Folding during ribosomal biosynthesis of yellow fluorescent protein: Investigations by fluorescence spectroscopy and NMR spectroscopy

Toshitaka Tajima :田島 俊孝

Department of Structural and Molecular Biology
University College London

A Thesis submitted for the degree of
Doctor of Philosophy

31st July 2014

Acknowledgement

I would like to express all my best gratitude to: Doctor Ramadan Abuknesha, Doctor Kenneth Deans Bruce, Professor Richard Cammack, Doctor Yu Wai Chen, Professor John Christodoulou, Professor Snezana Djordjevic, Professor Mahmoud Halablab, Doctor D. Fleming Hansen, Professor Robert Hider, Professor Masahiko Okuni, Professor Jonathan G. F. Powell, Professor Peter Quinn, Professor Surjit Kaila Singh Srai, Professor Brian John Sutton, Professor Michele Vendruscolo, Professor Gabriel Waksman, Doctor Tamisuke Watanuki, Professor Masaki Yanagishita, Doctor Miki Yokoyama, all my dearest friends and families, and finally Dori and my mother Doctor Yoko Cecilia Tajima.

Declaration

I, Toshitaka Tajima, declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The work herein was carried out while I was a graduate student at University College London, Research department of Structural and Molecular Biology, under the supervisions of Professor John Christodoulou, Professor Snezana Djordjevic, and Professor Gabriel Waksman.

Thesis.

Antithesis.

Synthesis … (G.W.F. Hegel)

Thesis again? Oh dear!

Contents

Chapter 3 Result I - RNCs of YFP

Chapter 4 Result II - C-terminal truncations of YFP

Abbreviations

1D - 1 dimensional

2D - 2 dimensional

3D - 3 dimensional

4,4'-Dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt - bis-ANS

Calmodulin - CaM

Carr Purcell Meiboom Gill - CPMG

Circular dichroism - CD

Coomassie brilliant blue - CBB

Cryo-electron microscopy - cryoEM

Cyan fluorescent protein - CFP

Cystic fibrosis transmembrane conductance regulator - CFTR

Distilled and deionised water - ddH_2O

Dithiothreitol - DTT

Domain 5 of the gelation factor from *Dictyostelium discoideum* - Dom5

Escherichia coli - *E.coli*

Ethidium bromide - EtBr

Exchange spectroscopy - EXSY

Excited state proton transfer - ESPT

Fast protein liquid chromatography - FPLC

Fluorescence resonance energy transfer - FRET

Fluorescent protein - FP

Fourier transform infrared - FT-IR

Green fluorescent protein - GFP

Hetero-nuclear Multiple Quantum Coherence - HMQC

Hetero-nuclear single quantum coherence - HSQC

Isopropyl β-D-1-thiogalactopyranoside - IPTG

Luminol enhancer and Peroxide - ELT

Luria Bertani - LB

Messenger RNA - mRNA

Methionine aminopeptidase - MAP

Molecular dynamics - MD

Multi-wavelength dispersion - MAD

Nascent chain - NC

Nuclear magnetic resonance - NMR

Nuclear Overhauser Effect - NOE

Peptidyl deformylase - PDF

Peptidyl transferase centre - PTC

Pulse field gradient - PFG

Red fluorescent protein - RFP

Residual dipolar coupling - RDC

Ribosome nascent chain complex - RNC

Secretory monitoring protein - SecM

Signal recognition particle - SRP

Sodium dodecyl sulfate - SDS

Stimulated echo - STE

Tetramethylethylenediamine - TEMED

Transfer RNA - tRNA

Ultra violet - UV

Wild type - wt

Yellow fluorescent protein - YFP

Abstract

During biosynthesis on the ribosome, the nascent polypeptide chain emerges in a vectorial fashion, one amino acid at a time and it has its first opportunity to sample conformational space. Study of structural and molecular biology has shown that nascent chains (NC) can co-translationally acquire its native structure, however how an emerging NC forms the native conformation is a fundamental question in modern biology. Yellow fluorescent protein (YFP), a red shifted yellow emission variant of green fluorescent protein (GFP) that is consisted of 238 amino acid residues is highly distinct because of its fast maturation of fluorophore and its rapid folding kinetics. In addition, YFP has been extensively studied using X-ray crystallography, nuclear magnetic resonance (NMR), and biophysics, which make it an ideal system for examining protein folding on the ribosome. In this study, therefore, we created a series of ribosome-nascent chain complex (RNC) of YFP to simulate the progressive emergence of NC. We also developed an *in vitro* transcription translation system to generate isotopically-labelled RNCs for both biophysics and NMR spectroscopy. To complement the RNC study, we also created a co-translational folding mimetic using C-terminal truncations of isolated YFP to evaluate folding *in vitro*. Using these approaches, the co-translational folding of YFP was observed using both spectroscopies of fluorescence and NMR at a residue specific level. Here we demonstrate that native folding can take place very close to the ribosome in which YFP is separated by 19 amino acids from the ribosome's peptidyl transferase centre (PTC), and that YFP can tolerate the absence of up to 14 amino acids from its C-terminus and still acquire a native fold.

Chapter 1

Introduction

1.1 Studies of protein folding and misfolding

1.1.1 Protein folding and misfolding

Proteins are the 'elastic and flexible workhorses' of the cell that start to perform their vital functions once they have achieved their biologically active fold. The human body produces over 50,000 unique proteins and feature within an elaborate network of interactions within the cell. Within the cell, therefore, the production, activity and turnover of proteins is carefully regulated, and is typically referred to as protein homeostasis or 'proteostasis'. Central to proteostasis, therefore, is the capacity of proteins to efficiently acquire their native fold in a timely manner.

In 1973, Anfinsen demonstrated that a protein's amino acid sequence has the information required to drive the protein to adopt its biologically-active fold, which is typically its lowest-energy state [1]. Following this ground-breaking demonstration, protein folding studies have aimed at trying to examine how the arrangement of a linear chain of amino acids can form complex 3D structures of proteins of different topologies and that are involved in different biological roles. These studies typically involve the reversible denaturation and renaturation processes of isolated proteins and are typically studied *in vitro* and using biophysical and structural approaches, though a diverse array of studies also examine protein folding *in vivo*, by examining patterns of expression, and the analysis of folding rates under different environments and conditions [2], as

well as theoretical studies and *in silico* approaches using molecular dynamics (MD) simulations such as Monte Carlo simulations that explore distributions of energy states [3]. Together these studies aim to characterise the different states populated on a folding pathway including investigation of folding transition states [4] and their potential influence on the stability [5], as well as the formation of obligatory or non-obligatory intermediate states [6].

Over the years many models have been proposed for how proteins fold such as nucleation-condensation, the framework model and hydrophobic collapse [7], however the more modern view of folding is one of an energy landscape or 'folding funnel' (Figure1.1). In this model, polypeptide chains explore a range of conformations within a biased energy funnel leading to native structure, which is usually the most energetically limited and optimised state. During the folding process, the polypeptide chain may access intermediate states, often referred as 'easily' collapsed states, i.e. compact heterogeneous structures without persistent secondary or tertiary structure. These states are typically less stable than the native state, and can assist in the folding process, or under some circumstances can lead to kinetically trapped species, that are prone to misfolding leading with abnormal folding, which may reconfigure to thermodynamically stable ordered structures, some of which are may also result in aggregation as illustrated in Figure 1.1.

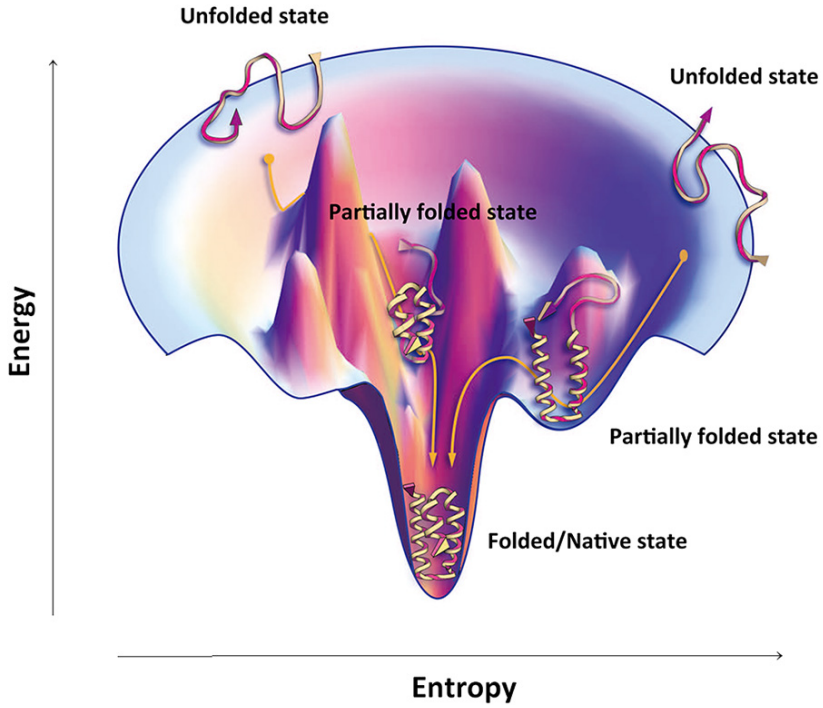


Figure 1.1 An image of protein folding process during *in vitro* re-folding on the corresponding free-energy landscape [8]. It is described that NCs confront a number of different stages during the biosynthesis, and that depending on the states of entropy and energy they change their conformational stage from unfolded/partially folded to folded native state.

A number of human diseases are known to be caused by abnormal protein misfolding that results in aggregation. A range of diseases including a number of neurodegenerative diseases such as Parkinson's, Alzheimer's, and Huntington's diseases are all associated with misfolded proteins such as α-synuclein, a-beta and huntingtin, respectively [9]. Other examples of diseases include the serpinopathies, which are implicated in liver cirrhosis, angioedema and emphysema [10], motor neurone disease [11] and cystic fibrosis [12]. In each of these cases, DNA mutations destabilise the protein and result in incorrect synthesis and/of folding and results in the production and accumulation of aggregates that overwhelm the cell's quality control pathways and lead to cellular

toxicity. It is very important, therefore, to understand during the synthesis how proteins acquire their biologically active folds, while avoiding the unfavourable events of misfolding.

1.2 The ribosome and the protein folding process

1.2.1 The ribosome and protein translation

Central to molecular biology is the dogma in which DNA is known to be transcribed to mRNA, and in which mRNA is used to for the translation of proteins, and within this process, the ribosome is critical in both the process of synthesis and folding of polypeptide chains. The structure and function of the ribosome has been studied by several high-resolution methods, e.g. X-ray crystallography [13] and cryo-electron microscopy (cryoEM) [14], and has been complemented by a plethora of biochemical studies. The structural dynamics of the translating ribosome remain the subject of intense research and more recently studies using cryoEM [15, 16], and NMR spectroscopy [17, 18, 19] have provided well-studied structural information about the ribosome and the NC.

The ribosome is a large macromolecular complex which consists of 58 different proteins and three large RNA molecules. In prokaryotes, ribosomal particles account for a large proportion (about 20%) of the cell mass. The ribosome consists of two subunits: in eukaryotes, these subunits, 40S and 60S [20], and in prokaryotes have 30S and 50S subunits [21] are found respectively. As mentioned, the key role of the 70S ribosomal complex is to translate the messenger RNA (mRNA) which is transcribed from the genetic DNA template via RNA polymerase [22] as illustrated in Figure 1.2.

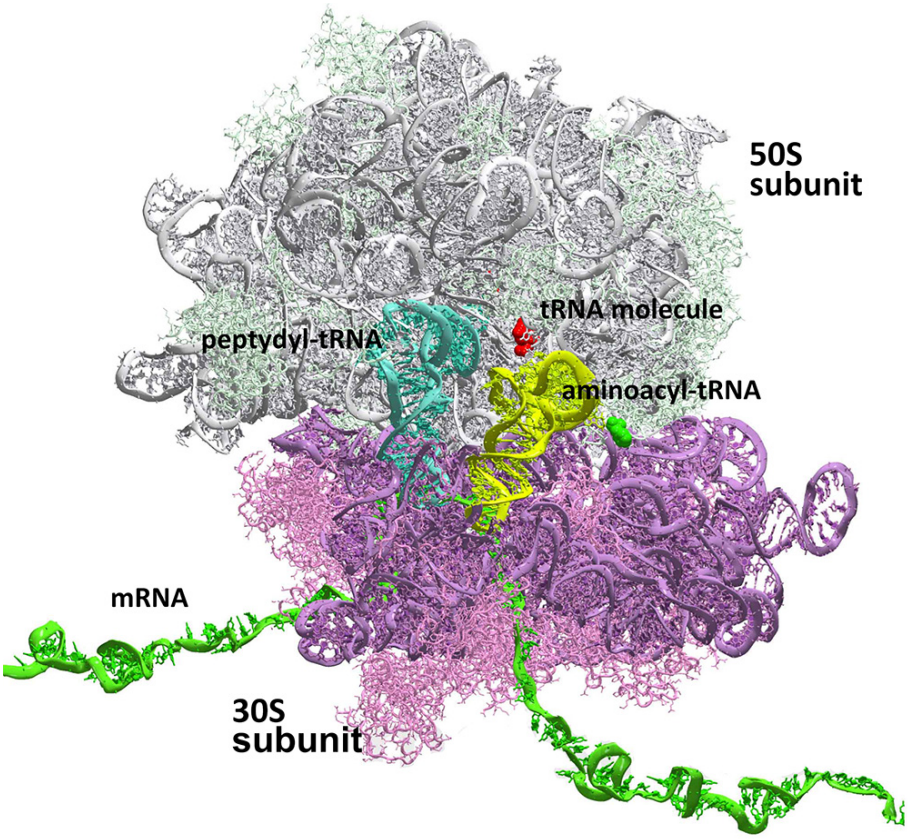


Figure 1.2 Representation of the complex of the ribosome (70S), **50S subunit** (colored gray), **30S subunit** (purple), **peptidyl-transfer RNA** (tRNA) (blue), **aminoacyl-tRNA** (yellow), **tRNA molecule** (red), **and translating mRNA** (green). One of the advantages to use RNCs is that interaction between 70S and targeting molecule can be observed as well as the folding process of NC and structural contribution of the ribosome. *Image quoted from the Sanbonmatsu team [23].

In *Escherichia coli* (*E.coli)*, the 30S subunit is in responsible for mRNA recognition and decoding and it is composed of one RNA molecule (16S, 1540 nucleotides) and 21 proteins [24]. On the other hand, the 50S subunit (composed of two RNA molecules of 120 (5S) and 2900 (23S) nucleotides and 34 proteins) is responsible for peptide bond formation. During translation a GTP-driven process, the decoding of mRNA requires the recruitment of amino-acylated tRNA in complex with elongation factor EF-Tu, by the ribosomal stalk protein, L7/L12 to the A-site of the PTC. A peptide bond is then formed within the PTC, and this rate of elongation is approximately about 15 - 22 peptide bonds formations per second [25]. Translocation of the ribosome then follows in which the ribosome moves relative to the mRNA, in which the tRNA present in the A-site moves to the P-site, and the deacylated tRNA moves to the E-site, which then exits the ribosome. The next mRNA triplet is then available within the A-site, which enables the next round of decoding and elongation to take place [26].

The 50S subunit also enables the safe passage of the nascent protein through the ribosomal tunnel, which is about 100Å long and between 10-20Å in width and largely formed by RNA and three ribosomal proteins, L4, L22, and L23 as illustrated in Figure 1.3. The exit tunnel is said to be hydrophilic to allow the unhindered path of a NC. Recent cryoEM studies also show that during the translation, that the NC remains largely extended within the exit tunnel. The NC also has the capacity to form at least an α-helical structure [27] in some regions of the exit tunnel, in particular in the so-called 'vestibule' which is approximately 80Å from the PTC, and which is the widest area of the tunnel c.a. 20Å. This is also supported by biophysical studies using FRET (fluorescence resonance energy transfer) [28] and cysteine-modification 'pegylation' experiments [29], which also show that the NC can adopt simple structures such as α-helices and hairpin, and which also show the full length of synthesised residues within

the exit tunnel [30].

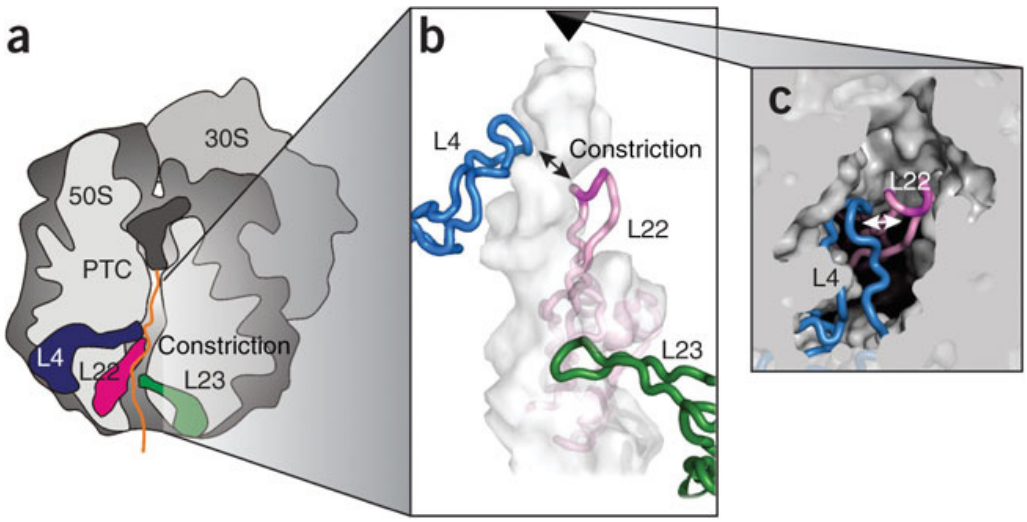


Figure 1.3 Anatomy of the exit tunnel of 70S ribosome. (a) Sliced image of the 70S ribosome consisted of 50S and 30S subunits, PTC site, and the tunnel: L4 (blue), L22 (magenta), and L23 (green) proteins which contribute folding during the translation. (b) Geography of the exit tunnel based on simulation with L4, L22, and L23. SecM stalling sequence that is commonly used to hold RNC state specifically make an interaction with L22 protein, and stalling surface of L22 is highlighted. (c) A view of exit tunnel with L4 and L22 proteins from PTC site [31].

1.2.2 Protein folding during biosynthesis: co-translational protein folding

Both structural and biophysical studies have shown that limited folding of the NC occurs within the ribosomal exit tunnel, and that the majority of folding takes place outside of the ribosome. Therefore it asks the question: what is the destiny of the NC as it begins to emerge from the exit tunnel? As was described in Section 1.2.1, the different states accessible for a protein in the folding pathway has been studied

extensively *in vitro* by denaturing and renaturing the entire polypeptide chain to understand its intrinsic propensity to occupy intermediate states.

Unlike protein folding *in vitro*, there are significant differences to the way in which proteins adopt structure *in vivo,* which takes place in a crowded cellular environment, c.a. 300mg/ml [32]. There are a number of biochemical, biophysical and structural studies which have demonstrated that during protein synthesis and as the NC emerges, it can begin to form structure while it is still attached to the ribosome. This process is known as co-translational folding. This process is said to enable efficient protein folding, as it re-shapes the energetic landscape otherwise experienced by an isolated protein, through the formation of early intermediate folded states involving N-terminus of the nascent polypeptide chain, which can serve as nuclei for the rest of the sequence to collapse [31].

Protein folding as it occurs on the ribosome, therefore, presents a very different environment for the emerging NC. The vectorial emergence of the NC is related to the rate of translation, c.a., 2 - 20 amino acids per second [25, 33], and which is influenced by the use of codons. For example, as has been shown for SufI, the substitution of codons for rarely-used condons dramatically alters the folding capacity of the protein [34]. In addition, the emerging NC can interact with auxiliary factors, such as molecular chaperones. The first molecular chaperone that can interact with a NC is trigger factor (TF), which binds at the base of the ribosomal exit tunnel to ribosomal protein L23. TF is a 47 kDa protein which is said to form a ‘crouching dragon’ structure, and which arches over the exit from the ribosome. The manner in which TF modulates NC folding are unclear, however two possible mechanisms include one in which TF forms a cage in

which the NC can fold, and alternatively, by forming direct interactions with the NC, as it threads through [35].

In addition to molecular chaperones, there are other auxiliary factors that bind at the ribosomal exit tunnel and compete for the emerging NCs. For example, methionine aminopeptidase (MAP) and peptidyl deformylase (PDF), are two enzymes which are involved in the removal of the N-terminal initiator methionine [36]. In addition, NCs with signal sequences can be targeting co-translationally to the endoplasmic reticulum (or the equivalent membranes in prokaryotes) by the action of the signal recognition particle (SRP). It is clear therefore that as the NC emerges from the ribosomal exit that it enters a very crowded and chaotic environment and the dynamic process of folding is a complex event.

1.2.3 Experimental evidence for co-translational protein folding

Initial biochemical studies of co-translational folding studies showed that NCs are able to acquire enzymatic activity and bind conformational antibodies [37]. One of the first demonstrations using biophysical studies to describe the co-translational folding firefly luciferase as a nascent polypeptide chain showed the sequence determination of the folding in this protein in which the N-terminal 190 residues shown to form, resulting in an increased rate of folding compared with the *in vitro* refolding from denaturant. A native-like intermediate state observed during co-translational folding was not observed when the released protein was denatured and refolded *in vitro*, which indicates that this intermediate is specified to emerge as the nascent polypeptide chains during co-translational folding [38]. In addition, these studies showed that the NC state of luciferase acquires its activity within the scale of seconds during the synthesis, but the

acquisition of the activity upon refolding process from a chemically denatured state is considerably slower as the scale of minutes [39].

Other fluorescence spectroscopy strategies have also been essential in illustrating the properties of co-translational folding. In particular, the use of FRET, for example, has shown that transmembrane bound sequences to sample α-helical secondary structural elements while confined to the ribosomal tunnel [40]. CryoEM observations of the translating ribosome tunnel strongly support this hypothesis [41]. It has been frequently described that folding of the N-terminal region occurs upon exit from the ribosomal tunnel [42]. A variety of studies have used biophysical methods to provide the evidence for co-translational folding upon the exit of the RNC.

1.2.4 Structural studies of co-translational folding

As NC is highly dynamic upon emergence from the exit tunnel, the observation of its structural properties have alluded electron density methods such as X-ray crystallography or cryoEM. In contrast, NMR spectroscopy, with the technique having the ability to provide information on highly dynamic system, has recently been shown to be capable of providing high-resolution structural and dynamical information of emerging NCs. The capacity to study NCs in this manner has been made possible through the ability to generate stable RNCs. To generate RNCs, the NC is arrested during translation and more recently, this has been achieved successfully, using a stalling sequence from the secretory monitoring [43] protein (SecM) in *E.coli*, which includes a sequence of 17 amino acids within which a pattern of nine residues interacts with the ribosome tunnel (more specifically with proteins L4, L22 and RNA 23S) [44]. In particular, this stalling strategy has been applied to an immunoglobulin domain

system (Ig2-RNC, composed of domain 5 and 6 of the gelation factor from *Dictyostelium discoideum,* Dom5*).* The NMR investigation of this RNC showed that the N-terminal domain of the RNC (Dom5) achieve its native fold during synthesis while still being tethered to the ribosome through the C-terminal domain of the NC [19] (For this YFP study, please see Section 3.1.1 in Result 1 Chapter for more details). More recently, these studies have been extended to examine Ig2-RNCs of different lengths, in order to examine the progressive folding of the Dom5 protein (Launay H*,* PhD thesis, 2011) can generate different length of polypeptide chains by adding extra sequences and it enables to observe native state of folding process.

The success in examining co-translational folding by NMR and cryo-EM have come from the ability to generate significant quantities of RNCs using both *in vivo* [18] and *in vitro* approaches [19]. Of these strategies, *in vitro* approaches using cell-free technology is very powerful as an effective way to produce the RNCs which enables rapid sample preparation and more control over the components within the reaction and offers the prospect of generating RNCs for a range of studies.

1.3 Cell-free protein synthesis

1.3.1 Coupled transcription-translation protein synthesis

Since Nirenberg and Mattheath opened the door in 1961 [45], cell-free protein synthesis has been a powerful tool for understanding how mRNA is translated by the ribosome into functional polypeptides. This has expanded to the ready availability of expression systems both from prokaryotic (*E.coli*) and eukaryotic (wheat-germ, rabbit reticulocyte lysate, yeasts) sources, and have expanded a technical renaissance with applications for the production of proteins for both functional and structural studies. Such significant improvements for over four decades have led to productivity of up to milligram quantities of targeted product expression [46].

Many proteins are difficult to express *in vivo* pose significant issues for structural studies, and include a variety of examples including toxic proteins, membrane proteins, complex assemblies and proteins that exhibit insolubility issues. A significant advantage of cell-free protein synthesis therefore, is that it offers the capacity to manipulate reaction conditions for such systems [47]. Indeed, significant progress has made within this technology, such that it is now possible to obtain proteins post-translational modifications, e.g. disulfide bonds, as well as integral membrane proteins, correctly imbedded into a lipid bilayer and which are notoriously difficult to generate using *in vivo* approaches. Another significant advantage of cell-free is the ready manipulation of the reaction components and conditions in a reaction which permits automation and miniaturisation; this has been applied fields of protein arrays, *in vitro* evolution and multiplexed real-time labelling, among others [46, 48, 49, 50]. More recently, progress in cell-free has also resulted in the development of the PURE system,

in which individual components of the reaction are added separately, which enables near complete control over the entire reaction [51].

Cell-free protein synthesis is possible as it has been shown that the components of the cell can be reconstituted *in vitro* and in which protein biosynthesis can occur in the absence of the confines of being contained within a cellular membrane. In its simplest form, translation using a DNA template to the resulting polypeptide can be accomplished using a crude lysate from any given organism which provides the essential components for translation, e.g. ribosomes, RNA polymerase, accessory enzymes, tRNA, and other factors in combination with exogenously added DNA to the mRNA template, amino acids and an energy supply. The *in vitro* translation scheme can be carried out in a 'coupled' manner (transcription and translation occurring within the reaction, using a DNA template) or in an uncoupled manner (only translation, using an mRNA template). In general, coupled systems result in higher protein yields, and are generally more efficient [46]. Unlike an uncoupled reaction, the coupled reaction is supplemented with additional NTPs and a highly processive RNA polymerase, such as those encoded by T7, T3 or SP6 bacteriophages. Use of DNA plasmid or PCR products as the templates has also resulted in the emergence of a variety of new applications for *in vitro* expression system, which, for example, can circumvent the need for cloning into a relevant vector. Cell-free systems can be used in either a 'batch' mode or a 'dialysis' mode; in the batch mode the reaction is within a single compartment, while in the dialysis mode, usually two compartments are used, which enables reaction by-products to be removed and energy sources to be regenerated more easily.

Indeed, any organism can potentially be used as a resource for the preparation of a cell-free protein expression system, although the most popular hosts are *E.coli*, wheat

germ and rabbit reticulocytes [52, 53]. The choice of the cell species depends on purpose and types of protein, but in general *E.coli* based systems provide higher yields and expression for structural studies. The protein yields of *E.coli* based systems range from a few micrograms to milligrams per reaction depending on the protein and the reaction format, for example batch and dialysis modes. Eukaryotic based systems, although having a reduced expression rate than the *E.coli* system, provide a better platform for functional studies particularly for post-translationally modified proteins. Protein yields of the rabbit reticulocyte lysate are typically in the microgram scale per reaction volume, while by comparison, the yield from a wheat germ extract system is typically two orders of magnitude greater [52, 54].

1.3.2 Cell-free and effective labelling for NMR spectroscopy

In some structural studies, the use of selectively labelled proteins is particularly important. For example, selenomethionine labelling is a frequently used technique by x-ray protein crystallographers to facilitate protein phasing by multi-wavelength dispersion (MAD). Similarly, NMR spectroscopy which is able to report atomic resolution information, requires isotopic labelling of amino acids. Specifically in the case of NMR, isotopic labelling, e.g. ^{1}H - ^{15}N and ^{1}H - ^{13}C, can be a costly procedure and in some instances, selective labelling of specific amino acids can be problematic *in vivo*, in which scrambling can take place. This is typically a result of isotopically labelled amino acids that are added to the growth media, which are used as precursors in amino acid metabolism and biosynthesis pathways within the cell. The distinct advantage of cell-free therefore, is that such pathways are not present and the addition of amino acids are user-defined, therefore selective labelling can occur in an extremely efficient manner. Cell-free systems can, therefore, be used in a high-throughput manner

that is not possible using *in vivo* approaches alone [55, 56, 57].

1.3.3 Further applications of cell-free

The ability to selectively manipulate the cell-free environment is enabling new applications to be developed. For example, RNA silencing is a rapidly growing field with, RNAi or siRNAs [58]. In addition cell-free technology is making a significant impact in protein folding studies. A fraction of proteins obtained by both *in vitro* and *in vivo* systems are relatively insoluble or frequently misfolded [59], however it has been shown that the addition of detergents or chaperones to the reaction mixture has been successful in overcoming these challenges [59]. Cell-free has also been instrumental in generating RNCs for structural studies, in particular a variety of different complexes have been generated for cryoEM to study folding, quality control mechanisms and translocation [60]. Also, a trial of hybrid systems composed of lysates from different energy sources and components, such as from *archaea*, e.g. *Pyrococcus furiosus*, may provide further understanding for folding context [61].

1.4 Fluorescence and fluorescent proteins for study of biochemistry

1.4.1 Physical basis of fluorescence

Fluorescence is a phenomenon which can be observed when an orbital electron of a molecule, the atom relaxes to its ground state and in doing so, can emit a photon of light after they are excited by a higher quantum state by types of energy, for example by X-rays, ultra violet (UV) or other visible spectra. When a molecule releases the excess amount of free energy as electromagnetic radiation, under the visible spectra the electromagnetic radiation shows fluorescent properties as an emission. An excitation peak for a molecule is described as the highest energy state that a molecule is excited by specific wavelength by an energy source and emission peak is described as the state that a molecule releases excess free energy after excited by the different types of source.

Fluorescence associated with a molecule can be described by four elements (i) Extinction coefficient, (ii) Quantum yield, (iii) Stoke's shift (iv) wavelength of emission:

i) Extinction coefficient

The mass attenuation coefficient is a measurement of how strongly a chemical species or substance absorbs or scatters light at a given wavelength, (per unit mass) [62]. The extinction coefficient is measured by following equation:

$$I = Ioe^{-\left(\frac{\pi}{p}\right)pl}$$

(equation 1.1)

Equation 1.1 Where I_0 is the original intensity of the beam, I is the intensity of the beam at distance into the substance, e is Euler's number, about 2.718, μ is the attenuation coefficient.)

A molecule or atom which holds a high extinction coefficient is more sensitive to than lower extinction coefficient molecule.

ii) Quantum yield (Φ)

The recovery yield of the fluorescence excited by a source is described by the quantum yield. If we assume that a molecule or atom absorbs 1 mol of energy source (e.g. a light source), the molecule releases 0.5 mol of fluorescence as an emission peak. In this case, the quantum yield can be estimated as 50% recovery : Φ= 0.5 (The maximum strength of the quantum yield (Φ) is 1.0 (100%)). A value of higher quantum yield describes the larger extract recovery rate from an energy source [62].

$$\Phi = \frac{\text{Number of photons emitted}}{\text{Number of photons absorbed}}$$

(equation 1.2)

Equation 1.2 Φ = Quantum yield, *mol.of photons can be calculated by; the strength of a light, number of vibration speed, time for irradiation, Planck's constant, and Avogadro constant)

iii) Stoke's shift

The Stoke's shift is the difference between positions of the band maxima of the absorption (excitation) and emission spectra of the same electronic transition. A large Stoke's shift means that there is a large frequency gap between 'excitation' and 'emission' states, so that large gap provides large energy scale [63], e.g. GFP absorbs at 395 nm and releases its intrinsic green fluorescence at 509 nm as emission, on the other hand YFP absorbs at 514 nm and releases intrinsic yellow fluorescence at 527 nm. Thus although GFP and YFP differ by 3 residues including a single mutation on the fluorophore, there is a large difference of the Stoke's shift between GFP and YFP [63].

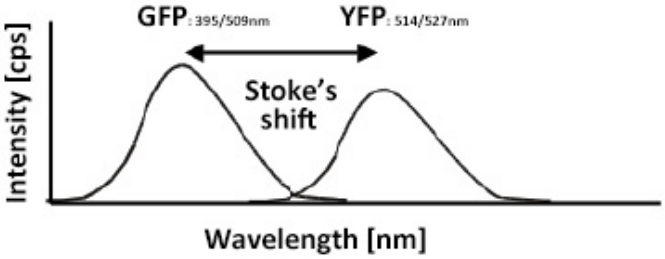


Figure 1.4 A typical Stoke's shift representing the difference of peak intensity between two different molecules. Excitation and emission of GFP are: 395/509 nm, of YFP are 514/527 nm, therefore the stoke's shift between two FPs is about 18 nm in their emission peaks.

iv) Wavelength of emission

When a molecule absorbs (or is excited by) a photon excitation, it gains energy from a source and enters an excited state. One way for the system to enter the state of relax is to emit a photon, thus losing its energy as emission. Wavelength can be described by following equations:

$$E = h\gamma, or\ E = \frac{hc}{\lambda} \quad \text{(equation 1.3)}$$

Equation 1.3 As before, where E is energy, frequency is γ, wavelength is λ, c is the speed of light, and h is Planck's Constant. In this equation, as λ increases hc is divided by the wavelength, therefore the energy decreases. When the emission is high, the energy is low, so that it can be described that losing its energy means the molecule emits better [62].

The intrinsic fluorescence from protein molecules can mainly be classified to two types: (i) fluorescence from three different amino acids, and (ii) fluorescence from excited photons in the aromatic rings. Most protein include tryptophan (W), tyrosine (Y), phenylalanine (F), that contain an aromatic ring in the structure, and that releases fluorescence by excitation (mostly excited by light), and the fluorescence of a folded protein (excluding fluorescent proteins (FPs)) is a mixture of the fluorescence from individual aromatic rings from these three types of amino acids. Most of the intrinsic fluorescence emissions of a folded protein are due to the excitation of tryptophan residues, and some emissions are due to tyrosine and phenylalanine [64].

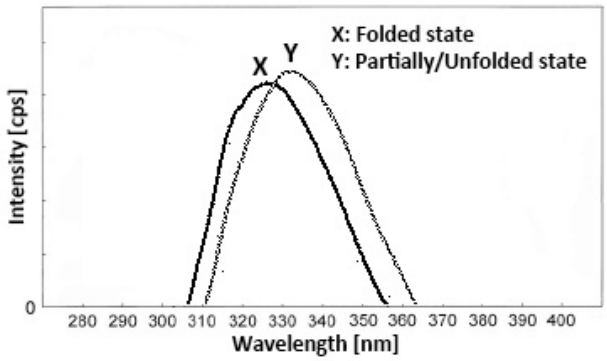


Figure 1.5 Excitation (λ_{ex} = 280nm) **and emission peaks of tryptophan, tyrosine, and phenylalanine between app. 300 to 350nm**. A typical emission (λ_{em}) of YFP that contains 26 aromatic ring compounds is 320 to 325nm (X: YFP, folded state, black), however as YFP molecule becomes unfolded state, λ_{em} has gradually been shifted 325 to 330 nm (Y: YFP, partially folded/unfolded states, grey).

Among these three amino acids that releases intrinsic fluorescence, tryptophan has a wavelength of maximum absorption of 280 nm and an emission peak that ranges from 300 to 350 nm and which is typically at a maximum at 340 nm, for a folded protein. The fluorescence arising from tryptophan residues is often used as a diagnostic tool of the conformational state of a protein and is particularly powerful probe as most proteins contain only one or a few tryptophan residues. Its fluorescence is, therefore, a very sensitive probe for analysis of the conformational state of individual tryptophan residues [64].

1.4.2 FPs in structural and molecular biology

FPs are classified as the members of modified proteins that share the unique property of being self-sufficient to form a visible wavelength fluorophore from a sequence of amino acids within their own polypeptide sequence [65]. The first demonstration of FPs

were observed in the fluorescence particles observed in bioluminescent organs of *Aequorea victoria* jellyfish (*phylum Cnidaria*, class Hydrozoa) as noted by Davenport and Nicol in 1955 [66], but it was Osamu Shimomura who identifed source of the fluorescence within a fluorophore in the particles of a protein in *Aequorea victoria*. Shimomura and his colleagues isolated the proteins from *Aequorea* and named the molecule as ‘aequorin’ that showed the fluorescence emission spectrum under the specific wavelength in the protein. Shimomura wrote in 1962, "A protein giving solutions that look slightly greenish in sunlight though only yellowish under tungsten lights, and exhibiting a very bright, greenish fluorescence in the ultraviolet of a Mineralite (a handheld ultraviolet lamp), has also been isolated from ‘squeezates’ (Shimomura et al. 1962) [67]. Through the research of aequorin, it was then discovered that energy transfer from aequorin to this GFP was due to the *in vivo* luminescence of *Aequorea*. Later, the aequorin and protein were re-named the ‘fluorophore’ [65] (or chromophore) and ‘GFP’ [65], and a number of similar products were discovered from several species, e.g. *Obelia* (*phylum Cnidaria*, class *Hydrozoa*) and the sea pansy *Renilla reniformis* (*phylum Cnidaria*, class *Anthozoa*) [65]. The chemical contribution of the fluorophore was still uncertain until 1979 when Shimomura correctly determined the fluorophore as a 4-(*p*-hydroxybenzylidene)-5-imidazoli-dinone moiety covalently linked within the polypeptide chain [65].

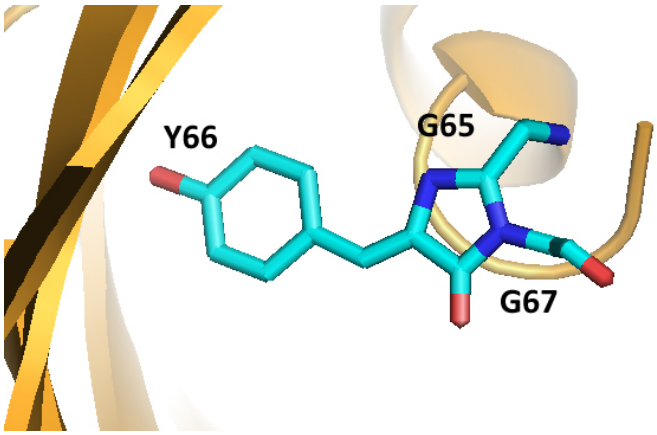


Figure 1.6 The fluorophore of YFP that inherits the original GFP fluorophore from residues: 65 to 67 but holds a mutation of S65G. The mutation, including 4 mutations (described in **Figure 1.8**), results in differences in the visible spectrum (from green to yellow) and excitation and emission wavelength (from $\lambda_{ex/em}$: 395/475 to 509 nm to $\lambda_{ex/em}$: 514nm to 527 nm.) Exported by MacPymol.

The formation of the fluorophore in the FPs starts with the nucleophilic G67 amide nitrogen attacking the electrophilic serine 65 carbonyl carbon to form a 5-membered ring in the main chain of the protein. The resulting tetrahedral hemiaminal intermediate undertakes an elimination of water to form a second intermediate. At the final stage, the Cα-Cβ bond of tyrosine 66 oxidises itself to a double bond with consumption of molecular oxygen and generation of hydrogen peroxide [68]. The double bond simultaneously converts the 5-membered ring into an aromatic formation and puts it into a 'bonding' with the aromatic phenol ring of the tyrosine side chain. Formation of fluorophore is, therefore, spontaneous only within the context of the FP in the β-sheet structure where steric constraints force the peptide into a tight turn conformation [69] and the side chains of highly conserved residues, such as R96 and E222 and which are positioned to facilitate the reaction, e.g. hydrogen bonds, between the fluorophore and R96 or E222.

In the wild type (wt) *Aequorea* GFP, the fluorophore exists as an equilibrium mixture of the neutral phenol (absorbance λ_{ex}: 395 nm, extinction coefficient: 25,000 M-1cm^{-1}) and anionic phenolate (absorbance λ_{ex}: 475 nm, extinction coefficient: 9,500 M-1cm^{-1}) [70, 71]. Regardless of whether the excitation peak is at 397 nm or 475 nm, the fluorescence emission occurs from the anionic phenolate species (fluorescence λ_{em}: 504 nm) with a high quantum yield of 0.79 which is considerably high [71]. Excitation of the neutral phenol species occurs fast excited state proton transfer (ESPT) [72] of the phenol proton to an internal hydrogen bond [73]. Nowadays, the variants of *Aequorea* GFP, e.g. YFP, red fluorescent protein (RFP), and cyan fluorescent protein (CFP), with the individual equilibrium states and the peaks of excitation and emission that are shifted to the phenol [74, 75] or phenolate [76] are widely applied and studied in biochemistry. In particular, these fluorescence proteins are commonly introduced into living cells to the C- or N- termini of proteins to visualise their location and dynamics during expression, using fluorescence microscopy [77, 78, 79].

1.4.3 Structural properties of GFP and YFP

There are many different types of family members of GFP-like proteins that share a barrel structure, and similar maturation of the fluorophore, but which have very distinct fluorescent properties, resulting in different colours. The colour of FPs is determined by the composition of the fluorophore and several interactions between the core and the rest part of the molecules [80].

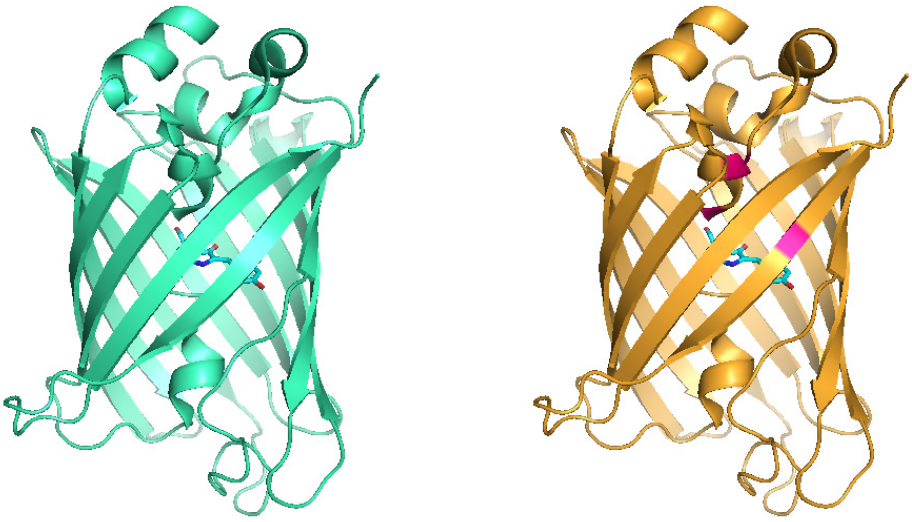


Figure 1.7 Comparison between GFP (left) and YFP (right). 4 mutations: S65G, V68L, S72A, and T203Y, occur differences not only visible emission but also maturation rate of folding among 2 FPs. Mutations on YFP are highlighted in pink as described in **Figure 1.8**.

One of the red shifted variants of GFP, YFP [81], has a specific and unique properties with shifted peaks of excitation and emission at 514 and 527 nm (absorbance λ_{ex}: 397/475 nm to 514 nm, fluorescence λ_{em}: 504 nm to 527nm). YFP sustains three types of mutations including S65G on the fluorophore. The mutation of S65G prevents the formation of a hydrogen bond between S65 and E222 (as described in previous Section 1.4.2). Second, the mutation T203Y forms additional polarisation to the fluorophore on YFP. T203Y also forms a phenol ring between the fluorophore (called π π-stacking) and T203 which induces the excited energy state. As a result, the excitation peak of YFP is shifted to 514 nm. As CFP which holds a mutation of Y66W, YFP is also commonly applied to FRET as a genetically encoded sensor. By the mutations of S65G and T203Y, whereas, YFP holds the fast fluorophore maturation, and may result in a positive effect to formation of the fluorophore on YFP where dehydration and oxidation occurs as GFP [82] Figure 1.8 demonstrates the amino acids conformations of GFP and

YFP (the sequences of GFP and YFP were obtained from NCBI: GFP locus = AAB47852, and YFP locus = ACS44346). As demonstrated in Figure 1.8, there are also mutation of V68L and S72A on YFP sequences, but the functions of these mutations are still in subject.

GFP

```
MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT (50)
GKLPVPWPTL VTTFSYGVQC FSRYPDHMKR HDFFKSAMPE GYVQERTIFF (100)
KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV (150)
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY (200)
LSTQSALSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK (238)
```

YFP

```
MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT (50)
GKLPVPWPTL VTTFGYGLQC FARYPDHMKR HDFFKSAMPE GYVQERTIFF (100)
KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV (150)
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY (200)
LSYQSALSKD PNEKRDHMVL LEFVTAAGIT HGMDELYK (238)
```

Figure 1.8 Full residues of GFP and YFP amino acid sequences. Mutation of YFP is highlighted in pink, and the region of the fluorophpres 65 to 67 are highlighted in grey. NCBI quotation Locus: GFP = AAB47852, YFP = ACS44346.

Most FPs are known to be relatively intolerant to changes in pH and ionic concentrations [83], and changes its structural and physical conformation in such hostile environments [84]. This tendency is more severe in YFP, and by the reason conversion of the states from matured/folded to the nascent polypeptide chain can be easily occurred to YFP by changing pH and ionic concentrations. Venus, a genetically

modified variant of YFP has been reported to maintain a tolerance to hostile environments, e.g., high pH and ionic concentration, and the mutations of Venus (to YFP): F46L, F64L, M153T, V163A, S175G, are planed to sustain the acceleration of the dehydration and oxidation that maintain the stability of the molecule. The beauty of YFP is not only its tolerance to different aqueous environments, but its rapid fluorophore maturation and very high quantum yield (more visible bright intrinsic yellow fluorescence) [78]. (in this research study, the sequence: locus = ACS44346, has been applied to all mature YFP (wt) sequence not Venus).

1.5 NMR applications to structural biology and protein folding

1.5.1 NMR studies for protein structure, dynamics, and folding

NMR is a powerful tool for providing structural and dynamical information on various states accessible to proteins at a residue-specific level. While traditionally the technique is associated with studies of proteins less than 30 kDa the development of advanced isotopic labelling strategies and TROSY pulse sequences [85] has enabled very large proteins and assemblies to be studied for example, the 670 kDa proteasome core particle [86], the 900 kDa GroEL/ES complex [87] and protein folding on the ribosome [88].

An important characteristic of NMR experiments is the observation and study of chemical shifts of resonances within 2D ^{1}H - ^{15}N and ^{1}H - ^{13}C correlation spectra, which are typically recorded either as HSQC (heteronuclear single quantum coherence) or HMQC (heteronuclear multiple quantum coherence) experiments. These spectra are 'fingerprints' of the folding of proteins and provide the starting point for residue specific analyses. Cross peak chemical shifts can provide backbone (e.g. NH amide groups) or side chain (CH groups) information and are defined by the chemical environment in which the amino acid resides within the structure of a protein. For example, in the unique carbon/proton chemical shift associated with each CH group within side chains. The carbons with a single proton bound to them, for example the CαH of most amino acids (except glycine) or the CγH in Leucine, can generate single resonance. Second, carbons with two protons attached to them, for example the H-Cβ-H group in Leucine or Glutamate, will, as a result of this, generate two C-H resonances. The resonances from a CH_2 group are distinct if the chemical environments of each

proton are different. Methyl groups as found in leucine, isoleucine, valine, alanine and methionine generate only one distinct C-H resonance because of the free rotation around the central carbon. This free rotation means all protons occupy an equivalent average chemical environment. On the other hand, nitrogen/proton correlation experiments always generate a single cross peak for each residue because of the amide N-H in each amino acid [89].

1.5.2 NMR approaches to protein folding

NMR spectroscopy is a useful technique to understand protein folding at a residue specific level. For example, 1D ^{1}H or 2D ^{1}H - ^{15}N correlation spectra can be used as tools to rapidly assess folding of a protein. A 1D ^{1}H NMR spectrum demonstrates a signal for each proton, and well-dispersed ^{1}H amide chemical shift values between 6.5 - 10 ppm are typical of the folded conformations of proteins, while poorly-dispersed signals between 7.9 - 8.6 ppm are indicative of disordered states.

2D ^{1}H - ^{15}N correlation spectra containing a cross peak for each amide group can be a model to observe the secondary and tertiary structure of a protein with residual specific resolution. In 2D spectra, each cross peak is defined by position of chemical shift, and changes in the chemical shifts when a perturbation occurs can indicate conformational changes within the targeting proteins.

Similar to 1D spectra, the dispersion of chemical shifts in 2D ^{1}H - ^{15}N correlation spectra provides information about the tertiary structure of the protein, where presence of poorly-dispersed resonances between 7.9 - 8.6 ppm are typically indicative of molten globule or unfolded states of the proteins. Changes in the intensity or chemical shifts of

each cross peaks with temperature or chemical denaturants can also be recorded to investigate the unfolding process of the proteins. Additionally, ^{1}H - ^{13}C correlation spectra, in which cross peaks are observed for each covalently bonded pair of ^{1}H and ^{13}C atoms (e.g. side chain methyl groups), can be also monitored to probe the structure and folding of a protein for example based on the dispersion of methyl ^{1}H chemical shifts. However, the dispersion of cross peaks in 2D ^{1}H - ^{13}C spectra of unfolded proteins is poor. Therefore intrinsically disordered proteins, which lack folded structure, are better studied by ^{1}H - ^{15}N spectra due to the greater ^{15}N chemical shift dispersion [90].

In order to obtain further understanding of structure and dynamics of a protein, chemical shift assignment of cross peaks in the 2D ^{1}H - ^{15}N spectra is required. To obtain a chemical shift assignments, 3D ^{1}H, ^{15}N, ^{13}C NMR experiments are required to assign each cross peaks to corresponding amino acid residues. These experiments provide rich information about intra and inter residue connectivity from a nucleus, e.g. Hα, Cα, Cβ, and C', in residues (*i-1*) or *i* to the associated amide NH resonances. Having established these connections between the nuclei, a sequential assignment of the backbone resonances can be obtained through connectivity patterns using the amino acid sequence of a protein. Chemical shift of H, NH, Hα, Cα, Cβ, and C' nuclei determined by the 3D assignment experiments can be also used to analyse the secondary structure of a protein by comparison to the estimated random coil of chemical shift values for a given residue type [91, 92].

In addition to the study of secondary structure within a protein, the tertiary and quaternary structure can also be determined by the measurement of inter-proton distances [93]. These inter-proton distances can be obtained by use of nuclear

overhauser effect (NOE) experiments, such as the 3D NOESY-HSQC which gives cross peaks between nearby (<5Å) residues that can be used as ^{1}H - ^{1}H distance restraints in structure calculations. Additional experiments can be recorded to provide dihedral angle and hydrogen bond restraints, which can also be incorporated into structure calculation software.

The measurement of residual dipolar couplings (RDCs) has become a new and powerful method for obtaining structural restraints [94, 95, 96]. The measurement of RDCs between two nuclei can provide information on the distance between the nuclei and their orientation relative to the molecular alignment in an anisotropic medium such as phage or lipid bicelles. In comparison to NOEs, where the distance dependence is $1/r^6$ (radius), the distance dependence of the RDCs is $1/r^3$, therefore this can provide longer-range information on nuclei more distant within the structure.

An important method to study the size of proteins in folded and denatured states, as well as in RNCs, is measurement of translational diffusion using pulse field gradient (PFG) NMR experiments. Diffusion coefficients are useful to estimate the hydrodynamic radius of a species according to the Stokes-Einstein equation [97]:

$$D = \frac{kT}{6\pi\eta R} \qquad \text{(equation 1.4)}$$

Equation 1.4 where k is the Boltzman constant ($k = 1.38\times10^{-23} JK^{-1}$), T is the temperature in Kelvin, and the viscosity of water at 283K (10℃) $\eta = 1.305\times10^{-3}$ Pa s^{-1}.

Diffusion coefficients and hydrodynamic radii have previously been measured for a range of proteins in both native or unfolded conformations, and using these data, relationship between hydrodynamic radius (R_h, in nm) and number of residues (N) can be defined. These are shown in Equation 1.5 for folded state and Equation 1.6 for denatured state [97].

$$R_h = 2.21N^{0.29} \quad \text{(equations 1.5)}$$

$$R_h = 2.21N^{0.57} \quad \text{(equations 1.6)}$$

NMR diffusion is particularly useful approach to distinguish not only the isolated proteins but also to resolve a mixture of species in solution according to the difference in their diffusion coefficients [98]. Particularly, the ^{15}N stimulated echo (XSTE) diffusion experiment [99] can be used to distinguish between two different ^{15}N-labelled species in solution and have been applied subsequently in this study for C-terminal truncations, which will be described in Result 2 Chapter with multiple variants in solution.

1.5.3 NMR applications for studying dynamics of proteins

NMR spectroscopy is a particularly useful technique to observe the behaviour of the native state conformation. From pico seconds to hours, the ability to observe dynamics of polypeptides on different time scales is one of the advantages to use NMR spectroscopy. In fact, NMR allows the study of many biological processes such as protein folding, enzyme catalysis, ligand binding, and other biological active processes, using NMR active nuclei as reporters of motion within the macromolecule.

1.5.3.1 Fast processes - pico to nano seconds

^{15}N nuclear spin relaxation rate measurement is the most widely used set of experiments for observing protein dynamics by NMR. Relaxation of a spin refers to the process by which the nuclear magnetisation returns to the equilibrium state after an RF pulse is applied, and its rate is determined by the magnitude of fluctuations in the magnetic field at the nucleus. Typically, R_2, R_1 and heteronuclear NOE experiments for ^{1}H - ^{15}N nuclei are recorded which provide information about fast processes from pico to nanosecond timescales that describe rotational diffusion and fast loop motions. The NOE reports on the mobility of the protein, where a value of 0.8 corresponds to a rigid structure, and a value of -2 for a flexible structure. Relaxation in R_2 experiment has additional contributions from chemical exchange and is more sensitive to slower motions on micro to milli seconds timescales as followed [100].

1.5.3.2 Slow process - micro seconds to hours

In addition to R_2 relaxation measurements, Carr Purcell Meiboom Gill (CPMG) relaxation dispersion experiments have become a major technique to observe slow protein dynamics on the millisecond timescale. A key feature of relaxation dispersion measurements is to enable the study of protein conformation and dynamics in high energy 'excited' states as well as the ground state. Relaxation dispersion experiments provide information about the millisecond timescale of protein behaviours, e.g. ligand binding, enzyme catalysis, and protein folding [101, 102].

Slower processes that cannot be studied by relaxation dispersion can be investigated by exchange spectroscopy (EXSY) and real-time NMR. Exchange occurring in EXSY experiments gives rise to additional cross peaks, the intensity of which enables

calculation of the rate of exchange. Real-time NMR can also provide information about the process of protein folding and unfolding, on time scales upwards from 1 second. Alternatively H-D exchange is an approach to observing protein behaviour by real-time NMR, using measurements of amide exchange that occurs at slower than second time scales to obtain information about the structure and the stability of the protein, and rates of local unfolding [103].

1.6 Aims of the study

At present, our knowledge of protein folding as it occurs on the ribosome is limited to a few examples that use either biochemical, biophysical or structural techniques on a particular system. In order to gain a better understanding of the mechanisms underlying co-translational folding. The aim of this Thesis is to develop a model for the co-translational folding properties of YFP using RNCs and C-terminal truncations as an *in vitro* mimic. By exploiting YFP's unique fluorescence spectral properties together with NMR spectroscopy, it will permit a first-hand insight into both the structure and dynamical properties of this protein during synthesis.

Chapter 2

Materials and Methods

2.1 Plasmid design for study of YFP RNCs and truncation C-terminal truncations

2.1.1 RNC constructs of YFP17, YFP34, and YFP55

The RNC plasmids YFP17, YFP34, and YFP55 were designed and generated by Dr Lisa D Cabrita and their construction will be described here. Full length sequence derived from *Aequorea victoria* YFP was subcloned as an NheI-SpeI fragment into three different pLDC vectors, pLDC-17, pLDC-34 and pLDC-55. Each vector has a pET21b (+) backbone, and a hexahistidine tag at the N-terminus. Each vector has the 17 amino acid SecM stalling sequence: 150 - 167 (FSTPVWISQAQGIRAGP), at the end of C-terminus. Each of the vectors has different linkers derived from the SecM protein [43, locus: ENH55076]: pLDC-17 has no linking sequencing, pLDC-34 has residues: 102 - 123 (GCATTKGYRIDYAHFTPQASL) and pLDC-55 has residues: 102 - 143 (GCATTHLALLDTLSALLTQEGTPSEKGYRIDYAHFTPQASL for YFP55). DNA templates were designated as YFP17, YFP34, and YFP55, and give rise to NCs of 266, 283 and 303 residues respectively. Figure 2.1 shows a schematic of each of the constructs used in this study.

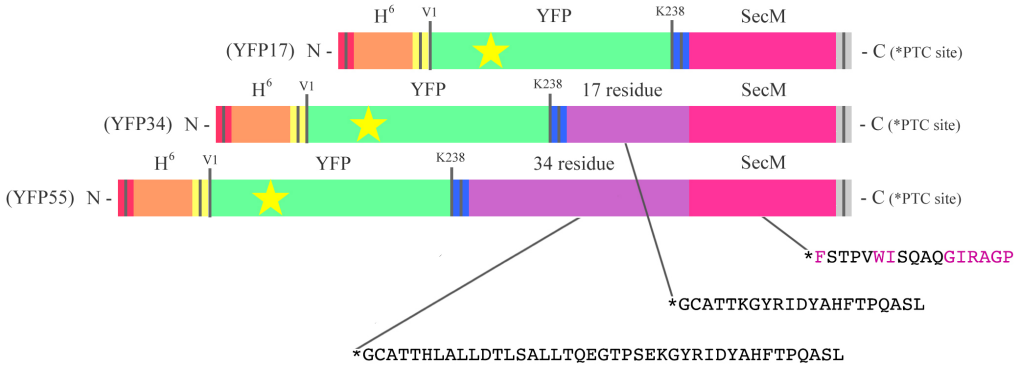


Figure 2.1 The constructs of YFP-RNCs preparation. In each construct, the SecM stalling sequence (17 residues: 150 - 167) was added to create a holding interaction with L4/L22 ribosomal proteins within the tunnel (the part that creates interaction with L4/L22 proteins on SecM, pink). Parts of SecM sequence were also added to YFP34 and YFP55 (17 and 34 residues, purple) as the linker based on the residue: 102 -149. More details are described in **A.1** to **A.3** in Appendix Chapter (*NdeI site: red, NheI site: yellow, stop codon: grey, SpeI site: pale, and EcoRI site: blue - from left to right).

Name of construct	*YFP17*	*YFP34*	*YFP55*
Molecular weight	30 kDa	32 kDa	34.1 kDa
Number of amino acids	266	283	303

Table 2.1 Basic information of YFP17, YFP34, and YFP55 as NC states. Further detail is described in Section **A.4** in Appendix Chapter.

2.1.2 Midiprep DNA plasmid preparation

0.5 µl of each of the plasmids: YFP17, YFP34, and YFP55 (500 ng/µl) were transformed into 50 µl of DH5α gold competent cells (Invitrogen, US) by heat-shock at 42℃ for 45 seconds followed by 5 minutes on ice. Pre-warmed SOC medium (Novagen, US) was added and incubated at 37℃, 72 rpm shaking for 15 minutes. 200 µl of the transformed cells were plated onto Luria Bertani (LB) agar plates containing ampicilin at 100 µg/ml, and the plates were incubated at 37℃ overnight. A single colony from the plate was picked for each plasmid and put into 3 ml of LB media containing 100 µg/ml ampicilin. The LB cultures were then incubated at 37℃, 180 rpm shaking for 3 hours (typically OD absorbance at 600 nm (A_{600}) was 0.1). After 3 hours pre-culture, 300 µl of these mixtures were placed into 300 ml of LB media containing 100 µg/ml ampicilin, and the new mixtures were incubated at 37℃, 180 rpm shaking until A_{600} reached 2.4 to 3.0, typically after 20 hours.

After 20 hours incubation, the culture was centrifuged at 5,000 rpm for 30 minutes at 4℃ typically in 1.8 - 2.0 g pellet obtained from 300 ml cultures. Plasmid preparations were performed via the QIAfilter Plasmid Midi Kit, using their proprietary buffers (P_1, P_2, P_3, Qiagen MidiPrep Kit, US). Transformed and pelleted competent cells were resuspended in 8 ml of chilled P_1 buffer of Qiagen MIDI prep. The mixtures were placed in 50 ml falcon tubes, and 8 ml of P_2 buffer was added and was vigorously inverted to completely dissolve and were then incubated at room temperature. After 7 minutes incubation, chilled P_3 buffer was added, the mixtures were then poured into filter cartridges, and were incubated for 10 minutes (Usage of P_1, P_2, and P_3 buffers followed the protocol of QIAfilter Plasmid MidiPrep Kit protocol). After 10 minutes incubation, the mixture was separated into the debris and the lysate (via the plungers),

and the plasmids were washed and eluted with 60 ml of QC (wash) buffer and 7.5 ml of QF (elution) buffer.

For a typical cell-free reaction, in particular for the purposes of preparing an NMR sample of an RNC, a high quantity of plasmids is required, e.g., 50 µg to 100 µg for a 5 ml cell-free reaction, so the manufacturer's instructions were modified slightly and eluted the plasmids by a protocol which resulted in a greater yield of plasmid than the original instruction: we used 5.25 ml of ice-cold iso-propanol (70% vol.) which was added to 7.5 ml of the elution mixtures, and the mixtures were incubated at -80℃ for 15 minutes. Then the mixtures of DNA plasmids and isopropanol were centrifuged at 4℃ in 1.5 ml sterilised tubes, 20 minutes for 13,200 g. The pellets of DNA plasmids were washed with 75% (v/v) ethanol and (25% (v/v) distilled and deionised water (ddH_2O). 35 - 55 µg of DNA plasmid was finally obtained from 300ml LB culture. Each DNA plasmid was diluted to 500 ng/µl for use in a cell-free reaction.

2.1.3 DNA plasmids of isolated YFP C-terminal truncations

YFP (wt) and the C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, and f(0) are illustrated in Figure 2.2, and contain the coding region corresponding to the N-terminal 238 amino acid residues of full length YFP. The modifications were created by the insertion of a stop codon: taa, into designated position using site-directed mutagenesis. The construct of fluorophore (0) was modified by removing the fluorophore residues: 65 - 67, via PCR.

The coding region of YFP, defined as (wt) in this thesis, [81, PDB code: 1MYW, NCBI locus: ACS44346], was amplified via PCR and was inserted into the pET21b (+)

vector backbone (generated by Dr Lisa D Cabrita). Various truncation mutants of YFP were generated based on the YFP (wt) template adding a stop codon TAA, for C11, C12, C13, C14, C15, C16, C17, and C24, and removing the fluorophore residues: 65 - 67, for f(0) via PCR using above YFP (wt) template with the primers described in Table 2.2. All DNA amplification was performed using the following conditions: Reactions were performed in 34 µl volume (10 µl of DNA template (100 ng of YFP (wt)), 0.75 µl of forward and reverse primers (10 µM, total 1.5 µl), 5 µl of dNTPs (2 mM), 6.25 µl of 2x buffer of Novagen, 10.75 µl of autoclaved ddH_2O, and 0.5 µl of KOD Hot Start DNA Polymerase (3U/µl)). The annealing step was performed using a cycle of 1 × at 94℃ for 30 seconds, 36 × at 98℃ for 10 seconds, 55℃ for 7 minutes, and 68℃ for 10 minutes, 1 × at 4℃ for 10 minutes.

After PCR, the reaction mixtures were digested by DpnI digestion enzyme to remove the original methylated template for transformation, and were run on 0.8% w/v agarose gel (containing ethidium bromide (EtBr)) to observe the digested product against a 1 kB ladder (TrackIt, Invitrogen). The full sequences of these nine YFP C-terminal truncations (including YFP (wt)) are shown in Sections A.5 to A.14 in Appendix Chapter.

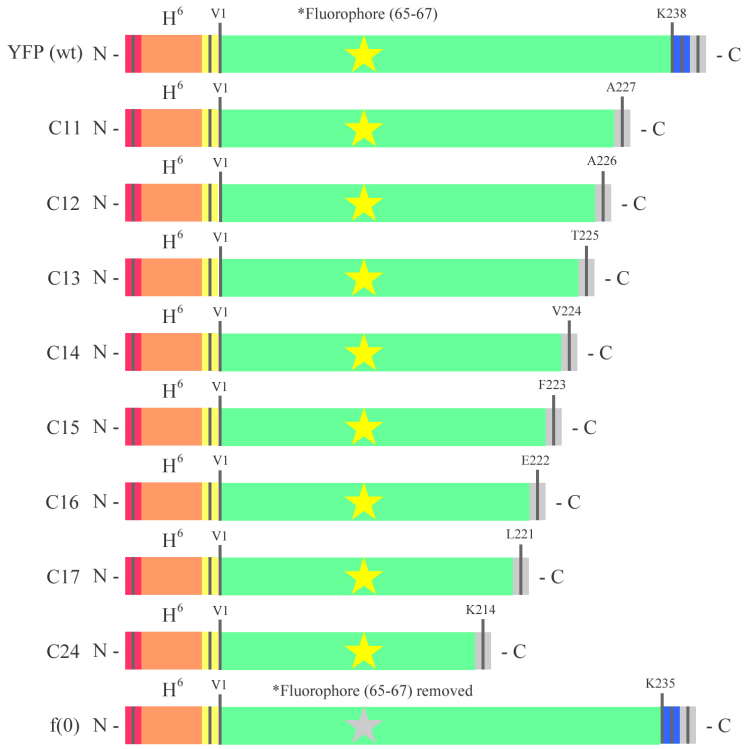


Figure 2.2 Protein sequences of the YFP C-terminal truncations, C11 to f(0) used in this study topic. A stop codon: `taa`, was added to each C-terminal truncation at the designated positions, e.g., A227 on C11. The residues comprising the fluorophore: 65 - 67, were deleted in the f(0) C-terminal truncation shown by the grey star as changed to the yellow for the presence of the fluorophore. More details of the DNA constructs and amino acid residue are described in Section **A.5** to **A.14** in Appendix Chapter (*NdeI site: red, NheI site: yellow, EcoRI site: blue, and stop codon: grey).

2.1.4 Purity check of DNA plasmids and PCR products

At each step, obtained DNA plasmids and PCR products were examined via Nano-Drop (Thermo Scientific, UK) and spectrophotometer (Shimazu Industry, Japan), absorptions at 260 nm for DNA and 280 nm for proteins via spectra of UV absorbance.

Name	*Forward*	*Reverse*
C11	5′ gtg acc gcc gcc taa ggg atc act ctc′3	5′ gag agt gat ccc taa ggc ggc ggt cac′3
C12	5′ ttc gtg acc gcc taa gcc ggg atc act′3	5′ agt gat ccc ggc taa ggc ggt cac gaa′3
C13	5′ gag ttc gtg acc taa gcc gcc ggg atc′3	5′ gat ccc ggc ggc taa ggt cac gaa ctc′3
C14	5′ ctg gag ttc gtg taa acc gcc gcc ggg′3	5′ ccc ggc ggc ggt taa cac gaa ctc cag′3
C15	5′ ctg ctg gag ttc taa gtg acc gcc gcc′3	5′ ggc ggc ggt cac taa gaa ctc cag cag′3
C16	5′ gtc ctg ctg gag taa ttc gtg acc gcc′3	5′ ggc ggt cac gaa taa ctc cag cag gac′3
C17	5′ atg gtc ctg ctg taa gag ttc gtg acc′3	5′ ggt cac gaa ctc taa cag cag gac cat′3
C24	5′ ccc aac gag aag taa cgc gat cac atg′3	5′ cat gtg atc gcg taa ctt ctc gtt ggg′3
f(0)	5′ ccc acc ctc gtg acc acc ttc (-) ctg cag tgc ttc gcc cgc tac′3	5′ gta gcg ggc gaa gca ctg cag (-) gaa ggt ggt cac gag ggt ggg′3

Table 2.2 Oligonucleotide primers used to create YFP C-terminal truncations shown in **Figure 2.2** via PCR. Full sequences of YFP C-terminal truncations and more detailed primer sequences are described in Section **A. 5** to **A.14** in Appendix Chapter.

2.2 Optimisation of cell-free reaction for RNC studies

For the production of RNCs by cell-free, two strategies were used: an S30 (RIKEN, Yokohama, Japan) extract and a commercial available kit, RTS500 kit (Roche, US)

2.2.1 Minimum performance and optimisation for nutrition rich environments (RTS100 and RTS500)

	RTS500 (10 µl scale)	*RTS500 (5 ml scale for NMR)*
e.coli lysate	4.8 µl (48%)	2,400 µl (48%)
reaction mix	2 µl (20%)	1,000 µl (20%)
^{15}N - ^{13}C aas (2 mM)	2.4 µl (24%)	1,200 µl (24%)
^{15}N tryptophan (2 mM)	0.2 µl (2%)	100 µl (2%)
RNA guard (20U/µl)	0.2 µl (2%)	100 µl (2%)
DNA plasmid	200 ng in 0.4µl (4%)	100 µg in 200 µl (4%) = 500 ng/µl
Total volume	10 µl	5,000 µl

Table 2.3 Optimised cell-free reaction condition for RTS100 and RTS500 kits.

The manufacture recommends performing a reaction at a minimum 50 µl using the RTS100 kit, however the reactions were scaled down from 50 µl to 10 µl scale in a test reaction with the provided GFP vector with an unlabelled amino acid source. The RTS500 is a larger-scale kit (for 5 ml) and the trial expressions were also performed on a 10 µl scale (in a 1.5 ml eppendorf). The optimised protocol for expression in both the 10 µl and 5 ml scale is shown in Table 2.3. Concentrations for selectively labelled ^{15}N -

^{13}C amino acids. ^{15}N tryptophan, and RNA guard (the RNAse inhibitor, Invitrogen, US) were as described in the manufacture protocol.

2.2.2 Preparations of amino acids and other cell-free reaction ingredients

^{15}N - ^{13}C amino acid mixture (Spectra Stable Isotopes, Sigma-Aldrich, UK) and ^{15}N tryptophan (^{15}N tryptophan, Isotech, Switzerland) were dissolved in 60 mM Hepes and 0.7 M HCl, and were then filtered by 0.22 μm filter. Isotopically labelled amino acids mixture (^{15}N - ^{13}C) and tryptophan (^{15}N) were at fixed concentration (2 mM) for each in reconstitution buffer of RTS500. RNA guard was diluted in autoclaved ddH_2O, and was finally fixed concentration 10 units/μl as the suppliers recommended.

2.2.3 Ultra high-speed centrifugation via sucrose cushion for separation of RNC and free NC samples

RNCs were separated from released NCs using a 30% w/v sucrose cushion in tico-buffer (30% sucrose (w/v), 10 mM Hepes at pH7.5, 30 mM NH_4Cl, 6 mM $MgCl_2$, and 1 mM beta-mercaptanol). Following the cell-free reaction, the sample was diluted with 90 μl ice-cold tico-buffer coloured, and the mixture was subjected to a 1 ml 30% w/v sucrose cushion at 4℃ for 3 hours at 96,000 rpm. After the centrifugation, approximately 150 μl of the top layer was carefully placed to 1.5 ml tube. The 150 μl layer was expected to contain the released free NCs through dividing the volume of reaction mixture. By adding of 600 μl methanol (4 vol.), 150 μl chloroform (1 vol.), and 450 μl ddH_2O in order to precipitate the free NC. Both pellets of RNCs and free NCs were visualised by immunoblot.

2.2.4 Cell-free reaction with S30 at RIKEN Systems and Structural Biology Center

In a collaboration with Dr Chie Takemoto, Dr Yutaka Mutoh, and Professor Shigeyuki Yokoyama (RIKEN, SSBC, Japan), the use of the S30 cell-free expression system was performed in order to obtain the RNCs of YFP17, YFP34, and YFP55.

In trial expression, YFP34 construct showed the highest expression compared with other RNCs constructs, so that a large scale RNC sample of YFP34 was performed in dialysis mode. 1 ml of S30 cell-free reaction, mixture contained 20 µl of autoclaved ddH_2O, 373 µl of ^{15}N - ^{13}C labelled LMCP-tRNA, 10 µl of 17.5 mg/ml tRNA, 10 µl of 5%NaN_3, 5.8 µl of 1.6 M Mg $(OAc)_2$, 75 µl of 20 mM ^{15}N - ^{13}C labelled amino acid - Tyr and 10 mM DTT, 66 µl of 3.75 mg/ml Creatine Kinase, 6.66 µl of 10 mg/ml T7 RNA polymerase, 300 µl of S30 extract (essentially a pure 70S ribosome), and 10 mg of purified YFP34 plasmid dissolved in 132 ml of autoclaved ddH_2O. The reaction was performed in 10 ml of dialysis mixture; 2.358 ml of ddH_2O, 3.733 ml of ^{15}N - ^{13}C labelled LMCP-tRNA, 100 µl of NaN_3, 3 ml of S30 buffer (10 mM Tris-acetate buffer (pH 8.2) containing 14 mM $Mg(OAc)_2$, 60 mM) KOAc, 1 mM DTT, and 0.5 ml/l 2-mercaptoethanol), 58 µl of 1.6 M $Mg(OAc)_2$, 750 µl of 20 mM ^{15}N - ^{13}C labelled amino acid - Tyr and 10 mM DTT, at 30°C, 300 rpm shaking for 60 minutes incubation.

After 60 minutes of this S30 reaction, the YFP34 cell-free reaction was quenched by adding of 3 ml of ice cold tico-buffer, and the mixture of 1 ml reaction and 3 ml tico-buffer was then centrifuged at 95,000 rpm, 4°C for 210 minutes in 5 ml of 30% w/v sucrose cushion in 10 mM Hepes at pH 7.5, 30 mM NH_4Cl, 20 mM $MgCl_2$, and 2 mM 2-mecaptanol to pellet the ribosomal method including for YFP34. After

centrifugation, the pellet was re-suspended and dissolved in 270 µl of tico-buffer and 30 µl of D_2O at pH 7.4. The RNC sample of YFP34 was then analysed by NMR spectroscopy.

2.2.5 Cell-free reaction with RTS500

In order to modify YFP-RNCs by the Roche system for large scale, a batch mode cell-free reaction for production of our model YFP-RNCs was performed (will be described in Sections 3.2 to 3.4 in Result 1 Chapter). The reaction condition were at 30°C, 750 rpm shaking for 60 minutes speed in a 50 ml falcon tube. The ingredients were described in Table 2.5. The reaction was quenched by adding an equal volume of ice cold tico-buffer (10 mM Hepes at pH 7.5, 30 mM NH_4Cl, 6 mM $MgCl_2$, and 1 mM beta-mercaptanol).

2.3 Centrifugation and sucrose gradient for purification

2.3.1 RNCs separation by centrifugation and sucrose gradient

The large scale cell-free reaction mixtures were layered onto 10 - 35% sucrose gradients in tico-buffer (10 mM Hepes at pH 7.5, 30 mM NH_4Cl, 6 mM $MgCl_2$, and 1 mM 2-mercaptanol) as ultra-centrifuged at 4°C for 16 hours in an SW28 rotor (Beckman Coulter, Fullerton, US). Sucrose gradients were then fractionated via a sucrose gradient fractionation Foxy Jr. via optical absorbance at 254 nm. All fractions were also monitored by SDS-PAGE with silver staining, and fractions corresponding to RNCs and 70S ribosome were pooled and concentrated to 5 ml via a 100 kDa Amicon Ultra centrifugation concentrator (Millipore, UK). This step was repeated three times in order to remove excess sucrose which could interfere with 2D ^{1}H - ^{13}C HMQC spectra in NMR spectroscopy. The RNC sample was finally concentrated to a final volume of 250 - 270 µl and 3.8 - 5 µM concentration. The concentration of ribosome was estimated by assuming that: 1OD A_{260} is equal to 24 pmole/ml.

2.4 Biochemical analysis of RNCs and YFP C-terminal truncations

2.4.1 SDS-PAGE and coomassie blue staining

Sample from cell-free reaction, affinity and gel filtration chromatography, and sucrose gradient, have been run SDS-PAGE for a purity check. The molecular weights of YFP17, YFP34, and YFP55 as the NC states were 30, 32, and 34.1 kDa respectively as described in Table 2.1. RNCs and released NCs were visualised by 12% w/v precast SDS-PAGE gel (1.5M Tris-HCl at pH 8.8, ddH_2O, 30% acrylamide, 10% sodium dodecyl sulfate (SDS), 10% ammonium peroxodisulfate (APS), and tetramethylethylenediamine (TEMED), and 4% stacking gel (0.5 M Tris-HCl at pH 6.8, ddH_2O, 30% acrylamide, 10% SDS, 10% APS, and TEMED, Invitrogen)). RNCs, free NCs, and protein samples were prepared with PBS, 2x Laemlli buffer (Invitrogen, US), and 0.1 M dithiothreitol (DTT), and were heated at 95℃ for 5 minutes before loading onto the stacking gel. The gels were typically run at 75 V constant (15 to 25 mA) for 15 minutes and was then run at 180 V constant (15 to 25 mA) for 36 minutes further in ultra pure 10x Tris-Tricine SDS running buffer (Invitrogen, US) with ddH_2O. After SDS-PAGE, the gel was stained with either silver staining or coomassie brilliant blue (CBB). For checking the purity of ribosome particles, silver staining method (Silver Quest, Invitrogen, US) was used following the manufacturer's instruction. And for checking the availability of proteins (and sometimes for ribosome and RNCs), The CBB staining (0.05% CBB (w/v), 40% methanol, 7% acetic acid, and ddH_2O) and the destaining solutions (10% methanol, 7% acetic acid, and ddH_2O) were used.

2.4.2 Immunoblot analysis for hexahistidine (H^6) tagged protein samples

Immunoblot (western blot) method was applied to detect RNCs. RNC samples were run on a 12% w/v SDS-PAGE gel (in Section 2.4.1). After SDS-PAGE, the his-tagged (H^6) RNCs, free NCs, and protein samples were transferred to a nitrocellulose membrane at 25 V constant (100 to 120 mA) for 90 to 120 minutes in the transfer buffer (25 mM Tris, 10mM Glycine, 20% Methanol, 0.01% SDS, and ddH_2O) at 4℃ or equivalent temperature conditions, e.g. on ice. For anti-His immunoblot detection of samples, blocking membrane was washed by TBS-T (25 mM Tris at pH 7.5, 150 mM NaCl, 0.05% Tween-20, and ddH_2O) and soaked in 5% skimmed milk or Penta His-HRP conjugate blocking solution (Qiagen, US) in 200 ml TBS-T at room temperature for 1 hour. The membrane was then washed with TBS-T, and the his-tagged sample on the membrane was blocked via a 1/2,500 Penta His-HRP primary antibody (Qiagen, US) in TBS-T at room temperature for 60 minutes at 100 rpm shaking speed. After primary blocking, the membrane was re-washed by TBS-T for at least three times, and Penta His-HRP primary antibody on the membrane was soaked and visualised by luminol enhancer and Peroxide (ELT) Super Signal Westo Pico solution mixture (Thermo Scientific, UK) at room temperature for 5 minutes. The blocking image was transferred to Kodak BioMax film (Kodak, Japan) and detected with an LAS-3000 imaging system (FUJIFILM, Japan).

2.5 Expression of isolated YFP (wt) and YFP C-terminal truncations

2.5.1 Transformation of DNA plasmids to competent cells

0.2 - 0.5 μl plasmids of these YFP C-terminal truncations (100 ng/μl) were transformed into 50 μl of DH5α competent cells (Invitrogen, US) by heat-shock at 42℃ for 45 seconds followed by 3 minutes on ice. 200 μl of SOC medium (Novagen, US) was added and incubated at 37℃, 240 rpm shaking for 60 minutes. 250 μl of transformed cells mixture was plated onto LB agar containing ampicilin at 100 μg/ml, and the plates were incubated at 37℃ for overnight. A single colony was picked and put into 50 ml of LB medium which contained 100 μg/ml ampicilin. LB culture was then incubated at 37℃, 240 rpm shaking for overnight. After the overnight incubation, the transformed cells mixture was centrifuged at 4,000 rpm and pelleted.

2.5.2 Expression of YFP and C-terminal truncations

Once DNA sequences were confirmed, each plasmid was transformed into BL21 Gold (DE3) competent cells (Invitrogen, US) and grown for overnight at 37℃, 240 rpm shaking. 20ml of this culture was put into fresh M9 medium culture (47.8 mM sodium phosphate dibasic at pH 8.0, 22 mM monopotassium phosphate, 8.56 mM NaCl, 0.4% ^{15}N ammonium chloride, 0.4% ^{13}C glucose, 100 mM calcium chloride, 100 mM magnesium sulfate at 37℃, 240 rpm shaking until A_{600} reached 0.4 to 0.6 (took approximately 3 hours). Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 500 μM was added to the LB medium to induce the production of proteins. For growth of six C-terminal truncations: C11, C12, C13, C14, C24, and f(0)

for 3 hours at 37℃, 240 rpm shaking, while for three C-terminal truncations: C15, C16, C17, for 24 hours at 16℃, 240 rpm shaking. Cells were harvested by centrifugation at 4℃, 4,500 rpm for 20 minutes, and typically 3.6 to 4.19 g pellets were obtained from 1 litre of M9 culture for each construct.

	Molecular weight	*No. of residues*	*Ext. coefficient (ε)*
YFP (wt)	26.7 kDa	238	22015
C11	25.5 kDa	227	22015
C12	25.4 kDa	226	22015
C13	25.4 kDa	225	22015
C14	25.3 kDa	224	22015
C15	25.2 kDa	223	22015
C16	25.1 kDa	222	22015
C17	24.9 kDa	221	22015
C24	24.1 kDa	214	22015
f(0)	26.5 kDa	235	22015

Table 2.4 Basic information of YFP C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, and f(0). Further detail is described in Section **A.4** in Appendix Chapter.

2.5.3 Purification of YFP and C-terminal truncations

Cell pellets of YFP (wt) and the C-terminal truncations were resuspended in 20 ml lysis buffer (20 mM Tris-HCl at pH 8.0, 100 mM NaCl, and a protease inhibitor tablet (Roche, UK)). Cells were pressured with a constant cell disruption system at 24 kpsi at 4℃. The sample was then sonicated (15 seconds sonication and 45 seconds

incubation repeated 6 times), and centrifuged at 17,000 rpm, 4℃, for 30 minutes.

The cell lysis was applied to 5 ml of Ni-NTA resin (Qiagen, US) in chromatography columns, and incubated at 4℃ for overnight. After washing with Talon base buffer for equilibration (20 mM Tris-HCl at pH 8.0, 100 mM NaCl), YFP (wt) and the C-terminal truncations were eluted via Talon elution buffer (20 mM Tris-HCl at pH 8.0, 100 mM NaCl, and 100 mM imidazole). Typical concentrations obtained for each construct were 4.8 - 10 mg/ml, and from Talon metal affinity chromatography, 125 - 165 mg of YFP (wt) and the C-terminal truncations.

YFP (wt) and the C-terminal truncations were then subjected to size exclusion chromatography using the fast protein liquid chromatography (FPLC) was applied. The size exclusion buffer used (25 mM Na_2HPO_4 at pH 8.0, 150 mM NaCl) with 1 mM EDTA and DTT with a Superdex S200 16/60 column (GE Healthcare, US) on a AKTA prime plus (Amersham Bioscience, UK). The column volume was 120 ml. Each protein sample was injected into the Superdex column via AKTA prime plus at room temperature and elution samples were collected in 5ml fraction after 60ml equilibration. From gel filtration chromatography, 0.75 - 5.8 mg of YFP (wt) and C-terminal truncations were typically obtained.

2.6 NMR Spectroscopy

2.6.1 NMR spectroscopy for RNCs sample

All RNC samples and 70S ribosome were generally concentrated to 12 μM, prepared in tico-buffer (10 mM Hepes at pH 7.5, 30 mM NH_4Cl, 6 mM $MgCl_2$, and 1 mM 2-mercaptanol) at pH volumes of pH 7.68 (YFP17), pH 7.59 (YFP34), and pH 7.89 (YFP55). NMR data were recorded on a Bruker 700 MHz spectrometer (UCL) or a Varian s900c-MHz spectrometer (RIKEN, SSBC, Yokohama campus), both equipped with triple resonance cryogenic probeheads. As a lock solvent, 10% v/v D_2O was present in all NMR samples. ^{1}H - ^{15}N hetero-nuclear single quantum coherence (HSQC (SOFAST-HSQC)) and ^{1}H - ^{13}C hetero-nuclear multiple quantum coherence (HMQC) spectra were recorded using a 300 ms recycling delay with 1,024 × 64 complex points collected in the ^{1}H and ^{15}N dimensions, respectively, and 1,024 transients per increment. ^{15}N-edited XSTE diffusion measurements [104] with varied diffusion coding gradient strengths (3 ms total duration) were recorded by using a diffusion delay of 200 ms with 2,048 transients. These two experiments, SOFAST-HSQC and X-STE diffusion measurements, were carried out in an interleaved manner, allowing the integrity of the RNC to be ascertained. These experiments were also recorded with a reconstituted mature YFP (wt) / 70S sample for comparison in final concentration of 5 μM for both samples. ^{1}H PFG-LED diffusion measurements [104] were also recorded for RNC samples, to monitor the translational diffusion of the 70S ribosome by using the observable proton resonances of the stalk L7/L12 proteins [18]. HNCO experiments (^{1}H - ^{15}N or ^{1}H - ^{13}C correlation projections) were recorded and revealed that the observed ^{1}H - ^{15}N correlations in the SOFAST-HSQC spectra could not arise from free amino acids because the presence of a peptide bond is the prerequisite for detection of

HiNiCOi-1 correlations [105]. High-resolution ^{1}H - ^{15}N SOFAST-HSQC spectra (2,048 × 200 complex points) of mature YFP (wt) was recorded at several temperatures ranging from 10° to 25°C to facilitate assignment of the resolved cross-peaks in RNCs of YFP17, YFP34, and YFP55 for SOFAST-HSQC spectra by following the minimum chemical shift displacement criterion. All NMR data including ^{1}H - ^{13}C HMQC spectra were processed and analysed by XWIN-NMR (Bruker BioSpin, Switzerland), NMRPipe [106], and Sparky [107] software packages. The linewidth analyses were carried out by using the two-dimensional cross-peak fitting routine in Sparky, assuming Gaussian line shapes with visual inspection of the fitting quality in each case. Cross peak of YFP C-terminal truncations were analysed to obtain the mean values of linewidths. Thirty-two non overlapping cross-peaks that are commonly present in the spectrum of RNCs of YFP before and after the addition of puromycin.

2.6.2 NMR spectroscopy for isolated YFP C-terminal truncations

YFP C-terminal truncations were prepared in tico-buffer (10 mM Hepes at pH 7.5, 30 mM NH_4Cl, 6 mM $MgCl_2$, and 1 mM 2-mercaptanol) and pH was measured before recording. They were measured pH by pH meter: pH 7.58 (C11), pH 7.67 (C12), pH 7.78 (C13), pH 7.60 (C14), pH 7.76 (C15), pH 7.53 (C16), pH 7.51 (C17), pH 7.67 (C24), pH 7.78 (f(0)), and pH 7.59 (YFP (wt). NMR data were recorded on a Bruker 700 MHz spectrometer. As a lock solvent, 10% v/v D_2O was present in all NMR samples. ^{1}H - ^{15}N SOFAST-HSQC spectra were recorded by using a 300-ms recycling delay with 1,024 × 64 complex points collected in the ^{1}H and ^{15}N dimensions, respectively, and 1,024 transients per increment. ^{15}N -edited X-STE diffusion measurements with varied diffusion coding gradient strengths (3 ms total duration) were recorded by using a diffusion delay of 200 ms with 2,048 transients. These three

experiments, SOFAST-HSQC, HMQC, and XSTE diffusion measurements, were carried out in an interleaved manner, allowing the integrity of the YFP C-terminal truncations to be ascertained. High-resolution ^{1}H - ^{15}N HSQC spectra (2,048 × 200 complex points) of the YFP C-terminal truncations was recorded at several temperatures ranging from 10° to 25°C to facilitate assignment of the resolved cross-peaks in the YFP C-terminal truncations SOFAST-HSQC spectra by following the minimum chemical shift displacement criterion. And also in the spectrum of YFP (wt) / C-terminal truncations were used to obtain the ^{1}H - and ^{15}N weighted chemical shift displacement given by:

$$\Delta\delta NH = \sqrt{(\Delta N)^2 + (\delta H)^2} \quad \text{(equation 2.1)}$$

Equation 2.1 ΔδNH is given by difference of chemical shift in ^{1}H and ^{15}N from original protein sequence in 2D spectrum of NMR.

All NMR data was processed and analysed by XWIN-NMR, NMRPipe, and Sparky software packages as described (the details will be described in Result 2 Chapters).

2.7 Fluorescence Spectroscopy

2.7.1 Intrinsic yellow fluorescence data collection on YFP-RNC and isolated C-terminal truncations

The intrinsic yellow fluorescence, extrinsic fluorescence, and anisotropy data excited by both non-polarised and polarised light for RNCs and YFP C-terminal truncations were recorded using a FluoroMax3 Holiba spectrophotometer (Horiba, Japan). YFP absorbs at 514 nm and releases emission as fluorescence at 527 nm [108]. The intrinsic yellow fluorescence data for all RNCs and YFP C-terminal truncations was obtained using the following conditions: acquisition 450 - 600 nm, increment 1 nm, integration 0.1 s, excitation slit 2.5 nm, emission slit 2.5 nm (for higher concentration sample, e.g. more than 5 μM, the slits for excitation and emission were 1 nm / 1 nm), and excitation peak at 514 nm [108]. In order to obtain intrinsic fluorescence from tryptophan, YFP C-terminal truncations were also excited at 280 nm with emission collected from 280 - 350 nm.

2.7.2 Extrinsic fluorescence data collection

Extrinsic fluorescence data for RNCs and YFP C-terminal truncations was obtained using 4,4'-Dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt (bis-ANS) typical scans ranged from 425 - 545 nm with typical excitation slit 2.5 nm, emission slit 2.5 nm (for higher concentration sample, e.g., more than 5 μM, the slits for excitation and emission were 1 nm / 1 nm,), with excitation of bis-ANS at 390 nm (each sample was labelled with 5 times concentration of bis-ANS at 4℃ for 1 hour, and then scanned) [109].

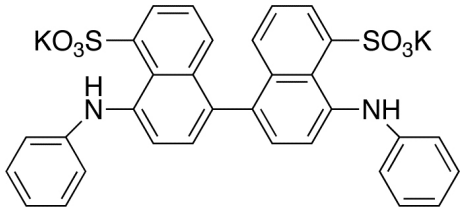


Figure 2.3 Schematic illustration of bis-ANS, 4,4'-Dianilino-1,1'-binaphthyl-5,5'-disulfonic acid dipotassium salt, molecular weight (MW) of the product is 672.85 [110].

2.7.3 Anisotropy data collection

The basic set up for measuring of fluorescence anisotropy is shown in Figure 2.4. In this experiment, the polarised (non-dispersed) light irradiated polarised light sources and full spectra of fluorescence emission is collected by four different angles: I_{HH} (horizontal/horizontal), I_{HV} (horizontal/vertical), I_{VH} (vertical/horizontal), and I_{VV} (vertical/vertical). Having these four different I-values, G- and R-values can be obtained by the Perrin equations described below. And finally calculated R-value stands for rotation and flexibility of NCs on each RNC [111].

$$R = \frac{I_{VV} - G \times I_{VH}}{I_{VV} + 2 \times G \times I_{VH}} \quad \text{(equations 2.2)}$$

$$G\ (\lambda_{em}) = \frac{I_{HV}}{I_{HH}} \quad \text{(equations 2.3)}$$

Equation 2.2 - 2.3 G- and R- values are obtained from emission spectra in four different angles at 527nm obtained from polarised excitation source: I_{HH} (horizontal/horizontal), I_{HV} (horizontal/vertical), I_{VH} (vertical/horizontal), and I_{VV} (vertical/vertical), as described in **Figure 2.4**.

For the data of anisotropy excited by polarised light in four different angles: I_{HH}, I_{HV}, I_{VH}, and I_{VV}, was obtained using following conditions: acquisition 450 - 600 nm, increment 1 nm, integration 0.1 s, excitation slit 2.5 nm, emission slit 2.5 nm (for higher concentration sample, e.g. more than 5 μM, the slits for excitation and emission were 1 nm / 1 nm,), and excitation peak at 514 nm with polarised light source with auto polariser function of FluoroMax3.

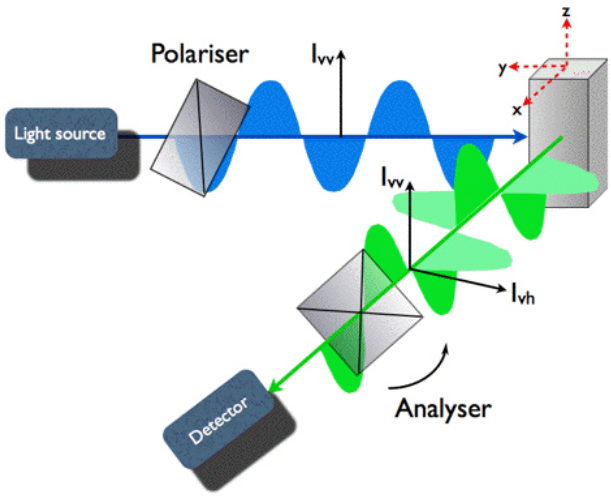


Figure 2.4 Fluorescence anisotropy. Polarised light (non-dispersed) excites a molecules (pink cuvette) in four directions, I_{HH}, I_{HV}, I_{VH}, and I_{VV}, the excited molecule releases fluorescence emission in four directions as they excited. Each peak collected and then G and R-values obtained by the Perrin equations described in above [111]. Figure is taken from a website [112].

Chapter 3 Result I - RNCs of YFP

A biophysical and structural characterisation of the co-translational protein folding characteristics of YFP

3.1 Behaviour of YFP on the ribosome

3.1.1 YFP as a model for examining protein folding on the ribosome

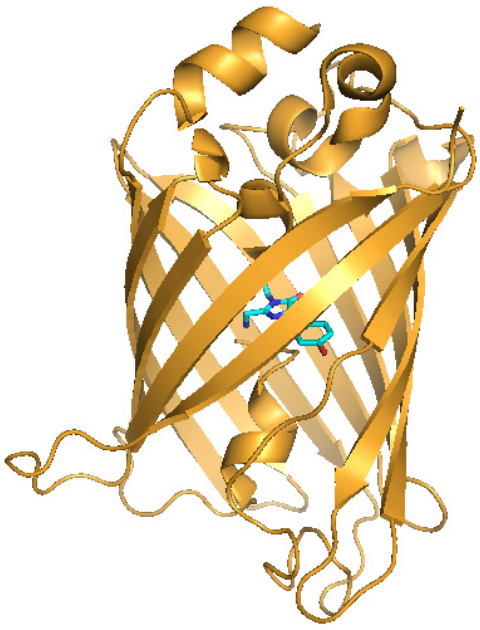


Figure 3.1 Structure of fully folded Venus, a variant of YFP. Structure solved by X-ray crystallography (2.2Å), Rekas et al, (2002) PDB Code: 1MYW. The fluorophore, residue 65 to 67, is shown in ball and stick (highlighted in blue, green, and cyan).

As previously detailed in Section 1.4.2 to 1.4.3 in Introduction Chapter, YFP is a FP and has properties such as a fast maturation rate associated with its fluorophore and comparable folding kinetics which means that it has significant advantages compared to other related FPs, e.g. GFP or RFP [79] for folding studies. During its process of

folding, the 238 amino acids of YFP acquires a structure which comprising of 4 α-helices and 11 β-strands which encapsulate the fluorophore (amino acids 65 to 67) in a β-barrel-like structure as shown in Figure 3.1. Once YFP has adopted its biologically-active state, it is able to fluoresce with an emission peak at in the visible range at 527 nm, following excitation at 514 nm.

The folding process of the FPs, as characterised by denaturation studies of GFP show that it is complex, involving for example, very slow folding kinetics [113, 114], and in which there is the formation of intermediates [115, 116] and the subsequent slow maturation of its fluorophore [117]. In contrast, folding studies of YFP [118] have shown that it is extremely thermodynamically stable, even more so than GFP, in which its midpoint of unfolding occurs in 5.8 M guanidinium chloride. YFP, however, is more environmentally sensitive, for example to mildly acidic conditions and chloride ions compared to GFP [84]. Hydrogen-deuterium exchange experiments using NMR spectroscopy also show that the protein has a very stable core formed by strands 1 - 2 and 4 - 6 as well as 10 and 11, but interestingly, strand 7 is said to be the least stable and previously identified to be one of the last strands to form the native state [119]. The protein, however, is more sensitive to mildly acidic conditions and chloride ions compared to GFP [84]. The intriguing folding properties of YFP and its established NMR properties, such as complete assignments of its backbone and side chains [120] makes YFP a very favourable model to study the process of co-translational protein folding as it occurs on the ribosome.

3.1.2 RNCs to study co-translational protein folding

For studies of co-translational protein folding, a variety of structural and biophysical methods are used (Section 1.1 in Introduction Chapter), and central to these studies is the generation of RNCs. Using the ability to arrest translation, it is possible to capture NC at different stages during the synthesis on the ribosome, and this enables protein folding of emerging NCs to be monitored at equilibrium. To achieve homogeneous translational-arrest, peptidic motifs are typically used and in particular, sequences such as SecM [43], TnaC [121], and ErmC [122] have been applied successfully to generate RNCs for structural studies. The ability of these peptidic motifs to enable translational-arrest comes from their ability to interact specifically with the ribosome. For example, recent structural studies of SecM-arrested RNCs [16] show that translational-arrest is achieved by specific interactions made between the motif and the ribosomal exit tunnel, in particular with ribosomal proteins L4 and L22, which result in a relay within the rRNA network towards the PTC (Figure 1.3 in Section 1.2.1 in Introduction Chapter); These events are also coupled with an altered geometry of the PTC which together encourage translational arrest [16].

In particular, for NMR studies of co-translational folding, The SecM stalling sequence has been used to successfully to generate RNCs of Dom5 [19] as well as barnase [123] and SH3 [124] for NMR studies of co-translational folding. Within our laboratory, Dom5-RNCs have been the subject of study, to a number of co-translational folding studies, in which RNCs of different lengths have been used to monitor the progressive emergence and folding of NCs as followed by NMR spectroscopy. These data show that Dom5 has the capacity to adopt a native fold while it is tethered to the ribosome [19] and that folding takes place at a significant distance, at least for 43 amino acids from the

PTC (Cabrita et al., (2014). Emergence and folding of RNCs during biosynthesis at high-resolution using NMR spectroscopy, manuscript in preparation).

In the co-translational folding studies of Dom5, the RNCs used for NMR are produced using *in vivo* expression in *E.coli* [123, 124]. While this method is robust and generates very large amounts of material, it is laborious and time consuming. An alternative method using a coupled transcription-translation cell-free protein synthesis, method offers a number of advantages, as detailed in Section 1.3 in Introduction Chapter. Specifically for RNCs, cell-free strategies using an *E.coli* extract is attractive because it enables rapid sample preparation, and importantly compared to *in vivo* strategies, it offers the distinct advantage of being able to specifically label the 70S ribosome-bound NCs with NMR-active ^{1}H - ^{15}N and ^{1}H - ^{13}C stable isotopes and importantly without any background labelling of the ribosome (as described in Section 1.3.1 to 1.3.3 in Introduction Chapter). Background labelling of the ribosome arises from labelling of the ribosomal protein L7/L12, which is the only region readily observable using conventional NMR techniques, and using the *in vivo* approach, its labelling can vary between 5 - 15% (Cassaignau A et al., 2014, A strategy for co-translational folding studies of RNCs using NMR spectroscopy, Manuscript in preparation). In addition, the adaptability of cell-free system means that it also permits non-natural amino acids such as those that are fluorophore-linked and others that can be chemically tagged (e.g. biotinylation) to be added to the reaction, enabling modified proteins to be generated. Developing a robust cell-free system for the production of RNCs therefore offers significant advantages for the study of co-translational protein folding.

As shown in Section 1.5 in Introduction Chapter, NMR spectroscopy is a powerful technique that provides the unique ability to offer residue-specific information

regarding the structure and dynamics of proteins and it has been shown that it has the capacity to report on dynamic NCs [35, 123, 124]. It is however, a relatively insensitive technique and typically requires large amounts (milligram quantities) of material. By contrast, fluorescence spectroscopy is highly sensitive and has the potential to offer detailed information regarding structural transitions within proteins as an ensemble and in this way, is a powerful complement to NMR spectroscopy. Together these strategies can offer the prospect of very detailed information of co-translational protein folding.

3.1.3 Aims of the chapter

YFP is an ideal candidate for the study of protein folding on the ribosome as it has an intricate 3D structure and complex folding mechanism *in vitro*. It presents an intriguing question of how the β-barrel structure is formed on the ribosome and how the fluorophore is created. Indeed, the ability to exploit YFP's natural function to fluoresce permits a unique method to monitor structure acquisition.

In this study, therefore, the unique fluorescent capacity of YFP will be used in parallel with NMR spectroscopy to provide residue-specific information on how this protein folds co-translationally on the ribosome. For this purpose, three SecM-stalled YFP-RNCs have been created in order to represent the progressive emergence of YFP from the ribosome (Figure 3.2). In addition, in order to generate RNCs that would be suitable for both structural and biophysical studies and offer the prospect of developing modified RNCs, a coupled transcription-translation cell-free approach was developed.

3.2 Development and optimisation of a cell-free system to produce RNCs of YFP for biophysical and structural studies

3.2.1 Design of YFP-RNCs

To develop suitable RNCs that would enable co-translational folding to be monitored by both fluorescence and NMR spectroscopies, RNC constructs of YFP were generated in the pLDC vector [19], in which the 17 amino acid SecM stalling sequence (FSTPVWISQAQGIRAGP) [16, 43] (generated by Dr Lisa D Cabrita) is used to induce translational-arrest. RNC constructs were designed in which the YFP protein which has a hexa-histidine tag at its N-terminus and is separated from the PTC by an inert, unstructured 55 amino acid linker sequence derived from residues of the SecM protein and two amino acids from the cloning site (EcoRI) at the C-terminus (referred to here as the linker sequence). By manipulating the length of this linking sequence this gave rise to three constructs, YFP17, YFP34 and YFP55, in which the YFP NC at its C-terminus was linked to the ribosome at distances of 19, 36 and 57 amino acids (Figure 3.2). In this Chapter the YFP-RNCs of linker lengths 19, 36 and 57 will herein be referred to as YFP17, YFP34 and YFP55 respectively.

The linking sequence allows the YFP protein to reside at different distances from the ribosome's PTC, which therefore has the capacity to provide information of how YFP progressively acquires its tertiary structure on the ribosome. It is expected that YFP17, with an additional 17 amino acids would partly reside in the ribosomal exit tunnel, which is said to accommodate up to 30 amino acids in an extended conformation [125, 126, 127]; YFP34 would permit YFP to reside just outside the tunnel, while YFP55 would be expected to have YFP extended some distance beyond the PTC and the

ribosomal exit tunnel.

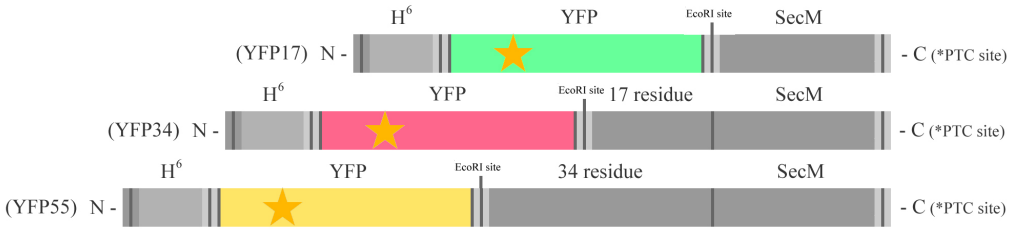


Figure 3.2 A schematic showing image of the different RNC constructs generated for YFP for co-translational folding studies by fluorescence and NMR. RNC constructs of YFP were generated in which the SecM sequence at the C-terminus was used for translational arrest and an N-terminal his tag is used for purification. To mimic the progressive emergence and folding of the YFP, different linker lengths derived from the SecM protein were used. Three YFP-RNCs were studied, in which YFP is separated from the PTC by a linker of 19, 36 and 57 amino acids; this linker includes an EcoRI site (2aa) and the SecM sequence (which is varied in length between 17, 34 and 55 amino acids). Also indicated by the yellow star is the fluorophore, which represents three residues 65 to 67.

3.3 Evaluating the expression of YFP - RNCs

3.3.1 Optimisation trials of cell-free for the overexpression of isolated YFP

To evaluate the expression of the three YFP-RNCs, DNA constructs described in Figure 3.2 were used as templates using the RTS system (Roche), a commercially-prepared *in vitro* coupled transcription-translation system. The RTS system has two reaction scales, RTS100 and RTS500 which are typically used for analytical and preparative expression scales respectively. In the first instance, the recommended conditions (30℃, 1 hour, 900 rpm, detailed in Table 2.3 and in Section 2.2.1 in Materials and Methods Chapter) for both the RTS100 and RTS500 systems that allow overexpression of proteins, were evaluated using a 10 μl reaction scale. After the reactions, the samples were applied directly to SDS-PAGE and YFP expression was analysed using an anti-His immunoblot (Figure 3.3). As there was no separation of the RNCs from released NC in this experiment, the expression of YFP is very likely to reflect the presence of released, isolated NCs as well as any RNCs (Section 3.3.2 to 3.4.1 for RNC purification).

As shown in Figure 3.3, there was clear expression of the YFPs in the three linker lengths of YFP17, YFP34, and YFP55, in which the bands of corresponding to the approximate molecular weight of released YFP of 30, 32, and 34.1 kDa respectively, as observed on the anti-his immunoblot (Section 2.4.2 in Materials and Methods Chapter). This is also evident from a distinct laddering effect observed between the three YFP constructs, in which the slight increase in the molecular weight is consistent with the addition of the linking sequence (Figure 3.2). In addition, there is also the presence of

lower molecular weight species, at c.a. 27 and 29 kDa that are evident across the RNC samples, in particular the YFP34 and YFP55 samples (Figure 3.3). It is not entirely clear the origin of these additional bands, though it is possible that the anti-His antibody is recognizing the presence of histidine-rich proteins on the ribosome in a non-specific manner; alternatively, the bands may also correspond to truncations product, this is indeed possible for the longer linkers in YFP34 and YFP55, which are known to correspond to disordered structure [128] and may therefore may be more susceptible to proteolysis.

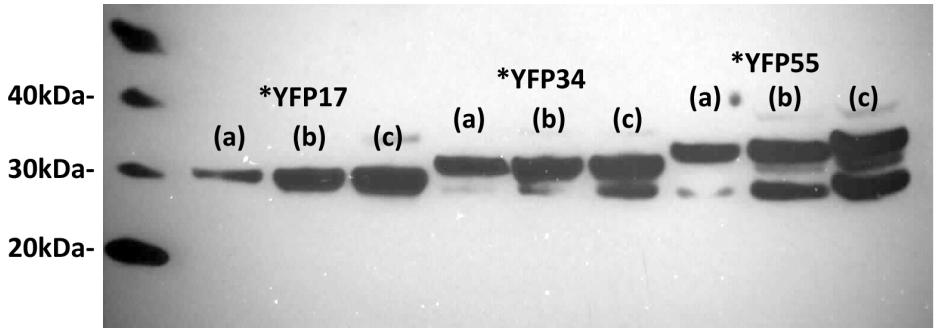


Figure 3.3 Evaluating cell-free expression of YFP17, YFP34, and YFP55 as probed by an anti-His immunoblot. Each construct YFP17, YFP34, and YFP55, were synthesised in three different conditions: **(a)** recommended conditions by the supplier (Roche) for RTS100 and RTS500 **(b)** the amount of DNA plasmid template was increased from 100 ng (2%) to 200 ng (4%) in total volume of 10 µl scale reaction, IPTG added. **(c)** the amount of DNA plasmid template as increased in condition **(b)**, while the amount of labelled by ^{15}N and ^{13}C amino acids was increased from 1 mM (12%) to 2 mM (24%) and IPTG was added. The temperature and shaking speed did not appear to affect the level of expression, and thus the conditions used for synthesis was set at 30℃, 900 rpm. Lanes 1, 4, and 7 show RNC samples synthesised by supplier recommended protocol. Lanes 2, 5, and 8 show RNC samples synthesised in DNA rich environment from 100 ng to 200 ng in total reaction. Lanes 3, 6, and 9 show RNC samples synthesised in both DNA and ^{15}N and ^{13}C labelled amino acids rich environment from 1mM to 2mM in total volume.

This initial experiment demonstrated the utility to generate released, isolated YFP. To evaluate on a qualitative basis, the ability to improve the expression of released YFP (using the YFP-RNC template), several parameters (temperature, DNA template concentration, shaking speed and the effects of additives) were next examined in a systematic manner. The rationale was by improving the overall capacity to express YFP as a released NC, that it would then permit high levels of YFP-RNCs to be generated.

3.3.1.1 Reaction temperature and shaking speed during incubation

The supplier (Roche) recommends evaluating a range of temperatures (25℃ to 37℃) to optimise protein expression, and therefore the reaction temperatures of 25, 30 and 35℃ were trialed. It was found that under these temperature conditions that there was no difference in the overall expression and thus the reaction temperature was set to 30℃ for all subsequent experiments, which is also similar to that used for *in vivo* expression of RNCs (Cassaignau A et al., 2014). It is also recommended that the reactions are incubated with shaking at 900 rpm, however it was found that at least with the reaction scale used (10 μl), there was no difference in the expression when the shaking speed was reduced to 750 rpm. The shaking speed, however, is an important consideration for preparative scale reaction volumes (i.e. 5 ml) that are required for RNCs for NMR using the RTS500 system, thus in line for the recommendations for such a reaction scale, the shaking speed was maintained at 900 rpm for 10 μl scale and changed at 750 rpm for the RTS500 5ml reaction for NMR spectroscopy (the latter speed reflecting speed limitations associated with the use of 50ml tubes required for these experiments).

3.3.1.2 Template concentration and the effects of additives

First, the effect of template concentration was examined. It was anticipated that the DNA concentration was important for optimal expression as the cell-free reaction depends, in part, on the amount of DNA template available and thus this was examined next. The amount of DNA template was therefore increased from the recommended concentration of 100 ng/µl (2% of the reaction volume) to 200 ng/µl (4%). It was found that when the DNA template was used at twice the original concentration, this improved the overall expression, as shown in Figure 3.3. The improvement in the expression is likely to result from the fact that the extent to which protein translation can occur depends to a very significant degree on the presence and quality of template; greater quantities of DNA can provide more opportunities for initiating transcription and thus in principle, generate more mRNA for translation.

In addition to the concentration of the DNA template, the nature of the vector was also explored as a possible contributor to improving protein expression. Specifically, the vector backbone for the RNCs (pLDC) contains a lac operon. Ordinarily under conditions of *in vivo* expression in *E.coli*, the lac operon is inhibited by the absence of lactose and in which a lactose-derivative IPTG can 'induce' protein expression. Thus to evaluate the possibility of the influence of including IPTG on protein expression, IPTG was added to the reaction at a concentration of 500 mM, together with an RNase inhibitor. The addition of IPTG, however, did not cause any effect to level of expression between (b) and (c) as observed in Figure 3.3. This may suggested, therefore, that transcription initiation was not affected by any effects associated with an inhibited lac operon.

To investigate obtaining cell-free sample, the effect of amino acid concentration was examined by small scale. Another component of the cell-free reaction and one that is also user-controlled is the addition of amino acids, which is recommended to be used at a concentration of 1mM with the RTS500 reaction scale. The amino acids were investigated because specifically for the production of isotopically-enriched proteins, isotopically-labelled amino acids are used in place of unlabelled amino acids. In addition, unlike the unlabelled amino acids in the Roche kit which individually, are at equivalent concentrations, the isotopically-labelled amino acids are supplied as a blend (Isotec, Sigma, A.15 in Appendix Chapter), and thus the amino acids are in different proportions. An understanding of the concentration effects of the isotopically-labelled amino acid mix is therefore important. It was found that when the concentration of labelled amino acids (^{15}N and ^{13}C) was increased from 1 mM (12% of the reaction volume) to 2 mM (24%) together with a fixed DNA concentration of 200 ng/μl, the expression of each of the isolated YFPs was comparable to that of reactions with 100 ng/μl DNA and 1 mM amino acids, which perhaps indicates that a maximum expression level is obtained under these conditions. Interestingly, it was found that the expression of the isotopically labelled YFPs also required the presence of an RNAse inhibitor in the reactions; the reason for this was not entirely clear, but it is possible that the labelled amino acids, which are derived from an algal lysate (Isotec, US) may contain trace amounts of RNAses that easily degrade mRNA. The presence of the RNAse inhibitor improved the mRNA lifetime within the reaction to ensure that enough mRNA was available for translation.

From the analytical trials systematically evaluating the components of the RTS500 cell-free system to enhance YFP expression, the conditions as described in Table 2.3 in Section 2.2.1 in Materials and Methods Chapter, gave rise to optimal expression of

isolated, isotopically labelled YFP.

3.3.2 Optimisation of conditions for the production of YFP-RNCs

Following the establishment of a set of conditions that permitted the overexpression of isotopically labelled isolated YFP, the ability to generate and recover YFP-RNCs was investigated next. RNC samples of YFP17, YFP34, and YFP55 were synthesised under the optimised conditions (as shown in Table 2.3 in Section 2.2.1 in Materials and Methods Chapter). In this instance, after 1 hour expression, each sample was loaded to a 30% (w/v) sucrose cushion to separate ribosome-bound NCs (present within the pellet) and free NCs (in the supernatant).

As demonstrated in Figure 3.4, it can be clearly observed that the released NCs for each YFP construct show a series of bands roughly approximating to the predicted molecular weights (Section A.4 in Appendix Chapter), in which YFP17, YFP34, and YFP55 are calculated to be 30, 32, and 34.1 kDa respectively. These released NCs were used to evaluate the bands observed for the corresponding ribosome-bound NCs of each of the YFPs. Typically, RNCs observed by SDS-PAGE are also characterised by an additional band-shift of approximately 17 kDa, which corresponds to the presence of the bound P-site prolyl-tRNA. Observing the tRNA-bound form of RNCs requires observation using low pH conditions (pH<6.0), and it is also highly sensitive to the high pH (pH 8.8) conditions used for electrophoresis (Cassaignau A et al, 2014). As these YFP-RNCs are in pH 8.8 loading dye, the tRNA-bound band-shift is not observable, however, the presence of ribosome-bound NCs is ascertained by the fact that the YFP is detected within the ribosomal pellet following the 30% w/v sucrose cushion. Together, therefore as observed in Figure 3.4, the molecular weights of the RNCs are identical to

those of the isolated proteins, which confirms that full-length YFP-RNCs are being produced within the RTS500 expression system.

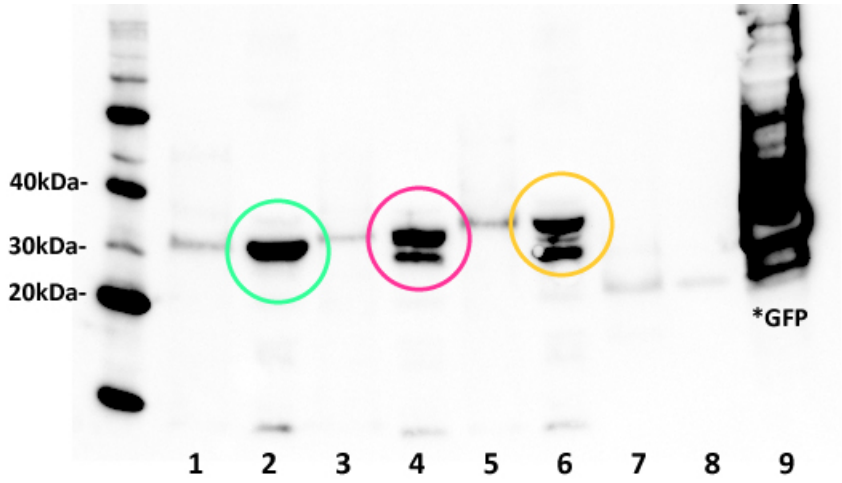


Figure 3.4 Cell-free RNC samples and free NCs of YFP17, YFP34, and YFP55 along with released GFP expressed using the equivalent conditions (1 hour at 30℃, and 900 rpm). Once synthesised by the cell-free expression system (RTS500), each sample was visualised by an anti-his immunoblot after purification of 30% w/v sucrose cushion. The immunoblot demonstrates that YFP17 (30 kDa, circled in green), YFP34 (32 kDa, circled in pink), and YFP55 (34.1 kDa, circled in yellow) were expressed and visualised at designated position. Lanes 1 and 2 demonstrate free NC and RNC of YFP17, lanes 3 and 4 demonstrate free NC and RNC of YFP34, and lanes 5 and 6 demonstrate free NC and RNC of YFP55 respectively. Lane 9 shows GFP (26.7 kDa) synthesised under same condition (not purified further by sucrose cushion), which served as a positive control.

3.4 Optimisation of a strategy to recover RNCs from cell-free reactions

3.4.1 A two step strategy for recovering RNCs

An important consideration in the study of RNCs is that the signals obtained by fluorescence or NMR spectroscopies reflect an exclusively ribosome-bound NC. Although NMR is a relatively insensitive technique, which requires micromolar concentrations of protein; the detection of small amounts of released NC can therefore be difficult to detect. On the other hand, fluorescence can detect nano and even femtomolar concentrations of protein by intrinsic fluorescence is possible as it is a highly sensitive technique. It is therefore possible that within the same sample using fluorescence detection, that small levels of released NCs can be detected as well as RNCs and thus it would be more difficult to distinguish the fluorescence arising from any released NC relative to RNCs. In order to be confident that the recovery of highly purified RNCs were free of any released species, the purification of the RNCs was evaluated further.

In the samples described in Section 3.3.2, a 30% w/v sucrose cushion subject to centrifugation conditions of 100,000 rpm, 90 minutes was used to separate the RNCs from the released NC. To ensure that this methodology was an effective separation strategy, a coloured dye was added to the top of a 30% w/v sucrose cushion and the diffusion of the dye was evaluated after centrifugation as demonstrated in Figure 3.5 (b). Given that the presence of the dye was localised to the top of the cushion, it suggests that a suitable separation between the pellet and supernatant has occurred. When this approach was applied to RNC samples (i.e. by adding a dye to the sample), there was

also very little diffusion of the supernatant, observed and thus the released NCs were confined to the top of the cushion and distinct from the species within the pellet. The RNC itself using a 10 μl reaction scale typically gave rise to a very small and clear gel pellet that was almost undetectable by eye.

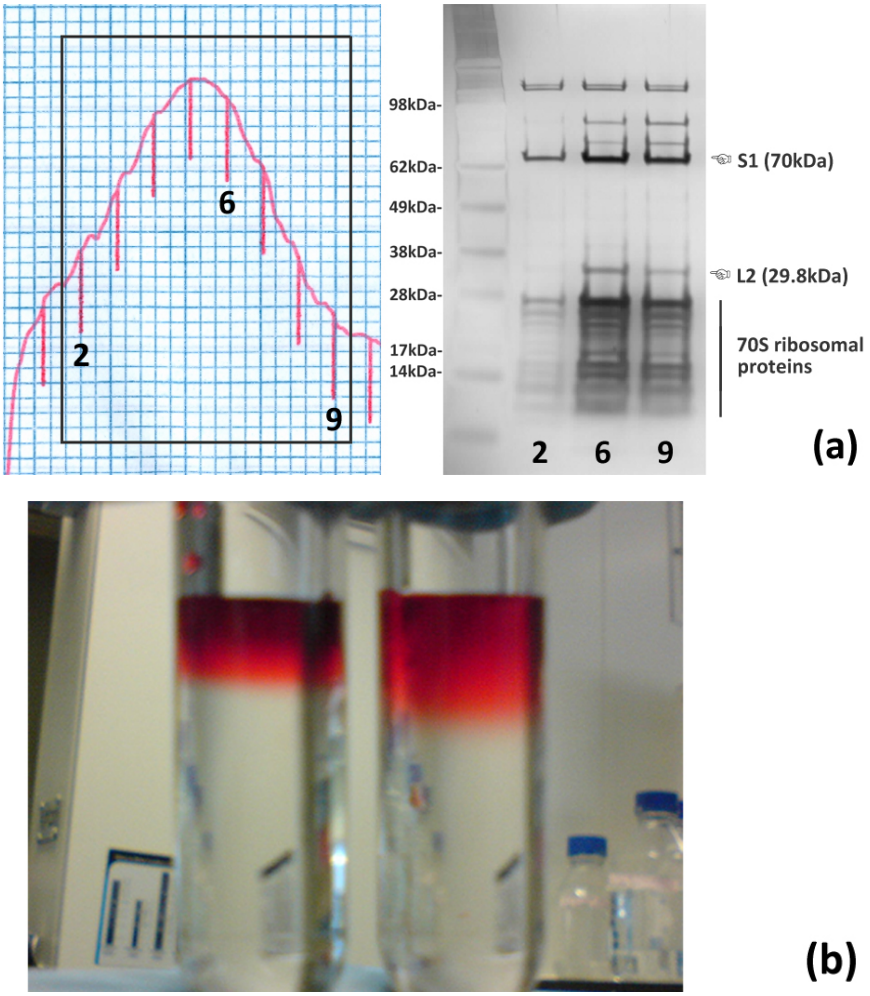


Figure 3.5 Biophysical aspects of RNCs. **(a) Sucrose fractionation profile via 10-35% gradient**. Sucrose gradient profile of 70S ribosome (RTS500) in tico-buffer and SDS-PAGE of fractions, showing different manner of ingredients. Fraction 6 shows the highest purity of 70S ribosome by existence of S1 and several ribosomal proteins. In contrast, fractions 2 and 9 show a slight impurity by missing several bands that are observed in fraction 6, e.g. S1, L2 and 70S ribosomal proteins. **(b) Sucrose cushion images before (left) and after (right) ultra high-speed centrifugation**. Before high-speed ultra sucrose cushion, the top layer maintains original volume 100 µl mixtures of cell-free reaction dissolved in red coloured tico-buffer. After the centrifugation, the top layer spreads free NCs on the layer, and RNCs and 70S ribosome pelleted.

3.4.1.1 Evaluating the recovery of RNC using immunoblot analysis

Although the centrifugation results using a dye showed successful separation, to evaluate further the possibility that a single centifugation step may not be sufficient for separating free NCs from the RNCs, an additional purification step was employed for each of the YFP-RNC lengths. As a control for the RNCs, isolated GFP was also used to evaluate the efficiency of separation.

In this instance, 20 µl cell-free reactions were performed and each of the samples were subjected to the cushion step to recover the pellets containing the RNCs and supernatant containing released NCs illustrated in Figure 3.6 (detailed further in Section 2.2.3 in Materials and Methods Chapter). The supernatant which was separated from the pellet, is referred to as 'f1' while the pellet was resuspended in 20 µl of tico-buffer. 10 µl of the resuspended was retained as 'r1' and the remaining 10 µl was then subjected to a second sucrose cushion step using same volume and conditions as in the previous step, to recover the second pellet 'r2' and supernatant 'f1'. The r2 pellet was resuspended pellet in 10 µl of tico-buffer. This was used to evaluate therefore, the efficiency of the additional sucrose cushion step in removing any potential released species (f1) that were carried over into the pellet (r1) in the first sucrose cushion immunoblot and intrinsic YFP fluorescence analyses (Section 3.4.1.2) were used.

As observed in Figure 3.6, each of the RNC samples showed significant differences after the first and second sucrose cushion steps. Using YFP17 as an example (Figure 3.4 (a)), after the first sucrose cushion step, released YFP17-NC and YFP17 were observed in fractions f1 and r1 respectively corresponding to a band of approximately 30 kDa. The almost equal proportions of protein in both fractions suggests that either significant

release of the NC occurs during centrifugation, or possibly, that there is a significant proportion of released species generated during expression. Following the second sucrose cushion step, the r2 band is clearly observable (circled in green), however there is no band in present in the f2 fraction. This suggests indicates therefore, that any released YFP17-NC is very effectively removed after the first centrifugation step and similarly, there appears to be no further release of NCs as a result of the centrifugation step itself. This pattern of behaviour was observed for each of the RNCs and strongly suggests overall, that there is very efficient separation taking place during purification

Intriguingly, the isolated GFP gave rise to some interesting results relative to the RNCs. When it was subjected to the same purification strategy as described above, a characteristic 30 kDa band was observed in both fractions s1 and p1 (equivalent to 'f1' and 'r1' respectively, in the RNC experiment) and similarly after the second centrifugation step, GFP was again recovered in s2 and p2 (equivalent to 'f2' and 'r2'). These results are most unusual, as it is not expected that GFP would be found in the ribosomal pellet p1, and similarly that after a second centrifugation step, that there should not be any GFP within either s2 or p2. The coloured dye experiment (Figure 3.5 (b) in Section 3.4.1) suggested that there was ample separation between a pellet and supernatant after even one centrifugation step.

In the case of GFP, its expression is far greater than that of the RNCs, and thus is possible that some diffusion of the GFP occurred during the first centrifugation step resulting in the presence of protein in p1. The fact that GFP is present again p2, following the second centrifugation step may suggest again, that separation was not efficient, however as roughly equal proportions of GFP were found in both s2 and p2, it suggests more likely that separation is not an issue, but rather, released GFP in high

concentrations, can associate with the ribosome. In light of this result, it highlights even further the absolute need for stringent separation of RNCs and released NCs of YFP.

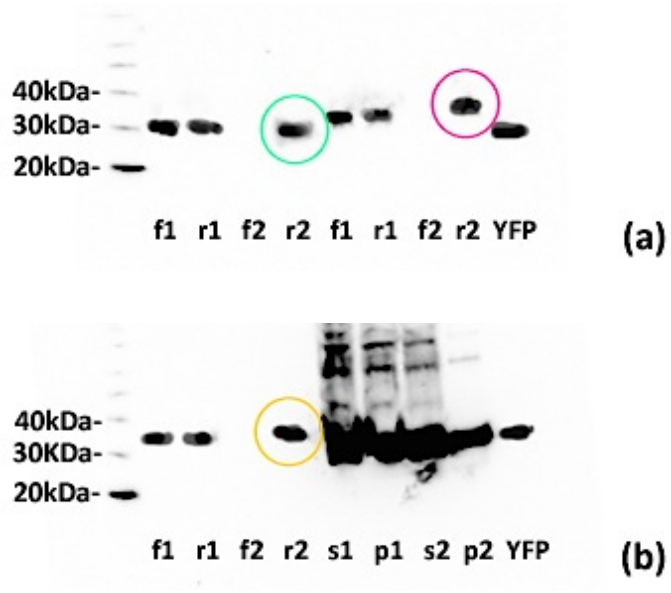


Figure 3.6 (a) Separation samples of YFP17 and YFP34 using a double sucrose cushion centrifugation step. To obtain highly purified RNC samples of YFP17 (30 kDa, circled in green) and YFP34 (32 kDa, circled in pink), two sucrose cushion steps were applied, and f1, r1, f2, and r2 samples were obtained and visualised by an anti-his immunoblot. f1 and f2 samples are the free NCs from first and second purification steps, r1 and r2 samples are RNCs from first and second purification steps respectively. (**b**) **Separation samples of YFP55** (34.1 kDa, circled in yellow) **and isolated GFP** (26.7 kDa) **samples using a double sucrose cushion centrifugation step**. Same purification steps were applied to YFP55 as YFP17 and YFP34. In this immunoblot, same separation step was repeated with isolated GFP expressed by RTS500 in same condition as s1 (supernatant 1), p1 (pellet 1), s2 (supernatant 2), and p2 (pellet 2). GFP construct was not added SecM sequence at the end of C-terminus, so the final product after double ultra high-speed sucrose cushion was not RNC but mature protein (p2). In order to estimate the occupancy of RNCs, 1µM (50 ng) concentration of mature YFP (wt) was also indicated (26.7 kDa, circled in grey). The r2 samples of YFP17, YFP34, YFP55 were finally applied and analysed for fluorescence spectroscopy as the final product of RNCs.

	Concentration of 70S	*Concentration of YFP (Raw data)*	*Occupancy of RNCs (%)*
YFP17 f1	-	1.36 µM (32.78)[*std:24.12]	-
YFP17 r1	2.02 µM	0.98 µM (23.61)	48.5%
YFP17 f2	-	0.01 µM (0.29)	-
YFP17 r2	1.79 µM	0.8 µM (19.20)	44.6%
YFP34 f1	-	1.03 µM (23.67)[*std:22.96]	-
YFP34 r1	1.99 µM	1.08 µM (24.87)	54.5%
YFP34 f2	-	0.003 µM (0.07)	-
YFP34 r2	2.38 µM	1.24 µM (29.86)	52.1%
YFP55 f1	-	1.06 µM (26.63)[*std:25.24]	-
YFP55 r1	2.01 µM	0.99 µM (25.01)	49.3%
YFP55 f2	-	0.01 µM (0.18)	-
YFP55 r2	1.86 µM	0.91 µM (22.94)	48.9%
1uM YFP	-	1 µM [*as std: 24.12, 22.96, 25.24]	-

Table 3.1 Summary of occupancies of RNCs after 1st and 2nd ultra high-speed sucrose cushion using as calculated using densitometry analysis of bands on an anti-his immunoblot (ImageJ software [129]).

To complement the immunoblot analysis in Figure 3.6 (a) and (b) purified, isolated YFP was used as a control to estimated the occupancies of the YFP-RNCs based on the concentration of YFP and 70S ribosome, using densitometry measurements by Image-J software [129] together with the concentration of ribosome as measured by absorbance at 260 nm. These results as presented in Table 3.1 describes that the occupancies of the YFP constructs were on average 50% for r1 and 48% for r2 respectively. This suggests the ribosomes are generally well occupied with NCs, but more importantly, the second centrifugation step does not result in any significant loss of ribosome-bound NCs, and thus the integrity of the RNCs following this purification strategy can be assured.

3.4.1.2 Evaluating the recovery of RNCs using intrinsic YFP fluorescence

To complement further the immunoblot analysis, the intrinsic yellow fluorescence each of the YFP samples, f1, r1, f2, and r2, was analysed in which the fluorescence emission intensity of each of the samples was recorded at 527 nm following the excitation of the YFP fluorophore at 514 nm. Figure 3.7 demonstrates the raw intensities of each sample for YFP17, YFP34, and YFP55 along with the fluorescence intensity of purified YFP at different protein concentrations from 50 nM to 3 μM. Examining the raw intensities, it can be seen that there were notable differences in the recovery of fluorescence in each of the YFP-RNCs before and after the two centrifugation steps. Overall, the r2 samples showed a decreased fluorescence intensity across the all the RNC samples and in addition there was a very weak fluorescence observable in the f2 samples, suggest that either the second purification step removed trace amounts of f1 from the sample, or that perhaps additional release of the NCs was occurring during the second purification step. To explore these data further, the intrinsic yellow fluorescence at 527 nm was also then used a measure of the occupancy of the

RNCs as described in Table 3.2.

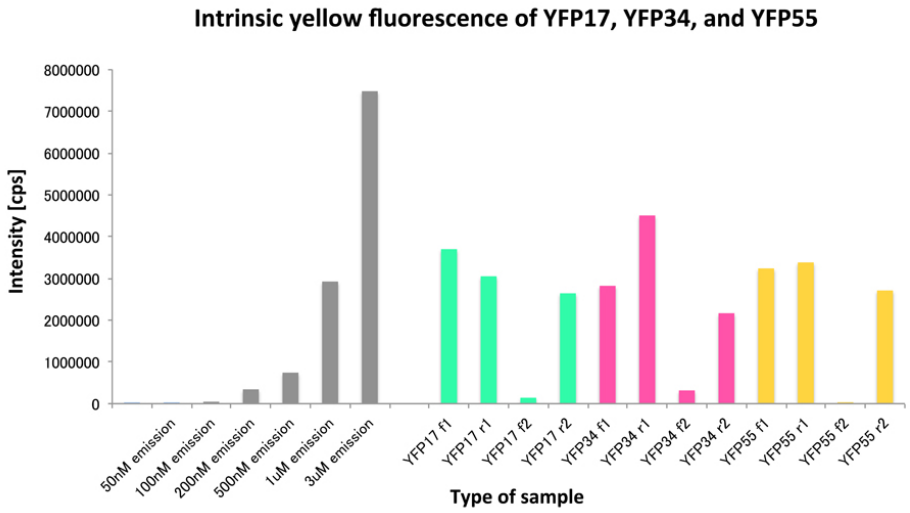


Figure 3.7 Intrinsic yellow fluorescence of YFP17, YFP34, and YFP55 comparing to the standard curves of mature YFP in different concentration. Each sample: f1, r1, f2, and r2 samples purified by double ultra high-speed sucrose cushion, was excited at 514 nm and collected its emission peaks at 527 nm.

Table 3.2 demonstrates the occupancy of r1 and r2 samples as detected by intrinsic yellow fluorescence. These results show that apart from YFP34 which showed a large discrepancy in the occupancies of r1 and r2 samples of 94.3% and 39.4% respectively, YFP17 and YFP55 showed minor differences between r1 and r2 samples, respectively of about 63.9% and 63.1% (YFP17) and 70.7% and 62.1% (YFP55). Indeed in both the YFP17 and YFP55 samples, there is the observation of fluorescence in f2. When the fluorescence intensity observed in r2 is added to f1, this roughly equates to the fluorescence observed in r1 (Table 3.2) suggesting that the RNCs have remained stable during purification. However in the case of YFP34 a similar result is not obtained, in which f2 and r2 equate to approximately 1.14 µM, compared to the original r1 sample of 1.87 µM. It possible that perhaps in YFP34 it is a less stable RNC sample and thus is

subject to more release during purification. The disparity observed in r1 and f2 and r2 might reflect possible aggregation of the released NC which may therefore result in a lower intensity value being observed.

Comparing the results obtained here and in the immunoblot (Section 3.4.1.1), it is clear that there are significant differences in the RNC occupancies obtained for the two strategies. For example, YFP17 showed 48.5% for r1 and 44.6% for r2 samples in occupancy detected by immunoblot. On the other hand, the construct showed 63.9% for r1 and 63.1% for r2 samples in occupancy detected by intrinsic yellow fluorescence. Occupancies as detected by fluorescence were typically higher compared to analysis by immunoblot and in addition, it appeared to be more sensitive, as there was trace fluorescence detected in the f2 samples (Figure 3.7), which were not obvious as detected by immunoblot (Figure 3.6). The analysis of fluorescence arising from the RNC, is however, prone to more error compared to that of the released NC. This is because the fluorescence of the RNC is likely to be influenced by the conformation of the RNC, while in the released NC, it would be expected that that YFP is folded and thus fluoresces in an equivalent manner. As a result, both methods were thus used to evaluate the RNCs, in which the occupancy was typically evaluated by immunoblot, but the presence of any released species was detected using sucrose cushion and measuring the fluorescence of the released fraction.

	Concentration of 70s	*Conc. (Intensity) of fluorescence*	*Occupancy of RNCs (%)*
YFP17 f1	-	1.55 μM	-
YFP17 r1	2.02 μM	1.29 μM	63.9%
YFP17 f2	-	0.127 μM	-
YFP17 r2	1.79 μM	1.13 μM	63.1%
YFP34 f1	-	1.2 μM	-
YFP34 r1	1.99 μM	1.87 μM	94.3%
YFP34 f2	-	0.2 μM	-
YFP34 r2	2.38 μM	0.94 μM	39.4%
YFP55 f1	-	1.37 μM	-
YFP55 r1	2.01 μM	1.42 μM	70.7%
YFP55 f2	-	0.08 μM	-
YFP55 r2	1.86 μM	1.15 μM	62.1%

Table 3.2 Summary of occupancies of RNCs after 1st and 2nd ultra high-speed sucrose cushion detected by intrinsic yellow fluorescence at 527 nm by fluorescence spectroscopy.

The results of the RNC occupancies following the second sucrose centrifugation step also demonstrate the overall stability of the RNCs. Indeed, the values indicate that the YFP-RNCs retain high occupancies, as there was no observation of any further released NC after the second purification step. The first and second purification steps retained the very positive effect for obtaining occupancy of highly purified RNCs, r2 samples, and this is evident from the analyses of immunoblot and fluorescence. Even though there was a difference between immunoblot and fluorescence detection methods for the occupancies, an average of these values was taken. Therefore it can be estimated that YFP-RNCs have been recovered have occupancies of: 53.85% for YFP17, 45.75% for YFP34, and 55.5% for YFP55. (percentage was obtained by; (% detected by imunoblot + % detected by fluorescence) / 2). In total, YFP55 showed the highest occupancy based on this calculation method.

3.5 Analysis of YFP-RNCs using fluorescence-based techniques

3.5.1 Intrinsic yellow fluorescence analysis of YFP17, 34 and 55

After the cell-free expression and recovery of the YFP-RNCs was optimised, the fluorescent properties of the RNCs were evaluated in a quantitative manner to understand how the fluorescence could be used a probe for the folded conformation in each of the RNC lengths. Based on previous results in Tables 3.1 and 3.2, the double sucrose cushion purification step was used to recover the r2 samples for YFP17, YFP34, and YFP55 and these were used in the following analyses.

Figure 3.8 (a) and (b) illustrates respectively, the emission spectra and the intensities obtained for the intrinsic yellow fluorescence of each RNC samples, alongside samples of with 250 μM isolated YFP and 1 μM of purified 70S ribosome. It can be clearly observed that as the length of linker sequences increased, the intensities of intrinsic yellow fluorescence also increased respectively. For instance, YFP17 showed c.a. 6,000 cps of intrinsic yellow fluorescence at 527 nm where peak of emission spectra observed. Similarly, YFP34 showed over 8,000 cps of intrinsic yellow fluorescence at 527 nm, and YFP55 showed c.a. 10,000 cps of intrinsic fluorescence. A negligible fluorescence signal was detected from 70S ribosome (less than 500 cps), which is expected, as YFP's fluorophore is both unique and distinct. When the emission intensity of the RNC is compared to an isolated YFP concentration standard of 250 nM (8,000 cps) it appears that YFP34 and YFP55 have fluorescence intensities that are nearly similar (within error) to that observed for 250 nM, while for YFP17, the intensity is approximately 25% lower.

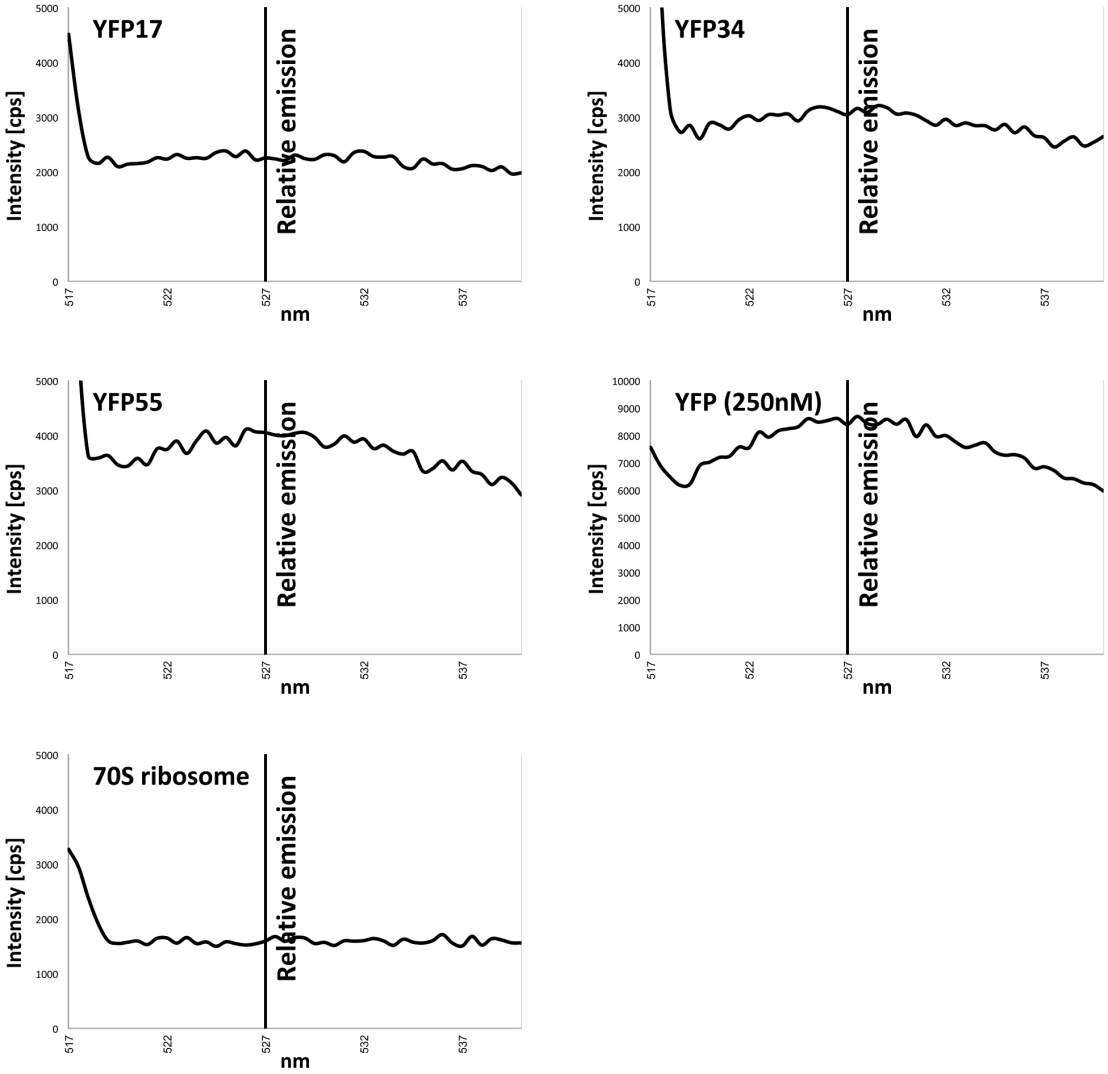


Figure 3.8 (a) Absorption raw spectra of YFP17, YFP34, YFP55, 250nM YFP, and 70S ribosome. Each sample was excited at 514 nm and was collected the spectra range between 515 to 540 nm. The emission peak, λ_{em}: at 527 nm which was highlighted as 'Relative emission' at designated position 527 nm.

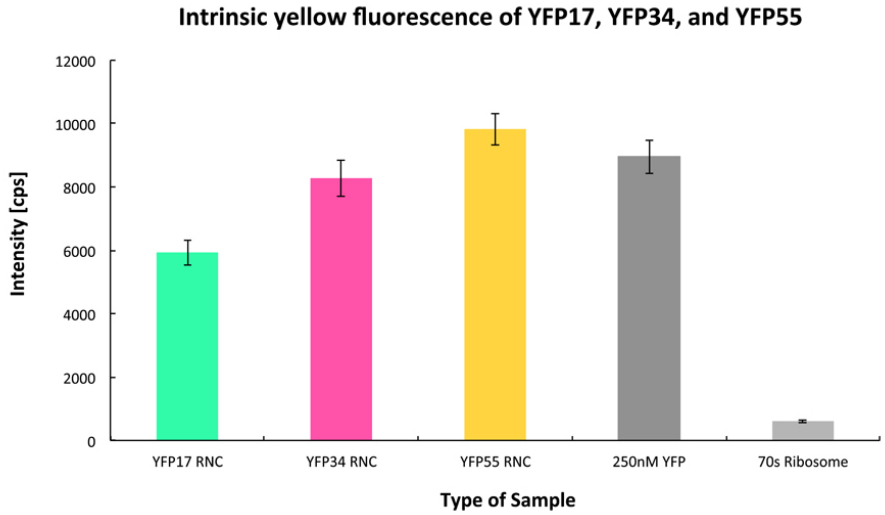


Figure 3.8 (b) Intrinsic yellow fluorescence of YFP17, 34, and 55. After the double sucrose cushion purification step, the RNCs corresponding to YFP17, YFP34, and YFP55 were excited at 514 nm and collected its emission, with the peak of fluorescence intensity recorded at peaks of 527 nm. A relative intensity of intrinsic yellow fluorescence can be observed from YFP34 and YFP55. On the other hand, YFP17 which was not expected to show its intrinsic yellow fluorescence demonstrated a clear intensity about 6,000 cps as demonstrated in Figure 3.7.

The differences observed in the intrinsic yellow fluorescence intensities appear therefore to be an intrinsic property of the conformation of RNC, rather than the concentration of the RNCs itself, as the NC occupancy and RNC concentration was consistent across all samples (Table 3.1). The NCs of YFP in each RNC is therefore in a different environment at each linker length, and therefore it is possible that the YFP either shows partial folding, or there are other events taking place on the ribosome which alters the environment of the fluorophore within the YFP, for example, the NC is interacting with the ribosome's surface and may quench the fluorescence. In such scenario, YFP17 could in this instance be considered as being partial folded.

3.5.2 Extrinsic fluorescence analysis of YFP17, YFP34, and YFP55

To explore the relative differences in fluorescence intensity in each of the RNCs which may relate to partial folding, the extrinsic fluorescence properties of each RNC was evaluated using the dye bis-ANS. bis-ANS is a dye which binds specifically to hydrophobic patches in proteins and is often used as a probe to evaluate partially folded states, such as molten globules [130, 131]. bis-ANS was added to each RNC, as well as samples of YFP alone and YFP in the presence of 70S ribosome, and emission spectra ranging from 425 nm to 540 nm was collected, following excitation at 390 nm in Figure 3.9.

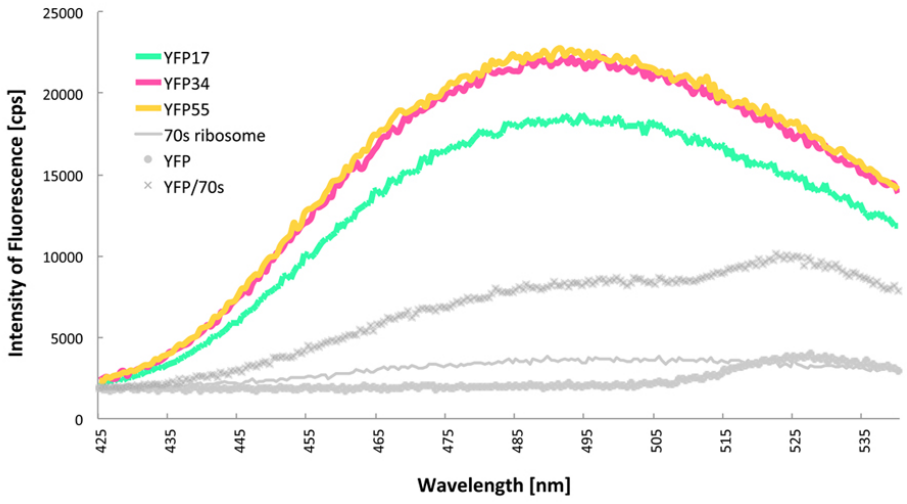


Figure 3.9 Extrinsic fluorescence of YFP17, YFP34, and YFP55. RNC samples, r2, after the double sucrose cushion steps were labelled by bis-ANS which interacts with hydrophobic surface of the protein for 30 minutes incubation on ice. Each RNC was excited at 390 nm and collected its emission spectra between 425 to 540 nm range. To observe the behavior of RNCs, 1 μM of 70S ribosome, mature YFP, and mixture of YFP and 70S ribosome was also treated in same manner and excited and collected their emission spectra.

By examining the peak of fluorescence at 490 nm, the relative difference in bis-ANS binding can be evaluated for each of the RNCs. It can be seen that YFP alone and 70S ribosomes show fluorescence intensities of 2,000 cps and 3,600 cps, indicating a low affinity towards ANS. It is expected that because YFP is folded, that there would be very little binding of bis-ANS, which is observed. In the RNCs, however, there was enhanced binding of the bis-ANS. Based upon the intrinsic yellow fluorescence from RNCs of YFP17, YFP34, and YFP55 in Section 3.4.1.2 to 3.5.1, it has been shown that YFP17 can fluoresce, suggesting that at this linker length, YFP17 has the capacity to adopt a native-like fold, or that it is possibly in equilibrium with a partially folded state in which enough of a population of native structure is available such that the fluorophore can be formed.

The capacity of YFP17 RNC to bind bis-ANS therefore, may suggest that folding to the native state is incomplete, such that there may be some exposure of hydrophobic surfaces on the molecule (Figure 2.5 in Section 2.7.2 in Materials and Methods Chapter); this is possible, as 11 amino acids of YFP17 would be expected to be present within the tunnel. In contrast to YFP17, both YFP34 and YFP55 showed enhanced ANS fluorescence about 22,000 and 22,500 cps respectively. This was an unexpected result because if YFP17 shows the propensity to adopt a native fold, then as the linker extends, the expectation is that capacity for YFP to fold to the native state would increase, and thus its affinity for ANS would decrease. Indeed, the intrinsic yellow fluorescence demonstrated in Figures 3.7 and 3.8 indicate that as the linker length increases, the fluorescence intensity also increases. The capacity for the RNCs to bind bis-ANS therefore does not exclusive relate to the presence and conformation of YFP alone; it is possible that the linker may be forming a type of structure that may bind to bis-ANS, which is evident from the similar fluorescence values obtained for YFP34 and YFP55.

3.5.3 Anisotropy analysis of YFP17, YFP34, and YFP55

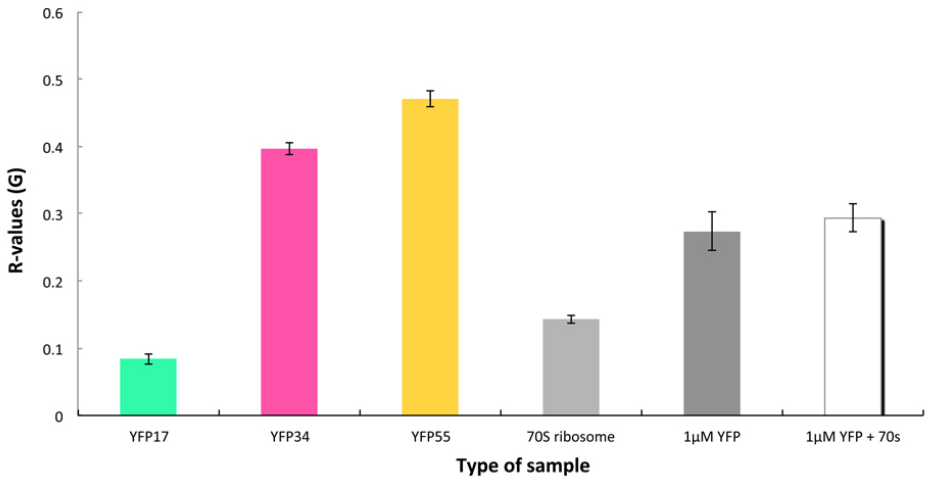


Figure 3.10 Fluorescence anisotropy of RNCs of YFP17, YFP34, and YFP55. R-value for each construct based on G-value obtained by the application of the polarised light in four different angles, I_{HH}, I_{HV}, I_{VH}, I_{VV}, is evaluated and demonstrated. 1 μM of 70S ribosome, mature YFP, and the mixture of YFP and 70S ribosome were also examined by the anisotropy method, and showed R-values respectively.

To understand the differences in fluorescence behavior of the RNCs further, the dynamics of the RNCs was examined using fluorescence anisotropy measurements. Using polarising filters and a method described in Section 2.7.3 in Materials and Methods Chapter, spectra for the calculation of I-, G-, and R-values for each of the constructs were collected by excitation at 514nm and collecting the fluorescence emission at 527 nm.

Using the Perrin equation, the R-values were evaluated as shown in Figure 3.10 (described in Figure 2.4 in Materials and Methods Chapter 2.7.3). As shown in this figure, it is shown that as the linker sequences between YFP (wt) and SecM increased,

or the distance between the NCs of YFP and the exit tunnel of 70S ribosome extended, R-values also increased in each RNC construct. For example, YFP17 showed the lowest R-value which is approximately less than 0.1. On the other hand, YFP34 had an R-value of approximately 0.4, and YFP55 a value of 0.48. In contrast, isolated YFP in the presence and absence of 70S ribosomes had lower R-values of 0.28 to 0.295 respectively, while the ribosome alone had an R-value of 0.15; this value is excepted to be low as the ribosome shows very little fluorescence intensity following an excitation at 514 nm.

The increase in anisotropy values that have been observed for the RNCs appeared to correlate with the linker length, and as a result, reflects the rotational properties of the NC. The higher R-values observed for YFP34 and YFP55 (0.4 and 0.48 respectively) suggest that NCs of YFP at these linker lengths have significant motional freedom compared to YFP17 (0.1). These data are also therefore likely to describe the environment of the YFP; in YFP17, the NC is close to the exit tunnel of the 70S ribosome, while YFP34 and YFP55 are at a distance from the ribosome and correlate well to the linking sequences associated with these RNC constructs in Figure 3.2. The R-value for YFP55 was significantly different to that of isolated YFP, with values of 0.48, the longest linker RNC and isolated YFP (0.28, respectively) showed a large difference. It would be anticipated that the isolated YFP would have the higher R-value as it is not tethered to the ribosome. The fact that it does not, may suggest that perhaps a higher order conformation of YFP exists in solution; indeed YFP is known to naturally dimerise [78], and this may give rise to the ‘lower’ R-values. The similar R-values obtained for isolated YFP in the presence of ribosomes (0.295) may also suggest a potential interaction with the ribosomal surface.

3.6 NMR studies of the RNCs of YFP34 and YFP55

3.6.1 Cell-free preparation of RNC samples for NMR spectroscopy

To complement and enhance the biophysical studies of the RNCs, NMR studies were next considered, in order to provide residue-specific information on both the structure and dynamics of the RNCs. Initially, two RNCs, YFP34 and YFP55 were selected for study as it was expected that the linker lengths would offer favourable motional properties as suggested by the anisotropy studies (Figure 3.10), which are important for ensuring that the NC has enough flexibility to be observable by NMR.

Using 1ml reaction volume using an S30 system (in a collaboration with RIKEN SSBC group, Section 2.2.4 in Materials and Methods Chapter) and a 5ml RTS500 reaction (Section 2.2.5 in Materials and Methods Chapter), isotopically labelled RNCs of YFP34 and YFP55 were generated respectively and were purified as described in Sections 2.2.4 and 2.2.5 in Materials and Methods Chapter, and the series of optimisation in Figures 3.3 to 3.6. To evaluate the integrity of the samples, the fluorescence emission spectra of RNCs were evaluated and gave rise to intensities of 5,427 cps for YFP34 and 13,768 cps for YFP55 respectively. When taking into the account the ribosome concentrations which were 0.85 μM for YFP34 and 2.7 μM for YFP55 respectively, the occupancies of YFP34 and YFP55 calculated by intrinsic yellow fluorescence were about 18% for YFP34 and 45% for YFP55. Based on these occupancies, the concentrations of NC were 0.15 μM for YFP34 and 1.2 μM for YFP55 respectively.

3.6.2 2D properties of YFP (wt) and YFP34 and YFP55 in ^{1}H - ^{15}N

To initiate NMR studies of YFP-RNCs, isolated YFP (wt) labelled by ^{15}N and ^{13}C was prepared and evaluated using 2D SOFAST NMR, and the resonances were identified using the published backbone assignment of YFP (wt) [120, BMRB Entry: 15826]. Venus [81], a variant of YFP (also used in these RNC studies), was already assigned [120, BMRB Entry: 15826], so that the assignment of Venus enabled the spectra of the purified YFP to be evaluated. Figure 3.11 shows 2D spectrum of isolated mature YFP (wt) in ^{1}H - ^{15}N and its corresponding assignment. Within this 2D correlation spectrum, a number of well-solved cross peaks can be identified: 153 cross peaks in the ^{1}H - ^{15}N spectrum as demonstrated in Figure 3.11. Similarly, 2D spectrum of isolated mature YFP (wt) in ^{1}H - ^{13}C in which 93 well-resolved cross peaks could be observed and assigned confidently is demonstrated in Figure 3.13 in Chapter 3.63. Of these cross peak, 21 are observable in the methyl region (1.5 to -1.0 ppm in ^{1}H and 10 to 30 ppm in ^{13}C), which is typically associated with folded structure (Figure 3.13). The observable resonances in both the ^{15}N and ^{13}C correlation spectra were used to subsequently study the structural behaviour of YFP34 and YFP55 RNCs.

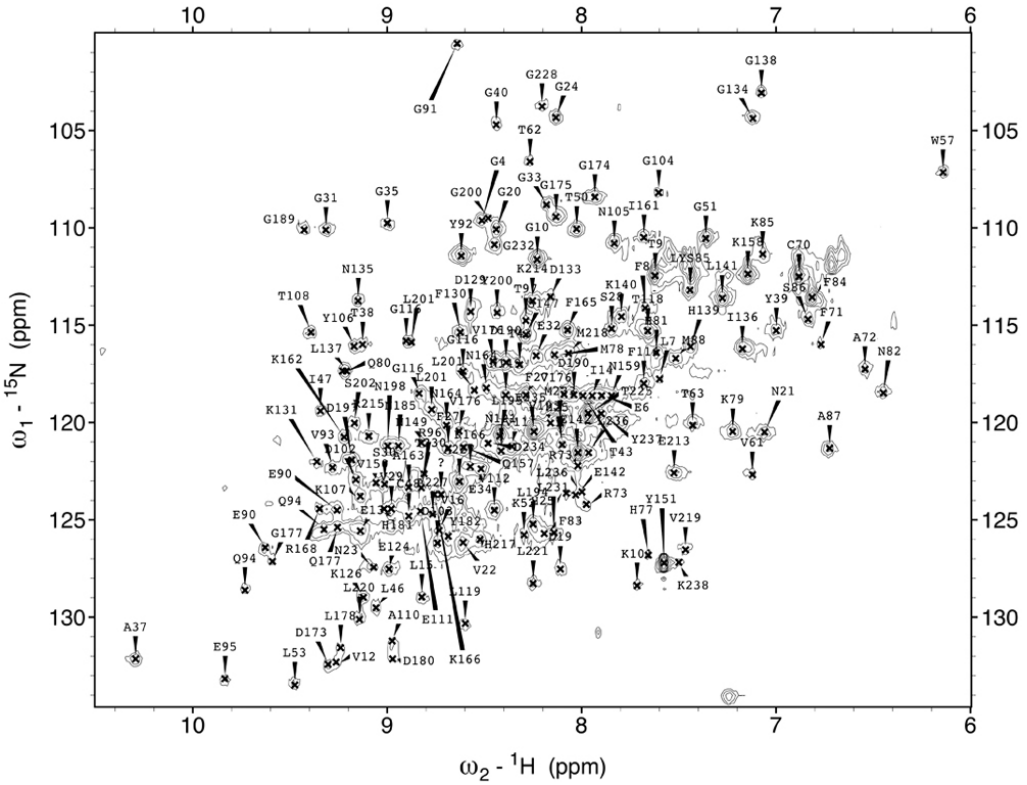
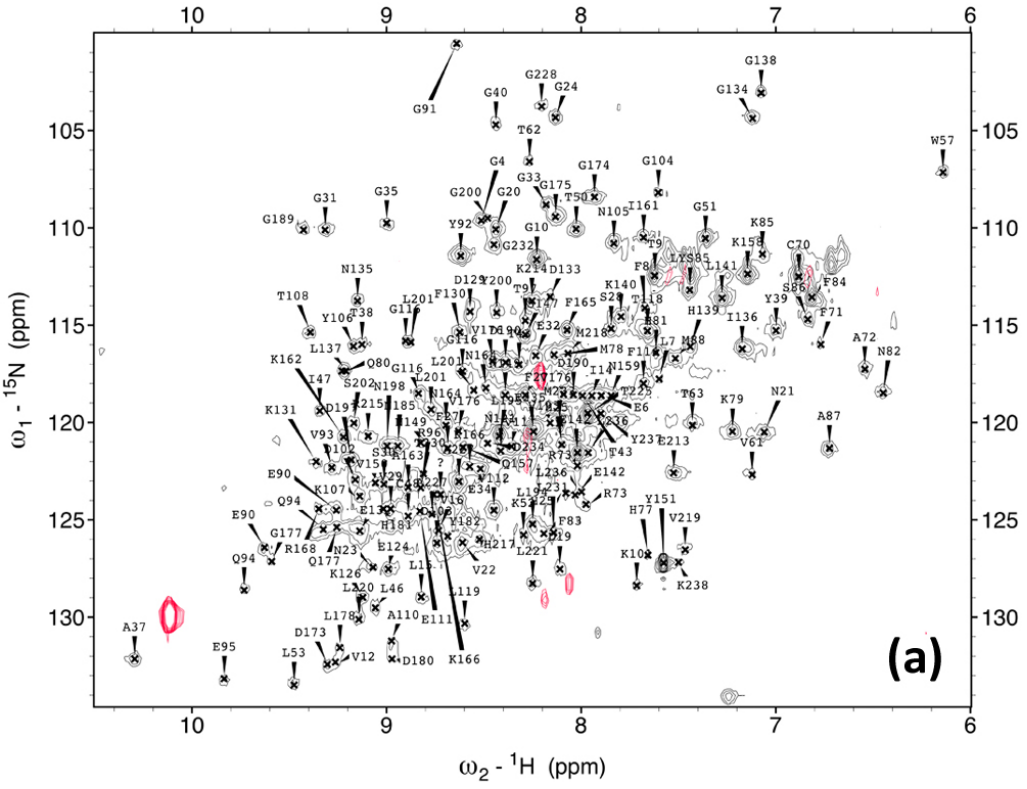


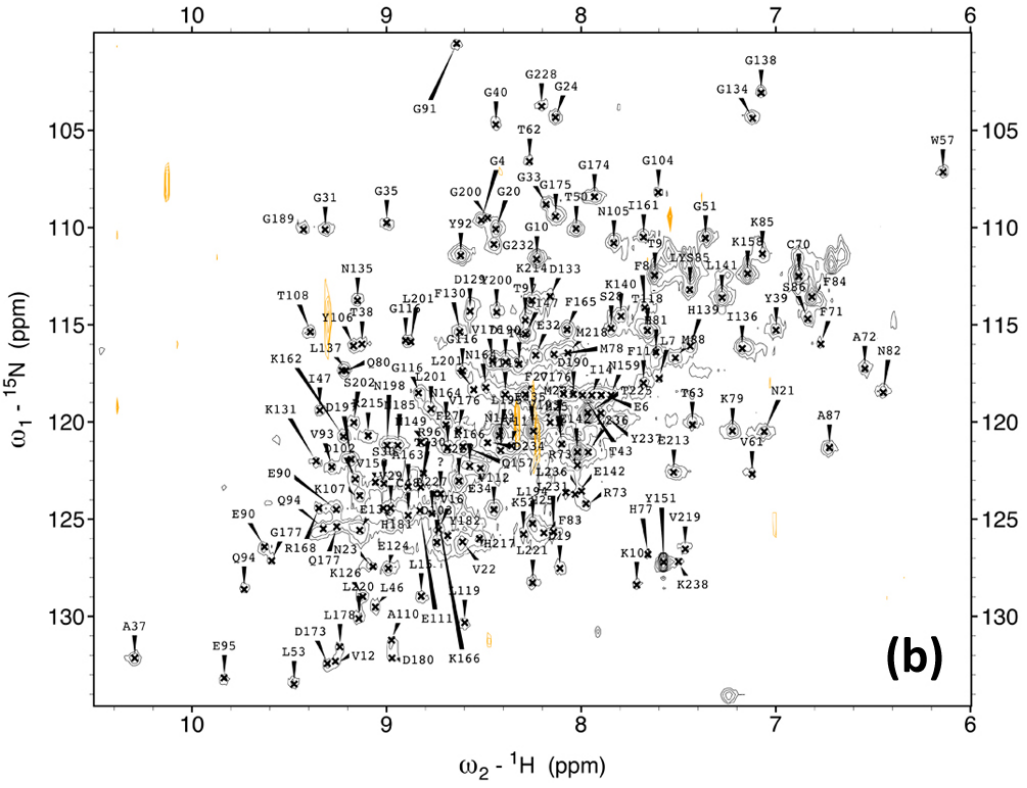
Figure 3.11 ^{1}H - ^{15}N 2D spectrum of YFP (wt). Using assignment of Venus (PDB code: 1MYW, BMRB entry 15826), each cross peak of mature YFP (wt) was assigned by the Sparky software. 153 cross peaks were observed and assigned in the range of 10.5 to 6 ppm in ^{1}H and 100 to 135 ppm in ^{15}N (this figure fully covers the ranges of ppm). The YFP was at a concentration of 50 μM and recorded at 25℃.

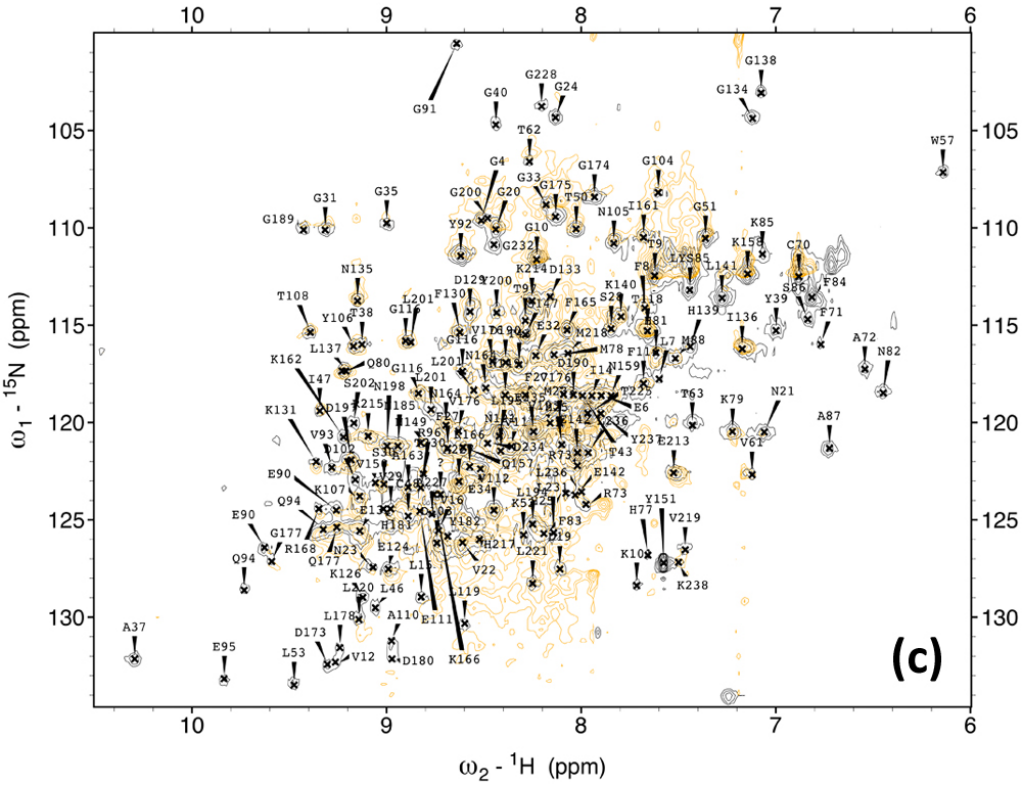
3.6.2.1 Structural analysis of 2D ^{15}N SOFAST-HSQC of YFP-RNCs by NMR spectroscopy

Using the assignment of the isolated YFP (wt), the ^{15}N labelled RNCs of YFP34 and YFP55 were evaluated by comparison as shown Figure 3.12 (a) and 3.12 (b), in which YFP34 is shown in pink and YFP55 in yellow. The data for each spectra was recorded in 3 hours at 25℃, and the final concentration for each sample was 50 μM for YFP (wt), 0.85 μM for YFP34, and 2.7 μM for YFP55 respectively. In both the RNCs, weak intensity signals within the correlation spectra were observed, and in which only few cross peaks of YFP34 (5 resonances, Figure 3.12 (a)) and YFP55 (9 resonances, Figure 3.12 (b)) could be unambiguously observed. It appears however, that these RNC resonances did not overlay with the isolated YFP (wt) spectrum, even though the released NC of YFP55 overlaid with isolated YFP (Fig 3.12(c)). The high intensity of these observable cross peaks, which are also dispersed within a narrow range in the ^{1}H dimension suggests that they are very likely arising from the disordered flexible linker.

The absence in the RNC spectra of any resonances arising from YFP was unexpected. It is possible that the low RNC occupancies, giving rise to NC concentrations of 0.15 μM for YFP34 and 1.2 μM for YFP55 respectively. Another possibility is that there are interactions between the YFP-NCs and the surface of the ribosome. Such an interaction would effectively immobilise YFP and result in broadening beyond detection of observable signals. The low occupancy as observed by intrinsic yellow fluorescence of YFP34 was also mirrored in 2D correlation spectrum, in which weak cross peaks, particularly in ^{1}H - ^{15}N sample was observed, as described in Figure 3.12 (a). In this figure, only 5 cross peaks could be readily identified and assigned in ^{1}H - ^{15}N spectrum in final sample was 0.15 μM.


(a)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)


(b)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)


10
9
8
7
6
105
110
115
120
125
130
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)
(c)

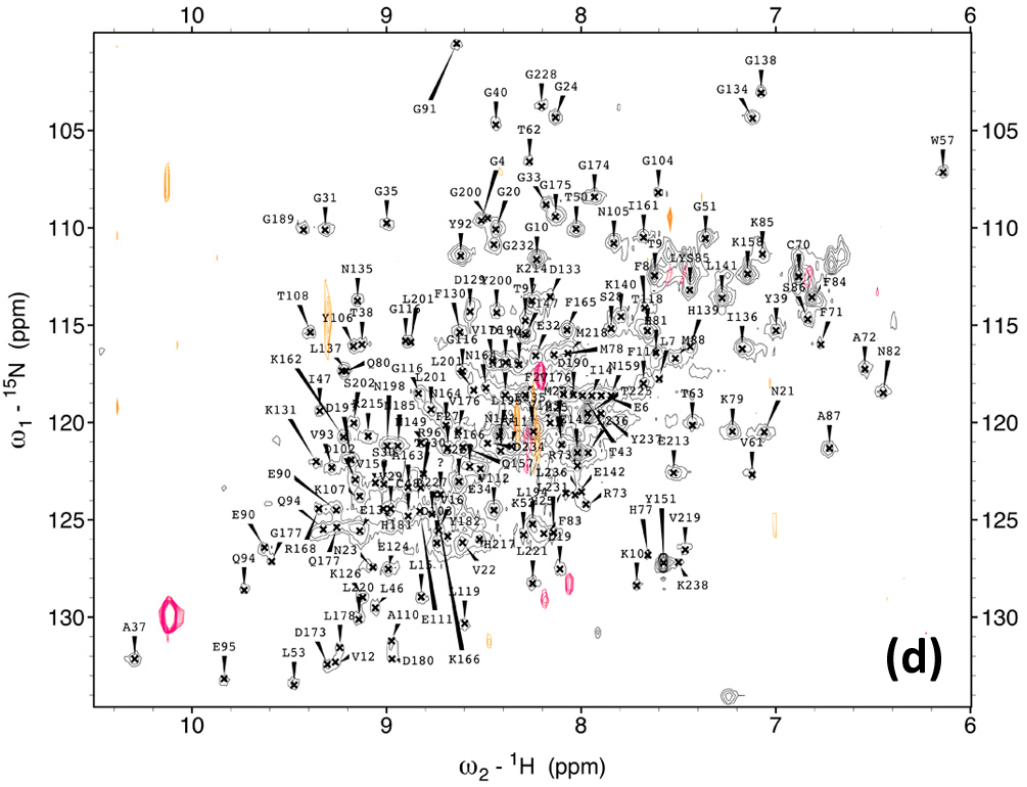


Figure 3.12 1H - 15N 2D spectra of each YFP-RNCs. (a) 2D spectrum of YFP34 (pink). Due to the low occupancy of YFP34 in final sample, only 5 peaks were observed and demonstrated. Using the assignment of Venus, each observable cross peak within the RNC spectra was compared by attribution. (b) 2D spectrum of YFP55 (yellow), which was similarly compared by attribution as for YFP34. (c) 2D spectrum of NC of YFP55 (yellow) overlaid to 2D spectrum of YFP (wt, grey). This sample was purified from in vivo RNC purification (by the courtesy of Dr Lisa D Cabrita). (d) 2D spectra of YFP34 and YFP55 overlaid with YFP (wt).

While the spectrum of YFP34 was very weak, the 2D spectrum of YFP55 was relatively evaluated and found to maintain different spectral properties. As shown in ^{1}H - ^{15}N spectrum in Figure 3.12 (b), YFP55 shows 9 cross peaks. In total, these 5 and 9 cross peaks from the RNCs did not correspond with both isolated YFP (wt) in Figure 3.11 and the free NC of YFP55 (Figure 3.12 (c)) that shows a number of cross peak similarity to YFP (wt). (NC of YFP55 was purified from *in vivo* RNC purification, by courtesy of Dr Lisa D Cabrita). 2D spectra of YFP34 and YFP55 in ^{1}H - ^{15}N were also compared to the spectrum of the isolated YFP (wt) in Figure 3.12 (d). Both YFP34 and YFP55 show an overall lower intensity in the cross peaks as they were evaluated in very low concentrations in 0.85 and 2.7 μM. Also, the actual concentrations of NC were 0.15 μM for YFP34 and 1.2 μM for YFP55 respectively. When overlaid these 2D spectra of YFP34 and YFP55 with the isolated YFP (wt) spectrum, these 5 and 9 cross peaks are observable in ^{1}H - ^{15}N spectra of YFP 34 and YFP55.

However, none of these cross peaks do not behave corresponding to folded cross peaks from YFP (wt). This may indicate that NCs of YFP attach to the surface of 70S ribosome which is consisted of the gigantic molecular weight, as a result, NCs of YFP34 and YFP55 did not correspond well to show magnetic frequency to the spectroscopy due to the strong interaction to the 70S ribosome, and the flexible linker sequences of RNCs which were also labelled by ^{15}N and ^{13}C isotopes caused the magnetic resonances as well as NCs. This could be considered as the difference of the linker sequences as illustrated in Figure 3.2; 5 cross peaks for YFP 34 which is shorter than YFP55 that showed 9 cross peaks. This unsimilarity of 2D spectra from two RNCs was also observed in ^{1}H - ^{13}C spectra, and will also be described in next chapter.

3.6.3 2D properties of YFP (wt) and YFP34 and YFP55 in ^{1}H - ^{13}C

Although the spectra of ^{15}N labelled YFP34 and YFP55 were too weak for a thorough analysis, the enhanced sensitivity possible in the carbon dimension would enable 2D spectrum of ^{1}H - ^{13}C labelled RNCs to be observed. As demonstrated in Figure 3.13, ^{1}H - ^{13}C 2D spectrum of YFP (wt) was also evaluated with Venus assignment [120, BMRB Entry: 15826]. In this figure, 93 of well-solved cross peaks of YFP (wt) were observed and assigned corresponding to Venus. This 2D spectrum was also used to subsequently study the structural behaviour of YFP34 and YFP55.

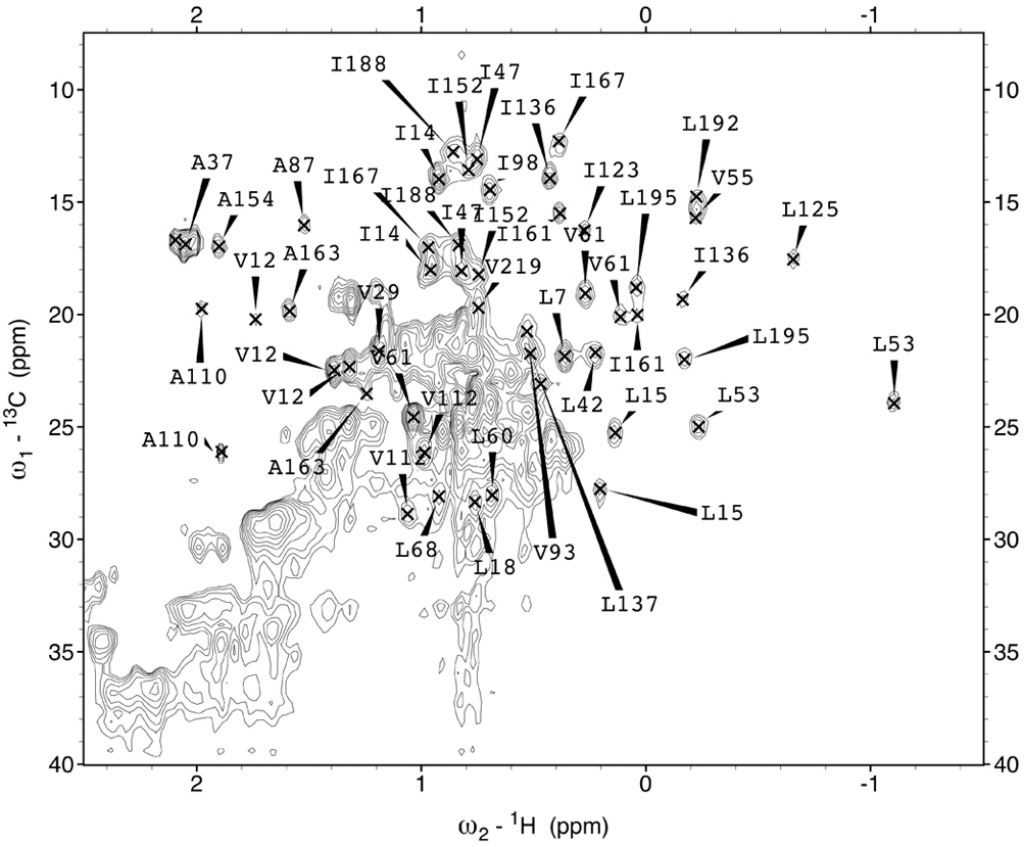
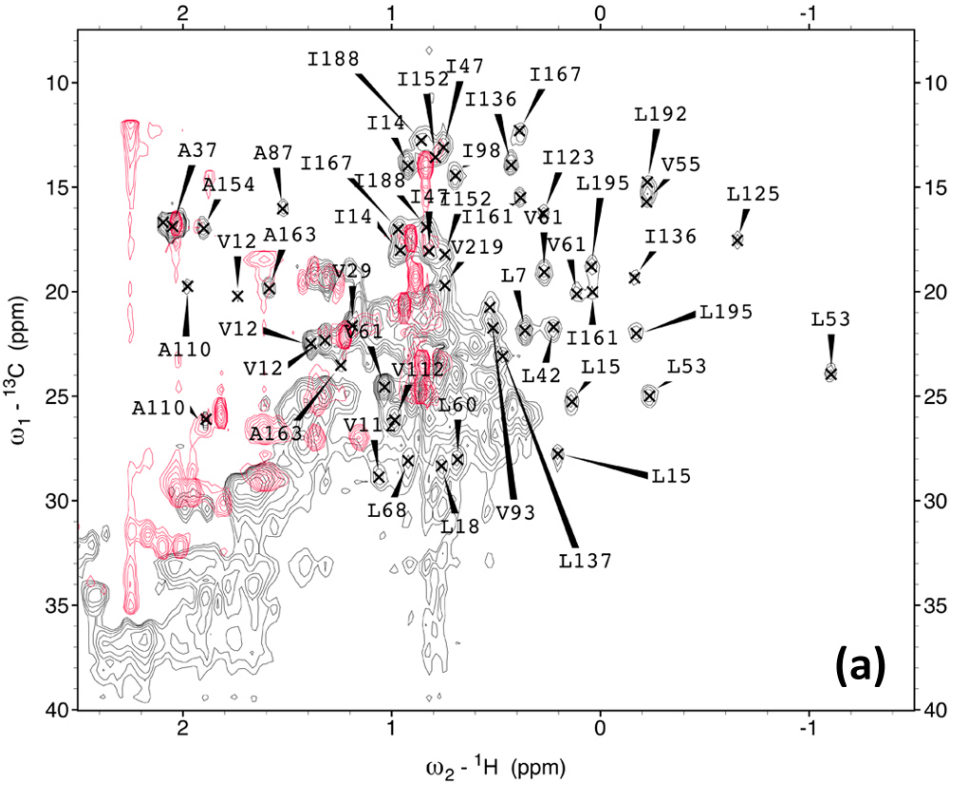
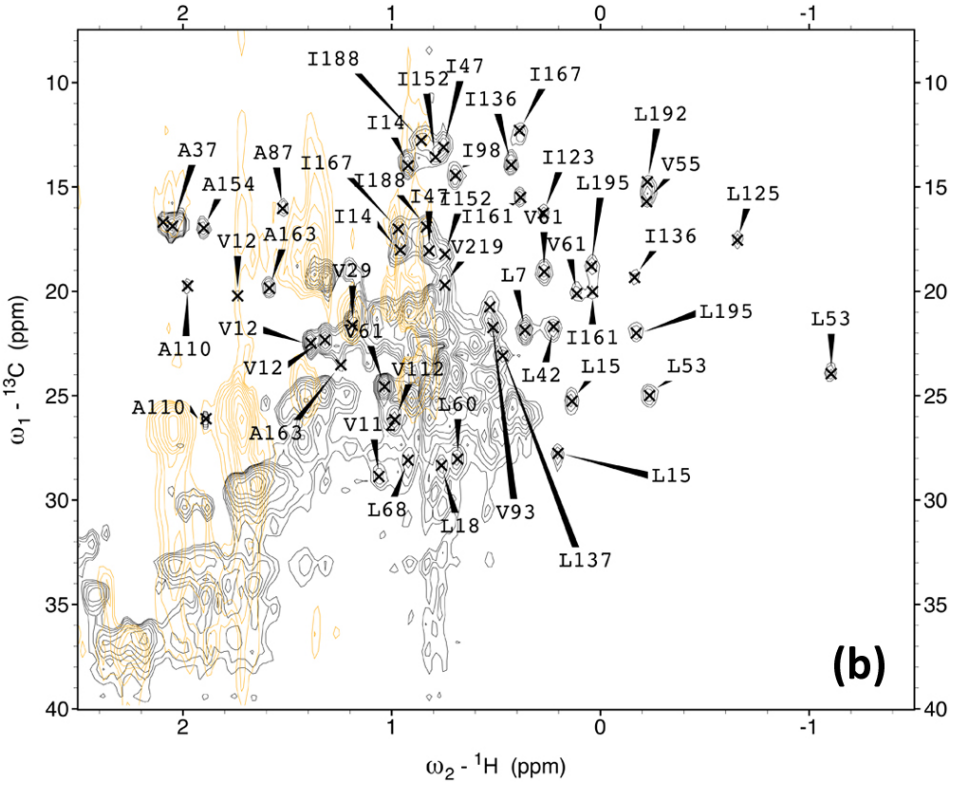


Figure 3.13 ^{1}H - ^{13}C 2D spectrum of YFP (wt). Using assignment of Venus (PDB code: 1MYW, BMRB entry 15826), each cross peak of mature YFP (wt) was assigned by the Sparky software. 93 cross peaks were observed and assigned in the range of 2.5 to -1.5 ppm in ^{1}H and 7.5 to 40 ppm in ^{13}C covering the methyl-group containing residues of YFP (this figure fully covers the ranges of ppm).

3.6.3.1 Analysis of 2D ^{13}C HMQC of YFP-RNCs by NMR spectroscopy

In the spectrum of isolated YFP (wt) in Figure 3.13, 21 cross peaks which are commonly observed in the fully folded proteins are observable in the methyl region in the range of 1.5 to -1.0 ppm in ^{1}H and 10 to 30 ppm in ^{13}C in Figure 3.13. On the other hand, the spectrum of YFP34 demonstrated in Figure 3.14 (a) shows a large number of cross peaks located in the region between 2.0 and 0.0 ppm in ^{1}H dimension, and 10 and 35 ppm in ^{13}C dimension, a region associated with disordered structure. A similar set of resonances also within a near identical region of the spectrum was also evident in YFP55 in Figure 3.14 (b). In neither the spectra for YFP34 or YFP55 were any resonances observed within the methyl region of the spectrum, which is consistent with the observations in the ^{15}N spectra. It must be also noted that the 2D spectra of YFP34 and YFP55 in both ^{1}H - ^{15}N and ^{1}H - ^{13}C were recorded in different environments and facilities, i.e. YFP34 recorded at RIKEN, Yokohama (900 Mhz, Varian) and YFP55 recorded at UCL (700 Mhz, Bruker). Together these results indicate that the observable resonances for the RNC arise from the disordered linker region and as observed in the ^{15}N correlation spectra, that the absence of folded resonances arising from YFP is likely to be the result of interactions with the ribosomal surface.


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V61
V29
L7
V12
V61
L195
L53
A110
I161
V12
V112
L42
L15
L53
A110
L60
A163
V112
L15
L68
V93
L18
L137
(a)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V61
V29
L7
V12
V61
L195
L53
A110
I161
V12
V112
L42
L15
L53
A110
L60
A163
V112
L15
L68
V93
L18
L137
(b)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)

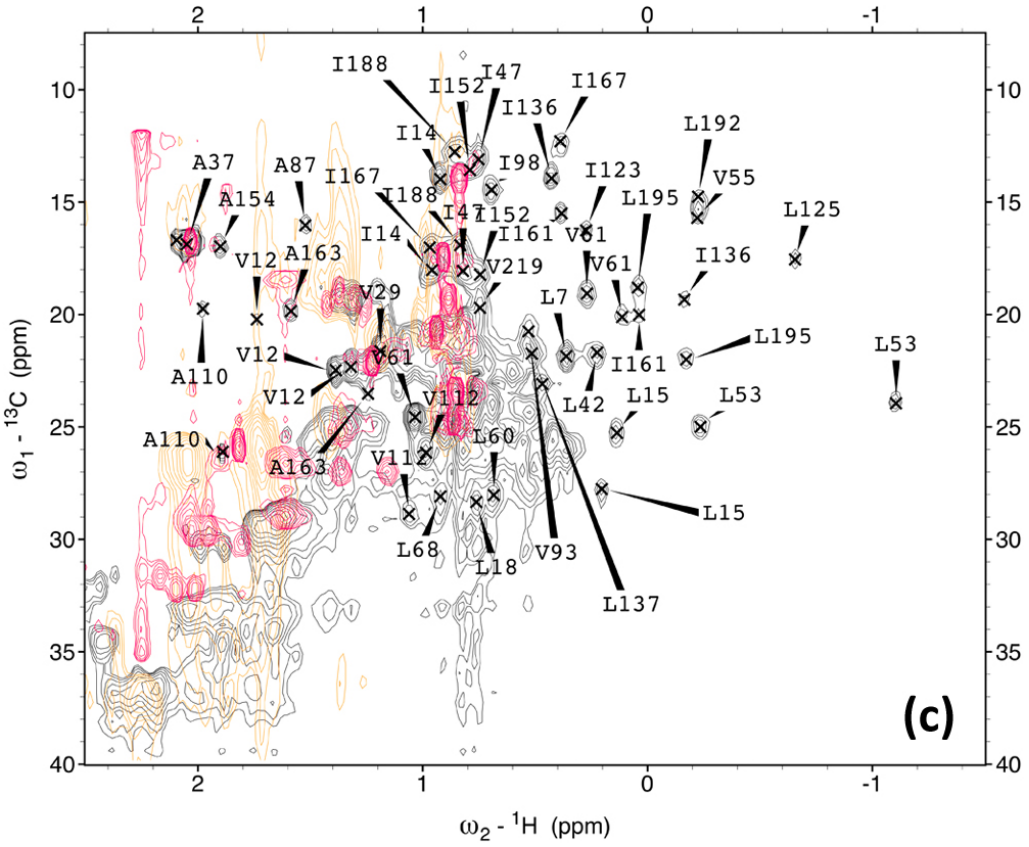


Figure 3.14 ^{1}H - ^{13}C 2D spectra of each YFP-RNCs. (**a**) 2D spectrum of YFP34 (pink). (**b**) 2D spectrum of YFP55 (yellow). Due to the low occupancy of YFP34 and YFP55 in final samples, a few peaks was observed and demonstrated. (**c**) 2D spectra of YFP34 and YFP55 overlaid on YFP (wt) with Venus assignment.

As observed in ^{1}H - ^{15}N 2D spectra of RNCs in Figure 3.12, most cross peaks of YFP34 and YFP55 in ^{1}H - ^{13}C did not show a similarity to YFP (wt) which is completely folded, and some of the peaks obtained and recorded could be considered as the resonances that were caused the magnetic field from the labelled linker sequences. However, YFP34 shows more than 30 observable cross peaks and this is more than in case of YFP55 which shows less than 30 peaks. Having an idea that the enhanced possible sensitivity of the carbon dimension by 2 or 3 methyl binding between ^{1}H and ^{13}C which results in more cross peaks than ^{15}N, this does not correspond to the case of

^{1}H - ^{15}N 2D analysis in Chapter 3.6.2.1. This must be, however, noticed that 2D spectra of YFP34 and YFP55 in both ^{1}H - ^{13}C and ^{1}H - ^{15}N were recorded in different environments and facilities, i.e. YFP34 recorded at RIKEN, Yokohama (900 Mhz, Varian) and YFP55 recorded at UCL (700 Mhz, Bruker).

3.6.4 Structural analysis of YFP-RNCs by NMR spectroscopy

These ^{1}H - ^{15}N and ^{1}H - ^{13}C spectra of YFP34 and YFP55 (Figures 3.12 (a), 3.14 (a), 3.12 (b), and 3.14 (b)), do not show a noticeable similarity to the 2D spectra of isolated YFP (wt) evaluated in Figures 3.11 and 3.13. Only 2D spectrum of released NCs of YFP55 in ^{1}H - ^{15}N described in Figure 3.12 (c) shows a noticeable similarity to 153 of identified cross peaks which is considered as folded YFP (wt). This sample was obtained from an *in vivo* preparation, and the correlation spectrum of the released YFP55 shows that the NC is folded. This result may indicate that YFP can acquire a native fold when they bound to the 70S ribosome, because both YFP34 and YFP55 showed a clear intrinsic yellow fluorescence (Figures 3.7 and 3.8).

In Figure 3.12 (d), 5 for YFP34 and 9 for YFP55 cross peaks in ^{1}H - ^{15}N spectra were observed, however a similarity to 2D spectrum of YFP (wt) overlaid is difficult to distinguish. Further, cross peaks from YFP34 and YFP55 do not correspond each other, therefore a similarity between 2D spectra of YFP34 and YFP55 can also not be found. In ^{1}H - ^{13}C 2D spectra in Figure 3.14, it can be shown that some of the cross peaks from YFP34 and YFP55 show a similarity located in the region between .0 and 0.6 ppm in ^{1}H dimension, and 13 and 26 ppm in ^{13}C dimension in Figure 3.14 (c). Apart from these regions, some of corresponding cross peaks can be also found in full range of this 2D spectrum. Nevertheless, most cross peaks do not correspond to the region assigned by

YFP (wt) in both ^{1}H - ^{15}N and ^{1}H - ^{13}C, and this could be considered as the first assumption that these 2D spectra were obtained from the linker sequences which were also isotopically labelled.

Overall, these data indicate that the sensitivity of the YFP-RNCs is overall low in NMR spectroscopy. This may be due to the relatively low occupancy of the RNCs, but another possibility is that the RNCs interact with the large ribosomal surface which may interfere the magnetic resonance of NCs. In total, these results obtained from the combined NMR results in ^{15}N and ^{13}C do not support the observation demonstrated by intrinsic yellow fluorescence (Figures 3.7 and 3.8) that at the RNCs state the NCs of YFP34 and YFP55 are folded.

3.7 Discussion and concluding remarks

3.7.1 Observing intrinsic yellow fluorescence from YFP17 construct

This study showed the development of a cell-free system for the generation homogeneous, isotopically-labelled YFP-RNCs of different lengths and in which the integrity of the samples was confirmed using biochemistry and intrinsic yellow fluorescence. This enabled both fluorescence and NMR to study the structural and dynamical characteristics of YFP as it is being synthesised on the ribosome. These data indicate that YFP can indeed form a folded structure whilst bound to the ribosome when it is extended 36 and 57 amino acids from the PTC (YFP34 and YFP55 respectively).

When YFP is separated from the PTC by 19 amino acids (YFP17), the fluorescence data showed that it was able to adopt a native-like structure as shown by intrinsic yellow fluorescence, but that perhaps a population of partially-folded species was also present, as shown by the reduced fluorescence intensity as well as and bis-ANS binding. The ability for YFP17 to fluoresce was unexpected, as the ribosomal exit tunnel is expected to hold 30 amino acids in an extended form [132, 133] and thus at least 11 amino acids from YFP are within the tunnel as demonstrated in Figure 3.15.

It was found that the sensitivity of the isotopically-labelled YFP-RNCs was overall very low for NMR spectroscopy; the YFP itself was not readily observable as a bound-NC, but only upon release from the ribosome. When bound to the ribosome, however, only the linker region in the RNC is observable. This effect may be due to the relatively low occupancy of the RNCs, but another possibility is that the YFP-RNCs themselves interact with the large ribosomal surface which may interfere with the

signals arising from the NCs. Indeed, the latter is consistent with the fluorescence anisotropy data (Section 3.5.3), which may suggest potential interactions, as well as the double sucrose cushion, which showed that large amounts of the related FP or GFP can associate with the ribosome (Section 3.4.1). Despite NMR being unable in this instance to provide reliable structural data on the RNCs, the intrinsic yellow fluorescence experiments (Figures 3.7 and 3.8) demonstrate that YFP has the capacity to adopt native structure close to the ribosomal exit.

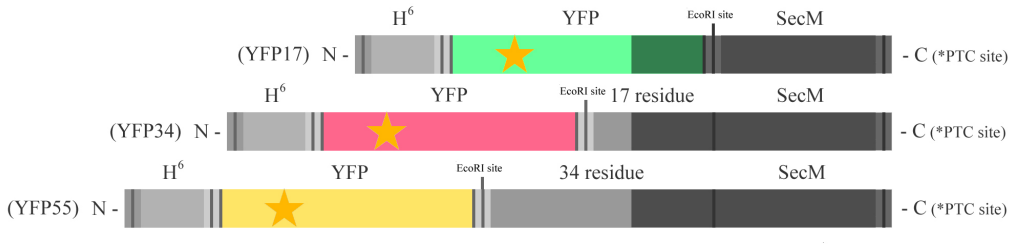


Figure 3.15 YFP-Provisional image of RNCs as expected to feature within the 70S ribosome. When bound to the ribosome, part of the NC will reside within the ribosomal exit tunnel as the tunnel can hold 30 amino acids in an extended form. Shown is a schematic of each of the YFP constructs and the regions of the NC that are expected to be present in the tunnel (shown by the shaded area) and those that are outside. In YFP55 and YFP34 the entire YFP is expected to be outside of the tunnel, however in the case of YFP17, up to 11 amino acids are expected to be buried.

Previous studies also described the deletion of 11 amino acid residues from C-terminus resulted in the loss of intrinsic green fluorescence in GFP [134] and thus the expectation was of a similar result to be observed in YFP17, i.e. 'zero' fluorescence. The fact that there is observable fluorescence intensity in YFP17 is indeed a surprise, as at least 11 amino acids of YFP are expected to be retained within the exit tunnel of the 70S ribosome, yet YFP in this construct is able to fold. The results presented in Figures 3.5

and 3.6 confirm that the fluorescence observed in YFP17 does not arise from the presence of released YFP within the sample and rather, an inherent property of the RNC. As previously shown for studies of isolated GFP fluorescence, deletion of 11 [134] or 12 [135] amino acid from the C-terminus resulted in the complete loss of intrinsic green fluorescent behavior, because the protein is not able to adopt its native fold at that state. However as shown by a related study of isolated GFP study the deletion of 10 [136] amino acid from the C-terminus (expressed in the cells) resulted in a decrease in 56% intrinsic green fluorescence compared to mature full length EGFP [137].

There is, therefore, a narrow tolerance for truncations within the FP family, but it appears that in YFP on the ribosome, the absence of 11 amino acids can be tolerated. These suggest therefore that there are possible differences either between YFP and GFP or indeed in the capacity for YFP to fold on the ribosome compared to *in vitro*. To explore these prospects, in the next chapter, the relationship between co-translational folding of YFP and the appearance of intrinsic yellow fluorescence will be evaluated using YFP C-terminal truncations as characterised by fluorescence and NMR spectroscopies.

Chapter 4 Result II - C-terminal truncations of YFP

Development of an *in vitro* model for co-translational protein folding of YFP using C-terminal truncations

4.1 Protein folding studies and truncations

As mentioned in Sections 1.1.1 to 1.2.3 in Introduction Chapter, the process of co-translational protein folding is the process by which proteins fold within living systems, yet the mechanisms by which this process occurs at a structural and dynamical level remains to be understood. While RNCs presents a folding scenario which more closely resembles the process as it occurs within cells, their study can be challenging as they are difficult to produce. The use of C-terminal truncations, however, is an alternative and highly complementary strategy that mimics the progressive emergence of a NC and importantly, reports on the type of structure and conformational changes that would be possible.

4.1.1 C-terminal truncations and previous studies

Previous studies have used C-terminal truncations to describe the length dependence of folding in of a number of model systems, such as the 153 residues protein, apomyoglobin, which is comprised of an all α-helical structure. A series of C-terminal truncations of different lengths of 36, 77, and 119 amino acids were evaluated by far-UV circular dichroism (CD) spectroscopy, Fourier transform infrared (FT-IR) and fluorescence [138]. This study revealed that low amounts of random coil existed at all lengths, and that at a short length a non-native β-sheet structure was adopted. This was found to have the propensity to misfold and adopt amyloid-like properties. As the

fragment length increased to 119 amino acids, the β-sheet structure interconverted to the native α-helical structure, as observed in the full-length protein. This study also showed that folding as observed within the increasing protein fragments was distinct from urea denaturation, in which only α-helical structure was present at all stages during folding/unfolding of apomyogloblin [138]. A similar set of observations were found in fragments of the 64 residue Chymotrypsin inhibitor 2, also showed that at a length of 53 to 60 amino acids, molten globule-like structures were observed, and which were not evident in folding pathway under denaturing conditions [139]. It was only when the polypeptide chain was at a length of 62 to 63 amino acids that a native-like structure was observed, presumably due to the correct packing of side-chains.

Studies of other small proteins with αβ structures (staphylococcus nuclease and barnase) have shown different types of folding behaviour and in which case several residues at the C-terminus were required for native-like folding to be observed. For example, C-terminal fragments of the 149 residue staphylococcus nuclease comprising of 103, 112, 128, and 136 amino acids were each found to have decreasing levels of nuclease activity as the length of the polypeptide chain decreased. This correlated with CD spectroscopy, which revealed secondary structure in all four C-terminal truncations s, which decreased as the fragment length was shortened. This study showed that native-like structure formation was dependent on the length of the chains, in which less than 136 residues were required [140]. Barnase (110 amino acids), was found to have residual secondary structure in fragments of 22 and 79 amino acids, arising from the first α-helix$_1$. When the chain was increased to 105 amino acids, there was an ensemble of conformations in which intermediate, parallel secondary and tertiary structure formation was observable by CD and fluorescence. In addition, characterisation by NMR showed that that the α-helix$_1$ was forming non-native interactions as the chain

length increased, however in the presence of 105 amino acids, the structure was gradually stabilised to complete the final native fold [141].

Overall, these studies described above illustrate how C-terminal truncations can be used to map the chain length dependence. Importantly, identifying the types of structure that are possible as observed using C-terminal truncations will assist in correlating and understand folding as it occurs on the ribosome. Indeed, as shown with denaturation studies of firefly luciferase [39] and with antibody recognition of NCs of the P22 tail-spike protein [142], the structure formed by a protein during folding *in vitro* can differ to that observed on the ribosome.

Furthermore, possible differences between *in vitro* folding as described by C-terminal truncations and folding on the ribosome are currently being explored by studies of isolated Dom5 (Waudby CA, Karyadi ME et al). The protein has been subjected to a series of truncations in which 2-21 amino acids have been progressively deleted. Studies using CD spectroscopy and NMR have revealed that the protein can completely unfold when 9 amino acids have been removed from the C-terminus, corresponding to its most C-terminal beta-strand. However, in the absence of 8 amino acids, the protein can begin to form structure and in the absence of between 4 to 6 amino acids, the protein adopts a native-like intermediate. These studies indicate that that isolated Dom5 can support a significant degree of structure in the absence of its complete sequence. In contrast, NMR studies using a series of Dom5-RNCs of different lengths, however, indicate that entire sequence of Dom5 must be at a significant distance from the PTC prior to acquiring structure (Cabrita LD, Launay H et al). These studies of isolated Dom5 and of Dom5-RNCs represent the first structural insight into co-translational folding and further comparisons using studies of this kind are needed.

4.1.2 Aims of the chapter

As shown in Chapter 3, the co-translational folding properties of YFP was explored by generating RNCs of increasing linker lengths, enabling the protein to be captured at different stages of folding during its emergence from the ribosome. Intriguingly, YFP17, which was expected to be unfolded (Section 3.5.1 in Result 1 Chapter), showed that it was in fact able to adopt native structure, as shown by the observation of intrinsic yellow fluorescence.

Therefore, to complement the studies of the YFP-RNCs presented in Chapter 3, C-terminal truncations were generated. These were studied to explore the requirement of minimum amino acid sequence for YFP to acquire its native fold structure that provides the intrinsic yellow fluorescence, in order to further understand the molecular mechanisms of its folding *in vitro* and how it occurs on the ribosome as demonstrated by the three types of RNCs. A biophysical approach using fluorescence spectroscopy and NMR was applied in combination to probe the structural characteristics of the C-terminal truncations.

4.2 Production of C-terminal truncations of YFP as a model for co-translational folding

4.2.1 Design of C-terminal truncations of YFP

Based on previous results described Sections 3.1 to 3.5 in Result 1 Chapter, C-terminal truncations of the isolated YFP protein were created to understand further how YFP acquires its biologically-active structure as it emerges from the ribosome during synthesis. The idea underlying the C-terminal truncations is that these constructs would permit 'a vanishing point' of intrinsic yellow fluorescence on YFP sequence to be observed, and would therefore report the minimum amino acid sequence of YFP that is required for the protein to adopt its native fold. Using site-directed mutagenesis, 9 C-terminal truncations were generated by inserting a stop codon at different distances from the C-terminus (K238), to progressively shorten the YFP protein. From the C-terminus between 11 and 24 amino acids were removed, resulting in constructs named C11, C12, C13, C14, C15, C16, C17, C24 (Figure 2.2 in Materials and Methods Chapter 2.1.3). The C24 truncation represents the largest number of amino acids removed from the C-terminus which correspond to the final β-strand of the YFP structure. To complement these contructs, an additional C-terminal truncation f(0), was created in which the fluorophore-associated residues (65 - 67) were removed. The removal of the fluorophore enables the YFP to remain stable and monomeric but is unable to fluoresce and therefore serves as a control for observing fluorescence in the C-terminal truncations.

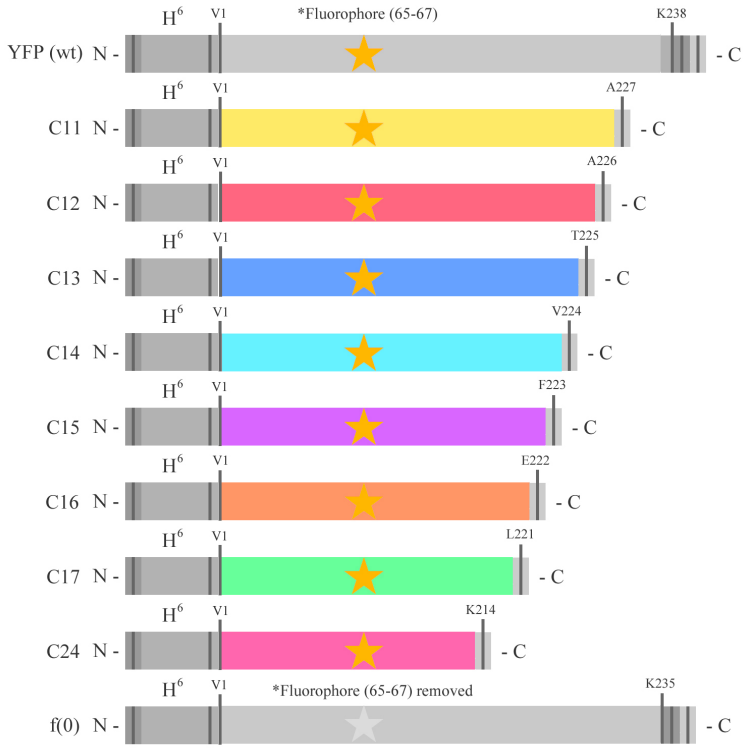


Figure 4.1 Protein sequences of the C-terminal truncations, C11 to f(0) used in this study topic. A stop codon: `taa`, was added to each C-terminal truncation at the designated positions, e.g., A227 on C11. The residues comprising the fluorophore: 65 - 67, were deleted in the f(0) C-terminal truncation shown by the grey star as changed to the yellow for the presence of the fluorophore. More details of the DNA constructs and amino acid residue are described in Section A.5 to A.14 in Appendix Chapter.

4.2.2 Recombinant expression and purification of C-terminal truncations in *E.coli*

For the production of the recombinant C-terminal truncations, each of the DNA constructs was expressed in BL21 (DE3) *E.coli* and purified using Ni-NTA (via the N-terminal His tag) and size exclusion chromatography (Sections 2.5.2 and 2.5.3 in Materials and Methods Chapter). Using SDS-PAGE analysis, it was found that there were differences in the expression behavior of YFP (wt) compared to the C-terminal truncations as demonstrated in Figure 4.2. The C-terminal truncations C11, C12, C13, C14, and f(0) C-terminal truncations were found to express solubly under the equivalent conditions used for YFP (wt) (37℃, for 12 hours), however C15, C16, C17, and C24 required a lower induction temperature of 16℃, for a period of 24 hours. This difference in induction temperature used for the different C-terminal truncations was due to tendency of the constructs C15, C16, C17, and C24 to naturally aggregate. Indeed, when the expression of these constructs was trialed at 37℃, several unexpected products, e.g. a dimer, trimer higher order aggregates, as well as truncation products were frequently observed when the purified material was subjected to FPLC, size exclusion chromatography (Figure 4.3). When the lower temperature conditions (16℃) were used instead, as this was found to result in the isolation of monomeric material, presumably due to the low temperature which slows down protein expression and therefore promotes favourable folding conditions (Figure 4.3).

Following purification, the C-terminal truncations were examined by SDS-PAGE (Figure 4.2) to assess their homogeneity. Typically, the migration of C-terminal truncations corresponded very well to their expected molecular weights (as predicted from their amino acid sequences) (Section A.4 in Appendix Chapter for molecular

weights). For example, the predicted molecular weights of C11 and YFP (wt) are 25.5 kDa and 26.7 kDa respectively and the difference in molecular weights between the two proteins, (approximately 1.2 kDa), can be observed in Figure 4.2 as a slight shift in the distance migrated by the C11 band relative to YFP (wt). Similarly, it can be also observed that with each progressive deletion at the C-terminus migrated further into the gel, and confirming that a smaller protein had been generated. Although C15, C16, C17, and C24 migrated on the gel at a similar distance, difference of molecular weight can be distinguished by band of f(0) C-terminal truncation which stands next to these C-terminal truncations.

In order to explore further physical properties of the isolated C-terminal truncations, both fluorescence and NMR spectroscopies were subsequently carried out. For the samples for NMR spectroscopy, the C-terminal truncations were prepared with a uniform isotopic ^{13}C and ^{15}N labelling strategy in Section 2.5.2 in Materials and Methods Chapter.

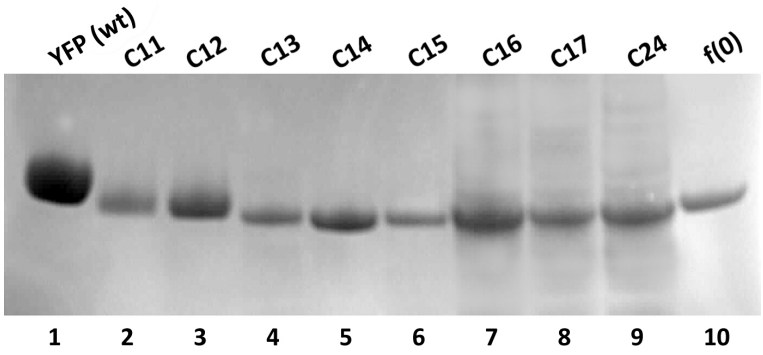
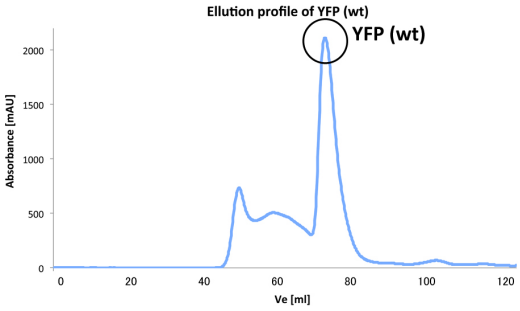
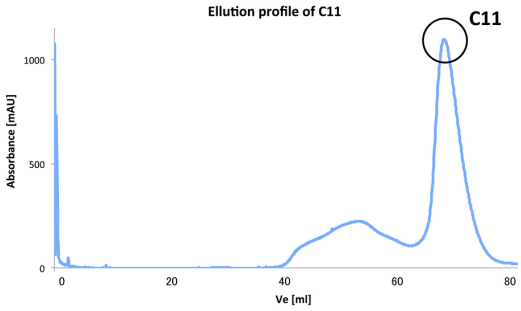
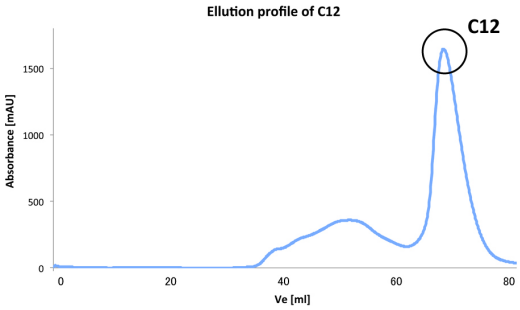
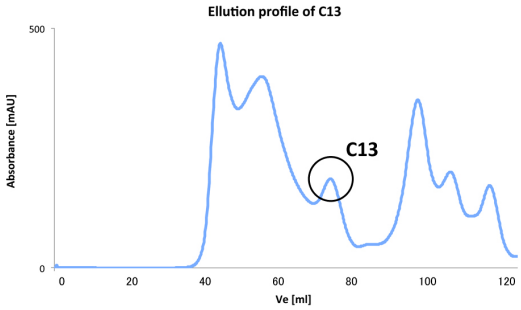


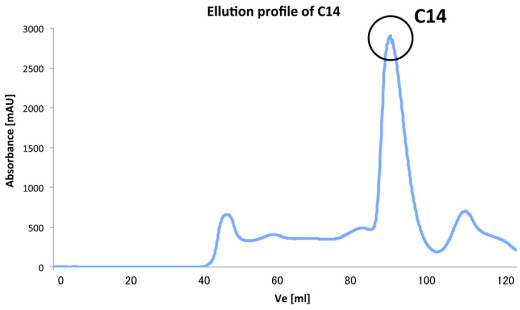
Figure 4.2 SDS-PAGE of pure and soluble C-terminal truncations from lanes 2 to 10: C11, C12, C13, C14, C15, C16, C17, C24, and f(0) with full length of YFP (wt) on lane 1. Dissolved and visualised by SDS-PAGE and CBB staining method. Each C-terminal truncation can be observed by their difference of molecular weight.

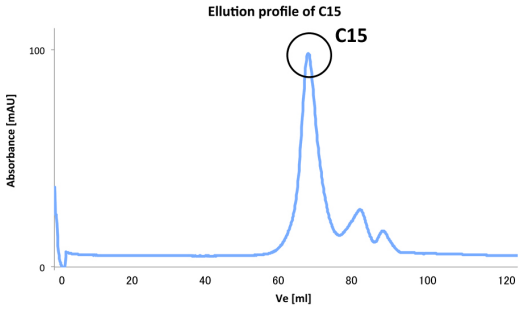

Ellution profile of YFP (wt)
YFP (wt)
Absorbance [mAU]
Ve [ml]


Ellution profile of C11
C11
Absorbance [mAU]
Ve [ml]


Ellution profile of C12
C12
Absorbance [mAU]
Ve [ml]


Ellution profile of C13
C13
Absorbance [mAU]
Ve [ml]


Ellution profile of C14
C14
Absorbance [mAU]
Ve [ml]


Ellution profile of C15
C15
Absorbance [mAU]
Ve [ml]

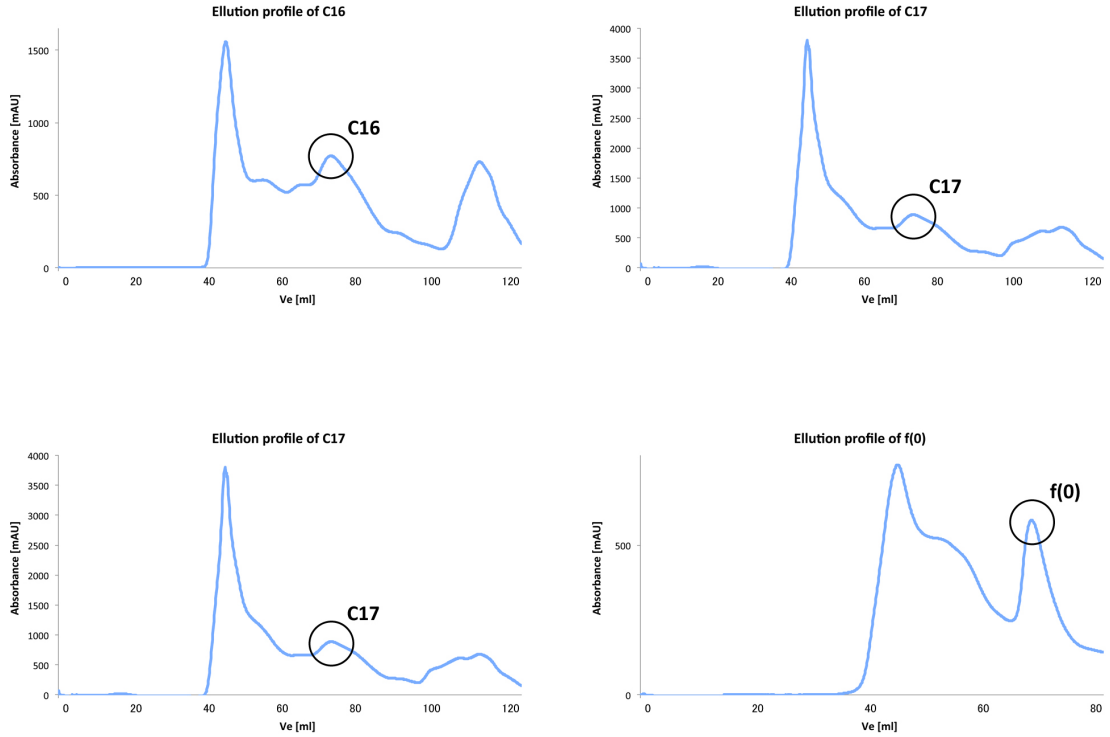


Figure 4.3 Elution profiles of YFP (wt) and C-terminal truncations:C11, C12, C13, C14, C15, C16, C17, C24, and f(0) by AKTA Prime plus FPLC with Superdex S200 16/60 column. Molecular weights of YFP (wt) and C-terminal truncations are between 26.7 kDa and 24.1 kDa. Most proteins were eluted in *Ve* of 70 to 80 ml. The V_o of the column was 45ml.

4.3 Intrinsic yellow fluorescence analysis of C-terminal truncations

4.3.1 Analysis of intrinsic yellow fluorescence for C-terminal truncations

As described in Section 1.4.2 in Introduction Chapter, the ability for YFP to fluoresce can be a criterion for the study of protein folding of NCs emerging from the ribosome. Indeed the ability for YFP to fluoresce in the visible light range was used as an initial means of assessing the 9 C-terminal truncations compared to the mature folded YFP (wt) as well as f(0), which is indistinguishable in its properties from YFP (wt) except that it is unable to fluoresce [81]. Figure 4.4 illustrates visible colour differences in each of the C-terminal truncations, in which C11, C12, C13, and C14 show a bright yellow colour that is comparable to that of the control YFP (wt). In C15, there is a decrease in the intensity of the yellow colour, which is subsequently absent in C16, C17, C24, and f(0), indicating that the fluorophore of YFP was not formed.

Following on from these initial observations, the intensity of intrinsic yellow fluorescence was recorded for each C-terminal truncation as well as for the control proteins YFP (wt) and f(0). Each sample was excited at 514 nm, and the maximum fluorescence intensity was collected at its peaks of emission spectra at 527 nm. As observed in Fig 4.5, with each truncation, there is a progressive decrease in the fluorescence emission at 527 nm, but no shifts were observed. The C15 truncation showed a significant decrease in fluorescence intensity, while in C16 to C24, there was no observable fluorescence, with an emission profile similar to f(0).

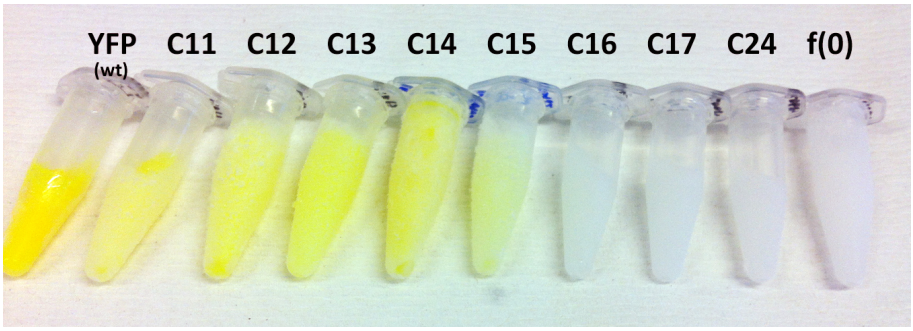
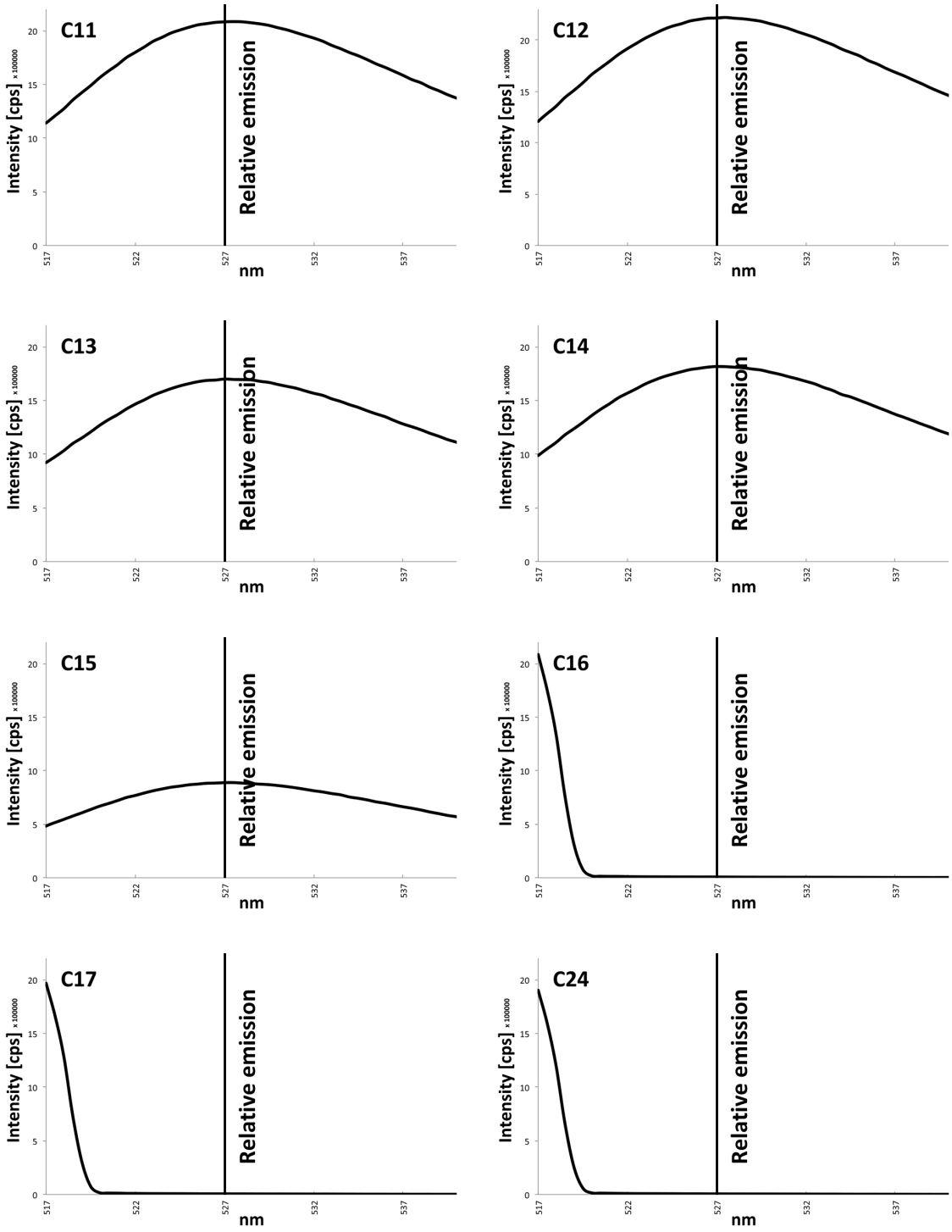


Figure 4.4 Observable colour differences of C-terminal truncations, C11 to f(0). Each of the YFP constructs, which are at a concentration of 5 μM, show observer differences in the characteristic yellow colour that is observed for YFP (wt) (far left). The yellow colour decreases in intensity as the C-terminal truncation increases, with C15 showing a weak yellow colour and in C16 there is no detectable colour; C16 shows similar colour characteristics as f(0) YFP, which is unable to show yellow fluorescence.


C11
C12
C13
C14
C15
C16
C17
C24
Intensity [cps] x 100000
Relative emission
nm
517
522
527
532
537
0
5
10
15
20

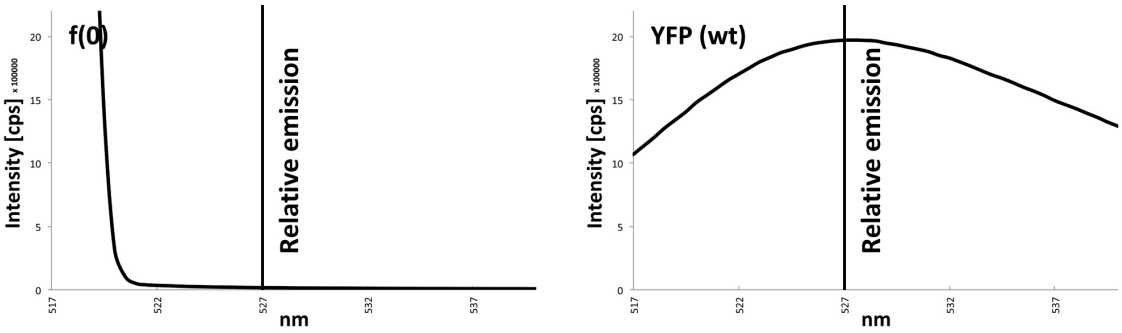


Figure 4.5 (a) Full spectra of C-terminal truncations evaluated by range of 515 to 540nm (x-axis) **and intensity** [cps] (y-axis). Each C-terminal truncation was excited at 514 nm (λ_{ex}) and expected emission peaks at 527 nm (λ_{em}) were marked as 'relative emission'.

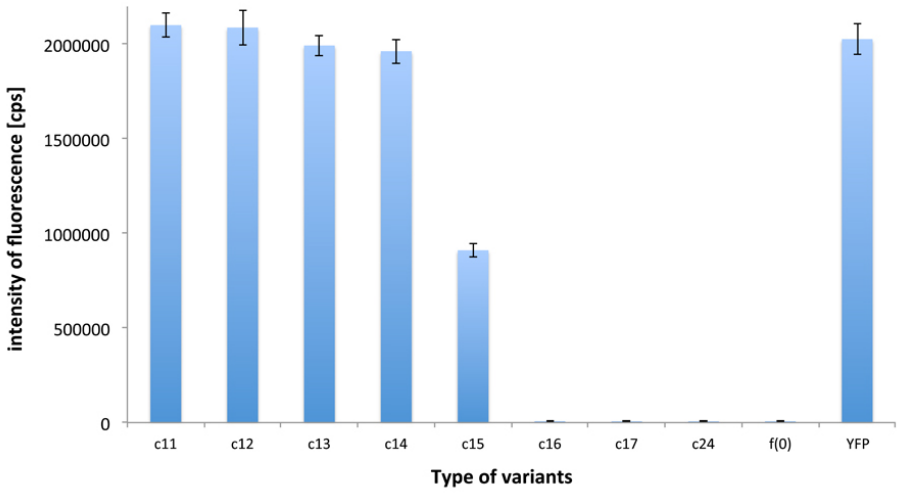


Figure 4.5 (b) Intensity of intrinsic yellow fluorescence of C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, f(0), and YFP (wt) constructs. Excitation was 514 nm (λ_{ex}) and emission was collected at 527 nm (λ_{em}). Each peak of intrinsic yellow fluorescence for the C-terminal truncations is plotted at 527 nm where all the samples showed peak intensities as emission. As observed previously by the visible colour difference, C11, C12, C13, and C14 showed intrinsic yellow fluorescence that was comparable to YFP (wt). In contrast, the C15 construct showed 44.9% +/-1.8% intrinsic yellow fluorescence compared to mature YFP (wt), and C16, C17, C24, and f(0) constructs did not show intrinsic yellow fluorescence as observed in **Figure 4.4** by visible difference under the day light condition.

The averaged fluorescence intensities for YFP (wt) and for each of the truncations was excited at 514 nm and recorded its emission peak at 527 nm and shown in Figure 4.5 (a). These data in Figure 4.5 (b) show that within error, the C-terminal truncations C11, C12, C13 and C14 displayed YFP fluorescence intensity that was similar to that of YFP (wt). In contrast, the fluorescence intensity of the C15 truncation was approximately 44.9% +/-1.8% of intensity that was observed in the emission spectrum of folded mature YFP (wt), an observation which is consistent with the visible colour differences observed in Figure 4.4. Similarly, in the shorter constructs, C16, C17, C24 as well as f(0), there was no observable intrinsic yellow fluorescence displayed at an excitation wavelength of 514 nm. These data demonstrate that the formation of the fluorophore, which is formed by residues 65 - 67 are strongly dependent upon the formation of YFP's beta-barrel; indeed the removal of the C-terminal β-strand (which lies between residues F223 to E222 results in the loss of intrinsic yellow fluorescence of YFP.

4.4 Intrinsic fluorescence analysis of C-terminal truncations

4.4.1 Intrinsic fluorescence using excitation of aromatic residues at 280 nm

To complement the intrinsic YFP fluorescence studies, the intrinsic fluorescence arising from the aromatic residues tryptophan, tyrosine and phenylalanine was also explored to examine the properties of YFP using alternative structural probes and to understand further the conformational properties of the C-terminal truncation C-terminal truncations. YFP (wt) contains 1 tryptophan (W107), 12 tyrosine (Y49, Y66 on the fluorophore, Y74, Y92, Y106, Y143, Y145, Y151, Y182, Y200, Y203, and Y237) and 13 phenylalanine (F8, F27, F46, F54, F61, F83, F84, F99, F100, F114, F130, F165, and F223) residues, which cover the entire molecule, as illustrated in Figure 4.6: although an excitation of 280nm will excite each of these residues, of these, the tryptophan residues have the highest quantum yield and therefore the fluorescence of the protein reflects the environment surrounding the tryptophan residues.

For each C-terminal truncation was excited at 280 nm, and the intrinsic fluorescence emission was collected between 300 - 400 nm as shown in Figure 4.7. In YFP (wt), the maximum peak of fluorescence: λ_{em} is approximately 309 nm, which indicates that the aromatic residues are in a buried environment which is typically associated with a folded protein. The λ_{em}, however, is slightly red shifted in each of the C-terminal truncations. There was a 2 nm shift (311 nm) in C16 and C17, the two C-terminal truncations whose intrinsic yellow fluorescence was similar to that of YFP (wt). Interestingly, a similar λ_{em} of 311 nm was observed for C15, which displayed a 44.9% +/-1.8% decrease in YFP fluorescence which suggests that the aromatic residues are in

similar environments to that observed for wild-type and that the protein is therefore in a folded conformation.

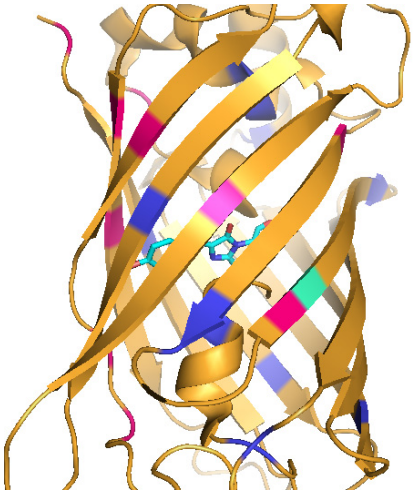


Figure 4.6 Image of YFP (wt) indicating aromatic amino acid residues in the structure. Single tryptophan (W107) is highlighted in green, 12 tyrosine (Y49, Y66 on the fluorophore, Y74, Y92, Y106, Y143, Y145, Y151, Y182, Y200, Y203, and Y237) are highlighted in magenta, and 13 phenylalanine (F8, F27, F46, F54, F61, F83, F84, F99, F100, F114, F130, F165, and F223) are highlighted in light-blue. The fluorophore, residue 65 to 67, is shown in ball and stick (blue, green, and cyan).

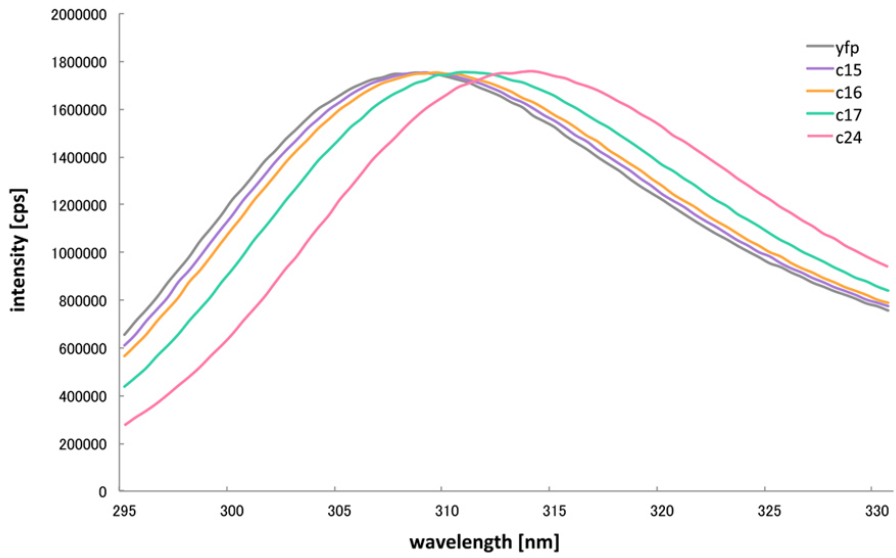


Figure 4.7 Intensity of fluorescence of C-terminal truncations, C15 to C24 excited at 280 nm. A gradual peak shift can be observed for mature YFP (wt, grey), C15 (purple), C16 (orange), C17 (light green), and C24 (pink). Excitation at 280 nm for all the constructs resulted in λ_{em} that gradually shifted from 309 nm as shown for YFP (wt) to 315 nm (C24). C15 showed 311 nm, C16 showed 311.5 nm, C17 showed 312 nm respectively. C11, C12, C13, and C14, which showed full intrinsic yellow fluorescence demonstrated in **Figure 4.5** which are indistinguishable from YFP, therefore these truncations were excluded from this figure.

C24 showed a highest emission peak at approximately 314 to 315 nm respectively, which is a difference of 5nm relative to YFP (wt). Typically for a folded protein, the λ_{em} at an excitation of 280 nm is c.a. 330 - 335 nm, and in its unfolded state this can be shifted as much 15 - 20 nm to c.a. 350 nm [143]. Although the difference of emission peaks shift appears to be small for the C15, C16, C17 and C24, these constructs also showed negligible YFP intrinsic fluorescence. It is likely therefore that these small shifts in λ_{em} correspond to large structural changes in the protein. The truncation at the C-terminus is expected to at least result in the unfolding of the protein, as the absence of particular amino acids results in the loss of critical contacts, as has been observed

previously for other proteins such as the immunoglobulin domain Dom5 (Waudby CA, Karyadi ME et al., unpublished observations). However, it is also possible that the absence of these amino acids may significantly compromise the structural integrity of YFP such that it aggregates, and this may also account for these small λ_{em} shifts.

4.5 Extrinsic fluorescence analysis of C-terminal truncations

4.5.1 Extrinsic fluorescence properties of the C-terminal truncations as probed using bis-ANS

To explore further the structural properties of the C-terminal truncations, they were labelled with a five fold molar excess of concentrated bis-ANS, a dye which binds specifically to exposed hydrophobic patches, and is therefore a useful probe of partially folded conformations.

Figure 4.8 shows the bis-ANS emission spectra of C-terminal truncations C15, C16, C17, and C24, all of which showed absent or significantly reduced intrinsic yellow fluorescence (Section 4.2). As shown in Figure 4.8, it can be observed that as the length of deletion from C-terminus increases, λ_{em} at 490 nm also increases, which indicates that there is a greater exposure of hydrophobic region of the protein, which may indicate that the protein may be partly folded; both completely folded or unfolded proteins do not bind bis-ANS as in both cases, regions of hydrophobicity are buried or completely exposed, respectively. In this instance YFP (wt) showed approximately 135,000 cps of intensity at 490 nm. in contrast, C15, C16, and C17 gradually increased their peak intensities at 490 nm: 150,000 cps, 210,000 cps, and 235,000 cps respectively. And C24 showed the highest peak intensity at 490 nm for 373,000 cps.

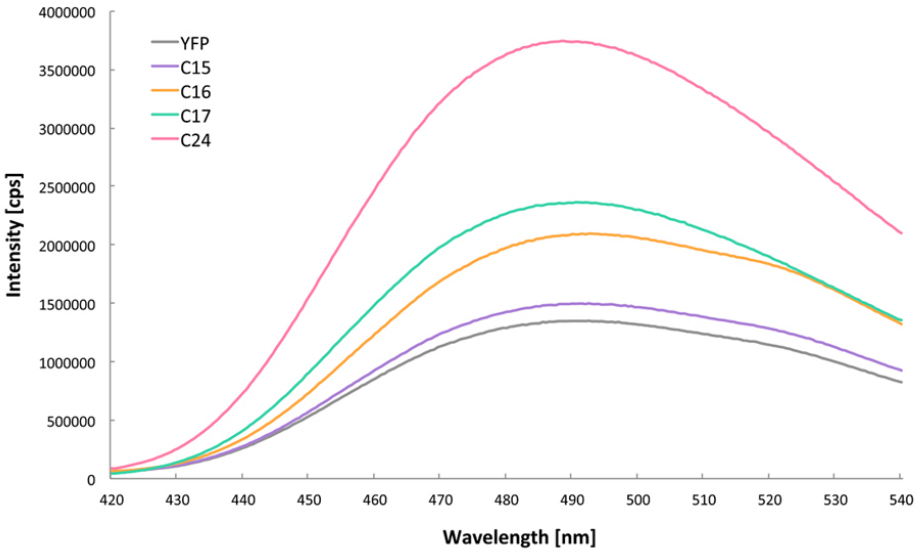


Figure 4.8 Extrinsic emission fluorescence spectra of C-terminal truncations with bis-ANS: C15 (purple), C16 (orange), C17 (light green), and C24 (pink) with mature YFP (wt, grey). A gradual increasing of approximate peak intensity at 490nm can be observed in each construct. For example, C15 showed a minor difference with YFP (wt), but C16 showed higher intensity than C15. C16 showed an increasing to C15, and C17 and C24 showed similar increasing. This gradual increase corresponds to the length of deletion from the C-terminus. C11, C12, C13, and C14, which showed full intrinsic yellow fluorescence demonstrated in Figure 4.5 showed zero difference to YFP (wt), so these truncations were excluded from this figure.

These bis-ANS data were compared with the intrinsic YFP fluorescence data and revealed some intriguing correlations: For example, as described in Figure 4.8, C15 C-terminal truncation showed a similar ability to bind bis-ANS as the mature YFP (wt), but displayed approximately a 50% decrease in the intrinsic yellow fluorescence compared to mature folded YFP (wt). In the case of the other C-terminal truncations, C16, C17, and C24 these each showed a gradual increase in their ability to bind bis-ANS, demonstrating the progressive exposure of hydrophobic regions, as the length of the deletion from C-terminus increased. In particular, C24 showed the highest

extrinsic fluorescence comparing to the other C-terminal truncations. This indicates that in this shortest C-terminal truncation most likely adopts a high population of a conformation with significant exposure of hydrophobic patches, or indeed it may also simply bind more ANS molecules than the other C-terminal truncations. It is likely, given the spectral properties observed in both the extrinsic and intrinsic fluorescence, that that as the C-terminus of YFP is truncated, that it results in the formation of an alternative conformation. Given that bis-ANS does not typically associate with unfolded proteins, the truncations are either partly folded or aggregated.

The bis-ANS binding of the C-terminal truncations was compared to the results obtained for the RNCs of YFP17, YFP34, and YFP55 (Section 3.5.2 in Result 1 Chapter). For instance, RNCs did not show similar bis-ANS binding compared to YFP, and they increased extrinsic fluorescence as they extended extra linker sequences as demonstrated in Figure 3.10 in Section 3.5.3 in Result 1 Chapter. In contrast, C-terminal truncations showed increase of intensity of extrinsic fluorescence as they were truncated the residues from the C-terminus. These mean that RNCs showed increase of hydrophobic surfaces as they were exposed to the surface of the ribosome, and that C-terminal truncations showed those of increasing as they lost the residues of the final β-strand from the C-terminus. Indeed, it is clear that YFP does not form a partially-folded conformation as observed C16, C17, and C24. It is possible that the YFP on the ribosome may be either adopting a folded or unfolded state at these different linker lengths (17, 34 and 55 amino acids), as the ability to bind bis-ANS is not observable in either of these conformations. While it is not possible to definitively distinguish between either of these outcomes, NMR studies of the 105 amino acid immunoglobulin-like protein ddFLN5 shows that a linker of at least 47 amino acids is needed for the protein to acquire its native fold (Cabrita LD, Launay H et al.,

unpublished observations). It is therefore likely that YFP is not folded at 17 and 34 amino acids, but may have acquired its native fold with a linker of 55 amino acids.

4.6 Structural analysis of the C-terminal truncations using NMR spectroscopy

4.6.1 Strategies for analysis

As described in Section 2.5.2 in Materials and Methods Chapter, each C-terminal truncation, C11, C12, C13, C14, C15, C16, C17, C24, and f(0), was uniformly labelled with both ^{15}N and ^{13}C isotopes and a series of 1D and 2D NMR experiments were recorded for each including ^{1}H - ^{15}N SOFAST-HSQC, ^{1}H - ^{13}C HMQC and translational diffusion experiments at 25°C. The data acquisition parameters for the C-terminal truncations were equivalent to those used for the RNCs as described in Section 3.6.1 in Result 1 Chapter.

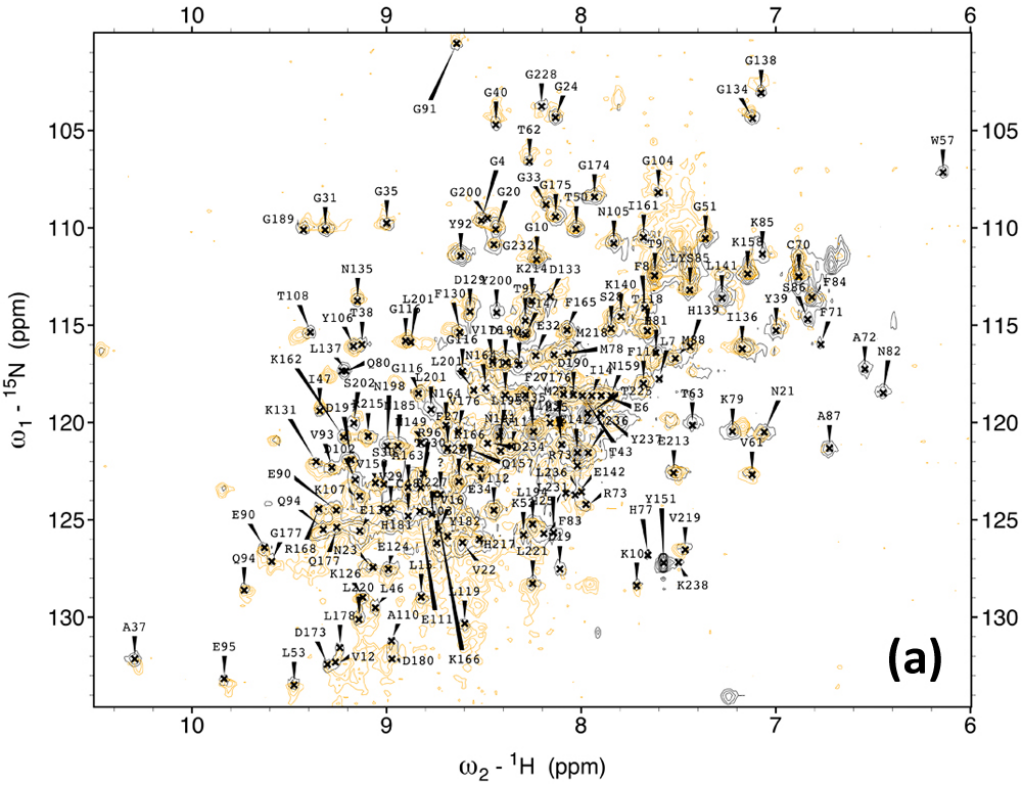
4.6.2 2D analysis of the C-terminal truncations using ^{15}N SOFAST-HSQC experiments

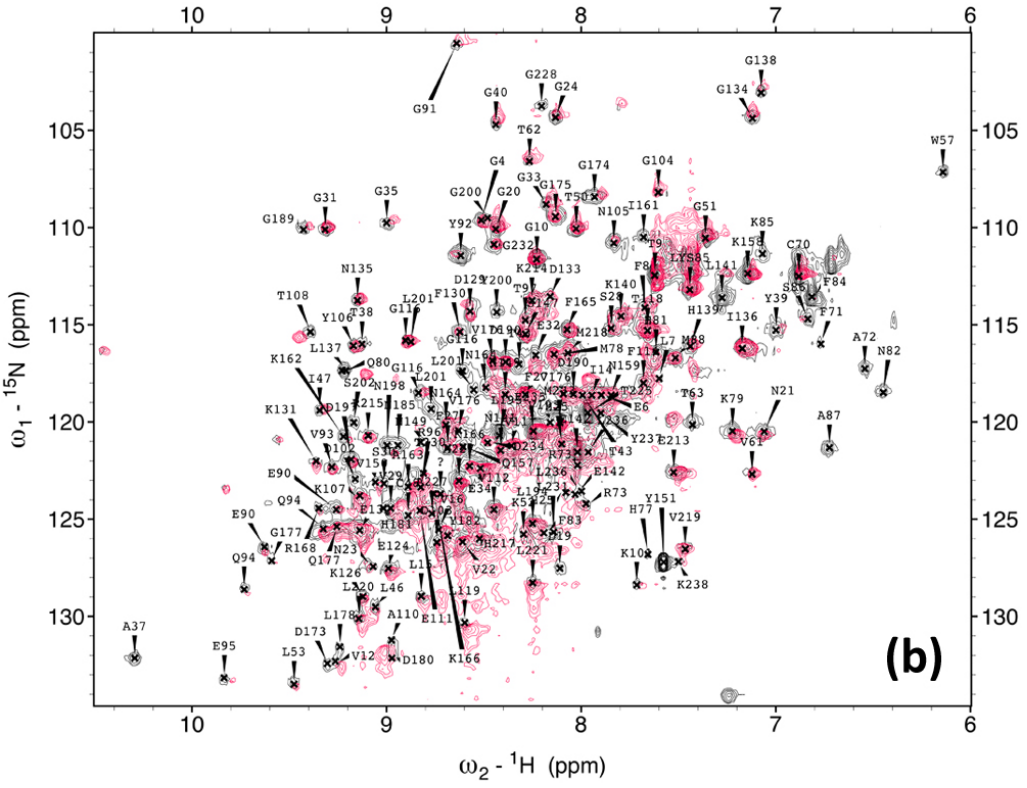
The structural properties of the C-terminal truncations, C11, C12, C13, C14, C15, C16, C17, C24, and f(0), were evaluated using 2D SOFAST correlation spectra and were assigned in both ^{1}H - ^{15}N and ^{1}H - ^{13}C in 2D using assignments that were transferred from YFP (wt), Venus [81], a phenotype of YFP, was already assigned [120, BMRB Entry: 15826] as demonstrated in Figure 3.11 in Section 3.6.2 in Result 1 Chapter.

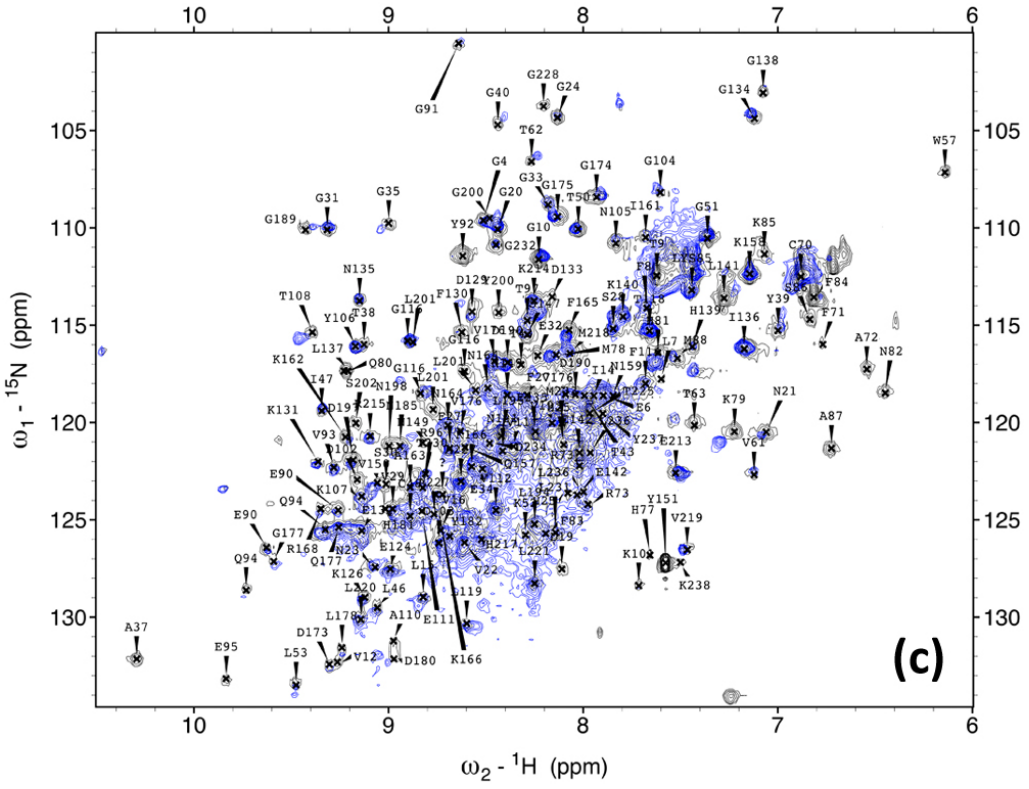
Figure 4.9 shows 2D ^{1}H - ^{15}N correlation spectra of C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, f(0) overlaid with YFP (wt). Comparing to YFP (wt) spectra, only a minor difference was found on 2D spectra of C11, C12, C13, and C14. In contrast to the truncations C11 to C14, in 2D correlation spectra of C15 (Figures 4.9 (e)) shows observable differences compared to YFP (wt), in which there were a number

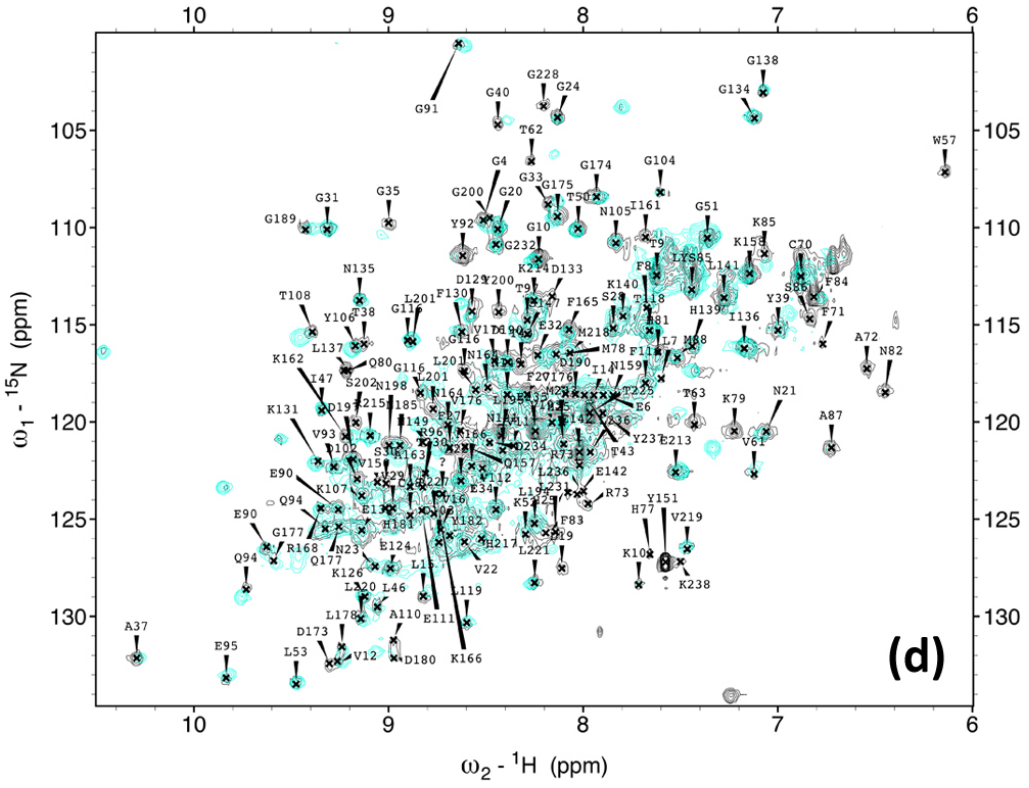
of peaks that were not observable in the spectrum in Figures 4.9 (e). When compared to YFP(wt), 13 peaks from C15 were absent, which were in addition to the deleted 15 amino acids from C-terminus: (VTAAGITHGMDELYK). There were observable similarities between cross peaks as demonstrated in Figure 4.9, and will be studied in more detail by analysis by chemical shift in Section 4.5.4.

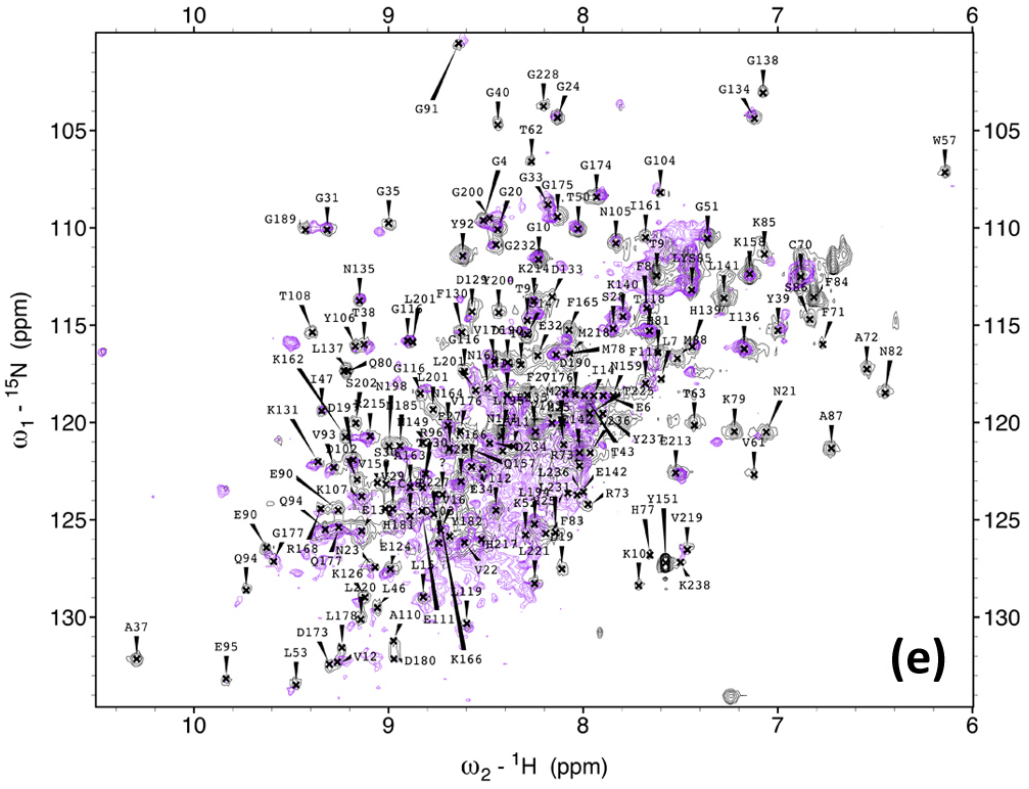
In addition, very significant differences were also observed between YFP (wt) and the C-terminal truncations: C16, C17, C24 (Figure 4.9 (f,g,h). Indeed, the very large overlap of resonances observed within the spectra, with a chemical shift range of 10 to 6ppm (^{1}H dimension), 100 to 135 ppm (^{15}N dimension). These spectra show characteristics of an aggregated protein, and for this reason, the ^{1}H - ^{15}N spectral characteristics of these C-terminal truncations were not assigned for chemical shift and peak intensity analysis.

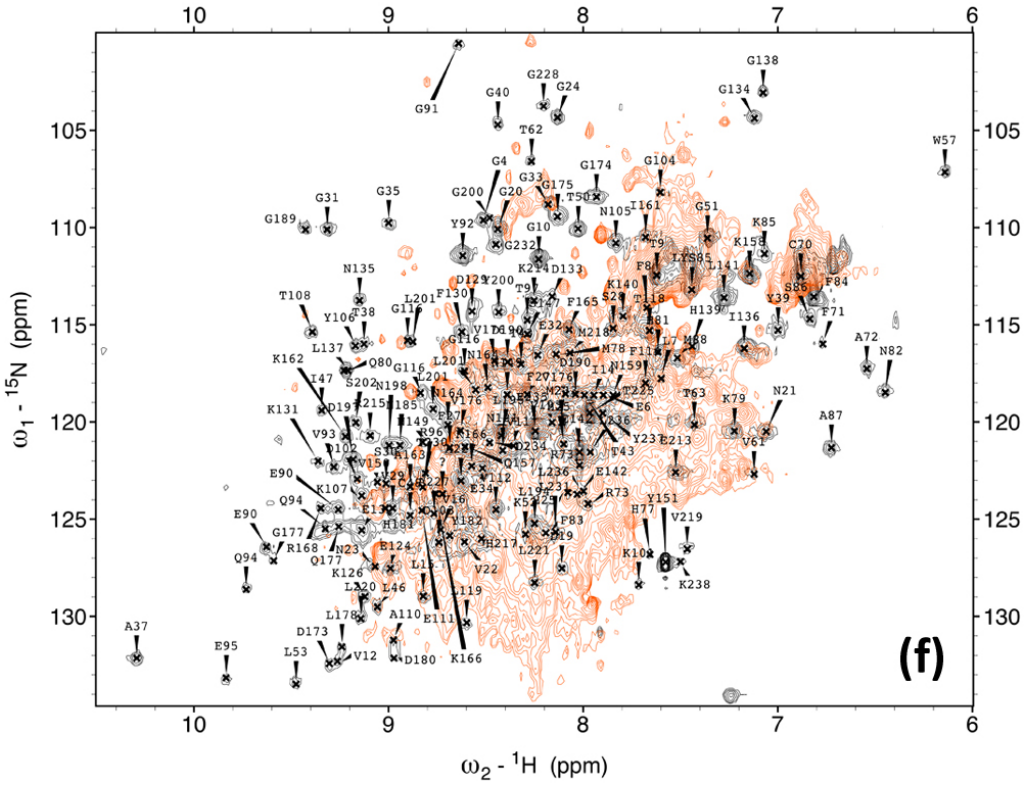

(a)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)

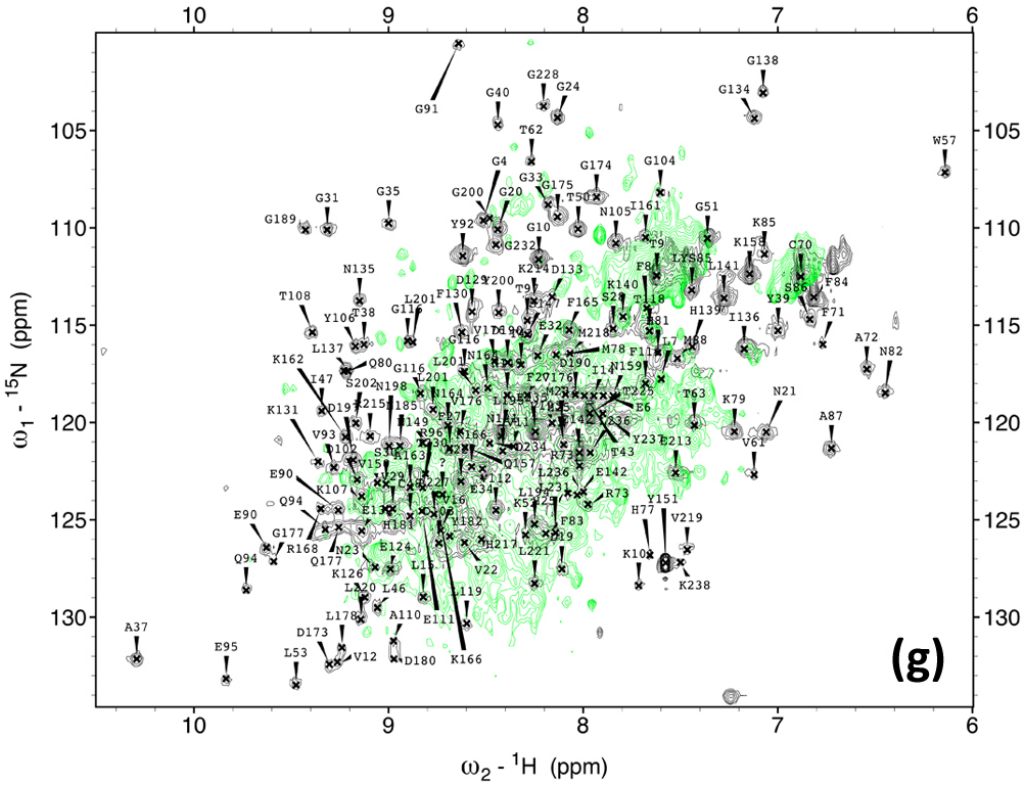

(b)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)

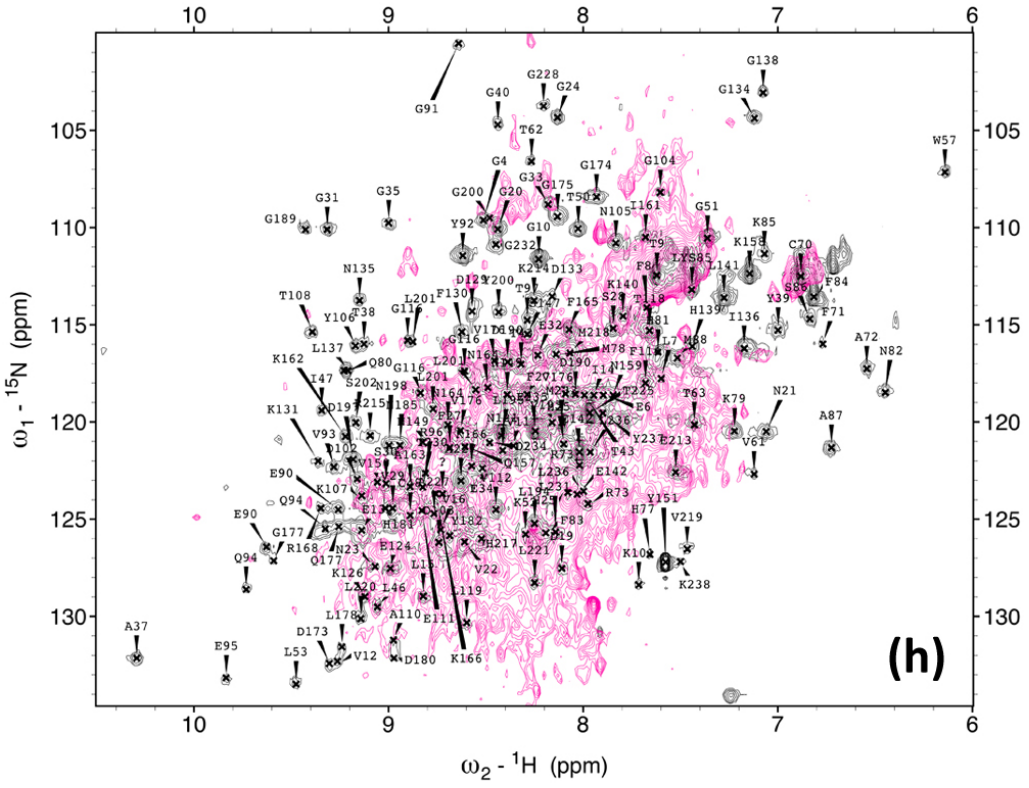

(c)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)


(d)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)


(e)
ω1 - 15N (ppm)
ω2 - 1H (ppm)


(f)
ω1 - 15N (ppm)
ω2 - 1H (ppm)


(g)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)


(h)
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)

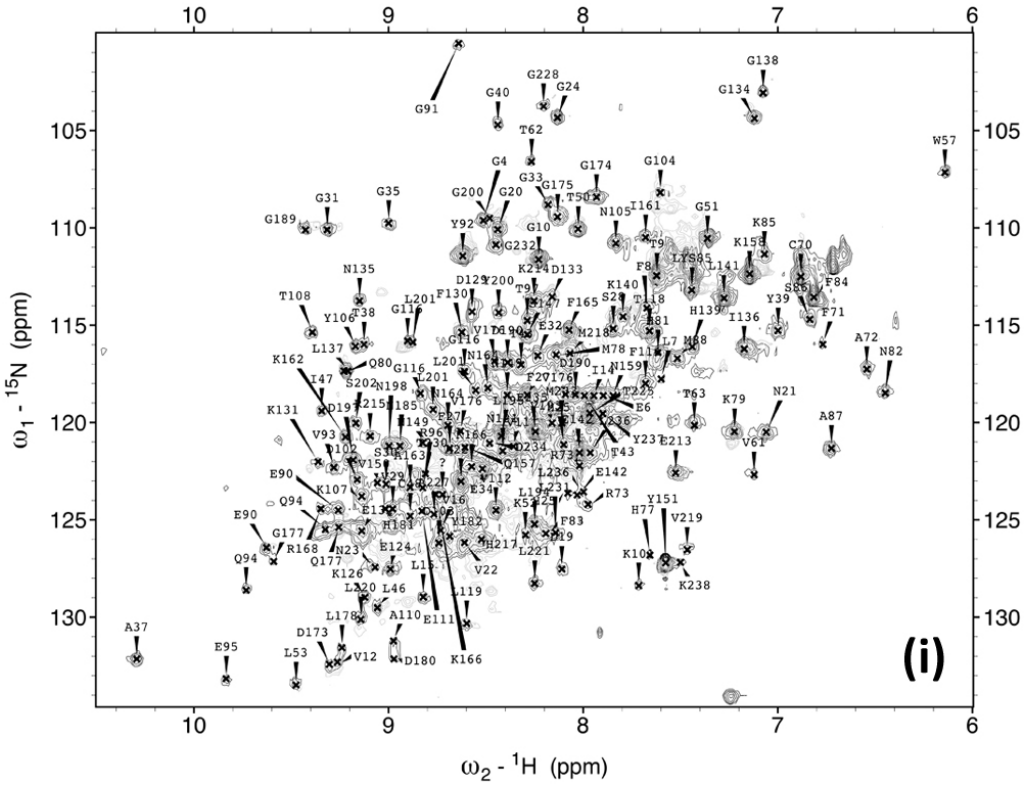


Figure 4.9 ^{1}H - ^{15}N 2D correlation spectra of YFP C-terminal truncations and assigned by attribution using YFP (Venus). (**a**) ^{1}H -^{15}N 2D correlation spectra of C11 (yellow), (**b**) C12 (red), (**c**) C13 (blue), (**d**) C14 (turquoise), (**e**) C15 (purple), (**f**) C16 (orange), (**g**) C17 (light green), (**h**) C24 (pink), and (**i**) f(0) (dark grey) with assignment.

4.6.3 Chemical shift analysis between folded mature YFP and C-terminal truncations using Venus assignment

The resonances arising from each of the C-terminal truncations in the correlation spectra was by attribution using Venus protein [81], BMRB code: 15826 [120]. In YFP 125 peaks in ^{1}H - ^{13}C correlation spectra are observed and 93 of these are well-resolved and readily identified within the C-terminal truncations. Similarly, 183 cross peaks in ^{1}H - ^{15}N chemical shifts are readily observable and 153 of these are observable within C-terminal truncations.

Figure 4.10 demonstrates the range of chemical shift differences between YFP and C11, C12, C13, C14, C15, and f(0). The truncations C16, C17, and C24 truncations were omitted from further analysis (See Section 4.6.2). The difference in chemical shift, ΔδNH, was calculated by the Equation 2.1 [144].

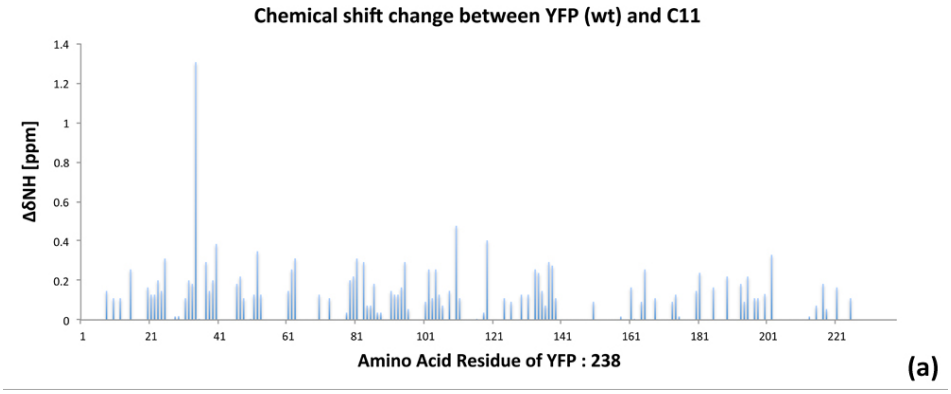
$$\Delta\delta NH = \sqrt{(\Delta N)^2 + (\delta H)^2} \quad \text{(equation 2.1)}$$

Equation 2.1 ΔδNH is given by difference of chemical shift in ^{1}H and ^{15}N from original protein sequence in 2D spectrum of NMR.

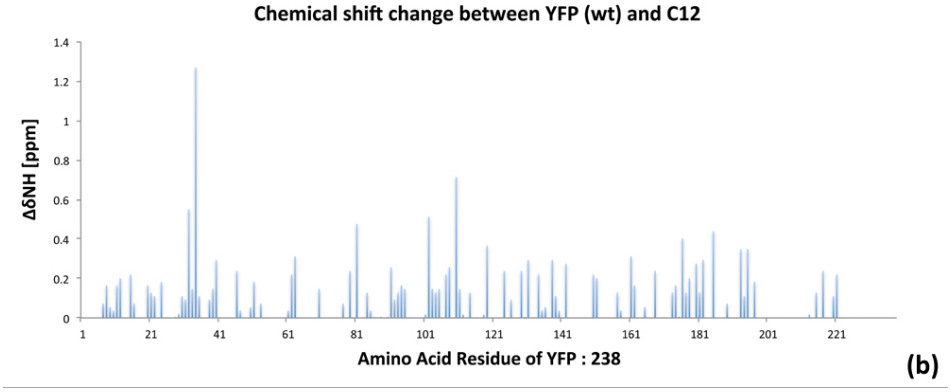
Following Figure 4.10, the majority of chemical shift differences (ΔδNH) in each C-terminal truncation range between 0.2-0.4 in C11, C12, and C13 which indicate very minor local differences comparing to YFP (wt). In C14, larger changes are apparent with chemical shifts and which are more obvious in C15, in which 6 residues in the C15 mutant showed values greater than 1.0 and is not native-like in structure. Moreover, the data for C14 and C15 show similarities in the chemical shift differences in which the same 6 residues: V29, E34, C48, I161, K166, V176, exceeded the value of 1.0 (Figure 4.10 (e)). This suggests that the by truncating the YFP by 14 and 15 amino acids at the

C-terminus that key interactions are disrupted within these residues, such that further truncations result in unfolding and/or aggregation of the protein.

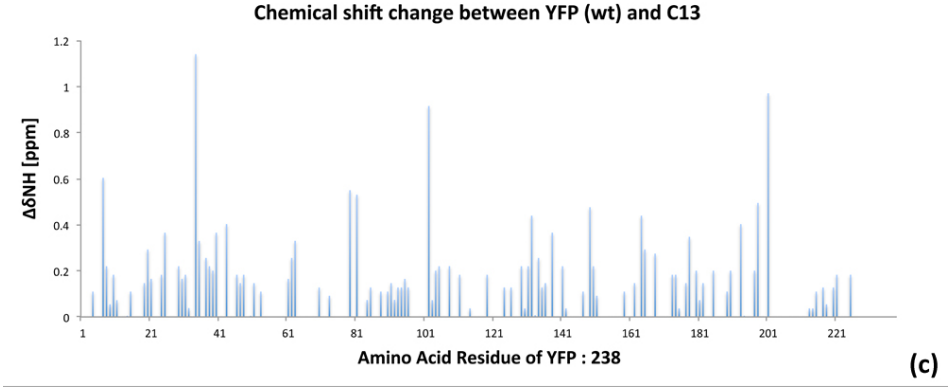
By comparison, analysis of the f(0) C-terminal truncation showed large scale of differences in the values observed for some amino acids, which is very likely the result of the deletion of residues 65 to 67. It is clear that deletion of residues associated with the central core of the protein, that a large scale conformational change is observed. This observation is supported by 2D spectrum in ^{1}H - ^{15}N data shown in Figure 4.9 (i), in which some cross peaks have gradually shifted in both ^{1}H - ^{15}N and ^{1}H - ^{13}C. As the assignment of f(0) is not known, the identities of these resonances cannot be ascertained.


Chemical shift change between YFP (wt) and C11
ΔδNH [ppm]
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
21
41
61
81
101
121
141
161
181
201
221
Amino Acid Residue of YFP : 238

(a)


Chemical shift change between YFP (wt) and C12
ΔδNH [ppm]
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
21
41
61
81
101
121
141
161
181
201
221
Amino Acid Residue of YFP : 238

(b)


Chemical shift change between YFP (wt) and C13
ΔδNH [ppm]
1.2
1
0.8
0.6
0.4
0.2
0
1
21
41
61
81
101
121
141
161
181
201
221
Amino Acid Residue of YFP : 238

(c)

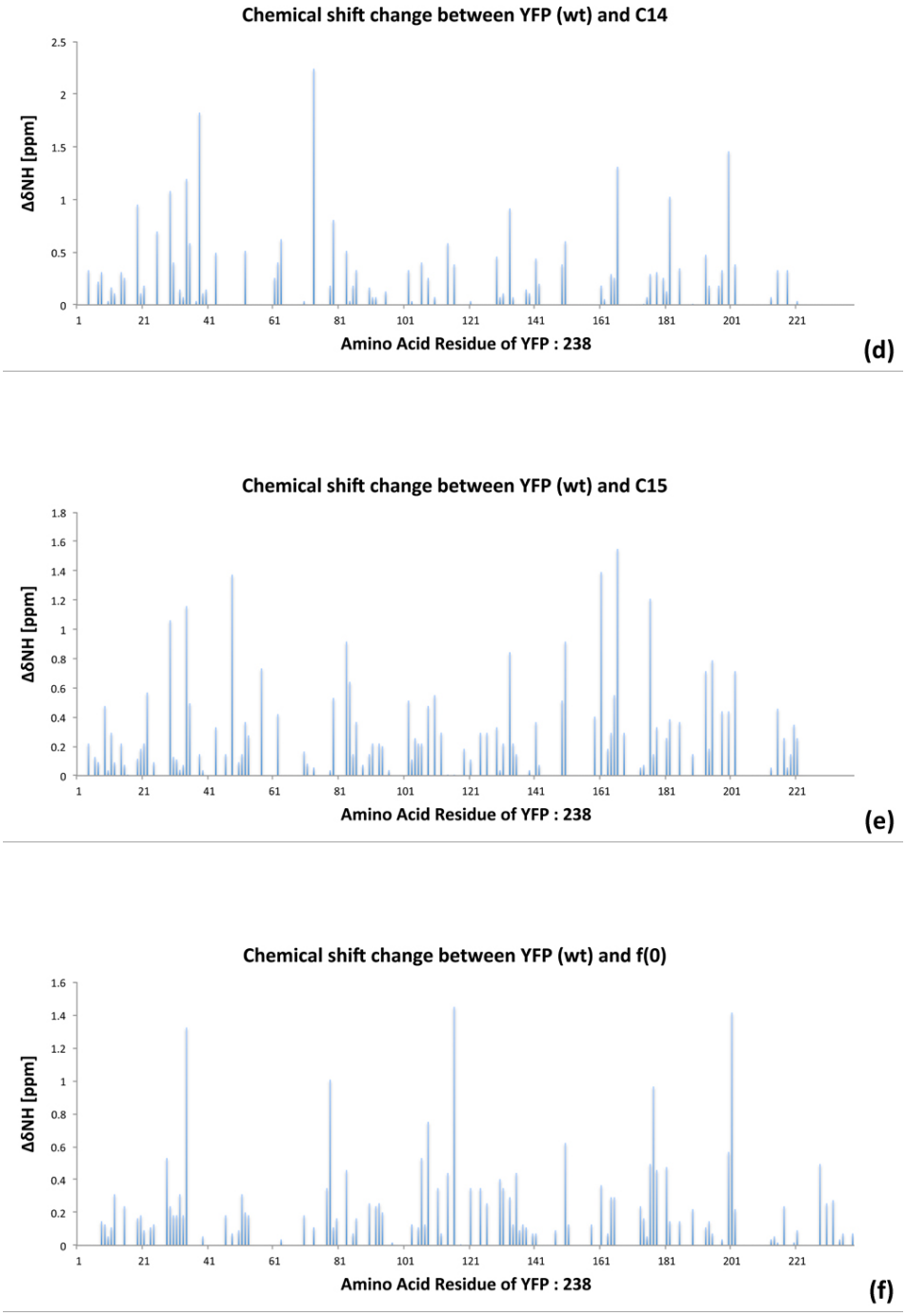


Figure 4.10 Chemical shift differences (ΔδNH) C-terminal truncations: C11 (**a**), C12 (**b**), C13 (**c**), C14 (**d**), C15 (**e**), and f(0) (**f**), comparing to mature YFP (wt) (based on Venus assignment [120]). X-axis shows the amino acid sequence of YFP (238), and Y-axis shows the ΔδNH, which demonstrate the cross peak movement between YFP (wt) and C11, C12, C13, C14, C15, and f(0).

4.6.4 Intensity analysis of ^{1}H - ^{15}N correlation spectra

The signal intensities of each cross peak for C-terminal truncations: C11, C12, C13, C14, C15, and f(0), were analysed as these are reflective of the relaxation properties of a protein and can therefore provide complementary information to that observed by chemical shift analysis (Section 2.6.2 in Materials and Methods Chapter). Figures 4.11 demonstrates the signal intensity of well-resolved cross peaks in C11, C12, C13, C14, C15, and f(0) respectively: 137 cross peaks for C11, 133 cross peaks for C12, 132 cross peaks for C13, 131 cross peaks for C14, 128 cross peaks for C15, and 127 cross peaks for f(0) C-terminal truncations. Overall, there was a lower average peak intensity for the truncations compared to YFP (wt), in which the wild-type shows an average intensity that is 7 times greater than that of observed each of the truncations. The truncations themselves showed very similar intensity values (Fig 4.11), suggesting that perhaps they were experiencing similar dynamical properties. It is clear however, that these differences compared to YFP (wt) suggests that the truncations have very different dynamical properties compared to full-length YFP.

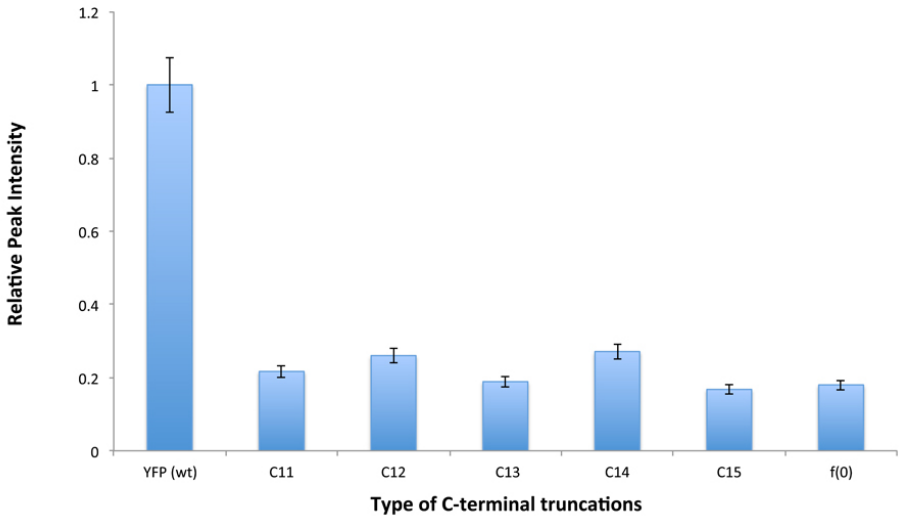


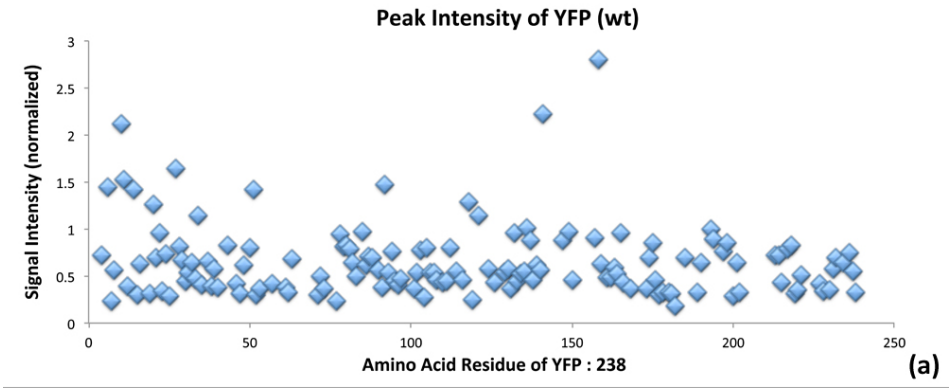
Figure 4.11 The average peak intensity of YFP (wt) and C-terminal truncations: C11, C12, C13, C14, C15, and f(0) in ^{1}H - ^{15}N 2D spectra. Signal intensity from each assigned peak by Sparky was exported and graphed. X-axis stands for species of C-terminal truncations, and Y-axis stands for normalised signal intensities (divided by value of YFP (wt)).

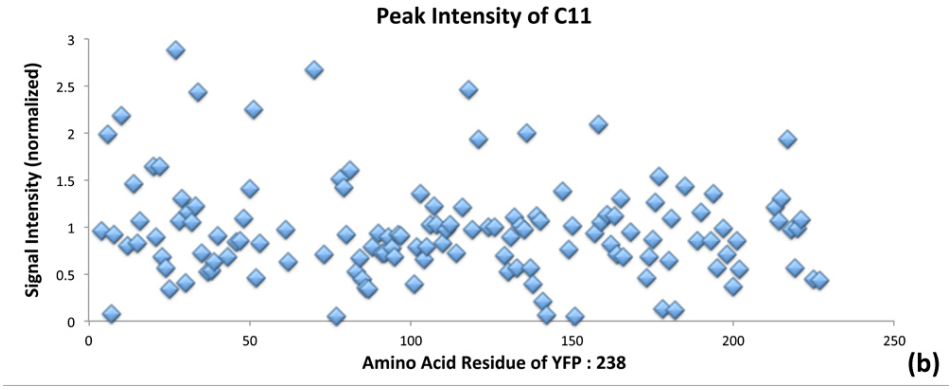
At a residue-specific level for YFP (wt), there was a fairly uniform set of intensities across the amino acid sequence, however there were differences in several residues T9, C70, F84, E142, Y151, K158, L195, E225, M233, and E235 which appeared to give rise to above-average intensities (Figure 4.12). When mapped to the YFP structure, residues T9, E142, L195, E225, and E235 are positioned to the flexible regions of YFP (wt), not involved in the α-helices and β-strands.

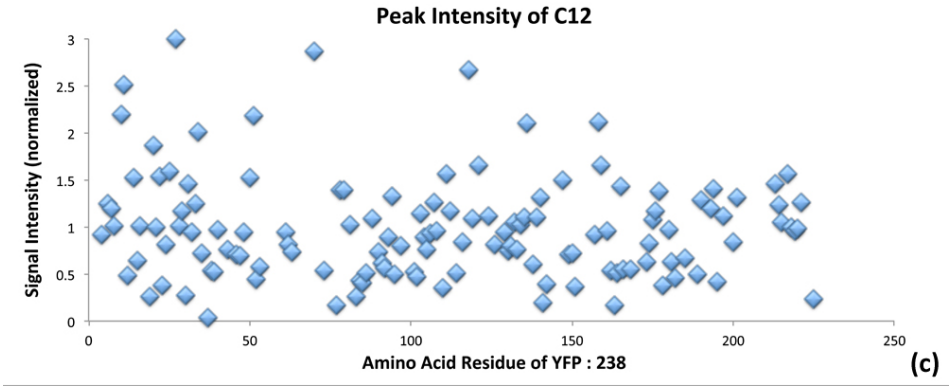
Peak intensity of T9, E142, L195, E225, and E235 positioned to the flexible domains of YFP (wt) which did not form the α-helices and β-strands. C70, F84, Y151, K158, and E225 positioned in the α-helices and β-strands which were firmly folded. So these higher peaks from flexible domains can be considered as the flexibility of YFP in linker domains among the α-helices and β-strands.

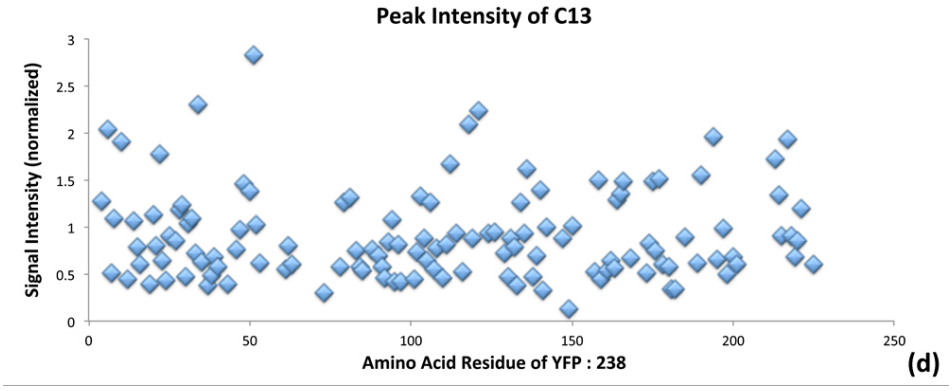
Compared to range of signal of YFP (wt) in Figure 4.12 (a), the signal intensity of each C-terminal truncations: C11, C12, C13, C14, C15, and f(0) showed a considerable range of intensity values across the sequence, showing that potentially the molecule at these different truncations may be experiencing different dynamical characteristics compared to YFP (wt). Interestingly, despite the spread of intensity values there were also some residues for which the intensities were far above the average in the truncations C11 to C15 compared to YFP (wt), namely residues T9, C70, K158 and C12. In addition, residues M233 and E235 in the f(0) also show changes that are greater than that for the average. The region associated with these residues maintain a close link to maintaining the fluorophore, and in this f(0) mutant, the fluorophore-associated residues are deleted. It is possible therefore, that the enhanced intensities observed within these residues in the f(0) mutant reflect changes surrounding the M233 and E235 residues to compensate for the absence of the fluorophore, and maintain structural integrity.

On the contrary, T9 and C70, also K158 on C12, showed more higher intensity than averaged range between 0.368 (500,000) and 7.35 (10,000,000) in each C-terminal truncations: C11, C12, C13, C14, C15, and f(0). Further, M233 and E235 on f(0) variant showed more higher signal intensity than the other residues though the region of the fluorophore: 65 to 67 residues, were deleted, so that this could be a scenario that YFP can sustain its own well-formed structure without the existence of the region of fluorophore. In total, the C-terminal truncations did not show the strong elasticity in the value of their signal intensity and behaved the native like proteins comparing to the mature folded YFP (wt) (The values of C-terminal truncations were divided by 1,360,886: I_{C11} - $I_{f(0)}$ / *6*).


Peak Intensity of YFP (wt)
Signal Intensity (normalized)
Amino Acid Residue of YFP : 238
(a)


Peak Intensity of C11
Signal Intensity (normalized)
Amino Acid Residue of YFP : 238
(b)


Peak Intensity of C12
Signal Intensity (normalized)
Amino Acid Residue of YFP : 238
(c)


Peak Intensity of C13
Signal Intensity (normalized)
Amino Acid Residue of YFP : 238
(d)

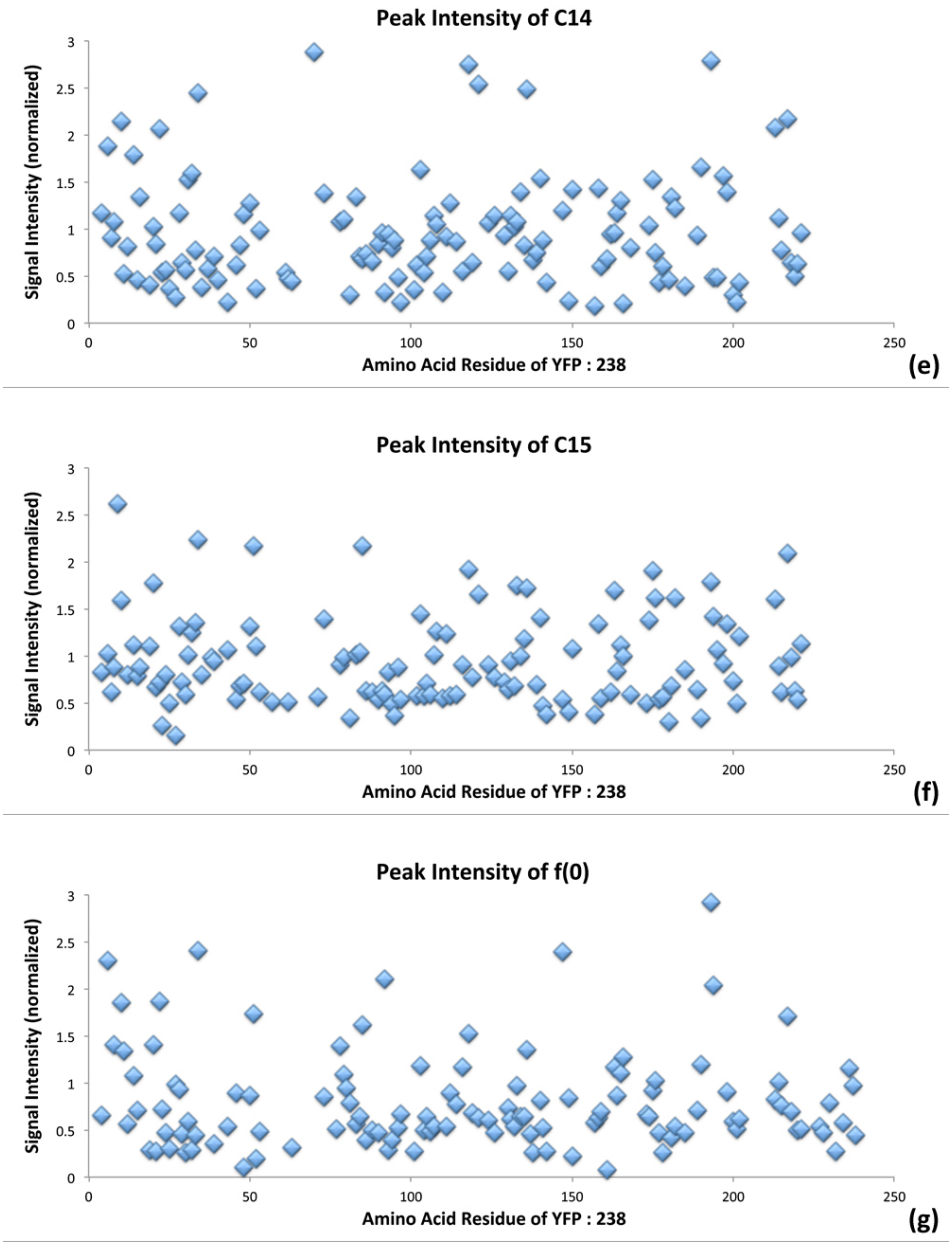


Figure 4.12 The peak intensity of YFP (wt) and C-terminal truncations: (**a**) YFP (wt), (**b**) C11, (**c**) C12, (**d**) C13, (**e**) C14, (**f**) C15, and (**g**) f(0) peak intensities in ^{1}H - ^{15}N 2D spectra. Signal intensity from each assigned peak by Sparky was exported and graphed. X-axis shows the amino acid residues, and the Y-axis shows for the raw intensity for each residue.

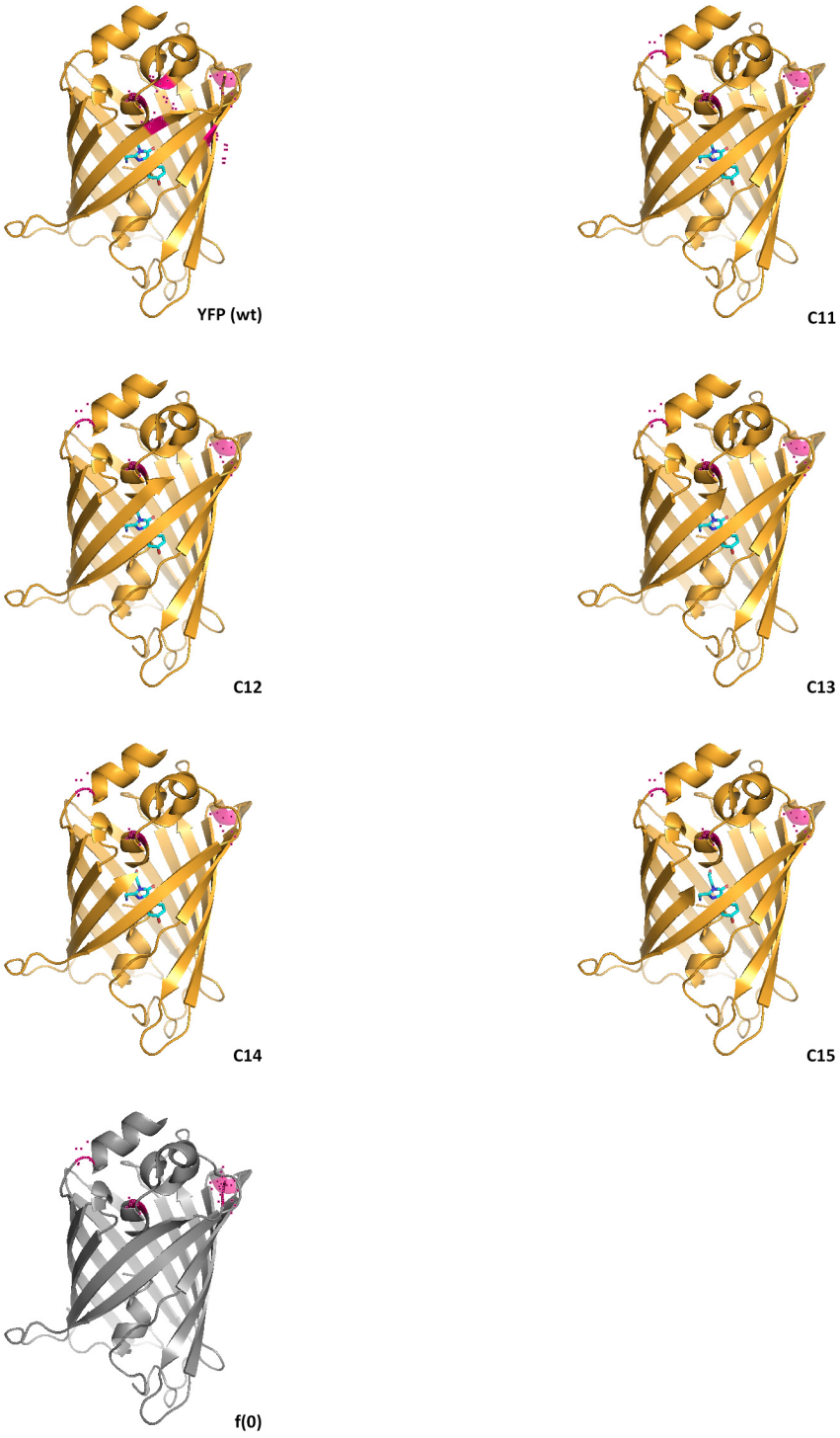


Figure 4.12 (**h**) **Images of YFP (wt) and C-terminal truncations**: C11, C12, C13, C14, C15, and f(0). Residues that showed more higher intensity than average are highlighted in pink. For YFP (wt): C70, F84, Y151, K158, and E225, for C11 to C15: T9, C70, and K158, for f(0): T9, C70, K158, M233, and E235 are highlighted respectively.

4.6.5 Observation of changes in both cross peak intensity and chemical shift

To further illustrate the changes in cross peak intensity observed within the C-terminal truncations, the cross peak of Tyr106 was randomly chosen. Figure 4.13 (a) illustrates 2D spectra of superposition at Tyr106 of ^{1}H - ^{15}N from C11 to f(0) truncations. The intensity values from Tyr106 are plotted in Figure 4.13 (b), and which shows they change across each of the variants. It was observed that the cross peaks Tyr106 of C11, C12, C13 and C14 displayed very similar intensity values in contrast, however, the intensity of the Tyr106 of C15 variant decreased by ca. 43%. In the shorter constructs, C16, C17, C24 showed no cross peaks on these ^{1}H - ^{15}N 2D spectra as shown in Figure 4.9 (f - h). Interestingly, the gradual decrease in cross peak intensity for Tyr106 is mirrored in the decrease in intrinsic yellow fluorescence intensity shown in Figure 4.5. Also, it can be observed that Thr38 next to Tyr106 disappeared from C-terminal truncations although it clearly appeared on YFP (wt) spectrum.

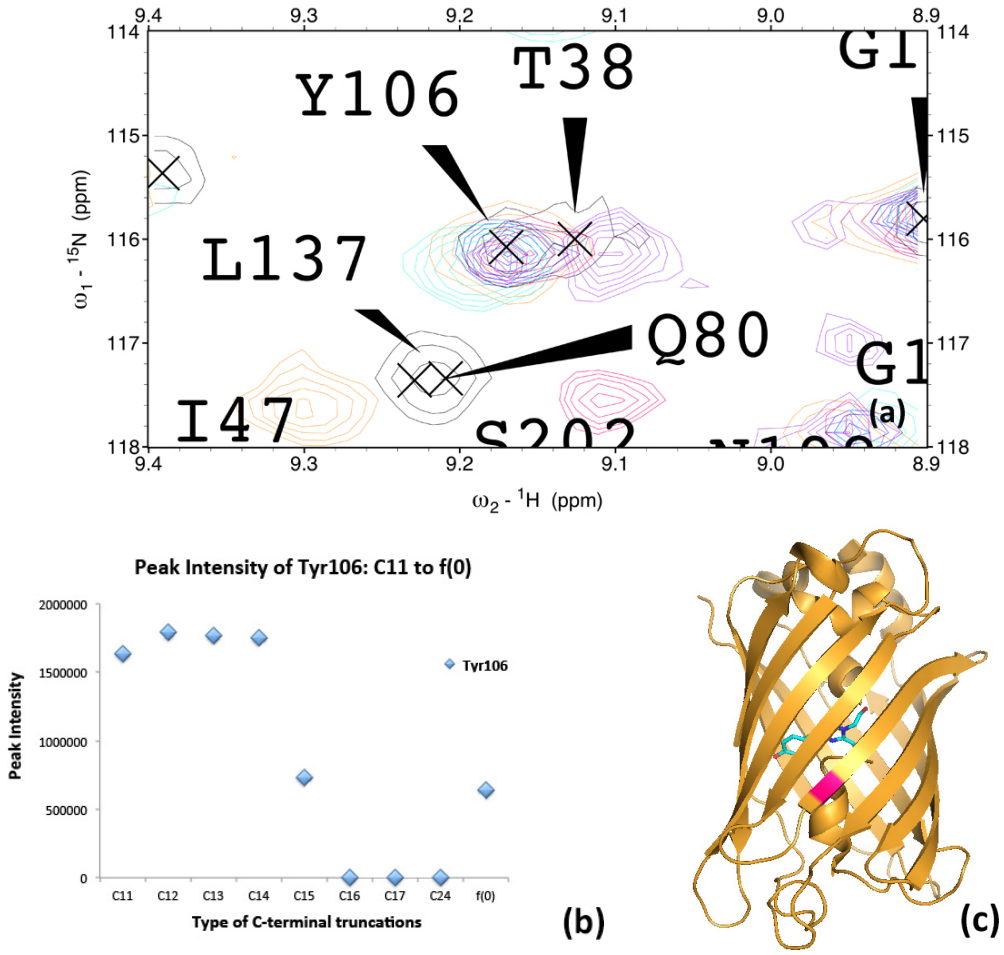
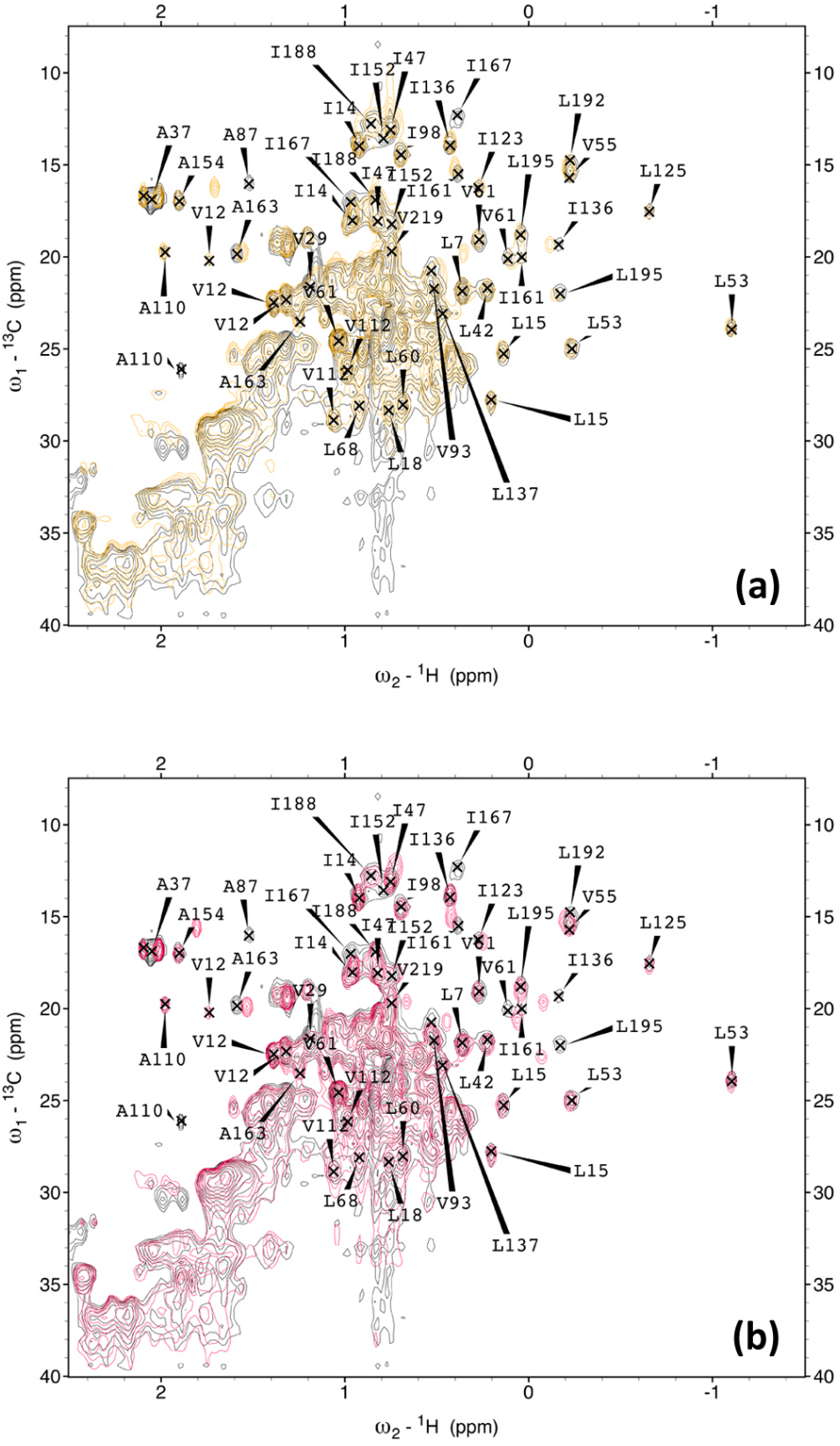
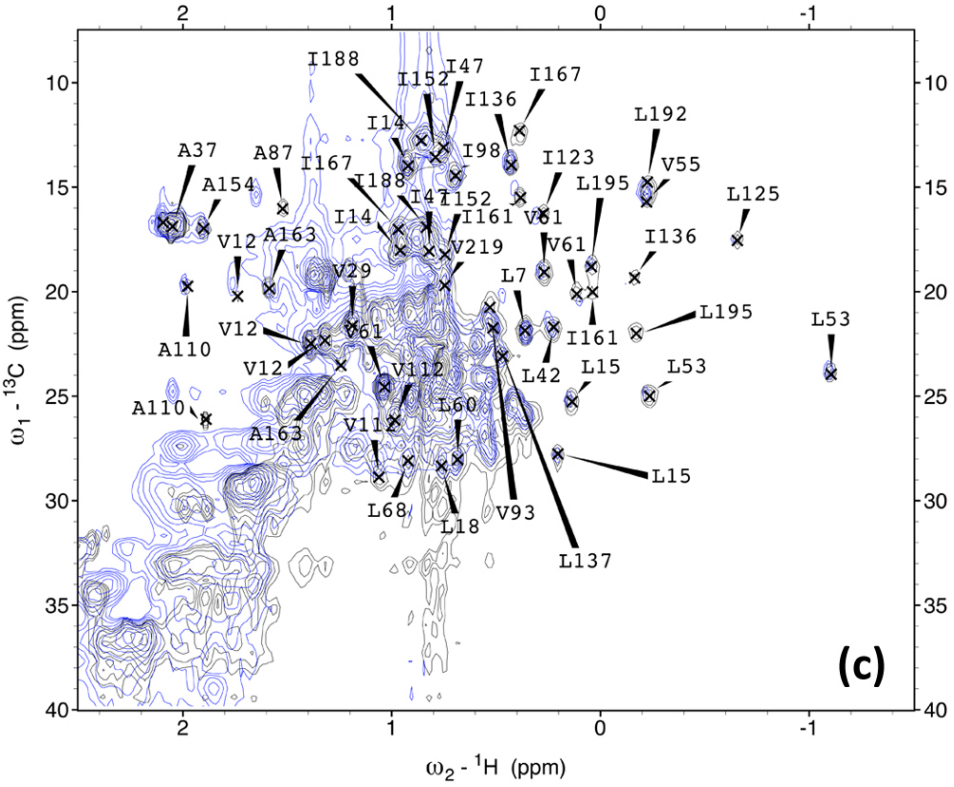


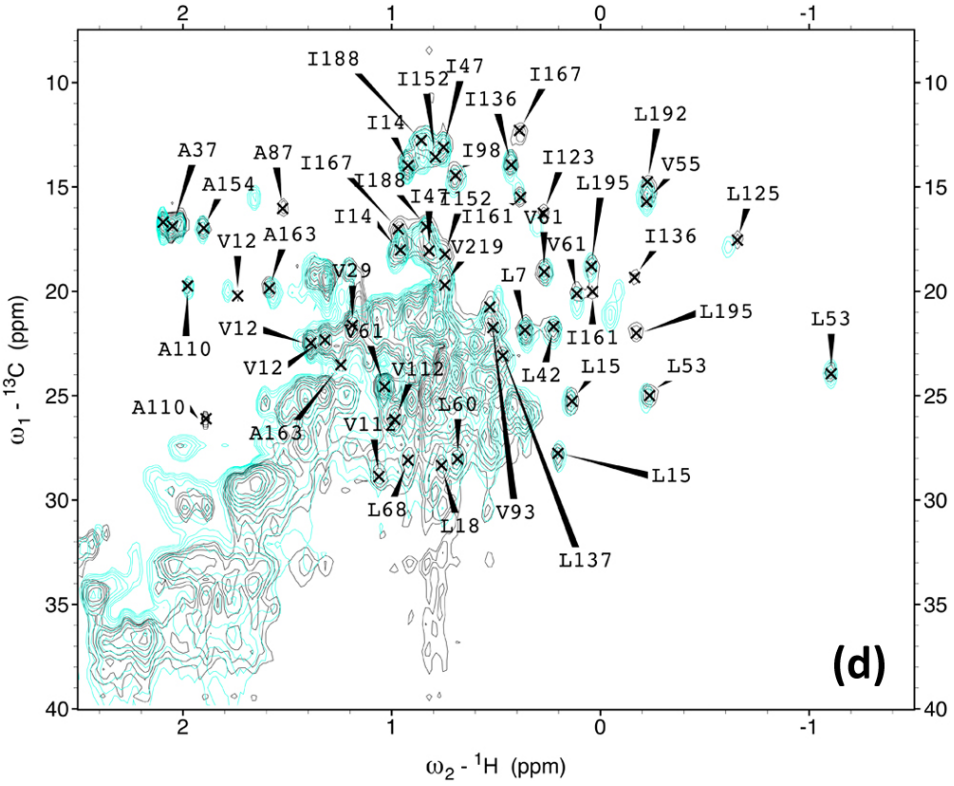
Figure 4.13 (a) Superposition of 2D ^{1}H - ^{15}N spectra of C-terminal truncations at cross peak of Tyr106. C11 (yellow), C12 (red), C13 (blue), C14 (turquoise), C15 (purple), f(0) (dark grey), and YFP (wt, grey) focusing Tyr106 at 8.5 to 9.8 ppm for ^{1}H, 112 to 121 ppm. Using assignment of Venus, each observable cross peak was assigned by the Sparky software. (**b**) **Peak intensity of Tyr106**: C11, C12, C13, C14, C15, C16, C17, C24, and f(0). (**c**) Image of YFP (wt) indicating Tyr106 (highlited in pink) which is situated in 5th β-strand.

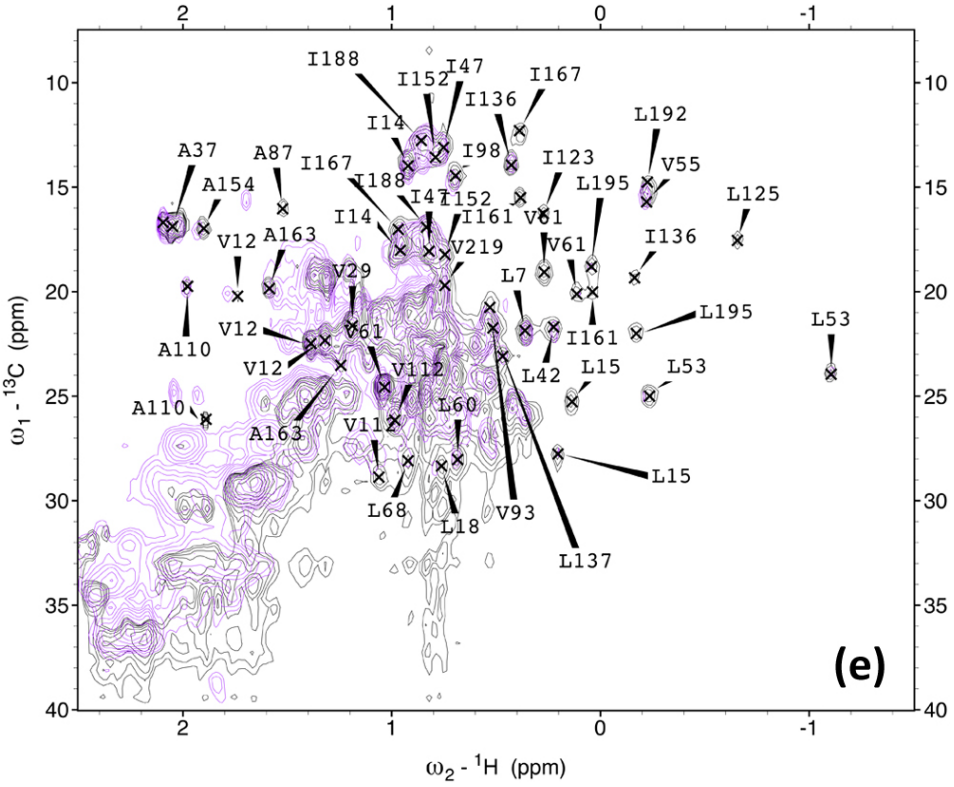
4.6.6 2D analysis of the C-terminal truncations using ^{13}C HMQC experiments

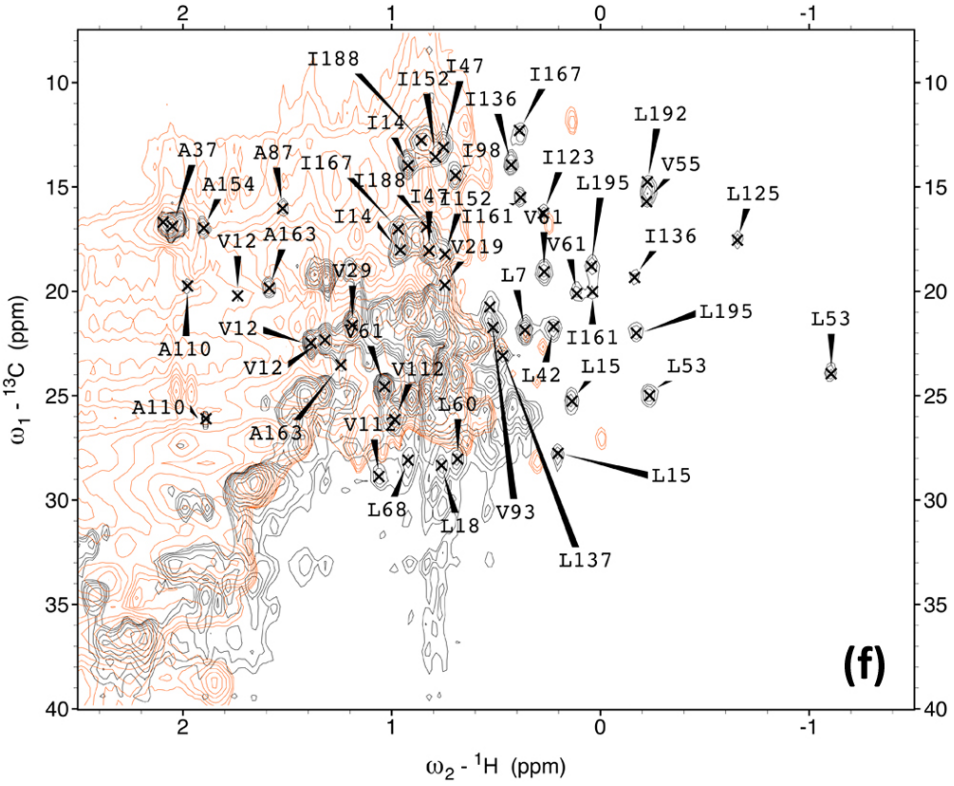
Figure 4.14 demonstrates matched ^{1}H - ^{13}C correlation spectra of the methyl C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, f(0), and mature YFP (wt), in which the methyl side chains of leucine, isoleucine, valine and alanine are observable. Methyl side chains are typically contained buried and thus are useful probes of a protein's structure and these will be explored in more detail in Section 4.6.1. Similar to what was observed within the ^{1}H - ^{15}N correlation spectra in Figure 4.9, the truncations C11, C12, C13, C14, and C15, overlaid with YFP (wt) in Figure 4.14 (a - e). In contrast, it is shown that in spectra of C16, C17 and C24, the well-dispersed cross peaks have almost disappeared in Figure 4.14 (f - h), most likely a result of aggregation.

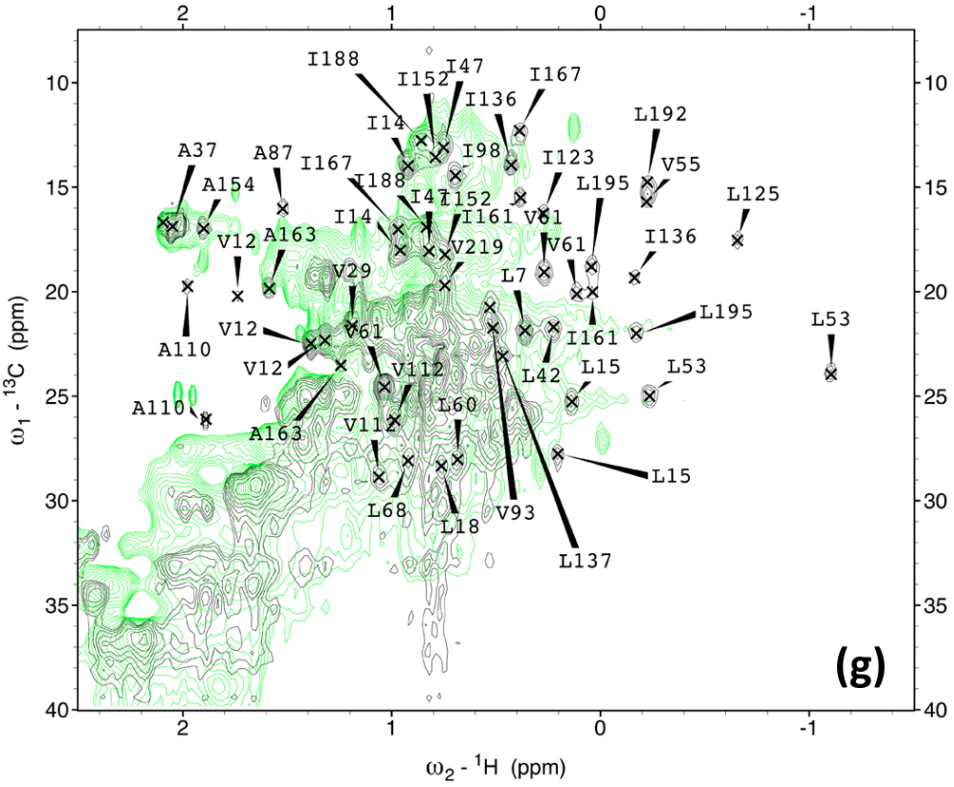

(a)
(b)
ω_1 - ^{13}C (ppm)
ω_2 - ^{1}H (ppm)
I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
I47
I152
L195
L125
I14
I161
V61
V12
A163
V219
I136
V29
L7
V12
V61
L195
L53
A110
I161
V12
V112
L42
L15
L53
A110
A163
L60
V112
L15
L68
V93
L18
L137

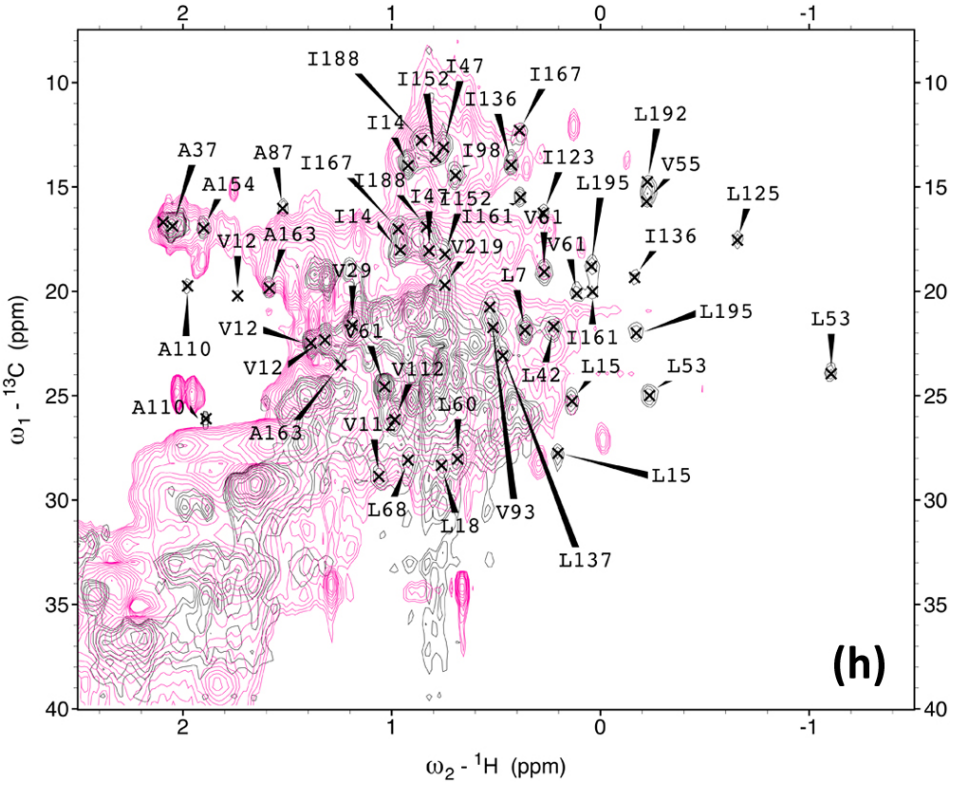

I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V61
V29
L7
L195
L53
V12
V61
I161
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(c)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V61
V29
L7
L195
L53
V12
V61
I161
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(d)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V29
V61
L7
V12
V61
I161
L195
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(e)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
I14
V12
A163
V219
I136
V29
V61
L7
V12
V61
I161
L195
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(f)
2
1
0
-1
10
15
20
25
30
35
40
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I167
I98
I123
V55
A154
I188
L195
L125
I47
I152
I161
V81
I14
V12
A163
V219
V61
I136
V29
L7
L195
V12
V61
I161
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(g)
ω_1 - ^{13}C (ppm)
ω_2 - ^{1}H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I167
I98
I123
V55
A154
I188
L195
L125
I47
I152
I161
V81
I14
V12
A163
V219
V61
I136
V29
L7
L195
V12
V61
I161
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
L18
V93
L137
(h)
ω_1 - ^{13}C (ppm)
ω_2 - ^{1}H (ppm)

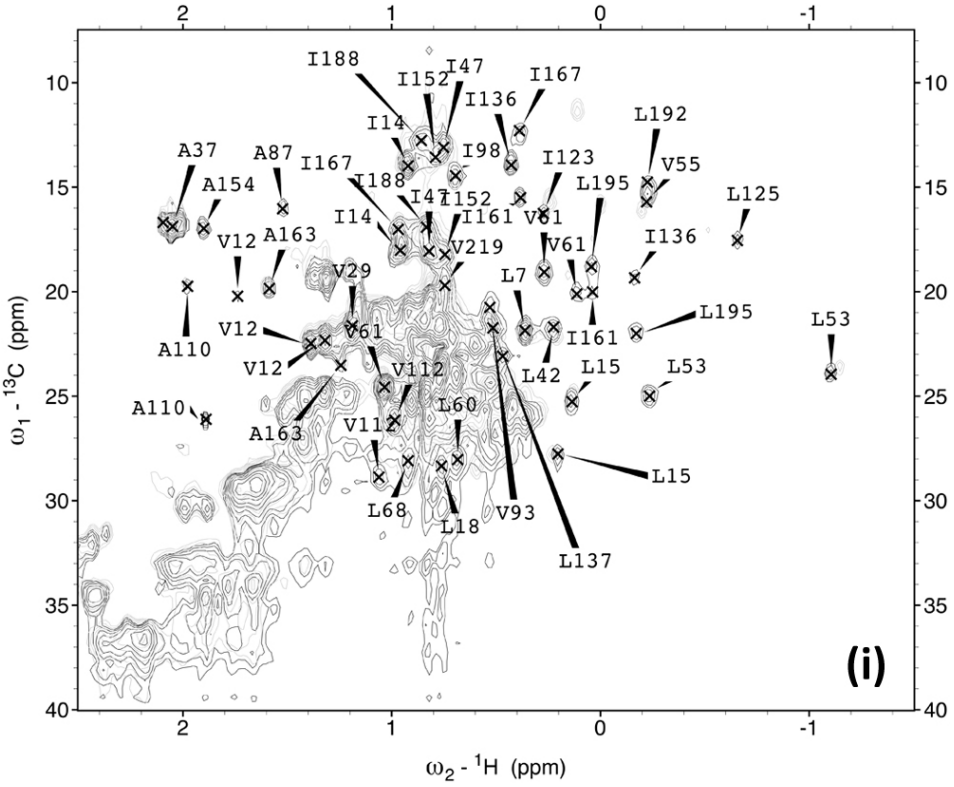


Figure 4.14 ^{1}H - ^{13}C 2D correlation spectra of C-terminal truncations in (**a**) ^{1}H - ^{13}C 2D spectra of C11 (yellow), (**b**) C12 (red), (**c**) C13 (blue), (**d**) C14 (turquoise), (**e**) C15 (purple), (**f**) C16 (orange), (**g**) C17 (light green), (**h**) C24 (pink), and (**i**) f(0) (grey) with Venus assignment.

4.6.7 Translational diffusion properties of the C-terminal truncations

The determination of hydrodynamic radii of proteins on diffusion measurements demonstrates a simple but solid understanding for folding conformations. To measure the hydrodynamic radii of YFP C-terminus truncations, 1D ^{15}N-XSTE NMR diffusion measurement [145] was performed at 25℃. The diffusion coefficient and hydrodynamic radius (r_h) were calculated by the Stejskal - Tanner equation [146]:

$$I = I_0 \exp(-D\ (\Delta - \frac{\delta}{3})\ (\gamma\ \delta\ G)^2)$$

(equations 4.2)

$$rh = \frac{kT}{6\pi\upsilon D}$$

(equations 4.3)

Equation 4.2 - 3 Stejskal - Tanner equation. Diffusion co-efficient of C-terminal truncations were evaluated by this equation as demonstrated in **Figure 4.15** and **Table 4.1**. Some part of equation was modified from original formula for data set obtained.

The signal intensity for C11 was plotted as demonstrated in Figure 4.15 and which gives rise to a diffusion coefficient of 1.08 × 10^{-10} m^2s^{-1} which corresponds to a hydrodynamic radius of 2.24 nm. This value is lower than values obtained from YFP (wt) which showed diffusion coefficient of 1.08 × 10^{-10} m^2s^{-1} and hydrodynamic radius of 2.26 nm. This could potentially be attributed to the loss of 11 amino acid residues from the last β-strand on YFP (wt). Similarly, 1D ^{15}N XSTE experiments were subsequently recorded for C12, C13, C16, C17, C24, and f(0) variants and overall, they gave rise to very similar diffusion coefficients as obtained for YFP (wt), with an average value of 1.075 × 10^{-10} m^2s^{-1} which correspond to 27.8 to 30.5 kDa in their molecular weight. For C14 and C15, the diffusion coefficients were higher compared to YFP (wt), giving rise to smaller hydrodynamic radii of 2 and 1.74 nm respectively. It is

possible that in these truncations, there is a compaction that is observed.

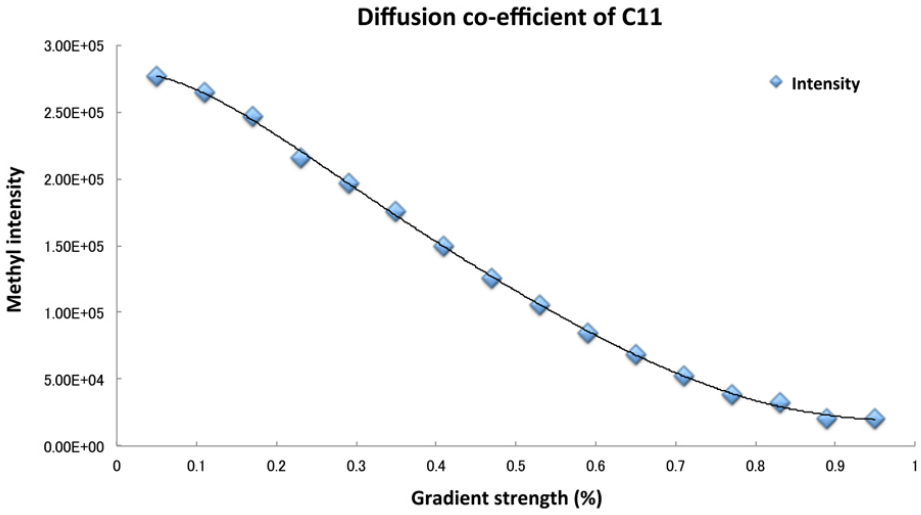


Figure 4.15 A plot of the decay in signal intensity in ^{15}N-XSTE diffusion experiment of C11. Methyl intensity (y-axis) versus the relative gradient strength, G% (x-axis), for C-terminal truncation, C11. Data is fitted to the Stejskal-Tanner equation.

In the case of lower diffusion coefficient to YFP (wt), there is a possibility that the variants were not in monomeric state or unfolded. Yet, most C-terminal truncations except C14 and C15 demonstrated comparable diffusion coefficient of 1.075×10^{-10} m^2s^{-1} in average. The denatured state diffusion coefficients of C14 and C15 were determined to be 1.22×10^{-10} m^2s^{-1} and 1.4×10^{-10} m^2s^{-1} corresponding to hydrodynamic radii of 2 and 1.74 nm were also determined by the Stejskal - Tanner equation, but these fast diffusion coefficients are not comparable to the value determined for YFP (wt) and other variants which may correspond only differences of 1.4 to 1.5 kDa molecular weights, but show that these denatured state are more compact than expected for the proteins of the same lengths completely denatured.

Variants	*Diffusion coefficient*	r_h *(nm)*	*Expected M.W.*
YFP (wt)	$1.08 \times 10^{-10}\ m^2s^{-1}$	2.26 nm	28.6 kDa
C11	$1.09 \times 10^{-10}\ m^2s^{-1}$	2.24 nm	27.8 kDa
C12	$1.08 \times 10^{-10}\ m^2s^{-1}$	2.26 nm	28.6 kDa
C13	$1.07 \times 10^{-10}\ m^2s^{-1}$	2.28 nm	29.5 kDa
C14	$1.22 \times 10^{-10}\ m^2s^{-1}$	2 nm	19 kDa
C15	$1.4 \times 10^{-10}\ m^2s^{-1}$	1.74 nm	11.9 kDa
C16	$1.06 \times 10^{-10}\ m^2s^{-1}$	2.3 nm	30.5 kDa
C17	$1.08 \times 10^{-10}\ m^2s^{-1}$	2.26 nm	28.6 kDa
C24	$1.07 \times 10^{-10}\ m^2s^{-1}$	2.28 nm	29.5 kDa
f(0)	$1.08 \times 10^{-10}\ m^2s^{-1}$	2.26 nm	28.6 kDa

Table 4.1 Diffusion coefficient and expected molecular weight of C-terminal truncations, C11, C12, C13, C14, C15, C16, C17, C24, f(0), and YFP (wt).

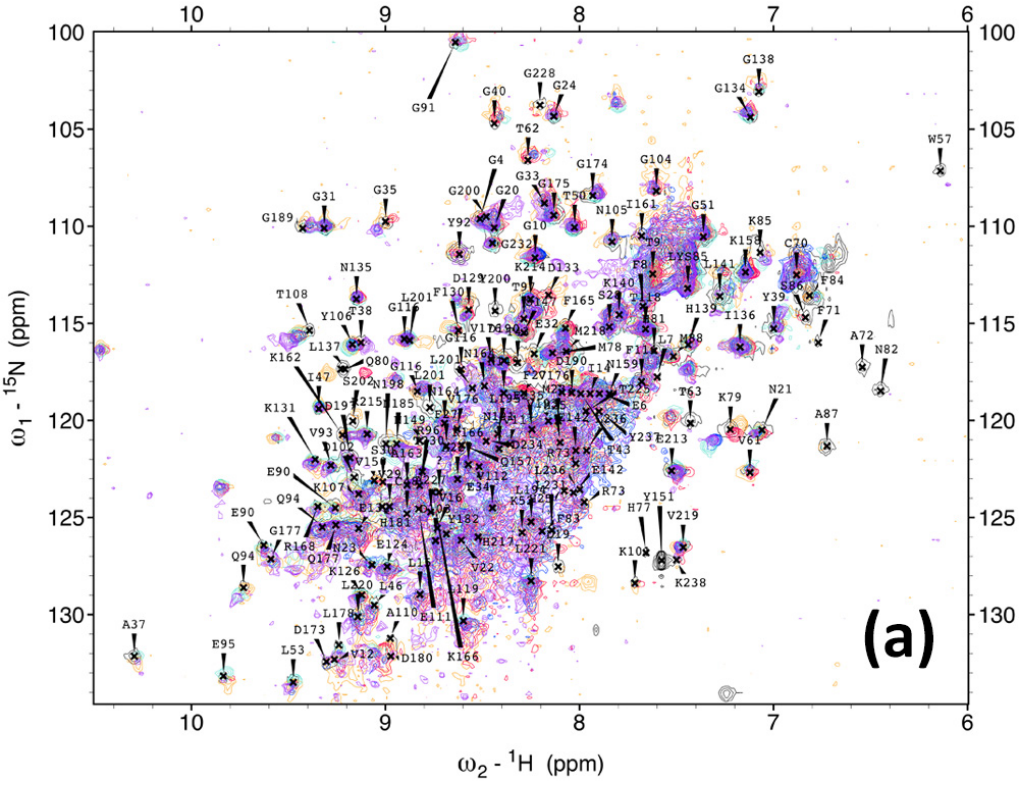
4.7 Minimum amino acids sequence requirements for YFP to acquire structure: rationalising fluorescence and NMR observations

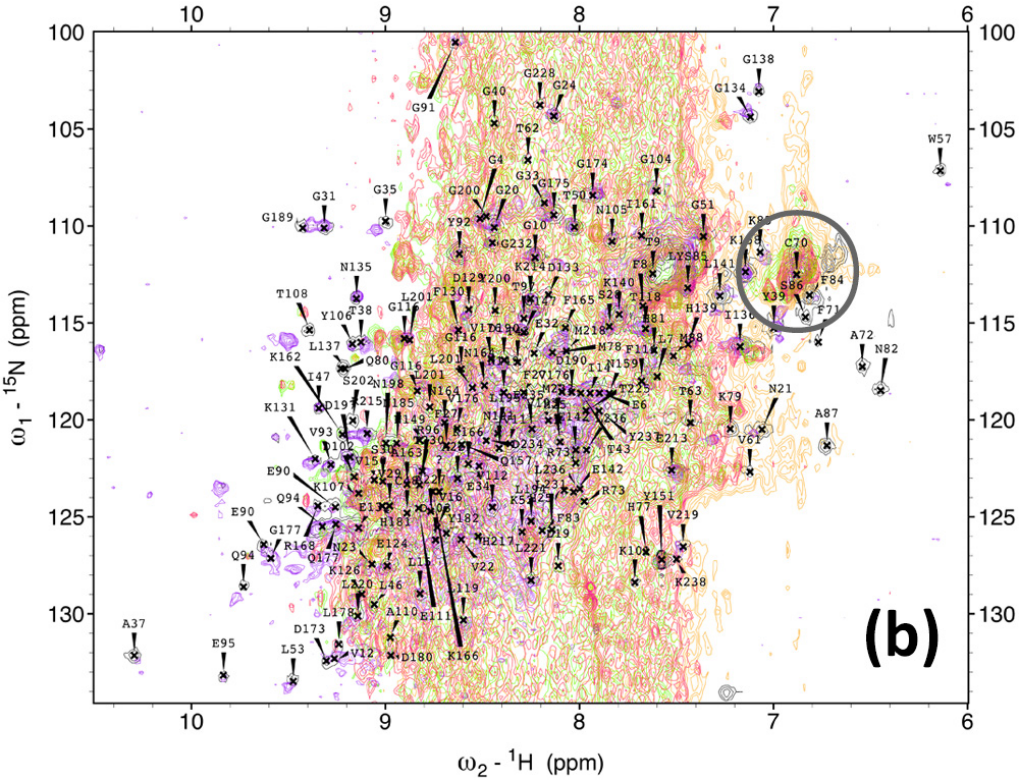
4.7.1 Comparison of the C-terminal truncations using ^{1}H - ^{15}N and ^{1}H - ^{13}C 2D correlations

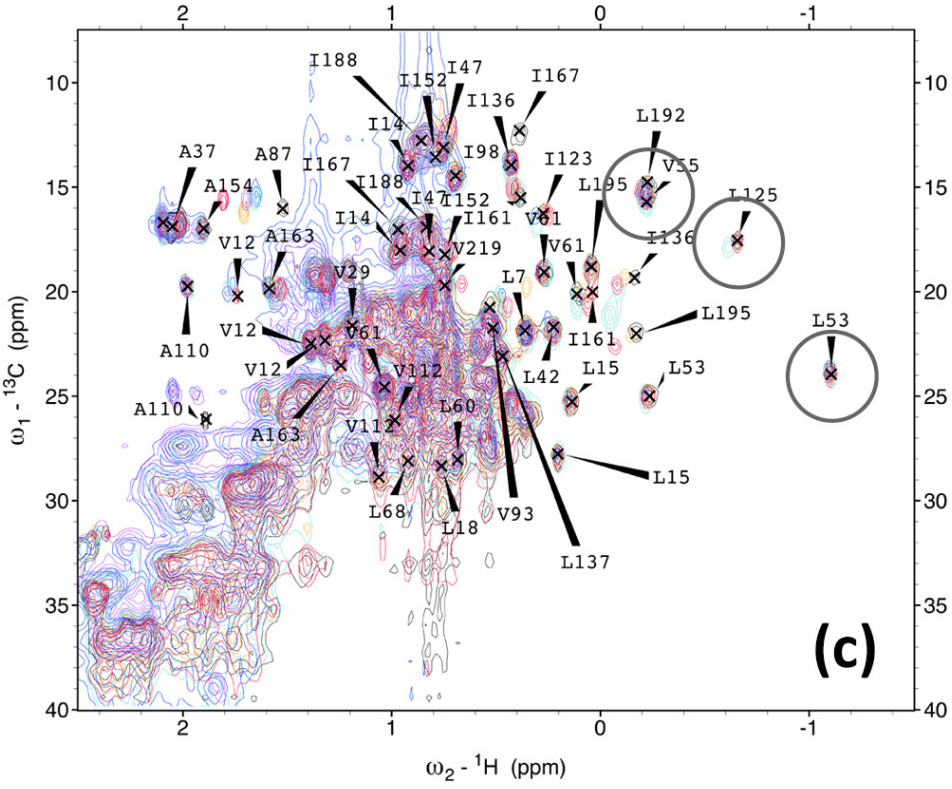
As has previously been demonstrated by spectroscopies of fluorescence and NMR, C15 truncation behaved as a key factor to show the intrinsic yellow fluorescence and to maintain the solid structure of YFP. The structural differences observed in C16, C17, and C24 are obvious to YFP (wt), C11, C12, C13, and C14, and C15 truncation stands on the boundary of folding and intrinsic yellow fluorescence of YFP. At least, it can be expected that as demonstrated by GFP [82], deletion or conformational change of electrostatic interactions between G65 and E222, G67 and R96 may cause a possible structural change on YFP as happen on GFP.

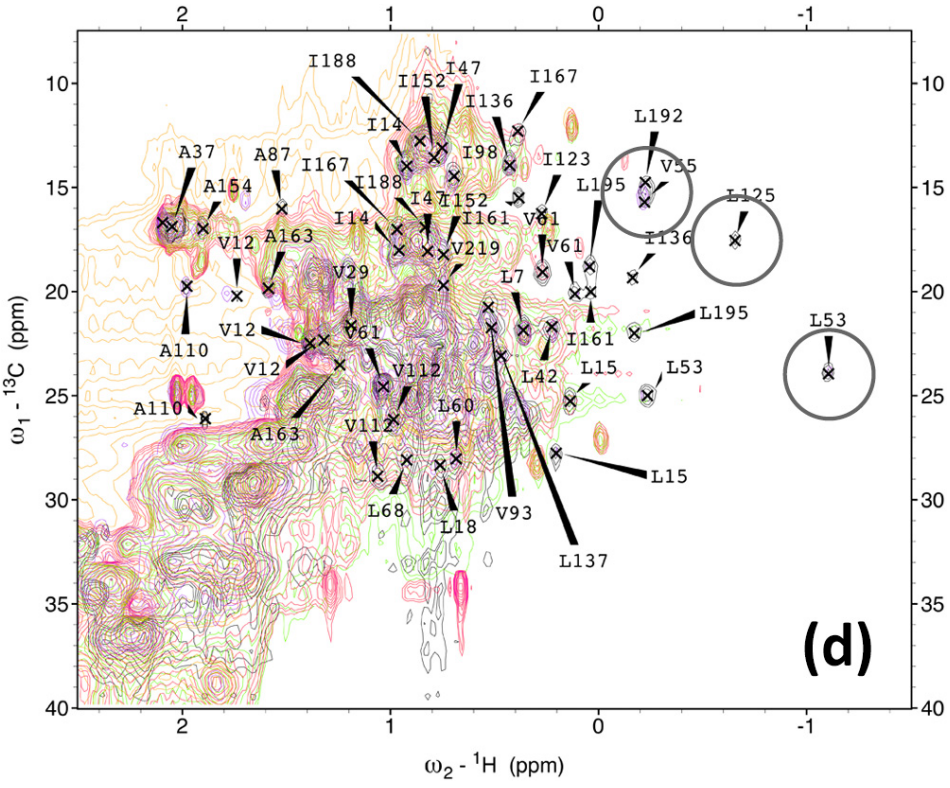
Following to the 2D spectra of C-terminal truncations, it has been shown that observable similarity between C11, C12, C13, C14, C15 and YFP (wt) can be found in both ^{1}H - ^{15}N and ^{1}H - ^{13}C. However, as intrinsic yellow fluorescence revealed, difference between C16, C17, C24 and YFP (wt) are more obvious as demonstrated in Figures 4.9 and 4.14. Figure 4.16 illustrates observable similarity between C11 and C15, and difference between C15 and C24. For example, Figure 4.16 (a), truncations C11 to C15 show observable similarity to YFP (wt) in 2D ^{1}H -^{15}N spectrum. Most cross peaks correspond to original YFP (wt) assignment, yet some cross peaks vanished particularly in C15 showed 44.9% +/-1.8% of intrinsic yellow fluorescence. On the other hand, Figure 4.16 (b), truncations C16 to C24 show a noticeable difference to YFP (wt). C16

to C24 truncations only share C70 cross peak with YFP (wt) which is circled in grey in Figure 4.16 (b).


(a)
ω1 - 15N (ppm)
ω2 - 1H (ppm)


(b)
ω1 - 15N (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
V12
A163
I14
V219
I136
V61
V29
L7
L195
V12
V61
I161
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
V93
L18
L137
(c)
ω1 - 13C (ppm)
ω2 - 1H (ppm)


I188
I152
I47
I167
I136
I14
L192
A37
A87
I98
I123
V55
A154
I167
I188
L195
L125
I47
I152
I161
V61
V12
A163
I14
V219
I136
V61
V29
L7
L195
V12
V61
I161
L53
A110
V12
V112
L42
L15
L53
A110
A163
V112
L60
L15
L68
V93
L18
L137
(d)
ω1 - 13C (ppm)
ω2 - 1H (ppm)

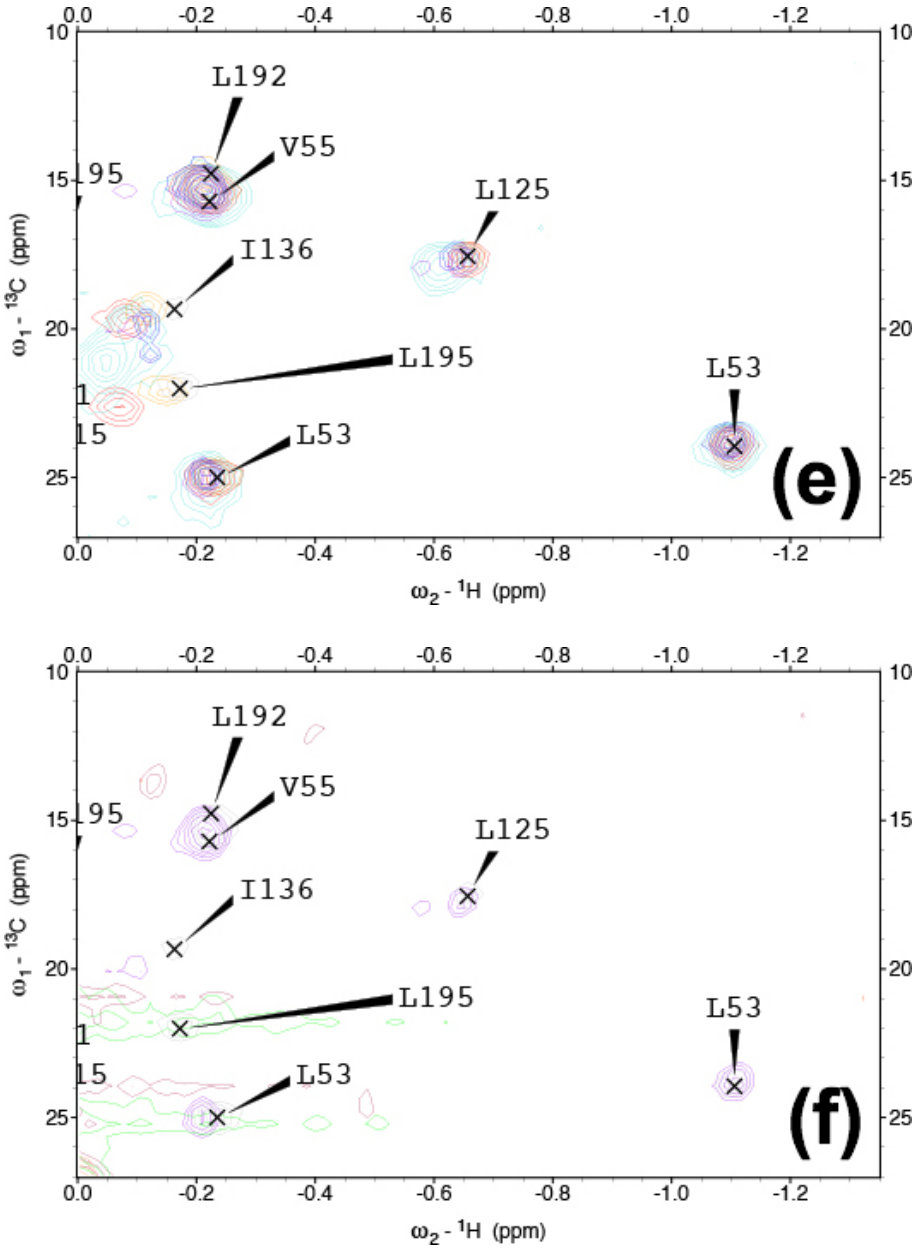


Figure 4.16 Comparison of the C-terminal truncations using ^{1}H - ^{15}N and ^{1}H - ^{13}C 2D correlations. (**a**) ^{1}H - ^{15}N 2D spectra of C11 (yellow), C12 (red), C13 (blue), C14 (turquoise), and C15 (purple) overlaid with YFP (wt, grey) with Venus assignment. (**b**) ^{1}H - ^{15}N 2D spectra of C15 (purple), C16 (orange), C17 (green), and C24 (pink) overlaid with YFP (wt). (**c**): ^{1}H - ^{13}C 2D spectra of C11, C12, C13, C14, overlaid with C15 and YFP (wt). (d): ^{1}H - ^{13}C 2D spectra of C15, C16, C17, and C24 overlaid with YFP (wt), (**e, f**) Superpositions of 2D ^{1}H - ^{13}C spectra of under and post C15.

Similarity and difference between C11 to C15 and C16 to C24 can be also observed in 2D correlation spectra of ^{1}H - ^{13}C in this figure. In figure 4.16 (c), C11 to C15 show well-assigned cross peaks compared to YFP (wt). As observed in Figure 4.16 (c), some cross peaks evanesced from YFP (wt) or shift to positive or negative values in ^{1}H or ^{13}C 2D spectra, but in total they still seem to be folded. This is particular because C11 to C15 show cross peaks from L53, V55, L125, and L192 (circled in grey). The methyl region located between 0.5 and - 1.5 ppm in ^{1}H dimension, and 10 and 30 ppm in ^{13}C dimension indicate the cross peaks from folded proteins [19, 88, 147], therefore these cross peaks may indicate that C11 to C15 are folded as observed in YFP (wt). Figure 4.16 (d) shows not only obvious difference of the cross peaks of C16 to C24 to YFP (wt) but also evanescence of L53, V55, L125, and L192. It can be clearly observed that YFP (wt) and truncations C11 to C15 maintain these methyl cross peaks, yet C16 to C24 truncations loss these.

4.8 Discussion and concluding remarks

4.8.1 The boundary of intrinsic yellow fluorescence between C14 to C16: under and post C15

From these results, now it can be estimated that glutamine, number 222 plays an important key role for maintain the solid structure of YFP as demonstrated on GFP [82]. In this study, a hydrogen bond between G65 and E222 on C15, and the effect of the hydrogen bond to F223 on C16 construct are now to be subjected. Particularly about molecular function between fluorophore on GFP and E222, Barondeau et al., have shown a result by X-ray crystallography that there are several primarily electrostatic interactions between G65 and E222, G67 and R96, e.g. hydrogen bonds or Van der Waals interactions that maintain its intrinsic green fluorescence [82]. Especially, these interactions between E222 and G65 (described as mutation S65G) on the fluorophore of GFP are essential to support its intrinsic green fluorescence at atomic resolution standard. Having say that YFP is a mutant of GFP including S65G mutation, these interactions between G65 and E222 are still to work showing intrinsic yellow fluorescence on YFP molecule either, or it can be concluded that the function of E222 on YFP and interactions between G65 and E222 that are critical to understand folding of YFP (however, in original Venus assignment [120], G65 and E222 were not assigned in both 2D ^{1}H - ^{15}N and ^{1}H - ^{13}C spectra, so this YFP C-terminal truncation study could not explore further in NMR).

Even though C16 variant still maintain E222 at the very end of C-terminus, the construct did not appear to show its intrinsic yellow fluorescence. Yet C15 maintains both E222 and F223 at the end of C-terminus showing half intrinsic yellow fluorescence about 44.9% +/-1.8%. These 2 amino acids, E222 and F223, therefore, can be

considered as the key residues to observe the conformational change between F223 at the end of C15 and E222 of C16 on YFP that show ‘half’ and ‘zero’ intrinsic yellow fluorescence at residual specific level.

Overall, YFP C-terminal truncations can be classified to the three phenotypes: C11 to C14 (and f(0)) as native like fold states, C15 as molten globule state, and C16 to C24 as unfolded and exposed for aggregation states (classified as ‘under’ and ‘post’ C15). Not only NMR but also behaviour observed by fluorescence of intrinsic yellow, extrinsic, and aromatic residue excitation indicate that distinctive conformational changes occurred between under C15 (C11 to C14) and post C15 (C16 to C24) truncations. This YFP C-terminal truncation study conclusively demonstrated an understanding about the vanishing point of intrinsic yellow fluorescence that is sustained by a ‘healthy’ fluorophore formation, and a minimum sequence requirement demonstrated by C15 and C16 that maintain F223 and E222 at the C-terminus respectively.

Chapter 5 Discussion

Behaviour of YFP-RNCs and C-terminal truncations: boundary of intrinsic yellow fluorescence

5.1 The folding characteristics of YFP-RNCs by fluorescence and NMR spectroscopy

The members of FPs family are well studied and applied in the field of modern biochemistry because of their favourable biochemical characteristics, e.g. efficient folding, rapid maturation of the fluorophore, high fluorescent quantum yields, and an ability to tolerate a range of pH and ion conditions. It is for these reasons that YFP is also particularly attractive in the study of protein folding. To explore co-translational protein folding on the ribosome, a cell-free approach was developed for YFP to produce translationally-arrested RNCs: YFP17, YFP34, YFP55, for both NMR and biophysical studies. The folding of YFP on the ribosome was, therefore, explored by generating three RNC lengths that represent 'folding snapshots' of the protein. In addition, isolated YFP C-terminal truncations: C11, C12, C13, C14, C15, C16, C17, C24, and f(0), were also created to compare the folding behaviour of YFP both as it occurs on the ribosome, using a mimic of an isolated protein.

As shown in Chapter 3, a cell-free approach was developed which permitted isotopically labelled RNC samples to be generated, which would be suitable for NMR as well as fluorescence spectroscopy. Two cell-free based strategies were examined, an S30 extract and a commercial RTS500 kit, and using each of these, it was possible to generate analytical as well as preparative quantities of high quality RNCs for both

spectroscopies of fluorescence and NMR respectively.

Initially, RNCs of YFP17, YFP34, and YFP55 in which YFP is at distances of 19, 36 and 57 amino acid residues were evaluated using fluorescence spectroscopy using both intrinsic and extrinsic approaches. Fluorescence approaches showed that for each of the RNCs, a clear intrinsic yellow fluorescence was observed, which suggests that the YFP was capable of adopting a native fold at distances ranging from 19 to 57 amino acid residues from the PTC site suggesting that folding is possible very close to the ribosome's exit tunnel. Interestingly, the fluorescence intensity appeared to correlate with the increasing linker length, in which YFP34 and YFP55 had a significantly higher intensity values relative to YFP17 was much lower. Further evidence to support these observations were found in fluorescence anisotropy results which correlated well to the rotational ability of the YFP-NCs fused to the ribosome; YFP17 gave rise to higher R-values compared to that of YFP55. The YFP fluorescence in each of the RNCs was clearly observable, suggesting a native fold, although the intensity of YFP17 appeared a relative quenche suggesting a possibility that was partially folded. These RNCs were therefore investigated using bis-ANS experiments. It was anticipated that limited ANS fluorescence would be observed for the folded YFP34 and YFP55, and that a small degree of fluorescence might be observable for YFP17. These ANS data, however, showed an unexpected correlation, in which fluorescence increased linker length extended. This suggested that very likely the YFP was indeed folded, however there was partial structure being formed within the linker, which gave rise to the ANS fluorescence.

To further explore, the structural properties of the YFP-RNCs, NMR spectroscopy was also investigated, as it can offer residue-specific detail, however this was found to be

challenging. This was in part due to the low intensity signals in both the ^{1}H - ^{15}N and ^{1}H - ^{13}C dimensions that were observed for the YFP-RNCs, in which it was not possible to assess the data on a residue-by-residue basis. As the NMR signal is entirely dependent on the concentration of the bound NC, one possibility for these low intensity may be the result of the stability of the samples during the purification process: 10-30%, w/v sucrose gradient, 12 hours, which gave rise to variable levels of occupancy: 18 % for YFP34 and 44% for YFP55. Another contribution to the low intensity signals, may also be the result of interactions between NCs and the 70S ribosomal surface which have huge surfaces. Despite the limitations in behaviour of RNCs in NMR, it was possible to compare the overall structure of YFP55 with isolated YFP (wt) and it was found YFP55 shared many hallmarks of having a folded conformation, as a number of well-dispersed YFP resonances were observed. These NMR data therefore tells that YFP can adopt its native conformation even close to the ribosome.

5.2 Behaviour of YFP C-terminal truncations - Function of E222 for YFP folding

The capacity for YFP to show any fluorescence close to the ribosome is also interesting as the shortest linker, YFP17 was also found to be folding-competent. These results were especially intriguing in the context of the expected capacity of the ribosomal exit tunnel. The ribosomal exit tunnel holds approximately 30 amino acids in an extended conformation, and in YFP17, at least c.a. 11 amino acids from YFP would therefore be present in the tunnel. It suggests that at least part of the YFP sequence can reside within the tunnel and the rest of NC can still adopt significant degree of structure so that some native-folding can take place. To understand these results further, RNC studies were complemented with studies using C-terminal truncations of isolated YFP

(wt) in order to reproduce folding process *in vitro* environment.

Having YFP C-terminal truncations, it was found using both fluorescence and NMR that YFP could tolerate the progressive removal of its C-terminus in which a native-like conformation as observed using intrinsic yellow fluorescence was observed when up to 14 amino acids were removed: under C15. It was also found, however, that upon removal of 15 amino acids: C15, resulted in a decrease in intrinsic yellow fluourescence intensity (44.9 +/- 1.8% compared to YFP (wt)). The removal of 16 amino acids and up to 24 amino acids: post C15, resulted in the entire loss of intrinsic yellow fluorescence; this was likely due to the mutations destabilising the structure, as post C15 resulted in proteins that were highly aggregation-prone demonstrated by NMR.

Further analysis of the C-terminal truncations indicated some intriguing insights regarding the capacity to support its native-structure. As mentioned, the removal of 15 amino acid residues resulted in the losses of function and structure, and further truncations, post C15, resulted in the misfolding and aggregation. This provides some interesting insights into the ability for YFP to adopt structure; the 15 amino acid residues of the protein correspond to the final β-strand. C-terminal truncation study has shown that a stop codon just after E222 on very end of C-terminus C16 resulted in the complete losses of function and structure corresponding to the loss of 16 amino acids; the loss of part of the β-strand, therefore, disrupts the primarily electrostatic contributions that come from G65 as well as the contributions from an interaction between G67 and R96 which forms a part function of the fluorophore to sustain conformational folding [82].

5.3 Comparisons to the other studies

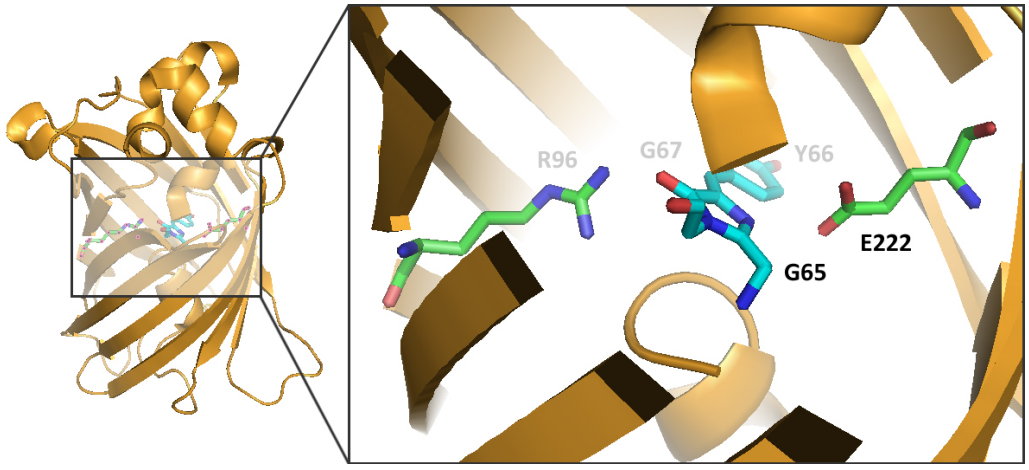


Figure 5.1 A schematic of possible primary electrostatic interactions, e.g. hydrogen bonds, Van der waals interactions, between G65 and E222. Having study of GFP [82], interactions between G67 and R96 on YFP is also expected to sustain the conformational fold of YFP and to show intrinsic yellow fluorescence from the fluorophore which is consisted of G65, Y66, and G67 [81].

As described in previous sections, Barondeau et al., demonstrated in this paper that E222 has a hydrogen bond to S65G which is a part of functional group interactions between the R96, E222, and T62 carbonyl oxygen atoms and the fluorophore formations (in this paper, GFP was expressed as S65G mutation) [82]. Previous GFP studies also showed that there is a hydrogen bond between S65 and E222 on GFP [148] which is subject of a possible role of the final proton acceptor on the fluorophore of GFP at ESPT [149], and that E222 forms a hydrogen bond to S205 which is in the hydrogen bond chain with protonated fluorophore on YFP [150]. These may suggest that E222 on YFP has a primary role to maintain folding and function of the fluorophore.

When the YFP-RNCs are evaluated in the context of these C-terminal truncations, it suggests that in YFP17, there is a sufficient amount of the protein sequence that is outside of the tunnel to permit native contacts to be formed. Further linker lengths of the YFP i.e. one in which up to 15 amino acids of YFP reside in the tunnel, would permit the 'vanishing point' of intrinsic yellow fluorescence and provide more information of how structure is acquired co-translationally for YFP. It appears that YFP can fold very close to the ribosomal exit tunnel, and these results contrast those to recent studies of Dom5 (Cabrita et al., unpublished data; Launay H, PhD Thesis, 2011). The Dom5, which forms a β-sheet sandwich was found to have a very different minimum sequence length for its folding; the 105 amino acids Dom5 needed to entirely exposed and at a significant distance for about more than 40 amino acid residues away from the ribosome's PTC site. On the other hand, 12 deletion of C-terminal truncation of Dom5 ($ddFln_{646-738}$) showed a number of same cross peaks to isolated Dom5 in ^{1}H - ^{15}N and ^{1}H - ^{13}C in 2D spectra. Beyond that, C-terminal truncations of Dom5 gradually loss the folded conformation (Karyadi ME, PhD thesis). These suggest that there are inherent differences between the capacity for emerging NCs of different sequence to adopt structure.

The folding of the YFP and C-terminal truncations can also be compared to a previous folding study of FPs on the ribosome. Kelkar et al., studied the folding of four different FPs using RNCs in different length of tether lengths on cystic fibrosis transmembrane conductance regulator (CFTR) N-terminus of 15, 22, 40, and 103 residues, and it was shown by fluorescent spectroscopy that 15 amino acid length showed 0-5%, 22 amino acid length showed 5-10%, 40 amino acid length showed 70%, and 103 amino acid length showed 100% of intrinsic red fluorescence from mCherry respectively (the results were compared to the maximal intensity). By exploring the result, they

concluded that 15 amino acid length of tether was too short for folding to be completed. Indeed, with the 15 amino acid tether, there was only observable fluorescence when the RNCs were released from the ribosome with an RNAse digestion. Again this highlights the importance of the final β-strand on the C-terminus in folding all FPs share in their conformational structure. And intriguingly in the Kelkar study, it was shown that the YFP must be at least 40 amino acids away from the ribosome for folding to be observed. One of the important issues in the Kelkar study is the possibility that the CFTR tether may form secondary structure, and may in fact, bring YFP closer to the ribosome than anticipated [151]; indeed it has been shown previously that NCs can form at least an α-helical structure within the ribosomal tunnel [27]. Therefore, in the case of the CFTR tether, if it is compacting with the ribosomal exit tunnel, this may limit the amount of YFP sequence that is available outside.

Another feature observed from this study is that the minimum sequence requirements for YFP to acquire its native conformational structure differ to that of GFP. Indeed, as demonstrated by several previous studies [134, 135, 136], using C-terminal truncations of isolated GFP an 10 to 12 amino acid deletion resulted in the loss of intrinsic green fluorescence owing to the loss of the part of the final β-strand on the molecule and causing unfolding of the protein (in these studies, it was shown that 10 deletion resulted in 56%, and 11 to 12 deletions resulted in zero intrinsic green fluorescence). It suggests that despite the high homology between GFP and YFP that there are inherent differences in their stability and their ability to fold. It is therefore pertinent to understand the folding of different family members on a case-by-case basis.

5.4 Future study

The results presented in this study show the development of an *E.coli* cell-free system for the production of RNCs, and this study would be the first example in which a dual biophysics and NMR approach has been used to monitor the folding of RNCs. The strategy for YFP also highlighted the differences in stability requirements for RNCs, in which biophysical studies which are relatively quick (minutes scale) meant that the RNCs could be studied easily following a double sucrose cushion purification step in a small scale. In contrast, NMR studies required RNCs to be purified using a more laborious strategy and required hours of data acquisition, and the overall occupancy and integrity of the RNCs were therefore compromised. Future NMR studies of YFP-RNCs will therefore need more careful consideration (e.g. different choice of linker) to ensure that the RNCs are stable and remain stable during the experiment. In addition, the possibility that YFP may be interacting with the ribosome, as described in Sections 3.6.2 and 3.6.3 in Chapter 3 may also require further evaluation of buffer conditions to potentially minimise these interactions.

The study of co-translational folding of YFP is thus still ongoing, however, this study has expanded a number of avenues for future study, particularly in the development of cell-free based approaches. During the course of this doctrate study, the cell-free strategies have permitted pilot studies of other RNCs to be conducted in both biophysics and NMR, including α-synuclein and ddFLN have been already expressed by the cell-free expression with the collaboration with colleagues within the laboratory. The flexibility of the cell-free strategy also means that alternative, more complex isotopic labeling strategies can be employed more readily. For example, selective isotopic labeling of methyl containing side-chains in Dom5-RNCs are being combined with

TROSY-based NMR experiments to directly monitor folded structure in a highly sensitive way. This is complicated by the presence of background ribosomal resonances; this problem would be overcome by using cell-free, in which the ribosomes are NMR-silent.

In addition, cell-free approaches for RNCs will permit the incorporation of non-natural as well as fluorescence amino acids that would complement the structural studies using NMR. YFP is unique in that it has a fluorescence that has an obvious distinction from the ribosome, and the incorporation of fluorescent amino acids with different spectral properties into NCs would permit similar biophysical studies to take place. Currently, cell-free approaches are being used to generate different ribosome-associated complexes for structural biology. For example, structural studies have examined complexes such as TF-RNC [152], SRP-RNC [153], and calmodulin (CaM)-RNC [154]. In more ambitious terms, the capacity to initiate translation in a very controlled manner and would ,therefore, permit initiate NC folding in real time, and thus offer the advantage of studying protein folding kinetics. This would present an incredibly exciting opportunity to complement the equilibrium 'folding snapshots' as observed by NMR spectroscopy. Also, an understanding about an interaction between G65 and E222 at ESPT is highly expected for future study of YFP fluorophore formation using X-ray crystallography or MD simulations with C15 and C16 C-terminal truncations in high resolution.

Chapter 6 References

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

[2] Ellis RJ, Hartl FU. (1999) Principles of protein folding in the cellular environment. Curr Opin Struct Biol. 1999 Feb;9(1):102-10.

[3] Onuchic JN. (1997) Theory of Protein Folding: The Energy Landscape Perspective. Annu. Rev. Phys. Chem. 1997. 48:545–600

[4] Li Z, Scheraga HA. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci U S A. 1987 Oct;84(19):6611-5.

[5] Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. (1989) Mapping the transition state and pathway of protein folding by protein engineering. Nature. 1989 Jul 13;340(6229):122-6.

[6] De S, Krishnadev O, Srinivasan N, Rekha N. (2005) Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. BMC Struct Biol. 2005 Aug 16;5:15.

[7] Fersht A. (1999) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman and Company.

[8] Dill KA, MacCallum JL. (2012) The protein-folding problem, 50 years on. Science. 2012 Nov 23;338(6110):1042-6.

[9] Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci. 1999 Sep;24(9):329-32.

[10] Elliot PR, Pei XY, Dafforn TR, Lomas DA (2000) Topography of a 2.0 A structure of a1-antitrypsin reveals targets for rational drug design to prevent conformational disease. Protein Science (2000), 9:1274–1281.

[11] Shaw P (2005) Molecular and cellular pathways of neurodegeneration in motor neurone disease. J Neurol Neurosurg Psychiatry. Aug 2005; 76(8): 1046–1057.

[12] Sato S, Ward CL, Krouse ME, Wine JJ, Kopito RR. (1996) Glycerol reverses the misfolding phenotype of the most common cystic fibrosis mutation. J Biol Chem. 1996 Jan 12;271(2):635-8.

[13] Ban N, Nissen P, Hansen J, Moore PB, Steizt TA. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289, 905-20.

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

[15] Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache JP, Villa E, Trabuco LG, Becker T, Mielke T, Schulten K, Steitz TA, Beckmann R. (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. Science. 2009 Dec 4;326(5958):1412-5.

[16] Bhushan S, Hoffmann T, Seidelt B, Frauenfeld J, Mielke T, Berninghausen O, Wilson DN, Beckmann R. (2011) SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center. PLoS Biol. 2011; 9(1):e1000581

[17] Arenz S, Ramu H, Gupta P, Berninghausen O, Beckmann R, Vázquez-Laslop N, Mankin AS, Wilson DN. (2014) Molecular basis for erythromycin-dependent ribosome stalling during translation of the ErmBL leader peptide. Nat Commun. 2014 Mar 24;5:3501.

[18] Christodoulou J, Larsson G, Fucini P, Connell SR, Pertinhez TA, Hanson CL, Redfield C, Nierhaus KH, Robinson CV, Schleucher J, Dobson CM. (2004) Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. PNAS. 101, 10949-10954.

[19] Hsu ST, Fucini P, Cabrita LD, Launay H, Dobson CM, Christodoulou J. (2007). Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. Proc Natl Acad Sci U S A. 2007 Oct 16;104(42):16516-21.

[20] Wilson DN, Doudna Cate JH. (2012) The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol. 2012 May 1;4(5). pii: a011536.

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

[22] Crick F. (1970) Central Dogma of Molecular Biology. Nature Vol.227 August 1970.

[23] http://www.t6.lanl.gov/kys/

[24] Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. Science. 2006 Sep 29;313(5795):1935-42. Epub 2006 Sep 7.

[25] Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. (2000) The structural basis of ribosome activity in peptide bond synthesis. Science. 2000 Aug 11;289(5481):920-30.

[26] Steitz TA. (2008) A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol. 2008 Mar;9(3):242-53.

[27] Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, Berninghausen O, Wilson DN, Beckmann R. (2010) α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. Nat Struct Mol Biol. 2010 Mar;17(3):313-7.

[28] Woolhead CA, Johnson AE, Bernstein HD. (2006) Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. Mol Cell. 2006 Jun 9;22(5):587-98.

[29] Lu J, Hua Z, Kobertz WR, Deutsch1 C. (2011) Nascent Peptide Side-chains Induce Rearrangements in Distinct Locations of the Ribosomal Tunnel. J Mol Biol. Aug 12, 2011; 411(2): 499–510.

[30] Woolhead CA, McCormick PJ, Johnson AE. (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. Cell. 2004 Mar 5;116(5):725-36.

[31] Kramer G, Boehringer D, Ban N, Bukau B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol. 2009 Jun;16(6):589-97. doi: 10.1038/nsmb.1614.

[32] Hartl FU, Hayer-Hartl M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science. 2002 Mar 8;295(5561):1852-8.

[33] Lafontaine DL, Tollervey D. (2001) The function and synthesis of ribosomes. Nat Rev Mol Cell Biol. 2001 Jul;2(7):514-20.

[34] Zhang G, Ignatova Z. (2009) Generic Algorithm to Predict the Speed of Translational Elongation: Implications for Protein Biogenesis. PLoS ONE;2009, Vol. 4 Issue 4,

[35] Rutkowska A, Mayer MP, Hoffmann A, Merz F, Zachmann-Brand B, Schaffitzel C, Ban N, Deuerling E, Bukau B. (2008) Dynamics of trigger factor interaction with translating ribosomes. J Biol Chem. 2008 Feb 15;283(7):4124-32. Epub 2007 Nov 28.

[36] Goemaere E, Melet A, Larue V, Lieutaud A, Alves de Sousa R, Chevalier J, Yimga-Djapa L, Giglione C, Huguet F, Alimi M, Meinnel T, Dardel F, Artaud I, Pagès JM. (2012) New peptide deformylase inhibitors and cooperative interaction: a combination to improve antibacterial activity. J Antimicrob Chemother. 2012 Jun;67(6):1392-400.

[37] Sandikci A, Gloge F, Martinez M, Mayer MP, Wade R, Bukau B, Kramer G. (2013) Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding. Nat Struct Mol Biol. 2013 Jul;20(7):843-50.

[38] Zhang G, Hubalewska M, Ignatova Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat Struct Mol Biol. 2009 Mar;16(3):274-80.

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

[40] Ziv G, Haran G, Thirumalai D. (2005) Ribosome exit tunnel can entropically stabilize alpha-helices. Proc Natl Acad Sci U S A. 2005 Dec 27;102(52):18956-61. Epub 2005 Dec 15.

[41] Frauenfeld J, Gumbart J, Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, Schulten K, Beckmann R. (2011) Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nat Struct Mol Biol. 2011 May;18(5):614-21.

[42] Gilbert RJ, Fucini P, Connell S, Fuller SD, Nierhaus KH, Robinson CV, Dobson CM, Stuart DI. (2004) Three-dimensional structures of translating ribosomes by Cryo-EM. Mol Cell. 2004 Apr 9;14(1):57-66.

[43] Gumbart J, Schreiner E, Wilson DN, Beckmann R, Schulten K. (2012) Mechanisms of SecM-mediated stalling in the ribosome. Biophys J. 2012 Jul 18;103(2):331-41. doi: 10.1016/j.bpj.2012.06.005. Epub 2012 Jul 17.

[44] Zengel JM, Jerauld A, Walker A, Wahl MC, Lindahl L. (2003) The extended loops of ribosomal proteins L4 and L22 are not required for ribosome assembly or L4-mediated autogenous control. RNA. 2003 Oct;9(10):1188-97.

[45] Nirenberg MW, Mattheath JH. (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A. 1961 Oct 15;47:1588-602.

[46] Kigawa T. Yabuki T. Matsuda N. Matsuda T. Nakajima R. Tanaka A. Yokoyama S. (2004) Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. Journal of Structural and Functional Genomics 5: 63–68.

[47] Findeis MA, Whitesides GM. (1987) Protein synthesis in cell-free reticulocyte lysates on multi-hour incubation. Appl Biochem Biotechnol. 1987 Oct;15(3):169-89.

[48] Seki E, Matsuda N, Yokoyama S, Kigawa T. (2007) Cell-free protein synthesis system from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation, Analytical Biochemistry 377 (2008) 156–161.

[49] Hodgman CE, Jewett MC. (2013) Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. Biotechnol Bioeng. 2013 Apr 23. doi: 10.1002/bit.24942.

[50] Spirin AS. (2004) High-throughput cell-free systems for synthesis of functionally active proteins. Trends Biotechnol. 2004 Oct;22(10):538-45.

[51] Shimizu Y, Ueda T. (2010) PURE technology. Methods Mol Biol. 2010;607:11-21.

[52] Morita EH, Sawasaki T, Tanaka R, Endo Y, Kohno T. (2003) A wheat germ cell-free system is a novel way to screen protein folding and function. Protein Sci. 2003 Jun;12(6):1216-21.

[53] Oliver CL, Boyd CD. (1985) *In vitro* translation of messenger RNA in a rabbit reticulocyte lysate cell-free system. Methods Mol Biol. 1985;2:145-55.

[54] Katzen F, Chang G, Kudlicki W. (2005) The past, present and future of cell-free protein synthesis. Trends Biotechnol. 2005 Mar;23(3):150-6.

[55] Kigawa T, Muto Y, Yokoyama S. (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. J Biomol NMR. 1995 Sep;6(2):129-34.

[56] Vinarov DA, Lytle BL, Peterson FC, Tyler EM, Volkman BF, Markley JL. (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. Nat Methods. 2004 Nov;1(2):149-53.

[57] Ozawa K, Wu PS, Dixon NE, Otting G. (2006) ^{15}N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. FEBS J. 2006 Sep;273(18):4154-9. Epub 2006 Aug 23.

[58] Wakiyama M, Kaitsu Y, Yokoyama S. (2006) Cell-free translation system from Drosophila S2 cells that recapitulates RNAi. Biochem Biophys Res Commun. 2006 May 19;343(4):1067-71. Epub 2006 Mar 23.

[59] Niwa T, Kanamori T, Ueda T, Taguchi H. (2012) Global analysis of chaperone effects using a reconstituted cell-free translation system. Proc Natl Acad Sci U S A. 2012 Jun 5;109(23):8937-42. doi: 10.1073/pnas.1201380109. Epub 2012 May 21.

[60] Schaffitzel C, Ban N. (2007) Generation of ribosome nascent chain complexes for structural and functional studies. Journal of Structural Biology 158: 463-47.

[61] Hethke C, Bergerat A, Hausner W, Forterre P, Thomm M. (1999) Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of Pyrococcus transcription. Genetics. Aug 1999; 152(4): 1325-1333.

[62] Valeur B. (2001). Molecular Fluorescence: Principles and Applications, Wiley-VCH. ISBN 352729919X,

[63] Abbyad P, Childs W, Shi X, Boxer SG. (2007) Dynamic Stokes shift in green fluorescent protein variants. Proc Natl Acad Sci U S A. 2007 Dec 18;104(51):20189-94.

[64] Vivian JT, Callis PR. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. Biophys. J. 80 (5): 2093–109.

[65] Shimomura O. (1979) Structure of the Chromophore of *Aequorea* Green Fluorescent Protein, FEBS Lett., 104, 220-222.

[66] Davenport D, Nicol JAC. (1955) Luminescence in Hydromedusae. Proc. R. Soc. Lond. B 29 November 1955 vol. 144 no. 916 399-411

[67] Shimomura O, Johnson FH, Saiga Y. (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from Luminous Hydromedusan, Aequorea, J. Cell. Comp. Physiol., 59, 223-239.

[68] Zhang L, Patel HN, Lappe JW, Wachter RM. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein, J. Am. Chem. Soc., 128, 4766-4772.

[69] Branchini BR, Nemser AR, Zimmer M. (1998) A computational analysis of the unique protein-induced tight turn that results in posttranslational chromophore formation in green fluorescent protein, J. Am. Chem. Soc, 120, 1-6.

[70] Morise H, Shimomura O, Johnson FH, Winant J. (1974) Intermolecular energy transfer in the bioluminescent system of *Aequorea*, Biochemistry, 13, 2656-2662.

[71] Patterson GH, Knobel SM, Sharif WD, Kain SR, Piston DW. (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy, Biophys. J., 73, 2782-2790.

[72] Chattoraj M, King BA, Bublitz GU, Boxer SG. (1996) Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer, Proc. Natl. Acad. Sci. U.S.A., 93, 8362-8367.

[73] Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein, Proc. Natl. Acad. Sci. U.S.A., 94, 2306-2311.

[74] Tsien RY. (1998) The green fluorescent protein, Annu. Rev. Biochem., 67, 509-544.

[75] Zapata-Hommer O, Griesbeck O. (2003) Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP, BMC Biotechnol., 3, 5.

[76] Heim R, Cubitt AB, Tsien RY. (1995) Improved green fluorescence, Nature, 373, 663-664.

[77] Misteli T, Spector DL. (1997) Applications of the green fluorescent protein in cell biology and biotechnology. Nat Biotechnol. 1997 Oct;15(10):961-4.

[78] Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol. 2002 Jan;20(1):87-90.

[79] Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol. 2004 Dec;22(12):1567-72. Epub 2004 Nov 21.

[80] Miyawaki A, Nagai T, Mizuno H. (2003) Mechanisms of protein fluorophore formation and engineering. Curr Opin Chem Biol. 2003 Oct;7(5):557-62.

[81] Agata R, Jean-Rene A, Nagai T. Miyawaki A, Ikura M. (2002) Crystal Structure of Venus, a Yellow Fluorescent Protein with Improved Maturation and Reduced Environmental Sensitivity. The Journal of Biological Chemistry. 277:50573–50578.

[82] Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. Proc Natl Acad Sci U S A. 2003 Oct 14;100(21):12111-6.

[83] Jayaraman S, Haggie P, Wachter RM, Remington SJ, Verkman AS. (2000) Mechanism and cellular applications of a green fluorescent protein-based halide sensor. J Biol Chem. 2000 Mar 3;275(9):6047-50.

[84] Hsu ST, Blaser G, Behrens C, Cabrita LD, Dobson CM, Jackson SE. (2009) Folding Study of Venus Reveals a Strong Ion Dependence of its Yellow Fluorescence under Mildly Acidic Conditions. J Biol Chem. 2010 February 12; 285(7): 4859–4869.

[85] Kay LE, Ikura M, Tschudin R, Bax A. (2011) Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. J Magn Reson. 2011 Dec;213(2):423-41.

[86] Sprangers R, Kay LE. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007 Feb 8;445(7128):618-22. Epub 2007 Jan 21.

[87] Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K. (2002) NMR analysis of a 900K GroEL GroES complex. Nature. 2002 Jul 11;418(6894):207-11.

[88] Waudby CA, Launay H, Cabrita LD, Christodoulou J. (2013) Protein folding on the ribosome studied using NMR spectroscopy. Prog Nucl Magn Reson Spectrosc. 2013 Oct;74:57-75.

[89] Cavanagh J, Fairbrother WJ, Palmer III, AG, Rance M, Skelton NJ. (2007) Principle and Practice Protein NMR spectroscopy, Elsevier Academic Press.

[90] Bermel W, Bruix M, Felli IC, Kumar MVV, Pierattelli R, Serrano S. (2013) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR. 2013 Mar;55(3):231-7.

[91] Wishart DS, Sykes BD, Richards FM. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol. 1991 Nov 20;222(2):311-33.

[92] Wang L, Markley JL. (2009) Empirical correlation between protein backbone ^{15}N and ^{13}C secondary chemical shifts and its application to nitrogen chemical shift re-referencing. J Biomol NMR. 2009 Jun;44(2):95-9.

[93] Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. FEBS J. 2011 Mar;278(5):687-703.

[94] Losonczi JA, Andrec M, Fischer MW, Prestegard JH. (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson. 1999 Jun;138(2):334-42.

[95] Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci U S A. 2007 Jun 5;104(23):9615-20. Epub 2007 May 29.

[96] Robustelli P, Kohlhoff K, Cavalli A, Vendruscolo M. (2010) Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. Structure. 2010 Aug 11;18(8):923-33.

[97] Choy WY, Mulder FA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol. 2002 Feb 8;316(1):101-12.

[98] Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry. 1999 Dec 14;38(50):16424-31.

[99] Augustyniak R, Ferrage F, Paquin R, Lequin O, Bodenhausen G. (2011) Methods to determine slow diffusion coefficients of biomolecules: applications to Engrailed 2, a partially disordered protein. J Biomol NMR. 2011 Jul;50(3):209-18.

[100] Orekhov VY, Korzhnev DM, Kay LE. (2004) Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. J Am Chem Soc. 2004 Feb 18;126(6):1886-91.

[101] Zeeb M, Balbach J. (2004) Protein folding studied by real-time NMR spectroscopy. Methods. 2004 Sep;34(1):65-74.

[102] Baldwin AJ, Kay LE. (2009) NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol. 2009 Nov;5(11):808-14.

[103] Gal M, Mishkovsky M, Frydman L. (2006) Real-time monitoring of chemical transformations by ultrafast 2D NMR spectroscopy. J Am Chem Soc. 2006 Jan 25;128(3):951-6.

[104] Ferrage F, Zoonens M, Warschawski DE, Popot JL, Bodenhausen G. (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. J Am Chem Soc. 2003 Mar 5;125(9):2541-5.

[105] Sattler M, Schleucher J, Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc
34:93–158.

[106] http://spin.niddk.nih.gov/NMRPipe/

[107] http://www.cgl.ucsf.edu/home/sparky/

[108] Shi X, Basran J, Seward HE, Childs W, Bagshaw CR, Boxer SG. (2007) Anomalous negative fluorescence anisotropy in yellow fluorescent protein (YFP 10C): quantitative analysis of FRET in YFP dimers. Biochemistry. 2007 Dec 18;46(50):14403-17. Epub 2007 Nov 21.

[109] Shi Y, Fan DJ, Li SX, Zhang HJ, Perrett S, Zhou JM. (2007) Identification of a potential hydrophobic peptide binding site in the C-terminal arm of trigger factor. Protein Sci. 2007 Jun;16(6):1165-75.

[110] Hawe A, Poole R, Jiskoot W. (2010) Misconceptions over Förster resonance energy transfer between proteins and ANS/bis-ANS: Direct excitation dominates dye fluorescence. Anal Biochem. 2010 Jun 1;401(1):99-106.

[111] Marcel AM, Martin vandeVen M, Ulises AU, Valeur B. (2013) Fluorescence anisotropy measurements in solution: Methods and reference materials (IUPAC Technical Report) 589 Pure Appl. Chem. Vol. 85, No. 3, pp. 589–608, 2013.

[112] http://www.photobiology.info/Visser-Rolinski.html

[113] Fukuda H, Arai M, Kuwajima K. (2000) Folding of green fluorescent protein and the cycle3 mutant. Biochemistry. 2000 Oct 3;39(39):12025-32.

[114] Huang JR, Hsu ST, Christodoulou J, Jackson SE. (2008) The extremely slow-exchanging core and acid-denatured state of green fluorescent protein. HFSP J. 2008 Dec;2(6):378-87.

[115] Enoki S, Saeki K, Maki K, Kuwajima K. (2004) Acid denaturation and refolding of green fluorescent protein. Biochemistry. 2004 Nov 9;43(44):14238-48.

[116] Jackson SE, Craggs TD, Huang JR. (2006) Understanding the folding of GFP using biophysical techniques. Expert Rev Proteomics. 2006 Oct;3(5):545-59.

[117] Reid BG, Flynn GC. (1997) Chromophore formation in green fluorescent protein. Biochemistry. 1997 Jun 3;36(22):6786-91.

[118] Hsu ST, Blaser G, Jackson SE. (2009) The folding, stability and conformational dynamics of beta-barrel fluorescent proteins. Chem Soc Rev. 2009 Oct;38(10):2951-65.

[119] Seifert MH, Ksiazek D, Azim MK, Smialowski P, Budisa N, Holak TA. (2002) Slow exchange in the chromophore of a green fluorescent protein variant. J Am Chem Soc. 2002 Jul 10;124(27):7932-42.

[120] Hsu ST, Behrens C, Cabrita LD, Dobson CM. (2009) ^{1}H, ^{15}N and ^{13}C assignments of yellow fluorescent protein (YFP) Venus. Biomol NMR Assign. 2009 Jun;3(1):67-72.

[121] Trabuco LG, Harrison CB, Schreiner E, Schulten K. (2010) Recognition of the regulatory nascent chain TnaC by the ribosome. Structure. 2010 May 12;18(5):627-37.

[122] Hue KK, Bechhofer DH. (1991) Effect of ermC leader region mutations on induced mRNA stability. J Bacteriol. 1991 Jun;173(12):3732-40.

[123] Rutkowska A, Beerbaum M, Rajagopalan N, Fiaux J, Schmieder P, Kramer G, Oschkinat H, Bukau B. (2009) Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. FEBS Lett. 2009 Jul 21;583(14):2407-13.

[124] Eichmann C, Preissler S, Riek R, Deuerling E. (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. Proc Natl Acad Sci U S A. 2010 May 18;107(20):9111-6.

[125] Voss NR, Gerstein M, Steitz TA, Moore PB. (2006) The geometry of the ribosomal polypeptide exit tunnel. J Mol Biol. 2006 Jul 21;360(4):893-906. Epub 2006 May 30.

[126] Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, Berninghausen O, Wilson DN, Beckmann R. (2010) alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. Nat Struct Mol Biol. 2010 Mar;17(3):313-7.

[127] Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W. (2008) Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nat Struct Mol Biol. 2008 May;15(5):494-9.

[128] Evans MS, Ugrinov KG, Frese MA, Clark PL. (2005) Homogeneous stalled ribosome nascent chain complexes produced *in vivo* or *in vitro*. Nat Methods. 2005 Oct;2(10):757-62.

[129] http://imagej.nih.gov/ij/

[130] Hawe A1, Sutter M, Jiskoot W. (2008) Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res. 2008 Jul;25(7):1487-99.

[131] Vanderheeren G, Hanssens I, Noyelle K, Van Dael H, Joniau M. (1998) The perturbations of the native state of goat alpha-lactalbumin induced by 1,1'-bis(4-anilino-5-naphthalenesulfonate) are Ca2+-dependent. Biophys J. 1998 Nov;75(5):2195-204.

[132] Nakatogawa H, Ito K. (2002) The ribosomal exit tunnel functions as a discriminating gate. Cell. 2002 Mar 8;108(5):629-36.

[133] O'Brien EP, Hsu ST, Christodoulou J, Vendruscolo M, Dobson CM. (2010) Transient tertiary structure formation within the ribosome exit port. J Am Chem Soc. 2010 Dec 1;132(47):16928-37.

[134] Kim HK, Kaang BK. (1998) Truncated green fluorescent protein mutants and their expression in Aplysia neurons. Brain Res Bull. 1998 Sep 1;47(1):35-41.

[135] Dopf J, Horiagon TM. (1996) Deletion mapping of the *Aequorea* victoria green fluorescent protein. Gene. 1996;173(1 Spec No):39-44.

[136] Li X, Zhang G, Ngo N, Zhao X, Kain SR, Huang CC. (1997) Deletions of the *Aequorea* victoria green fluorescent protein define the minimal domain required for fluorescence. J Biol Chem. 1997 Nov 7;272(45):28545-9.

[137] Zhang G, Gurtu V, Kain SR. (1996) An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. Biochem Biophys Res Commun. 1996 Oct 23;227(3):707-11.

[138] Chow CC, Chow C, Raghunathan V, Huppert TJ, Kimball EB, Cavagnero S. (2003) Chain length dependence of apomyoglobin folding: structural evolution from misfolded sheets to native helices. Biochemistry. 2003 Jun 17;42(23):7090-9.

[139] De Prat Gay G, Ruiz-Sanz J, Neira JL, Corrales FJ, Otzen DE, Ladurner AG, Fersht AR. (1995) Conformational pathway of the polypeptide chain of chymotrypsin inhibitor-2 growing from its N terminus in vitro. Parallels with the protein folding pathway. J Mol Biol. 1995 Dec 15;254(5):968-79.

[140] Griko YV, Gittis A, Lattman EE, Privalov PL. (1994) Residual structure in a staphylococcal nuclease fragment. Is it a molten globule and is its unfolding a first-order phase transition? J Mol Biol. 1994 Oct 14;243(1):93-9.

[141] Neira JL, Fersht AR. (1999) Acquisition of native-like interactions in C-terminal fragments of barnase. J Mol Biol. 1999 Mar 26;287(2):421-32.

[142] Kreisberg JF, Betts SD, Haase-Pettingell C, King J. (2002) The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. Protein Sci. 2002 Apr;11(4):820-30.

[143] Lakowicz JR. (1999) Principles of Fluorescence Spectroscopy ISBN 0306460939

[144] Nyon MP, Segu L, Cabrita LD, Lévy GR, Kirkpatrick J, Roussel BD, Patschull AO, Barrett TE, Ekeowa UI, Kerr R, Waudby CA, Kalsheker N, Hill M, Thalassinos K, Lomas DA, Christodoulou J, Gooptu B. (2012) Structural dynamics associated with intermediate formation in an archetypal conformational disease. Structure. 2012 Mar 7;20(3):504-12.

[145] Ferrage F, Zoonens M, Warschawski DE, Popot JL, Bodenhausen G. (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. J Am Chem Soc. 2003 Mar 5;125(9):2541-5.

[146] Sinnaeve D. (2012) The Stejskal - Tanner equation generalized for any gradient shape - an overview of most pulse sequences measuring free diffusion. Concepts in Magnetic Resonance Part A, Volume 40A, Issue 2, pages 39–65.

[147] Cabrita LD, Dobson CM, Christodoulou J. (2010) Protein folding on the ribosome. Curr Opin Struct Biol. 2010 Feb;20(1):33-45.

[148] Jung G, Wiehler J, Zumbusch A. (2005) The photophysics of green fluorescent protein: influence of the key amino acids at positions 65, 203, and 222. Biophys J. 2005 Mar;88(3):1932-47. Epub 2004 Dec 21.

[149] Lill MA, V Helms. (2002) Proton shuttle in green fluorescent protein studied by dynamic simulations. Proc. Natl. Acad. Sci. USA. 99:2778–2781.

[150] Zhong S, Navaratnam P, Santos-Sacchi J. (2014) A Genetically-Encoded YFP Sensor with Enhanced Chloride Sensitivity, Photostability and Reduced pH Interference Demonstrates Augmented Transmembrane Chloride Movement by Gerbil Prestin (SLC26a5). PLoS ONE, 01/2014; 9(6):e99095.

[151] Kelkar DA, Khushoo A, Yang Z, Skach WR. (2012) Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. J Biol Chem. 2012 Jan 20;287(4):2568-78.

[152] Merz F, Boehringer D, Schaffitzel C, Preissler S, Hoffmann A, Maier T, Rutkowska A, Lozza J, Ban N, Bukau B, Deuerling E. (2008) Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. EMBO J. 2008 Jun 4;27(11):1622-32.

[153] Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C. (2011) Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor. Nat Struct Mol Biol. 2011 Jan;18(1):88-90.

[154] Lamprou P, Kempe D, Katranidis A, Büldt G, Fitter J. (2014) Nanosecond dynamics of calmodulin and ribosome-bound nascent chains studied by time-resolved fluorescence anisotropy. Chembiochem. 2014 May 5;15(7):977-85.

Chapter A Appendix

A.1 RNC of YFP17 (pLDC-17, full sequence) designed by Dr Lisa D Cabrita.

CAT**ATG**CATCACCATCACCATCATGCTAGCgtgagcaagggcgaggagctgttcaccgg
ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgt
ccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccttcggctacggcctgca
gtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgc
ccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacc
cgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcat
cgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatc
cgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccc
catcggcgacggccccgtgctgctgcccgacaaccactacctgagctaccagtccgccc
tgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcc
gccgggatcactctcggcatggacgagctgtacaagGAATTCTTCAGCACGCCCGTCTG
GATAAGCCAGGCGCAAGGCATCCGTGCTGGCCCT**TAA**

HMHHHHHHASVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTA
AGITLGMDELYKEFFSTPVWISQAQGIRAGP(*stop)

A.2 RNC of YFP34 (pLDC-34, full sequence) designed by Dr Lisa D Cabrita.

CATATGCATCACCATCACCATCATGCTAGCgtgagcaagggcgaggagctgttcaccgg
ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgt
ccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccttcggctacggcctgca
gtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgc
ccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacc
cgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcat
cgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatc
cgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccc
catcggcgacggccccgtgctgctgcccgacaaccactacctgagctaccagtccgccc
tgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcc
gccgggatcactctcggcatggacgagctgtacaagGAATTCAAGGGTTATCGCATTGA
TTATGCGCATTTTACCCCACAAGCAAAAAGTCTCTTCAGCACGCCCGTCTGGATAAGCC
AGGCGCAAGGCATCCGTGCTGGCCCTTAA

HMHHHHHHASVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTA
AGITLGMDELYKEFGCATTKGYRIDYAHFTPQASLFSTPVWISQAQGIRAGP(*stop)

A.3 RNC of YFP55 (pLDC-55, full sequence) designed by Dr Lisa D Cabrita.

CATATGCATCACCATCACCATCATGCTAGCgtgagcaagggcgaggagctgttcaccgg
ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgt
ccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccttcggctacggcctgca
gtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgc
ccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacc
cgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcat
cgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagcc
acaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatc
cgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccc
catcggcgacggccccgtgctgctgcccgacaaccactacctgagctaccagtccgccc
tgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcc
gccgggatcactctcggcatggacgagctgtacaagGAATTCCATCTTGCATTACTGGA
TACGCTCAGCGCGCTGCTGACCCAGGAAGGCACGCCGTCTGAAAAGGGTTATCGCATTG
ATTATGCGCATTTTACCCCACAAGCAAAAAGTCTCTTCAGCACGCCCGTCTGGATAAGC
CAGGCGCAAGGCATCCGTGCTGGCCCTTAA

HMHHHHHHASVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTFGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKI
RHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTA
AGITLGMDELYKEFGCATTHLALLDTLSALLTQEGTPSEKGYRIDYAHFTPQASLFSTP
VWISQAQGIRAGP(*stop)

A.4 Basic informations of RNCs and YFP C-terminal truncations

Name of constructs	*YFP17*	*YFP34*	*YFP55*
Molecular weights of nascent chains	30kDa	32kDa	34.1kDa
Number of amino acids	266	283	303
Extinction coefficients (ε)	29005	31985	31985
Tag	H_6	H_6	H_6

Table 7.1 Basic information of YFP17, YFP34, and YFP55.Extension coefficient (ε) for each construct was calculated in Expasy, ProtParam tool (http://web.expasy.org/protparam/).

	M.W.	*No. of aa.s*	*Ext. coefficient (ε)*	*Tags*	*Last amino acid*
C11	25.5kDa	227	22015	H_6	Ala
C12	25.4kDa	226	22015	H_6	Ala
C13	25.4kDa	225	22015	H_6	Thr
C14	25.3kDa	224	22015	H_6	Val
C15	25.2kDa	223	22015	H_6	Phe
C16	25.1kDa	222	22015	H_6	Glu
C17	24.9kDa	221	22015	H_6	Leu
C24	24.1kDa	214	22015	H_6	Lys
f(0)	26.5kDa	235	22015	H_6	Lys
YFP (wt)	26.7kDa	238	23505	H_6	Lys

Table 7.2 Basic information of YFP variants, C11, C12, C13, C14, C15, C16, C17, C24, and f(0). Extension coefficient (ε) for each construct was calculated in Expasy, ProtParam tool (http://web.expasy.org/protparam/).

A.5 YFP (wt) (238aa)

*DNA sequence and amino acid sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1–20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21–40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41–60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61–80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81–100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101–120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121–140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141–160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161–180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181–200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201–220)

ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221–238)

VSKGEELFT GVVPILVELD (1–20)

GDVNGHKFSV SGEGEGDATY (21–40)

GKLTLKFICT TGKLPVPWPT (41–60)

LVTTFGYGLQ CFARYPDHMK (61–80)

QHDFFKSAMP EGYVQERTIF (81–100)

FKDDGNYKTR AEVKFEGDTL (101–120)

VNRIELKGID FKEDGNILGH (121–140)

KLEYNYNSHN VYIMADKQKN (141–160)

GIKVNFKIRH NIEDGSVQLA (161–180)

DHYQQNTPIG DGPVLLPDNH (181–200)

YLSYQSALSK DPNEKRDHMV (201–220)

LEFVTAAGIT HGMDELYK (221–238)

A.6 YFP C-terminal truncation - C11 (227aa, insertion of stop codon at 227-228aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag ttc gtg acc gcc gcc taa ggg atc act ctc ggc atg gac gag ctg tac aag (221-227)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LEFVTAA(*stop)GIT HGMDELYK (221-238)

Forward: 5′ - gtgaccgccgcctaagggatcactctc - 3′

Reverse: 5′ - gagagtgatcccttaggcggcggtcac - 3′

(*Tm: 74.57℃, GC content: 62.96%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.7 YFP C-terminal truncation - C12 (226aa, insertion of stop codon at 226-227aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag ttc gtg acc gcc taa gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-226)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LEFVTA(*stop)AGIT HGMDELYK (221-226)

Forward: 5′ - ttcgtgaccgcctaagccgggatcact - 3′

Reverse: 5′ - agtgatcccggcttaggcggtcacgaa - 3′

(*Tm: 76.46℃, GC content: 59.26%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.8 YFP C-terminal truncation - C13 (225aa, insertion of stop codon at 225-226aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag ttc gtg acc taa gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-225)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LEFVT(*stop)AAGIT HGMDELYK (221-225)

Forward: 5′ - gag ttc gtg acc taa gcc gcc ggg atc

Reverse: 5′ - gat ccc ggc ggc tta ggt cac gaa ctc

(*Tm: 75.76℃, GC content: 62.96%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.9 YFP C-terminal truncation - C14 (224aa, insertion of stop codon at 224-225aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag ttc gtg taa acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-224)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LEFV(*stop)TAAGIT HGMDELYK (221-224)

Forward: 5′ - ctg gag ttc gtg taa acc gcc gcc ggg

Reverse: 5′ - ccc ggc ggc ggt tta cac gaa ctc cag

(*Tm: 79.66℃, GC content: 66.67%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.10 YFP C-terminal truncation - C15 (223aa, insertion of stop codon at 223-224aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag ttc taa gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-223)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LEF(*stop)VTAAGIT HGMDELYK (221-223)

Forward: 5′ - ctg ctg gag ttc taa gtg acc gcc gcc

Reverse: 5′ - ggc ggc ggt cac tta gaa ctc cag cag

(*Tm: 75.33℃, GC content: 62.96%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.11 YFP C-terminal truncation - C16 (222aa, insertion of stop codon at 222-223aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg gag taa ttc gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-222)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

LE(*stop)FVTAAGIT HGMDELYK (221-222)

Forward: 5′ - gtc ctg ctg gag taa ttc gtg acc gcc

Reverse: 5′ - ggc ggt cac gaa tta ctc cag cag gac

(*Tm: 73.65℃, GC content: 59.26%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.12 YFP C-terminal truncation - C17 (221aa, insertion of stop codon at 221-222aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (201-220)

ctg taa gag ttc gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEKRDHMV (201-220)

L(*stop)EFVTAAGIT HGMDELYK (221)

Forward: 5′ - atggtcctgctgtaagagttcgtgacc - 3′

Reverse: 5′ - ggtcacgaactcttacagcaggaccat - 3′

(*Tm: 68.50℃, GC content: 51.85%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.13 YFP C-terminal truncation - C24 (214aa, insertion of stop codon at 214-215aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc ggc tac ggc ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-80)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (81-100)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (101-120)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (121-140)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (141-160)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (161-180)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (181-200)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag taa cgc gat cac atg gtc ctg (201-214)

ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (221-238)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTFGYGLQ CFARYPDHMK (61-80)

QHDFFKSAMP EGYVQERTIF (81-100)

FKDDGNYKTR AEVKFEGDTL (101-120)

VNRIELKGID FKEDGNILGH (121-140)

KLEYNYNSHN VYIMADKQKN (141-160)

GIKVNFKIRH NIEDGSVQLA (161-180)

DHYQQNTPIG DGPVLLPDNH (181-200)

YLSYQSALSK DPNEK(*stop)RDHMV(201-214)

LEFVTAAGIT HGMDELYK (-)

Forward: 5′ - cccaacgagaagtaacgcgatcacatg - 3′

Reverse: 5′ - catgtgatcgcgttacttctcgttggg - 3′

(*Tm: 71.80℃, GC content: 51.85%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.14 C-terminal truncation - YFP f(0) (235aa, fluorophore deletion between 65-67aa)

*DNA sequence, amino acid sequence, and primer sequence.

gtg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc (1-20)

gac gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc acc tac ggc (21-40)

aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg ccc gtg ccc tgg ccc acc ctc (41-60)

gtg acc acc ttc (ggc tac ggc) ctg cag tgc ttc gcc cgc tac ccc gac cac atg aag cag (61-77)

cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc (78-97)

aag gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac acc ctg gtg (98-117)

aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag (118-137)

ctg gag tac aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac ggc (138-157)

atc aag gtg aac ttc aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac (158-177)

cac tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac aac cac tac (178-197)

ctg agc tac cag tcc gcc ctg agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg (198-217)

ctg gag ttc gtg acc gcc gcc ggg atc act ctc ggc atg gac gag ctg tac aag (218-235)

VSKGEELFT GVVPILVELD (1-20)

GDVNGHKFSV SGEGEGDATY (21-40)

GKLTLKFICT TGKLPVPWPT (41-60)

LVTTF(GYG)LQ CFARYPDHMK (61-77)

QHDFFKSAMP EGYVQERTIF (78-97)

FKDDGNYKTR AEVKFEGDTL (98-117)

VNRIELKGID FKEDGNILGH (118-137)

KLEYNYNSHN VYIMADKQKN (138-157)

GIKVNFKIRH NIEDGSVQLA (158-177)

DHYQQNTPIG DGPVLLPDNH (178-197)

YLSYQSALSK DPNEKRDHMV (198-217)

LEFVTAAGIT HGMDELYK (218-235)

Forward: 5′ - cccaccctcgtgaccaccttcctgcagtgcttcgcccgctac - 3′

Reverse: 5′ - gtagcgggcgaagcactgcaggaaggtggtcacgagggtggg - 3′

(*Tm: 81.16℃, GC content: 66.67%, Ref: Primer3, MIT, http://frodo.wi.mit.edu/primer3/)

A.15 Amino acid composition for ^{15}N - ^{13}C amino acid mixture (Spectra Stable Isotopes, Sigma-Aldrich, UK)

Amino acid	*Avarage relative mole*
Asx (B) Aspartic acid or Asparagine	10.2%
Thr (T) Threonine	6.2%
Ser (S) Serine	4.4%
Glx (Z) Glutamine or Glutamic acid	10.1%
Pro (P) Proline	4.7%
Gly (G) Glysine	10.8%
Ala (A) Alanine	14.0%
Val (V) Valine	8.1%
Met (M) Methionine	2.7%
Ile (I) Isoleucine	5.8%
Leu (L) Leucine	9.6%
Tyr (Y) Tyrosine	0.8%
Phe (F) Phenylalanine	4.4%
His (H) Histidine	0.9%
Lys (K) Lysine	3.9%
Trp (W) Tryptophan	0.0% *^{15}N labeled Trp was added to RTS500.
Arg (R) Arginine	2.9%
Cys (Cys) Cysteine	0.4%

A.16 Script for nmrPipe, 2D ^{1}H - ^{15}N SOFAST-HSQC processing

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

A.17 Script for nmrPipe, 2D ^{1}H - ^{13}C HMQC (written by Dr John Kirkpatrick, SMB, UCL)

```
# Basic 2D Phase-Sensitive Processing:
#     Cosine-Bells are used in both dimensions.
#     Use of "ZF -auto" doubles size, then rounds to power of 2.
#     Use of "FT -auto" chooses correct Transform mode.
#     Imaginaries are deleted with "-di" in each dimension.
#     Phase corrections should be inserted by hand.


#!/bin/csh

nmrPipe -in test.fid                                          \
#| nmrPipe -fn SOL                                             \
| nmrPipe -fn POLY -time                                       \
| nmrPipe    -fn SP -off 0.5 -pow 1 -c 0.5                   \
| nmrPipe    -fn ZF -zf 1 -auto                               \
| nmrPipe    -fn FT -auto                                      \
| nmrPipe    -fn PS -p0 2.1 -p1 8.2 -di                        \
| nmrPipe    -fn TP                                            \
| nmrPipe    -fn SP -off 0.5 -pow 2 -c 1.0                   \
| nmrPipe    -fn ZF -zf 1 -auto                               \
| nmrPipe    -fn FT -auto                                      \
| nmrPipe    -fn PS -p0 -96 -p1 180 -di                       \
| nmrPipe    -fn POLY -auto -ord 2                                  \
| nmrPipe    -fn TP                                            \
| nmrPipe    -fn POLY -auto -window 32                         \
      -verb -ov -out test_jk.ft2
```