Structural Studies on yeast eIF5A using Biomolecular NMR and Molecular Dynamics

A thesis submitted in fulfilment of the requirements for the

degree of

Master of Science (Chemistry)

of

Rhodes University

by

Lester Takunda Sigauke

B.Sc. (Hons.) (Rhodes University)

April 2015

ABSTRACT

Eukaryotic initiation factor 5A, eIF5A, is a ubiquitous eukaryotic protein that has been shown to influence the translation initiation of a specific subset of mRNAs. It is the only protein known to undergo hypusination in a two-step post translational modification process involving deoxyhypusine synthase (DHS) and deoxyhypusine hydroxylase (DOHH) enzymes. Hypusination has been shown to influence translation of HIV-1 and HTLV-1 nuclear export signals, while the involvement of active hypusinated eIF5A in induction of IRES mediated processes that initiate pro-apoptotic process have inspired studies into the manipulation of eIF5A in anti-cancer and anti-diabetic therapies. eIF5A oligomerisation in eukaryotic systems has been shown to be influenced by hypusination and the mechanism of dimerisation is RNA dependent. Nuclear magnetic resonance spectroscopy approaches were proposed to solve the structure of the hypusinated eIF5A in solution in order to understand the influence of hypusination on the monomeric arrangement which enhances dimerisation and activates the protein. Cleavage of the 18 kDa protein monomer by introduction of thrombin cleavage site within the flexible domain was thought to give rise to 10 kDa fragments accessible to a 600 MHz NMR spectrometer. Heteronuclear single quantum correlation experiments of the mutated isotopically labelled protein expressed in *E. coli* showed that the eIF5A protein with a thrombin cleavage insert, $eIF5A_{Thr}$, was unfolded. *In silico* investigations of the behaviour of eIF5A and $eIF5A_{Thr}$ models in solution using molecular dynamics showed that the mutated model had different solution dynamics to the native model. Chemical shift predictors were used to extract atomic resolution data of solution dynamics and the introduction of rigidity in the flexible loop region of eIF5A affected solution behaviour consistent with lack of *in vivo* function of $eIF5A_{Thr}$ in yeast. Residual dipolar coupling and T_1 relaxation times were calculated in anticipation of the extraction of experimental data from RDC and relaxation dispersion experiments based on HSQC measurable restraints.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and extend gratitude to the following:

- **My Supervisor Dr K. A. Lobb:**

Dr Lobb, you have been the constant and unchanging source of direction and inspiration for not only this project but also my process of development as a researcher and critical thinker. You have equipped me in ways I could only dream about and I am grateful for your insight, innovation and your guidance.

- **My Co-Supervisor Prof R. Dorrington:**

I am grateful for the opportunity that you gave me to explore my passions and curiosity within the Life Sciences. I have grown tremendously under your mentorship and the exposure that you gave me while I was a student in your lab. It is my hope that the work that we have produced will assist you in the ambitious and exciting projects you have in Lab 417.

- **Fellow labmates of 417 class of 2012:**

Specifically to Gareth, thanks for taking time out of your research to teach me about proteins and gels. I learnt a lot and I am confident that I will be able to build well upon the foundations that you laid. To Gwyneth, thank you for keeping me sane and for helping me to find my feet in a world that at the time was foreign to me. Jonathan and Tendai, thanks for the laughs and late night chats about how to change the world.

- **My family:**

Thank you for releasing me into the wilderness that is Grahamstown. Thank you for checking up on me, supporting me, laughing with me, crying with me and praying with me. I appreciate the support, encouragement and patience, albeit sometimes. I look forward to celebrating with you.

- **My Grahamstown family and friends:**

Thank you for creating a home from home for me and many countless others. Thank you for sharing this journey with me.

- **God:**

Without whom I would be an absolute failure, lacking perspective, self-seeking, insecure and purposeless. Thank you for inviting me into your story and allowing me the space to discover my academic and spiritual gifting. Thank you for encouraging me to lay down my ambition for you and for the service of people.

TABLE OF CONTENTS

ABSTRACT	**II**
ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
LIST OF FIGURES	VIII
LIST OF TABLES	X
LIST OF ABBREVIATIONS	XI

Chapter 1: Literature Review	**1**
1.1 INTRODUCTION	2
1.2 EIF5A	2
1.2.1 General Functions	4
1.2.1.1 Cell Functions	4
1.2.1.2 ***Disease functions***	6
1.2.2 eIF5A hypusination and Post-translation modification	10
1.2.3 eIF5A Structure	12
1.2.3.1 Structural studies on eIF5A	12
1.2.3.2 eIF5A mechanisms of oligomerisation	15
1.3 MOLECULAR BIOPHYSICS AND STRUCTURAL BIOLOGY	16
1.3.1 Nuclear Magnetic Resonance (NMR)	19
1.3.1.1 Principles of NMR	20
1.3.1.2 Applications of NMR	25

1.3.2 Homology Modelling	30
1.3.3 Molecular dynamics in structural biology	33
1.3.3.1 Structure calculations	33
1.3.3.2 NMR and molecular dynamics in probing disorder	34
1.4 KNOWLEDGE GAP AND AIMS OF THE PROJECT	35

Chapter 2: Production of recombinant eIF5A$_{Thr}$ for structural studies	**37**
2.1 INTRODUCTION	38
2.2 MATERIALS AND METHODS	39
2.2.1 Bacterial strains, recombinant plasmids and culture conditions	41
2.2.2 Protein expression and isotopic labelling	42
2.2.3 Purification of his-tagged eIF5A$_{Thr}$	42
2.2.4 Thrombin cleavage	43
2.2.5 NMR analysis	43
2.3 RESULTS AND DISCUSSION	45
2.3.1 pJC6 construct validation	45
2.3.2 Protein Expression	47
2.3.3 Thrombin cleavage	52
2.3.4 NMR Analysis	53
2.4 CONCLUSION	56

Chapter 3: Homology Modelling and Molecular Dynamics	**57**
3.1 INTRODUCTION	58
3.2 MATERIALS AND METHODS	60
3.2.1 Homology modelling	60
3.2.2 Molecular Dynamics	64
3.3 RESULTS AND DISCUSSION	66
3.3.1 Homology Modelling	66
3.3.2 Molecular Dynamics	71
3.4 CONCLUSION	79
Chapter 4: Molecular Dynamics as a means to study solution Dynamics	**81**
4.1 INTRODUCTION	82
4.2 MATERIALS AND METHODS	88
4.2.1 Predicted Chemical Shifts and synthetic spectra	88
4.2.2 Back-calculation of Residual Dipolar Coupling	90
4.2.3 T_1 relaxation and τ_c from MD simulations	91
4.3 RESULTS AND DISCUSSION	92
4.3.1 Chemical Shift Prediction	92
4.3.1.1 Steered Molecular Dynamics	99
4.3.2 Residual Dipolar Coupling	102
4.3.3 T_1 relaxation and τ_c from MD simulations	111
4.4 CONCLUSION	119
Chapter 5: Concluding Remarks	117
References	119

Appendices	**140**
APPENDIX A: GROWTH MEDIA	141
APPENDIX B: BUFFERS	142
APPENDIX C: SDS-PAGE	143
APPENDIX D: NMR PULSE PROGRAMS	145

LIST OF FIGURES

Label	Description	Page no.
Figure 1-1:	Mechanism of action of deoxyhypusine synthase on eIF5A-Lys.	11
Figure 1-2:	Illustrations of yeast eIF5A chain A crystal structure (PDB ID: 3ER0).	14
Figure 2-1:	Advances in NMR spectroscopy and isotopic labelling in extending the size limit of biomacromolecular solution NMR (Yu, 1999).	38
Figure 2-2:	Validation of expression plasmid by restriction analysis	46
Figure 2-3:	Analysis of heterologous protein expression of eIF5AThr from transformed *E. coli* cells using 15% SDS-PAGE and staining with coomassie.	47
Figure 2-4:	Overexpression of $eIF5A_{Thr}$ grown in ^{15}N- labelled minimal media analysed on 15% SDS PAGE.	48
Figure 2-5:	Purification of $eIF5A_{Thr}$ using Nickel affinity chromatography.	50
Figure 2-6:	UV trace of purification profile	51
Figure 2-7:	SDS-PAGE analysis of the products of thrombin cleavage of 6xhis-eIF5Athr for the cleavage reaction.	52
Figure 2-8:	WATERGATE ^{1}H water suppression at o1p of 4.701 ppm.	53
Figure 2-9:	^{1}H-^{15}N HSQC spectrum of $eIF5A_{Thr}$ in 10% D_2O obtained from a 600 MHz spectrometer.	54
Figure 3-1:	Block diagram describing the protocol for MD simulations using GROMACS 4.5.5.	64
Figure 3-2:	Target to template Sequence alignment of 3ER0 chain A and B with PG57.	68
Figure 3-3:	Potential energy of position restrained MD.	72
Figure 3-4:	Root Mean Squared Deviation 1ns plot.	75
Figure 3-5:	Radius of gyration from 1ns of molecular dynamic simulations.	76
Figure 3-6:	Structures of unfolded PG57 showing the development of secondary structure after 10 ps, 100 ps and 1 ns dynamics.	78
Figure 3-7:	Ramachandran plot of models during 1ns of molecular dynamics simulations.	79
Figure 4-1:	Block diagram for the generation of synthetic spectra from MD trajectories.	89
Figure 4-2:	Hα Chemical Shift standard deviations for folded and unfolded eIF5AThr and	93

	folded eIF5A after 1 ns dynamics.	
Figure 4-3:	Model of $eIF5A_{Thr}$ showing domains of deviation.	95
Figure 4-4:	Overlaid ^{15}N-edited HSQC synthetic spectra.	106
Figure 4-5:	Overlaid spectra for time averaged $eIF5A_{Thr}$ models and experimental data (purple).	107
Figure 4-6:	$eIF5A_{Thr}$ steered molecular dynamics conformations.	100
Figure 4-7:	Steered Molecular Dynamics Analysis.	100
Figure 4-8:	Overlaid ^{15}N-HSQC spectra.	101
Figure 4-9:	Predicted residue RDC standard deviation.	103
Figure 4-10:	Residue 43, β domain, ^{1}H -^{15}N RDC over 1 ns MD simulation.	104
Figure 4-11:	Residue 66, γ domain, ^{1}H-^{15}N RDC over 1 ns MD simulation	106
Figure 4-12:	Residue 77, ^{1}H ^{15}N RDC over 1 ns MD simulation.	107
Figure 4-13:	Residue 93, S region, ^{1}H-^{15}N RDC over 1 ns MD simulation.	108
Figure 4-14:	Non-decoupled ^{15}N HSQC cross peak pairs of *eIF5A* folded conformation over 1 ns, in anisotropic solution.	110
Figure 4-15:	Non-decoupled ^{15}N HSQC cross peak pairs of $eIF5A_{Thr}$ folded conformation over 1 ns, in anisotropic solution.	111
Figure 4-16:	Residue specific predicted spin-lattice relaxation and correlation times of native eIF5A folded model simulations.	112
Figure 4-17:	Residue specific predicted spin lattice relaxation and correlation times of mutant $eIF5A_{Thr}$ model simulations.	112
Figure 4-18:	Residue specific spin-lattice relaxation and correlation times of folded $eIF5A_{Thr}$.	113
Figure D-1:	Proton spectra of $eI5A_{Thr}$ using p3919gp WATERGATE suppression.	145

LIST OF TABLES

Label	Description	Page no.
Table 1-1:	Crystal structures of eIF5A homologues available in PDB.	13
Table 1-2:	Magnetic properties of spin ½ nuclei relevant for biomacromolecular NMR (James, 1998a).	25
Table 2-1:	Genotypes of E. coli strains and a description of plasmid DNA used in this study.	40
Table 3-1:	Protein structures retrieved from query search using PG57 sequence ranked according to sequence identity with PG57.	67
Table 3-2:	Evaluation table for the best three models of PG57 generated from MODELLER model build.	69
Table 3-3:	Analysis table of ProSA quality assessment for PG57 unfolded, PG57 folded and 3ER0 folded conformation.	70
Table 3-4:	Analysis of structural integrity during MD assessed by the ProSA Z-score calculation and PROCHECK Ramachandran plots.	73
Table 3-5:	Residue specific model quality generated from the ProSA quality assessment calculation.	74

LIST OF ABBREVIATIONS

Abbreviation	Expansion
3D	3 dimensional
ATP	adenosine triphosphate
B	corrected magnetic field
B_o	field of strength
Cdc25c	cell division cycle 25 homolog c gene
Cdk1	cyclin B-bound CDC2
CHARMM-22	Chemistry at Harvard Macromolecular Mechanics 22
COSY	correlation spectroscopy
DDX3	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3
DFMO	DL-alpha-difluoromethylornithine
DG	distance geometry
DHS	deoxyhypusine synthase
DNA	deoxyribonucleic acid
DOHH	deoxyhypusine hydroxylase
DOHH	DOHH gene
eEF2	translation elongation factor 2
eIF1A	17 kDa eukaryotic initiation factor 1A
eIF2	eukaryotic initiation factor 2
eIF3	eukaryotic translation initiation factor 3
eIF4D	eukaryotic translation initiation factor 4D
eIF4E	eukaryotic translation initiation factor 4E
eIF5A	eukaryotic translation initiation factor 5A
$eIF5A_{Thr}$	eIF5A containing the thrombin cleavage site
G1/S	growth 1/synthesis
G2	Gap 2 phase
GHz	Giga hertz
GST	glutathione S-transferase
GTP	guanosine-5'-triphosphate
HEAT-repeats	huntingtin elongation factor 3E a subunit of protein phosphatase 2A and the target of rapamycin repeat
His	Histidine
HIV-1	human immunodeficiency virus type 1
hRIP	human rev interacting protein
HSQC	heteronuclear single quantum correlation
HTLV-1	human T-cell leukemia virus type 1
ID	identity
IF-M2B	Initiation factor M2B
IRES	internal ribosomal entry site
J	scalar coupling
kDa	kilo Dalton

LB	Luria Bertani
MD	molecular dynamics
Met-tRNA	methionyl-tRNA
Met-tRNAi	Met-tRNA complexed with eIF2
mGTP	2'(3')-O-(N-methylanthraniloyl)guanosine 5'-triphosphate
MHz	Mega hertz
mRNA	messenger RNA
N	population distribution
NAD	nicotinamide adenine dinucleotide
NES	nuclear export signals
NMR	Nuclear magnetic resonance
NO	nitric oxide
NOESY	nuclear Overhauser spectroscopy
NOS2	nitric oxide synthase gene
nt	nucleotide
OB-fold	oligonucleotide binding fold
p53	tumor suppresor p53
PCS	pseudocontact shifts
PDB	protein databank
PDGF2	platelet derived growth factor subunit 2
poly(A)-tail	poly adenosine tail
RDC	residual dipolar coupling
Rev	regulator of virion mRNA in HIV
Rex	regulator of virion mRNA in HTLV-1
rMD	restrained molecular dynamics
RMSD	root mean squared deviation
RNA	ribonucleic acid
Rnase A	ribonuclease A
$\hat{S}$	Spin Quantum Number
Sam68	Sarcoma associated in mitosis of 68 kDa
SH3 domain	Sarcoma homology 3 domain
siRNA	small interfering RNA
S-phase	synthesis phase
T_1	Longitudinal recovery of the magnetization along the Z axis
T_2	Relaxation in the transverse x-y plane
T-cell	thymus white blood cells
TNF-α	tumor necrosis factor alpha
TOCSY	total correlation spectroscopy
tRNA	transfer RNA
tTGase	tissue transglutaminase
UTR	3'-untranslated region
UV	Ultraviolet
Xpo1	nuclear export receptor

γ	nuclei-specific gyromagnetic responsiveness factor
σ	chemical shielding factor due to the atomic electronic oscillations
ω_o	Larmor frequency of precession of the nucleus around the axis of the magnetic field
OD_{600nm}	optical density at 600 nm
nm	nano meter
rpm	revolutions per minute
IPTG	isopropyl β-D-1-thiogalactopyranoside
SDS	sodium dodecyl sulphate
Tris-Cl	tris(hydroxylmethyl)aminomethane chloride
SDS-PAGE	sodium dodecyl sulphate poly acrylamide gel electrophoresis
FF IMAC	fast flow immobilised metal affinity chromatography
FPLC	fast performance liquid chromatography
CaCl2	calcium chloride
NaCl	sodium chloride
FID	free induction decay
LVPRGS	leucine valine proline arginine glycine and serine
bp	base pairs
kB	kilo bases
FT	flow through
DTT	dithiothreitol
ppm	parts per million
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular simulation computer program package
QM	quantum mechanics
MM	molecular mechanics
BLAST	basic local alignment tools
RCSB	The Research Colloboratory for Structural Bioinformatics
DOPE	Discrete Optimised Potential Energy
DOPE-Z	normalised DOPE score
LINCS	Linear Constraint Solver
DSC	Discrimination of protein secondary structure class
EM	energy minimisation
ns	nano second
ps	pico second
τ_c	Rotational reorientation
D	dipolar coupling
D_{max}	maximum dipolar coupling
CPMG	Carr, Purcell, Meiboom and Gill
PME	Particle-Mesh Ewald
Rg	radius of gyration

Chapter 1: Literature Review

1.1 INTRODUCTION

1.2 EIF5A

1.2.1 General Functions

1.2.1.1 Cell Functions

1.2.1.2 Disease functions

1.2.2 eIF5A hypusination and Post-translation modification

1.2.3 eIF5A Structure

1.2.3.1 Structural studies on eIF5A

1.2.3.2 eIF5A mechanisms of oligomerisation

1.3 MOLECULAR BIOPHYSICS AND STRUCTURAL BIOLOGY

1.3.1 Nuclear Magnetic Resonance (NMR)

1.3.1.1 Principles of NMR

1.3.1.2 Applications of NMR

1.3.2 Homology Modelling

1.3.3 Molecular dynamics in structural biology

1.3.3.1 Structure calculations

1.3.3.2 NMR and molecular dynamics in probing disorder

1.4 KNOWLEDGE GAP AND AIMS OF THE PROJECT

1.1 INTRODUCTION

A strategic research infrastructure collaboration between the South African Department of Science and Technology and the National Research Foundation of South Africa enabled the commissioning of a Bruker 600 MHz Nuclear Magnetic Resonance spectrometer with protein capabilities at Rhodes University. Since the commissioning of the instrument in 2006 the spectrometer has been used primarily for routine small molecule analyses and informing ligand-receptor interactions through saturation transfer. The room temperature, triple resonance Bruker microprobe allows us to extend the applications of our spectrometer to develop the capacity for protein analysis and structure determination.

The Dorrington Research Group within the Department of Microbiology and Biochemistry at Rhodes has been involved in structure-function studies on *Saccharomyces cerevisiae* (yeast) eIF5A in order to elucidate its role in regulating the cell cycle (Gentz et al., 2009). This 21 kDa protein falls within the accessible NMR size limit and offers an opportunity for protein analysis using a 600 MHz NMR spectrometer at Rhodes University.

1.2 eIF5A

Eukaryotic translation initiation factor 5A (eIF5A) is a ubiquitous translation initiation factor that is highly conserved in eukaryotes, from yeast to mammalian cells with homologues observed in archaea (Bommer et al., 1991; Mémin et al., 2014; Park, 1987). eIF5A is the only protein known to contain the unique hypusine [N^{ϵ}-(4-amino-2-hydroxybutyl)-lysine] amino acid (Paulo et al., 2014). This amino acid is formed during post-translational modification, which occurs through the sequential transformation of a conserved lysine residue with a spermidine precursor in the presence of deoxyhypusine synthase (DHS) and deoxyhypusine synthase (DOHH) enzymes (Kang et al., 1993; Park, 2008).

eIF5A has gone through several nomenclature changes since it was first discovered in 1976 initially being identified as IF-M2B and later eIF4D before being named as eIF5A (Cooper et al., 1983; Kemper et al., 1976; Park et al., 1996). eIF5A was initially isolated from rabbit reticulocyte cells as IF-M2B with an α and β variant. It was thought to be a ribosomal protein due to its small size, lack of independent function and loose association to the 40 S and 60 S Ribosomal subunits (Kemper et al., 1976). Comparison of IF-M2B's physical features with

those of other ribosomal proteins disproved this hypothesis and it was suggested that IF-M2B may be a translation initiation factor instead (Kemper et al., 1976). Due to its *in vitro* stimulation of methionyl-puromycin peptide bond synthesis, IF-M2B, later known as eIF-4D was thought to be involved in the initiation of protein synthesis (Gordon et al., 1987; Kang & Hershey, 1994; Park et al., 1996). Early evidence from subcellular distribution studies, its influence of protein biosynthetic rate and structural conservation across phylogenetic genre confirmed eIF5A's role in the cytoplasmic interactions of ribosomes with signalled mRNA fragments (Benne & Hershey, 1978; Bommer et al., 1991; Cooper et al., 1983; Pain, 1996).

When Kang & Hershey (1994) performed an important series of experiments related to eIF5A depletion in yeast, *S. cerevisiae*, they found that only 70% of total protein synthesis was achieved in eIF5A depleted cells. They concluded that eIF5A was not a factor essential for general protein synthesis but essential for translation initiation of a selected few mRNAs. Depletion of eIF5A in yeast resulted in the reversible arrest of cells in the G1-S phase of the cell cycle (Kang & Hershey, 1994; Park et al., 2010). Because the S-phase consists of DNA replication it was suggested that activated eIF5A was necessary for its initiation either directly or by affecting the translation of proteins that are involved in the initiation of DNA replication (Park et al., 2010). An indirect eIF5A role is plausible because hypusinated eIF5A is shown to be pumped out of the nucleus of the cell in response to interaction with nuclear export signal exportin 4 in higher eukaryotes (S. B. Lee et al., 2009; Lipowsky et al., 2000).

In mammalian cells eIF5A depletion resulted in a 5% decrease in total protein synthesis. It was thought that eIF5A was essential for protein synthesis in stress conditions through its direct interaction with phosphorylated eEF2, although the exact mechanism of its influence on initiation in these cells is undefined (Li et al., 2010). Inhibition of eIF5A hypusination in yeast cells helped to identify genes whose expression is hypusine dependent including methionine adenosyltransferase and cytochrome-c oxidase, enzymes necessary for cell proliferation (Hanauske-Abel et al., 1995). This study provided evidence for the indirect role of eIF5A in cell proliferation and proposed a selective influence on mRNA turnover and decay (Zuk & Jacobson, 1998).

These observations and the fact that eIF5A, unlike other initiation factors, is present in the nucleus at hyperstoichiometrical levels (Chattopadhyay et al., 2008) and is not necessary in

the formation of translation initiation complexes suggests that its influence on translation is not through the cap-dependant mechanism that dominates eukaryotic protein translation initiation (Lipowsky et al., 2000). Other mitigating evidence of an eIF5A cap-dependent influence on translation initiation arise from the fact that unlike most initiation factors, the majority of eIF5A interacts only transiently with ribosomes (Li et al., 2004; Thomas et al., 1979). Although eIF5A has been shown to be able to stimulate the synthesis of methionyl-puromycin synthesis *in vitro* (Myung Hee Park et al., 1996), it is now accepted that it is a translation factor of a specific subset of mRNA's (Kaiser, 2012; Lee et al., 2010; Zanelli et al., 2006). Although it has been 38 years since its first discovery, the fundamental function of eIF5A and its role in eukaryotic translation through modulation of ribosome activity is still unclear.

1.2.1 General Functions

1.2.1.1 Cell Functions

Eukaryotic cells undergo a series of sequential events as part of their growth and differentiation cycles. The cell cycle is divided into four distinct phases and begins with the cell in the G1 phase (Growth 1) which is the main growth phase and where synthesis of cytoplasmic factors necessary for DNA synthesis occurs. The next phase is the S-phase (Synthesis) where DNA synthesis occurs and is followed by the second growth interval (G2) which prepares the cell for mitotic division (Schnier et al., 1991). At every stage of the cycle there are certain checks and triggers, such as the Cdc25C adaptor protein, that need to be switched on and off depending on the internal or external state of the cell (Reinhardt & Yaffe, 2009). In mammalian cells it was found that inhibition of eIF5A activation results in reversible arrest of cells in the G1-S phase prior to crossing the restriction point of the cell cycle (Myung Hee Park, 1987; Voet & Voet, 2004). It was suggested that activated eIF5A was necessary for the initiation of DNA replication either directly or by affecting the translation of proteins that are involved in the initiation of DNA replication (Kang & Hershey, 1994). Inhibition of eIF5A hypusination results in the disappearance of mRNA species that code for methionine adenosyltransferase and cytochrome-c oxidase, enzymes necessary for cell proliferation, and reappears when hypusination is activated, providing evidence for the indirect role of eIF5A on cell proliferation and proposed a direct influence on mRNA turnover and decay (Hanauske-Abel et al., 1995; Kim, Hung, Yokota, Kim, & Kim, 1998). An

indirect role is plausible because hypusinated eIF5A is shown to be pumped out of the nucleus of the cell in response to exportin 4, thus it is plausible that hypusinated eIF5A plays an direct role in the translocation of specific mRNA species out of the nucleus (Lee et al., 2009; Lipowsky et al., 2000).

In yeast, growth was only slightly impaired by disruptions of the eIF5A gene (J. H. Park et al., 2006). An E66K mutation in the *Schizosaccharomyces pombe* DOOH, which inhibits hypusination, results in temperature sensitive growth resulting in abnormal distribution and morphology of mitochondria as a result of the absence of hypusinated eIF5A (Weir & Yaffle, 2004). This suggested a potential role for hypusinated eIF5A in the translation of proteins that were involved in protecting the cells from hyperthermia induced stress in *S. pombe* (Gosslau et al., 2009).

Because of the diversity of eIF5A cellular functions observed it was suggested that these occurred because of its influence on the translation of a few specific mRNAs (Caraglia et al., 2013; Rossi et al., 2013). It was proposed that rather than being involved in cap-dependent translation initiation, eIF5A was involved in the translation of specific mRNA sequences that possess internal ribosome entry sites (IRES) within the 5' UTR regions of their mRNA instead of a 5' polyamine cap (Fernández-Miragall, López de Quinto, & Martínez-Salas, 2009; H. A. Kang & Hershey, 1994). eIF5A's role in translation via an alternative mechanism was validated by the fact that p53 expression, a well known apoptotic inducer shown to be translated in an IRES dependant manner, was controlled by the expression and presence of eIF5A (Li et al., 2004; Taylor et al., 2007). Platelet derived growth factors are regulated by IRES mediated translation initiation (Bernstein, Sella, & Le, 1997). When hypusinated eIF5A was shown to be essential for translation of proteins necessary in early mouse embryonic development it was thought that this control was due to the regulation of eIF5A on PDGF2s which are regulated by IRES mediated translation initiation (Charlton, 2012; Nishimura, Lee, Park, & Park, 2012).

1.2.1.2 Disease functions

HIV

Hypusinated eIF5A has been shown to influence the replication of human immunodeficiency virus type 1 (HIV-1) and human T-cell leukemia virus type 1 (HTLV-1) by interacting with RNA transport factors, Rev (regulator of virion RNA) in HIV-1 and Rex in HTLV-1 (Bevec & Hauber, 1997; Katahira et al., 1995). These proteins were first found to bind to specific response elements present in the viral RNA enhancing their translocation out of the nucleus into the cytosol of the host (Ruhl et al., 1993). In order for efficient translocation to occur nuclear export receptor, Xpo1, interacts with nuclear export signals (NES) from Rev or Rex, in a leucine dependent manner (Bevec & Hauber, 1997). It has been shown that interaction of these NES's with Xpo1 is dependant on the activation of Rev or Rex by sequence-specific binding and interaction with hypusinated eIF5A (Kaiser, 2012; Zanelli & Valentini, 2007).

Another function of eIF5A, which may provide a hint towards its manipulation as a drug target for HIV therapies, relates to a study performed by Liu et al., (2011) which showed its ability to enhance viral RNA translation. Through ectopic expression, siRNA knockdown and transcomplementation assays, eIF5A along with other cofactors such as Sam68, hRIP and DDX3, were described as being essential co-factors in HIV-1 RNA translation. When small interfering RNA (siRNA) for eIF5A was introduced, expression was inhibited suggesting a specific role for eIF5A in the IRES mediated translation initiation of HIV-1 RNA. An IRES based bicistronic reporter gene assay was used to validate the role of eIF5A in cap-independent translation of HIV-1 proteins (Liu et al., 2011).

When ciclopirox, a topical fungicide, and deferiprone, a metal iron chelator, were used as DOHH inhibitors, formation of mature eIF5A was inhibited successfully. It was seen that at the level of HIV-1 translation initiation, viral gene expression was inhibited in human periferal blood mononuclear cells that were cloned for HIV-1 expression (Hoque et al., 2009). Use of DHS inhibitor guanylhydrazone or DHS RNAi treatment in a series of multi-drug resistant viruses resulted in significant inhibition of viral replication highlighting the role of eIF5A in the metabolism of essential viral mRNAs (Kaiser, 2012). These results highlight the potential for hypusine inhibitors as potential analogues for efficient anti-HIV-1 drug therapies that are less prone to developing drug resistance (Liu et al., 2011).

Cancer

eIF5A has been shown to be involved in the induction of cell growth and regulation of the cell-cycle (Caraglia et al., 2001; Kang & Hershey, 1994). Indirectly it has been shown to influence apoptotic processes and more recently carcinogenesis (Guan et al., 2004; Sheng, et al., 1998). In humans most cancer therapies attempt to induce targeted cell death or inhibit growth through restricted cellular proliferation (Caraglia et al., 2001). Because of its link to these strategies targeted for cancer therapies, eIF5A is being studied quite extensively as a potential anti-neoplastic agent with significant attempts made at preventing its activation through hypusination regulated by polyamine synthesis (Kerscher et al., 2010).

Successful studies that investigated the potential to regulate cell proliferation by targeting eIF5A included the successful use of a DHS inhibitor, D, L-α-difluoromethylornithine (DFMO), in Chinese hamster ovary cells and 2-(4-hydroxy-toluene-3-yl)-4,5-dihydro-5-carboxythiazole (Carboxythiazole) in human T-lymphocytes (Lalande, 1990; Park, 1987). These compounds were shown to result in reversible cell cycle arrest in the G1-S boundary or late G1 phase consistent with inhibition of DOHH activity (Park, 1987).

Tissue transglutaminase (tTGase) has been shown to inhibit cell proliferation and induce pro-apoptotic events in transfected cells (Caraglia et al., 2001). tTGase catalyzes post-translational modifications of proteins through the formation of covalent crosslinks between γ–carboxamido groups on glutamine and ε-amino groups on lysine endoresidues. Its mechanism of action has been well documented but however its precise role in induction of apoptosis has proven difficult to resolve (Yamaguchi & Wang, 2006). It has been shown that tTGase can interfere with polyamine synthesis and hypusine levels in cells by forming γ-glutamyl polyamine derivatives with hypusinated eIF5A. This reduction in levels of active eIF5A coincided with an increase in apoptotic processes and a 50% decrease in cell proliferation. This consequence proposes a role for active eIF5A in preventing apoptosis in cancer cells (Caraglia et al., 2001). When interferon-α and cytosine arabinoside were investigated for their potential use in cancer treatment, it was found that growth inhibition and apoptosis induction paralleled with reduced hypusine levels suggesting that these drugs were acting on eIF5A as their target (Caraglia et al., 1999).

A potential indication of how eIF5A regulates apoptosis can be seen in the necessity of eIF5A for the expression and activity of p53 in apoptosis induced cells (Taylor et al., 2007). p53 has been identified as a major component of the DNA damage response pathway and can protect cells from DNA damage by activating damage repair mechanisms, arresting cell growth and initiating apoptosis (Rahman-Roblick et al., 2007). In 50% of human cancers, p53 has been found to be defective suggesting a critical role in the regulation of tumour suppression (Bourdon, 2007). When apoptosis was induced in cells using the apoptosis inducer, Actinomycin D, overexpression of both eIF5A and p53 were found. However when apoptosis was induced in the presence of eIF5A specific small interfering RNA suppressing the expression of eIF5A, apoptosis was hindered with no accumulation of p53 observed (Taylor et al., 2007). It is thought that eIF5A participation in p53 dependant apoptosis is regulated independently of hypusination because hypusine inhibitors have little effect on apoptosis. Therefore it is plausible that unhypusinated eIF5A had an increased interaction with p53. The interaction of eIF5A with p53 was shown to be regulated by an eIF5A binding protein syntenin which did not bind in the presence of hypusine (Li et al., 2004). Whether eIF5A has a role in other p53 tumour suppressive roles is still to be investigated conclusively.

The mechanism of action of eIF5A is complicated by evidence for its role in cell proliferation as well as in cell death (Li et al., 2004). It is plausible from the experiments done previously that its influence on cell proliferation is as a result of its involvement in the translation of specific mRNAs that are controlled by the hypusinated form while its pro-apoptotic roles arise due to the activity of the p53 interacting with the unhypusinated state (Caraglia et al., 2001; De Benedetti & Harris, 1999; Nishimura et al., 2005). Rapid translocation of eIF5A to the nucleus from the cytoplasm in the event of induction of apoptosis from tumour necrosis factor-α confirms a directed role for the unhypusinated protein in the nucleus (Taylor et al., 2007). It is still unclear precisely how p53 and eIF5A interact in the nucleus to give rise to apoptosis however strong evidence suggests that unhypusinated eIF5A as the eIF5A-1 isoform forms a complex with syntenin (Caraglia et al., 2013).

Diabetes

eIF5A hypusination has been identified as a critical regulator of cytokine signalling during the inflammatory response in pancreatic islet β-cells (Maier et al., 2010). The systemic or

local release of pro-inflammatory cytokines has been attributed to the dysfunction and destruction of islet β-cells which are responsible for the production of insulin (Hauber, 2010). In type 1 and type 2 diabetes, it has been shown that induction of necrotic and pro-apoptotic pathways from inflammatory responses, through either an autoimmune defect or insulin resistance due to adiposity, cause islet β-cell destruction (Hauber, 2010; Maier, Ogihara, et al., 2010). A role of eIF5A in the translation of a pro-inflammatory tumour necrosis factor (TNF-α) by the regulation of its translation without affecting its mRNA levels further highlights a role for eIF5A in proinflammatory processes (Paulo et al., 2014).

In mouse models with lipopolysaccharide-induced sepsis, treatment with eIF5A-specific siRNA was shown to improve survival rates while treatment with control siRNA showed significantly higher levels of cytokine expression and higher fatality rates (Moore et al., 2008). This suggested an eIF5A role in the regulation of pro-inflammatory pathways. The mechanism of resistance of inflammation was described as being similar to that experienced in HIV-1 infection, where hypusinated eIF5A is shown to enhance the translation and translocation of pathogenic mRNA through IRES mediated translation mechanisms (Maier, et al., 2010). It is envisaged that the hypusinated eIF5A is essential for the translation of pro-inflammatory proteins and by reducing their translation, their effects are minimized.

When streptozotocin was used as a well-known β-cell destructive agent, it was shown that both hypusination inhibition and mRNA silencing of eIF5A protected the mice from insulin resistance and islet β-cell destruction (Maier et al., 2010). It has since been shown that eIF5A interacts with the *NOS2* mRNA in a hypusine dependant manner shuttling it out of the nucleus and permitting its translation and elongation. The expression of inducible nitric oxide synthases catalyses the production of chemical signalling NO which is a well-known pro-inflammatory signal which eventually permanently destroys the islet cells (Maier et al., 2010). Thus it is plausible to conclude that the DHS/eIF5A pathway presents a potential drug target in inflammation induced diseases such as diabetes, through its hypusine dependent regulation of inflammation induced pro-apoptotic processes through IRES mediated translation initiation (Hauber, 2010).

1.2.2 eIF5A hypusination and post-translation modification

As discussed above most eIF5A functions are hypusine dependent and hypusination modification occurs on the lysine residue located in the N-terminal loop between strands β3 and β4 (Figure 1-2B; Tong et al., 2009). Hypusination is a stepwise transformation involving the aforementioned 2 enzymes, DHS and DOHH (Cooper et al., 1983). In yeast *S. cerevisiae*, a disruption of the *DHS* gene proved to be lethal while a *DOHH* null strain (missing the gene encoding DOHH) mutant grew at a slower rate than its wild type strain (Nishimura et al., 2012). The influence of DHS on eIF5A activity was also shown through the effects of a known DHS inhibitor such as guanydylhydrazone. This restricted the cells in their non-degenerative stages of the cell cycles, consistent with the absence of active eIF5A which resulted in enlarged cells (Kang et al., 2007; Nishimura et al., 2012).

DHS is responsible for the catalysis of the first step of eIF5A activation, transforming a lysine residue to deoxyhypusine using spermidine as a substrate (Figure 1-2). Dehydrospermidine, formed from the NAD-dependent dehydrogenation of spermidine, is the precursor for the transfer of a 4-aminobutyl moiety onto the ε-amino group of Lys 329 of active DHS (Nishimura et al., 2012; Ruhl et al., 1993). DHS catalyzes the cleavage and transfer of this precursor to the eIF5A lysine forming a covalent enzyme-imine on the conserved lysine (Cano et al., 2008; Park et al., 1996). Enzyme-bound NADH reduces this intermediate imine group to deoxyhypusine (Clement et al., 2003; Nishimura et al., 2012). DHS exists as a homotetramer with 4 active sites from 4 identical 43 kDa monomers of 329 amino acids each (Kang, et al., 2007). DHS exhibits absolute specificity towards the eIF5A substrate while a narrow preference for spermidine analogues as substrate donors is shown (Park, 2008).

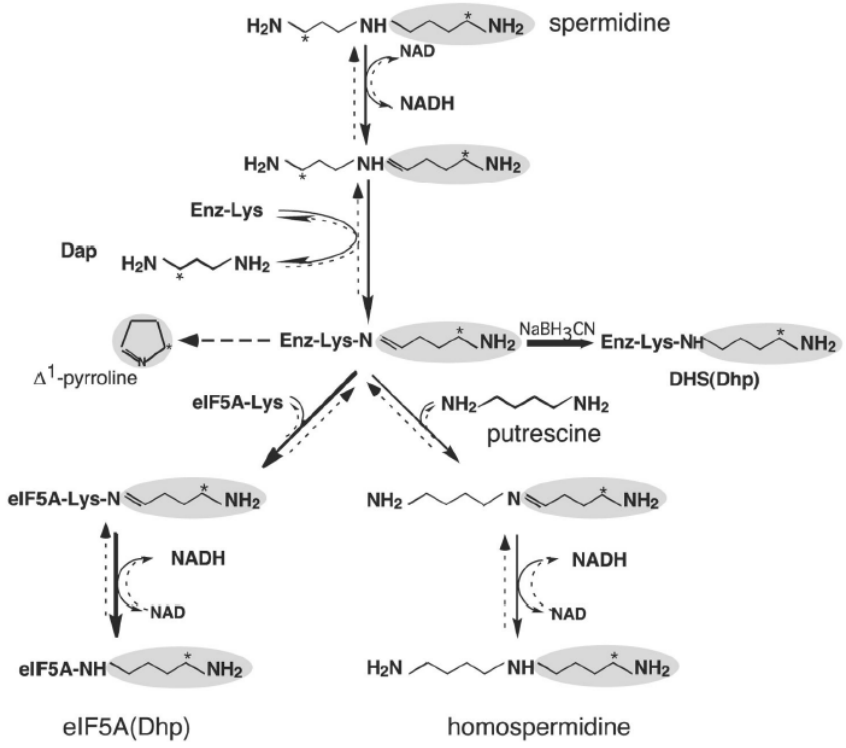


Figure 1-1: Mechanism of action of deoxyhypusine synthase on eIF5A-Lys. Spermidine is unsaturated and nucleophilic substitution of diaminopropane with basic lysine transfers labelled diaminobut-1-ene onto DHS. An equilibrium that favours nucleophilic substitution of diaminobut-1-ene onto eIF5A-Lys is enhanced by the fast NADH mediated reduction to form eIF5A-deoxyhypusine. Side reactions involve formation of homospermidine in the presence of putrescine while pyrroline can form at high temperatures. (Schematic diagram obtained from M. H. Park, (2008)).

Conversion of the deoxyhypusine moiety to hypusine occurs when DOHH catalyzes the nucleophilic hydroxylation of the 2' carbon of deoxyhypusine to yield hypusine (Tong et al., 2009). Homology models guided by sequence profiles and circular dichroism identified DOHH as being a superhelical structure consisting of 8 catalytic HEAT-repeats. These flexible scaffolds formed symmetrical Fe^{2+}-containing dyads of 4 HEAT motifs which could wrap around the incoming eIF5A and enhance its oxidation (Kang et al., 2007; Neuwald, 2000). DOHH activity is sensitive to iron chelators and molecular oxygen, showing that its chemical mechanism of action to be similar to that of metal-dependant double stranded β-helix enzymes despite having significant structural differences (Park et al., 2006).

In addition to hypusination, eIF5A in eukaryotic cells undergoes a series of post translational modification reactions involving phosphorylation and acetylation. A single serine residue in the N-terminal region of yeast eIF5A has been shown to undergo post-translational phosphorylation (Kang et al., 1993), which influences apoptotic induction in rapidly dividing yeast cells (Li et al., 2004). The yeast eIF5A Lys 47 was identified as a critical acetylation residue and a series of mutational analyses confirmed that a basic charge at the 47 position was critical for eIF5A hypusination (Cano et al., 2008) by enhancing protein-protein and protein-nucleic acid interactions (Choudhary et al., 2009; Lee et al., 2009).

1.2.3 eIF5A Structure

eIF5A has been described as having both nucleic acid as well as protein affinity based on its polarized charge distribution and secondary structure arrangements (Kim et al., 1998; Peat, Newman, Waldo, Berendzen, & Terwilliger, 1998; Tong et al., 2009). Xu and Chen (2001) performed systemic evolution of ligands using the exponential enrichment technique (SELEX) to identify specific mRNA target sequences that were recognised by eIF5A (Xu & Chen, 2001). They showed that although most RNA binding proteins target specific structural elements, eIF5A showed sequence specific binding at RNA binding sites (Zanelli & Valentini, 2007). A better understanding of the structure of eIF5A and the influence of hypusination on its conformational changes will provide significant insight into its function and interactions with ribosomes and other binding partners. This will better inform understanding of its cellular function and assist in the development drug therapies that exploit its behaviour (Charlton, 2012; Kaiser, 2012; J. H. Park et al., 2011).

1.2.3.1 Structural studies on eIF5A

Most of the structures of eIF5A and its homologues deposited in protein databanks are derived from crystals analysed by X-ray diffraction. The protein used for these structural studies is unhypusinated because the *E. coli* expression systems used to produce sufficient quantities of protein lack the enzymes essential for hypusination of eIF5A. Nevertheless these unhypusinated structures have given valuable information about the monomeric arrangement of the protein in its unhypusinated state (Kim et al., 1998). Recent developments have allowed hypusinated eIF5A to be produced using recombinant techniques in *E. coli* by the development of a polycistronic vector that allows the co-

expression of active eIF5A with DOHH and DHS human coding sequences (J. H. Park et al., 2011). By ensuring that the three proteins contained unique and compatible origins of replication after induction, the proteins produced using the modular polycistronic vector were shown to be active and successfully hypusinated. These developments provide significant advancement in the quest towards solving the structure of hypusinated active eIF5A. Strong sequence homology is shared between homologues of the eIF5A gene within Eukarya and Archaea (Kyrpides & Woese, 1998; J. H. Park et al., 2011) while elongation factor P has been identified as a bacterial analogue with numerous structural similarities to eukaryotic IF5A (J.H. Park et al., 2012; Rossi et al., 2014). Human eIF5A genes have relatively minor differences when compared to the yeast genes and human eIF5A-1 has been shown to be able to support growth in yeast eIF5A null strains (Cano et al., 2008). The HEX-1 *Neurospora crassa* orthologue shares 23% sequence homology with yeast eIF5A. Despite the low sequence similarity and functional divergence these analogues share significant monomeric structure similarity (P. Yuan et al., 2003).

Table 1- 1: Crystal structures of eIF5A homologues available in PDB

Species	Gene ID	ID	Resolution
Leishmania braziliensis	eIF5A	1x6O	1.60 Å
Pyrobaculum aerophilum	aIF5A	1BKB	1.75 Å
Neurospora crassa	HEX1	1KHI	1.78 Å
Leishmania mexicana	eIF5A	2EIF	1.80 Å
Homo sapiens	eIF5A1	3CPF	2.50 Å
Leishmania mexicana	eIF5A	1XTD	2.70 Å
Saccharomyces cerevisiae	eIF5A	3ER0	3.35 Å

Within the eukaryote IF5As the crystal structures show that monomeric eIF5A consists of 2 distinct anti-parallel beta sheet domains that are connected by a flexible hinge (J. H. Park et

al., 2012) (Figure 1-2A). Secondary structure matching queries classified the basic N–terminal domain as being an SH3-related barrel commonly found in mRNA binding ribosomal proteins and the C–terminal domain as an oligonucleotide binding fold (Parreiras-e-Silva et al., 2007; Tong et al., 2009). The basic N-terminal domain contains 6 β-strands, with a single turn α-helix between strands β1 and β2 in human, with PDB ID 3CPF, but absent in yeast, with PDB ID 3ER0, variants. The smaller C-terminal OB-fold domain contains five beta strands which form two anti-parallel β sheets which fold across one another forming a β sandwich enclosing a hydrophobic pocket (Kim et al., 1998)(Figure 1-2 B). Within this OB fold a 3 turn α-helix between β9 and β10 is present in the human and yeast homologues, which is absent in the *Lieshmania mexicana*, 2EIF (Tong et al., 2009)(Figure 1-2B).

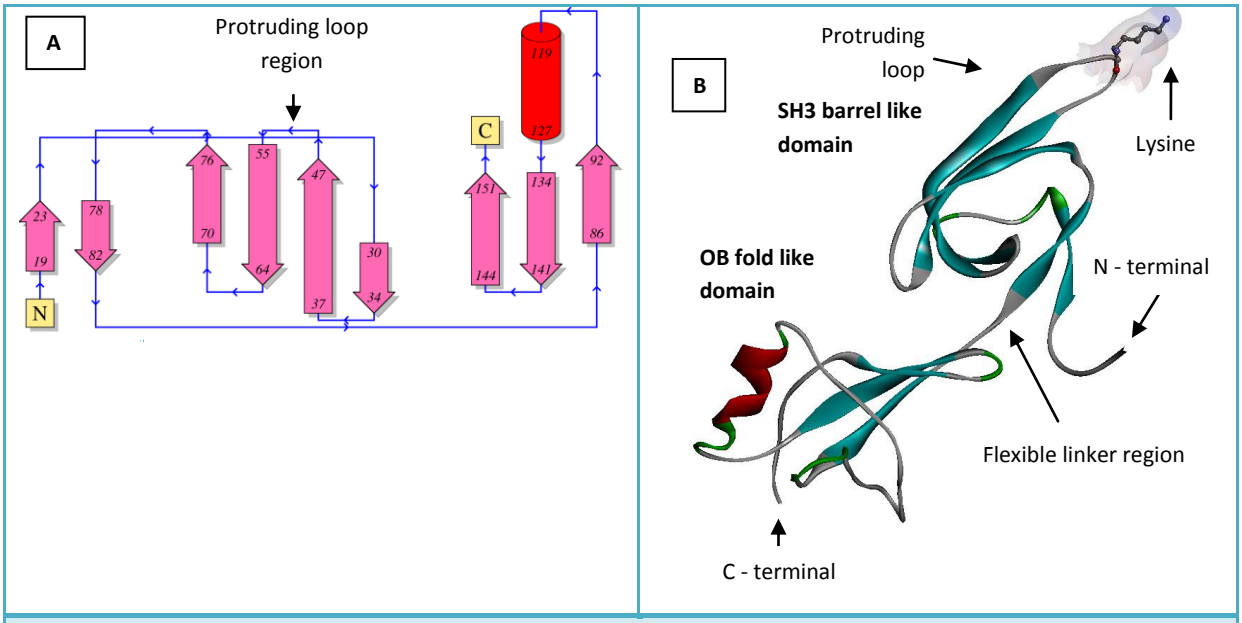


Figure 1-2: Schematic illustrations of yeast eIF5A chain A crystal structure (PDB ID: 3ER0) A. 2D topology diagram obtained from PDBSum PROMOTIF v3.0. Residue numbers indicated start and end of secondary structure. 9 Beta strands are computed and illustrated as arrows and the single helix as a tube. B. 3D diagram obtained from Discovery Studio Visualizer 3.1. Coloured by secondary structure with Lysine (51) displayed as ball and stick with VDW surface highlighting charge distribution.

The protruding loop between strands β3 and β4 in eukayrotes contains the highly conserved STSKTG**K**HGHAK consensus sequence that contains the hypusinated lysine (K51) (Figure 1-2B) (Cooper et al., 1983; Seung Bum Lee et al., 2010). Besides possessing the hypusination site, it has been proposed that residues in this region are essential for activity of

hypusinated protein through enhancing RNA binding, protein-protein interactions and hypusine orientation (Cano et al., 2008; Gentz et al 2009). Mutational studies showed that an acetylated lysine downstream of the hypusination site prevents DHS binding (Choudhary et al., 2009). Although bacterial EFP does not undergo hypusination, a unique lysine residue within a well conserved region complimentary to that in eukaryotes undergoes post-translational modification involving the addition of a β-lysine moiety to lysine (Park et al., 2012). This β-lysylation has been shown to be analogous to that of hypusination of eIF5A and is essential for bacterial cell growth (Rossi et al., 2013).

1.2.3.2 eIF5A mechanisms of oligomerisation

Native eIF5A obtained from rabbit reticulocytes has a molecular weight of 25.8 kDa based upon its K_d observed from gel exclusion under neutral isotonic conditions while under denaturing conditions molecular weights of between 17 and 19 kDa were observed. This implies a strong tendency of eIF5A to self-associate into dimmers (Kemper et al., 1976). Equilibrium sedimentation of human eIF5A in neutral isotonic conditions showed that protein purified from erythrocytes formed multiple oligomeric species consisting of dimers, tetramers and hexamers (Chung, Park, Folk, & Lewis, 1991). A study on yeast demonstrated the oligomerisation of native hypusinated eIF5A. Gel filtration of different concentrations showed that this dimerisation was not from concentration-dependant aggregation. Furthermore, mutation studies on Lys 51 and gel filtration in the presence of RNase A, showed that dimerisation was dependant on hypusination and RNA binding (Gentz, Blatch, & Dorrington, 2009). Gentz et al (2009) made use of the structural similarity of HEX-1 with yeast eIF5A and its well defined oligomeric structure to model the oligomeric state of yeast eIF5A in order to identify the potential residues critical for dimerisation. The dimer models obtained showed a strong possibility of structure stabilisation through residues that are likely to form salt bridges and hydrogen bonds with K51.

A follow-up study by Charlton (2012) on a K51R mutant of the yeast eIF5A protein produced contradictory results. This mutation resulted in a lethal phenotype although the oligomeric state of the mutant protein measured by gel filtration was similar to that of the wild-type protein (Charlton, 2012). Dias et al., (2013) were able to show that oligomerization of yeast eIF5A *in vitro* and *in vivo* was RNA dependent and independent of ionic interactions. Based

on size exclusion chromatography, GSTpull-down assays and Small-angle X-ray scattering experiments they were able to show that RNA dependent dimerisation was independent of hypusination in protein over-expressed in recombinant bacteria and yeast systems (C. A. O. Dias et al., 2013).

Gentz was able to show that the integrity of the yeast eIF5A dimer purified from overexpression in *E. coli* was dependent on different mechanisms compared to that obtained from the yeast cells. In contrast to the yeast, this dimer was not dependent on hypusination and although it was similarly disrupted with addition of reducing agents, unlike in yeast, mutations of Cys 39 abolished dimerisation (Gentz, 2008). *In vivo* mutation studies confirmed the structural role of hypusination in dimerisation and function. Charlton found that G50A, K48D, H52A and K56A yeast produced eIF5A mutants with mutations within the vicinity of the unstructured extended loop region , which had been identified as giving rise to yeast cell defects perhaps due to loss of eIF5A activity, were shown through *in vivo* assays to be hypusinated (Cano et al., 2008; Charlton, 2012). Biochemical analyses to investigate the oligomeric state of these mutants revealed that dimerisation was impaired by these mutations (Charlton, 2012). There is a need for structural data of the hypusinated protein interacting with an RNA moiety stabilising the loose dimer in order for the function of this elusive protein to be better understood (Gentz, 2008). Analysis of the yeast produced hypusinated protein using glutathione S-transferase pull-down assays followed by mass spectrometry confirmed Gentz's observations of RNA dependent dimerisation, but they also showed that in the absence of hypusination dimerisation of eIF5A is possible. Although they conclude that hypusination is not important for dimerisation the authors appear to not be aware of the influence of the different mechanisms of dimerisation on eIF5A function.

1.3 MOLECULAR BIOPHYSICS AND STRUCTURAL BIOLOGY

One of the goals of structural genomics initiatives is to contribute towards a greater understanding of protein function and propose an accurate biochemical and biological role of a protein within the organism derived from its structure (Thornton et al., 2000; Xiang, 2006). Currently the 2 main techniques that have been employed to solve the complete 3 dimensional structures of macromolecules at the atomic level are X-ray crystallography of

solid single crystals and nuclear magnetic resonance spectroscopy (NMR) for samples in their solution (Marius Clore & Gronenborn, 1999). The main limitations that arise in X-ray crystallography arise due to the difficulty to obtain sufficient heavy atom derivatives that assist in determining the phases of the reflections accurately. Multiple-anomalous dispersion has been developed to facilitate accurate phase determination but it has limited applications without suitable crystals (Kim, Hung, Yokota, Kim, & Kim, 1998). Although significant breakthroughs have occurred in crystallization techniques it is also very difficult to obtain suitably shaped crystals that diffract the irradiated beams to a sufficient resolution in most macro-molecules (Marius Clore & Gronenborn, 1999). These limitations can be clearly illustrated by noting the relatively fewer numbers of protein X-ray structures solved compared to the number of available protein sequences identified.

The determination of high resolution 3-D structures of proteins celebrated its 50th anniversary (NIGMS, 2011) since Kendrew and Perutz published the first high resolution structure of myoglobin and haemoglobin using the novel isomorphous replacement X-ray crystallography technique (Kendrew, 1963; Kendrew et al., 1958). Advancements in the fields of Molecular Biology and Genetics contributed to the understanding of protein primary and secondary structures. In 1962, Rossmann and Blow showed the possibility of solving unknown protein structures from previously solved structures of similar sequences using molecular replacement which formed the basis of modern homology modelling techniques (Rossmann & Blow, 1962). Electron microscopy and neutron diffraction showed that the sensitivity of crystallography techniques could be improved and they used them to extend their resolution (Finch & Stretton, 1964; Schoenborn, 1969).

In 1972, Jackson et al. showed the possibility to clone individual genes into enzymatically cleaved DNA molecules (Jackson, Symonst, & Berg, 1972). Cohen, later showed that these DNA molecules could contain heterogeneous genes that could produce functional proteins in host cells (Cohen, 2013). This contributed significantly to the rapid development of recombinant DNA technologies which resulted in a dramatic increase in the quantities of protein available for structural studies (Berman, 2008). Improvement of data acquisition and advances in the study of solution dynamics of proteins through NMR have enhanced the feasibility and accuracy of structure determination contributing to the large diversity of

structures solved and published (Berman, 2008; Phillips, Wlodawer, Yevitz, & Hodgson, 1976; Wagner & Wüthrich, 1978).

Ramachandran *et. al*, used the conformations of allowed conformations of amino-acids and excluded conformations that enabled close non-bonded contacts to generate theoretical structures (Ramachandran, Ramakrishnan, & Sasisekharan, 1963). Conformational analysis represented by an analytical function of atomic positions, chemical connectivity and interatomic distances of small molecules was extended to proteins to develop theoretical models for understanding protein structure and function (Levitt, 1976; Ramachandran et al., 1963). These techniques proved to be computationally expensive and irreproducible. This was due to the relatively larger number of atoms present in proteins than in smaller molecules, the difficulty to account for properties to be computed at higher room temperatures where proteins are stable and the numerous variables that needed to considered when calculating the overall energies of proteins as opposed to small organics (Gibson & Scheraga, 1969; Levitt & Warshel, 1975). In 1976, Levitt showed that approaches towards calculating time-averaged forces and simplified conformational energy calculations of globular structures could assist in the understanding of structure and conformation. They were able to successfully apply this work to give detailed descriptions of molecular geometry using suitably parameterized force-fields and appropriate energy minimization procedures to efficiently prepare folding simulations of a pancreatic trypsin inhibitor (Levitt, 1976). These techniques paved the way for applications of structure determination in understanding enzyme-substrate binding, protein-protein interactions and the changes in tertiary and quaternary structure.

The first programs that were used to visualize protein structures were developed in 1976 by Langridge at the University of California (Monmaney, 2005). They combined existing data and wrote software that was able to visualize structures with colour displays to differentiate the different atoms (Langridge, 1974). In 1981 the first 3D illustrations of the full range of protein structures were shown by Jane Richardson (NIGMS, 2011). Using earlier work done by Thornton and Chothia to generate two-dimensional topology diagrams, Richardson systematically devised rules, relationships and modified visualization software to develop

ribbon Richardson diagrams for the schematic comprehension of protein organization and structure (Richardson, 1981, 2000).

The development of visualisation software, innovations in molecular biology, rapid advancement in hardware and technologies have all contributed equally to the interdisciplinary field of structural biology. One of the largest protein structure databanks, The Protein Databank, was started in Brookhaven National Laboratory in 1971 under the sponsorship of Walter Hamilton as collaborations between crystallographers increased (Berman, 2008; Berman et al., 2000). Advancements in computing hardware and software allowed Meyer to write software that would allow graphical and geometrical evaluation of stored atomic co-ordinates in a common format (Meyer, 1997). They envisaged an efficient way to collect and distribute data of solved protein structures which began to increase as groups around the world found it easier to determine structures (Berman, 2008; Meyer, 1997). Currently the database contains over 88,000 protein structures solved by X-ray, NMR, electron microscopy, homology modelling and combinations of techniques (RCSB, 2013). Although only 10 % of structures deposited in PDB were solved using NMR compared to 90 % by X-ray techniques, there is a growing significance placed on the application of NMR because of its ability to study protein structures and dynamics in solution (Keeler, 2005; Powers, Moy, Siegel, & Mobilio, 2001; Xiong, 2006).

1.3.1 Nuclear Magnetic Resonance (NMR)

Nuclear Magnetic Resonance spectroscopy (NMR) is a non-destructive physical technique that is routinely used to characterize simple organic compounds as solids or in solution (Keeler, 2005). Its applications have been extended to study the structure, kinetics and dynamics of larger bio-macromolecules in solution (Rule & Hitchens, 2006). Ever since Wüthrich and co-workers solved the solution conformation of the proteinase inhibitor IIA peptide using two dimensional nuclear Overhauser enhancement spectroscopy, intense interest has surrounded the use of NMR for biophysics and structural biology applications (Rule & Hitchens, 2006; Williamson, Havel, & Wüthrich, 1985; Xin, 2003). Innovative developments through improved methodology in manipulation of pulse programs and hardware that gives rise to higher magnetic fields have contributed to the advancement of this technique (Delaglio, 2001; Griesinger & Sattler, 1999). Numerous contributions from

molecular biology have made it possible to specifically, efficiently and routinely isotopically enrich protein samples increasing their accessibility to studies (Kainosho et al., 2006; Ohki & Kainosho, 2008).

Bio-macromolecular NMR can be used to generate solution structures of proteins less than 40 – 60 kDa to atomic resolution (Cavanagh, Wayne, Palmer III, Rance, & Skelton, 2007). Although spectra have been obtained for proteins larger than 100 kDa, X-ray crystallography is the preferred tool for high resolution structure determination of molecules above the NMR upper limit (Fernandez & Wider, 2006; Snyder et al., 2005). It is arguable that high resolution protein structures solved through NMR are more superior to X-ray methods because the conformations are obtained in native like conditions (Cavanagh et al., 2007; Wüthrich, 2003). Partially folded polypeptide chains and segments are usually difficult to crystallize, and visualize through diffraction methods (Cavanagh et al., 2007; Kai J Kohlhoff, Robustelli, Cavalli, Salvatella, & Vendruscolo, 2009; Rule & Hitchens, 2006). Crystal lattices have the tendency to enhance non-native intermolecular contacts of partially folded segments that would otherwise be dis-ordered in solution (Wüthrich, 2003). Because bio-molecules studied using NMR are in solution it has been shown to be able to resolve ambiguous structures that arise due to crystal packing affects from crystallography (Ugwu & Apte, 2004). NMR has consequently been the preferred method for resolving partially folded protein segments (Wüthrich, 2003). However extensive work by the Northeast Structural Genomics Consortium in 2 pilot studies consisting of 420 proteins, showed that NMR and X-ray crystallography are not opposite tools used to investigate structure but can be used together to obtain better and more reliable structural data (Snyder et al., 2005; Yee et al., 2005).

1.3.1.1 Principles of NMR

NMR spectroscopy relies on spins of NMR active nuclei in the presence of a magnetic field and is based on the interaction of their magnetic momenta and the external magnetic field (James, 1998a; Savukov et al., 2007). This interaction generates more than one Zeeman energy spin state and transitions between these states can be induced and detected through electronic circuitry of the probe (Cavanagh et al., 2007; Rule & Hitchens, 2006). The energy difference, ΔE, between the states is nuclei specific and is related to the Larmor

frequency of precession of the nucleus around the axis of the magnetic field, ω_o, in the presence of a field of strength $\mathbf{B_o}$ (Reich, 2010). The ω_o frequency of the nuclei is influenced by a corrected magnetic field *B* based on *σ*, where *σ* is the chemical shielding factor due to the atomic electronic oscillations, and the nuclei-specific gyromagnetic responsiveness factor, *γ*.

$$\Delta E = E_\beta - E_\alpha = \gamma\hbar B; \ \omega_o = \gamma B; B = B_o - \sigma \quad (1.1)$$

Chemical shifts

The energy of the transition between the 2 spin states described by equation 1.1, is dependent on the nucleus *γ*, and the local magnetic field corrected by the screening, *σ*. Shielding from molecular electron density generates a magnetic field proportional to $\mathbf{B_o}$ which alters the field experienced by the nucleus (Reich, 2010). This correction is approximately a million times smaller than the frequency of the transition and can be used to describe the chemical environment or functional group of the particular resonance (James, 1998).

Chemical shifts of carbon, nitrogen and proton nuclei in proteins have been shown to be sensitive to local arrangements where factors such as hydrogen bonds, electronegativity, ring currents and the backbone conformations influence the measured chemical shifts (Kohlhoff et al., 2009). Secondary structure can readily be determined from chemical shifts of assigned resonances obtained from high resolution NMR spectra (Cavalli et al., 2007; Peti et al., 2001). Growing interest is developing around development of rapid protocols that will allow fast accurate determination of tertiary structure from the measurement of chemical shifts (Neal et al., 2003). When the chemical shifts of all applicable atom types are considered, a precise picture of the molecule's configuration and geometry can be derived when limitations due to covalent bonding and steric overlap are imposed (Kohlhoff et al., 2009). Whether a particular protein is folded or unfolded can be easily determined from its chemical shift patterns. Unfolded conformations adopt random coil dynamics in solution while folded conformations possess diverse and persistent interactions which result in sparse deshielded and shielded resonances (Fleming & Rose, 2008; Peti et al., 2001).

Coupling

Apart from the influence of chemical shielding on the chemical shift of a resonance, coupling between nuclei that are able to interact with each other can influence the energy of transitions between spin states in an NMR active nucleus. Two types of coupling that are of interest in this study are those that arise from scalar (*J*) interactions or from dipole-dipole interactions (Reich, 2010). Correlated nuclei in Fermi contact with each other experience indirect, spin-spin, scalar coupling through their *s* orbitals. Polarization of the *s* orbital of nucleus A occurs through the orientation of its spin within the magnetic field. This polarization in turn affects the electron distribution of a nucleus B directly or indirectly bonded to nucleus A. The perturbed electron distribution of nucleus B influences the chemical shielding of nucleus B and results in a small measurable difference in its resonance (Teng, 2013).

A dipolar coupling exists between 2 nuclear spins which are connected through space. This coupling depends on the distance between the 2 nuclei and experiences an angular dependence between the internuclear vector and the external magnetic field (Keeler, 2005; Reich, 2010). As the molecule tumbles in solution this dipole-dipole coupling averages to 0 and is not detected but in anisotropic solutions its influence can be detected as a residue specific dipole-dipole coupling. If the alignment frame is defined from the molecular frame and the laboratory frame, the time-averaged molecular orientation of a dipolar interaction with respect to the alignment frame can be determined by transforming the internuclear vector within the axes of the molecular frame (Cavanagh et al., 2007; Teng, 2013).

Relaxation

There are 2 main phenomena that are responsible for thermal equilibrium magnetization recovery through relaxation (Keeler, 2005). Following perturbation, the equilibrium of the system is restored through longitudinal relaxation of the energy level populations and transverse relaxation of the phase incoherence (Cavanagh et al., 2007; Reich, 2010). Only those mechanisms that are influential to spin ½ nuclei in diamagnetic biological macromolecules will be discussed.

Longitudinal relaxation: Longitudinal recovery of the magnetization along the Z axis is called T_1 spin-lattice relaxation and occurs through the interaction of the excited spins with their

lattice or surroundings (Carbajo & Neira, 2013b; Cavanagh et al., 2007; James, 1998; Rule & Hitchens, 2006). T_1 relaxations arise from oscillations of temporary fields within their lattice that fluctuate at rates comparable to the Larmor frequencies, ω_o, of the excited nuclei (Reich, 2010). These oscillations are generated from molecular motions such as isotropic tumbling of the molecule in solution or vibrations of internuclear bonds influencing inter-molecular and intra-molecular field fluctuations (Cavanagh et al., 2007; Wüthrich, 2003). Examples of some of these interactions that give rise to T_1 relaxations due to molecular motions are spin-rotation relaxation, chemical shift anisotropy relaxation, quadrupolar relaxation, scalar relaxation and Dipole-Dipole relaxation (James, 1998; Reich, 2010; Savukov et al., 2007). Quadrupolar relaxation is only relevant when nuclei that have non spherical nuclei charge distribution are present (Reich, 2010) and will not be discussed further here.

Rotating electronic spins generate a local magnetic field over a molecule or part of a molecule (Plonus, 1978). As these molecules rotate they create rotational quantum levels and if fluctuations of the generated fields oscillate at frequencies comparable to ω_o, they can initiate relaxation of nearby nuclei (Cavanagh et al., 2007; Nkari & Prestegard, 2010; Reich, 2010). This is described as spin-rotation relaxation and contributes to the T_1 relaxation of nuclei. Chemical shift anisotropy arises due to the changes of the orientation of a molecule with magnetic field. When these changes are comparable to ω_o relaxation of the nucleus occurs (James, 1998; Keeler, 2005). This mechanism is influenced by γ and $\mathbf{B_o}$ and correlates with isotropic molecular tumbling (Nkari & Prestegard, 2010; Reich, 2010). Re-orientations of the molecules with respect to $\mathbf{B_o}$ fluctuate the chemical shifts of the nuclei contributing to T_1 relaxation phenomena (Reich, 2010).

Relaxation can arise as a result of fluctuating magnetic field due to the changing dipole-dipole coupling between directly bonded magnetic nuclei. This coupling is influenced by the angular relationship between the external magnetic field and the internuclear vectors and as the molecules tumbles these relationships change. In rapidly tumbling molecules or vectors fluctuations are no longer proportional to ω_o and thus relaxation is less effective. When fluctuations are slower, relaxation is more efficient and results in a shorter T_1 (Cavanagh et al., 2007; Keeler, 2005; Reich, 2010).

Transverse relaxation: Relaxation in the transverse x-y plane occurs in order to decrease the phase coherence of nuclei introduced by the excitation (Reich, 2010). In the absence of a pulse, different nuclei within the system precess at different rates and frequencies dependent on their chemical environment influencing their local magnetic field. This magnetic field inhomogeneity results in a dephasing of the net magnetization vector within the transverse plane at equilibrium (Teng, 2013). T_2 relaxation is termed spin-spin relaxation and is defined as the time it takes in order for the equilibrium incoherence to be restored after a pulse (Cavanagh et al., 2007; James, 1998). T_2 relaxation occurs through chemical exchange mechanisms that arise as a result of swapping of chemical shifts and coupling constants between excited nuclear spins as well as fluctuating transient magnetic fields (Carbajo & Neira, 2013b; Reich, 2010).

Sensitivity

The population, *N*, distribution between two Zeeman states, α and β, when the spin system is unperturbed by a radio frequency pulse, is described by Boltzmann's distribution. Because the energy difference between the 2 different spin states, influenced by the size of the local magnetic field, *B*, and the responsiveness of the nuclei in that field, γ, is small the relative population differences between the states of allowed transitions is of the order of 1 part in 10^6 for a typical nucleus in modern spectrometers.

$$N_\beta / N_\alpha = e^{-\Delta E/\kappa T} = e^{-\gamma\hbar B/\kappa T} \approx 1 - \gamma\hbar B/\kappa T \qquad (1.2)$$

A change in polarisation is as a consequence of the net transitions of spins as a result of the RF pulse which supplies energy to overcome the barrier between the states (Hobbie, 1988). Due to the equal probability of transitions in either direction the population inversion is low and as a result NMR is frequently described as an insensitive physical technique (Keeler, 2005; Savukov et al., 2007).

Table 1-2: Magnetic properties of spin ½ nuclei relevant for biomacromolecular NMR (James, 1998)

Nucleus	Spin Quantum Number ($\hat{S}$)	Natural Abundance (%)	Gyromagnetic Ratio (γ) (10^{-7} rad/T sec)	Sensitivity (% vs ^{1}H)
^{1}H	½	99.9844	26.7520	100.00
^{13}C	½	1.108	6.7265	1.59
^{15}N	½	0.365	-2.7108	0.104
^{19}F	½	100	25.167	83.3
^{31}P	½	100	10.829	6.63

Because of the increase spectral overlap observed in protein NMR, the use of heteronuclear nuclei and transfer of magnetization has been used to improve the resolution (Frydman, Scherf, & Lupulescu, 2002; Griesinger & Sattler, 1999). The table above describes the magnetic properties of NMR active nuclei present in proteins. Because of the low natural abundance of ^{13}C and ^{15}N isotopes, enrichment approaches were significant in extending the applications of NMR based on the use of multidimensional techniques. Due to its large natural abundance and magnetic responsiveness, ^{1}H nuclei are often used in detection (Cavanagh et al., 2007).

1.3.1.2 Applications of NMR

Structure determination

Most protein structures solved by experimental techniques to date have been solved by single crystal X-ray crystallography techniques (Berman, 2008). When the PDB was accessed recently over 80% of solved structures where by X-ray crystallography while only 10 % were through NMR derived restraints. The three major problems that arise when studying proteins in solution through NMR are signal overlap, fast transverse relaxation and limited solubility (Fernandez & Wider, 2006). As a result of these limitations the size distribution of most structures solved by NMR is centred around 10 kDa (Ohki & Kainosho, 2008; Wuthrich & Wider, 2003). The use of ^{13}C and ^{15}N isotope labelling creates efficient magnetization

transfer and allows measurement of multidimensional spectra which improves the resolution of spectra for larger molecules solved by NMR (Cavanagh et al., 2007). Deuteration of non-exchangeable protons has been shown to improve spectra resolution allowing applications of transverse relaxation optimized spectroscopy (TROSY) which thus contributes to the NMR molecular weight range accessible to NMR (Ohki & Kainosho, 2008; Wider, 2000; Wuthrich & Wider, 2003).

Structure determination using NMR generates indirect data related to the structure of the protein which can be incorporated indirectly as structural restraints to calculate a final structure (Fernandez & Wider, 2006). The general strategy for structural determination using NMR is characterised by i) protein preparation, ii) NMR measurements, iii) assignment of NMR signals to individual atoms, iv) identification of conformational and structural restraints and finally, v) structure calculations to satisfy conformational and structural restraints (Wider, 2000). High concentrations of stable homogenous protein samples are essential for good quality spectra to be recorded with good sensitivity and resolution (Folkers, van Buuren, & Kaptein, 2004). Aggregation of protein samples can significantly impair NMR measurements and the use of NMR inactive buffers and surfactants are used to maintain high stability and solubility at critical concentrations (Lesley, 2001; Wider, 2000).

Homonuclear methods

For smaller proteins of about 10 kDa it is possible to detect and efficiently assign spin-systems of amino acids and map the relationships of their connectivities based on [^{1}H, ^{1}H]-COSY experiments. This experiment measures through bond, scalar coupled, interactions between protons up to 3 bonds apart. Each amino acid has a unique chemical shift cross peak pattern which allows it's identification (Wider, 2000). [^{1}H-^{1}H]-NOESY experiments allow through space correlations between pairs of protons within neighbouring spin systems to be collected and compared to the known amino acid sequence (Carbajo & Neira, 2013a). For larger proteins this is more difficult and a Total Correlation Spectroscopy (TOCSY) spectrum can allow the identification of amino acids based on their fingerprint spin-systems. TOCSY experiments can link resonances at opposite ends of spin systems efficiently thus identifying the amino acid types in larger proteins (Cavanagh et al., 2007; Keeler, 2005).

Homo-nuclear dipole-dipole cross-relaxation experiments such as Nuclear Overhauser Effect Spectroscopy (NOESY) are used to collect distance geometry and conformational information in the form of an integrated peak intensity of the proton-proton cross peak which gives indications of relative distances. Many unambiguous NOESY distances, which are limited to within 5 Å, are collected over a long time period and are used to prepare a protein structure that satisfies all the distance restraints defined (Wider, 2000)..

Heteronuclear methods

Larger proteins require 2D and 3D NMR techniques to overcome extensive signal overlap and decreased sensitivity in order to accurately resolve chemical signals (Wüthrich, 2003). Triple resonance experiments such as HNCO and HNCA are complementary experiments routinely used to collect scalar-coupled correlations across peptide-bonds of [^{13}C, ^{15}N]-labelled protein samples in structure determination (Carbajo & Neira, 2013a). The HNCO experiment generates a characteristic proton cross peak that arises from the transfer of magnetization from an amide proton to the amide nitrogen, which then correlates with the carbonyl carbon of the preceding amino acid which formed the peptide bond (Carbajo & Neira, 2013a). The HNCA experiment is similar to the HNCO experiment but makes use of the small intra-residual one bond [^{15}N-^{13}C$^{\alpha}$] scalar coupling to transfer magnetization from the amide nitrogen to the alpha carbon of the same amino acid. The α-proton from the α carbon of the preceding amino acid is also detected in HNCA due to a comparable inter-residual two bond [^{15}N-^{13}C$^{\alpha}$] scalar ($^{1}J_{NC\alpha}$) coupling which can allow transfer of some of the amide nitrogen magnetization (Cavanagh et al., 2007). Where the 2 α carbons have degenerate and indistinguishable chemical shifts an HN(CO)CA experiment can be performed (Carbajo & Neira, 2013a). This allows the intra-residual and preceding α-carbons to be distinguished through only detecting the α-carbon that was accessed through a magnetization path that utilised the preceding carbonyl carbon in the peptide bond (Cavanagh et al., 2007). By linking the cross peaks obtained from the HNCA and HNCO experiments, the cross-peaks giving rise to the peptide bonds can be identified sequentially (Herrmann, Guntert, & Wuthrich, 2002; Keeler, 2005).

The information obtained from the HNCA, HNCO, TOCSY and NOESY experiments is used to build a 3D protein model that satisfies all the conformational restraints and connectivities

(Herrmann et al., 2002) Other structural restraint data such as residual dipolar couplings of proteins in different alignment media and paramagnetic ion induced pseudocontact shifts can be incorporated into structure calculations when NOESY data is scarce and ambiguous for larger proteins (Carbajo & Neira, 2013a; Kai Jochen Kohlhoff, 2008).

Residual Dipolar Couplings Pseudocontact shift methods

Residual dipolar couplings (RDCs) of an assigned protein can be measured rapidly and used to calculate N-H^N, C^α-H^N, C^α-N and C^α-H^α internuclear vectors relative to the molecular frame of the protein (Delaglio, Kontaxis, & Bax, 2000). The RDC's measured can give information about the restricted orientation of an internuclear vector relative to the external magnetic field. In order to use this data to define the orientation of a vector specifically, RDC's of many inter-nuclear vectors are collected to create large data sets of orientations and are used to unambiguously define the specific geometries (Carbajo & Neira, 2013a; Rohl & Baker, 2002). When specific vectors can be confidently set as rigid, specific orientations and geometries of other vectors can be calculated giving long range structural information (Roberts, 2000; Rohl & Baker, 2002). RDC restraints can also be used with ab initio structure determination algorithms of proteins by incorporating them into database searches which screen protein fragments that contain predicted RDC's that match measured values. This approach can be used to predict phi and psi backbone angle constraints and alignment tensors which can eventually generate an acceptable structure at high resolution (Delaglio et al., 2000; Delaglio, 2001; Rohl & Baker, 2002).

In proteins that contain paramagnetic ions, long range structural restraints from the dipolar interactions between the unpaired metal electrons and the resonating nuclei can be observed using relatively simple and fast NMR experiments (Pintacuda, Park, Keniry, Dixon, & Otting, 2006). Paramagnetic ions can induce hyperfine shifts arising from a combination of Fermi contact shifts, originating from scalar coupling of nuclei within a few bonds of the metal, and relaxation enhancement of nuclei within 40 Å of the paramagnetic ion (Jensen et al., 2006). Paramagnetic enhancement of longitudinal and transverse relaxation is inversely proportional to the 6^{th} power of the distance of the nuclei from the unpaired electron of the ion, and can be used to collect distance restraints from the shift observed analogous to the measurement of NOE's (Bertini et al., 2001). To define the dipolar interactions, electron spin

density matrices are generated to account for spin polarization and delocalisation of unpaired electron spin-density of the paramagnetic atoms into ligand orbitals. These more accurate descriptions of the dipolar interactions help to estimate the effective distance inferred by interaction observed (Jensen et al., 2006; Wilkens et al., 1998).

Fermi contact shifts are less influential on the measured pseudocontact shifts (PCS) because they are susceptible to fast exchange with solvent waters. PCS measurements of the differences in chemical shifts of the paramagnetic and the diamagnetic cases are more accessible for structural studies because they are not affected by these fast exchanges with the solvent waters (Jensen et al., 2006). It has been shown that when 3D knowledge of a protein is available it is possible to record sufficient PCS data from a routine ^{15}N-HSQC for accurate structure refinement (Pintacuda et al., 2006). They can readily measure long-range interactions, θ and φ angles defining the orientation of the magnetic susceptibility tensor of the metal with respect to the resonating nucleus (Graham et al., 2011; Pintacuda et al., 2006).

High-throughput screening and small molecule interactions

NMR techniques can be applied to the drug discovery process to assist in the screening and evaluation of compounds for applications to drug discovery (Roberts, 2000). NMR has been used as a fast technique to detect sensitive intra-molecular interactions and give information at atomic resolution (Diercks, Coles, & Kessler, 2001). Proteins that have been discovered purely on a genomics basis without immediate availability of bioassays can be screened through NMR's versatility (Petsko, 2006; Pineda-lucena, 2006). Determination of binding constants as well as accurate identification of exactly which segments of the ligand are interacting with the target are possible (Dalvit et al., 2000).

Although methods such as chemical shifts mapping, diffusion rate transfer, transverse relaxation, and nuclear Overhauser effects have been developed for NMR based screening they can be used for function studies of proteins (Hajduk, Meadows, & Fesik, 1999). These methods are able to identify different binding sites and affinities of weakly binding peptides, lipids, RNA, other proteins and ligands on macromolecular targets thus evaluating the

interactions of proteins (Hajduk et al., 1999; Kallen et al., 1991; Zheng et al., 1996). (Betz, Saxena, & Schwalbe, 2006).

1.3.2 Homology Modelling

Ever since the work of Ramachandran et al., (1963) and Gibson and Scheraga (1969), protein structure prediction has been the dream for physicists, mathematicians, computer scientists and computational chemists alike (Xiang, 2006). Better understanding of protein stability coupled with significant advances in computational competence has allowed this dream to be accessible. With an increase in the demand of structural information to better understand protein function, tedious experimental techniques to solve protein structures are being replaced by faster computational techniques that bridge the gap between known sequences and predicted structures (Stahl, Guba, & Kansy, 2006; Xiang, 2006).

Protein structure prediction can be divided into 2 main types of approaches classified according to the methods employed to predict structure from the known sequence (Xiong, 2006). The first method, homology based modelling, proposes that the three dimensional structure of a protein can be found through its sequence similarity to one or more proteins of known structures (Xiang, 2006). The justification of this is found in the evolutionary relationships of protein sequences classified by structural genomics initiatives that aim to structurally characterize all proteins systematically (Goldsmith-fischman & Honig, 2003; Xiang, 2006). Protein families with evolutionary relationships have been shown to contain significant structural similarities. Where a similarity of at least 30 % has been shown with an already solved structure, it is possible to have a meaningful structure that has a resolution of 3.5 Å, comparable to that of low resolution X-rays (Goldsmith-fischman & Honig, 2003; Shi, Blundell, & Mizuguchi, 2001; Söding, Biegert, & Lupas, 2005). When Issar and Gaur investigated the nitrogenase enzyme of *Pseudomonas putida* they were able to show higher resolutions which allowed 90 % of side-chain atoms to be modelled to RMSD confidence levels of 1 Å when similarities were above 40 % (Issar & Gaur, 2013).

The second method attempts to solve the difficult problem of structure prediction from *ab initio* first principles similar to energy minimization and optimisation employed in small molecules. By using physical chemistry techniques to generate energy landscapes it is

possible to sample all the potential conformations available (Xiang, 2006). The difficulty with this approach, apart from it being potentially time consuming, is that the energy difference between folded and unfolded conformations is relatively small and the global minimum selected may be different to the true conformation (Peter G Wolynes, Eaton, & Fersht, 1998; Zwanzig, Szabo, & Bagchi, 1992). Levinthal in 1969 showed that there are a large number of possible conformations that an unfolded polypeptide chain can adopt due to the large number of degrees of freedom inherent in each amino acid. He contrasted this process which happens rapidly *in vitro* and postulated that several intermediate steps must direct the protein and thus a random conformational search description of folding was insufficient (Zwanzig et al., 1992). The prediction of protein structure based on first principles has thus faced difficult challenges because of the difficulty of developing force-fields that allow rapid sampling and mimicry of the fine-tuned process of folding (Levitt, 1976; P. G. Wolynes, Eaton, & Fersht, 2012). It has now been shown experimentally that folding routes that a protein can follow are often dependant on other factors and conditions that are essential for preventing aggregation (Wang, Nema, & Teagarden, 2010). These often stabilize unfavourable, inaccessible conformers and allow the proteins to access final conformations which may have been excluded through conformational sampling (Xiang, 2006).

Despite the difficulty experienced in determining protein structure from first principles there is growing interest in groups aiming to study and simulate protein folding (L. E. Kay, 2009; Vallurupalli, Hansen, & Kay, 2008; Wolynes et al., 1998). However the most favourable technique for the prediction of protein structure currently is homology modelling (Xiang, 2006; Xiong, 2006). Although *ab initio* methods have proven to be relatively less successful than Homology modelling, the principles have been applied to structure refinement protocols in homology modelling (Goldsmith-fischman & Honig, 2003). Side chains and conformations of loop regions that do not contain sequence similarity with template molecules have proven to be difficult to predict (Fiser, Do, & Sali, 2000a). *Ab initio* methods based on high-temperature molecular dynamics have been shown to assist conformational searches of proteins. They are used to generate large numbers of loop conformations which are easily screened by evaluating the sum of their normalized energy

functions based on restraints such as bond angles and bond lengths, to yield more accurate conformations of proteins (Bruccoleri & Karplus, 1990). Fiser et al., (2000) showed that for 4 residue loops using a CHARMM-22 forcefield based on 500 independent optimizations, they were able to have 100 % of main-chain N, Ca, C and O atoms with an RMSD confidence <2 Å with an average accuracy of 0.59 $\pm$ 0.05Å. They also showed that the accuracy of their prediction was influenced by the accuracy of the force-field more than the extent of conformational sampling (Fiser et al., 2000a).

Structural genomics has contributed significantly to the field of structural biology. The main goal of international structural genomics initiatives is to experimentally determine enough structures in order for the other structures to be defined by homology modelling (Goldsmith-fischman & Honig, 2003). Although significant progress has been made to solve protein structures quickly and accurately using homology modelling techniques coupled with *ab initio* approaches to loop conformations, this technique is still dependant heavily on the accuracy of the template structures (Xiang, 2006). Improvements in experimentally solved structures will contribute towards more accurate structures solved by homology modelling (Wüthrich, 2003; Xiang, 2006; Xiong, 2006). A limitation common to both homology modelling and crystallography is the inability to accurately determine structural fluctuations. NMR approaches have been shown to provide essential restraints such as residual dipolar couplings that can be incorporated into molecular dynamics simulations to accurately characterise conformational fluctuations of proteins in solutions (Montalvao, De Simone, & Vendruscolo, 2012). Residual anisotropic magnetic interactions and chemical shift perturbations have recently been used to probe invisible, excited state conformations of proteins which exist transiently, informing mechanisms and refining understanding of protein function (Vallurupalli et al., 2008). Combinations of experimental techniques to understand and predict protein structure and dynamics could provide a wealth of information for structural genomics initiatives intending to relate structure to function (NIGMS, 2011; Thornton et al., 2000).

1.3.3 Molecular dynamics in structural biology

1.3.3.1 Structure calculations

Collected structural restraint data from X-ray crystallography, NMR and homology modelling is incorporated into iterative structural calculations. These employ computer programs that satisfy as many of the experimental restraints as possible in order to generate 3D structures of the molecular model structures (Carbajo & Neira, 2013a; Ernst, 1992; Fernandez & Wider, 2006; Wüthrich, 2003). Approaches to structure calculations employ either distance geometry (DG) based restraints or restrained molecular dynamics (rMD) methods (Ernst, 1992). DG based algorithms can search for conformations using interatomic distances to select configurations of the amino acids that satisfy the distance constraints. They may also restrict torsional angles and by adjusting a random starting conformation minimizing the differences of measured experimental distances with the proposed structure a satisfactory structure can be determined (Wider, 2000).

Restricted molecular dynamics relies on the use of classical molecular dynamics to search the energy landscape of conformations using force-fields supplemented by energy terms based on the NMR derived constraints as input parameters for solving Newton's equation of motion (Ernst, 1992; Schwieter et al., 2003; Wider, 2000). When the system is heated it overcomes energy barriers and after cooling 3D conformations are accessed which were previously obstructed due to restrictions. rMD or simulated annealing has the advantage of sampling a larger range of conformational space and incorporation of NMR derived restraints speeds up the process by limiting the degrees of freedom (Delaglio et al., 2000; Delaglio, 2001). Predicted NMR restraints can also be incorporated into structural calculations using DG or rMD simulations at various stages to guide the conformational convergence (Kohlhoff et al., 2009).

In practice structural calculations employ a combination of both algorithms. For an efficient rMD calculation the structure that best satisfies the structural restraints of a DG protocol is used as a starting structure for the rMD step. Different starting structures that satisfy the structural restraints to an accepted degree from the DG are chosen (Carbajo & Neira, 2013a; Wider, 2000). Because the NMR, X-ray or homology modelling derived restraints used for the DG give a description of a range of possible values as opposed to exact structural

restraints, structure calculations typically generate ensembles of possible structures with each minimizing the violations of constraints satisfactorily (Wider, 2000). Typically 100 globally minimized conformers are generated by evaluation of the energy function that is used to guide the simulated annealing process (Wüthrich, 2003). The structures that carry the smallest residual error are chosen to define the precision of the NMR ensemble structure generated. This precision is expressed by an averaged pairwise RMSD of the backbone atoms. Generally an RMSD value in the range of 1 Å indicates a successful well defined structure (Carbajo & Neira, 2013a).

1.3.3.2 NMR and molecular dynamics in probing disorder

Variable precision at different portions of the protein chain are observed as a result of increased disorder which can be classified as either being as a result of static disorder or dynamic disorder (Wüthrich, 2003). Static disorder can be accounted for by the scarcity of NMR restraints that give rise to reduced precision. The increasing use of long-distance restraints from PCS's has been shown to reduce these and improve the accuracy of the NMR structures (Graham et al., 2011; Jensen et al., 2006). Dynamic disorder occurs due to fast nanosecond and sub-nanosecond intra-molecular motions within the protein, difficult to resolve with both X-ray crystallography and protein NMR (Wüthrich, 2003).

Recently it has been shown that back-calculated RDC data from instantaneous conformations of proteins during MD simulations, can be used to extract alignment tensors corresponding to conformational fluctuations of proteins in their native states. By penalizing deviations between experimental and calculated RDC values these tensors are used to restrain protein conformations in molecular dynamics simulations to determine native state dynamics (Montalvao et al., 2012). These approaches which incorporate MD and NMR methods which employ residue specific relaxation times have been shown to be able to solve the structures of low population, invisible excited states of a protein, at high resolutions (Vallurupalli et al., 2008).

Measurement of RDC's can help to define native state dynamics of up to 1 microseconds of partially folded proteins or portions of proteins that exhibit large sub-nanosecond dynamic disorder (Lange et al., 2008). By probing these energy landscapes, valuable structural

information of different states arising from dynamic disorder can be resolved accurately and verified experimentally (Vallurupalli et al., 2008). An extensive molecular biophysics and structural biology approach to structure determination and protein dynamics can inform a better understanding of a protein's structure and proposed function.

1.4 KNOWLEDGE GAP AND AIMS OF THE PROJECT

The manner in which protein structures can oscillate between transitional states can be influenced by numerous different post-translational modifications and as a result many challenges arise in predicting function from structure (Laskowski et al. 1998; C. H. Lee et al., 2001; Thornton et al., 2000). Conclusive evidence does suggest that hypusination is essential for eIF5A function but however it is yet unknown how this mechanism influences the dimerisation of eIF5A and thus influences interactions with other proteins such as syntenin and elongation factor 2 giving rise to the myriad of proposed functions of this ubiquitous protein.

Recent developments surrounding solving the 3D structure of human eIF5A in solution have been limited to the unhypusinated state of the protein. The solid crystal monomeric structure of human eIF5A has been determined while the backbone 1H, ^{15}N and 13C chemical shift assignments of eIF5A have also been identified (Tong et al., 2009; J. Yuan, Jiang, Jin, Zhang, & Yan, 2009). Previous work surrounding solving the structure of eIF5A have resolved around over-expression of eIF5A in bacteria cells and solving a single structure of the crystallized unhypusinated protein. Complexes of eIF5A with targets such as RNA or elongation factors have not been developed owing to the difficulty to generate hypusinated proteins in large quantities. Highly pure, functional, hypusine-containing eIF5A has been produced in E. coli cells by employing a polycistronic vector in large quantities for structural studies (J. H. Park et al., 2011).

Biophysical and computational techniques that relate the time averaged behaviour of the protein in solution can be used to probe the structure dynamics of eIF5A in solution rapidly and accurately. Recent developments in biomacromolecular NMR such as pseudocontact shifts, residual dipolar couplings and transverse relaxation optimised spectroscopy can be used to extend the size range of 600 MHz NMR spectrometers for protein structural studies.

The aims of our project are to use NMR restraints to study the solution behaviour of eIF5A and validate the cleavage approach to obtain terminal fragments of eIF5A for structure determination. By understanding the influence of hypusination on eIF5A solution behaviour the potential for eIF5A to be used as a drug target in anticancer, anti-HIV-1 and anti-inflammatory diseases will be justified (Caraglia et al., 2013; Kerscher et al., 2010; Maier, Tersey, et al., 2010).

Chapter 2: Production of recombinant eIF5A$_{Thr}$ for structural studies

2.1 INTRODUCTION

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains, recombinant plasmids and culture conditions

2.2.2 Protein expression and isotopic labelling

2.2.3 Purification of his-tagged eIF5A$_{Thr}$

2.2.4 Thrombin cleavage

2.2.5 NMR analysis

2.3 RESULTS AND DISCUSSION

2.3.1 pJC6 construct validation

2.3.2 Protein Expression

2.3.3 Thrombin cleavage

2.3.4 NMR Analysis

2.4 CONCLUSION

Chapter 2: Production of recombinant eIF5A$_{Thr}$ for structural studies

2.1 INTRODUCTION

Generating experimental evidence that will lead to the understanding the oligomeric state of hypusinated eIF5A poses a challenge to the structural biologist. With technological advances surrounding the application of nuclear magnetic resonance (NMR) spectroscopy in biological problems the opportunity arises to use this technique to probe the oligomeric state and structure of this elusive protein in solution. NMR has been shown to be capable of solving structures of up to 100 kDa on 1GHz machines (Kwan *et al.,* 2011; Tugarinov *et al.*, 2005; Yu, 1999). Techniques such as relaxation optimized spectroscopy, residual dipolar coupling and pseudocontact shifts have previously been explored as options to extract experimental structural restraints at modest field strengths from HSQC pulse programs (Cavanagh et al., 2007; Graham et al., 2011; Lange et al., 2008; Vallurupalli et al., 2008).

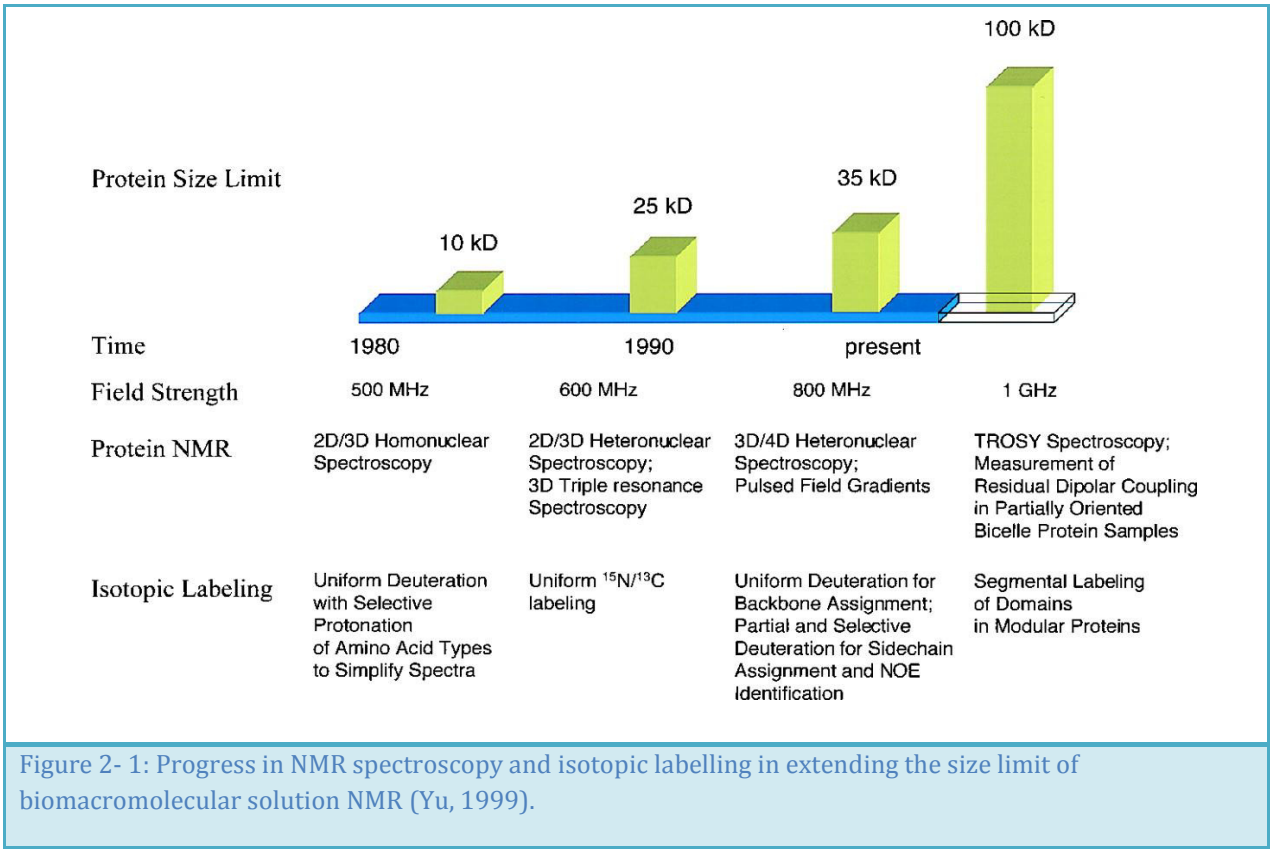


Figure 2- 1: Progress in NMR spectroscopy and isotopic labelling in extending the size limit of biomacromolecular solution NMR (Yu, 1999).

The second challenge is related to the expression of sufficient quantities of the hypusinated version of the protein structural studies. *E. coli* expression systems, routinely used to produce large quantities of protein, result in misfolded, but soluble, unhypusinated versions of the protein (Gentz *et al*., 2009). Successful expression of hypusinated protein by means of co-transfection of the eIF5A, DHS and DOHH vectors in mammalian target cell systems allowed the overexpression of hypusinated and active eIF5A protein in mammalian systems (Lee et al., 2009). Following on from this successful expression, overexpression of hypusinated eIF5A through co-expression in *E. coli* has been achieved providing a means of access to large quantities of isotopically labelled protein in its native, hypusinated and RNA bound dimeric state for structural studies (Park et al., 2011).

As discussed in Chapter 1, the sensitivity and resolution of NMR data is directly proportional to the bulk magnetization and inversely proportional to the size of the protein investigated (Keeler, 2005). In order to optimise this study using the 600 MHz NMR spectroscope it was decided to use two peptides, representing the N-terminus and C-terminus. Charlton (2012) developed an expression vector for over-production of yeast protein *TIF51A* ($eIF5A_{Thr}$) in *E. coli.* The expressed eIF5A contains an introduced thrombin cleavage site within the flexible linker region between the N-terminal domain carrying the conserved, hypusinated K50, and the C-terminal domain carrying the nucleotide binding domain (Figure 1-1 and Figure 2-2). The protein is tagged at its N-terminus with a hexa-histidine tag to enable nickel affinity purification of the recombinant protein. The thrombin cleavage site enables separation of the N-terminal and C-terminal domain producing peptides of 10.35 kDa (the 6 x His – tagged N – terminal domain) and an 8.03 kDa C-terminal domain, both of which are amenable to NMR structural studies at 600 MHz, than the full length eIF5A.

The purpose of the research in this chapter was to conduct NMR studies on the N-terminal domain of the unhypusinated eIF5A protein. $eIF5A_{Thr}$ was over expressed using *E. coli* in isotopically labelled medium and subjected to preliminary NMR analysis.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains, recombinant plasmids and culture conditions

Relevant bacterial strains and plasmid constructs used in this study are listed in table 2-1 and were obtained from Jane Charlton (Rhodes University, South Africa). Recombinant

plasmids were hosted in *E. coli* DH5α cells while expression of eIF5A$_{Thr}$ was carried out in *E. coli* BL21 (DE3) cells. All *E. coli* cultures were grown in Luria Bertani (LB) medium containing 125 µg/ml of ampicillin. Competent cells used for transformation with expression vectors, pJC6 and pT7-7 were prepared following the protocol of Hanahan & Harbor (1983).

Table 2-1: Genotypes of E. coli strains and a description of plasmid DNA used in this study

	GENOTYPE / DESCRIPTION		**REFERENCE**
Strain			
DH5α (*E. coli*)	*F⁻, ϕ80lacZΔM15, Δ(lacZYA-argF)U169, deoR, recA1*	Amp$^-$	(Hanahan & Harbor, 1983)
BL21 (DE3) (*E. coli*)	*F⁻, ompT, hsdS$_a$(r_B-, m_B-), dcm, gal, λ(DE3)*	Amp$^-$	(Studier & Moffat, 1986)
Plasmid			
pT7-7 6xHis	Bacterial expression backbone vector with T7 promoter and 6xHis-tag	Amp^R	(Idicula, Blatch, Cooper, & Dorrington, 2002)
pJC6	Bacterial expression vector derived from pT7-7 6xHis-tag containing the yeast *TIF51A* gene with a thrombin cleavage site inserted into the flexible linker region of eIF5A (eIF5A$_{Thr}$)	Amp^R	(Charlton, 2012)

Plasmid DNA was extracted from *E. coli* DH5α cells using a High Pure Plasmid Isolation Kit (Roche™) according to the manufacturer's instructions. The integrity of pJC6 and pT7-7 was confirmed by restriction analysis *Sal* I and *Hinc* II and analysed by agarose gel electrophoresis (Sambrook, Fritsch, & Maniatis, 1989). The restriction fragments were stained with SYBR® Safe (Invitrogen) and visualized using a UV transilluminator.

2.2.2 Protein expression and isotopic labelling

The pJC6 recombinant plasmid containing the gene that codes for the hexahistidine His-tagged eIF5A$_{Thr}$ protein, and the pT7-7 plasmid constructs were transformed into *E. coli* BL21 (DE3) cells. The transformed cells were incubated into 5 ml of Luria broth containing 125 µg/ml ampicillin and the overnight culture was inoculated into 200 ml of LB and grown at 37 °C with agitation at 200 rpm until an OD_{600nm} of 0.6 – 0.8 was reached. Protein expression was induced at this point by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 µM. Cultures were incubated for 6 hours after induction. The final cell culture suspension was harvested at 4 °C by centrifugation at 10 000 x g for 10 minutes. A 1 OD_{600nm} unit of cell extract was prepared by resuspending pellets in appropriate volume of 0.01 M phosphate buffered saline and 2 x sample buffer (10 % glycerol, 125 mM Tris-Cl pH 6.8, 4 % SDS, 10 % β-mercaptoethanol, 0.01 % bromophenol blue). The samples were boiled for 5 minutes before thawing on ice prior to loading on the gel. Sodium dodecylsulphate polyacrylamide electrophoresis (SDS–PAGE) using a 15 % bis-acrylamide resolving gel and a 4 % stacking gel was used to analyse the aliquots between a 150 volts for approximately 50 minutes. Coomassie Brilliant Blue stain (Laemmli, 1970) was used to visualize the gel. An unstained, US_7 protein molecular weight marker was used to determine the sizes of proteins and confirm the presence of eIF5A$_{Thr}$ during SDS-PAGE gel electrophoresis.

The method for the overexpression of isotope enriched ^{15}N singly labelled eIF5A$_{Thr}$ in *E. coli* BL21 (DE3) cells was followed according to Marley's approach which showed reduced isotope costs and achieved efficient isotope labelling (Marley, Lu, & Bracken, 2001; J. Yuan et al., 2009). For isotopic labelling of eIF5A$_{Thr}$, 2 L of *E. coli* BL21 (DE3) cells transformed with pJC6, were grown in LB containing 125 µg/ml of ampicillin until an OD_{600nm} of 0.6 – 0.8 was reached before they were harvested from using centrifugation at 4 °C under 5,000 x g for 30 minutes. The cell pellets were washed in 1 L of 1 x M9 salt solution (potassium phosphate, sodium phosphate, sodium chloride) containing 125 µg/ml of ampicillin with no nitrogen or carbon sources and incubated for 30 minutes at room temperature with agitation. The cells were harvested by centrifugation at 4 °C at 5,000 x g for 30 minutes. The cell pellet was resuspended in 2 L of ^{15}N isotopically enriched M9 minimal medium (M9 salts, boric acid,

manganese sulphate, zinc sulphate, ammonium molybdate, potassium iodide, copper sulphate, iron chloride, thiamine, pyridoxine, magnesium sulphate, calcium chloride, glucose, and ^{15}N-ammonium chloride (Cambridge Isotope Laboratories)) containing 125 µg/ml ampicillin. Our M9 minimal media contained optimised vitamins and trace elements necessary to support high cell densities and improved expression yields (Marley et al., 2001). Cultures were grown at 37 °C with shaking at 200 rpm for 1 - 2 hours until the cells experienced an OD_{600nm} increase of 0.1 units. Protein expression was induced by the addition of IPTG to a final concentration of 1 µM. Cultures were allowed to grow for 4 hours after induction with OD_{600nm} readings being taken at regular intervals. 1 ml of culture was set aside after each interval reading and the cells were harvested by centrifugation at 10 000 x g for 1 minute. The final cell culture was harvested for protein purification after 4 hours of incubation in the presence of IPTG.

2.2.3 Purification of his-tagged eIF5A$_{Thr}$

Cells were harvested from the 2 L of culture by centrifugation under 10 000 x g at 4 °C for 10 minutes. The supernatant was carefully decanted and the pellets were resuspended in 50 ml of lysis buffer (500 mM sodium chloride, 40 mM imidazole, 1 x proteinase inhibitor cocktail tablet (Roche), 10 mg of RNase A (Roche), 50 µg/ml of lysozyme enzyme in 50 ml of 20 mM phosphate buffer at pH 7.4). The sample was incubated at 37 °C for 10 minutes before the cells were subjected to sonication on ice at 4 °C using a hand-held 3 mm probe (Vibra Cell ^{TM}) at 70 mA for a total of 20 minutes with 10 second intervals. The cell debris was pelleted by centrifugation of the cell lysate at 4 °C under 5, 000 x g for 30 minutes. The supernatant was filtered through a sterile 0.2 µm syringe filter (Pall Life Sciences) and labelled as clear lysate. The pellet was resuspended in phosphate buffered saline (sodium chloride, potassium chloride, sodium phosphate, potassium phosphate, pH 7.4). The cell lysate was purified using nickel affinity chromatography through a 5 ml His-trapTM FF IMAC column (GE Healthcare) over a 0 – 100 % gradient of elution buffer (500 mM sodium chloride, 500 mM imidazole in 20 mM phosphate buffer at pH 7.4) using a Fast Performance Liquid Chromatography (Äkta FPLC^{TM} GE Healthcare) pump system at a flow rate of 5 ml/min. Protein elution was monitored by measuring UV absorbance at 280 nm and 5 ml fraction were collected. 5ul of each fraction was monitored through SDS-PAGE analysis and the

fractions containing protein were grouped after validation by SDS-PAGE analysis as described in 2.2.2. Fractions that contained eIF5A$_{Thr}$ were pooled together and dialysed against NMR sample buffer (1 mM of dithiothreitol and 100 mM sodium phosphate buffer at pH 6.8) using Amicon Ultr-10 Ultracentrifugal filters (Millipore). Contaminating protein was removed by using Ultr-30 filters before the pure protein sample was dialysed and concentrated into a 150 µl aliquot of 10 % D_2O NMR sample buffer. Protein concentrations were determined using the NanoDrop^{TM} 2000 (Thermo Scientific) with adjusted molecular weight (18.3 kDa) and extinction co-efficient (2980 M^{-1} cm^{-1}) factors obtained for the ExPASy Proteomics server.

2.2.4 Thrombin cleavage

Restriction grade human thrombin (Novagen^{TM}) at a concentration of 1 U per mg of purified eIF5A$_{Thr}$) was used to cleave 50 µl of purified protein (0.25 mM) in cleavage buffer (150 mM NaCl, 2.5 mM $CaCl_2$ in 20 mM phosphate buffer at pH 8.4) overnight (17 hours) after agitation. SDS – PAGE was used to analyse cleavage of the protein.

2.2.5 NMR analysis

In order to perform efficient heteronuclear single quantum correlation, HSQC, NMR experiments, 50 µl of 1.05 µg/µl of pure protein in 10% deuterated NMR sample buffer was concentrated to 10 – 20 µl under a vacuum at 30 ^{o}C. The protein sample was transferred to a 1 mm NMR tube. After changing the probe-head the pre-amp to amplifier connections were checked to ensure that they matched the nuclei to be investigated before performing temperature stability checks. Deuterium was chosen as the solvent lock during acquisitions and the probe-head was tuned and matched appropriately for uniform magnetization. Preparation of the system was completed by reading a preset optimized shim file.

Sample validation was achieved through a short proton acquisition with water suppression achieved through a phase sensitive, spin echo type, WATERGATE pulse sequence within a double pulse field gradient (P16), illustrated in Appendix D. Presaturation of all protons is attained through a non-selective 90^{o} pulse (P1). The first of the symmetric pulsed field gradients (P16) dephases all saturated resonances. After the delay for gradient recovery (D16) lapses the WATERGATE multiple pulse series, which consists of six 90^{o} pulses, is initiated. During gradient recovery the solute and solvent proton spins precess at different

rates influenced by their chemical shifts. When the first 90^{o} pulse of the series is initiated all the spins are rotated by 90^{o}. Within the pulse series the pulse lengths (P27) are scaled by a factor of 3, 9, 19, 19, 9 and 3 where the underlined numbers are pulses that have their phases inverted (ph 4). Each pulse in this series, however, is separated by twice the delay time for binomial water suppression (D19). This means that at the end of the pulse train the water signals are dephased efficiently while the solute spins are inverted. Although some of the solute spins at this point are unsaturated after the last 90^{o} inverted pulse the majority of them can be refocused by the second pulsed field gradient (P16). The second pulsed field gradient only refocuses the solute spin signals for detection while the majority of water proton spins are unperturbed and remain dephased. Detection on the proton channel occurs 4 µs after the delay for gradient recovery (D16). Despite saturation exchange effects, residual water signals due to poor shimming and the presence of spin diffusion within the solute protons, a water suppressed proton spectrum allows the detection of low concentration solute species within the vicinity of the strong water resonance (Cavanagh et al., 2007; Rule & Hitchens, 2006).

Prior to starting the HSQC acquisition, careful interactive optimizations of solvent dependant parameters were performed by minimizing the FID. Predetermined acquisition parameters were used with the receiver gain while the numbers of scans were customised to suit the protein concentration. It is well known that whenever water suppression is achieved through pre-saturation, decrease in sensitivity for the solute occurs. If the solute contains exchangeable protons, the suppressed solvent protons, which are no longer NMR active, can exchange with these resulting in loss of protein signals (Carbajo & Neira, 2013b; Cavanagh et al., 2007). We opted to analyse our labelled protein using a carbon decoupled, ^{15}N-edited, HSQC which incorporates a WATERGRATE water suppression scheme within its pulse sequence as described in Appendix 2.4. This HSQC pulse sequence, refocuses the water magnetization with a, flip back pulse, before initiating the water suppression prior to detection acquisition. This improves sensitivity by minimizing gradient saturation of water during WATERGATE and thus decreases the chances of saturation exchange upon detection (Mori et al., 1995). The effects of preparation and water suppression are described in detail within the results section below. While pulse programs can be verified from the Appendix.

2.3 RESULTS AND DISCUSSION

The objective of the experiments described in this chapter was to produce ^{15}N labelled $eIF5A_{Thr}$ protein that could be used to generate the N – terminal fragment suitable for biomolecular NMR studies. In order to do this it was necessary to validate the pJC6 plasmid, encoding the $eIF54_{Thr}$, optimize expression of protein in complete medium and formulate a minimal medium broth that would allow the achievement of at least 95 % of ^{15}N and ^{13}C isotope enrichment of protein. The thrombin cleavage site (LeuValProArg↓GlySer) introduced within the flexible linker region of eIF5A enabled the separation of the N-terminal fragments of $eIF5A_{Thr}$ for structural studies.

2.3.1 pJC6 construct validation

The integrity of the pJC6 expression vector was confirmed by restriction endonuclease analysis using the enzymes which digest the vector at specific sites. The results are illustrated in Figure 2-2.

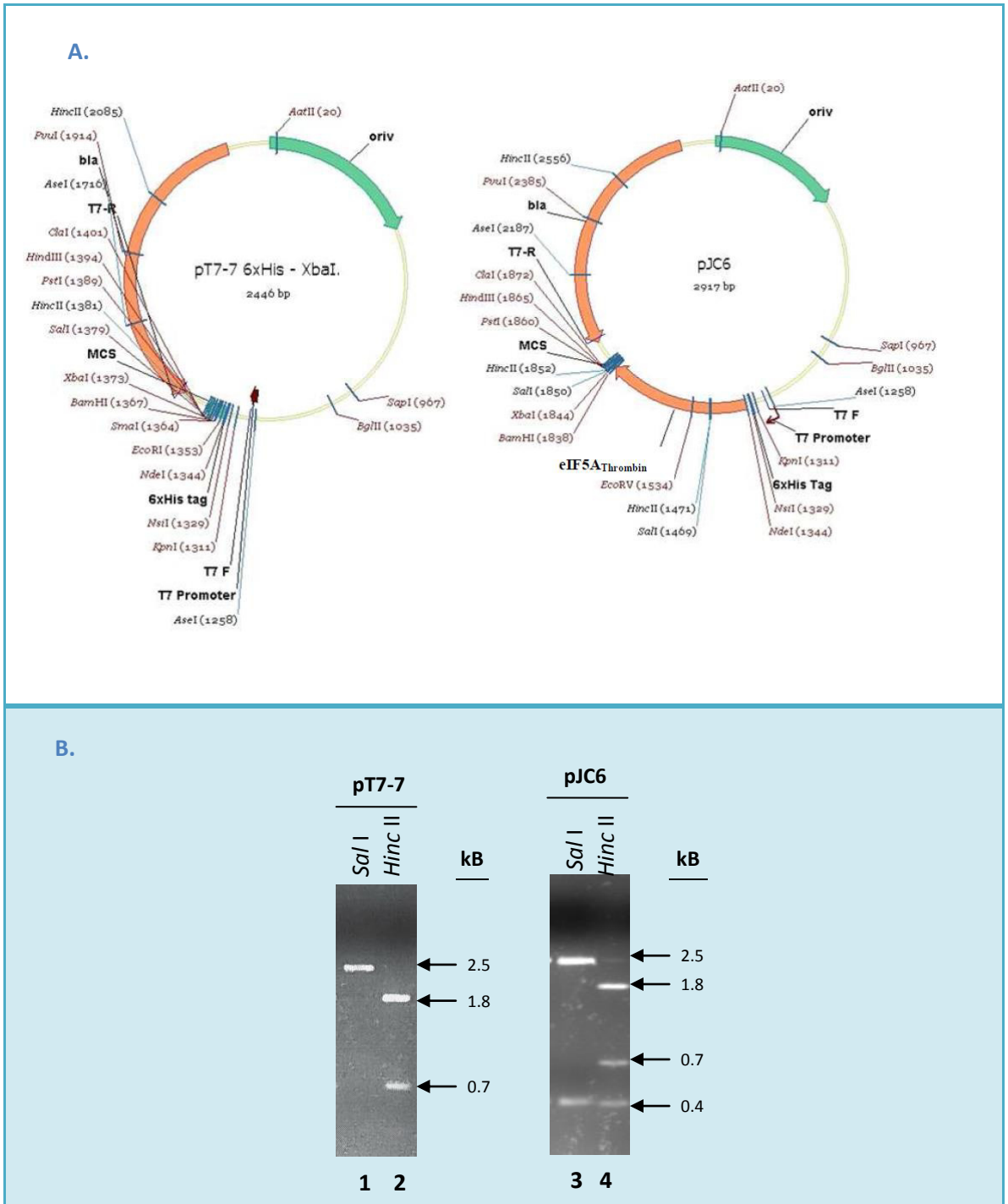


Figure 2-2: Validation of expression plasmid by restriction analysis. Panel A: Restriction maps of bacterial expression vectors provided by Charlton (2012) and used in production of eIF5A$_{Thr}$ in *E. coli* cells. Bla: β-lactanase, T7 Promoter: T7 bacteriophage RNA polymerase promoter, MCS: multiple cloning site, eIF5A$_{Thr}$: *TIF51A* gene with LVPRGS thrombin cleavage site. Panel B: Restriction analysis of expression vectors using 1 % agarose gel electrophoresis. The size of the bands corresponding to the expected molecular weights are shown with arrows on the right of the gel. Enzymes used for digestion are indicated above the lanes. *Hinc* II digestion was in buffer B while *Sal* I digestion was in buffer D. Samples were incubated at 37 °C for 1 hour to allow complete digestion.

The pT7-7 vector plasmid has a single *Sal* I site and 2 *Hinc* II sites. Digestion with *Sal* I produced a single fragment migrating at 2500 bp (Figure 2-2 B, Lane 1). Digestion with *Hinc* II produced 2 fragments 700 bp and 1746 bp (Figure 2-2B, Lane 2). The expression vector pJC6 gave rise to 2 fragments of 2 500 bp and 400 bp when *Sal* I was used (Figure 2-3 B, Lane 3). When *Hinc* II was used to digest pJC6, 3 bands of 381 bp, 704 bp and 1832 bp which correspond to the predicted sizes from the restriction map (Figure 2-2A) were obtained. This confirmed that pJC6 contained the *TIF51A* coding for the eIF5A$_{Thr}$ protein.

2.3.2 Protein Expression

First it was necessary to confirm that pJC6 produced eIF5A$_{Thr}$ protein. This was done by transforming pJC6 into *E. coli* BL21 *E. coli* cells and inducing expression from the T7 promoter in the presence of IPTG. The optimum conditions for expression of eIF5A$_{Thr}$ were determined for cells growing in LB broth (complete medium) and minimal medium.

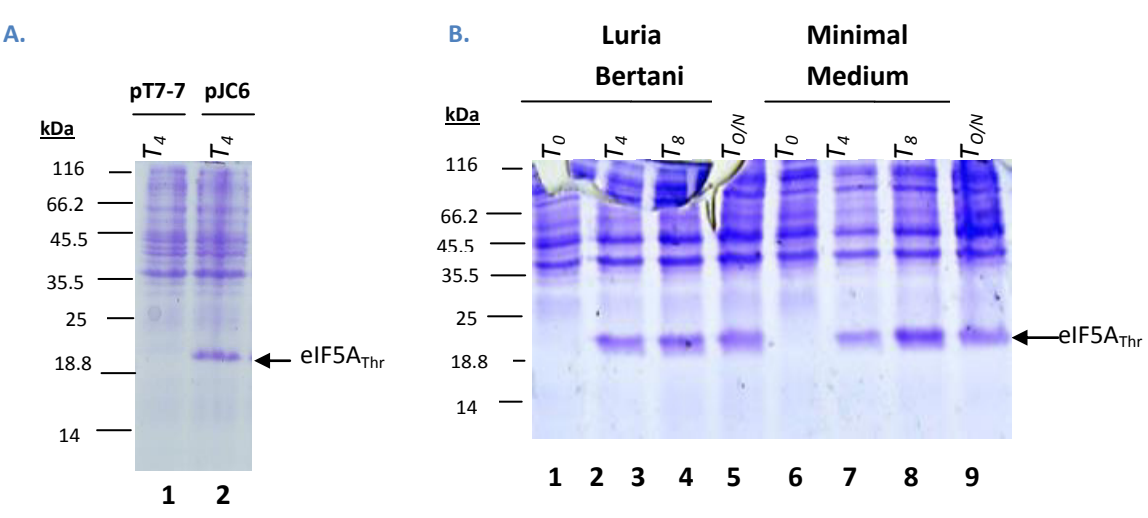


Figure 2-3: Analysis of heterologous protein expression of eIF5AThr from transformed *E. coli* cells using 15% SDS-PAGE and staining with coomassie. Panel A: Control expression experiment. *E. coli* transformed with PT7-7 (Lane 1 = eIF5A$_{Thr}^{-}$) and pJC6 (Lane 2 = eIF5A$_{Thr}^{+}$). Induction allowed to proceed for 4 hours. Panel B: Time course induction. Expression in ^{14}N Luria Bertani broth and ^{15}N minimal media broth. T_0 = 0 hours with respect to induction with IPTG. $T_{O/N}$ = induction allowed to proceed overnight after induction with IPTG.

After 4 hours of induction a protein band with a relative molecular weight of 21 kDa was observed in cells transformed with pJC6 for the (Figure 2-3 A, Lane 2). The size of this protein is consistent with that observed by Charton (2012). Since the 21 kDa band was not

present in cells transformed with the vector, pT7-7 (Figure 2-3 A, lane 1) it was concluded that pJC6 expressed 6xhis- eIF5A$_{Thr.}$ Expression experiments showed that expression in isotopic minimal medium was comparable to that observed in the LB broth (Figure 2-3 B).

^{15}N-labelling of 6xhis-eIF5A$_{Thr}$

The next step was to produce ^{15}N-labelled 6xhis-eIF5A$_{Thr}$. This was done by first growing up a sufficient amount of biomass (growth in LB medium to an OD_{600nm} of between 0.6 – 0.8.). This was followed by a transfer of the cells to a minimal medium containing $^{15}NH_4Cl$ after cell washing (Figure 2-4 Panel A, experiment T_{-2}, Lane 1).

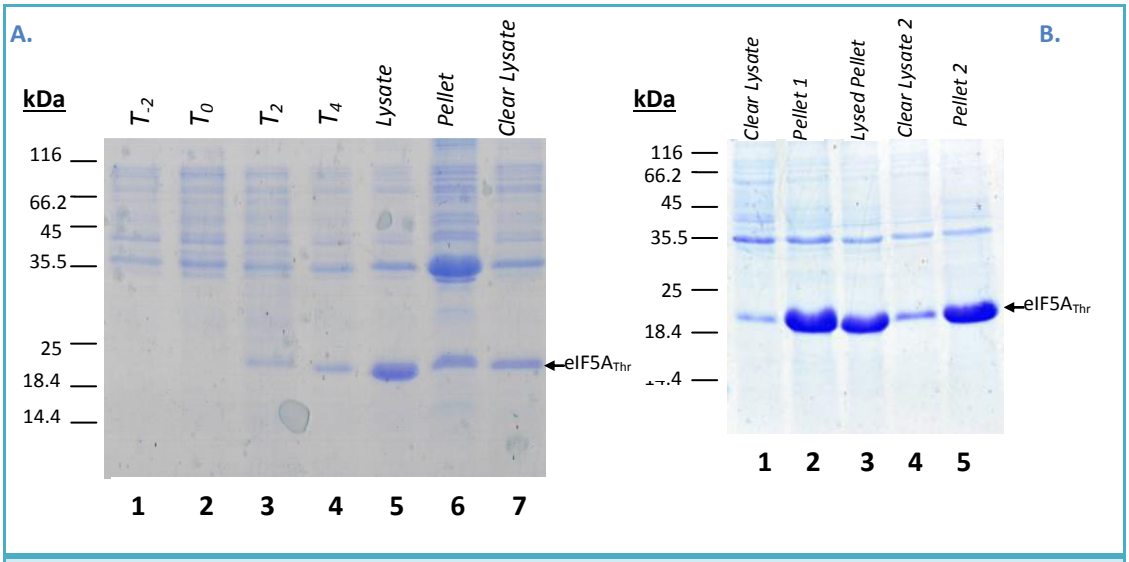


Figure 2-4: Overexpression of eIF5A$_{Thr}$ grown in ^{15}N- labelled minimal media analysed on 15% SDS PAGE. Panel A: Overexpression time course and lysis. Lane 1 = T_{-2} (2 hours prior to induction. Start of growth recovery stage); Lane 2 = T_0 (0 hours after induction. Induction point after growth recovery stage.); Lane 3 = T_2; Lane 4 = T_4; Lane 5 = Lysate; Lane 6 = Pellet; Lane 7 = Clear Lysate. Lane 8 – 9 = 15% SDS PAGE Gel of 8 hr IPTG induction of pT7-7(Lane 8) control and pJC6 (Lane 9) transformed BL21 *E. coli* cells. Panel B: Extended Lysis of pellet (Panel A, Lane 6). Pellet 1 is Pellet fraction from Panel A lane 6. Clear Lysate 2 was then added to Clear Lysate (Panel A, Lane 7).

T_0 denotes the cell culture in minimal media at the point of induction with IPTG. After induction of eIF5A$_{Thr}$ in minimal media, the protein band was first observed 2 hours (T_2) post induction with IPTG, Lane 3, and the concentration increased up to 4 hours (T_4) when the cells were harvested, Lane 4. There is an increase in the concentration of eIF5A$_{Thr}$ between 2

hours and 4 hours of induction, Lane 3 and Lane 4, without an increase in concentration of cellular proteins. This is consistent with the increased metabolic load placed on cells during overexpression (Sahdev, Khattar, & Saini, 2008). After harvesting, the cell pellets were lysed using lysozyme enzyme and pulsed on ice. The lysate, Lane 5, was analysed by SDS-PAGE before the soluble protein was separated from cell debris by centrifugation. The soluble cytosolic eIF5A$_{Thr}$ was observed to be present in the clear lysate supernatant fraction, Lane 7. A substantial amount of expressed eIF5A$_{Thr}$ protein was present in the pellet fraction, Lane 6. We proposed that this was due to *in vivo* disordered aggregates of protein which formed insoluble inclusion bodies as well as due to incomplete cell lysis (Upadhyay, Murmu, Singh, & Panda, 2012). Previous studies of N-terminal hexahistidine tags have shown that they have a protein specific negative impact on protein solubility which enhances intracellular misfolding and aggregation (Woestenenk et al., 2004). After resuspending pellet in lysis buffer and repeating lysis, eIF5A$_{Thr}$ present in the cell debris fraction, pellet, due to incomplete lysis was extracted out, Figure 2-4 Panel B. From the lysed pellet, Lane 3 after lysis the clear lysate obtained, Lane 4, contained eIF5A$_{Thr}$ protein.

The 6xhis-eIF5A$_{Thr}$ was purified using nickel affinity chromatography. A small scale purification was done on the unlabelled protein. SDS-PAGE analysis showed the presence of the 21 kDa 6xhis-eIF5A$_{thr}$ protein in the soluble fraction of the cell free lysate (Figure 2-5 A, lane 1) and the absence of the 21 kDa band in the flow-through from the column indicated binding of the protein to the Nickel beads (Figure 2-5 A, lane 2). The 21 kDa protein was absent in the wash steps because the 6xhis-eIF5A$_{Thr}$ binds to the column strongly to the column (Figure 2-5, lanes 3 and 4). The addition of 500 mM imidazole, which is a competitor for binding to Ni^{2+} resulted in the elution of 6xhis-eIF5A$_{Thr}$ (Figure 2-5 A, lane 5). Also present in the eluted fractions were some proteins of higher molecular weight, suggesting the presence of contaminating *E. coli* proteins.

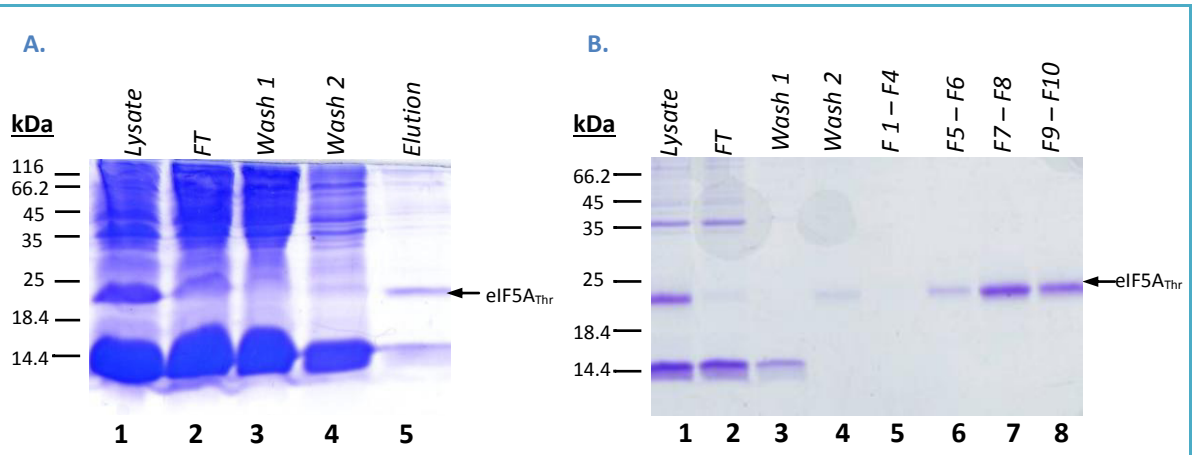


Figure 2- 5: SDS-PAGE of purification of eIF5A$_{Thr}$ using nickel affinity chromatography. Panel A: SDS-PAGE analysis of protein fractions from small scale purification of unlabelled eIF5A$_{Thr}$ using His-trap column. Lysate (Lane 1) is the cell debris free clear lysate from lysis stage, FT (Lane 2) is flow through obtained from loading onto column, Wash 1 and Wash 2 (Lane 3 and 4) are obtained from wash steps using binding buffer with low imidazole, Elution (Lane 5) is obtained from washing with elution binding with high imidazole content. Panel B: SDS-PAGE of protein fractions from gradient elution purification of labelled eIF5A$_{Thr}$ using His-trap column. Lanes 1 – 4 are as Panel A; Lanes 5 – 8 are fractions obtained from gradient elution, where F1 = 5 ml Fraction 1 after initiation of elution gradient.

Having confirmed that 6xhis-eIF5A$_{thr}$ could be purified by nickel affinity chromatography, the large scale purification of the ^{15}N-labelled protein was performed. The system utilizes a gradient elution with imidazole rather than the step-wise protocol used for the small scale protocol, which reduces the co-elution of contaminating *E.* coli proteins with the His-tagged protein of interest. SDS-PAGE analysis of the purification fractions showed the presence of the 21 kDa ^{15}N-labelled 6xhis-eIF5A$_{Thr}$ in the soluble fraction of the cell free lysate (Figure 2-5 B, lane 6–8). Some protein was observed in the flow through (Figure 2-5 B, Lane 2), indicating that not all the protein was bound to the column. A small amount of 6xhis-eIF5A$_{Thr}$ was also observed in the wash fractions (Figure 2-5, lanes 4) but overall, the majority of the protein remained bound to the Ni-affinity beads in the column. Addition of the imidazole gradient resulted in elution of the 6xhis-eIF5A$_{Thr}$ (Figure 2-5 B, lanes 6, 7 and 8, representing fractions 5 to 10).

The UV active component of eluent consisting of imidazole and protein was traced by absorbance at 280 nm during the automated purification protocol.

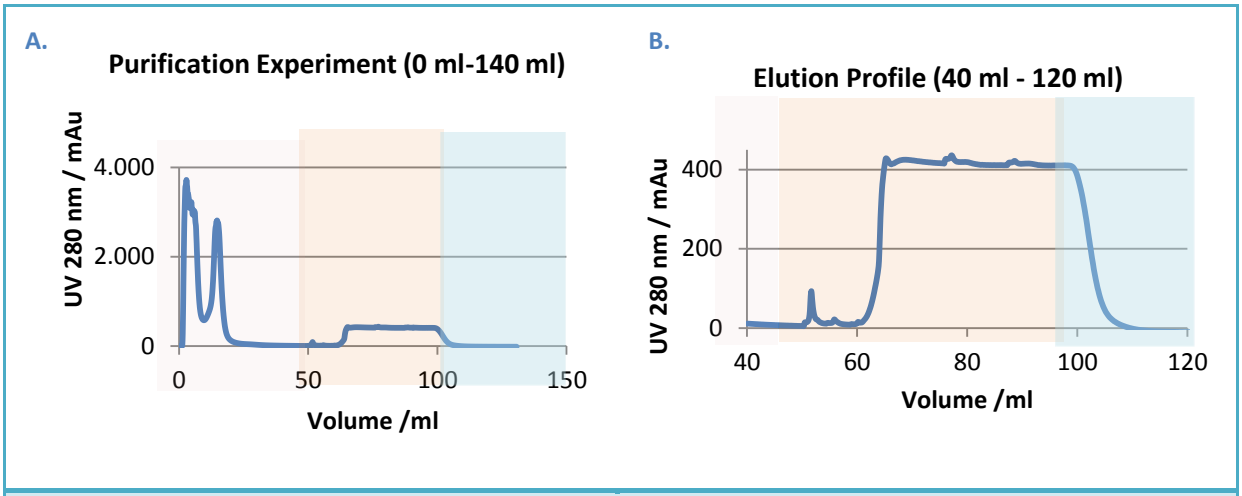


Figure 2-6: UV trace of purification profile. Pink blocks describe the low imidazole wash steps between 0 and 50 ml. A spike at 55 ml is consistent with non-specific bindings eluting at a high imidazole concentration. Orange describes the elution stage using a 0 -100 % of elution buffer containing 500 mM of imidazole gradient over 50 ml while the blue describes the 100 % elution buffer recovery wash steps. Panel A: UV trace over over entire experiment Panel B: High resolution elution stage of purification protocol.

Between 0 and 25 ml of elution volume significant absorbtion at 280 nm is observed (Figure 2-6 A). This is consistent with absorbtion at 280 nm due to unbound proteins in the eluent during the flowthrough and first wash stages (Figure 2-5 B, lane 2 and 3). A spike at 55 ml during the early gradient elution stages is consistent with proteins with weak and non-specific binding interactions with the column (Figure 2-6 B) and were confirmed to be 6xhis eIF5A$_{Thr}$ protein by SDS-PAGE (Figure 2-5 B, lane 4). UV active components obtained after 60 ml were shown to contain protein, Fraction 5 – 10 (Figure 2-6 B). No protein was observed after 80 ml of eluent although significant absorbtion was recorded consistent with the increased imidazole content in recovery wash stages.

A total of 0.21 mg of ^{15}N-lablled protein was obtained in 200 µl of sample buffer after buffer exchange removed the imidazole and 30 kDa molecular cut-off filters were used for the removal of aggregated protein. An aliquot of 50 µl of protein in 10 % D_2O NMR sample buffer containing DTT to avoid aggregation was concentrated to 10 µl under vacuum and gentle heating for preparation of the NMR sample. The remainder was used for thrombin cleavage experiments.

2.3.3 Thrombin cleavage of 6xhis-eIF5A$_{Thr}$

As discussed earlier, one of the aims of this project was to use thrombin cleavage to produce to smaller eIF5A peptides for NMR studies. Accordingly, purified 6xhis-eIF5A$_{thr}$ was treated with thrombin protease. SDS-PAGE analysis showed cleavage of the 21 kDa 6xhis-eIF5A$_{thr}$ protein into two peptides with a relative molecular weight of 10.35 kDa and 8.03 kDa respectively (Figure 2.7, compare lane 1 with lane 2).

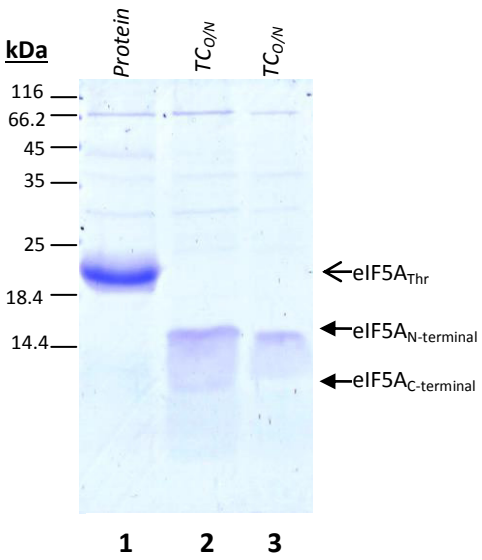


Figure 2-7: SDS-PAGE analysis of the products of thrombin cleavage of 6xhis-eIF5A$_{thr}$ for the cleavage reaction. Lane 1 = purified protein with no thrombin in sodium phosphate buffer; Lane 2 and Lane 3 are duplicate overnight cleavage experiments

2.3.4 NMR Analysis

About 52 µg of protein in 10 µl gave a concentration of about 0.25 mM of protein. NMR water suppression and HSQC experiments are described below.

Water Suppression

The solvent used for the experiments was 90% H_2O/10% D_2O in order to prevent deuterium exchange and had a large signal relative to the protein. Efficient suppression of these water signals was necessary in order to mitigate signal broadening allowing detection of low concentration solutes.

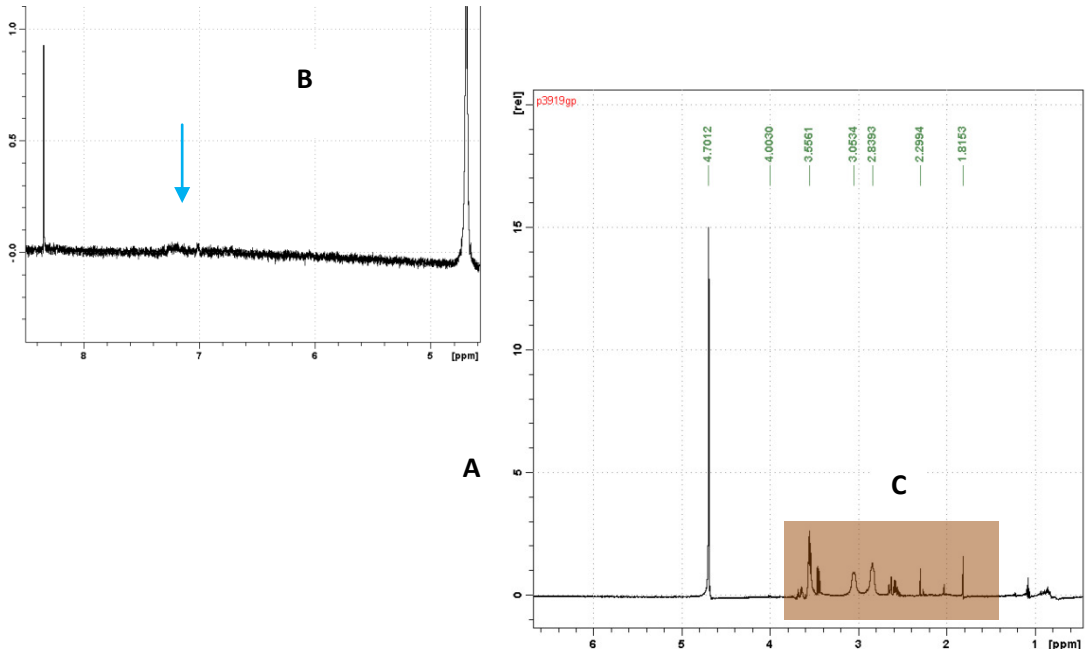


Figure 2-8: WATERGATE ^1H water suppression at o1p of 4.701 ppm. Panel A: Water resonance at 4.7012 ppm. Panel B: Downfield deshielded resonances within the protein amide bond regions. Insert C: Upfield shielded resonances of aliphatic protons from traces of buffer and dithiothreitol surfactant.

Figure 2-8 shows the spectrum obtained from water suppression through a WATERGATE pulse sequence. Insert A shows the entire spectrum of the protein with some residual water signal. The downfield region, Insert B, is expected to contain the more deshielded amide proton signals. Because larger molecules have larger viscosities their correlation times decrease resulting in a decreased spin-spin relaxation and more efficient relaxation

mechanisms. The small broad signals observed downfield of the water resonance are indicative of low concentrations of large sized solutes confirming the presence of protein in solution. The upfield region, Insert C, under close inspection shows three signals at 3.56, 3.05 and 2.83 ppm of relatively low intensity. These signals are consistent with the presence of aliphatic solutes. The reduced dithiothreitol added to enhance solubility of eIF5A$_{Thr}$ as well as traces of Tris-buffer could contribute to the unresolved signals in this region. When the HSQC experiments are performed no transfer of magnetization through labelled ^{15}N from these solutes confirms that they are not associated with the protein.

^{15}N-edited HSQC

The decoupled ^{15}N-HSQC was run over 54 hours and the data processed at the console using the *fib* Topspin command. The spectrum was phased manually and is illustrated in Figure 2-9.

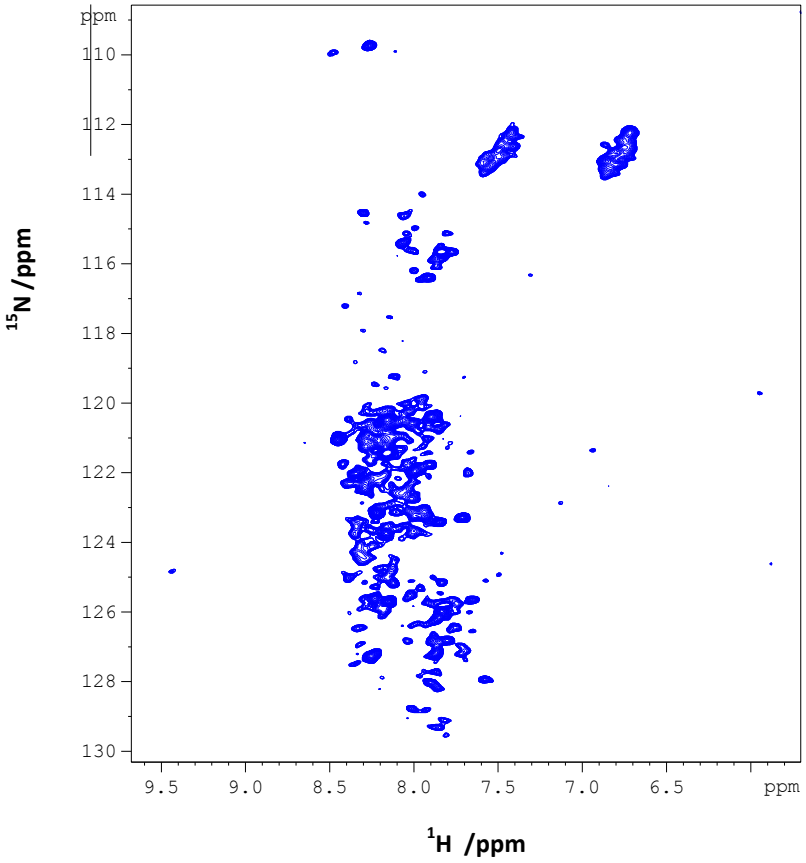


Figure 2-9: ^{1}H-^{15}N HSQC spectrum of eIF5A$_{Thr}$ in 10% D_2O obtained from a 600 MHz spectrometer.

The F1 dimension describes the nitrogen axis and the spectrum is described by the clear box. The solid box F2 dimension describes the proton axis and corresponding spectra. Interactive phasing was done in order to display the modulated resonances of the protein from the background noise signals with the relative intensity identified from the contour map. Topspin peak identifier identified 128 cross peaks within our 152 amino acid protein when we expected to see 146 cross peaks from an ^{15}N-edited HSQC spectrum. Intermediate conformational chemical exchange, signal overlap and saturation exchange with water could account for the 12% loss of signal.

The backbone amides proton peaks are compressed within the 6 – 8 ppm range consistent with random coils . This implies that the eIF5A$_{Thr}$ produced using the protocol we described was unfolded (P. Rossi et al., 2010). This observation is consistent with the formation of inclusion bodies which form readily in the presence of unfolded proteins. It was shown by Sahdev et al., (2005) that growth of proteins expressed in *E. coli* at lower temperatures reduced the toxic effects observed at 37 °C resulting in increased expression yields and protein activities. The reduced toxicity was primarily due to the decreased inclusion body formation which is temperature dependant. Higher yields were explained by the fact that some of the chaperones necessary for protein folding experience an increase in their expression at lower temperatures thus increasing protein stability and promoting the correct folding patterns. Degradation losses are also reduced at low temperatures because the proteases are less active at lower temperatures (Sahdev et al., 2008). In a study to generate backbone and side chain assignments of human eIF5A, induction at 25 °C gave high yields of correctly folded protein (J. Yuan et al., 2009). When the activity of the over expressed protein was investigated, induction of the hypusinated protein from a polycistronic vector that contained the modifying enzymes showed similar temperature dependence. At 37 °C rapid expression was observed but with low hypusine modification probably due to misfolding of the eIF5A or the modifying enzymes. At 18 °C however the rate of synthesis and modification were reduced due to a decreased metabolic rate (J. H. Park et al., 2011).

Folding is anticipated to occur within the NMR sample tube during the acquisition of NMR data. However because of the presence of the surfactant the rate at which this does so is slower.

2.4 CONCLUSION

Overexpression was performed in ^{15}N labelled M9 minimal media before the cells were harvested, lysed and purified using Ni^{2+} affinity chromatography. The protein was concentrated and analysed on a 600 MHz NMR spectrometer. Expressing the recombinant protein in minimal media using 99 % ^{15}N – ammonium chloride as the sole nitrogen source allowed us to achieve sufficient levels of labelling necessary for heteronuclear NMR structure studies. From 2 L of cell culture we obtained 0.21 mg of soluble pure ^{15}N-labelled protein concentrated to 200 µl. Charlton (2012) obtained 12 mg of unlabelled protein from 200 ml of *E. coli* while we were only able to obtain 2.36 mg prior to concentration and buffer exchange.

From the HSQC chemical shift peak pattern the protein that was expressed using our protocol appeared to be unfolded. Although folded eIF5A is produced routinely in *E. coli* BL21 cells, the influence of the insertion of the LVPRGS cleavage site on the global dynamics and folding of eIF5A in solution could not been evaluated in this study. In the case where the global folding of eIF5A is not influenced by the cleavage site, the expression protocol would have to be modified for later investigations. However if it is found that the insertion of the cleavage site does indeed have an effect on the global folding nature of eIF5A, alternative approaches towards the study of eIF5A without the cleavage site on a 600 MHz NMR would need to be investigated. These approaches would include longer acquisition times and the use of Residual dipolar coupling, lanthanide induced pseudo contact shifts, deuteration and manipulation of relaxation dispersion.

Chapter 3: Homology Modelling and Molecular Dynamics

3.1 INTRODUCTION

3.2 MATERIALS AND METHODS

3.2.1 Homology modelling

3.2.2 Molecular Dynamics

3.3 RESULTS AND DISCUSSION

3.3.1 Homology Modelling

3.3.2 Molecular Dynamics

3.4 CONCLUSION

Chapter 3: Homology Modelling and Molecular Dynamics

3.1 INTRODUCTION

Given the limited experimental NMR data available for analysis, computational strategies including molecular dynamics were employed to understand the influence of the cleavage site on the eIF5A structure and solution dynamics. In order to do this it was envisaged to conduct an *In silico* approach towards the investigation of the influence of the insertion of the 6 residue cleavage site on the overall structure integrity and solution dynamics of unhypusinated eIF5A monomer.

Ever since work by Rossmann and Blow confirmed the hypothesis that many large protein molecules are made up of similar subunits, extensive work has since been done to solve protein structures of biomolecules based on their sequence identity and similarity (Ramachandran et al., 1963; Rossmann & Blow, 1962; Xiang, 2006). Fast computational techniques that are used to generate predicted protein structures employ 2 main approaches towards selecting successfully constructed models. Comparative sequence homology based modelling techniques and *ab initio* first principle based techniques are generally used in concert with each other. Comparative homology modelling uses sequence homology from experimentally solved structures (such as obtained from x-ray structures) to guide structure determination while a*b initio* techniques rely on satisfying energy restraints of a force-field that obtains a global minimum structure by searching through the conformational landscape available (P. G. Wolynes et al., 2012; Xiang, 2006; Xiong, 2006). Improvements in accuracy of *ab initio* structures from proteins that do not have experimentally determined homologues can be achieved through the incorporation of experimental restraints from circular dichroism and NMR to guide structure building (Montalvao et al., 2012). We aim to define a predicted structure of the mutated eIF5A with the thrombin cleavage domain using homology modelling approaches, rather than an *ab initio* approach.

A description of the means by which protein sequences, after synthesis, adopt their native low energy folded conformation is crucial to the understanding of protein function (Wolynes

et al., 2012). Due to the large number of degrees of freedom available to intermediates along the folding route, Levinthal proposed that protein folding was directed and followed a well-defined pathway which lead to its rapid folding to the native state (Zwanzig et al., 1992). However experimental investigations into protein folding and unfolding showed that different folding scenarios exist which direct folding mechanisms (Dobson, Šali, & Karplus, 1998). The pathway adopted is directed by the amino acid sequence and satisfies thermodynamic and kinetic requirements. From experimental investigations it was found that protein folding may be described as following an energy landscape as opposed to a folding pathway. The trajectories followed for a particular sequence are independent and in two instances of the same protein with identical sequences do not necessarily follow the same route in order to adopt their identical native state (Chan & Dill, 1998; Rimratchada, McLeish, Radford, & Paci, 2014). A particular sequence has an energy landscape that describes the general routes that are followed by its proteins to access their native state (Rimratchada et al., 2014).

Although it is generally accepted that the folding routes that proteins follow often involve numerous factors (such as chaperones) that stabilise seemingly inaccessible conformations, homology modelling has been able to accurately predict native protein structures to comparably high resolutions (Fiser, Do, & Sali, 2000b; Gershenson & Gierasch, 2011; Hartl, Bracher, & Hayer-Hartl, 2011). This is made possible by the fact that the overall tertiary structure of a protein is determined by its sequence (Xiang, 2006). Although there are numerous permutations of conformation that a protein can adopt in nature, a limited number of families of folds are allowed exist (Dobson et al., 1998; Lewis et al., 2013; Pieper et al., 2011). In this Chapter we aim to use molecular dynamics to simulate the solution behaviour of EIF5A protein models and relate the introduction of rigidity into the flexible domain in $EIF5A_{Thr}$ with its solution dynamics.

The principle of molecular dynamics (MD) simulations is that one can employ a particular numerical integration algorithm of Newton's equation of motion in small time steps for a defined inter-atomic potential to generate atomic trajectories from a system of interest within certain conditions (Berendsen, van der Spoel, & van Drunen, 1995; J. Li, 2005). For our case the system was set-up, parameters defined and the sample prepared within the

Groningen Machine for Chemical Simulations (GROMACS) package before running the simulation.

GROMACS is a flexible molecular dynamics program that implements a parallel message-passing method to improve speed and efficiency by allowing inter-processor communication (Berendsen et al., 1995; Pronk et al., 2013a). Its functionality is based on the GROMOS package and allows the rapid simulation of bio(macro)molecules in solution on single workstations, parallel servers and processors that utilize multithreading (Hess et al., 2008; Lindahl et al., 2001). Because GROMACS to date has a lot of algorithmic optimizations and is fast at calculating non-bonded interactions its applications have extended from biological systems such as proteins, lipids and nucleic acids to non-biological polymeric systems (Pronk et al., 2013b). GROMACS allows investigations into protein – ligand interactions based on electrostatics, free energy calculations related to solvation, membrane simulations of proteins within lipid bi-layers as well as QM/MM *ab initio* simulations. We implemented MD simulations in order to generate trajectories which were used to predict time averaged NMR spectra.

3.2 MATERIALS AND METHODS

Based on the experimental HSQC cross peak pattern observed it was hypothesized that eIF5A$_{Thr}$ had adopted an unfolded conformation in solution with loss of tertiary and perhaps some secondary structure. To test this hypothesis we generated models of PG57 eIF5A$_{THR}$ in a fully unfolded denatured conformation and a homology model in its folded conformation. We evaluated these structures and simulated their molecular dynamics over 1ns. Their trajectories were used to generate time averaged backbone chain ^{15}N-HSQC spectra using a structure-based protein chemical shift predictor CamShift (Kohlhoff et al., 2009). Perl scripts were written in order to efficiently process the trajectories produced from GROMACS with CamShift calculations in order to analyse the large amounts of data obtained.

3.2.1 Homology modelling

The four basic steps of a concerted homology modelling protocol involve template identification, sequence alignment, model building and the incorporation of *ab initio* algorithms in refinement and validation.

Template identification

Template identification is the process by which the sequence of an already experimentally solved structure is identified as a suitable template from query sequences of the protein to model, which is searched for within protein databanks (Blundell et al., 2006). Generally a satisfactory template is selected as having the highest percentage identity to the query sequence, or that which bears a structure that has the highest resolution. Depending on the application, the ideal structure may contain co-crystallized cofactors or ligand structures (Blundell et al., 2006; Özlem Tastan Bishop, de Beer, & Joubert, 2008).

Heuristic search engines help to survey the protein database and identify the most suitable templates rapidly and accurately. The most common programs employed use variants of the BLAST computational algorithms (Oladele, Bamigbola, & Bewaji, 2009). Basic local alignment search tools, (BLAST), based search tools are database search tools that use local alignment algorithms that are capable of detecting weak but biologically significant sequence similarities and optimise speed. In our approach we used the BLAST algorithm to search the RCSB PDB using the sequence of eIF5A$_{THR}$ from the PG57 plasmid insert prepared by Gentz (2008) as query sequences. From this database search, 6 sequences were retrieved that showed high sequence similarities with our PG57 query sequence. The sequence similarities and identity matrix were calculated using an all-against-all matrix from BioEdit and the sequence similarities obtained with BLAST and BioEdit were compatible.

Alignment

After the most similar sequences have been identified the query sequence is then aligned before the model is built. Alignment can either be based on structural homologies in distantly related homologues or it can be based on multiple sequence homologies (Özlem Tastan Bishop et al., 2008). In instances where sequence homologies exist, versions of Clustal, MUSCLE and TCoffee are normally used for alignment. They accomplish accurate multiple sequence alignment of sequences with varying degress of computational cost and complexity (Edgar, 2004; Larkin et al., 2007). FUGUE and 3DCoffee are alternative programs that manipulate the fact that structure is more conserved than sequence is. They are able to create position-specific substitution matrices that relate a particular sequence with a compatible fold in distantly related homologues (O'Sullivan, Suhre, Abergel, Higgins, & Notredame, 2004).

A FASTA file of the sequences retrieved by the BLAST database search was prepared before submitting the file for alignment. We performed a sequence homology based multiple sequence alignment using the ClustalW2 online server (EMBL-EBI, 2013). The alignment was done without iterations maintaining the default gap opening penalty at 10.0 and a gap extension penalty of 0.2. The GONNET Matrix was used to measure the differences among the amino acid types because they had strong evolutionary relationships.

Model building

Transfer of information from the query sequence to the target sequence occurs through the process of model building after the sequences have been aligned. The four main methods employed in model building are rigid body assembly (SCHRODINGER, SWISS-MODEL and PrISM), segment matching (SEGMOD), spatial restraint (MODELLER and DRAGON) and artificial evolution (NEST) (Xiang, 2006). Spatial restraints modelling using MODELLER 9v10 was chosen to build our folded eIF5A$_{THR}$ homology model of PG57. Spatial restraints allow the calculation of energy terms based on satisfaction of spatial restraints and distance optimizations (Özlem Tastan Bishop et al., 2008; Söding et al., 2005). Constraint space programming is used to solve for loop regions that contain gaps and mutations within the queries that are not found in the templates (Schwede, Kopp, Guex, & Peitsch, 2003).

Prior to building the model the atomic coordinates of the template were prepared from a 3ER0 PDB file which was edited to include only one chain and renumbered. This chain sequence was then realigned with PG57 to prepare the target-template input alignment file in PIR format for MODELLER. The MODELLER execution python script, shown in Appendix 3, was edited to reference the appropriate alignment file, template structure and target sequence. In this method, by using optimized probability density functions obtained from different initial conformations, 100 models are built iteratively from which the highest quality structure is selected. Before the final structure is selected MODELLER refines the probable structures and assesses the reliability of the models ranking them according to a score of their atomic distance–dependence from a generalized native state.

Refinement and validation

After the generation of a model, structure refinement precedes model validation and evaluation. In order to refine the proposed structure, molecular mechanics methods can be used to eliminate sub-optimal conformations (Özlem Tastan Bishop et al., 2008). MODELLER

uses a CHARMM based force field which uses a sum of individual terms such as bond lengths, valence angles, dihedral angles and backbone torsional corrections to satisfy stereochemical spatial restraints related to internal and non-bonded contributions of a particular conformation (Brooks et al., 2009; Sali & Blundell, 1993). This force field is used during restrained energy minimizations which obtains the near native low free energy structures, global minima, by surveying the energy landscape of all potential conformations. This automated process of refinement on these 100 structures generates 100 proposed structures and these are evaluated based on their Discrete Optimised Potential Energy (DOPE) score (Brooks et al., 2009; Brooks et al., 1983). The DOPE score is a knowledge-based statistical potential that has been shown to be the best scoring function for evaluating structures obtained from X-ray crystal structures (Shen & Sali, 2006).

After the proposed structure is selected it is further validated for inconsistencies and verified for errors. Programs used for model verification and validation can either be based on stereochemistry or based on the matches between sequence and structure (Tastan Bishop et al., 2008; Söding et al., 2005; Xiang, 2006). Protein stereochemistry and geometries were checked using PROCHECK which compared the model generated with standard measures of geometry observed from known structures in the PDB (Laskowski, 2009). We were able to identify improper protein stereochemistry; symmetry; geometry associated with bond lengths, bond angles, torsion angles and chirality; as well as the structural packing (Laskowski et al., 1998; Xiang, 2006). Matches between sequence and structure were verified by Verify3D and ProSA which are able to distinguish incorrectly folded segments which need further refinement of their overall structure, by measuring the match between the sequence and structure, and assigning a score for each residue related to how well it fits its current environment (Xiang, 2006). Verify3D evaluates the area of the residue that is buried, the fraction of side chain areas that are covered by polar atoms and the local secondary structure to which the residue belongs based on an overall profile score for the structure (Luthy, Bowie, & Eisenberg, 1992). Human inspection is often incorporated in the further validation of structural models to improve the confidence of the final probable protein structure (Özlem Tastan Bishop et al., 2008).

The best structure was selected using the MODELLER normalised DOPE-Z Score, ProCHECK and Verify3D. The best PG57 structure was visually compared with the best 3ER0 structure

generated by MODELLER when it was aligned with its crystal structure version. Fully unfolded PG57 eIF5A$_{THR}$ was also prepared simply by converting the sequence to structure in Discovery Studio and saving the file as a PDB after cleaning the geometry. The unfolded PG57 conformation was also analysed using ProSA.

3.2.2 Molecular Dynamics

GROMACS version 4.5.5 was used for molecular dynamic simulations of unfolded PG57, folded PG57 and the 3ER0 structures (Apol et al., 2010; Pronk et al., 2013b). Due to limited computational resources only 1ns of solution dynamics were simulated using the protocol described below in Figure 3-1.

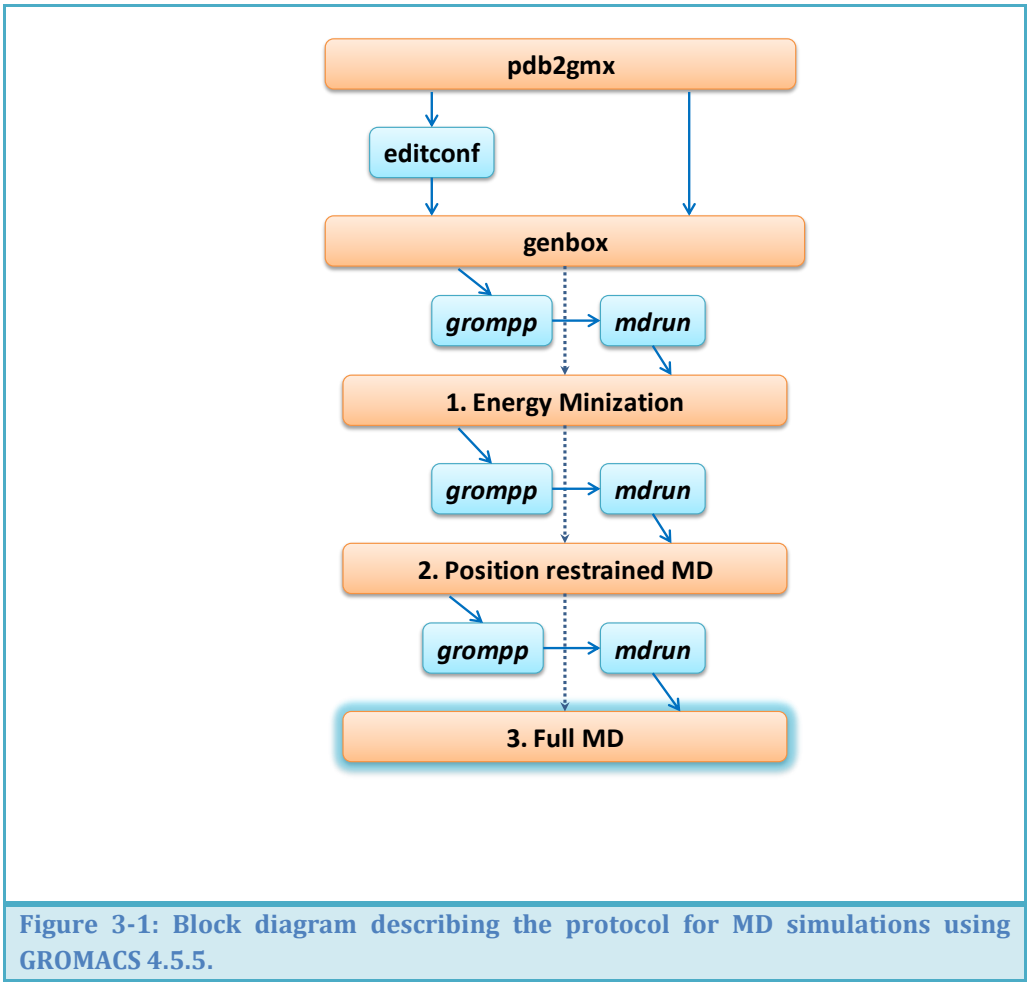


Figure 3-1: Block diagram describing the protocol for MD simulations using GROMACS 4.5.5.

MD simulations within GROMACS require the preparation of adequate GROMACS format structure files (.gro) and topology files (.top) from single model PDB input files. These

GROMACS files are created by the pdb2gmx script that invites the selection of the appropriate force field and water type to be used during the simulations (Apol et al., 2010). The .gro structure files are extended .pdb files that have been converted into Gromos87 file format to allow the inclusion of atom velocities and concatenation for trajectories. The .top topology files contain complete descriptions of interactions that occur within atoms in the system. They store information related to the atom types, bonds, atom pairs, bond angles, dihedrals and the referenced .itp file which defines the system to be studied and its participating force field (Apol et al., 2010). We chose to use the charge and main chain torsion optimized AMBER-03 forcefield in co-operation with the TIP3P water model for our dynamics.

After generating the .gro and .top files, they become the inputs for the next step. In this step a user defined box is created (editconf) before the protein of interest is placed at its centre and solvated (genbox) (Apol et al., 2010). The solvated protein defined by the new topology and structure files is pre-processed (grompp) to produce input files for the molecular dynamics run. Pre-processing converts the molecular description of the system into an atomic description with interaction descriptions contained in the .top file being converted to constraints for MD. Pre-processing reads in the parameters defined for the MD run from the .mdp file and corrects the system, described by the .top and .gro files, to set the net acceleration to zero with respect to the input parameters. This generates a .tpr output file. The .tpr files are binary portable files that contain information about the pre-processed system topology, parameters of the run, starting coordinates and their velocities. The .tpr file and the .gro file become the input files for the dynamics run and are later used to generate the trajectory files of the simulation from the trjconv script.

There are a total of three runs that are executed for every structure with two of them, energy minimization and position restrained dynamics, preparative steps. Before each MD run the topology files and structure files are pre-processed according to the parameter file that defines the step, as described above (Apol et al., 2010). Energy minimization is the first non-dynamic preparative step, performed in order to relax any bad contacts of the protein with solvent molecules. Steepest decent methods were implemented for as many steps are required to converge the system in energy minimization for the 3ER0 and PG57 folded and

the PG57 unfolded structures. Dynamic preparation of the solvent molecules is performed by position restrained MD which optimizes the solvent system prior to the full dynamics run. To do this, the protein is restrained within fixed reference positions while the solvent molecules are allowed to traverse with the forces applied to the system. To ensure that sufficient minimization had occurred potential energies were computed and results are described below. Convergence of energies will indicate a stable system. The unfolded PG57 model required significantly longer position restrained molecular dynamics runs in order to minimize the Linear Constraint Solver (LINCS) errors. After performing 10ps of position restrained molecular dynamics (100 ps for unfolded), full 1ns molecular dynamics of the system were simulated and the trajectories data stored within trajectory .tpr and .trr files. The trajectories were analysed and the behaviour of the protein during simulation monitored.

Alternative approaches to obtaining unfolded models were explored by using unfolding interactive molecular dynamics. This approach will be discussed in Chapter 4.

3.3 RESULTS AND DISCUSSION

3.3.1 Homology Modelling

Template identification

The table below shows the PDB code, the source organism, resolution of structures, amino acid lengths, sequence similarity and sequence identity of sequences retrieved from BLAST search using PG57 eIF5A protein of 169 amino acid lengths.

Table 3-1: Protein structures retrieved from query search using PG57 sequence ranked according to sequence identity with PG57.

PDB Code	Source Organism	Protein	Resolution (Å)	Amino Acid Length	Sequence Identity to PG57	Sequence Similarity to PG57
3ER0_A	*Saccharomyces Cerevisiae*	eIF5A	3.35	167	93.1	93.1
1X6O	*Leishmania Braziliensis*	eIF5A	1.6	174	46.7	64.4
2EIF_A	*Methanocaldococcus jannaschii DSM 2661*	eIF5A	1.8	136	23.3	44.8
1EIF_A	*Methanocaldococcus jannaschii DSM 2661*	eIF5A	1.9	135	23.3	44.2
1KHI_A	*Neurospora crassa*	HEX1	1.78	176	21.7	47.2

From the table above, although 3ER0_A had a low resolution it was chosen as the template source for alignment and model building. The sequence from PG57 had very little difference with 3ER0 and the difference was only with the leucine (L), valine (V), proline (P), arginine (R) thrombin cleavage site. The high sequence identity, 46.7 %, and ideal resolution, 1.6 Å, of 1X6O were mitigated by the fact that it originates from a protozoan species and is found as a monomer, implying significant structural and functional differences. *M. jannaschii*, eIF5A, and *N. crassa*, HEX1 (1KHI), structures were excluded because they showed less admirable sequence identities perhaps influenced by the shorter amino acid chain lengths for the eIF5A variants. The HEX1 protein (1KHI) is a dimeric protein involved in the formation of septal spores of filamentous fungi (Tenney et al., 2000). Although it does not share sequential homology, it has been shown to share significant structural and functional homology with eIF5A. It possesses an α-helix and an oligonucleotide binding fold within its C-terminal domains while a similar SH3 barrel β-sheet N-terminal domain exists similar to that in eIF5A. Because of these similarities HEX-1 dimerisation arrangement has been used to infer the eIF5A arrangement (Gentz et al., 2009). However due to the lack of evidence

supporting the existence of the same eIF5A crystal lattices that are present in HEX-1, these inferals are inadequate to account for eIF5A tertiary structure.

Sequence Alignment

Multiple sequence alignment, from ClustalW, confirms the selection of 3ER0 as a suitable template for model building. The target to template sequence alignment file is shown in Figure 3.2 and it was used as input for model building. Predicted secondary structure elements generated within Discovery Studio using the Discrimination of protein secondary structure class (DSC) prediction method are also shown. There is strong sequence homology and the predicted secondary structure elements correlate despite the LVPRGS thrombin cleavage site mutation introduced.

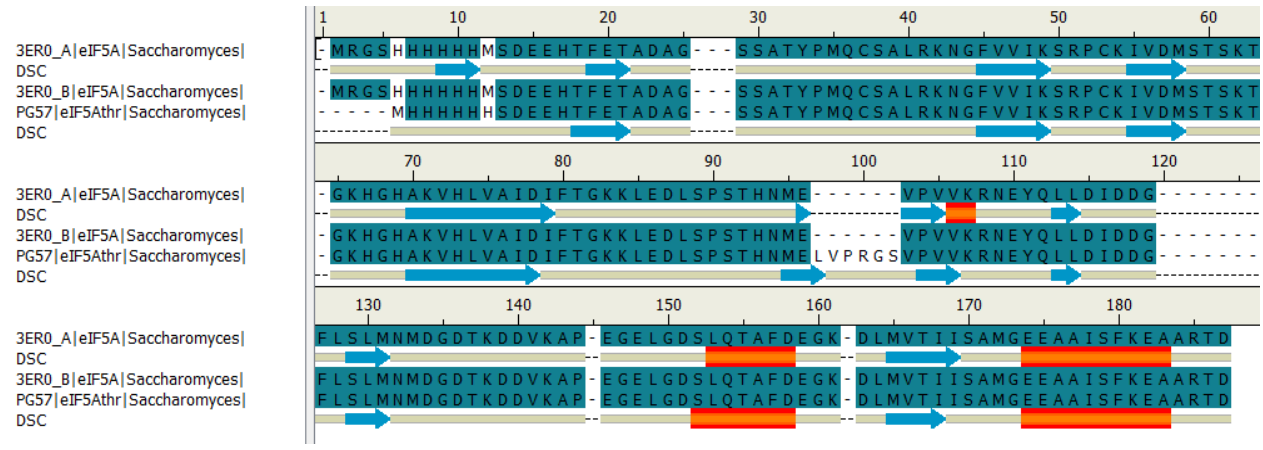


Figure 3-2: Target to template Sequence alignment of 3ER0 chain A and B with PG57. Strong alignment is coloured green and the DSC row illustrates predicted secondary structure. Blue arrows denote β strands while orange illustrates the α-helices. Grey codes for loops.

Model Building and Validation

Of the 100 generated structures from MODELLER, after refinement and ranking, the top 3 structures (based on the DOPE scores) from the PG57 structure ensemble were 20, 77 and 98. Stereochemical properties were evaluated using the PDBSum pictorial database that performs PROCHECK calculations and generates plots as part of its analyses. We can extract the residues located within the most favoured ABL regions of the Ramachandran plot of the model. A sequence-structure match score was generated by Verify3D. Structure 77 was chosen as the best structure based on its performance in the validation and analysis tests, obtaining the lowest combined overall rank over all of the tests.

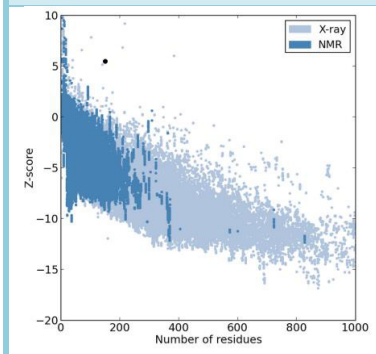
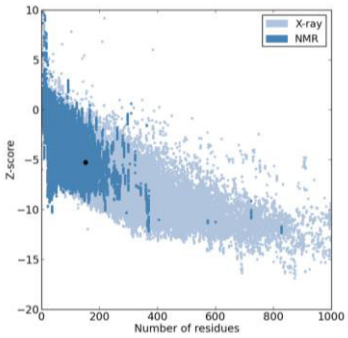
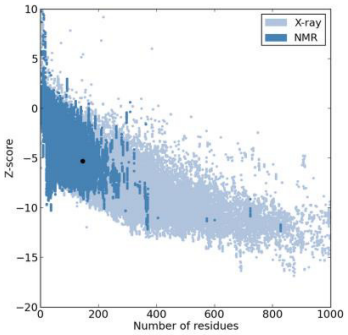
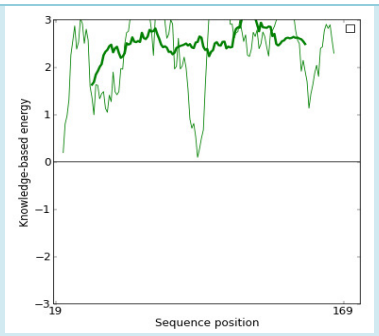
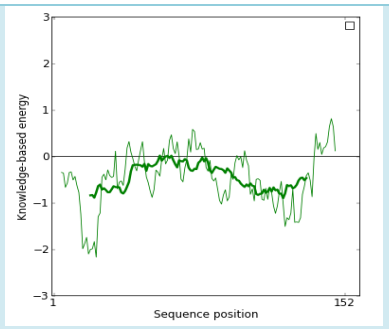
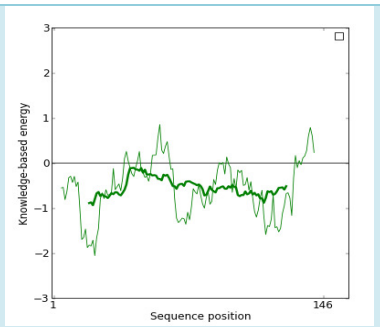
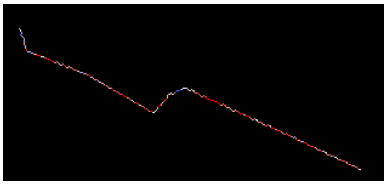
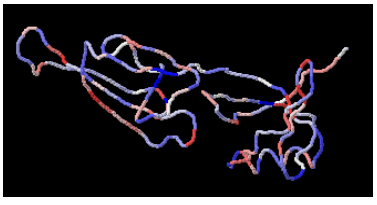
Table 3-2: Evaluation table for the best three models of PG57 generated from MODELLER model build

	MODELLER	**PROCHECK**		**VERIFY3D**	**OVERALL**
Structures	DOPE Score *(Rank)*	% in ABL *(Rank)*	AVG. G – Factor *(Rank)*	Overall Score *(Rank)*	Total Rank Score
(# 20)	-11231 *(2)*	75.8 *(3)*	-0.45 *(1)*	45.8*(3)*	9
(# 77)	-11300 *(1)*	76.5*(2)*	-0.51*(3)*	49.7*(1)*	7
(# 98)	-11222 *(3)*	77.3 *(1)*	-0.46 *(2)*	48.1*(2)*	8

ProSA

More detailed energy plots were prepared for structure 77 to evaluate its feasibility to be used as a structure in further analysis and molecular dynamics and to identify regions of unfavoured conformations. Below is a comparison of the ProSA energy plots from the unfolded eIF5A, the folded eIF5A$_{Thr}$ and the native 3ER0 X-ray crystal structure of monomeric eIF5A obtained from model building. The fully unfolded structure was to be subjected to partial folding and NMR analysis of its initial state was not done, thus the initial expected unfavourable quality was not deemed to be of concern.

Table 3-3: Analysis table of ProSA quality assessment for PG57 unfolded, PG57 folded and 3ER0 folded conformation

PG57 UNFOLDED MODEL	PG57 FOLDED MODEL	3ER0 FOLDED MODEL
Overall model Quality Z-score against number of residues for experimental techniques		
		
Z-score of 5.47	Z – score of -5.29	Z – score of -5.39
Local model quality illustrated by single residue energies plot		
Average over		
10 residues	40 residues	
		
High energy values	Low energy values	
Secondary structure arrangement with residues coloured by residue energy		
Low Energy	High Energy	
		
Unfolded high energy model	Folded low energy conformations	

Acceptable Z-scores of -5.29 and -5.39 are within the same range for PG57 structure number 77 and 3ER0 respectively. The residue energy plots do not show significantly high energies for specific amino acids when energy is averaged over 40 residues. However for 10 residues, positive spikes are observed between residues 65-75 and 130-140 for PG57 folded and 50-60 and 125-130 in 3ER0. Their 3D arrangements exhibit almost identical secondary structure arrangements with minor deviations within the flexible loop region containing the LVPRGS mutation in the folded PG57 mutant. These values are indicative of reasonably accurate structures (Wiederstein & Sippl, 2007). A strong resemblance to the conformation of 3ERO observed in the secondary structure analysis justifies the homology model generated. As expected the unfolded model shows an unfavourable positive overall quality score of 5.47. High residue energy values are seen consistently along the structure for the unfolded model consistent with a highly unfavoured non-native conformation.

3.3.2 Molecular Dynamics

GROMACS contains many different tools for analysis of simulations. We performed detailed analyses on the full 1ns trajectories in order to evaluate how well the crystal structures maintained their structural integrity within simulation, how the size of the protein was affected during the simulation and whether backbone torsion angles were within allowed regions within simulations.

Potential energies were analysed after minimization, during the position restrained molecular dynamics in order to evaluate whether the systems had been adequately minimized prior to running full simulations. GROMACS generates a portable energy file during the simulation which contains all the energy terms calculated during the simulation. For our purposes we extracted data related to the potential energy for each structure over 500 steps for 1 ps of position restrained dynamics and the data is illustrated in Figure 3-3.

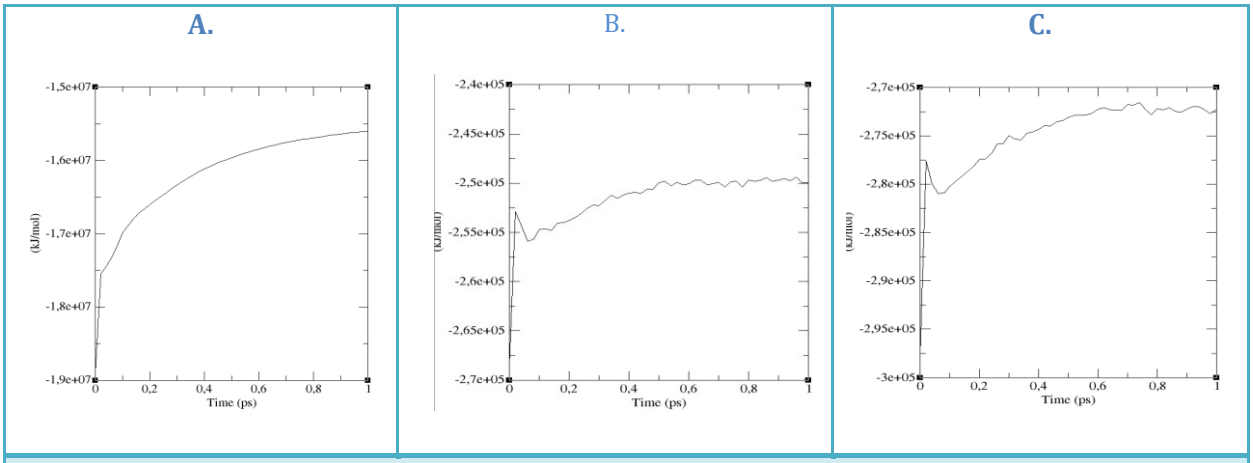


Figure 3-3: Potential energy of position restrained MD. Panel A: PG57 unfolded Model; Panel B: PG57 folded model; Panel C: 3ER0 folded model.

After the simulations were run the trajectories and structures were analysed in order to evaluate whether the simulations improved the qualities of the structures obtained from model building. An energy minimization step of steepest gradient was employed on the structures after the simulation. From this minimization step, the overall quality for the unfolded model decreased as seen in Table 3-4 below. This is reflected in the lower percentage of 67.9% for residues within their favourable orientations after EM. After minimization the folded 3ER0 model observes an increase in the percentage of residues within the ABL region to 83.6%. The folded PG57 model experiences a drop in the percentage of residues within the ABL regions despite the decrease in the Z-score after EM.

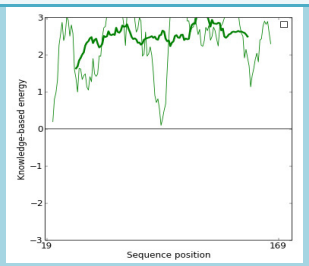
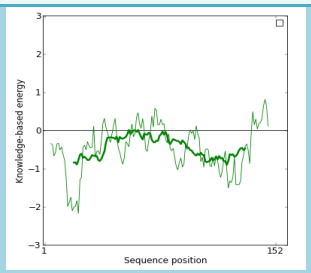
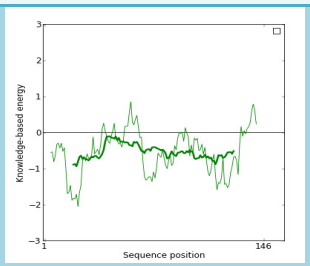
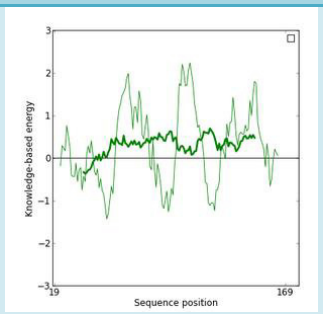
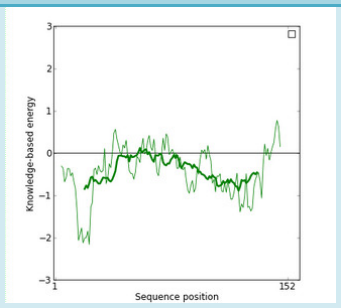
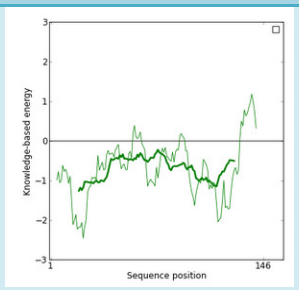
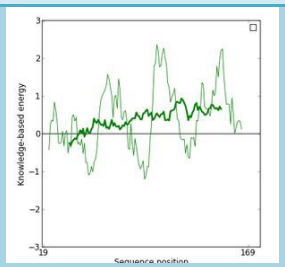
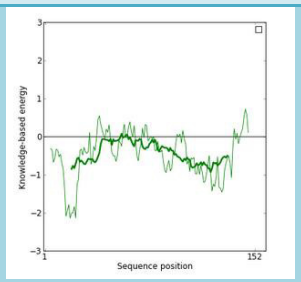
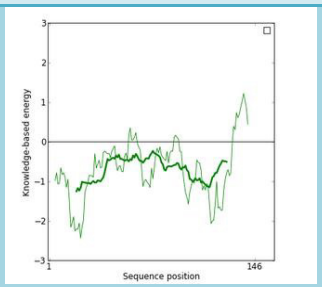
We used ProSA to calculate the overall model quality and the residue specific quality, while PROCHECK was used to generate the Ramachandran plots to obtain the percentage of residues that possess favourable phi-psi angles.

Table 3-4: Analysis of structural integrity during MD assessed by the ProSA Z-score calculation and PROCHECK Ramachandran plots.

Property	Time	PG57 Unfolded	PG57 Folded	3ER0 Folded
Overall model quality (Z-score)	0.00 ps	5.47	-5.29	-5.39
	1.00 ns	-0.57	-5.13	-5.81
	EM after 1.00 ns	-0.15	-5.2	-5.78
Percentage of Residues in ABL regions	0.00 ps	0.00%	76.5%	59.4%
	1.00 ns	80.2%	75.8%	78.9%
	EM after 1.00 ns	67.9%	72.7%	83.6%

From Table 3-4 after 1 ns of simulation the 3ER0 model and the unfolded PG57 model adopt more favourable conformations. The unfolded model adopts significant stabilisation of -0.57 after 1 ns of simulation from a Z-score of 5.47 before simulation. The folded PG57 model experienced slight destabilisation to -5.13 from -5.29 as a result of simulations. The Ramachandran data supports the observation of an increase in model quality observed in the unfolded and 3ER0 models after simulation. For the unfolded PG57 model, at 0 ps of simulation, none of the residues are within their favoured orientations but after 1 ns of simulation 80.2% of the residues possess optimal torsions. For the folded 3ER0 model initially there are 59.4% of its residues within their favoured orientations but after simulation 78.9% of residues obtain favour torsions. This improvement in quality is reflected in the Z-score decreasing from -5.39 to -5.81 for the folded 3ER0 model. For the folded PG57 model, simulations reduced the percentage of residues within the favoured ABL regions by a fraction.

Table 3-5: Residue specific model quality generated from the ProSA quality assessment calculation

Time	PG57 Unfolded	PG57 Folded	3ER0 Folded
0.00 ps			
1.00 ns			
EM after 1.00 ns			

Residue specific model quality assessments made by ProSA shown in Table 3-5, illustrate consistent trends as those observed in the overall model quality scores. Obvious differences in the residue specific qualities are observed for the unfolded model after 1ns dynamics. Although a decrease in the overall model quality is present in the energy minimized folded PG57 structure, these differences are barely noticeable in the knowledge based residue specific energy between the starting structure, 1ns simulation and the post-dynamics energy minimization step. Plots of the averages calculated over 10 residues (thin green) do not show much variation between the energies of structures after 0 ps and 1.00 ns for the folded conformations. This implies a strong structural integrity within the dynamics simulation and the final energy minimization step. Although residues are destabilized out of

the favourable torsion orientations, it appears from the knowledge based energy calculations, other factors such as hydrogen bonds and solvent interactions are introduced within the structure to satisfy thermodynamic costs to conformation. Energy minimization is a step in which the solvated system is rapidly minimized in order to reduce strong repulsive forces. It is possible that an unfavourable overall geometry is accessed as a result of improving some parts of the system at the expense of others during the final minimization. The molecular dynamics simulations reduce the unfavourable interactions introduced during solvation and model building.

Root Mean Squared Deviation (RMSD) Analysis

The root mean squared deviation was used to compare the structure from a trajectory with its reference structure obtained upon the completion of the position restrained molecular dynamics. This RMSD was computed in a size-independent manner after optimal rigid body superposition was performed. Scaling of the radius of gyration and using equal principle moments of inertia allowed a univers*al* size-independent comparison since overall differences in size and ellipticity were reduced (Maiorov & Crippen, 1995). The Cα atom coordinates were chosen for comparison in order to allow a detailed whole protein structural analysis over 1ns of simulation, Figure 3-4.

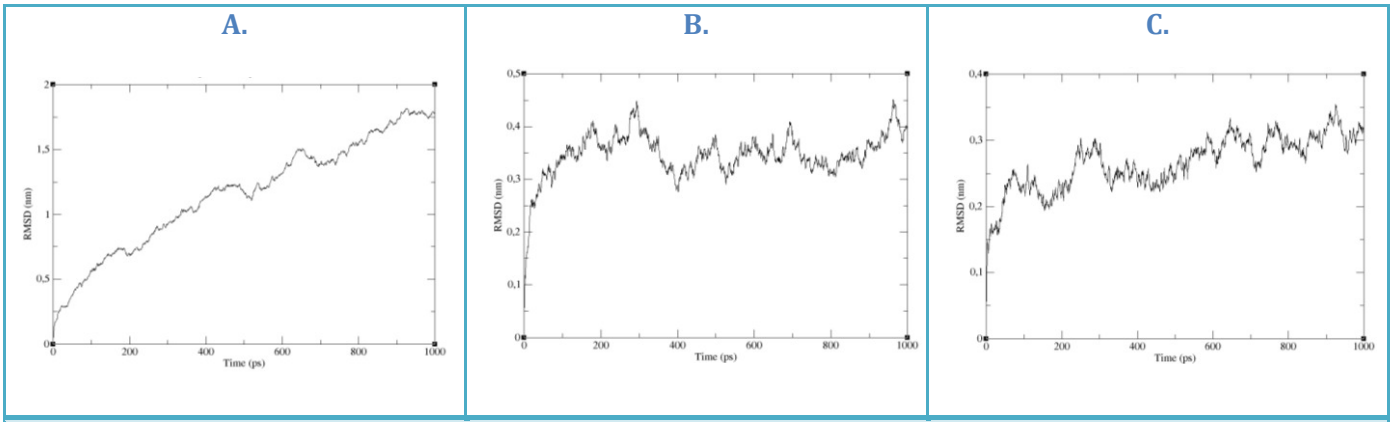


Figure 3-4: Root Mean Squared Deviation 1ns plot. Panel A: PG57 unfolded; Panel B: PG57 folded; 3ER0 folded models.

The folded conformations (Figure 3-4 Panel B and Panel C) maintain strong geometric similarity throughout the dynamics with RMSD values lying well below 1 nm for overall

similarity in folding motifs of our proteins. It is well known that biological molecules exhibit specific processes that occur over a wide range of time scales (Karplus & McCammon, 2002). The rapid fluctuations observed within the folded conformations can be attributed to local motions such as atomic fluctuations, side chain motions and loop motions within the folded motifs.

Irreversible structural deviations are observed within the unfolded conformation (Figure 3-4 A). This can be attributed to large-scale motions that occur rapidly within the unfolded unstable conformation. Other contributions to the larger RMSD are folding processes that may have occurred during simulations. Size analyses expressed as the radius of gyration can shed light on the folding degree observed within 1 ns of the unfolded simulation.

Radius of gyration

The extent of the protein's span is computed by its radius of gyration. The radius of gyration is calculated by taking the mean square of the distance of all the atoms from the centre of mass of the molecule and is proportional to the RMSD between the structures (Apol et al., 2010). We computed the radius of gyration of a molecule about its principal axis during the simulation over 1 ns of dynamics and the data was plotted in Figure 3-5.

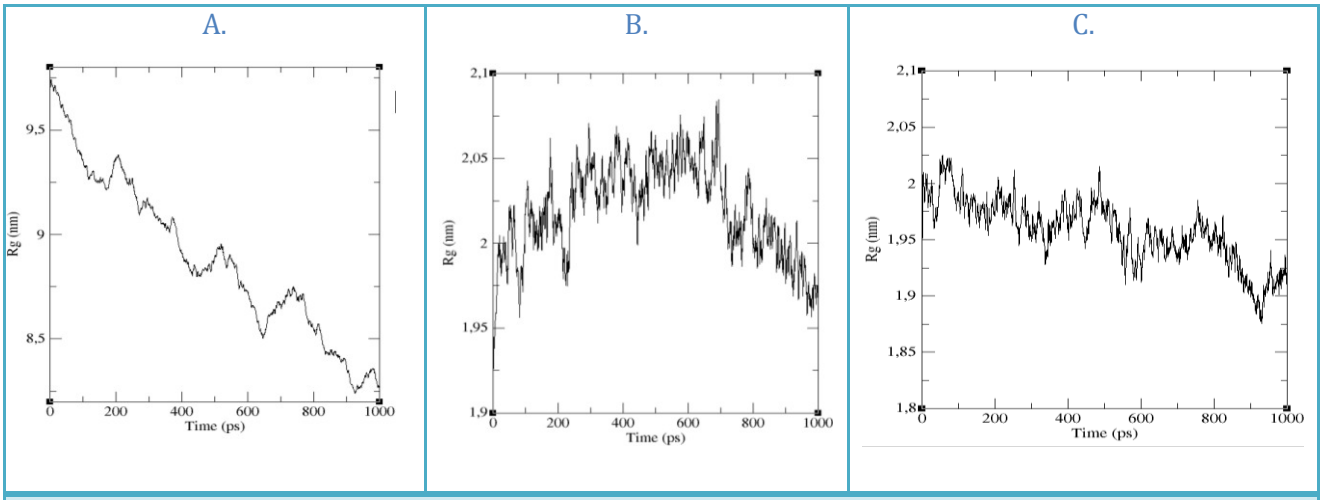


Figure 3-5: Radius of gyration from 1ns of molecular dynamic simulations. Panel A: PG57 unfolded; Panel B: PG57 folded; Panel C: 3ER0 folded models.

Over simulations of 1ns, the folded conformations experience rapid fluctuations in their distances. In the 3ER0 folded conformation a 3% overall drop in radial size is observed over

1 ns while a periodic fluctuation of about 250 ps is observed (Figure 3-5 C). This periodic fluctuation is consistent with rigid body domain motions associated with hinge bending around the flexible loop region of the protein (Karplus & McCammon, 2002). No overall size change is observed within the PG57 folded structure over 1ns of simulation (Figure 3-5 B). The periodic fluctuation observed in 3ER0 appears is not obvious over 1 ns in the folded PG57 structure (Figure 3-5 B and C). There is however a broad fluctuation noticeable over 1ns. Longer simulations could give information about the nature of that fluctuation. Increased rigidity within the flexible hinge region of the protein which arises from the mutation present in PG57 would result in a lengthening of the 250 ps hinge bending observed in the native 3ER0 model.

An obvious size reduction of 8% is observed from Figure 3-5 A in the unfolded conformation over the 1 ns. This rapid size reduction can be used to account for the relatively large RMSD observed later in the simulation compared to earlier. This propensity to contract is consistent with the protein folding code which attempts to understand the protein folding problem. In this code they postulated that the conformation adopted by a protein is as a result of the satisfaction of thermodynamic factors with hydrophobic interactions being the major driving forces (Dill, Ozkan, Shell, & Weikl, 2008). Our unfolded conformation of PG57 has not been given its optimum initial conformation and will continue to contract spontaneously. This will continue until it adopts a conformation with secondary structures that satisfy all the other interactions such as van der Waals forces, electrostatics and hydrogen bonding (Wolynes et al., 2012). Whether this structure will be the native conformation is dependent on the folding route. If the energy surface describing the folding mechanism is rough the folding trajectory is likely to contain stable intermediates which may trap the protein preventing further folding unless sufficient energy is supplied to overcome these barriers (Chan & Dill, 1998; Dobson et al., 1998).

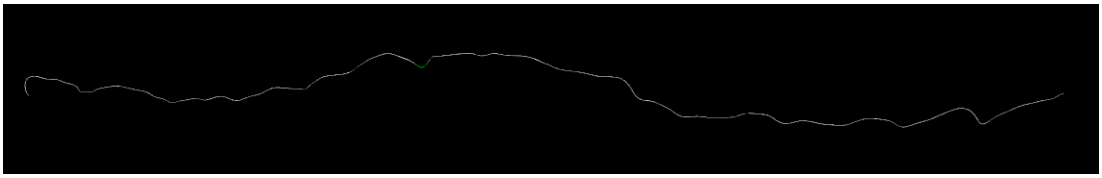
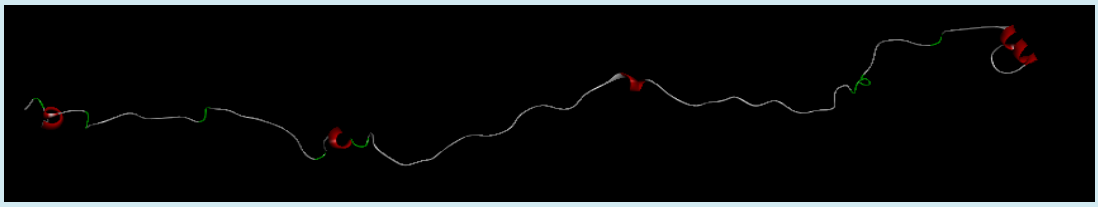
Time	Structure
10 ps	
100 ps	
1 ns	

Figure 3-6: Structures of unfolded PG57 showing the development of secondary structure after 10 ps, 100 ps and 1 ns dynamics.

In the unfolded conformation after 1 ns of simulation we were able to detect secondary structure formation in addition to continued contraction as shown in Figure 3-6. Longer simulations of the unfolded protein dynamic would allow a structure representative of an energy minima to be approached which ideally would resemble the native structure provided no chaperone proteins are involved in the folding mechanism. These complementary enzymes direct the folding reaction by stabilisation intermediate conformations that are unfavoured allowing the protein to be exposed to previously inaccessible conformations (Chan & Dill, 1998).

Ramachandran Plots

The Ramachandran analysis of the trajectories obtained over 1 ns of simulation allows us to monitor whether the backbone torsion phi and psi angles are maintained within allowed

regions most of the time. We generated plots for all of the residues for each structure during the simulation and plotted their specific values on the same phi and psi axes. The data is illustrated in Figure 3-7.

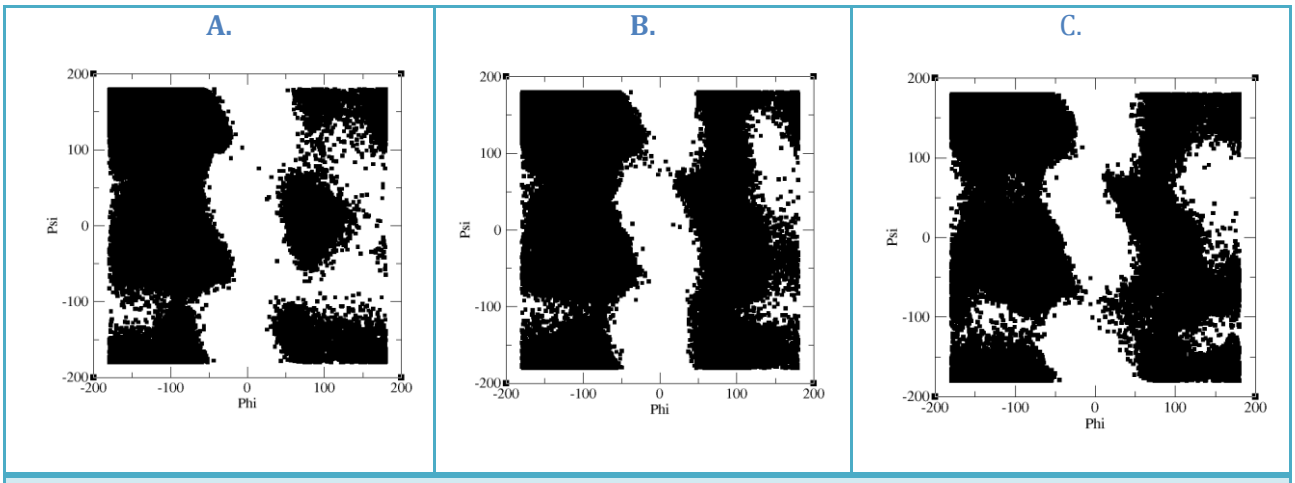


Figure 3-7: Ramachandran plot of models during 1ns of molecular dynamics simulations. Panel A: PG57 unfolded; Panel B: PG57 folded; Panel C: 3ER0 folded.

When using the RAMPAGE program to evaluate the Ramachandran plots it can be seen that for our structures more than 95% of the amino acids generate phi/psi angles that are within acceptable limits (Lovell et al., 2003). Although the PG57 unfolded conformation undergoes the most significant structure changes, it however maintains the highest structural integrity during the simulations and its dynamics. This data gives us confidence that the forces and restraints defined by the forcefields and the parameters during simulations are reliable for accurately simulating the solution dynamics of our models in water.

3.4 CONCLUSION

We were able to generate homology models of the mutated eIF5A$_{Thr}$ based on the native eIF5A obtained from the X-ray structure of 3ER0. We were able to simulate the solution dynamics of folded and unfolded eIF5A$_{Thr}$ and folded eIF5A using the AMBER-03 forcefield within the GROMACS software package over 1 ns of dynamics.

Potential energy convergence plots indicated stable systems that were adequately prepared and minimized before the full unrestrained dynamics over 1 ns were carried out. Observations from gyration analyses of the folded structures during simulations are consistent with findings made by Kim et al. (1998). They proposed a rotation of 7^{o} about the axis along the length of the monomeric structure within the flexible linker region of *Methanococcus jannaschii* eIF5A (Kim et al., 1998). Whether the rigidity observed during simulations of the folded eIF5A$_{Thr}$ from PG57 could have prevented it from folding appropriately after synthesis is still yet to be unambiguously proved. This possibility is however unlikely although it should not be overlooked until an accurate solution structure of eIF5A$_{Thr}$ is obtained.

It is well known that protein sequence can influence the folding routes that proteins take to arrive at their native structures (Dill et al., 2008; Wolynes et al., 2012). It is unclear at this stage what influence the mutation has on destabilizing the intermediates along the folding route towards the folded structure of eIF5A$_{THR}$ towards achieving a conformation observed from other crystal structures. Future studies that allow the MD simulations of the unfolded proteins to proceed beyond 1 ns for both the mutated eIF5A$_{Thr}$ and native eIF5A could give insight into the folding rates of the different homologues. This information could be used to decipher the misfolding experienced by the mutant *in vivo*.

Chapter 4: Molecular Dynamics as a means to study solution Dynamics

4.1 INTRODUCTION

4.2 MATERIALS AND METHODS

4.2.1 Predicted Chemical Shifts and synthetic spectra

4.2.2 Back-calculation of Residual Dipolar Coupling

4.2.3 T_1 relaxation and τ_c from MD simulations

4.3 RESULTS AND DISCUSSION

4.3.1 Chemical Shift Prediction

4.3.1.1 Steered Molecular Dynamics

4.3.2 Residual Dipolar Coupling

4.3.3 T_1 relaxation and t_c from MD simulations

4.4 CONCLUSION

Chapter 4: Molecular Dynamics as a means to study solution Dynamics

4.1 INTRODUCTION

There is considerable effort that has gone into attempts at minimizing the experimental data needed to generate protein structures in a fast and reliable manner, thus expanding the applications of experimental techniques to unresolved structures and conformations. In particular some groups have been able to use chemical shift data alone while others have used residual dipolar coupling data integrated with chemical shifts to extend biophysical applications to solutions of native, non-native and membrane-bound proteins (Cavalli et al., 2007; Delaglio, 2001; Pintacuda et al., 2006; Rohl & Baker, 2002). It is our hope from this investigation that extracting NMR restraints obtained from solution dynamics of folded eIF5A and its mutant eIF5A$_{Thr}$ would allow us to further interrogate the structural integrity of eIF5A during 1 ns of molecular dynamics in appropriate solvent.

Interest is growing around the use of chemical shifts, RDC data and relaxation dispersion to probe structures of elusive, low populated, excited protein states (Carbajo & Neira, 2013b; L. E. Kay, 2009; Vallurupalli et al., 2008). Chemical shifts obtained from NMR have been shown to provide a wealth of information related to protein secondary structure and backbone dihedral angles (Cavanagh et al., 2007; Herrmann, Guntert, & Wuthrich, 2002). Applications to proteins whose tertiary structures have not been determined experimentally have become feasible with the development of accurate and robust chemical shift predictors (Kohlhoff et al., 2009). These have been used to generate predicted shift data from experimentally determined structures. This data is then used to refine the relationship between conformational space and chemical shift informing structure searches that are based on the measured chemical shifts patterns (Cavalli et al., 2007; Kohlhoff et al., 2009; Montalvao et al., 2012). Residual dipolar couplings (RDC's) have been incorporated into protein structure prediction through algorithms that use RDC's as a means of searching for fold homology while attempting multiple fold recognition. They have also been used as restraints for model building in sequence homology based structure prediction (Delaglio, 2001; Güntert, 2009; Rohl & Baker, 2002). Experimentally derived measurements such as

thermodynamic data can be extracted from differentiable relaxation dispersion experiments and can be used to infer the behaviour of proteins in different exchange processes (Vallurupalli et al., 2008).

Because of the strong relationship between protein structure, conformation and chemical shifts, one of the aims of this project was to obtain experimental solution data of eIF5A using NMR protocols. It was intended to provide experimental evidence which seeks to justify the dynamics of the monomer in order to validate the influence of hypusination and dimerisation on its behaviour. By extending the applications of NMR at 600 MHz using lanthanide induced pseudocontact shifts, it had been hoped that the structure of the hypusinated dimer would be solved in future studies.

Due to sensitivity to molecular structure and its conformation, fluctuations of native states reduce the precision of NMR applications to structure studies. Despite the presence of such fluctuations and discrepancies, chemical shifts are readily measurable with great accuracy (Neal et al., 2003; Shen et al., 2009). The experimental shift patterns from unresolved proteins can be used to search the conformational space available from solved structures of ground state conformations within data banks whose predicted chemical shifts have positive matches (Cavalli et al., 2007; Kohlhoff et al., 2009). A major obstacle however towards this is the complicated nature of the dependency between the conformation and chemical shifts as well as the intrinsic dynamic nature of proteins in solution (Herrmann et al., 2002; Wishart & Case, 2002). Computational tools such as ShiftY, TALOS, ShiftS, PROSHIFT, SPARTA, SHIFTX and CamShift have been developed to accurately predict the chemical shifts of a protein from its experimentally solved structure (Cornilescu et al., 1999; Kohlhoff et al., 2009; Neal et al., 2003; Wishart et al., 1997; Xu & Case, 2001). Chemical shift predictors have been successfully implicated in the generation of reasonably accurate protein structures from incomplete chemical shift data in structures with limited structural fluctuations. This extends the potential to use NMR for structure determination of large proteins and protein complexes with sparse data (Shen et al., 2009).

CamShift is a fast and accurate low computational cost tool that was developed in order to predict chemical shifts from three-dimensional protein structures using rapid measurement of inter-atomic distances (Kohlhoff et al., 2009). Because of its speed, robustness and low

computational cost, CamShift has found applications in determination of highly dynamic non-native protein conformations.

$$\delta_a^{pred} = \delta_a^{rc} + \sum_{b,c} \alpha_{bc}\, d_{bc}^{\beta_{bc}} \qquad (4.1)$$

Equation 3.1 describes the formula used to generate the predicted chemical shift, δ_a^{pred}, of atom *a*, from its random coil chemical shift, δ_a^{rc}, and secondary, conformational contribution to chemical shift defined by the distance between atoms *b* and *c*, d_{bc}. The random coil value is the chemical shift value determined from short disordered polymers. While the secondary chemical shift values arise due to non-covalent influences such as solvent interactions, ionization constants, ring orientations and hydrogen bond interactions (Wishart & Case, 2002). In CamShift a series of atom pairs of *b* and *c* are chosen from within the vicinity of *a*, and their sums give rise to the overall conformational contribution to the chemical shift of atom *a* within a specified range. The parameter α is dependent on their atom types which are governed by the atomic species, the residue types atomic type within the residues and the atomic hybridisation states. The distance weighting is parameterised by β, which considers whether the atoms *b* and *c* are bound or unbound. A combination of the conformational contribution and the random coil contribution gives the value for the predicted chemical shift of atom *a*, observed in that particular conformation (Kohlhoff et al., 2009).

Although chemical shifts provide limited structural information, orientation dependent RDC data provides restraints that possess a strong conformational dependence (Montalvao et al., 2012). RDC couplings occur as a result of the interaction between two magnetically susceptible nuclei that are partially aligned within a magnetic field. The strength of the coupling, $D_{1,2}$, between the two nuclei, *1* and *2,* is dependent on the orientation of the angle, ϑ, between the internuclear vector and magnetic field, where $D_{max} = -\mu_0\gamma_1\gamma_2 h/8\pi^3 r^3$ is the maximal value for the dipolar coupling, for magnetic constant, μ_0, for two nuclei, with magnetogyric ratios, $\gamma_1 \text{and } \gamma_2$, separated by a fixed internuclear distance, r.

$$D_{1,2} = D_{\max}\langle\frac{1}{2}(3\cos^2\vartheta - 1)\rangle \quad (4.2)$$

In an isotropic solution, where the solute tumbles randomly, the time dependent orientation of the angle ϑ samples all the possible values equally and thus over time the $D_{1,2}$ is reduced to zero. However in the presence of anisotropy, introduced through liquid crystalline media such as bicelles, phages and strained and unstrained polyacrylamide gels, partial alignment results in a measurable non-zero value of the time averaged dipolar contribution to the scalar coupling (Teng, 2013).

In the case of proteins it is necessary to define the measured RDC in terms of coordinate transformations. Here the principle axis system or the internuclear vector of the nuclear spin interaction is oriented with fixed angles, φ_i and φ_j, to the molecular reference frame which is oriented with respect to the laboratory frame of the magnetic field in a time dependent manner. A traceless, real, and symmetric, Cartesian 3 x 3 Saupe order matrix, A_{ij}, of rank 2 is used to define order parameters which describe the orientation of the molecular alignment frame with respect to the laboratory field under the influence of the alignment media. Where δ_{ij} is the Kronecker delta function and ψ_i is the angel between the molecular reference axis, *i, j* = (x, y, z), and the z axis of the laboratory frame, an irreducible representation of the Saupe order matrix is $A_{ij} = \frac{1}{2}\left(3cos\psi_i cos\psi_j - \delta_{ij}\right)$. When the protein is partially aligned, a non-zero time dependent alignment tensor $\langle A_{ij}\rangle$ is obtained from an averaging of all the translations and rotations of the order matrix contributing to the measurable RDC. The dipolar coupling, *D*, of an internuclear vector within a protein is described as a sum of the product of the alignment tensor and the orientation of the internuclear vector.

$$D = D_{\max}\sum_{ij}\langle A_{ij}\rangle \cos\varphi_i \cos\varphi_j \quad (4.3)$$

The residual dipolar coupling analysis tool, REDCAT, was chosen as a tool to perform the RDC back-calculations. Implemented in C/C++ programming language, REDCAT is a portable and user friendly analysis tool that allowed us to augment our analyses of the molecular

dynamics rapidly and efficiently. Based on the *orderten_svd* program, REDCAT contains additional features that allow the rapid back-calculation of RDC's from a co-ordinate file and the selection of an arbitrary alignment tensor with compatible Euler angles provided the internuclear vectors are treated as rigid (Valafar & Prestegard, 2004). As these parameters have strong conformational dependence they are very sensitive to protein dynamics. Non-native excited states of proteins are relatively short-lived and therefore require sensitive techniques for observation. RDC data can provide dynamic experimental data of elusive conformations (Lange et al., 2008; Nkari & Prestegard, 2010). RDC data can also be used in structure determination protocols to describe ensemble conformations of the ground states that satisfy the time averaged RDC restraints (Montalvao et al., 2012).

The rate at which the magnetization induced by a pulse returns to the equilibrium bulk magnetization, $\mathbf{B_o}$, is influenced by the rate of relaxation mechanisms (Reich, 2010). Relaxation of magnetization within the z-axis is influenced by interaction of spin magnetization and its surrounding environment. This mechanism of relaxation is termed spin-lattice relaxation and can be measured by an inverse recovery NMR experiment (Cavanagh et al., 2007). Restoration of equilibrium magnetization along the x and y axis occurs through the loss of phase coherence introduced by the pulse. In the absence of a pulse, different nuclei within the system precess at different rates and frequencies dependent on their chemical environment influencing their local magnetic field. This magnetic field inhomogeneity results in a dephasing of the net magnetization vector within the transverse plane at equilibrium. The time it takes in order for the equilibrium incoherence to be restored is termed the T_2 spin-spin relaxation (Cavanagh et al., 2007; James, 1998). The rate of recovery of this incoherence after a pulse is influenced by the interaction of the nuclear spins with each other, and can be measured quite readily by the Carr-Purcell-Meiboom-Gill (CPMG), spin-echo NMR experiment (Kay, 2009).

In order for T_1 relaxation to occur, magnetic field fluctuations between the surroundings and the spin ½ nuclei are necessary (Kay, Torchia, & Bax, 1989; Keeler, 2005). As the molecule tumbles the magnetic fields change and if these fluctuations occur at the Larmor precession frequency, v_o, several sources of magnetic flux can be affected resulting in an observable T_1 relaxation (Carbajo & Neira, 2013b). Rapid tumbling of molecules in solution can give rise to

fluctuating magnetic fields by influencing chemical shifts of nuclei and scalar couplings affecting directly bonded nuclei which are dependent on their orientation with the magnetic field (Cavanagh et al., 2007).

Advances in labelling techniques have allowed the isolation of specific spin systems which can be used to probe the dynamics of proteins based on the relaxation rates of large sets of backbone nuclei (Kay, 2008). By incorporation of the CPMG pulse sequences to obtain spin-spin relaxation rates, during relaxation dispersion NMR experiments, the extraction of qualitative structural information of conformers involved in exchange processes is possible (Vallurupalli et al., 2008). Measurements of the rates of longitudinal relaxation have in the past been used to study the dynamics and mobility of protein domains due to the strong dependence of correlation time on relaxation (Barbato, 1992; Kay, Torchia, & Bax, 1989). By calculating the correlation times of different amide vectors during molecular dynamics we intend to calculate the predicted T_1 relaxations during dynamics based on an expression derived by Hertz relating the rotational reorientation with the relaxation time of a carbon atom (Hertz, 1967).

To complement the data obtained from experimental NMR protocols it was intended to extract and calculate theoretical NMR data from the molecular dynamics simulations. By calculating theoretical chemical shifts and RDC's, it is possible to plot time-averaged NMR spectra that are corrected for dipolar coupling. We intend on extending theoretical tools based on static structures to dynamic systems by calculating time-averaged data from conformers extracted during molecular dynamics simulations. Analysis of the RDC oscillations observed during the simulations allows for a high resolution interrogation of the behaviour of the dipolar interaction and hence the molecular bond. Measurement of the amide T_1 on individual residues can probe the mobility of specific domains based on their dynamics and chemical shift correlations within that region.

4.2 MATERIALS AND METHODS

4.2.1 Predicted Chemical Shifts and synthetic spectra

Although it only relies on the influence of measurable distance dependant restraints, which can be calculated rapidly, CamShift performs just as well as other tools for chemical shift prediction (Robustelli, Stafford, & Palmer, 2012; Y. Shen & Bax, 2010). For our applications it allows the rapid analysing of a few thousand frames of dynamics and has significantly less computational cost (Kohlhoff et al., 2009). CamShift was used within a bash terminal to generate chemical shift tables for backbone N, H_N, H_α, C_α and C_β atom types from co-ordinate files in appropriate PDB format. In our applications we extracted coordinates of the MD trajectories into PDB format at constant intervals obtained from the 1 ns molecular dynamic simulation. The large, temporary multi-model pdb's were used to generate multiple chemical shift tables which were stored while the multi-model pdb's could be deleted after chemical shift calculations since disk space in the analysis was a challenge. Statistical analysis scripts were written and multidimensional arrays were used to store and retrieve relevant data from the multi-model chemical shift data files accurately and reliably.

In order to plot time-averaged NMR spectra, average chemical shifts for each atomic type over 1 ns were computed for each residue of the model. UNIX C-shell scripts and NMRWish TCL scripts were created and executed sequentially. As shown in Figure 4-1, *###_1ns_combinedfid.pl* scripts allow the extraction and simultaneous calculation of average chemical shifts by calling *pdb2nmr.pl*. Predicted RDC's were converted to spectra by introducing the appropriate coupling while augmenting and diminishing the chemical shifts by calling *pdb2rdc.pl* also from within *###_1ns_combinedfid.pl*. Using the simulation-to-spectra protocol described in the Figure 4-1 flow diagram, respective HSQC spectra are plotted from the resulting data.

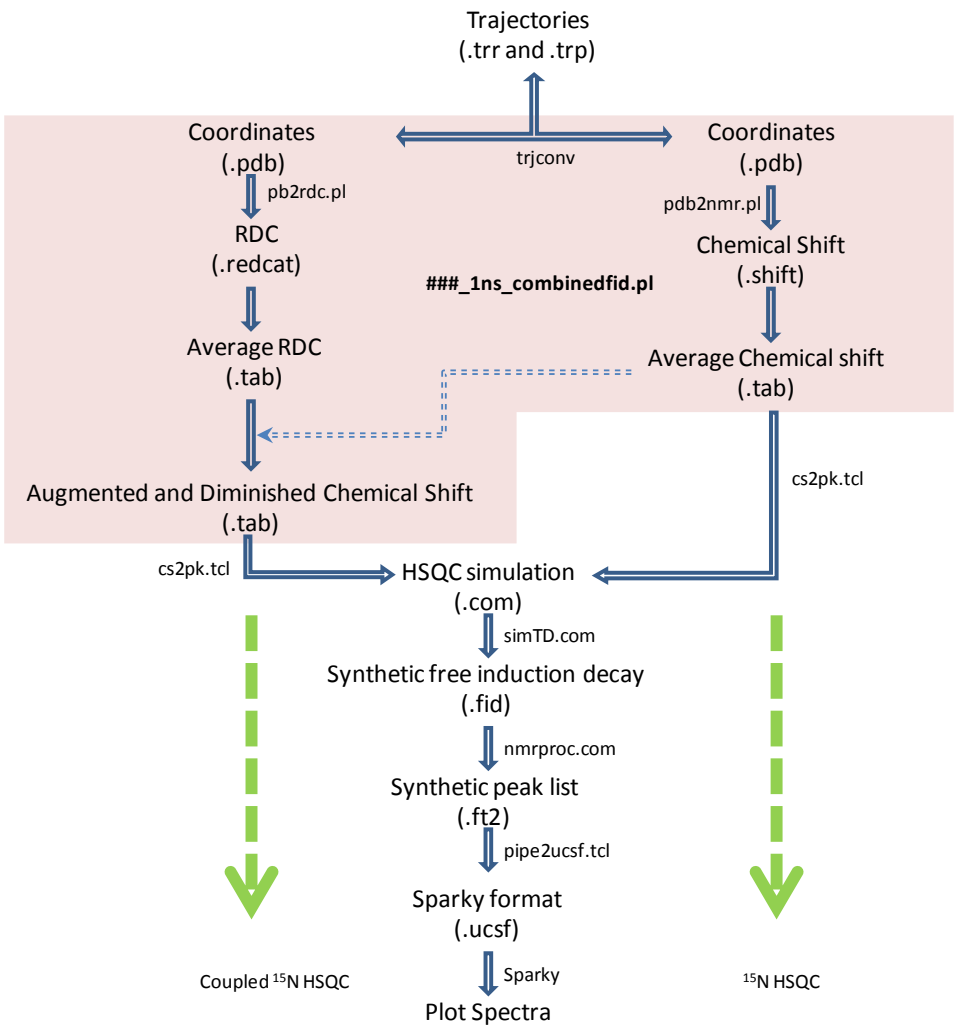


Figure 4-1: Block diagram for the generation of synthetic spectra from MD trajectories.

The average chemical shift tables in CamShift format were converted to NMRPipe format by the *cs2pk.tcl* script which prepared the output simulation script (Delaglio, 2001). Simulation of N-edited HSQC experiments with appropriate exponential decay and phase correction was specified by the *simTD.com* and *nmrproc.com* scripts respectively (Delaglio, 2001). NMRDraw and Sparky, from the NMRPipe suite of programs, were used to generate synthetic spectra from the calculated chemical shifts. NMRDraw allows us to produce a synthetic free induction decay (FID) for a ^{15}N-edited HSQC spectrum by running the simulation script with the formatted chemical shift data. This FID is then convoluted by a fourier transform specified by the *nmrproc.com* script to generate a synthetic peak lists for

the HSQC spectrum. The convoluted FID data is then converted into .ucsf format from pipe format in order for all the peaks in the synthesized spectrum to be visualized and phased within Sparky.

4.2.2 Back-calculation of Residual Dipolar Coupling

Back-calculation of the RDCs for a particular structure can be easily done using the residual dipolar coupling analysis tool, REDCAT (Valafar & Prestegard, 2004). From the coordinate file that gives information of the internuclear vectors and an arbitrarily chosen alignment tensor that defines the orientation of the molecular frame with respect to the laboratory frame, RDC data for the internuclear vector can be calculated. Predicted RDC data validated with experimentally determined data can be used to monitor protein dynamics.

A representative conformer from a single model was used to prepare a REDCAT input file that was set-up to calculate dipole moments of NH internuclear vectors from valid residues. A principle order tensor $\mathcal{S}_{xx}, \mathcal{S}_{yy}, \mathcal{S}_{zz}$ and corresponding Euler angles were used to define the alignment tensor and the orientation of the alignment frame with the molecular frame. Co-ordinate files in PDB format were extracted from the molecular dynamics trajectories and the models were used to define the orientation of the internuclear vector of valid residues outlined by input file. The internuclear NH vectors of each residue were calculated with respect to the molecular frame. The internuclear vectors were used to calculate the RDCs of the protein in the conformation adopted at the particular point of the molecular dynamics simulation. Analysis of how the theoretical RDC data changes with time can even give insight into processes that occur during the course of the molecular dynamics. To do this we used standard UNIX commands (*cat*) to concatenate the RDC output files and pipe them to a filter which globally searches for a regular expression (*grep*) matching a specific amino acid. This prints out the order tensor and Euler angels which describe the internuclear vector and its orientation to the alignment tensor at each 0.5 ps interval of the simulation. The RDC data along with the parameters that define the internuclear vector were stored in a REDCAT data file containing multiple sets of data.

Synthetic time averaged HSQC without decoupling of hydrogen nuclei were plotted to show the effects of $^{1}J_{NH}$ scalar coupling with residual dipolar coupling in simulated anisotropic

media. The coupled cross peaks were obtained by converting the calculated RDC into a total coupling contribution assuming the $^{1}J_{NH}$ scalar coupling to be constant throughout at 90 Hz, for future studies this could be done to a greater degree of accuracy by calculating the $^{1}J_{NH}$ scalar coupling based on the instantaneous NH geometries. This coupling was converted into parts per million when $\mathbf{B_o}$ is 600MHz. The chemical shifts were augmented and diminished by half of this coupling appropriately and the two resulting synthetic FID's were generated before the cross peaks were plotted separately and overlapped in SPARKY. The combination is the simulated coupled ^{15}N HSQC in oriented media. See Supplementary data for the script.

4.2.3 T_1 relaxation and τ_c from MD simulations

Based on work by Gasyna who determined the spin-lattice relaxation time of ^{13}C using inversion recovery, we calculated expected T_1 relaxation times of ^{15}N nuclei during the simulated dynamics of our models (Gasyna & Jurkiewicz, 2004). Rotational reorientation, τ_c, has been shown to have a strong dominance over the observed T_1 due to its influence on fluctuations of dipole-dipole interactions. We used τ_c as a parameter to define the expected T_1 relaxations of ^{15}N-^{1}H internuclear vectors as

$$\frac{1}{T_1} = N\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2 \gamma_N^2 \gamma_H^2}{r_{NH}^6}\ \tau_c \qquad (4.4)$$

Where μ_0 is the permeability of free space set at 1.25663706 x 10^{-7} Hm^{-2} and N is the number of directly bonded hydrogen atoms, set to 1 for amides. γ_N and γ_H are the gyromagnetic ratios of nitrogen and hydrogen respectively while $r_{NH}$is the N – H distance (Hertz, 1967). The r_{NH} and the τ_c were obtained from the atomic coordinates contained in the trajectory files where the correlations were measured between successive models of the simulation. By employing the Math::Trig module within the Perl script the r_{NH} was calculated based on the length of the vector between the coordinates of the two atoms within the map. The τ_c was found from the rate of change of the angle between this vector in successive intervals defined in radians per second. To do this the angular difference in the vectors over the time between intervals was calculated and expressed in seconds. For the sake of simplicity we did not consider the effects of translation motions on the τ_c within the

simulation. For each residue the τ_c calculated was expressed as an average over the entire length of the simulation before it was used to calculate the spin – lattice relaxation T_1.

4.3 RESULTS AND DISCUSSION

4.3.1 Chemical Shift Prediction

Predicted chemical shifts of structures at specific time intervals were generated in order to monitor their variability within the course of the molecular dynamics simulation, and to better gauge a predicted chemical shift on a longer timescale. Statistical analysis scripts written in perl were used to calculate standard deviations and the average chemical shift observed for each ^{1}H-^{15}N nucleus pair over the duration of the dynamics simulation. A multidimensional array was populated and used to store and retrieve shift data readily from each residue at every interval of the simulation. These chemical shift averages and deviations were calculated from data contained in this array. In order to analyse the backbone chain dynamics we extracted and calculated chemical shift standard deviations for the Hα atoms. Deviations in chemical shift can infer a unique NMR perspective on structural fluctuations and dynamics because CamShift predictions are determined primarily from the rapid measurement of inter-atomic distances.

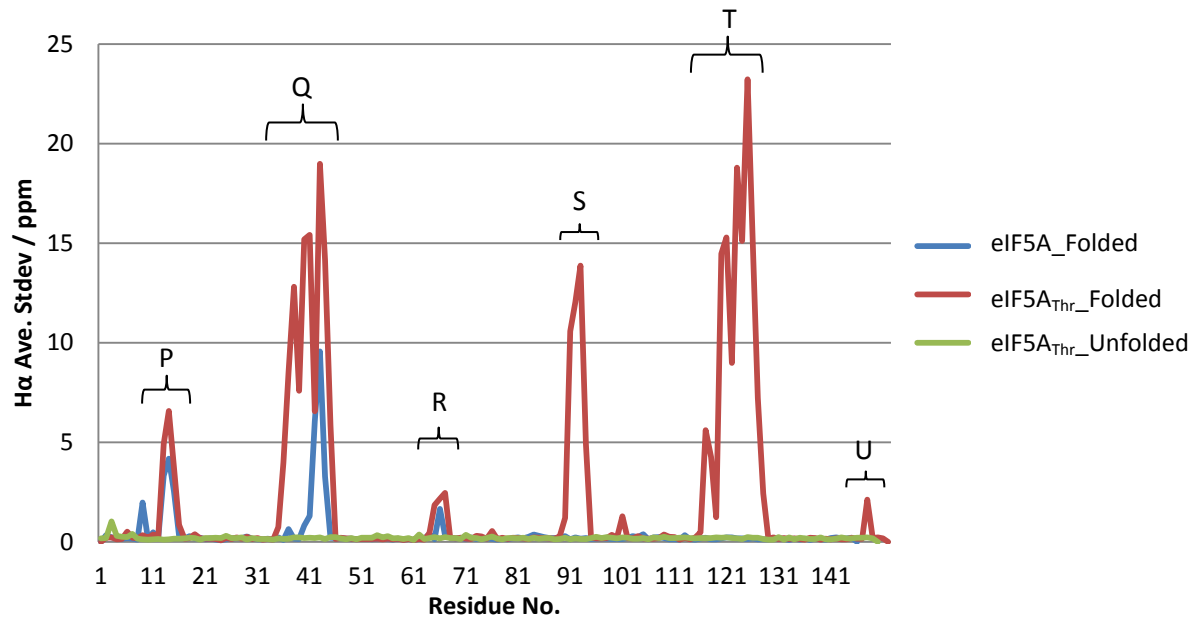


Figure 4-2: Hα Chemical Shift standard deviations for folded and unfolded eIF5AThr and folded eIF5A after 1 ns dynamics.

For the purpose of discussion it is convenient to label certain flexible regions of the protein structure P(9 - 16), Q(33 – 43), R(63 – 68), S(90 – 95), T(117 – 128) and U(148). From the chemical shift standard deviations of dynamic simulations calculated over 1 ns, from Figure 4-2, the eIF5A$_{Thr}$ folded model (red) shows the most significant chemical shift deviations and between 81 – 141 deviation pattern does not match the eIF5A model structure. Whether these observations give indications that introduction of the LVPRGS site reduced macromolecular stability is worth exploring further through studies that investigate structural integrity between different models such as perturbation response scanning (Atilgan & Atilgan, 2009).

The eIF5A$_{Thr}$ model experiences the same pattern of deviation as eIF5A within the P, Q and regions. The P region matches the β strands of the sandwich domains of the N-terminals which experience weak intermittent hydrogen bonding. The first 10 residues preceding the P region experience limited deviation because of their increased disorder which enables them to adopt random coil dynamics with relatively small or negligible deviation due to

effects of surrounding protein. The 10 residue stretch region of the Q region matches the conserved loop region that contains the lysine group where hypusination will ultimately occur. It is possible that eventual hypusination of the highly mobile lysine group in this protruding region enhances the reach of this domain and, combined with its flexibility, allows it to lock onto the corresponding hypusine residue from a second monomer enabling dimerisation. Mutations within this neighbouring region which have allowed hypusination but limited the functionality of eIF5A, perhaps limited functionality by influencing the mobility of this region as a whole, reducing the likelihood of dimerisation and thus lowering the presence of active dimeric hypusinated protein (Charlton, 2012; J. H. Park et al., 2011; C F Zanelli & Valentini, 2007). Experimental chemical shift experiments could shed light into the implications of these observations as an alternative way of tracing relative mobility of the Q region with different mutations, but this would require hypusinated protein.

The R region shows relatively minor chemical shift deviations for the residues that extrude out of the loop region joining β strand 4 and strand 5 within the SH3 β barrel N-terminal domain (Figure 4-2 and Figure 4-3). Inspection of the dynamics suggested that the deviation observed resulted from the movement of the N and C terminal domains relative to each other through the flexible linker region, consistent with previous studies (Kim et al., 1998). The strength of hydrogen bonding within the β sheet fold of this region limits internal motions, but there may be a contribution to chemical shift deviation if rapid side-chain motions are considered (typical of Cα – Hα bonds). A chemical shift deviation of similar magnitude to the R region is observed in the U region of the mutated eIF5A$_{Thr}$ monomer but this observation is absent in the native monomer. The source of this deviation may arise due to the mobility of the loop region between U region and the rest of the protein. As this loop region is free to rotate and move it occasionally brings the residues of the U region into close proximity with the rest of the protein.

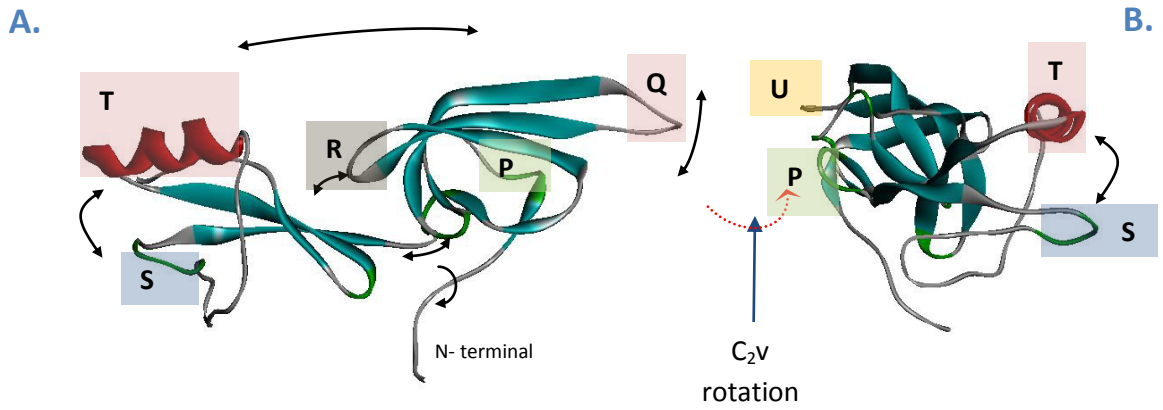


Figure 4-3: Model of eIF5A$_{Thr}$ showing domains of deviation. Structure B shows a view after a C_2V rotation of A, depicting the wobble of δ with respect to ε domain residues.

Other regions of deviation that are absent in the native monomer but obvious in the mutant are those observed in the S and T regions. As seen from Figure 4-3 the S region can move towards the T region residues and thus cause a chemical shift perturbation influencing their deviation. From Figure 4-2, the deviation observed in the T region affects more residues than the S region inferring that this is not the only source of variation. As seen in Figure 4-3 the T residues all lie within the α helical region of eIF5A$_{Thr}$ which exhibit fluctuations in chemical shift that are absent in the native model. From the extent of deviation it can be inferred that this is an unstable region with significant fluctuations in chemical environment observed during the simulation. This degree of deviation is not observed in the native eIF5A and is not expected in most α-helices, which are stabilised by extensive inter-residual hydrogen bonding (Lindhout et al., 2004). We infer that this deviation represents an area of instability with eIF5A$_{Thr}$ within the helix during molecular dynamics.

In order to plot spectra that mimic the observed ^{15}N HSQC experiment, average H and N chemical shifts were calculated and used to plot time averaged NMR spectra.

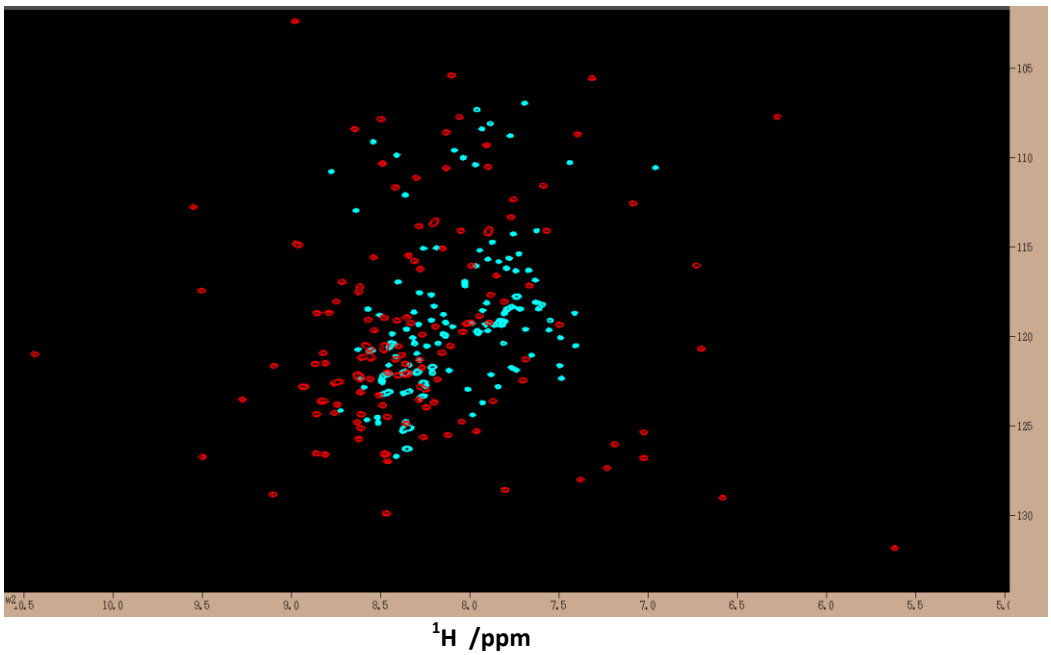


Figure 4-4: Overlaid ^{15}N-edited HSQC synthetic spectra. eIF5A$_{Thr}$ folded (red) and unfolded (cyan) from time-averaged predicted chemical shift data.

The overlay spectra in Figure 4-4 from the folded, ordered, and the unfolded, disordered, conformations of the mutated proteins show a distinct difference in their chemical shift patterns. The unfolded conformation, coloured in cyan, consistently resembles the random coil chemical shift pattern where proton resonances of the degenerate cross peaks occur within a 7 – 9 ^{1}H ppm cluster region. The chemical shift pattern observed from the folded conformation shows a distinctively dispersed scatter of chemical shifts. The chemical shifts are clustered between 6.5 and 9.5 ppm with some outliers as far as 3.0 and 12.0 ppm, not shown.

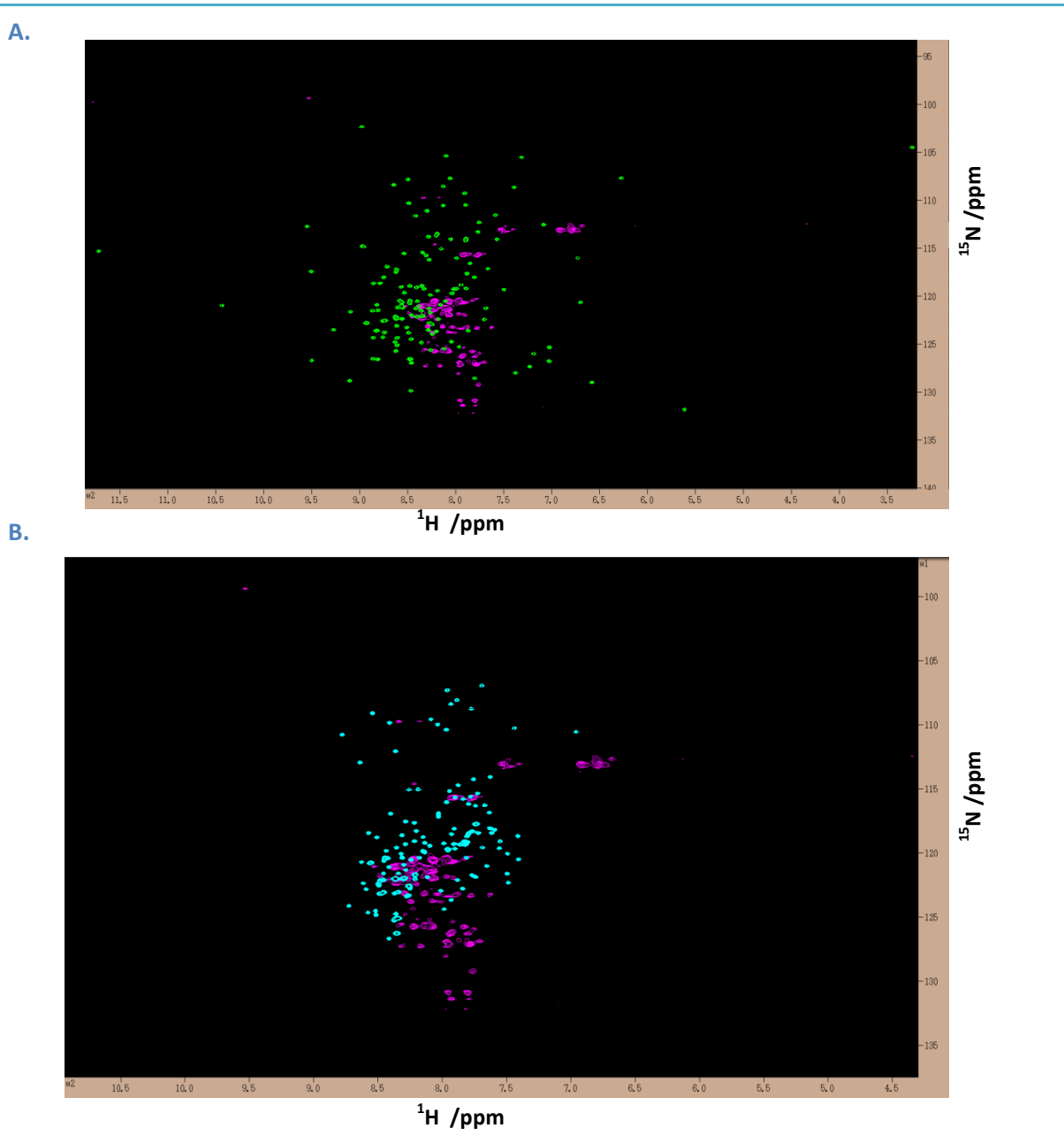


Figure 4-5: Overlaid spectra for time averaged eIF5A$_{Thr}$ models and experimental data (purple). Panel A: eIF5A$_{Thr}$ folded (green) and experimental data (purple). Panel B: eIF5A$_{Thr}$ unfolded (cyan) and experimental data (purple).

The predicted time-averaged spectra, over dynamics, for folded (Figure 4-5 A) and unfolded (Figure 4-5 B) conformations of eIF5A$_{Thr}$ were displayed overlaid over the experimentally observed spectra. Figure 4-5 B shows that the experimental data (purple) matches the chemical shift pattern for the unfolded conformations (cyan) of eIF5A$_{Thr}$ somewhat better

than the scattered folded conformation (green), Figure 4-5 A. The experimental data appeared to be slightly shielded in their ^{1}H resonances compared to the predicted shifts. An obvious deshielding of 5 ppm and in some cases 10 ppm is observed for the ^{15}N resonances in Figure 4-5 B of experimental compared to predicted data. A small portion of this deshielding might be attributed to solvent effects from the 0.1M phosphate ions in the experimental sample buffer, but really points towards differences in the CamShift method with respect to the this particular experimental system. Solutes within the buffer can stabilize the resonance of the amide bonds enhancing the deshielded state (Mujika et al., 2006).

In the development of CamShift and later versions, the effects of solvents, δ_{solv}, were investigated. They found that the contribution of δ_{solv} on Hα chemical shifts within a range of 5 Å was negligible (Kohlhoff, 2008). The CamShift model thus does not include solvent effects and hence our predicted spectra were limited as result. Although during our MD simulation we chose a Particle-Mesh Ewald (PME) method to calculate the influence of long-range electrostatic interactions, we chose to solvate our proteins by TIP3P water molecules without any additional solutes and charges. We did not account for precise electrostatic effects by the solvents on the dynamics of our protein. A more precise model during dynamics would include additional solute ions within the solution at a concentration comparable to the sample buffer solutions.

Despite these interventions the influence of these solutes on the chemical shifts would be overlooked due to lack of incorporation into the CamShift model in generating predicted chemical shifts. An alternative program that could be used to optimize the observations by considering electrostatics of solvent effects on chemical shifts is the Random Forest algorithm. In this algorithm additional terms which include those that compute solvent exposure are fitted into the parameters that influence the predicted chemical shift (Breiman, 2001; Hamelrych, 2005). Despite the inconsistencies introduced by the dynamics and the absence of the solvent effects on chemical shifts, the CamShift predicted spectra performs reasonably well. The spectra observed provided added evidence to the conclusions from Chapter 2 which propose that the protein produced by induction at 37 °C was largely unfolded.

4.3.1.1 Steered Molecular Dynamics

Although the overlaid chemical shift spectra for the unfolded folding simulations (Figure 4-5 Panel B) showed a close match to the experimental chemical shifts, it was thought that by unfolding the folded conformer an improvement to the match could be obtained. Figure 3-6, from Chapter 3, shows the progression of folding observed during the *in silico* molecular dynamic simulations of the unfolded model. The appearance of secondary structure within the unfolded 1 ns model (Figure 3-6), justifies a theory in support of folding occurring during the simulations. By undertaking an interactive steered molecular dynamics procedure to simulate unfolding of the folded model an unfolded intermediate conformation that matched the experimental data more efficiently can be accessed.

The GROMACS simulation package was used to perform the steered molecular dynamics based on the Lemkul and Bevan (2010) approach. The parameter files were adjusted in order to allow a force constant of 1000 kJ $mol^{-1}nm^{-2}$ on the terminal Asp 152 residue maintaining a constant pulling velocity of 10 nm per ns in the z-dimension only. Umbrella sampling was incorporated in order to calculate the acceleration and thus a potential of mean force of the system during each 0.002 ps time-step interval. Periodic boundary conditions were permitted in all directions while the long-range dispersion correction of (Lemkul & Bevan, 2010). The steered molecular dynamics simulations proceeded successfully and were able to simulate 790 ps of solution dynamics before boundary box restrictions terminated the simulations.

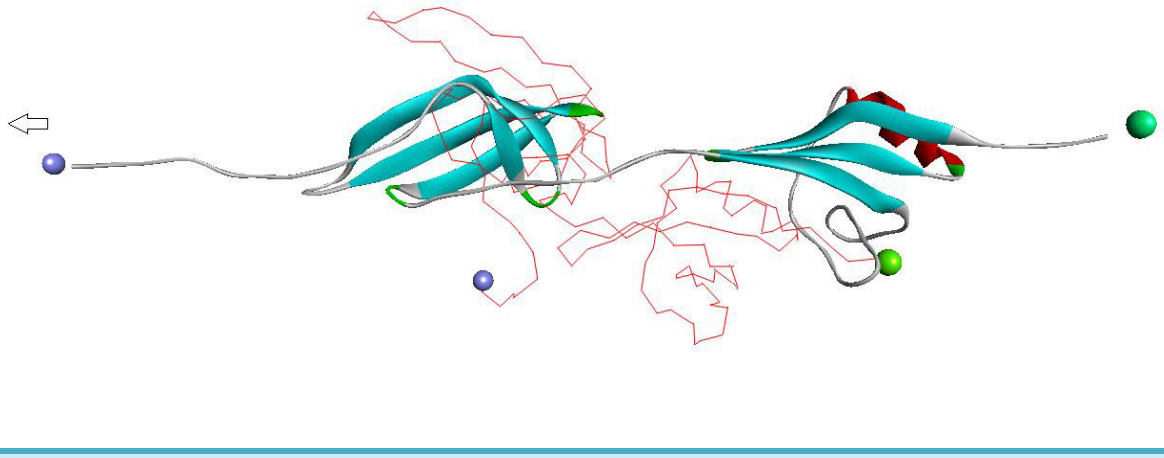


Figure 4-6: eIF5A$_{Thr}$ steered molecular dynamics conformations. Folded eIF5A$_{Thr}$ at start of steered dynamics (coloured as red, structure displayed as wire). Unfolding folded eIF5A$_{Thr}$ after 790 ps (coloured as secondary structure, structure displayed as ribbon). Green spheres (ASP 152) are reference points for pulling; Purple spheres (ASP 1) residues that are pulled by force constant indicated as white arrow.

Figure 4-6 illustrates the structures of the PG57 folded model under the 1000 kJ mol-1 nm-2 force constant. The radius of gyration as shown in Figure 4-7 A increased by 50 % from 2 nm to 3 nm in 790 ps of steered simulations. Figure 4-7 B shows that the potential energy during the simulations was stable and fluctuated at -2.06 x 10^7 kJ/mol.

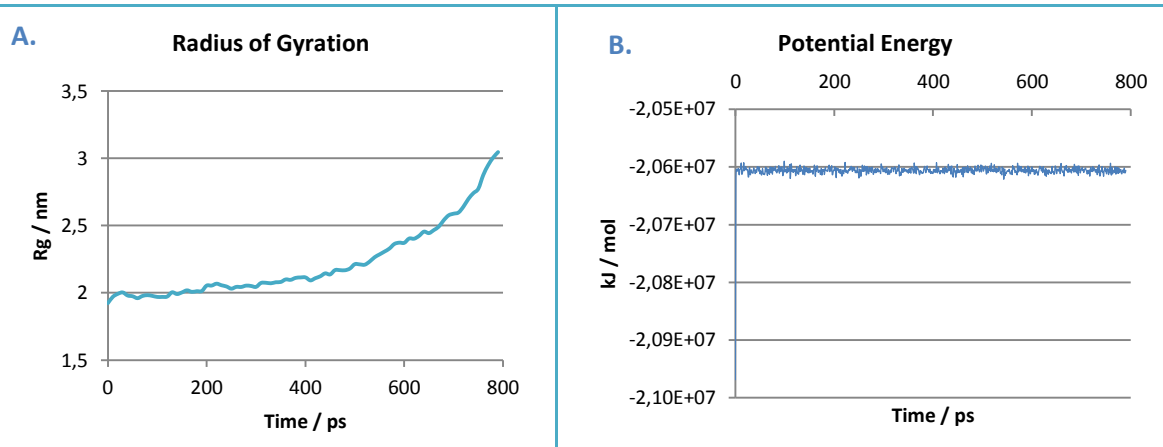


Figure 4-7: Steered Molecular Dynamics Analysis. Panel A: Radius of gyration over 790 ps of simulation. Panel B: Potential Energy of Protein over the 790 ps of simulation.

Conformations of the unfolding model were extracted in frames predicted chemical shifts of each conformation were extracted using CamShift before the synthetic FID's were calculated in NMR Pipe and spectra plotted in Sparky as outlined above. In order to analyse the change in the chemical shift signatures average chemical shifts were obtained for each 80 ps time interval during the simulation. Spectra from the first 80 ps and the last 80 ps were extracted and overlaid over the experimental HSQC.

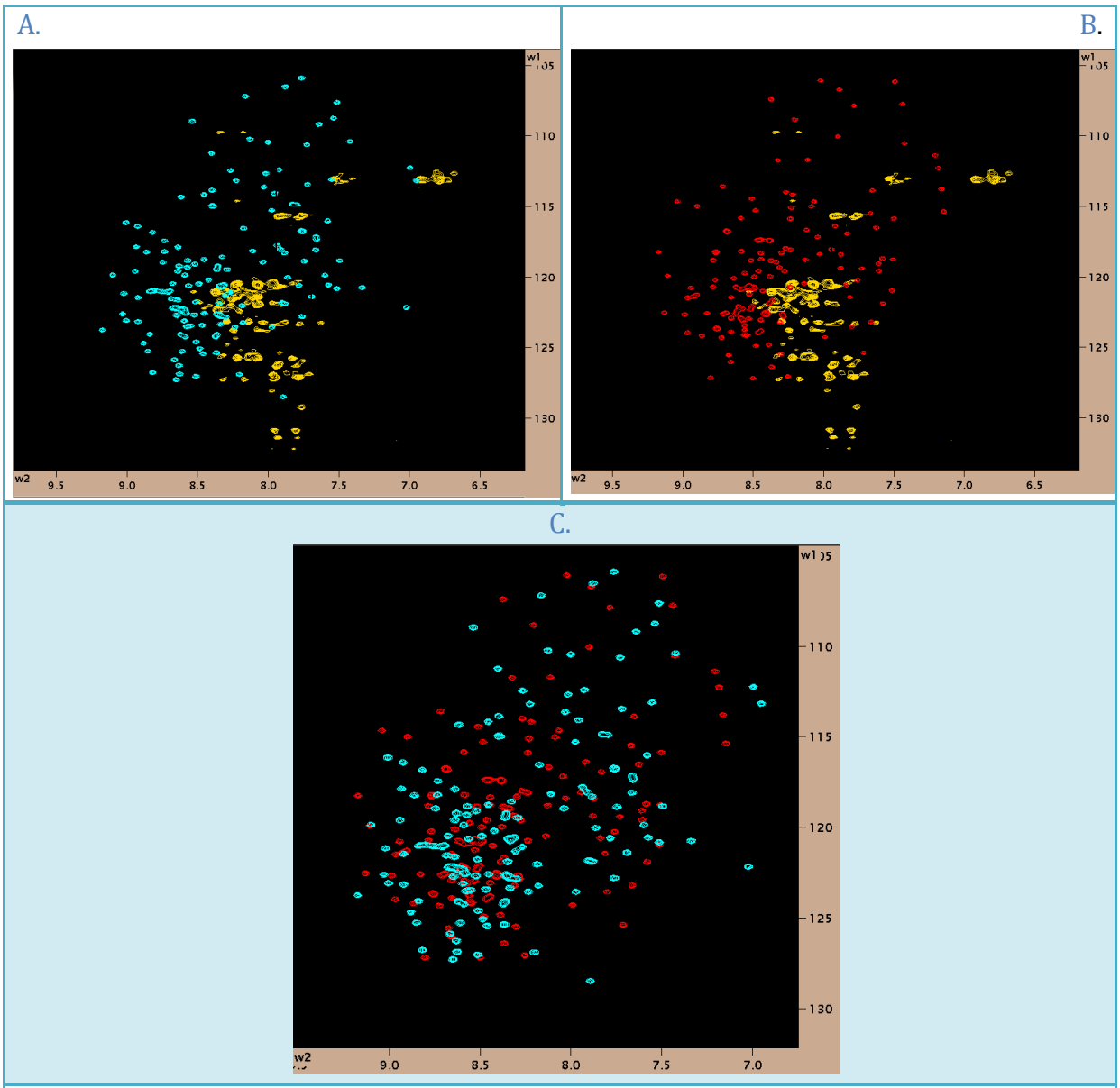


Figure 4-8: Overlaid ^{15}N-HSQC spectra. Panel A: Average predicted spectra from structures obtained from first 80 ps simulation (cyan) overlaid experimental HSQC observations (gold). Panel B: Average predicted spectra from structures obtained from the last 80 ps of simulation (red) overlaid experimental HSQC observations (gold). Panel C: Average spectra from first 80 ps (red) overlaid an average spectrum from last 80 ps (cyan). Sparky was used to display spectra and edit contours for visualisation.

After 790 ps of unfolding Figure 4-8 B and C show that the unfolding of eIF5A$_{Thr}$ in solution makes the ^{1}H ppm chemical shift range narrower compared to the fully folded model. The partially unfolded model Figure 4-8 A appears to be more dispersed than the experimentally observed spectra. This is consistent with a fully folded model. We propose that extended unfolding steered molecular dynamics could allow us to access an intermediate conformation that matches the experimentally observed spectra. In our situation the best match for the experimental data was obtained from the unfolded folding scenario.

4.3.2 Residual Dipolar Coupling

The principle diagonal order tensors $\mathcal{S}_{xx}, \mathcal{S}_{yy}, \mathcal{S}_{zz}$ were arbitrarily chosen to define the alignment tensor as 2e-4; 5e-4, and -7e-4 respectively while the Euler angles which define the orientation were set at -40°, -50° and 60°. In our studies the maximal splitting resonances when the vector aligns perfectly with the magnetic field, D_{max}, is modulated by the distance of the nuclei. The nuclei specific parameters, $\mu_0\gamma_1\gamma_2 h/8\pi^3$, for ^{15}N and ^{1}H protons is -24.35 kHz where the distance *r* is measured from the co-ordinate file of the conformation at each consistent time interval. The internuclear vectors were used to calculate ^{15}N – ^{1}H RDCs for every residue after each interval within the 1 ns molecular dynamics simulation.

Fluctuations of the RDC arise due to variations in the interaction of the vector with the magnetic field influenced by the molecular tumbling and stretching vibrations inherent with the identity of the residue and its chemical environment. In isotropic media the effects of the fluctuations arising due to molecular tumbling will result in an overall zero contribution to the observed coupling. In anisotropic media however fluctuations due to molecular tumbling are reduced resulting in an observable shift related to the cumulative coupling, combining with the scalar coupling due to interaction with the magnetic field. Experimental measurements of this dipole represent a time averaged moment but however back calculated data (i.e. predicted RDC data) can give an instantaneous value which can be extended to picoseconds (ps) time scales in molecular dynamics. These instantaneous and ps timescale RDC's are dependent on the interaction of the vector with the alignment tensor

as well as the internuclear distance between the ^{15}N and ^{1}H. Time averaged fluctuations observed over µs and ms intervals, in experiments, can be used to probe slow macromolecular dynamics making experimental investigations of previously invisible, excited, intermediate states possible (Montalvao et al., 2012).

Statistical analysis programs written in perl were used to prepare, access and analyse the RDC data. Average standard deviations experienced over 1 ns of simulation for each residue of the unfolded and folded eIF5A$_{Thr}$ and the folded eIF5A without the thrombin cleavage were plotted below with the average deviation indicated.

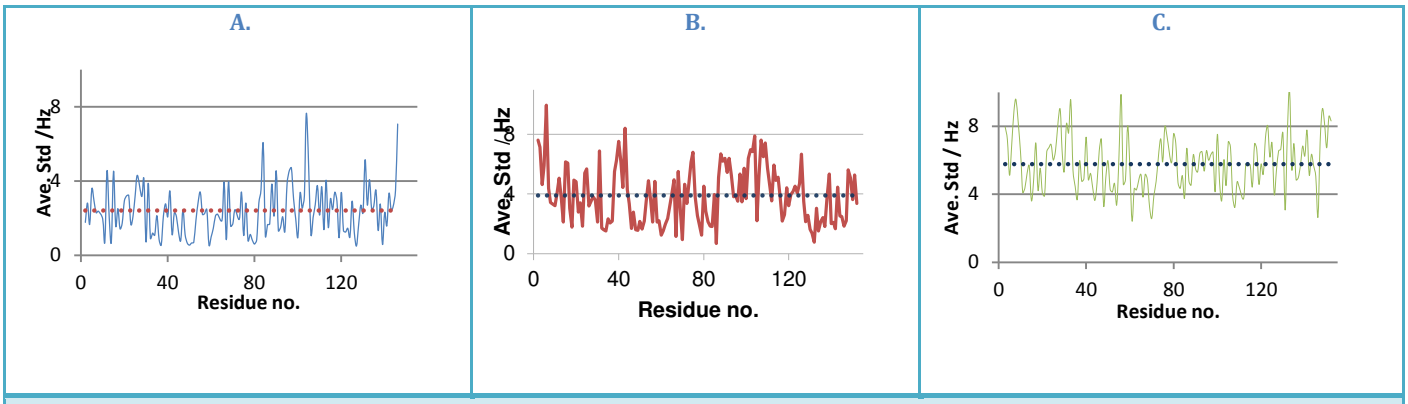


Figure 4-9: Predicted residue RDC standard deviation. Panel A - eIF5A folded; Panel B - eIF5A$_{Thr}$ folded; Panel C - eIF5A$_{Thr}$ unfolded residue. Dotted line indicates average macromolecular standard deviation.

From the plot of the residue specific average deviation over 1ns of simulation the differences in the behaviours of the internuclear vectors can be observed. It can be seen that different residue types have different average standard deviations observed over 1ns of simulation. These differences arise from fluctuations associated with the different vibration modes of the amide bonds and their chemical environments. Different amino acids allow different degrees of freedom due to the resonance stabilization of the amide bond and steric effects from the side chains. The exact hydrogen bonding present has a partial influence on the magnitude and fluctuation of the residual dipolar coupling. Perpetual inter-residue hydrogen bonds stabilize the vector in a particular orientation while transient bonding with solvent waters results in rapid fluctuations due to inherent instability.

The stability of the ordered structures decreased the degree of fluctuations of the NH internuclear vectors. The folded native protein, Figure 4-9A, observed the smallest residue average deviation of 2.41 Hz of residual dipolar couplings while the mutated folded model, Figure 4-9B, observed an average of 3.97 Hz. For the two mutant models with identical residues, the degrees of fluctuation of each of the residues over 1 ns of simulation was different. The unfolded model, Figure 4-9C, experienced the largest residue average deviation of 5.7 Hz. The propensity for random coil dynamics and intermittent hydrogen bonding with solvent waters resulted in the larger deviation observed for the unfolded conformation over 1 ns of simulation. Increased order and steady hydrogen bonding involved with the N-H vectors resulted in smaller average deviations for the ordered folded models.

The interaction of the N-H vector with the alignment tensor relative to the other vectors in the protein does not change rapidly due to structural integrity within the 1 ns timescale while the contributions from rapid vibrations are noticeable over ps timescales. The RDCs of a query amino acid from eIF5A and eIF5A$_{Thr}$ folded proteins were extracted and dynamic data over 1 ns was plotted as shown below for further scrutiny.

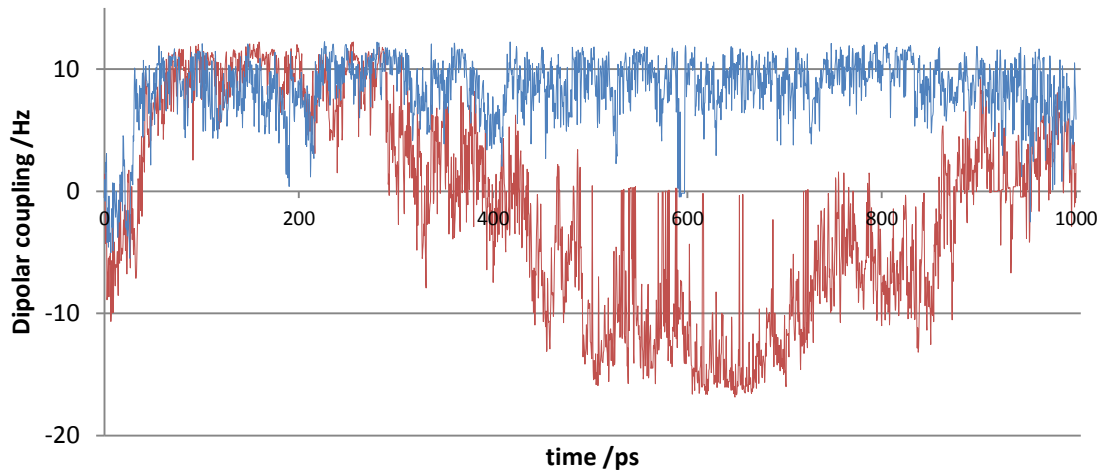


Figure 4-10: Residue 43, β domain, ^{1}H-^{15}N RDC over 1 ns MD simulation. eIF5A$_{Thr}$ (Red) and eIF5A (blue).

Analysis of the oscillations of the β domain lysine residue 43 from the mutated and native protein gave some insightful observations, Figure 4-10. This residue undergoes significant structural fluctuations identified from the Hα chemical shift deviations and obvious variations between the mutated model (red) and the native model (blue) Dipolar couplings are visible. The averaged standard deviation over 1ns of simulation for the mutated protein is 8.2 Hz while that for the native model is 2.4 Hz. Over 1 ns of simulation a large movement over about 600 ps is present in the mutated protein but absent in the native protein. The magnitude of the deviation of rapid fluctuations resulting from the oscillating orientations of the internuclear vector in the mutant appears to be twice the size of the native model.

A change in orientations of the internuclear vector with respect to the alignment frame is observed as a change in the RDC noticeable over longer timescales. The observed 600 ps movement is due to slow processes that are allowed in the mutant and restricted in the native. The increased intensity of the rapid fluctuations is due to the decreased order within this internuclear vector in the mutant. The decreased stability in this loop region results in intermittent hydrogen bonding and thus increased deviation as the size of the internuclear vector is also allowed to fluctuate. Apart from oscillations in the orientation of this vector, fluctuations associated with the inter-atomic distances contribute to the increased intensity of residual dipolar couplings.

The differences in the relative mobility of the hypusination consensus region between the mutated protein and the native protein are consistent with the observations made from the predicted chemical shift analysis and previous mutation studies (Cano et al., 2008; Dias et al., 2008). Decreased stability and increased mobility in this region influences the formation of the dimerized tertiary structure of the hypusinated protein.

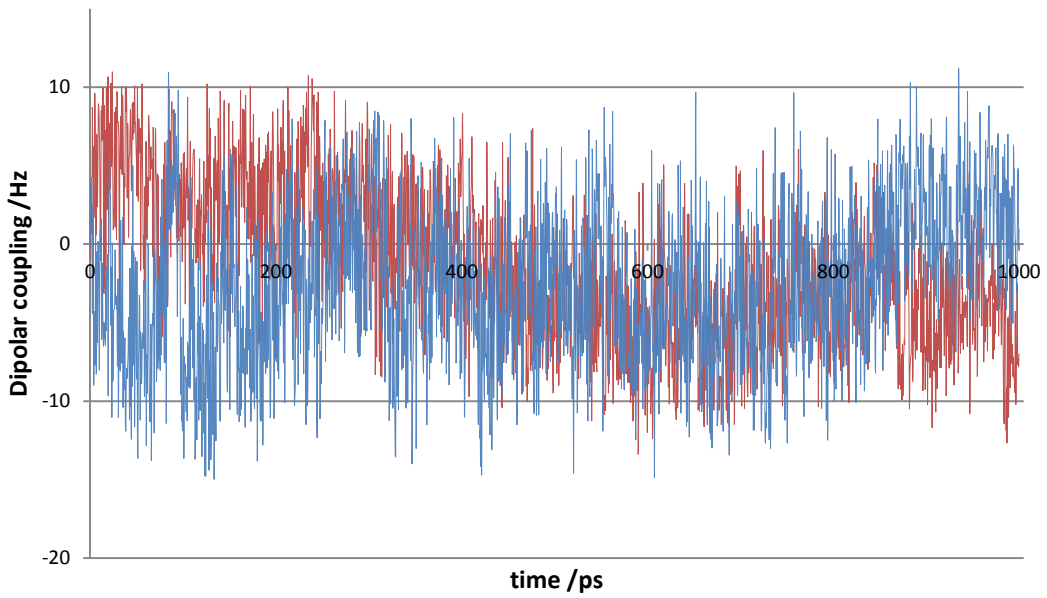


Figure 4-11: Residue 66, γ domain, ^{1}H-^{15}N RDC over 1 ns MD simulation. eIF5AThr (Red) and eIF5A (blue).

Figure 4-11 shows the residual dipolar coupling over 1 ns of simulation for the threonine 66 residue of the R region in the mutated and the native models. The mutant model experiences a slightly larger average standard deviation of 4.9 Hz over 1 ns of simulation while the native experiences an average deviation of 3.9 Hz. These observations are consistent with the observations from the chemical shift deviations where the mutant experienced a larger deviation over 1 ns. Rapidly changing hydrogen bonds between threonine 66 and the second β strand of the C terminal OB fold domain contributes to the fluctuation observed in the native and mutant models. Longer simulations can be run to analyze the slower processes in molecular dynamics associated with the movement of the 2 terminal domains relative to each other.

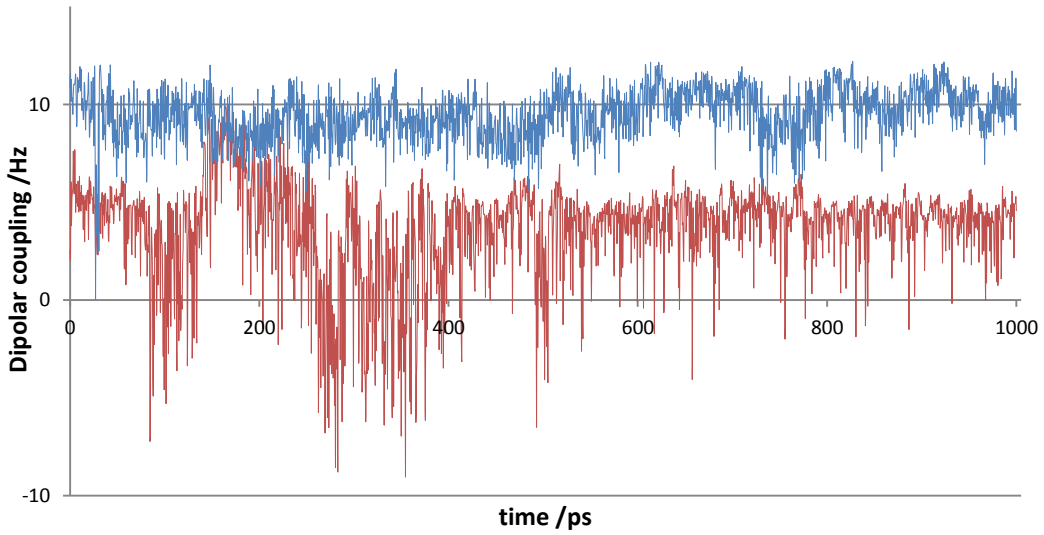


Figure 4-12: Residue 77, ^{1}H-^{15}N RDC over 1 ns MD simulation. eIF5AThr (Red) and eIF5A (blue).

Figure 4-12 shows the RDC data of 2 amino acids that are located at the same position on the models but have different identities due to the introduction of the LVPRGS thrombin cleavage site in the mutant model. Residue 77 represents a valine amino acid in the mutated protein while an arginine residue is extracted from the native protein at that position. Although their chemical shift deviations from Figure 4-2, imply similar structural fluctuations, due to their similar chemical environments, their RDC data reflect differences in the fluctuations of the N-H internuclear vectors. The average RDC over 1 ns of simulation for arginine from the native protein is 10.4 Hz while that for valine in the native model is only 3.64 Hz. This difference reflects the sensitivity of the residual dipolar coupling on the amino acid type as well as the chemical environment and orientation of the particular internuclear vector. The RDC standard deviation observed over 1ns for the arginine in the native model is 0.79 Hz while the valine from the mutated protein experiences a deviation of 2.59 Hz.

Arginine from the native protein is basic and polar and is likely to form multiple hydrogen bonds that stabilise the orientation of the N-H internuclear vector (Borders et al., 1994). On the contrary the non-polar valine experiences less hydrogen bonding and is thus less

ordered. A major contribution to the larger valine deviation is the fluctuation observed within the initial 400 ps of the simulation but absent after 500 ps. This infrequent conformational correction lasts for 300 ps. We can attribute this fluctuation to minimizations of the structure within the flexible loop region. During model building the modelling of this region introduced errors due to strains which are eventually relieved by the unrestrained dynamics.

A shift in the principle axis can cause the calculated RDC for the different amino acids in different models to be different. However because the alignment tensor observed within the R region and residue 77 is unchanged a change in the orientation of the vector and the nature of the amino acid is more likely to affect the RDC observed.

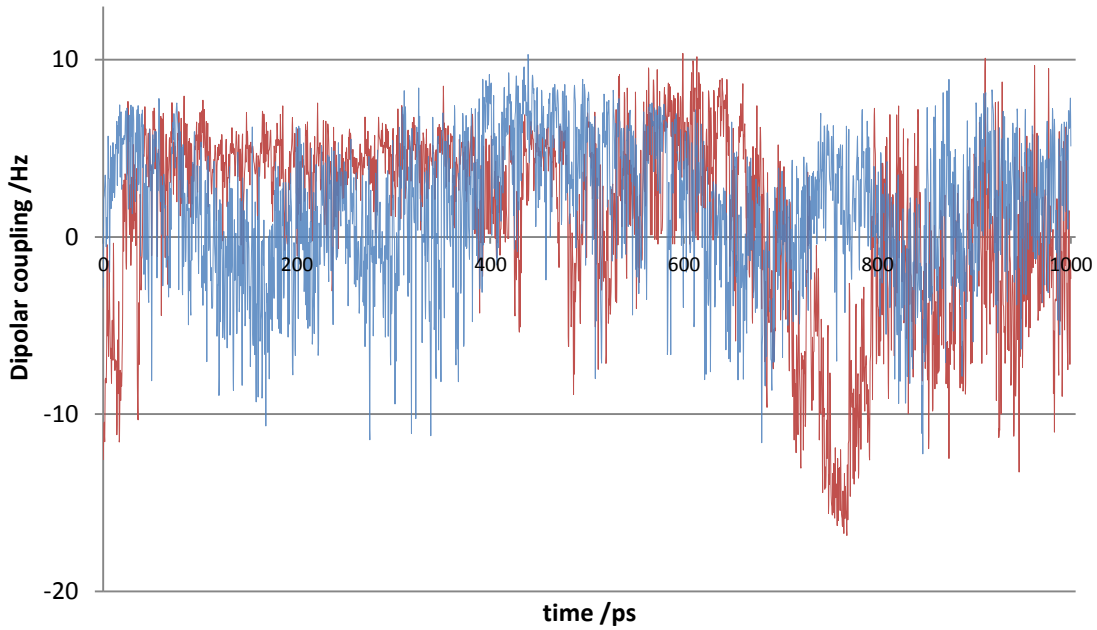


Figure 4-13: Residue 93, S region, ^{1}H-^{15}N RDC over 1 ns MD simulation. $eIF5A_{Thr}$ (Red) and eIF5A (blue).

From Figure 4-13 analysis of the S region gives insight into the nature of the dynamics observed in this region. From Figure 4-13 over 1 ns of simulation significant differences were observed in the chemical shift deviations between the native and the mutant proteins

at residue 93. This difference was attributed to the mobility and instability of the T α-helix domain which was absent in the native model. The RDC data shows that the glycine residue of the mutant protein had an average deviation of 5.3 Hz while its mean Dipolar coupling was 1.0 Hz. The native model's arginine, at a similar location, had an average Dipolar coupling of 2.7 Hz while its deviation was 2.1 Hz over the duration of the simulation. Although there is a two-fold difference observed with the deviation this does not match the nearly 100-fold difference observed from the chemical shift deviations in Figure 4-2. The difference in fluctuation observed between the chemical shift deviation and the dipolar coupling deviation can be explained by the limitation of residual dipolar coupling data in detecting long range non-bonded effects. Because the chemical shift data is distance dependant and can be influenced by fluctuations as far as 5 Å apart, they are more sensitive to structural fluctuations involving domains that are further away. The RDCs do not experience the influence of these fluctuations on the magnitude of the coupling.

The difference in deviation and magnitude of the RDC within the mutants and the native models in this region can be attributed to the instability and thus flexibility of this domain within the mutant absent in the native model. Because of the different secondary structures the mutant S region is able to move and thus affecting the orientation of the internuclear vector of residues under that fluctuation. Glycine residues are well known sources of structural flexibility due to the absence of sidechains that restrict the backbone dihedral torsions (Steinert et al., 1991). Thus it is plausible that the larger rapid dipolar coupling deviation observed within glycine 93 but absent in arginine arose from the increased degrees of freedom of these torsions and dihedrals along the models' backbone.

In order to measure the residual dipolar couplings between two nuclei of interest, an HSQC may be run without decoupling, or alternately two separate HSQC experiments are run. Where the two are run, spectra which detect the in-phase magnetizations and anti-phase magnetizations are collected from coupled experiments which employ spin-state selective filters (Andersson et al., 1998). From the difference of the cross peaks within the proton resonance in the presence and absence of alignment media the size of the residual dipolar addition to the heteronuclear $^{1}J_{NH}$ scalar coupling can be estimated experimentally.

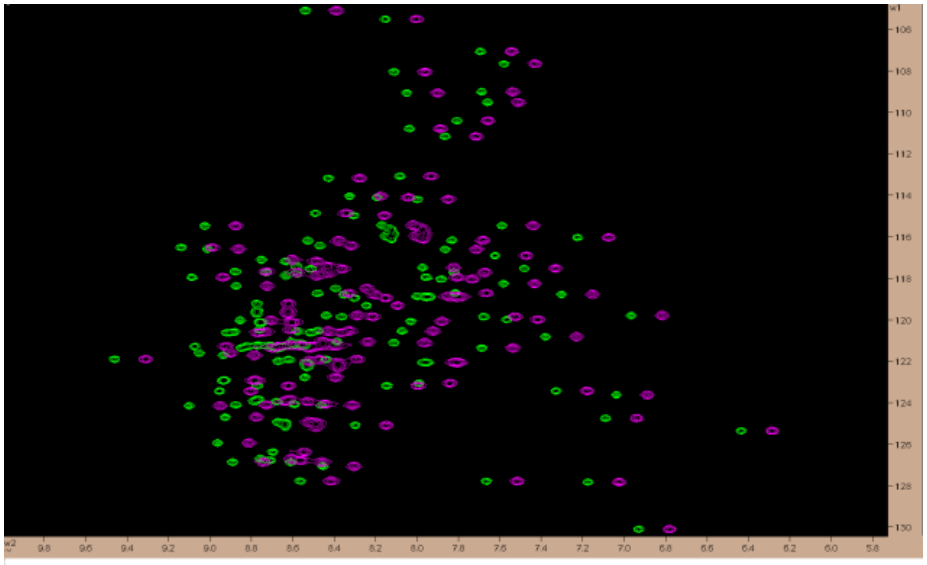


Figure 4-14: Non-decoupled ^{15}N HSQC cross peak pairs of *eIF5A* folded conformation over 1 ns, in anisotropic solution. Augmented (green) and diminished (purple) signals.

Synthetic time averaged HSQC spectra were plotted to simulate the effects of coupling of the heteronuclear $^{1}J_{NH}$ scalar coupling when the average RDC over 1ns of simulations were calculated (Figure 4-14 and Figure 4-15). The coupled cross peaks were obtained by converting the calculated RDC into a total coupling contribution when the $^{1}J_{NH}$ scalar coupling was 90 Hz. This coupling was converted into parts per million when $\mathbf{B_o}$ is 600MHz. The chemical shifts from decoupled spectra were augmented and diminished by half of the coupling appropriately and the pair of synthetic FID's were generated before the cross peaks were plotted separately and overlapped in SPARKY. The augmented proton resonance cross peak pairs can be used to describe the anti-phase and in-phase magnetization contribution to the scalar coupling of coupled heteronuclear spectra.

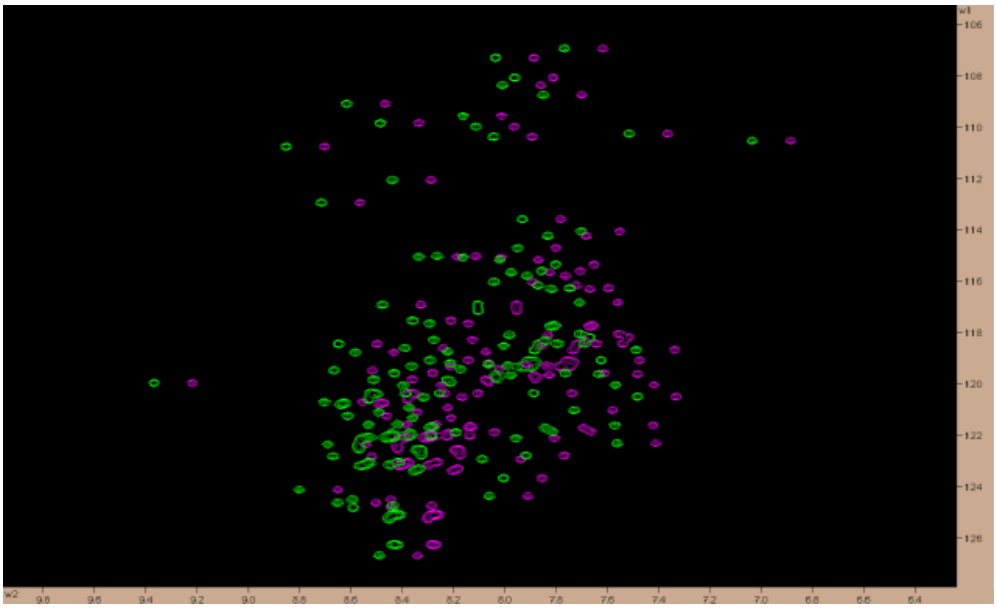


Figure 4-15: Non-decoupled ^{15}N HSQC cross peak pairs of *eIF5A*$_{Thr}$ folded conformation over 1 ns, in anisotropic solution. Augmented (green) and diminished (purple) signals.

4.3.3 T_1 relaxation and τ_c from MD simulations

A Perl script was written in order to allow the calculation of the average τ_c experienced over 1 ns of molecular dynamics simulations. The T_1 observed as a result of these fluctuations is a reflection of the efficiency of relaxation processes that occur during the 0.5 ps intervals. These processes are influenced to a larger extent by molecular tumbling although contributions from the mobility of the domain and the fluctuations of the internuclear vector should not be overlooked. The folded and unfolded mutated protein models and the folded native protein structures were analysed by the protocol and the results are described briefly below.

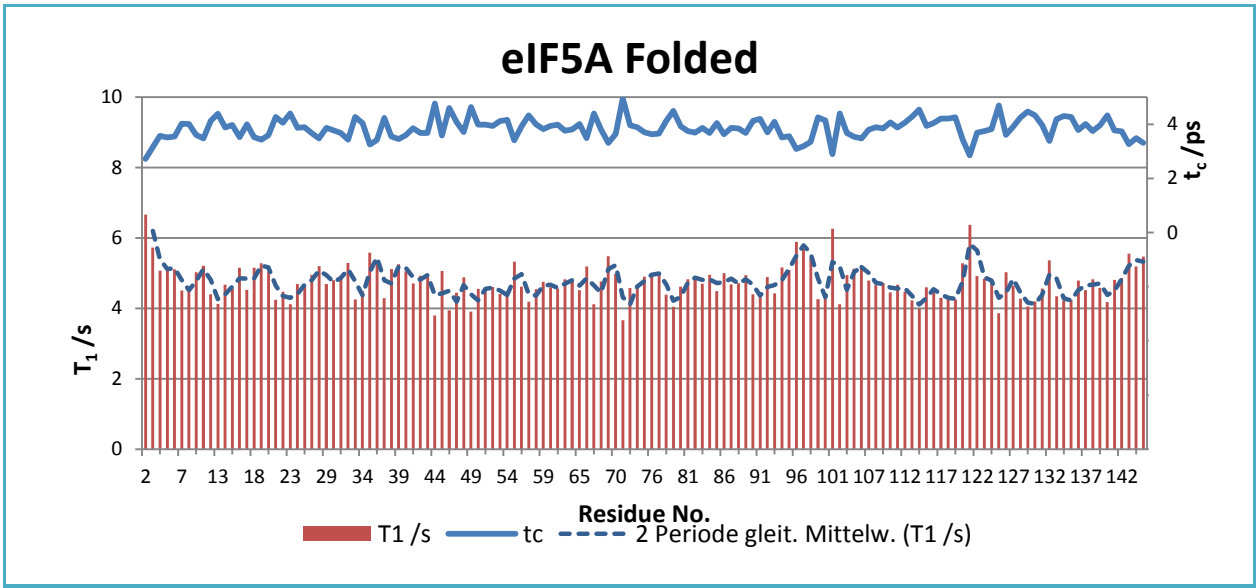


Figure 4-16: Residue specific predicted spin-lattice relaxation and correlation times of native eIF5A folded model simulations.

Non-zero spin-lattice relaxation and correlation times of non-proline residues, were calculated and the values from the folded eIF5A$_{Thr}$ are illustrated in Figure 4-16. An obvious observation from the data is the inverse relationship between the calculated τ_c and the T_1 relaxation time, traced by the 2 period moving point average consistent with equation 4.4.

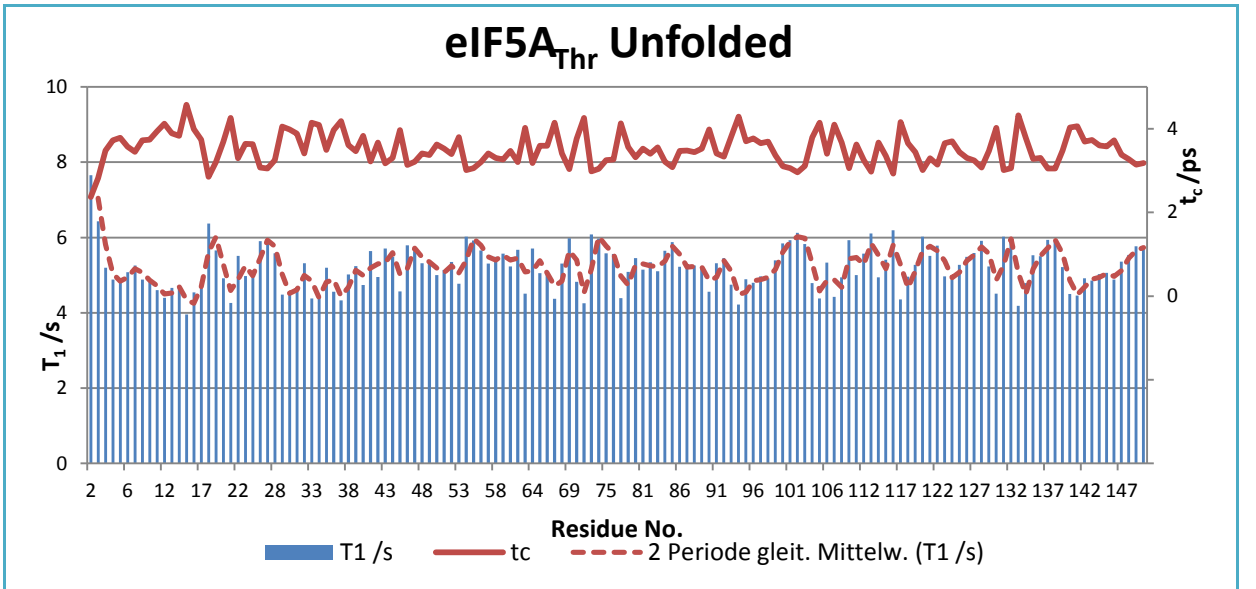


Figure 4-17: Residue specific predicted spin lattice relaxation and correlation times of mutant eIF5A$_{Thr}$ model simulations.

Figure 4-16 and Figure 4-17, the folded native model and the mutant unfolded models have a similar predicted T_1 pattern obtained from their simulated dynamics. The folded native model had an average T_1 of 4.59 s and an average τ_c of 3.73 ps while the unfolded mutant model had an average T_1 of 5.04 s and an average τ_c of 5.04 ps. Their residues experienced a similar standard deviation of 21 % for both their T_1 and τ_c.

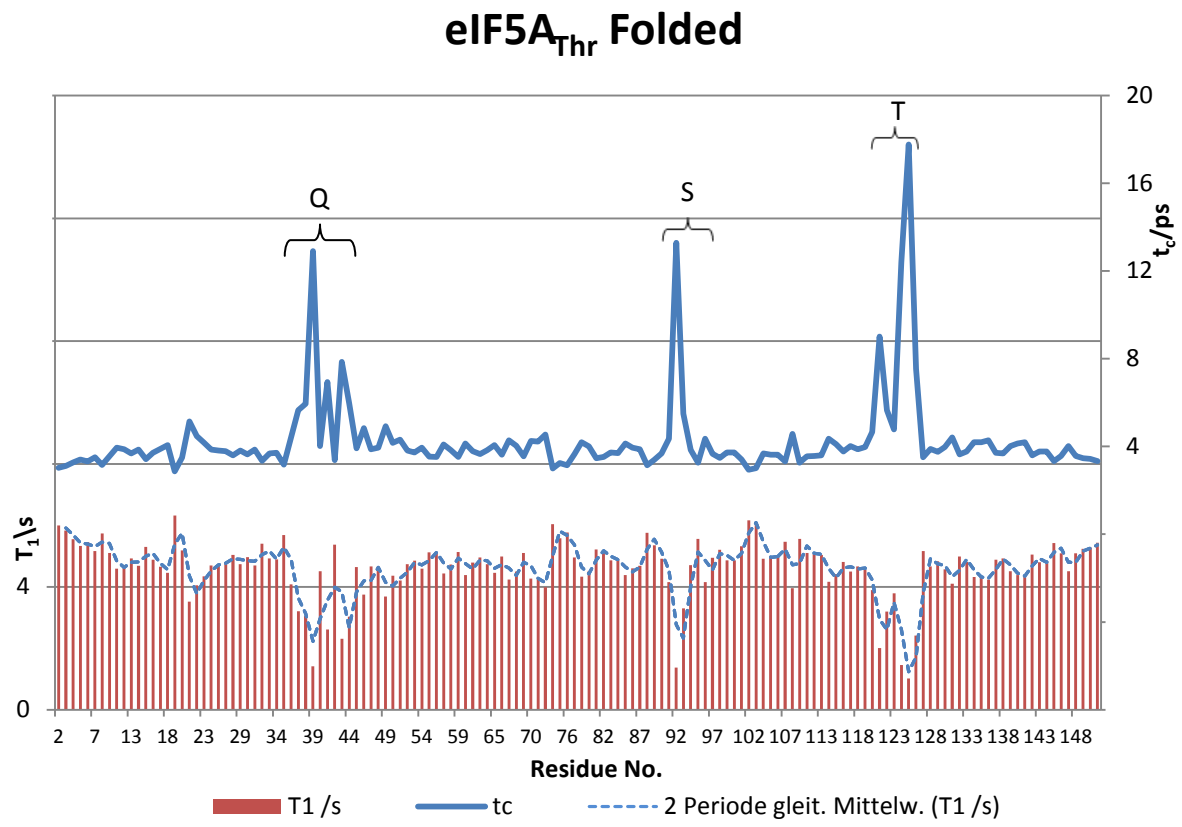


Figure 4-18: Residue specific spin-lattice relaxation and correlation times of folded eIF5A$_{Thr}$. Bracket Q, S and T define the 33-43, 90-95, 117-128 residue regions respectively, Figure 4-2.

However for the folded eIF5A$_{Thr}$ mutant, from Figure 4-18, significant differences are noticeable within residues that were identified as being within the Q, R and the S regions from the HA chemical shift standard deviations, Figure 4-2. The mean T_1 is 4.43 s and the mean τ_c over all residues is 4.09 ps with a standard deviation of 30% and 50% respectively. From the τ_c data we can see that residue 40 has a longer τ_c of 12.9 ps while residue 92 has

13.28 ps. Residue 125 records the longest τ_c of 17.77 ps which is equivalent to deviation of 430 % from the mean.

The longer correlation times can be accounted for by structural restraints that limit the rate of reorientation of the vectors within these domains. When these observations are compared to observations made from measurements of RDC's and chemical shifts of residues within the β and δ domains it was found that residues in these domains possessed significantly more deviations and thus they experienced more dynamic fluctuations. The observations implied that the domains were less stable and had larger degrees of freedom to oscillate influencing the distance dependant chemical shifts and the orientation dependant residual dipolar couplings.

From our observations of the predicted τ_c we can observe a limitation to the methodology we used to calculate the τ_c. The limitation is the in ability to account for translation or domain dependant fluctuations within the calculation of molecular tumbling which influences relaxation. When the movement of the domain, which contains the internuclear vector in question, occurs it moves independently of the rest of the macromolecule. This movement may reduce the observed angle that the vector underwent within the time-step. Reducing the size of this angle reduces the velocity of the vector reducing its tumbling rate.

Another significant limitation to our protocol is that we only consider vector specific tumbling without considering the influence of molecular tumbling (the motion of the protein as a whole). In reality due to the decreased stability and increased mobility of these domains, we would expect the rates of change in magnetic fluctuations to increase with decreased stability. These would increase the rate of relaxation resulting in a shorter T_1. In our case we observe an increase in the T_1 of residues which are in mobile domains based on our approximation of τ_c. It is not sufficient for us to estimate the τ_c because its calculation is not a trivial task and requires knowledge of different kinds of motions as well as the shape of the molecule (Teng, 2013). When backbone dynamics of a 16 kDa protein were carried out using ^{15}N relaxation measurements based on the inverse detected heteronuclear NMR, the τ_c's observed in solution were within 6 and 7 ns while the average T_1 was 503 ± 36 ms (Barbato, 1992). Our study obtained residue specific τ_c's in the range of 4 ps and thus the estimates for T_1 in the range of 4 s were longer. The residue specific fluctuations without

tumbling showed longer relaxation times and thus the fluctuations did not efficiently quench the magnetisation.

It can be seen from our results that fluctuations of internuclear vectors contribute towards the overall relaxation of a specific residue. Dominant contributions to relaxation arise from the result of molecular tumbling. Although we are able to calculate a value for τ_c and T_1, due to limitations we were not able to exhaustively determine the τ_c contribution, as a result of dynamics, to the T_1 relaxation, nor is there experimental data for comparison.

Stoke's law relates the viscosity of the medium, η_w, the effective hydrodynamic radius of the molecule estimated from its molecular mass, r_H, the Boltzmann constant, k_B, and the temperature, T, to the isotropic rotational correlation time, τ_c, of an approximately spherical globular protein (Lewis E Kay, 2008; Teng, 2013).

$$\tau_c = 4\pi\eta_w r^3_H/(3k_BT)$$

A more accurate estimate of T_1 from τ_c, would rely on an estimate of the hydrodynamic radius of the protein from its molecular mass. However obtaining τ_c, of internuclear vectors assists in understanding the contributions of residue specific dynamics dependant on the mobility of the domain.

4.4 CONCLUSION

The protocol developed to analyse the molecular dynamics simulations was able to extend the applications of static chemical shift and residual dipolar coupling predictors to dynamic systems. Camshift generated chemical shift peak tables and applications of basic statistical analyses were used in order to describe residue interactions occurring over the 1 ns of molecular dynamics simulations. REDCAT algorithms were also incorporated in analyses and calculations of $^{15}N - {}^{1}H$ RDCs for the models during the 1 ns of simulations were performed. By applying delicate data extraction from the atomic coordinates of trajectory files non-zero T_1 relaxations contributions for each residue can be calculated giving access to prediction of structural dynamics comparable to relaxation dispersion experiments. By incorporating the REDCAT and CamShift data with NMR Pipe tools, synthetic spectra were generated from convoluting synthetic free induction decays obtained from NMR Draw. These synthetic

spectra can be used to observe differences in the folded or unfolded models influenced by their solution behaviour.

Analysis of chemical shift data, RDCs and T_1 relaxation times gave insight into the solution behaviour of folded eIF5A and $eIF5A_{Thr}$, folded and unfolded, mutants. Average Hα chemical shift standard deviations showed a significant difference in the behaviour of the folded mutant model compared to the native folded model. This data suggests a structural destabilisation of the mutant due to the introduction of the thrombin cleavage site in the $eIF5A_{Thr}$ model. Overlaid chemical shift spectra for the unfolded model matched the experimental HSQC data more than the folded mutant spectra. In order to access the conformation of the unfolded model maintained in solution giving rise to the experimental HSQC spectra, it was envisaged that an unfolding steered dynamics experiment could achieve this. From 780 ps of steered dynamics a slight improvement in the match of the folded mutant model with the experimental spectra was observed.

Over 1 ns of simulation we were able to relate conformational sampling through rapid dynamics with the calculated residual dipolar coupling in order to get a better understanding of the influence of conformation and orientation on RDCs. Back-calculated RDC and predicted T_1 relaxation data allowed further analysis of the eIF5A solution behaviour and while providing restraints for future NMR experimental data on eIF5A. Because of their sensitivity to conformation and their ease of measurement RDC data can be used to provide restraints for MD simulations which can probe excited invisible states of proteins.

By measuring and analysing NMR specific restraints from molecular dynamics trajectories it was shown that the folded $eIF5A_{Thr}$ mutant does not behave in a similar manner in solution as the native folded eIF5A.

Chapter 5: Concluding Remarks

Extensive work has gone into understanding the role of post-translational hypusination of eukaryotic eIF5A on its influence on the regulation of cell growth, cell differentiation and IRES mediated protein synthesis. In order to capitalise on the 600 MHz NMR spectrometer for the solution study of eIF5A it was thought that by preparing 10 kDa of the eIF5A terminal domains, solution structures could be obtained by using NMR approaches for sequential assignment. Isotopically labelled eIF5A containing the thrombin cleavage site was successfully expressed in *E. coli* and cleavage was successfully achieved. NMR ^{15}N-HSQC experiments performed on the uncleaved unhypusinated *E. coli* produced eIF5A monomer gave rise to a peak signature that is consistent with a natively unfolded protein. These observations were consistent with functional analyses of the eIF5A$_{Thr}$ protein that showed that the mutant protein was unable to substitute for the native protein during complementation assays in yeast although it was hypusinated (Charlton, 2012).

Folded and unfolded models of eIF5A$_{Thr}$ were prepared using standard homology modelling techniques based on the folded eIF5A crystal structures. Analysis of the solution behaviour of the folded native and mutated eIF5A models using molecular dynamics showed that indeed the introduction of rigidity within the flexible loop region of eIF5A affected the macromolecule solution behaviour of the eIF5A monomer. NMR restraints such as chemical shifts were calculated from the molecular dynamics trajectories in order to probe the solution behaviour at a higher atomic resolution. Predicted HSQC spectra for the unfolded mutant model matched the experimental HSQC experimental data better than the predicted spectra from the folded mutant models.

Preliminary *in silico* unfolding experiments were performed in order to attempt to match experimental chemical shift signatures to probable solution structures. Unfolding using steered molecular dynamics improved the match of the folded mutant model HSQC peak signature to the experimental HSQC data. Despite this improvement folding experiments of the unfolded model over 1 ns of simulation showed a closer match to the experimental data. It is presumed that the structure maintained in solution can be accessed using the folding molecular dynamics technique as opposed to the unfolding technique due to computational restrictions associated with the protocols applied in steered molecular

dynamics. Steered molecular dynamics is computationally expensive and requires the use of umbrella sampling between time-steps compared to the Newtonian undirected simulated molecular dynamics. Future molecular dynamics experiments of the unfolded models could verify the introduction of rigidity on protein folding pathway adopted in solution by the mutant protein. These observations could shed light on the necessity of flexibility in this region of the protein for the eIF5A function and dimerisation.

Predicted residual coupling data were calculated in anticipation of the experimental RDC data obtained at 600 MHz. RDC experiments provide improved sensitivity at 600 MHz and allow the extraction of structural data for eIF5A suitable for structure studies. It was shown that dynamic data can be extracted from analysis of the N-H vectors obtained from the calculation of RDCs. T_1 relaxation times coupled with the fluctuation of N-H vectors obtained from RDC data allowed the probing of probing dynamics and showed the potential for the use of relaxation dispersion experiments to provide insight into eIF5A protein solution dynamics.

Cleavage of eIF5A into its monomers is unnecessary because at standard temperatures the 600 MHz NMR spectrometer was able to extract well resolved HSQC spectra. The ability to produce hypusinated eIF5A in *E. coli* using a polycistronic vector improves the likelihood of obtaining a solution structure of hypusinated eIF5A monomer in solution. RDC, pseudocontact shift and relaxation dispersion experiments performed on the protein monomer under 600 MHz have potential to extract structural and solution behaviour of eIF5A in its hypusinated and unhypusinated state.

References

Andersson, P., Nordstrand, K., Sunnerhagen, M., Liepinsh, E., Turovskis, I., & Otting, G. (1998). Heteronuclear correlation experiments for the determination of one-bond coupling constants. *Journal of Biomolecular NMR*, *11*(4), 445–450. doi:10.1023/A:1008206212145

Apol, E., Apostolov, R., Berendsen, H. J. C., van Buuren, A., Bjelkmar, Pã¤., van Drunen, R., … Lindahl, E. (2010). GROMACS Getting Started. *www.gromacs.org*.

Atilgan, C., & Atilgan, A. R. (2009). Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Computational Biology*, *5*(10), e1000544. doi:10.1371/journal.pcbi.1000544

Barbato, G. (1992). Backbone Dynamics of Calmodulin Studied by 1 5, 5269–5278.

Benne, R., & Hershey, J. W. (1978). The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *Journal of Biological Chemistry*, *253*(May), 3078–3087.

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1), 43–56. Retrieved from http://www.sciencedirect.com/science/article/pii/001046559500042E

Berman, H. M. (2008). The Protein Data Bank: a historical perspective. *Acta Crystallographica2*, *A64*, 88–95.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–42. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102472&tool=pmcentrez&rendertype=abstract

Bernstein, J., Sella, O., & Le, S.-Y. (1997). PDGF2/c-sis mRNA Leader Contains a Differentiation-linked Internal Ribosomal Entry Site (D-IRES). *Journal of Biological Chemistry*, *272*(14), 9356–9362. doi:10.1074/jbc.272.14.9356

Bertini, I., Donaire, a, Jiménez, B., Luchinat, C., Parigi, G., Piccioli, M., & Poggi, L. (2001). Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D9k. *Journal of Biomolecular NMR*, *21*(2), 85–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11727989

Betz, M., Saxena, K., & Schwalbe, H. (2006). Biomolecular NMR: a chaperone to drug discovery. *Current Opinion in Chemical Biology*, *10*(3), 219–25. doi:10.1016/j.cbpa.2006.04.006

Bevec, D., & Hauber, J. (1997). Eukaryotic Initiation Factor 5A Activity and HIV-1 Rev Function. *Neurosignals*, *6*(3), 124–133. doi:10.1159/000109118

Blundell, T. L., Sibanda, B. L., Montalvão, R. W., Brewerton, S., Chelliah, V., Worth, C. L., … Burke, D. (2006). Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1467), 413–23. doi:10.1098/rstb.2005.1800

Bommer, U. A., Lutsch, G., Stahl, J., & Bielka, H. (1991). Eukaryotic initiation factors eIF-2 and eIF-3: interactions, structure and localization in ribosomal initiation complexes. *Biochimie*, *73*(7), 1007–1019. Retrieved from http://www.sciencedirect.com/science/article/pii/030090849190142N

Borders, C. L., Broadwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldridge, A. M., & Pett, V. B. (1994). A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Science : A Publication of the Protein Society*, *3*(4), 541–8. doi:10.1002/pro.5560030402

Bourdon, J.-C. (2007). p53 Family isoforms. *Current Pharmaceutical Biotechnology*, *8*(6), 332–6. Retrieved from /pmc/articles/PMC3523268/?report=abstract

Breiman, L. (2001). Random forests. *Maachine Learning*, *41*, 5–32.

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., … Karplus, M. (2009). CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, *30*(10), 1545–614. doi:10.1002/jcc.21287

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217. doi:10.1002/jcc.540040211

Bruccoleri, R., & Karplus, M. (1990). Conformational sampling using high-temperature molecular dynamics. *Biopolymers*, *29*(14), 1847–1862.

Cano, V. S. P., Jeon, G. A., Johansson, H. E., Henderson, C. A., Park, J. H., Valentini, S. R., … Park, M. H. (2008). Mutational analyses of human eIF5A-1--identification of amino acid residues critical for eIF5A activity and hypusine modification. *The FEBS Journal*, *275*(1), 44–58. doi:10.1111/j.1742-4658.2007.06172.x

Caraglia, M., Marra, M., Giuberti, G., Budillon, A., Lentini, A., Beninati, S., & Abbruzzese, A. (2001). The role of eukaryotic initiation factor 5A in the control of cell proliferation and apoptosis Review Article. *Amino*, 91–104.

Caraglia, M., Park, M. H., Wolff, E. C., Marra, M., & Abbruzzese, a. (2013). eIF5A isoforms and cancer: two brothers for two functions? *Amino Acids*, *44*(1), 103–9. doi:10.1007/s00726-011-1182-x

Caraglia, M., Tagliaferri, P., Budillon, A., & Abbruzzese, A. (1999). Post-Translational Modifications of Eukaryotic Initiation Factor-5A (eIF-5A) as a New Target for anti-cancer therapy. *Advances in Experimental Medicine and Biology*, *472*, 187–198. doi:10.1007/978-1-4757-3230-6

Carbajo, R. J., & Neira, J. L. (2013a). Biomolecular NMR. In *NMR for Chemists and Biologists* (pp. 83–115). Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-6976-2

Carbajo, R. J., & Neira, J. L. (2013b). The Basis of Nuclear Magnetic Resonance Spectroscopy. In *NMR for Chemists and Biologists* (pp. 1–29). Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-6976-2

Cavalli, A., Salvatella, X., Dobson, C. M., & Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9615–20. doi:10.1073/pnas.0610313104

Cavanagh, J., Wayne, J. F., Palmer III, A. G., Rance, M., & Skelton, N. J. (2007). *Protein NMR Spectroscopy: Principles and Practice*. (N. Gracy, L. Han, J. Ochs, & A. Russums, Eds.) (Second Edi.). London: Elsevier Inc.

Chan, H. S., & Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins*, *30*(1), 2–33. doi:10.1002/(SICI)1097-0134(19980101)30:1<2::AID-PROT2>3.0.CO;2-R

Charlton, J. (2012). *Understanding the biomolecular interactions involved in dimerisation of the Saccharomyces cerevisiae eukaryotic translation initiation factor*. *Masters of Science*. Rhodes University.

Chattopadhyay, M. K., Park, M. H., & Tabor, H. (2008). Hypusine modification for growth is the major function of spermidine in Saccharomyces cerevisiae polyamine auxotrophs grown in limiting spermidine. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6554–9. doi:10.1073/pnas.0710970105

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., … Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science (New York, N.Y.)*, *325*(5942), 834–40. doi:10.1126/science.1175371

Chung, S. I., Park, M. h., Folk, J. E., & Lewis, M. S. (1991). Eukaryotic initiation factor 5A: the molecular form of the hypusine-containing protein from human erythrocytes. *Biochimica et Biophysica Acta (BBA)*. Retrieved February 27, 2014, from http://ac.els-cdn.com/016748389190490Q/1-s2.0-016748389190490Q-main.pdf?_tid=3accb646-9fb8-11e3-92a1-00000aacb362&acdnat=1393510116_705ca94277485b7f7fb4c0cf3286012a

Clement, P. M. J., Henderson, C. A., Jenkins, Z. A., Smit-McBride, Z., Wolff, E. C., Hershey, J. W. B., … Johansson, H. E. (2003). Identification and characterization of

eukaryotic initiation factor 5A-2. *European Journal of Biochemistry*, *270*(21), 4254–4263. doi:10.1046/j.1432-1033.2003.03806.x

Cohen, S. N. (2013). DNA cloning: A personal view after 40 years. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15521–9. doi:10.1073/pnas.1313397110

Cooper, H., Park, M., Folk, J., Safer, B., & Braverman, R. (1983). Identification of the hypusine-containing protein Hy+ as translation initiation factor eIF-4D. *Proceedings of the National Academy of Sciences USA*, *80*, 1854.

Cornilescu, G., Delaglio, F., & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, *13*, 289–302.

Dalvit, C., Pevarello, P., Tatò, M., Veronesi, M., Vulpetti, a, & Sundström, M. (2000). Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *Journal of Biomolecular NMR*, *18*(1), 65–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11061229

De Benedetti, A., & Harris, A. L. (1999). eIF4E expression in tumors: its possible role in progression of malignancies. *The International Journal of Biochemistry & Cell Biology*, *31*(1), 59–72. Retrieved from http://www.sciencedirect.com/science/article/pii/S1357272598001320

Delaglio, F. (2001). Protein Structure Modelling using Chemical Shifts and Residual Dipolar Coupling Homology. *Thesis*, 105. Retrieved from http://spin.niddk.nih.gov/NMRPipe/

Delaglio, F., Kontaxis, G., & Bax, A. (2000). Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings. *Journal of American Chemical Society*, (122), 2142–2143.

Dias, C. a O., Cano, V. S. P., Rangel, S. M., Apponi, L. H., Frigieri, M. C., Muniz, J. R. C., … Valentini, S. R. (2008). Structural modeling and mutational analysis of yeast eukaryotic translation initiation factor 5A reveal new critical residues and reinforce its involvement in protein synthesis. *The FEBS Journal*, *275*(8), 1874–88. doi:10.1111/j.1742-4658.2008.06345.x

Dias, C. A. O., Garcia, W., Zanelli, C. F., & Valentini, S. R. (2013). eIF5A dimerizes not only in vitro but also in vivo and its molecular envelope is similar to the EF-P monomer. *Amino Acids*, *44*(2), 631–44. doi:10.1007/s00726-012-1387-7

Diercks, T., Coles, M., & Kessler, H. (2001). Applications of NMR in drug discovery. *Combinatorial Chemistry*, 285–291.

Dill, K. A., Ozkan, S. B., Shell, M. S., & Weikl, T. R. (2008). The Protein Folding Problem. *Annual Reviews of Biophysics*, *37*(June), 289–316.

Dobson, C. M., Šali, A., & Karplus, M. (1998). Protein Folding: A Perspective from Theory and Experiment. *Angewandte Chemie International Edition*, *37*(7), 868–893. doi:10.1002/(SICI)1521-3773(19980420)37:7<868::AID-ANIE868>3.0.CO;2-H

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113. doi:10.1186/1471-2105-5-113

EMBL-EBI. (2013). ClustalW2 < Multiple Sequence Alignment. Retrieved June 24, 2013, from http://www.ebi.ac.uk/Tools/msa/clustalw2/

Ernst, R. R. (1992). NUCLEAR MAGNETIC RESONANCE FOURIER. *Nobel Lecture*, 12–57.

Fernandez, C., & Wider, G. (2006). NMR Spectroscopy of Large Biological Macromolecules in Solution. In J. L. Arrondo & A. Alonso (Eds.), *Avanced Techniques in Biophysics, Springer Series in Biophysics* (Vol 10., pp. 89–128). Springer.

Fernández-Miragall, O., López de Quinto, S., & Martínez-Salas, E. (2009). Relevance of RNA structure for the activity of picornavirus IRES elements. *Virus Research*, *139*(2), 172–82. doi:10.1016/j.virusres.2008.07.009

Finch, J. T., & Stretton, A. O. (1964). The structure of the "polyheads" of T4 bacteriophage. *Journal of Molecular Biology*, *10*(570-575).

Fiser, A., Do, R. K. G., & Sali, A. (2000a). Modelling of loops in protein structures. *Protein Science*, *9*, 1753–1773.

Fiser, A., Do, R. K., & Sali, A. (2000b). Modeling of loops in protein structures. *Protein Science : A Publication of the Protein Society*, *9*(9), 1753–73. doi:10.1110/ps.9.9.1753

Fleming, P. J., & Rose, G. D. (2008). Conformational properties of unfolded proteins. *Protein Science Encyclopedia*.

Folkers, G. E., van Buuren, B. N. M., & Kaptein, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. *Journal of Structural and Functional Genomics*, *5*(1-2), 119–31. doi:10.1023/B:JSFG.0000029200.66197.0c

Frydman, L., Scherf, T., & Lupulescu, A. (2002). The acquisition of multidimensional NMR spectra. *PNAS*, *99*(25), 15858–15862.

Gasyna, Z. L., & Jurkiewicz, A. (2004). Determination of Spin—Lattice Relaxation Time Using 13C NMR. An Undergraduate Physical Chemistry Laboratory Experiment. *Journal of Chemical Education*, *81*(7), 1038. doi:10.1021/ed081p1038

Gentz, P. M. (2008). Towards understanding the mechanism of dimerisation of Saccharomyces cerevisiae eukaryotic translation initiation factor 5A. *Doctoral Thesis*, (January).

Gentz, P. M., Blatch, G. L., & Dorrington, R. A. (2009). Dimerization of the yeast eukaryotic translation initiation factor 5A requires hypusine and is RNA dependent. *The FEBS Journal*, *276*(3), 695–706. doi:10.1111/j.1742-4658.2008.06817.x

Gershenson, A., & Gierasch, L. M. (2011). Protein folding in the cell: challenges and progress. *Current Opinion in Structural Biology*, *21*(1), 32–41. doi:10.1016/j.sbi.2010.11.001

Gibson, K. D., & Scheraga, H. A. (1969). Minimization of polypeptide energy, VII. Second derivatives and statistical weights of energy minima for deca-L-Alanine. *Proceedings of the National Academy of Sciences*, *4*(2), 242–245.

Goldsmith-fischman, S., & Honig, B. (2003). Structural genomics : Computational methods for structure analysis. *Protein Science*, *12*, 1813–1821. doi:10.1110/ps.0242903.1999

Gordon, E. D., Mora, R., Meredith, S. C., Lee, C., & Lindquist, S. L. (1987). Eukaryotic Initiation Factor 4D, the Hypusine-containing Protein, Is Conserved among Eukaryotes. *The Journal of Biological Chemistry*, *262*(34). Retrieved from http://www.jbc.org/content/262/34/16585.full.pdf

Gosslau, A., Jao, D. L.-E., Butler, R., Liu, A. Y.-C., & Chen, K. Y. (2009). Thermal killing of human colon cancer cells is associated with the loss of eukaryotic initiation factor 5A. *Journal of Cellular Physiology*, *219*(2), 485–93. doi:10.1002/jcp.21696

Graham, B., Loh, C. T., Swarbrick, J. D., Ung, P., Shin, J., Yagi, H., … Otting, G. (2011). DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconjugate Chemistry*, *22*(10), 2118–25. doi:10.1021/bc200353c

Griesinger, C., & Sattler, M. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *34*, 93–158.

Guan, X.-Y., Fung, J. M.-W., Ma, N.-F., Lau, S.-H., Tai, L.-S., Xie, D., … Sham, J. S. T. (2004). Oncogenic role of eIF-5A2 in the development of ovarian cancer. *Cancer Research*, *64*(12), 4197–200. doi:10.1158/0008-5472.CAN-03-3747

Güntert, P. (2009). Automated structure determination from NMR spectra. *European Biophysics Journal : EBJ*, *38*(2), 129–43. doi:10.1007/s00249-008-0367-z

Hajduk, P. J., Meadows, R. P., & Fesik, S. W. (1999). NMR-based screening in drug discovery. *Quarterly Reviews of Biophysics*, *32*(3), 211–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11194565

Hamelrych, T. (2005). An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. *Proteins*, *59*, 38–48.

Hanahan, D., & Harbor, C. S. (1983). Studies o n T r a n s f o r m a t i o n o f Escherichia coli with P l a s m i d s. *Journal of Molecular Biology*, *166*, 557–580.

Hanauske-Abel, H. M., Slowinska, B., Zagulska, S., Wilson, R. C., Staiano-Coico, L., Hanauske, A.-R., … Szabo, P. (1995). Detection of a sub-set of polysomal mRNAs associated with modulation of hypusine formation at the G1-S boundary Proposal of a role for eIF-5A in onset of DNA replication. *FEBS Letters*, *366*(2), 92–98. Retrieved from http://www.sciencedirect.com/science/article/pii/001457939500493S

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324–32. doi:10.1038/nature10317

Hauber, J. (2010). Revisiting an old acquaintance: role for eIF5A in diabetes. *The Journal of Clinical Investigation*, *120*(6), 1806–8. doi:10.1172/JCI43237

Herrmann, T., Guntert, P., & Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*, *319*(1), 209–228.

Hertz, H. G. (1967). Chapter 5 Microdynamic behaviour of liquids as studied by NMR relaxation times. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *3*, 159–230. Retrieved from http://www.journals.elsevierhealth.com/periodicals/jpnmrs/article/0079-6565(67)80015-3/abstractref

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447. doi:10.1021/ct700301q

Hobbie, R. K. (1988). *Intermediate Physics for Medicine and Biology* (2nd ed.). Wiley.

Hoque, M., Hanauske-Abel, H. M., Palumbo, P., Saxena, D., D'Alliessi Gandolfi, D., Park, M. H., … Mathews, M. B. (2009). Inhibition of HIV-1 gene expression by Ciclopirox and Deferiprone, drugs that prevent hypusination of eukaryotic initiation factor 5A. *Retrovirology*, *6*, 90. doi:10.1186/1742-4690-6-90

Idicula, A. M., Blatch, G. L., Cooper, T. G., & Dorrington, R. A. (2002). Binding and activation by the zinc cluster transcription factors of Saccharomyces cerevisiae. Redefining the UASGABA and its interaction with Uga3p. *The Journal of Biological Chemistry*, *277*(48), 45977–83. doi:10.1074/jbc.M201789200

Issar, S., & Gaur, R. K. (2013). Homology modelling of Nitrogenase enzyme encoded by NIFHDK gene cluster of Pseudomonas putida. *International Journal of Advanced Research*, *1*(8), 722–732.

Jackson, D. A., Symonst, R. H., & Berg, P. (1972). Biochemical Method for Inserting New Genetic Information into DNA of, *69*(10), 2904–2909.

James, T. L. (1998). Chapter 1 Fundamentals of NMR. In U. of California (Ed.), *Fundamentals of NMR* (pp. 1–31). San Francisco.

Jensen, M. R., Hansen, D. F., Ayna, U., Dagil, R., Hass, M. a S., Christensen, H. E. M., & Led, J. J. (2006). On the use of pseudocontact shifts in the structure determination of

metalloproteins. *Magnetic Resonance in Chemistry : MRC*, *44*(3), 294–301. doi:10.1002/mrc.1771

Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M., & Güntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature*, *440*(7080), 52–57. doi:10.1038/nature04525

Kaiser, A. (2012). Translational control of eIF5A in various diseases. *Amino Acids*, *42*(2-3), 679–84. doi:10.1007/s00726-011-1042-8

Kallen, J., Spitzfaden, C., Zurini, M. G. M., Wider, G., Widmer, H., Wuthrich, K., & Walkinshaw, M. D. (1991). Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy. *Nature*, *353*, 276–271.

Kang, H. a, Schwelberger, H. G., & Hershey, J. W. (1993). Translation initiation factor eIF-5A, the hypusine-containing protein, is phosphorylated on serine in Saccharomyces cerevisiae. *The Journal of Biological Chemistry*, *268*(20), 14750–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8325852

Kang, H. A., & Hershey, J. W. (1994). Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of Saccharomyces cerevisiae. *J. Biol. Chem.*, *269*(6), 3934–3940. Retrieved from http://www.jbc.org/content/269/6/3934.short

Kang, K. R., Kim, Y. S., Wolff, E. C., & Park, M. H. (2007). Specificity of the deoxyhypusine hydroxylase-eukaryotic translation initiation factor (eIF5A) interaction: identification of amino acid residues of the enzyme required for binding of its substrate, deoxyhypusine-containing eIF5A. *The Journal of Biological Chemistry*, *282*(11), 8300–8. doi:10.1074/jbc.M607495200

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*(9), 646–52. doi:10.1038/nsb0902-646

Katahira, J., Ishizaki, T., Sakai, H., Adachi, A., Yamamoto, K., & Shida, H. (1995). Effects of translation initiation factor eIF-5A on the functioning of human T-cell leukemia virus type I Rex and human immunodeficiency virus Rev inhibited trans dominantly by a Rex mutant deficient in RNA binding. *J. Virol.*, *69*(5), 3125–3133. Retrieved from http://jvi.asm.org/content/69/5/3125.short

Kay, L. E. (2008). Using relaxation dispersion NMR spectroscopy to determine structures of excited , invisible protein states, 113–120. doi:10.1007/s10858-008-9251-5

Kay, L. E. (2009). Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding, *96*(March), 2045–2054. doi:10.1016/j.bpj.2008.12.3907

Kay, L. E., Torchia, D. a, & Bax, a. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, *28*(23), 8972–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2690953

Keeler, J. (2005). *Understanding NMR Spectroscopy* (1st Editio.). Chichester: John Wiley and Sons, Ltd.

Kemper, W. M., Berry, K. W., & Merrick, W. C. (1976). Purification and properties of rabbit reticulocyte protein synthesis initiation Purification and Properties of Rabbit Reticulocyte Synthesis Initiation Factors M2Balpha and M2Bbeta. *Journal of Biological Chemistry*, *251*(18), 5551–5557.

Kendrew, J. C. (1963). Myoglobin and the structure of proteins. *Science (New York, N.Y.)*, *139*(3561), 1259–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10958321

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. W., & Phihips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray anaysis. *Nature*, (666), 181.

Kerscher, B., Nzukou, E., & Kaiser, a. (2010). Assessment of deoxyhypusine hydroxylase as a putative, novel drug target. *Amino Acids*, *38*(2), 471–7. doi:10.1007/s00726-009-0406-9

Kim, K. K., Hung, L. W., Yokota, H., Kim, R., & Kim, S. H. (1998). Crystal structures of eukaryotic translation initiation factor 5A from Methanococcus jannaschii at 1.8 A resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10419–24. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27909&tool=pmcentrez&rendertype=abstract

Kohlhoff, K. J. (2008). *Protein Chemical Shifts as Structural Restraints in Molecular Dynamics Simulations*. University of Cambridge.

Kohlhoff, K. J., Robustelli, P., Cavalli, A., Salvatella, X., & Vendruscolo, M. (2009). Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *Journal of the American Chemical Society*, *131*(39), 13894–5. doi:10.1021/ja903772t

Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F., & Mackay, J. P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS Journal*, *278*(5), 687–703. doi:10.1111/j.1742-4658.2011.08004.x

Kyrpides, N. C., & Woese, C. R. (1998). Universally conserved translation initiation factors. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(1), 224–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18182&tool=pmcentrez&rendertype=abstract

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685. doi:10.1038/227680a0

Lalande, M. (1990). A reversible arrest point in the late G1 phase of the mammalian cell cycle. *Experimental Cell Research*, *186*(2), 332–339. Retrieved from http://www.sciencedirect.com/science/article/pii/001448279090313Y

Lange, O. F., Lakomek, N.-A., Farès, C., Schröder, G. F., Walter, K. F. a, Becker, S., … de Groot, B. L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science (New York, N.Y.)*, *320*(5882), 1471–5. doi:10.1126/science.1157092

Langridge, R. (1974). Interactive three-dimensional computer graphnics and molecular biology. *Federal Proceedings*, *33*(12), 2332–5.

Larkin, M. a, Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. a, McWilliam, H., … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, *23*(21), 2947–8. doi:10.1093/bioinformatics/btm404

Laskowski, R. a, MacArthur, M. W., & Thornton, J. M. (1998). Validation of protein models derived from experiment. *Current Opinion in Structural Biology*, *8*(5), 631–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9818269

Laskowski, R. A. (2009). PDBsum new things. *Nucleic Acids Research*, *37*(Database issue), D355–9. doi:10.1093/nar/gkn860

Lee, C. H., Um, P. Y., & Park, M. H. (2001). Structure-function studies of human deoxyhypusine synthase: identification of amino acid residues critical for the binding of spermidine and NAD. *The Biochemical Journal*, *355*(Pt 3), 841–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1221802&tool=pmcentrez&rendertype=abstract

Lee, S. B., Park, J. H., Folk, J. E., Deck, J. A., Pegg, A. E., Sokabe, M., … Park, M. H. (2010). Inactivation of eukaryotic initiation factor 5A (eIF5A) by specific acetylation of its hypusine residue by spermidine/spermine acetyltransferase 1 (SSAT1). *Journal of Biochemistry*, *433*(1), 205–213. doi:10.1042/BJ20101322.Inactivation

Lee, S. B., Park, J. H., Kaevel, J., Sramkova, M., Weigert, R., & Park, M. H. (2009). Biochemical and Biophysical Research Communications The effect of hypusine modification on the intracellular localization of eIF5A. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 1–6. doi:10.1016/j.bbrc.2009.04.049

Lemkul, J. A., & Bevan, D. R. (2010). Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *The Journal of Physical Chemistry. B*, *114*(4), 1652–60. doi:10.1021/jp9110794

Lesley, S. a. (2001). High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expression and Purification*, *22*(2), 159–64. doi:10.1006/prep.2001.1465

Levitt, M. (1976). Rapid Simulation of Protein Conformations of Protein Folding. *Journal of Molecular Biology*, *104*, 59–107.

Levitt, M., & Warshel, A. (1975). Computer Simulation of protein folding. *Nature*, *253*, 694–698.

Lewis, T. E., Sillitoe, I., Andreeva, A., Blundell, T. L., Buchan, D. W. A., Chothia, C., … Orengo, C. (2013). Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. *Nucleic Acids Research*, *41*(Database issue), D499–507. doi:10.1093/nar/gks1266

Li, A.-L., Li, H.-Y., Jin, B.-F., Ye, Q.-N., Zhou, T., Yu, X.-D., … Zhang, X.-M. (2004). A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. *The Journal of Biological Chemistry*, *279*(47), 49251–8. doi:10.1074/jbc.M407165200

Li, C. H., Ohn, T., Ivanov, P., Tisdale, S., & Anderson, P. (2010). eIF5A promotes translation elongation, polysome disassembly and stress granule assembly. *PloS One*, *5*(4), e9942. doi:10.1371/journal.pone.0009942

Li, J. (2005). Basic Molecular Dynamics. In S. Yip (Ed.), *Handbook of Materials Modelling* (pp. 565–588). Netherlands: Springer Netherlands.

Lindahl, E., Hess, B., & Spoel, D. van der. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modeling Annual*, *7*(8), 306–317. doi:10.1007/s008940100045

Lindhout, D. a, Litowski, J. R., Mercier, P., Hodges, R. S., & Sykes, B. D. (2004). NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers*, *75*(5), 367–75. doi:10.1002/bip.20150

Lipowsky, G., Bischoff, F. R., Schwarzmaier, P., Kraft, R., Kostka, S., Hartmann, E., … Görlich, D. (2000). Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. *The EMBO Journal*, *19*(16), 4362–71. doi:10.1093/emboj/19.16.4362

Liu, J., Henao-Mejia, J., Liu, H., Zhao, Y., & He, J. J. (2011). Translational regulation of HIV-1 replication by HIV-1 Rev cellular cofactors Sam68, eIF5A, hRIP, and DDX3. *Journal of Neuroimmune Pharmacology : The Official Journal of the Society on NeuroImmune Pharmacology*, *6*(2), 308–21. doi:10.1007/s11481-011-9265-8

Lovell, S. C., Davis, I. W., Iii, W. B. A., Bakker, P. I. W. De, Word, J. M., Prisant, M. G., … Richardson, D. C. (2003). Structure Validation by C alpha Geometry: phi , psi and C beta Deviation. *Proteins: Structure, Function, and Genetics*, *450*(August 2002), 437–450.

Luthy, R., Bowie, J., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(March), 83 – 85.

Maier, B., Ogihara, T., Trace, A. P., Tersey, S. A., Robbins, R. D., Chakrabarti, S. K., … Thompson, J. E. (2010). The unique hypusine modification of eIF5A promotes islet β cell inflammation and dysfunction in mice. *The Journal of Clinical Investigation*, *120*(6), 2156.

Maier, B., Tersey, S. A., & Mirmira, R. G. (2010). Hypusine: a new target for therapeutic intervention in diabetic inflammation. *Discovery Medicine*, *10*(50), 18–23.

Maiorov, V. N., & Crippen, G. M. (1995). Size-independent comparison of protein three-dimensional structures. *Proteins*, *22*(3), 273–83. doi:10.1002/prot.340220308

Marley, J., Lu, M., & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, *20*(1), 71–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11430757

Mémin, E., Hoque, M., Jain, M. R., Heller, D. S., Li, H., Cracchiolo, B., … Mathews, M. B. (2014). Blocking eIF5A Modification in Cervical Cancer Cells Alters the Expression of Cancer-Related Genes and Suppresses Cell Proliferation. *Cancer Research*, *74*(2), 552–62. doi:10.1158/0008-5472.CAN-13-0474

Meyer, E. F. (1997). The first years of the Protein Data Bank. *Protein Science : A Publication of the Protein Society*, *6*(7), 1591–7. doi:10.1002/pro.5560060724

Monmaney. (2005). 35 Who Made a Difference: Robert Langridge. *Smithsonian.com*.

Montalvao, R. W., De Simone, A., & Vendruscolo, M. (2012). Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. *Journal of Biomolecular NMR*, *53*(4), 281–92. doi:10.1007/s10858-012-9644-3

Moore, C. C., Martin, E. N., Lee, G., Taylor, C., Dondero, R., Reznikov, L. L., … Scheld, W. M. (2008). Eukaryotic translation initiation factor 5A small interference RNA-liposome complexes reduce inflammation and increase survival in murine models of severe sepsis and acute lung injury. *The Journal of Infectious Diseases*, *198*(9), 1407–14. doi:10.1086/592222

Mori, S., Abeygunawardana, C., Johnson, M. O., & Vanzijl, P. C. M. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of Magnetic Resonance, Series B*, *108*(1), 94–98.

Mujika, J. I., Matxain, J. M., Eriksson, L. a, & Lopez, X. (2006). Resonance structures of the amide bond: the advantages of planarity. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, *12*(27), 7215–24. doi:10.1002/chem.200600052

Neal, S., Nip, A. M., Zhang, H., & Wishart, D. S. (2003). Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *Journal of Biomolecular NMR*, *26*(3), 215–240. doi:10.1023/A:1023812930288

Neuwald, A. F. (2000). HEAT Repeats Associated with Condensins, Cohesins, and Other Complexes Involved in Chromosome-Related Functions. *Genome Research*, *10*(10), 1445–1452. doi:10.1101/gr.147400

NIGMS, N. I. for G. M. S. U. (2011). 50 Years of Protein Structure Determination Timeline. Retrieved November 2, 2013, from http://publications.nigms.nih.gov/psi/timeline_text.html

Nishimura, K., Lee, S. B., Park, J. H., & Park, M. H. (2012). Essential role of eIF5A-1 and deoxyhypusine synthase in mouse embryonic development. *Amino Acids*, *42*(2-3), 703–10. doi:10.1007/s00726-011-0986-z

Nishimura, K., Murozumi, K., Shirahata, A., Park, M. H., Kashiwagi, K., & Igarashi, K. (2005). Independent roles of eIF5A and polyamines in cell proliferation. *Biochemical Journal*, (385). Retrieved from http://www.biochemj.org/bj/385/bj3850779.htm

Nkari, W. K., & Prestegard, J. H. (2010). NMR Resonance Assignments of Sparsely Labeled Proteins: Amide Proton Exchange Correlations in Native and Denatured States. *Journal of American Chemical Society*, *131*(14), 5344–5349. doi:10.1021/ja8100775.NMR

O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., & Notredame, C. (2004). 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *Journal of Molecular Biology*, *340*(2), 385–95. doi:10.1016/j.jmb.2004.04.058

Ohki, S., & Kainosho, M. (2008). Stable isotope labeling methods for protein NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *53*(4), 208–226. doi:10.1016/j.pnmrs.2008.01.003

Oladele, T. O., Bamigbola, O. M., & Bewaji, C. O. (2009). On efficiency of sequence alignment algorithms. *AFrican Scientist*, *10*(1), 9–14.

Özlem Tastan Bishop, A., de Beer, T. A. P., & Joubert, F. (2008). Protein homology modelling and its use in South Africa. *South African Journal of Science*, *104*(February), 2–6.

Pain, V. M. (1996). Initiation of Protein Synthesis in Eukaryotic Cells. *European Journal of Biochemistry*, *236*(3), 747–771. doi:10.1111/j.1432-1033.1996.00747.x

Park, J. H., Aravind, L., Wolff, E. C., Kaevel, J., Kim, Y. S., & Park, M. H. (2006). Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(1), 51–6. doi:10.1073/pnas.0509348102

Park, J. H., Dias, C., Lee, S. B., Valentini, S. R., Sokabe, M., Fraser, C. S., & Park, M. H. (2011). Production of active recombinant eIF5A: reconstitution in E.coli of eukaryotic hypusine modification of eIF5A by its coexpression with modifying enzymes. *Protein Engineering, Design & Selection : PEDS*, *24*(3), 301–9. doi:10.1093/protein/gzq110

Park, J.-H., Johansson, H. E., Aoki, H., Huang, B. X., Kim, H.-Y., Ganoza, M. C., & Park, M. H. (2012). Post-translational modification by β-lysylation is required for activity of Escherichia coli elongation factor P (EF-P). *The Journal of Biological Chemistry*, *287*(4), 2579–90. doi:10.1074/jbc.M111.309633

Park, M. H. (1987). Regulation of Biosynthesis of Hypusine in Chinese Hamster Ovary Cells. *Journal of Biological Chemistry*, (262), 12730–12734.

Park, M. H. (2008). The Post-Translational Synthesis of a Polyamine-Derived Amino Acid, Hypusine, in the Eukaryotic Translation Initiation Factor 5A (eIF5A). *Journal of Biochemistry*, *139*(2), 161–169.

Park, M. H., Joe, Y. A., Kang, K. R., Lee, Y. B., & Wolff, E. C. (1996). The polyamine-derived amino acide hypusine: its post-translational formation in eIF-5A and its role in cell proliferation. *Amino Acids*, *10*, 109–121.

Park, M. H., Nishimura, K., Zanelli, C. F., & Valentini, S. R. (2010). Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids*, *38*(2), 491–500. doi:10.1007/s00726-009-0408-7

Parreiras-e-Silva, L. T., Gomes, M. D., Oliveira, E. B., & Costa-Neto, C. M. (2007). The N-terminal region of eukaryotic translation initiation factor 5A signals to nuclear localization of the protein. *Biochemical and Biophysical Research Communications*, *362*(2), 393–398. Retrieved from http://www.sciencedirect.com/science/article/pii/S0006291X07016786

Paulo, O., Jr, D. A., Toledo, T. R., Rossi, D., Rossetto, D. D. B., Faria, T., … Roberto, S. (2014). Hypusine Modification of the Ribosome-binding Protein eIF5A , a Target for New Anti-Inflammatory Drugs : Understanding the Action of the Inhibitor GC7 on a Mur- ine Macrophage Cell Line. *Current Pharmaceutical Design*, *20*(2), 284–292.

Peat, T. S., Newman, J., Waldo, G. S., Berendzen, J., & Terwilliger, T. C. (1998). Structure of translation initiation factor 5A from Pyrobaculum aerophilum at 1.75 å resolution. *Structure*, *6*(9), 1207–1214. Retrieved from http://www.sciencedirect.com/science/article/pii/S0969212698001208

Peti, W., Smith, L. J., Redfield, C., & Schwalbe, H. (2001). Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. *Journal of Biomolecular NMR*, *19*(2), 153–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11256811

Petsko, G. A. (2006). An Introduction to Modeling Structure. In *Current Protocols in Bioinformatics*. John Wiley and Sons, Inc.

Phillips, J. C., Wlodawer, A., Yevitz, M. M., & Hodgson, K. O. (1976). Applications of synchrotron radiation to protein crystallography: preliminary results. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(1), 128–32.

Pieper, U., Webb, B. M., Barkan, D. T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., … Sali, A. (2011). ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Research*, *39*(Database issue), D465–74. doi:10.1093/nar/gkq1091

Pineda-lucena, A. (2006). *NMR in Drug Discovery*.

Pintacuda, G., Park, A. Y., Keniry, M. a, Dixon, N. E., & Otting, G. (2006). Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein-protein

complexes. *Journal of the American Chemical Society*, *128*(11), 3696–702. doi:10.1021/ja057008z

Plonus, M. A. (1978). *Applied electromagnetics*. New York: Mcgraw-Hill Inc.,.

Powers, R., Moy, J. F., Siegel, M. M., & Mobilio, D. (2001). Methods of Structure-Based Drug Design Using MS/NMR. *United States Patent Application*. United States of America: American Home Products Corporation.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., … Lindahl, E. (2013a). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, *29*(7), 845–54. doi:10.1093/bioinformatics/btt055

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., … Lindahl, E. (2013b). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, *29*(7), 845–54. doi:10.1093/bioinformatics/btt055

Rahman-Roblick, R., Roblick, U. J., Hellman, U., Conrotto, P., Liu, T., Becker, S., … Wiman, K. G. (2007). p53 targets identified by protein expression profiling. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(13), 5401–6. doi:10.1073/pnas.0700794104

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, *7*(1), 95–99. doi:10.1016/S0022-2836(63)80023-6

RCSB. (2013). The Protein DataBank. Retrieved from http://www.rcsb.org/pdb/statistics/holdings.do

Reich, H. J. (2010). Relaxation in NMR Spectroscopy. In *Structure Determination using NMR* (pp. 1–11). Madison: University of Wisconsin.

Reinhardt, H. C., & Yaffe, M. B. (2009). Kinases that control the cell cycle in response to DNA damage: Chk1, Chk2, and MK2. *Current Opinion in Cell Biology*, *21*(2), 245–55. doi:10.1016/j.ceb.2009.01.018

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advanced Protein Chemistry*, *34*, 167–339.

Richardson, J. S. (2000). Early ribbon drawings of proteins. *Nature Structural Biology*, *7*(8), 624–5. doi:10.1038/77912

Rimratchada, S., McLeish, T. C. B., Radford, S. E., & Paci, E. (2014). The role of high-dimensional diffusive search, stabilization, and frustration in protein folding. *Biophysical Journal*, *106*(8), 1729–40. doi:10.1016/j.bpj.2014.01.051

Roberts, G. (2000). Applications of NMR in drug discovery. *Drug Discovery Today*, *5*(6), 230–240. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10825729

Robustelli, P., Stafford, K. a, & Palmer, A. G. (2012). Interpreting protein structural dynamics from NMR chemical shifts. *Journal of the American Chemical Society*, *134*(14), 6365–74. doi:10.1021/ja300265w

Rohl, C. a, & Baker, D. (2002). De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *Journal of the American Chemical Society*, *124*(11), 2723–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11890823

Rossi, D., Galvão, F. C., Bellato, H. M., Boldrin, P. E. G., Andrews, B. J., Valentini, S. R., & Zanelli, C. F. (2013). eIF5A has a function in the cotranslational translocation of proteins into the ER. *Amino Acids*, 645–653. doi:10.1007/s00726-013-1618-6

Rossi, D., Kuroshu, R., Zanelli, C. F., & Valentini, S. R. (2014). eIF5A and EF-P: two unique translation factors are now traveling the same road. *Wiley Interdisciplinary Reviews. RNA*, *5*(2), 209–22. doi:10.1002/wrna.1211

Rossi, P., Swapna, G. V. T., Huang, Y. J., Aramini, J. M., Anklin, C., Conover, K., … Montelione, G. T. (2010). A microscale protein NMR sample screening pipeline. *Journal of Biomolecular NMR*, *46*(1), 11–22. doi:10.1007/s10858-009-9386-z

Rossmann, M. G., & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, (15), 24–31.

Ruhl, M., Himmelspach, M., Bahr, G. M., Hammerschmid, F., Jaksche, H., Wolff, B., … Bevec, D. (1993). Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation. *The Journal of Cell Biology*, *123*(6 Pt 1), 1309–20. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2290910&tool=pmcentrez&rendertype=abstract

Rule, G. S., & Hitchens, T. K. (2006). *Fundamentals of Protein NMR Spectroscopy*. (R. Kaptein, Ed.). Dordrecht: Springer.

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, *307*(1-2), 249–64. doi:10.1007/s11010-007-9603-6

Sali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, *234*, 779–815.

Sambrook, J., Fritsch, E. D., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual 2*. Cold Spring Harbor, New York: Schnie.

Savukov, I. M., Seltzer, S. J., & Romalis, M. V. (2007). Detection of NMR signals with a radio-frequency atomic magnetometer. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*, *185*(2), 214–20. doi:10.1016/j.jmr.2006.12.012

Schnier, J., Schwelberger, H. G., Smit-McBride, Z., Kang, H. A., & Hershey, J. W. (1991). Translation initiation factor 5A and its hypusine modification are essential for cell viability in the yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.*, *11*(6), 3105–3114. doi:10.1128/MCB.11.6.3105

Schoenborn, B. P. (1969). Neutron diffraction analysis of myoglobin. *Nature*, *224*(5215), 143–6.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13), 3381–3385. doi:10.1093/nar/gkg520

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*, *160*(1), 65–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12565051

Shen, M.-Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science : A Publication of the Protein Society*, *15*(11), 2507–24. doi:10.1110/ps.062416606

Shen, Y., & Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR*, *48*(1), 13–22. doi:10.1007/s10858-010-9433-9

Shen, Y., Vernon, R., Baker, D., & Bax, A. (2009). De novo protein structure generation from incomplete chemical shift assignments. *Journal of Biomolecular NMR*, *43*(2), 63–78. doi:10.1007/s10858-008-9288-5

Sheng, H., Shao, J., Morrow, J., Beauchamp, R. D., & DuBois, R. N. (1998). Modulation of apoptosis and Bcl-2 expression by prostaglandin E2 in human colon cancer cells. *Cancer Research*, *58*(2), 362–366. Retrieved from http://www.scopus.com/record/display.url?eid=2-s2.0-0031984810&origin=inward&txGid=6028FD6E927C659BA188B681A1BD555E.53bsOu7mi7A1NSY7fPJf1g%3a1

Shi, J., Blundell, T. L., & Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology*, *310*(1), 243–57. doi:10.1006/jmbi.2001.4762

Snyder, D. A., Chen, Y., Denissova, N. G., Acton, T., Aramini, J. M., Ciano, M., … Rajan, P. A. (2005). Comparisons of NMR spectral quality and success in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *Journal of the American Chemical Society*, *127*(47), 16505–16511.

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–8. doi:10.1093/nar/gki408

Stahl, M., Guba, W., & Kansy, M. (2006). Integrating molecular design resources within modern drug discovery research: the Roche experience. *Drug Discovery Today*, *11*(7-8), 326–33. doi:10.1016/j.drudis.2006.02.008

Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S. Q., Haynes, S. R., & Steven, a C. (1991). Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *International Journal of Biological Macromolecules*, *13*(3), 130–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1716976

Studier, F. W., & Moffat, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective highlevel expression of cloned genes. *Journal of Molecular Biology*, *189*, 113–130.

Taylor, C. A., Sun, Z., Cliche, D. O., Ming, H., Eshaque, B., Jin, S., … Thompson, J. E. (2007). Eukaryotic translation initiation factor 5A induces apoptosis in colon cancer cells and associates with the nucleus in response to tumour necrosis factor α signalling. *Experimental Cell Research*, *313*(3), 437–449. Retrieved from http://www.sciencedirect.com/science/article/pii/S0014482706003892

Teng, Q. (2013). Basic Principles of NMR. In *Structural Biology* (pp. 1–63). Boston, MA: Springer US. doi:10.1007/978-1-4614-3964-6

Tenney, K., Hunt, I., Sweigard, J., Pounder, J. I., McClain, C., Bowman, E. J., & Bowman, B. J. (2000). Hex-1, a gene unique to filamentous fungi, encodes the major protein of the Woronin body and functions as a plug for septal pores. *Fungal Genetics and Biology : FG & B*, *31*(3), 205–17. doi:10.1006/fgbi.2000.1230

Thomas, A., Goumans, H., Amesz, H., Benne, R., & Voorma, H. O. (1979). A Comparison of the Initiation Factors of Eukaryotic Protein Synthesis from Ribosomes and from the Postribosomal Supernatant. *European Journal of Biochemistry*, *98*(2), 329–337. doi:10.1111/j.1432-1033.1979.tb13192.x

Thornton, J. M., Todd, a E., Milburn, D., Borkakoti, N., & Orengo, C. a. (2000). From structure to function: approaches and limitations. *Nature Structural Biology*, *7 Suppl*(november), 991–4. doi:10.1038/80784

Tong, Y., Park, I., Hong, B.-S., Nedyalkova, L., Tempel, W., & Park, H.-W. (2009). Crystal structure of human eIF5A1: insight into functional similarity of human eIF5A1 and eIF5A2. *Proteins*, *75*(4), 1040–5. doi:10.1002/prot.22378

Tugarinov, V., Choy, W., Orekhov, V. Y., & Kay, L. E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(3), 622–7. doi:10.1073/pnas.0407792102

Ugwu, S. O., & Apte, S. P. (2004). Conformational Stability. *Pharmaceutical Technology*, (March), 1–22.

Upadhyay, A. K., Murmu, A., Singh, A., & Panda, A. K. (2012). Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in Escherichia coli. *PloS One*, *7*(3), e33951. doi:10.1371/journal.pone.0033951

Valafar, H., & Prestegard, J. H. (2004). REDCAT: a residual dipolar coupling analysis tool. *Journal of Magnetic Resonance*, *167*(2), 228–241. Retrieved from http://www.sciencedirect.com/science/article/pii/S1090780703004361

Vallurupalli, P., Hansen, D. F., & Kay, L. E. (2008). Structures of invisible , excited protein states by relaxation dispersion NMR spectroscopy. *PNAS*, *105*(33), 11766–11771.

Voet, D., & Voet, J. G. (2004). *Biochemistry* (3rd ed.). New York: John Wiley and Sons, Inc.

Wagner, G., & Wüthrich, K. (1978). Dynamic model of globular protein conformations based on NMR studies in solution. *Nature*, *275*, 247–248. doi:doi:10.1038/275247a0

Wang, W., Nema, S., & Teagarden, D. (2010). Protein aggregation—Pathways and influencing factors. *International Journal of Pharmaceutics*, *390*(2), 89–99. doi:10.1016/j.ijpharm.2010.02.025

Wider, G. (2000). Structure Determination of Biological Macromolecules in Solution Using NMR spectroscopy. *BioTechniques*, *29*, 1278–1294.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(Web Server issue), W407–10. doi:10.1093/nar/gkm290

Wilkens, S. J., Xia, B., Volkman, B. F., Weinhold, F., Markley, J. L., & Westler, W. M. (1998). Inadequacies of the Point-Dipole Approximation for Describing Electron - Nuclear Interactions in Paramagnetic Proteins: Hybrid Density Functional Calculations and the Analysis of NMR Relaxation of High-Spin Iron(III) Rubredoxin. *Journal of Physical Chemistry B*, *102*(42), 8300–8305.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology*, *182*(2), 295–315. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3839023

Wishart, D. S., & Case, D. A. (2002). Use of chemical shifts in macromolecular structure determination. *Methods in Enzymology*, *338*, 3–34. Retrieved from http://www.sciencedirect.com/science/article/pii/S0076687902382144

Wishart, D. S., Watson, M. S., Boyko, R. F., & Sykes, B. D. (1997). Automated 1H and 13C chemical shift prediction using the BioMagResBank. *Journal of Biomolecular NMR*, *10*(4), 329–336. doi:10.1023/A:1018373822088

Woestenenk, E. A., Hammarström, M., van den Berg, S., Härd, T., & Berglund, H. (2004). His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors. *Journal of Structural and Functional Genomics*, *5*(3), 217–229. doi:10.1023/B:jsfg.0000031965.37625.0e

Wolynes, P. G., Eaton, W. a, & Fersht, A. R. (1998). Chemical physics of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 11037–11038. doi:10.1073/pnas.1215733109

Wolynes, P. G., Eaton, W. a., & Fersht, a. R. (2012). From the Cover: Chemical physics of protein folding. *Proceedings of the National Academy of Sciences*, *109*(44), 17770–17771. doi:10.1073/pnas.1215733109

Wüthrich, K. (2003). NMR studies of structure and function of biological macromolecules (Nobel Lecture). *Journal of Biomolecular NMR*, *27*(1), 13–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15143746

Wuthrich, K., & Wider, G. (2003). Transverse relaxation – optimized NMR spectroscopy with biomacromolecular structures in solution †. *Magnetic Resonance in Chemistry : MRC*, *41*, 80–88. doi:10.1002/mrc.1280

Xiang, Z. (2006). Advances of Homology Protein Structure Modelling. *Current Protein Pept Sci.*, *7*(3), 217–227.

Xin, C. (2003). COMPUTER AIDED DRUG DESIGN : DRUG TARGET DIRECTED IN SILICO APPROACHES COMPUTER AIDED DRUG DESIGN : DRUG TARGET DIRECTED IN SILICO APPROACHES BY.

Xiong, J. (2006). *Essential Bioinformatics*. Cambridge University Press.

Xu, A., & Chen, K. Y. (2001). Hypusine is required for a sequence-specific interaction of eukaryotic initiation factor 5A with postsystematic evolution of ligands by exponential enrichment RNA. *The Journal of Biological Chemistry*, *276*(4), 2555–61. doi:10.1074/jbc.M008982200

Xu, X.-P., & Case, D. A. (2001). Automated prediction of 15N, 13Cα, 13Cβ and 13C′ chemical shifts in proteins using a density functional database. *Journal of Biomolecular NMR*, *21*(4), 321–333. doi:10.1023/A:1013324104681

Yamaguchi, H., & Wang, H.-G. (2006). Tissue transglutaminase serves as an inhibitor of apoptosis by cross-linking caspase 3 in thapsigargin-treated cells. *Molecular and Cellular Biology*, *26*(2), 569–79. doi:10.1128/MCB.26.2.569-579.2006

Yee, A. a, Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., … Arrowsmith, C. H. (2005). NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society*, *127*(47), 16512–7. doi:10.1021/ja053565+

Yu, H. (1999). Extending the size limit of protein nuclear magnetic resonance. *Proceedings of the National Academy of Sciences*, *96*(2), 332–334. doi:10.1073/pnas.96.2.332

Yuan, J., Jiang, N., Jin, C., Zhang, X., & Yan, X. (2009). Backbone and sidechain 1 H, 15 N, and 13 C assignments of the human eIF5A. *Biomolecular NMR Assignments*, *3*(1), 25–28.

Yuan, P., Jedd, G., Kumaran, D., Swaminathan, S., Shio, H., Hewitt, D., … Swaminathan, K. (2003). A HEX-1 crystal lattice required for Woronin body function in Neurospora crassa. *Nature Structural Biology*, *10*(4), 264–70. doi:10.1038/nsb910

Zanelli, C. F., Maragno, A. L. C., Gregio, A. P. B., Komili, S., Pandolfi, J. R., Mestriner, C. A., … Valentini, S. R. (2006). eIF5A binds to translational machinery components and affects translation in yeast. *Biochemical and Biophysical Research Communications*, *348*(4), 1358–1366. Retrieved from http://www.sciencedirect.com/science/article/pii/S0006291X06017724

Zanelli, C. F., & Valentini, S. R. (2007). Is there a role for eIF5A in translation? *Amino Acids*, *33*(2), 351–8. doi:10.1007/s00726-007-0533-0

Zheng, J., Cahill, S. M., Lemmon, M. A., Fushman, D., Schlessinger, J., & Cowburn, D. (1996). Identification of the binding site for acidic phospho- lipids on the PH domain of dynamin: implications for stimulation of GTPase activity. *Journal of Molecular Biology*, (255), 14–21.

Zuk, D., & Jacobson, A. (1998). A single amino acid substitution in yeast eIF-5A results in mRNA stabilization. *The EMBO Journal*, *17*(10), 2914–25. doi:10.1093/emboj/17.10.2914

Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal ' s paradox. *Proceedings of the National Academy of Sciences USA*, *89*(January), 20–22.

Appendices

APPENDIX A: GROWTH MEDIA 149

APPENDIX B: BUFFERS 150

APPENDIX C: SDS-PAGE 151

APPENDIX D: NMR PULSE PROGRAMS 153

APPENDIX A: GROWTH MEDIA

Luria-Bertani broth:

10 g tryptone

5 g yeast extract

5 g NaCl

15 g agar*

Make up to 1 litre using ddH_2O

Autoclaved

M9 Minimal Media (per Litre):

M9 Salts (400 ml of 5 x)

- ✓ 64g $Na_2H_2P0_4.7H_2O$
- ✓ 15g KH_2PO_4 (anhydrous)
- ✓ 2.5g NaCl

Dissolve in 500 ml of ddH_2O and adjust pH to 7.2

Make up to 1 litre and Autoclave

$^{15}NH_4Cl$ (20 ml of 10%)

Adjust pH to 7.2 and autoclave

Glucose (40 ml of 20%)

Autoclave

Trace metal mix (20 ml)

- ✓ Boric acid (50 mg/l)
- ✓ $MnSO_4.2H_2O$ (40 mg/l)
- ✓ $ZnSO_4$ (40 mg/l)
- ✓ $(NH_4)_6Mo_2O_{24}.4H_2O$ (20 mg/l)
- ✓ KI (10 mg/l)
- ✓ $CuSO_4$ (4 mg/l)

Dissolve in ddH_2O and adjust pH to 7.2

Autoclave

- ✓ $FeCl_3$ (20 mg in 20 ml)

Autoclave

Cool metal mix to 50 °C before adding $FeCl_3$

Vitamin mix (1 ml)

- ✓ 10 % Thiamine and Pyridoxine

  Filter sterilize

1 M $MgSO_4$ (4 ml)

Autoclave

1 M $CaCl_2$ (0.4 ml)

Autoclave

Cool and make up to 2 litres using ddH_2O

APPENDIX B: BUFFERS

Phosphate buffer saline:

4g NaCl

0.1g KCl

0.72g NaH_2PO_4

0.12g KH_2PO_4

Dissolve in 400 ml adjust pH to 7.4 using NaOH

Make up to 500 ml using ddH2O and autoclave.

Lysis Buffer:

20 mM phosphate buffer

500 mM NaCl

40 mM imidazole

pH 7.4

Elution Buffer

20 mM phosphate buffer

500 mM NaCl

500 mM imidazole

pH 7.4

NMR Sample buffer:

100 mM Phosphate buffer pH 6.8

1 mM for every 2 mg/ml of protein

90 % H_2O and 10 % D_2O

Cleavage Buffer:

20 mM of phosphate buffer or Tris – Cl buffer at pH 8.4

150 mM of NaCl

2.5 mM of $CaCl_2$

APPENDIX C: SDS PAGE

Reagents	SDS resolving gel (15 %)	**S**DS stacking gel (4 %)
30 % acrylamide/bisacrylamide (29:1)	5 ml	0.66 ml
ddd H2O	1.1 ml	3.58 ml
1 M Tris-HCl pH 8.8	3.75 ml	-
1 M Tris-HCl pH 6.8	-	0.625 ml
10 % SDS	100 µl	50 µl
10 % APS	37.5 µl	71.25 µl
TEMED	8.25 µl	8.25 µl

2x SDS-PAGE sample buffer (10 ml):

1.25 ml 1 M Tris-HCl, pH 6.8

4 ml 10 % SDS

1 ml β-mercaptoethanol

1ml glycerol

0.001 g bromophenol blue

2.75 ml ddH2O

2x SDS-PAGE running buffer (1 L):

12 g Tris-HCl

57.6 g glycine

4 g SDS

1 L ddH2O

Destain I (Per 1 L):

500 ml methanol

100 ml glacial acetic acid

400 ml ddH2O

Destain II (Per 1 L):

80 ml methanol

70 ml glacial acetic acid

880 ml ddH2O

Coomassie stain (1 L):

2.6 g Coomassie R125

Dissolve in 500 ml and makeup to 1 L using Destain I

Filter through Buchner funnel to remove insoluble particles

APPENDIX D: NMR PULSE PROGRAMS

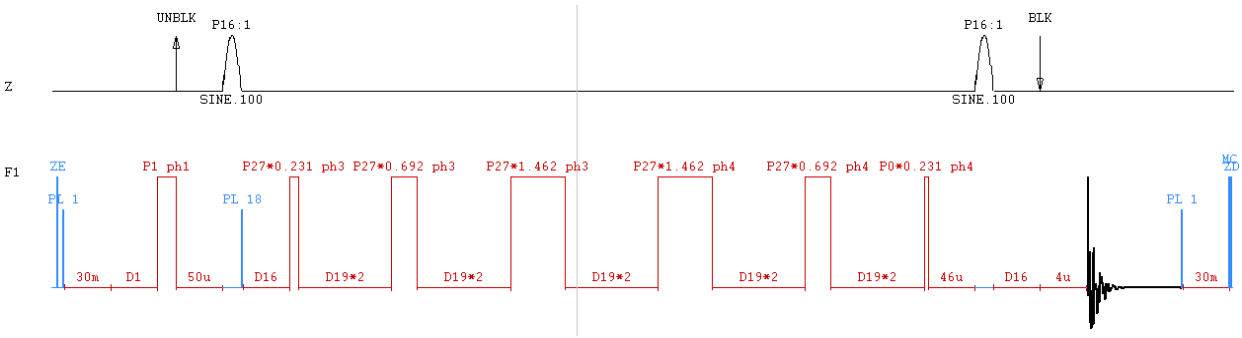


Figure D-1: Proton spectra of eI5A$_{Thr}$ using p3919gp WATERGATE suppression.