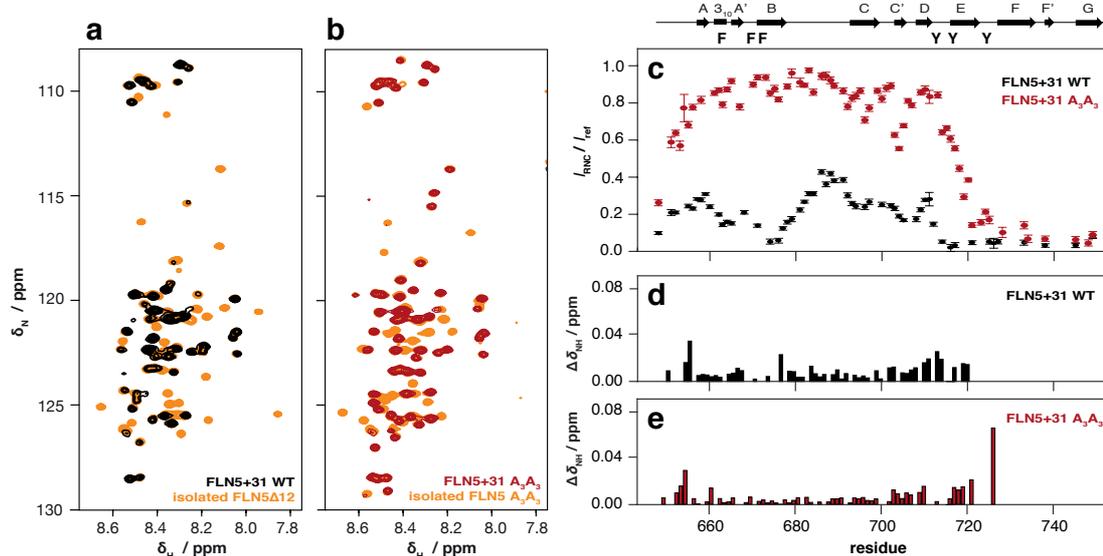


Figure 3.1 2D NMR spectroscopy of FLN5+31 WT and A_3A_3 RNC mutant. 2D 1H - ^{15}N SOFAST-HMQC spectra of (a) FLN5+31 WT RNC (black) overlaid with isolated FLN5 Δ 12, a disordered FLN5 variant (orange); and (b) FLN5+31 A_3A_3 RNC (red) overlaid with isolated FLN5 A_3A_3 (orange). (c) Relative SOFAST-HMQC signals of FLN5+31 WT (black) and A_3A_3 (red) RNC compared with an isolated reference protein, with schematic above showing secondary structure elements of full-length FLN5 and positions of phenylalanine and tyrosine to alanine mutations. Errors derive from spectral noise. (d) Amide chemical shift changes of signals of FLN5+31 WT (black) and A_3A_3 (red) RNC compared with an isolated reference protein. Data recorded and analysed by A.M.E. Cassaignau (in preparation).

When these six aromatic residues were each mutated to alanine residues (A.M.E. Cassaignau, *et al.*, in preparation), resulting in the FLN5+31 A_3A_3 RNC, 1H - ^{15}N SOFAST-HMQC signal intensities in the N-terminus recovered to close to those of the isolated FLN5 protein (Figure 3.1b-c). This indicated a loss of interaction and binding of

disordered FLN5 to the slowly tumbling ribosome (relative to WT RNC). Moreover, the resonances progressively broadened toward the C-terminus, as observed for WT RNC (Cabrita et al. 2016), and this was attributed, in part, to the reduction in mobility of the nascent chain due to its tethering to the ribosome, and also to persistent nascent chain interactions with the ribosome near the exit vestibule, or local changes in structure. In both FLN5+31 A₃A₃ and WT RNCs, the reduction in SOFAST-HMQC intensities was accompanied by small but systematic chemical shift changes, anti-correlated with signal intensities, and most prominently toward the C-terminus (Figure 3.1d-e). This data indicates that these regions were in a different chemical environment and were therefore structurally different than when isolated from the ribosome.

Additionally, as previously discussed, the disordered FLN5+31 RNC is in principle able to adopt a folded conformation as the complete domain was found to be outside of the exit tunnel. Because FLN5 is able to populate native-like states even when truncated when in isolation from the ribosome, this raises the possibility of some intra-domain, native-like contacts to occur within the domain, in addition to non-native contacts between the nascent chain C-terminal linker and the ribosomal surface. Indeed hints of such structural changes are corroborated with observed chemical shift changes (discussed above).

3.1.2 Structural studies of RNCs by NMR spectroscopy

Because of the possibility of some native and non-native-like structure acquisition to occur on the ribosome, in conformations distinct to those found in isolation, we were motivated to gain further chemical shift information on the nascent chain as these are the most readily and accurately measurable NMR parameters. We were particularly interested in measuring changes in ¹³C shifts (relative to isolated protein), as these are sensitive to secondary structure acquisition (detailed in section 3.2.3), and also in measuring residual dipolar couplings (RDCs) as a complementary approach to gain structural information. The substantial reduction in experimental sensitivity of higher order, multi-dimensional NMR and other experiments of greater complexity have so far excluded their application to RNCs. However, the development of PLRE methods to increase the sensitivity of NMR experiments of RNC without changes in chemical shifts (Chapter 2) presented an opportunity to explore such inherently insensitive, but information-rich experiments. In addition, because of increased mobility due to reduced interaction of FLN5+31 A₃A₃ RNC with the ribosome, we were interested in exploiting its increased NMR sensitivity in structural studies. In this Chapter, we describe the development of a strategy to obtain, at high resolution and with enhanced sensitivity, 3D

NMR experiments and RDC measurements. This will enable detailed structural descriptions of the emerging nascent chain from the ribosome, contributing to the currently limited understanding of co-translational protein folding outside of the exit tunnel. We discuss the theoretical considerations of this study below.

3.2 Theory and applications of 3D NMR

3.2.1 Triple-resonance NMR experiments

NMR spectra contain a rich source of information about electronic shielding and thus the chemical environment of individual nuclei. All this information can in principle be extracted from 1D experiments, although the complexity and particularly size of the biomolecules typically studied renders this practically impossible. Resolving the spectral information across two, three or even more orthogonal dimensions can minimize spectral crowding of often broad, weak signals for large molecules (due slow tumbling and therefore fast T_2 relaxation), so that more detailed interpretations become accessible. While homonuclear multi-dimensional experiments such as homonuclear correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) have remedied such problems for smaller proteins (less than 100 residues), resonance degeneracy and cross-peak overlaps that remain for larger systems can only be circumvented through heteronuclear 3D (or higher dimension) experiments (Bax 2011).

By selectively isotopically labelling proteins with ^{13}C and ^{15}N , the scalar couplings $^1J_{\text{NH}}$, $^1J_{\text{CH}}$, $^1J_{\text{CC}}$ and $^1J_{\text{CN}}$ can be used to establish connectivity between pairs of nuclei. The relatively large size and uniformity of these scalar couplings and the nuclei's arrangements on the amide backbone, enable the design of multiple, efficient one-bond transfers of magnetization, rather than by a single step through the smaller, more variable conformation-dependent 3J couplings used in TOCSY and COSY experiments (Table 3.1). By initiating magnetization on ^1H and then successively evolving magnetization on each dimension before final detection on the initial nucleus ('out-and-back'), crucial sequential connectivity information can be obtained, in what are termed triple resonance experiments (Kay et al. 1990; Bax 2011).

There is now a whole array of different triple-resonance experiments (Table 3.1), differing on the magnetization transfer pathways and indirect evolutions such that chemical and structural information can be extracted from a combination of different nuclei along both the amide backbone and residue side-chains, specifically $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$, $^1\text{H}^{\alpha}$, $^{13}\text{C}^{\alpha}$, $^1\text{H}^{\beta}$, and $^{13}\text{C}^{\beta}$. Using different combinations of triple-resonance experiments in tandem has therefore become the ubiquitous strategy for residue assignments of

resonances for large (> 10 kDa) proteins (Ikura et al. 1990). For example, the HN(CA)CO experiment (Clubb et al. 1992) correlates the amide ^1H and ^{15}N resonances of one residue with its $^{13}\text{C}'$ chemical shift but also that of the preceding residue (Table 3.1); the HNC0 experiment (Kay et al. 1990) is therefore used in parallel as only the $^{13}\text{C}'$ chemical shift of the preceding residue is observed (Table 3.1), and this enables sequential assignment of NMR resonances. Additional pairs of experiments are often implemented in parallel to avoid ambiguities in the spectra.

3.2.2 Sensitivity of 3D NMR experiments

In addition to exploiting the larger one-bond rather than multiple-bond J -couplings to transfer magnetization between nuclei (section 3.2.1), the sensitivity of the triple-resonance experiments described above are optimised through the use of 'out-and-back' transfers of magnetization to and from ^1H nuclei, which have a large gyromagnetic ratio. These approaches compensate, to some extent, the losses of signal due to inherent inefficiencies (and the many number) of magnetization transfer steps, the continuing signal decay resulting from transverse relaxation, and the reduction of $\sqrt{2}$ in sensitivity with each spectral dimension (as both the imaginary and real components of the signal must be sampled). Where differences in J -couplings are relatively small, for example between $^1J_{\text{C}\alpha}$ and $^2J_{\text{C}\alpha}$, these manifest in multiple, inter-residue coherence pathways, contributing further to reductions in the overall sensitivity of each observed coherence (Clore and Gronenborn 1994). The HNC0 experiment is therefore the most sensitive triple-resonance experiment; the relative theoretical sensitivities of other pulse sequences are summarized in Table 3.1.

Experiment	Correlations observed (relative sensitivity*)	Magnetisation transfer pathway	J couplings	References
HNCO	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}$ (100%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}}$	(Kay et al. 1990)
HN(CA)CO	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_i$ (13%) ${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}$ (4%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}^\alpha}$ ${}^1J_{\text{C}^\alpha\text{C}^\alpha}$ ${}^2J_{\text{NC}^\alpha}$	(Clubb et al. 1992)
HNCA	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_i^\alpha$ (50%) ${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}^\alpha$ (15%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}^\alpha}$ ${}^2J_{\text{NC}^\alpha}$	(Kay et al. 1990)
HN(CO)CA	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}^\alpha$ (71%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}}$ ${}^1J_{\text{C}^\alpha\text{C}^\alpha}$	(Grzesiek and Bax 1992a)
HNCACB	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_i^\alpha - {}^{13}\text{C}_i^\beta$ (α : 4%; β : 1.7%) ${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}^\alpha - {}^{13}\text{C}_{i-1}^\beta$ (α : 1.3%; β : 0.5%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}^\alpha}$ ${}^1J_{\text{C}^\alpha\text{C}^\beta}$ ${}^2J_{\text{NC}^\alpha}$	(Wittekind and Mueller 1993)
HN(CO)CACB	${}^1\text{H}_i^{\text{N}} - {}^{15}\text{N}_i - {}^{13}\text{C}_{i-1}^\alpha - {}^{13}\text{C}_{i-1}^\beta$ (α : 13%; β : 9%)		${}^1J_{\text{NH}}$ ${}^1J_{\text{NC}}$ ${}^1J_{\text{NC}^\alpha}$ ${}^1J_{\text{C}^\alpha\text{C}^\beta}$	(Grzesiek and Bax 1992b)

Table 3.1 Summary of selected triple-resonance experiments to obtain chemical information on C' , C^α and C^β . In the magnetization transfer pathways, arrows indicate the transfer of coherence between nuclei whose chemical shift is evolved (black circles) and nuclei where coherence is transferred between, but whose chemical shift is not evolved (open circles). Scalar couplings: ${}^1J_{\text{NH}} \sim 92$ Hz, ${}^1J_{\text{NC}} \sim -15$ Hz, ${}^1J_{\text{NC}^\alpha} \sim -11 - -7$ Hz, ${}^1J_{\text{C}^\alpha\text{C}^\alpha} \sim 55$ Hz, ${}^1J_{\text{C}^\alpha\text{C}^\beta} \sim 35$ Hz, ${}^2J_{\text{NC}^\alpha} \sim 4-9$ Hz. *Theoretical relative sensitivities derived by analysis of the trajectory of magnetization during the pulse sequences. Adapted from references (Sattler and Schleucher 1999; Cavanagh et al. 2010).

The different approaches to enhance NMR sensitivity discussed in Chapter 2 can also be readily applied to triple-resonance experiments. For example, the PLRE agent NiDO2A has previously been employed in HNCA experiments to accelerate longitudinal relaxation and thereby increase the experimental sensitivity by up to 86% for the IDP α -synuclein (Cai et al. 2006). While the work described here has so far only utilized the PLRE-induced sensitivity enhancements on ^{15}N -labelled samples, these previous studies have demonstrated that the larger increases in transverse relaxation rate of ^{13}C (due its larger gyromagnetic ratio relative to ^{15}N) are outweighed by the increases in ^1H longitudinal relaxation at optimised PLRE agent concentrations. Longitudinal relaxation-optimisation within 3D NMR experiments can also be achieved by using the series of band-selective excitation short transient (BEST) pulse sequences (Schanda et al. 2006). Analogous to the SOFAST-HMQC scheme (Schanda et al. 2005), in the BEST experiments, efficient longitudinal relaxation is achieved through dipolar interactions between the selectively excited amide spins and the remaining, unperturbed aliphatic protons; and this enables fast repetition rates (Schanda et al. 2006). Because PLRE effects are significantly and synergistically enhanced in combination with longitudinal relaxation-optimised experiments (Chapter 2), we were motivated to explore the use of NiDO2A with such pulse schemes, and in particular, when applied to RNCs and other large biomolecular complexes.

3.2.3 Using chemical shifts to predict secondary structure

Significant advances have recently been made in using NMR chemical shifts in protein structure determination, for example, their use as replica-averaged restraints in MD simulations to determine a structural ensemble of FLN5 RNCs (Cabrita et al. 2016). However, 3D protein structures of large macromolecular assemblies such as RNCs calculated exclusively using NMR measurements is lacking due to the insensitivity of complex 3D NMR experiments for large molecules and also the high demands of computational power.

Instead, more readily accessible information from chemical shifts can be gained without or before such involved MD simulations are performed using predictions of protein secondary structure. Indeed there is empirical evidence that backbone chemical shifts are correlated with secondary structure formation (Dalgarno et al. 1983; Wishart and Sykes 1994) characterised by the polypeptide torsion backbone angles ϕ and ψ (Spera and Bax 1991), and this has led to the use of the 'secondary chemical shift' $\Delta\delta_S$, defined as $\Delta\delta_S = \delta_{\text{obs}} - \delta_{\text{r.c.}}$, where δ_{obs} is the observed chemical shift, and $\delta_{\text{r.c.}}$ is the corresponding value expected for the same residue in a random coil peptide. The value of

the secondary chemical shift, most notably of H^α , C^α and C^β nuclei, exhibits characteristic patterns for secondary structure elements: secondary chemical shifts of C^α nuclei in α -helices are typically deshielded ($\Delta\delta_S > 0$) and are shielded ($\Delta\delta_S < 0$) when found in β -strands. The reverse is true for C^β nuclei: those in α -helices are typically shielded, and are deshielded in β -strands.

These calculations formed the basis of the chemical shift index (CSI), in which each residue within a protein sequence is assigned a value corresponding to α -helix, β -sheet or random coil conformations, based on their secondary chemical shifts (Wishart and Sykes 1994). More recent work (the $\delta 2D$ method) has improved on this technique by predicting the secondary structure populations within disordered conformations of proteins, which are highly heterogeneous and can only be represented as an ensemble with distribution of different structures (Camilloni et al. 2012).

3.3 Theory and applications of residual dipolar couplings

As a highly complementary approach to observing chemical shift changes, we explored how residual dipolar couplings within RNCs could be measured at high sensitivity. RDCs are an NMR property that can be used in the determination of protein structure and dynamics across a wide range of time-scales (Tolman and Ruan 2006). In contrast to the nuclear Overhauser effect (NOE), which scales as distance r^{-6} and so reports on interactions between distances of $< 5 \text{ \AA}$, RDCs provide angular dependent structural information, with greater spectroscopic sensitivity. They have therefore proven particularly useful as restraints in the refinement of structure determinations of proteins and determining the relative orientations of domains (Kanelis et al. 2001).

3.3.1 The physical origin of residual dipolar couplings

The physical origin of RDCs arises from direct magnetic interaction between a pair of nuclear magnetic moments, which depend on the angle between the internuclear vector and the external magnetic field, and the distance between the nuclear spins r_{AB} (Figure 3.2a). In isotropic solution, molecules behave according to Brownian motion, sampling all possible orientations with equal probability; thus, the dipolar interactions average to zero. However, in anisotropic media, small but non-zero time-averaged dipolar couplings can then be measured between pairs of nuclei, the strength of which is given by (Tolman and Ruan 2006):

$$D_{IS} = -\frac{\mu_0 h}{16\pi^3} \cdot \frac{S\gamma_I\gamma_S A_a}{r_{AB}^3} \left\{ (3 \cos^2 \theta - 1) + \frac{3}{2} A_r (\sin^2 \theta \cos 2\phi) \right\} \quad (3.1)$$

where μ_0 is the permeability in a vacuum, h is Planck's constant, S is the generalized order parameter, γ_I and γ_S are the gyromagnetic ratios of the nuclei, A_a and A_r are

components of the molecular alignment tensor, and ϕ and θ are the angles between the principle axis of the molecular alignment tensor and the external field. If the residual dipolar coupling is weak, then the dipolar coupling Hamiltonian is reduced to:

$$\hat{H}_D = 2\pi D_{\text{IS}} I_z S_z \quad (3.2)$$

This Hamiltonian has the same form as that of the scalar coupling. Therefore, under aligned conditions, the apparent scalar coupling observed experimentally is given by $(J + D)$ (Cavanagh et al. 2010).

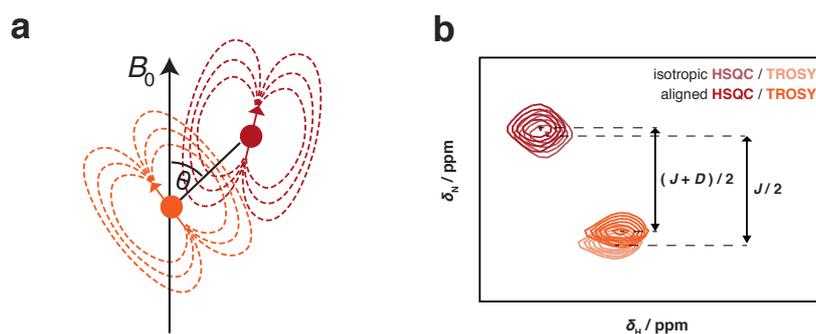


Figure 3.2 Residual dipolar couplings. (a) Schematic of the dipolar interaction between two magnetic dipoles where θ is the angle between the internuclear vector and the external magnetic field B_0 . (b) Schematic of the measurement of the dipolar coupling constant D using HSQC and TROSY experiments of isotropic and aligned samples.

3.3.2 Alignment methods

There is a range of different chemical and biochemical approaches to induce partial alignment of molecules to enable RDC measurements. For example, alignment by steric interactions can be achieved through the use of polyacrylamide gels (Sass et al. 2000), phospholipid bicelles (Tjandra 1997), or polyethylene glycol-based bilayers (Rückert and Otting 2000). Alignment can also be attained through paramagnetism of a protein by, for instance, linking the protein to a metal-chelating tag (Wöhnert et al. 2003).

Another approach is the use of filamentous bacteriophage that also introduces electrostatic interactions, in addition to steric forces, to partially orient molecules in solution (Clore and Starich 1998). Commonly used and also utilised in this study is Pf1, from *Pseudomonas aeruginosa*, which forms long rods with a negatively charged surface ($pI \sim 4.0$). The coat protein forms an α -helical structure parallel along the long axis of the phage. The repeating carbonyl groups are thought to be the source of the phage's large anisotropic magnetic susceptibility and enable the formation of liquid crystal media that induce alignment of biomolecules in solution (Hansen et al. 1998).

3.3.3 Measuring residual dipolar couplings

Dipolar coupling constants are measured indirectly using sets of data obtained under both isotropic and anisotropic conditions. Under isotropic conditions, dipolar couplings D average to zero such that only the pure scalar coupling J remains. However, when measured under anisotropic conditions, both the scalar and dipolar couplings are non-zero, enabling the calculation of the dipolar coupling constant by taking the difference between the two measurements (de Alba and Tjandra 2002).

The experimental methods to measure dipolar coupling constants are well established. In proteins, RDCs between several pairs of nuclei on the peptide plane can be measured; the most common and sensitive measurements utilize the N-H^N coupling. These can be extracted by recording in-phase and anti-phase J -coupled 2D NMR experiments, the sum and difference of which produce only the upfield and downfield components of a doublet (Ottiger et al. 1998). Another approach is the use of HSQC and TROSY experiments (Figure 3.2b), which have been found to be most sensitive in applications of RDC measurements to ribosomal particles (X. Wang, *et al*, in preparation).

3.3.4 Applications of residual dipolar couplings

As previously discussed, RDCs are commonly used to refine existing well-folded protein structures, such as those obtained by X-ray crystallography (Kanelis et al. 2001), and has also been applied in the study of large systems, such as the 360-kDa 20S proteasome core particle in combination with the methyl-TROSY effect (Ruschak and Kay 2012).

In conformationally heterogeneous samples, such as intrinsically disordered proteins, RDCs report on all orientations of inter-nuclear vectors over the ensemble of conformers interchanging on a timescale up to milliseconds. Structural elements such as α -helices adopt biased alignments to the external magnetic field, enabling analysis of the disordered protein conformations (Kosol et al. 2013). While there are challenges associated with determining the minimum ensemble size, RDC data can be used to either select computationally derived conformers according to their agreement with experimental data or in restrained MD simulations by introducing preferential conformational sampling (Salmon et al. 2010).

Previous work within our group has been undertaken to measure the RDCs of ribosomal particles including the 70S *E. coli* ribosome. While such measurements of large macromolecular assemblies present significant challenges in the sample preparation, data acquisition and computational analysis and modelling, it was demonstrated that successful alignment of 70S ribosomes could be achieved using filamentous bacteriophage (X. Wang, *et al*, in preparation). High-resolution spectra were acquired of

the aligned sample, and this enabled the measurement of RDCs of the flexible bL12 stalk protein of the 70S ribosome, which were subsequently used in MD simulations for structure refinement of the ribosome-bound domain. These studies have also been extended to RNCs of α -synuclein, in which a highly negatively charged mutant was produced to reduce electrostatic interactions with the ribosomal surface and thereby increase the mobility and observability of the nascent chain by NMR spectroscopy (A. Deckert, *et al.*, in preparation). In this Chapter, we explore how these methods can be extended more generally and with enhanced sensitivity to other RNC systems that interact more significantly with the ribosomal surface.

3.4 Results

3.4.1 Production of $^{13}\text{C}/^{15}\text{N}$ -labelled RNCs

To explore the structure of ribosome-bound nascent chains by 3D NMR spectroscopy, we used the FLN5+31 RNC as a model system. In addition to the wild-type, we also produced the mutant FLN5+31 A₃A₃ RNC because the nascent chain interactions with the ribosomal surface were largely removed by mutation of six phenylalanine and tyrosine to alanine residues, resulting in increased mobility and almost 2-fold increase in 2D SOFAST-HMQC sensitivity. This provided us a more NMR-sensitive RNC system to study. In addition, we were motivated to acquire carbon chemical shifts of both RNCs to analyse in more detail how tethering and proximity to the ribosome perturb the folding and structure of the nascent chains, and how loss of interaction with the ribosomal surface alters these effects.

RNCs with uniformly $^{15}\text{N},^{13}\text{C}$ -labelled nascent chains were generated *in vivo* in *E. coli* and purified at high yields as previously described (Cassaignau et al. 2016). During the course of NMR data acquisition, the sample was continuously monitored biochemically by western blot in parallel with SORDID diffusion measurements (Figure 3.3), to ensure that the resonances observed derived exclusively from ribosome-bound species. Because of the low concentrations (10 μM) and short lifetimes of the samples (often less than 20 h) and long times required for 3D NMR experiments, a new sample was made for each triple-resonance experiment, and where required, several samples were produced and the total data summed to generate the final data set.

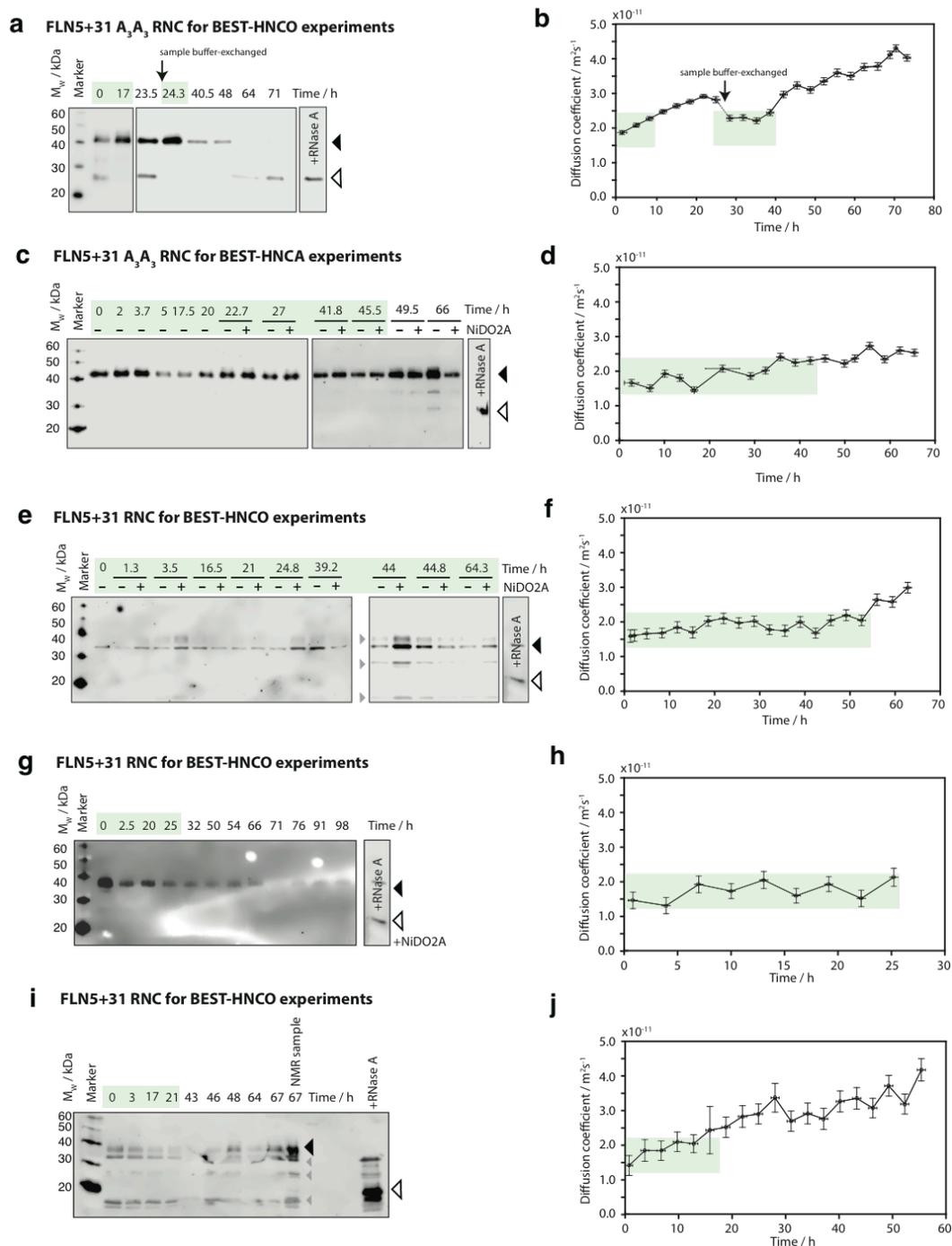


Figure 3.3 Monitoring of FLN5+31 A₃A₃ (a-d) and WT (e-j) RNC sample stability during NMR data acquisition. (a, c, e, g, i) Western blot analyses of RNCs detected using anti-His antibodies. Upshifted bands (~40 kDa, indicated by black triangles) correspond to the tRNA-bound form of the RNC in which the nascent chain is attached to the ribosome, with release of the nascent chain (and tRNA) monitored by the lower band (~23 kDa, indicated by the white triangles). Background bands deriving from the ribosome are indicated by grey triangles. (b, d, f, h, j) Diffusion coefficients of the NMR sample monitored by 1D SORDID experiments. Highlighted areas (green) indicate the time during which the RNC remains intact. Experiments recorded during this time were summed together to produce the final data set.

3.4.2 3D NMR of FLN5+31 A₃A₃ RNC

We initially explored the longitudinal relaxation-optimised BEST series of experiments (Schanda et al. 2006) of the FLN5+31 A₃A₃ RNC to determine the achievable experimental sensitivity of 3D NMR spectroscopy of large macromolecular complexes. We began by acquiring the BEST-HNCO experiment, as the HNCO experiment is the most sensitive triple-resonance pulse scheme (Table 3.1). The data were recorded using a 700-MHz spectrometer rather than at a higher magnetic field to reduce transverse relaxation effects of the carbonyl carbon, as its relaxation by chemical shift anisotropy (CSA) is proportional to the square of the field strength (Rule and Hitchens 2006).

The RNC sample was stable for a total of ~ 18 h (Figure 3.3a-b), enabling the acquisition of 12 short (1.5 h each) 3D BEST-HNCO experiments recorded throughout this period (interleaved by sample stability controls) which were summed to produce the final spectrum (Figure 3.4a-e). This permitted us to observe almost all the expected resonances in the HNCO spectrum of the FLN5+31 A₃A₃ RNC. Substantial broadening of residues occurred between G716 to G750, and residues within the flexible FLN6 linker did not give rise to observable cross-peaks, consistent with data from previous SOFAST-HMQC experiments of WT FLN5+31 RNC (Cabrita et al. 2016) and FLN5+31 A₃A₃ RNC (Figure 3.1c) (A.M.E. Cassaignau, *et al.*, in preparation). The most C-terminal residue observable in the HNCO was E725, 56 residues from the PTC, a greater number than what is expected to be sequestered within the exit tunnel. In addition, by comparing the signal intensities of the RNC with those obtained for isolated FLN5 A₃A₃ RNC (Figure 3.5f), we found significant reduction in intensity beyond V717. Such observations of WT RNC have been attributed to both a reduction in mobility of the tethered nascent chain and also binding of the nascent chain to the slowly tumbling ribosome (Cabrita et al. 2016); and this suggested specific interactions with the ribosomal surface persisted despite removal of stretches of aromatic residues. The slight signal decrease toward the N-terminus of FLN5 is likely due to binding of the histidine purification tag to the ribosomal surface as well. The remaining protein sequence did not show significant attenuation in signal relative to isolated protein. The mutations of aromatic residues to alanines therefore largely eliminated specific contacts with the ribosomal surface, enabling a substantial increase in SNR and, when combined with sensitivity gains from the BEST experiment, the acquisition of the first 3D spectrum of an RNC.

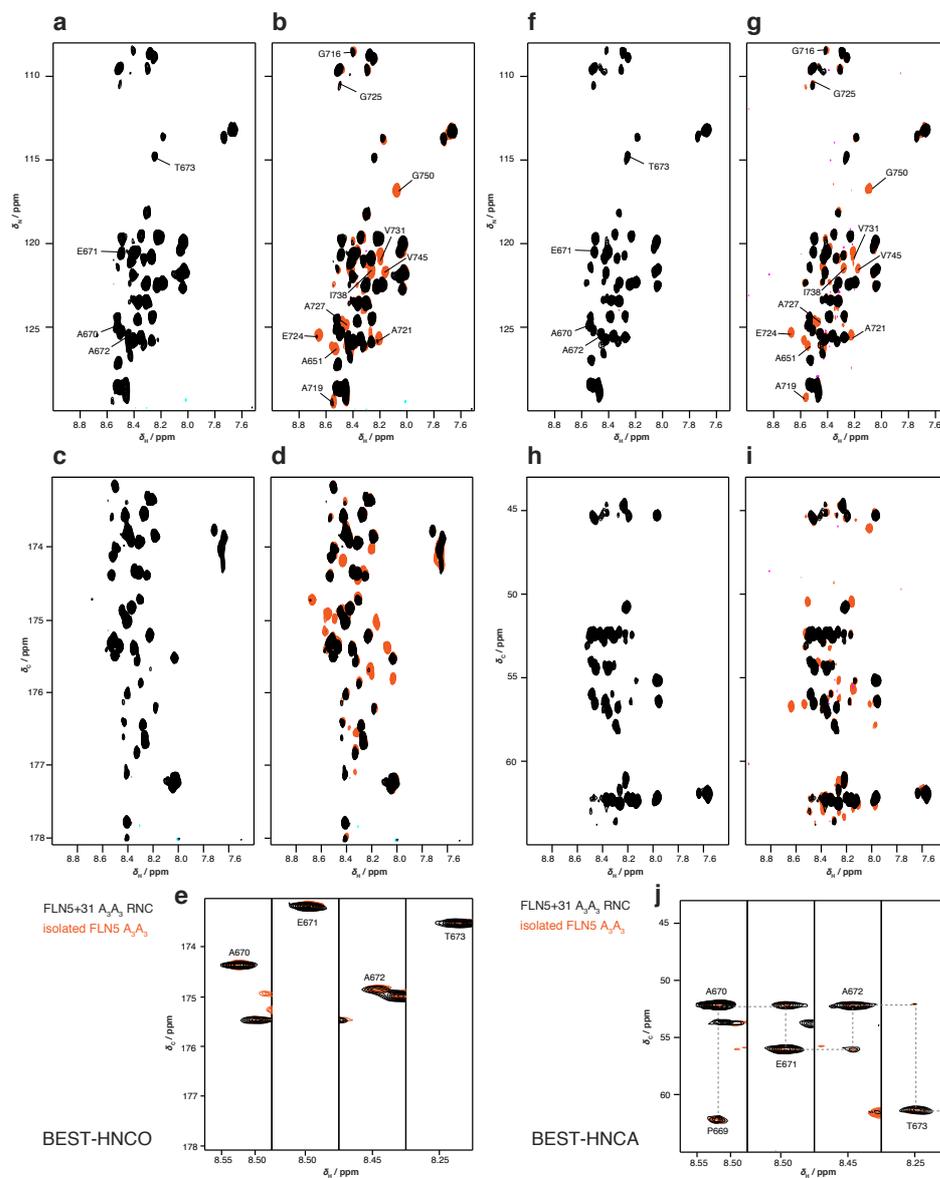


Figure 3.4 3D NMR spectroscopy of FLN5+31 A₃A₃ RNC using (a-e) BEST-HNCO and (f-j) BEST-HNCA experiments, recorded using a 700-MHz and 800-MHz spectrometer respectively, and at 10°C. (a, f) ¹H/¹⁵N projections of the 3D spectrum, highlighting the positions of the slices shown in e and j, and (b, g) overlaid with the corresponding 3D spectrum of isolated FLN5 A₃A₃ (orange), highlighting selected resonances that were broadened significantly in the RNC spectrum. (c, h) ¹H/¹³C projections of the 3D spectrum, and (d, i) overlaid with the corresponding 3D spectrum of isolated FLN5 A₃A₃ (orange). (e) ¹H/¹³C slices of the HNCO spectrum of RNC (black) and isolated FLN5 A₃A₃ (orange). (j) ¹H/¹³C slices of the HNCA spectrum of RNC (black) and isolated FLN5 A₃A₃ (orange), with sequential connectivity between residues shown with dashed lines.

We measured the $^1\text{H}^{\text{N}}$, ^{15}N and $^{13}\text{C}'$ chemical shifts of the HNCO resonances of the RNC, and compared these with the isolated protein (Figure 3.5a-c). Uncertainties were measured as half the ratio of the linewidth to SNR (Materials and Methods), and were uniformly small for most resonances, indicating that the SNR was sufficient for accurate measurement of the chemical shifts. The changes in chemical shifts were small ($\Delta\delta_{\text{H}} < 0.02$ ppm, $\Delta\delta_{\text{N}} < 0.2$ ppm, $\Delta\delta_{\text{C}'} < 0.1$ ppm), suggesting that the conformation of FLN5 A₃A₃ remains largely unchanged when ribosome-bound. However, we did observe small but systematic increases in $\Delta\delta_{\text{H}}$ toward the C-terminus. In addition, the $^{13}\text{C}'$ chemical shifts of the RNC were generally deshielded relative to isolated protein, indicating a possible change in secondary structure population to more helical conformations.

To investigate these structural changes further, we recorded BEST-HNCA experiments of FLN5+31 A₃A₃ RNC (Figure 3.4f-j) to obtain C $^{\alpha}$ chemical shifts, which are more sensitive probes of changes in secondary structure. We used the HNCA experiment rather than the HN(CO)CA experiment to retain sequential connectivity information and to observe two resonances for each residue. This approach enabled two independent measurements of the C $^{\alpha}$ chemical shifts in principle and allowed us to validate the transfer of assignments from isolated FLN5 A₃A₃ to the RNC.

The BEST-HNCA experiment was recorded within ~ 20 h. Despite the lower sensitivity of the HNCA experiment compared to the HNCO experiment (Table 3.1) due to the smaller J -coupling constant and parallel magnetization transfer pathways to the starting and preceding residue (Kay et al. 1990), we again observed almost all the expected nascent chain resonances with sufficient sensitivity to accurately measure their chemical shifts (Figure 3.4f-j and Figure 3.5e). For the majority of residues (70%), we detected the corresponding pair of HNCA cross-peaks and this enabled us to validate the residue assignments of the RNC spectra by sequential connectivity (Figure 3.4j) and also to obtain C $^{\alpha}$ chemical shifts for 3 proline residues (P669, P690, P698). We measured the C $^{\alpha}$ chemical shifts, and where two peaks were observed for the same residue, we used both measurements and calculated their average weighted to the cross-peak signal intensity. Comparison of the chemical shifts with those of the disordered, isolated FLN5 A₃A₃ protein (Figure 3.5e) showed that the perturbations were small ($\Delta\delta_{\text{C}^{\alpha}} < 0.2$ ppm), but in contrast to the C' chemical shift changes, were both deshielded and shielded along the protein sequence.

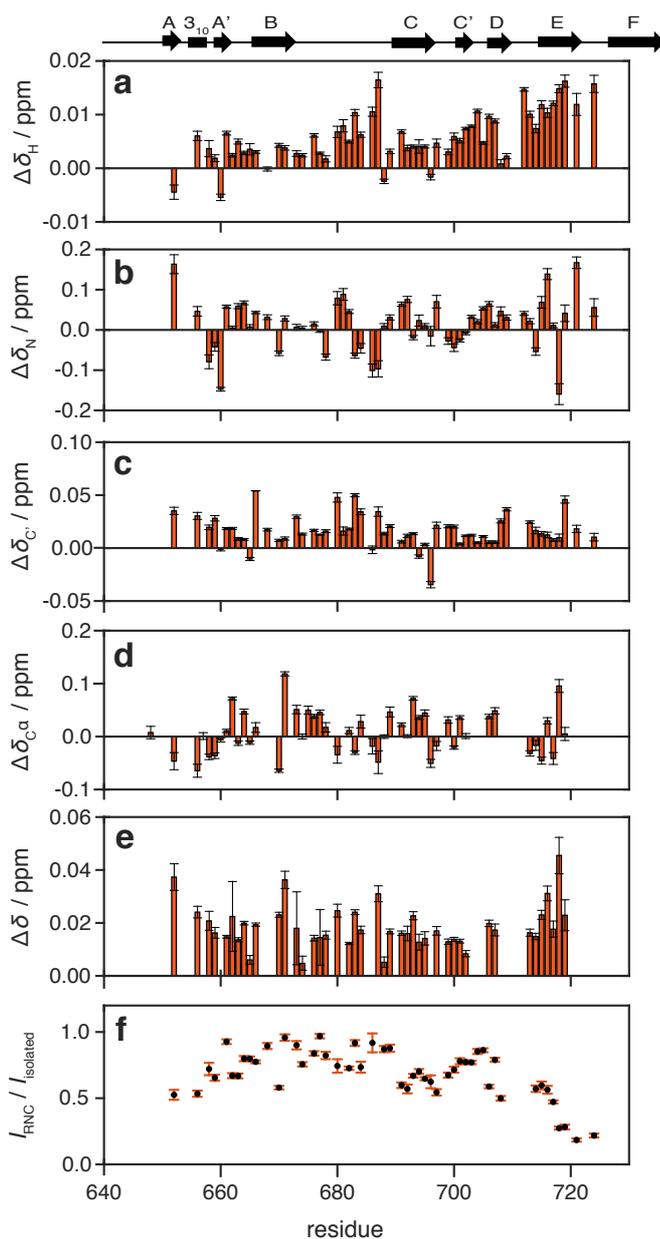


Figure 3.5 Analysis of HNCO and HNCA experiments of FLN5+31 A₃A₃ RNC. Residue-specific changes in (a) amide ¹H, (b) amide ¹⁵N, (c) carbonyl ¹³C, and (d) alpha ¹³C chemical shifts of FLN5+31 A₃A₃ RNC relative to isolated FLN5 A₃A₃. (e) Chemical shift perturbations of FLN5+31 A₃A₃ RNC relative to isolated FLN5 A₃A₃ incorporating all of the available amide ¹H and ¹⁵N, and carbonyl and alpha ¹³C chemical shifts, with chemical shift changes for each nucleus scaled accordingly. Uncertainties of chemical shifts were calculated using measured linewidths. (see Materials & Methods). (f) Signal attenuation of BEST-HNCO experiment of FLN5+31 A₃A₃ RNC relative to isolated FLN5 A₃A₃ recorded under the same conditions. Errors correspond to spectral noise. Schematic above shows secondary structure elements of full-length WT FLN5.

Because the chemical shift changes were small for all nuclei measured by the BEST-HNCO and BEST-HNCA experiments, we combined all the chemical shifts to determine their cumulative effect. When summed together (and weighting for different nuclei, see Materials and Methods), we found that the chemical shift perturbations, while remaining small, showed systematic increases toward both the N- and C-termini (Figure 3.5e). This could be attributed to binding of the histidine purification tag and specific interactions of the FLN5 C-terminus to the ribosomal surface respectively, and therefore correlate with our observations of the HNCO signal attenuation, as discussed earlier.

We also translated the set of measured nascent chain chemical shifts into probabilities of occupancy of secondary structure elements, using the δ 2D method

(Camilloni et al. 2012). The results showed that the nascent chain was predominantly in a random coil conformation with some polyproline II structure, characteristic of disordered proteins (Figure 3.6a). The nascent polypeptide was also found to populate extended β -sheet conformations throughout the protein sequence, indicating that the disordered nascent polypeptide still possessed some residual structure. We next sought to compare the ribosome-bound and isolated forms of FLN5 A₃A₃ as a direct measurement of the effect of tethering to the ribosome on the protein structure (Figure 3.6b-c). The comparison between different protein states also reduces the error associated with the use of reference chemical shifts of random coils for calculating secondary structure populations. As broadening of RNC resonances resulted in only a subset of measurable chemical shifts, we used the same set of cross-peaks for the isolated conformation to obtain quantitative comparisons without the effects of systematic errors. The data showed that the changes in the secondary structure population of FLN5 A₃A₃ outside of the exit tunnel when ribosome-bound were all within the uncertainties of the δ 2D method. This might be expected, as the conformation of FLN5 A₃A₃ remains disordered when isolated from the ribosome (in contrast to the wild-type) due to disruption of the hydrophobic core by mutations of the aromatic residues. However, as chemical shifts of the FLN5 A₃A₃ C-terminus (beyond E725) and FLN6 linker could not be measured, this analysis does not preclude secondary structure formation in segments of FLN5 within the ribosomal exit tunnel and vestibule that interact with ribosomal proteins and RNA.

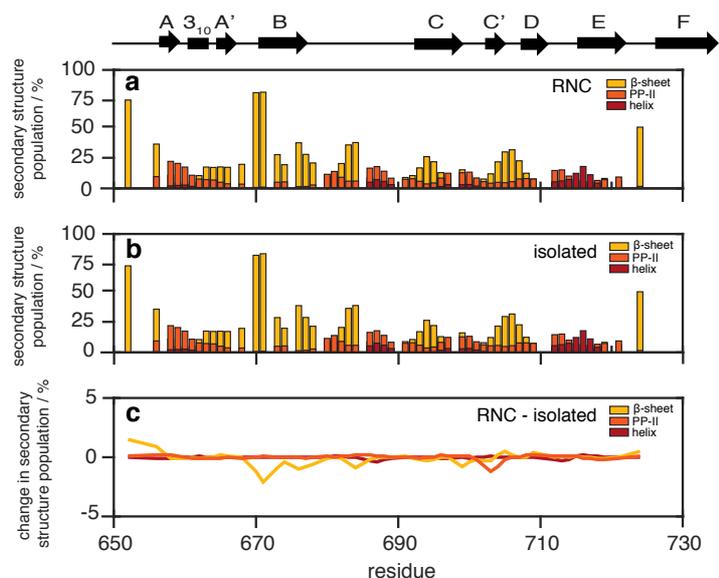


Figure 3.6 Comparison of secondary structure populations between isolated and ribosome-bound FLN5+31 A₃A₃, calculated with the δ 2D method (Camilloni et al. 2012) using the amide backbone chemical shifts measured by BEST-HNCO and BEST-HNCA experiments. Secondary structure populations of (a) RNC and (b) isolated protein, and (c) their difference. Schematic above shows secondary structure elements of full-length WT FLN5.

3.4.3 PLRE studies on BEST-HNCO and BEST-HNCA experiments of FLN5 Δ 12

The NMR experiments of FLN5+31 A₃A₃ RNC demonstrated that the accurate measurement of carbon chemical shifts by 3D NMR was feasible and that this could be used to determine secondary structure populations of ribosome-bound species. We therefore sought to apply these methods to study wild-type FLN5+31 RNC, which adopts a disordered conformation when ribosome-bound, but when isolated from the ribosome, folds into its native, globular FLN5 structure, as discussed in the introduction of this chapter. We were motivated to understand whether native contacts were also present when ribosome-bound such that a shift towards more folded conformations could be detected. The predominant challenge with NMR of FLN5+31 RNC is the approximately 4-fold reduction in sensitivity (relative to FLN5+31 A₃A₃ RNC) due to specific interactions with the ribosomal surface. This translates into a ~16-fold increase in experimental time required to achieve the same sensitivity observed for FLN5+31 A₃A₃ RNC that cannot be readily attained with the short lifetimes of RNC samples. We therefore explored the use of the PLRE agent NiDO2A in triple-resonance experiments to significantly reduce the total experimental times required to accurately measure carbon chemical shifts.

We first tested the PLRE-effect on HNCO and HNCA experiments with the isolated FLN5 Δ 12 protein, a C-terminally truncated variant of FLN5 that was previously found to have a similar disordered conformation as the FLN5+31 RNC (Cabrita et al. 2016). BEST-

HNCO and BEST-HNCA experiments were acquired as 2D spectra in the ^1H - ^{13}C dimensions on uniformly ^{15}N , ^{13}C -labeled samples of FLN5 Δ 12 in the presence and absence of 40 mM NiDO2A PLRE agent (Figure 3.7a, d). As seen in Chapter 3 and in previous studies (Cai et al. 2006; Theillet et al. 2011), the PLRE agent did not induce chemical shift changes and importantly, shift changes also did not occur in the ^{13}C dimension for both carbonyl and C^α nuclei. In addition, only marginal line broadening was observed. However, analysis of the spectra showed a uniform distribution of sensitivity enhancements across the protein sequence on addition of 40 mM NiDO2A, with a mean residue-specific increase of 1.31 ± 0.17 (s.d.) and 1.43 ± 0.20 (s.d.) in the BEST-HNCO and BEST-HNCA experiments respectively.

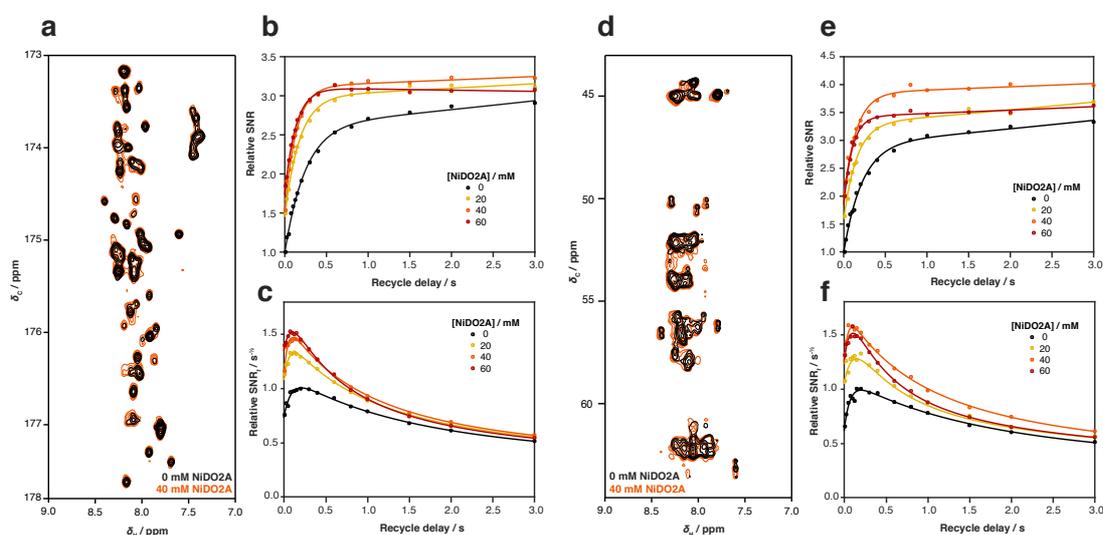


Figure 3.7 The effect of paramagnetic longitudinal relaxation enhancement on the sensitivity of (a-c) HNCO and (d-f) HNCA experiments of FLN5 Δ 12. Overlay of (a) ^1H - ^{15}N - ^{13}C HNCO and (d) ^1H - ^{15}N - ^{13}C HNCA spectra recorded as 2D experiments in the absence (black) and presence (orange) of 40 mM NiDO2A. Relative signal-to-noise ratio of (b) 1D ^1H - ^{15}N - ^{13}C HNCO and (e) 1D ^1H - ^{15}N - ^{13}C HNCA experiments as a function of the recycle delay at NiDO2A concentrations of 0-60 mM. Relative sensitivity (signal-to-noise ratio per unit time $^{1/2}$) of (c) 1D ^1H - ^{15}N - ^{13}C HNCO and (f) 1D ^1H - ^{15}N - ^{13}C HNCA experiments as a function of the recycle delay at NiDO2A concentrations of 0-60 mM.

To further quantify the experimental sensitivity enhancements and to optimize the concentration of PLRE agent, a series of 1D ^1H - ^{15}N - ^{13}C BEST-HNCO and BEST-HNCA experiments were acquired with the recycle delay varied between 10 ms and 3 s, and the amide envelope was integrated to calculate the SNR (Figure 3.7b, e). The experimental sensitivities were derived by division of the SNR by the square root of the total time for each experiment (Figure 3.7c, f). Similar to PLRE studies of SOFAST-HMQC experiments

of isolated proteins described in Chapter 3, we found that progressive addition of NiDO2A to samples of FLN5 Δ 12 increased the sensitivity of BEST-HNCO experiments by up to 1.47-fold with 40 mM NiDO2A. The sensitivity of BEST-HNCA experiments was also enhanced by increasing concentrations of PLRE agent by up to 1.51-fold with 40 mM NiDO2A. These data were concomitant with increases in the measured ^1H longitudinal relaxation rates, and with the measured transverse relaxation rates of ^1H and ^{15}N spins that increased only marginally on addition of increasing NiDO2A concentrations (Table 2.2). Titration of additional NiDO2A to a final concentration of 60 mM, resulted in less than 5% improvement in the sensitivity of both experiments relative to 40 mM NiDO2A due to increasing transverse relaxation rates. At the concentrations of NiDO2A tested, the sensitivity was primarily limited by the rate of longitudinal relaxation during the recycle delay when magnetization returns to equilibrium, but at higher concentrations (60 mM) and longer recycle delays, a sufficiently fast R_1^{H} was achieved and the signal was attenuated to a greater extent by rapid transverse relaxation. This effect resulted in a shift to a shorter optimum recycle delay (from 200 ms to 75 ms for the BEST-HNCO experiment), and diminishing returns in sensitivity improvements on addition of higher NiDO2A concentrations. In summary, at the optimum concentration of 40 mM NiDO2A, the sensitivity of BEST-HNCO and BEST-HNCA experiments of disordered FL5 Δ 12 was improved by 50%, reducing experimental times by 2.25-fold to achieve the same sensitivity without significant T_2 -associated broadening effects.

[NiDO2A] / mM	$R (^{15}\text{N}) / \text{s}^{-1}$		$R (^1\text{H}) / \text{s}^{-1}$		
	R_1	R_2	R_1 hard	R_1 amide-selective	R_2
0	1.75 ± 0.09	5.16 ± 0.76	1.05 ± 0.07	9.39 ± 0.91	22.74 ± 1.37
20	1.79 ± 0.08	5.33 ± 3.05	2.29 ± 0.08	10.14 ± 0.86	22.33 ± 1.73
40	1.82 ± 0.05	5.25 ± 1.38	3.96 ± 0.06	10.59 ± 0.69	21.41 ± 0.93
60	1.86 ± 0.04	5.26 ± 1.35	5.25 ± 0.21	11.63 ± 0.73	22.74 ± 0.63

Table 3.2 Measured ^1H and ^{15}N longitudinal and transverse relaxation rates of FLN5 Δ 12 as a function of NiDO2A concentration.

3.4.4 3D NMR of FLN5+31 RNC

We next produced samples of uniformly $^{13}\text{C},^{15}\text{N}$ -labelled ribosome-bound FLN5+31 nascent chains, and tested the effect of NiDO2A on the sensitivity of BEST-HNCO experiments. Using 1D experiments and integrating the amide envelope, we found a 1.5-fold increase in sensitivity on addition of 40 mM NiDO2A (Figure 3.8i), in line with those

obtained with isolated FLN5Δ12. A series of 3D BEST-HNCO experiments were therefore recorded with the NiDO2A-doped RNC sample, and when a reduction in RNC occupancy was detected (by SORDID diffusion measurements), data acquisition was halted and a new sample was prepared. The data sets of 3 samples were combined to produce the final HNCO spectrum, with a total experimental time of ~95 h (Figure 3.8a, f, h).

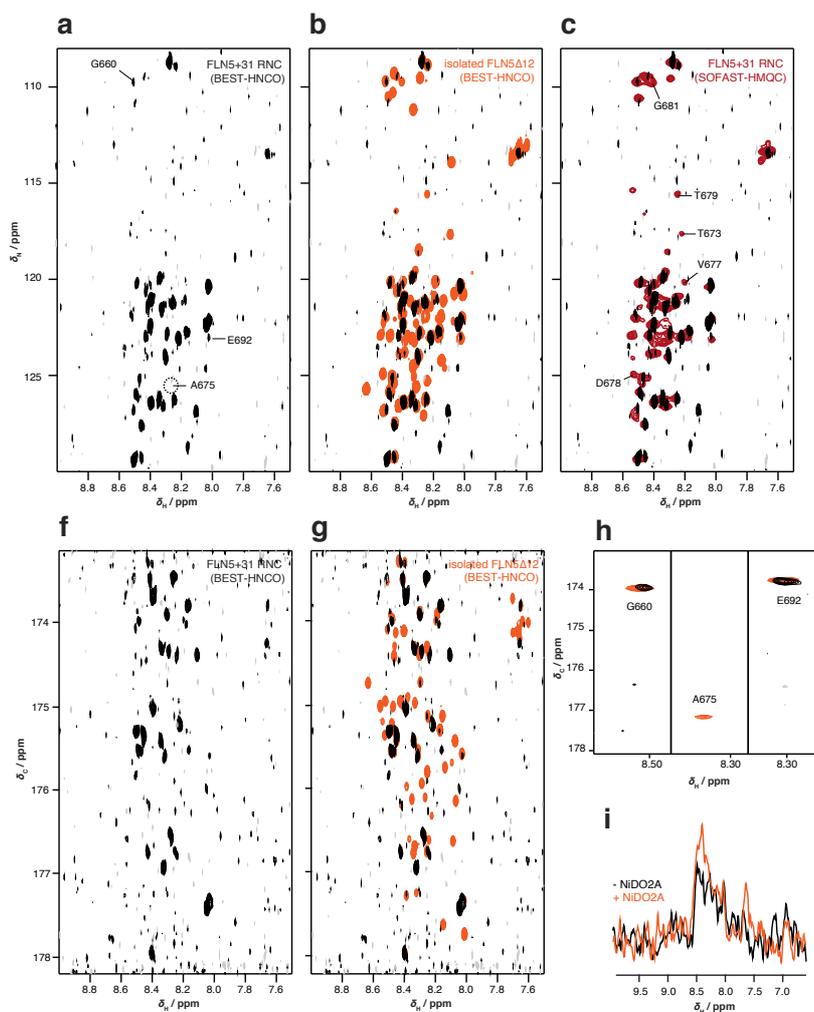


Figure 3.8 BEST-HNCO experiments of FLN5+31 RNC using PLRE, recorded using a 700-MHz spectrometer at 10°C. (a) $^1\text{H}/^{15}\text{N}$ projection of the BEST-HNCO spectrum, highlighting the positions of the slices shown in e, and (b) overlaid with the corresponding HNCO of isolated FLN5Δ12 (orange) highlighting selected resonances that were broadened significantly and undetectable in the RNC spectrum, and (c) overlaid with the SOFAST-HMQC spectrum of FLN5+31 RNC (red) highlighting resonances that were expected but not observable in the HNCO spectrum (f) $^1\text{H}/^{13}\text{C}$ projection of the HNCO spectrum, and (g) overlaid with the corresponding HNCO of isolated FLNΔ12 (orange). (h) $^1\text{H}/^{13}\text{C}$ slices of the HNCO spectrum of RNC (black) and isolated FLN5 A3A3 (orange). (i) 1D HNCO experiments in the absence (black) and presence (orange) of 40 mM NiDO2A.

The 3D HNC0 spectrum of the RNC showed 41 discernable resonances that all derived from the nascent chain and which could be mapped onto FLN5Δ12 cross-peaks (Figure 3.8b, g). These represented 66% of the resonances expected for the RNC (when compared to cross-peaks observable and that can be accurately measured in SOFAST-HMQC spectra, Figure 3.8c). HNC0 resonances of the nascent chain sequence that were difficult to discern from the noise were predominantly in the C-terminus beyond residue D709 and the region between V662 and V682. Indeed these regions were significantly broadened in the SOFAST-HMQC spectra due to interactions with the ribosomal surface (Figure 3.9e), and their absence in the HNC0 spectrum is therefore likely due to the restricted sensitivity of the experiment.

The observable resonances had an average residue-specific SNR of 10.9 ± 3.5 (s.d.) which enabled the accurate measurement of the $^1\text{H}^{\text{N}}$, ^{15}N and $^{13}\text{C}'$ chemical shifts of 25 residues. Comparison with the HNC0 of isolated FLN5Δ12 showed only small changes in chemical shift ($\Delta\delta_{\text{H}} < 0.01$ ppm, $\Delta\delta_{\text{N}} < 0.2$ ppm, $\Delta\delta_{\text{C}} < 0.05$ ppm, Figure 3.9a-d) with those of the carbonyl carbon showing weak bias to deshielded chemical shifts, corresponding to a shift to more helical conformations. When combined, the chemical shift perturbations seemed to show some anti-correlation with the signal attenuation, with the largest changes occurring towards both termini (where the signal was significantly broadened) (Figure 3.9d-e). The resonances of regions unobservable in the obtained HNC0 spectrum due to low sensitivity are therefore also those that are likely to be most perturbed when ribosome-bound, an indication of which is seen in the $^1\text{H}^{\text{N}}$ and ^{15}N chemical shift changes measured by SOFAST-HMQC experiments (Figure 3.1). Additional sensitivity gained by recording data of a further RNC sample, equivalent to 20-40 h additional experimental time, is likely to result in observation of these resonances by consideration of the total acquisition time used to acquire the HNC0 spectrum of FLN5+31 A₃A₃ RNC, the difference in sensitivity between the RNCs, and the enhancement in sensitivity provided by addition of NiDO2A.

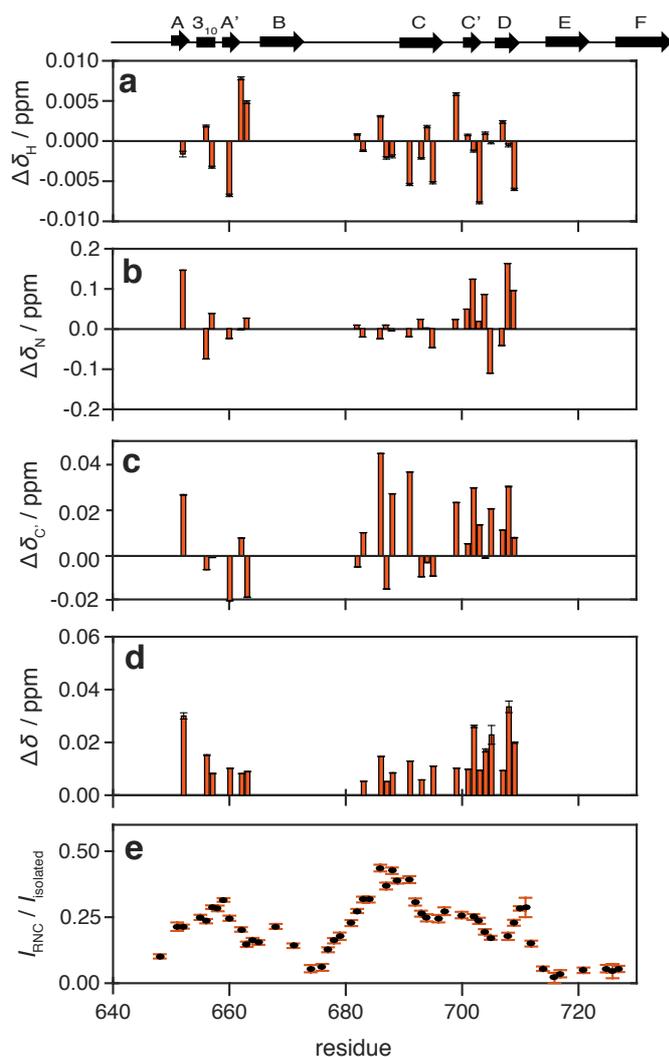


Figure 3.9 Analysis of the HNCOC experiment of FLN5+31 RNC. Residue-specific changes in (a) amide ^1H , (b) amide ^{15}N , and (c) carbonyl ^{13}C chemical shifts of FLN5+31 RNC relative to isolated FLN5 Δ 12. (e) Chemical shift perturbations of FLN5+31 RNC relative to isolated FLN5 Δ 12 incorporating all of the available amide ^1H and ^{15}N , and carbonyl ^{13}C chemical shifts, with chemical shift changes for each nucleus scaled accordingly (see main text). (f) For reference, signal attenuation of SOFAST-HMQC experiment of FLN5+31 RNC relative to isolated FLN5 Δ 12 (SOFAST-HMQC data recorded and analysed by A.M.E. Cassaignau). Schematic above shows secondary structure elements of full-length WT FLN5.

3.4.5 Towards residual dipolar coupling measurements of FLN5+31 RNC

The ability to use PLRE to enhance the sensitivity of 3D NMR experiments and thereby accurately measure RNC carbon chemical shifts with only minimal T_2 -associated broadening effects encouraged us to explore other structural techniques that have eluded NMR studies of RNCs due to their inherent insensitivity. In this light, we have begun some preliminary work in developing a strategy to measure RDCs of the FLN5+31 RNC using PLRE-induced sensitivity improvements.

As previously discussed, work has recently been undertaken to successfully extract RDC measurements from both unoccupied 70S *E. coli* ribosomes (X. Wang, *et al.*, in preparation) and also those harboring α -synuclein nascent chains (A. Deckert, *et al.*, in preparation). In contrast to FLN5+31 RNC, in both cases, the relatively long sample lifetimes (approximately one week rather than 1-2 days for FLN5 RNCs) and minimal interactions with the ribosomal surface were highly beneficial to the spectroscopic

sensitivity and were exploited to measure RDCs. To apply these methods to FLN5+31 RNCs would therefore require additional sensitivity gains, and we therefore sought to extend our PLRE studies to these experiments

We generated uniformly ^{15}N -labelled ribosome-bound FLN5+31 and aligned the RNC in bacteriophage at a concentration of 10 mg/mL, which was found to be sufficient for alignment of the disordered α -synuclein RNC to measure RDCs without significant broadening effects associated with high concentrations of alignment media (A. Deckert, *et al*, in preparation). The alignment was monitored by measuring the quadrupolar splitting of deuterium signals arising from ^1H - ^2H -O molecules in the NMR sample buffer. The splitting was determined as 4.7 Hz (Figure 3.9a), corresponding to a Pf1 concentration of ~ 4 mg/mL. The discrepancy between the Pf1 concentrations is due to the difficulty in accurately pipetting the viscous phage stock solution and unreliable absorbance measurements to determine the phage concentration, and these therefore require optimisation in future investigations.

We proceeded with acquiring sets of HSQC and TROSY experiments which, when obtained for both isotropic and anisotropic samples, enable the measurement of the RDCs (section 3.3.3). We were particularly interested in how the PLRE agent NiDO2A would affect the resonances of aligned samples as both the nickel complex and alignment in phage inevitably introduce signal broadening associated with increased transverse relaxation when titrated to NMR samples at high concentrations. We acquired phase-cycled ^1H - ^{15}N HSQC experiments of FLN5+31 in bacteriophage for ~ 10.5 h in both the absence (Figure 3.10e) and presence of 40 mM NiDO2A (Figure 3.10f), during which the sample stability was monitored by SORDID diffusion measurements and determined to be stable for a total of ~ 55 h (Figure 3.10b). On addition of NiDO2A, we observed an increase of 1.83-fold in HSQC-sensitivity, using a recovery delay of 250 ms (Figure 3.10c) that was uniform across the protein sequence and readily observable in the 2D spectra. We combined all the 2D HSQC experiments acquired during the lifetime of the RNC sample (total experimental time of ~ 36 h, Figure 3.10g) and were able to observe all of the expected nascent chain resonances that importantly did not show significantly increased line widths even in the presence of alignment media.

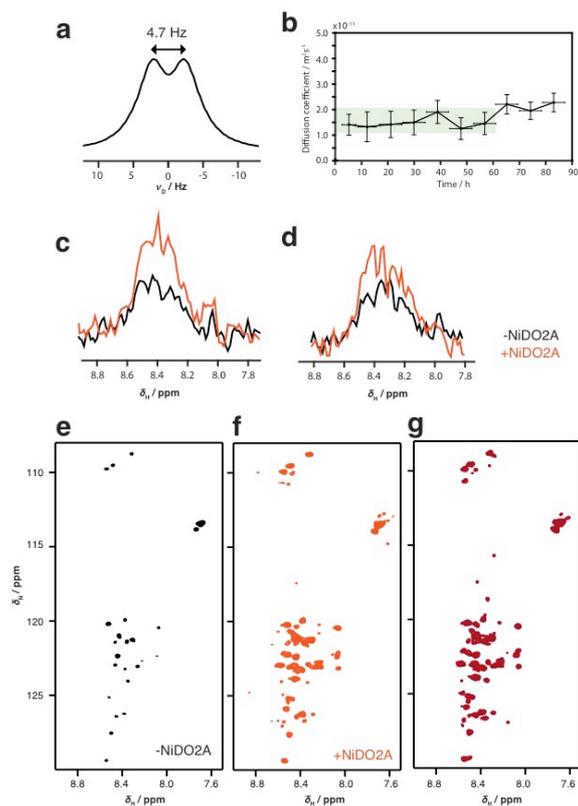


Figure 3.10 Preliminary investigations of residual dipolar coupling measurements of FLN+31 RNC using PLRE; experiments acquired on a 700-MHz spectrometer at 10°C. (a) Measurement of quadrupolar splitting of ^2H NMR signal following sample alignment with bacteriophage. (b) Monitoring of FLN5+31 RNC sample stability during NMR data acquisition using 1D SORDID experiments to determine the diffusion coefficient. Highlighted area (green) indicates the time during which the RNC remains intact. Experiments recorded during this time were summed together to produce the final data set. (c) 1D ^1H - ^{15}N HSQC and (d) 1D ^1H - ^{15}N TROSY spectra of FLN5+31 RNC aligned with phage in the absence (black) and presence (orange) of 40 mM NiDO2A, total experimental time of ~ 10.5 h each. 2D ^1H - ^{15}N HSQC spectra of FLN5+31 aligned with phage in (e) the absence and (f) the presence of 40 mM NiDO2A. (g) 2D ^1H - ^{15}N HSQC spectrum of FLN5+31 aligned with phage, total experimental time of ~ 36 h.

We also acquired 2D ^1H - ^{15}N TROSY experiments of the same aligned RNC sample, but these did not result in any observable cross-peaks. We attribute this to the lower sensitivity of TROSY experiments, which retains only one of four multiplet components of the ^1H - ^{15}N moiety (Pervushin et al. 1997). However, using 1D TROSY experiments, we found that the addition of 40 mM NiDO2A improved the sensitivity by 1.38-fold (Figure 3.10d). As the 2D HSQC resonances did not show significant changes in the line widths, we expect this to be the case for TROSY spectra, particularly when considering the TROSY effect (Pervushin et al. 1997).

We have shown how the PLRE effect can be incorporated into HSQC and TROSY experiments to improve the experimental sensitivity whilst crucially and despite the

addition of the phage alignment media, retaining the line widths of the resonances such that chemical shifts can still be accurately measured. We propose that future RDC measurements of FLN5+31 RNC incorporate NiDO2A into the sample buffers immediately at the start of NMR data acquisition, the use of BEST-HSQC and BEST-TROSY experiments, and ensure that the time allocated for HSQC and TROSY experiments is proportioned according to their PLRE-enhanced sensitivities, to maximize the experimental sensitivity that can be achieved from a single RNC sample and which will enable measurement of RDCs.

3.5 Discussion

As previously described, the low achievable concentrations, short lifetimes, and slow tumbling of purified RNC samples are significant challenges in their study by NMR spectroscopy, and are systemic of the study of other such large macromolecular assemblies, for example the 900-kDa GroEL-GroEL complex (Fiaux et al. 2002). In these NMR studies, the assignment of NMR resonances have relied on a 'divide and conquer' approach in which either monomeric units or, in the case of RNCs, the nascent chain, are assigned as small component units isolated from the larger complex, and the corresponding assignment of NMR resonances are transferred to the complete assembly. While this strategy has enabled the study of supra-molecular complexes by 2D NMR (Frueh et al. 2013), it relies on the fact that resonances do not undergo extensive shift changes when in complex with the complete structure. The studies described here show that at least some subtle chemical shift changes can occur for nascent chains when bound to the ribosome (in the case of the FLN5+31 A₃A₃ RNC), and it likely that more significant chemical shift changes can occur for the wild-type FLN5+31 RNC, particularly in ¹³C^α and ¹³C^β nuclei that are highly sensitive to secondary structure formation and within regions that are significantly broadened due to interactions with the ribosomal surface. The use of direct 3D NMR measurements on large complexes is therefore beneficial in both validating resonance assignments and in predicting secondary structure populations.

The reduction in nascent chain interactions with the ribosomal surface and resulting increase in mobility, by mutagenesis of aromatic residues to alanines, enabled the acquisition of the first 3D spectra of an RNC, each from a single sample preparation, using the longitudinal relaxation-enhanced BEST-HNCO and BEST-HNCA experiments. The FLN5+31 A₃A₃ RNC did not show significant changes in the measureable chemical shifts or secondary structure populations, indicating that tethering to the ribosome does not significantly affect nascent chain residues well outside of the exit tunnel. These experiments formed the basis of the development of a strategy to study how the

structures of co-translational folding intermediates are perturbed by tethering to the ribosome. Application of the PLRE agent NiDO2A to 3D NMR experiment provided up to 50% increase in sensitivity and thereby enabled accurate measurement of ~40% of the $^1\text{H}^N$, ^{15}N and $^{13}\text{C}'$ chemical shifts of wild-type FLN5+31 RNC, with much of the remaining resonances at just above the noise level. While further sample preparations and HNCO experiments can provide additional sensitivity (by effectively increasing the total number of scans acquired) to measure the outstanding chemical shifts, alternative methods to increase sensitivity, such as the use of non-uniform weighted sampling (Waudby and Christodoulou 2012) and deuteration of NMR samples (see Chapter 4) (Sattler and Fesik 1996), are likely to be highly beneficial. Indeed the measurement of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts of FLN5+31 RNC presents further spectroscopic challenges. We therefore propose that these techniques in combination with PLRE and the more sensitive BEST-HN(CO)CA (Grzesiek and Bax 1992a) and BEST-HN(CO)CACB experiments (Grzesiek and Bax 1992b) (Table 3.1) should be used in these endeavours.

Recent studies of the co-translational folding of FLN5 by NMR spectroscopy raised interesting questions on how the ribosome tunes the structure acquisition of multi-domain proteins and stabilises the formation of disordered proteins; the successful application of 3D NMR spectroscopy to FLN5 RNCs described here raises the prospect of furthering our understanding of this process, and these are widely applicable to other macromolecular assemblies too. Sensitivity enhancement by titration of PLRE agents has now also been extended to other NMR investigations of FLN5 RNCs such as in cross-correlated relaxation rate measurements (A.M.E. Cassaignau, C.A. Waudby, *et al.*, in preparation), and we have here described preliminary studies in applying such sensitivity-improved methods to RDC measurements that can provide orthogonal information on the structure and dynamics of ribosome-bound nascent chains.

In this Chapter, we have shown how the application of PLRE can be extended to more complex NMR experiments to extract structural information on large macromolecular assemblies that have hitherto been unattainable due to limitations in spectroscopic sensitivity. We chose the use of chemical shift and RDC measurements as examples in this work to demonstrate the applications of PLRE to RNCs as they can readily be used as either validations or restraints in MD simulations to bias force fields to the native structure, even when only sparse data is available (Robustelli et al. 2010). We anticipate that these approaches will enable unprecedented insights into complex structure formation of the dynamic nascent chain outside of the exit tunnel that have so far eluded other high resolution techniques such as X-ray crystallography and cryo-EM.

Chapter 4

***In vivo* observations of ribosome-nascent chain complexes by in-cell NMR spectroscopy**

4.1 Introduction

NMR spectroscopy is unique in its ability to observe biological macromolecules *in situ* in their native environment and under near-physiological conditions, and this has already been exploited extensively in magnetic resonance imaging (MRI) for instance. Its application of biomolecular NMR to cellular samples has relied on the ability to observe only those nuclei that are 'NMR-active' by selective isotopic labelling (section 4.3.3), enabling discrimination of the species under investigation from the unlabelled, crowded, cellular environment. A similar strategy has also been successfully demonstrated in *in vitro* NMR observations of RNCs that were produced *in vivo* and where residues from the nascent chain account for <1% of the total ribosomal particle. This therefore suggested to us that NMR studies of RNCs would be amenable to in-cell investigations. We describe the development of a strategy for such studies in this Chapter. The ability to observe RNCs *in vivo* offers enticing prospects into understanding how the native, cellular environment perturbs the structural and dynamic features of co-translational protein folding that have yet to be explored, due in large part to the relative insensitivity of NMR of large systems. To deconstruct the different factors that play a role in altering co-translational folding, the use of reconstituted cytosol has enabled high-resolution, NMR experiments to be performed within a cellular mimic under more controlled conditions. In this work, we describe structural perturbations of disordered FLN5 in reconstituted cytosol, which were found to represent the in-cell data well. By comparisons with BSA and Ficoll as artificial crowding agents, we were able to attribute the structural changes to both excluded volume effects and non-specific interactions. We believe such macromolecular crowding effects are likely to have structural and biological implications to its ribosome-bound analogue.

4.1.1 *In vivo* approaches to study protein structure and folding

The basis of our current understanding of *de novo* protein folding derives predominantly from refolding studies of denatured, purified proteins in dilute buffers. However, as described in Chapter 1, there are significant deviations to the biological process, not least the vectorial emergence of the ribosome and the crowded cellular milieu (Cabrita et al. 2010). Significant progress has been made in recent years to understand how these factors affect protein folding in the cell, but are often studied independently and are only beginning to converge together in more recent studies, concomitant with technological advances, both experimentally and computationally.

Indeed, in parallel with *in vitro* studies of translationally stalled RNCs (as previously discussed in this work), experimental techniques have been developed enabling observation of biological molecules inside cells. The earliest work on studying protein structure *in vivo* utilized chemical cross-linking and enzymatic digestion. Challenges associated with selectivity have been alleviated more recently, but the technique still involves disruptive cell lysis and subsequent analysis *in vitro* (Kaake et al. 2014).

In contrast, recent advances in fluorescence microscopy techniques are enabling studies of protein structure and dynamics directly within living cells. These typically involve labelling the protein under investigation by fluorescence, such as FRET (Ebbinghaus et al. 2010) or using biarsenical fluorescein dye (FLAsH) (Ignatova and Gierasch 2004), or introducing mutations to modulate fluorescence and folding rates in green fluorescent protein (Merkel and Regan 2000). These studies have been able to report on protein stability and folding, and when combined with single particle tracking, enable detection of conformations of individual proteins (Sakon and Weninger 2010; König et al. 2015), with each reporting on how the cellular environment, most notably macromolecular crowding, perturbs *in vitro* observations of structure and dynamics. In recent years, *in vivo* fluorescence imaging has been applied to single ribosomes (English et al. 2011), building on the foundations from studies of fluorescently-labelled ribosomes performed in purified *in vitro* transcription and translation systems (Perez and Gonzalez 2011). Single-particle tracking of ribosomes has enabled *in vivo* investigations of the stringent response machinery (English et al. 2011), and it is likely that applications of this approach will enable future co-translational folding studies in living cells, which are already beginning to be explored *in vitro* (Uemura et al. 2008).

Studies using *in vivo* fluorescence techniques have given unique insights into important biological processes, but generally report on global changes in biomolecular structures. Instead, in-cell NMR spectroscopy has emerged as a complementary

technique enabling higher, near-atomic resolution investigations of structure and dynamics of biological molecules in living cells.

4.2 Applications of in-cell NMR spectroscopy

Advances in experimental sensitivity and resolution have revolutionized the study of large biological complexes by NMR spectroscopy, as previously discussed. The ability to observe these low sensitivity, rapidly relaxing systems have enabled the application of NMR to studies of biological molecules in living cells, which have similar spectroscopic challenges. The emerging field of in-cell NMR has also been driven by technological advances in sample preparation (discussed in section 4.3), and ultimately by the desire to study protein structure under more physiologically relevant conditions, as evidence grows on how the crowded cellular environment can induce changes in conformation, stability and dynamics of proteins.

An example of this is shown in a landmark in-cell NMR study, which for, the first time, determined a protein structure *de novo* and exclusively from NMR data acquired in living cells (Sakakibara et al. 2009). This study showed that sensitivity-enhanced triple-resonance and NOE experiments could rapidly be acquired on short-lived bacterial samples, and subsequently be used in calculating the 3D, *in vivo* structure of proteins, in this case, of THA1718, from *Thermus thermophilus* HB8. While its structure remained largely the same as that *in vitro*, subtle differences were observed in the dynamic, metal-binding loop regions that were attributed to the effects of viscosity and molecular crowding and that may have implications to its biological function (Sakakibara et al. 2009). Other methods for structure determination that have traditionally been the preserve of *in vitro* NMR studies are now also being utilized in living cells. In NMR studies of the Streptococcal protein G B1 domain (GB1), purified protein was tagged with a paramagnetic lanthanide complex before injection into oocytes. This allowed the direct measurement of pseudocontact shifts (PCSs) and RDCs from a single set of 2D NMR experiments, which were later implemented in structure calculations (Müntener et al. 2016).

However, rather than complete *de novo* structure determination, the vast majority of in-cell NMR experiments have focused on comparing intracellular structure, dynamics and stability with observations made *in vitro*, delineating how the physical properties of the intracellular environment can cause structural changes (Serber and Dötsch 2001; Freedberg and Selenko 2014). Although changes in for example viscosity are modest, with estimates between 1.2-1.4 times that of water (Bicknese et al. 1993), macromolecular crowding effect and interactions with the cellular components

contribute most significantly to differences observed between *in vitro* and in-cell protein conformation, but also protein dynamics (on the pico- to nanosecond time-scale), as monitored by NOE measurements (Bryant et al. 2005).

A study on the IDP FlgM showed how proteins can adopt distinct conformations *in vitro* and *in vivo* (Dedmon et al. 2002). While in-cell HSQC spectra of uniformly ¹⁵N-labelled FlgM showed only resonances from the N-terminus, cross-peaks deriving from the C-terminus were not observable. This was in contrast to the *in vitro* NMR spectrum in which the entire protein sequence was detected, suggesting that the C-terminus had gained structure within cells and thereby become ‘invisible’ (section 4.3.3.1). This was later verified by further HSQC and circular dichroism (CD) experiments, the latter of which showed increased population of α -helical conformation on addition of 2.5 M glucose to mimic the crowded cellular environment (Dedmon et al. 2002).

Crowding reagents such as glucose, Ficoll, BSA and lysozyme have also been used in numerous other studies to recreate the effects of the cytosolic environment, as it enables easy and quantitative manipulation of macromolecular crowding effects without the technical constraints of in-cell NMR (Miklos et al. 2011). These crowders introduce varying levels of non-specific interactions and excluded volume effects, against which intracellular conditions can be compared to (Theillet et al. 2016). However, as the cytosol is inhomogeneous, consisting of ~4000 different proteins with different size, charge and other structural properties, a more appropriate approach might be the use of a ‘reconstituted cytosol’, in which *E. coli* cells are lysed, lyophilized and resuspended at different concentrations (Sarkar et al. 2013). Indeed, it was found that destabilization of the globular protein chymotrypsin inhibitor 2 (CI2) occurred to a greater extent in cytosol than in homogeneous protein crowders as a result of weak, non-specific interactions (Sarkar et al. 2013). Studies using such crowding agents and reconstituted cytosol have therefore enabled a reductionist approach to understanding the effect of the cellular environment, providing intermediate conditions between those in *in vitro* and in-cell studies but with greater control of the sample conditions and with higher sensitivity.

4.3 Technical aspects of in-cell NMR spectroscopy

The production of in-cell NMR samples is critical for acquiring high-quality and reliable in-cell spectra, and there are primarily two approaches: either by overexpression of a protein in cell cultures or by delivering (most commonly by injecting) purified protein into cells; and both techniques have proven successful in numerous in-cell studies.

4.3.1 In-cell sample preparation by overexpression

The most common method for producing in-cell samples, also used in this study, is by overexpressing the target protein using an inducible plasmid in bacteria (Sakakibara et al. 2009), yeast (Bertrand et al. 2012), insect (Hamatsu et al. 2013), or more recently human (HEK293T) cells (Barbieri et al. 2016) with appropriate isotopes. This method is also the most convenient as it is often the strategy to produce isotopically labeled samples for *in vitro* NMR samples but with omission of the purification steps (Sakakibara et al. 2009). Proteins are therefore advantageously studied in the same cells (and native environment) in which they were expressed without further manipulation of the sample. This also allows for the investigation of proteins that are difficult to purify or are unstable *in vitro*; for example, this has recently been used to study the aggregation-prone 43-KDa transactive response DNA binding protein (TDP-43) in cell extracts (A. Deckert & X. Wang, UCL).

As cells are cultured in media enriched with isotopes, high levels of background signal might be expected as all cellular components could become isotopically labelled. However, in practice, only a small number of distinguishable background resonances appear (Xu et al. 2014), the extent of which depends on the labelling strategy employed (section 4.3.3).

4.3.2 In-cell sample preparation by delivering purified proteins into cells

By delivering purified proteins into cells, background signals arising from isotopic labelling become negligible, with contributions arising only from natural abundance ^{13}C (1.1%). This approach typically involves injecting the target protein produced in *E. coli* into large, easily manipulated cells, such as *Xenopus laevis* oocytes. The process is tedious and produces inherent variability between cells, although automatic robotic injection has facilitated greater consistencies, in both localization and quantity, in producing such samples (Selenko et al. 2006). Alternative methods also include delivery into eukaryotic cells by cell-penetrating peptides (CPPs) covalently linked to the protein that are subsequently removed by endogenous enzymatic or autonomous reductive cleavage (Inomata et al. 2009); permeabilising cells by the bacterial toxin streptolysin O (SLO) and resealing by calcium (II) ions (Ogino et al. 2009); and electroporation of purified, isotopically-enriched into cultured mammalian cells (Theillet et al. 2016).

All these methods involve expression and purification of proteins before subsequent insertion into the cell. During sample preparation, the protein has therefore already undergone cellular processes such as folding, oxidization, co- and post-translational modifications and co-factor binding, complicating to some extent the

interpretation of their eventual study in foreign cells. Furthermore, a highly concentrated stock solution of the target protein is required: for example, to prepare oocyte samples with an intracellular concentration of 50 μM , the starting protein concentration must be 1 mM (Serber et al. 2007). These methods are therefore particularly unsuitable for proteins and complexes that are unstable, aggregate or misfold at high concentrations, including RNCs, which are maximally 10-15 μM and stable for only several hours when purified.

4.3.3 Selective isotopic labelling strategies

In addition to the method delivery of labelled proteins into living cells, the choice of isotopic labelling forms an important part of the experimental design. This is particularly the case in the current study where both the conformation of the ribosome-bound nascent chain (Cassaignau et al. 2016) and non-specific interactions with the cytosolic environment in cells (Xu et al. 2014) can place constraints on the detectability of the species under investigation, and these are discussed below.

4.3.3.1 Uniform ^{15}N -labelling

Uniform ^{15}N -labelling allows the study of backbone amide groups across a large dispersion of chemical shifts, as they are particularly sensitive to the local amino acid sequence. This labelling strategy is therefore particularly suited to resolving resonances deriving from disordered states including those that are ribosome-bound (Cabrita et al. 2016; Cassaignau et al. 2016). When applied to in-cell samples, uniform ^{15}N -labelling has however yielded mixed results. Although background labelling is only minimal, resulting in just a small number of signals from small molecule metabolites or peptides, both specific and non-specific intermolecular interactions with cytosolic species can occur. This results in faster transverse relaxation and significant broadening of resonances to such an extent that only background signals appear in spectra (Serber et al. 2001; Xu et al. 2014). While some proteins have been successfully detected in-cell by uniform ^{15}N -labelling, such as α -synuclein (Waudby et al. 2013a; Theillet et al. 2016) and TTHA1718 (Sakakibara et al. 2009), others (particularly globular proteins) such as ubiquitin (Xu et al. 2014) result in no discernable resonances and require alternative labelling for in-cell observation. The sensitivity can be improved by combining ^{15}N -labelling with deuteration to reduce the deleterious effects of transverse relaxation (section 4.3.3.3).

4.3.3.2 Uniform and selective ^{13}C -labelling

Although ^{13}C -labelling can provide more sensitive probes of tertiary structure formation, uniform ^{13}C -labelling of in-cell samples results in significant background signals due to

the greater number of C-H than N-H groups in cellular molecules (Serber et al. 2007). Significant reductions in background resonances and thus unambiguous assignment of protein residues can be achieved by selective labelling of methyl groups, for example the δ -methyl groups of isoleucines (Serber et al. 2004). Selective methyl labelling also offers large enhancements in NMR sensitivity due to the threefold degeneracy of methyl protons which effectively increases the concentration of each group (relative to, for example, the backbone amides), and because methyl groups are localised at the peripheries of dynamic side chains with more favourable (slower) relaxation (Tugarinov and Kay 2005). Further sensitivity enhancements are possible when combined with the methyl-TROSY effect, which involves the cancellation of neighbouring intra-methyl ^1H - ^1H and ^1H - ^{13}C dipole-dipole relaxation interactions resulting in slower relaxation (Tugarinov et al. 2003). Using this selective labelling strategy, ribosome-bound FLN5 in folded conformations have been successfully observed, despite some restrictions on mobility as a result of tethering to the ribosome (Cabrita et al. 2016).

4.3.3.3 Deuteration

Selective methyl labelling, as described above (section 4.3.3.2), relies on uniform deuteration with selective protonation at specific sites (in this case, certain methyl groups). By replacing protons with deuterons, the relaxation properties of the remaining protons improve as dipolar relaxation pathways are largely eliminated (Sattler and Fesik 1996). Deuteration also reduces the dipolar interactions between for example ^{13}C or ^{15}N and their directly bound ^1H . As the gyromagnetic ratio of ^2H is ~ 6.5 times smaller than that of ^1H , the relaxation rates of the heteronuclei are therefore scaled to $(\gamma_{\text{D}}/\gamma_{\text{H}})^2 \sim 0.02$, resulting in narrower line widths and increased sensitivity (Sattler and Fesik 1996). Deuteration has therefore permitted the study of larger systems *in vitro*, which suffer from increased relaxation rates caused by slower overall tumbling in solution. In-cell applications of deuteration to for instance ^{15}N -labelled calmodulin, have also detected improvements in resolution and sensitivity after sufficient time was apportioned to back-exchange labile amides with protons (Xu et al. 2014).

4.3.3.4 ^{19}F -Labelling

As the NMR-active (spin $\frac{1}{2}$) ^{19}F is not naturally occurring but can be detected with high sensitivity (83% of ^1H), ^{19}F labelling is an attractive option to obtain in-cell spectra with minimal background signal, and spectra in only 1D are readily interpretable (Li et al. 2010). However, incorporation of ^{19}F into (aromatic) residues by fluorinated analogues can change the chemical properties of the molecules under investigation (Brindle et al.

1989), and preliminary spectra of ^{19}F -labelled RNCs *in vitro* (acquired by J. Paton, UCL) have proven challenging to interpret.

4.3.4 Further technical aspects of in-cell NMR experiments

An essential aspect of in-cell NMR is to ensure that the resonances observed derive exclusively from intracellular species. While many studies employ a strategy in which in-cell samples are compared with those obtained from supernatants after centrifugation of the sample (Sakakibara et al. 2009; Theillet et al. 2016), an alternative and non-invasive method involves the use of NMR diffusion measurements. By exploiting the confinement of species within cells and the consequent restriction of translational diffusion, pulse-field gradients (with a 300 ms diffusion delay) can be used to quantify or exclusively select for intracellular signals (Waudby et al. 2012).

A final but equally crucial quality control step for in-cell samples is ensuring that the cells are alive during data acquisition. In-cell samples are typically stable for several hours depending on the type of cells and experimental conditions used, before depletion of nutrients and/or oxygen, and the build-up of acidic metabolic products, induce cell death (Barbieri et al. 2016). The viability of *E. coli* cells can most easily be verified by spreading dilutions of the NMR sample before and after data acquisition onto LB agar plates and counting the number of colonies formed (Serber et al. 2007).

4.3.5 From *in vitro* to in-cell NMR studies of RNCs

With similar spectroscopy challenges to in-cell NMR (Sakakibara et al. 2009; Robinson et al. 2012), much of the strategy employed for the *in vitro* studies of RNCs by NMR spectroscopy is likely to be highly amenable to a transition to in-cell studies. In particular, both approaches require the large-scale production of selectively labelled RNCs in *E. coli*, and the use of rapid acquisition, sensitivity-enhanced pulse sequences in combination with continuous quality controls of the sample (Cassaignau et al. 2016). We therefore sought to develop a strategy to study RNCs by in-cell NMR spectroscopy.

In addition to the short lifetimes of in-cell samples, which require constant monitoring of cell leakage and viability (section 4.3.4), in-cell NMR samples of RNC are likely to be further complicated by the inherent instability of the ribosomal complexes. Release of the nascent chain via spontaneous translocation (Fredrick and Noller 2003) or the action of trace amounts of proteases results in degradation of purified samples of RNCs (Cassaignau et al. 2016); the presence of additional species such as release factors (Tsai et al. 2014) that are removed during rigorous purification steps are likely to additionally contribute to the release of ribosome-bound nascent chains within living in-cell samples. The key challenge in studying RNCs by NMR is therefore ensuring that the

observed resonances originate from ribosome-bound rather than free nascent polypeptides. While the large discrepancy in diffusion rates between these two states can be exploited *in vitro* to monitor RNC stability (Chapter 3), confinement in *E. coli* cells prevents any measurement of translational diffusion to make this method amenable *in vivo*. An alternative strategy to monitor RNC occupancy must therefore involve exploiting other differences in the properties of released and bound nascent chains.

In this Chapter, we describe how this could be achieved using FLN5+31 RNC as a model in-cell system. As discussed in previous Chapters, the FLN5+31 RNC consists of the complete FLN5 sequence linked to a 31-residue sequence derived from FLN6 and the SecM motif, and is in a disordered conformation despite FLN5 folding in isolation (Cabrita et al. 2016). Analysis of residue-specific broadening of NMR resonances of FLN5+31 RNC suggested specific interactions between the nascent chain (particularly the C-terminal residues required for complete FLN5 folding) and the ribosomal surface, which prevented the folding of the nascent chain FLN5 domain (Cabrita et al. 2016). We therefore hypothesized that upon release and in absence of the ribosome that the nascent chain would change into a folded conformation (both *in vitro* and *in vivo*). This conformational change would therefore result in concomitant and significant changes in the resonances observed in NMR spectra with respect to those of the disordered RNC, and offering a possible strategy in discriminating between ribosome-bound and released nascent chains. To begin to understand the physiological implications of the cellular environment to co-translational protein folding, we also describe experiments in reconstituted cytosol (section 4.2) that allow greater sensitivity and resolution of NMR spectra. We show how the structure of disordered FLN5 is perturbed by the cytosolic environment, which is likely to have implications to the conformation of FLN5+31 *in vivo*.

4.4 Results

4.4.1 Developing a strategy to observe RNCs in living cells by in-cell NMR

4.4.1.1 In-cell NMR sample preparation and quality control

In-cell NMR samples were produced by initially growing *E. coli* in unlabelled rich media before expression was induced in minimal media containing the appropriate isotopes (Figure 4.1). Although switching media between the growth and protein expression phase does not reduce background signals in the eventual NMR spectra, increased sensitivity has previously been observed using this strategy (Serber et al. 2001); this is likely due to higher protein expression levels as a result of replenishing nutrients and removing metabolic products. In-cell samples were prepared as a 50% (w/v) cell slurry

in fresh unlabelled minimal media supplemented with additional glucose (at a final concentration of 2% (w/v)) to compensate for faster glucose consumption because of the high cell density ($OD_{600nm} > 100$) of the NMR sample (Barbieri et al. 2016).

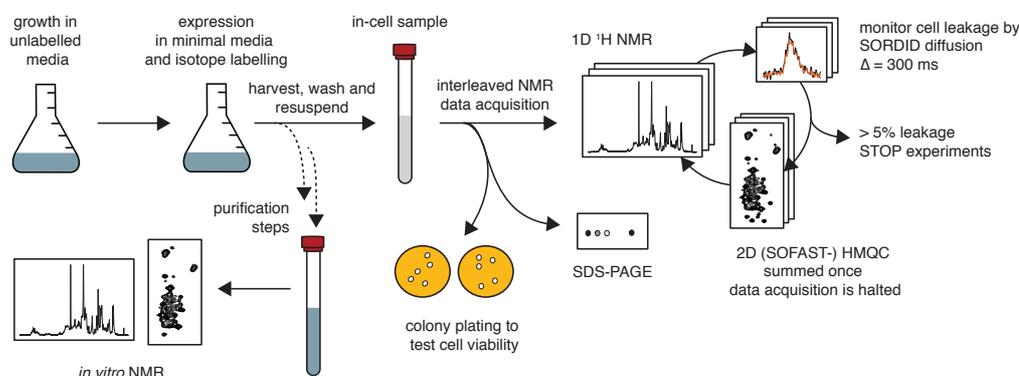


Figure 4.1 General scheme of in-cell NMR. Cells are initially grown in unlabelled rich media before expression and isotopic labelling in minimal media. Whereas *in vitro* NMR involves purification of the protein, in-cell NMR requires only washing of the cells to remove extracellular species and allow hydrogen back-exchange (where required) before immediate NMR data acquisition. NMR experiments are recorded in an interleaved manner such that cell leakage is continuously monitored by 1D SORDID diffusion measurements, and experiments are halted when >5% of protein is extracellular. The in-cell sample is also checked by colony plating to test cell viability and biochemically by SDS-PAGE.

To ensure that the NMR resonances observed derived exclusively from intracellular species, cells were washed several times after harvesting with unlabelled minimal media, and experiments were recorded in an interleaved manner, in which a series of short 2D correlation spectra were acquired between rapid 1D SORDID diffusion measurements recorded using a 300 ms diffusion delay to monitor protein leakage (section 4.3.4, Figure 4.1). The 2D data were collected until the fraction of extracellular protein exceeded 5%, and then combined to generate a final data set. The viability of the bacterial cells in the in-cell sample was tested by plating colony tests before data acquisition, and showed no discernable difference when repeated after typically 12 h of NMR experimental time at 10°C.

4.4.1.2 In-cell NMR of ^{15}N -labelled FLN5 Δ 12

We initially began by acquiring an in-cell NMR spectrum of uniformly ^{15}N -labelled FLN5 Δ 12, an unfolded variant of FLN5 that in previous studies (Cabrita et al. 2016) was found to give rise to NMR resonances closely matching those of the FLN5+31 RNC. By acquiring 2D ^1H - ^{15}N correlation spectra using the sensitive SOFAST-HMQC experiment (Figure 4.2a), well-resolved resonances were observed that could unambiguously be

attributed to intracellular FLN5 Δ 12 when overlaid with the same spectrum obtained for FLN5 Δ 12 *in vitro* (Figure 4.2b). Additional resonances were observed in the in-cell spectrum that likely derived from small molecules such as metabolites, peptides or free amino acids as these signals were approximately 8-fold narrower in line width and up to ~200-fold higher in intensity than those of FLN5 Δ 12. These cross-peaks have also previously been observed in other in-cell NMR studies (Serber et al. 2001; Xu et al. 2014). Nevertheless, the in-cell NMR spectrum shows that FLN5 Δ 12 remains in a similar conformation *in vivo* (compared with *in vitro*), and indeed that unfolded conformations of FLN5 do not interact significantly with the cellular environment so as to be broadened beyond detection. We did observe broadening of residues, attributable to the magnetically inhomogeneous cellular sample (Waudby et al. 2013a), but more significant broadening of at least two resonances (714G, 716G), which were also observed for RNCs *in vitro*, indicated residue-specific interactions with cellular components (see also section 4.4.2). Nevertheless, well-resolved and unambiguous resonances could be detected in-cell for the remaining residues by uniform ^{15}N -labelled, in contrast to other proteins such as ubiquitin that become ‘invisible’ in-cell (Xu et al. 2014). If FLN5+31 RNC adopts the same disordered conformation inside the cell as *in vitro*, a similar spectrum might therefore be expected.

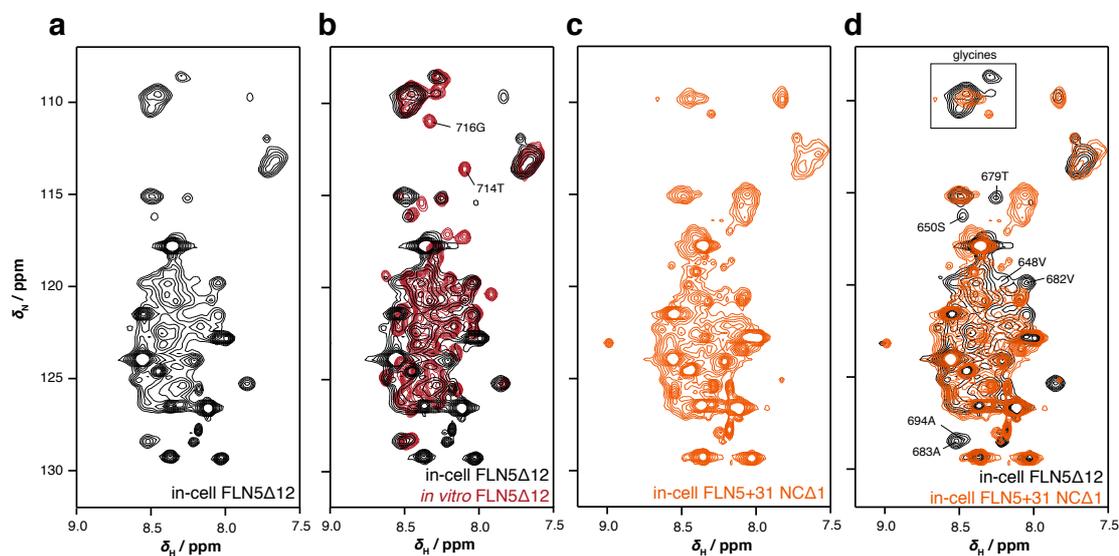


Figure 4.2 In-cell NMR of isolated FLN5 constructs. 2D ^1H - ^{15}N SOFAST-HMQC spectrum (a) of in-cell FLN5 Δ 12, and (b) overlay with the spectrum obtained *in vitro*. 2D ^1H - ^{15}N SOFAST-HMQC spectrum, highlighting resonances that were significantly broadened *in vivo* (c) of in-cell of FLN5+31 NCA1, and (d) overlay with in-cell FLN5 Δ 12 (orange) highlighting FLN5 Δ 12 resonances that are distinct from FLN5+31 NCA1 (black).

4.4.1.3 In-cell NMR characterisation of ^{15}N -labelled FLN5+31 NCA1

A key question in this work is how we can distinguish intracellular nascent chains that are ribosome-bound from those that are released from the ribosome by NMR spectroscopy. To address this, we truncated the FLN5+31 RNC sequence by one residue, removing the C-terminal proline of the SecM stalling sequence (P166). Within SecM-stalled ribosomes, P166-tRNA remains at the A-site without bond formation with G165-tRNA at the P-site. The proline is crucial to stalling due to its intrinsically slow peptide bond formation, and its removal results in complete relief of translational stalling (Nakatogawa and Ito 2002). This C-terminal truncation, termed FLN5+31 NCA1, is therefore an excellent mimic of ribosome-released FLN5+31 nascent chains.

Producing an in-cell sample using the same conditions employed for FLN5 Δ 12, we acquired an in-cell 2D ^1H - ^{15}N SOFAST-HMQC spectrum of FLN5+31 NCA1 (Figure 4.2c). In addition to the same background signals previously observed (section 4.4.1.2), the spectrum showed resonances that were narrowly dispersed within the direct dimension, indicating a degree of disorder within FLN5+31 NCA1. We attributed this to the 30-residue linker consisting of N-terminal residues of FLN6 and SecM (omitting the C-terminal proline). When overlaid with the in-cell spectrum of FLN5 Δ 12 (Figure 4.2d), we found that despite the narrow dispersion of both spectra, a total of at least 10 resonances could be identified that did not overlap with those of FLN5 Δ 12. These predominantly arose from glycine residues due to their distinct chemical shift in the ^{15}N dimension, but also from other amino acids (Figure 4.2d). These data showed that FLN5+31 nascent chains released from the ribosome give rise to a different ^1H - ^{15}N correlation spectrum than that of disordered FLN5 species.

The in-cell NMR spectrum of FLN5+31 NCA1 (Figure 4.2c) did not show any disordered FLN5 resonances and only a small total number of cross-peaks were observable. We therefore hypothesized that the FLN5 domain was folded (as expected initially, section 4.3.5) and ‘invisible’ by ^{15}N -labelling due to rapid transverse relaxation caused by non-specific interactions with cellular components, as seen with previous in-cell studies of full-length FLN5 (Waudby et al. 2012). To investigate this further, we lysed the in-cell sample by freeze-thawing several times and repeated the NMR experiments. The 2D ^1H - ^{15}N SOFAST-HMQC spectrum of the lysed sample (Figure 4.3b) showed significant reduction in line widths of all resonances and remarkably, a large number of additional cross-peaks that were unobservable in the in-cell spectrum (Figure 4.3c). These cross-peaks were well-dispersed across the proton dimension, and because all the resonances of the *in vitro*, folded FLN5 domain could easily be mapped onto the spectrum (Figure 4.3f), this demonstrated that FLN5 was indeed in a native-like folded

conformation in FLN5+31 NCA1. Most of the remaining cross-peaks that did not match those of *in vitro* FLN5 were present in the in-cell spectrum, and could also be assigned to residues of FLN6 (Figure 4.3d). Finally, resonances that could not be assigned to either FLN5 or FLN6 could instead be assigned to either small molecule metabolites (section 4.4.1.2) or residues of the disordered SecM stalling sequence. The latter was evidenced by the presence of a strongly deshielded cross-peak, characteristic of a tryptophan residue, and an additional resonance within the ^{15}N -upshifted chemical shift range (~ 110 ppm) of disordered glycine residues (Figure 4.3d).

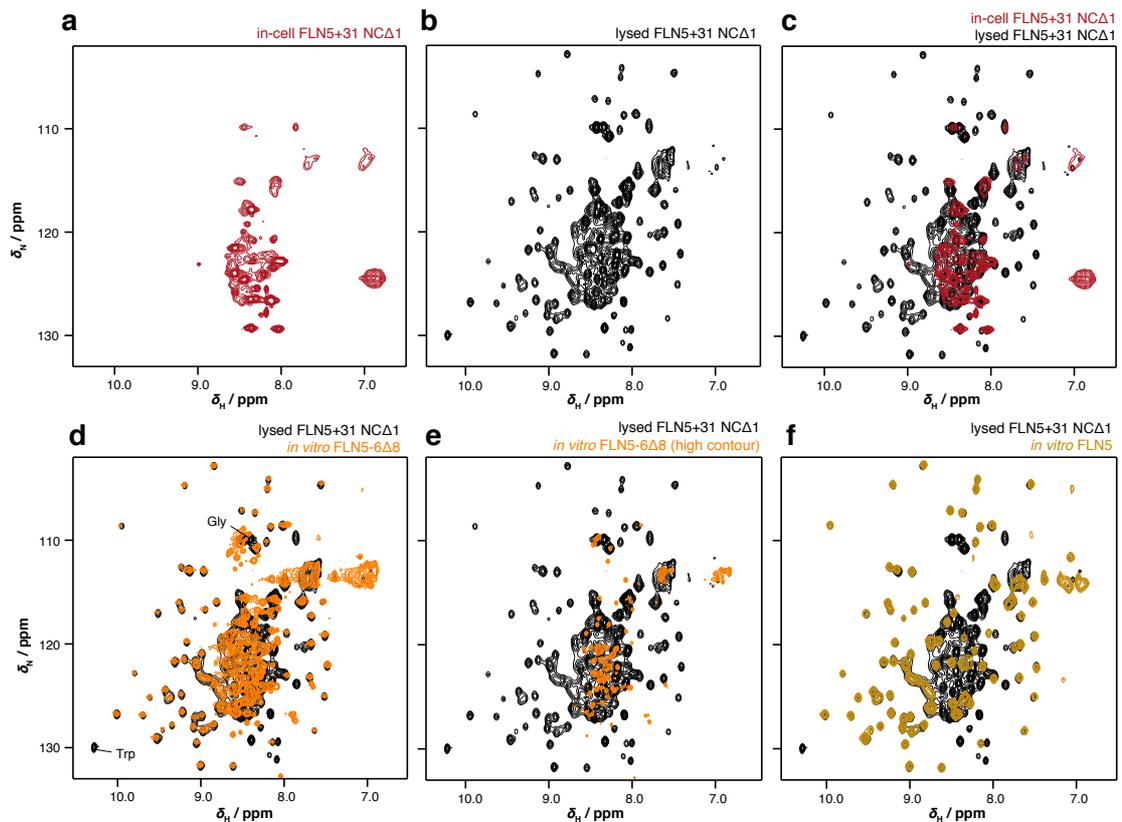


Figure 4.3 NMR characterisation of FLN5+31 NCA1. 2D ^1H - ^{15}N SOFAST-HMQC spectrum of (a) in-cell +31 NCA1; and of (b) in-cell FLN5+31 NCA1 following cell lysis, overlaid with (c) the spectrum obtained prior cell lysis, the *in vitro* spectrum of FLN5-6 Δ 8 at (d) low and (e) high contour levels to show the disordered regions, and (f) the *in vitro* spectrum of FLN5. Resonances attributed to the tryptophan and glycine residues in the SecM sequence are indicated in (d).

4.4.1.4 In-cell NMR of folded FLN5 states by selective methyl labelling

As we were unable to detect intracellular folded conformations of FLN5 by uniform ^{15}N -labelling, we prepared deuterated in-cell samples of FLN5+31 NCA1 in which the isoleucine C δ methyl groups were selectively $^{13}\text{CH}_3$ -labelled. By exploiting the methyl-TROSY effect intrinsic to ^1H - ^{13}C HMQC experiments to reduce transverse relaxation

effects (section 4.3.3.2), we were able to observe at least two isoleucine resonances (I674 and I695) dispersed from the background signal and with identical chemical shifts to those of FLN5 *in vitro* (Figure 4.4a). Following repeated freeze-thaw cycles, the 2D ^1H - ^{13}C HMQC spectrum of the resulting lysed cell sample (Figure 4.4b) showed resonances with reduced line-widths, enabling clear observation of four FLN5 isoleucine cross-peaks. These experiments confirmed that on release from the ribosome, FLN5+31 nascent chains adopt a native-like folded FLN5 domain that is observable in-cell by selective methyl labelling.

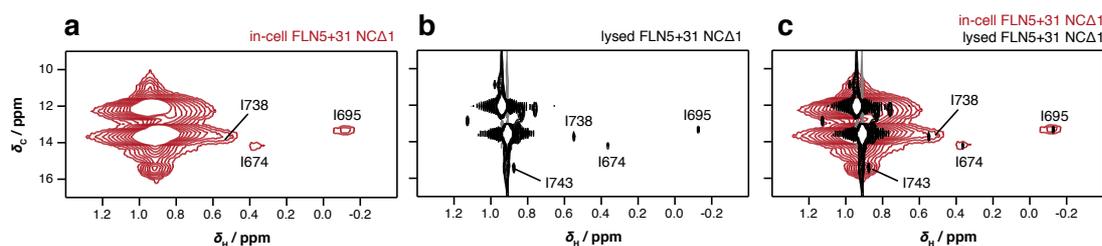


Figure 4.4 In-cell NMR of Ile- δ 1- $^{13}\text{CH}_3$ -labelled FLN5+31 NCA1. 2D ^1H - ^{13}C HMQC spectrum of (g) in-cell FLN5+31 NCA1; and of (h) in-cell FLN5+31 NCA1 following cell lysis, overlaid with (i) the spectrum obtained prior cell lysis.

In summary, the experiments of FLN5+31 NCA1 (Figure 4.3 and Figure 4.4) showed that the FLN5 domain folds on release of the ribosome leaving only the linker residues disordered. These in-cell spectra serve as references as to how we might expect the in-cell FLN5+31 RNC sample to appear when released from the ribosome. If alternative resonances appear in the in-cell FLN5+31 RNC spectrum, these should therefore derive from ribosome-bound species. If the FLN5+31 RNC remains in the same conformation as *in vitro*, we would therefore observe an in-cell spectrum closely matching that of FLN5 Δ 12.

4.4.1.5 Production of perdeuterated in-cell samples expressed at high cell densities

As previously discussed, a significant spectroscopic challenge in working with macromolecular complexes such as RNCs by NMR spectroscopy is the large molecular weight of the ribosome resulting in slow tumbling and thus fast transverse relaxation. As we have already observed in the case of the globular FLN5 domain within FLN5+31 NCA1, the increased viscosity of the cytosol (approximately twice that of water (Brindle et al. 1989)) and transient interactions with cellular components further contribute to such transverse relaxation effects in-cell, significantly reducing the possibility of

observing in-cell resonances of RNCs. However, we were able to reduce the result of such T_2 -associated effects by selective methyl labelling within a deuterated background (section 4.4.1.4), so we began to explore how deuteration could also increase the sensitivity of uniformly ^{15}N -labelled in-cells samples for the study of disordered (RNC) conformations.

The sample requirements for in-cell NMR are similar to those of *in vitro* NMR studies of RNCs, as both require large quantities of selectively labelled species within a dense background (the ribosome or cellular environment). Using a similar strategy as for the production of deuterated, selectively methyl-labelled RNCs *in vitro*, we initially tested the sensitivity enhancements of ^{15}N -labelled isolated FLN5 Δ 12 in perdeuterated cells. We acquired in-cell 2D ^1H - ^{15}N SOFAST-HMQC spectra (Figure 4.4b-f) and observed significant improvements in both sensitivity and resolution compared to a non-deuterated in-cell sample that had similar expression levels (Figure 4.4a). In particular, by taking seven well-dispersed resonances (Figure 4.4b-c), we found a reduction in mean average line-widths from 89.8 to 70.6 Hz and a mean residue-specific sensitivity increase of 3.0 ± 0.9 (SD); similar enhancements in sensitivity by deuteration were previously observed in studies of NmerA within *E coli* cells grown in an algae extract-derived rich media (Serber et al. 2001). As all the expected resonances were observed in the spectra, this indicated that the washing step in protonated media prior to data acquisition was sufficient for back-exchange to occur within deuterated cells that had expressed FLN5 Δ 12.

As an aside, we reliably yielded ~ 200 mg per 25 mL cell culture using this high cell density growth method for deuteration. This small culture size was sufficient for an in-cell NMR sample and we were therefore able to limit the costs of deuteration (as this can often restricts its application in protein expression and in-cell NMR).

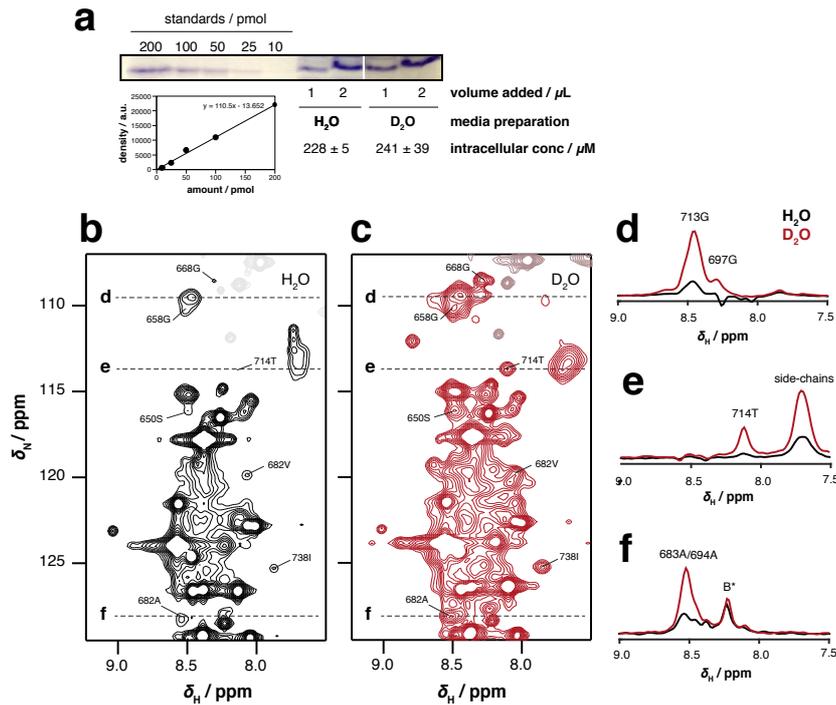


Figure 4.5 Deuteration of uniformly ^{15}N -labelled samples expressed in *E. coli* at high densities. (a) Estimations of expression levels of FLN5 Δ 12 in cells grown in MDG/EM9 prepared with $^1\text{H}_2\text{O}$ and $^2\text{H}_2\text{O}$, by densitometric analysis of coomassie-stained SDS-PAGE using a standard curve from known amounts of purified FLN5 Δ 12. 2D ^1H - ^{15}N SOFAST-HMQC of in-cell FLN5 Δ 12 expressed in MDG prepared with (b) $^1\text{H}_2\text{O}$, (c) $^2\text{H}_2\text{O}$, normalised for concentration of FLN5 Δ 12 as determined by densitometric analysis; dashed lines indicate locations of one-dimensional slices as shown in (d)-(f), labelled with their assignments (B^* , background signal) and highlighted resonances were used to determine mean residue-specific sensitivity enhancements and line-width reductions.

4.4.1.6 *In vivo* observations of RNCs by in-cell NMR

With reference spectra of ribosome-released nascent chains at hand (sections 4.4.1.3 and 4.4.1.4), together with NMR sensitivity-enhanced sample preparations to observe both intracellular disordered and folded proteins, we progressed to producing in-cell samples of the FLN5+31 RNC. We initially recorded a 2D ^1H - ^{15}N SOFAST-HMQC experiment of a uniformly ^{15}N -labelled FLN5+31 RNC in perdeuterated *E. coli* cells (Figure 4.6a). We observed resonances indicative of an unfolded conformation, but sensitivity was significantly reduced compared to spectra of isolated proteins likely due to reduced tumbling of a ribosome-bound species and the reduced protein concentration. We once again attributed the broad linewidths of the in-cell resonances to the magnetically inhomogeneity of the dense cellular sample (Waudby et al. 2013a). The most resolved region of the spectrum was within the chemical shift range of glycine residues, and was not obscured by background resonances deriving from other intracellular species were also found in the same region of the spectrum (and also at low contour levels in in-cell

samples of the isolated proteins) (Figure 4.6a). To ascertain whether the observed resonances derived from ribosome-bound species, we compared the chemical shifts of the glycine residues with those of FLN5+31 N Δ 1 (Figure 4.6f-g). The resonances of FLN5+31 N Δ 1 from lysed cells did not overlap well with those of the in-cell spectrum of FLN5 +31 RNC, suggesting that a population of released nascent chains was not observable and that instead, the cross-peaks derived from intracellular ribosome-bound species. Indeed comparisons with the disordered FLN5 variant FLN5 Δ 12 (Figure 4.6d-e) showed better overlap of the glycine resonances, particularly of 658G, 660G, 681G and 687G. Similarly, these four glycine cross-peaks overlapped with those of purified RNC *in vitro* (Figure 4.6c). Resonances of 700G and 697G were well-resolved in the *in vitro* spectra of disordered FLN5 and RNC, but undetectable in the in-cell RNC spectrum (Figure 4.6c). It is interesting to note that these two residues are closer towards the C-terminus of the nascent chain than the four observable glycines. This is perhaps further evidence of a ribosome-bound species, as C-terminal residues are more restricted in mobility due to tethering to the ribosome (possessing reduced rotational correlation times) resulting in faster transverse relaxation, and this is exacerbated *in vivo*. We attribute the small chemical shift changes to $^1\text{H}/^2\text{H}$ isotope effects, which typically result in upfield shifted ^{15}N chemical shifts of < 0.1 ppm for glycines (Jaravine et al. 2004), consistent with our observations.

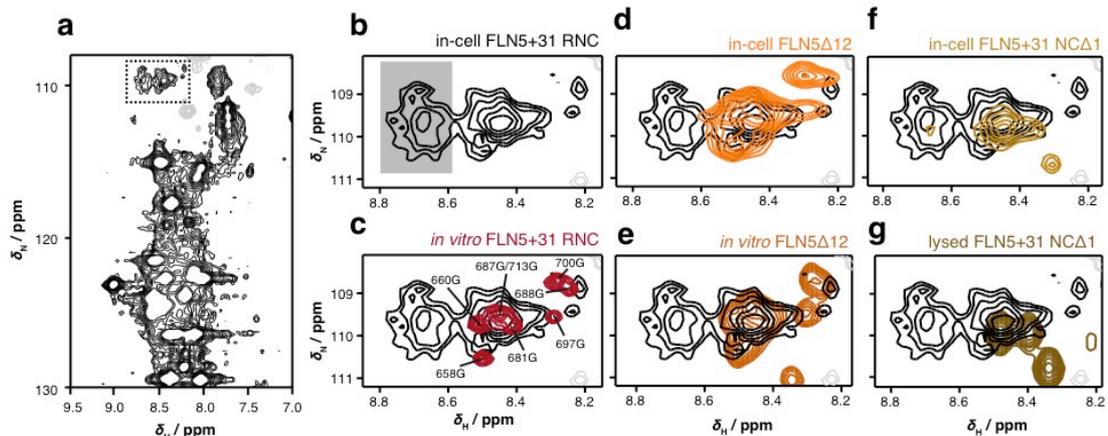


Figure 4.6 Identification of in-cell FLN5+31 RNC NMR resonances. (a) 2D ^1H - ^{15}N SOFAST-HMQC spectrum of in-cell FLN5+31 RNC, box indicates glycine region and location of spectra b-g. Grey-shaded box in b indicates background resonances also found in other in-cell spectra at low contour levels. Overlays of 2D ^1H - ^{15}N SOFAST-HMQC spectrum of (b) in-cell FLN5+31 RNC with (c) the *in vitro* spectrum of FLN5+31 RNC (acquired at 950 MHz by A. Cassaignau, UCL); (d) with in-cell FLN5 Δ 12; (e) with *in vitro* FLN5 Δ 12; with FLN5+31 N Δ 1, (f) in-cell and (g) following cell lysis.

4.4.2 Structural insights of FLN5 Δ 12 in living cells, reconstituted cytosol and crowding agents

To gain insights into how the cellular environment affects the structure of disordered FLN5 at high resolution and in more controlled conditions, we returned to *in vitro* NMR studies using various crowding agents and then validated these results by comparing with in-cell NMR data. Firstly, we produced unlabelled reconstituted cytosol as a mimic of the cytoplasm in an approach similar to one previously described (Sarkar et al. 2013). Cells were grown by the same method used for in-cell samples, then lysed by French press, lyophilized and resuspended at different concentrations. The fraction (by mass) of protein in the reconstituted cytosol was determined by bicinchoninic acid (BCA) assay and found to be $65.4 \pm 7.5\%$, consistent with the protein content of *E. coli* cells (Pedersen 1978). We then titrated 25, 37 and 50 g dry weight/L of reconstituted cytosol into NMR samples of FLN5 Δ 12, and recorded 2D ^1H - ^{15}N SOFAST-HMQC experiments. Although we only detected small chemical shift changes ($\delta_{\text{NH,average}} = 0.021 \pm 0.011$ (S.D.) ppm in 50 g/L lysate), we observed progressive reduction in NMR signal intensity on addition of increasing cytosolic material (Figure 4.7a), particularly in the regions of the A', B, C and E strands of the folded FLN5 sequence. We found a similar pattern of residue-specific broadening in the 2D ^1H - ^{15}N SOFAST-HMQC spectrum of the in-cell sample of FLN5 Δ 12 both prior and following cell lysis (Figure 4.7e), the latter of which enabled higher resolution spectra to be obtained. These results validated the use of reconstituted cytosol at 25-50 g/L as an appropriate reconstructed cellular environment.

The non-uniformity of the NMR signal reductions across the amino acid sequence prompted us to investigate further the structural changes of FLN5 Δ 12 in the cytosolic environment. We repeated the ^1H - ^{15}N SOFAST-HMQC experiments using the artificial crowding agents BSA (Figure 4.7c) and Ficoll (Figure 4.7d). Although both crowders induced signal broadening of FLN5 Δ 12 mirroring the general pattern observed in cytosol, the signal reductions induced by BSA were more pronounced, particularly in the A', B and C strand regions, and was thus more reflective of the cytosol data. Whereas Ficoll introduces predominantly excluded volume effects due to its relative inert nature, the globular protein BSA introduces additional non-specific interaction forces (Wang et al. 2012; Seeliger et al. 2013). Analysis of our data therefore suggested that the reduction of SOFAST-HMQC signal intensities was the result of mostly excluded volume effects combined with, to a lesser extent, non-specific interactions.

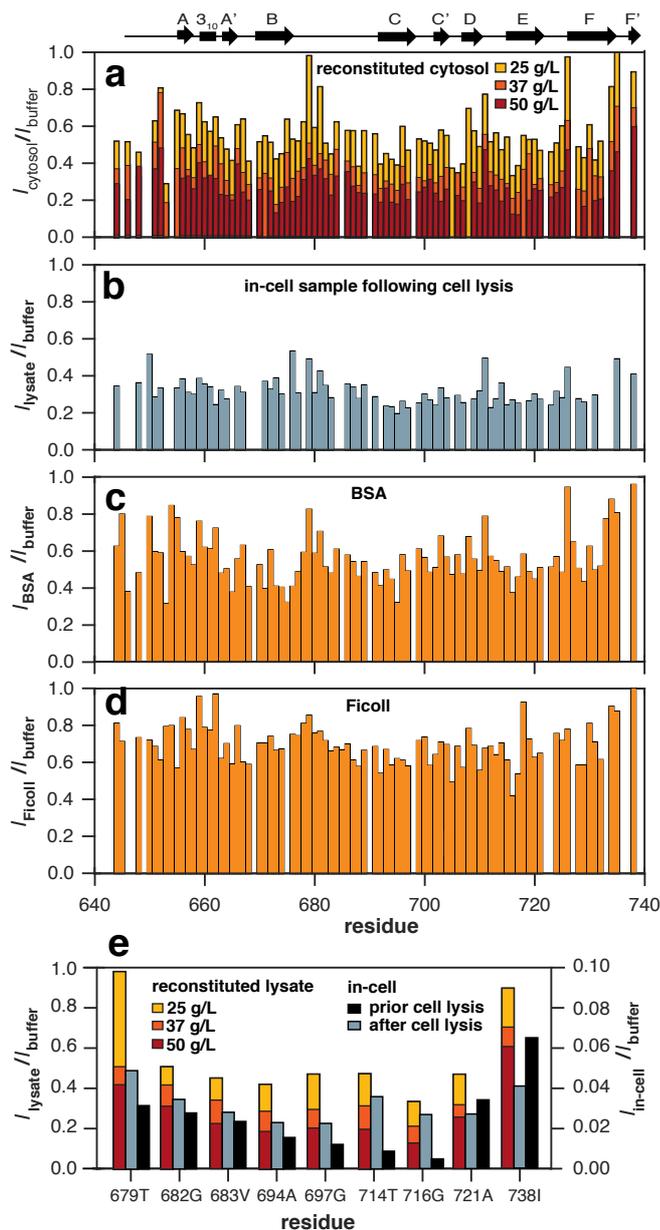


Figure 4.7 Residue-specific analysis of NMR signal intensities of FLN5 Δ 12 in cells and crowded environments. (a) Shown for comparison, amide chemical shift differences between folded, full-length FLN5 and unfolded FLN5 Δ 12, with schematic above showing secondary structure elements of full-length FLN5. Relative intensities of ^1H - ^{15}N SOFAST-HMQC resonances of FLN5 Δ 12 (b) in reconstituted cytosol at 25 (yellow), 37 (orange) and 50 g dry weight/L (red); (c) in-cell sample following cell lysis; (d) in 200 g/L BSA; and (e) in 200 g/L Ficoll. (f) Relative intensities of ^1H - ^{15}N SOFAST-HMQC resonances of FLN5 Δ 12 that could unambiguously be assigned in the in-cell spectrum; relative intensities in reconstituted lysate (at 25 (yellow), 37 (orange), 50 g dry weight/L (red)) and in-cell prior (black) and after cell lysis (blue). Note that the scale for in-cell relative signal intensities is magnified ten-fold.

Such macromolecular crowding effects have been suggested to perturb folding equilibria (Ådén and Wittung-Stafshede 2014); for example, in-cell studies have shown that the folding equilibrium of disordered FlgM is shifted by interactions with cytosolic components such that folded conformations are populated *in vivo* (Dedmon et al. 2002). However, in this case, the change in structure was likely to be small because only minor chemical shift changes were observed. Nevertheless, these data suggested there could be structural implications in the ribosome-bound analogue, which could be resolved in future applications of reconstituted cytosol to NMR samples of FLN5+31 RNC. We also propose that the approaches detailed in Chapter 3 to determine secondary structure populations of nascent chains is likely to be highly beneficial in these studies.

4.5 Discussion

4.5.1 FLN5+31 RNC remains disordered *in vivo* despite crowding effects

We have developed a sensitivity-enhanced in-cell NMR approach to observing different intracellular FLN5 conformations using a perdeuterated, dual labelling strategy. The use of ^{15}N -labelling permitted *in vivo* observations of disordered FLN5 species (Figure 4.2a) and much of the flexible, disordered 30-residue FLN6/SecM linker (Figure 4.2c), while selective labelling of the isoleucine δ 1-methyl groups in combination with the methyl-TROSY effect reduced transverse relaxation effects to enable detection of FLN5 domain in a folded conformation (Figure 4.4). Because of large conformational differences between ribosome-released and bound nascent chains, this allowed us to identify intracellular nascent chain resonances deriving from ribosome-bound species.

Although the full FLN5 domain is fully emerged from the ribosomal tunnel in the FLN5+31 RNC (Cabrita et al. 2016) and is able to fold in isolation upon release from the ribosome (sections 4.4.1.3 and 4.4.1.4), our in-cell NMR experiments (section 4.4.1.6) reveal that the RNC remains disordered in the cell. While macromolecular crowding effects induced by the cytosol or artificial crowding agents can perturb the structure and folding equilibrium of disordered FLN5 conformations (section 4.4.2), such forces are insufficient to completely fold the FLN5 domain while the nascent chain remains tethered to the ribosome. This suggests that despite the large number of different proteins and nucleic acids at high concentrations in the cell, the interactions between the FLN5 nascent chain and the ribosome that restrict folding are highly specific. However, experiments in reconstituted cytosol suggest that at least some structural changes within disordered FLN5 constructs can occur as a result of both excluded volume effects and non-specific interactions, although they have yet to be applied to RNCs. We have previously attempted to introduce RNCs into 'fresh' (non-lyophilised) lysate, but this resulted in partial release of the nascent chain from the RNC; we have more recently optimised the preparation of reconstituted cytosol and preliminary small-scale studies have shown that RNC breakdown is negligible on addition of the cytosol (data not shown). We also propose that, in addition to producing several more in-cell RNC samples (and summing together the spectra to generate a final data set) to increase sensitivity, detection of RNC by SDS-PAGE should be used in future investigations to verify these in-cell data, although we advise that precaution should be taken to minimise disruption of RNC integrity during cell lysis of such small volume samples.

4.5.2 Approaches to identifying ribosome-bound and released species

In this study, we have exploited the conformational change of a nascent chain upon release from the ribosome to identify intracellular ribosome-bound species, resolving a key challenge of working with RNCs by in-cell NMR. Indeed because released FLN5 nascent chains were predominantly in a folded conformation that was undetectable by ^{15}N -labelling during the complete time-course of data acquisition, this significantly reduced background resonances arising from released species in spectra of cells expressing RNC. In cases where conformational differences are not observed between ribosome-bound and released forms, such as the IDP α -synuclein (Deckert et al. 2016), the use of reconstituted cytosol as an environment intermediate between *in vitro* and in-cell is still likely to be highly beneficial in understanding the implications of the cellular milieu to co-translational protein folding (Sarkar et al. 2013).

4.5.3 Future perspectives

The use of STINT-NMR involves sequential, time-controlled expression of two proteins in a single bacterial cell (Burz et al. 2006). This would enable co-expression of another protein to monitor interactions with the nascent chain. For example, this might include the study of trigger factor, a ribosome-associated molecular chaperone which has previously been studied in *in vitro* NMR studies of RNCs (Deckert et al. 2016), in cells lacking the chaperone, enabling a more complete description of co-translational protein folding under physiological conditions. Furthermore, knowledge of the actions of trigger factor *in vivo* remains very limited, with only some evidence that trigger factor requires longer nascent chains for full recruitment in cells (Oh et al. 2011). In-cell NMR studies of RNCs and their interactions with trigger factor offer some prospects into shedding light onto *in vivo* mechanisms of these molecular chaperones (and others), and their influence in cellular protein folding.

In conclusion, we have shown how sensitivity enhancement by isotopic labelling strategies in combination with parallel in-cell studies of ribosome-released analogues enable the study of ribosome-nascent chain complexes in living cells. The *in vitro* approaches described in Chapter 4 to determine the conformations populated by RNCs are likely to be amenable to in-cell and reconstituted cytosol studies with continuing advances in NMR sensitivity. We believe the strategies described here will significantly progress the study of co-translational protein folding to ones under increasingly more physiologically relevant conditions, enabling a more comprehensive understanding of this fundamental biological process.

Chapter 5

Concluding remarks

While recent X-ray crystallography and cryo-EM studies have revealed detailed structural descriptions of the ribosome (Schmeing and Ramakrishnan 2009), comparatively little structural information is known on the dynamic nascent chain as it emerges from the ribosomal exit tunnel and begins to sample complex, native-like conformations. In this work, we have described strategies to study the structure of translationally stalled ribosome-nascent chain complexes, both *in vitro* and in living cells, by NMR spectroscopy. NMR is a high-resolution technique that is uniquely suited to studying such inherently flexible and dynamic structures, although its application to RNCs is limited by low spectroscopic sensitivity deriving from the slow tumbling of the ribosome, and the low maximum achievable concentrations and short lifetimes of purified samples (Waudby et al. 2013b).

In Chapter 2, we have described the use of the PLRE effect to increase the sensitivity of NMR measurements of RNCs. By combining longitudinal relaxation-optimised pulse sequences with the paramagnetic agent NiDO₂A, recovery times were significantly reduced, resulting in a 4.5-fold enhancement in the sensitivity of diffusion NMR measurements and 1.9-fold increase in SOFAST-HMQC sensitivity of a ribosome-bound nascent chain. Importantly, these sensitivity improvements did not measurably compromise RNC sample integrity or structure.

We exploited these significant PLRE-induced improvements in NMR sensitivity in Chapter 3 to explore more advanced NMR experiments that have previously eluded RNC studies. Triple resonance experiments, acquired using the longitudinal relaxation-optimised BEST series of experiments (Lescop et al. 2007), enabled chemical shift measurements of carbon nuclei C' and C^α of RNCs, and subsequent calculations of secondary structure populations of the nascent chain. As an orthogonal measure of RNC structure, we worked towards applying PLRE in residual dipolar coupling measurements, and found enhancements of up to 50% in sensitivity in both HSQC and TROSY experiments without T_2 -induced significant broadening of line widths, even in samples aligned with bacteriophage.

Finally, in Chapter 4, we have described a strategy to study RNCs within living cells. A key result was distinguishing ribosome-bound nascent chains from those

released from the ribosome in cells, by exploiting conformational (and thus spectroscopic) differences and isotopic labeling strategies, because confinement in the cell rendered diffusion NMR measurements, as probes of ribosome attachment, impossible. Using sensitivity enhancements gained by growing bacterial cells in deuterated media, we were therefore able to study RNCs *in vivo* at a residue-specific level.

The model RNC system used throughout this work was FLN5+31 RNC, an immunoglobulin domain linked by 31 residues (deriving from FLN6 and the SecM stalling sequence) from the PTC of the ribosome, but which remains in an unfolded conformation (Cabrita et al. 2016). Its disorder is particularly noteworthy as the complete FLN5 domain at 31 residues from the PTC is fully emerged from the ribosomal exit tunnel (as determined by PEGylation assays) and is able to fold in isolation from the ribosome even with the truncation of up to four C-terminal residues. In fact, a total of at least 42 additional linker residues to the FLN5 domain are required before a folding transition is observed (Cabrita et al. 2016). The folded state of the nascent chain is therefore likely to be destabilized on the ribosome (at least at short linker lengths due to steric effects), while resonance broadenings observed in NMR spectra of FLN5+31 (relative to signal intensities of disordered FLN5 isolated from the ribosome), attributed to transient interactions with the overall negatively charged ribosomal surface, likely result in stabilization of the unfolded state (Cabrita et al. 2016).

However, these conclusions do not preclude the presence of any residual structure with native-like contacts or local structure formation within FLN5+31 RNC. This is corroborated with small but systematic chemical shift changes, anti-correlated with resonance broadenings across the primary sequence (see Chapter 3). To investigate this further, we therefore sought to use triple resonance experiments to measure secondary populations of the nascent chain. However, the low spectroscopic sensitivity of FLN5+31 RNC limited our chemical shift measurements to only 66% of the expected resonances. The remaining resonances are expected to be more informative, with larger chemical shifts in the carbon dimension (concomitant with previously measured changes in ^{15}N and ^1H chemical shifts) corresponding to larger differences in secondary structure populations, relative to disordered FLN5 off the ribosome. Chemical shifts, and RDC measurements, can be readily used as either validations or restraints in MD simulations (Robustelli et al. 2010), and we anticipate that with continuing work, greater structural insights can be gained into this co-translational folding intermediate and therefore how the ribosome can modulate the folding landscape and accessible conformations of an emerging nascent chain.

Our initial observations of FLN5+31 RNC in living cells described in Chapter 4 showed that the nascent chain also adopted a disordered conformation on the ribosome *in vivo*. This suggests that transient interactions with the ribosomal surface, which actively prevent FLN5 from folding 31 residues from the PTC (as previously described), are preserved in the crowded cellular environment. Furthermore, insights from our NMR studies in crowded solutions indicate that disordered FLN5 experiences both excluded volume effects and transient interactions with cellular components. In-cell NMR of FLN5 off the ribosome (FLN5 Δ 12 and FLN5+31 N Δ 1) showed that interactions with the membrane were likely to be particularly prevalent because resonance broadening was significantly reduced upon lysis, an observation previously found in in-cell NMR of the disordered α -synuclein (Waudby et al. 2013a; Theillet et al. 2016). The in-cell stability of the nascent chain is therefore dependent on the balance of the opposing effects of repulsive excluded volume forces (which favors compact, folded conformations) and attractive interactions with the cytosol and the ribosome (which favor disordered states) (Wang et al. 2012).

The most immediate question for future studies is therefore how such effects contribute to the overall co-translational folding of FLN5 *in vivo*. While the disordered conformation of FLN5+31 in-cell indicates a folding 'offset' on the ribosome *in vivo*, do additional factors in the cell shift the folding mid-point of FLN5 RNCs relative to that in buffer, to either more compact (by excluded volume) or expanded states (by cellular interactions)? The in-cell NMR strategy developed in this work enables selective observation of only disordered states using uniform ^{15}N -labelling, and can therefore be readily applied to FLN5 RNCs of varying linker lengths. By monitoring the signal intensities of the well-resolved glycine region of 2D correlation spectra, the population of disordered RNCs can therefore be measured across different lengths away from the PTC, in a strategy similar to which has been applied *in vitro* (Cabrita et al. 2016). In parallel, NMR studies of RNCs in crowded solutions (as described for isolated FLN5 Δ 12 in Chapter 4) may also be of particular interest to delineate cellular effects on nascent chain folding. More specifically, the use of 3D NMR to measure ^{13}C chemical shift perturbations and calculate secondary structure populations on the ribosome in various crowded solutions could offer residue-specific insights into changes in structure. When combined with MD simulations (Robustelli et al. 2010), additional measurements such as of RDCs are likely to be of further benefit in elucidating structural perturbations. Three-dimensional NMR experiments and RDCs have already been measured in in-cell NMR of isolated proteins (Waudby et al. 2013a; Müntener et al. 2016), and with continuing improvements in spectroscopic sensitivity and sample preparations, the prospect of applying these

measurements to RNCs *in vivo*, rather than in reconstituted cytosol or crowding agents, is becoming increasingly more probable.

While the use of prokaryotic cells in in-cell NMR has many advantages, such as ease of handling and manipulation of cells to produce isotopically labelled RNCs (Cassaignau et al. 2016), there is growing interest in the use of eukaryotic cells for more meaningful, physiologically relevant conditions in which to study protein structure and dynamics. The most commonly used methods involve delivery of purified proteins into the cell by, for example, microinjection into large *X. laevis* oocytes (Selenko et al. 2006) or electroporation into mammalian cells (Theillet et al. 2016). However, because of the high concentrations of purified protein required ($>100 \mu\text{M}$), it is likely that instead, intracellular expression strategies for in-cell NMR of RNCs will be of greater interest. Indeed, this approach also enables the study of proteins in the same cellular environment in which they have been produced, and this has so far been applied to yeast (Bertrand et al. 2012), insect (Hamatsu et al. 2013), and human kidney cells (Barbieri et al. 2016). Nevertheless, the in-cell NMR approaches developed in this work to study RNCs *in vivo* are widely applicable to all cell types, and should therefore be able form the foundations of any future endeavors into alternative cell lines. With the ability to study *de novo* protein folding in increasingly physiologically relevant environments, we believe a more comprehensive understanding this fundamental biological process is now within closer reach.

Chapter 6

Materials and methods

All standard reagents were of the highest grade available and purchased from Sigma-Aldrich and Fischer Scientific, unless stated otherwise.

6.1 Molecular biology and biochemistry

6.1.1 Plasmids and bacterial strains

FLN5+31 RNC and isolated FLN5 protein constructs were cloned into the pLDC vector as previously described (Cabrita et al. 2009; Cabrita et al. 2016) (see Appendix) and include an N-terminal hexa-histidine purification tag. RNC constructs also consist of a C-terminal SecM stalling sequence in place of a stop codon (Gumbart et al. 2012). The FLN5 A₃A₃ mutants were generated by site-directed mutagenesis in which six phenylalanine and tyrosine residues were substituted with alanine residues: F665A, F672A, F675A, Y715A, Y719A and Y727A (A. Caissaignau *et al*, in preparation). The FLN5+31 NCA1 mutant was produced by substitution of the final SecM proline residue with a stop codon by site-directed mutagenesis from the corresponding RNC template DNA (section 6.1.4). Isolated α -synuclein was expressed from a pT7-7 vector, a kind gift from Prof Peter Lansbury (Harvard University).

All clones contained genes for ampicillin resistance and a T7 promoter enabling protein expression to be induced by isopropyl- β -D-1-thiogalactopyranoside (IPTG). All DNA and protein sequences can be found in the Appendices.

Cloning was performed using *E. coli* DH5 α (Invitrogen) competent cells. Expression of proteins, ribosomes and RNC was carried out in *E. coli* BL21-GOLD(DE3) competent cells (Stratagene). The bacterial strains were purchased or prepared in-house as described below (section 6.1.3).

6.1.2 Growth media and antibiotic

Filter-sterilised antibiotic, typically ampicillin (100 μ g/mL) was added to media and agar plates where required. For the production of ribosomes and competent cells, tetracycline (10 μ g/mL in 70% (v/v) ethanol) was added as the antibiotic.

Lysogeny broth (LB) media (5 g/L yeast extract, 10 g/L tryptone, 10g/L NaCl) was used for routine bacterial growths, such as starter cultures, transformations and

expression of unlabelled protein, and autoclaved prior use. For agar plates, the media was supplemented with agar (15 g/L).

For the production of competent cells, super optimal broth (SOB) media (5 g/L yeast extract, 20 g/L tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄) was used as the growth media, and autoclaved prior use.

For the expression of isotopically labelled isolated protein, minimal media (M9) was used as the growth media, with ¹⁵NH₄Cl and/or ¹³C-glucose added as the sole nitrogen and carbon sources respectively (Table 6.1).

Reagent	Final concentration	Sterilisation
Na ₂ HPO ₄ ·7H ₂ O	64 g/L	Autoclave
KH ₂ PO ₄	15 g/L	Autoclave
NaCl	5 g/L	Autoclave
MgSO ₄	2 mM	Filter
CaCl ₂	100 uM	Filter
100 X BME vitamins	1 X	Filter
D-glucose*	0.4% (w/v)	Autoclave
NH ₄ Cl†	0.1% (w/v)	Filter

Table 6.1 Composition of M9 media. *For uniform ¹³C-labelling, 0.2% (w/v) ¹³C-glucose was added. †For uniform ¹⁵N-labeling, ¹⁵NH₄Cl was added.

Ribosome-nascent chain complexes were expressed by an initial high cell density growth step in MDG media, followed by expression of the nascent chain in optimised minimal media EM9 with ¹⁵NH₄Cl and/or ¹³C-glucose for uniform isotopic labelling.

Ribosomes were produced using cultures grown in MDG media. Where uniform ¹⁵N-labeling of ribosomes was required, MDG media was prepared without ¹⁴NH₄Cl and L-aspartic acid; instead, ¹⁵NH₄Cl was added to the media as the sole nitrogen source.

Reagent	Final concentration	Sterilisation
Na ₂ HPO ₄	25 mM	Autoclave
KH ₂ PO ₄	25 mM	Autoclave
NH ₄ Cl*	50 mM	Autoclave
Na ₂ SO ₄	5 mM	Autoclave
L-aspartic acid pH 7.0*	0.2% (w/v)	Autoclave
Trace metals†	0.02% (v/v)	Filter
MgSO ₄	2 mM	Filter
D-glucose	0.4% (w/v)	Filter

Table 6.2 Composition of MDG growth media. *For ¹⁵N-labelled ribosomes, L-aspartic acid was omitted and ¹⁵NH₄Cl was added. †50 mM FeCl₂ (dissolved in 0.1M HCl), 20 mM CaCl₂, 1 mM MnCl₂, 1 mM ZnSO₄, 2 mM CoCl₂, 2 mM CuCl₂, 2 mM NiCl₂, 2 mM Na₂MoO₄, 2 mM Na₂SeO₃, 2 mM H₃BO₃.

Reagent	Final concentration	Sterilisation
Na ₂ HPO ₄	7.1 g/L	Autoclave
KH ₂ PO ₄	3.4 g/L	Autoclave
NaCl	584 mg/L	Autoclave
MgSO ₄	5 mM	Filter
CaCl ₂	200 μM	Filter
100 X BME vitamins	0.25 X	Filter
Trace metals	0.025% (v/v)	Filter
D-glucose*	0.4% (w/v)	Autoclave
NH ₄ Cl†	0.1% (w/v)	Filter

Table 6.3 Composition of EM9 expression media. *For uniform ¹³C-labelling, 0.2% (w/v) ¹³C-glucose was used. †For uniform ¹⁵N-labeling, ¹⁵NH₄Cl was used.

6.1.3 Production of *E. coli* competent cells

Competent *E. coli* cells were prepared using the Hanahan method (Hanahan 1983). SOB media (3 mL), inoculated with a single colony from a streaked plate, was incubated at 37°C for 16-18 h. The starter culture (1 mL) was then used to inoculate SOB (200 mL) in a 1 L flask and incubated at 37°C until OD_{600nm} reached 0.5-0.6. The culture was collected in cold, sterile tubes and chilled on ice for 10 min before harvesting by centrifugation (4000 rpm for 10 min at 4°C). The cell pellet was gently resuspended in transformation buffer 1 (67 mL, Table 6.4), chilled on ice for 30 min and then centrifuged as previously described. The cell pellet was then resuspended in transformation buffer 2 (16 mL, Table

6.4) and incubated on ice for 15 min. The cells were distributed into 50 μ L aliquots on ice, flash-frozen with liquid nitrogen and stored at -80°C .

Reagent	Final concentration	
	Transformation buffer 1*	Transformation buffer 2†
RbCl	100 mM	10 mM
MnCl ₂	50 mM	-
CaCl ₂	10 mM	75 mM
Glycerol	15% (w/v)	15% (w/v)
KOAc	30 mM	-
MOPS	-	10 mM

Table 6.4 Composition of transformation buffers 1 and 2 used for the production of competent cells. *Adjusted to pH 5.8 with acetic acid. †Adjusted to pH 6.8 with NaOH.

6.1.4 Site-directed mutagenesis

The QuickChange site-directed mutagenesis kit was used to modify the FLN5+31 RNC construct to generate the FLN5+31 NCA1 mutant (see Appendices for primers used).

6.1.4.1 Polymerase chain reaction (PCR)

Amplification of DNA was performed by polymerase chain reaction (PCR) using a Corbin Thermal Cycler, KOD polymerase (Novagen) and the oligonucleotides (Eurofins) described in Appendix A. The reaction mixture and temperature cycles typically used are detailed in Table 6.5 and Table 6.6 respectively. The PCR products were initially verified using agarose gel electrophoresis (section 6.1.7).

Reagent	Final concentration	Volume added (μL)
Template DNA (50 ng/ μ L)	2 ng/ μ L	1.00
5' primer (10 μ M)	300 nM	0.75
3' primer (10 μ M)	300 nM	0.75
MgSO ₄ (25 mM)	1.5 mM	1.50
dNTPs (2 mM)	200 μ M	2.50
10 X KOD buffer	1 X	0.50
KOD DNA polymerase (3 u/ μ L)	0.06 u/ μ L	0.50
Autoclaved ddH ₂ O	-	17.50

Table 6.5 Typical composition of PCR mixture used.

Step	Temperature	Time	Cycles
Initialisation	95°C	2 min	X 1
Denaturation	95°C	20 s	
Annealing	55°C	10 s	X 20
Elongation	70°C	3 min	
Final elongation	70°C	5 min	X 1

Table 6.6 Typical PCR temperature cycle used.

6.1.4.2 DpnI digest

After PCR amplification, remaining template DNA was digested with DpnI enzyme by mixing the PCR product (20 uL), BSA (0.25 uL) and DpnI (1 uL) in reaction buffer (total reaction volume of 25 uL). The mixture was incubated at 37°C for 3 h before heat inactivation of the enzyme at 65°C for 20 min.

6.1.5 Transformation

To 50 µL thawed competent cells, 1 µL (~50 ng) DNA was added and placed on ice for 30 min. The cells were subsequently heat shocked by incubating at 42°C for 45 seconds, and then allowed to recover on ice for 2 min. LB media (1 mL) pre-warmed to 37°C was added to the cells and incubated at 37°C with shaking. After 1 h, ~200 uL of the cell culture was spread onto pre-warmed LB (with ampicillin) agar plates. The plates were incubated for 14-18 h at 37°C.

6.1.6 DNA purification and quantification

DNA was purified from cultures (5 mL) of a single colony of transformed DH5α cells grown overnight at 37°C using the QiaPrep Spin miniprep kit (Qiagen) following the manufacturer's instructions.

Purified DNA was quantified by measuring the absorbance at 260 nm and calculated using the following relationship: $1 \text{ OD}_{260} \equiv 50 \text{ ng}/\mu\text{L}$.

6.1.7 Agarose gel electrophoresis

PCR products were verified by agarose gel electrophoresis. Agarose (0.8% (w/v)) was dissolved by heating in TAE buffer (4.84 g/L Tris, 1.142% (v/v) glacial acetic acid, 50 mM EDTA, pH 8.0). The solution was then cooled to ~45°C before 1 µg/mL ethidium bromide was added and the gel cast. DNA samples were prepared with loading dye (6X: 150 g/L Ficoll 400, 2.5 g/L Orange G dye). Electrophoresis was performed at 100 V for ~25 min before imaging under ultra-violet light.

6.1.8 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate (SDS) polyacrylamide gels were cast using the recipe described in Table 6.7. Samples were prepared with loading dye (Table 6.8) and boiled for 1 min before loading onto the gels. SDS-PAGE gels were run at 200 V for 40 min in MOPS–SDS running buffer (Alfa Aesar).

Reagent	Resolving gel	Stacking gel
Bis-tris (1.25 M) pH 6.8	2.86 mL	1.00 mL
Acrylamide (30% (w/v))	4.00 mL	0.88 mL
APS (25% (w/v))	60 uL	9 uL
TEMED	4 uL	9 uL
ddH ₂ O	3.14 mL	3.00 mL

Table 6.7 Composition of an SDS-PAGE gel

Unless immunodetection was required, gels were then visualised by staining with Coomassie stain (0.05% (w/v) Brilliant Blue R-250, 40% (v/v) ethanol, 10% (v/v) glacial acetic acid) followed by destaining solution (40% (v/v) ethanol, 10% (v/v) glacial acetic acid). For ribosomal samples, gels were typically silver-stained (SilverQuest, Invitrogen) according to the manufacturer's instructions.

Reagent	Final concentration
Glycerol	30% (v/v)
Bis-tris pH 5.7	250 mM
Bromophenol Blue	0.04% (w/v)
DTT	0.8% (w/v)
SDS	8% (w/v)

Table 6.8 Composition of SDS-PAGE loading dye

6.1.9 Immunodetection

6.1.9.1 Western blot

The SDS-PAGE gel was stacked with nitrocellulose membrane, soaked in transfer buffer (6 g/L Tris, 14.4 g/L glycine, 0.01% (w/v) SDS, 20% (v/v) methanol) and placed into the transfer chamber. Transfer was performed at 250 mA for 2 h at room temperature.

The membrane was then incubated with primary antibody (Table 6.9) at 4 °C for at least 2 h and then washed in TBST [0.05% (w/v) Tween-20 in Tris-buffered saline (TBS) (2.42 g/L Tris, 11.69 g/L NaCl, pH 7.4)] 4 times for 15 min each at room temperature. Where necessary, these steps were repeated with secondary antibody.

Finally, the membrane was incubated with chemiluminescence substrate (SuperSignal West Pico (Femto where required) substrate, Pierce) and developed using a Fujifilm LAS-1000 scanner with between 2-10 min exposure times depending on chemiluminescence intensities.

	Blocking buffer	Antibody
Anti-His	1% (v/v) casein in TBS	1:5000 penta-His HRP conjugate (Qiagen) in TBST
Anti-αSyn	5% (w/v) milk powder in TBST	Primary: 1:1000 mouse anti- α Syn in blocking buffer Secondary: 1:1000 anti-mouse IgG in blocking buffer
Anti-SecM	5% (w/v) BSA in TBST	Primary: 1:10000 rabbit anti-SecM (a kind gift from Prof Bernd Bukau (University of Heidelberg)) in 2% (w/v) BSA in TBST Secondary: 1:20000 anti-rabbit in 2% (w/v) BSA in TBST

Table 6.9 Buffer compositions for western blot immunodetection.

6.1.9.2 Densitometric analysis

Densitometric analysis was performed using ImageJ software of the chemiluminescence intensities of known protein standards.

6.1.10 Protein expression and purification

6.1.10.1 Expression of isolated proteins

A single colony of transformed cells was used to inoculate 5 mL LB media. After incubating overnight at 37°C with shaking, the starter culture was used to inoculate 1 L M9 media prepared with $^{15}\text{NH}_4\text{Cl}$ (for uniform ^{15}N -labeling) and/or ^{13}C -glucose (for uniform ^{13}C -labeling) to an initial $\text{OD}_{600\text{nm}}$ of ~ 0.1 .

The culture was grown at 37°C with shaking until the $\text{OD}_{600\text{nm}}$ reached ~ 0.6 . IPTG was added to the culture to a final concentration of 1 mM to induce protein expression, and the culture was incubated at 37°C for a further 4 h with shaking. The cells were harvested by centrifuging at 4000 rpm for 30 min, and the subsequent cell pellet was flash-frozen with liquid nitrogen and stored at -20°C.

6.1.10.2 Purification of isolated FLN5 constructs

The buffers used in this purification are detailed in Table 6.10.

Buffer	Composition
Lysis	25 mM Na ₂ HPO ₄ , 2M NaCl, 2% (w/v) Triton X-100, 0.2 µg/mL DNase I, 2 mg/mL lysozyme, pH 8.0
Equilibration	25 mM Na ₂ HPO ₄ , 500 mM NaCl, pH 8.0
Wash	25 mM Na ₂ HPO ₄ , 500 mM NaCl, 25 mM imidazole, pH 8.0
Elution	25 mM Na ₂ HPO ₄ , 500 mM NaCl, 250 mM imidazole, pH 8.0
SEC	25 mM Na ₂ HPO ₄ , 150 mM NaCl, pH 8.0
Ion-exchange buffer A	20 mM Tris-HCl, pH 8.0
Ion-exchange buffer B	20 mM Tris-HCl, 1 M NaCl, pH 8.0
Tico	10 mM HEPES, 30 mM NH ₄ Cl, 6 mM MgCl ₂ , 2 mM BME, 1 mM EDTA, pH 7.5

Table 6.10 Buffer compositions for the purification of isolated FLN5 constructs.

Thawed cells were resuspended in lysis buffer and lysed by sonication on ice, using 10 cycles of 10 s sonication followed by 20 s of rest. The cellular debris was pelleted at 19000 rpm for 1 h at 4°C and then passed through a 0.45 µm filter.

The cell lysate was incubated at 4°C with nickel-nitrilotriacetic acid (Ni-NTA) resin (washed with equilibration buffer) for 1 h to allow binding. The resin was washed to baseline with wash buffer, monitored by Bradford assay (40 µL added to 160 µL Bradford reagent), and the protein subsequently eluted with elution buffer. The elution fractions were analysed by SDS-PAGE, and the purest fractions pooled, buffer-exchanged into SEC buffer and concentrated to less than 5 mL using a 10 kDa molecular weight cut-off concentrator.

Further purification was performed by running the sample on a HiLoad 16/60 Superdex S75 column (GE Healthcare) in SEC buffer and collecting 1 mL fractions. The fractions were analysed by SDS-PAGE, and the pure fractions pooled, buffer exchanged into Tico buffer and concentrated.

When required, an additional anion-exchange purification step was performed prior to gel filtration, in which the protein sample was buffer exchanged into ion-exchange buffer A and loaded onto a 5 mL HiTrap DEAE column (GE Healthcare) pre-equilibrated with the same buffer. The column was washed with 60 mL ion exchange buffer A and the protein was eluted using a 60 mL linear salt gradient from 0.1-1.0 M NaCl using ion exchange buffer B at 5 mL/s, collected in 1.5 mL fractions.

The concentration of the final sample was determined from absorption at 280 nm, using the extinction coefficient of $\epsilon = 5960 \text{ M}^{-1}\text{cm}^{-1}$ (for FLN5 A₃A₃, $\epsilon = 1490 \text{ M}^{-1}\text{cm}^{-1}$) calculated from the amino acid sequence. The purified samples were flash-frozen with liquid nitrogen and stored at -80°C.

6.1.10.3 Purification of isolated α -synuclein constructs

The buffers used in this purification are detailed in Table 6.11.

Buffer	Composition
Lysis	100 mM Tris-HCl, 10 mM EDTA, 0.2 $\mu\text{g}/\text{mL}$ DNase I, pH 8.0
Ion exchange buffer A	25 mM Tris-HCl, 2 mM BME, pH 7.7
Ion exchange buffer B	25 mM Tris-HCl, 2 mM BME, 1 M NaCl, pH 7.7
Phosphate buffer	40 mM Na ₂ HPO ₄ , 150 mM NaCl, pH 6.6

Table 6.11 Buffer compositions for the purification of isolated α Syn constructs.

Thawed cells were resuspended in lysis buffer supplemented with a protease inhibitor tablet (Sigma-Aldrich) and lysed by sonication on ice, using 10 cycles of 10 s sonication followed by 20 s of rest. The cellular debris was pelleted at 19000 rpm for 1 h at 4°C.

The cell lysate was boiled for 20 min and the resultant aggregated *E. coli* proteins separated from the lysate by centrifuging at 13500 rpm for 20 min at 4°C. The supernatant was stirred at 4°C after addition of streptomycin sulphate (10 mg/mL) to remove nucleic acids. The sample was spun at 13500 rpm for 20 min at 4°C. To the supernatant, ammonium sulphate (400 mg/mL) was added and the sample stirred for 30 min at 4°C to selectively precipitate the protein. The sample was centrifuged again at 13500 rpm for 20 min at 4°C and the resultant supernatant was discarded. The pellet was resuspended in minimal volume of ion exchange buffer A and dialysed against water for 16-18 h.

The sample was loaded onto a 5 mL HiTrap Q Sepharose column (GE Healthcare) pre-equilibrated in ion exchange buffer A. The column was washed to baseline using ion exchange buffer A, after which the protein was eluted using a 60 mL linear salt gradient from 0-600 mM NaCl with ion exchange buffer B at 5 mL/s, collected in 1.5 mL fractions. The fractions were analysed by SDS-PAGE, and the pure fractions pooled, buffer exchanged into phosphate buffer and concentrated using a 10-KDa molecular weight cut-off concentrator.

The concentration of the final sample was determined from absorption at 280 nm, using the extinction coefficient of isolated WT α -synuclein ($\epsilon = 5960 \text{ M}^{-1}\text{cm}^{-1}$) or isolated Y125W α -synuclein ($\epsilon = 9970 \text{ M}^{-1}\text{cm}^{-1}$) calculated from the amino acid sequence. The purified samples were flash-frozen with liquid nitrogen and stored at -80°C .

6.1.11 Ribosome and ribosome-nascent chain expression and purification

6.1.11.1 Expression of ribosomes

Ribosomes were produced by growing *E. coli* cells in a high-density growth medium, with tetracycline as the antibiotic in each media: firstly, a starter culture of 5 mL LB was inoculated with a single colony from an LB agar plate streaked with BL21(DE3) cells. After incubation at 37°C for 3-4 h, the starter culture was used to inoculate MDG media to an $\text{OD}_{600\text{nm}}$ of 0.005; $^{14}\text{NH}_4\text{Cl}$ and L-aspartic acid, or $^{15}\text{NH}_4\text{Cl}$ for uniform ^{15}N -labelled samples, were added as the nitrogen sources. The cell culture was incubated at 30°C for 16-22 h until a final $\text{OD}_{600\text{nm}}$ of ~ 5 , and then harvested by centrifuging at 4000 rpm for 30 min at 4°C . The cell pellet was flash-frozen and stored at -20°C .

6.1.11.2 Purification of ribosomes

The buffers used in this purification are detailed below.

Buffer	Composition
Lysis	50 mM HEPES, 1M KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 5 mM EDTA, 0.2 $\mu\text{g}/\text{mL}$ DNase I, 2 mg/mL lysozyme, pH 7.5
Cushion	35% (w/v) sucrose, 50 mM HEPES, 1 M KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 5 mM EDTA, 5 mM ATP, 0.1% protease inhibitor*, pH 7.5
Resuspension	50 mM HEPES, 1 M KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 5 mM EDTA, 0.1% protease inhibitor*, pH 7.5
Gradient	10/35% (w/v) sucrose, 50 mM HEPES, 1 M KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 5 mM EDTA, 0.1% protease inhibitor*, pH 7.5
Butyl column buffer A	20 mM HEPES, 1.5 M NH_4SO_4 , 400 mM KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 0.1% protease inhibitor*, pH 7.5
Butyl column buffer B	20 mM HEPES, 400 mM KOAc, 12 mM $\text{Mg}(\text{OAc})_2$, 2 mM BME, 0.1% protease inhibitor*, pH 7.5
Tico	10 mM HEPES, 30 mM NH_4Cl , 12 mM MgCl_2 , 2 mM BME, 1 mM EDTA, 0.1% protease inhibitor*, pH 7.5

Table 6.12 Buffer compositions for the purification of ribosomes. *1 protease inhibitor tablet dissolved in 1 mL ddH₂O = 100%

Thawed cells were resuspended in lysis buffer supplemented with 1 protease inhibitor tablet (Sigma Aldrich). The cells were lysed by passing through a French press 4-6 times. The cellular debris were pelleted by centrifuging at 19000 rpm for 1 h at 4°C. The lysate was layered onto cushion buffer and centrifuged at 35000 rpm for 14 h (or 44000 rpm for 4 h) at 4°C using a Type 45Ti rotor (Beckman Coulter).

The resulting pellets were resuspended in minimal volume of resuspension buffer and further purified using a linear 10-35% sucrose gradient prepared with gradient buffers and centrifuged at 22000 rpm for 15 h at 4°C using an SW28 rotor (Beckman Coulter). The sucrose gradients were fractionated in 1 mL fractions, monitored by absorbance at 254 nm and assessed by SDS-PAGE. Fractions containing pure 70S ribosome-containing fractions were pooled, buffer exchanged into Tico buffer and concentrated using 100 kDa molecular weight cut-off concentrators.

Alternatively to sucrose gradients, the cushion pellets were resuspended in minimal volume of butyl column buffer A and further purified by hydrophobic interaction chromatography: the sample was loaded onto a 5 mL HiTrap Butyl column (GE Healthcare) pre-equilibrated in butyl column buffer A. After washing to baseline, ribosomes were eluted using a 60 mL linear salt gradient from 0.1-1.0 M NH_4SO_4 using butyl column buffer B at 5 mL/s, collected in 1.5 mL fractions. The fractions were assessed by SDS-PAGE, and the pure fractions pooled, buffer exchanged into Tico buffer and concentrated.

The concentration of the final sample was determined from absorption at 260 nm, using the extinction coefficient $\epsilon = 4.2 \times 10^7 \text{ M}^{-1}\text{cm}^{-1}$ (where $A_{260} = 1$ corresponds to 24 pmol/mL). The purity of the sample was assessed by determining A_{260}/A_{280} ; a pure ribosome sample has $A_{260}/A_{280} = 2.0$. The purified samples were flash-frozen with liquid nitrogen and stored at -80°C.

6.1.11.3 Expression of ribosome-nascent chain complexes

Ribosome-nascent chain complexes were expressed by initially growing *E. coli* cells in MDG media before expressing in EM9 media. Firstly, a starter culture of 5 mL LB was inoculated with a single colony of transformed cells and incubated at 37°C. After 4-6 h, the starter culture was used to inoculate MDG media to an $\text{OD}_{600\text{nm}}$ of 0.005.

The cell culture was incubated for ~18 h at 30°C and then harvested by centrifugation (4000 rpm for 30 min at 4°C). The cell pellet was washed twice, each time by gently resuspending in EM9 salts followed by centrifuging at 4000 rpm for 30 min at 4°C. The cells were then resuspended in EM9 media containing the appropriate isotopes and IPTG (final concentration of 1 mM) to induce expression. After incubating for 10 min

at 30°C, rifampicin (final concentration of 150 µg/L prepared in DMSO) was added and culture was allowed to incubate for a further 30 min at 30°C. Finally, the cells were harvested by spinning at 4000 rpm for 30 min at 4°C, flash-frozen with liquid nitrogen and stored at -20°C.

6.1.11.4 Purification of ribosome-nascent chain complexes

The buffers used in this purification are detailed below.

Buffer	Composition
Lysis	50 mM HEPES, 1M KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 0.2 µg/mL DNase I, 2 mg/mL lysozyme, 10% protease inhibitor*, pH 7.5
Cushion	30% (w/v) sucrose, 50 mM HEPES, 1 M KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 5 mM ATP, 0.1% protease inhibitor*, pH 7.5
Resuspension	50 mM HEPES, 1 M KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 0.1% protease inhibitor*, pH 7.5
Equilibration	50 mM HEPES, 500 mM KOAc, 6 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 0.1% protease inhibitor*, pH 7.5
Wash	50 mM HEPES, 500 mM KOAc, 6 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 5 mM ATP, 0.1% protease inhibitor*, pH 7.5
Elution	50 mM HEPES, 500 mM KOAc, 6 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 250 mM imidazole, 0.1% protease inhibitor*, pH 7.5
Gradient	10/35% (w/v) sucrose, 50 mM HEPES, 1 M KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 5 mM EDTA, 0.1% protease inhibitor*, pH 7.5
Butyl column buffer A	20 mM HEPES, 1.5 M NH ₄ SO ₄ , 400 mM KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 0.1% protease inhibitor*, pH 7.5
Butyl column buffer B	20 mM HEPES, 400 mM KOAc, 12 mM Mg(OAc) ₂ , 2 mM BME, 0.1% protease inhibitor*, pH 7.5
Tico	10 mM HEPES, 30 mM NH ₄ Cl, 12 mM MgCl ₂ , 2 mM BME, 1 mM EDTA, 0.1% protease inhibitor*, pH 7.5

Table 6.13 Buffer compositions for the purification of RNC. *1 protease inhibitor tablet dissolved in 1 mL ddH₂O = 100%

Thawed cells were resuspended in lysis buffer supplemented with 1 protease inhibitor tablet. The cells were lysed by passing through a French press 4-6 times. The cellular debris were pelleted by centrifuging at 19000 rpm for 1 h at 4°C. The lysate was layered

onto cushion buffer and centrifuged at 35000 rpm for 14 h (or 44000 rpm for 4 h) at 4°C using a Type 45Ti rotor (Beckman Coulter).

The resulting pellets were resuspended in minimal volume of resuspension buffer and incubated at 4°C with nickel-iminodiacetic acid (Ni-IDA) resin (washed with equilibration buffer) for 1 h to allow binding. The resin was washed to baseline with wash buffer followed by equilibrium buffer and the RNC subsequently eluted with elution buffer; each step was monitored by Bradford assay (40 µL added to 160 µL Bradford reagent). The eluate was buffer exchanged into equilibration buffer to remove imidazole and concentrated using 100 kDa molecular weight cut-off concentrators.

The sample was further purified using a linear 10-35% sucrose gradient prepared with gradient buffers and centrifuged at 22000 rpm for 15 h at 4°C using an SW28 rotor (Beckman Coulter). The sucrose gradients were fractionated in 1 mL fractions, monitored by absorbance at 254 nm and assessed by SDS-PAGE. Fractions containing pure 70S ribosome-containing fractions were pooled, buffer exchanged into Tico buffer and concentrated.

Alternatively to sucrose gradients, the sample was subjected to another sucrose cushion step (as previously described) before resuspension in minimal volume of butyl column buffer A and further purification by hydrophobic interaction chromatography. The sample was loaded onto a 5 mL HiTrap Butyl column (GE Healthcare) pre-equilibrated in butyl column buffer A. After washing to baseline, ribosomes were eluted using a 60 mL linear salt gradient from 0.1-1.0 M NH_4SO_4 using butyl column buffer B at 5 mL/s, collected in 1.5 mL fractions. The fractions were assessed by SDS-PAGE, and the pure fractions pooled, buffer exchanged into Tico buffer and concentrated.

The concentration of the final sample was determined from absorption at 260 nm, using the extinction coefficient $\epsilon = 4.2 \times 10^7 \text{ M}^{-1}\text{cm}^{-1}$ (where $A_{260} = 1$ corresponds to 24 pmol/mL). The purity of the sample was assessed by determining A_{260}/A_{280} ; a pure ribosomal sample has $A_{260}/A_{280} = 2.0$. The purified samples were flash-frozen with liquid nitrogen and stored at -80°C.

6.1.11.5 Small-scale expression and purification of RNCs

The RNCs were expressed in 50 mL cultures using the protocol described in section 6.1.11.3. The cells were resuspended in lysis buffer (BugBuster (primary amine-free) Protein Extract Reagent (Novagen, Inc) supplemented with 25 mM MgCl_2 , 5 mL/g cells) and incubated on ice for 30 min. After pelleting (14000 rpm, 15 min at 4°C), the supernatant was layered on top of sucrose cushion buffer (Table 6.13) and centrifuged at 100000 rpm for 1 h at 4°C using a TLA 110 rotor (Beckman Coulter). The pellets were

resuspended in Tico buffer (Table 6.13) and the concentration of the sample was determined by absorption at 260 nm (section 6.1.11.4). The sample was flash-frozen and stored at -80°C.

6.1.12 Preparation of in-cell NMR samples

NMR samples of cells in which the protein or RNC of interest had been expressed were prepared following the protocols described in sections 6.1.10.1 and 6.1.11.3 with some adjustments as detailed below.

For uniform ^{15}N -labelling of samples, cells were initially grown in LB media until OD \sim 0.6 and then washed with M9 salts before resuspension in M9 media with $^{15}\text{NH}_4\text{Cl}$ (0.1% (w/v)) and IPTG (1 mM) to induce expression.

For uniform $^2\text{H},^{15}\text{N}$ -labelling of samples, cells were progressively adapted into deuterated isotopes using a stepwise approach: a 5 mL starter culture of LB was used to inoculate 5 mL of MDG media prepared in 80% $^2\text{H}_2\text{O}$. After incubating overnight at 37°C, the resulting culture was used to inoculate 5 mL of MDG media prepared in 100% $^2\text{H}_2\text{O}$ with $^{12}\text{C-d}_7$ -glucose (0.1% (w/v)) and unlabelled glucose (0.1% (w/v)). The culture was grown at 37°C overnight after which it was used to inoculate the final starter culture of MDG media prepared in 100% $^2\text{H}_2\text{O}$ with $^{12}\text{C-d}_7$ -glucose (0.2% (w/v)). Once incubated for 6-8 h at 37°C, the starter culture was used to inoculate 50 mL of MDG media prepared in 100% $^2\text{H}_2\text{O}$ with $^{12}\text{C-d}_7$ -glucose (0.2% (w/v)) which was incubated for at least 16 h at 37°C (ensuring that the final OD_{600 nm} exceeded 2.0). The cells were centrifuged at 4000 rpm for 20 min at 4°C, and resuspended into EM9 media prepared in 100% $^2\text{H}_2\text{O}$ with $^{12}\text{C-d}_7$ -glucose (0.2% (w/v)) and $^{15}\text{NH}_4\text{Cl}$ (0.1% (w/v)). After incubating the culture at 30°C for 1 h, IPTG (1 mM) was added and expression allowed to proceed for 2 h at 30°C.

The production of uniform $^2\text{H-Ile}\delta_1$ - $^{13}\text{CH}_3$ -labelled samples, in which the δ_1 methyl group of isoleucine side chains were selectively protonated, was achieved using the same approach as described above for uniform $^2\text{H},^{15}\text{N}$ -labelling, except $^{15}\text{NH}_4\text{Cl}$ was replaced with $^{14}\text{NH}_4\text{Cl}$ and the isoleucine precursor 2-ketobutyric-4- $^{13}\text{CH}_3$,3,3-d $_2$ acid in the EM9 expression media.

Immediately after expression, cells were harvested by centrifugation at 4000 rpm for 20 min at 4°C. The cell pellet was washed twice, each time by gently resuspending in EM9 salts followed by centrifuging at 4000 rpm for 10 min at 4°C. The cell pellet was resuspended as a 50% (w/v) slurry in unlabelled EM9 media containing 2% (w/v) glucose (or $^{12}\text{C-d}_7$ -glucose in perdeuterated samples). For $^2\text{H-Ile}\delta_1$ - $^{13}\text{CH}_3$ -labelled samples, the cell pellet was instead washed with EM9 salts prepared in $^2\text{H}_2\text{O}$ at 10°C and resuspended in unlabelled EM9 media prepared in $^2\text{H}_2\text{O}$ with 2% (w/v) $^{12}\text{C-d}_7$ -glucose.

To test the viability of cells, 10 μL of the NMR sample were taken before and after data acquisition and diluted 10⁸-fold in LB media of which 100 μL was plated onto each of three LB agar plates and incubated for 16 h at 37°C.

6.1.13 Preparation of lysed in-cell NMR samples

After NMR acquisition, the in-cell sample was centrifuged at 14000 rpm for 10 min at 4°C and the resulting supernatant removed. The cell pellet was resuspended in unlabelled EM9 media (of the same volume as the discarded supernatant). The cells were lysed by freeze-thawing 6-8 times in which each time the cell pellet was flash-frozen in liquid nitrogen and allowed to thaw in a beaker of water. The cellular debris was pelleted by centrifuged at 14000 rpm for 15 min at 4°C. The lysate was immediately used for further NMR data acquisition using the same parameters as the in-cell sample for direct comparison.

6.1.14 Preparation of reconstituted cytosol

Lyophilised *E. coli* lysate was prepared using a similar approach to a previously described protocol for reconstituted cytosol (Smith et al. 2013).

A starter culture of 5 mL LB was inoculated with a single colony from an LB agar plate streaked with BL21(DE3) cells. The starter culture was then used to inoculate 500 mL of unlabelled MDG media to a starting OD₆₀₀ of ~0.005 after which it was incubated for ~18 h at 37°C. The cell pellet was washed twice, each time by gently resuspending in EM9 salts followed by centrifuging at 4000 rpm for 30 min at 4°C. The cells were then resuspended in unlabelled EM9 media containing the appropriate isotopes and IPTG (final concentration of 1 mM). After incubating for 10 min at 30°C, rifampicin (final concentration of 150 $\mu\text{g}/\text{L}$ prepared in DMSO) was added and culture was allowed to incubate for a further 30 min at 30°C. Finally, the cells were harvested by spinning at 4000 rpm for 30 min at 4°C, flash-frozen with liquid nitrogen and stored at -20°C.

Thawed cells were resuspended in Tico buffer (Table 6.13) and lysed by passing through a French press 4-6 times. The cellular debris were pelleted by centrifuging at 19000 rpm for 1 h at 4°C, and the supernatant passed through a 0.22 μm filter. The lysate was dialysed (3 kDa molecular weight cut-off) against Tico buffer at 4°C for 72 h. The lysate was then lyophilised, resuspended into Tico buffer and the pH adjusted to 7.5. The lysate was then aliquoted into quantities for each NMR sample which were lyophilised and stored at -80°C. The protein concentration of the reconstituted cytosol was determined using a BCA assay kit according to the manufacturer's instructions.

6.1.15 Preparation of NiDO2A

DO2A [1,4,7,10-tetraazacyclododecane-1,7-bis(acetic acid)] was purchased from Macrocyclics, Inc. (Dallas, Texas, USA) as a lyophilized salt (H₂DO2A.4HCl). A 5% molar excess of DO2A (200 mg) was mixed with anhydrous nickel (II) chloride (57 mg) and dissolved in 5 mL of deionized water. The solution was adjusted to neutral pH, coinciding with a colour change from blue to purple (Cai et al. 2006). Salt and excess DO2A were removed by Dowex Retardion 11A8 ion-exchange resin (Sigma-Aldrich UK) packed into a column and connected to an ÄKTA FPLC system. The absorption at wavelength 545 nm and the conductivity were monitored for the elution of NiDO2A and excess salt, respectively. Desalting followed by lyophilisation and redissolving of the sample in deionised water was repeated two to three times, and again immediately before use in NMR experiments. The concentration of NiDO2A stock solution (determined by measuring dry mass of NiDO2A before dissolving) was adjusted to a final concentration of 0.5–1.0 M and added to NMR samples as required. The extinction coefficient of NiDO2A at 545 nm was determined as $\epsilon = (7.51 \pm 0.36) \text{ M}^{-1} \text{ cm}^{-1}$.

6.1.16 Preparation of residual dipolar coupling measurement samples

NMR samples for RDC measurements were prepared by aligning in bacteriophage Pf1 (ASLA Biotech Ltd.). To buffer exchange, Pf1 was centrifuged at 95000 rpm for 1 h at 5°C, and resuspended into Tico buffer (Table 6.13) to produce the stock solution. The concentration of phage was determined by absorption at 270 nm, using the extinction coefficient $\epsilon = 2.25 \text{ ml cm}^{-1} \text{ mg}^{-1}$. The appropriate volume of the stock solution was then added to the NMR sample using a glass pipette.

6.2 NMR spectroscopy

6.2.1 NMR experimental conditions

NMR experiments were recorded on a 700-MHz Bruker Avance III, 800-MHz Bruker Avance III HD spectrometer (both equipped with a TXI cryogenic probe), or a 500MHz Bruker Avance III spectrometer equipped with a TXI room temperature probe, all with a unidirectional gradient coil generating a maximum gradient of 0.55 T m⁻¹.

NMR samples of isolated α -synuclein were prepared in 40 mM Na₂HPO₄, 150 mM NaCl, pH 6.6, supplemented with 10% (v/v) D₂O as a lock solvent and 0.01% (w/v) DSS as a reference standard. All other *in vitro* samples were prepared in Tico buffer (Table 5.13) and supplemented with 10% (v/v) D₂O and 0.001% (w/v) DSS. All NMR experiments were recorded using 5 mm diameter Shigemi tubes; this is particularly important in NMR diffusion experiments as the reduced sample height both inhibits the

onset of convection (Chung et al. 2011). NMR experiments of FLN5, FLN5 RNC and ribosomes described in the PLRE investigations (Chapter 3) were performed at 25°C. All other *in vitro* experiments were performed at 10°C.

In-cell NMR samples in protonated, unlabelled EM9 media (section 6.1.12) were supplemented with 10% (v/v) DSS and 0.001% (w/v) DSS. For ^2H -Ile δ 1-[^{13}C] δ 1-labelled samples (section 6.1.12), DSS was added at a final concentration of 0.004% (w/v). In-cell samples were transferred to a Shigemi tube using a glass pipette, and the cells were allowed to settle at the bottom of the tube before NMR data acquisition at 10°C. The inner plunger of the Shigemi tube was not used.

6.2.2 NMR data acquisition, processing and analysis

All NMR spectra were referenced to DSS (Wishart et al. 1995), processed with nmrPipe (Delaglio et al. 1995) and imported into CCPN Analysis (Vranken et al. 2005) or MATLAB (R2014b, The MathWorks Inc.) for analysis.

6.2.2.1 1D ^1H experiments

One-dimensional ^1H spectra were acquired using excitation sculpting for water suppression, with spectral widths of 20 ppm and 16384 time-points for an acquisition time of ~600 ms, and a recycle delay of 1 s. Spectra were processed using cosine-squared window functions.

6.2.2.2 ^1H - ^{15}N SOFAST-HMQC experiments

Two-dimensional ^1H - ^{15}N SOFAST-HMQC spectra (Schanda et al. 2005) were acquired with 128 complex points and sweep widths of 23 (α -synuclein, FLN5 Δ 12, FLN5+31 RNC, FLN5+31 N Δ 1), 33 (FLN5), or 32 ppm (ribosomes) in the indirect ^{15}N dimension; and 1024 points and sweep widths of 20 (α -synuclein and all FLN5 constructs) or 14 ppm (ribosomes) in the direct (^1H) dimension, corresponding to acquisition times of ~50 ms. 2D spectra were recorded with a 50 ms inter-scan recovery delay. Using the same parameters, one-dimensional ^1H - ^{15}N SOFAST HMQC spectra were acquired as the first increment of the 2D spectra, or where required, a whole series of ^1H - ^{15}N SOFAST HMQC spectra were acquired as pseudo-2D experiments, with recovery times varying from 50 ms to 1 s. Spectra were processed using cosine-squared window functions. Exponential window functions were used in the direct dimension for ribosome and RNC spectra.

6.2.2.3 Relaxation rate measurements

Proton longitudinal and transverse relaxation rates were measured using ^1H - ^{15}N HSQC experiments incorporating initial inversion-recovery and spin-echo elements

respectively. Effective longitudinal relaxation rates were measured following both hard and amide-selective square (400–560 ms) inversion pulses. ^{15}N longitudinal and transverse relaxation rates were measured using standard Bruker library sequences. In all cases, experiments were acquired as pseudo-2D experiments, and 1D integrals of the amide region were fitted to determine approximate relaxation rates averaged over all residues in the protein.

6.2.2.4 PLRE investigations of NMR diffusion measurements

^{15}N XSTE and SORDID diffusion experiments were acquired with diffusion delays Δ varied between 110 and 350 ms. The gradient strength G was varied (between 0.28 and 0.53 T m $^{-1}$) to obtain a constant echo attenuation I/I_0 in all experiments, by maintaining a constant value of the product $G^2(\Delta - \delta/3 - \tau/2) \approx G^2\Delta$ according to the Stejskal–Tanner equation ((Stejskal and Tanner 1964; Wu et al. 1995)):

$$\frac{I}{I_0} = \exp[-D \gamma^2 \sigma^2 G^2 \delta^2 (\Delta - \delta/3 - \tau/2)] \quad (6.1)$$

where D is the diffusion coefficient, γ is the gyromagnetic ratio, δ is the length of the encoding and decoding gradient pulses ($\delta = 4$ ms), σ is the shape factor of the gradient pulses ($\sigma = 0.9$ for the trapezoidal gradient shapes used in this work), and τ is the delay between the bipolar gradient pulses. In order to avoid damage to the probe due to the high rate of repetition in SORDID experiments, the time-averaged power in the gradient coil, $\bar{P} \propto \frac{1}{T_{\text{scan}}} \int_0^{T_{\text{scan}}} I(t)^2 dt$, where current I is directly proportional to the applied gradient strength G , was considered and limited according to the probe specification. For the RNC, we sought to compare constant echo attenuation between XSTE with $\Delta = 100$ ms ($G = 5$, 95% G_{max}) and, to meet the limits of \bar{P} , a longer diffusion delay of $\Delta = 190$ ms was used for SORDID experiments ($G = 10.4$, 69.5% G_{max}).

XSTE spectra were acquired with 64 scans, 1024 complex points and sweep width of 20 (α -synuclein and FLN) or 15 ppm (ribosomes and RNC) in the ^1H dimension, a recovery delay of 1 s, and acquisition times of 51.25 ms (α -synuclein and FLN5) or 48.79 ms (ribosomes and RNC). SORDID experiments were recorded using 256 (α -synuclein and FLN5), 64 (ribosomes), or 480 scans (RNC), 1024 complex points and sweep width of 20 (α -synuclein and FLN5) or 15 ppm (ribosomes and RNC) in the ^1H dimension, and acquisition times of ~ 50 ms. By considering the effect of overlapping scans, additional phase cycling was introduced in the SORDID experiment to select heteronuclear coherences and improve solvent suppression (detailed in Figure 2.1c). Diffusion spectra were processed using cosine-squared window functions, or an exponential window

function for ribosome and RNC samples, linear baseline correction and solvent suppression filters.

Following integration of amide regions of diffusion spectra, diffusion coefficients D were calculated using the Stejskal-Tanner equation (Equation 5.1). The standard deviation of the noise integral was calculated using all diffusion spectra from each sample.

The experimental sensitivity (signal-to-noise ratio per unit time, SNR_t), was calculated as:

$$\text{SNR}_t = \frac{\text{total amide envelope integral}}{\text{standard deviation of noise integral}} \times \frac{1}{\sqrt{N_{\text{scan}} T_{\text{scan}}}} \quad (6.2)$$

where N_{scan} is the number of scans and T_{scan} is the total time for one scan. The values for experimental sensitivity were subsequently normalized according to the maximum SNR_t obtained for XSTE experiments in the absence of NiDO2A. The resulting SNR_t data points, plotted over varying Δ , were fitted to their theoretical sensitivities derived by detailed analysis of the trajectory of magnetization during the pulse sequences:

$$\text{SNR}_{t, \text{XSTE}} = \frac{A \exp(-4R_2^H \tau) \exp(-4R_2^N \tau) \exp(-R_1^N T_N)}{\sqrt{N_{\text{scan}} T_{\text{scan}}}} \times [1 - \exp(-R_1^H T_{\text{rec}})] \quad (6.3)$$

$$\text{SNR}_{t, \text{SORDID}} = \frac{2A \exp(-4R_2^H \tau) \exp(-4R_2^N \tau) \exp[-2\delta'(R_1^N + R_1^H)] \exp[-(\Delta - 2\delta' - 6\tau)R_1^N]}{\sqrt{2 N_{\text{scan}} T_{\text{scan}}}} \times [1 - \exp(-R_1^H T_{\text{rec}})] \quad (6.4)$$

where A is a scaling factor, T_N is the length of period during which magnetization is stored on ^{15}N nuclei, T_{rec} is the recovery time, τ is the delay for INEPT transfer ($\tau = |1/4J_{\text{NH}}| = 2.72$ ms, for $J_{\text{NH}} \approx -92$ Hz), and δ' is the length of the delay as described in the pulse sequence (Fig. S1). An additional factor of 2 is included in the theoretical SORDID sensitivity to account for each diffusion delay extending over two scans. The measured transverse relaxation rates R_2^H and R_2^N were directly inputted into the fit (as their exponential factors only attenuate the scaling), and a global fitting was used to determine the individual longitudinal relaxation rates R_1^H and R_1^N of each sample, and the shared amplitude A between each diffusion experiment. Errors were calculated by bootstrapping of residuals (Efron and Tibshirani 1994).

6.2.2.5 NMR diffusion measurements of RNCs

To monitor attachment of the nascent chain to the ribosome, ^{15}N SORDID experiments were acquired using the same parameters described in section 6.2.2.4, and processed using exponential window functions, linear baseline correction and solvent suppression filters. Where sensitivity of the spectra was low, several diffusion spectra were summed together and the total spectrum processed before further analysis. Following integration of amide regions of diffusion spectra, diffusion coefficients D were calculated using the Stejskal-Tanner equation (Equation 5.1).

6.2.2.6 In-cell NMR diffusion measurements

To monitor cell leakage of in-cell samples, ^{15}N SORDID experiments or 1D diffusion-edited ^1H - ^{13}C HMQC spectra (for ^{15}N - and ^{13}C -labelled samples respectively) were acquired with a diffusion delay of 300 ms, and encoding and decoding gradient lengths of 4 ms, with gradient strengths of 5% and 95% of the maximum gradient strength. Experiments were recorded with 1024 complex points, sweep width of 15 ppm, and an acquisition time of ~ 50 ms. Spectra were processed using cosine-squared window functions.

6.2.2.7 ^1H - ^{15}N - ^{13}C BEST-HNCA and BEST-HNCO experiments

Three-dimensional ^1H - ^{15}N - ^{13}C BEST-HNCA experiments were recorded on the highest field spectrometer available (800-MHz) equipped with a cryogenic probe to maximize sensitivity. Experiments were acquired with 2560 points and sweep width of 16 ppm in the direct (^1H) dimension; 48 points and sweep width of 23 ppm in the ^{15}N dimension; and 40 points and sweep width of 23 ppm in the ^{13}C dimension; corresponding to acquisition times of ~ 50 ms. Spectra were acquired with a recovery delay of 150 ms, and processed using cosine window functions.

Three-dimensional ^1H - ^{15}N - ^{13}C BEST-HNCO experiments were recorded on a 700-MHz spectrometer equipped with a cryogenic probe to balance maximizing sensitivity and reducing the effects of transverse relaxation of carbonyl carbons at high field strengths (see main text, Chapter 4). Spectra were acquired with 1024 points and sweep width of 15 ppm in the direct (^1H) dimension; 40 points and sweep width of 23 ppm in the ^{15}N dimension; and 40 points and sweep width of 5 ppm in the ^{13}C dimension; corresponding to acquisition times of ~ 50 ms. Spectra were acquired with a recovery delay of 200 ms, and processed using cosine window functions and linear prediction.

Using the same parameters, 2D ^1H - ^{15}N - ^{13}C BEST-HNCA and HNCO spectra of the ^1H - ^{13}C plane were acquired by setting only one time-point in the ^{15}N dimension. Similarly, a series of 1D ^1H - ^{15}N - ^{13}C BEST-HNCA and HNCO spectra were acquired with the same

parameters and as the first increment of the 3D spectra with recovery times varying from 50 ms to 3 s. Spectra were processed using cosine window functions.

Where two C α resonances were observable for one residue i , the chemical shift was measured by taking the average of the chemical shifts deriving from the resonances of the i th and the $(i+1)$ th residue, and weighted to the signal intensity I :

$$\delta_{C\alpha,average} = \frac{I_{C\alpha,i} \delta_{C\alpha,i} + I_{C\alpha,i+1} \delta_{C\alpha,i+1}}{I_{C\alpha,i} + I_{C\alpha,i+1}} \quad (6.5)$$

The chemical shift perturbation of the RNC was then measured for each nucleus ($^1\text{H}^N$, ^{15}N , $^{13}\text{C}'$ and $^{13}\text{C}\alpha$, denoted X below) and residue as the difference between the RNC chemical shift and the isolated protein chemical shift:

$$\Delta\delta_X = \delta_{X,RNC} - \delta_{X,isolated} \quad (6.6)$$

The complete chemical shift change of the RNC relative to the isolated protein was then calculating using all the available chemical shift changes for the residue:

$$\Delta\delta = \sqrt{[(\Delta\delta_H)^2 + (0.2 \Delta\delta_N)^2 + (0.3 \Delta\delta_{C'})^2 + (0.3 \Delta\delta_{C\alpha})^2]} \quad (6.7)$$

Scaling factors are introduced to weight the relative chemical shifts of different nuclei (Williamson 2013). Uncertainties in the chemical shifts were calculated using the linewidth LW and signal intensity I of the resonance:

$$\delta_{error} = \frac{1}{2} \cdot \frac{LW}{I} \quad (6.8)$$

and propagated through calculations of chemical shift changes (Douglas and Donald 1971).

6.2.2.8 RDC measurements

Alignment of the sample was monitored by measuring the deuterium splitting from 1D ^2H NMR spectra. Two-dimensional ^1H - ^{15}N HSQC and TROSY experiments were acquired with 1024 complex points over a sweep width of 15 ppm in the direct (^1H) dimension and 128 points over 23 ppm in the indirect (^{15}N) dimension, corresponding to an acquisition time of 48.7 ms. Spectra were acquired with a 250 ms recovery delay, and processed using exponential window functions in the direct dimension and cosine window functions in the indirect dimension. Using the same parameters, one-dimensional ^1H - ^{15}N HSQC and TROSY spectra were acquired as the first increment of the 2D spectra.

A.2.2 FLN5

	His-tag	641	651
	M H H H	H H H A S K P A P S	A E H S Y A E G E G
	ATGCATCACCAT	CACCATCAGCTAGCAAACCAGCCCATCC	GCTGAACACTCTTATGCTGAAGGTGAAGGT
661	L V K V F D N A P A	671 E F T I F A V D T K	681 G V A R T D G G D P
	TTAGTCAAAGTATTTGATAATGCCCCAGCT	GAATTCACCTATTTTCGCCGTTGACACTAAA	GGTGTGCTCGTACCGATGGTGGTATCCA
691	F E V A I N G P D G	701 L V V D A K V T D N	711 N D G T Y G V V Y D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT	TTAGTCGTTGATGCCAAAGTTACCGATAAC	AATGACGGTACTTATGGTGTGTCTATGAT
721	A P V E G N Y N V N	731 V T L R G N P I K N	741 M P I D V K C I E G
	GCCCCAGTTGAAGGTAACATAATGTTAAT	GTCACCCTCCGTGGTAATCCAATCAAAAAT	ATGCCAATCGATGTCAAATGCATTGAAGGT
751	*		
	TAA		

A.2.3 FLN5Δ12

	His-tag	641	651
	M H H H	H H H A S K P A P S	A E H S Y A E G E G
	ATGCATCACCAT	CACCATCAGCTAGCAAACCAGCCCATCC	GCTGAACACTCTTATGCTGAAGGTGAAGGT
661	L V K V F D N A P A	671 E F T I F A V D T K	681 G V A R T D G G D P
	TTAGTCAAAGTATTTGATAATGCCCCAGCT	GAATTCACCTATTTTCGCCGTTGACACTAAA	GGTGTGCTCGTACCGATGGTGGTATCCA
691	F E V A I N G P D G	701 L V V D A K V T D N	711 N D G T Y G V V Y D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT	TTAGTCGTTGATGCCAAAGTTACCGATAAC	AATGACGGTACTTATGGTGTGTCTATGAT
721	A P V E G N Y N V N	731 V T L R G N P I *	
	GCCCCAGTTGAAGGTAACATAATGTTAAT	GTCACCCTCCGTGGTAATCCAATCTAA	

A.2.4 FLN5 A₃A₃

	His-tag	641	651
	M H H H	H H H A S K P A P S	A E H S Y A E G E G
	ATGCATCACCAT	CACCATCAGCTAGCAAACCAGCCCATCC	GCTGAACACTCTTATGCTGAAGGTGAAGGT
661	L V K V A D N A P A	671 E A T I A A V D T K	681 G V A R T D G G D P
	TTAGTCAAAGTAGCGGATAATGCCCCAGCT	GAAGCCACTATGCCCGCCGTTGACACTAAA	GGTGTGCTCGTACCGATGGTGGTATCCA
691	F E V A I N G P D G	701 L V V D A K V T D N	711 N D G T A G V V A D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT	TTAGTCGTTGATGCCAAAGTTACCGATAAC	AATGACGGTACTGCGGGTGTGTGCGGGAT
721	A P V E G N A N V N	731 V T L R G N P I K N	741 M P I D V K C I E G
	GCCCCAGTTGAAGGTAACGCAATGTTAAT	GTCACCCTCCGTGGTAATCCAATCAAAAAT	ATGCCAATCGATGTCAAATGCATTGAAGGT
751	*		
	TAA		

A.2.5 FLN5+31 RNC

	His-tag	641	FLN5	651
	M H H H	H H H A S K P A P S		A E H S Y A E G E G
	ATGCATCACCAT	CACCATCACGCTAGCAAACCAGCCCATCC		GCTGAACACTCTTATGCTGAAGGTGAAGGT
661	L V K V F D N A P A	671 E F T I F A V D T K		681 G V A R T D G G D P
	TTAGTCAAAGTATTGATAATGCCCCAGCT	GAATTCACTATTTTCGCCGTTGACACTAAA		GGTGTGCTCGTACCGATGGTGGTATCCA
691	F E V A I N G P D G	701 L V V D A K V T D N		711 N D G T Y G V V Y D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT	TTAGTCGTTGATGCCAAAGTTACCGATAAC		AATGACGGTACTTATGGTGTGTCTATGAT
721	A P V E G N Y N V N	731 V T L R G N P I K N		741 M P I D V K C I E G
	GCCCCAGTTGAAGTAACTACAATGTTAAT	GTCACCCCTCCGTGGTAATCCAATCAAAAAAT		ATGCCAATCGATGTCAAATGCATTGAAGGT
751 FLN6 linker residues	A N G E D S S F G S	761 F T E L F S T P V W	SecM stalling motif	771 I S Q A Q G I R A G
	GCCCAATGGTGAAGATTCATCATTGGTTCA	TTCACTGAGCTTTCAGCACGCCCGTCTGG		ATAAGCCAGGGCAAGGCATCCGTGCTGGC
781	P *			
	CCTTAA			

A.2.6 FLN5+31 A₃A₃ RNC

	His-tag	641	FLN5	651
	M H H H	H H H A S K P A P S		A E H S Y A E G E G
	ATGCATCACCAT	CACCATCACGCTAGCAAACCAGCCCATCC		GCTGAACACTCTTATGCTGAAGGTGAAGGT
661	L V K V A D N A P A	671 E A T I A A V D T K		681 G V A R T D G G D P
	TTAGTCAAAGTAGCGGATAATGCCCCAGCT	GAAGCCACTATGCCCGCGTTGACACTAAA		GGTGTGCTCGTACCGATGGTGGTATCCA
691	F E V A I N G P D G	701 L V V D A K V T D N		711 N D G T A G V V A D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT	TTAGTCGTTGATGCCAAAGTTACCGATAAC		AATGACGGTACTCCGGTGTGTGTCGGAT
721	A P V E G N A N V N	731 V T L R G N P I K N		741 M P I D V K C I E G
	GCCCCAGTTGAAGTAACCCCAATGTTAAT	GTCACCCCTCCGTGGTAATCCAATCAAAAAAT		ATGCCAATCGATGTCAAATGCATTGAAGGT
751 FLN6 linker residues	A N G E D S S F G S	761 F T E L F S T P V W	SecM stalling motif	771 I S Q A Q G I R A G
	GCCCAATGGTGAAGATTCATCATTGGTTCA	TTCACTGAGCTTTCAGCACGCCCGTCTGG		ATAAGCCAGGGCAAGGCATCCGTGCTGGC
781	P *			
	CCTTAA			

A.2.7 FLN5+31 NCΔ1

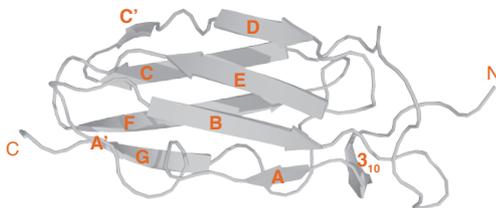
	His-tag	641	FLN5	651	
	M H H H	H H H A S	K P A P S	A E H S Y A E G E G	
	ATGCATCACCAT	CACCATCAGCTAGCAAACCAGCCCATCC		GCTGAACACTCTTATGCTGAAGGTGAAGGT	
661	L V K V F D N A P A	671	E F T I F A V D T K	681	G V A R T D G G D P
	TTAGTCAAAGTATTTGATAATGCCCCAGCT		GAATTCACTATTTTCGCCGTTGACACTAAA		GGTGTTCGTCGTACCGATGGTGGTGATCCA
691	F E V A I N G P D G	701	L V V D A K V T D N	711	N D G T Y G V V Y D
	TTTGAAGTCGCTATCAATGGTCCAGATGGT		TTAGTCGTTGATGCCAAAGTTACCGATAAC		AATGACGGTACTTATGGTGTGTCTATGAT
721	A P V E G N Y N V N	731	V T L R G N P I K N	741	M P I D V K C I E G
	GCCCCAGTTGAAGGTAAC TACAATGTTAAT		GTCACCCTCCGTGGTAATCCAATCAAAAAT		ATGCCAATCGATGTCAAATGCATTGAAGGT
751 FLN6 linker residues		761	SecM stalling motif	771	
A N G E D S S F G S		F T E L F S T P V W		I S Q A Q G I R A G	
GCCCAATGGTGAAGATTCATCATTCCGGTTCA		TTCACTGAGCTCTTCAGCACGCCCGTCTGG		ATAAGCCAGGCCGAAGGCATCCGTGCTGGC	
781					
*					
TAA					

A.2.8 Primers for SecMΔ1 mutation

FWD
ATC CGT GCT GGC TAA TAA CCA TGG ACC

REV
GGT CCA TGG TTA TTA GCC AGC ACG GAT

A.3 FLN5 crystal structure



Secondary structure
elements are labelled in
bold. PDB: 1QFH

A. 4 NMR processing scripts

NMR data were converted and processed using nmrPipe (Delaglio et al. 1995). Example scripts that were typically used are shown below.

A.4.1 Script for conversion and processing of 1D ¹H NMR experiments

An example script is shown below for the conversion and processing of 1D ¹H NMR experiments and subsequent conversion into a text file for importing into MATLAB (R2014b, The MathWorks Inc.) for analysis. Other experiments that were also acquired as a 1D NMR experiment, such as ¹H-¹⁵N SOFAST-HMQC, HNC0 and HNCA experiments were processed using similar scripts.

```

bruk2pipe -in ./fid \
  -bad 0.0 -aswap -AMX -decim 1418.6667 -dspfv 20 -grpdly 67.9896545410156 \
  -xN 16384 \
  -xT 8192 \
  -xMODE DQD \
  -xSW 14097.744 \
  -xOBS 700.233 \
  -xCAR 4.916 \
  -xLAB 1H \
  -ndim 1 \
  -out ./test.fid -verb -ov

nmrPipe -in test.fid \
| nmrPipe -fn SP -off 0.5 -end 1.00 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 58.0 -p1 0.00 -di -verb \
  -ov -out test.ft1

pipe2txt.tcl -index ppm test.ft1 > data.txt

```

A.4.2 Script for conversion and processing of pseudo-2D NMR experiments

An example script is shown below for the conversion and processing of ^1H - ^{15}N SORDID diffusion measurements and subsequent conversion into a text file for importing into MATLAB (R2014b, The MathWorks Inc.) for analysis. Other experiments that were also acquired as a pseudo-2D NMR experiment, such as relaxation rate measurements were processed using similar scripts.

```

bruk2pipe -in ./ser \
  -bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdly 67.9868774414062 \
  -xN 1024 -yN 2 \
  -xT 512 -yT 2 \
  -xMODE DQD -yMODE Real \
  -xSW 10504.202 -ySW 2.000 \
  -xOBS 700.253 -yOBS 1.000 \
  -xCAR 4.917 -yCAR 0.000 \
  -xLAB 1H -yLAB TAU \
  -ndim 2 -aq2D States \
  -out ./test.fid -verb -ov

nmrPipe -in test.fid \
| nmrPipe -fn SOL \
| nmrPipe -fn EM -lb 10.0 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 108.00 -p1 0.00 -di -verb \
| nmrPipe -fn BASE -nw 5 -nl -1ppm 7ppm 11ppm \
  -ov -out test.ft1

pipe2txt.tcl -index ppm test.ft1 > data.txt

```

A.4.3 Script for conversion, addition and processing of several pseudo-2D ^1H - ^{15}N SORDID diffusion experiments

An example script is shown below for the conversion and addition of two ^1H - ^{15}N SORDID diffusion experiments together. Each summed data set was then processed and converted into a text file for importing into MATLAB (R2014b, The MathWorks Inc.) for analysis.

```

set speclist = (103 108 111 114 117 120 123 126 129 132 135 138 141 144 147 150)

set a = 9
foreach i ($speclist)
@ b = $a + 1

mkdir {$a}
mkdir {$speclist[$a]}_{$speclist[$b]}

addNMR -in1 ../{$speclist[$a]}/test.fid -in2 ../{$speclist[$b]}/test.fid -out
./{$a}/test.fid

cd ./{$a}

bruk2pipe -in ./ser \
  -bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdlly 67.9868774414062 \
  -xN          1024 -yN          2 \
  -xT          512 -yT          2 \
  -xMODE       DQD -yMODE       Real \
  -xSW         10504.202 -ySW     2.000 \
  -xOBS        700.253 -yOBS     1.000 \
  -xCAR        4.917 -yCAR      0.000 \
  -xLAB        1H -yLAB        TAU \
  -ndim        2 -aq2D        States \
  -out ./test.fid -verb -ov

nmrPipe -in test.fid \
| nmrPipe -fn SOL \
| nmrPipe -fn EM -lb 10.0 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 108.00 -p1 0.00 -di -verb \
| nmrPipe -fn BASE -nw 5 -nl -1ppm 7ppm 11ppm \
  -ov -out test.ft1

cd ../
end

set speclist = (1 3 5 7 9 11 13 15)

mkdir sordid_1/
mkdir sordid_2/
rm -f sordid_1/*.txt
rm -f sordid_2/*.txt

set a=0
foreach i ($speclist)
@ a = $a + 1
cd ./{$i}

@ b = $a + 9
readROI -ndim 1 -in test.ft1 -x 1H -dy TAU 2 0 -out 2.dat
pipe2txt.tcl -index ppm 2.dat > ../sordid_2/D_{$b}.txt
readROI -ndim 1 -in test.ft1 -x 1H -dy TAU 1 0 -out 1.dat
pipe2txt.tcl -index ppm 1.dat > ../sordid_1/D_{$b}.txt

cd ../
end

cd sordid_1/
cat *.txt > dif_1
cd ../sordid_2/
cat *.txt > dif_2
cd ../

```

A.4.4 Script for conversion and processing of 2D experiments

An example script is shown below for the conversion and processing of ^1H - ^{15}N SOFAST-HMQC experiments. Other 2D experiments such as ^1H - ^{15}N HSQC, ^1H - ^{15}N TROSY and ^1H - ^{13}C HMQC experiments, were processed using similar scripts.

```

bruk2pipe -in ./ser \
  -bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdl 67.9868774414062 \
  -xN 1024 -yN 128 \
  -xT 512 -yT 64 \
  -xMODE DQD -yMODE States-TPPI \
  -xSW 10504.202 -ySW 1632.387 \
  -xOBS 700.253 -yOBS 70.964 \
  -xCAR 4.928 -yCAR 118.250 \
  -xLAB HN -yLAB 15N \
  -ndim 2 -aq2D States \
  -out ./test.fid -verb -ov

nmrPipe -in test.fid \
| nmrPipe -fn SP -off 0.5 -end 1.00 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 260.00 -p1 0.00 -di -verb \
| nmrPipe -fn EXT -left -sw \
| nmrPipe -fn TP \
| nmrPipe -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 -90.00 -p1 180.00 -di -verb \
  -ov -out test.ft2

```

A.4.5 Script for conversion, addition and processing of several 2D experiments to a single data set

An example script is shown below for the conversion of multiple 2D ^1H - ^{15}N SOFAST-HMQC experiments before addition of all the FIDs. The summed data set was then processed. Other 2D experiments which required addition of multiple data sets, such as ^1H - ^{15}N HSQC, ^1H - ^{15}N TROSY and ^1H - ^{13}C HMQC experiments, were processed using similar scripts.

```

set speclist = (10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58)

set a=1

foreach i ($speclist)
cd ./${speclist[$a]}

bruk2pipe -in ./ser \
  -bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdl 67.9868774414062 \
  -xN 1024 -yN 128 \
  -xT 512 -yT 64 \
  -xMODE DQD -yMODE States-TPPI \
  -xSW 10504.202 -ySW 1632.387 \
  -xOBS 700.253 -yOBS 70.964 \
  -xCAR 4.928 -yCAR 118.250 \
  -xLAB HN -yLAB 15N \
  -ndim 2 -aq2D States \
  -out ./test.fid -verb -ov

cd ../
@ a = $a + 1
end

addNMR -in1 ./10/test.fid -in2 ./13/test.fid -out ./test.fid
set a=1
foreach i ($speclist)
@ b = $a + 1
addNMR -in1 ./${speclist[$a]}/test.fid -in2 ./test.fid -out ./test.fid
@ a = $a + 1

```

end

```
nmrPipe -in test.fid \
| nmrPipe -fn EM -lb 10.0 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 260.00 -p1 0.00 -di -verb \
| nmrPipe -fn EXT -left -sw \
| nmrPipe -fn TP \
| nmrPipe -fn SP -off 0.5 -end 1.00 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 -90.00 -p1 180.00 -di -verb \
-ov -out test.ft2
```

A.4.6 Script for conversion and processing of 3D NMR experiments

```
bruk2pipe -in ./ser \
-bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdlly 67.9868774414062 \
-xN 1024 -yN 40 -zN 40 \
-xT 512 -yT 20 -zT 20 \
-xMODE DQD -yMODE Echo-AntiEcho -zMODE States-TPPI-N \
-xSW 10504.202 -ySW 1632.387 -zSW 880.592 \
-xOBS 700.263 -yOBS 70.965 -zOBS 176.111 \
-xCAR 4.922 -yCAR 118.248 -zCAR 175.601 \
-xLAB HNx -yLAB 15Ny -zLAB 13C \
-ndim 3 -aq2D States \
-out ./ft/test%03d.fid -verb -ov

xyz2pipe -in $TEMP/test%03d.fid -x -verb \
| nmrPipe -fn POLY -time \
| nmrPipe -fn SOL \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 145.6 -p1 0 -di -verb \
| nmrPipe -fn EXT -x1 5.5ppm -xn 11ppm -sw \
| pipe2xyz -out $TEMP/test%03d.ft3 -x

xyz2pipe -in $TEMP/test%03d.ft3 -z -verb \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto -di \
| pipe2xyz -out $TEMP/test%03d.ft3 -z -inPlace

xyz2pipe -in $TEMP/test%03d.ft3 -y -verb \
| nmrPipe -fn LP -fb -ps0-0 \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 -90 -p1 0.0 -di -verb \
| pipe2xyz -out $TEMP/test%03d.ft3 -y -inPlace

#Hilbert inverse Fourier Transformation
xyz2pipe -in $TEMP/test%03d.ft3 -z -verb \
| nmrPipe -fn HT -auto \
| nmrPipe -fn FT -inv \
| nmrPipe -fn ZF -auto -inv \
| nmrPipe -fn LP -fb \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 0 -p1 0 -di -verb \
| pipe2xyz -out $TEMP/test%03d.ft3 -z -inPlace

xyz2pipe -in $TEMP/test%03d.ft3 -x -verb \
| nmrPipe -fn POLY -auto \
| pipe2xyz -out $TEMP/test%03d.ft3 -x -inPlace

xyz2pipe -in $TEMP/test%03d.ft3 -out ./final.ft
proj3D.tcl
```

A.4.7 Script for conversion, addition and processing of several 3D NMR experiments into a single data set

```

set speclist = (10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58)

set a=1

foreach i ($speclist)
cd ./${speclist[$a]}

bruk2pipe -in ./ser \
  -bad 0.0 -aswap -AMX -decim 1904 -dspfv 20 -grpdl 67.9868774414062 \
  -xN 1024 -yN 40 -zN 40 \
  -xT 512 -yT 20 -zT 20 \
  -xMODE DQD -yMODE Echo-AntiEcho -zMODE States-TPPI-N \
  -xSW 10504.202 -ySW 1632.387 -zSW 880.592 \
  -xOBS 700.263 -yOBS 70.965 -zOBS 176.111 \
  -xCAR 4.922 -yCAR 118.248 -zCAR 175.601 \
  -xLAB HNx -yLAB 15Ny -zLAB 13C \
  -ndim 3 -aq2D States \
  -out ./ft/test%03d.fid -verb -ov

cd ../
@a = $a + 1
end

addNMR -in1 ./10/ft/test%03d.fid -in2 ./13/ft/test%03d.fid -out ./ft/test%03d.fid
set a=1
foreach i ($speclist)
@a = $a + 1
addNMR -in1 ./${speclist[$a]}/ft/test%03d.fid -in2 ./ft/test%03d.fid -out
\ ./ ft/test%03d.fid
@a = $a + 1
end

xyz2pipe -in $TEMP/test%03d.fid -x -verb \
| nmrPipe -fn POLY -time \
| nmrPipe -fn SOL \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 145.6 -p1 0 -di -verb \
| nmrPipe -fn EXT -x1 5.5ppm -xn 11ppm -sw \
| pipe2xyz -out $TEMP/test%03d.ft3 -x

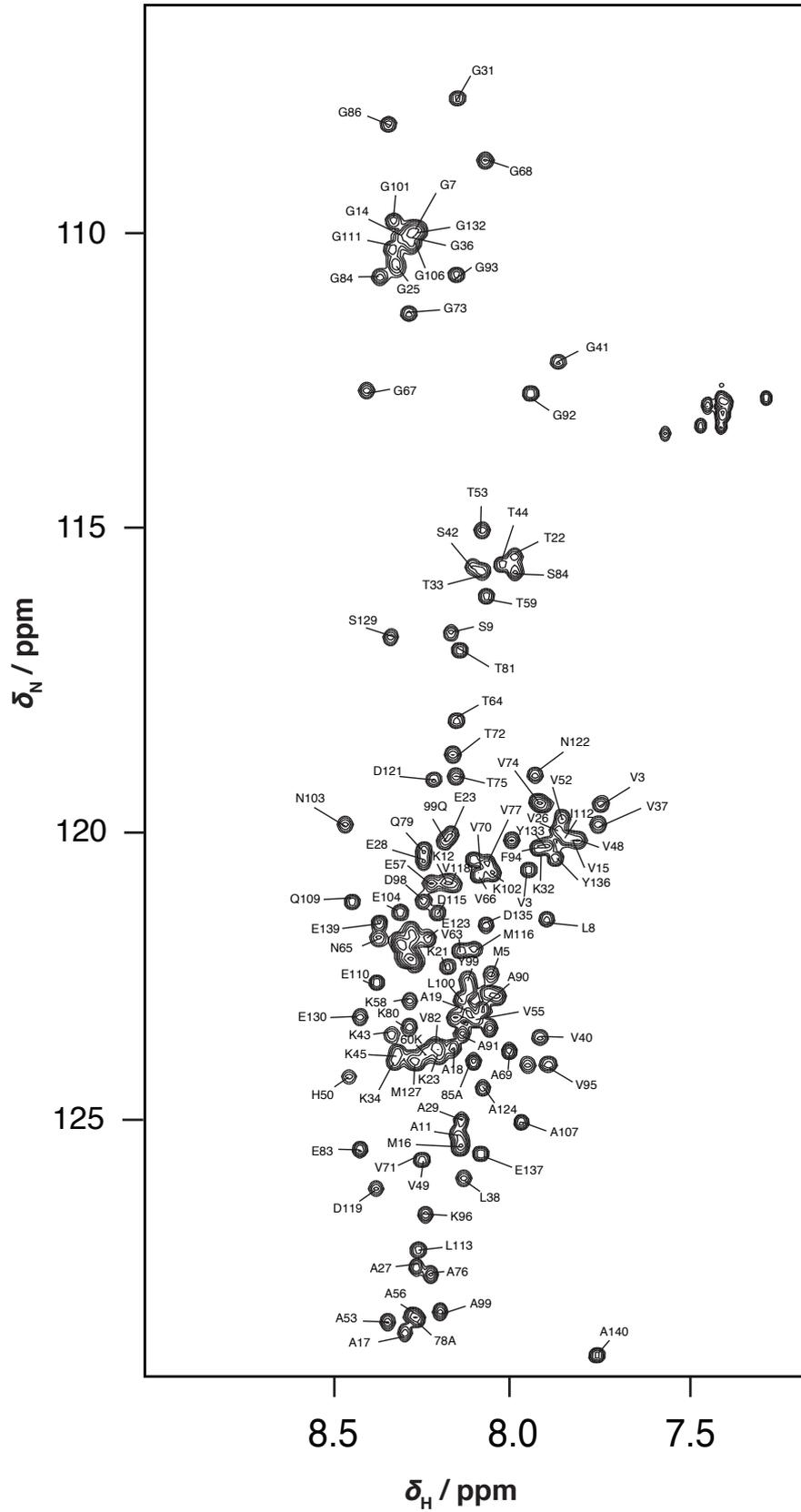
xyz2pipe -in $TEMP/test%03d.ft3 -z -verb \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto -di \
| pipe2xyz -out $TEMP/test%03d.ft3 -z -inPlace

xyz2pipe -in $TEMP/test%03d.ft3 -y -verb \
| nmrPipe -fn LP -fb -ps0-0 \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -auto \
| nmrPipe -fn PS -p0 -90 -p1 0.0 -di -verb \
| pipe2xyz -out $TEMP/test%03d.ft3 -y -inPlace

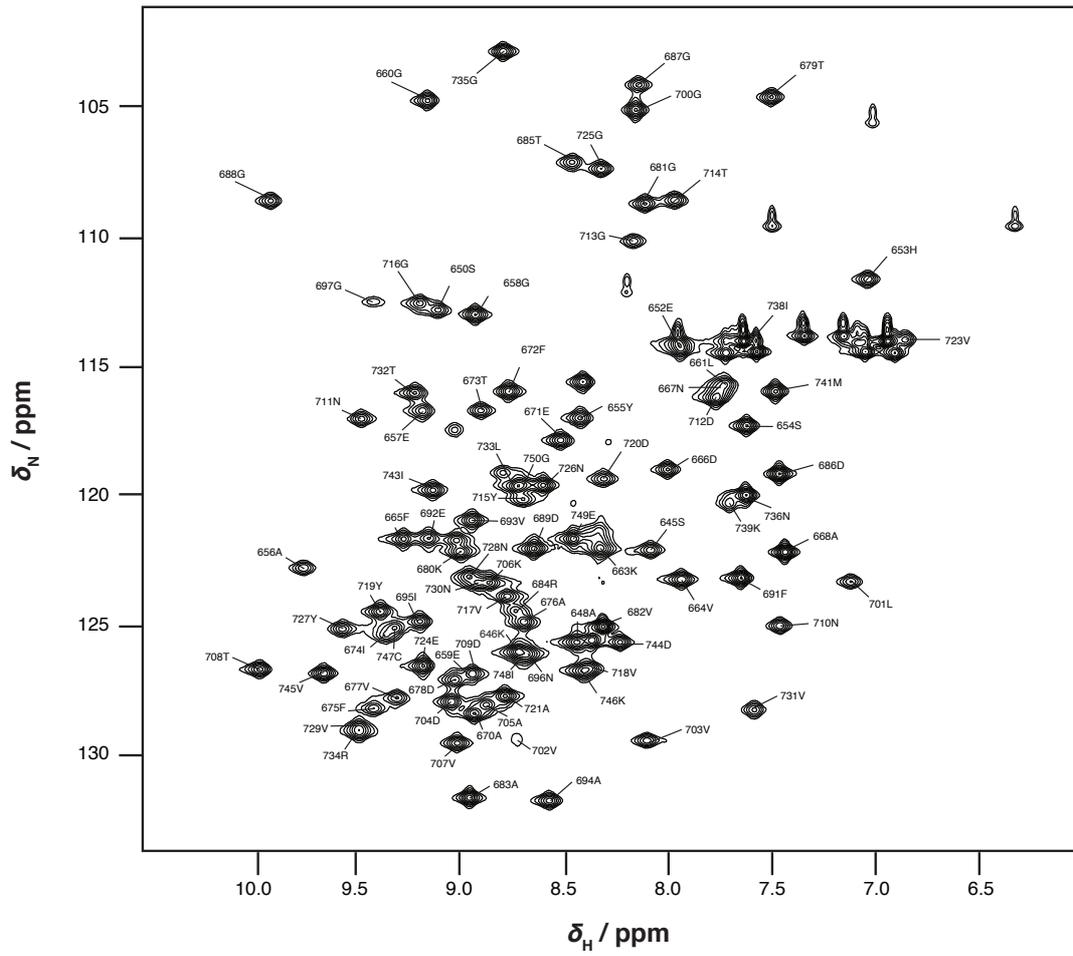
#Hilbert inverse Fourier Transformation
xyz2pipe -in $TEMP/test%03d.ft3 -z -verb \
| nmrPipe -fn HT -auto \
| nmrPipe -fn FT -inv \
| nmrPipe -fn ZF -auto -inv \
| nmrPipe -fn LP -fb \
| nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 0 -p1 0 -di -verb \
| pipe2xyz -out $TEMP/test%03d.ft3 -z -inPlace

xyz2pipe -in $TEMP/test%03d.ft3 -x -verb \

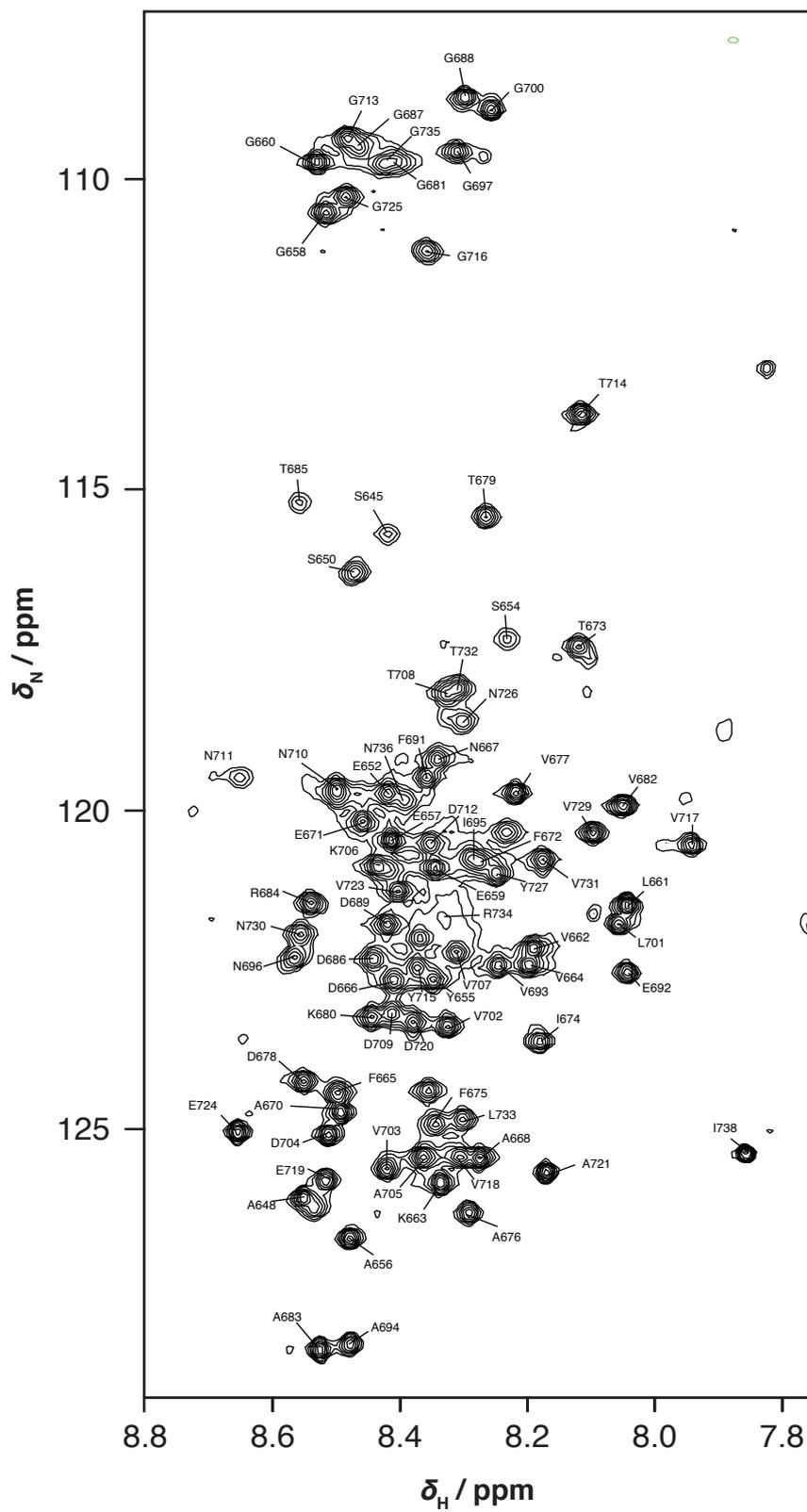
```

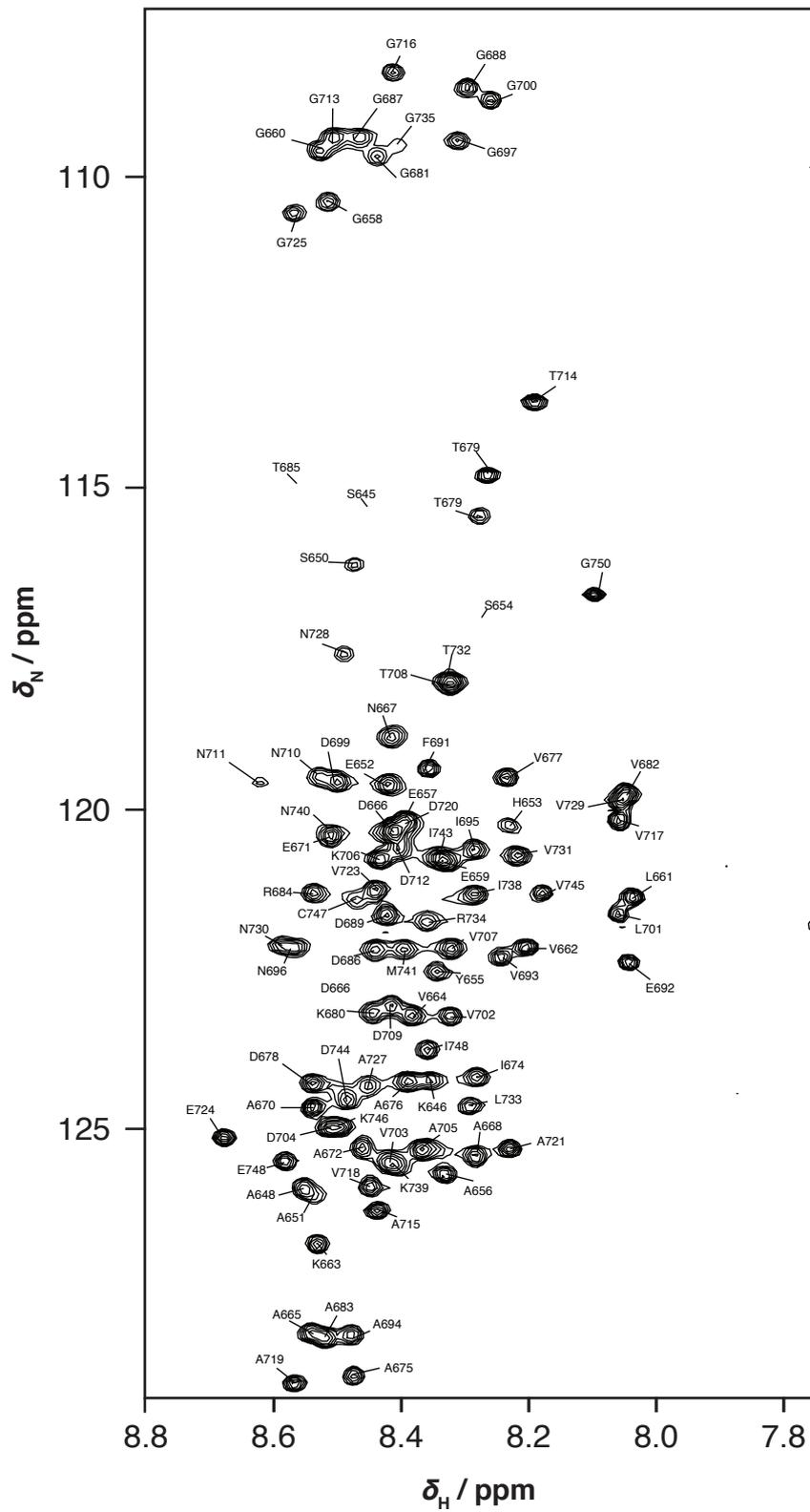

A.5.2 α -Synuclein

A.5.3 FLN5



A.5.4 FLN5Δ12



A.5.5 FLN5 A₃A₃

References

- Agashe VR, Guha S, Chang H-C, et al (2004) Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117:199–209.
- Akopian D, Shen K, Zhang X, Shan S-O (2013) Signal recognition particle: an essential protein-targeting machine. *Annu Rev Biochem* 82:693–721. doi: 10.1146/annurev-biochem-072711-164732
- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.
- Augustyniak R, Ferrage F, Damblon C, et al (2012a) Efficient determination of diffusion coefficients by monitoring transport during recovery delays in NMR. *Chem Commun (Camb)* 48:5307. doi: 10.1039/c2cc30578j
- Augustyniak R, Ferrage F, Damblon C, et al (2012b) Efficient determination of diffusion coefficients by monitoring transport during recovery delays in NMR. *Chem Commun (Camb)* 48:5307. doi: 10.1039/c2cc30578j
- Augustyniak R, Ferrage F, Paquin R, et al (2011) Methods to determine slow diffusion coefficients of biomolecules: applications to Engrailed 2, a partially disordered protein. *Journal of Biomolecular NMR* 50:209–218. doi: 10.1007/s10858-011-9510-8
- Ádén J, Wittung-Stafshede P (2014) Folding of an Unfolded Protein by Macromolecular Crowding in Vitro. *Biochemistry* 140403112343003. doi: 10.1021/bi500222g
- Baldwin AJ, Anthony-Cahill SJ, Knowles TPJ, et al (2008) Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution NMR diffusion measurements. *Angew Chem Int Ed Engl* 47:3385–3387. doi: 10.1002/anie.200703915
- Barbet-Massin E, Huang C-T, Daebel V, et al (2015) Site-Specific Solid-State NMR Studies of “Trigger Factor” in Complex with the Large Ribosomal Subunit 50S. *Angewandte Chemie-International Edition n/a–n/a*. doi: 10.1002/anie.201409393
- Barbieri L, Luchinat E, Banci L (2016) Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc* 11:1101–1111. doi: 10.1038/nprot.2016.061
- Bax A (2011) Triple resonance three-dimensional protein NMR: before it became a black box. *J Magn Reson* 213:442–445. doi: 10.1016/j.jmr.2011.08.003
- Becker T, Bhushan S, Jarasch A, et al (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326:1369–1373. doi: 10.1126/science.1178535
- Berndt U, Oellerer S, Zhang Y, et al (2009) A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc Natl Acad Sci USA* 106:1398–1403. doi: 10.1073/pnas.0808584106
- Bernini A, Venditti V, Spiga O, Niccolai N (2009) Progress in Nuclear Magnetic Resonance Spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* 54:278–289. doi: 10.1016/j.pnmrs.2008.10.003
- Bertini I, Luchinat C, Parigi G (2001) *Solution NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*. Elsevier Science BV

- Bertrand K, Reverdatto S, Burz DS, et al (2012) Structure of proteins in eukaryotic compartments. *J Am Chem Soc* 134:12798–12806. doi: 10.1021/ja304809s
- Bicknese S, Periasamy N, Shohet SB, Verkman AS (1993) Cytoplasmic viscosity near the cell plasma membrane: measurement by evanescent field frequency-domain microfluorimetry. *Biophysj* 65:1272–1282. doi: 10.1016/S0006-3495(93)81179-2
- Bloembergen N, Purcell EM, Pound RV (1948) Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Physical Review* 73:679–712. doi: 10.1103/PhysRev.73.679
- Borgia MB, Borgia A, Best RB, et al (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474:662–665. doi: 10.1038/nature10099
- Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W (2008) Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat Struct Mol Biol* 15:494–499. doi: 10.1038/nsmb.1402
- Brindle K, Williams S-P, Boulton M (1989) ¹⁹F NMR detection of a fluorine-labelled enzyme in vivo. *FEBS Letters* 255:121–124. doi: 10.1016/0014-5793(89)81073-7
- Bryant JE, Lecomte JTJ, Lee AL, et al (2005) Protein dynamics in living cells. *Biochemistry* 44:9275–9279. doi: 10.1021/bi050786j
- Buhr F, Jha S, Thommen M, et al (2016) Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Molecular Cell* 61:341–351. doi: 10.1016/j.molcel.2016.01.008
- Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat Methods* 3:91–93. doi: 10.1038/nmeth851
- Cabrita LD, Cassaignau AME, Launay HMM, et al (2016) A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol* 23:278–285. doi: 10.1038/nsmb.3182
- Cabrita LD, Dobson CM, Christodoulou J (2010) Protein folding on the ribosome. *Current Opinion in Structural Biology* 20:33–45. doi: 10.1016/j.sbi.2010.01.005
- Cabrita LD, Hsu S-TD, Launay H, et al (2009) Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci USA* 106:22239–22244. doi: 10.1073/pnas.0903750106
- Cai S, Seu C, Kovacs Z, et al (2006) Sensitivity Enhancement of Multidimensional NMR Experiments by Paramagnetic Relaxation Effects. *J Am Chem Soc* 128:13474–13478. doi: 10.1021/ja0634526
- Camilloni C, De Simone A, Vranken WF, Vendruscolo M (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* 51:2224–2231. doi: 10.1021/bi3001825
- Cassaignau AME, Launay HMM, Karyadi M-E, et al (2016) A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat Protoc* 11:1492–1507. doi: 10.1038/nprot.2016.101
- Cavanagh J, Fairbrother WJ, Arthur G Palmer I, et al (2010) *Protein NMR Spectroscopy*. Academic Press
- Chang CA, Chen C-Y, Chen H-Y (1999) Determination of Stability Constants of Metal Complexes by Capillary Electrophoresis. *Journal of the Chinese Chemical Society* 46:519–528.

- Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366. doi: 10.1146/annurev.biochem.75.101304.123901
- Chow CC, Chow C, Raghunathan V, et al (2003) Chain length dependence of apomyoglobin folding: structural evolution from misfolded sheets to native helices. *Biochemistry* 42:7090–7099. doi: 10.1021/bi0273056
- Christodoulou J, Larsson G, Fucini P, et al (2004) Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. *Proc Natl Acad Sci USA* 101:10949–10954. doi: 10.1073/pnas.0400928101
- Chung K-C, Yu H-Y, Ahn S-D (2011) Convection Effects on PGSE-NMR Self-Diffusion Measurements at Low Temperature: Investigation into Sources of Induced Convective Flows. *Bulletin of the Korean Chemical Society* 32:1970–1974. doi: 10.5012/bkcs.2011.32.6.1970
- Clark PL, King J (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem* 276:25411–25420. doi: 10.1074/jbc.M008490200
- Clore GM, Gronenborn AM (1994) NMR of Proteins. *Protein Science*
- Clore GM, Iwahara J (2009) Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem Rev* 109:4108–4139. doi: 10.1021/cr900033p
- Clore GM, Starich MR (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses.
- Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^1H N, ^{15}N , and ^{13}C chemical shifts in ^{15}N , ^{13}C -labelled proteins. *Journal of Magnetic Resonance* (1969) 97:213–217. doi: 10.1016/0022-2364(92)90252-3
- Dalgarno DC, Levine BA, Williams R (1983) Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins.
- de Alba E, Tjandra N (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy* 40:175–197. doi: 10.1016/S0079-6565(01)00042-5
- de Prat Gay G, Ruiz-Sanz J, Neira JL, et al (1995) Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N terminus in vitro. Parallels with the protein folding pathway. *Journal of Molecular Biology* 254:968–979. doi: 10.1006/jmbi.1995.0669
- Deckert A, Waudby CA, Wlodarski T, et al (2016) Structural characterization of the interaction of α -synuclein nascent chains with the ribosomal surface and trigger factor. *Proc Natl Acad Sci USA*. doi: 10.1073/pnas.1519124113
- Dedmon MM, Patel CN, Young GB, Pielak GJ (2002) FlgM gains structure in living cells. *Proc Natl Acad Sci USA* 99:12681–12684. doi: 10.1073/pnas.202331299
- Delaglio F, Grzesiek S, Vuister GW, et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6:277–293.
- Dhar A, Samiotakis A, Ebbinghaus S, et al (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci USA* 107:17586–17591. doi: 10.1073/pnas.1006760107
- Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29:7133–7155.

- Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. *Annu Rev Biophys* 37:289–316. doi: 10.1146/annurev.biophys.37.092707.153558
- Douglas AS, Donald MW (1971) *Principles of instrumental analysis*. Holt, Rinhart, Winston, New York
- Dunkle JA, Cate JHD (2010) Ribosome structure and dynamics during translocation and termination. *Annu Rev Biophys* 39:227–244. doi: 10.1146/annurev.biophys.37.032807.125954
- Ebbinghaus S, Dhar A, McDonald JD, Gruebele M (2010) Protein folding stability and dynamics imaged in a living cell. *Nat Methods* 7:319–323. doi: 10.1038/nmeth.1435
- Efron B, Tibshirani RJ (1994) *An Introduction to the Bootstrap*. Chapman & Hall
- Eichmann C, Preissler S, Riek R, Deuerling E (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc Natl Acad Sci USA* 107:9111–9116. doi: 10.1073/pnas.0914300107
- Eletsky A, Moreira O, Kovacs H, Pervushin K (2003) A novel strategy for the assignment of side-chain resonances in completely deuterated large proteins using ¹³C spectroscopy. *Journal of Biomolecular NMR* 26:167–179.
- English BP, Haurlyuk V, Sanamrad A, et al (2011) Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc Natl Acad Sci USA* 108:E365–73. doi: 10.1073/pnas.1102255108
- Ernst RR, Anderson WA (1966) Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev Sci Instrum* 37:93. doi: 10.1063/1.1719961
- Evans MS, Sander IM, Clark PL (2008) Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *Journal of Molecular Biology* 383:683–692. doi: 10.1016/j.jmb.2008.07.035
- Evans MS, Ugrinov KG, Frese M-A, Clark PL (2005) Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods* 108:757–762. doi: 10.1038/nmeth790
- Faure G, Ogurtsov AY, Shabalina SA, Koonin EV (2016) Role of mRNA structure in the control of protein folding. *Nucleic Acids Res*. doi: 10.1093/nar/gkw671
- Fernández C, Wider G (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Current Opinion in Structural Biology* 13:570–580. doi: 10.1016/j.sbi.2003.09.009
- Ferrage F, Zoonens M, Warschawski DE, et al (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J Am Chem Soc* 125:2541–2545. doi: 10.1021/ja0211407
- Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* 418:207–211. doi: 10.1038/nature00860
- Frank J, Agrawal RK (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406:318–322. doi: 10.1038/35018597
- Frank J, Gonzalez RL (2010) Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu Rev Biochem* 79:381–412. doi: 10.1146/annurev-biochem-060408-173330
- Fredrick K, Noller HF (2003) Catalysis of ribosomal translocation by sparsomycin. *Science*

- 300:1159–1162. doi: 10.1126/science.1084571
- Freedberg DI, Selenko P (2014) Live Cell NMR. *Annu Rev Biophys* 43:171–192. doi: 10.1146/annurev-biophys-051013-023136
- Frueh DP, Goodrich AC, Mishra SH, Nichols SR (2013) NMR methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology* 23:734–739. doi: 10.1016/j.sbi.2013.06.016
- Frydman J, Erdjument-Bromage H, Tempst P, Hartl FU (1999) Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol* 6:697–705. doi: 10.1038/10754
- Fulle S, Gohlke H (2009) Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. *Journal of Molecular Biology* 387:502–517. doi: 10.1016/j.jmb.2009.01.037
- Gershenson A, Gierasch LM (2011) Protein folding in the cell: challenges and progress. *Current Opinion in Structural Biology* 21:32–41. doi: 10.1016/j.sbi.2010.11.001
- Gil S, Hošek T, Solyom Z, et al (2013) NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. *Angew Chem Int Ed Engl*. doi: 10.1002/anie.201304272
- Gong F, Yanofsky C (2002) Instruction of translating ribosome by nascent peptide. *Science* 297:1864–1867. doi: 10.1126/science.1073997
- Grzesiek S, Bax A (1992a) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)* 96:432–440. doi: 10.1016/0022-2364(92)90099-S
- Grzesiek S, Bax A (1992b) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* 114:6291–6293. doi: 10.1021/ja00042a003
- Gumbart J, Schreiner E, Wilson DN, et al (2012) Mechanisms of SecM-mediated stalling in the ribosome. *Biophys J* 103:331–341. doi: 10.1016/j.bpj.2012.06.005
- Hamatsu J, O'Donovan D, Tanaka T, et al (2013) High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc* 135:1688–1691. doi: 10.1021/ja310928u
- Han J-H, Batey S, Nickson AA, et al (2007) The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* 8:319–330. doi: 10.1038/nrm2144
- Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166:557–580. doi: 10.1016/S0022-2836(83)80284-8
- Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5:1065–1074. doi: 10.1038/4176
- Hartl FU, Bracher A (2011) Molecular chaperones in protein folding and proteostasis.
- Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16:574–581. doi: 10.1038/nsmb.1591
- Helm L (2006) Relaxivity in paramagnetic systems: Theory and mechanisms. *Progress in Nuclear Magnetic Resonance Spectroscopy* 49:45–64. doi: 10.1016/j.pnmrs.2006.03.003

- Hiller S, Wider G, Etezady-Esfarjani T, et al (2005) Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. *Journal of Biomolecular NMR* 32:61–70. doi: 10.1007/s10858-005-3070-8
- Hoffmann A, Bukau B, Kramer G (2010) Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta* 1803:650–661. doi: 10.1016/j.bbamcr.2010.01.017
- Holtkamp W, Kokic G, Jäger M, et al (2015) Cotranslational protein folding on the ribosome monitored in real time. *Science* 350:1104–1107. doi: 10.1126/science.aad0344
- Hsu S-TD, Cabrita LD, Fucini P, et al (2009) Structure, Dynamics and Folding of an Immunoglobulin Domain of the Gelation Factor (ABP-120) from *Dictyostelium discoideum*. *Journal of Molecular Biology* 388:865–879. doi: 10.1016/j.jmb.2009.02.063
- Hsu S-TD, Fucini P, Cabrita LD, et al (2007) Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc Natl Acad Sci USA* 104:16516–16521. doi: 10.1073/pnas.0704664104
- Huber D, Rajagopalan N, Preissler S, et al (2011) SecA interacts with ribosomes in order to facilitate posttranslational translocation in bacteria. *Molecular Cell* 41:343–353. doi: 10.1016/j.molcel.2010.12.028
- Hyberts SG, Arthanari H, Wagner G (2012) Applications of non-uniform sampling and processing. *Top Curr Chem* 316:125–148. doi: 10.1007/128_2011_187
- Ignatova Z, Gierasch LM (2004) Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci USA* 101:523–528. doi: 10.1073/pnas.0304533101
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of ^1H , ^{13}C , and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29:4659–4667.
- Inomata K, Ohno A, Tochio H, et al (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 457:106–109. doi: 10.1038/nature07839
- Ito K (2005) Ribosome-based protein folding systems are structurally divergent but functionally universal across biological kingdoms. *Molecular Microbiology* 57:313–317. doi: 10.1111/j.1365-2958.2005.04696.x
- Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2: Evidence for a two-state transition. *Biochemistry* 30:10428–10435.
- Jaravine VA, Cordier F, Grzesiek S (2004) Quantification of H/D Isotope Effects on Protein Hydrogen-bonds by $^3\text{J}_{\text{NC}}$ and $^1\text{J}_{\text{NC}}$ Couplings and Peptide Group ^{15}N and $^{13}\text{C}'$ Chemical Shifts. *Journal of Biomolecular NMR* 29:309–318. doi: 10.1023/B:JNMR.0000032516.87434.35
- Johnson CS Jr (1999) Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34:203–256.
- Kaake RM, Wang X, Burke A, et al (2014) A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells. *Mol Cell Proteomics* 13:3533–3543. doi: 10.1074/mcp.M114.042630
- Kaiser CM, Chang H-C, Agashe VR, et al (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* 444:455–460. doi: 10.1038/nature05225
- Kanelis V, Forman-Kay JD, Kay LE (2001) Multidimensional NMR methods for protein structure determination. *IUBMB life* 52:291–302. doi: 10.1080/152165401317291147

- Karplus M, Weaver DL (1994) Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* 3:650–668. doi: 10.1002/pro.5560030413
- Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. *J Magn Reson* 89:496–514. doi: 10.1016/0022-2364(90)90333-5
- Kelkar DA, Khushoo A, Yang Z, Skach WR (2012) Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. *J Biol Chem* 287:2568–2578. doi: 10.1074/jbc.M111.318766
- Kiefhaber T (1995) Kinetic traps in lysozyme folding. *Proc Natl Acad Sci USA* 92:9029–9033.
- Kim SJ, Yoon JS, Shishido H, et al (2015) Translational tuning optimizes nascent protein folding in cells. *Science* 348:444–448. doi: 10.1126/science.aaa3974
- Kimchi-Sarfaty C, Oh JM, Kim I-W, et al (2007) A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528. doi: 10.1126/science.1135308
- Kleckner IR, Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *BBA - Proteins and Proteomics* 1814:942–968. doi: 10.1016/j.bbapap.2010.10.012
- Kleizen B, van Vlijmen T, de Jonge HR, Braakman I (2005) Folding of CFTR is predominantly cotranslational. *Molecular Cell* 20:277–287. doi: 10.1016/j.molcel.2005.09.007
- Komar AA (2009) A pause for thought along the co-translational folding pathway. *Trends in Biochemical Sciences* 34:16–24. doi: 10.1016/j.tibs.2008.10.002
- Kosol S, Contreras-Martos S, Cedeño C, Tompa P (2013) Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* 18:10802–10828. doi: 10.3390/molecules180910802
- Kosolapov A, Deutsch C (2009) Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol* 16:405–411. doi: 10.1038/nsmb.1571
- Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes—a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* 46:131–155. doi: 10.1016/j.pnmrs.2005.03.001
- König I, Zarrine-Afsar A, Aznauryan M, et al (2015) Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods* 12:773–779. doi: 10.1038/nmeth.3475
- Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* 16:589–597. doi: 10.1038/nsmb.1614
- Land A, Zonneveld D, Braakman I (2003) Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J* 17:1058–1067. doi: 10.1096/fj.02-0811com
- Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance* 187:163–169. doi: 10.1016/j.jmr.2007.04.002
- Li C, Wang G-F, Wang Y, et al (2010) Protein 19F NMR in *Escherichia coli*. *J Am Chem Soc* 132:321–327. doi: 10.1021/ja907966n
- Li C, Wang Y, Pielak GJ (2009) Translational and rotational diffusion of a small globular protein

- under crowded conditions. *J Phys Chem B* 113:13390–13392. doi: 10.1021/jp907744m
- Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520. doi: 10.1126/science.1208351
- Lu J, Deutsch C (2005a) Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* 44:8230–8243. doi: 10.1021/bi050372q
- Lu J, Deutsch C (2005b) Folding zones inside the ribosomal exit tunnel. *Nat Struct Mol Biol* 12:1123–1129. doi: 10.1038/nsmb1021
- Lucas LH, Larive CK (2004) Measuring ligand-protein binding using NMR diffusion experiments. *Concepts in Magnetic Resonance* 20A:24–41. doi: 10.1002/cmr.a.10094
- Madl T, Güttler T, Görlich D, Sattler M (2011) Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angew Chem Int Ed Engl* 50:3993–3997. doi: 10.1002/anie.201007168
- Maier R, Eckert B, Scholz C, et al (2003) Interaction of trigger factor with the ribosome. *Journal of Molecular Biology* 326:585–592.
- Makeyev EV, Kolb VA, Spirin AS (1996) Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Letters* 378:166–170.
- Markwick PRL, Malliavin T, Nilges M (2008) Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol* 4:e1000168. doi: 10.1371/journal.pcbi.1000168
- Merkel JS, Regan L (2000) Modulating Protein Folding Rates in Vivo and in Vitro by Side-chain Interactions between the Parallel Strands of Green Fluorescent Protein. *Journal of Biological Chemistry* 275:29200–29206. doi: 10.1074/jbc.M004734200
- Miklos AC, Sarkar M, Wang Y, Pielak GJ (2011) Protein crowding tunes protein stability. *J Am Chem Soc* 133:7116–7120. doi: 10.1021/ja200067p
- Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* 34:601–611. doi: 10.1016/j.tibs.2009.07.004
- Mulder FAA, Bouakaz L, Lundell A, et al (2004) Conformation and dynamics of ribosomal stalk protein L12 in solution and on the ribosome. *Biochemistry* 43:5930–5936. doi: 10.1021/bi0495331
- Müntener T, Häussinger D, Selenko P, Theillet F-X (2016) In-Cell Protein Structures from 2D NMR Experiments. *J Phys Chem Lett* 7:2821–2825. doi: 10.1021/acs.jpcllett.6b01074
- Nakatogawa H, Ito K (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell* 108:629–636.
- Neira JL, Fersht AR (1999) Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. *Journal of Molecular Biology* 285:1309–1333. doi: 10.1006/jmbi.1998.2249
- Netzer WJ, Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388:343–349. doi: 10.1038/41024
- Nilsson OB, Hedman R, Marino J, et al (2015) Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *CellReports* 12:1533–1540. doi: 10.1016/j.celrep.2015.07.065
- Ogino S, Kubo S, Umemoto R, et al (2009) Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-

- Forming Toxin, Streptolysin O. *J Am Chem Soc* 131:10834–10835. doi: 10.1021/ja904407w
- Oh E, Becker AH, Sandikci A, et al (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* 147:1295–1308. doi: 10.1016/j.cell.2011.10.044
- Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Current Opinion in Structural Biology* 14:70–75. doi: 10.1016/j.sbi.2004.01.009
- Ortega A, García de la Torre J (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* 127:12764–12765. doi: 10.1021/ja053080l
- Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 131:373–378. doi: 10.1006/jmre.1998.1361
- Otting G (2010) Protein NMR Using Paramagnetic Ions. *Annu Rev Biophys* 39:387–405. doi: 10.1146/annurev.biophys.093008.131321
- Pedersen S (1978) Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* 14:179–190. doi: 10.1016/0092-8674(78)90312-4
- Perez CE, Gonzalez RL (2011) In vitro and in vivo single-molecule fluorescence imaging of ribosome-catalyzed protein synthesis. *Curr Opin Chem Biol* 15:853–863. doi: 10.1016/j.cbpa.2011.11.002
- Pervushin K, Riek R, Wider G, 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* 94:12366–12371.
- Pervushin K, Vögeli B, Eletsky A (2002) Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *J Am Chem Soc* 124:12898–12902.
- Rantaharju J, Mareš J, Vaara J (2014) Spin dynamics simulation of electron spin relaxation in Ni(2+) (aq). *Journal of Chemical Physics* 141:014109. doi: 10.1063/1.4885050
- Rauch T, Hundley HA, Pfund C, et al (2005) Dissecting functional similarities of ribosome-associated chaperones from *Saccharomyces cerevisiae* and *Escherichia coli*. *Molecular Microbiology* 57:357–365. doi: 10.1111/j.1365-2958.2005.04690.x
- Robinson KE, Reardon PN, Spicer LD (2012) In-cell NMR spectroscopy in *Escherichia coli*. *Methods Mol Biol* 831:261–277. doi: 10.1007/978-1-61779-480-3_15
- Robustelli P, Kohlhoff K, Cavalli A, Vendruscolo M (2010) Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* 18:923–933. doi: 10.1016/j.str.2010.04.016
- Rovnyak D, Hoch JC, Stern AS, Wagner G (2004) Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *Journal of Biomolecular NMR* 30:1–10. doi: 10.1023/B:JNMR.0000042946.04002.19
- Rule GS, Hitchens TK (2006) *Fundamentals of Protein NMR Spectroscopy*. Springer
- Ruschak AM, Kay LE (2012) Proteasome allostery as a population shift between interchanging conformers. *Proceedings of the National Academy of Sciences* 109:E3454–62. doi: 10.1073/pnas.1213640109

- Rutkowska A, Mayer MP, Hoffmann A, et al (2008) Dynamics of trigger factor interaction with translating ribosomes. *J Biol Chem* 283:4124–4132. doi: 10.1074/jbc.M708294200
- Rückert M, Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* 122:7793–7797. doi: 10.1021/ja001068h
- Saio T, Guan X, Rossi P, et al (2014) Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* 344:1250494. doi: 10.1126/science.1250494
- Sakakibara D, Sasaki A, Ikeya T, et al (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 457:102–105. doi: 10.1038/nature07814
- Sakon JJ, Weninger KR (2010) Detecting the conformation of individual proteins in live cells. *Nat Methods* 7:203–205. doi: 10.1038/nmeth.1421
- Salmon L, Nodet G, Ozenne V, et al (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132:8407–8418. doi: 10.1021/ja101645g
- Sanz JM, Fersht AR (1993) Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry* 32:13584–13592.
- Sarkar M, Smith AE, Pielak GJ (2013) Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci USA* 110:19342–19347. doi: 10.1073/pnas.1312678110
- Sass HJ, Musco G, Stahl SJ, et al (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *Journal of Biomolecular NMR* 18:303–309.
- Sattler M, Fesik SW (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* 4:1245–1249.
- Sattler M, Schleucher J (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution.
- Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR* 33:199–211. doi: 10.1007/s10858-005-4425-x
- Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* 128:9042–9043. doi: 10.1021/ja062025p
- Schlesinger AP, Wang Y, Tadeo X, et al (2011) Macromolecular Crowding Fails To Fold a Globular Protein in Cells. *J Am Chem Soc* 133:8082–8085. doi: 10.1021/ja201206t
- Schlunzen F, Wilson DN, Tian P, et al (2005) The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure* 13:1685–1694.
- Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461:1234–1242. doi: 10.1038/nature08403
- Schuwirth BS, Borovinskaya MA, Hau CW, et al (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310:827–834. doi: 10.1126/science.1117230
- Seeliger J, Werkmüller A, Winter R (2013) Macromolecular crowding as a suppressor of human IAPP fibril formation and cytotoxicity. *PLoS ONE* 8:e69652. doi: 10.1371/journal.pone.0069652
- Seidelt B, Innis CA, Wilson DN, et al (2009) Structural insight into nascent polypeptide chain-

- mediated translational stalling. *Science* 326:1412–1415. doi: 10.1126/science.1177662
- Selenko P, Serber Z, Gadea B, et al (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc Natl Acad Sci USA* 103:11904–11909. doi: 10.1073/pnas.0604667103
- Serber Z, Dötsch V (2001) In-cell NMR spectroscopy. *Biochemistry* 40:14317–14323.
- Serber Z, Ledwidge R, Miller SM, Dötsch V (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J Am Chem Soc* 123:8895–8901.
- Serber Z, Selenko P, Hänsel R, et al (2007) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc* 1:2701–2709. doi: 10.1038/nprot.2006.181
- Serber Z, Straub W, Corsini L, et al (2004) Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *J Am Chem Soc* 126:7119–7125. doi: 10.1021/ja049977k
- Siller E, DeZwaan DC, Anderson JF, et al (2010) Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *Journal of Molecular Biology* 396:1310–1318. doi: 10.1016/j.jmb.2009.12.042
- Smith AE, Sarkar M, Young GB, Pielak GJ (2013) Amide proton exchange of a dynamic loop in cell extracts. *Protein Science* 22:1313–1319. doi: 10.1002/pro.2318
- Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C-alpha and C-beta ¹³C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492. doi: 10.1021/ja00014a071
- Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618–622. doi: 10.1038/nature05512
- Stejskal EO, Tanner JE (1964) Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *Journal of Chemical Physics* 288:288–292.
- Thanaraj TA, Argos P (1996) Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci* 5:1973–1983. doi: 10.1002/pro.5560051003
- Theillet F-X, Binolfi A, Bekei B, et al (2016) Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature* 530:45–50. doi: 10.1038/nature16531
- Theillet F-X, Binolfi A, Liokatis S, et al (2011) Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins. *Journal of Biomolecular NMR* 51:487–495. doi: 10.1007/s10858-011-9577-2
- Tjandra N (1997) Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* 278:1111–1114. doi: 10.1126/science.278.5340.1111
- Tolman JR, Ruan K (2006) NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem Rev* 106:1720–1736. doi: 10.1021/cr040429z
- Tsai A, Kornberg G, Johansson M, et al (2014) The Dynamics of SecM-Induced Translational Stalling. *Cell Reports* 7:1521–1533. doi: 10.1016/j.celrep.2014.04.033
- Tsai C-J, Sauna ZE, Kimchi-Sarfaty C, et al (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *Journal of Molecular Biology* 383:281–291. doi: 10.1016/j.jmb.2008.08.012

- Tsalkova T, Odom OW, Kramer G, Hardesty B (1998) Different conformations of nascent peptides on ribosomes. *Journal of Molecular Biology* 278:713–723. doi: 10.1006/jmbi.1998.1721
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^1H [bond] ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125:10420–10428. doi: 10.1021/ja030153x
- Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1:749–754. doi: 10.1038/nprot.2006.101
- Tugarinov V, Kay LE (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem* 6:1567–1577. doi: 10.1002/cbic.200500110
- Uemura S, Iizuka R, Ueno T, et al (2008) Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic Acids Res* 36:e70. doi: 10.1093/nar/gkn338
- van den Berg B, Ellis RJ, Dobson CM (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J* 18:6927–6933. doi: 10.1093/emboj/18.24.6927
- Voss NR, Gerstein M, Steitz TA, Moore PB (2006) The geometry of the ribosomal polypeptide exit tunnel. *Journal of Molecular Biology* 360:893–906. doi: 10.1016/j.jmb.2006.05.023
- Vranken WF, Boucher W, Stevens TJ, et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. doi: 10.1002/prot.20449
- Wang Y, Sarkar M, Smith AE, et al (2012) Macromolecular crowding and protein stability. *J Am Chem Soc* 134:16614–16618. doi: 10.1021/ja305300m
- Waudby CA, Camilloni C, Fitzpatrick AWP, et al (2013a) In-Cell NMR Characterization of the Secondary Structure Populations of a Disordered Conformation of α -Synuclein within *E. coli* Cells. *PLoS ONE* 8:e72286. doi: 10.1371/journal.pone.0072286.s003
- Waudby CA, Christodoulou J (2012) An analysis of NMR sensitivity enhancements obtained using non-uniform weighted sampling, and the application to protein NMR. *J Magn Reson* 219:46–52. doi: 10.1016/j.jmr.2012.04.013
- Waudby CA, Launay H, Cabrita LD, Christodoulou J (2013b) Progress in Nuclear Magnetic Resonance Spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* 74:57–75. doi: 10.1016/j.pnmrs.2013.07.003
- Waudby CA, Mantle MD, Cabrita LD, et al (2012) Rapid Distinction of Intracellular and Extracellular Proteins Using NMR Diffusion Measurements. *J Am Chem Soc* 134:11312–11315. doi: 10.1021/ja304912c
- Wegrzyn RD, Deuerling E (2005) Molecular guardians for newborn proteins: ribosome-associated chaperones and their role in protein folding. *Cell Mol Life Sci* 62:2727–2738. doi: 10.1007/s00018-005-5292-z
- Wetlaufer DB (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA* 70:697–701.
- Wickramasinghe NP, Kotecha M, Samoson A, et al (2007) Sensitivity enhancement in (^{13}C) solid-state NMR of protein microcrystals by use of paramagnetic metal ions for optimizing (^1H) $T(1)$ relaxation. *J Magn Reson* 184:350–356. doi: 10.1016/j.jmr.2006.10.012
- Wilkins DK, Grimshaw SB, Receveur V, et al (1999) Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* 38:16424–16431. doi: 10.1021/bi991765q

- Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73:1–16. doi: 10.1016/j.pnmrs.2013.02.001
- Wishart DS, Sykes BD (1994) Chemical shifts as a tool for structure determination. *Meth Enzymol* 239:363–392.
- Wittekind M, Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *Journal of Magnetic Resonance*
- Woolhead CA, McCormick PJ, Johnson AE (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* 116:725–736.
- Wöhnert J, Franz KJ, Nitz M, et al (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. *J Am Chem Soc* 125:13338–13339. doi: 10.1021/ja036022d
- Wright CF, Teichmann SA, Clarke J, Dobson CM (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* 438:878–881. doi: 10.1038/nature04195
- Wu D, Chen A, Johnson CS (1995) An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *Journal of Magnetic Resonance* 115:260–264.
- Xu G, Ye Y, Liu X, et al (2014) Strategies for Protein NMR in *Escherichia coli*. *Biochemistry* 53:1971–1981. doi: 10.1021/bi500079u
- Zhang G, Hubalewska M, Ignatova Z (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology* 16:274–280. doi: 10.1038/nsmb.1554
- Zhang Z, Smith PES, Frydman L (2014) Reducing acquisition times in multidimensional NMR with a time-optimized Fourier encoding algorithm. *Journal of Chemical Physics* 141:194201. doi: 10.1063/1.4901561
- Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *Journal of Molecular Biology* 222:599–620.
- Ziv G, Haran G, Thirumalai D (2005) Ribosome exit tunnel can entropically stabilize alpha-helices. *Proc Natl Acad Sci USA* 102:18956–18961. doi: 10.1073/pnas.0508234102