**This document is downloaded from DR-NTU (https://dr.ntu.edu.sg)
Nanyang Technological University, Singapore.**

NMR study of eukaryotic translation termination

Wong, Leo E

2013

Wong, L. E. (2013). NMR study of eukaryotic translation termination. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/54845

https://doi.org/10.32657/10356/54845

Downloaded on 20 Mar 2024 19:24:09 SGT

NMR STUDY OF EUKARYOTIC TRANSLATION TERMINATION

WONG LEO E

SCHOOL OF BIOLOGICAL SCIENCES

2013

NMR Study of Eukaryotic Translation Termination

WONG LEO E

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirements for the degree of
Doctor of Philosophy

2013

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Associate Professor Konstantin Pervushin for the opportunity to work on this Ph.D. thesis, his constant encouragement and support throughout my research, his patience, motivation, enthusiasm and immense knowledge. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I am grateful to the School of Biological Sciences, Nanyang Technological University, Singapore for providing a research scholarship without which this work would not have been possible. This research was also supported by Singapore Ministry of Education AcRF grant T207B3205. I would like to acknowledge Professor Vladislav Orekhov, Professor Lev Kisselev and Professor Ludmila Frolova for their collaborations on the projects.

It was my pleasure to be able to work with Dr. James Masse, Dr. Victor Jaravine, Dr. Li Yan and Ms. Shubhadra Pillay, whom I am indebted to for their contributions to different parts of the projects. I would also like to thank other colleagues for their helps in one way or another: Ms. Bai Yang, Mrs. Liu Rubing, Ms. Margaret Phillips, Ms. Rachel Lim Sing Mei, Ms. Zhao Jing, Mr. Wei Lei, Mr. Edward Tan, Mr. Bertrand Russell, Dr. Xiao Zhou, Dr. Alistair Irvine, Dr. David Libich, Dr. Ye Hong, Mrs. Chua Geok Lin, Mrs. Sebanti Gupta, Mr. Mukesh Mahajan, and Mr. Alok Patra. The critical comments and suggestions from the examiners are very much appreciated.

I dedicate this thesis to my father and mother whom I owe all that I have become today. I thank my sister and my relatives for their love and support.

List of Abbreviations

aa-tRNA	aminoacyl-tRNAs
CARA	Computer Aided Resonance Assignment
CSP	chemical shift perturbation
DC	decoding center
DQF-COSY	double-quantum filtered correlation spectroscopy
eIF	eukaryotic initiation factor
eRF1	eukaryotic class I release factor
eRF3	eukaryotic class II release factor
FID	free induction decay
GTPase	guanosine triphosphate hydrolase
H44	helix 44 of 16S or 18S rRNA
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-thiogalactopyranoside
MDD	multidimensional decomposition
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	messenger RNA
NOESY	nuclear Overhauser effect spectroscopy
pre-TC	pre-termination ribosomal complex
PTC	peptidyl transferase center
RNA	ribonucleic acid
RRF	ribosome recycling factor
RRL	rabbit reticulocyte lysate
S/N	signal-to-noise ratio
TA	targeted acquisition
rRNA	ribosomal RNA
tRNA	transfer RNA
TROSY	transverse relaxation optimized spectroscopy

Contents

Summary 4

Chapter 1: Introduction 5

1.1 The ribosome 5

1.2 The translation process 10

1.3 Translation termination – Class I release factors 13

1.4 Mechanism of termination by bacterial RF1 and RF2 14

1.5 Class II release factors 18

1.6 Interactions between eRF1 and eRF3 19

1.7 Models of stop codon recognition by eRF1 20

1.8 Medical implications of premature termination codon 24

1.9 Automatic resonance assignment of NMR spectra 25

1.10 Fast NMR data acquisition 28

1.11 Aims of the study 30

Chapter 2: Materials and Methods 33

2.1 Materials and reagents 33

2.2 Sample and dataset for demonstration of targeted acquisition 34

2.2.1 Apo-CcmE 34

2.2.2 MDD reconstruction of 3D spectra 34

2.2.3 Parameters for *AutoLink II* 35

2.3 Expression plasmids for different constructs of eRF1 36

2.4 Expression and purification of protein samples 37

2.5 NMR spectroscopy 37

2.6 Structure calculations 38

2.7 NMR structural characterization of RNA 39

2.8 *In vitro* RF activity assay of chimeric eRF1s 40

Chapter 3: Automatic Assignment of Protein Backbone Resonances from Nonlinear Sampled NMR Data 41

3.1 Abstract 41

3.2 *Psyte* and *AutoLink II* programs 41

3.3 Targeted time domain data acquisition and MDD reconstruction 43

3.4 Automatic determination of spin systems by *Psyte* 46

3.5 Advancement of *Psyte* over threshold filtering 48

3.6 Consistency of residue specific assignment by *AutoLink II* 50

3.7 Robustness of *AutoLink II* in residue specific assignment.................................. 52

3.8 Completeness of residue specific assignment for targeted acquisition 55

3.9 Discussion.. 58

Chapter 4: Selectivity of Stop Codon Recognition is Modulated by Multiple Conformations of GTS Loop in eRF1 62

4.1 Abstract 62

4.2 The solution structures of wt N-domain and Q^{122}FM(Y)F^{126} 62

4.3 The GTS loop adopts distinct conformations in wt N-domain and Q^{122}FM(Y)F^{126} 65

4.4 Multiple conformations of the GTS loop and its implication............................. 69

4.5 Pico- to nanosecond dynamics of wt N-domain and Q^{122}FM(Y)F^{126} 71

4.6 C127 mutants of eRF1 exhibiting omni-, bi-, and unipotent specificity 72

4.7 Lack of interaction between N-domain and pentaribonucleotides containing stop codon 74

4.8 Discussion.. 75

Chapter 5: Interactions between N-domain and Mimics of the Decoding Region of Ribosomal Helix 44.. 80

5.1 Abstract 80

5.2 Design of 15-mer RNA mimicking the decoding region of H44 80

5.3 NMR structural characterization of the 15-mer RNAs....................................... 82

5.4 Paromomycin binds to the internal loop 85

5.5 Helix α1 of N-domain interacts with the 15-mer RNAs 86

5.6 Role of the internal loop in its interaction with N-domain................................ 89

5.7 Interactions between N-domain and ribosome... 91

5.8 Interactions between N-domain and C-domain of eRF1 94

5.9 Binding of 15-mer RNA displaces C-domain from the NC-complex................... 97

5.10 Discussion.. 99

5.10.1 Interactions between N-domain and mimics of the decoding region of H4499

5.10.2 A model of N-domain bound to the pre-termination complex 101

Chapter 6: Domain Interactions of NM-domain 105

6.1 Abstract 105

6.2 Absence of Inter-domain Contact in NM-domain.. 105

6.3 Precipitation of NM-domain by 15-mer-UAA 109

6.4 Discussion.. 109

Chapter 7: Conclusions 112
References 115
Appendix A: *Psyte*'s Algorithm 127
Appendix B: HSQC Spectra of wt N-domain and $Q^{122}FM(Y)F^{126}$ 131

Summary

Translation termination is critical. Many diseases are caused by nonsense mutation. Hence, understanding termination could pave the way to strategic interference of the process with medical benefit. Unlike in bacteria, eukaryotic translation termination is more complex. Exact mechanism of stop codon recognition by class I release factor eRF1 and the cooperative role of class II release factor eRF3 remain obscure. By solving the solution structures of both wild-type N-domain of human eRF1 exhibiting omnipotent specificity and its mutant with UGA-unipotency, we found the conserved GTS loop adopting alternate conformations. We propose that structural variability in the GTS loop may underline the switching between omnipotency and unipotency of eRF1. In addition, we showed the specific binding of a 15-mer RNA oligonucleotide mimicking the decoding region of 18S rRNA helix 44 to helix α1 of N-domain, on the interface that is shielded partially by C-domain in full-length eRF1. The 15-mer RNA displaces C-domain from the non-covalent NC-complex, suggesting an imperative domain rearrangement in eRF1 during which N-domain accommodates itself into ribosomal A site. On another hand, we also demonstrated the feasibility of targeted acquisition of NMR data that is dynamically controlled by the completeness of automatic backbone resonances assignment, in an effort to accelerate structural study of biomolecules by NMR spectroscopy.

Chapter 1: Introduction

1.1 The ribosome

Ribosome is the sophisticated cellular machinery that drives the translation process – a pivot in the central dogma of molecular biology. It is astonishing how ribosome continuously receives different protein factors and amino acid carriers to orchestrate faithful decoding of a messenger ribonucleic acid (mRNA) into its polypeptide chain. Ribosome is ancient and ubiquitous in living cells. It was first described by George Emil Palade under the electron microscope as "… component of the ground substance of the cytoplasm which is particulate in nature and small in size." (Palade, 1955) Despite being small for an organelle, its sheer size and complexity was enough to deter any scientist back in the 1970's from attempting to look into its macromolecular structure, except perhaps Ada Yonath (Pennisi, 1999).

The prokaryotic ribosomes (70S, in the Svedberg unit), in both Archaea and Eubacteria, consist of a small (30S) and a large (50S) subunit. The small subunit is composed of a 16S ribosomal ribonucleic acid (rRNA) and 21 proteins; the large subunit is composed of a 5S and a 23S rRNA, as well as 31 proteins. The slightly larger eukaryotic ribosome (80S) consists of a 40S and a 60S subunit, respectively. Its small subunit has a 18S rRNA and 33 proteins; while its large subunit is composed of three rRNAs, *i.e.* 5S, 5.8S and 28S, as well as about 46 proteins. In general, ribosomes have three pockets to accommodate different transfer RNAs (tRNAs), *i.e.* A site for aminoacyl-tRNAs (aa-tRNAs), P site for peptidyl-tRNA, and E site for exiting deacylated tRNA, as well as two tunnels for the traversing mRNA and the nascent polypeptide chain, respectively (Figure 1.1). Apart from the overall structural framework, there are three major sites on the rRNAs that play the catalytic “ribozyme” roles: (i) peptidyl transferase center (PTC) that catalyzes both peptide bond formation and peptidyl-tRNA hydrolysis, (ii) decoding center (DC) that monitors cognate codon-anticodon base pairing at A site, and (iii) GTPase-activating center (GAC) (specifically the sarcin-ricin loop (SRL)) that triggers guanosine triphosphate (GTP) hydrolysis in many translational GTPases. SRL resides on the 23S or 28S rRNA near the entrance to A site.

Although archaeal ribosomes have a size and components similar to those of their bacterial counterparts, most of their ribosomal proteins and translation factors have primary sequences closer to the eukaryotes (Lecompte et al., 2002). Furthermore, hybrid ribosomes of archaeon *Sulfolobus solfataricus'* and the yeast *Saccharomyces cerevisiae*'s large and small subunits were fully functional in protein synthesis *in vitro*, in contrast to the combination of *Escherichia coli*'s and the yeast's ribosomal subunits (Altamura et al., 1986). The mammalian mitochondrial ribosome has a molecular weight similar to the bacteria’s, but interestingly, it forms a 55S particle instead (consisting of a 28S and a 39S subunit, respectively) and contains half as much rRNA and nearly twice as much protein compared to the bacteria’s (O'Brien, 2002).

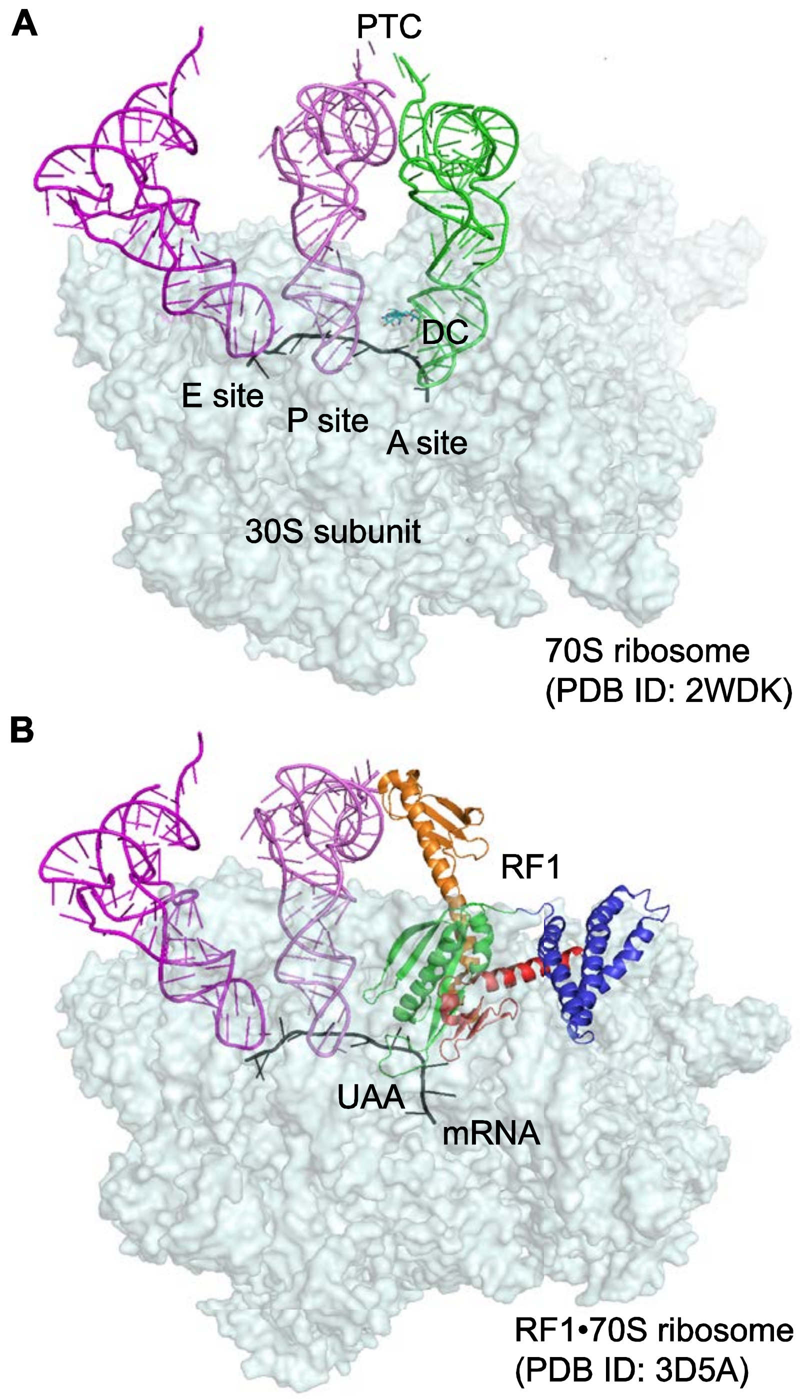

A
PTC
DC
E site
P site
A site
30S subunit
70S ribosome
(PDB ID: 2WDK)
B
RF1
UAA
mRNA
RF1•70S ribosome
(PDB ID: 3D5A)

Figure 1.1 Crystal structures of 70S ribosome. (A) The 30S subunit together with mRNA and tRNAs were shown. Paromomycin was bound to DC (cyan) of the 30S subunit. (B) The 30S subunit together with mRNA, tRNAs, and release factor RF1 bound to its cognate stop codon were shown.

Solving the atomic structure of ribosomal complex together with the template mRNA, peptidyl-tRNA, aa-tRNA, and the relevant translation factors arrested at different stages of translation would lead us directly to understanding how the massive protein synthesis machinery works. Furthermore, the high-resolution structures of ribosomal complexes with a variety of antibiotics also explain their respective inhibitory mechanisms. Successive publications of the high-resolution crystal structures of prokaryotic 30S and 50S subunits in 2000 marked a monumental milestone in biology (Ban et al., 2000, Schluenzen et al., 2000, Wimberly et al., 2000). A year later, Harry Noller's group improved on their previous effort and successfully tackled the whole 70S ribosome containing both mRNA and tRNAs at 5.5 Å (Yusupov et al., 2001). Those breakthroughs laid the critical foundation for subsequent structural works on 70S ribosome at higher resolution (Schuwirth et al., 2005, Selmer et al., 2006), ribosomal subunits containing different antibiotics (reviewed in (Hermann, 2005)), and functional complex of 70S ribosome bound with different translation factors, studied by both X-ray crystallography and cryo-electron microscopy (cryo-EM) (reviewed in (Schmeing and Ramakrishnan, 2009)). This series of remarkable success in the field of ribosome research culminates in the latest reports of high-resolution crystal structures of eukaryotic ribosome (Ben-Shem et al., 2011) and its individual subunits (Rabl et al., 2011, Klinge et al., 2011). Detailed analysis comparing the structures of both 70S and 80S ribosomes as well as their implicated significance has been reviewed by (Klinge et al., 2012) and (Melnikov et al., 2012). A list of representative crystal structures of ribosome and ribosomal complexes during the past twelve years are summarized in Table 1.1.

PDB ID	**Description**	**References**
1FKA	30S subunit, *Thermus thermophilus*, 3.3 Å	(Schluenzen et al., 2000)
1J5E	30S subunit, *Thermus thermophilus*, 3.0 Å	(Wimberly et al., 2000)
1FFK	50S subunit, *Haloarcula marismortui*, 2.4 Å	(Ban et al., 2000)
1GIX, 1GIY	70S ribosome, tRNAs, mRNA, *Thermus thermophilus*, 5.5 Å	(Yusupov et al., 2001)
2AVY 2AW4	70S ribosome, tRNAs, mRNA, *Escherichia coli*, 3.5 Å	(Schuwirth et al., 2005)
2J00 2J01	70S ribosome, tRNAs, mRNA, paromomycin, *Thermus thermophilus*, 2.8 Å	(Selmer et al., 2006)
2WRN 2WRO	70S ribosome, tRNAs, mRNA, aa-tRNA:EF-Tu, paromomycin, *Thermus thermophilus*, 3.6 Å	(Schmeing et al., 2009)
2WRI 2WRJ	70S ribosome, tRNAs, mRNA, EF-G, fusidic acid, *Thermus thermophilus,* 3.6 Å	(Gao et al., 2009)
4GD1 4GD2	70S ribosome, classical or hybrid P/E state of tRNA binding, mRNA, *Escherichia coli,* 3.2 Å	(Dunkle et al., 2011)
3UZK 3UZL	70S ribosome, cognate or near-cognate A site-tRNA, mRNA, paromomycin, *Thermus thermophilus*, 3.3 Å	(Demeshkina et al., 2012)

3D5A 3D5B	70S ribosome, tRNAs, mRNA, RF1:UAA, *Thermus thermophilus*, 3.2 Å	(Laurberg et al., 2008)
3F1E 3F1F	70S ribosome, tRNAs, mRNA, RF2:UAA, *Thermus thermophilus*, 3.0 Å	(Korostelev et al., 2008)
2WH1 2WH2	70S ribosome, tRNAs, mRNA, RF2:UGA, *Thermus thermophilus*, 3.5 Å	(Weixlbaumer et al., 2008)
3MR8 3MS1	70S ribosome, tRNAs, mRNA, RF1:UAG, *Thermus thermophilus*, 3.6 Å	(Korostelev et al., 2010)
2XZM	40S subunit, eIF1, *Tetrahymena thermophila*, 3.9 Å	(Rabl et al., 2011)
4A18 4A1E	60S subunit, eIF6, cycloheximide (a eukaryotic-specific inhibitor), *Tetrahymena thermophila*, 3.5 Å	(Klinge et al., 2011)
3U5B- 3U5E	80S apo-ribosome, *Saccharomyces cerevisiae*, 3.0 Å	(Ben-Shem et al., 2011)

Table 1.1 Representative crystal structures of ribosome and ribosomal complexes. These high resolution 3D structures have contributed significantly to our understanding of the different stages of the translation process in ribosome at the atomic level.

1.2 The translation process

The universal translation process consists of four major stages: initiation, elongation, termination and recycling (Schmeing and Ramakrishnan, 2009, Rodnina and Wintermeyer, 2009, Melnikov et al., 2012) (a movie of the process can be viewed at http://www.mrc-lmb.cam.ac.uk/ribo/homepage/movies/translation_bacterial.mov). Besides the coding mRNA and a vast pool of specifically charged tRNAs, there is a myriad of protein factors involved in coordinating the complicated process of protein translation with utmost accuracy. During the initiation stage, a number of initiation factors (e)IFs work together to recruit mRNA and the small subunit, to position fMet-tRNA^{fMet} (in bacteria) or Met-tRNA^{Met} (in eukaryotes) to the start codon at P site, and to assemble the initiation complex with the large ribosomal subunit. Nonetheless, the whole initiation stage is much more intricate and differences of the process at the molecular level between eukaryotes and prokaryotes has been reviewed (Sonenberg and Hinnebusch, 2009).

The elongation stage comprises repeating cycles of a sequence of sub-steps, resulting in the addition of a single amino acid to the peptidyl-tRNA after each cycle. First, a GTP-bound elongation factor EF-Tu (in bacteria) or eEF1A (in eukaryotes) delivers aa-tRNAs to the ribosomal A site. Then, selection of cognate aa-tRNA (so-called decoding step) occurs twice at DC, separated in-between by hydrolysis of GTP in the aa-tRNA-bound elongation factor. Following tRNA selection, the cognate aa-tRNA accommodates itself into the A site pocket to facilitate peptidyl transfer. The GTP hydrolysis between the tRNA selection acts as a kinetic proofreading mechanism to enhance decoding fidelity. In bacteria, A1492, A1493, and G530 of 16S rRNA at DC monitors the geometry of cognate codon-anticodon base pairing, which induces a ‘domain closure’ of 30S subunit (Ogle et al., 2001, Ogle et al., 2002, Ogle and Ramakrishnan, 2005). The ‘domain closure’ is also induced by near-cognate aa-tRNA (Demeshkina et al., 2012). Furthermore, cognate codon-anticodon base pairing also correlates with higher forward rates of GTPase activation and accommodation, thereby further enhancing tRNA selection (reviewed in (Rodnina and Wintermeyer, 2001, Zaher and Green, 2009)). During the decoding process, aa-tRNA undergoes structural rearrangement in relations to the ribosome, observable as distinct Förster resonance energy transfer (FRET) states representing the codon-anticodon base pairing, GTPase-activated, and fully accommodated states (Blanchard et al., 2004). After the accommodation of cognate aa-tRNA, the peptidyl-transferase reaction occurs at PTC, in which the α-amino group of the aa-tRNA nucleophilically attacks the ester carbon of the peptidyl-tRNA to form a new peptide bond. Ribosome increases the reaction rate by at least 10^5-fold (Sievers et al., 2004), while the underlying reaction mechanism is still under dispute. After peptidyl transfer, a "ratcheting" movement occurs between the subunits, followed by forward translocation of tRNAs and mRNA catalyzed by EF-G (in bacteria) or eEF2 (in eukaryotes).

Translation terminates when any of the nonsense or stop codons, *i.e.* UAA, UAG and UGA, reaches A site. This step is catalyzed by release factors (refer to Section 1.3). Lastly, dissociation of the large subunit in bacteria requires ribosome recycling factor (RRF) and EF-G; while in eukaryotes, it is kinetically coupled to translation termination and involves ABC-type ATPase ABCE1 (Shoemaker and Green, 2011, Becker et al., 2012). Beyond the described clockwork process, all living cells deploy a rescue mechanism when the translating mRNA is broken in one way or another, whereby translation will be resumed and the aberrant mRNA and polypeptide will be disposed of accordingly. In bacteria, trans-translation involves a protein-RNA complex called tmRNA-SmpB (reviewed in (Felden and Gillet, 2011)); while a bunch of sophisticated pathways constitute the whole mRNA surveillance system in eukaryotes (reviewed in (van Hoof and Wagner, 2011)).

1.3 Translation termination – Class I release factors

Class I release factors (RFs), *i.e.* RF1 and RF2 in bacteria and eRF1 in eukaryotes, mediate translation termination via recognition of the stop codon at DC and catalysis of the peptidyl-tRNA ester bond hydrolysis at PTC (Figure 1.1, B) (reviewed in (Kisselev et al., 2003)). These protein factors were first successfully isolated from *E. coli* S30 extract (Capecchi, 1967) and rabbit reticulocyte lysate (Goldstein et al., 1970), respectively, showing that a protein instead of tRNA was responsible to decode the stop codons. A termination assay based on the measurement of polypeptide release efficiency from reconstituted ribosomal intermediate (Caskey et al., 1968) was instrumental in delineating the functions of RFs. Using the *in vitro* RF assay, "bipotent" RF1 and RF2 were shown to recognize different groups of stop codons, namely UAA/UAG and UAA/UGA, respectively (Scolnick et al., 1968). Conversely, the "omnipotent" eRF1 is able to decode all three stop codons (Beaudet and Caskey, 1971, Frolova et al., 1994), with the exceptions of a group of variant-code organisms (refer to Section 1.7). RF1 and RF2 do not share any sequence similarity with eRF1 (Frolova et al., 1994), except for the GGQ motif that is critical for both to cleave the ester bond (Frolova et al., 1999). Furthermore, the crystal structures of eRF1 (Song et al., 2000), RF2 (Vestergaard et al., 2001) and RF1 (Shin et al., 2004) also show that RF1/RF2 are structurally distinct from eRF1, except for the long helix leading to the GGQ loop in eRF1's M-domain and RF1/RF2's domain 3 (Figure 1.2). On the other hand, archaeal class I RF, aRF1, is omnipotent and share sequence homology with eRF1 (Kisselev et al., 2003).

1.4 Mechanism of termination by bacterial RF1 and RF2

Naturally, the first question to ask about bacterial class I RFs is how the protein recognizes stop codons at the decoding center. When the tripeptide motifs, *i.e.* Pro-Ala/Val-Thr (PxT) in RF1 and Ser-Pro-Phe (SPF) in RF2, were discovered as the determinants of their respective stop codon specificity, a "tripeptide anticodon" hypothesis was proposed (Ito et al., 2000). It was suggested that the first and third amino acids of the tripeptides independently discriminate the second and third purine bases, respectively (Ito et al., 2000, Nakamura et al., 2000). The role of the tripeptides in stop codon recognition is supported by the high-resolution crystal structures of RF1-bound (Laurberg et al., 2008, Korostelev et al., 2010) and RF2-bound (Weixlbaumer et al., 2008, Korostelev et al., 2008) ribosomal complexes that include their respective cognate stop codons. However, the mRNA-protein interactions appear to be much more extensive (Figure 1.2, A).

From the molecular dynamics free-energy calculations based on those structures, the mechanisms for RF1 and RF2 to discriminate between different stop codons and against sense codons were proposed (Sund et al., 2010). The study had shown that the interaction energetics between RF1/RF2 and their cognate stop codons are large enough to account for the stop codon decoding accuracy. The first uridine U1 is stabilized by hydrogen bonding to the backbone carbonyl and amide group of the glycine and glutamate residues at the tip of helix α5 and the side-chain of T186 of the tripeptide loop. For the second stop codon position, the glutamate (E128) side-chain of RF2 together with D131, R191 and S195 of its tripeptide loop form a fine-tuned recognition switch that allows distinct interactions with both A2 and G2, while in RF1 the glutamate (E119) forms a more stable ion pair with R182 and hence only A2 is stabilized by hydrogen bonding to P184 and T186 of its tripeptide loop. The backbone carbonyl of T186 in RF1 also contributes to discrimination against G2 as its interaction becomes repulsive. For the third stop codon position, A3 is stabilized by bifurcated hydrogen bonding to the side-chain of T194/T203 of RF1/RF2, which is facing away from the tripeptide loop. In RF1, a key water molecule, which bridges the base of the third stop codon and the phosphate group of U531 of 16S rRNA, can potentially flip its orientation to interact with either A3 or G3. Conversely, in RF2, the space for the water molecule is blocked. Furthermore, Q181 of RF1, which is absent in RF2, selects for G3 by hydrogen bonding. The stacking of G530 of 16S rRNA against the base of the third codon position does not contribute to discriminate between different codons. Hence, the role played by G530, A1492, and A1493 in tRNA selection is irrelevant for stop codon recognition by RF1 and RF2. Due to the absence of sequence homology with RF1/RF2, stop codon recognition by eRF1 is expected to be different (refer to Section 1.7).

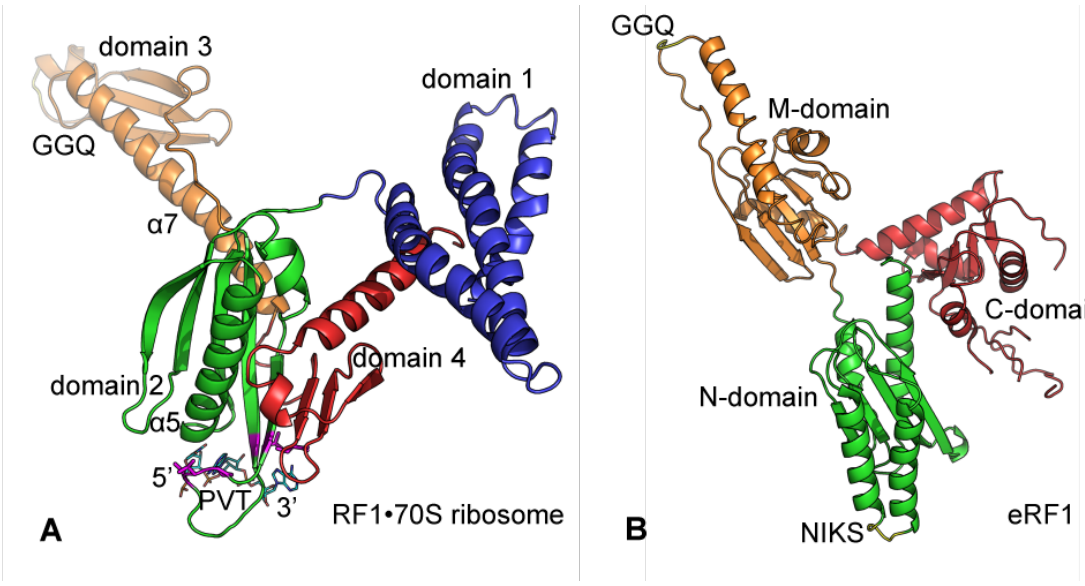


Figure 1.2 Domain architectures of RF1 bound to 70S ribosome and eRF1. (A) Ribosome-bound RF1 shows a tertiary structure that resembles tRNA. The interactions between RF1 and the stop codon can be seen from the complex structure (PDB ID: 3D5A). The amino acids that are in direct interactions with the stop codon are colored magenta. (B) The crystal structure of eRF1 shows three distinct domains arranged in a Y-shape (PDB ID: 1DT9).

The second important question would be how the GGQ loop catalyzes hydrolysis of peptidyl-tRNA. This might not have been answered exactly since the crystal structures only reveal the product state, in which the glutamine residue is positioned close to the ribose of A76 (the last nucleotide) of the deacylated P site tRNA. A model was proposed such that the glutamine is coordinating a water molecule to mount nucleophilic attack on the ester bond while A2451 of 23S rRNA at PTC is stabilizing 2'-OH of A76 (Weixlbaumer et al., 2008). The two glycine residues of the GGQ motif are critical to confer flexibility to the loop for adopting a drastic conformation. Correspondingly, movement of U2585 of 23S rRNA away from clashing with the GGQ loop is also important to expose the ester bond to nucleophilic attack. Later, by solving the crystal structure of similar ribosomal complex but with a aa-tRNA analog that resists hydrolysis, the role played by the glutamine residue in the catalytic reaction was redefined to be forming a catalytic pocket that accommodates a single water molecule (Jin et al., 2010). This is in agreement with the finding that the glutamine residue excludes nucleophiles larger than water from entering the active site (Shaw and Green, 2007). Since mutations of the conserved glutamine in GGQ have only a modest effect, the glutamine's side-chain is probably not involved directly in the catalysis. Conversely, its backbone amide group might form a hydrogen bond to the 3'-OH of A76 stabilizing the transition state of the product.

Translation termination is highly accurate: erroneous polypeptide release at sense codon occurs at an estimated frequency of 1 in 10^5 in *E. coli* (Jorgensen et al., 1993, Freistroffer et al., 2000). Although stop codon recognition by the RFs exhibits remarkable specificity, the catalytic rate of peptidyl-tRNA hydrolysis is also affected by the A site codon, indicating a coupling between the two reactions. When RF1 binds to the stop codon, DC undergoes structural changes that involve flipping out of A1493 (in h44 of 16S rRNA) and stacking of A1913 (in H69 of 23S rRNA) onto the vacated space in h44 (Laurberg et al., 2008). Concurrently, a conformational rearrangement of the "switch loop" connecting domains 3 and 4 of RF1 occurs to extend helix α7 by two helical turns as well as to provide it the flexibility for reorientation, while the "switch loop" is being stabilized by residues 1914-1915 of H69. The differential interactions of RFs and the ribosome between cognate and near-cognate termination complexes clearly indicates an underlying coupling mechanism from stop codon recognition to peptidyl-tRNA hydrolysis (He and Green, 2010). Therefore, the polypeptide release process in bacteria proceeds via allosteric regulations of class I RFs by both mRNA and rRNAs. Although the global architecture of eRF1 resembles that of ribosome-bound RF1/RF2 as well as a tRNA (Figure 1.2), the mechanism of translation termination in eukaryotes may not be analogous and thereby remains obscure.

1.5 Class II release factors

Class II RFs, *i.e.* RF3 in bacteria and eRF3 in eukaryotes, are translational GTPases. They do not recognize stop codons but are able to stimulate polypeptide release efficiency of class I RFs via GTP hydrolysis (Zhouravleva et al., 1995, Frolova et al., 1996, Freistroffer et al., 1997). Both RF3 and eRF3 have 3D structures similar to EF-Tu (Kong et al., 2004, Gao et al., 2007). However, their functional roles in translation termination are distinct (reviewed in (Kisselev et al., 2003)). eRF3 is encoded by essential genes, while RF3 is not. RF3 neither enhances the affinity of class I RFs to the ribosome nor improves the fidelity of stop codon decoding (Freistroffer et al., 2000). Instead, it binds to the post-termination ribosomal complex in GDP-bound form and exchanges GDP with GTP (Zavialov et al., 2001), and then actively removes class I RFs from the complex for recycling (Freistroffer et al., 1997 , Gao et al., 2007). On the other hand, eRF1•eRF3•GTP enters the pre-termination ribosomal complex (pre-TC) as a ternary complex (Kobayashi et al., 2004, Mitkevich et al., 2006), and subsequent GTP hydrolysis in eRF3 significantly enhances polypeptide release efficiency, in comparison to eRF1 alone (Alkalaeva et al., 2006). In addition, replacement of the GTP with non-hydrolyzable analog inhibits polypeptide release completely. Hence, the stimulatory and regulatory roles of the GTP hydrolysis show that cooperativity between eRF1 and eRF3 in translation termination is likely to be multi-step, and the detailed mechanism remains to be elucidated.

1.6 Interactions between eRF1 and eRF3

The complex structure of eRF1 and eRF3 may shed light on the mechanistic role of eRF3 in enhancing polypeptide release efficiency of eRF1. In the crystal structures of human and *Schizosaccharomyces pombe* eRF1 in complex with domains 2 and 3 of their respective eRF3, C-domain of eRF1 is interacting with domain 3 of eRF3 via mainly a hydrophobic interface, and the N- and M-domain of eRF1 in the human eRF1•eRF3-23 complex displays a bent conformation (Cheng et al., 2009). The small-angle X-ray scattering (SAXS) study on the full length eRF1•eRF3 complex showed that M-domain of eRF1 is in contact with domains 2/3 as well as the switch region of eRF3 GTPase domain, and that the arginine residue R189/R192 (*S. pombe*/human numbering) on M-domain may play a part in stimulating the GTPase activity of eRF3 (Cheng et al., 2009).

The archaeal class I release factor (aRF1) is highly homologous to eRF1, and is able to release polypeptide at all three stop codons in mammalian ribosome (Dontsova et al., 2000). However, archaea do not have a distinct protein that plays the role of eRF3, but instead their tRNA carrier GTPase protein aEF1α is able to bind to aRF1 in a GTP-dependent manner via interface similar to eRF1•eRF3 complex (Saito et al., 2010). Similarly, aEF1α binds to aPelota, a protein structurally homologous to aRF1/eRF1 that mediates the so-called no-go decay of mRNA surveillance. Recently, a crystal complex structure of aRF1 and GTP-bound aEF1α from *Aeropyrum pernix* has been solved, offering a high-resolution view on the interaction interfaces between aRF1 and aEF1α (Kobayashi et al., 2012). The domain arrangement of aRF1 in the complex is similar to free aRF1 (Saito et al., 2010), which is in the bent conformation. Domain 3 of aEF1α is interacting with C-domain of aRF1 in similar way to the human eRF1•eRF3-23 complex (Cheng et al., 2009). Domain 2 of aEF1α is in contact with the tip region of aRF1 M-domain. However, part of the loop including the GGQ motif was proteolytic degraded during crystal formation, indicating exposure of the GGQ loop to the solvent. Domain 1 of aEF1α (the GTPase domain) interacts with aRF1 M-domain via an extensive network of hydrogen bonding. Furthermore, residue H94, which is gated by V15 and I70 from the γ-phosphate of GTP, is coordinating together with R68 and D96 a water molecule proposed to be catalyzing GTP hydrolysis. The aRF1•aEF1α•GTP complex displays an overall tertiary structure resembling the aa-tRNA•EF-Tu•GTP complex, supporting the theory of mimicry between translation elongation and termination. In addition, a pair of electrostatic interaction between E134 (domain 1) and R413 (domain 3) of aEF1α is also found in EF-Tu in the aa-tRNA•EF-Tu•GDPNP complex.

1.7 Models of stop codon recognition by eRF1

On eukaryotic translation termination, the crystal structure of eRF1 contributes significantly to our understanding of its functions (Song et al., 2000). In terms of structure and function, eRF1 is composed of three distinct protein domains performing specific roles: N-domain recognizes the stop codon at DC, M-domain triggers hydrolysis of peptidyl-tRNA at PTC, and C-domain forms a complex interface with eRF3 (Figure 1.2, B) (Frolova et al., 2000, Ito et al., 1998, Eurwilaichitr et al., 1999). Based on the “tRNA mimicry” hypothesis, the tRNA-like domain architecture of eRF1 in the crystal structure suggests that the NIKS tetrapeptides at the tip of N-domain might serve as equivalence to the tRNA anticodon (Song et al., 2000). Furthermore, the role of the NIKS motif in stop codon recognition is supported by photoactivatable crosslinking and mutagenesis experiments (Chavatte et al., 2002, Frolova et al., 2002). However, a conclusion has yet to be drawn since several models derived from different computational and experimental approaches have been proposed over the years that are not entirely compatible to each other (Figure 1.3).

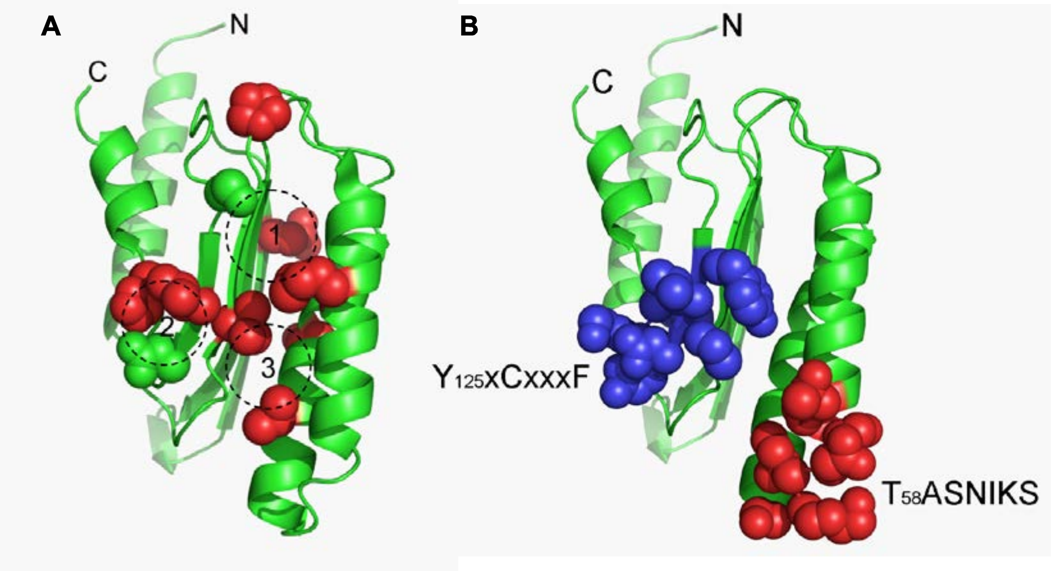


Figure 1.3 Models of stop codon recognition. (A) The cavity model states that the highlighted regions form three pockets to accommodate the nucleobases of stop codon. (B) The linear peptide anticodon model states that the TASNIKS motif is responsible to decode the stop codon via direct interactions, while the nonlinear model states that in addition to the TASNIKS motif, the YxCxxxF motif is also required for stop codon recognition.

Variant-code organisms are eukaryotes that have one or two of the universal stop codons being reassigned to a sense codon. Reasonably, eRF1s of variant-code organisms should have lost the ability to decode a specific subset of the stop codons, even in condition without competition from nonsense-suppressor tRNAs. Indeed, chimeric eRF1s composed of N-domain from some of the variant-code organisms (*e.g.* ciliates) and the MC-domain of eRF1 from universal-code organisms (*e.g.* human or yeast) retain stop codon specificity typical for ciliate eRF1s, demonstrating unequivocally that N-domain of eRF1 directly decodes the stop codon (Ito et al., 2002, Salas-Marco et al., 2006, Lekomtsev et al., 2007, Eliseev et al., 2010). Therefore, phylogenetic analysis on the diverse eRF1 amino acid sequences from both universal- and variant-code organisms would lead us to the answer regarding which part of N-domain actually confers the specificity of stop codon recognition.

From the phylogenetic studies, the NIKS motif is found to be highly conserved among universal-code organisms but frequently diverges in variant-code organisms (Knight and Landweber, 2000, Lozupone et al., 2001, Kim et al., 2005). It was later proposed that the TASNIKS motif forms a flexible element that can assume a tight or relaxed conformation to interact directly with different stop codons (Muramatsu et al., 2001). However, besides the NIKS motif, various substitutions on the YxCxxxF motif were also found to affect stop codon recognition specificity, leading to the proposal of a nonlinear model in which stop codon recognition is modulated by positive and negative determinants (Frolova et al., 2002, Seit-Nebi et al., 2002, Kolosov et al., 2005). Subsequently in the photoactivatable crosslinking experiments both the TASNIKS and YxCxxxF motifs were found to be crosslinked to the stop codon

(Bulygin et al., 2010). On the other hand, a genetic screening from random mutations had uncovered several residues distributed across the core of N-domain that individually shifted the bias of stop codon recognition specificity (Bertram et al., 2000). Based on those results, a cavity model was proposed such that each of the nucleobases of the stop codon would fit onto the individual pockets defined (Figure 1.3, A). The cavity model was found to be in agreement with a study on the evolutionary rates of amino acid changes in N-domain (Inagaki et al., 2002).

From a broader perspective, the different models of stop codon recognition similarly involve many residues covering an extensive surface area on one side of N-domain. The experimental approaches from which the models were derived are deemed to have different limitations. eRF1 is an essential gene, hence *in vivo* genetic screening would not be able to retrieve the mutations that would have abolished stop codon recognition completely. Furthermore, from the *in vitro* RF assay, the effect on stop codon specificity by many of the point mutations varied depending on the substituted amino acids, rendering deduction impossible. Lastly, crosslinking potentially spans a finite distance in the vicinity of the photoactivatable analog, and the downstream step of identifying the crosslinked peptide product is challenging. Therefore, the proposed models of stop codon recognition should be subjected to testing by other experimental approaches.

1.8 Medical implications of premature termination codon

Premature termination codon is a nonsense mutation that occurs within the open reading frame of a gene. It is found in about ten percent of inherited diseases, causing the degradation of the mRNA template and the production of a truncated polypeptide that has either reduced or null functionality. The classic examples are Duchenne muscular dystrophy, cystic fibrosis, and hemophilia. Many cancers are also related to the acquired premature termination codon mutation on tumor suppressor genes. For instance, 24% of all known mutations in *APC* gene are premature termination codon mutation (Laurent-Puig et al., 1998).

Nonsense-mediated decay (NMD) is the cellular mRNA surveillance pathway that specifically targets mRNA template harboring premature termination codon for degradation. The UPF proteins, *i.e.* encoded by *UPF1* (*smg-2* in *Caenorhabditis elegans*), *UPF2* (*smg-3*), and *UPF3* (*smg*-4) are major factors mediating NMD in both exon junction complex (EJC)-dependent and –independent pathways (Kervestin and Jacobson, 2012). During pre-mRNA splicing, various factors including Upf2 and Upf3 are recruited to form an EJC at 20-24 nucleotides upstream from the exon-exon junction. On the other hand, Upf1 is associated with eRF1•eRF3 complex, and form the so-called SURF complex together with Smg1, Smg8, and Smg9. When premature termination codon is localized at 50-55 nucleotides upstream of the exon-exon junction, a stalled ribosome allows the interactions between Upf1 and Upf2, leading to the activation of Upf1. Subsequently, Upf1 recruits decay enzymes to promote endonucleolytic cleavage, decapping followed by 5’ to 3’ decay or deadenylation followed by 3’ to 5’ decay of the aberrant mRNA. However, if readthrough of premature termination codon occurs, the EJC will be dissociated after the first round of translation, and the aberrant mRNA is stabilized.

As an attempt to elevate the survival rate of patients with genetic disorders caused by premature termination codon, researchers have tried with success of using aminoglycosides that target the ribosome to induce readthrough of premature termination codon in mouse models of the genetic diseases (Bidou et al., 2012). However, due to the nephrotoxicity and ototoxicity of certain aminoglycosides like gentamicin, they are not considered as suitable treatment. Likewise, non-aminoglycosides compounds like negamycin and tylosin have been shown to induce readthrough in mouse models (Arakawa et al., 2003, Zilberberg et al., 2010). Most interestingly, a new drug ataluren (PTC124) was identified via a drug screening to selectively induce a higher readthrough for UGA in the middle of the gene but not at the correct termination position, in both human muscle cells and mouse models (Welch et al., 2007). PTC124 has no antibacterial function, appears to be safe, and can be delivered orally. However, disappointing results from recent trials for Duchenne muscular dystrophy has casted doubts on the potential of the drug.

1.9 Automatic resonance assignment of NMR spectra

Resonance assignment is typically recognized as an essential intermediate benchmark in NMR analysis of biomolecules with many downstream steps leading to 3D structure reconstructions being automated (Nilges et al., 1997, Herrmann et al., 2002a, Herrmann et al., 2002b). The use of this benchmark as a target for NMR data acquisition (Jaravine and Orekhov, 2006) would result in a streamlined process greatly contributing to the efficiency of NMR structure determination.

In the last few years, several attempts at automating resonance assignment have demonstrated considerable progress (Atreya et al., 2000, Tian et al., 2001, Pristovsek et al., 2002, Coggins and Zhou, 2003, Hitchens et al., 2003, Jung and Zweckstetter, 2004, Langmead and Donald, 2004, Eghbalnia et al., 2005a, Eghbalnia et al., 2005b, Lin et al., 2005, Masse and Keller, 2005, Masse et al., 2006, Wu et al., 2006). In the most favorable cases, sequence specific resonance assignment and structural NOE assignment can be done in parallel with structure calculations (Grishaev and Llinas, 2004, Grishaev et al., 2005, Takeda et al., 2007). Based on the algorithm of relative hypothesis prioritization (RHP) that is aimed to mimic reasoning by a spectroscopist, *AutoLink* has demonstrated enough robustness and efficiency to tackle the NP-hard type of problem in assigning backbone resonances (Figure 1.4) (Masse and Keller, 2005).

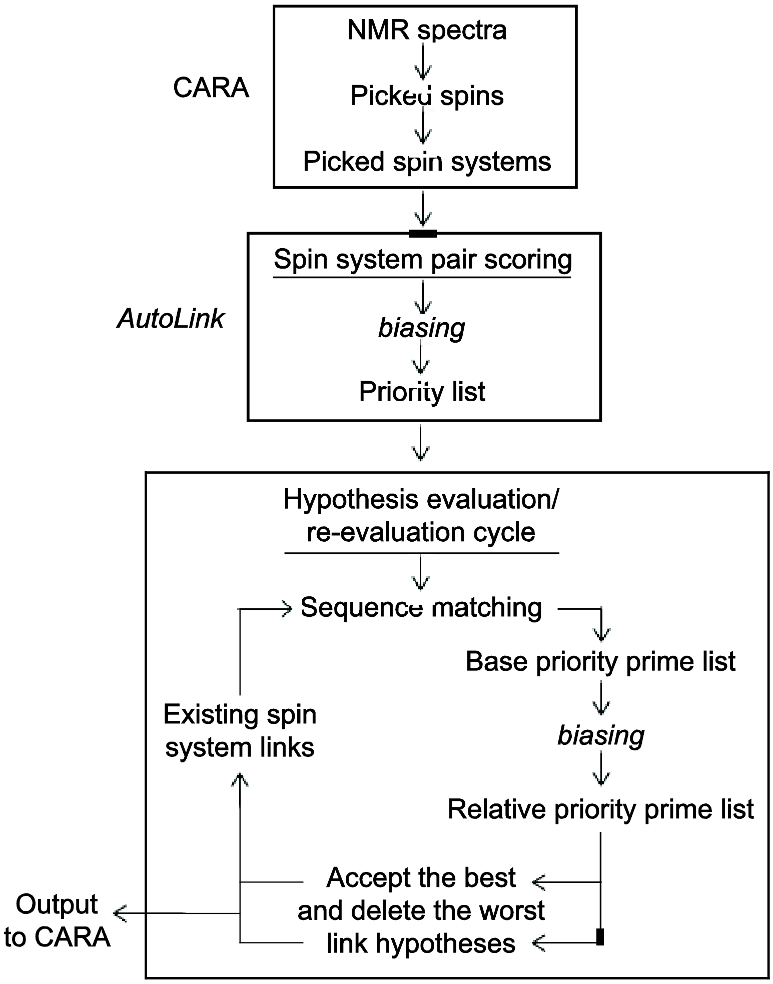

CARA
NMR spectra
Picked spins
Picked spin systems
AutoLink
Spin system pair scoring
biasing
Priority list
Hypothesis evaluation/
re-evaluation cycle
Sequence matching
Base priority prime list
biasing
Relative priority prime list
Existing spin system links
Accept the best and delete the worst link hypotheses
Output to CARA

Figure 1.4 Schematic diagram of the algorithm implemented in *AutoLink*. The input to *AutoLink* is a list of spin systems – a collection of spins/chemical shifts belonging to individual amino acid. Based on the matching of chemical shifts, as many as possible pairs of spin systems are generated. Each pair of spin systems is scored by a user-defined function and different *biasings* – score penalty, forming a list of "link hypothesis" that is ordered by their scores. The priority list enters the cycle of hypothesis evaluation and re-evaluation, which serves to modify the link hypothesis scores based on the acceptance or rejection of other link hypotheses after each cycle. In the first step, a base priority prime list is formed by combining the link hypotheses (and their scores) into different fragments, and scored by matching to the amino acid sequence. The scores are further modified by different *biasings*, which are described in (Masse and Keller, 2005). From the relative priority prime list, a certain numbers, as chosen by the user, of the top link hypotheses are accepted as linked spin systems, while some at the bottom of the list are deleted. The cycle terminates either at

the chosen maximum number of cycles or when no more link hypotheses are determinable.

Despite the impressive efforts directed to this problem, there is often still enough complexity left to require at least some human assistance. This situation is further aggravated when resonance assignment is set as a target for NMR spectral acquisition with less stable proteins that are prone to degradation during measurement. In this case, not only should the spectrum analysis be well advanced to deal with a host of problems such as missing and spurious cross-peaks, wide amplitude variation of the detectable resonances (*e.g.* due to dynamic line broadening), raise of signals from degradation products, but fast data acquisition methods might become essential.

1.10 Fast NMR data acquisition

Recently several accelerated acquisition schemes of multidimensional NMR spectra were developed, which can be used to control real-time data acquisition targeted to as complete as possible assignment of NMR resonances. For concentrated solutions of small- and medium-sized proteins (*i.e.* in data sampling limited cases), “projection reconstruction” was introduced (Kupce and Freeman, 2004). Adaptive selection of the tilt-angles was proposed helping to optimize the time usage of spectrometer (Kupce and Freeman, 2004, Eghbalnia et al., 2005a, Eghbalnia et al., 2005b, Hiller et al., 2005, Hiller et al., 2007). A robust scheme can be created by combining the basic features of projection reconstruction and “fast” data acquisition approaches, *e.g.* a spatially encoded (Frydman et al., 2003) and relaxation-optimized approach (Pervushin et al., 2002), fast pulsing techniques (Kupce and Freeman, 2007), combined with "ultrafast" spectroscopy (Schanda et al., 2005, Gal et al., 2007, Mishkovsky et al., 2007). The other reconstruction methods include the use of G-matrix Fourier transform-NMR (Kim and Szyperski, 2003) and *J* couplings networks (Atreya et al., 2007), covariance spectroscopy of higher dimensions (Zhang and Bruschweiler, 2004, Snyder et al., 2007a, Snyder et al., 2007b), fast Fourier transforms of non-equispaced data (Marion, 2005), 2D Fourier transformations of arbitrarily sampled NMR datasets (Kazimierczuk et al., 2006b, Kazimierczuk et al., 2006a), filter diagonalization methods (Mandelshtam et al., 1998, Mandelshtam, 2000, Armstrong et al., 2005), maximum entropy reconstructions (Rovnyak et al., 2004, Frueh et al., 2006), and multidimensional decomposition algorithm (MDD) (Orekhov et al., 2001, Luan et al., 2005).

The fundamental assumption of MDD is that an experimental *n*D NMR dataset can be optimally approximated by tensor product of 1D shapes along the *n* dimensions. The decomposition problem for a 3D dataset is to minimize the following function:

$$\min \sum_{ijk} \left[S_{ijk} - \sum_{m=1}^{M} a^m * F1_i^m * F2_j^m * F3_k^m \right]^2 + \lambda \sum_{m=1}^{M} (a^m)^2$$

, where *S* is the dataset, indices *i*, *j*, and *k* identify the grid points, *M* is the total number of components (which is equivalent to the total number of un-identical 1D shapes), a^m is the amplitude of a component, $F1_i^m$, $F2_j^m$ and $F3_k^m$ are numbers representing elements of the normalized shapes, and the second summation is for Tikhonov's regularization. For an input matrix *S* of size

$I*J*K$, the number of parameters for optimization is $M*(I + J + K - 2)$. Later, the MDD procedure was modified to analyze sparse datasets:

$$\min \sum_{ijk} G_{ijk} * \left[S_{ijk} - \sum_{m=1}^{M} a^m * F1_i^m * F2_j^m * F3_k^m \right]^2 + \lambda \sum_{m=1}^{M} (a^m)^2$$

, where $G_{ijk} = 1$ for a recorded and $G_{ijk} = 0$ for an omitted data point in the NMR experiment (Orekhov et al., 2003). This improved method is able to output shapes that contain all elements.

Relying on the MDD reconstruction from sparsely acquired NMR data, feedback looping of data acquisition and analysis was conceptualized by introducing a targeted acquisition (TA) scheme for real-time NMR spectroscopy (Jaravine and Orekhov, 2006). TA/MDD offers a possibility to combine adaptive, non-regular data sampling in *n*-dimensions with full or partial reconstruction of incompletely sampled 3D spectra or hyper-dimensional spectra (Jaravine et al., 2006, Jaravine et al., 2008), resulting in concurrent data accumulation, processing, and monitoring of spectral quality.

1.11 Aims of the study

Riding on the tremendous success of ribosome X-ray crystallography since the turn of the century, the mechanisms behind the translation process as well as the inhibitory mechanism of various ribosome-targeting antibiotics have been understood from the atomic structures of ribosomal complexes reconstituted in different ways. Of course, cryo-EM and pre-steady-state kinetics also contribute significantly to the progress of the field. In particular, the puzzle of translation termination in bacteria seems to have been completed by the high-resolution crystal structures of termination complex with either RF1 (Laurberg et al., 2008, Korostelev et al., 2010) or RF2 (Weixlbaumer et al., 2008, Korostelev et al., 2008) bound to their respective cognate stop codons, and the cryo-EM structure of the post-termination ribosome bound with RF3 in the GTP form (Gao et al., 2007). The only missing piece that remains to be confirmed is the exact catalytic mechanism of peptidyl-tRNA hydrolysis.

eRF1 had been identified and isolated 18 years ago (Frolova et al., 1994). Over the years, we have gained a significant amount of knowledge from the crystal structure of eRF1 (Song et al., 2000), *in vitro* reconstitution of termination complex, biochemical and genetic studies on a repertoire of eRF1 mutants, *etc.* about the functional motifs of eRF1 as well as the cooperativity between eRF1 and eRF3. Nevertheless, two key questions still beg for answers: (i) how does the omnipotent eRF1 recognize all three stop codons, and (ii) how does eRF3 enhance the polypeptide release efficiency of eRF1. In spite of the recent report of a cryo-EM structure of mammalian pre-termination complex bound with eRF1•eRF3, the rather low-resolution model could only provide limited insights (Taylor et al., 2012).

In collaboration with Professor Ludmila Frolova, who could provide support on the *in vitro* reconstitution experiments, we set out to answer the first question by a structural approach, *i.e.* solving the structures of wild type eRF1 N-domain and its mutants that exhibit different stop codon selectivity. Since solution NMR spectroscopy is a sensitive method to probe molecular interactions, we were also studying the potential interactions between N-domain and different structural RNAs. In-line with the structural study, we were also developing computer-assisted methodology in biomolecular NMR, hoping to speed up the process of structural elucidation.

Chapter 2: Materials and Methods

2.1 Materials and reagents

Isotope enriched 99% $^{15}N-NH_4Cl$ salts, 99% U-^{13}C-glucose, 99% deuterium oxide, and 2,2-dimethyl-2-silapentane-5-sulfonate sodium salts (DSS) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA USA) and ISOTEC of Sigma Aldrich (St. Louis, MO USA). Bacto^{TM} Tryptone and Bacto^{TM} Yeast Extract were purchased from BD (Franklin Lakes, NJ USA). Ampicillin sodium salt and Chloramphenicol were purchased from AppliChem GmbH (Darmstadt, Germany). Sodium azide, phenylmethanesulfonyl fluoride (PMSF) and 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) were purchased from Sigma Aldrich (St. Louis, MO USA). 5-ml HisTrap HP columns and 5-ml HiTrap Desalting columns were purchased from GE Healthcare Life Sciences (Piscataway, NJ USA). Centrifugal filter tubes were purchased from EMD Millipore (Billerica, MA USA) and Sartorius AG (Göttingen, Germany). Bacteriophage *Pf1* was purchased from Hyglos GmbH (Bernried, Germany). Paromomycin sulfate salt was purchased from Sigma Aldrich (St. Louis, MO USA). Rabbit reticulocyte lysate system (nuclease treated) was purchased from Promega (Fitchburg, WI USA). Other analytical grade reagents and solvents were purchased from Merck KGaA (Darmstadt, Germany), AppliChem GmbH (Darmstadt, Germany), AppliChem Inc. (St. Loius, MO USA), and Sigma Aldrich (St. Louis, MO USA).

2.2 Sample and dataset for demonstration of targeted acquisition

2.2.1 Apo-CcmE

Apo-CcmE-His6 (residues 30–159) was expressed and purified as described (Enggist et al., 2002). The sample was dialyzed against 300 mM NaCl, 50 mM sodium phosphate, pH 7.2, and was subsequently concentrated to 0.5 mM. After one-week incubation at 45 ºC, MALDI-TOF spectra showed essentially complete removal of the structurally flexible C-terminal fragment (residues 131–159) and presence of short peptides as degradation products manifested as additional small cross-peaks in the [^{1}H,^{15}N]-HSQC spectrum. Thus, in the time course typical for NMR measurements, the protein exhibited (i) small deviation of chemical shifts in apo-CcmE (L30–H130) lacking C-terminus in comparison to the full length construct, (ii) broadening beyond detection of a set of resonances located primarily in loop regions, and (iii) emergence of additional cross-peaks due to degradation. This sample was judged as a suitable model for automatic resonance assignment of a protein on the background of emerging artifacts, *e.g.* degradation products and missing cross-peaks, typically encountered in real protein samples.

2.2.2 MDD reconstruction of 3D spectra

Five triple-resonance spectra commonly used for protein backbone assignment were acquired on a Varian Inova 600 MHz spectrometer using the incremental non-uniform sampling scheme (INUS) (Jaravine et al., 2006). 25% (for HNCACB, HN(CA)CO, and ct-HNCA) or 11% (for HNCO and HN(CO)CA) of the regular indirect dimension grid points were sampled using the INUS schedule (Table 2.1). The time domain data was converted into the *nmrPipe* format (Delaglio et al., 1995) and the directly detected dimension was processed in the conventional way. In the indirect dimensions (t_1 and t_2), NMR data were processed using R-MDD (Jaravine et al., 2006) as is implemented in the *MDDnmr* program (Tugarinov et al., 2005). The reconstructed spectra were inspected using CARA (cara.nmr.ch). The program *MDDnmr* is available by a direct request to Prof. Vladislav Orekhov or Dr. Victor Jaravine.

Spectrum	Acquisition time, hours	Spectral width in ω_1 (^{13}C), Hz	Spectral width in ω_2 (^{15}N), Hz	Number of points in regular grid	
	Total: 40			t_1 (^{13}C)	t_2 (^{15}N)
HNCO	5 (11%)*	2100	2500	120	60
HN(CO)CA	5 (11%)	4527	2500	90	60
ct-HNCA	10 (25%)	4527	2500	120	60
HN(CA)CO	10 (25%)	2100	2500	120	60
HNCACB	10 (25%)	12071	2500	120	60

*In brackets is percent of acquisition time needed to sample the full regular grid.

Table 2.1 Details of the set of 3D spectra acquired with incremental non-uniform sampling scheme (INUS) for MDD reconstruction.

2.2.3 Parameters for *AutoLink II*

The automatic backbone resonances assignment program *AutoLink II*, which includes the *Psyte* module for automatic spin systems identification, can be downloaded from http://www.autolink.nmr-software.org/. Under the common workflow, *Psyte* output is fed to the residue specific assignment module of *AutoLink II*. Several control parameters were modified during the process, which include the score combination equation (“eq.”), the numbers of links and unlinks per round (“#link/r” and “#unlink/r”, respectively), and dynamic determinacy bias (“ddb”). Residue specific assignment was completed within ten to twenty minutes in the following execution order. Firstly, “eq.” = (CA*CB*CO)^0.33, “#link/r” = 1000, “#unlink/r” = 10, “ddb” = 0.5. After that, the unassigned spin systems were unlinked, and *AutoLink II* was executed once more with “eq.” = auto, “#link/r” = 3, “#unlink/r” = 2, “ddb” = 1.0. The numbers of unlinks per round in the first step was taken as 10% of the numbers of amino acids in the protein sequence. This execution scheme is aimed to assign spin systems with high confidence level and no ambiguity in the first step, followed by more stringent criterion in the second step. Such a scheme was able to reduce the amount of calculation time dramatically without compromising accuracy of the assignment result.

2.3 Expression plasmids for different constructs of eRF1

The DNA fragments encoding the wild-type and the $Q^{122}FM(Y)F^{126}$ mutant of N-domain (residues 1-142), the NM-domain (residues 1-275), and the C-domain (residues 275-437) of human eRF1 with a C-terminal hexahistidine tag were cloned into pET23(+) vector (Novagen) under the phage T7 RNA polymerase promoter (Kononenko et al., 2008) (courtesy of Prof. Lev Kisselev). The K18QK19Q mutant of N-domain was generated by site-directed mutagenesis using the KOD DNA polymerase and DpnI protocol with a pair of forward (*i.e.* 5’-GGAGATCTGGAAGATCCAGCAGCTCATTAAGAGC-3’) and reverse (*i.e.* 5’-GCCTCCAAGCTCTTAATGAGCTGCTGGATCTTCC-3’) primers.

2.4 Expression and purification of protein samples

Uniformly ^{13}C, ^{15}N-labeled wt N-domain and Q^{122}FM(Y)F^{126} were expressed in *E. coli* BL21 Rosetta(DE3) or CodonPlus (DE3)-RIL host cells cultured in minimal media (M9) utilizing $^{15}NH_4Cl$ (1.0 g/l) and $^{13}C_6$-glucose (2.0 g/l) as the sole nitrogen and carbon sources. Uniformly ^{2}H, ^{13}C, ^{15}N-labeled NM-domain was produced using ^{2}H2O (99% d-enrichment) as the solvent. All recombinant proteins were purified from cell lysates utilizing a 5-ml HisTrap HP column (GE Healthcare), and further purified using three 5-ml HiTrap Desalting columns (GE Healthcare) connected in series. NMR samples contained 0.1-1.0 mM protein in 20 mM MES, 100 mM KCl, 2 mM DTT at pH 6.0. For all experiments involving C-domain, 3 mM of β-mercaptoethanol was added to prevent oxidation of the cysteine thiol groups.

2.5 NMR spectroscopy

All NMR spectra were acquired using 600, 700, or 900 MHz Bruker Avance II spectrometers. Chemical shifts were referenced to DSS directly for ^{1}H and indirectly for ^{13}C and ^{15}N spins. The NMR data were processed using TopSpin 2.0 (www.bruker-biospin.com) and analyzed using CARA (cara.nmr.ch). ^{1}H, ^{15}N, and ^{13}C resonances of wt N-domain, Q^{122}FM(Y)F^{126}, and NM-domain were assigned using 3D TROSY-HNCA and TROSY-HNCACB. Side-chain ^{1}H and ^{13}C were assigned using iterative analysis of the 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HMQC spectra coupled with structure calculations. The assignment process was facilitated by comparison with chemical shifts deposited in the Biological Magnetic Resonance Data Bank (www.bmrb.wisc.edu) for individual domains (Oda et al., 2004, Ivanova et al., 2006, Mantsyzov et al., 2007). Reverse labeling of phenylalanine (Kelly et al., 1999) and the dual amino acid-selective ^{13}C-^{15}N labeling technique (Yabuki et al., 1998) were employed to resolve ambiguous assignments in $Q^{122}FM(Y)F^{126}$. Residual dipolar couplings of wt N-domain and $Q^{122}FM(Y)F^{126}$ were abstracted from the chemical shifts of TROSY and anti-TROSY cross-peaks in isotropic and anisotropic solvent conditions, respectively. Partial alignment of the proteins was induced by addition of 10 mg/ml bacteriophage *Pf1* (Hyglos GmbH). The axiality and rhombicity of the alignment tensor were calculated using PALES (Zweckstetter, 2008). Transverse relaxation time T_2 was obtained from the measurements with eight relaxation delays, *i.e.* 12.5, 25, 50, 62.5, 87.5, 112.5, 156.25, and 200 ms. The spectra measuring ^{1}H–^{15}N NOE were acquired with a 2-s relaxation delay, followed by a 3-s period of proton saturation. In the absence of proton saturation, the spectra were recorded with a relaxation delay of 5 s. The exponential curve fitting and data analysis were carried out using Origin 8 (Origin Lab).

2.6 Structure calculations

NOE distance restraints for the calculated structures of wt N-domain and $Q^{122}FM(Y)F^{126}$ were obtained from ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HMQC spectra, respectively. Backbone dihedral angle restraints (*i.e.* φ and ψ) were derived from the backbone $^{13}C'$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$, $^{1}H_\beta$ chemical shift values using TALOS (Cornilescu et al., 1999). Structure calculations were performed using CYANA 3.0 (Guntert et al., 1997, Herrmann et al., 2002b) and visualized using MOLMOL (Koradi et al., 1996) and PyMOL (Delano Scientific). Quality of the final structures was assessed using PROCHECK-NMR (Laskowski et al., 1996). The homology models of the N-domains of *Euplotes aediculatus* and *Stylonychia mytilus* were calculated by the I-TASSER server (Roy et al., 2010, Zhang, 2008). The Template Modeling (TM)-score of both models as compared to the structure of wt N-domain were calculated by the TM-align server (Zhang and Skolnick, 2005).

2.7 NMR structural characterization of RNA

All of the synthesized RNA oligonucleotides were purchased from 1st BASE (Singapore). The lyophilized RNA oligonucleotides were dissolved in diethylpyrocarbonate (DEPC)-treated water, and diluted into the NMR buffer supplemented with 2-5 mM $MgCl_2$. The RNA samples were incubated at 80 °C for two minutes followed by cooling to room temperature. Chemical shift assignments of ^1H from the nucleobases of 15-mer-UAA were based on the standard 2D NOESY (mixing time of 350 ms), DQF-COSY, and natural abundance ^{13}C-HSQC all acquired at 25°C, as well as ^{15}N-SOFAST-HMQC (Schanda et al., 2005) without ^{15}N-decoupling measured at 20°C. Assignments of ^1H from the other RNA constructs were based on their respective 2D NOESY spectra. H1' protons were only partially assigned. 1D ^{31}P-spectra with ^1H-decoupling were used to distinguish between double-stranded helical and stem-loop structures. In the N-domain/RNA titration experiments, the ^1H resonances of RNA were monitored as a function of added ^{13}C, ^{15}N-labeled wt N-domain to RNA, and *vice versa*. Suppression of ^1H resonances stemming from ^{13}C, ^{15}N-labeled protein was achieved using *J*-coupling filter followed by gradient purge. Chemical shift perturbations were calculated by $\mathrm{CSP} = \sqrt{\{(\Delta\delta H)^2 + (0.14\times\Delta\delta N)^2\}}$. The binding affinity can be estimated from the CSP values either using the Scatchard plot method (Fielding, 2003), or by curve fitting to $CSP = CSP_{max} / (2[P]) \times ([L] + [P] + K_d - \sqrt{([L] + [P] + K_d)^2 - 4[L][P]})$, where [L] and [P] are total concentrations of ligand and protein, respectively.

2.8 *In vitro* RF activity assay of chimeric eRF1s

Cloning and point mutagenesis of chimeric *Euplotes aediculatus*/human and *Stylonychia mytilus*/human eRF1 genes, expression and purification of chimeric eRF1 proteins were described (Lekomtsev et al., 2007, Eliseev et al., 2010). The eRF1 activity was measured in an *in vitro* system suggested by Caskey *et al.* (Caskey et al., 1974). Rabbit reticulocyte ribosomes were isolated and purified as described (Seit-Nebi et al., 2002). These experiments were performed by the group of Prof. Ludmila Frolova.

Chapter 3: Automatic Assignment of Protein Backbone Resonances from Nonlinear Sampled NMR Data

3.1 Abstract

A major advancement in biomolecular NMR would be the acquisition of sufficient data in the shortest time possible. Fast NMR technique will be useful to tackle problems like unstable sample and allow a more extensive set of experiments to be measured within certain time constraint. Multidimensional decomposition (MDD) is an algorithm to reconstruct spectra from nonlinear and sparsely sampled data, opening an avenue for minimal data acquisition time (Orekhov et al., 2003, Tugarinov et al., 2005, Jaravine et al., 2006). *AutoLink* is a robust algorithm for automatic sequential resonance assignment of biopolymers based on the idea of relative hypothesis prioritization (Masse and Keller, 2005). By combining both algorithms, we proposed the method of targeted acquisition of NMR data that is dynamically controlled by the completeness of automatic backbone resonances assignment. As demonstrated using the 12 kDa heme chaperone protein CcmE, 90 and 98% of manually assignable resonances were automatically assigned within 10 and 40 hours of nonlinear sampling of five 3D NMR spectra, respectively. This study shows that automatic resonance assignment might serve as a guiding criterion for optimal run-time allocation of NMR resources.

3.2 *Psyte* and *AutoLink II* programs

We introduce *Psyte*, a new module of the program *AutoLink II*, to work with the semi-automated resonance assignment program CARA. Both *Psyte* and *AutoLink II* are written by Dr. James Masse in Lua programming language. A schematic diagram and description of the algorithm are provided in Appendix A.

In brief, *Psyte* is developed for the recognition of spin systems within multidimensional NMR spectra. The module combines expert knowledge, systematic rules, and competition-based (non-monotonic) decision-making processes in order to group chemical shifts extracted from spectra into spin systems. *Psyte* resolves spectral overlap by deconvolution of peak models derived directly from the NMR spectra, but does not require any well-resolved peak to derive its models as is implemented in *XEASY* (Bartels et al., 1995). Additionally, *Psyte* cross compares spectra in order to validate decisions made on individual spectrum. The generated spin system grouping hypotheses are verified by establishing recursion loops to the spectral intensities.

Psyte successfully removes most spectral artifacts. The program's artifact detectors can be divided into two main categories, specific and non-specific. Specific artifact recognizers are designed to identify artifacts of a known type, *e.g.* due to truncation of time-domain signal, presence of strong solvent resonance, *etc*. The non-specific artifact detectors recognize artifacts by comparing the observed peaks against the program's "expert knowledge" of what the spectra are supposed to look like. These non-specific artifact recognizers rate NMR spectra according to *Psyte*'s cognitive understanding of what their data and artifact content are and adjust the spectrum interpretation accordingly.

Due to the fact that the results of its analysis are spin systems rather than purely ungrouped peaks, *Psyte*'s output is served as input for the downstream resonance assignment programs like *AutoLink II*. For the most difficult cases, *Psyte* has been designed such that it can work iteratively with a user and account for user modifications to the CARA repository, so that a spectroscopist can help the program in its analysis if cases are found where the program is prone to error.

3.3 Targeted time domain data acquisition and MDD reconstruction

An apo form of medium-sized protein apo-CcmE (Enggist et al., 2002) was used in course of the analysis. A fingerprint 2D [^{1}H-^{15}N]-projection of 3D HNCO spectrum reconstructed from 11% of sampling shows the majority of previously assigned resonances as well as missing and spurious cross-peaks (Figure 3.1). Five triple resonance spectra typically employed for backbone resonance assignment were measured using INUS schemes (Jaravine et al., 2006) with maximum numbers of sampled t_1 and t_2 hyper-complex time domain points in 3D spectra reaching 11% or 25% of the regular grid (see details in Table 1 in Chapter 2). In order to simulate run-time progression in the acquisition of NMR signal, MDD reconstructions were performed using selected quartets of 1D FIDs representing hyper-complex data points in t_1 and t_2 extracted in accordance with an INUS schedule from the pre-recorded 11% and 25% sampled INUS datasets. Therefore, we obtained 16 time-snapshots of the complete dataset available at 1.5 hours, 3 hours, 4.5 hours, and so on, after the start of acquisition. Analysis presented here in Chapter 3 was performed on five out of the 16 snapshots, *i.e.* 1.4, 2.8, 4.1, 5.5, and 11% for HNCO and HN(CO)CA, and 3.1, 6.3, 9.4, 12.5, and 25% for ct-HNCA, HN(CA)CO, and HNCACB.

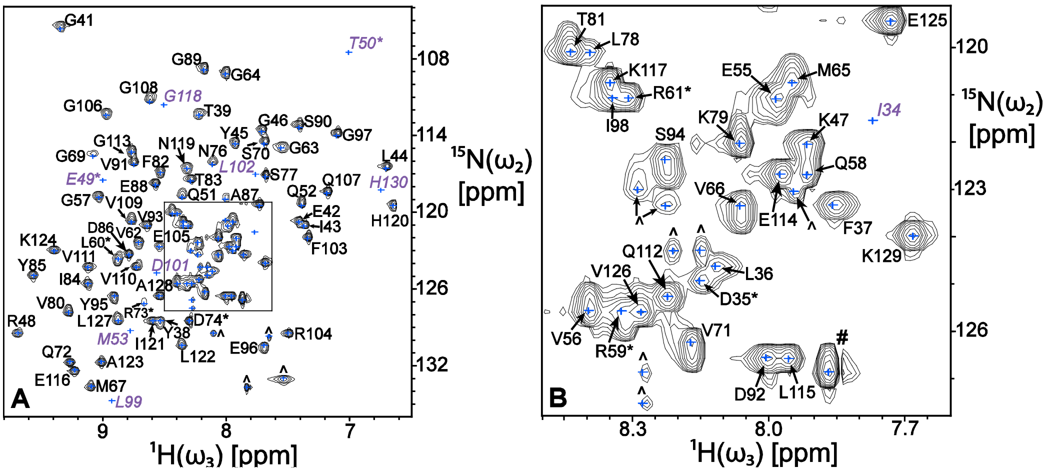


Figure 3.1 Assignment of the backbone amide chemical shifts of apo-CcmE. (A) [^{1}H-^{15}N]-projection of the 3D HNCO spectrum reconstructed from 11% of time domain data (see Table 1 in Chapter 2). (B) Inset of (A). All cross-peaks are marked by blue crosses, while those that are present in the reference HSQC spectrum but are not detected in the MDD-reconstructed spectra are indicated by magenta *italics*. Residues of which ambiguous assignment was obtained from *Psyte*/*AutoLink II* are indicated by asterisk. Cross-peaks due to degradation fragments are marked by “^”.

Figure 3.2 (A) shows the numbers of cross-peaks identified in the individual MDD-reconstructed spectra as a function of the progressively sampled grid. Direct inspection shows that HNCO and HN(CO)CA experiments have reached the targeted numbers of cross-peaks after sampling of *ca.* 8% of the grid, and hence, can be dropped from the acquisition schedule, dedicating the spectrometer resources to other more demanding experiments. For a closer look into the resolution and sensitivity of the resonance peaks in the ^{13}C-dimension of the MDD-reconstructed spectra, resonances of residue E105 in HNCACB, HN(CA)CO, and HNCO spectra are shown as an example (Figure 3.2, C, D, and E). As is expected in MDD reconstruction, the peak position and the line-width remained constant with the progression of data sampling, while gradual gain of spectral sensitivity occurred as the sampling level increased. That means a certain minimal percentage of data sampling is required for a given resonance peak to appear in the spectrum.

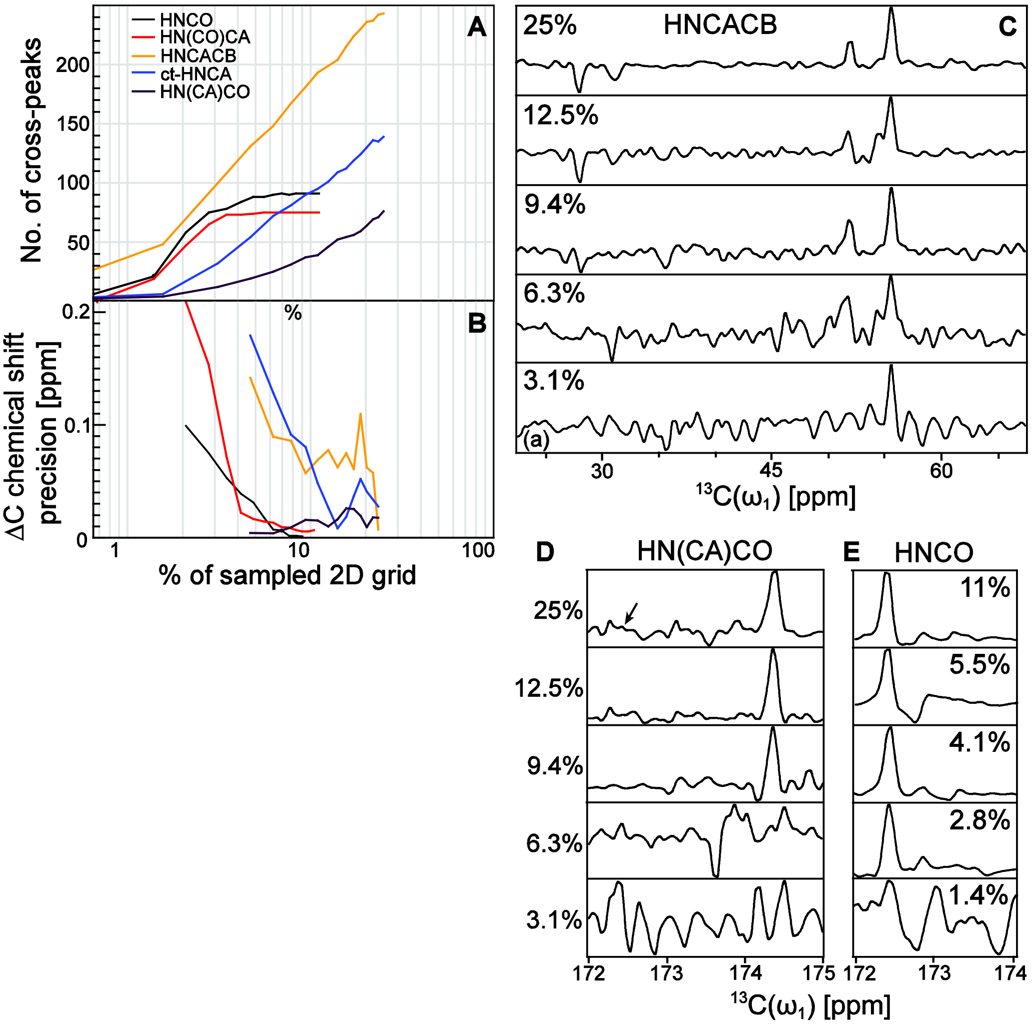


Figure 3.2 Progressive increase of S/N ratio and chemical shift precision in MDD-reconstructed spectra with increasing percentage of sampling. (A) The numbers of cross-peaks identified in the MDD-reconstructed 3D spectra versus fraction of sampled time domain. In this representation, the 100% corresponds to the number of sampled points spanning the full regular grid (see Table 1 in Chapter 2). (B) The precision of identified cross-peak positions versus fraction of sampled time domain. 1D traces along the ^{13}C-dimension of (C) HNCACB, (D) HN(CA)CO, and (E) HNCO spectra taken at δ^{15}N = 122.69

ppm and δ^{1}H = 8.54 ppm (*i.e.* residue E105) at various percentage of sampled time domain relative to the full grid. The arrow in (D) indicates the position of an expected inter-residual cross-peak still absent in HN(CA)CO spectrum sampled at 25% of the full grid.

3.4 Automatic determination of spin systems by *Psyte*

In the first step, *Psyte* groups resonances detected in a set of 3D heteronuclear spectra into spin systems. Due to the complexity inherent to multidimensional spectra reconstructed from sparse data, the spin systems obtained do not always correspond exactly to those typically identified by a spectroscopist. In the case of apo-CcmE, some extra spin systems were created from ambiguous regions in the spectra that contain only artifact resonances, and they were generally deficient in the relevant spins. Nonetheless, these spurious spin systems had rarely been assigned to the amino acid sequence by downstream logic in *AutoLink II*, owing to their low link hypothesis scores and competition against link hypotheses involving only real spin systems.

The amount of information directed to *AutoLink II* for residue specific assignment in the form of spin systems created by *Psyte* is dependent on different combinations of MDD-reconstructed spectra (Figure 3.3). Progressively sampled spectra, according to the corresponding percentage of sampling as indicated in Figure 3.2, were assembled into 13 sets forming three characteristic groups. Inspection of the identified spin systems in the first group (sets 1-6) indicates that for all reasonably high levels of data sampling, the numbers of created spin systems remain approximately constant (*i.e.* between 90 and 110) with the tendency of abstracting more spin systems from better sampled spectra. The fact that more noisy spectra do not result in higher numbers of spin systems created demonstrates the high tolerance of *Psyte*'s algorithm to presence of noise and spectral artifacts.

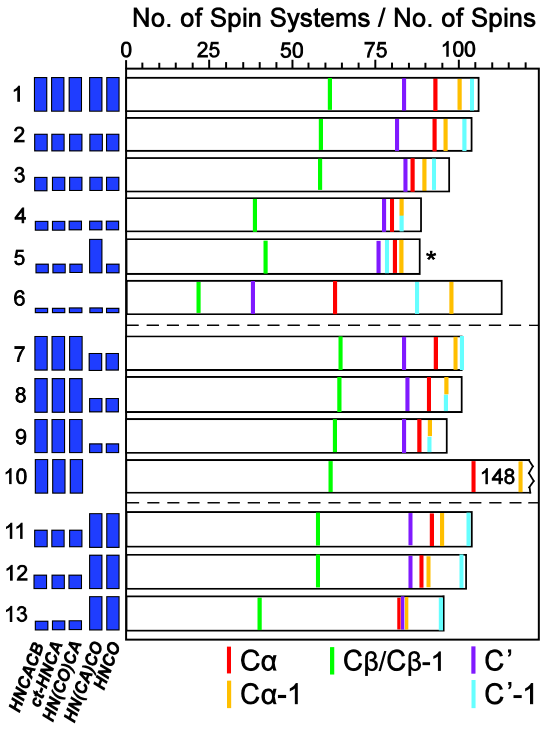


Figure 3.3 Analysis of automatic determination of spins and spin systems by *Psyte*. The numbers of spin systems (horizontal bars) and the numbers of the respective spins (colored lines) identified in 13 sets of MDD-reconstructed 3D spectra. The percentage of data sampling of each individual spectrum in the set is indicated by vertical full bars. In set 1, the full bars correspond to HNCACB (25%), ct-HNCA (25%), HN(CO)CA (11%), HN(CA)CO (25%), and HNCO (11%). In set 10, 148 spins systems were determined. Set 5 marked by asterisk contains HNCO sampled at 2.8% and HN(CA)CO sampled at 25% of the full time domain grid.

It is noteworthy that at the very low level of sampling represented by set 6 (*i.e.* 3.1% for HNCACB, HN(CA)CO, and ct-HNCA; 1.4% for HNCO and HN(CO)CA), significantly higher number of spurious spin systems can be picked with reduced total numbers of Cα/$C\alpha_{-1}$, and C’ spins. We identified the generally reduced S/N ratio in HN(CA)CO spectrum as the cause for this behavior of the algorithm. A selective increase in the sampling level of only this spectrum (set 5) drives the numbers of identified spins and spin systems closer to optimum found in sets 1-3. Therefore, besides reaching a minimal sampling level for all spectra (*ca.* above 6%), some individual spectra with low sensitivity must be preferentially sampled for larger proportion of time domain data, in order to remove or alleviate spin system abstraction “bottlenecks”. Based on this observation, a readjustment of the general sampling scheme can be performed (see Discussion in Chapter 3).

3.5 Advancement of *Psyte* over threshold filtering

Here we demonstrated the shortcoming of linear threshold-based peak-picking in spins abstraction from MDD-reconstructed spectra. We utilized a standard 2D- and 3D-picker for CARA on a series of HNCACB spectra. The 3D-picker requires prior determination of the amide groups from [^{1}H-^{15}N]-projection of the triple resonance spectrum. Clearly, the numbers of amide groups and peaks abstracted by the 2D- and 3D-picker, respectively, are dependent on the respective threshold values for both 2D and 3D spectra (Figure 3.4, A and B). As the thresholds increase, false peaks are being excluded and the numbers of amide group and peaks drop exponentially. However, the optimal numbers of amide groups and peaks do not fall on the curves where threshold values are sufficiently high. Using the HNCACB spectrum reconstructed from 25% of data sampling as an example, we found that threshold values of 80 and 20,000 for 2D and 3D, respectively, were optimal to abstract a reasonably good but not error-free set of spin systems (Figure 3.4, C). Upon inspection of the spin system list, three spin systems were picked from high-amplitude noise and eight spin systems that are assignable by *Psyte*/*AutoLink II* were missed (*i.e.* E55, Q58, R61, V66, S70, I84, V91, and D92). The missing of R61 is caused by the inability of the 2D-picker to discern spin systems with very close ^{15}N chemical shifts. On the other hand, *Psyte* was able to tackle such situation (Figure 3.4, D). The example shows overlap of peaks between I98 (broadened) and R61 (low amplitude). Even though there are four positive peaks in the ^{13}C-dimension of H/N 46, the standard 2D- and 3D-picker identified them as spins from a single spin system. Therefore, the outcome of automatic peak-picking based on linear threshold filtering is challenged by the dynamic range of the spectrum.

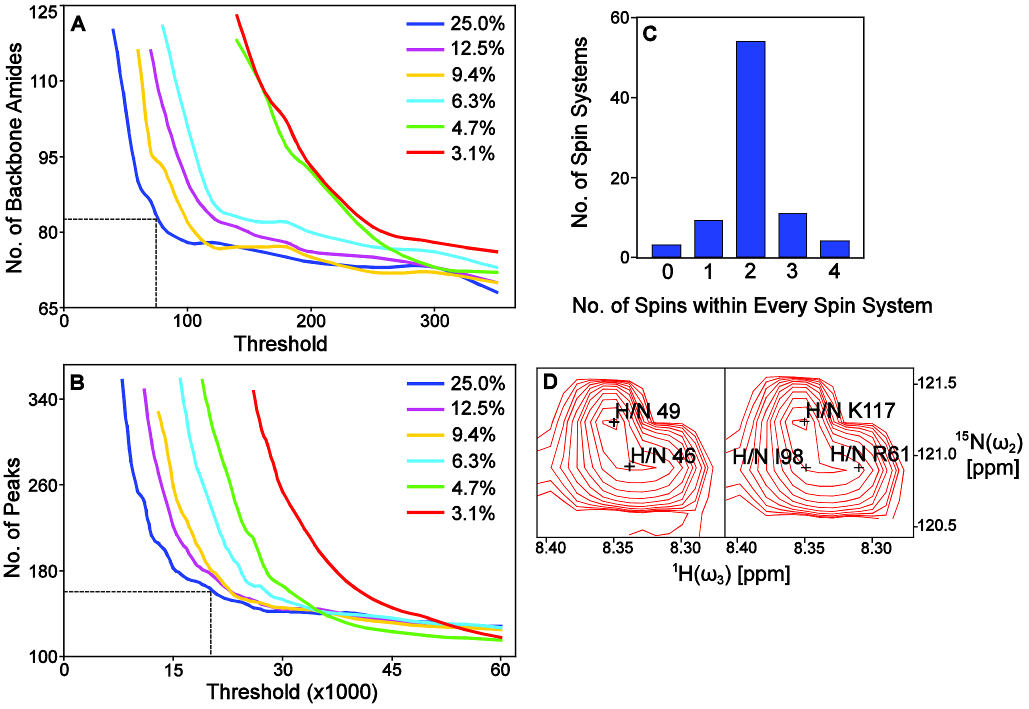


Figure 3.4 Performance of 2D- and 3D-picker based on threshold filtering. (A) The numbers of backbone amides identified from HNCACB spectra of various sampling levels versus the intensity thresholds set in the 2D-picker. 81

backbone amides were picked when the threshold was set to 80. (B) The total numbers of peaks identified from HNCACB spectra of various sampling levels versus the intensity thresholds set in the 3D-picker. 166 peaks were picked when the threshold was set to 80 and 20,000 for 2D- and 3D-picker, respectively. (C) From the list of 166 spins being picked, the numbers of spin systems which falls into categories of different numbers of spins within every spin system is shown. 54 spin systems contain two spins each (*i.e.* Cα and $C\alpha_{-1}$) and nine spin systems contain one spin each. This list was deemed reasonably close to the optimal upon direct inspection. (D) From the list of spin systems mentioned in (C), a single spin system (*i.e.* H/N 46) was determined using the standard 2D- and 3D-picker, while there were in fact two resolved spin systems, which were correctly picked up and assigned by *Psyte*/*AutoLink II* as residues I98 and R61.

3.6 Consistency of residue specific assignment by *AutoLink II*

Residue specific assignment was achieved by *AutoLink II* using the amino acid sequence of apo-CcmE (L30-H130) and the spin systems identified by *Psyte* as the input to the automatic assignment module. Due to the random nature of the order of generated hypothesis on spin system groupings, slightly different lists of spin systems were generated even for small variation of control parameters. This variability of spin system determination is subsequently propagated to variations in residue specific assignment (Figure 3.5), which are typically found for residues with low spectral S/N ratio such as residues R59 and R73 (Figure 3.1). There are instances where slightly different chemical shifts were assigned to the same spin, to variable extent between different spins. Therefore, several lists of spin systems were generated using the same set of spectra followed by separate attempts of residue specific assignment (Figure 3.5). The discrepancies in the assignment between the two (or more) spin systems lists helped to isolate problematic residues and regions in the spectra, thus allowing spectroscopists to either spot the wrong assignments (*e.g.* R59), or assign more residues after manually inspecting the relevant spin systems and cross-peaks in the spectra (*e.g.* D35). This logic can be implemented in the future development of *AutoLink II*. In addition, *Psyte* can also be tweaked to minimize the variability in spin system determination in such a way that optimizes the outcome of *Psyte*/*AutoLink II*.

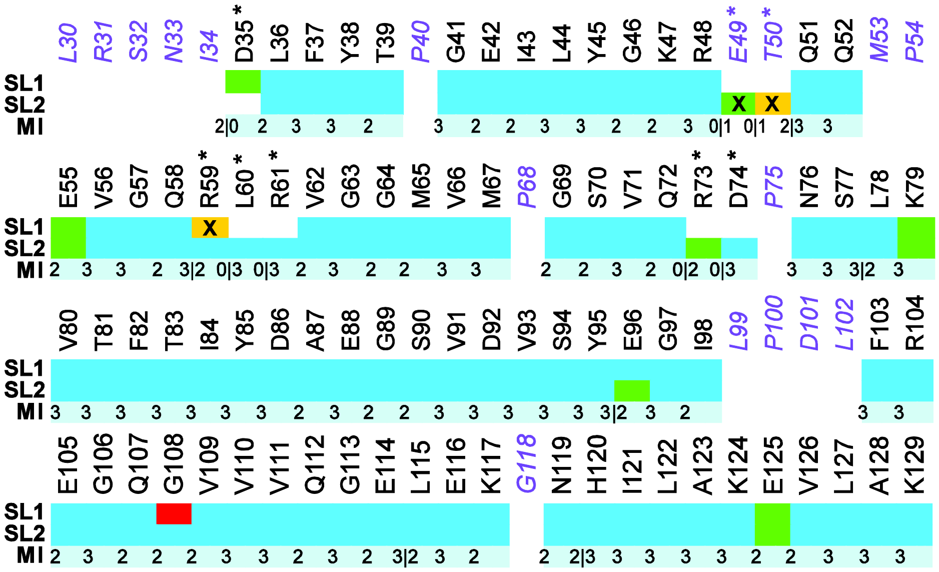

L30 R31 S32 N33 I34 D35* L36 F37 Y38 T39 P40 G41 E42 I43 L44 Y45 G46 K47 R48 E49* T50* Q51 Q52 M53 P54
SL1
SL2
X X
MI 2|0 2 3 3 2 3 2 2 3 3 2 2 3 0|1 0|1 2|3 3
E55 V56 G57 Q58 R59* L60* R61* V62 G63 G64 M65 V66 M67 P68 G69 S70 V71 Q72 R73* D74* P75 N76 S77 L78 K79
SL1
SL2
X
MI 2 3 3 2 3|2 0|3 0|3 2 3 2 2 3 3 2 2 3 2 0|2 0|3 3 3 3|2 3
V80 T81 F82 T83 I84 Y85 D86 A87 E88 G89 S90 V91 D92 V93 S94 Y95 E96 G97 I98 L99 P100 D101 L102 F103 R104
SL1
SL2
MI 3 3 3 3 3 3 3 2 3 2 2 3 3 3 3 3 3|2 3 2 3 3
E105 G106 Q107 G108 V109 V110 V111 Q112 G113 E114 L115 E116 K117 G118 N119 H120 I121 L122 A123 K124 E125 V126 L127 A128 K129
SL1
SL2
MI 2 3 2 2 2 3 3 2 3 2 3|2 3 2 2 2|3 3 3 3 3 2 2 3 3 3

Figure 3.5 Overview of residue specific assignment of backbone resonances by *Psyte/AutoLink II*. Two spin systems lists (SL) determined in separate executions of *Psyte* using 25% and 11% INUS sampled spectra (see Table 1 in Chapter 2). The *AutoLink* sequence fit score (Masse and Keller, 2005) of individual residue is represented by the following color scheme: 0.81-1.00 (blue), 0.71-0.80 (green), 0.61-0.70 (yellow), and 0.60 and below (red). Cross (X) represents wrong assignment as compared to the reference manual assignment (Enggist et al., 2002). The asterisk (*) indicates discrepancy in resonance assignment between SL1 and SL2 (*e.g.* R59 and D35). After manual inspection (MI) of identified spin systems in *Psyte* output, these discrepancies can be resolved resulting in the final accurate assignment colored turquoise.

The numbers of inter-residue matches of $C\alpha_{-1}/C\alpha$, $C\beta_{-1}/C\beta$, and C'_{-1}/C' chemical shifts are listed below the amino acid sequence representing the availability of connectivity information within the assigned spin systems in SL1 and SL2. Residues in magenta *italics* are designated as ‘not-assignable’, due to fact that they are either proline residues or the corresponding cross-peaks are broadened beyond detection in all MDD-reconstructed 3D spectra.

AutoLink II provides a graphical output of assigned residues as well as statistics associated with the quality of assignment as described previously (Masse and Keller, 2005). The sequence fit scores of the automatically achieved assignment show that almost all assigned chemical shifts score higher than 0.80 due to generally good spectral S/N ratio (Figure 3.5). Even though the wrong assignments in this particular instance could be identified by their low sequence fit scores (*i.e.* R59 in SL1, and E49 and T50 in SL2), such benchmark was not applicable for the sets of spectra with lower levels of data sampling as much more ambiguity were abound. In those cases, consistency of residue specific assignment between two spin systems lists was employed as the benchmark (Figure 3.7).

3.7 Robustness of *AutoLink II* in residue specific assignment

AutoLink II had demonstrated its robustness in several difficult cases. The peak position of $C\alpha_{-1}$ of G108 assigned by *Psyte* in SL1 was slightly off-center and did not match well with that of Cα of Q107, as shown by the extremely low sequence fit score (Figure 3.5). Nonetheless, the assignment was correctly judged by *AutoLink II* since it was supported by other link hypothesis and did not result in violation of the overall assignment result. In general, assignment of the various spectral features due to MDD reconstruction artifacts to the amino acid sequence was never observed. Furthermore, resonances due to degradation products (Figure 3.1) were also not assigned by *AutoLink II*. Although the assignments of residues D35, R73, and D74 were problematic as the resonances of L36 and R73 had been broadened to great extent (Figure 3.1), yet they were correctly assigned by *AutoLink II* in SL1 and SL2, respectively.

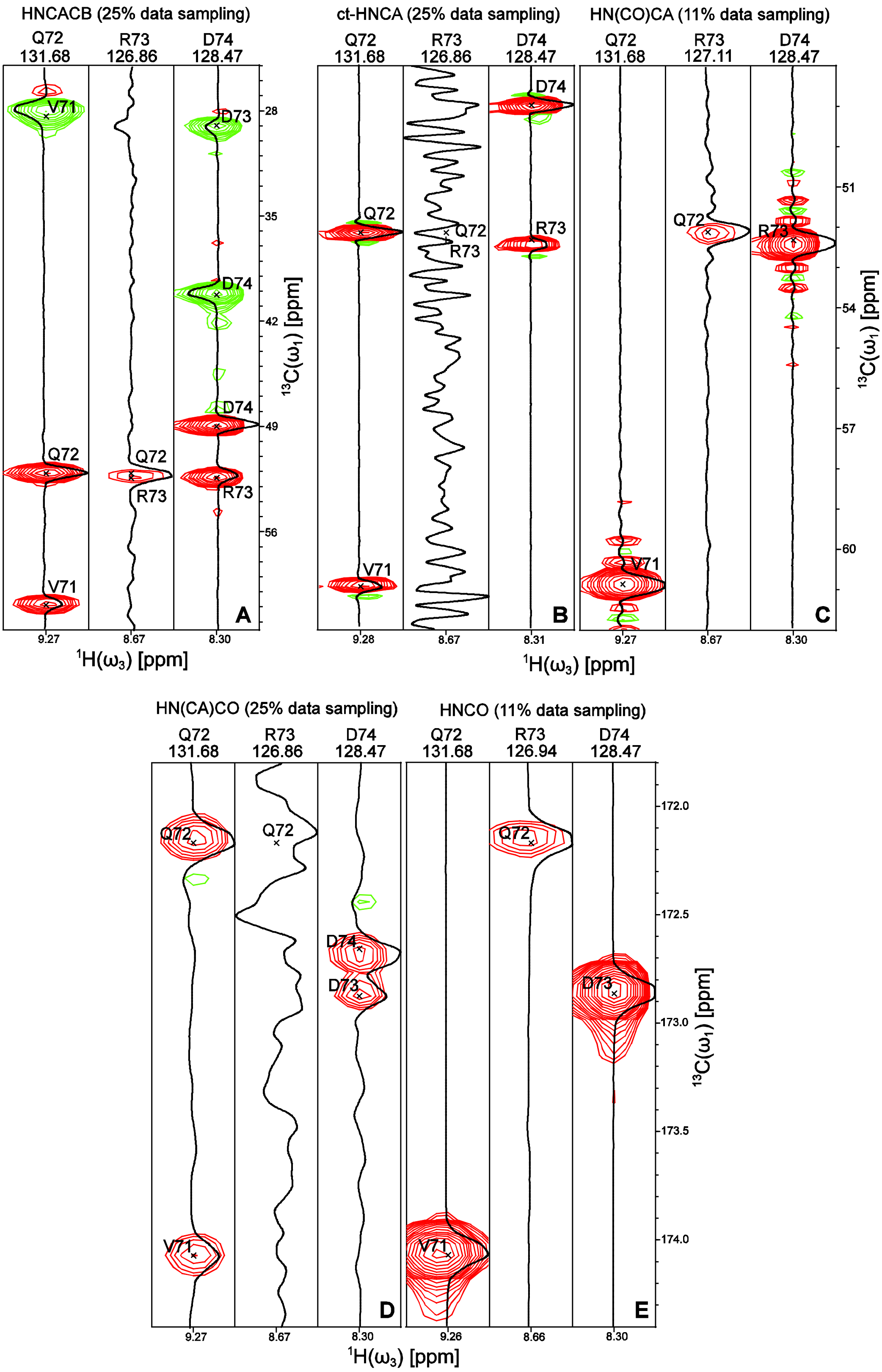


Figure 3.6 Assignment of R73 and D74 with limited information. 2D contour plot strips and the corresponding 1D traces along the ^{13}C-dimension of MDD-reconstructed spectra of (A) HNCACB, (B) ct-HNCA, (C) HN(CO)CA,

(D) HN(CA)CO, and (E) HNCO for residues Q72, R73, and D74 taken at their respective δ^{15}N and δ^{1}H. Resonances of R73 in both ct-HNCA and HN(CA)CO were not observable at all.

The capability of *Psyte/AutoLink II* to assign R73 and D74 from the five spectra in set 1 (*i.e.* 25% for HNCACB, ct-HNCA, and HN(CA)CO, and 11% for HN(CO)CA and HNCO) illustrates the robustness of the algorithm emulating expert reasoning as well as the implementation of targeted acquisition of multiple spectra, in dealing with seriously broadened resonances and degenerated chemical shifts (Figure 3.6). The identity of the degenerated Cα and Cα$_{-1}$ resonances of R73 could be deduced from the HNCACB and HN(CO)CA spectra, even though the resonances had disappeared completely in the ct-HNCA spectrum (Figure 3.6, A, B, and C). Similarly, the C' and C'$_{-1}$ resonances of R73 were completely lost in the HN(CA)CO spectrum (Figure 3.6, D). Given the limited information (*i.e.* without the Cα connectivity and the C' chemical shift of R73), the assignment of both R73 and D74 was still achievable in SL2.

3.8 Completeness of residue specific assignment for targeted acquisition

Our aim is to achieve real-time targeted acquisition of NMR data assisted by automatic assignment of protein backbone resonances. Feasibility of real-time targeted acquisition scheme applicable to proteins prone to degradation was tested off-line by attempting automatic resonance assignment using pre-recorded and MDD-reconstructed spectra at various percentages of time domain sampling (*i.e.* sets 1 to 13 in Figures 3.3 and 3.7). This bootstrapping approach was used to explore robustness of the process and to define how much data is needed to reach the target as it allows to consider many scenarios, which otherwise would require real-time recording of all of them, which is technically hard taking into account fast degradation of the protein sample. After a suitable schedule of acquisition and assignment has been established, we simulated real-time coupling of data acquisition and automatic assignment. The results show that the real-time control is computationally feasible with metric time needed to reconstruct the corresponding 3D spectra using a Linux station equipped with four central processing units always not exceeding about half of NMR spectrometer acquisition time needed to partially and simultaneously sample the five 3D spectra. Every assignment attempt takes one to two hours depending on the convergence properties of the problem at hand, and typically runs in parallel with MDD reconstructions, each on a separate central processing unit node.

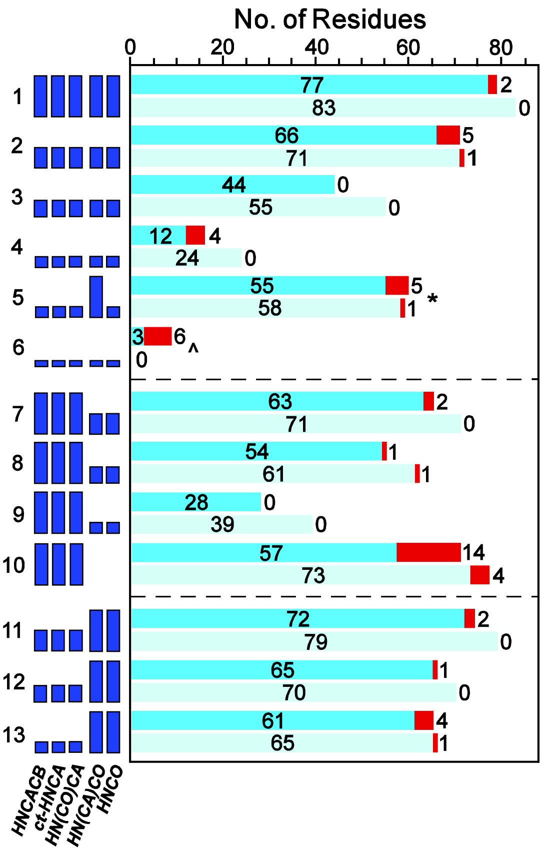


Figure 3.7 Completeness of residue specific assignment by *Psyte*/*AutoLink II* is restricted by spectrum with low S/N ratio. The numbers of consistently assigned residues (blue), the maximum numbers of wrong assignments (red), and the total numbers of assigned residues after manual inspection of identified spin systems in *Psyte* output (turquoise) achieved in 13 different sets of MDD-reconstructed spectra at various levels of data sampling (same as in Figure 3.3). Set 5 marked by asterisk contains HNCO spectrum sampled at 2.8% and HN(CA)CO spectrum sampled at 25% of the full grid. For set 6 marked by "^", no correct assignment is achieved.

Contribution of individual spectrum to the overall assignment result was investigated by fixing the level of data sampling for HNCACB, ct-HNCA, and HN(CO)CA spectra, while reducing that for HNCO and HN(CA)CO spectra progressively (*i.e.* sets 7 to 10), and *vice versa* (*i.e.* sets 11 to 13) (Figure 3.7). The group with fixed levels of data sampling for HNCO and HN(CA)CO spectra (*i.e.* sets 11 to 13) clearly outperformed the other group (*i.e.* sets 7 to 9). This indicates that the percentage of data from which HNCO and HN(CA)CO spectra were reconstructed could restrict the completeness of assignment, and hence, representing the "assignment bottleneck". In fact, the most significant limiting factor found in this study was the quality of HN(CA)CO spectra – a case which can be justified by the drastic improvement in assignment completeness of set 5 over that of set 4 (Figure 3.7). Therefore, the percentage of time domain sampling of individual spectrum should be guided by the inherent relative sensitivity of the respective NMR experiments. Upon closer inspection, we have also found that ct-HNCA and HNCACB spectra are complementing each other to provide adequate Cα connectivity information even at situations with very low S/N ratio in the spectra reconstructed from very low percentage of data sampling (*i.e.* sets 11 to 13).

3.9 Discussion

We have developed a method of automatic resonance assignment process with *Psyte*/*AutoLink II* to handle specific problems arising in real-time automatic analysis of NMR data (Fiorito et al., 2006). The complexity of the problem has been effectively reduced by the introduction of two levels of analysis. First, *Psyte* sorts out variations of chemical shifts among the 3D spectra and determines a single representative spin system list. At the second level, *Autolink II* attempts to assign the backbone resonances to the amino acid sequence while simultaneously identifying spin systems that may arise from artifacts and degradation products. In order to systematically explore limits of applicability and robustness of such a combination as well as for the purpose of establishing an optimal schedule of data acquisition, this method was applied to the apo form of medium-sized protein apo-CcmE, of which exhaustively analyzed multidimensional NMR spectra and 3D structure are available (Enggist et al., 2002). We find this protein a relevant model since the NMR spectra of this protein, in addition to a set of resonances stemming from the structurally defined core, feature signals from identifiable degradation products as well as resonances broadened by conformational exchange located in loops and flanking regions. Overall, 83 ^{1}H-^{15}N cross-peaks were identified as assignable (among which five exhibiting significantly reduced intensity), 11 were assigned to proteolytic degradation fragments showing reduced intensity in all connected spin systems and 13 backbone resonances were broadened beyond detection (Figure 3.1). We demonstrated that in the absence of human intervention and with the use of the optimal spectrum sampling scheme, 95% of the assignable resonances can be stably assigned.

Previously the number of cross-peaks picked from reconstructed spectra was set as target for data acquisition (Jaravine et al., 2006). This acquisition termination criterion is shown to be sufficient for chemically stable proteins with homogenous distribution of cross-peak intensities throughout the amino acid sequence. However, the application of a spectrum-wide global noise threshold (as implemented in most standard peak-pickers) on spectra from degrading proteins might result in missing assignments, especially at lower sampling levels. This is due to the presence of low intensity resonances stemming from dynamic regions of protein, which could be missed by threshold filtering, as well as the presence of low molecular weight degradation products giving rise to spurious cross-peaks. In addition, the noise in the MDD-reconstructed spectra is not anymore uniformly distributed throughout the spectrum, but rather appears in bands centered at ^{1}H and ^{15}N resonances along indirectly detected dimensions (Figure 3.8). On this basis, an appropriate estimate of the spectrum-wide threshold value becomes difficult to obtain (Figure 3.4).

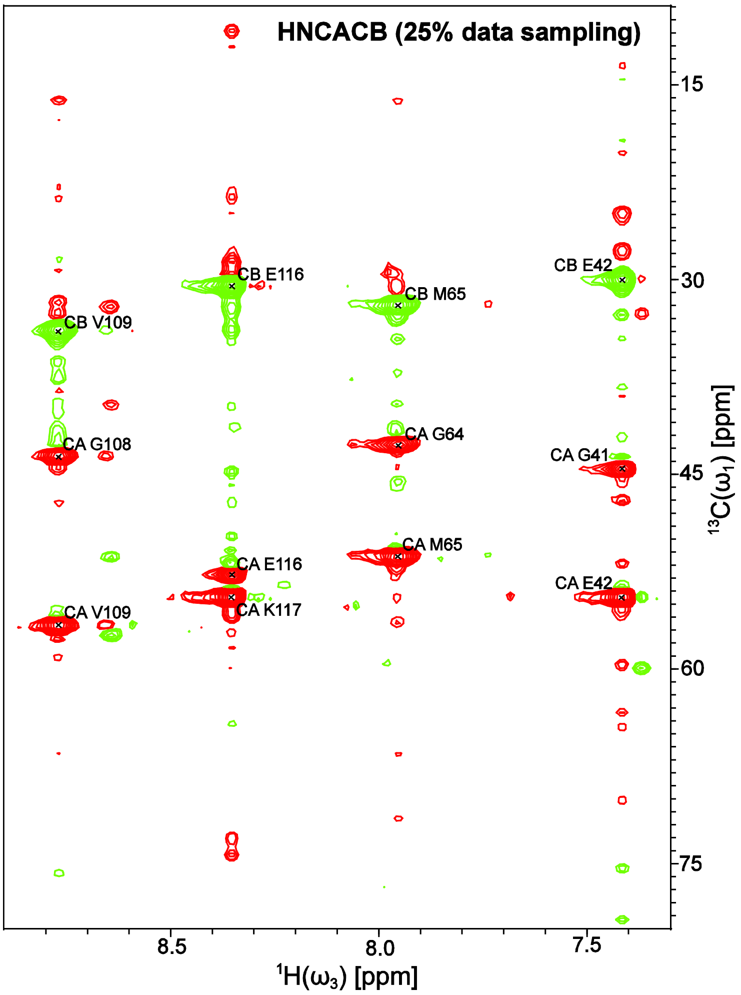


Figure 3.8 Noise band in a MDD-reconstructed spectrum. [^{1}H,^{13}C]-plane at δ^{15}N = 120.761 ppm of HNCACB spectrum reconstructed from 25% of the full time domain grid.

The critical novelty in our current approach is to replace linear threshold-based peak picking with hypothesis-driven recursive construction of spin systems. We note that the use of previous implementations of *AutoLink* required manual grouping of peaks from several peak-lists into a single list of spin systems, which is deemed impossible in a real-time data acquisition. In essence, reliable and complete identification of cross-peaks requires *a priori* knowledge of expected signals together with extensive noise filtering procedures implemented in the module *Psyte*. This observation led us to conclude that simple abstraction of complex spectrum to a list of cross-peaks might suffer from inherent inability to correctly decompose peak clusters in the absence of higher order analysis information (Figure 3.4, D).

In conclusion, our method provides a flexible criterion to optimally allocate NMR resources, based on the completeness and accuracy of achievable assignment as well as the level of protein degradation. Since degradation of biomolecules might limit the accumulation of spectral S/N ratio with the overall NMR time invested, devising optimal strategies to guide data acquisition represents a new and important avenue in the field of automation of NMR structure determination.

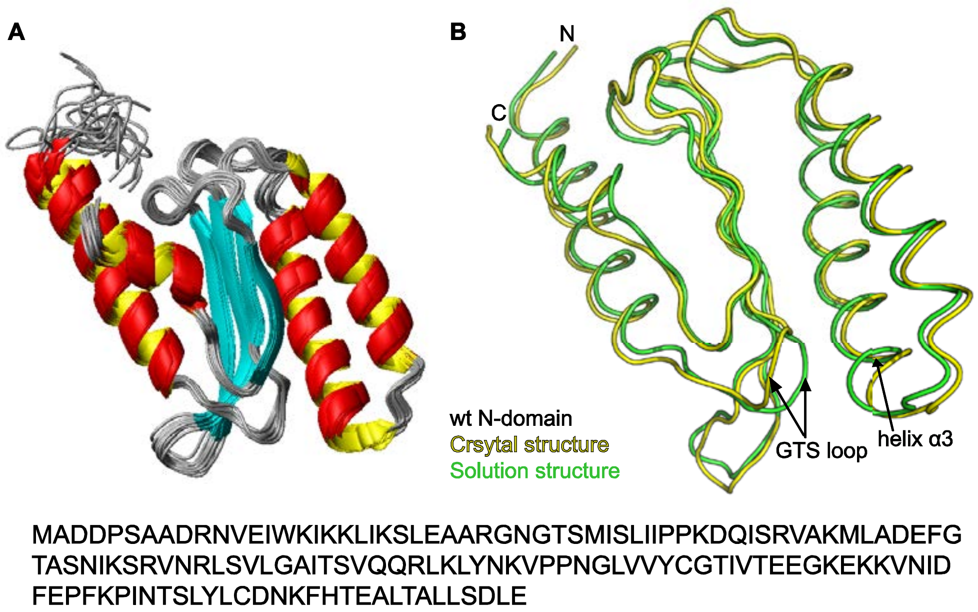
Chapter 4: Selectivity of Stop Codon Recognition is Modulated by Multiple Conformations of GTS Loop in eRF1

4.1 Abstract

A mutant of human eRF1 with four amino acid substitutions in N-domain (*i.e.* T122Q, S123F, L124M, and L126F), corresponding to eRF1 of *Stylonychia* (a genus of ciliate) with UGA as its only stop codon (Figure 4.7), was shown to exhibit strong UGA unipotency (Lekomtsev et al., 2007). To understand the structural effect of the quadruple mutations that lead to unipotency for UGA in the mutant eRF1, we determined the solution structures of wild-type (wt) N-domain and the UGA-unipotent mutant (denoted as $Q^{122}FM(Y)F^{126}$ henceforth), and compared them with the corresponding crystal structure of N-domain in full-length human eRF1 (Song et al., 2000). We found that those point mutations, while preserving the global structure of N-domain, perturb the hydrophobic core of N-domain and alter conformation of the strictly conserved GTS loop (residues 31 to 33) remote from the mutation sites. These results indicate that switching between omnipotency and unipotency of eRF1 may be modulated by distinct conformations of the GTS loop. Furthermore, the GTS loop is also a more dynamic part of N-domain on pico- to nanosecond timescale, suggesting it as a potential ligand binding site.

4.2 The solution structures of wt N-domain and $Q^{122}FM(Y)F^{126}$

The solution structures of both wt N-domain and $Q^{122}FM(Y)F^{126}$ were solved by NOE-derived distance restraints, and verified against measured residual dipolar couplings (RDCs) (courtesy of Dr. Li Yan and Ms. Shubhadra Pillay). The structure determination statistics of wt N-domain and $Q^{122}FM(Y)F^{126}$ are reported in Tables 4.1 and 4.2, respectively. As expected, the NMR solution structure of wt N-domain matches closely with the crystal structure of N-domain in full-length eRF1 (Figure 4.1). The only observed significant deviations are positioning of the N-terminal part of helix α3 and conformation of the GTS loop.



MADDPSAADRNVEIWKIKKLIKSLEAARGNGTSMISLIIPPKDQISRVAKMLADEFG
TASNIKSRVNRLSVLGAITSVQQRLKLYNKVPPNGLVVYCGTIVTEEGKEKKVNID
FEPFKPINTSLYLCDNKFHTEALTALLSDLE

Figure 4.1 Comparison between crystal and solution structures of wt N-domain. (A) The ensemble of 20 lowest energy NMR structures of wt N-domain. (B) Superposition of the crystal structure (green) (PDB ID: 1DT9) and the solution structure (yellow) of wt N-domain.

NOE restraints

Total unambiguous distance restraints	2139
Intra-residual	616
Sequential (| i – j | = 1)	622
Short-range (| i – j | ≤ 1)	1238
Medium (2 ≤ | *i* – *j* | ≤ 4)	418

Long-range ($\| i - j \| \geq 5$)	483
Hydrogen bond restraints	77
RMSD from the average atomic coordinates (residues 6-140, Å)	
Backbone atoms	0.78 ± 0.18
All heavy atoms	1.14 ± 0.16
Ramachandran analysis (%)	
Residues in most favored regions	66.9
Residues in additional allowed regions	27.2
Residues in generously allowed regions	5.8
Residues in disallowed regions	0.1

Table 4.1 Structural statistics for the final 20 conformers of wt N-domain. None of the structures exhibits distance violations greater than 0.2 Å or dihedral angle violations greater than 5°.

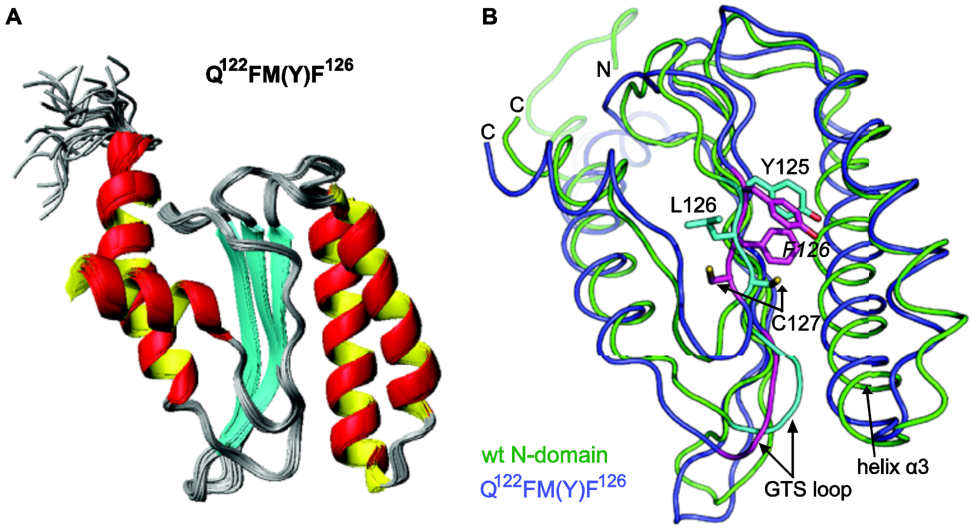
NOE restraints	
Total unambiguous distance restraints	3081
Intra-residual	824
Sequential ($\| i - j \| = 1$)	715
Short-range ($\| i - j \| \leq 1$)	1539
Medium ($2 \leq \| i - j \| \leq 4$)	619
Long-range ($\| i - j \| \geq 5$)	923
Dihedral angle restraints	256
Hydrogen bond restraints	75
RMSD from the average atomic coordinates (residues 6-140, Å)	
Backbone atoms	0.26 ± 0.05
All heavy atoms	0.68 ± 0.03
Ramachandran analysis (%)	
Residues in most favored regions	81.3

Residues in additional allowed regions	18.7
Residues in generously allowed regions	0.0
Residues in disallowed regions	0.0

Table 4.2 Structural statistics for the final 20 conformers of Q^{122}FM(Y)F^{126}. None of the structures exhibits distance violations greater than 0.2 Å or dihedral angle violations greater than 2°.

4.3 The GTS loop adopts distinct conformations in wt N-domain and Q^{122}FM(Y)F^{126}

The global structure of Q^{122}FM(Y)F^{126} is well superimposable with wt N-domain (Figure 4.2). Their structural differences are confined to β-strand β4 that contains the quadruple point mutations, as well as the GTS loop. Likewise, helix α3 of Q^{122}FM(Y)F^{126} is also repositioned relative to wt N-domain (Figure 4.2). Although the point mutations are spatially remote from the GTS loop, structural alteration that occurred to the GTS loop in Q^{122}FM(Y)F^{126} is evident from the difference in amide chemical shift for the GTS loop and several other nearby residues including C97 and T99 (Figure 4.3, A and B). Furthermore, different GTS loop conformations are also confirmed by clear differences in the NOEs patterns between wt N-domain and Q^{122}FM(Y)F^{126} (Figure 4.3, C).


A
Q122FM(Y)F126
B
N
C
C
Y125
L126
F126
C127
wt N-domain
Q122FM(Y)F126
GTS loop
helix α3

MADDPSAADRNVEIWKIKKLIKSLEAARGNGTSMISLIIPPKDQISRVAKMLADEFG
TASNIKSRVNRLSVLGAITSVQQRLKLYNKVPPNGLVVYCGTIVTEEGKEKKVNID
FEPFKPIN**Q**$_{122}$**FM**Y**F**$_{126}$CDNKFHTEALTALLSDLE

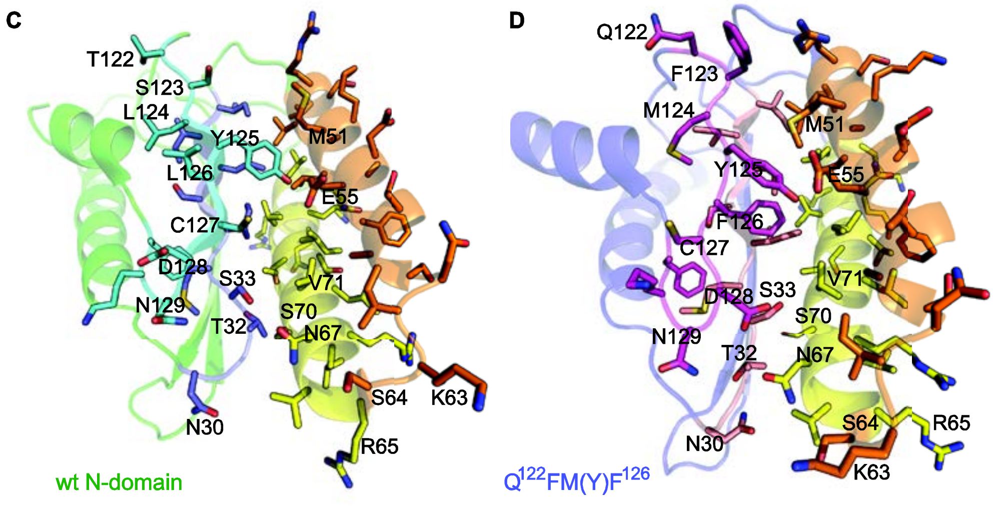

C
T122
S123
L124
Y125
M51
L126
E55
C127
D128
S33
V71
N129
S70
T32
N67
S64
K63
N30
R65
wt N-domain
D
Q122
F123
M124
M51
Y125
E55
F126
C127
V71
D128
S33
S70
N129
T32
N67
N30
S64
R65
K63
Q122FM(Y)F126

Figure 4.2 Structural comparison between Q^{122}FM(Y)F^{126} and wt N-domain. (A) The ensemble of 20 lowest energy NMR structures of Q^{122}FM(Y)F^{126}. (B) Superposition of the solution structures of wt N-domain (green) and Q^{122}FM(Y)F^{126} (blue), and the regions that are structurally distinct are labeled and highlighted in cyan (wt N-domain) and magenta (Q^{122}FM(Y)F^{126}). (C) Structure of wt N-domain showing the side-chains from the GTS loop and β1 (blue), β4 (cyan), α2 and the NIKS loop (orange), and α3 (yellow). (D) Structure of Q^{122}FM(Y)F^{126} showing the side-chains from the GTS loop and β1 (pink), β4 (magenta), α2 and the NIKS loop (orange), and α3 (yellow).

The alternative conformation of the GTS loop in Q^{122}FM(Y)F^{126} is maintained via an intricate propagation of hydrogen bonding perturbations from the mutation sites situated at the beginning of β-strand β4 that constitutes the hydrophobic core of N-domain (Figure 4.2). In wt N-domain, this β-strand forms a well-defined network of hydrogen bonds with the adjacent β-strand starting from L124 to D128, as seen from the alternating directions of the side-chains in consecutive residues. In Q^{122}FM(Y)F^{126}, the regular hydrogen-bonding network is disrupted starting from M124. Mutation L126F is critical for breaking the regularity, since the phenylalanine aromatic ring is found flipped to the opposite side of β-strand β4 relative to the side-chain position of L126, thus inflicting positional change of C127. As the side-chain of C127 is moved away from the hydrophobic core formed between the central β-sheet, helix α2, and helix α3 (Figure 4.4), the N-terminal part of helix α3 is repositioned closer to the GTS loop (Figure 4.2). The phenylalanine substitution at position 126 in three of the ciliates that are unipotent for UGA suggests that similar structural features may have causal effect on the UGA unipotency in those organisms (Figure 4.7). Nevertheless, the structural alteration in Q^{122}FM(Y)F^{126} is likely to be an additive effect from all of the four point mutations (Lekomtsev et al., 2007).

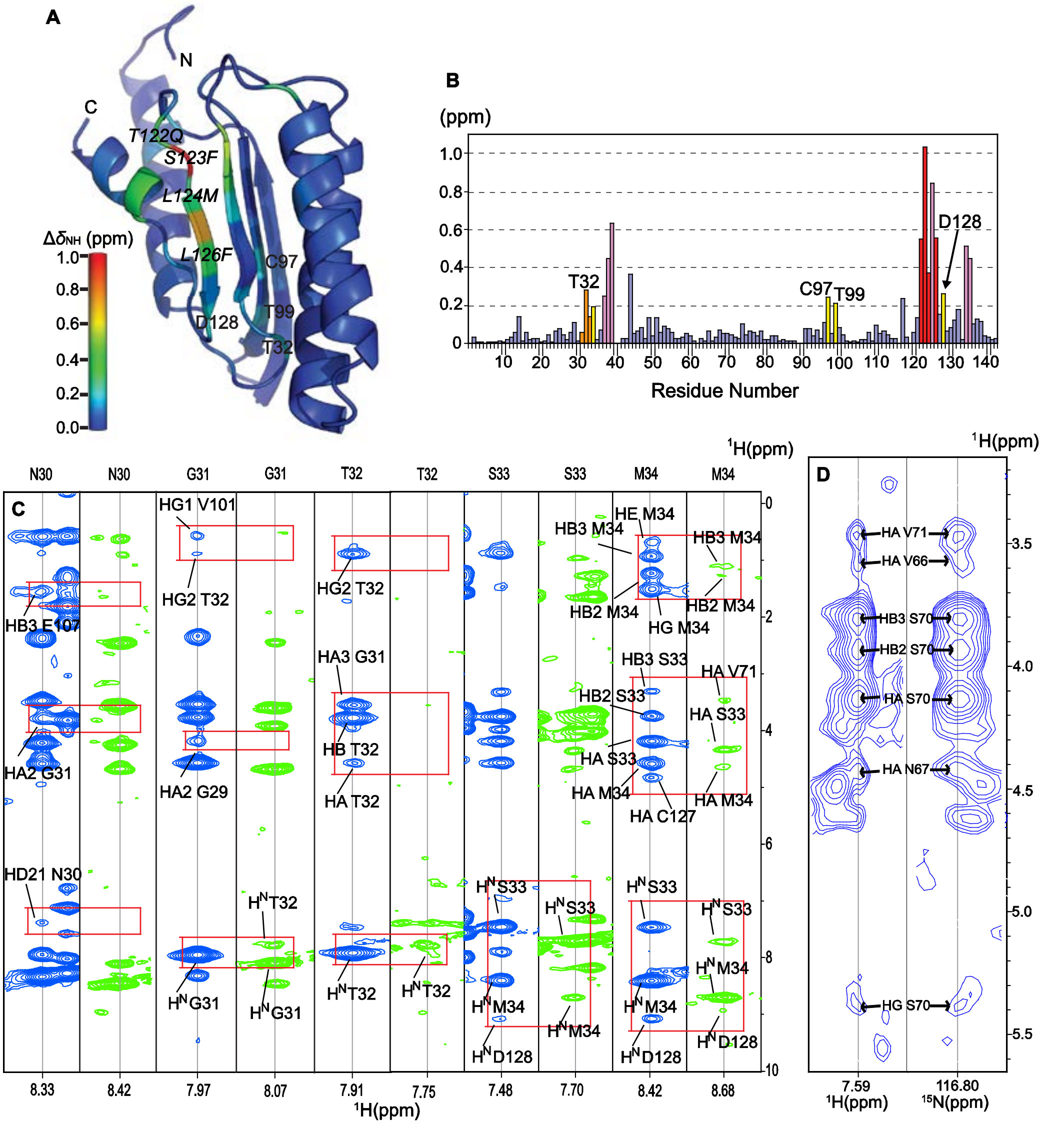
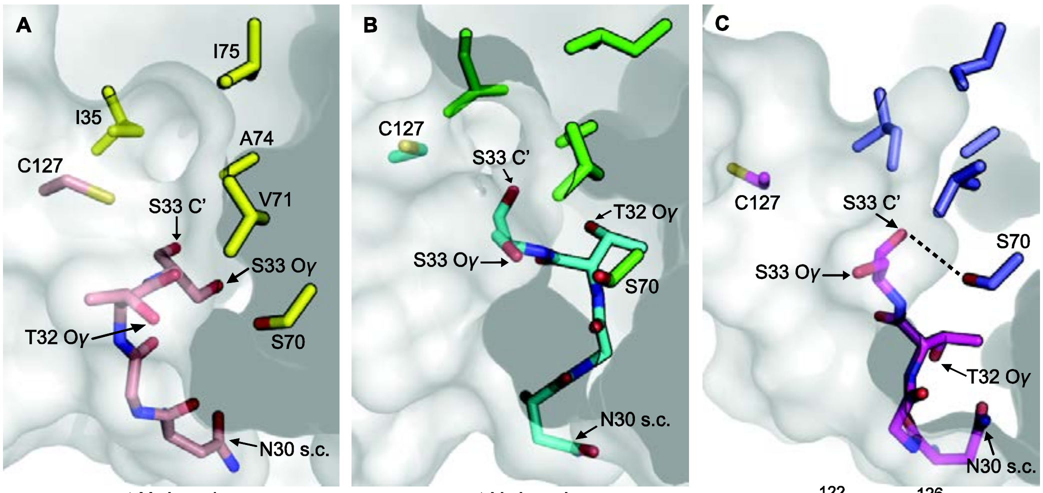


Figure 4.3 Distinct conformations of the GTS loop in wt N-domain and $Q^{122}FM(Y)F^{126}$. (A) Difference in amide chemical shift ($\Delta\delta_{NH} = \sqrt{\{(\Delta\delta H)^2 + (0.14\times\Delta\delta N)^2\}}$) between wt N-domain and $Q^{122}FM(Y)F^{126}$ are mapped onto the structure of wt N-domain according to the color scale. (B) Plot of the same amide chemical shift difference ($\Delta\delta_{NH}$) as (A), with specific residues highlighted as follows: point mutations (red), residues close to the point mutations in space (pink), the GTS loop (orange), and residues close to the GTS loop in space (yellow). (C) Differences in the NOE patterns as observed from the amides of N30 to M34 of wt N-domain (blue) and $Q^{122}FM(Y)F^{126}$ (green) show that their GTS loops have distinct conformations. (D) The NOE cross-peak of HG S70 as observed from the amide of S70 indicates participation of

the hydroxyl group in hydrogen bonding. This NOE cross-peak was not observed in wt Ndomain.

4.4 Multiple conformations of the GTS loop and its implication

Remarkably, the GTS loop adopts distinct conformations in structures solved at different situations, namely the crystal and solution structures of wt N-domain as well as the solution structure of $Q^{122}FM(Y)F^{126}$ (Figure 4.4). This implies that the GTS loop has the flexibility to adopt different conformations even within wt N-domain. In fact, the GTS loop in the solution structures has rather defined conformations, as backbone RMSD of the loop region (N30-M34) in wt N-domain and $Q^{122}FM(Y)F^{126}$ are 0.32 ± 0.28 Å and 0.11 ± 0.06 Å, respectively. This suggests that the observed GTS loop conformations might represent end-states in a complex equilibrium of different conformations.



wt N-domain
Crystal Structure

wt N-domain
Solution Structure

$Q^{122}FM(Y)F^{126}$
Solution Structure

Figure 4.4 Multiple conformations of the GTS loop in different structures. The GTS loop region (residues N30–S33) observed in the crystal structure of wt N-domain (A), the solution structure of wt N-domain (B), and the solution structure of $Q^{122}FM(Y)F^{126}$ (C). Positions of the residues that form the

hydrophobic core above the GTS loop are shown by their side-chains only colored by yellow (A), green (B), and blue (C) . Hydrogen bonding between the hydroxyl group of S70 and the carbonyl oxygen of S33 in $Q^{122}FM(Y)F^{126}$ is denoted by a dashed line. The distance between the hydrogen donor and the acceptor is 2.67 ± 0.09 Å. “s.c.” means side-chain of N30.

Close inspection reveals that side-chains of the individual residues have different solvent exposure in three structures, hinting at the possibility that alternative GTS loop conformations expose different functional chemical groups for interactions with the stop codons. Functional implication of the GTS loop in stop codon recognition had been reported, since T32A and T32V eRF1 mutants showed tendency towards UGA unipotency (Cheng et al., 2009). Furthermore, two individual point mutations in the GTS loop of eRF1 (*i.e.* T32A and S33A) were found to exhibit opposite effects on their release activity measured *in vitro* using fully reconstituted eukaryotic translation system, namely 32% UAA/30% UAG/75% UGA and 100% UAA/90% UAG/63% UGA, respectively (personal communication with Dr. Elena Alkalaeva). Those results suggest that T32 is not required for decoding UGA. Coincidentally, the side-chain of T32 in the structure of $Q^{122}FM(Y)F^{126}$ was found hidden from the solvent (Figure 4.4, C).

We have also found that in $Q^{122}FM(Y)F^{126}$, but not in wt N-domain, the resonance stemming from the hydroxyl proton of S70 can be observed (Figure 4.3, D), protected by potential hydrogen bonding to the carbonyl oxygen of S33 (Figure 4.4, C). Hence, the structure of $Q^{122}FM(Y)F^{126}$ indicates that S70 might involve in stabilization of the GTS loop in its alternative conformation.

4.5 Pico- to nanosecond dynamics of wt N-domain and Q^{122}FM(Y)F^{126}

The backbone ^{15}N relaxation rates of wt N-domain and Q^{122}FM(Y)F^{126} have shown that the GTS loop is relatively more dynamic on the sub-nanosecond timescale than the bulk of N-domain (Figure 4.5). It has been suggested that the ligand-binding sites are often found at or close to the flexible regions of proteins (Ishima and Torchia, 2000, Jarymowycz and Stone, 2006). The significantly different NOEs- and RDCs-derived average structures of the GTS loop in wt N-domain and Q^{122}FM(Y)F^{126} has already implied a rather complex equilibrium between conformations conferring different functionality. Being a dynamic "hotspot", we further suggest that the observed dynamics of the GTS loop may provide the necessary flexibility to switch between differently functionalized conformations upon interaction with the stop codon.

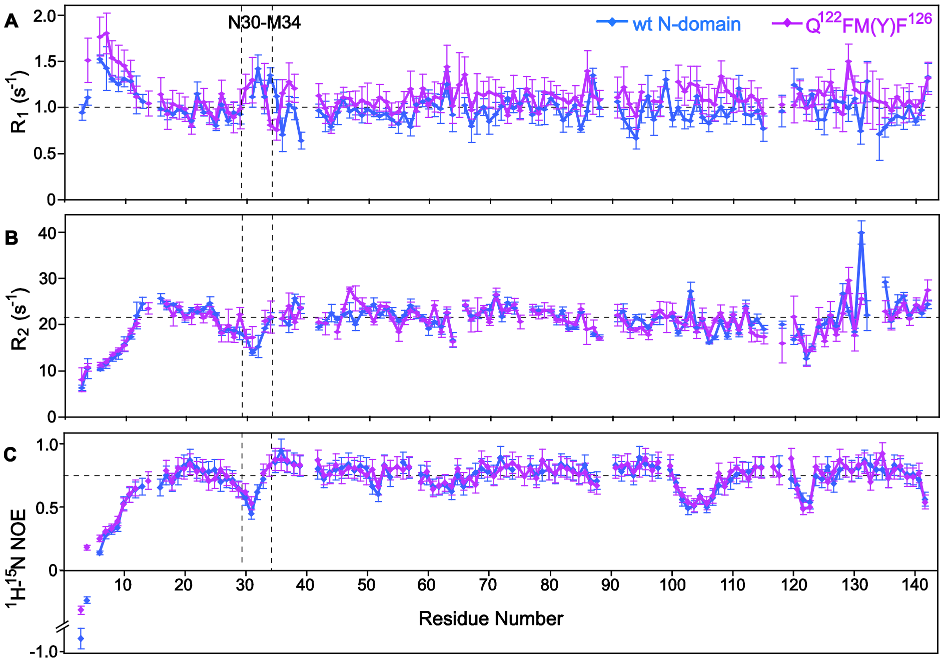


Figure 4.5 The dynamic properties of wt N-domain and $Q^{122}FM(Y)F^{126}$. Plots of the longitudinal relaxation rate R_1 (A), the transverse relaxation rate R_2 (B), and the heteronuclear $^1H,^{15}N$-steady-state NOE values (C) of the amide ^{15}N-nuclei of wt N-domain and $Q^{122}FM(Y)F^{126}$ measured at 25 °C. The GTS loop region (*i.e.* N30–M34) is highlighted. The standard error is indicated by the error bars, and the average values of the respective relaxation parameters (residues 16–142) are indicated by the horizontal dashed lines.

Apparently, the dynamic properties of wt N-domain and $Q^{122}FM(Y)F^{126}$ do not differ significantly from each other (Figure 4.5). Hence, this led us to conclude that although the switching between omnipotency and unipotency of eRF1 can be sufficiently explained by alteration of the GTS loop conformation, it is not reflected by the pico- to nanosecond dynamics. An elaborate investigation of the dynamics on other timescales might be useful to gain further insight.

4.6 C127 mutants of eRF1 exhibiting omni-, bi-, and unipotent specificity

Although the distinct GTS loop conformation in $Q^{122}FM(Y)F^{126}$ compared to wt N-domain implicates a functional role of the GTS loop in stop codon recognition, one may still ponder about the possibility that the substituted residues in the mutant abolishes their direct interactions with either UAA or UAG stop codon. To prove that the region of β-strand β4 of eRF1 (positions 122-128) has an indirect implication in stop codon decoding, the RF activity of two variant-code (*i.e. Euplotes* and *Stylonychia*) eRF1s with C127 mutations has been determined in an *in vitro* RF assay (courtesy of Prof. Ludmila Frolova). C127 is an invariant residue in the family of eRF1s, and is located neighbor to F126 in $\mathrm{Q}^{122}\mathrm{FM(Y)F}^{126}$ that possesses the same stop codon specificity as *Stylonychia* eRF1 (Figure 4.7). The sequence homology between human eRF1 and eRF1 of *Euplotes* and *Stylonychia* are 72.8 and 70.1%, respectively (Figure 4.6, A). Hence, their N-domains are likely to have very similar folds, as structural models of *Euplotes* and *Stylonychia*'s N-domain derived by homology modeling possess a TM-score of 0.92 and 0.94, respectively (Zhang and Skolnick, 2004, Zhang and Skolnick, 2005). A TM-score falls within the range of (0,1] and value of 1 represents the highest similarity.

A

Sequences of eRF1s compared	**Identity (%)**	**Identity + similarity (%)**
Homo sapiens / Euplotes aediculatus	59.6	72.8
Homo sapiens / Stylonychia mytilis	55.3	70.1
Euplotes aediculatus / Stylonychia mytilis	56.7	72.0

B

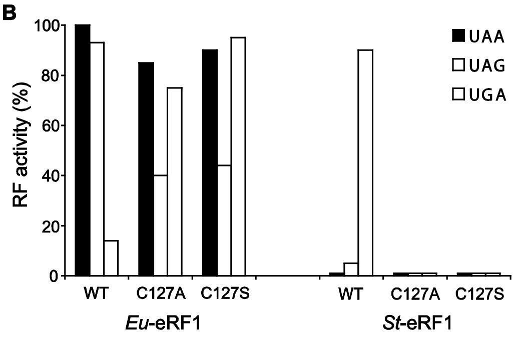


Figure 4.6 RF activities of C127 mutants of eRF1 with omni-, bi-, and unipotent specificity (courtesy of Prof. Ludmila Frolova). (A) Percent identity and similarity between human, *Euplotes*, and *Stylonychia* eRF1 obtained from pairwise sequence alignment of their amino acid sequences. (B) *In vitro* RF activity (Caskey et al., 1974) of chimeric eRF1 constructs containing the N-domain (positions 1–144) of *Euplotes* eRF1 (*i.e.* WT *Eu*-eRF1) or *Stylonychia* eRF1 (*i.e.* WT *St*-eRF1) as well as the C127A or C127S mutant of both *Eu*-eRF1 and *St*-eRF1. All eRF1 constructs contain MC-domain of human eRF1 (positions 145–437).

Human eRF1 is omnipotent, while *Euplotes* eRF1 decodes UAA and UAG stop codons only, and *Stylonychia* eRF1 is UGA-unipotent. It was shown earlier that C127A and C127S mutants of human eRF1 exhibited tendency towards UGA unipotency (Seit-Nebi et al., 2002), while C127S mutant of *Euplotes* eRF1 restored efficient recognition of UGA stop codon without changing of UAA and UAG stop codon decoding in the readthrough RF assay (Fan-Minogue et al., 2008). We have shown that C127A and C127S mutants of *Euplotes* eRF1 also restored recognition of UGA stop codon but RF activity towards UAG was reduced for both mutants (Figure 4.6, B). However, insertion of the same C127A and C127S mutations into *Stylonychia* eRF1 with UGA-only specificity caused total abolishment of RF activity towards UGA (Figure 4.6, B).

The different effects of the same C127 mutations on the recognition of UGA by human, *Euplotes* and *Stylonychia* eRF1s are a very hard argument in favor of the suggestion that C127 does not participate directly in UGA recognition.

4.7 Lack of interaction between N-domain and pentaribonucleotides containing stop codon

We have tried to titrate separately two pentaribonucleotides, *i.e.* 5'-AUAAA-3' and 5'-UGAAA-3', into wt N-domain up to five-fold molar excess. The absence of chemical shift perturbation on wt N-domain indicates lack of detectable interaction between them. Similarly, it was reported that interaction between stop codon-containing oligonucleotides and eRF1 was observed only in the presence of ribosome (Chavatte et al., 2003). In contrast to being free in solution, oligonucleotides may have significantly lower entropy at ribosomal A site, thereby favoring interaction with N-domain within the ribosome. On the other hand, in the context of conformational selections for binding, interactions between N-domain of eRF1 and the ribosomal A site may populate certain conformations of N-domain competent for binding to stop codons. Since N-domain has an overall stable tertiary structure, it is more probable for the flexible regions, *e.g.* the GTS loop, the loop connecting β-strands β2 and β3 (V101–E107), and the loop leading to β4 (I120–T122) (Figure 4.5), to sample the conformational space below certain energy bound. This scenario agrees with our argument that multiple conformations of the GTS loop modulate selectivity of stop codon recognition (4.8 Discussion).

4.8 Discussion

The strictly conserved GTS loop of N-domain has been implicated in decoding or in direct contact with the stop codon (Bulygin et al., 2010, Wang et al., 2010, Bulygin et al., 2011). The most significant insight from the finding of distinct GTS loop conformations in wt N-domain and $Q^{122}FM(Y)F^{126}$ is that the shift in stop codon selectivity might be determined by the structural changes that are critical for interactions of amino acid(s) with the stop codon. Having the same amino acid sequence as the wild-type at the positions 122-126, eRF1 with a single point mutation T32A had been shown to exhibit tendency towards UGA unipotency ((Cheng et al., 2009), personal communication with Dr. Elena Alkalaeva). Why does T32A mutant show similar attribute as $Q^{122}FM(Y)F^{126}$? The same question can be asked about the stop codon decoding by various eRF1 mutants investigated in previous studies, since in many cases the point mutations scattered across a large part of N-domain resulted in the same bias of stop codon selectivity. This paradox can be explained in two ways: (i) eRF1 mutant, which has lost interaction with one or two out of three nucleotides of a stop codon, can still support the peptide release for that particular stop codon but with reduced efficiency, probably because, one amino acid substitution is not enough to destroy completely the direct interaction with stop codon(s) (this might be the case for T32A mutant), and (ii) the part of described earlier point mutations, which have given rise to altered stop codon specificity, are responsible for structural modulation of N-domain regions that actually interact with the stop codon(s) (this might be the case for $Q^{122}FM(Y)F^{126}$).

Similar mutations of the highly conserved C127 residue in human, *Euplotes* and *Stylonychia* eRF1s evoke different responses: the decreased recognition for UAA and UAG for human eRF1, the appearance of UGA recognition for *Euplotes* eRF1, and its disappearance for *Stylonychia* eRF1 (Figure 4.6, B). This indicates that C127 does not participate directly in stop codon recognition as well as very likely the same for the other residues in the region of residues 122-126. It also shows that residues nearby to the stop codon recognition site can critically affect stop codon selectivity.

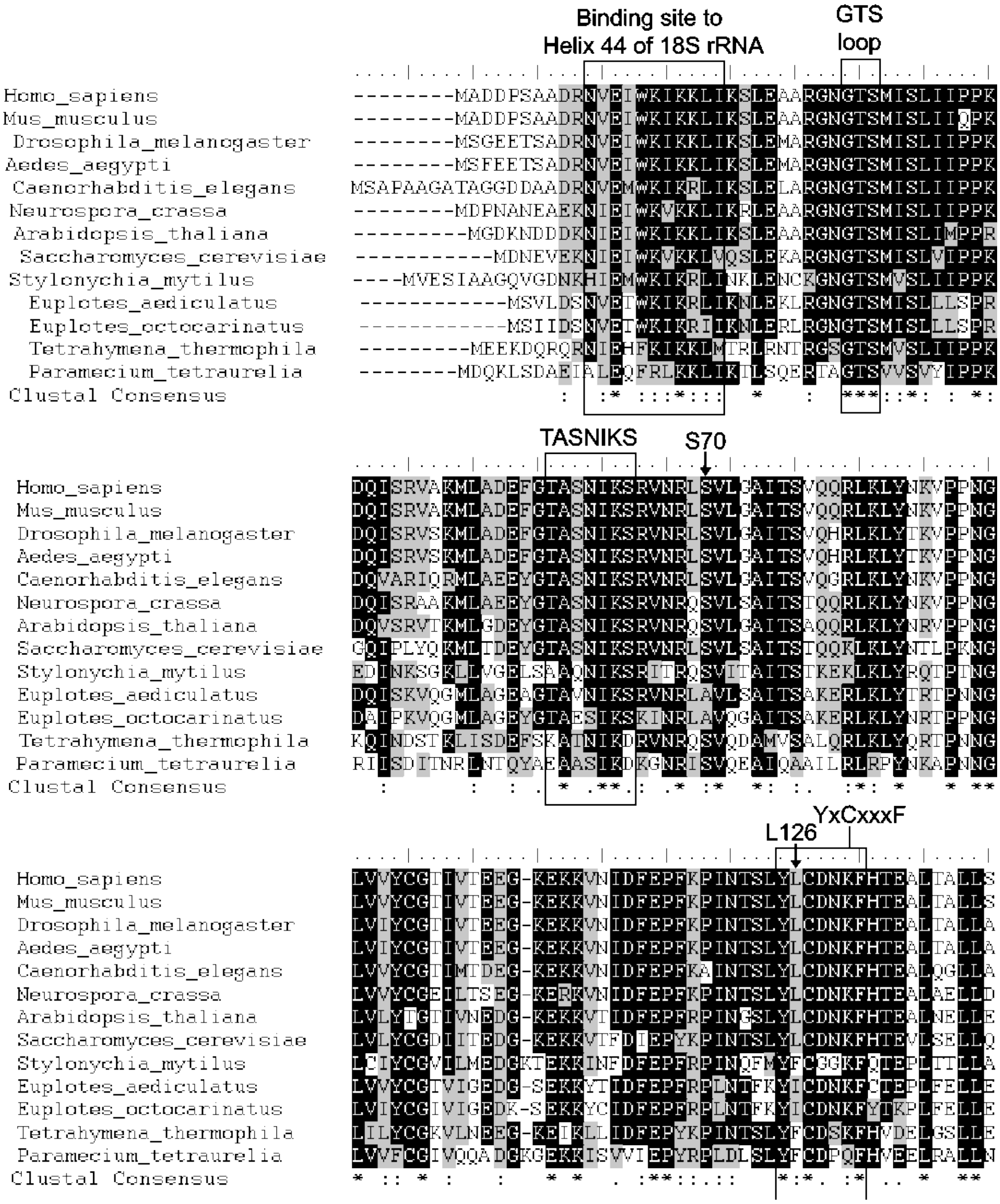

Binding site to Helix 44 of 18S rRNA
GTS loop
TASNIKS
S70
L126
YxCxxxF
Homo_sapiens
Mus_musculus
Drosophila_melanogaster
Aedes_aegypti
Caenorhabditis_elegans
Neurospora_crassa
Arabidopsis_thaliana
Saccharomyces_cerevisiae
Stylonychia_mytilus
Euplotes_aediculatus
Euplotes_octocarinatus
Tetrahymena_thermophila
Paramecium_tetraurelia
Clustal Consensus

Figure 4.7 Sequence conservation of the GTS loop and the putative motif for binding to Helix 44 of 18S rRNA (see Chapter 5). Multiple sequence alignment of the amino acid sequences of eRF1 N-domain from different eukaryotic organisms using ClustalW2 online server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Numbering of the residues follows that of human eRF1. The GTS loop is strictly conserved. The binding site to Helix 44 of 18S rRNA indicates a conserved motif, *i.e.* $(W/F)_{15}(K/R)_{16}(I/V/L)_{17}(K)_{18}(K/R)_{19}(L/I)_{20}$, with amino acids of the same nature occupying at each position. Substitution S70A occurs in *E. aediculatus* and *E. octocarinatus*, which are variant-code organisms with UAA and UAG as

stop codons. Conversely, substitution L126F occurs in *S. mytilus*, *T. thermophila*, and *P. tetraurelia*, all of which are unipotent for UGA.

In the context of our proposed mechanism, point mutation that alters selectivity of stop codon recognition is likely to modulate structure of the GTS loop, or even its capacity to switch between different conformations. Indeed, several point mutations were found to hit on the residues that constitute the hydrophobic core right above the GTS loop, *e.g.* I35, V71, V78, and C127 (Bertram et al., 2000, Cheng et al., 2009, Fan-Minogue and Bedwell, 2008). Remarkably, the width of this hydrophobic core is directly related to the differential positioning of helix α3 as observed in the structures of wt N-domain and $Q^{122}FM(Y)F^{126}$ (Figures 4.1 and 4.2). The width, measured as the distance between the amides of M34 and V71, is reduced from 7.68 Å and 7.38 ± 0.18 Å in the crystal and solution structures of wt N-domain, respectively, to 6.56 ± 0.13 Å in $Q^{122}FM(Y)F^{126}$. In light of these observations, it is attractive to hypothesize that repositioning of helix α3 in N-domain could occur during stop codon recognition as the GTS loop samples between different configurations. Interestingly, S70 on helix α3 was found to be involved in stabilization of the GTS loop in $Q^{122}FM(Y)F^{126}$ (Figure 4.4, C). Residue S70 is critical for UGA-decoding, as a point mutation S70A restricts human eRF1 to recognize UAA and UAG only. Conversely, the A70S substitution in *Euplotes* eRF1 changes its stop codon recognition from UAA and UAG bipotency to become an omnipotent one (Eliseev et al., 2010). These data verify the assumption that S70A substitution is associated with UGA reassignment (Liang et al., 2005).

Besides helix α3, helix α2 could also play a role in modulating the selectivity of stop codon recognition. First of all, M51 and E55 on helix α2 are able to alter stop codon recognition patterns (Bertram et al., 2000, Kolosov et al., 2005). Secondly, the TASNIKS motif was found to confer distinct requirement of eRF3 upon eRF1 on decoding UAA/UAG and UGA (Fan-Minogue et al., 2008). As T58 in the TASNIKS motif was observed to interact with the 15-mer RNA (see Chapter 5) and the NIKS motif had also been implicated in ribosome binding (Frolova et al., 2002), interactions between helix α2 and the ribosome is highly possible. Furthermore, P41 and P89, which could be critical for the formation of the β-turns that connect the core β-sheet to helices α2 and α3, were also found to affect stop codon recognition (Bertram et al., 2000, Velichutina et al., 2001). This hints at a higher than anticipated degree of complexity in the stop codon decoding mechanism of eRF1.

Chapter 5: Interactions between N-domain and Mimics of the Decoding Region of Ribosomal Helix 44

5.1 Abstract

Both N-domain and M-domain of eRF1 possess putative ribosomal binding sites (Kisselev et al., 2003). The genetic interactions between eRF1 and the decoding region of Helix 44 (H44) of 18S rRNA in the 40S subunit were reported (Velichutina et al., 2001). In the attempt to investigate whether direct interaction between N-domain and the decoding region of H44 is possible, we utilized a 15-mer RNA oligonucleotide that contains internal loops, mimicking the decoding region of H44 (Fourmy et al., 1996, Lynch and Puglisi, 2001b, Lynch et al., 2003, Kondo et al., 2006). By generating different mutants of the 15-mer RNA, we obtained constructs with different sizes of internal loop, and found that presence of a 2-nt internal loop is critical for strong binding to N-domain. These results indicate that helix α1 of N-domain potentially interacts with H44 at A site, and lysine residues K18 and K19 contribute significantly to the binding. Interestingly, the RNA-binding region is shielded partially by C-domain in solution, similar to what is observed from the crystal structure of full-length eRF1 (Song et al., 2000). In solution, the 15-mer RNA is able to displace C-domain from the non-covalent complex of N-domain and C-domain, suggesting an imperative domain rearrangement in eRF1 during which N-domain accommodates itself into the ribosomal A site.

5.2 Design of 15-mer RNA mimicking the decoding region of H44

We designed a series of double-stranded RNA oligonucleotides, aiming to mimic the decoding region of H44 in 18S rRNA (Figure 5.1). The 15-mer RNA constructs are C*i*- (*i.e.* inversion symmetry) duplexes containing either two 2-nt asymmetric internal loops at the symmetrical positions or a single large internal loop (Figure 5.1, B). It appears that the symmetric structures provide sufficient thermodynamic stability to accommodate a significant number of mismatches in the central region, enabling studies of the impact of variability in the internal loop length found in different organisms (*e.g. Tetrahymena thermophila* and *Thermomyces lanuginosus* in Figure 5.1, A) on binding to N-domain. The 7-mer RNA, which does not contain any internal loop, was used as a reference for the RNA binding study. In addition, in some constructs we incorporated 2'-*O*-methyl at C_5 position exhibiting an easily detectable ^{1}H NMR line (Clore et al., 1984) as a structural probe for interactions near the internal loop. This modification has been shown to cause negligible effect on the structure of double-stranded RNA (Popenda et al., 1997).

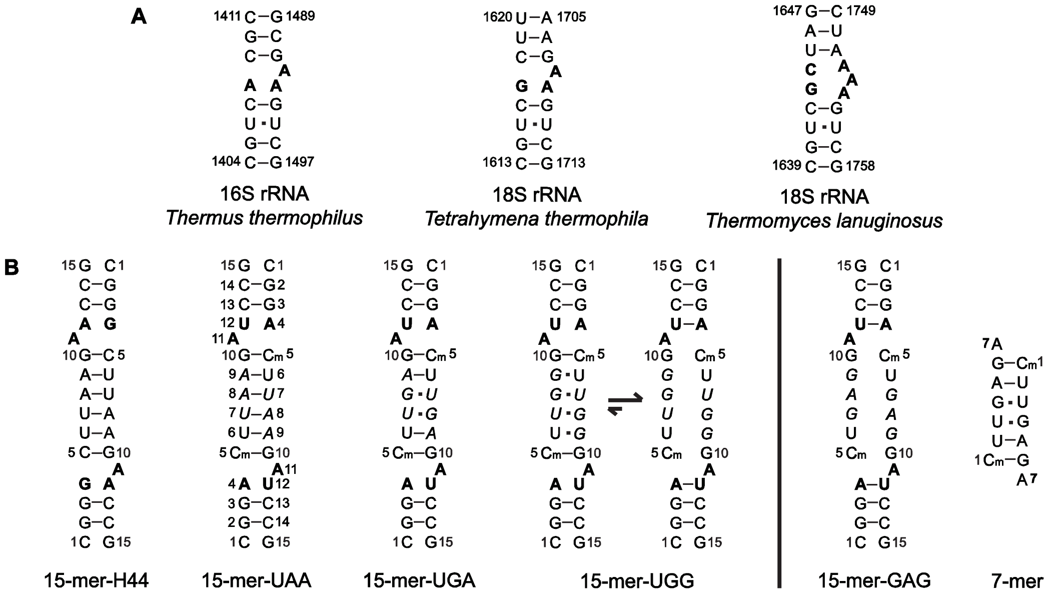

A
16S rRNA
Thermus thermophilus
18S rRNA
Tetrahymena thermophila
18S rRNA
Thermomyces lanuginosus
B
15-mer-H44
15-mer-UAA
15-mer-UGA
15-mer-UGG
15-mer-GAG
7-mer

Figure 5.1 15-mer RNAs mimicking the decoding region of H44. (A) Secondary structures of the highly conserved decoding region of H44 in 16S and 18S rRNAs. The internal loops are highlighted in bold; Watson-Crick and non-Watson-Crick base pairs are denoted by "–" and "•", respectively. (B)

Secondary structures of the RNA constructs used in our studies as characterized by NMR spectroscopy in solution. 15-mer-H44 contains two internal loops with the same sequence as the one in *Tetrahymena thermophila* 18S rRNA shown in (A). Four other RNA constructs are named according to their respective nucleotide substitutions in the central stem, which are denoted in *italics*. The 7-mer RNA does not contain any internal loop.

5.3 NMR structural characterization of the 15-mer RNAs

Chemical shift assignment of 15-mer-UAA was achieved using ^{15}N-SOFAST-HMQC, ^{13}C-HSQC, DQF-COSY, and NOESY spectra (Figure 5.2). Secondary structures of the 15-mer RNA constructs (Figure 5.1, B) were determined based on their respective NOESY spectra, as well as the assigned chemical shifts of 15-mer-UAA. All RNA constructs with nucleotide substitutions in the central stem adopt a double-stranded helical structure (Figure 5.2, A). Structural stability of the central stem can be inferred from the ^{31}P spectra as well, since resonance peaks from the backbone phosphate became broadened and less dispersed progressively from 15-mer-UAA to 15-mer-GAG. The canonical double-stranded helical structure can be clearly seen from the presence of three cross-peaks stemming from G_2's, G_3's, and G_{10}'s imino protons in the H1'-H5 region (5.0-6.2 ppm; right panel in Figure 5.2, E) (Heus and Pardi, 1991).

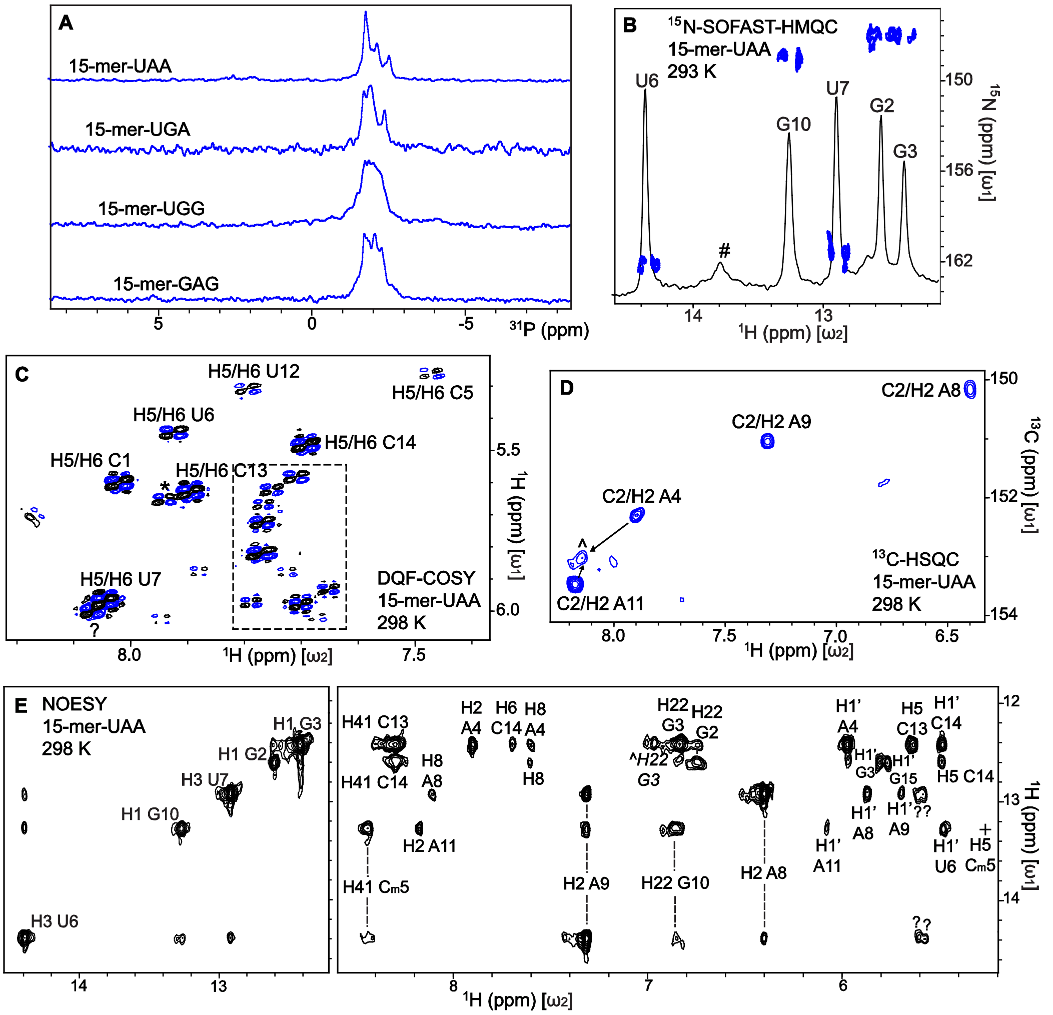


Figure 5.2 NMR spectral features and assignments of the 15-mer RNAs. (A) 1D ^{31}P spectra with proton decoupling during acquisition show that all mutant 15-mer RNAs have canonical double-stranded helical structure, rather than forming a single-stranded stem-loop structure (Legault and Pardi, 1994). Generally, the backbone phosphates within the loop of a RNA stem-loop structure exhibit chemical shifts that are as much as 2.5 ppm downfield from the collection of overlapping resonances belonging to the RNA stem part, which is obviously absent here. The ^{31}P chemical shift was not referenced. (B) The imino resonances of 15-mer-UAA can be easily assigned to either uridine or guanosine based on their ^{15}N chemical shifts (Furtig et al., 2003). The minor peak denoted by "#" only appeared at 20 °C. (C) The H5/H6 cross-peaks from the pyrimidines of 15-mer-UAA as observed in the DQF-COSY spectrum. Cross-peaks within the dashed box do not belong to the H5/H6, as NOE was not found at the corresponding positions in the NOESY spectrum. The minor

peak denoted by asterisk belongs to H5/H6 of C_{13} in another conformation when base pair A_4–U_{12} is closed (*e.g.* in 15-mer-UGG and 15-mer-GAG), demonstrating the dynamic behavior of the internal loop. (D) All expected C2/H2 cross-peaks from the adenines of 15-mer-UAA were observed in the ^{13}C-HSQC spectrum measured with the carbon carrier at 144.8 ppm. Although the minor peak denoted by caret could not be assigned, it is possible to have arisen from either A_{11} or A_4. Generally, C2/H2 of adenines in Watson-Crick base pairs (*i.e.* A_8 and A_9) exhibit chemical shifts that are relatively upfield to adenines in other situations. (E) Part of the NOESY spectrum of 15-mer-UAA in H_2O measured at 25 °C with a mixing time of 350 ms, of which cross-peaks (in the right panel) to the imino protons (in the left panel) are labeled according to chemical shifts on the ω_2 axis. Vertical dashed lines connect cross-peaks from the same protons. H22 of G_3 in the minor conformation is denoted by caret, while H5 of Cm_5 (denoted by "+") was attenuated by water suppression, which is otherwise observable on the other quadrant of the spectrum.

In 15-mer-UAA, five Watson-Crick base pairs were detected using imino proton resonances protected from exchange with solvent (left panel in Figure 5.2, E). Watson-Crick base pairing between A_4 and U_{12} was not observed. Conversely, a minor conformation associated with the closing base pair of A_4–U_{12} in 15-mer-UGG and 15-mer-GAG was observed, indicating a slow 'breathing' action of the internal loop on the ^{1}H chemical shift timescale (Figure 5.2, C, D, and E). This type of intramolecular dynamics is expected in RNA containing internal loops (Popenda et al., 2008). The secondary structure of 15-mer-UGA is similar to 15-mer-UAA. The non-Watson-Crick base pair of U7–G8 is confirmed by detection of the corresponding imino proton resonances (Figure 5.6, A) as well as strong NOEs between the two protons. For 15-mer-UGG and 15-mer-GAG, while an overall helical structure was maintained

(Figure 5.2, A), both of the RNA constructs are deemed to form a large internal loop, due to the lack of observable stable base pairing from the nucleotides in the central region (Figure 5.6, A). The 7-mer construct was expected to contain no internal loop, and this was confirmed by the observation of all expected base pairings (lower panel in Figure 5.6, D).

5.4 Paromomycin binds to the internal loop

Aminoglycosides are a class of antibiotics that bind to the decoding region of H44 in 16S rRNA (Moazed and Noller, 1987). In particular, paromomycin binds much stronger to the prokaryotic than to the eukaryotic ribosome, owing to several differences in the rRNA sequence of the internal loop (Velichutina et al., 2001, Fan-Minogue and Bedwell, 2008, Recht et al., 1999). Despite the fact that we had introduced several different nucleotide substitutions into the 15-mer RNA constructs, paromomycin still binds to the internal loop as expected (Figure 5.3). This indicates that the 15-mer RNA constructs (at least for 15-mer-UAA and 15-mer-UGA) possess some essential structural determinants sufficient for paromomycin binding, hence clearly supports the structural relevance of the 15-mer RNA constructs as mimics of the decoding region of H44.

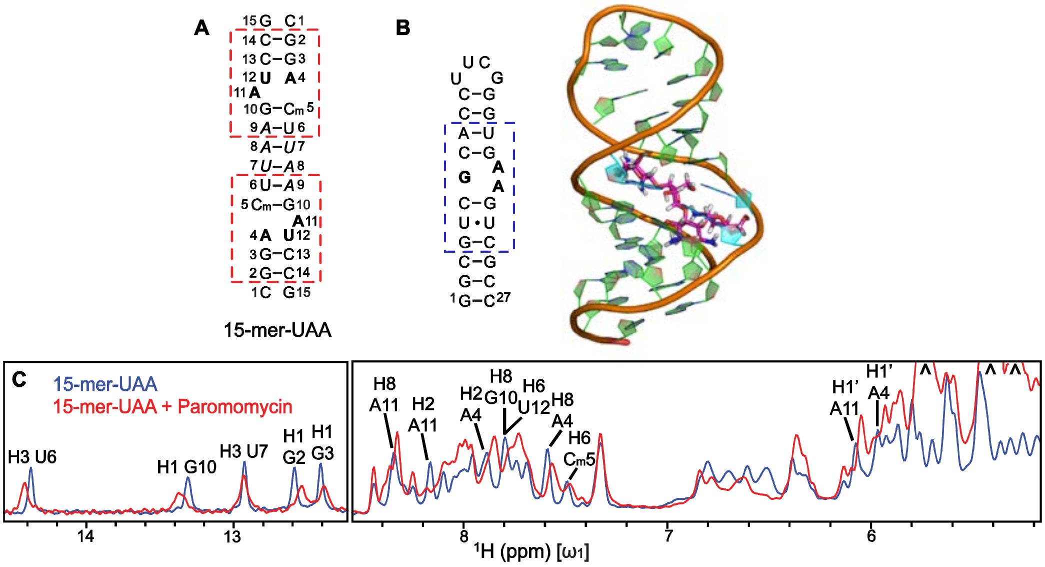

A
15G C1
14C–G2
13C–G3
12U A4
11A
10G–Cm5
9A–U6
8A–U7
7U–A8
6U–A9
5Cm–G10
A11
4A U12
3G–C13
2G–C14
1C G15
15-mer-UAA
B
U C
U G
C–G
C–G
A–U
C–G
A
G
A
C–G
U•U
G–C
C–G
G–C
1G–C27
C
15-mer-UAA
15-mer-UAA + Paromomycin
H3 U6
H1 G10
H3 U7
H1 G2
H1 G3
H8 A11
H2 A11
H2 A4
H8 G10
H6 U12
H8 A4
H6 Cm5
H1' A11
H1' A4
^
^
^
14
13
8
7
6
^{1}H (ppm) [ω1]

Figure 5.3 Paromomycin binds to the internal loop. (A) Paromomycin binds to the region of 15-mer-UAA designated by red dashed box as demonstrated in (C). (B) Region of the 27-mer oligoribonucleotides (Lynch and Puglisi, 2001a) designated by blue dashed box are in close contact with the bound paromomycin, as shown by its complex structure (PDB ID: 1FYP). (C) 1D proton spectra of 15-mer-UAA in H_2O before (blue) and after addition of paromomycin (red). Three peaks stemming from paromomycin are indicated by caret.

5.5 Helix α1 of N-domain interacts with the 15-mer RNAs

Titration of the 15-mer RNAs into N-domain showed a very specific binding interaction in the regime of fast exchange on chemical shift timescale (Figure 5.4). Among the 15-mer RNA constructs, the strongest binding 15-mer-UGA binds to N-domain with K_d of ~80 μM as estimated from NMR titration data (Figure 5.4, C). The binding interface spans from residues N11 to I21, covering most of helix α1 (Figure 5.4, B). In particular, several lysine residues within the binding region (*i.e.* K16, K18, and K19) are highly conserved and are likely to contribute to the interactions with RNA (see Figure 4.7 in Chapter 4). In addition, the aromatic side-chain of the conserved W15 could also be involved in a stacking interaction with the nucleic acid bases. These types of interactions are commonly found non-sequence-specific interactions in protein-RNA complexes (Morozova et al., 2006, Allers and Shamoo, 2001).

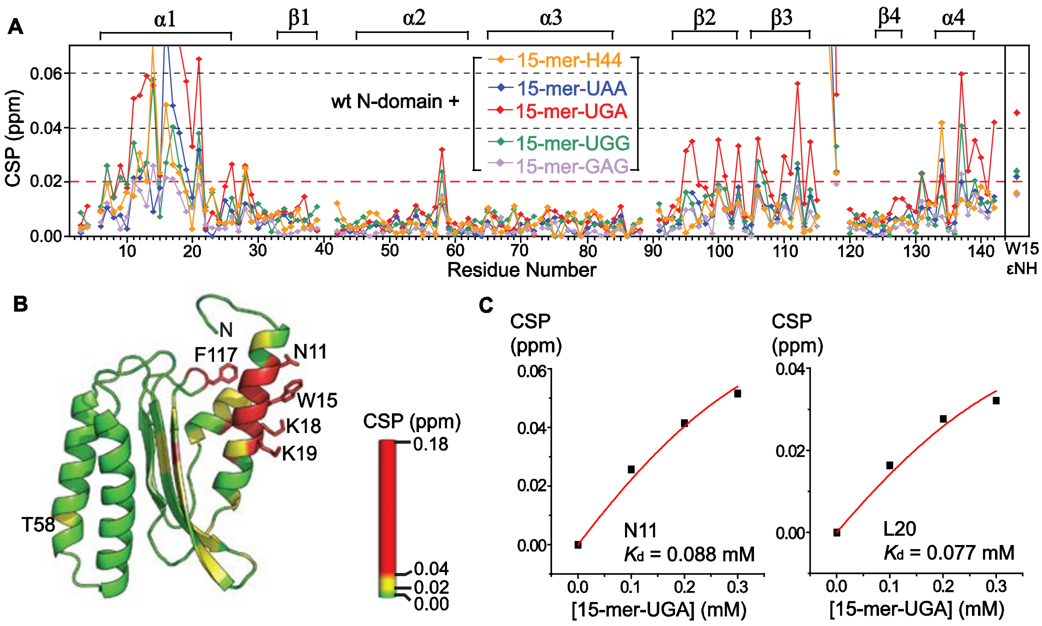


Figure 5.4 Helix α1 of N-domain interacts with the 15-mer RNAs. (A) Chemical shift perturbations (CSP) of the backbone amide resonances of ^{15}N-labeled N-domain upon titration with the 15-mer RNAs up to the RNA to protein molar ratio (R/P) as shown in Figure 5.6, A. The CSP values were obtained from a series of TROSY [^{1}H,^{15}N]-HSQC spectra by $\text{CSP} = \sqrt{\{(\Delta\delta\text{H})^2 + (0.14\times\Delta\delta\text{N})^2\}}$. The secondary structure of N-domain is indicated on top of the plots. (B) The CSP values for wt N-domain + 15-mer-UGA as shown in (A) are mapped onto the structure of wt N-domain according to the color scale. Side-chains of the significantly perturbed residues upon RNA binding are shown and labeled. (C) Binding isotherms derived from the CSP on N11 and L20 upon titration with 15-mer-UGA. The binding affinities were obtained by fitting the data points to the equation, $\text{CSP} = \text{CSP}_{\text{max}} / (2[\text{P}]) \times ([\text{L}] + [\text{P}] + K_{\text{d}} - \sqrt{\{([\text{L}] + [\text{P}] + K_{\text{d}})^2 - 4[\text{L}][\text{P}]\}})$, where [L] and [P] are total concentrations of ligand and protein, respectively.

Besides helix α1 some other residues of N-domain were also affected by RNA binding as detected by the chemical shift perturbations (CSP) (Figure 5.4, A and B). These observations can be explained by allosteric propagation of the structural perturbations at helix α1 directly involved in RNA binding to its immediate spatial proximity. Such an effect was manifested as a drastic CSP at F117, of which side-chain is in close contact with helix α1. Some residues with notable CSPs are located in helix α4, especially at and around T137, which is tightly packed against helix α1 with the side-chain of T137 facing I21 of helix α1. Another site with rather subtle CSP is located on the opposite side of helix α1 at T58 (Figure 5.4, A and B). Interestingly, in the crystal structure of eRF1 in complex with domain 2/3 of eRF3, an ATP molecule was found in proximity to T58 of eRF1, with N7 of adenine in contact with the threonine hydroxyl group (PDB ID: 3E1Y) (Cheng et al., 2009). This indicates a possibility of interactions between T58 and the flipped out adenine in the internal loop of the 15-mer RNAs.

We have generated a mutant of N-domain (*i.e.* K18QK19Q) to test the contribution of the lysine residues in its binding to 15-mer RNAs (Figure 5.5). Glutamine was chosen as the substitution for lysine since it has a long polar side-chain without the positive charge. As expected, the double point mutations have caused negligible structural changes to N-domain (Figure 5.5, A). Conversely, binding affinity of the mutant N-domain to 15-mer-UGA has been reduced significantly compared to wt N-domain (Figure 5.5, B, C, and D).

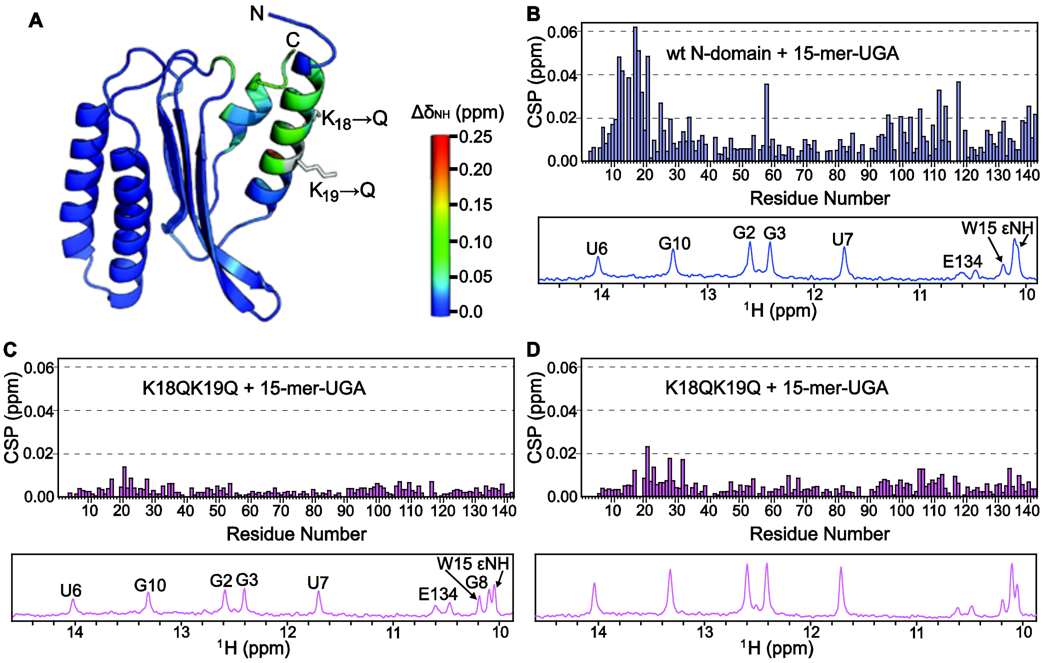


Figure 5.5 K18 and K19 are critical for binding to the 15-mer RNA. (A) Difference in amide chemical shift ($\Delta\delta_{NH} = \sqrt{\{(\Delta\delta H)^2 + (0.14\times\Delta\delta N)^2\}}$) between wt N-domain and K18QK19Q are mapped onto the structure of wt N-domain according to the color scale. (B) CSP of the backbone amide resonances of ^{15}N-labeled N-domain upon addition of equimolar amount of 15-mer-UGA. (C) and (D) CSPs of the backbone amide resonances of ^{15}N-labeled K18QK19Q upon addition of equimolar and two-fold molar excess of 15-mer-UGA, respectively.

5.6 Role of the internal loop in its interaction with N-domain

Upon binding to N-domain, all tested 15-mer RNA constructs induced specific and consistent CSP profiles, albeit with variable magnitudes reflecting different binding affinities (Figure 5.4, A). Specifically, 15-mer-GAG exhibits significantly weaker response, which is likely due to its thermodynamically unstable central stem. This implies that binding of the 15-mer RNA to N-domain does not depend strongly on the sequence of the internal loop, but rather on the structural variability of the double-stranded RNA.

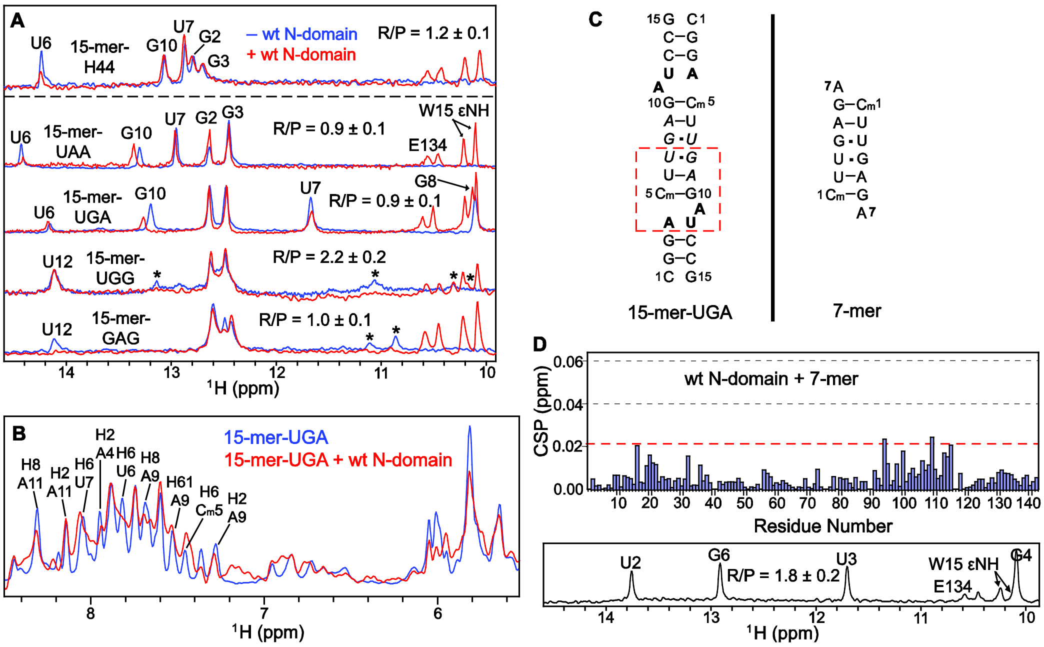


Figure 5.6 Role of the internal loop in its interaction with N-domain. (A) ^{1}H spectra showing the imino protons of the 15-mer RNAs in free (blue) and N-domain-bound form (red). The ^{15}N-split resonances stemming from ^{1}H^{N} E134 and ^{1}H$^{\varepsilon N}$ W15 of N-domain are indicated along with assignment of the imino protons of the RNAs as well as the RNA to protein molar ratio (R/P). Transient base pairing between nucleotides in the central stem of 15-mer-UGG and 15-mer-GAG is observed as the minor peaks denoted by asterisk. (B) ^{1}H spectra showing only resonances of 15-mer-UGA in free (blue) and N-domain-bound form (red) by selective suppression of amide and aromatic proton resonances stemming from the protein. (C) N-domain binds to the region of 15-mer-UGA designated by red dashed box as demonstrated in (A) and (B). (D) CSP of the backbone amide resonances of ^{15}N-labeled N-domain upon addition of about two-fold molar excess of 7-mer RNA.

We further investigated the region of the 15-mer RNA that binds to N-domain (Figure 5.6, A and B). To highlight the role of the internal loop in binding to N-domain, we tested the interaction using the 7-mer RNA, which comprises the central region of 15-mer-UGA but without the internal loops (Figure 5.6, C and D). The significantly attenuated binding affinity, in comparison to 15-mer-UGA, shows that the internal loop is required for specific binding to helix α1 of N-domain. Furthermore, only a specific set of resonances of the 15-mer RNA was perturbed upon interaction with N-domain (Figure 5.6, A and B). Based on that, the binding site can be mapped onto a region of 15-mer-UGA that includes the internal loop (Figure 5.6, C). Close inspection revealed that nucleotides A_9 and G_{10} seemed to experience a larger perturbation compared to other relevant nucleotides. We argue that if the central RNA stem was the only element required for the interaction, binding affinity of the 7-mer RNA to N-domain should be comparable to the 15-mer RNA, particularly 15-mer-UGA. Unfortunately, there was no observable inter-molecular NOE between N-domain and the 15-mer RNA in isotope-edited NOESY spectrum, possibly due to the micromolar range of binding affinity, hence hampering further elucidation of the complex. Nevertheless, we have established that the interaction between N-domain and RNA requires certain structural elements that include double-stranded helices and a 2-nt internal loop. These results strongly support the potential of N-domain to interact directly with the decoding region of H44.

5.7 Interactions between N-domain and ribosome

Having established the interactions between N-domain and mimics of the decoding region of H44, we tested if N-domain binds similarly to ribosome (Figure 5.7). Surprisingly, adding either *E. coli* ribosome (courtesy of Dr. Feng Shu) or rabbit reticulocyte lysate (RRL) (*i.e.* containing tRNAs, ribosome, amino acids, initiation, elongation, and termination factors) into N-domain at a low ribosome-to-N-domain molar ratio induced the same CSP profiles, indicating a similar binding mode to both prokaryotic and eukaryotic ribosomes. In spite of the rather small magnitude of CSP, the shifts are clearly identifiable at the given spectral resolution, and the shift directions are also consistent for both *E. coli* ribosome and RRL (Figure 5.7, D and E).

In the current experiments of which the amount of ribosome added was far from saturation, the TROSY [^{1}H,^{15}N]-HSQC spectra displayed shifts of the resonance peaks, resembling binding event in the fast exchange regime. However, other researchers had reported instead attenuation of amide resonance intensities for the residues on the binding interface with ribosome in other protein-ribosome interactions (Sette et al., 1999, Feng et al., 2011). Since ribosome is a megadalton particle that tumbles on a timescale several orders of magnitude slower than a protein, the NMR spectral feature of their interactions remains unknown and requires further investigation using a novel theoretical framework.

Since ribosomal A site is relatively conserved among all organisms and we have shown that N-domain interacts with *E. coli* ribosome (despite the nature of interactions being inconclusive), we tested if N-domain binds to the ribosome of *Thermus thermophilus* programmed with stop codon by sucrose cushion centrifugation (courtesy of Prof. Gao Yonggui). However, no interaction was detected by the method. Although the putative ribosomal binding site on N-domain is specific and binds to H44 mimics with micromolar affinity, the overall interactions between eRF1 and the pre-termination complex (pre-TC) might be more extensive and possibly involve other parts of eRF1.

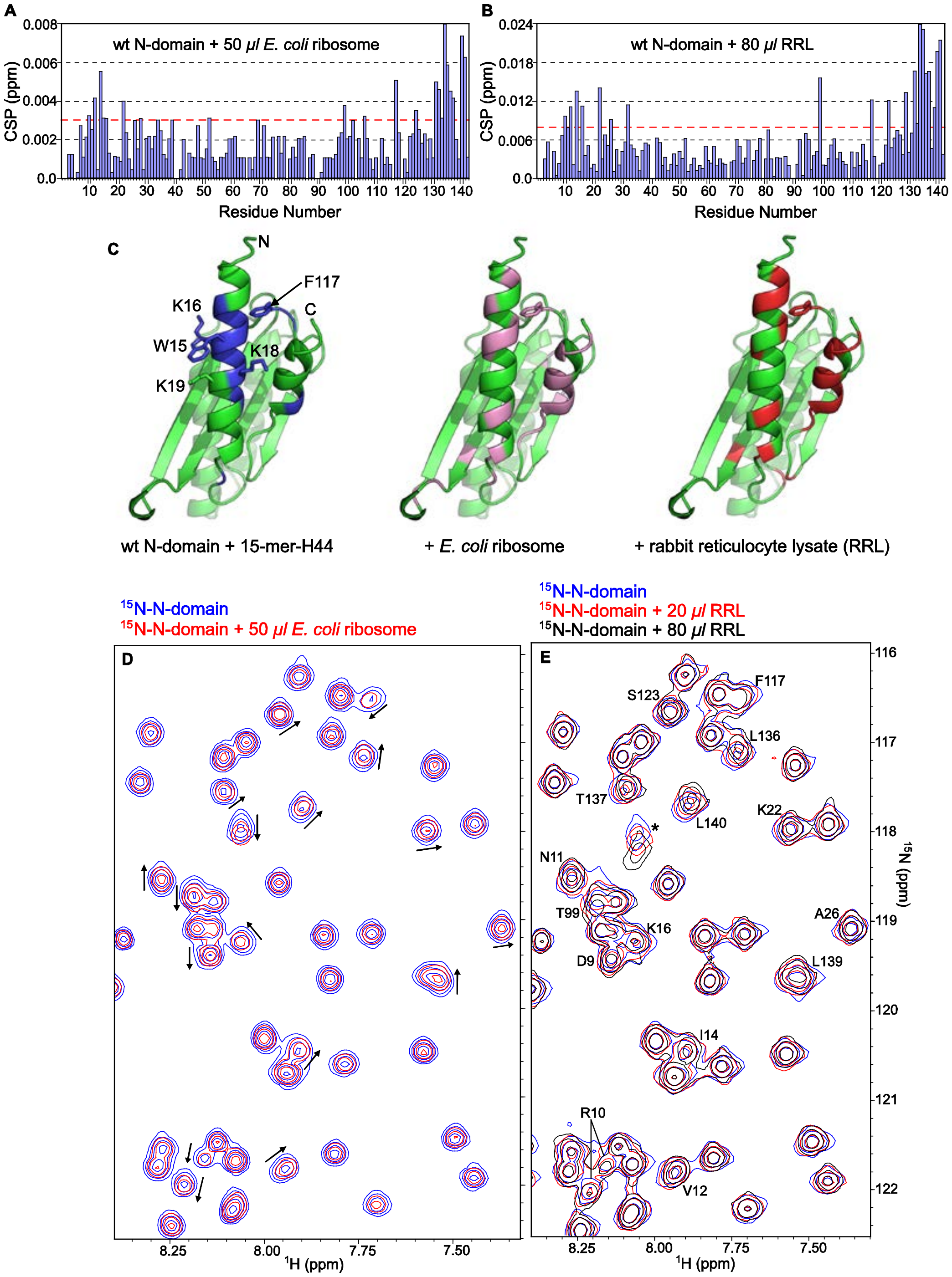


Figure 5.7 Interaction interfaces between N-domain and both prokaryotic and eukaryotic ribosomes. (A) and (B) CSPs of the backbone amide resonances of ^{15}N-labeled N-domain upon addition of 50 *μ*l and 80 *μ*l of *E. coli* ribosome and rabbit reticulocyte lysate (RRL), respectively. (C) Residues of N-domain that experienced the most CSPs upon addition of 15-mer-H44 (blue; Figure 5.4, A), *E. coli* ribosome (pink; Figure 5.7, A), and RRL (red; Figure 5.7,

B) are highlighted respectively. (D) and (E) Overlays of portion of the TROSY [^{1}H,^{15}N]-HSQC spectra of ^{15}N-labeled N-domain upon addition of *E. coli* ribosome and RRL, respectively. The arrows in (D) show the direction of peak shifts that are clearly identifiable from their 1D traces in the respective dimensions.

5.8 Interactions between N-domain and C-domain of eRF1

The 3D crystal structure of full-length eRF1 shows that its three protein domains extend outward from the center of mass to form a Y-shape tertiary structure, with C-domain exhibiting a considerable contact interface with N-domain (Song et al., 2000). In solution, TROSY NMR spectra of full-length eRF1 showed fewer than expected resonances, with the majority of cross-peaks broadened by conformational exchange (personal communication with Prof. Konstantin Pervushin). We traced the source of this conformational jitter to the interactions between N-domain and C-domain, since the NM-domain construct corresponding to a C-domain-truncated eRF1 lacks significant structural interactions between N-domain and M-domain (see Chapter 6). Interestingly, our model RNAs interact with helix α1 of N-domain that is found at the interface with C-domain in the crystal structure of eRF1 (Figure 5.8, C). We hypothesized that the 15-mer RNA and C-domain may compete for binding to N-domain, and hence, set out to test it by competitive binding experiments.

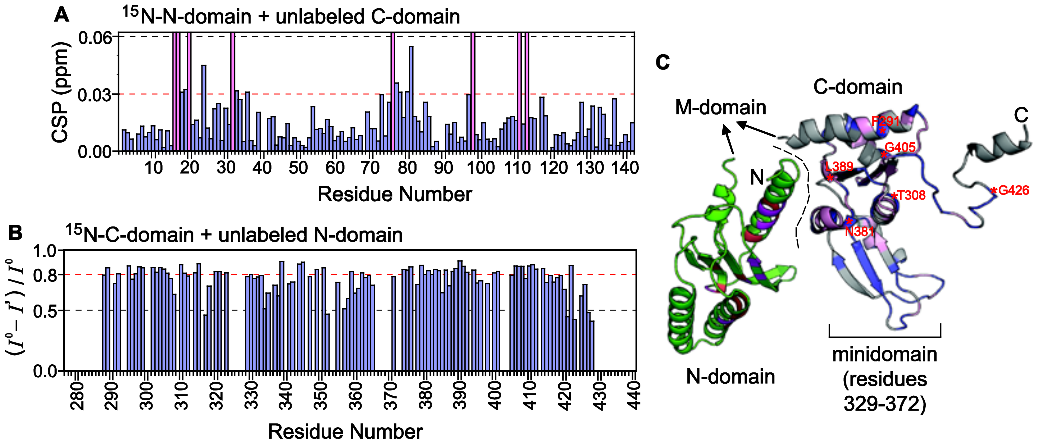


Figure 5.8 Complex formation between N-domain and C-domain. (A) CSP of the backbone amide resonances of ^{15}N-labeled N-domain upon addition of equimolar amount of unlabeled C-domain. The pink colour bars denote residues of which resonances are broadened beyond detection. (B) Reduction in signal intensity of the backbone amide resonances of ^{15}N-labeled C-domain upon addition of equimolar amount of unlabeled N-domain. (C) The complex model of N-domain and C-domain is reconstructed by replacing C-domain in the crystal structure of full-length eRF1 (PDB ID: 1DT9) with the NMR solution structure of the individual C-domain (PDB ID: 2KTV) (Mantsyzov et al., 2010), since the minidomain of C-domain has not been resolved in the crystal structure. The crystal structure of eRF1 shows that C-domain contacts N-domain at the interface denoted by a dashed line. In solution, the binding interface between N-domain and C-domain can be inferred from the significantly affected residues upon complex formation: residues of N-domain (green) with CSP ≥ 0.03 ppm are colored in red while those that were broadened beyond detection are colored in magenta according to (A); residues of C-domain (blue) with intensity reduction ≥ 0.8 are colored in pink while those that are not resolved are colored in grey according to (B). Several examples of the shifted cross-peaks as shown in Figure 5.9 are labeled by their residue numbers.

NMR titration experiments with ^{15}N-labeled N-domain and unlabeled C-domain, and *vice versa*, show that N-domain indeed interacts with C-domain at the interface that includes the contact area observed in the crystal structure of eRF1 (Figure 5.8). Majority of the amide resonances of C-domain were severely attenuated by line broadening upon N-domain binding (Figure 5.8, B and 5.10, B), while several amide resonances of N-domain were also disappeared when binding to C-domain (Figure 5.8, A). This indicates that the kinetic rate of binding is in the fast to intermediate exchange regime on NMR timescale, and the binding of N-domain may induce additional conformational exchange to C-domain on another timescale. This low affinity or transient nature of binding between N-domain and C-domain is supported by thermodynamically weak, entropy-driven interactions observed earlier by isothermal titration colorimetric studies (Kononenko et al., 2008). Nevertheless, CSP and extreme broadening of cross-peaks allowed us to map the binding interface on N-domain (Figure 5.8, C). The binding interface seems to be more extensive than can be predicted from the crystal structure of eRF1, which suggests that there are other modes of complex formation between N-domain and C-domain. In fact, variations in domain orientation were observed in the different crystal structures of eRF1 in complex with domain 2/3 of eRF3, as the relative orientations between N-domain and C-domain differ by 15°, 16°, and 30° for human complex versus free eRF1, human complex versus *Schizosaccharomyces pombe* complex, and *S. pombe* complex versus free eRF1, respectively (Cheng et al., 2009). It is interesting to note that part of the GTS loop was also significantly affected by the presence of C-domain (Figure 5.8, A and C).

5.9 Binding of 15-mer RNA displaces C-domain from the NC-complex

After establishing the non-covalent complex formation between N-domain and C-domain in solution, we investigated if the 15-mer RNA competes with C-domain for the same binding interface. Perturbations of the C-domain resonances were monitored in a competitive binding between the two domains and 15-mer RNA. Initially, addition of N-domain attenuated majority of the cross-peaks and shifted some of them (left panel in Figure 5.9, B). As 15-mer-UGA was titrated into the complex of N-domain and C-domain, the attenuated and shifted cross-peaks returned progressively to the initial state corresponding to free C-domain in solution (right panel in Figure 5.9, B). As we have already shown that the 15-mer RNA binds to N-domain, while C-domain does not interact with the 15-mer RNA at all (Figure 5.9, A), this result clearly demonstrates that the 15-mer-RNA is able to displace C-domain from the non-covalent complex of N-domain and C-domain. The competitive titration experiment was not performed on ^{15}N-labeled N-domain, since the resulting chemical shift perturbations would represent a summation of three states (*i.e.* free N-domain, N-domain + C-domain, and N-domain + RNA), which is more complicated than the current setup.

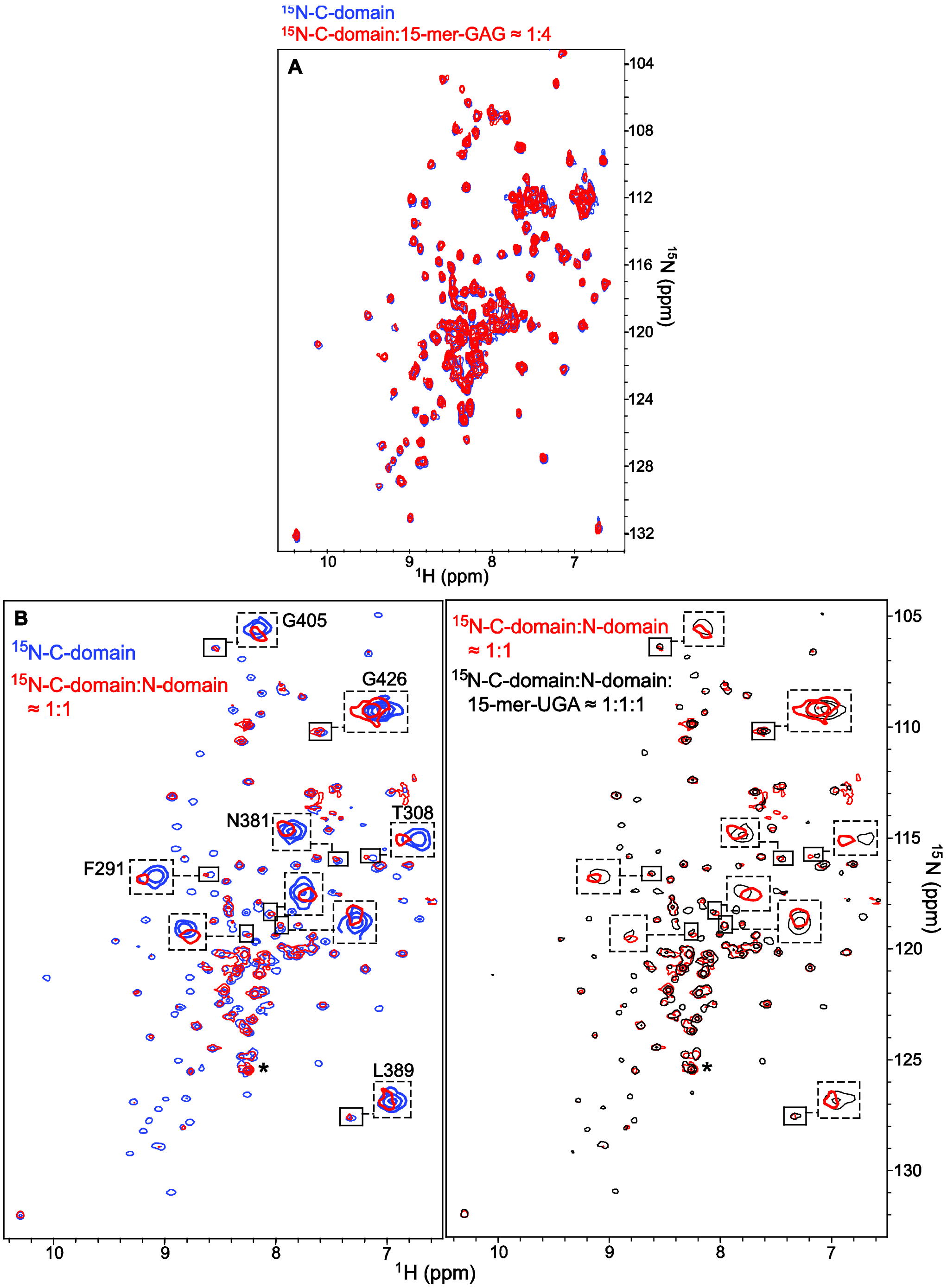


Figure 5.9 Binding of 15-mer RNA to N-domain displaces C-domain from the NC-complex. (A) Overlay of [^{1}H,^{15}N]-HSQC spectra of ^{15}N-labeled C-domain before (blue) and after addition of fourfold molar excess of 15-mer-GAG (red) shows that no perturbation had occurred to the C-domain resonances. (B) Overlay of TROSY [^{1}H,^{15}N]-HSQC spectra measured at different stages of the competitive binding experiment: free ^{15}N-labeled C-domain (blue), addition of equimolar amount of unlabeled N-domain to ^{15}N-labeled C-domain (red), addition of equimolar amount of 15-mer-UGA to the complex of unlabeled N-

domain and ^{15}N-labeled C-domain (black). Contour levels of the spectra are adjusted according to the cross-peak denoted by asterisk, which belongs to the His-tag of C-domain and is undisturbed at all three stages. Several examples of the shifted cross-peaks are highlighted by dashed boxes.

5.10 Discussion

5.10.1 Interactions between N-domain and mimics of the decoding region of H44

With the knowledge of the GTS loop being implicated in stop codon recognition, we seek to explore possible orientations of N-domain within the ribosomal A site by investigating interactions between N-domain and 18S rRNA. As genetic interactions between eRF1 and the decoding region of H44 of 18S rRNA had been reported (Velichutina et al., 2001), we decided to test if there is any direct interaction between them. Our model 15-mer RNA intended to mimic the decoding region of H44 outside of the ribosome does interact specifically with helix α1 of N-domain, and its binding affinity is significantly reduced in the absence of the internal loop. How far has this somewhat reductionist's approach achieved? Our results are well supported by previous studies and are able to provide novel insights. A truncated mutant of eRF1, eRF1$^{21\text{-}437}$, was shown to reduce significantly its own release activity, as well as the stimulating activity towards eRF3 GTPase, indicating that N-terminal deletion of eRF1 until residue I21 is enough to affect its binding to the ribosome (Frolova et al., 2000).

Aminoglycosides are known to reduce the fidelity of both elongation and termination of protein translation (Rospert et al., 2005). Although they bind much stronger to the prokaryotic than to the eukaryotic ribosome, the latter is still susceptible to the nonsense suppression effect of various aminoglycosides (Palmer et al., 1979, Singh et al., 1979). Furthermore, those drugs have been used effectively in alleviating diseases caused by premature termination codon (Rowe and Clancy, 2009). Most importantly, it was shown that the nonsense suppression induced by paromomycin in yeast is likely to be caused by the interference to the termination process, instead of compromising the selection of cognate tRNA (Salas-Marco and Bedwell, 2005). All these results support the hypothesis of direct interaction between N-domain and the decoding region of H44.

A1752G (*rdn15*), a point mutation in H44 of *S. cerevisiae* 18S rRNA (refer to 18S rRNA of *T. lanuginosus* in Figure 5.1, A for the nucleotide position), is able to rescue cell lethality caused by a mutant eRF1, Sup45p-P86A, at 37 °C (Velichutina et al., 2001). The point mutation P86A was thought to reduce Sup45p's efficiency in stop codon recognition, rather than inhibiting it completely. The rescue mechanism by *rdn15* is far from clear, but was postulated to prolong the residence time of the mutant eRF1 at the ribosomal A site. Interestingly, we have found that 15-mer-GAG, having a larger internal loop than the rest, binds significantly weaker to N-domain. As *rdn15* effectively reduces the native 3-nt to a 2-nt internal loop, the inverse correlation between N-domain-RNA binding affinity and the size of the internal loop provides a possible explanation for the complementary between *rdn15* and Sup45p-P86A.

5.10.2 A model of N-domain bound to the pre-termination complex

Prior to solving the high-resolution structures of eRF1-bound pre-termination complex (pre-TC), understanding of the mechanism of translation termination in eukaryotes will have to rely on piecing together biochemical, structural, and genetics data from different studies. Our structural study of N-domain led us to suggest that eRF1 might decode different stop codons by adopting distinct GTS loop conformations, thus implying direct access of the GTS loop to the stop codon. In addition, the data on N-domain-RNA interactions have shown that N-domain potentially interacts with H44 of 18S rRNA. Based on these results, we propose a structural model that encompasses currently known interactions between N-domain of eRF1 and A site of eukaryotic ribosome (Figure 5.10).

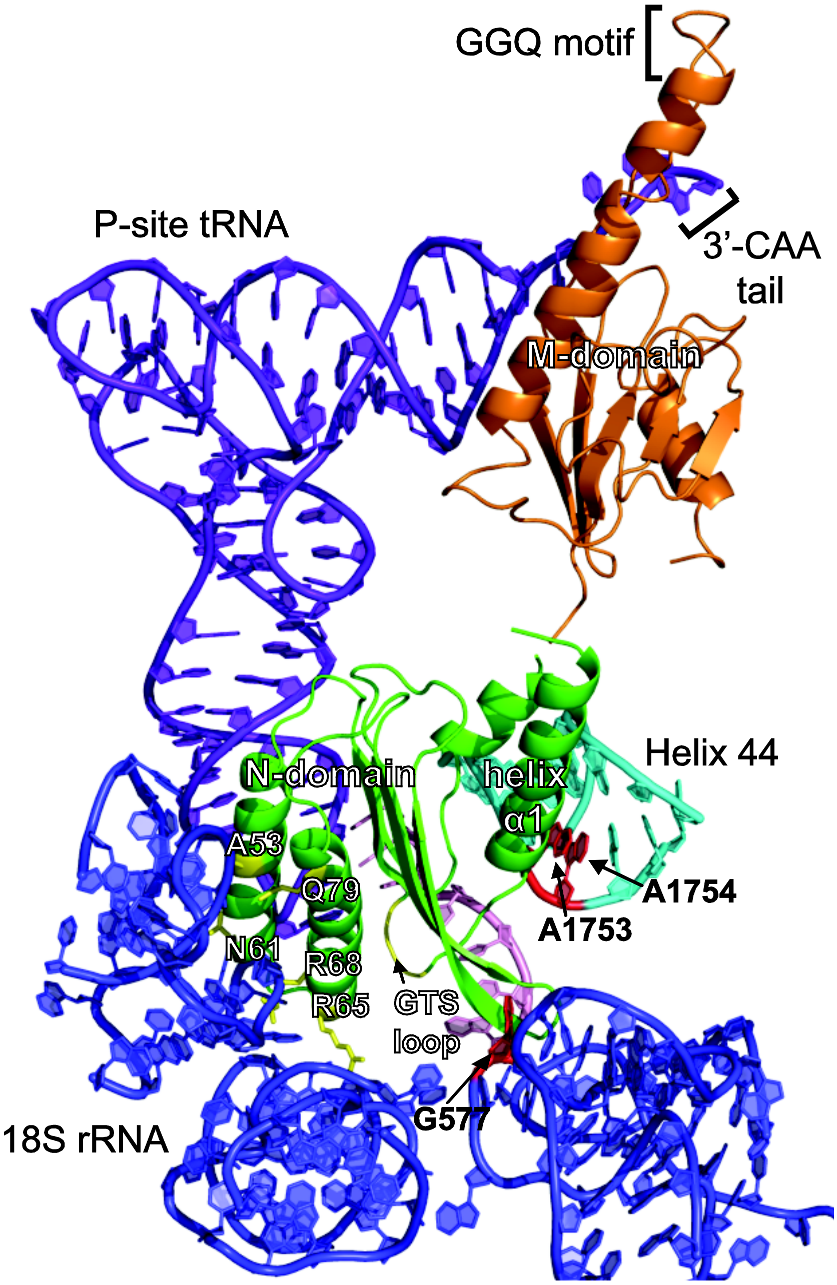

GGQ motif
P-site tRNA
3'-CAA tail
M-domain
Helix 44
N-domain
helix α1
A53
Q79
N61
R68
R65
GTS loop
A1754
A1753
G577
18S rRNA

Figure 5.10 A model of eRF1 NM-domain docked onto the A site of eukaryotic ribosome. NM-domain of eRF1 (green and orange) was manually docked onto the A site of 18S rRNA (blue) with P-site bound tRNA (purple) and mRNA (pink) (PDB ID: 3IZ7), based on the insights derived from the interactions between helix α1 of N-domain and the decoding region of H44 (cyan), as well as the putative role of the GTS loop in stop codon recognition. Steric clashes were prevented by visual monitoring, and no computational energy minimization was employed. The corresponding nucleotides critical for tRNA selection in prokaryotic ribosome and selected residues of eRF1 are highlighted in red and yellow, respectively.

The model shows that it is possible for the GTS loop to contact the stop codon while helix α1 is positioned next to the decoding region of Helix 44. Although helix α1 is not in the exact position to interact with Helix 44, a slight forward movement of the stop codon towards the P site will compensate for this discrepancy. Interestingly, it was reported that a 2-nt toeprint shift occurs when the eRF1•eRF3•GTP complex binds to the pre-TC (Alkalaeva et al., 2006). On the other side of N-domain, the side-chains of A53, N61, R65, R68, and Q79 are facing 18S rRNA. Residues R65 and R68 affect the binding of eRF1 to the ribosome (Frolova et al., 2002), while each of the point mutants, A53K, N61K and Q79K/R, was shown to substantially reduce the level of stop codon readthrough in comparison to wild-type, indicating enhanced ribosome binding due to the lysine or arginine substitution (Hatin et al., 2009). In our model, the GTS loop is close enough to the stop codon to allow photoactivatable cross-linking with the second and third stop codon positions (Bulygin et al., 2010). In the eRF1/pre-TC cross-linking experiments, the KSR loop (positions 63-65) and V66 were suggested to be in contact with the first stop codon position (Chavatte et al., 2002, Bulygin et al., 2010). Although within margins of cross-linking experiments, in our model these residues are not located in the direct proximity of the first uridine of the stop codon, thereby further experiments are required to resolve the issue.

With the orientation of N-domain in our model, a hinge motion between N-domain and M-domain would allow the GGQ motif to reach the 3’-CAA tail of P-site tRNA as had been suggested earlier (Cheng et al., 2009, Bulygin et al., 2010), while C-domain would be required to move away from helix α1 (Figure

5.10). The latter is well demonstrated by the competitive binding experiments (Figure 5.9). Hence, a major domain rearrangement between N-domain and C-domain is likely to occur in translation termination during which N-domain accommodates itself into the A site. A possible significance of the domain rearrangement in eRF1 is discussed in Chapter 6.

Chapter 6: Domain Interactions of NM-domain

6.1 Abstract

Peptidyl-tRNA hydrolysis can be uncoupled from GTP hydrolysis in the *in vitro* RF assay, since NM-domain of eRF1 is sufficient for stop codon-dependent peptide release (Frolova et al., 1996, Frolova et al., 2000). Nonetheless, the presence of eRF3 and GTP hydrolysis enhances significantly the efficiency of peptide release *in vitro* (Alkalaeva et al., 2006). The question is whether stop codon recognition by N-domain would propagate the signal to M-domain for peptidyl-tRNA hydrolysis. Isothermal titration calorimetric (ITC) measurement had shown that N-domain interacts with M-domain in entropy-driven way (Kononenko et al., 2008). Therefore, we are investigating the structural conformation of NM-domain in solution to gain insight into possible allostery regulation of the function of NM-domain.

6.2 Absence of Inter-domain Contact in NM-domain

The backbone amide and C_α chemical shifts of 259 out of 266 non-proline residues of NM-domain have been successfully assigned (Figure 6.1). The chemical shifts are very similar and comparable to the reported chemical shifts of individual N-domain (Oda et al., 2004) and M-domain (Ivanova et al., 2006) that had been deposited on the Biological Magnetic Resonance Bank (www.bmrb.wisc.edu). This gave the early hint on the absence of interaction interface between the two domains in the NM-domain construct.

NM-domain is susceptible to proteolytic degradation at the linker region (*i.e.* residues L140-S144). In the NM-domain sample without protease inhibitor, the M-domain part of the protein was completely hydrolyzed after two days at room temperature, while the N-domain part remained intact (Figure 6.2). Most importantly, the chemical shift of N-domain resonances remained unchanged, except for a few at the C-terminus. This indicates that besides the linker peptide N-domain does not interact with M-domain in the NM-domain construct, and hence they are likely to tumble independently in solution.

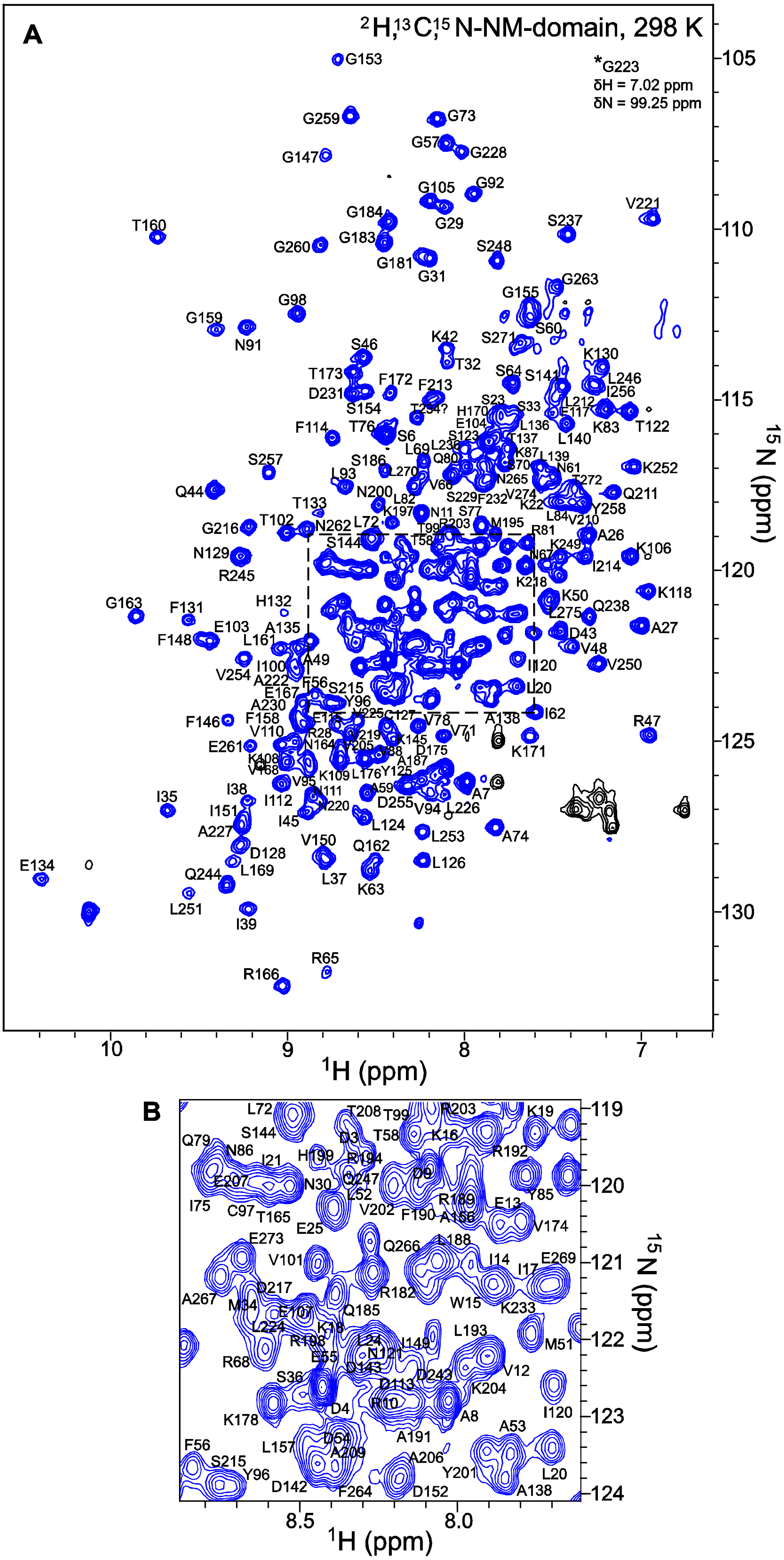

A
^{2}H,^{13}C,^{15}N-NM-domain, 298 K
*G223
δH = 7.02 ppm
δN = 99.25 ppm
G153
G259
G73
G57
G228
G147
G105
G92
G184
G29
V221
T160
S237
G183
G260
S248
G181
G31
G263
G155
G98
G159
N91
S271
S60
K42
S46
T32
K130
S64
S141
L246
T173
F172
F213
I256
D231
S23
S33
L212
S154
T234?
H170
E117
K83
T122
F114
T76
E104
L136
L140
S6
S123
T137
L69
L236
K87
L139
S186
Q80
S70
N61
L93
L270
N265
K252
S257
V66
T272
N200
S229
F232
V274
Q211
Q44
L82
K22
T133
K197
N11
S77
L84
V210
Y258
T102
N262
L72
T99
R203
M195
R81
A26
G216
S144
T58
K249
K106
N129
N67
I214
R245
K218
K118
K50
G163
H132
F131
L275
Q238
A135
D43
A27
E103
F148
L161
V48
I100
A49
V250
V254
I120
A222
F56
S215
L20
E167
A230
Y96
I62
F146
F158
E115
V225
C127
V78
A138
R47
V110
R28
V71
E261
N164
K219
K145
Y205
K171
K108
V88
D175
V168
A187
Y125
V95
K109
L176
A59
I38
I35
N111
A7
I112
D255
I151
N220
V94
L226
I45
L124
A227
L253
A74
V150
Q162
D128
E134
L126
L169
Q244
L37
K63
L251
I39
R65
R166
^{15}N (ppm)
105
110
115
120
125
130
10
9
8
7
^{1}H (ppm)
B
L72
T208
T99
R203
K19
Q79
S144
D3
T58
K16
N86
I21
H199
R194
R192
E207
N30
Q247
D9
I75
L52
R189
E13
Y85
C97
T165
V202
F190
A156
V174
E25
E273
Q266
L188
V101
I14
E269
I17
D217
R182
A267
M34
W15
K233
E107
Q185
L224
K18
L193
R198
L24
I149
N121
R68
E55
M51
D143
D243
S36
V12
D113
K204
D4
R10
I120
K178
A8
A53
D54
A191
F56
L157
A209
A206
S215
Y201
L20
Y96
D142
F264
D152
A138
119
120
121
122
123
124
^{15}N (ppm)
8.5
8.0
^{1}H (ppm)

Figure 6.1 Backbone amide resonances assignment of NM-domain. (A) and (B) TROSY [^{1}H,^{15}N]-HSQC spectrum of deuterated NM-domain measured at 700 MHz spectrometer, showing almost all expected resonances. G223 lies outside of the window shown.

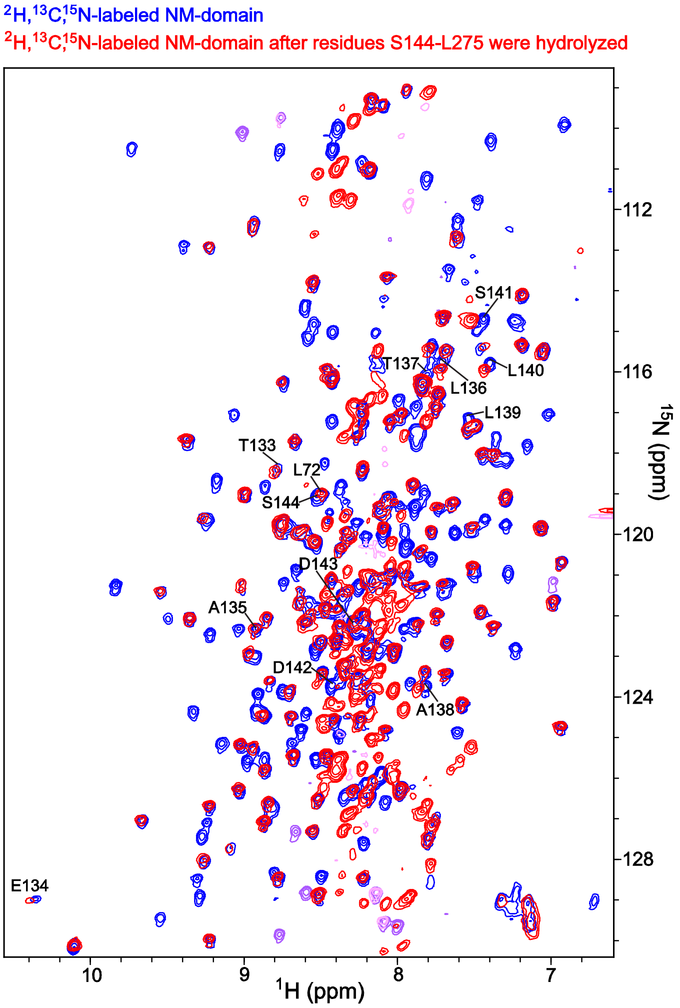


Figure 6.2 Absence of contact interface between N-domain and M-domain in the covalently linked NM-domain. N-domain and M-domain are connected covalently by linker peptide (residues L140-S144). Overlay of TROSY [^{1}H, ^{15}N]-HSQC spectra of deuterated ^{13}C,^{15}N-labeled NM-domain before (blue) and after M-domain was hydrolyzed two days later (red) show that the remaining N-domain resonances were minimally perturbed except the residue-labeled resonances, which are close to the C-terminus of N-domain.

6.3 Precipitation of NM-domain by 15-mer-UAA

The 15-mer RNA constructs bind to helix α1 of N-domain, signifying an important interaction between N-domain and the decoding region of H44 in 18S rRNA. We would like to investigate the response of whole NM-domain to 15-mer-UAA. Unfortunately, severe precipitation of the sample was observed upon addition of 15-mer-UAA into NM-domain, and S/N ratio of the resulting spectrum was drastically affected such that the result was no longer interpretable (Figure 6.3). The reduced sensitivity of the spectrum could be caused by either sample precipitation or chemical exchange-induced relaxation, but the issue was not resolved because the precipitation was too severe.

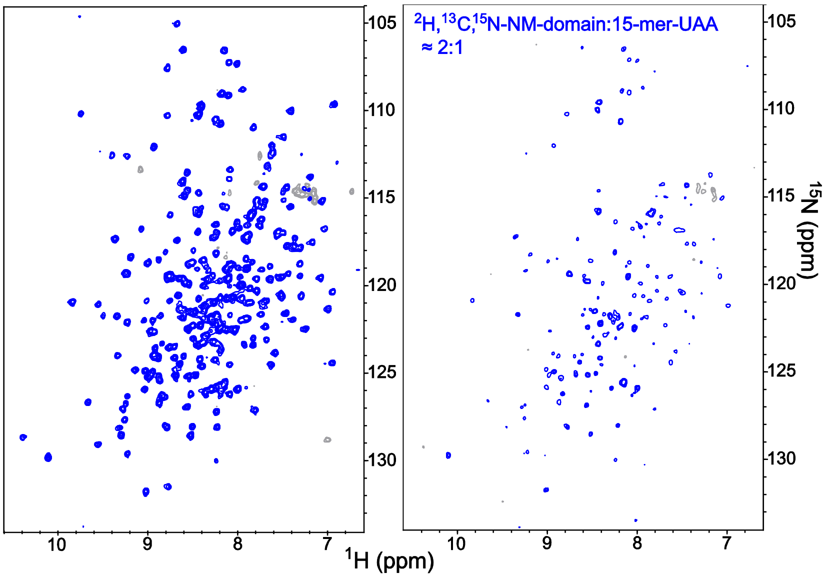


Figure 6.3 Sensitivity loss upon addition of 15-mer-UAA into NM-domain.

6.4 Discussion

In the crystal structure of full-length eRF1, there was no contact between N- and M-domains, except for the linker peptide (Figure 6.4, A). In solution, the lack of interaction interface also suggests that the two domains might tumble independently of each other. However, we still cannot rule out the possibility of allostery regulation via helix α4 of N-domain and the linker peptide. It would be interesting in the future to look into the inter-domain motion of NM-domain using other state-of-the-art experimental methods (Chen and Tjandra, 2008, Monkenbusch et al., 2010, Shapira and Prestegard, 2010).

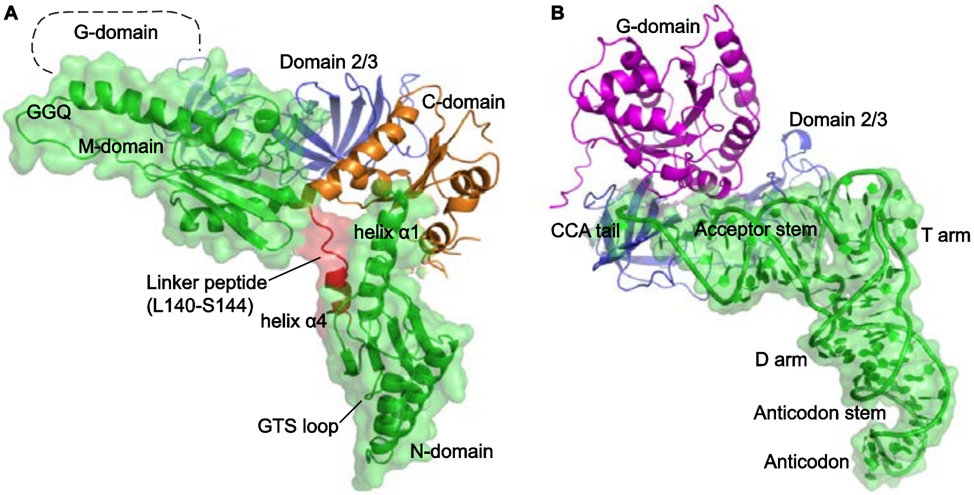

A
G-domain
Domain 2/3
C-domain
GGQ
M-domain
helix α1
Linker peptide
(L140-S144)
helix α4
GTS loop
N-domain
B
G-domain
Domain 2/3
CCA tail
Acceptor stem
T arm
D arm
Anticodon stem
Anticodon

Figure 6.4 Domain architecture of eRF1•eRF3 and tRNA•EF-Tu. (A) For clarity purpose, the molecular surface of N-domain and M-domain of eRF1 in complex with domain 2/3 of eRF3 (PDB ID: 3E1Y) is shown and colored green, while C-domain is colored orange. Residues labeled in Figure 6.2 are colored red. (B) 70S ribosome-bound tRNA•EF-Tu ternary complex trapped in the A/T state by non-hydrolyzable GTP analog (PDB ID: 2XQD).

Based on the available data, peptidyl-tRNA hydrolysis achieved by NM-domain alone may not be allosteric regulated by stop codon recognition. Instead, M-domain may possess certain ribosomal binding motif that guides the GGQ loop to PTC (Kisselev et al., 2003). On the other hand, these clues together with the findings presented in Chapter 5 further suggest that the domain rearrangement between N-domain and C-domain upon interaction with the decoding region of H44 might play a role in signal transduction that could accelerate peptidyl-tRNA hydrolysis. This would explain the significantly enhanced peptide release efficiency by eRF1•eRF3•GTP *in vitro*, in comparison to NM-domain alone (Alkalaeva et al., 2006).

The overall picture of the mechanism from stop codon recognition to peptidyl-tRNA hydrolysis is far from clear. First of all, it remains to be tested whether stop codon recognition is kinetically coupled to the proposed domain rearrangement in eRF1, since it had been shown that eRF1 can bind to the ribosome independent of stop codon (Frolova et al., 1996). Secondly, it is unclear what happens after the domain rearrangement. One possible reaction that would have been triggered by the domain rearrangement is GTP hydrolysis in eRF3, since the movement of C-domain of eRF1 could reposition domain 2/3 as well as G-domain of eRF3, relative to the catalytic sarcin-ricin loop (SRL) of the 60S subunit. The rationale behind this proposal is based on the general mechanism for activation of GTP hydrolysis in translational GTPases derived from the crystal structure of 70S ribosome-bound tRNA•EF-Tu in the A/T state (Voorhees et al., 2010, Valle et al., 2002) (Figure 6.4, B). One approach to prove this hypothesis is by measuring the GTPase activity of the reconstituted pre-TC with mutant eRF1 that has either (i) C-domain being cross-linked to N-domain, or (ii) N-domain-H44 interaction being disrupted.

Chapter 7: Conclusions

Two bodies of works have been presented in this thesis. In the first project, we aimed to test the computational feasibility of real-time assignment of backbone resonances of a protein in the shortest time possible. The project was motivated by the availability of novel NMR processing technique like MDD for spectrum reconstruction from nonlinearly sampled data, our in-house development of advanced automatic sequential assignment program *AutoLink*, as well as the calling for NMR methodology that tackles the problem of poor sample stability during the period of conventional 3D NMR experiments. We proposed the method of targeted acquisition of NMR data that is dynamically controlled by the completeness of automatic backbone resonances assignment. We had demonstrated that *Psyte*/*AutoLink II* is capable of producing accurate assignment of the backbone resonances based on five conventional 3D spectra reconstructed from different percentages of data sampling, thereby proving the point of targeted acquisition. Furthermore, the time saving factor of over 80% in that particular demonstration is also significant and widely applicable. We had further investigated in details the robustness of the automatic process using a bootstrap approach, and discovered the "assignment bottleneck" caused by NMR experiment of relatively lower sensitivity, and the impact of variability in *Psyte*'s outputs on the consistency of residue specific assignment done by *AutoLink II*. Based on these results, the guiding criterion for optimal run-time allocation of NMR resources between different experiments can be redefined with the inclusion of parameters describing the system-dependent sensitivity of the respective experiments (Senthamarai et al., 2010), and *Psyte*/*AutoLink II* can also be further improved by implementing an additional layer of logic that monitors the consistency of assignment. In conclusion, our proposed method of targeted acquisition of NMR data represents a viable strategy for NMR structural studies. My works in this project have paved the foundation for the second project on eRF1 and translation termination.

In the second project, we were studying the atomic structures of several mutants of N-domain of eRF1 to find out the reasons behind the different stop codon specificities in different mutants, as well as the interactions between N-domain and mimics of the decoding region of H44 to understand the binding of eRF1 to ribosome. One of the mutants, $Q^{122}FM(Y)F^{126}$, exhibits very strong UGA-only specificity. By comparing the solution structures of wild-type N-domain and $Q^{122}FM(Y)F^{126}$, and the crystal structure of N-domain in full-length eRF1, we observed distinct conformations of the GTS loop, that are most probably caused by the observed repositioning of helix $\alpha3$ and changes in the hydrophobic core right above the GTS loop propagated from the mutation sites. These structural insights together with the flexible nature of the GTS loop on sub-nanosecond timescale suggest that the GTS loop could potentially act as a switch to decode different stop codons. The conformational switching of protein backbone usually occurs on a slower timescale, and thus can be studied by other NMR relaxation experiments in the future. By consolidating our current data and the previous results from other researchers, we conclude that structural variability in the GTS loop may underline the switching between omnipotency and unipotency of eRF1, hence implying direct access of the GTS loop to the stop codon. Our findings are critical to resolve previous confusions inherent to the different models of stop codon recognition derived from other computational and experimental studies.

The importance in understanding the binding process of eRF1 to ribosome is reflected by the direct impact of aminoglycosides and mutations in the decoding region of H44 on eukaryotic translation termination. eRF1 has intrinsic affinity for ribosome, and several putative ribosomal binding sites on eRF1 had been proposed previously. Without relying on the isotopic labeling of RNA samples, we managed to design and characterize a series of double-stranded RNA constructs mimicking the decoding region of H44 with varied nucleotide sequence and length of the internal loops. Several conclusions can be drawn from the NMR titration experiments regarding the interactions between N-domain and mimics of H44: (i) the binding is rather weak (*i.e.* in micromolar range) but is specific to helix α1 of N-domain, (ii) the binding is critically dependent on the internal loop as well as its length, but not on its sequence, and (iii) similar interface on N-domain was perturbed by the addition of both prokaryotic and eukaryotic ribosomes. Most importantly, the oligoribonucleotides could compete with C-domain and displace it from a non-covalent complex of N-domain and C-domain, suggesting strongly that a domain rearrangement in eRF1 may occur during which N-domain accommodates itself into A site. The critical contribution of NMR spectroscopy is the feasibility to study weak but biologically significant interactions for dynamic processes. We reckon that our approach has successfully shed lights on parts of the highly complicated mechanism of translation termination in eukaryotes, and our docked model of eRF1 N-domain onto A site of 40S subunit would become useful.

References

ALKALAEVA, E. Z., PISAREV, A. V., FROLOVA, L. Y., KISSELEV, L. L. & PESTOVA, T. V. 2006. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell,* 125**,** 1125-36.

ALLERS, J. & SHAMOO, Y. 2001. Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J Mol Biol,* 311**,** 75-86.

ALTAMURA, S., CAMMARANO, P. & LONDEI, P. 1986. Archaebacterial and eukaryotic ribosomal subunits can form active hybrid ribosomes. *FEBS Lett,* 204**,** 129-33.

ARAKAWA, M., SHIOZUKA, M., NAKAYAMA, Y., HARA, T., HAMADA, M., KONDO, S., IKEDA, D., TAKAHASHI, Y., SAWA, R., NONOMURA, Y., SHEYKHOLESLAMI, K., KONDO, K., KAGA, K., KITAMURA, T., SUZUKI-MIYAGOE, Y., TAKEDA, S. & MATSUDA, R. 2003. Negamycin restores dystrophin expression in skeletal and cardiac muscles of mdx mice. *J Biochem,* 134**,** 751-8.

ARMSTRONG, G. S., MANDELSHTAM, V. A., SHAKA, A. J. & BENDIAK, B. 2005. Rapid high-resolution four-dimensional NMR spectroscopy using the filter diagonalization method and its advantages for detailed structural elucidation of oligosaccharides. *J Magn Reson,* 173**,** 160-8.

ATREYA, H. S., GARCIA, E., SHEN, Y. & SZYPERSKI, T. 2007. J-GFT NMR for precise measurement of mutually correlated nuclear spin-spin couplings. *J Am Chem Soc,* 129**,** 680-92.

ATREYA, H. S., SAHU, S. C., CHARY, K. V. & GOVIL, G. 2000. A tracked approach for automated NMR assignments in proteins (TATAPRO). *J Biomol NMR,* 17**,** 125-36.

BAN, N., NISSEN, P., HANSEN, J., MOORE, P. B. & STEITZ, T. A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science,* 289**,** 905-20.

BARTELS, C., XIA, T., BILLETER, M., GUNTERT, P. & WUTHRICH, K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR,* 6**,** 1-10.

BEAUDET, A. L. & CASKEY, C. T. 1971. Mammalian peptide chain termination. II. Codon specificity and GTPase activity of release factor. *Proc Natl Acad Sci U S A,* 68**,** 619-24.

BECKER, T., FRANCKENBERG, S., WICKLES, S., SHOEMAKER, C. J., ANGER, A. M., ARMACHE, J. P., SIEBER, H., UNGEWICKELL, C., BERNINGHAUSEN, O., DABERKOW, I., KARCHER, A., THOMM, M., HOPFNER, K. P., GREEN, R. & BECKMANN, R. 2012. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature,* 482**,** 501-6.

BEN-SHEM, A., GARREAU DE LOUBRESSE, N., MELNIKOV, S., JENNER, L., YUSUPOVA, G. & YUSUPOV, M. 2011. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science,* 334**,** 1524-9.

BERTRAM, G., BELL, H. A., RITCHIE, D. W., FULLERTON, G. & STANSFIELD, I. 2000. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA,* 6**,** 1236-47.

BIDOU, L., ALLAMAND, V., ROUSSET, J. P. & NAMY, O. 2012. Sense from nonsense: therapies for premature stop codon diseases. *Trends Mol Med,* 18**,** 679-88.

BLANCHARD, S. C., GONZALEZ, R. L., KIM, H. D., CHU, S. & PUGLISI, J. D. 2004. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol,* 11**,** 1008-14.

BULYGIN, K. N., KHAIRULINA, Y. S., KOLOSOV, P. M., VEN'YAMINOVA, A. G., GRAIFER, D. M., VOROBJEV, Y. N., FROLOVA, L. Y. & KARPOVA, G. G. 2011. Adenine and guanine recognition of stop codon is mediated by different N domain conformations of translation termination factor eRF1. *Nucleic Acids Res,* 39**,** 7134-46.

BULYGIN, K. N., KHAIRULINA, Y. S., KOLOSOV, P. M., VEN'YAMINOVA, A. G., GRAIFER, D. M., VOROBJEV, Y. N., FROLOVA, L. Y., KISSELEV, L. L. & KARPOVA, G. G. 2010. Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome. *RNA,* 16**,** 1902-14.

CAPECCHI, M. R. 1967. Polypeptide chain termination in vitro: isolation of a release factor. *Proc Natl Acad Sci U S A,* 58**,** 1144-51.

CASKEY, C. T., BEAUDET, A. L. & TATE, W. P. 1974. Mammalian release factor; in vitro assay and purification. *Methods Enzymol,* 30**,** 293-303.

CASKEY, C. T., TOMPKINS, R., SCOLNICK, E., CARYK, T. & NIRENBERG, M. 1968. Sequential translation of trinucleotide codons for the initiation and termination of protein synthesis. *Science,* 162**,** 135-8.

CHAVATTE, L., FROLOVA, L., LAUGAA, P., KISSELEV, L. & FAVRE, A. 2003. Stop codons and UGG promote efficient binding of the polypeptide release factor eRF1 to the ribosomal A site. *J Mol Biol,* 331**,** 745-58.

CHAVATTE, L., SEIT-NEBI, A., DUBOVAYA, V. & FAVRE, A. 2002. The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J,* 21**,** 5302-11.

CHEN, K. & TJANDRA, N. 2008. Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. *J Am Chem Soc,* 130**,** 12745-51.

CHENG, Z., SAITO, K., PISAREV, A. V., WADA, M., PISAREVA, V. P., PESTOVA, T. V., GAJDA, M., ROUND, A., KONG, C., LIM, M., NAKAMURA, Y., SVERGUN, D. I., ITO, K. & SONG, H. 2009. Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev,* 23**,** 1106-18.

CLORE, G. M., GRONENBORN, A. M., PIPER, E. A., MCLAUGHLIN, L. W., GRAESER, E. & VAN BOOM, J. H. 1984. The solution structure of a RNA pentadecamer comprising the anticodon loop and stem of yeast tRNAPhe. A 500 MHz 1H-n.m.r. study. *Biochem J,* 221**,** 737-51.

COGGINS, B. E. & ZHOU, P. 2003. PACES: Protein sequential assignment by computer-assisted exhaustive search. *J Biomol NMR,* 26**,** 93-111.

CORNILESCU, G., DELAGLIO, F. & BAX, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR,* 13**,** 289-302.

DELAGLIO, F., GRZESIEK, S., VUISTER, G. W., ZHU, G., PFEIFER, J. & BAX, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR,* 6**,** 277-93.

DEMESHKINA, N., JENNER, L., WESTHOF, E., YUSUPOV, M. & YUSUPOVA, G. 2012. A new understanding of the decoding principle on the ribosome. *Nature,* 484**,** 256-9.

DONTSOVA, M., FROLOVA, L., VASSILIEVA, J., PIENDL, W., KISSELEV, L. & GARBER, M. 2000. Translation termination factor aRF1 from the archaeon Methanococcus jannaschii is active with eukaryotic ribosomes. *FEBS Lett,* 472**,** 213-6.

DUNKLE, J. A., WANG, L., FELDMAN, M. B., PULK, A., CHEN, V. B., KAPRAL, G. J., NOESKE, J., RICHARDSON, J. S., BLANCHARD, S. C. & CATE, J. H. 2011. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science,* 332**,** 981-4.

EGHBALNIA, H. R., BAHRAMI, A., TONELLI, M., HALLENGA, K. & MARKLEY, J. L. 2005a. High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J Am Chem Soc,* 127**,** 12528-36.

EGHBALNIA, H. R., BAHRAMI, A., WANG, L., ASSADI, A. & MARKLEY, J. L. 2005b. Probabilistic Identification of Spin Systems and their Assignments including Coil-Helix Inference as Output (PISTACHIO). *J Biomol NMR,* 32**,** 219-33.

ELISEEV, B., KRYUCHKOVA, P., ALKALAEVA, E. & FROLOVA, L. 2010. A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. *Nucleic Acids Res,* 39**,** 599-608.

ENGGIST, E., THONY-MEYER, L., GUNTERT, P. & PERVUSHIN, K. 2002. NMR structure of the heme chaperone CcmE reveals a novel functional motif. *Structure,* 10**,** 1551-7.

EURWILAICHITR, L., GRAVES, F. M., STANSFIELD, I. & TUITE, M. F. 1999. The C-terminus of eRF1 defines a functionally important domain for translation termination in Saccharomyces cerevisiae. *Mol Microbiol,* 32**,** 485-96.

FAN-MINOGUE, H. & BEDWELL, D. M. 2008. Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. *RNA,* 14**,** 148-57.

FAN-MINOGUE, H., DU, M., PISAREV, A. V., KALLMEYER, A. K., SALAS-MARCO, J., KEELING, K. M., THOMPSON, S. R., PESTOVA, T. V. & BEDWELL, D. M. 2008. Distinct eRF3 requirements suggest alternate eRF1 conformations mediate peptide release during eukaryotic translation termination. *Mol Cell,* 30**,** 599-609.

FELDEN, B. & GILLET, R. 2011. SmpB as the handyman of tmRNA during trans-translation. *RNA Biol,* 8**,** 440-9.

FENG, S., LI, H., ZHAO, J., PERVUSHIN, K., LOWENHAUPT, K., SCHWARTZ, T. U. & DROGE, P. 2011. Alternate rRNA secondary structures as regulators of translation. *Nat Struct Mol Biol,* 18**,** 169-76.

FIELDING, L. 2003. NMR methods for the determination of protein-ligand dissociation constants. *Curr Top Med Chem,* 3**,** 39-53.

FIORITO, F., HILLER, S., WIDER, G. & WUTHRICH, K. 2006. Automated resonance assignment of proteins: 6D APSY-NMR. *J Biomol NMR,* 35**,** 27-37.

FOURMY, D., RECHT, M. I., BLANCHARD, S. C. & PUGLISI, J. D. 1996. Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science,* 274**,** 1367-71.

FREISTROFFER, D. V., KWIATKOWSKI, M., BUCKINGHAM, R. H. & EHRENBERG, M. 2000. The accuracy of codon recognition by polypeptide release factors. *Proc Natl Acad Sci U S A,* 97**,** 2046-51.

FREISTROFFER, D. V., PAVLOV, M. Y., MACDOUGALL, J., BUCKINGHAM, R. H. & EHRENBERG, M. 1997. Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J,* 16**,** 4126-33.

FROLOVA, L., LE GOFF, X., RASMUSSEN, H. H., CHEPEREGIN, S., DRUGEON, G., KRESS, M., ARMAN, I., HAENNI, A. L., CELIS, J. E., PHILIPPE, M. & ET AL. 1994. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature,* 372**,** 701-3.

FROLOVA, L., LE GOFF, X., ZHOURAVLEVA, G., DAVYDOVA, E., PHILIPPE, M. & KISSELEV, L. 1996. Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA,* 2**,** 334-41.

FROLOVA, L., SEIT-NEBI, A. & KISSELEV, L. 2002. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA,* 8**,** 129-36.

FROLOVA, L. Y., MERKULOVA, T. I. & KISSELEV, L. L. 2000. Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. *RNA,* 6**,** 381-90.

FROLOVA, L. Y., TSIVKOVSKII, R. Y., SIVOLOBOVA, G. F., OPARINA, N. Y., SERPINSKY, O. I., BLINOV, V. M., TATKOV, S. I. & KISSELEV, L. L. 1999. Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA,* 5**,** 1014-20.

FRUEH, D. P., SUN, Z. Y., VOSBURG, D. A., WALSH, C. T., HOCH, J. C. & WAGNER, G. 2006. Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. *J Am Chem Soc,* 128**,** 5757-63.

FRYDMAN, L., LUPULESCU, A. & SCHERF, T. 2003. Principles and features of single-scan two-dimensional NMR spectroscopy. *J Am Chem Soc,* 125**,** 9204-17.

FURTIG, B., RICHTER, C., WOHNERT, J. & SCHWALBE, H. 2003. NMR spectroscopy of RNA. *Chembiochem,* 4**,** 936-62.

GAL, M., SCHANDA, P., BRUTSCHER, B. & FRYDMAN, L. 2007. UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *J Am Chem Soc,* 129**,** 1372-7.

GAO, H., ZHOU, Z., RAWAT, U., HUANG, C., BOUAKAZ, L., WANG, C., CHENG, Z., LIU, Y., ZAVIALOV, A., GURSKY, R., SANYAL, S., EHRENBERG, M., FRANK, J. & SONG, H. 2007. RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors. *Cell,* 129**,** 929-41.

GAO, Y. G., SELMER, M., DUNHAM, C. M., WEIXLBAUMER, A., KELLEY, A. C. & RAMAKRISHNAN, V. 2009. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science,* 326**,** 694-9.

GOLDSTEIN, J. L., BEAUDET, A. L. & CASKEY, C. T. 1970. Peptide chain termination with mammalian release factor. *Proc Natl Acad Sci U S A,* 67**,** 99-106.

GRISHAEV, A. & LLINAS, M. 2004. BACUS: A Bayesian protocol for the identification of protein NOESY spectra via unassigned spin systems. *J Biomol NMR,* 28**,** 1-10.

GRISHAEV, A., STEREN, C. A., WU, B., PINEDA-LUCENA, A., ARROWSMITH, C. & LLINAS, M. 2005. ABACUS, a direct method for protein NMR structure computation via assembly of fragments. *Proteins,* 61**,** 36-43.

GUNTERT, P., MUMENTHALER, C. & WUTHRICH, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol,* 273**,** 283-98.

HATIN, I., FABRET, C., ROUSSET, J. P. & NAMY, O. 2009. Molecular dissection of translation termination mechanism identifies two new critical regions in eRF1. *Nucleic Acids Res,* 37**,** 1789-98.

HE, S. L. & GREEN, R. 2010. Visualization of codon-dependent conformational rearrangements during translation termination. *Nat Struct Mol Biol,* 17**,** 465-70.

HERMANN, T. 2005. Drugs targeting the ribosome. *Curr Opin Struct Biol,* 15**,** 355-66.

HERRMANN, T., GUNTERT, P. & WUTHRICH, K. 2002a. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR,* 24**,** 171-89.

HERRMANN, T., GUNTERT, P. & WUTHRICH, K. 2002b. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol,* 319**,** 209-27.

HEUS, H. A. & PARDI, A. 1991. Novel proton NMR assignment procedure for RNA duplexes. *J Am Chem Soc,* 113**,** 4360-61.

HILLER, S., FIORITO, F., WUTHRICH, K. & WIDER, G. 2005. Automated projection spectroscopy (APSY). *Proc Natl Acad Sci U S A,* 102**,** 10876-81.

HILLER, S., WASMER, C., WIDER, G. & WUTHRICH, K. 2007. Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. *J Am Chem Soc,* 129**,** 10823-8.

HITCHENS, T. K., LUKIN, J. A., ZHAN, Y., MCCALLUM, S. A. & RULE, G. S. 2003. MONTE: An automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins. *J Biomol NMR,* 25**,** 1-9.

INAGAKI, Y., BLOUIN, C., DOOLITTLE, W. F. & ROGER, A. J. 2002. Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. *Nucleic Acids Res,* 30**,** 532-44.

ISHIMA, R. & TORCHIA, D. A. 2000. Protein dynamics from NMR. *Nat Struct Biol,* 7**,** 740-3.

ITO, K., EBIHARA, K. & NAKAMURA, Y. 1998. The stretch of C-terminal acidic amino acids of translational release factor eRF1 is a primary binding site for eRF3 of fission yeast. *RNA,* 4**,** 958-72.

ITO, K., FROLOVA, L., SEIT-NEBI, A., KARAMYSHEV, A., KISSELEV, L. & NAKAMURA, Y. 2002. Omnipotent decoding potential resides in eukaryotic translation termination factor eRF1 of variant-code organisms and is modulated by the interactions of amino acid sequences within domain 1. *Proc Natl Acad Sci U S A,* 99**,** 8494-9.

ITO, K., UNO, M. & NAKAMURA, Y. 2000. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature,* 403**,** 680-4.

IVANOVA, E. V., KOLOSOV, P. M., BIRDSALL, B., KISSELEV, L. L. & POLSHAKOV, V. I. 2006. NMR assignments of the middle domain of human polypeptide release factor eRF1. *J Biomol NMR,* 36 Suppl 1**,** 8.

JARAVINE, V., IBRAGHIMOV, I. & OREKHOV, V. Y. 2006. Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. *Nat Methods,* 3**,** 605-7.

JARAVINE, V. A. & OREKHOV, V. Y. 2006. Targeted acquisition for real-time NMR spectroscopy. *J Am Chem Soc,* 128**,** 13421-6.

JARAVINE, V. A., ZHURAVLEVA, A. V., PERMI, P., IBRAGHIMOV, I. & OREKHOV, V. Y. 2008. Hyperdimensional NMR spectroscopy with nonlinear sampling. *J Am Chem Soc,* 130**,** 3927-36.

JARYMOWYCZ, V. A. & STONE, M. J. 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev,* 106**,** 1624-71.

JIN, H., KELLEY, A. C., LOAKES, D. & RAMAKRISHNAN, V. 2010. Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc Natl Acad Sci U S A,* 107**,** 8593-8.

JORGENSEN, F., ADAMSKI, F. M., TATE, W. P. & KURLAND, C. G. 1993. Release factor-dependent false stops are infrequent in Escherichia coli. *J Mol Biol,* 230**,** 41-50.

JUNG, Y. S. & ZWECKSTETTER, M. 2004. Mars -- robust automatic backbone assignment of proteins. *J Biomol NMR,* 30**,** 11-23.

KAZIMIERCZUK, K., KOZMINSKI, W. & ZHUKOV, I. 2006a. Two-dimensional Fourier transform of arbitrarily sampled NMR data sets. *J Magn Reson,* 179**,** 323-8.

KAZIMIERCZUK, K., ZAWADZKA, A., KOZMINSKI, W. & ZHUKOV, I. 2006b. Random sampling of evolution time space and Fourier transform processing. *J Biomol NMR,* 36**,** 157-68.

KELLY, M. J., KRIEGER, C., BALL, L. J., YU, Y., RICHTER, G., SCHMIEDER, P., BACHER, A. & OSCHKINAT, H. 1999. Application of amino acid type-specific 1H- and 14N-labeling in a 2H-, 15N-labeled background to a 47 kDa homodimer: potential for NMR structure determination of large proteins. *J Biomol NMR,* 14**,** 79-83.

KERVESTIN, S. & JACOBSON, A. 2012. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol,* 13**,** 700-12.

KIM, O. T., YURA, K., GO, N. & HARUMOTO, T. 2005. Newly sequenced eRF1s from ciliates: the diversity of stop codon usage and the molecular surfaces that are important for stop codon interactions. *Gene,* 346**,** 277-86.

KIM, S. & SZYPERSKI, T. 2003. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc,* 125**,** 1385-93.

KISSELEV, L., EHRENBERG, M. & FROLOVA, L. 2003. Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J,* 22**,** 175-82.

KLINGE, S., VOIGTS-HOFFMANN, F., LEIBUNDGUT, M., ARPAGAUS, S. & BAN, N. 2011. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science,* 334**,** 941-8.

KLINGE, S., VOIGTS-HOFFMANN, F., LEIBUNDGUT, M. & BAN, N. 2012. Atomic structures of the eukaryotic ribosome. *Trends Biochem Sci,* 37**,** 189-98.

KNIGHT, R. D. & LANDWEBER, L. F. 2000. The early evolution of the genetic code. *Cell,* 101**,** 569-72.

KOBAYASHI, K., SAITO, K., ISHITANI, R., ITO, K. & NUREKI, O. 2012. Structural basis for translation termination by archaeal RF1 and GTP-bound EF1alpha complex. *Nucleic Acids Res,* 40**,** 9319-28.

KOBAYASHI, T., FUNAKOSHI, Y., HOSHINO, S. & KATADA, T. 2004. The GTP-binding release factor eRF3 as a key mediator coupling translation termination to mRNA decay. *J Biol Chem,* 279**,** 45693-700.

KOLOSOV, P., FROLOVA, L., SEIT-NEBI, A., DUBOVAYA, V., KONONENKO, A., OPARINA, N., JUSTESEN, J., EFIMOV, A. & KISSELEV, L. 2005. Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res,* 33**,** 6418-25.

KONDO, J., URZHUMTSEV, A. & WESTHOF, E. 2006. Two conformational states in the crystal structure of the Homo sapiens cytoplasmic ribosomal decoding A site. *Nucleic Acids Res,* 34**,** 676-85.

KONG, C., ITO, K., WALSH, M. A., WADA, M., LIU, Y., KUMAR, S., BARFORD, D., NAKAMURA, Y. & SONG, H. 2004. Crystal structure and functional analysis of the eukaryotic class II release factor eRF3 from S. pombe. *Mol Cell,* 14**,** 233-45.

KONONENKO, A. V., MITKEVICH, V. A., DUBOVAYA, V. I., KOLOSOV, P. M., MAKAROV, A. A. & KISSELEV, L. L. 2008. Role of the individual domains of translation termination factor eRF1 in GTP binding to eRF3. *Proteins,* 70**,** 388-93.

KORADI, R., BILLETER, M. & WUTHRICH, K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph,* 14**,** 51-5, 29-32.

KOROSTELEV, A., ASAHARA, H., LANCASTER, L., LAURBERG, M., HIRSCHI, A., ZHU, J., TRAKHANOV, S., SCOTT, W. G. & NOLLER, H. F. 2008. Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A,* 105**,** 19684-9.

KOROSTELEV, A., ZHU, J., ASAHARA, H. & NOLLER, H. F. 2010. Recognition of the amber UAG stop codon by release factor RF1. *EMBO J,* 29**,** 2577-85.

KUPCE, E. & FREEMAN, R. 2004. Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc,* 126**,** 6429-40.

KUPCE, E. & FREEMAN, R. 2007. Fast multidimensional NMR by polarization sharing. *Magn Reson Chem,* 45**,** 2-4.

LANGMEAD, C. J. & DONALD, B. R. 2004. An expectation/maximization nuclear vector replacement algorithm for automated NMR resonance assignments. *J Biomol NMR,* 29**,** 111-38.

LASKOWSKI, R. A., RULLMANNN, J. A., MACARTHUR, M. W., KAPTEIN, R. & THORNTON, J. M. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR,* 8**,** 477-86.

LAURBERG, M., ASAHARA, H., KOROSTELEV, A., ZHU, J., TRAKHANOV, S. & NOLLER, H. F. 2008. Structural basis for translation termination on the 70S ribosome. *Nature,* 454**,** 852-7.

LAURENT-PUIG, P., BEROUD, C. & SOUSSI, T. 1998. APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res,* 26**,** 269-70.

LECOMPTE, O., RIPP, R., THIERRY, J. C., MORAS, D. & POCH, O. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res,* 30**,** 5382-90.

LEGAULT, P. & PARDI, A. 1994. 31P chemical shift as a probe of structural motifs in RNA. *J Magn Reson B,* 103**,** 82-6.

LEKOMTSEV, S., KOLOSOV, P., BIDOU, L., FROLOVA, L., ROUSSET, J. P. & KISSELEV, L. 2007. Different modes of stop codon restriction by the Stylonychia and Paramecium eRF1 translation termination factors. *Proc Natl Acad Sci U S A,* 104**,** 10824-9.

LIANG, H., WONG, J. Y., BAO, Q., CAVALCANTI, A. R. & LANDWEBER, L. F. 2005. Decoding the decoding region: analysis of eukaryotic release factor (eRF1) stop codon-binding residues. *J Mol Evol,* 60**,** 337-44.

LIN, H. N., WU, K. P., CHANG, J. M., SUNG, T. Y. & HSU, W. L. 2005. GANA--a genetic algorithm for NMR backbone resonance assignment. *Nucleic Acids Res,* 33**,** 4593-601.

LOZUPONE, C. A., KNIGHT, R. D. & LANDWEBER, L. F. 2001. The molecular basis of nuclear genetic code change in ciliates. *Curr Biol,* 11**,** 65-74.

LUAN, T., JARAVINE, V., YEE, A., ARROWSMITH, C. H. & OREKHOV, V. Y. 2005. Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. *J Biomol NMR,* 33**,** 1-14.

LYNCH, S. R., GONZALEZ, R. L. & PUGLISI, J. D. 2003. Comparison of X-ray crystal structure of the 30S subunit-antibiotic complex with NMR structure of decoding site oligonucleotide-paromomycin complex. *Structure,* 11**,** 43-53.

LYNCH, S. R. & PUGLISI, J. D. 2001a. Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *J Mol Biol,* 306**,** 1037-58.

LYNCH, S. R. & PUGLISI, J. D. 2001b. Structure of a eukaryotic decoding region A-site RNA. *J Mol Biol,* 306**,** 1023-35.

MANDELSHTAM, V. A. 2000. The multidimensional filter diagonalization method. *J Magn Reson,* 144**,** 343-56.

MANDELSHTAM, V. A., TAYLOR, H. S. & SHAKA, A. J. 1998. Application of the filter diagonalization method to one- and two-dimensional NMR spectra. *J Magn Reson,* 133**,** 304-12.

MANTSYZOV, A. B., IVANOVA, E. V., BIRDSALL, B., ALKALAEVA, E. Z., KRYUCHKOVA, P. N., KELLY, G., FROLOVA, L. Y. & POLSHAKOV, V. I. 2010. NMR solution structure and function of the C-terminal domain of eukaryotic class 1 polypeptide chain release factor. *FEBS J,* 277**,** 2611-27.

MANTSYZOV, A. B., IVANOVA, E. V., BIRDSALL, B., KOLOSOV, P. M., KISSELEV, L. L. & POLSHAKOV, V. I. 2007. NMR assignments of the C-terminal domain of human polypeptide release factor eRF1. *Biomol NMR Assign,* 1**,** 183-5.

MARION, D. 2005. Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. *J Biomol NMR,* 32**,** 141-50.

MASSE, J. E. & KELLER, R. 2005. AutoLink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. *J Magn Reson,* 174**,** 133-51.

MASSE, J. E., KELLER, R. & PERVUSHIN, K. 2006. SideLink: automated side-chain assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. *J Magn Reson,* 181**,** 45-67.

MELNIKOV, S., BEN-SHEM, A., GARREAU DE LOUBRESSE, N., JENNER, L., YUSUPOVA, G. & YUSUPOV, M. 2012. One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol,* 19**,** 560-7.

MISHKOVSKY, M., KUPCE, E. & FRYDMAN, L. 2007. Ultrafast-based projection-reconstruction three-dimensional nuclear magnetic resonance spectroscopy. *J Chem Phys,* 127**,** 034507.

MITKEVICH, V. A., KONONENKO, A. V., PETRUSHANKO, I. Y., YANVAREV, D. V., MAKAROV, A. A. & KISSELEV, L. L. 2006. Termination of translation in eukaryotes is mediated by the quaternary eRF1*eRF3*GTP*Mg2+ complex. The biological roles of eRF3 and prokaryotic RF3 are profoundly distinct. *Nucleic Acids Res,* 34**,** 3947-54.

MOAZED, D. & NOLLER, H. F. 1987. Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature,* 327**,** 389-94.

MONKENBUSCH, M., RICHTER, D. & BIEHL, R. 2010. Observation of protein domain motions by neutron spectroscopy. *Chemphyschem,* 11**,** 1188-94.

MOROZOVA, N., ALLERS, J., MYERS, J. & SHAMOO, Y. 2006. Protein-RNA interactions: exploring binding patterns with a three-dimensional superposition analysis of high resolution structures. *Bioinformatics,* 22**,** 2746-52.

MURAMATSU, T., HECKMANN, K., KITANAKA, C. & KUCHINO, Y. 2001. Molecular mechanism of stop codon recognition by eRF1: a wobble hypothesis for peptide anticodons. *FEBS Lett,* 488**,** 105-9.

NAKAMURA, Y., ITO, K. & EHRENBERG, M. 2000. Mimicry grasps reality in translation termination. *Cell,* 101**,** 349-52.

NILGES, M., MACIAS, M. J., O'DONOGHUE, S. I. & OSCHKINAT, H. 1997. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol,* 269**,** 408-22.

O'BRIEN, T. W. 2002. Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene,* 286**,** 73-9.

ODA, Y., MURAMATSU, T., YUMOTO, F., ITO, M. & TANOKURA, M. 2004. Backbone (1)H, (13)C and (15)N resonance assignment of the N-terminal domain of human eRF1. *J Biomol NMR,* 30**,** 109-10.

OGLE, J. M., BRODERSEN, D. E., CLEMONS, W. M., JR., TARRY, M. J., CARTER, A. P. & RAMAKRISHNAN, V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science,* 292**,** 897-902.

OGLE, J. M., MURPHY, F. V., TARRY, M. J. & RAMAKRISHNAN, V. 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell,* 111**,** 721-32.

OGLE, J. M. & RAMAKRISHNAN, V. 2005. Structural insights into translational fidelity. *Annu Rev Biochem,* 74**,** 129-77.

OREKHOV, V. Y., IBRAGHIMOV, I. & BILLETER, M. 2003. Optimizing resolution in multidimensional NMR by three-way decomposition. *J Biomol NMR,* 27**,** 165-73.

OREKHOV, V. Y., IBRAGHIMOV, I. V. & BILLETER, M. 2001. MUNIN: a new approach to multi-dimensional NMR spectra interpretation. *J Biomol NMR,* 20**,** 49-60.

PALADE, G. E. 1955. A small particulate component of the cytoplasm. *J Biophys Biochem Cytol,* 1**,** 59-68.

PALMER, E., WILHELM, J. M. & SHERMAN, F. 1979. Phenotypic suppression of nonsense mutants in yeast by aminoglycoside antibiotics. *Nature,* 277**,** 148-50.

PENNISI, E. 1999. The race to the ribosome structure. *Science,* 285**,** 2048-51.

PERVUSHIN, K., VOGELI, B. & ELETSKY, A. 2002. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *J Am Chem Soc,* 124**,** 12898-902.

POPENDA, L., ADAMIAK, R. W. & GDANIEC, Z. 2008. Bulged adenosine influence on the RNA duplex conformation in solution. *Biochemistry,* 47**,** 5059-67.

POPENDA, M., BIALA, E., MILECKI, J. & ADAMIAK, R. W. 1997. Solution structure of RNA duplexes containing alternating CG base pairs: NMR study of r(CGCGCG)2 and 2'-O-Me(CGCGCG)2 under low salt conditions. *Nucleic Acids Res,* 25**,** 4589-98.

PRISTOVSEK, P., RUTERJANS, H. & JERALA, R. 2002. Semiautomatic sequence-specific assignment of proteins based on the tertiary structure--the program st2nmr. *J Comput Chem,* 23**,** 335-40.

RABL, J., LEIBUNDGUT, M., ATAIDE, S. F., HAAG, A. & BAN, N. 2011. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science,* 331**,** 730-6.

RECHT, M. I., DOUTHWAITE, S. & PUGLISI, J. D. 1999. Basis for prokaryotic specificity of action of aminoglycoside antibiotics. *EMBO J,* 18**,** 3133-8.

RODNINA, M. V. & WINTERMEYER, W. 2001. Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu Rev Biochem,* 70**,** 415-35.

RODNINA, M. V. & WINTERMEYER, W. 2009. Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol,* 21**,** 435-43.

ROSPERT, S., RAKWALSKA, M. & DUBAQUIE, Y. 2005. Polypeptide chain termination and stop codon readthrough on eukaryotic ribosomes. *Rev Physiol Biochem Pharmacol,* 155**,** 1-30.

ROVNYAK, D., FRUEH, D. P., SASTRY, M., SUN, Z. Y., STERN, A. S., HOCH, J. C. & WAGNER, G. 2004. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson,* 170**,** 15-21.

ROWE, S. M. & CLANCY, J. P. 2009. Pharmaceuticals targeting nonsense mutations in genetic diseases: progress in development. *BioDrugs,* 23**,** 165-74.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* 5**,** 725-38.

SAITO, K., KOBAYASHI, K., WADA, M., KIKUNO, I., TAKUSAGAWA, A., MOCHIZUKI, M., UCHIUMI, T., ISHITANI, R., NUREKI, O. & ITO, K. 2010. Omnipotent role of archaeal elongation factor 1 alpha (EF1alpha in translational elongation and termination, and quality control of protein synthesis. *Proc Natl Acad Sci U S A,* 107**,** 19242-7.

SALAS-MARCO, J. & BEDWELL, D. M. 2005. Discrimination between defects in elongation fidelity and termination efficiency provides mechanistic insights into translational readthrough. *J Mol Biol,* 348**,** 801-15.

SALAS-MARCO, J., FAN-MINOGUE, H., KALLMEYER, A. K., KLOBUTCHER, L. A., FARABAUGH, P. J. & BEDWELL, D. M. 2006. Distinct paths to stop codon reassignment by the variant-code organisms Tetrahymena and Euplotes. *Mol Cell Biol,* 26**,** 438-47.

SCHANDA, P., KUPCE, E. & BRUTSCHER, B. 2005. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR,* 33**,** 199-211.

SCHLUENZEN, F., TOCILJ, A., ZARIVACH, R., HARMS, J., GLUEHMANN, M., JANELL, D., BASHAN, A., BARTELS, H., AGMON, I., FRANCESCHI, F. & YONATH, A. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell,* 102**,** 615-23.

SCHMEING, T. M. & RAMAKRISHNAN, V. 2009. What recent ribosome structures have revealed about the mechanism of translation. *Nature,* 461**,** 1234-42.

SCHMEING, T. M., VOORHEES, R. M., KELLEY, A. C., GAO, Y. G., MURPHY, F. V. T., WEIR, J. R. & RAMAKRISHNAN, V. 2009. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science,* 326**,** 688-94.

SCHUWIRTH, B. S., BOROVINSKAYA, M. A., HAU, C. W., ZHANG, W., VILA-SANJURJO, A., HOLTON, J. M. & CATE, J. H. 2005. Structures of the bacterial ribosome at 3.5 A resolution. *Science,* 310**,** 827-34.

SCOLNICK, E., TOMPKINS, R., CASKEY, T. & NIRENBERG, M. 1968. Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A,* 61**,** 768-74.

SEIT-NEBI, A., FROLOVA, L. & KISSELEV, L. 2002. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep,* 3**,** 881-6.

SELMER, M., DUNHAM, C. M., MURPHY, F. V. T., WEIXLBAUMER, A., PETRY, S., KELLEY, A. C., WEIR, J. R. & RAMAKRISHNAN, V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science,* 313**,** 1935-42.

SENTHAMARAI, R. R., KUPROV, I. & PERVUSHIN, K. 2010. Benchmarking NMR experiments: a relational database of protein pulse sequences. *J Magn Reson,* 203**,** 129-37.

SETTE, M., SPURIO, R., VAN TILBORG, P., GUALERZI, C. O. & BOELENS, R. 1999. Identification of the ribosome binding sites of translation initiation factor IF3 by multidimensional heteronuclear NMR spectroscopy. *RNA,* 5**,** 82-92.

SHAPIRA, B. & PRESTEGARD, J. H. 2010. Electron-nuclear interactions as probes of domain motion in proteins. *J Chem Phys,* 132**,** 115102.

SHAW, J. J. & GREEN, R. 2007. Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol Cell,* 28**,** 458-67.

SHIN, D. H., BRANDSEN, J., JANCARIK, J., YOKOTA, H., KIM, R. & KIM, S. H. 2004. Structural analyses of peptide release factor 1 from Thermotoga maritima reveal domain flexibility required for its interaction with the ribosome. *J Mol Biol,* 341**,** 227-39.

SHOEMAKER, C. J. & GREEN, R. 2011. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A,* 108**,** E1392-8.

SIEVERS, A., BERINGER, M., RODNINA, M. V. & WOLFENDEN, R. 2004. The ribosome as an entropy trap. *Proc Natl Acad Sci U S A,* 101**,** 7897-901.

SINGH, A., URSIC, D. & DAVIES, J. 1979. Phenotypic suppression and misreading Saccharomyces cerevisiae. *Nature,* 277**,** 146-8.

SNYDER, D. A., XU, Y., YANG, D. & BRUSCHWEILER, R. 2007a. Resolution-enhanced 4D 15N/13C NOESY protein NMR spectroscopy by application of the covariance transform. *J Am Chem Soc,* 129**,** 14126-7.

SNYDER, D. A., ZHANG, F. & BRUSCHWEILER, R. 2007b. Covariance NMR in higher dimensions: application to 4D NOESY spectroscopy of proteins. *J Biomol NMR,* 39**,** 165-75.

SONENBERG, N. & HINNEBUSCH, A. G. 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell,* 136**,** 731-45.

SONG, H., MUGNIER, P., DAS, A. K., WEBB, H. M., EVANS, D. R., TUITE, M. F., HEMMINGS, B. A. & BARFORD, D. 2000. The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell,* 100**,** 311-21.

SUND, J., ANDER, M. & AQVIST, J. 2010. Principles of stop-codon reading on the ribosome. *Nature,* 465**,** 947-50.

TAKEDA, M., IKEYA, T., GUNTERT, P. & KAINOSHO, M. 2007. Automated structure determination of proteins with the SAIL-FLYA NMR method. *Nat Protoc,* 2**,** 2896-902.

TAYLOR, D., UNBEHAUN, A., LI, W., DAS, S., LEI, J., LIAO, H. Y., GRASSUCCI, R. A., PESTOVA, T. V. & FRANK, J. 2012. Cryo-EM structure of the mammalian eukaryotic release factor eRF1-eRF3-associated termination complex. *Proc Natl Acad Sci U S A,* 109**,** 18413-8.

TIAN, F., VALAFAR, H. & PRESTEGARD, J. H. 2001. A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *J Am Chem Soc,* 123**,** 11791-6.

TUGARINOV, V., KAY, L. E., IBRAGHIMOV, I. & OREKHOV, V. Y. 2005. High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc,* 127**,** 2767-75.

VALLE, M., SENGUPTA, J., SWAMI, N. K., GRASSUCCI, R. A., BURKHARDT, N., NIERHAUS, K. H., AGRAWAL, R. K. & FRANK, J. 2002. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J,* 21**,** 3557-67.

VAN HOOF, A. & WAGNER, E. J. 2011. A brief survey of mRNA surveillance. *Trends Biochem Sci,* 36**,** 585-92.

VELICHUTINA, I. V., HONG, J. Y., MESECAR, A. D., CHERNOFF, Y. O. & LIEBMAN, S. W. 2001. Genetic interaction between yeast Saccharomyces cerevisiae release factors and the decoding region of 18 S rRNA. *J Mol Biol,* 305**,** 715-27.

VESTERGAARD, B., VAN, L. B., ANDERSEN, G. R., NYBORG, J., BUCKINGHAM, R. H. & KJELDGAARD, M. 2001. Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol Cell,* 8**,** 1375-82.

VOORHEES, R. M., SCHMEING, T. M., KELLEY, A. C. & RAMAKRISHNAN, V. 2010. The mechanism for activation of GTP hydrolysis on the ribosome. *Science,* 330**,** 835-8.

WANG, Y., CHAI, B., WANG, W. & LIANG, A. 2010. Functional characterization of polypeptide release factor 1b in the ciliate Euplotes. *Biosci Rep,* 30**,** 425-31.

WEIXLBAUMER, A., JIN, H., NEUBAUER, C., VOORHEES, R. M., PETRY, S., KELLEY, A. C. & RAMAKRISHNAN, V. 2008. Insights into translational termination from the structure of RF2 bound to the ribosome. *Science,* 322**,** 953-6.

WELCH, E. M., BARTON, E. R., ZHUO, J., TOMIZAWA, Y., FRIESEN, W. J., TRIFILLIS, P., PAUSHKIN, S., PATEL, M., TROTTA, C. R., HWANG, S., WILDE, R. G., KARP, G., TAKASUGI, J., CHEN, G., JONES, S., REN, H., MOON, Y. C., CORSON, D., TURPOFF, A. A., CAMPBELL, J. A., CONN, M. M., KHAN, A., ALMSTEAD, N. G., HEDRICK, J., MOLLIN, A., RISHER, N., WEETALL, M., YEH, S., BRANSTROM, A. A., COLACINO, J. M., BABIAK, J., JU, W. D., HIRAWAT, S., NORTHCUTT, V. J., MILLER, L. L., SPATRICK, P., HE, F., KAWANA, M., FENG, H., JACOBSON, A., PELTZ, S. W. & SWEENEY, H. L. 2007. PTC124 targets genetic disorders caused by nonsense mutations. *Nature,* 447**,** 87-91.

WIMBERLY, B. T., BRODERSEN, D. E., CLEMONS, W. M., JR., MORGAN-WARREN, R. J., CARTER, A. P., VONRHEIN, C., HARTSCH, T. & RAMAKRISHNAN, V. 2000. Structure of the 30S ribosomal subunit. *Nature,* 407**,** 327-39.

WU, K. P., CHANG, J. M., CHEN, J. B., CHANG, C. F., WU, W. J., HUANG, T. H., SUNG, T. Y. & HSU, W. L. 2006. RIBRA--an error-tolerant algorithm for the NMR backbone assignment problem. *J Comput Biol,* 13**,** 229-44.

YABUKI, T., KIGAWA, T., DOHMAE, N., TAKIO, K., TERADA, T., ITO, Y., LAUE, E. D., COOPER, J. A., KAINOSHO, M. & YOKOYAMA, S. 1998. Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J Biomol NMR,* 11**,** 295-306.

YUSUPOV, M. M., YUSUPOVA, G. Z., BAUCOM, A., LIEBERMAN, K., EARNEST, T. N., CATE, J. H. & NOLLER, H. F. 2001. Crystal structure of the ribosome at 5.5 A resolution. *Science,* 292**,** 883-96.

ZAHER, H. S. & GREEN, R. 2009. Fidelity at the molecular level: lessons from protein synthesis. *Cell,* 136**,** 746-62.

ZAVIALOV, A. V., BUCKINGHAM, R. H. & EHRENBERG, M. 2001. A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell,* 107**,** 115-24.

ZHANG, F. & BRUSCHWEILER, R. 2004. Indirect covariance NMR spectroscopy. *J Am Chem Soc,* 126**,** 13180-1.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9**,** 40.

ZHANG, Y. & SKOLNICK, J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins,* 57**,** 702-10.

ZHANG, Y. & SKOLNICK, J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res,* 33**,** 2302-9.

ZHOURAVLEVA, G., FROLOVA, L., LE GOFF, X., LE GUELLEC, R., INGE-VECHTOMOV, S., KISSELEV, L. & PHILIPPE, M. 1995. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J,* 14**,** 4065-72.

ZILBERBERG, A., LAHAV, L. & ROSIN-ARBESFELD, R. 2010. Restoration of APC gene function in colorectal cancer cells by aminoglycoside- and macrolide-induced read-through of premature termination codons. *Gut,* 59**,** 496-507.

ZWECKSTETTER, M. 2008. NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc,* 3**,** 679-90.

Appendix A: *Psyte*'s Algorithm

As with our previous programs, *AutoLink* (Masse and Keller, 2005) and *SideLink* (Masse et al., 2006), the *Psyte*'s algorithm is designed to do human-like reasoning in order to achieve its goal. The spectrum analysis performed by *Psyte* can be divided into two main sections (Figure I). Section 1 deals with the analysis of individual spectrum. Section 2 handles integration of the results from individual spectrum into coherent spin systems. Each section can be further divided into two main stages.

In section 1, the program focuses on a single spectrum. In the initial stage of section 1, the program identifies probable HSQC-dimension peaks by first scanning the spectrum for all local maxima and minima (if negative peaks are expected to be present), and subsequently reducing the peaks to projections by deleting the cross-peak frequencies. Projection peaks that are very close to each others are combined so that multiple cross-peaks from the same spin system do not contribute to extra spin systems. In the next stage of section 1, a high-resolution competition-based non-monotonic fitting algorithm is used to identify cross-peaks associated with each HSQC projection peak and also to more accurately determine the chemical shifts of the HSQC projection peaks.

Various non-monotonic artifact/noise filters are interpolated throughout both stages, in order to reduce the number of incorrectly identified spin systems. At this stage, however, there is no absolute restriction as to how many cross-peaks may be contained within each spin system, but rather this information is only used to help identify artifacts and noise. A more stringent restriction is implemented coordinately in section 2, during which the analysis of the individual spectrum is combined with the results from the analysis of other spectra.

In the first stage of section 2, each of the spin systems obtained from section 1 is deconstructed into individual peaks, which are then re-merged into spin systems. The target spin systems may emerge from other spectra or from the same spectrum, if more than one cross-peak per spin system is possible. This merging process is implemented as relative hypothesis prioritization (RHP) simulated non-monotonic logic, which is similar to the mechanisms *AutoLink* and *SideLink* employ for resonance assignment. Since the input to the RHP-logic is expected to contain many extra peaks (*i.e.* from noise and artifacts), it is common for the merged spin systems to contain errors. For this reason, *Psyte* employs a post-RHP recombination algorithm that utilizes the results of the RHP-logic to construct all reasonable combinations of peaks to form extra spin systems.

In the next stage of section 2, a complex series of non-monotonic correction algorithms analyze the resulting spin systems and make relatively small corrections in order to achieve a high degree of accuracy in spin system determination. In cases where the ambiguity of the peak grouping is high, *Psyte* reports all reasonably possible peak combinations as separate spin systems, of which to be sorted out during the residue specific assignment process by *AutoLink II*. Thus, the uncertainty in the spectrum analysis is effectively propagated into the resonance assignment process, preventing the accumulation of errors due to decisions made with too weak support.

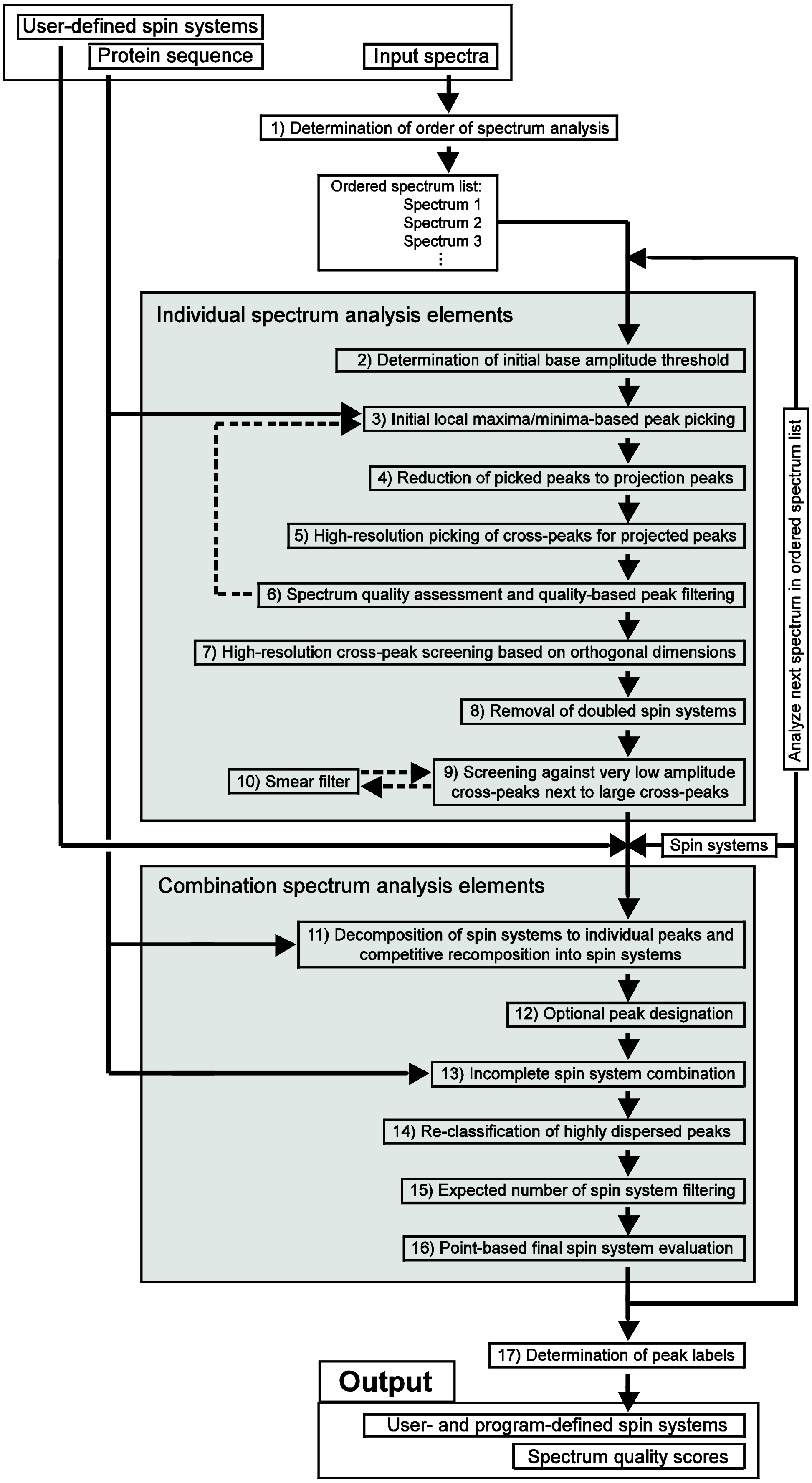

User-defined spin systems
Protein sequence
Input spectra
1) Determination of order of spectrum analysis
Ordered spectrum list:
Spectrum 1
Spectrum 2
Spectrum 3
⋮
Individual spectrum analysis elements
2) Determination of initial base amplitude threshold
3) Initial local maxima/minima-based peak picking
4) Reduction of picked peaks to projection peaks
5) High-resolution picking of cross-peaks for projected peaks
6) Spectrum quality assessment and quality-based peak filtering
7) High-resolution cross-peak screening based on orthogonal dimensions
8) Removal of doubled spin systems
10) Smear filter
9) Screening against very low amplitude cross-peaks next to large cross-peaks
Analyze next spectrum in ordered spectrum list
Spin systems
Combination spectrum analysis elements
11) Decomposition of spin systems to individual peaks and competitive recomposition into spin systems
12) Optional peak designation
13) Incomplete spin system combination
14) Re-classification of highly dispersed peaks
15) Expected number of spin system filtering
16) Point-based final spin system evaluation
17) Determination of peak labels
Output
User- and program-defined spin systems
Spectrum quality scores

Figure I Schematic diagram of the algorithm implemented in *Psyte* (courtesy of Dr. James Masse). Input consists of the spectra to be analyzed, the protein sequence, and optional user-defined spin systems. Output consists of spin systems and the program’s ratings of the input spectra. Spectral analysis proceeds in two main stages. The first stage (upper box) is primarily concerned with analysis of individual spectrum. The second stage (lower box) focuses on reconciling the results of the analysis of other spectra and the user’s optional spin system input.

Appendix B: HSQC Spectra of wt N-domain and $Q^{122}FM(Y)F^{126}$

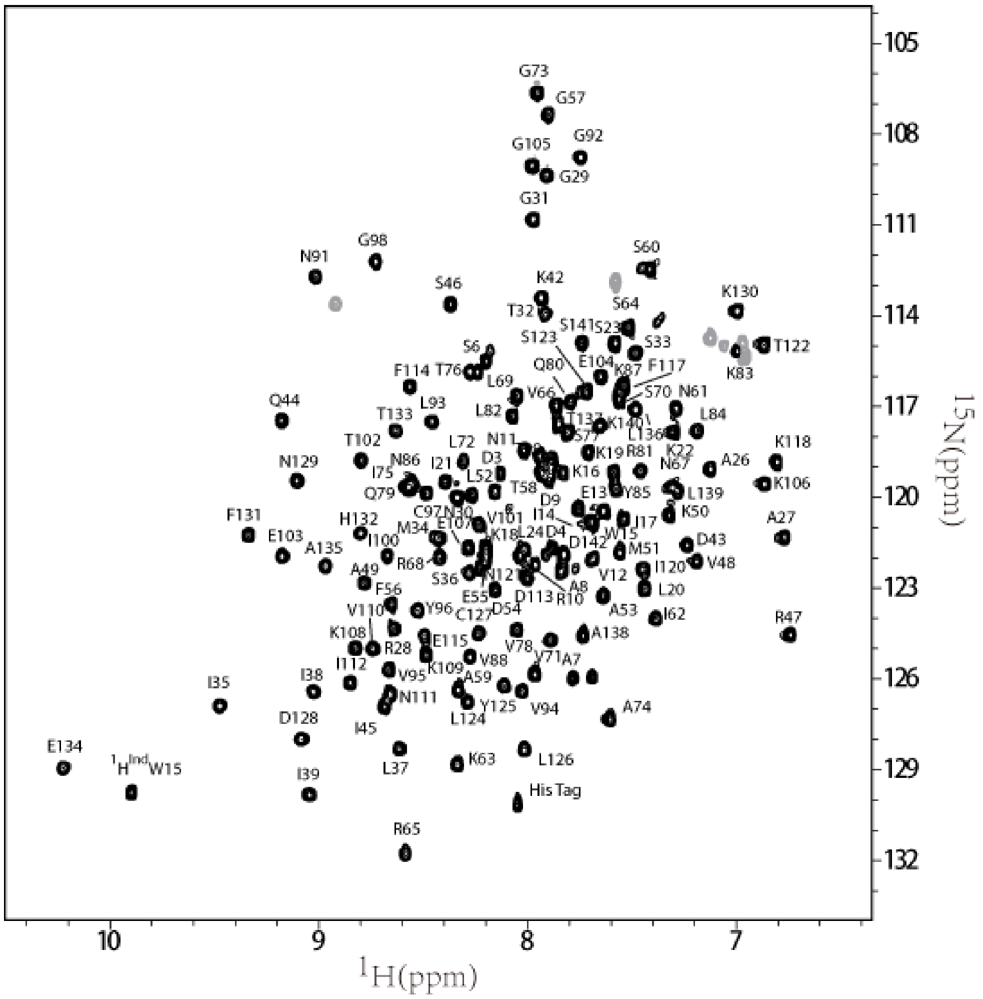


Figure II $^1H,^{15}N$-TROSY-HSQC of wt-N-domain at 298 K showing its residue-specific assignment.

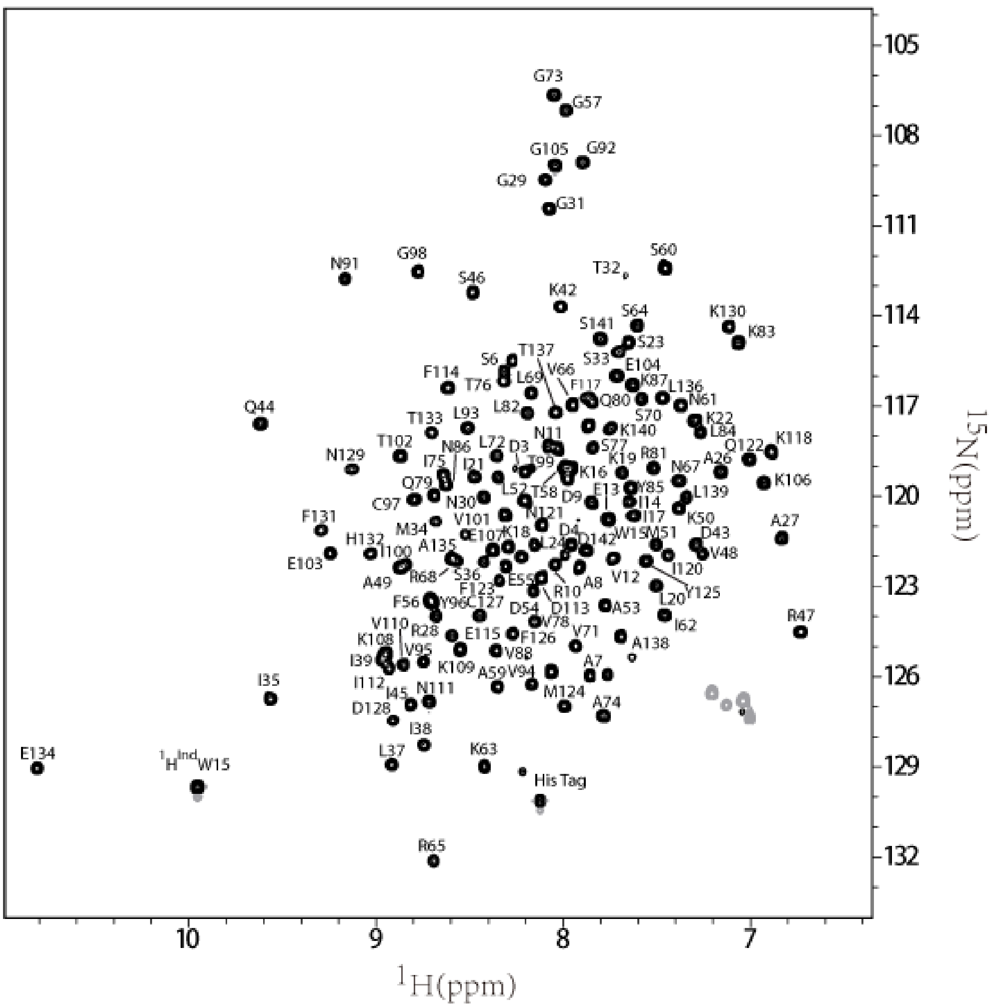


Figure III ^{1}H,^{15}N-TROSY-HSQC of Q^{122}FM(Y)F^{126} at 298 K showing its residue-specific assignment.